

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**ΑΝΑΛΥΣΗ ΑΚΟΛΟΥΘΙΩΝ ΕΙΚΟΝΩΝ ΓΙΑ
ΧΑΡΑΚΤΗΡΙΣΜΟ, ΠΡΟΕΠΙΣΚΟΠΗΣΗ ΚΑΙ
ΑΝΑΖΗΤΗΣΗ ΜΕ ΒΑΣΗ ΤΟ
ΠΕΡΙΕΧΟΜΕΝΟ**

Ιωάννης Σ. Αβρίθης

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ



Εθνικό Μετσόβιο Πολυτεχνείο

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννης Σ. Αβρίθης

Διπλ. Ηλεκτρολόγος Μηχανικός ΕΜΠ

ΑΝΑΛΥΣΗ ΑΚΟΛΟΥΘΙΩΝ ΕΙΚΟΝΩΝ ΓΙΑ ΧΑΡΑΚΤΗΡΙΣΜΟ, ΠΡΟΕΠΙΣΚΟΠΗΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Επιβλέπων Καθηγητής:

Στέφανος Κόλλιας, Καθηγητής ΕΜΠ

Τριμελής Συμβουλευτική Επιτροπή:

Στέφανος Κόλλιας, Καθηγητής ΕΜΠ

Χρήστος Χαλκιάς, Καθηγητής ΕΜΠ

Ηλίας Κουκούτσης, Επικ. Καθηγητής ΕΜΠ

Επταμελής Εξεταστική Επιτροπή

Στέφανος Κόλλιας
Καθηγητής ΕΜΠ

Ανδρέας Σταφυλοπάτης
Καθηγητής ΕΜΠ

Ηλίας Κουκούτσης
Επικ. Καθηγητής ΕΜΠ

Πέτρος Μαραγκός
Καθηγητής ΕΜΠ

Αθανάσιος Σκόδρας
Αν. Καθηγητής Παν.
Πατρών

Γεώργιος Τζιρίτας
Αν. Καθηγητής Παν.
Κρήτης

Αντώνιος Κωνσταντινίδης
Καθηγητής Παν. Λονδίνου

Αθήνα, Φεβρουάριος 2001

Περίληψη

Ανάλυση Ακολουθιών Εικόνων για Χαρακτηρισμό, Προεπισκόπηση και Αναζήτηση με Βάση το Περιεχόμενο

του Ιωάννη Αβρίθη

Η παρούσα διδακτορική διατριβή έχει κεντρικό αντικείμενο την επεξεργασία και ανάλυση εικόνων και ακολουθιών εικόνων για το χαρακτηρισμό και τη δεικτοδότηση του οπτικού τους περιεχομένου. Στόχος της είναι η συνεισφορά στην ανάπτυξη ενός αυτοματοποιημένου υπολογιστικού συστήματος το οποίο να έχει τη δυνατότητα ανάλυσης και αυτόματου χαρακτηρισμού οπτικοακουστικού υλικού, κατάτμησής του σε αντικείμενα, εξαγωγής περίληψης με σκοπό την προεπισκόπηση και πλοήγηση μέσα από το υλικό, καθώς και αναζήτησης και ανάκλησης με βάση το περιεχόμενο.

Η διατριβή απαρτίζεται από τέσσερα μέρη. Στο πρώτο μέρος γίνεται μια εισαγωγή στην ανάλυση ακολουθιών εικόνων, στην κατάτμηση και στην εξαγωγή αντικειμένων με βάση το χρώμα, την κίνηση, καθώς και το πεδίο βάθους στην περίπτωση των στερεοσκοπικών ακολουθιών. Με βάση το συνδυασμό των επιμέρους κατατμήσεων, προτείνεται μία τεχνική συγχώνευσης που επιτρέπει τον αξιόπιστο προσδιορισμό σημασιολογικών αντικειμένων. Το δεύτερο μέρος αναφέρεται στον αυτόματο χαρακτηρισμό του οπτικού περιεχομένου, με υπολογισμό διανυσμάτων περιγραφής που περιέχουν χαρακτηριστικά των αντικειμένων. Η πληροφορία του χαρακτηρισμού αποτελεί τη βάση για την εξαγωγή περίληψης η οποία υλοποιείται με την αυτόματη επιλογή ενός περιορισμένου συνόλου χαρακτηριστικών πλάνων και καρέ που περιγράφουν ικανοποιητικά το οπτικό περιεχόμενο. Η αναπαράσταση του επιλεγμένου υλικού με διανύσματα περιγραφής χρησιμοποιείται στη συνέχεια για αναζήτηση και ανάκληση με βάση το περιεχόμενο.

Στο τρίτο μέρος της διατριβής εξετάζεται το πρόβλημα της ανάλυσης και αναπαράστασης του περιγράμματος αντικειμένων με εφαρμογή στην ταξινόμηση και αναζήτηση αντικειμένων με βάση το σχήμα. Παρουσιάζεται μία πρωτότυπη τεχνική κανονικοποίησης περιγραμμάτων η οποία χωρίς απώλεια πληροφορίας παρέχει τη δυνατότητα αναλλοίωτης περιγραφής ως προς ένα σύνολο μετασχηματισμών. Τέλος, στο τέταρτο μέρος παρουσιάζεται μία τεχνική χρονικής κατάτμησης και συντακτικής ανάλυσης ακολουθιών τηλεοπτικών δελτίων ειδήσεων σε θεματικές ενότητες με χρήση της οπτικής πληροφορίας. Η τεχνική βασίζεται σε έναν προηγμένο αλγόριθμο αυτόματης ανίχνευσης και εντοπισμού ανθρώπινων προσώπων· η παραγόμενη πληροφορία χρησιμοποιείται επίσης για την υλοποίηση νέων, σημασιολογικών κριτηρίων αναζήτησης με βάση το περιεχόμενο. Η διατριβή καταλήγει σε έκθεση συμπερασμάτων και προτάσεων για ζητήματα που θα μπορούσαν να αποτελέσουν τη βάση για μελλοντική έρευνα.

Πίνακας Περιεχομένων

Πρόλογος	vii
Ευχαριστίες	ix
Εισαγωγή	1
1. Ερευνητικό Αντικείμενο	1
2. Υπάρχουσα Κατάσταση – Σχετικές Εργασίες	4
3. Περιγραφή Προτεινόμενων Τεχνικών	6
4. Συνεισφορά Διατριβής	9
5. Δομή Διατριβής	10
 ΜΕΡΟΣ Ι ΑΝΑΛΥΣΗ ΑΚΟΛΟΥΘΙΩΝ ΕΙΚΟΝΩΝ ΣΕ ΑΝΤΙΚΕΙΜΕΝΑ	 12
Κεφάλαιο 1 Κατάτμηση με Βάση το Χρώμα και την Κίνηση	13
1.1 Κατάτμηση Ακολουθιών Εικόνων σε Αντικείμενα	13
1.2 Αλγόριθμοι Κατάτμησης	14
1.3 Κατάτμηση με Βάση το Χρώμα	17
1.4 Κατάτμηση με Βάση την Κίνηση	19
 Κεφάλαιο 2 Κατάτμηση με Βάση το Πεδίο Βάθους	 22
2.1 Στερεοσκοπικές Ακολουθίες Βίντεο	22
2.2 Στερεοσκοπική Ανάλυση	23
2.3 Εκτίμηση Πεδίου Βάθους και Απόκλισης	25
2.4 Ανίχνευση Επικάλυψης	28
2.5 Κατάτμηση	30
 Κεφάλαιο 3 Συγχώνευση Αντικειμένων	 32
3.1 Ανίχνευση Σημασιολογικών Αντικειμένων	32
3.2 Αλγόριθμος Συγχώνευσης	33
3.3 Πειραματικά Αποτελέσματα	34
 ΜΕΡΟΣ ΙΙ ΧΑΡΑΚΤΗΡΙΣΜΟΣ, ΠΡΟΕΠΙΣΚΟΠΗΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ	 39
Κεφάλαιο 4 Αναπαράσταση Ακολουθιών Εικόνων	40
4.1 Ασαφή Ιστογράμματα	40
4.2 Μονοδιάστατη Ταξινόμηση	41
4.3 Πολυδιάστατη Ταξινόμηση	42
4.4 Διανύσματα Περιγραφής	45
4.5 Παράδειγμα Συνθετικής Ακολουθίας	46

Κεφάλαιο 5	Επιλογή Χαρακτηριστικών Πλάνων	50
5.1	Ανίχνευση και Χαρακτηρισμός Πλάνων	50
5.2	Ομαδοποίηση Πλάνων	52
5.3	Πειραματικά Αποτελέσματα	54
5.4	Αποτελέσματα σε Στερεοσκοπικές Ακολουθίες	55
Κεφάλαιο 6	Επιλογή Χαρακτηριστικών Καρέ με Βάση τη Χρονική Μεταβολή	60
6.1	Τροχιά Διανύσματος Περιγραφής	60
6.2	Μέθοδος Χρονικής Μεταβολής	61
6.3	Παράδειγμα Συνθετικής Ακολουθίας	62
6.4	Πειραματικά Αποτελέσματα	64
Κεφάλαιο 7	Βέλτιστη Επιλογή Χαρακτηριστικών Καρέ	67
7.1	Μέθοδος Συσχέτισης	67
7.2	Λογαριθμική Αναζήτηση	69
7.3	Αναζήτηση με Γενετικό Αλγόριθμο	71
7.4	Πειραματικά Αποτελέσματα	76
Κεφάλαιο 8	Αναζήτηση με Βάση το Περιεχόμενο	79
8.1	Μηχανισμός Αναζήτησης	79
8.2	Ανάδραση	82
8.3	Ενημέρωση Παραμέτρων	83
8.4	Πειραματικά Αποτελέσματα	85
ΜΕΡΟΣ ΙΙΙ	ΠΕΡΙΓΡΑΦΗ ΣΧΗΜΑΤΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ	88
Κεφάλαιο 9	Αναπαράσταση Περιγραμμάτων	89
9.1	Εισαγωγή	89
9.2	Προτεινόμενη Αναπαράσταση	92
9.3	Προϋποθέσεις – Περιορισμοί	93
9.4	B-Splines	95
Κεφάλαιο 10	Κανονικοποίηση Περιγραμμάτων	99
10.1	Ορθοκανονικοποίηση	99
10.2	Κανονικοποίηση Σημείου Αναφοράς	103
10.3	Κανονικοποίηση Περιστροφής και Κατοπτρισμού	105
10.4	Τελικά Στάδια Κανονικοποίησης	107
10.5	Αποδείξεις	108
Κεφάλαιο 11	Αναζήτηση με Βάση το Σχήμα – Αποτελέσματα	112
11.1	Εναλλακτικές Μέθοδοι	112
11.2	Ταξινόμηση και Αναζήτηση	114
11.3	Αποτελέσματα Κανονικοποίησης	117
11.4	Αποτελέσματα Ταξινόμησης και Αναζήτησης	123
ΜΕΡΟΣ ΙV	ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΔΕΛΤΙΩΝ ΕΙΔΗΣΕΩΝ	128
Κεφάλαιο 12	Εντοπισμός Προσώπων	129
12.1	Εισαγωγή	129
12.2	Μοντέλο Πιθανότητας Χρώματος Δέρματος	131

12.3 Εξαγωγή Τμημάτων Προσώπου	133
12.4 Πειραματικά Αποτελέσματα	140
Κεφάλαιο 13 Συντακτική Ανάλυση Δελτίων	145
13.1 Ανάλυση Δελτίων Ειδήσεων	145
13.2 Ταξινόμηση Πλάνων	147
13.3 Πειραματικά Αποτελέσματα	148
13.4 Αξιολόγηση	152
Κεφάλαιο 14 Προεπισκόπηση και Ανάκληση	154
14.1 Εξαγωγή Περίληψης και Προεπισκόπηση	154
14.2 Χαρακτηριστικά Προσώπων	156
14.3 Ανάκληση με Βάση τα Πρόσωπα	157
14.4 Πειραματικά Αποτελέσματα	159
Συμπεράσματα	162
Βιβλιογραφία	165
Ευρετήριο	174

Πρόλογος

Οι τεχνολογίες πολυμέσων έχουν διεισδύσει σε όλους σχεδόν τους τομείς της ανθρώπινης δραστηριότητας και πρόκειται να αλλάξουν ριζικά το τρόπο με τον οποίο διεξάγεται η πρόσβαση στην πληροφορία, η επικοινωνία, η εκπαίδευση, οι επαγγελματικές δραστηριότητες και η διασκέδαση. Οι τεχνολογίες αυτές δημιουργούν τη δυνατότητα ενσωμάτωσης και αλληλεπίδρασης ανάμεσα σε διαφορετικά μέσα, όπως για παράδειγμα η ομιλία, ο ήχος, η εικόνα, το βίντεο, το κείμενο και τα γραφικά. Καθώς η ψηφιοποίηση, κωδικοποίηση, αποθήκευση και μετάδοση εικόνων και βίντεο γίνεται όλο και πιο προσιτή, υπολογιστικά συστήματα και βάσεις δεδομένων αρχίζουν να αποθηκεύουν τεράστιο όγκο οπτικοακουστικής πληροφορίας και η πρόσβαση οποιουδήποτε χρήστη στην πληροφορία αυτή αρχίζει να γίνεται πραγματικότητα. Όμως, καθώς η τεχνολογία καταγραφής και αποθήκευσης δεδομένων προχωρά με γοργούς ρυθμούς, αρχίζει να γίνεται αισθητή η έλλειψη τεχνολογιών χαρακτηρισμού, δεικτοδότησης και ανάκλησης οπτικοακουστικής πληροφορίας από μεγάλα αρχεία.

Η ανάκληση πληροφορίας, σαν γνωστό αντικείμενο, υπάρχει εδώ και πολλά χρόνια. Μέχρι πρόσφατα όμως, οι μεθοδολογίες ανάκλησης εστιάζονταν στην κατανόηση πληροφορίας κειμένου, δηλαδή στην εξαγωγή λέξεων-κλειδίων, στην κατηγοριοποίηση ή στη δημιουργία περίληψης εγγράφων, πάντα με βάση το γραπτό κείμενο. Σε ένα «έγγραφο» πολυμέσων η πληροφορία ενσωματώνεται σε διάφορες μορφές μέσων, οι οποίες συνήθως συμπληρώνουν η μία την άλλη. Για την κατανόηση μίας ακολουθίας βίντεο για παράδειγμα, είναι απαραίτητη η ανάλυση όλων των διαθέσιμων τύπων δεδομένων: εικόνες από τα καρέ του βίντεο, κανάλια ήχου, κείμενα που μπορούν να εντοπισθούν στις εικόνες, καθώς και λέξεις που μπορούν να ανιχνευθούν από τον ήχο. Η διαδικασία ανάλυσης εμπλέκει συνήθως την κατάτμηση του εγγράφου σε αυτόνομες ενότητες με σημασιολογικό περιεχόμενο, την ταξινόμηση κάθε ενότητας σε μία προκαθορισμένη κατηγορία σκηνής, την αυτόματη εξαγωγή περίληψης και τη δεικτοδότηση του εγγράφου για την πλοήγηση στο περιεχόμενό του και την αποτελεσματική ανάκληση πληροφορίας.

Η δραστηριότητα της επιστημονικής κοινότητας, των διεθνών οργανισμών τυποποίησης και της βιομηχανίας είναι ιδιαίτερα έντονη προς την κατεύθυνση της δημιουργίας τεχνικών, προτύπων, προϊόντων και υπηρεσιών που θα διευκολύνουν ή ακόμη και θα αυτοματοποιήσουν τις διαδικασίες αυτές. Στο πλαίσιο της δραστηριότητας αυτής εντάσσεται και η πορεία του Εργαστηρίου Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων του Ε.Μ.Π., το οποίο υποστηρίζει επιστημονικά και τεχνολογικά την περιοχή της επεξεργασίας ακίνητων και κινούμενων εικόνων, καλύπτοντας θέματα σύλληψης, κωδικοποίησης, συμπίεσης, ανάλυσης, σύνθεσης, μετάδοσης, αποθήκευσης, ανακατασκευής, επίδειξης και διαχείρισης εικόνων, αναπτύσσοντας και διερευνώντας όλες τις εφαρμογές που άπτονται των θεμάτων αυτών. Πρόσφατα στο πλαίσιο της ερευνητικής εργασίας του εργαστηρίου εξετάζονται ζητήματα όπως η αυτόματη ανάλυση και κατανόηση βίντεο, η αναγνώριση προσώπων και ανθρώπινων εκφράσεων, καθώς και η μοντελοποίηση, ανάλυση και σύνθεση τρισδιάστατων σκηνών· τα πρακτικά αποτελέσματα των εργασιών αυτών εφαρμόζονται συνήθως στη σχεδίαση και υλοποίηση συστημάτων λογισμικού στο πλαίσιο ερευνητικών και αναπτυξιακών έργων για ε-

φαρμογές όπως η διαχείριση, ο χαρακτηρισμός, η αναζήτηση και η ανάκληση οπτικοακουστικών δεδομένων.

Η παρούσα διδακτορική διατριβή εντάσσεται στο γενικότερο πλαίσιο δραστηριοτήτων του εργαστηρίου και έχει κεντρικό αντικείμενο την επεξεργασία και ανάλυση εικόνων και ακολουθιών εικόνων για το χαρακτηρισμό και τη δεικτοδότηση του οπτικού τους περιεχομένου. Η ανάλυση ακολουθιών εικόνων αντιμετωπίζει τις μεγαλύτερες δυσκολίες σε σχέση με άλλα μέσα όπως ο ήχος ή το κείμενο, κυρίως λόγω του τεράστιου όγκου πληροφορίας που περιέχεται. Το πρώτο στάδιο ανάλυσης περιλαμβάνει τον προσδιορισμό αντικειμένων με βάση το χρώμα, την κίνηση, καθώς και το πεδίο βάθους στην περίπτωση των στερεοσκοπικών ακολουθιών. Πραγματοποιείται αυτόματος χαρακτηρισμός του οπτικού περιεχομένου με υπολογισμό διανυσμάτων περιγραφής, που περιέχουν χαρακτηριστικά των αντικειμένων και χρησιμοποιούνται για αναζήτηση με βάση το περιεχόμενο, καθώς και για εξαγωγή περίληψης, με την αυτόματη επιλογή χαρακτηριστικών πλάνων και καρέ. Εξετάζεται επίσης το πρόβλημα της ανάλυσης, αναπαράστασης και κανονικοποίησης του περιγράμματος αντικειμένων με κύριο στόχο την αναζήτηση με βάση το σχήμα· τέλος παρουσιάζεται μία εφαρμογή στη συντακτική ανάλυση δελτίων ειδήσεων, που βασίζεται στην αυτόματη ανίχνευση ανθρώπινων προσώπων. Σε κάθε στάδιο επεξεργασίας ιδιαίτερη σημασία δόθηκε στην ανάπτυξη γρήγορων αλγορίθμων, αφού η υπολογιστική πολυπλοκότητα αποτελεί βασικό περιορισμό στην ανάλυση ακολουθιών εικόνων.

Στο κείμενο της διατριβής έχει δοθεί έμφαση στην αναλυτική περιγραφή όλων των τεχνικών και του αναγκαίου τεχνικού υποβάθρου – ή στην παραπομπή σε άλλα συγγράμματα όπου αυτό δεν είναι δυνατό – ώστε να μην απαιτούνται εξειδικευμένες γνώσεις από τον αναγνώστη. Επίσης έχει καταβληθεί προσπάθεια ώστε η απόδοση των ξενόγλωσσων όρων στα Ελληνικά να είναι όσο το δυνατόν πιο ακριβής και να συμβαδίζει με την ορολογία άλλων συγγραφέων. Για την αποφυγή όμως οποιωνδήποτε παρερμηνειών, η Ελληνική απόδοση συνοδεύεται από την αντίστοιχη ξενόγλωσση κατά την πρώτη αναφορά κάθε όρου στο κείμενο της διατριβής. Το Ευρετήριο στο τέλος του συγγράμματος περιλαμβάνει ανεξάρτητη λίστα Ελληνικών και ξενόγλωσσων όρων.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους με βοήθησαν κατά τη διάρκεια της παρούσας διδακτορικής διατριβής, τα τελευταία τέσσερα χρόνια. Κατ' αρχήν θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή, κ. Στέφανο Κόλλια, για την πολύτιμη βοήθεια και υποστήριξή του, τόσο σε επιστημονικό, όσο και σε προσωπικό επίπεδο. Η συνεισφορά του στη διατριβή υπήρξε αποφασιστική, ενώ τα σχόλια και η εποικοδομητική κριτική του στο κείμενο της διατριβής βοήθησαν καθοριστικά στην ποιότητά της. Πάνω απ' όλα όμως τον ευχαριστώ για την ελευθερία που μου άφησε στην ερευνητική κατεύθυνση της διατριβής και στην ανάπτυξη και αξιοποίηση νέων ιδεών.

Θα ήθελα επίσης να ευχαριστήσω τους συναδέλφους μου στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων για τη βοήθειά τους στη δημιουργία ενός ευχάριστου περιβάλλοντος εργασίας, αλλά ιδιαίτερα εκείνους με τους οποίους συνεργάστηκα στο πλαίσιο της ερευνητικής μου εργασίας. Ειδικότερα είμαι υπόχρεος στο συνάδελφο Νίκο Τσαπατσούλη για την πολύτιμη προσφορά του στην εργασία που παρουσιάζεται στο τέταρτο μέρος της διατριβής, «Εφαρμογή στην Ανάλυση Δελτίων Ειδήσεων», καθώς και στο Γιάννη Ξηρουχάκη για τη σημαντική συμβολή του στην εργασία του τρίτου μέρους, «Περιγραφή Σχήματος Αντικειμένων». Επίσης θα ήθελα να εκφράσω την εκτίμησή μου προς τους συναδέλφους Τάσο Ντελόπουλο, Βασίλη Αλεξόπουλο, Γρηγόρη Παπαγεωργίου, Κώστα Καρπούζη, Νάσο Δροσόπουλο, Σπύρο Ιωάννου, Γιώργο Ακρίβα και Γαβριήλ Τσεχπενάκη για τις συνεργασίες που είχαμε κατά τη διάρκεια της διατριβής.

Επιθυμώ να αναγνωρίσω τη συνεισφορά του Καθηγητή Chas Girdwood του Independent Television Commission (ITC), υπεύθυνου του έργου ACTS MIRAGE, ο οποίος προσέφερε το τηλεοπτικό πρόγραμμα "Eye to Eye" που χρησιμοποιήθηκε στην πειραματική αξιολόγηση των τεχνικών ανάλυσης στερεοσκοπικών ακολουθιών στο πρώτο και δεύτερο μέρος της διατριβής. Επίσης ευχαριστώ τον Dr. Michael Jones του Compaq Cambridge Research Laboratory (CRL) για την ευγενική παροχή άδειας χρήσης της ερευνητικής βάσης δεδομένων που χρησιμοποιήθηκε στα πειράματα ανίχνευσης προσώπων του τέταρτου μέρους.

Οφείλω να εκφράσω τις βαθύτατες ευχαριστίες μου στη φίλη μου Όλγα Κουκλάκη για την ηθική υποστήριξη και τη συνεχή ενθάρρυνση που μου παρείχε στη διεξαγωγή της ερευνητικής μου δραστηριότητας, καθώς και για την υπομονή που επέδειξε κατά τη διάρκεια συγγραφής της διατριβής. Τέλος, είμαι ευγνώμων προς την οικογένειά μου για την ανεκτίμητη συμπαράστασή της σε όλη τη διάρκεια των σπουδών μου. Χωρίς την πολύπλευρη βοήθειά της η ολοκλήρωση της διδακτορικής διατριβής δε θα ήταν δυνατή.

Εισαγωγή

Στην εισαγωγή της διατριβής παρουσιάζεται αρχικά το ερευνητικό της αντικείμενο, οι πρόσφατες εξελίξεις και οι ανάγκες που προκύπτουν από αυτές. Στη συνέχεια πραγματοποιείται ανασκόπηση της πρόσφατης βιβλιογραφίας, αναφέρονται συνοπτικά οι υφιστάμενες εργασίες που σχετίζονται με τη διατριβή και τονίζονται οι ιδιαιτερότητες και τα μειονεκτήματα ορισμένων τεχνικών που οδηγούν στην ανάγκη βελτίωσής τους ή ανάπτυξης νέων. Ακολουθεί περιληπτική περιγραφή των τεχνικών που προτείνονται στο πλαίσιο της διατριβής, καθώς και ανάλυση της συνεισφοράς της διατριβής με αναφορές στις δημοσιεύσεις του συγγραφέα που σχετίζονται με κάθε προτεινόμενη τεχνική. Τέλος αναλύεται η δομή του συγγράμματος σε Μέρη και Κεφάλαια.

1. ΕΡΕΥΝΗΤΙΚΟ ΑΝΤΙΚΕΙΜΕΝΟ

Η ραγδαία ανάπτυξη των εφαρμογών πολυμέσων τα τελευταία χρόνια έχει προκαλέσει μία ανάλογη ζήτηση για αποτελεσματική αποθήκευση, διαχείριση, μετάδοση και πρόσβαση σε βάσεις δεδομένων πολυμέσων. Όλο και περισσότερη οπτικοακουστική πληροφορία γίνεται διαθέσιμη σε διάφορες μορφές ψηφιακών ή αναλογικών μέσων, όπως ακίνητες εικόνες, βίντεο, γραφικά, συνθετικά τρισδιάστατα μοντέλα, κείμενο, ήχος και ομιλία. Η εξάπλωση της χρήσης οπτικοακουστικού υλικού είναι ιδιαίτερα έντονη σε περιβάλλοντα ενημέρωσης, εκπαίδευσης, διασκέδασης, ιατρικής και τηλεπικοινωνιακών υπηρεσιών. Από την άλλη μεριά, σε ένα μεγάλο αριθμό εφαρμογών, όπως για παράδειγμα η αναγνώριση και κατανόηση εικόνων, η μετατροπή μέσων, η ανάκληση ή το φιλτράρισμα πληροφορίας, η οπτικοακουστική πληροφορία παράγεται, αποθηκεύεται, μεταδίδεται, αναζητείται και επαναχρησιμοποιείται από υπολογιστικά συστήματα. Ως συνέπεια, η ανάπτυξη νέων εργαλείων για το *χαρακτηρισμό* (annotation), τη *δεικτοδότηση* (indexing), την *εξαγωγή περίληψης* (summarization), την *προεπισκόπηση* (preview), την *πλοήγηση* (browsing), την *αναζήτηση* (search) και την *ανάκληση με βάση το περιεχόμενο* (content-based retrieval) έχει προσελκύσει αυξημένο ενδιαφέρον.

Η σημασία των εργαλείων αυτών είναι ακόμη μεγαλύτερη για την πλοήγηση σε βάσεις δεδομένων ψηφιακού βίντεο, και ιδιαίτερα στην περίπτωση που το υλικό είναι καταμετρημένο σε απομακρυσμένες πλατφόρμες, λόγω του τεράστιου όγκου οπτικοακουστικής πληροφορίας και των αντίστοιχων απαιτήσεων χωρητικότητας αποθήκευσης και εύρους ζώνης μετάδοσης που αυτός επιβάλλει. Παραδοσιακά, μία ακολουθία εικόνων, ή ακολουθία βίντεο, είτε σε ψηφιακή είτε σε αναλογική μορφή, αναπαρίσταται από ένα μεγάλο αριθμό διαδοχικών *καρέ* (frames), κάθε ένα από τα οποία είναι μία εικόνα και αντιστοιχεί σε μία δεδομένη χρονική στιγμή. Όμως αυτή η αναπαράσταση ακολουθιών εικόνων, που πηγάζει από τη διαδικασία αναλογικής αποθήκευσης, οδηγεί σε γραμμική (ακολουθιακή) πρόσβαση στο περιεχόμενο [150]. Παρόλο που η προσέγγιση αυτή είναι επαρκής για την εποπτική παρακολούθηση βί-

ντεο [62], έχει σημαντικούς περιορισμούς για τις νέες εφαρμογές πλοήγησης, δεικτοδότησης και αναζήτησης με βάση το περιεχόμενο.

Ο μόνος τρόπος πλοήγησης σε μία ακολουθία εικόνων με την αναπαράσταση αυτή είναι η ακολουθιακή ανίχνευση όλων των διαθέσιμων καρτέ, μια διαδικασία εξαιρετικά χρονοβόρα και αναποτελεσματική. Επιπλέον, η εκτέλεση αναζητήσεων σε ολόκληρο το οπτικοακουστικό υλικό είναι ανεπαρκής, λόγω της χρονικής συσχέτισης του οπτικού περιεχομένου και επομένως της πλεονάζουσας πληροφορίας [14]. Η ακολουθιακή αναπαράσταση είναι επίσης ανεπαρκής για τη διαχείριση μεγάλων αρχείων βίντεο, δεδομένου μάλιστα ότι καθημερινά παράγεται μεγάλος όγκος νέου οπτικοακουστικού υλικού που αρχειοθετείται· τα υφιστάμενα εργαλεία και αλγόριθμοι για αποτελεσματική οργάνωση και διαχείριση τέτοιων αρχείων είναι ακόμη περιορισμένα. Η ερευνητική εργασία στο αντικείμενο της δεικτοδότησης με βάση το περιεχόμενο σε βάσεις δεδομένων εικόνων ή βίντεο είναι έντονη, όπως αντικατοπτρίζεται σε δημοσιεύσεις σε πολυάριθμα συνέδρια και αφιερώματα σε κορυφαία περιοδικά του χώρου, όπως για παράδειγμα τα [2,59,121,122,123].

Η προσπάθεια που γίνεται στο πλαίσιο της διαδικασίας προτυποποίησης MPEG είναι επίσης χαρακτηριστική της ανάγκης ανάπτυξης νέων τεχνικών αναπαράστασης. Συγκεκριμένα, προγενέστερα πρότυπα κωδικοποίησης και συμπίεσης βίντεο, όπως τα MPEG-1/2 [63,64] και H.261/3 [24,68], κάθε ένα από τα οποία σχετίζεται με διαφορετικές εφαρμογές και έχει διαφορετικές απαιτήσεις χωρητικότητας αποθήκευσης και ταχύτητας μετάδοσης, αποτέλεσαν κλειδί για την εξάπλωση των οπτικοακουστικών εφαρμογών. Στο πρόσφατο πρότυπο MPEG-4 [65,117] δόθηκε έμφαση στην αποδοτική κωδικοποίηση οπτικοακουστικού υλικού με βάση τα αντικείμενα. Για πρώτη φορά στο πρότυπο αυτό δίνεται η δυνατότητα κωδικοποίησης και αναπαράστασης σε ανεξάρτητα *επίπεδα αντικειμένων βίντεο* (video object planes, VOPs), δίνοντας νέες διαστάσεις στη δημιουργία, στην πρόσβαση και στη *διαχείριση* (manipulation) του οπτικοακουστικού υλικού [27]. Επιπλέον, το αναπτυσσόμενο πρότυπο MPEG-7 [67] έχει βασικό στόχο την ανάπτυξη ενός ολοκληρωμένου πλαισίου για την *περιγραφή υλικού πολυμέσων* (multimedia content description) με τον προσδιορισμό ενός συνόλου από περιγραφείς οι οποίοι μπορούν να χρησιμοποιηθούν για την αναπαράσταση οποιουδήποτε τύπου πληροφορίας πολυμέσων [79,105].

Παράλληλα έχει αναπτυχθεί ένας αριθμός από πρωτότυπα υπολογιστικά συστήματα τα οποία παρέχουν τη δυνατότητα αναζήτησης και ανάκλησης με βάση το περιεχόμενο. Κάποια από αυτά τα συστήματα, συμπεριλαμβανομένων για παράδειγμα των Virage [52], QBIC [41], VisualSEEK [119], Photobook [99], MARS [106], Netra [82], και VideoQ [25], έχουν χρησιμοποιηθεί πειραματικά με επιτυχία τα τελευταία χρόνια και βρίσκονται ήδη στο στάδιο της εμπορικής εκμετάλλευσης. Στις περισσότερες περιπτώσεις, η πληροφορία του οπτικού περιεχομένου εξάγεται μέσω της κατάτμησης και μοντελοποίησης των αντικειμένων, και της επακόλουθης εξαγωγής χαμηλού επιπέδου χαρακτηριστικών των αντικειμένων όπως το χρώμα, η κίνηση, η υφή, το σχήμα, καθώς και η χρονική και χωρική συσχέτιση μεταξύ των αντικειμένων [1,8].

Η εμπειρία που αποκτήθηκε από τη χρήση τέτοιων συστημάτων δεικτοδότησης και ανάκλησης με βάση το περιεχόμενο δείχνει ξεκάθαρα ότι είναι απαραίτητη η ανάπτυξη νέων μορφών αναπαράστασης της οπτικοακουστικής πληροφορίας οι οποίες να ξεπερνούν την απλή αναπαράσταση των συμπεσμένων ή ασυμπίεστων καρτέ, ή ακόμη και την αναπαράσταση του MPEG-4 που βασίζεται στα αντικείμενα. Για το λόγο αυτό, ένα αναπόσπαστο τμήμα της

διαδικασίας προτυποποίησης MPEG-7 είναι ο καθορισμός ενός συνόλου από *Περιγραφείς* (Descriptors, D) και *Σχήματα Περιγραφής* (Description Schemes, DS) πληροφορίας πολυμέσων σε μία ειδική γλώσσα, τη *Γλώσσα Ορισμού Περιγραφής* (Description Definition Language, DDL) [66]. Παρά το γεγονός ότι οι περιγραφείς αυτοί δεν εξαρτώνται από το τρόπο με τον οποίο το οπτικοακουστικό υλικό παράγεται, κωδικοποιείται, αποθηκεύεται ή χρησιμοποιείται, κλειδί για τα μελλοντικά συστήματα πολυμέσων θα αποτελέσει ο βαθμός στον οποίο υψηλού επιπέδου σημασιολογική πληροφορία θα μπορεί να εξαχθεί αυτόματα για την επιτάχυνση της διαδικασίας χαρακτηρισμού.

Πρέπει επίσης να τονιστεί ότι τα περισσότερα από τα υφιστάμενα συστήματα δεικτοδότησης και ανάκλησης με βάση το περιεχόμενο περιορίζονται στις ακίνητες εικόνες, ενώ η επέκτασή τους για τη διαχείριση βάσεων δεδομένων βίντεο δεν είναι απλή, καθώς η εκτέλεση αναζήτησης σε κάθε καρέ μιας ακολουθίας βίντεο είναι χρονοβόρα και αναποτελεσματική. Οι απαιτήσεις αποθήκευσης του ψηφιοποιημένου οπτικοακουστικού υλικού, ακόμη και αν το υλικό είναι συμπίεσμένο, είναι πολύ μεγάλες και αντιμετωπίζονται με δυσκολία από τα περισσότερα υπολογιστικά συστήματα. Η κατάσταση χειροτερεύει όταν οι βάσεις δεδομένων είναι κατανομημένες σε απομακρυσμένες πλατφόρμες, επιβάλλοντας υψηλές απαιτήσεις ως προς το εύρος ζώνης μετάδοσης. Γι' αυτό το λόγο, πέρα από την ανάπτυξη αλγορίθμων για την αποτελεσματική σχεδίαση δικτύου μέσω της μοντελοποίησης των πηγών βίντεο, θα πρέπει επίσης να υλοποιηθούν νέες μέθοδοι για την αναπαράσταση του οπτικού περιεχομένου.

Παρά το γεγονός ότι η συντριπτική πλειοψηφία των αρχείων βίντεο αποτελείται από διδιάστατες (2-Δ) ακολουθίες, η χρήση *τριδιάστατων* (3-Δ) ακολουθιών βίντεο, οι οποίες συνήθως προέρχονται από στερεοσκοπικά συστήματα κάμερας, έχει αυξηθεί σημαντικά αφού οι ακολουθίες αυτές παρέχουν καλύτερη αναπαράσταση του οπτικού περιεχομένου και βελτιώνουν την επικοινωνία πολυμέσων. Οι τριδιάστατες ακολουθίες επιτρέπουν τη αποδοτικότερη διαχείριση των αντικειμένων με την εκμετάλλευση της πληροφορίας βάθους που προκύπτει από τη *στερεοσκοπική ανάλυση*. Επιπλέον, το πρόβλημα της κατάτμησης σε αντικείμενα αντιμετωπίζεται με μεγαλύτερη ακρίβεια, αφού τα αντικείμενα αποτελούνται συνήθως από περιοχές εικόνων που ανήκουν στο ίδιο επίπεδο βάθους [44]. Ένας μεγάλος αριθμός εφαρμογών, όπως για παράδειγμα η αυτόνομη επιτήρηση (surveillance), η δεικτοδότηση, η ανάκληση και η διαχείριση οπτικοακουστικού υλικού, μπορούν να ωφεληθούν από την τριδιάστατη αναπαράσταση. Γι' αυτό το λόγο συσκευές σύλληψης 3-Δ δεδομένων, στερεοσκοπικά συστήματα κάμερας και συστήματα τριδιάστατης απεικόνισης έχουν παρουσιάσει σημαντική ανάπτυξη και διάδοση, και συνεπώς τα αρχεία τριδιάστατου βίντεο αναμένεται να αυξηθούν με ταχείς ρυθμούς τα επόμενα χρόνια.

Μια πηγή οπτικοακουστικού υλικού στην οποία η δεικτοδότηση με βάση το περιεχόμενο αποκτά ιδιαίτερη σημασία είναι τα ψηφιακά αρχεία *τηλεοπτικών δελτίων ειδήσεων*, για δύο κυρίως λόγους. Πρώτον, η τυπική οργάνωσή τους σε θεματικές ενότητες που περιλαμβάνουν επαναλαμβανόμενη εμφάνιση παρουσιαστών, ρεπορτάζ και εξωτερικών λήψεων, επιτρέπει εύκολη *συντακτική ανάλυση* (parsing) [152], δηλαδή χρονική κατάτμηση σε στοιχειώδεις ενότητες. Δεύτερον, συστήματα πλήρως αυτοματοποιημένης δεικτοδότησης και χαρακτηρισμού δελτίων ειδήσεων θα είναι πολύτιμα για τους αναλυτές δεδομένων σε πρακτορεία ειδήσεων, δημοσιογραφικούς οργανισμούς, κρατικές υπηρεσίες, τηλεοπτικούς παραγωγούς και υπηρεσίες παροχής πληροφοριών [87], καθώς για το σκοπό αυτό χρησιμοποιούνται παραδοσιακά χειρονα-

κτικές μέθοδοι με υψηλό κόστος. Για τους παραπάνω λόγους στο πλαίσιο της διατριβής εξετάζεται η αυτόματη συντακτική ανάλυση δελτίων ειδήσεων με βάση την οπτική πληροφορία.

2. ΥΠΑΡΧΟΥΣΑ ΚΑΤΑΣΤΑΣΗ – ΣΧΕΤΙΚΕΣ ΕΡΓΑΣΙΕΣ

Για την αντιμετώπιση του μεγάλου όγκου πληροφορίας που εμπλέκεται στις ακολουθίες βίντεο μία αποτελεσματική τεχνική είναι η *εξαγωγή περίληψης* (summarization), δηλαδή η εξαγωγή μιας περιορισμένης ποσότητας πληροφορίας, η οποία όμως περιγράφει ικανοποιητικά το οπτικό περιεχόμενο των ακολουθιών. Κάποιες τεχνικές [97,149] προσανατολίζονται στην ανίχνευση αλλαγής πλάνων· μπορούν επομένως να χρησιμοποιηθούν ως ένα πρώτο στάδιο αλγορίθμων εξαγωγής περίληψης. Στην εργασία [89] έχει προταθεί η αναπαράσταση μιας ακολουθίας βίντεο με βάση την εξαγωγή καρτέ σε ομοιόμορφα χρονικά διαστήματα. Ο προτεινόμενος αλγόριθμος όμως δεν εκμεταλλεύεται ούτε την πληροφορία των πλάνων ούτε την οπτική ομοιότητα των καρτέ. Έτσι μπορεί να σημειωθεί απώλεια σημαντικών πλάνων μικρής διάρκειας, ενώ αντίθετα πλάνα με μεγάλη διάρκεια μπορούν να αντιπροσωπεύονται από πολλαπλά καρτέ παρόμοιου περιεχομένου, εισάγοντας πλεονάζουσα πληροφορία.

Η εκμετάλλευση της πληροφορίας των πλάνων και η επιλογή ενός καρτέ για την αναπαράσταση κάθε πλάνου έχει προταθεί στις εργασίες [5,120]. Όμως η επιλογή ενός και μόνο καρτέ δε μπορεί πάντα να παρέχει επαρκή πληροφορία για το οπτικό περιεχόμενο ενός πλάνου, ιδιαίτερα για τα πλάνα μεγάλης διάρκειας. Πρόσφατα έχουν παρουσιασθεί κάποιες άλλες τεχνικές [62,137] οι οποίες βασίζονται στην κατασκευή μιας συμπαγούς αναπαράστασης πλάνων ή ενός *μωσαϊκού εικόνων* (image mosaic). Συγκεκριμένα, στην εργασία [62] όλα τα καρτέ ενός πλάνου ευθυγραμμίζονται σε σχέση με το επικρατέστερο αντικείμενο, ενώ στην [137] απεικονίζεται μία πανοραμική άποψη των καρτέ σε ένα «μωσαϊκό». Παρά το γεγονός ότι μια τέτοια αναπαράσταση μπορεί να είναι πολύ αποτελεσματική για συγκεκριμένες εφαρμογές, όπως για παράδειγμα αθλητικά προγράμματα ή εσωτερικές παραγωγές, δεν μπορεί να εφαρμοσθεί σε πραγματικά πλάνα σύνθετου περιεχομένου στα οποία εμπλέκονται ταυτόχρονες μεταβολές του *φόντου* (background) και των αντικειμένων στο *προσκήνιο* (foreground) ή πολύπλοκες κινήσεις κάμερας.

Μία μέθοδος για την ανάλυση βίντεο και την κατασκευή μιας εικονικής περίληψης για την οπτική αναπαράσταση του περιεχομένου έχει προταθεί στην εργασία [150]. Η εργασία αυτή επικεντρώνεται στην διαμέριση μιας ακολουθίας βίντεο σε διαδοχικά τμήματα (θεματικές ενότητες, story units) και στη συνέχεια στην κατασκευή ενός “poster” για κάθε θεματική ενότητα, το οποίο βασίζεται στο περιεχόμενο των αντίστοιχων πλάνων· έτσι δεν επιλέγονται κάποια καρτέ από κάθε πλάνο. Μία εναλλακτική μέθοδος εξαγωγής περίληψης βασίζεται στην επιλογή πολλαπλών καρτέ [149] σε χρονικές στιγμές όπου το συσσωρευτικό άθροισμα ενός μέτρου απόστασης καρτέ ξεπερνά ένα προκαθορισμένο κατώφλι· η αυθαίρετη επιλογή του κατωφλίου όμως επιβάλλει σημαντικούς περιορισμούς στην τεχνική αυτή. Άλλες προσεγγίσεις για τη δεικτοδότηση βίντεο μέσω εξαγωγής περίληψης περιλαμβάνονται στις εργασίες [26,58,94].

Όσον αφορά στην αναζήτηση και ανάκληση οπτικοακουστικού υλικού, έχει παρατηρηθεί ότι ο παραδοσιακός *χαρακτηρισμός με λέξεις-κλειδιά* (keyword annotation) έχει το μειονέκτημα ότι, πέρα από τη χειρονακτική εργασία που απαιτείται για το χαρακτηρισμό, το πλούσιο περιεχόμενο των εικόνων και του βίντεο δε μπορεί να περιγραφεί αποτελεσματικά με χρήση

κειμένου μόνο [26,107]. Επιπλέον, όσο το μέγεθος των αρχείων βίντεο αυξάνει, η αναζήτηση με βάση το κείμενο γίνεται ανέφικτη, καθώς το οπτικό περιεχόμενο ερμηνεύεται διαφορετικά από κάθε σύστημα δεικτοδότησης [107]. Γι' αυτό το λόγο προτείνεται η αναζήτηση με βάση το οπτικό περιεχόμενο, δηλαδή χαρακτηριστικά όπως το χρώμα, η υφή, η κίνηση ή το σχήμα. Η προσέγγιση αυτή παρέχει καλύτερη οργάνωση και διαχείριση του υλικού, αφού η διαδικασία χαρακτηρισμού είναι αυτόματη.

Η ανάκληση στατικών έγχρωμων εικόνων με χρήση μοντέλων Markov έχει εξετασθεί στην εργασία [80], ενώ στην [19] έχει παρουσιασθεί μία προσέγγιση *ανάκλησης μέσω σχεδιαγράμματος*. Στην εργασία [138] έχει προταθεί μία μέθοδος ιεραρχικής ομαδοποίησης χρωμάτων για ανάκληση ακίνητων εικόνων με βάση το περιεχόμενο. Αντίστοιχα ένας μηχανισμός αναζήτησης σε βάσεις δεδομένων εικόνων με βάση το σχήμα έχει αναλυθεί στην εργασία [29], η οποία χρησιμοποιεί μία μέθοδο εμπνευσμένη από τις παραμορφώσεις μηχανικών συστημάτων. Όσον αφορά στην ανάκληση βίντεο, στην εργασία [94] έχει προταθεί μία μέθοδος ανάλυσης με τρισδιάστατα wavelets η οποία είναι χρήσιμη για προοδευτική δεικτοδότηση κίνησης. Μία τεχνική αυτόματης κατάτμησης και ανάκλησης με βάση την ανάλυση PCA (principal component analysis) σε ακολουθίες που έχουν υποστεί υποδειγματοληψία έχει αναφερθεί στην εργασία [53], ενώ η μοντελοποίηση και κατάτμηση σε αντικείμενα έχει εξετασθεί στην εργασία [45].

Ο χαρακτηρισμός του οπτικού περιεχομένου για σκοπούς δεικτοδότησης βασίζεται συνήθως σε χαμηλού επιπέδου χαρακτηριστικά όπως το χρώμα, η κίνηση, η υφή και το σχήμα. Πρόσφατα η χρήση του σχήματος του περιγράμματος αντικειμένων έχει αποκτήσει ιδιαίτερη σημασία για δύο λόγους: πρώτον, διότι παρέχει ένα ισχυρό εργαλείο για αναζήτηση και ανάκληση εικόνων με βάση το περιεχόμενο, χρησιμοποιώντας ένα μηχανισμό *αναζήτησης μέσω σχεδιαγράμματος / σκίτσου* [19], και δεύτερον, διότι στις τεχνικές κωδικοποίησης *δεύτερης γενιάς*, που βασίζονται στα αντικείμενα [110,128], η πληροφορία σχήματος ενσωματώνεται στα επίπεδα αντικειμένων με τη μορφή δυαδικής εικόνας (μάσκας) και μπορεί να χρησιμοποιηθεί για την πρόβλεψη ή χρονική παρεμβολή των τμημάτων μιας ακολουθίας [84].

Για την περιγραφή του σχήματος έχουν προταθεί πολυάριθμες τεχνικές στη βιβλιογραφία, όπως για παράδειγμα οι περιγραφείς Fourier [100], οι τροποποιημένοι περιγραφείς Fourier [108], η αυτόματη ανίχνευση σημείων ενδιαφέροντος [113], και οι αναλλοίωτες ποσότητες που βασίζονται στις ροπές [17], στο κυρτό περίβλημα [148] και στα έμμεσα πολυώνυμα [103]. Για την αντιμετώπιση των πιθανών μετασχηματισμών έχει προταθεί η μέθοδος της κανονικοποίησης για τον προσδιορισμό αναλλοίωτων ποσοτήτων στις εργασίες [104,115,126]. Επίσης για το ταίριασμα καμπυλών έχουν προταθεί τεχνικές όπως τα παραμορφούμενα πρότυπα [19,69] και τα μοντέλα ενεργών περιγραμμάτων [77].

Για την ανάλυση δελτίων ειδήσεων έχει αναπτυχθεί πρόσφατα σημαντικός αριθμός προτύπων συστημάτων που επιτρέπουν την αυτόματη ή ημι-αυτόματη ανάλυση και χαρακτηρισμό [22,86,93,152]. Τέτοια συστήματα καθιστούν δυνατή την πλοήγηση ειδήσεων με αλληλεπίδραση του χρήστη, την αναζήτηση με βάση το περιεχόμενο, καθώς και τη δημιουργία εφαρμογών *ειδήσεων κατ' απαίτηση* (news on demand, NoD). Τα περισσότερα απ' αυτά όμως βασίζονται σχεδόν αποκλειστικά σε *γλωσσική πληροφορία* (linguistic cues) και *πληροφορία κειμένου* (textual cues) που είναι διαθέσιμη από την παραγωγή των δελτίων. Είναι όμως κοινά αποδεκτό ότι η *οπτικοακουστική πληροφορία* θα πρέπει να διαδραματίζει σημαντικότερο ρόλο στο μέλλον [134], ενώ η ενσωμάτωση όλων των πληροφοριών περιεχομένου σε μια ενιαία πε-

ριγραφή για τη σημασιολογική κατάτμηση των ειδήσεων θα απαιτήσει μεθόδους τεχνητής νοημοσύνης.

3. ΠΕΡΙΓΡΑΦΗ ΠΡΟΤΕΙΝΟΜΕΝΩΝ ΤΕΧΝΙΚΩΝ

Στο πλαίσιο της διατριβής παρουσιάζεται ένας αριθμός προηγμένων τεχνικών για την αντιμετώπιση των προβλημάτων του αυτόματου χαρακτηρισμού, της δεικτοδότησης, της εξαγωγής περίληψης, της προεπισκόπησης και της αναζήτησης με βάση το περιεχόμενο σε βάσεις δεδομένων με εικόνες ακολουθίες βίντεο. Επίσης προτείνεται μία αποτελεσματική μέθοδος αναπαράστασης του σχήματος των αντικειμένων για ανάκληση με βάση το σχήμα και εφαρμόζεται η χρήση τεχνικών ανάλυσης οπτικού περιεχομένου σε ακολουθίες τηλεοπτικών δελτίων ειδήσεων.

Ανάλυση σε Αντικείμενα

Στο πρώτο στάδιο ανάλυσης πραγματοποιείται εντοπισμός των αντικειμένων σε εικόνες ή ακολουθίες εικόνων και εξαγωγή χαρακτηριστικών χαμηλού επιπέδου από τα αντικείμενα αυτά. Ακολουθείται η τεχνική της *κατάτμησης* εικόνων με βάση ένα κριτήριο ομοιομορφίας χαρακτηριστικών όπως το χρώμα, η κίνηση, η υφή, το σχήμα ή το πεδίο βάθους, έτσι ώστε κάθε τμήμα εικόνας που εξάγεται να αντιστοιχεί π.χ. σε περιοχή με ομοιόμορφο χρώμα. Για την κατάτμηση χρησιμοποιείται ο αλγόριθμος M-RSST που αποτελεί μια γρήγορη, πολυδιακριτική υλοποίηση του αλγορίθμου RSST και βασίζεται στην αναδρομική συγχώνευση περιοχών σε εικόνες διαδοχικών επιπέδων ανάλυσης. Ο αλγόριθμος αυτός εφαρμόζεται για κατάτμηση εικόνων με βάση το *χρώμα* και την *κίνηση*, ενώ στην ειδική περίπτωση των *στερεοσκοπικών ακολουθιών*, πραγματοποιείται ανίχνευση και εκμετάλλευση της πληροφορίας βάθους μέσω στερεοσκοπικής ανάλυσης, εκτίμησης του πεδίου απόκλισης και βάθους, καθώς και ανίχνευσης και αντιστάθμισης επικαλύψεων.

Η κατάτμηση με βάση το *πεδίο βάθους*, η οποία υλοποιείται και πάλι με τον αλγόριθμο M-RSST, δίνει μία πολύ καλή προσέγγιση των πραγματικών αντικειμένων που εμφανίζονται σε μία εικόνα, αφού τα τελευταία συνήθως αποτελούνται από περιοχές που βρίσκονται στο ίδιο επίπεδο βάθους. Όμως τα περιγράμματα των τμημάτων βάθους είναι αναξιόπιστα, ενώ αντίθετα η κατάτμηση με βάση το χρώμα δίνει ακριβή περιγράμματα αλλά συνήθως διαχωρίζει κάθε αντικείμενο σε περισσότερα από ένα τμήματα. Για το λόγο αυτό εισάγεται ένας *αλγόριθμος συγχώνευσης* των τμημάτων χρώματος με βάση την πληροφορία των τμημάτων βάθους. Έτσι οι δύο ανεξάρτητες κατατμήσεις συνδυάζονται και προσεγγίζονται ικανοποιητικά τα σημασιολογικά αντικείμενα μιας ακολουθίας. Η ίδια ακριβώς τεχνική μπορεί να χρησιμοποιηθεί για την συγχώνευση των τμημάτων χρώματος με τα τμήματα κίνησης.

Χαρακτηρισμός, Εξαγωγή Περίληψης και Ανάκληση

Η *αναπαράσταση* του οπτικού περιεχομένου εικόνων και ακολουθιών βίντεο πραγματοποιείται με τέτοιο τρόπο ώστε να επιτρέπεται η ενιαία περιγραφή όλων των διαθέσιμων χαρακτηριστικών και να διευκολύνονται οι συγκρίσεις εικόνων, καρέ ή πλάνων ακολουθιών. Χρησιμοποιούνται τα χαρακτηριστικά των τμημάτων που προκύπτουν από την κατάτμηση με βάση το χρώμα, την κίνηση και το πεδίο βάθους, περιλαμβάνοντας για παράδειγμα τη θέση, το μέγεθος, τις χρωματικές συνιστώσες, το μέσο διάνυσμα κίνησης και το βάθος του κάθε τμήματος.

Τα χαρακτηριστικά αυτά συνδυάζονται για την κατασκευή *διανυσμάτων περιγραφής*, τα οποία έχουν τη μορφή πολυδιάστατων ασαφών ιστογραμμάτων.

Για την εξαγωγή περίληψης ακολουθείται διαμέριση μιας ακολουθίας βίντεο σε διαδοχικά πλάνα, επιλογή ορισμένων *αντιπροσωπευτικών* ή *χαρακτηριστικών πλάνων* (representative shots) και στη συνέχεια η εξαγωγή ενός περιορισμένου αριθμού *χαρακτηριστικών καρτέ* (key frames) από τα επιλεγμένα πλάνα. Η επιλογή είναι ουσιαστικά μια διαδικασία *δειγματοληψίας με βάση το περιεχόμενο* (content-based sampling) [114]. Για την επιλογή χαρακτηριστικών καρτέ προτείνονται δύο διαφορετικές μέθοδοι: η μέθοδος *χρονικής μεταβολής* που βασίζεται στην επιλογή ακρότατων σημείων της τροχιάς του διανύσματος περιγραφής, και η μέθοδος *συσχέτισης* που βασίζεται στην ελαχιστοποίηση ενός κριτηρίου συσχέτισης μεταξύ των διανυσμάτων περιγραφής. Λόγω της πολυπλοκότητας της μεθόδου συσχέτισης, προτείνονται δύο εναλλακτικές τεχνικές ταχείας υλοποίησης, και συγκεκριμένα *λογαριθμική αναζήτηση* και *γενετικός αλγόριθμος*.

Η προτεινόμενη επιλογή χαρακτηριστικών πλάνων και καρτέ επιτρέπει εύκολη και ταχεία πλοήγηση σε μεγάλες βάσεις δεδομένων οπτικοακουστικού υλικού και αποτελεί ισχυρό εργαλείο εξαγωγής περίληψης και προεπισκόπησης. Για παράδειγμα, έχει παρατηρηθεί στην εργασία [150] ότι μία ακολουθία βίντεο διάρκειας 30 λεπτών αποτελείται συνήθως από 200 πλάνα. Επομένως, επιλέγοντας κατά μέσο όρο πέντε χαρακτηριστικά καρτέ από κάθε πλάνο, απαιτούνται μόνο 1,000 από τα 45,000 συνολικά καρτέ για την αναπαράσταση του οπτικού περιεχομένου. Επιπλέον, όσον αφορά στην αναζήτηση με βάση το περιεχόμενο, αντί για την εκτέλεση αναζήτησης σε όλα τα διαθέσιμα καρτέ μιας ακολουθίας, μπορεί κανείς να θεωρήσει μόνο τα επιλεγμένα χαρακτηριστικά καρτέ. Έτσι η αναζήτηση και ανάκληση πραγματοποιείται ταχύτερα και αποδοτικότερα, αφού η πλεονάζουσα πληροφορία απορρίπτεται.

Για το πρόβλημα της γενικευμένης *αναζήτησης* και *ανάκλησης* εικόνων και ακολουθιών με βάση το περιεχόμενο προτείνονται δύο εναλλακτικοί μηχανισμοί αναζήτησης, μέσω *παραδείγματος* και μέσω *σχεδιαγράμματος*, ενώ και στις δύο περιπτώσεις υιοθετείται μία τεχνική *ανάδρασης* από το χρήστη, η οποία καθιστά την αναζήτηση μία διαδικασία αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστικού συστήματος. Με αυτό τον τρόπο ο χρήστης απαλλάσσεται από το βάρος της επιλογής παραμέτρων ή της κατανόησης χαρακτηριστικών και αναπαραστάσεων χαμηλού επιπέδου· παράλληλα το υπολογιστικό σύστημα εφοδιάζεται με ένα εργαλείο για την αυτόματη αντιστοίχιση των χαρακτηριστικών χαμηλού επιπέδου σε υψηλού επιπέδου προδιαγραφές αναζήτησης, καθώς και για την αντιμετώπιση της υποκειμενικότητας με την οποία ο χρήστης καθορίζει τις προδιαγραφές αυτές.

Περιγραφή Σχήματος

Για την περιγραφή του περιγράμματος αντικειμένων, στο πλαίσιο της διατριβής προτείνεται μία μέθοδος *κανονικοποίησης* η οποία παρέχει περιγραφή *αναλλοίωτη σε affine μετασχηματισμούς* (affine-invariant) καθώς και μετασχηματισμούς παραμέτρου, χωρίς ταυτόχρονα να παρουσιάζει καμία απώλεια πληροφορίας σχήματος. Οι καμπύλες που αναπαριστούν τα περιγράμματα μοντελοποιούνται αρχικά με *B-Splines*· στη συνέχεια πραγματοποιείται σταδιακή κανονικοποίηση ως προς τους μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας, *κλίσης* (skew), σημείου αναφοράς, περιστροφής, κατοπτρισμού και φοράς διαγραφής. Η κανονικοποίηση είναι ένας γραμμικός μετασχηματισμός που πραγματοποιείται με χρήση πληροφορίας όπως οι ροπές μέχρι δευτέρου βαθμού και ο μετασχηματισμός Fourier κάθε καμπύλης.

Η προτεινόμενη μέθοδος κανονικοποίησης μπορεί να εφαρμοσθεί ως ένα στάδιο προεπεξεργασίας σε οποιαδήποτε τεχνική αναπαράστασης, ταξινόμησης, αναγνώρισης ή ανάκλησης με βάση το σχήμα, αφού ουσιαστικά αποσυσχετίζει το πρόβλημα της *affine-invariant* περιγραφής από την εξαγωγή χαρακτηριστικών και το ταίριασμα με πρότυπα. Το υπολογιστικό κόστος είναι αμελητέο, σε σχέση τουλάχιστον με το κόστος άλλων σταδίων επεξεργασίας, όπως π.χ. της κατάτμησης. Έτσι είναι δυνατή η ενσωμάτωση της μεθόδου σε συστήματα πραγματικού χρόνου για ανάκληση εικόνων ή ακόμη και κωδικοποίηση βίντεο. Το πρόβλημα της *ανάκλησης με βάση το σχήμα* αντιμετωπίζεται με τη χρήση τριών απλών μέτρων ομοιότητας των κανονικοποιημένων περιγραμμάτων, ενώ εξετάζεται και το συναφές πρόβλημα της *ταξινόμησης* σχημάτων σε προκαθορισμένες κατηγορίες, το οποίο αντιμετωπίζεται με χρήση νευρωνικού δικτύου.

Ανάλυση Δελτίων Ειδήσεων

Τέλος προτείνεται η εφαρμογή τεχνικών αυτόματης ανάλυσης βίντεο στην ειδική περίπτωση των *τηλεοπτικών δελτίων ειδήσεων*. Επειδή το επικρατέστερο τμήμα περιεχομένου των δελτίων ειδήσεων σχετίζεται με ανθρώπινες δραστηριότητες, οι εικόνες ανθρώπων και ιδιαίτερα προσώπων παίζουν σημαντικότερο ρόλο στην ανάλυση της δομής των δελτίων· έτσι η προτεινόμενη τεχνική στηρίζεται στην *ανίχνευση προσώπων*. Παρουσιάζεται μία πλήρως αυτοματοποιημένη διαδικασία ανίχνευσης και εντοπισμού προσώπων, η οποία αποτελείται από μία σειρά διαδοχικών σταδίων στα οποία ανιχνεύεται αρχικά το χρώμα του ανθρώπινου δέρματος και στη συνέχεια πραγματοποιείται κατάτμηση με βάση το χρώμα, συγχώνευση των τμημάτων δέρματος, ανάλυση του σχήματος των τμημάτων και επαλήθευση με χρήση μιας τεχνικής ταίριασματος με πρότυπο πρόσωπου.

Ο αξιόπιστος εντοπισμός ανθρώπινων προσώπων σε συνδυασμό με άλλα χαρακτηριστικά χρώματος και κίνησης σε μία ακολουθία δελτίων ειδήσεων χρησιμοποιείται για την *ταξινόμηση* των πλάνων της ακολουθίας σε προκαθορισμένες κατηγορίες, με χρήση ενός αριθμού απλών κανόνων. Με δεδομένο ότι στην τυπική οργάνωση των τηλεοπτικών ειδήσεων περιλαμβάνονται επαναλαμβανόμενες εμφανίσεις του παρουσιαστή εναλλάξ με άλλα είδη πλάνων, η ταξινόμηση αυτή οδηγεί άμεσα στη *συντακτική ανάλυση*, δηλαδή στη χρονική κατάτμηση των δελτίων σε στοιχειώδεις ενότητες ειδήσεων. Στη συνέχεια εξετάζεται η εφαρμογή των προτεινόμενων μεθόδων χαρακτηρισμού, δεικτοδότησης, εξαγωγής περίληψης, προεπισκόπησης και αναζήτησης με βάση το περιεχόμενο σε δελτία ειδήσεων, ενώ προτείνεται και ένας αριθμός από νέα *σενάρια αναζήτησης* που σχετίζονται με την ύπαρξη ανθρώπινων προσώπων, και συγκεκριμένα η αναζήτηση με βάση τα χαρακτηριστικά χρώματος, το μέγεθος και το πλήθος των προσώπων.

Άλλες Εργασίες

Αξίζει στο σημείο αυτό να αναφερθούν τρεις εργασίες του συγγραφέα που δημοσιεύθηκαν κατά τη διάρκεια της διδακτορικής διατριβής, οι οποίες όμως δεν εντάσσονται άμεσα στο ερευνητικό της αντικείμενο και γι' αυτό δεν περιλαμβάνονται στο παρόν σύγγραμμα. Συγκεκριμένα, στην εργασία [133] πραγματοποιείται εντοπισμός προσώπων με τη μέθοδο ταίριασματος με πρότυπο, ανίχνευση κίνησης, υπολογισμός του μετασχηματισμού Radon των διανυσμάτων κίνησης και κανονικοποίηση των καμπυλών που προκύπτουν με τελικό στόχο

τον αυτόματο χαρακτηρισμό *ανθρώπινων εκφράσεων*, η ταξινόμηση των οποίων πραγματοποιείται με νευρωνικό δίκτυο.

Στην εργασία [6] παρουσιάζεται μία πολυδιακριτική υλοποίηση ασαφούς ταξινόμησης εικόνων με χρήση νευρωνικού δικτύου και εφαρμογή στην ταξινόμηση εικόνων δορυφόρου σε χερσαίες και θαλάσσιες περιοχές. Τέλος στην εργασία [7] προτείνεται μία πρωτότυπη τεχνική σχεδίασης συστημάτων ανακατασκευής εικόνων από αισθητήρες υπερήχων. Η τεχνική βασίζεται στην εκπομπή και λήψη σημάτων ευρείας ζώνης που προέρχονται από *ψευδοτυχαίες ακολουθίες* (pseudorandom sequences), όμοια με τις τεχνικές *πολλαπλής προσπέλασης διαίρεσης κώδικα* (code division multiple access, CDMA) που χρησιμοποιούνται στις τηλεπικοινωνίες, και επιτυγχάνει σημαντική βελτίωση στην ταχύτητα ανακατασκευής και στην ανάλυση των εικόνων.

4. ΣΥΝΕΙΣΦΟΡΑ ΔΙΑΤΡΙΒΗΣ

Στο πλαίσιο της παρούσας διδακτορικής διατριβής χρησιμοποιούνται πολλοί γνωστοί αλγόριθμοι και τεχνικές. Σε πολλές περιπτώσεις προτείνονται νέες υλοποιήσεις που παρέχουν μικρότερο υπολογιστικό κόστος ή καλύτερα αποτελέσματα· σε άλλες πραγματοποιείται συνδυασμός υφιστάμενων τεχνικών ενώ τέλος πολλές από τις τεχνικές που προτείνονται είναι πρωτότυπες. Στην Ενότητα αυτή περιγράφεται λοιπόν η συνεισφορά του διατριβής στο ερευνητικό της πεδίο και επισημαίνονται οι δημοσιεύσεις του συγγραφέα που σχετίζονται με κάθε τεχνική.

Ο προτεινόμενος αλγόριθμος κατάρτησης M-RSST αποτελεί νέα υλοποίηση του γνωστού αλγορίθμου RSST [91]. Πρόκειται για μια γρήγορη, πολυδιακριτική υλοποίηση που βασίζεται στην αναδρομική συγχώνευση περιοχών σε εικόνες διαδοχικών επιπέδων ανάλυσης. Παρέχει πολύ μικρότερους χρόνους εκτέλεσης, όπως αναφέρεται στα αντίστοιχα αποτελέσματα, ενώ ταυτόχρονα φιλτράρει τα αντικείμενα ως προς το μέγεθος και επιτυγχάνει καλύτερα αποτελέσματα σε περιπτώσεις πολύπλοκης υφής. Παρουσιάστηκε για πρώτη φορά στην εργασία [35], η υλοποίησή του αναλύθηκε στην [8], ενώ χρησιμοποιήθηκε με επιτυχία για κατάρτηση και στις εργασίες [9,15,34,38,39,131].

Η στερεοσκοπική ανάλυση εικόνων για τη συσχέτιση του βάθους με την απόκλιση είναι γνωστή τεχνική [49], ενώ ο χρησιμοποιούμενος αλγόριθμος για των εκτίμηση πεδίου απόκλισης και βάθους αποτελεί παραλλαγή του αντίστοιχου αλγορίθμου των [135,136]. Η μέθοδος ανίχνευσης και αντιστάθμισης επικάλυψης βασίζεται σε μεγάλο βαθμό στην εργασία [50] και η συγχώνευση αντικειμένων είναι εμπνευσμένη από την εργασία [1], όπου όμως αναφέρεται συγχώνευση χρώματος και κίνησης μόνο, ενώ στο πλαίσιο της διατριβής πραγματοποιείται συγχώνευση χρώματος και βάθους. Όλες οι παραπάνω τεχνικές που εφαρμόζονται σε στερεοσκοπικές ακολουθίες προτάθηκαν στην εργασία [39] αλλά χρησιμοποιήθηκαν και στην [38].

Η μέθοδος αναπαράστασης οπτικού περιεχομένου με διανύσματα περιγραφής που προκύπτουν από ασαφή πολυδιάστατα ιστογράμματα είναι εξολοκλήρου πρωτότυπη· παρουσιάστηκε για πρώτη φορά στην εργασία [32] και από τότε χρησιμοποιήθηκε για σκοπούς χαρακτηρισμού, εξαγωγής περίληψης και ανάκλησης στις εργασίες [13,14,33,36] και αργότερα στις [8,31,34,35,37,38,39]. Η εξαγωγή περίληψης μέσω ομαδοποίησης έχει ξαναχρησιμοποιηθεί [3,74], αλλά η επιλογή χαρακτηριστικών πλάνων μέσω ομαδοποίησης με τον αλγόριθμο

Lloyd-Max [46] προτάθηκε αρχικά στην εργασία [14] και χρησιμοποιήθηκε επίσης στις [31,33,36,39].

Η ιδέα της εξαγωγής χαρακτηριστικών καρέ έχει προταθεί παλαιότερα στην εργασία [5], ενώ η επιλογή τους με βάση τις ιδιότητες της πολυδιάστατης καμπύλης τροχιάς του διανύσματος περιγραφής έχει προταθεί στην [30]. Όμως στο πλαίσιο της διατριβής χρησιμοποιήθηκε για το σκοπό αυτό το μέτρο καμπυλότητας της τροχιάς· η τεχνική αυτή δημοσιεύθηκε αρχικά στην εργασία [32], ενώ χρησιμοποιήθηκε και στις [8,14]. Η τεχνική λογαριθμικής αναζήτησης για την υλοποίηση της μεθόδου συσχέτισης είναι πρωτότυπη, δημοσιεύθηκε για πρώτη φορά στην εργασία [36] και ξαναχρησιμοποιήθηκε στην [37], αλλά και στην [8] όπου γίνεται εκτεταμένη σύγκριση με την τεχνική του γενετικού αλγορίθμου. Η τελευταία είναι επίσης πρωτότυπη: η πρώτη της δημοσίευση έγινε στην εργασία [33] και στη συνέχεια λόγω των σημαντικών επιδόσεων της χρησιμοποιήθηκε σε πολλές ακόμη εργασίες [8,31,34,38,39]. Η ανάκληση μέσω παραδείγματος και σχεδιαγράμματος, καθώς και η τεχνική της ανάδρασης είναι γνωστές [106,107]. Όμως η προτεινόμενη στο πλαίσιο της διατριβής τεχνική ενημέρωσης παραμέτρων προτάθηκε για πρώτη φορά στην εργασία [13] και χρησιμοποιήθηκε επίσης στην [35].

Η ιδέα της κανονικοποίησης για την εξαγωγή αναλλοίωτων ποσοτήτων έχει παρουσιασθεί στην εργασία [104] και αξιοποιηθεί για την κανονικοποίηση εικόνων στην [115]. Όμως η προτεινόμενη τεχνική κανονικοποίησης για τη δημιουργία πλήρως αναλλοίωτης αναπαράστασης 2-Δ κλειστών καμπυλών σε affine μετασχηματισμούς καθώς και μετασχηματισμούς παραμέτρου χωρίς απώλεια πληροφορίας είναι εξολοκλήρου πρωτότυπη. Η αρχική ιδέα της κανονικοποίησης δημοσιεύθηκε για πρώτη φορά στην εργασία [11], και χρησιμοποιήθηκε ακόμη στην [10]. Οι ιδιότητές της και οι λεπτομέρειες υλοποίησής της παρουσιάστηκαν στην [16] όπου χρησιμοποιήθηκε και για ανάκληση με βάση το σχήμα· αναλυτικότερη περιγραφή καθώς και αποδείξεις των σχετικών προτάσεων αναμένεται να δημοσιευθούν στην εργασία [12]. Η ταξινόμηση 2-Δ καμπυλών με χρήση B-Splines και νευρωνικού δικτύου παρουσιάστηκε στις εργασίες [10,11,145].

Ο συνδυασμός ανίχνευσης χρώματος δέρματος, προσαρμογής του χρωματικού μοντέλου, κατάτμησης με βάση το χρώμα, συγχώνευσης τμημάτων δέρματος και ανάλυσης σχήματος για εντοπισμό ανθρώπινων προσώπων είναι πρωτότυπος, αν και οι επιμέρους τεχνικές έχουν ξαναχρησιμοποιηθεί. Δημοσιεύθηκε αρχικά στην εργασία [131], ενώ ο επιπλέον συνδυασμός με χαρακτηριστικά υφής στο στάδιο επαλήθευσης (ταίριασμα με πρότυπο πρόσωπου) για μεγαλύτερη ακρίβεια ανίχνευσης παρουσιάστηκε στην εργασία [132]. Η προτεινόμενη τεχνική συντακτικής ανάλυσης δελτίων ειδήσεων με χρήση οπτικής πληροφορίας είναι επίσης πρωτότυπη και δημοσιεύθηκε στην εργασία [9]. Τα προτεινόμενα σενάρια αναζήτησης με βάση την πληροφορία προσώπων αποτέλεσαν το αντικείμενο των εργασιών [15,132].

5. ΔΟΜΗ ΔΙΑΤΡΙΒΗΣ

Η διατριβή απαρτίζεται από τέσσερα μέρη. Στο Μέρος I γίνεται μια εισαγωγή στην ανάλυση ακολουθιών εικόνων για τον προσδιορισμό αντικειμένων με βάση το χρώμα, την κίνηση, καθώς και το πεδίο βάθους στην περίπτωση των στερεοσκοπικών ακολουθιών. Στο Κεφάλαιο 1 δίνεται μία περιγραφή του αλγορίθμου M-RSST, ο οποίος εφαρμόζεται για κατάτμηση εικόνων με βάση το χρώμα και την κίνηση. Στο Κεφάλαιο 2 εξετάζεται η ειδική περίπτωση των στερεοσκοπικών ακολουθιών, στις οποίες εφαρμόζεται επιπλέον κατάτμηση με βάση το πεδίο

βάθους. Για το συνδυασμό των επιμέρους κατατιμήσεων προτείνεται στο Κεφάλαιο 3 μια τεχνική συγχώνευσης που επιτρέπει τον αξιόπιστο προσδιορισμό σημασιολογικών αντικειμένων.

Το Μέρος II αναφέρεται στο χαρακτηρισμό του οπτικού περιεχομένου, την εξαγωγή περίληψης και την αναζήτηση με βάση το περιεχόμενο. Ο υπολογισμός διανυσμάτων περιγραφής περιγράφεται στο Κεφάλαιο 4. Η πληροφορία χαρακτηρισμού αποτελεί τη βάση για την εξαγωγή περίληψης η οποία υλοποιείται με την αυτόματη επιλογή ενός περιορισμένου συνόλου χαρακτηριστικών πλάνων και καρτέ που περιγράφουν ικανοποιητικά το οπτικό περιεχόμενο. Η επιλογή χαρακτηριστικών πλάνων παρουσιάζεται στο Κεφάλαιο 5, ενώ η επιλογή χαρακτηριστικών καρτέ στα Κεφάλαια 6 και 7 για τις μεθόδους χρονικής μεταβολής και συσχέτισης αντίστοιχα. Η αναπαράσταση του επιλεγμένου υλικού με διανύσματα περιγραφής χρησιμοποιείται στη συνέχεια για αναζήτηση και ανάκληση με βάση το περιεχόμενο στο Κεφάλαιο 8.

Στο Μέρος III εξετάζεται το πρόβλημα της ανάλυσης και αναπαράστασης του περιγράμματος αντικειμένων με εφαρμογή στην ταξινόμηση και αναζήτηση αντικειμένων με βάση το σχήμα. Στο Κεφάλαιο 9 πραγματοποιείται μία βιβλιογραφική ανασκόπηση στις διαθέσιμες μεθόδους αναπαράστασης, ταξινόμησης, ταιριάσματος και αναγνώρισης περιγραμμάτων, δίνεται η γενική περιγραφή της προτεινόμενης μεθόδου κανονικοποίησης περιγραμμάτων, των ιδιοτήτων, προϋποθέσεων και περιορισμών της, και παρουσιάζεται η μοντελοποίηση καμπυλών με *B-Splines*. Στο Κεφάλαιο 10 δίνεται η αναλυτική περιγραφή και θεωρητική ανάλυση της μεθόδου κανονικοποίησης. Στο Κεφάλαιο 11 εξετάζεται το πρόβλημα της ταξινόμησης και ανάκλησης αντικειμένων με βάση το σχήμα τους και δίνονται παραδείγματα και πειραματικά αποτελέσματα τόσο για την κανονικοποίηση περιγραμμάτων όσο και για την αναζήτηση με βάση το σχήμα.

Τέλος, στο Μέρος IV παρουσιάζεται η εφαρμογή των μεθόδων ανάλυσης οπτικής πληροφορίας σε ακολουθίες τηλεοπτικών δελτίων ειδήσεων. Στο Κεφάλαιο 12 περιγράφεται η χρησιμοποιούμενη τεχνική αυτόματης ανίχνευσης και εντοπισμού ανθρώπινων προσώπων, ενώ στο Κεφάλαιο 13 η ταξινόμηση πλάνων και η συντακτική ανάλυση, δηλαδή η χρονική κατάτμηση των δελτίων σε στοιχειώδεις θεματικές ενότητες. Η παραγόμενη πληροφορία προσώπων χρησιμοποιείται για την υλοποίηση νέων, σημασιολογικών κριτηρίων αναζήτησης με βάση το περιεχόμενο που δίνονται στο Κεφάλαιο 14. Η διατριβή καταλήγει σε έκθεση συμπερασμάτων και προτάσεων για ζητήματα που θα μπορούσαν να αποτελέσουν τη βάση για μελλοντική έρευνα.

ΑΝΑΛΥΣΗ ΑΚΟΛΟΥΘΙΩΝ ΕΙΚΟΝΩΝ ΣΕ ΑΝΤΙΚΕΙΜΕΝΑ

Αντικείμενο του πρώτου Μέρους της διατριβής είναι ο εντοπισμός των αντικειμένων σε εικόνες ή ακολουθίες εικόνων και η εξαγωγή χαρακτηριστικών χαμηλού επιπέδου από τα αντικείμενα αυτά. Η ανάλυση σε αντικείμενα είναι απαραίτητη για το χαρακτηρισμό και την περιγραφή του οπτικού περιεχομένου με στόχο την ανάπτυξη εφαρμογών όπως η αυτόματη ανάλυση, ο χαρακτηρισμός, η δεικτοδότηση, η εξαγωγή περίληψης, η προεπισκόπηση και η αναζήτηση με βάση το περιεχόμενο· είναι ουσιαστικά το πρώτο βήμα για τον υπολογισμό ενός διανύσματος περιγραφής βάσει του οποίου μπορεί να εκτιμηθεί η ομοιότητα μεταξύ εικόνων, καρτέ, ή πλάνων μιας ακολουθίας. Για τον εντοπισμό των αντικειμένων ακολουθείται η τεχνική της *κατάτμησης* εικόνων με βάση ένα κριτήριο ομοιομορφίας χαρακτηριστικών όπως το χρώμα, η κίνηση, η υφή, το σχήμα ή το πεδίο βάθους, έτσι ώστε κάθε τμήμα εικόνας που εξάγεται να αντιστοιχεί π.χ. σε περιοχή με ομοιόμορφο χρώμα.

Στο πλαίσιο της διατριβής χρησιμοποιείται ο γενικός αλγόριθμος κατάτμησης M-RSST που αποτελεί μια γρήγορη, πολυδιακριτική υλοποίηση του αλγορίθμου RSST και βασίζεται στην αναδρομική συγχώνευση περιοχών σε εικόνες διαδοχικών επιπέδων ανάλυσης. Ο αλγόριθμος είναι ανεξάρτητος από την εξαγωγή των χαρακτηριστικών και μπορεί να χρησιμοποιηθεί για κατάτμηση με βάση οποιοδήποτε χαρακτηριστικό. Στο Κεφάλαιο 1 δίνεται μία συνοπτική περιγραφή του αλγορίθμου, ο οποίος στη συνέχεια εφαρμόζεται για κατάτμηση εικόνων με βάση το *χρώμα* και την *κίνηση*.

Στο Κεφάλαιο 2 εξετάζεται η ειδική περίπτωση των *στερεοσκοπικών ακολουθιών*, η χρήση των οποίων έχει αυξηθεί σημαντικά τα τελευταία χρόνια, λόγω της ανάπτυξης στερεοσκοπικών συστημάτων κάμερας και συσκευών τρισδιάστατης απεικόνισης. Στην περίπτωση αυτή είναι δυνατή η ανίχνευση και εκμετάλλευση της πληροφορίας βάθους μέσω στερεοσκοπικής ανάλυσης. Η κατάτμηση με βάση το *πεδίο βάθους*, η οποία υλοποιείται και πάλι με τον αλγόριθμο M-RSST, δίνει μία πολύ καλή προσέγγιση των πραγματικών αντικειμένων που εμφανίζονται σε μία εικόνα, αφού τα τελευταία συνήθως αποτελούνται από περιοχές που βρίσκονται στο ίδιο επίπεδο βάθους. Όμως τα περιγράμματα των τμημάτων βάθους είναι αναξιόπιστα, ενώ αντίθετα η κατάτμηση με βάση το χρώμα δίνει ακριβή περιγράμματα αλλά συνήθως διαχωρίζει κάθε αντικείμενο σε περισσότερα από ένα τμήματα. Για το λόγο αυτό παρουσιάζεται στο Κεφάλαιο 3 ένας *αλγόριθμος συγχώνευσης* των τμημάτων χρώματος με βάση την πληροφορία των τμημάτων βάθους. Έτσι οι δύο ανεξάρτητες κατατμήσεις συνδυάζονται και εξάγονται με ακρίβεια τα σημασιολογικά αντικείμενα μιας ακολουθίας. Η ίδια ακριβώς τεχνική μπορεί να χρησιμοποιηθεί για τη συγχώνευση των τμημάτων χρώματος με τα τμήματα κίνησης.

Κατάτμηση με Βάση το Χρώμα και την Κίνηση

Εφαρμογές όπως η αυτόματη ανάλυση, ο χαρακτηρισμός, η δεικτοδότηση, η εξαγωγή περίληψης, η προεπισκόπηση, καθώς και η αναζήτηση με βάση το περιεχόμενο από ένα σύνολο εικόνων ή ακολουθιών βίντεο, απαιτούν την περιγραφή του οπτικού περιεχομένου με ένα τρόπο που να επιτρέπει την ταχεία σύγκριση και τον υπολογισμό ενός μέτρου ομοιότητας μεταξύ εικόνων, καρτέ, ή πλάνων μιας ακολουθίας. Κάτι τέτοιο είναι εφικτό με την εξαγωγή ενός διανύσματος περιγραφής που περιέχει πληροφορία η οποία σχετίζεται με χαμηλού επιπέδου χαρακτηριστικά αντικειμένων όπως το χρώμα, η κίνηση, το πεδίο βάθους, το σχήμα και η υφή. Απαραίτητη λοιπόν για την ανάλυση του οπτικού περιεχομένου είναι η κατάτμηση σε αντικείμενα με βάση ένα από τα παραπάνω χαρακτηριστικά. Στο Κεφάλαιο αυτό περιγράφεται αρχικά ο γενικός αλγόριθμος κατάτμησης M-RSST που αναπτύχθηκε στο πλαίσιο της διατριβής. Πρόκειται για μια γρήγορη, πολυδιακριτική υλοποίηση του γνωστού αλγορίθμου RSST η οποία βασίζεται στην αναδρομική συγχώνευση περιοχών σε εικόνες διαδοχικών επιπέδων ανάλυσης. Στη συνέχεια ο αλγόριθμος εφαρμόζεται στην κατάτμηση εικόνων με βάση το χρώμα και την κίνηση των αντικειμένων, ενώ στα επόμενα Κεφάλαια χρησιμοποιείται για την κατάτμηση με βάση το πεδίο βάθους, την εξαγωγή σημασιολογικών αντικειμένων, την εξαγωγή του περιγράμματος και την ανίχνευση τμημάτων προσώπων σε ακολουθίες δελτίων ειδήσεων.

1.1 ΚΑΤΑΤΜΗΣΗ ΑΚΟΛΟΥΘΙΩΝ ΕΙΚΟΝΩΝ ΣΕ ΑΝΤΙΚΕΙΜΕΝΑ

Το πρώτο βήμα ανάλυσης του οπτικού περιεχομένου μιας εικόνας ή ακολουθίας εικόνων είναι η κατάτμηση (segmentation) σε αντικείμενα με σημασιολογικό περιεχόμενο (semantic meaning) και η εξαγωγή ουσιαστικής πληροφορίας που περιγράφει κάθε αντικείμενο. Ο στόχος της σημασιολογικής κατάτμησης (semantic segmentation) είναι να προσδιορισθεί η παρουσία ενός συνόλου περιοχών που αναπαριστούν γνωστά αντικείμενα. Αυτό όμως είναι ένα δύσκολο πρόβλημα, αφού απαιτεί εκ των προτέρων γνώση των αντικειμένων που πρέπει να ανιχνευθούν και επομένως μπορεί να λυθεί μόνο στο πλαίσιο μίας περιορισμένης κατηγορίας εφαρμογών. Για παράδειγμα, η ανίχνευση και εντοπισμός ανθρώπινων προσώπων μπορεί να επιτευχθεί με βάση τις γνωστές ιδιότητες των προσώπων, όπως το χρώμα και το σχήμα τους· έτσι σε εφαρμογές βιντεοτηλεφώνου ή ανάλυσης δελτίων ειδήσεων όπου τα πρόσωπα παίζουν σημαντικό ρόλο, η σημασιολογική κατάτμηση απλουστεύεται [131]. Αυτό ακριβώς είναι και το αντικείμενο του Μέρους IV της διατριβής. Όμως στη γενική περίπτωση όπου το περιεχόμενο είναι άγνωστο, η σημασιολογική κατάτμηση θεωρείται ακόμη ένα άλυτο πρόβλημα [95].

Στο πλαίσιο της διατριβής χρησιμοποιείται κατάτμηση με βάση το χρώμα σε κάθε εικόνα ή καρτέ ακολουθίας· στις περιπτώσεις των ακολουθιών εικόνων πραγματοποιείται επίσης κατάτμηση με βάση την κίνηση ενώ στις στερεοσκοπικές ακολουθίες και κατάτμηση με βάση το πεδίο βάθους. Παρά το γεγονός ότι σε ένα περιβάλλον ανάκλησης με βάση το περιεχόμενο η

σημσιολογική κατάτμηση θα ήταν ιδιαίτερα χρήσιμη, η κατάτμηση με βάση το χρώμα και την κίνηση παρέχει ένα ισχυρό εργαλείο αυτόματου χαρακτηρισμού για την εξαγωγή περιληψής όπου δεν απαιτείται μεγάλη ακρίβεια περιγραφής. Επιπλέον, στην περίπτωση των στερεοσκοπικών ακολουθιών παρουσιάζεται στο Κεφάλαιο 3 ένας αλγόριθμος συγχώνευσης των τμημάτων χρώματος με βάση την πληροφορία των τμημάτων βάθους, ο οποίος προσεγγίζει σε μεγάλο βαθμό τη σημσιολογική κατάτμηση, αφού τα πραγματικά αντικείμενα βρίσκονται συνήθως στο ίδιο επίπεδο βάθους. Ο ίδιος αλγόριθμος μπορεί να χρησιμοποιηθεί και για τη συγχώνευση τμημάτων χρώματος και κίνησης.

Κατά το στάδιο του χαρακτηρισμού η πληροφορία χρώματος, κίνησης και βάθους του κάθε τμήματος διατηρείται έτσι ώστε να παρέχεται ευέλικτη αναπαράσταση του οπτικού περιεχομένου, και η διαχείριση ή αναζήτηση κάθε είδους πληροφορίας να μην επηρεάζεται από τις υπόλοιπες. Συγκεκριμένα, το πλήθος τμημάτων χρώματος καθώς και η θέση, το μέγεθος και οι μέσες χρωματικές συνιστώσες τους χρησιμοποιούνται για την κατασκευή ενός διανύσματος περιγραφής χρώματος. Με όμοιο τρόπο, το πλήθος τμημάτων κίνησης καθώς και η θέση, το μέγεθος και το μέσο διάνυσμα κίνησής τους χρησιμοποιούνται για την κατασκευή ενός διανύσματος περιγραφής κίνησης. Ανάλογα πραγματοποιείται και ο χαρακτηρισμός των τμημάτων βάθους. Τα διανύσματα αυτά συνδυάζονται στη συνέχεια για την κατασκευή ενός συνολικού διανύσματος περιγραφής. Η παραπάνω διαδικασία χαρακτηρισμού περιγράφεται αναλυτικά στο Κεφάλαιο 4. Στο υπόλοιπο του παρόντος Κεφαλαίου περιγράφεται ο χρησιμοποιούμενος αλγόριθμος κατάτμησης, ο οποίος στη συνέχεια εφαρμόζεται για κατάτμηση με βάση το χρώμα και την κίνηση.

1.2 ΑΛΓΟΡΙΘΜΟΙ ΚΑΤΑΤΜΗΣΗΣ

Ο Αλγόριθμος RSST

Ως βάση για την κατάτμηση εικόνων στο πλαίσιο της διατριβής χρησιμοποιήθηκε ο αλγόριθμος *Recursive Shortest Spanning Tree* (RSST) [91]. Παρά τη σχετικά μεγάλη υπολογιστική του πολυπλοκότητα, θεωρείται ένα από τα πιο ισχυρά εργαλεία κατάτμησης εικόνων συγκρινόμενος με άλλες τεχνικές [92], όπως *color clustering*, *pyramidal region growing* καθώς και το μορφολογικό αλγόριθμο *watershed* [111]. Συγκριτική αξιολόγηση των παραπάνω αλγορίθμων στο πλαίσιο του προγράμματος COST211ter με χρήση μεγάλου αριθμού πειραμάτων σε ένα σύνολο δοκιμαστικών ακολουθιών βίντεο κατέληξε στο συμπέρασμα ότι ο RSST παρουσιάζει καλύτερες επιδόσεις [1]. Συγκεκριμένα, δίνει πάντα σωστές ακμές αντικειμένων και όσο το δυνατόν λιγότερα αυθαίρετα αντικείμενα. Ως προς το υπολογιστικό κόστος, ο RSST είναι επίσης ο ταχύτερος αλγόριθμος.

Το διάγραμμα ροής του αλγορίθμου φαίνεται στο Σχήμα 1-1 και η λειτουργία του περιγράφεται συνοπτικά στη συνέχεια. Κατά την *αρχικοποίηση* του αλγορίθμου η εικόνα εισόδου I , διαστάσεων $M_0 \times N_0$ pixels, χωρίζεται σε $M_0 \times N_0$ περιοχές (regions) ή *τμήματα* (segments) μεγέθους 1 pixel και για κάθε περιοχή παράγονται *σύνδεσμοι* (links) με τις τέσσερις γειτονικές της περιοχές. Γειτονικές μιας περιοχής X θεωρούνται οι περιοχές εκείνες που απέχουν 1 pixel από τη X και συνδέονται με τη X οριζόντια ή κατακόρυφα, αλλά όχι διαγώνια (4-connected). Σε κάθε σύνδεσμο ανατίθεται ένα *βάρος* (weight) ίσο με τη *απόσταση* (distance) των δύο περιοχών τις οποίες συνδέει. Η απόσταση αυτή μπορεί να ορισθεί με οποιοδήποτε τρόπο και καθορίζει ουσιαστικά την προτεραιότητα με την οποία οι περιοχές ενώνονται μεταξύ τους. Για παρά-

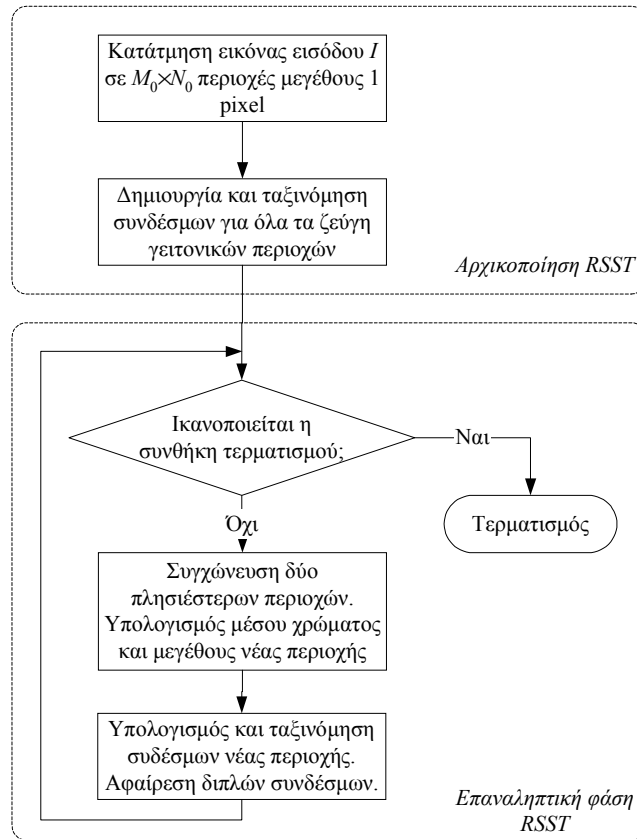
δειγμα, για τους σκοπούς της κατάτμησης με βάση το χρώμα η απόσταση $\delta_c(X,Y)$ μεταξύ δύο γειτονικών περιοχών X και Y συνήθως ορίζεται ως το γινόμενο της *Ευκλείδειας* απόστασης των μέσων χρωματικών συνιστωσών των δύο περιοχών και ενός όρου που είναι συνάρτηση του μεγέθους των δύο περιοχών σε pixels και ευνοεί τη συγχώνευση των μικρών περιοχών:

$$\delta_c(X,Y) = \|\mathbf{c}(X) - \mathbf{c}(Y)\| \frac{a(X)a(Y)}{a(X) + a(Y)} \quad (1-1)$$

όπου $\mathbf{c}(X)$ είναι ένα διάνυσμα διαστάσεων 3×1 που περιέχει τις τρεις μέσες χρωματικές συνιστώσες της περιοχής X και $a(X)$ είναι το μέγεθος της σε pixels. Χρησιμοποιώντας για παράδειγμα το χώρο *RGB* για την αναπαράσταση των χρωματικών συνιστωσών, η απόσταση $\delta_c(X,Y)$ γίνεται

$$\delta_c(X,Y) = \left[(c_R(X) - c_R(Y))^2 + (c_G(X) - c_G(Y))^2 + (c_B(X) - c_B(Y))^2 \right]^{1/2} \frac{a(X)a(Y)}{a(X) + a(Y)} \quad (1-2)$$

όπου $c_R(X)$, $c_G(X)$ και $c_B(X)$ είναι οι μέσες χρωματικές συνιστώσες *R*, *G* και *B* αντίστοιχα στο εσωτερικό της περιοχής X . Στη συνέχεια όλοι οι σύνδεσμοι ταξινομούνται κατά αύξουσα σειρά απόστασης, έτσι ώστε ο πρώτος σύνδεσμος με τη μικρότερη απόσταση να αντιστοιχεί στις πιο «κοντινές» περιοχές.



Σχήμα 1-1. Διάγραμμα ροής του αλγορίθμου RSST.

Κατά την επαναληπτική φάση του RSST πραγματοποιείται αναδρομική συγχώνευση (merging) επιλέγοντας κάθε φορά τις δύο γειτονικές περιοχές με τη μικρότερη απόσταση και εφαρμόζοντας τις ακόλουθες ενέργειες σε κάθε επανάληψη:

- (α) συγχωνεύονται οι δύο πλησιέστερες γειτονικές περιοχές

- (β) υπολογίζονται οι μέσες χρωματικές συνιστώσες και το μέγεθος σε pixels της νέας περιοχής που προήλθε από τη συγχώνευση
- (γ) επαναυπολογίζονται και ταξινομούνται τα βάρη των συνδέσμων της νέας περιοχής με όλες τις γειτονικές της
- (δ) αφαιρούνται οι τυχόν διπλοί σύνδεσμοι που προκύπτουν από τη συγχώνευση

Η επαναληπτική διαδικασία τερματίζεται όταν είτε ο συνολικός αριθμός περιοχών είτε η ελάχιστη απόσταση γίνει ίση με ένα προκαθορισμένο όριο ή *κατώφλι* (threshold). Το κατώφλι απόστασης είναι συνήθως προτιμότερο διότι το πλήθος τμημάτων που προκύπτει από την κατάρτηση εξαρτάται από τα αντικείμενα που υπάρχουν στην εικόνα.

Η υπολογιστική πολυπλοκότητα του RSST καθορίζεται σε μεγάλο βαθμό από την επιλογή του αλγορίθμου ταξινόμησης των βαρών, ο οποίος αποτελεί παράγοντα συμφόρησης. Για το λόγο αυτό προτείνεται στο πλαίσιο της διατριβής μία νέα *πολυδιακριτική υλοποίηση* (multiresolution implementation) του αλγορίθμου, που ονομάζεται Multiresolution RSST (M-RSST) και αναλύεται στη συνέχεια.

Ο Αλγόριθμος M-RSST

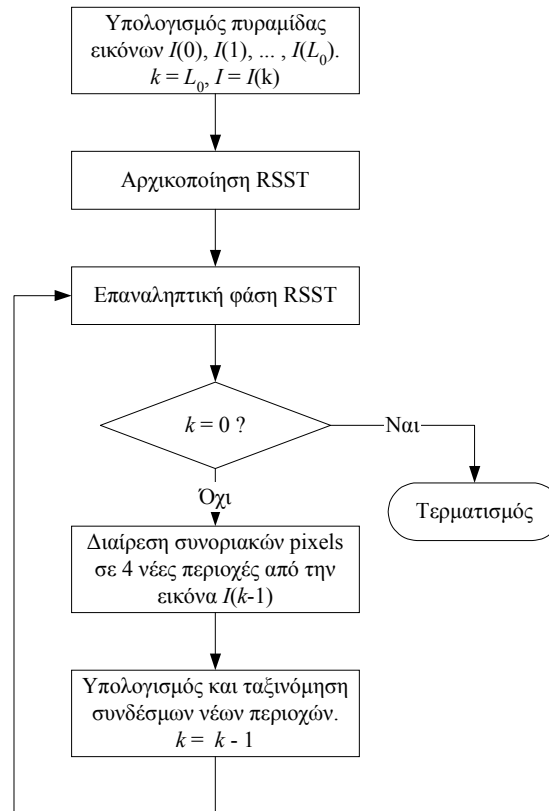
Ο αλγόριθμος M-RSST εφαρμόζει αναδρομικά τον αλγόριθμο RSST σε εικόνες αυξανόμενης ανάλυσης, όπως φαίνεται στο διάγραμμα ροής του Σχήματος 1-2. Αρχικά πραγματοποιείται ανάλυση της εικόνας εισόδου I σε πολλαπλές αναλύσεις (multiresolution decomposition) με χαμηλότερο επίπεδο ανάλυσης L_0 και κατασκευάζεται μία ιεραρχία εικόνων $I(0) = I, I(1), \dots, I(L_0)$. Έτσι δημιουργείται μία *κόλυρος πυραμίδα εικόνων* (truncated image pyramid) κάθε επίπεδο της οποίας έχει το ένα τέταρτο των pixels του αμέσως χαμηλότερου επιπέδου. Η αρχικοποίηση του αλγορίθμου RSST εφαρμόζεται στην εικόνα χαμηλότερης ανάλυσης $I(L_0)$ και στη συνέχεια ξεκινά μία επαναληπτική διαδικασία που περιλαμβάνει τα παρακάτω βήματα:

- (α) οι περιοχές της τρέχουσας εικόνας συγχωνεύονται αναδρομικά χρησιμοποιώντας την επαναληπτική φάση του αλγορίθμου RSST
- (β) τα συνοριακά pixels των περιοχών που προκύπτουν διαιρούνται σε τέσσερις νέες περιοχές το καθένα· οι χρωματικές συνιστώσες των νέων περιοχών προέρχονται από την εικόνα του αμέσως επόμενου υψηλότερου επιπέδου ανάλυσης
- (γ) υπολογίζονται και ταξινομούνται τα βάρη των νέων συνδέσμων των νέων περιοχών

Η παραπάνω διαδικασία συγχώνευσης / διαίρεσης (splitting) επαναλαμβάνεται μέχρι και την εικόνα υψηλότερης ανάλυσης $I(0)$. Η συγχώνευση των περιοχών της τελευταίας που προκύπτει από το βήμα (α) δίνει και το τελικό αποτέλεσμα κατάρτησης του αλγορίθμου M-RSST.

Όπως φαίνεται από την παραπάνω περιγραφή, η πραγματική συγχώνευση τμημάτων πραγματοποιείται στο χαμηλότερο επίπεδο ανάλυσης $I(L_0)$, όπου και ο συνολικός αριθμός τμημάτων είναι ο μικρότερος. Στα υπόλοιπα επίπεδα ανάλυσης ουσιαστικά προσδιορίζονται οι συνοριακές περιοχές κάθε τμήματος με μεγαλύτερη ακρίβεια, χωρίς να επηρεάζεται το εσωτερικό του. Επειδή σε κάθε συγχώνευση τμημάτων πραγματοποιείται ταξινόμηση των νέων βαρών συνδέσμων που προκύπτουν, η ταχύτητα καθορίζεται σε μεγάλο βαθμό από το πλήθος τμημάτων· έτσι είναι αναμενόμενο ο M-RSST να είναι σημαντικά ταχύτερος από τον RSST. Επιπλέον, λόγω της υποδειγματοληψίας, είναι φυσικό το αποτέλεσμα του M-RSST να μην

ταυτίζεται με εκείνο του RSST. Τα ζητήματα αυτά αναλύονται στα αποτελέσματα της επόμενης Ενότητας όπου εξετάζεται η κατάτμηση με βάση το χρώμα.



Σχήμα 1-2. Διάγραμμα ροής του αλγορίθμου M-RSST. Οι επιμέρους εργασίες «Αρχικοποίηση RSST» και «Επαναληπτική Φάση RSST» αναφέρονται στο Σχήμα 1-1.

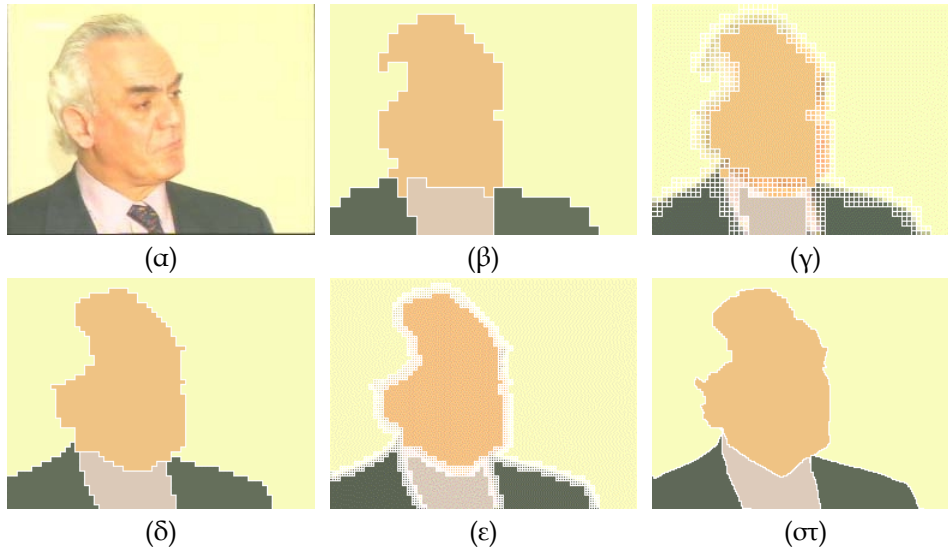
Σημειώνεται ότι ο αλγόριθμος M-RSST παρουσιάστηκε για πρώτη φορά στην εργασία [35], ενώ η υλοποίησή του αναλύθηκε στην [8]. Χρησιμοποιήθηκε επίσης με επιτυχία για κατάτμηση και στις εργασίες [9,15,34,38,39,131].

1.3 ΚΑΤΑΤΜΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΧΡΩΜΑ

Τα αποτελέσματα του προτεινόμενου αλγορίθμου M-RSST φαίνονται στο Σχήμα 1-3 για κατώφλι περιοχών ίσο με 5 και για αρχικό επίπεδο ανάλυσης $L_0 = 3$ (ισοδύναμο με blocks των 8×8 pixels). Μετά την εφαρμογή της επαναληπτικής φάσης RSST στην εικόνα $I(3)$ (Σχήμα 1-3β), τα συνοριακά pixels διαιρούνται σε τέσσερα νέα τμήματα το καθένα. Όπως φαίνεται στο Σχήμα 1-3γ, το πλήθος τμημάτων για την επόμενη επανάληψη RSST στο επίπεδο ανάλυσης 2 είναι σημαντικά μειωμένο σε σύγκριση με το αρχικό πλήθος τμημάτων του παραδοσιακού αλγορίθμου RSST στο ίδιο επίπεδο ανάλυσης. Το ίδιο ισχύει και για τα επόμενα, υψηλότερα επίπεδα ανάλυσης, όπως φαίνεται στο Σχήμα 1-3ε.

Αφού η ταχύτητα του RSST καθορίζεται κυρίως από το αρχικό πλήθος τμημάτων, είναι εμφανές ότι η προτεινόμενη υλοποίηση M-RSST είναι σημαντικά ταχύτερη από τον RSST. Η πολυπλοκότητα του M-RSST δεν μπορεί να υπολογισθεί με ευθύ τρόπο γιατί εξαρτάται το πλήθος, το σχήμα και το μέγεθος των περιοχών. Γι' αυτό το λόγο, ο μέσος χρόνος εκτέλεσης των δύο αλγορίθμων δίνεται συγκριτικά στον Πίνακα 1-1 για ένα σύνολο 200 έγχρωμων εικό-

νων σε τρεις διαφορετικές αναλύσεις. Σε όλες τις περιπτώσεις το αρχικό επίπεδο ανάλυσης είναι $L_0 = 3$, ενώ χρησιμοποιείται κατώφλι απόστασης για τον τερματισμό της διαδικασίας κατάρτησης, διότι δίνει καλύτερα αποτελέσματα και δεν χρειάζεται προσαρμογή για κάθε εικόνα. Οι χρόνοι εκτέλεσης αντιστοιχούν σε υλοποίηση των αλγορίθμων σε γλώσσα C, σε υπολογιστή Sun SparcStation-20. Παρατηρείται ότι ο λόγος βελτίωσης επηρεάζεται σημαντικά από τις διαστάσεις της εικόνας εισόδου, ενώ ο M-RSST είναι περίπου 400 φορές ταχύτερος από τον RSST για ένα τυπικό μέγεθος εικόνας 720×576 pixels.



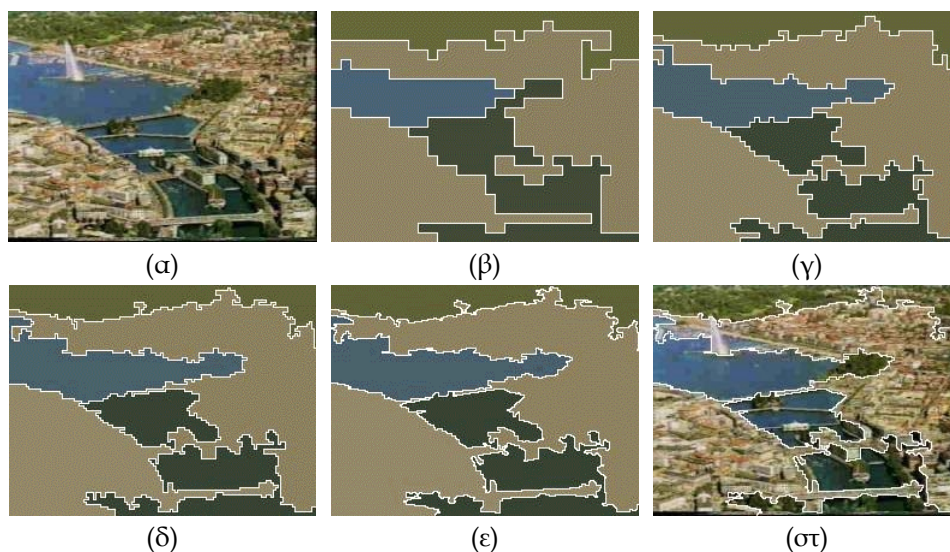
Σχήμα 1-3. Κατάτμηση με βάση το χρώμα, αλγόριθμος M-RSST. (α) Αρχική εικόνα. (β) Κατάτμηση στο επίπεδο ανάλυσης 3. (γ) Διαίρεση συνοριακών pixels στο επίπεδο 3. (δ) Κατάτμηση στο επίπεδο 2. (ε) Διαίρεση συνοριακών pixels στο επίπεδο 2. (στ) Τελική κατάτμηση.

Ανάλυση εικόνας	Μέσος χρόνος εκτέλεσης (sec)		Λόγος βελτίωσης
	RSST	M-RSST	
176×144 (QCIF)	5.65	0.13	43.46
352×288 (CIF)	44.21	0.38	116.35
720×576 (PAL)	534.22	1.36	392.81

Πίνακας 1-1. Μέσος χρόνος εκτέλεσης των αλγορίθμων RSST και M-RSST για ένα σύνολο 200 έγχρωμων εικόνων σε τρεις διαφορετικές αναλύσεις.

Επίσης, όπως παρατηρείται στο Σχήμα 1-3β, τα πολύ μικρά αντικείμενα αγνοούνται στην αρχική (χαμηλότερη) ανάλυση, και αφού δεν δημιουργούνται ούτε καταστρέφονται περιοχές σε κάθε επανάληψη, τα αντικείμενα αυτά εξαλείφονται σε όλα τα επίπεδα ανάλυσης (Σχήματα 1-3δ, στ). Για παράδειγμα, ακόμη και αν το κατώφλι περιοχών ήταν μεγαλύτερο από 5, οι λεπτομέρειες του προσώπου δεν θα δημιουργούσαν χωριστά τμήματα. Επομένως το αποτέλεσμα του M-RSST είναι διαφορετικό από εκείνο του παραδοσιακού RSST, όσον αφορά στα μικρά αντικείμενα. Αυτό το φιλτράρισμα ως προς το μέγεθος των αντικειμένων είναι πολλές φορές επιθυμητό καθώς έτσι αγνοούνται οι λεπτομέρειες και επιτυγχάνεται υψηλό επίπεδο αναπαράστασης περιεχομένου. Ένα ακόμη παράδειγμα φαίνεται στο Σχήμα 1-4, όπου η εικόνα ενός τοπίου διαχωρίζεται σωστά σε περιοχές κτιρίων, θάλασσα καθώς και δασικές περιοχές. Έτσι ο M-RSST αποφεύγει το φαινόμενο του *oversegmentation* (κατάτμηση σε υπερβολικά μεγάλο αριθμό περιοχών), και χειρίζεται σωστά εικόνες που περιέχουν πολύπλοκη υφή

(texture). Περισσότερα παραδείγματα κατάτμησης με βάση το χρώμα περιλαμβάνονται στα Κεφάλαια 3,11,12 και 13.



Σχήμα 1-4. Κατάτμηση με βάση το χρώμα, αλγόριθμος M-RSST. (α) Αρχική εικόνα. (β) Κατάτμηση στο επίπεδο ανάλυσης 3. (γ) Κατάτμηση στο επίπεδο 2. (δ) Κατάτμηση στο επίπεδο 1. (ε) Κατάτμηση στο επίπεδο 0 (τελική). (στ) Τελική κατάτμηση σε συνδυασμό με την αρχική εικόνα.

Τέλος, φαίνεται στα Σχήματα 1-3β,δ και 1-4β,γ,δ ότι ουσιαστικά μόνο τα περιγράμματα των τμημάτων αλλάζουν σε κάθε επανάληψη του αλγορίθμου, αφού ούτε δημιουργούνται νέα τμήματα ούτε καταστρέφονται τα υπάρχοντα. Είναι επομένως δυνατό να βρει κανείς το ακριβές σχήμα των αντικειμένων που περιέχονται σε μία εικόνα από το αποτέλεσμα της κατάτμησης στο υψηλότερο επίπεδο ανάλυσης όπως στα Σχήματα 1-3στ και 1-4ε. Επιπλέον, αυτό επιτυγχάνεται χωρίς τη χρήση ολόκληρων των εικόνων σε όλα τα επίπεδα ανάλυσης, αλλά μόνο τμημάτων τους στις συνοριακές περιοχές των αντικειμένων. Έτσι είναι δυνατή η αποτελεσματικότερη κατάτμηση ακολουθιών MPEG, χρησιμοποιώντας τα *blocks* κάθε εικόνας για την αρχικοποίηση του M-RSST. Στην περίπτωση αυτή δεν γίνεται πλήρης πολυδιακριτική ανάλυση κάθε εικόνας, αλλά πραγματοποιείται αποκωδικοποίηση ενός πολύ μικρού ποσοστού από *blocks* στις συνοριακές περιοχές των αντικειμένων, έχοντας ως αποτέλεσμα την πολύ γρήγορη υλοποίηση της κατάτμησης.

1.4 ΚΑΤΑΤΜΗΣΗ ΜΕ ΒΑΣΗ ΤΗΝ ΚΙΝΗΣΗ

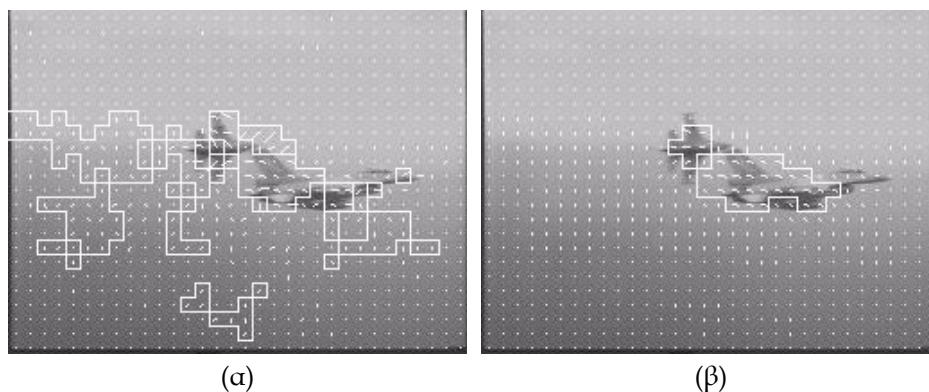
Για το πρόβλημα της *εκτίμησης κίνησης* (motion estimation) και κατάτμησης ακολουθιών εικόνων με βάση την κίνηση έχουν προταθεί αρκετές διαφορετικές τεχνικές, όπως τεχνικές που βασίζονται στη *διαφορά φωτεινότητας* μεταξύ διαδοχικών καρέ, τεχνικές εκτίμησης *οπτικής ροής* (optical flow), καθώς και τεχνικές ταυτόχρονης εκτίμησης και κατάτμησης κίνησης [127]. Κάθε μέθοδος έχει πλεονεκτήματα ή μειονεκτήματα που περιορίζουν τη χρήση της σε συγκεκριμένες εφαρμογές. Για παράδειγμα, οι τεχνικές ταυτόχρονης εκτίμησης και κατάτμησης κίνησης έχουν υψηλή υπολογιστική πολυπλοκότητα, ενώ οι τεχνικές που βασίζονται στη διαφορά φωτεινότητας μεταξύ διαδοχικών καρέ επηρεάζονται σημαντικά από το θόρυβο και τις μεταβολές φωτεινότητας. Οι τεχνικές εκτίμησης οπτικής ροής είναι αρκετά δημοφιλείς και χρησιμοποιούνται ευρύτατα στην κωδικοποίηση βίντεο αλλά και στην ανάλυση και κατανόηση εικόνων.

Λαμβάνοντας υπόψη τον υπολογιστικό φόρτο που απαιτείται για την ανάλυση μεγάλου αριθμού ακολουθιών βίντεο, αλλά και τις μειωμένες απαιτήσεις ακρίβειας στο πρόβλημα της επιλογής χαρακτηριστικών καρτέ, επιλέχθηκε στο πλαίσιο της διατριβής η εκμετάλλευση των διανυσμάτων κίνησης που είναι διαθέσιμα στις ακολουθίες MPEG. Έτσι δεν υπάρχει ανάγκη για εκτίμηση κίνησης και επιτογχάνεται ταχύτατη υλοποίηση. Όμως τα διανύσματα κίνησης των ακολουθιών MPEG υπολογίζονται με βάση την αποτελεσματικότερη κωδικοποίηση και δεν δίνουν ικανοποιητική εκτίμηση της πραγματικής κίνησης αφού περιέχουν μεγάλο ποσοστό θορύβου. Είναι λοιπόν απαραίτητο ένα επιπλέον βήμα επεξεργασίας για την εξομάλυνση των διανυσμάτων αυτών. Για το σκοπό αυτό επιλέγεται ένα φίλτρο median, λόγω της ικανότητάς του να διατηρεί τα περιγράμματα των αντικειμένων. Η κατάτμηση με βάση την κίνηση πραγματοποιείται διαιρώντας κάθε καρτέ σε περιοχές ομοιόμορφης κίνησης. Χρησιμοποιείται και πάλι ο αλγόριθμος M-RSST με αρχική ανάλυση όμοια με την ανάλυση των blocks της ακολουθίας MPEG, ενώ για τον υπολογισμό αποστάσεων μεταξύ περιοχών χρησιμοποιούνται διαφορές διανυσμάτων κίνησης αντί για χρωματικές διαφορές.

Για τους σκοπούς της κατάτμησης με βάση την κίνηση η απόσταση $\delta_m(X,Y)$ μεταξύ δύο γειτονικών περιοχών X και Y συνήθως ορίζεται, όμοια με την (1-1), ως το γινόμενο της Ευκλείδειας απόστασης των μέσων διανυσμάτων κίνησης των δύο περιοχών και ενός όρου που είναι συνάρτηση του μεγέθους των δύο περιοχών σε pixels και ευνοεί τη συγχώνευση των μικρών περιοχών:

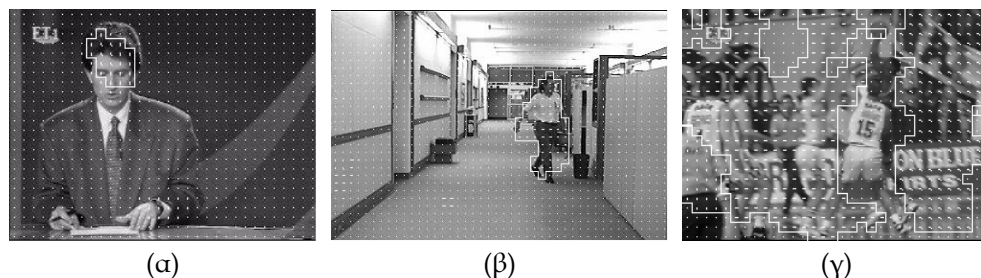
$$\delta_m(X,Y) = \|\mathbf{v}(X) - \mathbf{v}(Y)\| \frac{a(X)a(Y)}{a(X) + a(Y)} \quad (1-3)$$

όπου $\mathbf{v}(X)$ είναι το μέσο διάνυσμα κίνησης της περιοχής X , διαστάσεων 2×1 , και $a(X)$ είναι το μέγεθός της σε pixels. Στα πειράματα κατάτμησης με βάση την κίνηση χρησιμοποιείται κατώφλι απόστασης για τον τερματισμό της διαδικασίας κατάτμησης, όπως και στην περίπτωση της κατάτμησης με βάση το χρώμα.



Σχήμα 1-5. Κατάτμηση με βάση την κίνηση, αλγόριθμος M-RSST. (α) Κατάτμηση χωρίς εξομάλυνση διανυσμάτων κίνησης. (β) Κατάτμηση με εξομάλυνση.

Στο Σχήμα 1-5 φαίνονται τα αποτελέσματα της κατάτμησης με βάση την κίνηση για ένα καρτέ από ακολουθία τηλεοπτικών ειδήσεων. Είναι εμφανές από το Σχήμα 1-5α ότι χωρίς εξομάλυνση των διανυσμάτων κίνησης η κατάτμηση είναι λανθασμένη, ακόμη και σε περιοχές ομοιόμορφης φωτεινότητας χωρίς πραγματική κίνηση. Αντίθετα μετά την εξομάλυνση η κατάτμηση δίνει σωστά τα δύο αντικείμενα που κινούνται πραγματικά, όπως φαίνεται στο Σχήμα 1-5β.



Σχήμα 1-6. Παραδείγματα κατάτμησης με βάση την κίνηση, αλγόριθμος M-RSST. (α) Ομοιόμορφη κίνηση μικρού αντικείμενου (πρόσωπο παρουσιαστή) σε 'στατικό φόντο. (β) Σύνθετη κίνηση μικρού αντικείμενου σε 'στατικό φόντο. (γ) Συνδυασμός ακίνητων αντικειμένων σε κινούμενο φόντο (οριζόντια κίνηση κάμερας).

Μερικά ακόμη αποτελέσματα κατάτμησης με βάση την κίνηση παρουσιάζονται στο Σχήμα 1-6. Σε όλα τα παραδείγματα έχει πραγματοποιηθεί εξομάλυνση των διανυσμάτων κίνησης. Συγκεκριμένα, στο Σχήμα 1-6α φαίνεται το πρόσωπο ενός παρουσιαστή που κινείται κατακόρυφα· το αντικείμενο του προσώπου κινείται ομοιόμορφα ενώ το φόντο της εικόνας είναι ακίνητο. Στο Σχήμα 1-6β φαίνεται, και πάλι σε στατικό φόντο, ένας άνθρωπος που περπατά. Το τμήμα του ανθρώπου παρουσιάζει σύνθετη κίνηση αλλά και πάλι ανιχνεύεται ως ένα μόνο αντικείμενο. Τέλος στο 1-6γ εμφανίζεται μια περίπτωση γρήγορης οριζόντιας κίνησης της κάμερας. Η κάμερα ακολουθεί την κίνηση ενός ανθρώπου, ο οποίος παρουσιάζεται ακίνητος στην εικόνα, ενώ αντίθετα το φόντο φαίνεται να κινείται οριζόντια. Στην περίπτωση αυτή παρουσιάζονται κάποια προβλήματα σε κινούμενες περιοχές της εικόνας, κυρίως στα άκρα της, όπου λόγω ομοιόμορφου χρώματος η κίνηση δεν ανιχνεύεται σωστά και θεωρείται μηδενική. Όμως τα γενικά χαρακτηριστικά κίνησης της εικόνας εξαγονται σωστά σε όλες τις περιπτώσεις.

Κατάτμηση με Βάση το Πεδίο Βάθους

Η χρήση στερεοσκοπικών και γενικότερα τρισδιάστατων (3-Δ) ακολουθιών βίντεο έχει αυξηθεί σημαντικά τα τελευταία χρόνια, λόγω της ανάπτυξης στερεοσκοπικών συστημάτων κάμερας και συσκευών τρισδιάστατης απεικόνισης. Η ανάλυση 3-Δ ακολουθιών βίντεο επιτρέπει τον αποδοτικότερο χειρισμό αντικειμένων μέσω της ανίχνευσης και εκμετάλλευσης της πληροφορίας βάθους, αφού τα αντικείμενα συνήθως αποτελούνται από περιοχές που βρίσκονται στο ίδιο επίπεδο βάθους. Στο Κεφάλαιο αυτό παρουσιάζεται μια σύντομη ανασκόπηση της στερεοσκοπικής ανάλυσης εικόνων και περιγράφεται η προτεινόμενη μέθοδος εκτίμησης του πεδίου βάθους από ένα ζεύγος στερεοσκοπικών εικόνων. Η μέθοδος βασίζεται στον υπολογισμό του πεδίου απόκλισης με μία τεχνική παρόμοια με την ανίχνευση κίνησης σε ακολουθίες βίντεο. Στη συνέχεια πραγματοποιείται ανίχνευση και αντιστάθμιση των επικαλύψεων μεταξύ γειτονικών αντικειμένων, ενώ για την κατάτμηση με βάση το πεδίο βάθους εφαρμόζεται ο αλγόριθμος M-RSST. Τα αποτελέσματα της κατάτμησης συνδυάζονται στο επόμενο Κεφάλαιο με την κατάτμηση χρώματος ή και κίνησης για τον εντοπισμό αντικειμένων με σημασιολογικό περιεχόμενο.

2.1 ΣΤΕΡΕΟΣΚΟΠΙΚΕΣ ΑΚΟΛΟΥΘΙΕΣ ΒΙΝΤΕΟ

Παρά το γεγονός ότι η συντριπτική πλειοψηφία των αρχείων βίντεο αποτελείται από διδιάστατες (2-Δ) ακολουθίες, η χρήση τρισδιάστατων (3-Δ) ή στερεοσκοπικών ακολουθιών βίντεο, οι οποίες συνήθως προέρχονται από στερεοσκοπικά συστήματα κάμερας, έχει αυξηθεί σημαντικά αφού οι ακολουθίες αυτές παρέχουν καλύτερη αναπαράσταση του οπτικού περιεχομένου και βελτιώνουν την επικοινωνία πολυμέσων. Οι τρισδιάστατες ακολουθίες επιτρέπουν τη αποδοτικότερη διαχείριση των αντικειμένων με την εκμετάλλευση της πληροφορίας βάθους που προκύπτει από τη στερεοσκοπική ανάλυση. Επιπλέον, το πρόβλημα της κατάτμησης σε αντικείμενα αντιμετωπίζεται με μεγαλύτερη ακρίβεια, αφού τα αντικείμενα αποτελούνται συνήθως από περιοχές εικόνων που ανήκουν στο ίδιο επίπεδο βάθους [44].

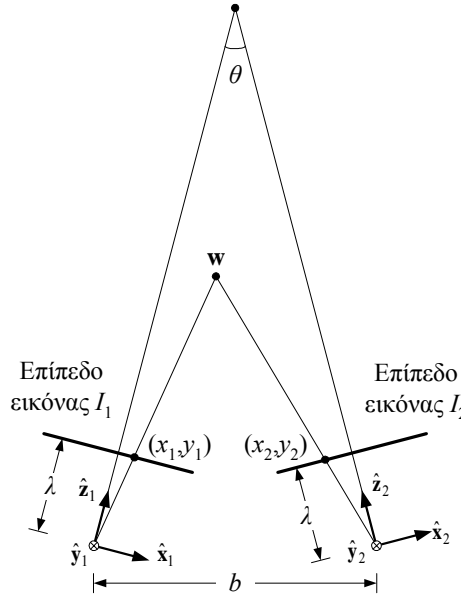
Ένας μεγάλος αριθμός εφαρμογών, όπως για παράδειγμα η αυτόνομη επιτήρηση (surveillance), η δεικτοδότηση, η ανάκληση και η διαχείριση οπτικοακουστικού υλικού, μπορούν να ωφεληθούν από την τρισδιάστατη αναπαράσταση. Γι' αυτό το λόγο συσκευές σύλληψης 3-Δ δεδομένων, στερεοσκοπικά συστήματα κάμερας και συστήματα τρισδιάστατης απεικόνισης έχουν παρουσιάσει σημαντική ανάπτυξη και διάδοση, και συνεπώς τα αρχεία τρισδιάστατου βίντεο αναμένεται να αυξηθούν με ταχείς ρυθμούς τα επόμενα χρόνια.

Η επεξεργασία και ανάλυση στερεοσκοπικών ακολουθιών που ακολουθείται στο πλαίσιο της διατριβής περιλαμβάνει τρία στάδια: (α) εκτίμηση πεδίου απόκλισης μέσω ταιριάσματος ενός στερεοσκοπικού ζεύγους εικόνων, (β) εκτίμηση πεδίου βάθους από το πεδίο απόκλισης μέσω στερεοσκοπικής ανάλυσης, και (γ) ανίχνευση και αντιστάθμιση επικάλυψης. Η στερεοσκοπική ανάλυση εικόνων για τη συσχέτιση του βάθους με την απόκλιση είναι γνωστή τεχνι-

κή [49], ενώ ο χρησιμοποιούμενος αλγόριθμος για των εκτίμηση πεδίου απόκλισης και βάθους αποτελεί παραλλαγή του αντίστοιχου αλγορίθμου των [135,136]. Η μέθοδος ανίχνευσης και αντιστάθμισης επικάλυψης βασίζεται σε μεγάλο βαθμό στην [50]. Όλες οι παραπάνω τεχνικές που εφαρμόζονται σε στερεοσκοπικές ακολουθίες προτάθηκαν στην εργασία [39] αλλά χρησιμοποιήθηκαν και στην [38].

2.2 ΣΤΕΡΕΟΣΚΟΠΙΚΗ ΑΝΑΛΥΣΗ

Όταν μία τρισδιάστατη σκηνή προβάλλεται στο επίπεδο μιας εικόνας μέσω μιας κάμερας, η πληροφορία του βάθους χάνεται, αφού η απεικόνιση αυτή είναι ένας μετασχηματισμός από τρισδιάστατο χώρο σε διδιάστατο. Συγκεκριμένα, μέσω της προοπτικής προβολής, κάθε 2-Δ σημείο της εικόνας αντιστοιχεί στο σύνολο όλων των συγγραμικών σημείων του 3-Δ χώρου που ανήκουν στη ευθεία η οποία διέρχεται από το σημείο αυτό και από το εστιακό σημείο της κάμερας [49]. Αν δεν είναι διαθέσιμη κάποια επιπλέον πληροφορία σχετικά με το πραγματικό 3-Δ σημείο που αντιστοιχεί στο σημείο της εικόνας, δεν είναι δυνατή η υλοποίηση ενός αντίστροφου μετασχηματισμού για την πλήρη ανάκτηση του 3-Δ σημείου.



Σχήμα 2-1. Γεωμετρία στερεοσκοπικού συστήματος κάμερας με συγκλίνοντες οπτικούς άξονες και προοπτική προβολή των 3-Δ σημείων στα επίπεδα των εικόνων.

Στις ακολουθίες εικόνων όμως, είναι δυνατή σε κάποιο βαθμό η ανάκτηση της πληροφορίας βάθους με βάση τις επικαλύψεις μεταξύ διαφορετικών περιοχών της εικόνας που βρίσκονται σε σχετική κίνηση μεταξύ τους [44]. Η εκτίμηση του βάθους είναι πολύ πιο αξιόπιστη με χρήση στερεοσκοπικής ανάλυσης, όταν είναι διαθέσιμες περισσότερες από μία απόψεις της ίδιας 3-Δ σκηνής από διαφορετικές κάμερες [75]. Η παρακάτω ανάλυση επικεντρώνεται στην περίπτωση του διοπτρικού (binocular) συστήματος κάμερας (όταν είναι διαθέσιμες δύο κάμερες), αλλά μπορεί εύκολα να επεκταθεί στη γενικότερη περίπτωση πολλαπλής κάμερας (multinocular ή multi-view) όπου βέβαια η εκτίμηση βάθους είναι ακόμη πιο αξιόπιστη.

Ας θεωρήσουμε ένα στερεοσκοπικό σύστημα με δύο κάμερες εστιακού μήκους (focal length) λ και απόστασης βάσης (baseline distance) b , όπως φαίνεται στο Σχήμα 2-1. Έστω ότι οι δύο οπτικοί άξονες συγκλίνουν υπό γωνία θ . Παράλληλοι οπτικοί άξονες μπορούν επίσης να χρη-

σιμποιηθούν [49], αλλά αυτό μπορεί να θεωρηθεί ειδική περίπτωση της παρακάτω ανάλυσης, θέτοντας $\theta = 0$. Τα δύο αντίστοιχα τοπικά συστήματα συντεταγμένων ορίζονται από τα μοναδιαία διανύσματα $\hat{x}_1, \hat{y}_1, \hat{z}_1$ και $\hat{x}_2, \hat{y}_2, \hat{z}_2$, ενώ οι δύο αρχές αξόνων βρίσκονται στα *εστιακά σημεία* (focal points), σε απόσταση λ από τα *επίπεδα εικόνας* (image planes) I_1 και I_2 αντίστοιχα. Χωρίς βλάβη της γενικότητας, έστω επίσης ότι το σύστημα συντεταγμένων που χρησιμοποιείται στην ανάλυση, δηλαδή το *σύστημα αναφοράς* (world coordinate system) ταυτίζεται με εκείνο της κάμερας 1 (αριστερής κάμερας). Τότε το σύστημα της κάμερας 2 (δεξιάς κάμερας) σχετίζεται με το πρώτο μέσω μιας περιστροφής κατά γωνία θ περί τον άξονα \hat{y}_1 και μιας μετατόπισης κατά b .

Για τη έκφραση των μετασχηματισμών μετατόπισης, περιστροφής και προοπτικής προβολής σε μορφή πινάκων χρησιμοποιούνται οι *ομογενείς συντεταγμένες* (homogeneous coordinates) αντί για τις Καρτεσιανές. Συγκεκριμένα, ένα 3-Δ σημείο \mathbf{w} με Καρτεσιανές συντεταγμένες (X, Y, Z) εκφράζεται σε ομογενείς συντεταγμένες ως $\mathbf{w} = [kX \ kY \ kZ \ k]^T$ όπου k είναι μία αυθαίρετη, μη μηδενική σταθερά. Η αντίστροφη μετατροπή σε Καρτεσιανές συντεταγμένες επιτυγχάνεται απλώς διαιρώντας τις τρεις πρώτες ομογενείς συντεταγμένες με την τέταρτη [49]. Με βάση τα παραπάνω, η *προοπτική προβολή* (perspective projection) του σημείου \mathbf{w} στο επίπεδο I_1 εκφράζεται ως

$$\mathbf{r}_1 = \mathbf{P} \mathbf{w} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/\lambda & 0 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} \quad (2-1)$$

όπου το διάνυσμα $\mathbf{r}_1 = [k_1X_1 \ k_1Y_1 \ k_1Z_1 \ k_1]^T$ περιέχει τις ομογενείς συντεταγμένες της προβολής του σημείου \mathbf{w} στο επίπεδο I_1 , και \mathbf{P} είναι ο *πίνακας προοπτικής προβολής* (perspective projection matrix) της κάμερας. Η παραπάνω σχέση προβολής είναι απλή διότι το σύστημα συντεταγμένων της κάμερας 1 συμπίπτει με το σύστημα αναφοράς. Με παρόμοιο τρόπο, προκύπτει ότι η προβολή $\mathbf{r}_2 = [k_2X_2 \ k_2Y_2 \ k_2Z_2 \ k_2]^T$ του σημείου \mathbf{w} στο επίπεδο I_2 εκφράζεται ως

$$\mathbf{r}_2 = \mathbf{P} \mathbf{R} \mathbf{T} \mathbf{w} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/\lambda & 0 \end{bmatrix} \begin{bmatrix} c & 0 & s & 0 \\ 0 & 1 & 0 & 0 \\ -s & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -bc' \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -bs' \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} \quad (2-2)$$

όπου οι πίνακες \mathbf{R} και \mathbf{T} είναι πίνακες περιστροφής και μετατόπισης αντίστοιχα και προκύπτουν από τη σχέση μεταξύ των δύο συστημάτων συντεταγμένων που αναφέρθηκε παραπάνω, ενώ $s = \sin\theta$, $c = \cos\theta$, $s' = \sin(\theta/2)$ και $c' = \cos(\theta/2)$. Χρησιμοποιώντας τις δύο παραπάνω σχέσεις, οι συντεταγμένες (x_1, y_1) και (x_2, y_2) της προβολής του σημείου \mathbf{w} στα επίπεδα I_1 και I_2 αντίστοιχα μπορούν να γραφούν συναρτήσει των συντεταγμένων (X, Y, Z) ως εξής: λ

$$x_1 = \lambda \frac{X}{Z}, \quad y_1 = \lambda \frac{Y}{Z} \quad (2-3)$$

$$x_2 = \lambda \frac{Xc + Zs - bc'}{-Xs + Zc + bs'}, \quad y_2 = \lambda \frac{Y}{-Xs + Zc + bs'} \quad (2-4)$$

ενώ οι συνιστώσες βάθους z_1 και z_2 είναι και οι δύο ίσες με λ και επομένως ανεξάρτητες από τις (X, Y, Z) . Επιπλέον, συνδυάζοντας τις Εξισώσεις (2-3) και (2-4) είναι δυνατό να εκφράσει κανείς το x_2 ως προς το x_1 και το y_2 ως προς το y_1 , απαλείφοντας τις συντεταγμένες X και Y :

$$x_2 = \lambda \frac{(\lambda s + x_1 c)Z - \lambda b c'}{(\lambda c - x_1 s)Z + \lambda b s'}, \quad y_2 = \lambda \frac{y_1 Z}{(\lambda c - x_1 s)Z + \lambda b s'} \quad (2-5)$$

Επομένως, αν το βάθος Z είναι γνωστό, η παραπάνω εξίσωση δίνει μία πολύ απλή σχέση ανάμεσα στις συντεταγμένες (x_1, y_1) και (x_2, y_2) ενός σημείου που προβάλλεται στις δύο κάμερες. Το αντίστροφο πρόβλημα, της εκτίμησης του βάθους από τη σχέση μεταξύ των (x_1, y_1) και (x_2, y_2) , αποτελεί αντικείμενο της επόμενης Ενότητας.

2.3 ΕΚΤΙΜΗΣΗ ΠΕΔΙΟΥ ΒΑΘΟΥΣ ΚΑΙ ΑΠΟΚΛΙΣΗΣ

Εκτίμηση Πεδίου Βάθους

Αν η αντιστοιχία μεταξύ δύο 2-Δ σημείων (x_1, y_1) και (x_2, y_2) είναι γνωστή, αν δηλαδή είναι γνωστό ότι προέρχονται και τα δύο από την προοπτική προβολή του ίδιου 3-Δ σημείου \mathbf{w} στα επίπεδα εικόνας I_1 και I_2 αντίστοιχα, τότε η εκτίμηση του βάθους Z είναι δυνατή από την Εξίσωση (2-5). Σημειώνεται μάλιστα ότι το Z είναι το βάθος στο σύστημα συντεταγμένων της κάμερας 1, αφού αυτό συμπίπτει με το σύστημα αναφοράς. Στην ειδική περίπτωση όπου $\theta = 0$ (παράλληλοι οπτικοί άξονες) η εξίσωση απλοποιείται σε $x_2 = x_1 - \lambda b/Z$, $y_2 = y_1$, και επομένως $Z = -\lambda b/(x_2 - x_1)$. Στη γενικότερη περίπτωση ($\theta \neq 0$), η εκτίμηση του Z βασίζεται στο διάνυσμα απόκλισης (disparity vector) μεταξύ των 2-Δ σημείων (x_1, y_1) και (x_2, y_2) . Συγκεκριμένα, το διάνυσμα απόκλισης $\mathbf{d}(x_1, y_1) = [d_x(x_1, y_1) \ d_y(x_1, y_1)]^T$ στη θέση (x_1, y_1) της κάμερας 1 ως προς την κάμερα 2 ορίζεται από τις σχέσεις

$$d_x = d_x(x_1, y_1) = x_2 - x_1 = \frac{[\lambda(\lambda s + x_1 c) - x_1(\lambda c - x_1 s)]Z - \lambda b(\lambda c' + x_1 s')}{(\lambda c - x_1 s)Z + \lambda b s'} \quad (2-6)$$

$$d_y = d_y(x_1, y_1) = y_2 - y_1 = \frac{[\lambda - (\lambda c - x_1 s)]y_1 Z - \lambda b s' y_1}{(\lambda c - x_1 s)Z + \lambda b s'} \quad (2-7)$$

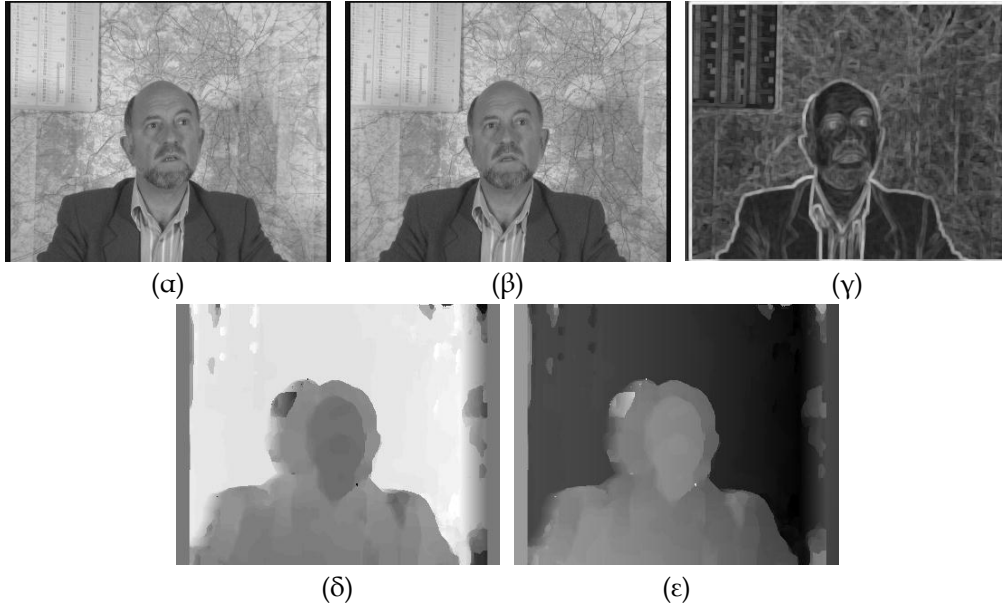
Επομένως, αν το διάνυσμα απόκλισης είναι γνωστό, οι Εξισώσεις (2-6) και (2-7) ανάγονται σε ένα γραμμικό σύστημα δύο εξισώσεων με ένα άγνωστο, το Z . Έτσι μία εκτίμηση ελαχίστων τετραγώνων είναι εφικτή με χρήση του ψευδοαντίστροφου του πίνακα διαστάσεων 2×1 που περιέχει τους συντελεστές του Z [136].

Με άλλα λόγια, αν τα διανύσματα απόκλισης είναι διαθέσιμα για όλα τα σημεία (x_1, y_1) στο επίπεδο I_1 , σχηματίζοντας ένα πεδίο απόκλισης (disparity field), τότε ένα αντίστοιχο πεδίο βάθους (depth field) μπορεί να υπολογισθεί με την παραπάνω διαδικασία. Παρά το γεγονός ότι ο υπολογισμός του βάθους από την απόκλιση είναι άμεσος και απλός, η εκτίμηση του πεδίου απόκλισης από τις εικόνες στα επίπεδα I_1 και I_2 είναι μία επίπονη διαδικασία η οποία απαιτεί την αντιστοίχιση κάθε σημείου (x_1, y_1) της εικόνας I_1 με ένα σημείο (x_2, y_2) της εικόνας I_2 , οδηγώντας σε υψηλό υπολογιστικό κόστος. Η διαδικασία αυτή είναι ανάλογη με την εκτίμηση κίνησης σε ακολουθίες εικόνων και περιγράφεται αναλυτικά παρακάτω.

Εκτίμηση Πεδίου Απόκλισης

Η εκτίμηση του πεδίου απόκλισης επιτυγχάνεται με τη χρήση ενός αλγορίθμου *block matching*, παρόμοιου με εκείνο που παρουσιάζεται στην εργασία [135]. Συγκεκριμένα, έστω ότι $I_1(x, y)$ και $I_2(x, y)$, συμβολίζουν αντίστοιχα τις τιμές των εικόνων στα επίπεδα I_1 και I_2 στη θέση $(x, y) \in F = \{1, \dots, M_0\} \times \{1, \dots, N_0\}$, όπου M_0, N_0 είναι οι διαστάσεις των δύο εικόνων. Οι εικόνες

I_1 και I_2 ονομάζονται επίσης και *αριστερό / δεξί κανάλι* (left/right channel) αντίστοιχα. Ο στόχος είναι να αντιστοιχηθεί κάθε pixel (x_1, y_1) της εικόνας I_1 με ένα pixel (x_2, y_2) της εικόνας I_2 , μέσα σε μία *περιοχή αναζήτησης* (search area) διαστάσεων $s_x \times s_y$ pixels γύρω από το σημείο (x_1, y_1) . Το διάνυσμα απόκλισης που προκύπτει από την αντιστοίχιση είναι $\mathbf{d}(x_1, y_1) = (x_2 - x_1, y_2 - y_1) \in B = \{-s_x, \dots, s_x\} \times \{-s_y, \dots, s_y\}$. Η κατακόρυφη διάσταση s_y της περιοχής αναζήτησης είναι πολύ μικρότερη από την οριζόντια, διότι η γωνία σύγκλισης θ είναι συνήθως μικρή και όλες οι αποκλίσεις είναι συνήθως κατά τον οριζόντιο άξονα. Σημειώνεται ότι στην ειδική περίπτωση των παράλληλων οπτικών αξόνων ($\theta = 0$) η περιοχή αναζήτησης ανάγεται σε ευθύγραμμο τμήμα ($s_y = 0$).



Σχήμα 2-2. Εκτίμηση πεδίου απόκλισης και βάθους για την ακολουθία *Claude*. (α) Αριστερό κανάλι. (β) Δεξί κανάλι. (γ) Συνάρτηση βάρους εξομάλυνσης. (δ) Πεδίο οριζόντιας απόκλισης. (ε) Πεδίο βάθους.

Το ταίριασμα ανάμεσα στα pixels (x_1, y_1) της I_1 με τα pixels (x_2, y_2) της I_2 , και ισοδύναμα η εκτίμηση της απόκλισης $\mathbf{d}(x_1, y_1)$, πραγματοποιείται μέσω της ακόλουθης ελαχιστοποίησης:

$$\mathbf{d}(x_1, y_1) = \arg \min_{\mathbf{u} \in B} J(\mathbf{u}, x_1, y_1), \quad \forall (x_1, y_1) \in F \quad (2-8)$$

όπου $\mathbf{u} = [u_x \ u_y]^T \in B$ είναι ένα *διάνυσμα μετατόπισης* (displacement vector) του σημείου (x_1, y_1) ως προς την εικόνα I_2 , και J είναι η ακόλουθη *συνάρτηση κόστους* (cost function):

$$J(\mathbf{u}, x_1, y_1) = D(\mathbf{u}, x_1, y_1) + S(\mathbf{u}, x_1, y_1), \quad \mathbf{u} \in B, (x_1, y_1) \in F \quad (2-9)$$

Ο πρώτος όρος, $D(\mathbf{u}, x_1, y_1)$, της συνάρτησης κόστους είναι μία *συνάρτηση σφάλματος* (block error function) που ορίζεται ως

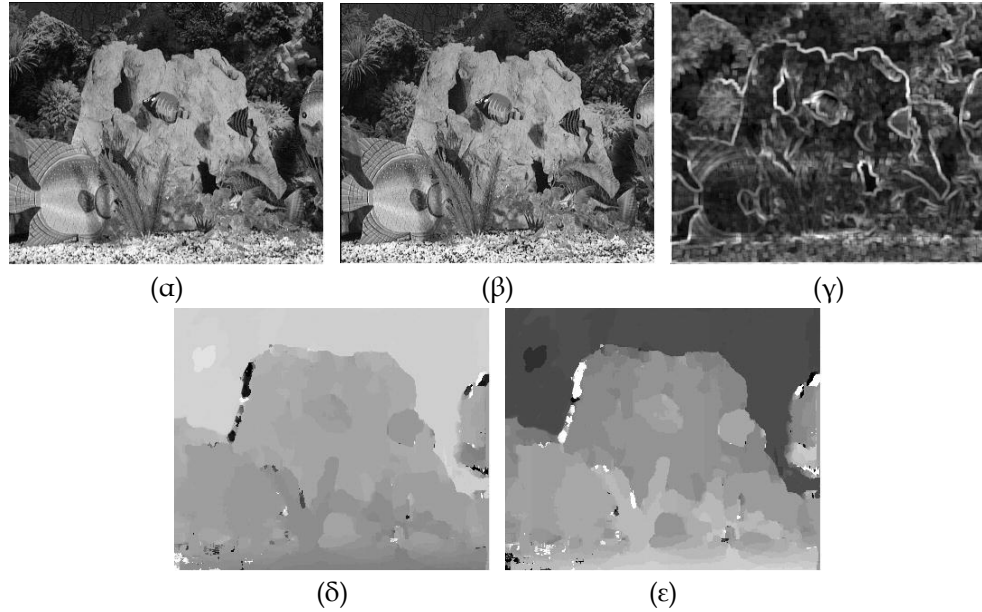
$$D(\mathbf{u}, x_1, y_1) = \sum_{x, y \in W} (I_2(x_1 + u_x + x, y_1 + u_y + y) - I_1(x_1 + x, y_1 + y))^2, \quad \mathbf{u} \in B, (x_1, y_1) \in F \quad (2-10)$$

όπου $W = \{-w, \dots, w\} \times \{-w, \dots, w\}$ είναι ένα τετράγωνο παράθυρο. Ο δεύτερος όρος, $S(\mathbf{u}, x_1, y_1)$, είναι μία *συνάρτηση εξομάλυνσης* (smoothness function) που χρησιμοποιείται για τη μείωση του θορύβου στην εκτίμηση του $\mathbf{d}(x_1, y_1)$ και ορίζεται ως

$$S(\mathbf{u}, x_1, y_1) = R(x_1, y_1) \sum_{\mathbf{v} \in N(x_1, y_1)} \|\mathbf{u} - \mathbf{v}\|^2, \quad \mathbf{u} \in B, (x_1, y_1) \in F \quad (2-11)$$

όπου $\|\cdot\|$ είναι η Ευκλείδεια νόρμα και $N(x_1, y_1) = \{\mathbf{d}(x_1 - 1, y_1), \mathbf{d}(x_1 - 1, y_1 - 1), \mathbf{d}(x_1, y_1 - 1), \mathbf{d}(x_1 + 1, y_1 - 1)\}$ είναι το σύνολο των διανυσμάτων απόκλισης όλων των pixels που είναι γειτονικά με το (x_1, y_1) . Τα διανύσματα αυτά έχουν ήδη υπολογισθεί από την Εξίσωση (2-8).

Η *συνάρτηση βάρους εξομάλυνσης* (smoothing weight function) $R(x_1, y_1)$ παίρνει χαμηλές τιμές σε περιοχές της εικόνας όπου η εκτίμηση της απόκλισης είναι αξιόπιστη, όπως οι ακμές και οι περιοχές πολύπλοκης υφής, και υψηλές τιμές σε περιοχές όπου η εκτίμηση δεν είναι αξιόπιστη, όπως περιοχές ομοιόμορφης φωτεινότητας. Η $R(x_1, y_1)$ υπολογίζεται βάσει της τοπικής διασποράς της εικόνας I_1 , και χρησιμοποιείται επίσης για την εκτίμηση του ακριβούς μεγέθους της περιοχής αναζήτησης B . Αφού σε περιοχές με υψηλή τιμή της $R(x_1, y_1)$ το πεδίο απόκλισης μεταβάλλεται ομοιόμορφα, μία μικρή περιοχή αναζήτησης είναι επαρκής. Αντίθετα, σε περιοχές με χαμηλή τιμή της $R(x_1, y_1)$ αναμένονται απότομες μεταβολές στην απόκλιση, οπότε απαιτείται μεγάλη περιοχή αναζήτησης. Συνεπάγεται ότι τα s_x και s_y μεταβάλλονται συναρτήσει της $R(x_1, y_1)$, καταλήγοντας σε γρήγορη υλοποίηση της διαδικασίας ελαχιστοποίησης.

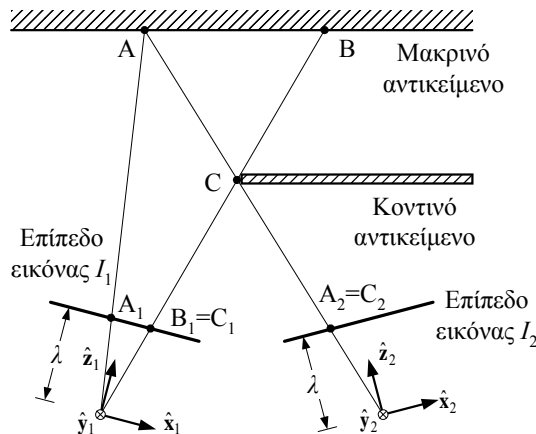


Σχήμα 2-3. Εκτίμηση πεδίου απόκλισης και βάθους για την ακολουθία Aqua. (α) Αριστερό κανάλι. (β) Δεξί κανάλι. (γ) Συνάρτηση βάρους εξομάλυνσης. (δ) Πεδίο οριζόντιας απόκλισης. (ε) Πεδίο βάθους.

Αποτελέσματα εκτίμησης του πεδίου απόκλισης και βάθους παρουσιάζονται στο Σχήμα 2-2 για την ακολουθία Claude. Συγκεκριμένα, στα Σχήματα 2-2α,β φαίνονται το αριστερό και δεξί κανάλι του καρέ #2 της ακολουθίας αντίστοιχα, ενώ στο Σχήμα 2-2γ φαίνεται η συνάρτηση βάρους εξομάλυνσης. Η κατακόρυφη απόκλιση είναι αμελητέα για την ακολουθία αυτή, αφού οι δύο κάμερες βρίσκονται στο ίδιο κατακόρυφο επίπεδο και η γωνία σύγκλισης θ είναι μικρή. Έτσι υπολογίζεται μόνο το πεδίο οριζόντιας απόκλισης $d_x(x_1, y_1)$, το οποίο και παρουσιάζεται στο Σχήμα 2-2δ, όπου οι λευκές περιοχές αντιστοιχούν σε θετική απόκλιση και οι γκριζες περιοχές σε σχεδόν μηδενική απόκλιση. Τέλος, το πεδίο βάθους εμφανίζεται στο Σχήμα 2-2ε, όπου οι μαύρες περιοχές αντιστοιχούν στο φόντο (μεγάλο βάθος) και οι γκριζες περιοχές στο προσκήνιο (μικρό βάθος). Σημειώνεται ότι και στα δύο πεδία οι σκιασμένες περιοχές βαθμιαίας μεταβολής έντασης (στην αριστερή πλευρά του προσώπου και στο δεξί όριο της εικόνας) οφείλονται στο φαινόμενο της επικάλυψης, το οποίο εξετάζεται στη συνέχεια. Αντίστοιχα αποτελέσματα για το καρέ #1 της ακολουθίας Aqua παρουσιάζονται στο Σχήμα 2-3.

2.4 ΑΝΙΧΝΕΥΣΗ ΕΠΙΚΑΛΥΨΗΣ

Η παραπάνω ανάλυση εκτίμησης της απόκλισης προϋποθέτει ότι σε κάθε σημείο της εικόνας I_1 αντιστοιχεί ένα σημείο της εικόνας I_2 . Όμως, επειδή κάθε κάμερα δίνει διαφορετική άποψη της 3-Δ σκηνής, μπορεί να υπάρχουν περιοχές της εικόνας I_1 που δεν εμφανίζονται στην I_2 γιατί επικαλύπτονται από άλλες περιοχές. Αυτή η περίπτωση ονομάζεται *επικάλυψη* (occlusion) και απεικονίζεται στο Σχήμα 2-4, όπου ένα κοντινό και ένα μακρινό αντικείμενο προβάλλονται στις δύο κάμερες.



Σχήμα 2-4. Επικάλυψη μακρινού αντικειμένου από κοντινότερο στην εικόνα I_2 .

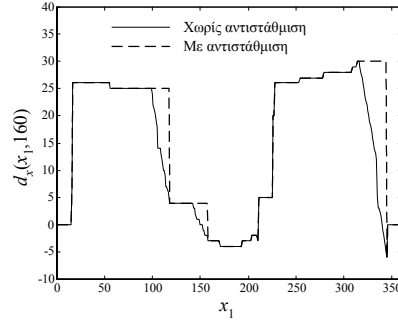
Όπως φαίνεται στο Σχήμα, ενώ το ευθύγραμμο τμήμα AB του μακρινού αντικειμένου είναι ορατό από την κάμερα 1 και προβάλλεται στο τμήμα A_1B_1 , στην κάμερα 2 καλύπτεται από το κοντινό αντικείμενο. Μόνο το σημείο A είναι ορατό στην κάμερα 2, το οποίο και προβάλλεται στο σημείο A_2 . Το αριστερό άκρο C του κοντινού αντικειμένου προβάλλεται στο σημείο $C_1 = B_1$ του επιπέδου I_1 και στο σημείο $C_2 = A_2$ του I_2 . Επιπλέον, έστω ότι $x_1(P)$ και $x_2(P)$ είναι η οριζόντια συνιστώσα της προβολής οποιουδήποτε 3-Δ σημείου P στο επίπεδο I_1 και I_2 αντίστοιχα, και $d_x(P) = x_2(P) - x_1(P)$ είναι η αντίστοιχη οριζόντια απόκλιση. Τότε το μήκος του ευθύγραμμου τμήματος A_1B_1 είναι $l = x_1(B) - x_1(A) = (x_2(A) - x_1(A)) - (x_2(C) - x_1(C)) = d_x(A) - d_x(C) > 0$, αφού είναι δεδομένο ότι $l > 0$, $x_1(B) = x_1(C)$ και $x_2(A) = x_2(C)$. Αυτό σημαίνει ότι για κάθε οριζόντιο ευθύγραμμο τμήμα που είναι ορατό από την κάμερα 1 αλλά όχι από την κάμερα 2, κατά τη διαγραφή του τμήματος με κατεύθυνση από αριστερά προς τα δεξιά παρουσιάζεται μείωση στην οριζόντια απόκλιση, ίση με το μήκος του τμήματος [50].

Όλες οι τιμές απόκλισης που προκύπτουν με τη διαδικασία ελαχιστοποίησης της Εξίσωσης (2-8) για τις περιοχές της εικόνας I_1 που δεν είναι ορατές στην I_2 δεν είναι αξιόπιστες και οδηγούν σε λανθασμένη εκτίμηση βάθους. Είναι σαφές επομένως ότι

- (α) οι περιοχές αυτές πρέπει να ανιχνευθούν
- (β) η επικάλυψη πρέπει να αντισταθμιστεί, αναθέτοντας κατάλληλες τιμές απόκλισης στις επικαλυπτόμενες περιοχές που ανιχνεύονται

Η πρώτη διαδικασία, η *ανίχνευση επικάλυξης* (occlusion detection), επιτυγχάνεται εντοπίζοντας περιοχές της I_1 στις οποίες η οριζόντια απόκλιση είναι φθίνουσα συνάρτηση της συνιστώσας x_1 και με κλίση περίπου ίση με -1 . Η κατακόρυφη απόκλιση δε λαμβάνεται υπόψη

αφού όλες οι αποκλίσεις είναι κυρίως κατά την οριζόντια κατεύθυνση όπως εξηγείται στα παραπάνω παραδείγματα. Η δεύτερη διαδικασία, η *αντιστάθμιση επικάλυψης* (occlusion compensation), επιτυγχάνεται διατηρώντας την τιμή απόκλισης σταθερή σε κάθε επικαλυπτόμενη περιοχή, και ίση με τη μέγιστη απόκλιση στην περιοχή αυτή. Έτσι κάθε επικαλυπτόμενη περιοχή ουσιαστικά συγχωνεύεται με τη γειτονική της περιοχή που έχει το μεγαλύτερο βάθος. Κάτι τέτοιο είναι αναμενόμενο αφού ένα αντικείμενο καλύπτεται από ένα άλλο μόνο όταν βρίσκεται πιο μακριά από την κάμερα.



(α)



(β)



(γ)

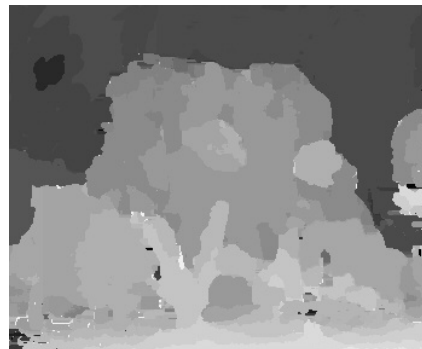


(δ)

Σχήμα 2-5. Ανίχνευση και αντιστάθμιση επικάλυψης για την ακολουθία Claude. (α) 1-Δ αντιστάθμιση για τη γραμμή 160 του πεδίου οριζόντιας απόκλισης. (β) Επικαλυπτόμενες περιοχές (μαύρο χρώμα). (γ) Αντισταθμισμένο πεδίο απόκλισης. (δ) Αντισταθμισμένο πεδίο βάθους.



(α)



(β)

Σχήμα 2-6. Ανίχνευση και αντιστάθμιση επικάλυψης για την ακολουθία Aqua. (α) Αντισταθμισμένο πεδίο οριζόντιας απόκλισης. (β) Αντισταθμισμένο πεδίο βάθους.

Η προτεινόμενη τεχνική ανίχνευσης και αντιστάθμισης επικάλυψης παρουσιάζεται στο Σχήμα 2-5 για το αριστερό κανάλι του καρέ #2 της ακολουθίας Claude. Η μονοδιάστατη περίπτωση φαίνεται πρώτα στο Σχήμα 2-5α, όπου εμφανίζεται το γράφημα της οριζόντιας απόκλισης $d_x(x_1, 160)$ συναρτήσεως της συνιστώσας x_1 για τη γραμμή της εικόνας I_1 με κατακόρυφη

συνιστώσα $y_1 = 160$, πριν και μετά την αντιστάθμιση επικάλυψης. Είναι σαφές ότι στα διαστήματα όπου η d_x αυξάνει ή είναι σταθερή η απόκλιση παραμένει ανέπαφη. Αντίθετα, τα διαστήματα φθίνουσας απόκλισης ανιχνεύονται ως επικαλυπτόμενα και εκεί η απόκλιση αντισταθμίζεται, αναθέτοντας τιμή απόκλισης ίση με εκείνη του γειτονικού μη επικαλυπτόμενου διαστήματος που βρίσκεται στα αριστερά του κάθε επικαλυπτόμενου διαστήματος. Οι επικαλυπτόμενες περιοχές του 2-Δ πεδίου οριζόντιας απόκλισης φαίνονται με μαύρο χρώμα στο Σχήμα 2-5β, ενώ το αντισταθμισμένο πεδίο απόκλισης και το αντίστοιχο πεδίο βάθους στα Σχήματα 2-5γ και δ αντίστοιχα. Παρατηρείται ότι, με εξαίρεση τα ανακριβή όρια (περιγράμματα) των αντικειμένων, το αντισταθμισμένο πεδίο βάθους παρέχει αξιόπιστο διαχωρισμό ανάμεσα στα κοντινά και μακρινά αντικείμενα. Παρόμοια αποτελέσματα φαίνονται στο Σχήμα 2-6 για το καρέ #1 της ακολουθίας Aqua. Σε κάθε περίπτωση το πεδίο βάθους είναι πιο αξιόπιστο μετά την αντιστάθμιση.

2.5 ΚΑΤΑΤΜΗΣΗ

Η κατάτμηση με βάση το πεδίο βάθους σε στερεοσκοπικές ακολουθίες πραγματοποιείται εφαρμόζοντας τον αλγόριθμο M-RSST στο αντισταθμισμένο πεδίο βάθους που προκύπτει από το στερεοσκοπικό ζεύγος κάθε καρέ της ακολουθίας. Για τους σκοπούς της κατάτμησης με βάση το πεδίο βάθους η απόσταση $\delta_d(X, Y)$ μεταξύ δύο γειτονικών περιοχών X και Y συνήθως ορίζεται, όμοια με την (1-1), ως το γινόμενο της Ευκλείδειας απόστασης των μέσων τιμών βάθους των δύο περιοχών και ενός όρου που είναι συνάρτηση του μεγέθους των δύο περιοχών σε pixels και ευνοεί τη συγχώνευση των μικρών περιοχών. Το πεδίο βάθους είναι μονοδιάστατο και επομένως χρησιμοποιείται η απόλυτη τιμή της διαφοράς βάθους:

$$\delta_d(X, Y) = |d(X) - d(Y)| \frac{a(X)a(Y)}{a(X) + a(Y)} \quad (2-12)$$

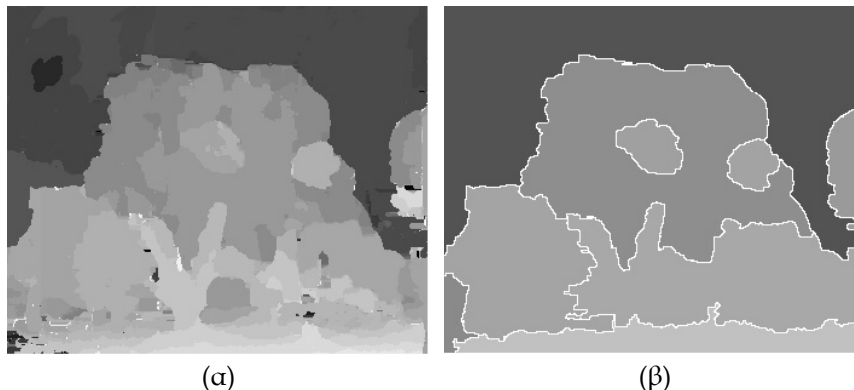
όπου $d(X)$ είναι οι μέση τιμή βάθους στο εσωτερικό της περιοχής X και $a(X)$ είναι το μέγεθος της σε pixels. Στα πειράματα χρησιμοποιείται κατώφλι απόστασης για τον τερματισμό της διαδικασίας κατάτμησης, όπως και στην περίπτωση κατάτμησης με βάση το χρώμα και την κίνηση. Επίσης υιοθετείται χαμηλότερο επίπεδο ανάλυσης $L_0 = 3$ για την αρχικοποίηση του αλγορίθμου, που αντιστοιχεί σε blocks μεγέθους 8×8 pixels.



Σχήμα 2-7. Κατάτμηση με βάση το πεδίο βάθους για την ακολουθία Claude. (α) Αντισταθμισμένο πεδίο βάθους. (β) Αποτέλεσμα κατάτμησης.

Στο Σχήμα 2-7 εμφανίζονται τα αποτελέσματα της κατάτμησης με βάση το πεδίο βάθους για την ακολουθία Claude. Συγκεκριμένα, στο Σχήμα 2-7α φαίνεται το αντισταθμισμένο πε-

δίο βάθους, ενώ στο Σχήμα 2-7β το τελικό στάδιο κατάτμησης για το υψηλότερο επίπεδο ανάλυσης $L_0 = 0$. Αντίστοιχα αποτελέσματα παρουσιάζονται στο Σχήμα 2-8 για την ακολουθία Aqua. Παρατηρείται ότι για την ακολουθία Claude παράγονται δύο αντικείμενα, που αντιστοιχούν στο φόντο και στο πρόσωπο που βρίσκεται στο προσκήνιο, ενώ για την ακολουθία Aqua παράγονται εννέα συνολικά αντικείμενα.



Σχήμα 2-8. Κατάτμηση με βάση το πεδίο βάθους για την ακολουθία Aqua. (α) Αντισταθμισμένο πεδίο βάθους. (β) Αποτέλεσμα κατάτμησης.

Και στις δύο περιπτώσεις οι περιοχές που ανιχνεύει ο αλγόριθμος κατάτμησης με βάση το πεδίο βάθους είναι πολύ κοντά στα πραγματικά αντικείμενα που υπάρχουν στις εικόνες. Αυτό είναι αναμενόμενο αφού κάθε αντικείμενο βρίσκεται συνήθως στο ίδιο επίπεδο βάθους. Όμως, λόγω της λανθασμένης εκτίμησης της απόκλισης, καθώς και της επικάλυψης μεταξύ αντικειμένων, οι οριακές περιοχές των αντικειμένων, επομένως και τα περιγράμματά τους είναι σχεδόν πάντα ανακριβή. Το ίδιο συμβαίνει και με την κατάτμηση με βάση την κίνηση. Αντίθετα η κατάτμηση με βάση το χρώμα δίνει πάντα σωστά περιγράμματα αντικειμένων, αλλά αποτυγχάνει να ανιχνεύσει τα πραγματικά αντικείμενα. Γι' αυτό το λόγο προτείνεται η συγχώνευση των αντικειμένων που προκύπτουν από διαφορετικές διαδικασίες κατάτμησης, ώστε να συνδυάζονται τα πλεονεκτήματα της κάθε μιας. Ο αλγόριθμος συγχώνευσης παρουσιάζεται στο επόμενο Κεφάλαιο.

Συγχώνευση Αντικειμένων

Τα τμήματα εικόνας που προκύπτουν από την κατάτμηση με βάση το πεδίο βάθους σε μια στερεοσκοπική ακολουθία βίντεο προσεγγίζουν ικανοποιητικά τα πραγματικά αντικείμενα της ακολουθίας, αφού τα τελευταία συνήθως αποτελούνται από περιοχές που βρίσκονται στο ίδιο επίπεδο βάθους· όμως τα περιγράμματα αυτών των τμημάτων είναι αναξιόπιστα λόγω της ανακρίβειας στην εκτίμηση του πεδίου απόκλισης και των ζητημάτων επικάλυψης. Από την άλλη μεριά η κατάτμηση με βάση το χρώμα δίνει ακριβή περιγράμματα αλλά συνήθως διαχωρίζει κάθε αντικείμενο σε περισσότερα από ένα τμήματα. Για το λόγο αυτό παρουσιάζεται ένας αλγόριθμος συγχώνευσης των τμημάτων χρώματος με βάση την πληροφορία των τμημάτων βάθους. Έτσι οι δύο ανεξάρτητες κατατμήσεις συνδυάζονται και εξάγονται με ακρίβεια τα σημασιολογικά αντικείμενα της ακολουθίας. Η ίδια ακριβώς τεχνική μπορεί να χρησιμοποιηθεί για την συγχώνευση των τμημάτων χρώματος με τα τμήματα κίνησης. Η αναπαράσταση του οπτικού περιεχομένου που προκύπτει από τη συγχώνευση είναι πιο αξιόπιστη και οδηγεί σε αποτελεσματικότερη εξαγωγή περίληψης καθώς και αναζήτησης με βάση το περιεχόμενο.

3.1 ΑΝΙΧΝΕΥΣΗ ΣΗΜΑΣΙΟΛΟΓΙΚΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ

Η εξαγωγή περίληψης μιας ακολουθίας βίντεο, καθώς και η αναζήτηση με βάση το περιεχόμενο, μπορούν να πραγματοποιηθούν πολύ πιο αποδοτικά και αξιόπιστα, αν το οπτικό περιεχόμενο περιγράφεται με βάση τα αντικείμενα που εμφανίζονται στην ακολουθία και έχουν σημασιολογικό (semantic) περιεχόμενο, δηλαδή πραγματικά αντικείμενα με φυσικό νόημα. Η σημασιολογική κατάτμηση έχει πρόσφατα προσελκύσει μεγάλο ενδιαφέρον στην ερευνητική κοινότητα, και ειδικότερο στο πλαίσιο των αναπτυσσόμενων προτύπων κωδικοποίησης MPEG-4 και MPEG-7 [1,51,85,117]. Παρά το γεγονός ότι κάποιες λύσεις έχουν προταθεί για συγκεκριμένες εφαρμογές (π.χ. βιντεοτηλέφωνο, δελτία ειδήσεων κλπ.) [23,44], η σημασιολογική κατάτμηση για ακολουθίες γενικού περιεχομένου θεωρείται ακόμη ένα άλυτο πρόβλημα [95].

Στις στερεοσκοπικές ακολουθίες βίντεο όμως, όπου η πληροφορία βάθους μπορεί να εκτιμηθεί με αρκετά μεγάλη ακρίβεια, αντικείμενα με σημασιολογικό περιεχόμενο μπορούν να αναγνωρισθούν αφού η επιφάνεια ενός πραγματικού αντικειμένου βρίσκεται συνήθως στο ίδιο επίπεδο βάθους. Για τη σωστή αναγνώριση πραγματικών αντικειμένων, με ταυτόχρονη διατήρηση του ακριβούς περιγράμματος κάθε αντικειμένου, προτείνεται η κατάτμηση των ακολουθιών με βάση το χρώμα, την κίνηση και το πεδίο βάθους, και στη συνέχεια η συγχώνευση των αντικειμένων που προκύπτουν από τις διαφορετικές κατατμήσεις.

Παρά το γεγονός ότι η κατάτμηση με βάση το πεδίο βάθους παρέχει καλύτερη αναπαράσταση το οπτικού περιεχομένου, αφού οι περιοχές που ανιχνεύει είναι πιο κοντά στα πραγματικά αντικείμενα, η κατάτμηση αυτή αποτυγχάνει να εκτιμήσει με ακρίβεια τις οριακές περιο-

χές των αντικειμένων και τα περιγράμματά τους. Αυτό οφείλεται στις ανακρίβειες της εκτίμησης της απόκλισης ανάμεσα στις εικόνες που προέρχονται από διαφορετικές κάμερες, καθώς και σε ζητήματα που σχετίζονται με την επικάλυψη των αντικειμένων. Αντίθετα, η κατάτμηση με βάση το χρώμα ανιχνεύει με μεγάλη ακρίβεια τα περιγράμματα των αντικειμένων αλλά δε μπορεί να αναγνωρίσει σωστά τα πραγματικά αντικείμενα, αφού συνήθως τα διαχωρίζει σε περισσότερα από ένα τμήματα [1]. Γι' αυτό το λόγο προτείνεται η *συγχώνευση* (fusion) των αντικειμένων που προκύπτουν από τα διαφορετικά είδη κατάτμησης σε μία κοινή κατάτμηση η οποία συνδυάζει τα πλεονεκτήματα της κάθε μιας.

Ο προτεινόμενος αλγόριθμος συγχώνευσης βασίζεται σε δύο διαφορετικές κατατμήσεις μιας εικόνας, μία με βάση το χρώμα και μία με βάση το πεδίο βάθους. Τα τμήματα που προκύπτουν ονομάζονται *τμήματα χρώματος* και *τμήματα βάθους* αντίστοιχα. Ο αλγόριθμος ουσιαστικά συγχωνεύει τα τμήματα χρώματος που ανήκουν (έχουν δηλαδή μεγαλύτερο ποσοστό επικάλυψης) στο ίδιο τμήμα βάθους. Μία ανάλογη διαδικασία μπορεί να πραγματοποιηθεί με τα τμήματα χρώματος και τα *τμήματα κίνησης* τα οποία προκύπτουν με κατάτμηση με βάση την κίνηση. Τα τμήματα κίνησης, όπως και τα τμήματα βάθους, δεν παρέχουν ακριβή περιγράμματα αλλά δίνουν καλή εκτίμηση των πραγματικών αντικειμένων αφού δύο τμήματα που κινούνται με τον ίδιο τρόπο συνήθως ανήκουν στο ίδιο αντικείμενο.

Στην πραγματικότητα, ο χρησιμοποιούμενος αλγόριθμος συγχώνευσης είναι εμπνευσμένος από την εργασία [1], όπου όμως αναφέρεται συγχώνευση χρώματος και κίνησης μόνο, ενώ στο πλαίσιο της διατριβής πραγματοποιείται κυρίως συγχώνευση χρώματος και βάθους. Ο αλγόριθμος προτάθηκε για πρώτη φορά στην εργασία [39] αλλά χρησιμοποιήθηκε και στην [38]. Σημειώνεται ότι η συγχώνευση τμημάτων χρώματος με βάση την κίνηση είναι ιδιαίτερα χρήσιμη όταν η πληροφορία βάθους δεν είναι διαθέσιμη. Η ακόλουθη παρουσίαση του αλγορίθμου αναφέρεται σε συγχώνευση με βάση το βάθος, αλλά μπορεί να εφαρμοσθεί χωρίς αλλαγές και για συγχώνευση με βάση την κίνηση.

3.2 ΑΛΓΟΡΙΘΜΟΣ ΣΥΓΧΩΝΕΥΣΗΣ

Ας υποθέσουμε ότι ένα καρέ μιας ακολουθίας διαχωρίζεται σε K^c τμήματα με βάση το χρώμα και σε K^d τμήματα με βάση το πεδίο βάθους. Έστω ότι τα τμήματα χρώματος συμβολίζονται με $S_i^c, i = 1, 2, \dots, K^c$, και τα τμήματα βάθους με $S_i^d, i = 1, 2, \dots, K^d$, αντίστοιχα. Τα τμήματα χρώματος S_i^c είναι συνεκτικά και δεν έχουν επικαλύψεις μεταξύ τους, δηλαδή $S_i^c \cap S_k^c = \emptyset$ για κάθε $i, k = 1, 2, \dots, K^c$, με $i \neq k$. Όμοια για τα τμήματα βάθους S_i^d ισχύει $S_i^d \cap S_k^d = \emptyset$ για κάθε $i, k = 1, 2, \dots, K^d$, με $i \neq k$. Έστω επίσης ότι με G^c και G^d συμβολίζονται τα σύνολα όλων των τμημάτων με βάση το χρώμα και το πεδίο βάθους αντίστοιχα:

$$G^c = \{S_i^c, i = 1, 2, \dots, K^c\} \quad (3-1)$$

$$G^d = \{S_i^d, i = 1, 2, \dots, K^d\} \quad (3-2)$$

Τα τμήματα χρώματος προβάλλονται στα τμήματα βάθους, έτσι ώστε να διατηρηθούν τα αντικείμενα που ανιχνεύονται από την πληροφορία βάθους, αλλά ταυτόχρονα και τα περιγράμματα που προκύπτουν με βάση το χρώμα. Για το σκοπό αυτό κάθε τμήμα χρώματος S_i^c σχετίζεται με ένα τμήμα βάθους με τέτοιο τρόπο ώστε η επιφάνεια επικάλυψης των δύο τμη-

μάτων να μεγιστοποιείται. Αυτό επιτυγχάνεται μέσω μιας *συνάρτησης προβολής* (projection function):

$$p(S_i^c, G^d) = \arg \max_{g \in G^d} \{a(g \cap S_i^c)\}, \quad i = 1, 2, \dots, K^c \quad (3-3)$$

όπου $a(\cdot)$ είναι η επιφάνεια, δηλαδή το πλήθος των pixels, ενός τμήματος. Με βάση την προηγούμενη εξίσωση, ορίζονται K^d σύνολα τμημάτων χρώματος, έστω C_i , $i = 1, 2, \dots, K^d$, κάθε ένα από τα οποία περιέχει όλα τα τμήματα χρώματος που προβάλλονται στο ίδιο τμήμα βάθους S_i^d :

$$C_i = \{g \in G^c : p(g, G^d) = S_i^d\}, \quad i = 1, 2, \dots, K^d \quad (3-4)$$

Στη συνέχεια κατασκευάζονται $K = K^d$ τμήματα S_i , $i = 1, 2, \dots, K$, κάθε ένα από τα οποία προκύπτει από τη συγχώνευση όλων των στοιχείων του αντίστοιχου συνόλου C_i :

$$S_i = \bigcup_{g \in C_i} g, \quad i = 1, 2, \dots, K \quad (3-5)$$

Τέλος, η συνολική κατάτμηση, G , ορίζεται ως το σύνολο όλων των παραπάνω τμημάτων:

$$G = \{S_i, i = 1, 2, \dots, K\} \quad (3-6)$$

Με άλλα λόγια, τα τμήματα χρώματος συγχωνεύονται κατά ομάδες σε $K = K^d$ νέα τμήματα με κριτήριο την ομοιότητα βάθους. Η τελική κατάτμηση που παράγεται από τον αλγόριθμο αποτελείται από τμήματα τα οποία περιέχουν τις ίδιες περίπου περιοχές της εικόνας, όπως και τα τμήματα βάθους, αλλά με ακριβή περιγράμματα που προέρχονται από τα τμήματα χρώματος.

3.3 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

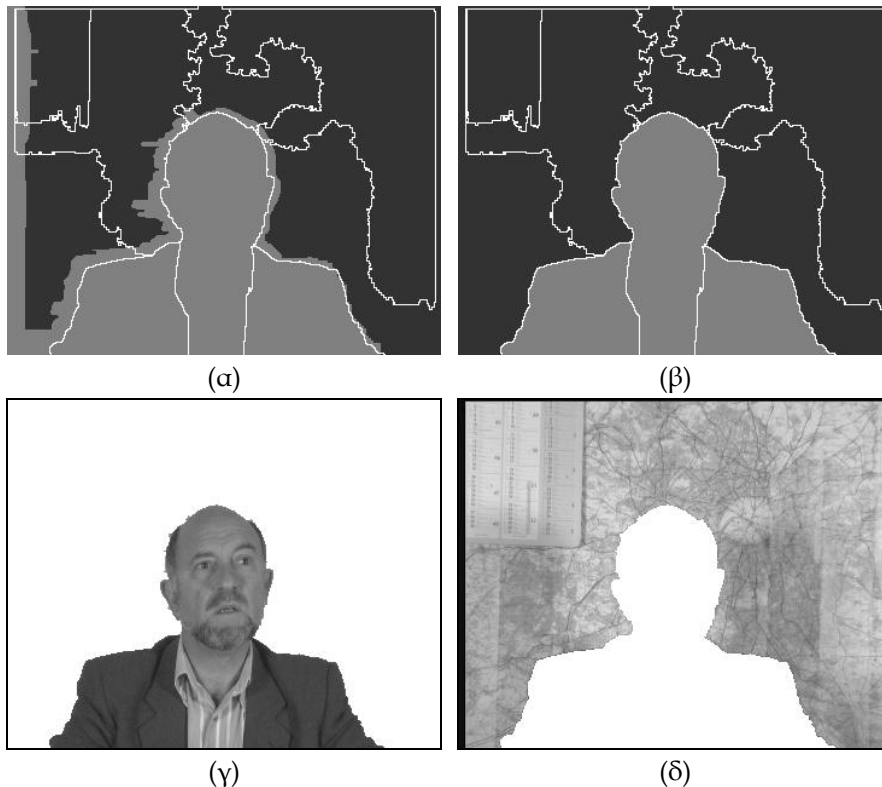
Πειραματικά αποτελέσματα της συγχώνευσης αντικειμένων παρουσιάζονται στα Σχήματα 3-1 και 3-2 για την ακολουθία Claude. Η εικόνα του αριστερού καναλιού του καρέ #2 της αρχικής ακολουθίας φαίνεται στο Σχήμα 3-1α. Αρχικά πραγματοποιείται ανεξάρτητη κατάτμηση με βάση το χρώμα και το πεδίο βάθους, όπως φαίνεται στα Σχήματα 3-1β,γ αντίστοιχα. Η κατάτμηση και στις δύο περιπτώσεις γίνεται με τον αλγόριθμο M-RSST, όπως έχει περιγραφεί αναλυτικά στις Ενότητες 1.3 και 2.5.



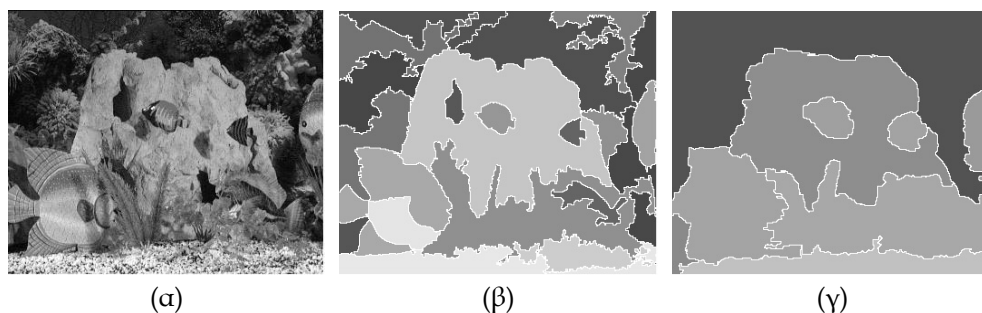
Σχήμα 3-1. Κατάτμηση ακολουθίας Claude. (α) Αριστερό κανάλι. (β) Κατάτμηση με βάση το χρώμα. (γ) Κατάτμηση με βάση το πεδίο βάθους.

Η κατάτμηση με βάση το βάθος συνδυάζεται στο Σχήμα 3-2α με τα περιγράμματα των τμημάτων χρώματος. Είναι προφανές ότι το πρόσωπο στο προσκήνιο της εικόνας αντιστοιχεί σε ένα μόνο τμήμα βάθους και σε τρία διαφορετικά τμήματα χρώματος, ενώ το φόντο της εικόνας σε ένα πάλι τμήμα βάθους και σε έξι τμήματα χρώματος. Είναι επίσης προφανές ότι

μόνο η κατάτμηση με βάση το πεδίο βάθους δίνει σωστά τα δύο αντικείμενα, ενώ η κατάτμηση με βάση το χρώμα διαχωρίζει τα αντικείμενα αυτά σε περισσότερα τμήματα αλλά διατηρεί τα ακριβή περιγράμματα τους. Ένα τμήμα εικόνας για κάθε αντικείμενο με σημασιολογικό περιεχόμενο, και μάλιστα με το ακριβές περίγραμμα, μπορεί να παραχθεί με τον προτεινόμενο αλγόριθμο συγχώνευσης.



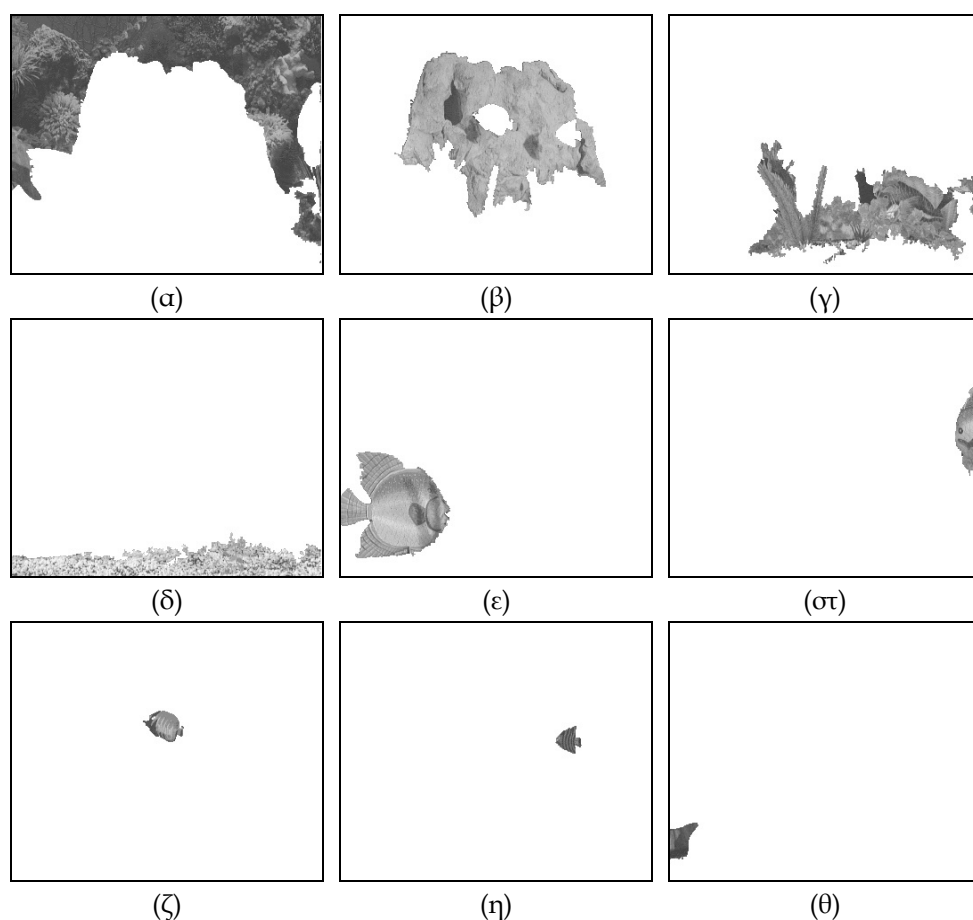
Σχήμα 3-2. Συγχώνευση αντικειμένων για την ακολουθία *Claude*. (α) Τμήματα βάθους σε συνδυασμό με περιγράμματα τμημάτων χρώματος. (β) Αποτέλεσμα συγχώνευσης, σε συνδυασμό με τα ίδια περιγράμματα. (γ) Κοντινό αντικείμενο (πρόσωπο, *Claude*). (δ) Μακρινό αντικείμενο (φόντο).



Σχήμα 3-3. Κατάτμηση ακολουθίας *Aqua*. (α) Αριστερό κανάλι. (β) Κατάτμηση με βάση το χρώμα. (γ) Κατάτμηση με βάση το πεδίο βάθους.

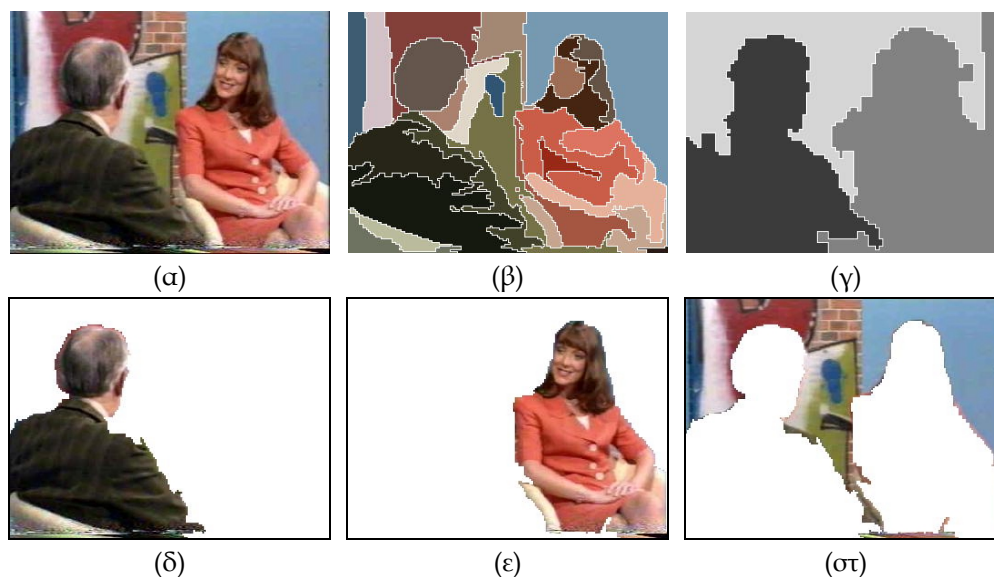
Στο Σχήμα 3-2β φαίνεται το αποτέλεσμα αυτής της συγχώνευσης, και πάλι σε συνδυασμό με τα περιγράμματα των τμημάτων χρώματος. Τα δύο πραγματικά αντικείμενα της εικόνας δηλαδή το κοντινό (πρόσωπο) και το μακρινό (φόντο), φαίνονται χωριστά στα Σχήματα 3-2γ,δ αντίστοιχα. Σημειώνεται ο διαχωρισμός της εικόνας στα δύο πραγματικά αντικείμενα γίνεται με πλήρως αυτοματοποιημένο τρόπο, χωρίς μάλιστα ο αριθμός των αντικειμένων να είναι εκ των προτέρων γνωστός.

Στα Σχήματα 3-3 και 3-4 παρουσιάζονται ανάλογα αποτελέσματα συγχώνευσης τμημάτων χρώματος και βάθους για την ακολουθία Aqua. Η εικόνα του αριστερού καναλιού του καρέ #1 της αρχικής ακολουθίας φαίνεται στο Σχήμα 3-3α. Και πάλι πραγματοποιείται ανεξάρτητη κατάτμηση με βάση το χρώμα και το πεδίο βάθους, όπως φαίνεται στα Σχήματα 3-3β,γ αντίστοιχα. Με τη συγχώνευση παράγονται εννέα συνολικά τμήματα, τα οποία απεικονίζονται στα Σχήματα 3-4α-θ. Παρατηρείται ότι και στην περίπτωση αυτή όλα σχεδόν τα τμήματα αντιστοιχούν σε πραγματικά αντικείμενα της εικόνας με σημασιολογικό περιεχόμενο. Εξαιρεση αποτελεί μόνο το τμήμα του Σχήματος 3-4θ, το οποίο αντιστοιχεί στο φόντο της εικόνας και διαχωρίζεται από το κυρίως φόντο του Σχήματος 3-4α γιατί και στην αρχική εικόνα τα δύο αντικείμενα δεν είναι συνεκτικά.



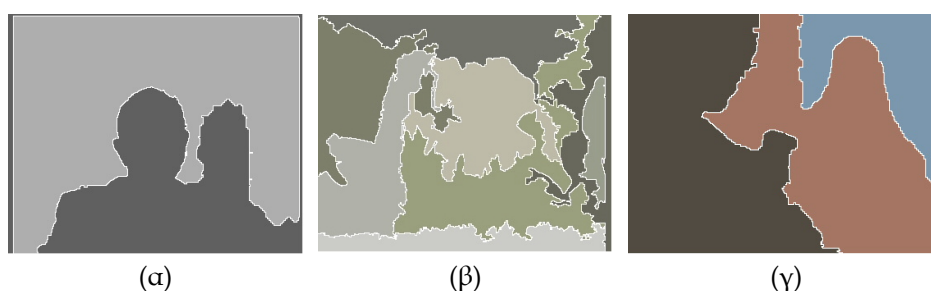
Σχήμα 3-4. Συγχώνευση αντικειμένων για την ακολουθία Aqua. Παρουσιάζονται χωριστά τα εννέα διαφορετικά αντικείμενα που εξάγονται από τον αλγόριθμο.

Ένα ακόμη παράδειγμα συγχώνευσης παρουσιάζεται στο Σχήμα 3-5 για ένα καρέ της στερεοσκοπικής ακολουθίας “Eye to Eye”, η οποία χρησιμοποιείται και στην Ενότητα 5.4 για την επιλογή χαρακτηριστικών πλάνων. Το αρχικό αριστερό κανάλι του καρέ φαίνεται στο Σχήμα 3-5α και παρουσιάζει τη σκηνή μιας συνέντευξης με δύο ανθρώπους σε εσωτερικό χώρο (στούντιο). Η κατάτμηση με βάση το χρώμα και το πεδίο βάθους φαίνεται στα Σχήματα 3-5β και γ αντίστοιχα. Με τη συγχώνευση δημιουργούνται συνολικά τρία τμήματα που απεικονίζονται στα Σχήματα 3-5δ-ςτ. Πρόκειται για τους δύο ανθρώπους που συμμετέχουν στη συνέντευξη και το τμήμα του φόντου. Επιβεβαιώνεται ότι και στην περίπτωση αυτή τα πραγματικά αντικείμενα της εικόνας εξάγονται με μεγάλη ακρίβεια.



Σχήμα 3-5. Συγχώνευση αντικειμένων για την ακολουθία "Eye to Eye". (α) Αριστερό κανάλι. (β) Κατάτμηση με βάση το χρώμα. (γ) Κατάτμηση με βάση το πεδίο βάθους. (δ) Κοντινό αντικείμενο #1. (ε) Κοντινό αντικείμενο #2. (στ) Μακρινό αντικείμενο (φόντο).

Πρέπει να σημειωθεί ότι η αναγνώριση των πραγματικών αντικειμένων δε μπορεί να επιτευχθεί μόνο με την πληροφορία χρώματος, γιατί συνήθως ένα αντικείμενο αποτελείται από πολλές περιοχές με διαφορετικά χαρακτηριστικά χρώματος. Αντίθετα με χρήση της πληροφορίας βάθους η αναγνώριση αυτή επιτυγχάνεται σε αρκετά μεγάλο βαθμό. Για να γίνει αυτό περισσότερο κατανοητό, πραγματοποιήθηκαν πειράματα κατάτμησης των τριών προηγούμενων ακολουθιών με βάση το χρώμα μόνο. Για να είναι μάλιστα συγκρίσιμα τα αποτελέσματα των δύο τεχνικών, χρησιμοποιήθηκε στον αλγόριθμο M-RSST κατώφλι αριθμού τμημάτων, έτσι ώστε η κατάτμηση με βάση το χρώμα να δώσει τον ίδιο αριθμό τμημάτων που δίνει και ο αλγόριθμος συγχώνευσης. Στα Σχήματα 3-6α,β,γ παρουσιάζονται τα αποτελέσματα κατάτμησης για τις ακολουθίες Claude, Aqua και "Eye to Eye" αντίστοιχα. Ο αριθμός τμημάτων για τις τρεις περιπτώσεις είναι 2 (Claude), 9 (Aqua) και 3 ("Eye to Eye"), ακριβώς όπως και στα Σχήματα 3-2γ-δ, 3-4α-θ, και 3-5δ-στ.



Σχήμα 3-6. Κατάτμηση με βάση το χρώμα μόνο, με περιορισμένο αριθμό τμημάτων. (α) Ακολουθία Claude. (β) Ακολουθία Aqua. (γ) Ακολουθία "Eye to Eye".

Τα αποτελέσματα που παρουσιάζονται στο Σχήμα 3-6 δεν είναι ικανοποιητικά αφού περιοχές που ανήκουν σε διαφορετικά αντικείμενα έχουν συγχωνευθεί, ενώ περιοχές του ίδιου αντικειμένου έχουν διαχωριστεί σε διαφορετικά τμήματα. Αντίθετα ο συνδυασμός πληροφορίας χρώματος και βάθους δίνει αποτελεσματική παράσταση του οπτικού περιεχομένου, δικαιολογώντας έτσι το επιπρόσθετο υπολογιστικό κόστος για την εκτίμηση και κατάτμηση με βάση το πεδίο βάθους. Σε κάποιες περιπτώσεις ασφαλώς, και ιδιαίτερα σε μακρινά πλάνα, η

απόκλιση μεταξύ των εικόνων της αριστερής και δεξιάς κάμερας είναι αμελητέα και επομένως η εκτίμηση βάθους είναι αναξιόπιστη. Οι περιπτώσεις αυτές ανιχνεύονται εύκολα διότι συνήθως οδηγούν σε ένα μόνο τμήμα με βάση το βάθος. Σε αυτές τις περιπτώσεις η κατάτμηση βάθους αγνοείται και λαμβάνεται υπόψη μόνο η κατάτμηση χρώματος και κίνησης.

ΧΑΡΑΚΤΗΡΙΣΜΟΣ, ΠΡΟΕΠΙΣΚΟΠΗΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Στο δεύτερο Μέρος της διατριβής εισάγεται το πρόβλημα της αναπαράστασης του οπτικού περιεχομένου εικόνων και ακολουθιών βίντεο με τέτοιο τρόπο ώστε να επιτρέπεται η ενιαία περιγραφή όλων των διαθέσιμων χαρακτηριστικών και να διευκολύνονται οι συγκρίσεις εικόνων, καρέ ή πλάνων ακολουθιών. Στη συνέχεια η αναπαράσταση αυτή χρησιμοποιείται για το χαρακτηρισμό και τη δεικτοδότηση του περιεχομένου με δύο κύριες εφαρμογές: την εξαγωγή περίληψης, με σκοπό την προεπισκόπηση οπτικοακουστικού υλικού, και την αναζήτηση με βάση το περιεχόμενο.

Στο Κεφάλαιο 4 παρατίθεται η προτεινόμενη μέθοδος *αναπαράστασης*, στην οποία χρησιμοποιούνται τα χαρακτηριστικά των τμημάτων που προκύπτουν από την κατάτμηση με βάση το χρώμα, την κίνηση και το πεδίο βάθους, περιλαμβάνοντας για παράδειγμα τη θέση, το μέγεθος, τις χρωματικές συνιστώσες, το μέσο διάνυσμα κίνησης και το βάθος του κάθε τμήματος. Τα χαρακτηριστικά αυτά συνδυάζονται για την κατασκευή *διανυσμάτων περιγραφής*, τα οποία έχουν τη μορφή πολυδιάστατων ασαφών ιστογραμμάτων.

Στο Κεφάλαιο 5 περιγράφεται το πρώτο στάδιο εξαγωγής περίληψης, το οποίο περιλαμβάνει την επιλογή ενός περιορισμένου αριθμού *χαρακτηριστικών πλάνων*, τα οποία όμως περιγράφουν ικανοποιητικά το οπτικό περιεχόμενο μιας ακολουθίας. Για το σκοπό αυτό πραγματοποιείται ανίχνευση και χαρακτηρισμός των πλάνων με συνολικά διανύσματα περιγραφής, και στη συνέχεια υλοποιείται ομαδοποίηση πλάνων με παρόμοιο περιεχόμενο σε ένα προκαθορισμένο αριθμό ομάδων και βέλτιστη επιλογή ενός αντιπροσωπευτικού πλάνου από κάθε ομάδα.

Το δεύτερο στάδιο εξαγωγής περίληψης είναι η επιλογή ενός περιορισμένου συνόλου *χαρακτηριστικών καρέ* από κάθε πλάνο, για την οποία προτείνονται δύο διαφορετικές μέθοδοι. Η μέθοδος *χρονικής μεταβολής* παρουσιάζεται στο Κεφάλαιο 6 και βασίζεται στην επιλογή ακρότατων σημείων της τροχιάς του διανύσματος περιγραφής, ενώ η μέθοδος *συσχέτισης* παρουσιάζεται στο Κεφάλαιο 7 και βασίζεται στην ελαχιστοποίηση ενός κριτηρίου συσχέτισης μεταξύ των διανυσμάτων περιγραφής.

Τέλος, το πρόβλημα της γενικευμένης *αναζήτησης* και *ανάκλησης* εικόνων και ακολουθιών με βάση το περιεχόμενο εξετάζεται στο Κεφάλαιο 8. Προτείνονται δύο εναλλακτικοί μηχανισμοί αναζήτησης, μέσω παραδείγματος και μέσω σχεδιαγράμματος, ενώ και στις δύο περιπτώσεις υιοθετείται μία τεχνική ανάδρασης από το χρήστη, η οποία καθιστά την αναζήτηση μία διαδικασία αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστικού συστήματος.

Αναπαράσταση Ακολουθιών Εικόνων

Για την αναπαράσταση του οπτικού περιεχομένου μιας ακολουθίας βίντεο είναι απαραίτητη η εξαγωγή διανυσμάτων περιγραφής για κάθε καρέ και κάθε πλάνο της ακολουθίας, με τέτοιο τρόπο ώστε τα διανύσματα να είναι ίσων διαστάσεων και να υπάρχει ένα προς ένα αντιστοιχία μεταξύ των στοιχείων διαφορετικών διανυσμάτων. Ο χαρακτηρισμός και η δεικτοδότηση μέσω των διανυσμάτων περιγραφής επιτρέπει τη γρήγορη και αποτελεσματική σύγκριση εικόνων, καρέ ή πλάνων ακολουθιών, και κατ' επέκταση την εξαγωγή περίληψης και την ανάκληση με βάση το περιεχόμενο. Για τη δημιουργία των διανυσμάτων περιγραφής χρησιμοποιούνται τα χαρακτηριστικά των τμημάτων που προκύπτουν από την κατάτμηση με βάση το χρώμα, την κίνηση και το πεδίο βάθους, περιλαμβάνοντας για παράδειγμα τη θέση, το μέγεθος, τις χρωματικές συνιστώσες, το μέσο διάνυσμα κίνησης και το βάθος του κάθε τμήματος. Όλα τα χαρακτηριστικά αυτά ταξινομούνται σε προκαθορισμένες κλάσεις και συγκεντρώνονται σχηματίζοντας ένα πολυδιάστατο ιστόγραμμα. Επιπλέον, για την αποφυγή λανθασμένης ταξινόμησης παρόμοιων χαρακτηριστικών σε διαφορετικές κλάσεις, η τεχνική γενικεύεται με την ασαφή αναπαράσταση της ταξινόμησης, καταλήγοντας έτσι σε ένα ασαφές πολυδιάστατο ιστόγραμμα. Τέλος τα διανύσματα περιγραφής προκύπτουν με τη συγκέντρωση όλων των στοιχείων του ιστογράμματος με προκαθορισμένη σειρά. Για την καλύτερη κατανόηση της μεθόδου δίνεται ένα αναλυτικό παράδειγμα εξαγωγής διανυσμάτων περιγραφής σε μια συνθετική ακολουθία βίντεο.

4.1 ΑΣΑΦΗ ΙΣΤΟΓΡΑΜΜΑΤΑ

Όλα τα χαρακτηριστικά που εξάγονται από την ανάλυση μιας ακολουθίας βίντεο σε αντικείμενα (όπως πλήθος, θέση, μέγεθος, χρώμα και κίνηση κάθε αντικειμένου) μπορούν να χρησιμοποιηθούν για να περιγράψουν το οπτικό περιεχόμενο του κάθε καρέ. Η περιγραφή γίνεται συνήθως με τη μορφή ενός διανύσματος χαρακτηριστικών (feature vector) ή διανύσματος περιγραφής. Όμως στη συγκεκριμένη περίπτωση το διάνυσμα περιγραφής δεν μπορεί να περιέχει άμεσα τα χαρακτηριστικά αυτά, καθώς το πλήθος τους διαφέρει από καρέ σε καρέ. Για παράδειγμα, ένα καρέ που αποτελείται από 20 αντικείμενα απαιτεί το διπλάσιο αριθμό από χαρακτηριστικά στοιχεία σε σχέση με ένα καρέ που αποτελείται από 10 αντικείμενα. Επιπλέον, επιλέγοντας μια τέτοια λύση, δεν υπάρχει καμία απολύτως αντιστοιχία μεταξύ των στοιχείων των διανυσμάτων δύο διαφορετικών καρέ, κάνοντας έτσι αδύνατη τη σύγκριση μεταξύ διανυσμάτων περιγραφής. Για να αντιμετωπισθεί αυτό το πρόβλημα προτείνεται στο πλαίσιο της διατριβής η ταξινόμηση των αντικειμένων ως προς τη θέση, το μέγεθος, το χρώμα την κίνηση ή το βάθος σε προκαθορισμένες κλάσεις, σχηματίζοντας έτσι ένα πολυδιάστατο ιστόγραμμα. Σύμφωνα με αυτή την τεχνική, κάθε στοιχείο του διανύσματος περιγραφής αντιστοιχεί σε μία συγκεκριμένη κλάση (που είναι ισοδύναμη με ένα bin του ιστογράμματος) και περιέχει το πλήθος των τμημάτων που ανήκουν στην κλάση αυτή.

Το μέγεθος των τμημάτων λαμβάνεται υπόψη αναθέτοντας διαφορετικές κλάσεις για μικρά και μεγάλα αντικείμενα, δηλαδή το μέγεθος θεωρείται ένα χαρακτηριστικό των αντικειμένων όπως και το χρώμα ή η κίνηση. Για παράδειγμα, ένα μεγάλο κινούμενο αντικείμενο ταξινομείται σε διαφορετική κλάση από ένα μικρό. Παρά το γεγονός ότι τα μεγάλα αντικείμενα μπορεί να θεωρούνται πιο σημαντικά από τα μικρά, η παραπάνω προσέγγιση εγγυάται ότι όλες οι πληροφορίες διατηρούνται, και έτσι σε ένα περιβάλλον ανάκλησης με βάση το περιεχόμενο, ο βαθμός σπουδαιότητας κάθε χαρακτηριστικού μπορεί να καθοριστεί από τον τελικό χρήστη, πιθανώς αναθέτοντας κατάλληλα βάρη στα στοιχεία του διανύσματος περιγραφής [35].

Επιπλέον, για να μειωθεί η πιθανότητα ταξινόμησης δύο παρόμοιων αντικειμένων σε διαφορετικές κλάσεις, προκαλώντας λανθασμένες συγκρίσεις, ανατίθεται ένας *βαθμός συμμετοχής* (degree of membership) κάθε αντικειμένου σε κάθε κλάση, οδηγώντας σε *ασαφή ταξινόμηση* (fuzzy classification) [76]. Στα συμβατικά ιστογράμματα, κάθε δείγμα – δηλαδή κάθε αντικείμενο, τμήμα ή περιοχή εικόνας – μπορεί να ανήκει σε μία μόνο κλάση (bin) του ιστογράμματος. Έτσι, δύο παρόμοια δείγματα που βρίσκονται σε αντίθετες πλευρές ως προς το όριο μεταξύ δύο κλάσεων θεωρείται ότι ανήκουν σε διαφορετικές κλάσεις. Αντίθετα, με την ασαφή ταξινόμηση, κάθε δείγμα επιτρέπεται να ανήκει σε περισσότερες από μία κλάσεις (ή και σε όλες), αλλά με διαφορετικό βαθμό συμμετοχής σε κάθε μία. Επομένως στο προηγούμενο παράδειγμα τα δύο παρόμοια δείγματα θα διέφεραν μόνο ελαφρώς στο βαθμό συμμετοχής τους σε σχέση με τις δύο γειτονικές κλάσεις. Επιπλέον, σε περιβάλλοντα ανάκλησης με βάση το περιεχόμενο, η ασαφής αναπαράσταση επιτρέπει στο χρήστη να πραγματοποιεί πολύπλοκες αναζητήσεις, όπως π.χ. να ψάχνει για *σχετικά* μεγάλα αντικείμενα μπλε χρώματος, που βρίσκονται *κοντά* στο κάτω μέρος της εικόνας.

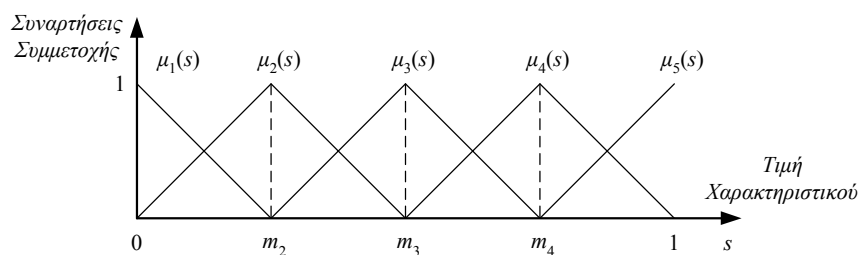
Σημειώνεται ότι η προτεινόμενη μέθοδος αναπαράστασης οπτικού περιεχομένου με διανύσματα περιγραφής που προκύπτουν από ασαφή πολυδιάστατα ιστογράμματα είναι εξολοκλήρου πρωτότυπη· παρουσιάστηκε για πρώτη φορά στην εργασία [32] και από τότε χρησιμοποιήθηκε για σκοπούς χαρακτηρισμού, εξαγωγής περίληψης και ανάκλησης στις εργασίες [13,14,33,36] και αργότερα στις [8,31,34,35,37,38,39].

4.2 ΜΟΝΟΔΙΑΣΤΑΤΗ ΤΑΞΙΝΟΜΗΣΗ

Ας θεωρήσουμε αρχικά την απλή περίπτωση ενός μονοδιάστατου χαρακτηριστικού s , που μπορεί να είναι π.χ. το μέγεθος ενός αντικειμένου. Το χαρακτηριστικό αυτό παίρνει τιμές σε ένα διάστημα το οποίο, χωρίς βλάβη της γενικότητας θεωρείται ότι είναι το $[0,1]$, δηλαδή το χαρακτηριστικό s είναι κανονικοποιημένο μεταξύ 0 και 1. Το διάστημα αυτό διαμερίζεται σε Q κλάσεις με τη χρήση Q *συναρτήσεων συμμετοχής* (membership functions) $\mu_n(s)$, $n = 1, 2, \dots, Q$. Για μία δεδομένη πραγματική τιμή s , το $\mu_n(s)$ υποδηλώνει το βαθμό συμμετοχής του χαρακτηριστικού s στη n -οστή κλάση. Οι συναρτήσεις συμμετοχής $\mu_n(s)$, $n = 1, 2, \dots, Q$ παίρνουν τιμές στο κλειστό διάστημα $[0,1]$, έτσι ώστε τιμές του $\mu_n(s)$ κοντά στη μονάδα (στο μηδέν) να δείχνουν ότι ο βαθμός συμμετοχής του χαρακτηριστικού s στη n -οστή κλάση είναι υψηλός (χαμηλός). Οι πλέον συνηθισμένες συναρτήσεις συμμετοχής είναι οι τριγωνικές, που ορίζονται ως

$$\mu_n(s) = \begin{cases} 1 - 2|s - m_n|/w, & |s - m_n| < w/2 \\ 0, & |s - m_n| \geq w/2 \end{cases} \quad (4-1)$$

για $n = 1, 2, \dots, Q$, όπου w είναι το πλάτος της βάσης κάθε τριγώνου και $m_n = (n - 1) / (Q - 1)$ είναι το κέντρο κάθε τριγώνου, έτσι ώστε $m_1 = 0$ και $m_Q = 1$. Ένα παράδειγμα ασαφούς ταξινόμησης με χρήση $Q = 5$ τριγωνικών συναρτήσεων συμμετοχής πλάτους $w = 2 / (Q - 1)$ παρουσιάζεται στο Σχήμα 4-1. Όπως φαίνεται στο Σχήμα αυτό, το πλάτος w ελέγχει το ποσοστό επικάλυψης μεταξύ διαδοχικών διαστημάτων ταξινόμησης, καθορίζοντας το πόσο ασαφής είναι η ταξινόμηση. Στο συγκεκριμένο παράδειγμα χρησιμοποιείται επικάλυψη 50%. Το ακριβές σχήμα και το ποσοστό επικάλυψης των συναρτήσεων $\mu_n(s)$ μπορεί να μεταβάλλεται σημαντικά [76]. Σύμφωνα όμως με πειραματικά αποτελέσματα, η επίδραση της επιλογής συναρτήσεων συμμετοχής στην εξαγωγή χαρακτηριστικών καρέ είναι πολύ περιορισμένη (εκτός από περιπτώσεις συνθετικών ακολουθιών εικόνων, όπως εξηγείται στην επόμενη Ενότητα). Επομένως επιλέχθηκαν τριγωνικές συναρτήσεις κυρίως λόγω των πολύ απλών υπολογισμών που απαιτούν.



Σχήμα 4-1. Παράδειγμα μονοδιάστατης ασαφούς ταξινόμησης με χρήση πέντε τριγωνικών συναρτήσεων συμμετοχής με ποσοστό επικάλυψης 50%.

Χρησιμοποιώντας τον παραπάνω τρόπο διαμέρισης, ένα ασαφές ιστόγραμμα μπορεί να κατασκευαστεί από ένα αριθμό δειγμάτων χαρακτηριστικών s_i , $i = 1, 2, \dots, K$. Κάθε δείγμα αντιστοιχεί σε ένα τμήμα εικόνας, όπου K είναι το πλήθος των τμημάτων. Η τιμή του ασαφούς ιστογράμματος, έστω $H(n)$, που αντιστοιχεί στην κλάση n , ορίζεται ως

$$H(n) = \frac{1}{K} \sum_{i=1}^K \mu_n(s_i), \quad n = 1, 2, \dots, Q \quad (4-2)$$

Πρέπει να παρατηρήσουμε ότι ο παραπάνω ορισμός είναι ισοδύναμος με τον ορισμό των παραδοσιακών ιστογραμμάτων όταν οι συναρτήσεις συμμετοχής παίρνουν μόνο δυαδικές τιμές (0 ή 1). Όμως, επειδή κάθε δείγμα μπορεί να έχει μη μηδενικό βαθμό συμμετοχής σε περισσότερες από μία κλάσεις, το ιστόγραμμα έχει νόημα ακόμη και όταν ο αριθμός των δειγμάτων είναι μικρός. Η ασαφής αναπαράσταση επομένως επιτρέπει την κατασκευή ιστογραμμάτων από ένα πολύ περιορισμένο σύνολο δεδομένων. Αυτό είναι πολύ σημαντικό αφού ο αριθμός αντικειμένων, K , σε μία εικόνα ή σε ένα καρέ από ακολουθία βίντεο, είναι συνήθως μικρότερος από το συνολικό αριθμό κλάσεων.

4.3 ΠΟΛΥΔΙΑΣΤΑΤΗ ΤΑΞΙΝΟΜΗΣΗ

Στη γενικότερη περίπτωση όπου χρησιμοποιούνται περισσότερα από ένα χαρακτηριστικά αντικειμένων, όπως μέγεθος, θέση, χρωματικές συνιστώσες, διανύσματα κίνησης και βάθος, κατασκευάζεται ένα πολυδιάστατο διάνυσμα χαρακτηριστικών τμήματος για κάθε αντικείμενο (τμήμα εικόνας). Επειδή έχουν προηγηθεί δύο διαφορετικές καταταμήσεις κάθε εικόνας, μία ως προς το χρώμα και μία ως προς την κίνηση, είναι διαθέσιμα δύο σύνολα τμημάτων: τα

τμήματα χρώματος S_i^c , $i = 1, 2, \dots, K^c$, και τα τμήματα κίνησης S_i^m , $i = 1, 2, \dots, K^m$, όπου K^c είναι το πλήθος τμημάτων χρώματος και K^m είναι το πλήθος τμημάτων κίνησης. Για κάθε τμήμα χρώματος S_i^c σχηματίζεται ένα διάνυσμα \mathbf{s}_i^c διαστάσεων $L^c \times 1$, ενώ για κάθε τμήμα κίνησης S_i^m σχηματίζεται ένα διάνυσμα \mathbf{s}_i^m διαστάσεων $L^m \times 1$:

$$\mathbf{s}_i^c = [\mathbf{c}^T (S_i^c) \mathbf{1}^T (S_i^c) a(S_i^c)]^T \quad (4-3)$$

$$\mathbf{s}_i^m = [\mathbf{v}^T (S_i^m) \mathbf{1}^T (S_i^m) a(S_i^m)]^T \quad (4-4)$$

όπου a είναι το μέγεθος ενός αντικειμένου, και $\mathbf{1}$ είναι ένα διάνυσμα διαστάσεων 2×1 που περιέχει τις συντεταγμένες του κέντρου βάρους του· το \mathbf{c} είναι ένα διάνυσμα διαστάσεων 3×1 που περιέχει τις τρεις μέσες χρωματικές συνιστώσες του, και τέλος \mathbf{v} είναι το μέσο διάνυσμα κίνησής του, διαστάσεων 2×1 . Έτσι, για τμήματα χρώματος έχουμε $L^c = 6$, ενώ για τμήματα κίνησης $L^m = 5$.

Στην περίπτωση των στερεοσκοπικών ακολουθιών όπου πραγματοποιείται και κατάτμηση με βάση το πεδίο βάθους, είναι διαθέσιμα και τα τμήματα βάθους S_i^d , $i = 1, 2, \dots, K^d$, όπου K^d είναι το πλήθος τους· έτσι για κάθε τμήμα βάθους S_i^d σχηματίζεται ένα διάνυσμα \mathbf{s}_i^d διαστάσεων $L^d \times 1$:

$$\mathbf{s}_i^d = [d(S_i^d) \mathbf{1}^T (S_i^d) a(S_i^d)]^T \quad (4-5)$$

όπου $d(S_i^d)$ είναι η μέση τιμή βάθους του τμήματος, και $L^d = 4$. Επιπλέον, αν πραγματοποιηθεί συγχώνευση τμημάτων χρώματος και βάθους, όπως αναφέρεται στο Κεφάλαιο 3, είναι δυνατός ο συνδυασμός των διανυσμάτων \mathbf{s}_i^c και \mathbf{s}_i^d σε ένα:

$$\mathbf{s}_i = [\mathbf{c}^T (S_i) d(S_i) \mathbf{1}^T (S_i) a(S_i)]^T \quad (4-6)$$

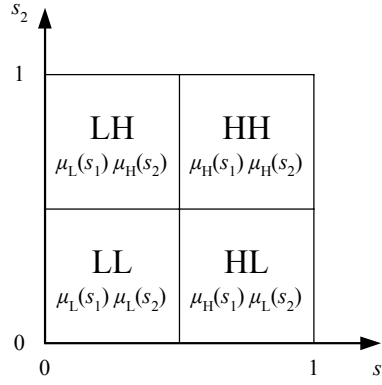
όπου τα τμήματα S_i , $i = 1, 2, \dots, K = K^d$ ορίζονται στην Εξίσωση (3-5). Η διαδικασία ταξινόμησης που παρουσιάζεται στη συνέχεια είναι ανεξάρτητη από το είδος των αντικειμένων. Επομένως, για απλότητα συμβολισμών, οι δείκτες c , m και d θα παραλείπονται στη συνέχεια· κάθε αντικείμενο, χρώματος, κίνησης ή βάθους, θα παριστάνεται ως S_i και θα περιγράφεται από το διάνυσμα \mathbf{s}_i , διαστάσεων $L \times 1$, όπου $L \in \{4, 5, 6, 7\}$, ανάλογα με τον τύπο του αντικειμένου. Επιπλέον το πλήθος τμημάτων θα συμβολίζεται με K , όπου $K = K^c$, $K = K^m$, ή $K = K^d$, και πάλι ανάλογα με τον τύπο των αντικειμένων.

Έστω ότι $\mathbf{s}_i = [s_{i,1} s_{i,2} \dots s_{i,L}]^T$, $i = 1, 2, \dots, K$, είναι το διάνυσμα που περιγράφει το αντικείμενο S_i , όπου K είναι το πλήθος των αντικειμένων σε μια εικόνα. Το πεδίο τιμών κάθε στοιχείου $s_{i,j}$, $j = 1, 2, \dots, L$, του διανύσματος \mathbf{s}_i διαμερίζεται σε Q περιοχές (κλάσεις) μέσω Q συναρτήσεων συμμετοχής $\mu_{n_j}(s_{i,j})$, $n_j = 1, 2, \dots, Q$. Όπως και στη μονοδιάστατη περίπτωση, για μια δεδομένη πραγματική τιμή του χαρακτηριστικού $s_{i,j}$, το $\mu_{n_j}(s_{i,j})$ υποδηλώνει το βαθμό συμμετοχής του στοιχείου $s_{i,j}$ στην κλάση με δείκτη n_j . Συγκεντρώνοντας τις κλάσεις με δείκτη n_j για όλα στοιχεία $j = 1, 2, \dots, L$, ορίζεται μία κλάση $\mathbf{n} = [n_1 n_2 \dots n_L]^T$ διάστασης L . Έτσι, ο βαθμός συμμετοχής κάθε διανύσματος \mathbf{s}_i στην κλάση \mathbf{n} μπορεί να ορισθεί ως το γινόμενο των συναρτήσεων συμμετοχής $\mu_{n_j}(s_{i,j})$ όλων των επιμέρους στοιχείων $s_{i,j}$ του \mathbf{s}_i στα αντίστοιχα στοιχεία n_j του \mathbf{n} :

$$\mu_{\mathbf{n}}(\mathbf{s}_i) = \prod_{j=1}^L \mu_{n_j}(s_{i,j}) \quad (4-7)$$

Για να ανήκει το διάνυσμα \mathbf{s}_i στην κλάση \mathbf{n} , θα πρέπει κάθε στοιχείο του, $s_{i,j}$, να ανήκει στην αντίστοιχη κλάση n_j . Επομένως οι συναρτήσεις συμμετοχής θα πρέπει να συνδυάζονται με τον τελεστή «ΚΑΙ», ο οποίος συνήθως υλοποιείται στα ασαφή συστήματα με την πράξη του πολλαπλασιασμού.

Ένα απλό παράδειγμα διδιάστατων διανυσμάτων απεικονίζεται στο Σχήμα 4-2. Έστω ότι ένα αντικείμενο S περιγράφεται από το διάνυσμα $\mathbf{s} = [s_1 \ s_2]^T$, και ότι χρησιμοποιούνται $Q = 2$ συναρτήσεις συμμετοχής $\mu_1(s_j)$ και $\mu_2(s_j)$ για να διαμερίσουν το πεδίο τιμών και των δύο στοιχείων $s_j, j = 1, 2$, του \mathbf{s} . Αφού η συνάρτηση $\mu_1(s_j)$ εκφράζει τις «χαμηλές» (low) τιμές του s_j και η $\mu_2(s_j)$ τις «υψηλές» (high), μπορούμε να ονομάσουμε τις δύο κλάσεις s_j “L” και “H”, και τις δύο συναρτήσεις συμμετοχής $\mu_L(s_j)$ και $\mu_H(s_j)$ αντίστοιχα. Οι διδιάστατες κλάσεις $\mathbf{n} = [n_1 \ n_2]^T$ μπορούν τότε να ονομάζονται “LL”, “LH”, “HL” και “HH”, ενώ ο βαθμός συμμετοχής του διανύσματος \mathbf{s} στην κλάση \mathbf{n} είναι $\mu_{\mathbf{n}}(\mathbf{s}) = \mu_{n_1}(s_1)\mu_{n_2}(s_2)$, ή, υπολογίζοντας όλους τους συνδυασμούς, $\mu_{LL}(\mathbf{s}) = \mu_L(s_1) \mu_L(s_2)$, $\mu_{LH}(\mathbf{s}) = \mu_L(s_1) \mu_H(s_2)$, $\mu_{HL}(\mathbf{s}) = \mu_H(s_1) \mu_L(s_2)$, και $\mu_{HH}(\mathbf{s}) = \mu_H(s_1) \mu_H(s_2)$.



Σχήμα 4-2. Παράδειγμα διδιάστατης ασαφούς ταξινόμησης με χρήση δύο συναρτήσεων συμμετοχής για κάθε διάσταση.

Με βάση τους παραπάνω ορισμούς, είναι δυνατό να κατασκευασθεί ένα πολυδιάστατο ασαφές ιστόγραμμα από τα δείγματα (αντικείμενα) των χαρακτηριστικών $\mathbf{s}_i, i = 1, 2, \dots, K$, ακριβώς όπως στη μονοδιάστατη περίπτωση. Η τιμή του ασαφούς ιστογράμματος, $H(\mathbf{n})$, ορίζεται με παρόμοιο τρόπο ως η μέση τιμή, για όλα τα δείγματα, των αντίστοιχων βαθμών συμμετοχής $\mu_{\mathbf{n}}(\mathbf{s}_i)$:

$$H(\mathbf{n}) = \frac{1}{K} \sum_{i=1}^K \mu_{\mathbf{n}}(\mathbf{s}_i) = \frac{1}{K} \sum_{i=1}^K \prod_{j=1}^L \mu_{n_j}(s_{i,j}) \quad (4-8)$$

Η τιμή του ιστογράμματος μπορεί λοιπόν να θεωρηθεί ως ο βαθμός συμμετοχής μιας ολόκληρης εικόνας (ή καρτέ) στην κλάση \mathbf{n} .

4.4 ΔΙΑΝΥΣΜΑΤΑ ΠΕΡΙΓΡΑΦΗΣ

Συγκεντρώνοντας τις τιμές $H(\mathbf{n})$ του πολυδιάστατου ιστογράμματος για όλες τις κλάσεις \mathbf{n} , δηλαδή για όλους τους συνδυασμούς δεικτών n_1, n_2, \dots, n_L , το διάνυσμα περιγραφής μιας εικόνας ή καρέ ορίζεται ως ένα διάνυσμα \mathbf{f} διαστάσεων $Q^L \times 1$ που περιέχει με κατάλληλη σειρά τις παραπάνω τιμές:

$$\mathbf{f} = [f_1 \ f_2 \ \dots \ f_{Q^L}]^T \quad (4-9)$$

Για τον υπολογισμό των επιμέρους στοιχείων $f_i, i = 1, 2, \dots, Q^L$, του διανύσματος περιγραφής, ορίζεται μία *συνάρτηση δεικτών* $z(\mathbf{n})$ που απεικονίζει κάθε μία από τις Q^L κλάσεις σε έναν ακέραιο μεταξύ 1 και Q^L :

$$z(\mathbf{n}) = 1 + \sum_{j=1}^L n_j Q^{L-j} \quad (4-10)$$

Χρησιμοποιώντας τη συνάρτηση δεικτών, τα στοιχεία του διανύσματος περιγραφής \mathbf{f} υπολογίζονται ως $f_{z(\mathbf{n})} = H(\mathbf{n})$ για όλες τις κλάσεις \mathbf{n} .

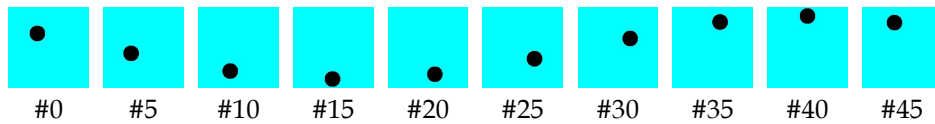
Επανερχόμενοι τώρα στον αρχικό ορισμό των διανυσμάτων περιγραφής των επιμέρους τμημάτων που περιέχονται σε μία εικόνα, προκύπτει ότι όλη η παραπάνω ανάλυση επαναλαμβάνεται δύο φορές: μία για τα τμήματα χρώματος S_i^c , που περιγράφονται από τα διανύσματα $\mathbf{s}_i^c, i = 1, 2, \dots, K^c$, και μία για τα τμήματα κίνησης S_i^m , που περιγράφονται αντίστοιχα από τα διανύσματα $\mathbf{s}_i^m, i = 1, 2, \dots, K^m$. Έτσι σχηματίζονται δύο ειδών διανύσματα περιγραφής εικόνων: το διάνυσμα περιγραφής χρώματος \mathbf{f}^c για τα τμήματα χρώματος και το διάνυσμα περιγραφής κίνησης \mathbf{f}^m για τα τμήματα κίνησης. Το συνολικό διάνυσμα περιγραφής, \mathbf{f} , μήκους $Q^L + Q^L$, που περιγράφει μία ολόκληρη εικόνα ή καρέ, σχηματίζεται συνενώνοντας τα διανύσματα \mathbf{f}^c και \mathbf{f}^m :

$$\mathbf{f} = [(\mathbf{f}^c)^T \ (\mathbf{f}^m)^T]^T \quad (4-11)$$

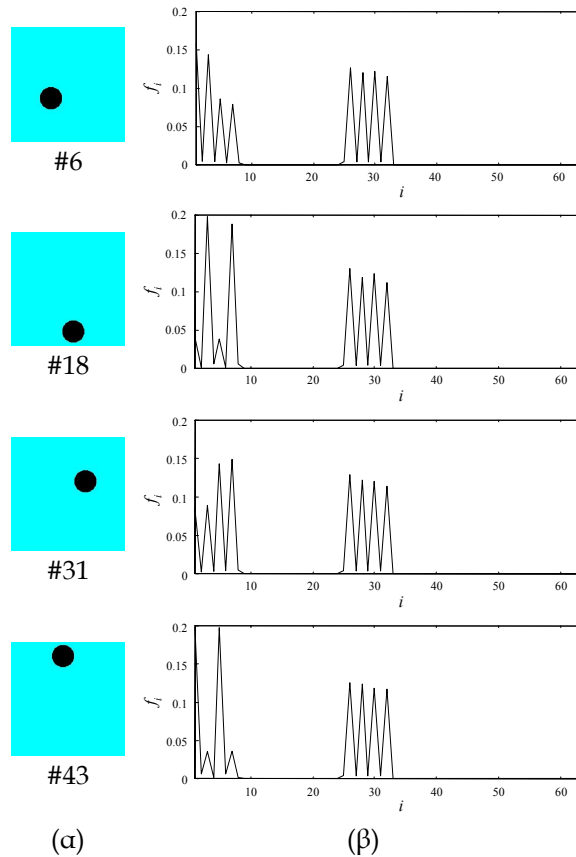
Επιπλέον, στην περίπτωση των στερεοσκοπικών ακολουθιών, είναι διαθέσιμο και το διάνυσμα περιγραφής βάθους \mathbf{f}^d , μήκους Q^{L^d} , το οποίο μπορεί επίσης να συνδυαστεί με τα \mathbf{f}^c και \mathbf{f}^m στην κατασκευή του συνολικού διανύσματος περιγραφής. Πρέπει να σημειωθεί ότι η διάσταση του διανύσματος περιγραφής, και συνεπώς η υπολογιστική πολυπλοκότητα της προτεινόμενης μεθόδου, αυξάνει εκθετικά ως προς το πλήθος Q των περιοχών διαμέρισης. Επιπλέον, ένας μεγάλος αριθμός τέτοιων περιοχών δεν βελτιώνει υποχρεωτικά την αποτελεσματικότητα της περιγραφής εικόνων, και επομένως της εξαγωγής περίληψης ή της ανάκλησης με βάση το περιεχόμενο. Απεναντίας, οδηγεί σε πολύ μεγάλο αριθμό κλάσεων και συνεπώς σε «θορυβώδη» ταξινόμηση, ενώ συνήθως τα περισσότερα στοιχεία του διανύσματος περιγραφής έχουν μηδενική τιμή. Τα πειράματα οδηγούν στο συμπέρασμα ότι μια λογική επιλογή ως προς την πολυπλοκότητα και την αποτελεσματικότητα είναι $Q = 3$.

Εκτός από τα χαρακτηριστικά των τμημάτων μιας εικόνας ή καρέ που προκύπτουν από τη διαδικασία κατάτμησης, μπορούν να συμπεριληφθούν στο διάνυσμα περιγραφής και *ολικά χαρακτηριστικά* (global features) που προκύπτουν από την ανάλυση της εικόνας ως σύνολο. Παραδείγματα τέτοιων χαρακτηριστικών είναι το συνολικό *ιστόγραμμα χρώματος* κάθε εικόνας ή η μέση *πολυπλοκότητα υφής* που μπορεί να εκτιμηθεί από τους συντελεστές DCT υψηλών συ-

χνοτήτων των blocks μιας ακολουθίας MPEG. Επιπρόσθετες ιδιότητες των τμημάτων μπορούν επίσης να συμπεριληφθούν στο διάνυσμα περιγραφής, όπως το σχήμα των περιγραμμάτων των αντικειμένων ή οι κεντρικές ροπές μεγάλου βαθμού. Σε αυτή την περίπτωση, το διάνυσμα περιγραφής θα αντιπροσώπευε καλύτερα το οπτικό περιεχόμενο των εικόνων, επιτρέποντας κάθε φορά επιλογή εκείνων των χαρακτηριστικών που θεωρούνται σημαντικότερα για την εξαγωγή περίληψης ή την ανάκληση με βάση το περιεχόμενο. Πρέπει όμως να σημειωθεί ότι η αναπαράσταση του οπτικού περιεχομένου με βάση τα διανύσματα περιγραφής είναι ανεξάρτητη από τους αλγόριθμους επιλογής χαρακτηριστικών πλάνων ή καρέ. Έτσι οποιαδήποτε μεταβολή στα διανύσματα περιγραφής μπορεί να γίνει χωρίς να επηρεάσει τους αλγόριθμους αυτούς.



Σχήμα 4-3. Συνθετική ακολουθία βίντεο μήκους 50 καρέ που χρησιμοποιείται για τα παραδείγματα των διανυσμάτων περιγραφής.



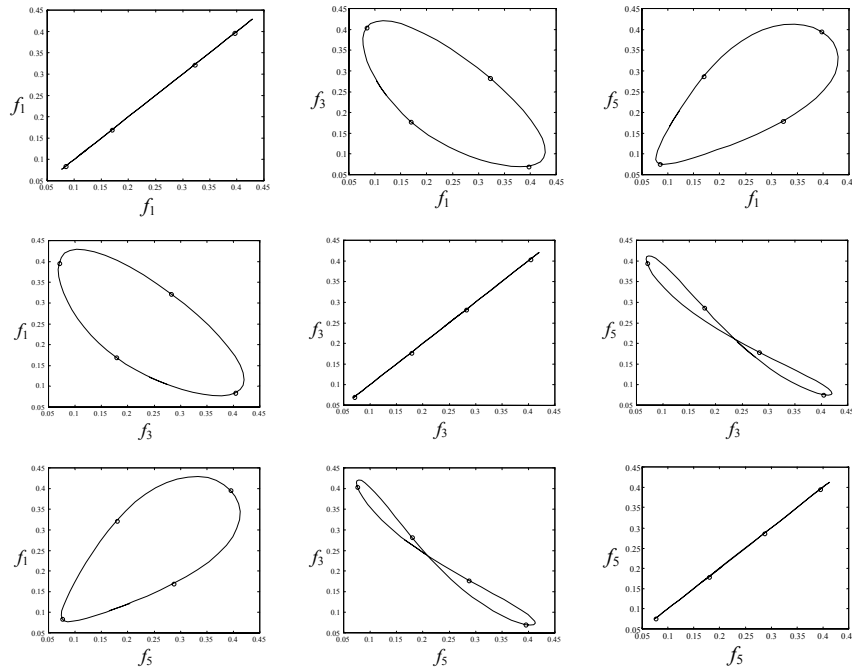
Σχήμα 4-4. (α) Τέσσερα επιλεγμένα καρέ της συνθετικής ακολουθίας του Σχήματος 4-3. (β) Διαγράμματα των συντελεστών f_i , $i = 1, 2, \dots, 64$ των αντίστοιχων διανυσμάτων περιγραφής.

4.5 ΠΑΡΑΔΕΙΓΜΑ ΣΥΝΘΕΤΙΚΗΣ ΑΚΟΛΟΥΘΙΑΣ

Για να γίνει περισσότερο κατανοητός ο τρόπος με τον οποίο τα προτεινόμενα διανύσματα περιγραφής αναπαριστούν την οπτική πληροφορία, δίνεται στη συνέχεια ένα παράδειγμα που προέρχεται από μία συνθετική ακολουθία βίντεο. Η ακολουθία αποτελείται από 50 καρέ ανάλυσης 256×256 pixels και απεικονίζει ένα συμπαγή μαύρο κύκλο ακτίνας 25 pixels που κινεί-

ται σε μια κατακόρυφη ελλειψοειδή τροχιά, μπροστά από ένα στατικό φόντο γαλάζιου χρώματος. Στο Σχήμα 4-3 φαίνονται 10 καρέ της ακολουθίας, των οποίων οι αριθμοί είναι ομοιόμορφα κατανεμημένοι μεταξύ 0 και 49 (#0, #5, #10, ..., #45). Για απλότητα θεωρούμε ότι το διάνυσμα περιγραφής περιέχει χαρακτηριστικά χρώματος μόνο και όχι κίνησης. Έστω λοιπόν ότι το διάνυσμα περιγραφής του καρέ με αριθμό k , $k = 0, 1, \dots, 49$, είναι $\mathbf{f}(k) = \mathbf{f}^c(k)$. Το πεδίο τιμών κάθε χαρακτηριστικού διαμερίζεται σε $Q = 2$ περιοχές με χρήση δύο τριγωνικών συναρτήσεων συμμετοχής με 50% επικάλυψη, και επομένως το μήκος του διανύσματος περιγραφής είναι $Q^{L^c} = 2^6 = 64$. Έτσι οι δείκτες των κλάσεων είναι $n_i \in \{1, 2\}$, $i = 1, 2, \dots, 6$, όπου το $n_i = 1$ αντιστοιχεί σε μια «χαμηλή» (“low”) τιμή, ενώ το $n_i = 2$ σε μια «υψηλή» (“high”) τιμή.

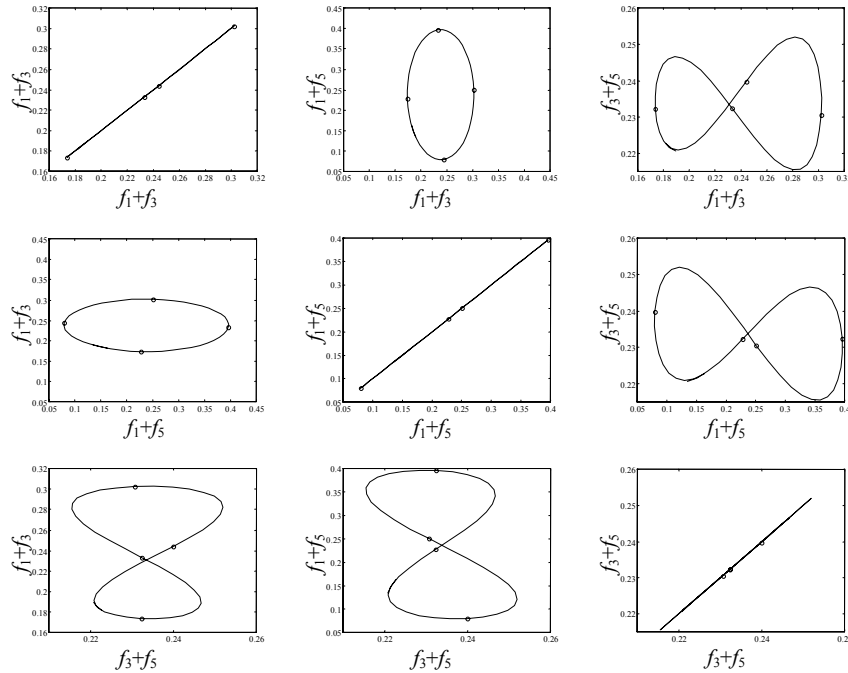
Το Σχήμα 4-4 απεικονίζει τέσσερα από τα καρέ της ακολουθίας, και συγκεκριμένα τα #6, #18, #31 και #43, μαζί με μία παράσταση των 64 στοιχείων του αντίστοιχου διανύσματος $\mathbf{f}(k)$. Στην πραγματικότητα τα καρέ αυτά έχουν επιλεγεί με τη μέθοδο εξαγωγής χαρακτηριστικών καρέ της Ενότητας 6.2. Μία πρώτη παρατήρηση είναι ότι το διάνυσμα περιγραφής μεταβάλλεται σε κάθε καρέ, παρά το γεγονός ότι η χρωματική σύνθεση της ακολουθίας παραμένει σταθερή. Αυτό είναι αναμενόμενο αφού το διάνυσμα περιγραφής περιέχει επίσης και γεωμετρικά χαρακτηριστικά των τμημάτων, και συγκεκριμένα το μέγεθος καθώς και την οριζόντια και κατακόρυφη θέση του κέντρου βάρους τους, οι οποίες μεταβάλλονται.



Σχήμα 4-5. Διδιάστατα γραφήματα όλων των συνδυασμών ζευγών από τα στοιχεία f_1 , f_3 και f_5 του διανύσματος περιγραφής.

Παρατηρείται επίσης ότι μόνο δύο ομάδες από στοιχεία του διανύσματος περιγραφής έχουν μη μηδενική τιμή. Η πρώτη ομάδα αντιστοιχεί στο μαύρο κύκλο ενώ η δεύτερη στην περιοχή του φόντου. Στο συγκεκριμένο παράδειγμα, οι δείκτες κλάσεων n_1 , n_2 και n_3 αντιστοιχούν στις χρωματικές συνιστώσες R , G και B , οι n_4 και n_5 στην οριζόντια και κατακόρυφη θέση (x και y , αντίστοιχα), ενώ ο n_6 στο μέγεθος κάθε αντικειμένου. Έτσι, οι κλάσεις \mathbf{n} της πρώτης ομάδας, για τις οποίες η συνάρτηση δεικτών $z(\mathbf{n})$ παίρνει τιμές στο σύνολο $\{1, 2, \dots, 8\}$, αντιστοιχούν σε «χαμηλές» τιμές κόκκινου, πράσινου και μπλε (μαύρος κύκλος). Παρομοίως, οι

κλάσεις \mathbf{n} της δεύτερης ομάδας, για τις οποίες η $z(\mathbf{n})$ παίρνει τιμές $\{25, 26, \dots, 32\}$, αντιστοιχούν σε «χαμηλές» τιμές κόκκινου, αλλά «υψηλές» τιμές πράσινου και μπλε (γαλάζιο φόντο). Η δεύτερη ομάδα που οφείλεται στο φόντο παραμένει ουσιαστικά σταθερή σε όλη τη διάρκεια της ακολουθίας, ενώ η πρώτη ομάδα που οφείλεται στο αντικείμενο του κύκλου μεταβάλλεται, αφού το αντικείμενο αυτό μετακινείται. Επιπλέον, οι άρτιες τιμές της $z(\mathbf{n})$ αντιστοιχούν σε «μεγάλα» αντικείμενα, ενώ οι περιττές σε «μικρά». Έτσι η ομάδα του φόντου έχει μηδενικές τιμές στις περιττές θέσεις δεικτών, ενώ η ομάδα του κύκλου στις άρτιες.



Σχήμα 4-6. Διδιάστατα γραφήματα όλων των συνδυασμών ζευγών από τα αθροίσματα στοιχείων $f_1 + f_3$, $f_1 + f_5$, και $f_3 + f_5$.

Στο Σχήμα 4-5 παρουσιάζονται γραφήματα των στοιχείων f_1 , f_3 και f_5 του διανύσματος περιγραφής. Κάθε γράφημα απεικονίζει ένα ζεύγος από τους παραπάνω συντελεστές και έχει το πρώτο μέλος του ζεύγους στον οριζόντιο άξονα και το δεύτερο στον κατακόρυφο. Κάθε σημείο ενός γραφήματος αντιστοιχεί σε ένα καρέ της ακολουθίας· έτσι παρουσιάζεται η χρονική μεταβολή του ζεύγους στοιχείων. Οι μικροί κύκλοι που εμφανίζονται στα γραφήματα αυτά – όπως και στα γραφήματα του Σχήματος 4-6 παρακάτω – αντιστοιχούν στην επιλογή χαρακτηριστικών καρέ με τη μέθοδο χρονικής μεταβολής και εξηγούνται στο Κεφάλαιο 6 Τα στοιχεία f_1 , f_3 και f_5 αντιστοιχούν σε ίδιες τιμές ως προς τους δείκτες n_1 , n_2 , n_3 και n_6 (δηλαδή συνιστώσες R , G , B και μέγεθος), ενώ διαφέρουν μεταξύ τους μόνο ως προς τους δείκτες n_4 και n_5 (θέση αντικειμένου x και y). Συγκεκριμένα, το f_1 αντιστοιχεί σε «χαμηλό» x και «χαμηλό» y (LL), το f_3 σε «χαμηλό» x και «υψηλό» y (LH), και το f_5 σε «υψηλό» x και «χαμηλό» y (HL), όμοια με το παράδειγμα διδιάστατης ασαφούς ταξινόμησης της προηγούμενης Ενότητας. Ουσιαστικά τα παραπάνω γραφήματα αποτελούν προβολές της τροχιάς του πολυδιάστατου διανύσματος περιγραφής στον υπόχωρο που ορίζεται από τα δύο επιλεγμένα στοιχεία.

Τέλος, στο Σχήμα 4-6 παρουσιάζονται γραφήματα παρόμοια με εκείνα του Σχήματος 4-5, αλλά για τα αθροίσματα στοιχείων $f_1 + f_3$, $f_1 + f_5$, και $f_3 + f_5$. Αφού το στοιχείο f_1 αντιστοιχεί σε «χαμηλό» y και το f_3 σε «υψηλό» y , με την άθροισή τους αφαιρείται η επίδραση της κατακό-

ρυφης θέσης y , και επομένως το άθροισμα $f_1 + f_3$ αναφέρεται σε απλώς σε «χαμηλό» x και είναι ανεξάρτητο από το y . Όμοια, το άθροισμα $f_1 + f_5$ αναφέρεται σε «χαμηλό» y και είναι ανεξάρτητο από το x . Το γράφημα που απεικονίζει το άθροισμα $f_1 + f_5$ ως προς το $f_1 + f_3$ παρουσιάζει ενδιαφέρον καθώς στην πραγματικότητα εκεί απεικονίζεται η κατακόρυφη θέση y του κυκλικού αντικειμένου ως προς την οριζόντια θέση x . Σε αυτή την περίπτωση επιβεβαιώνεται ότι η πληροφορία της ελλειπτικής τροχιάς του αντικειμένου έχει διατηρηθεί στο διάγραμμα περιγραφής.

Επιλογή Χαρακτηριστικών Πλάνων

Με την ανάλυση μιας ακολουθίας σε αντικείμενα σύμφωνα με τις μεθόδους κατάτμησης του Μέρους I και την παραγωγή διανυσμάτων περιγραφής για κάθε καρέ της ακολουθίας, είναι δυνατή η εξαγωγή περίληψης για την προεπισκόπηση καθώς και τη συμπαγή αναπαράσταση της ακολουθίας με σκοπό την αποτελεσματικότερη ανάκληση με βάση το περιεχόμενο. Στο παρόν Κεφάλαιο παρουσιάζεται το πρώτο στάδιο εξαγωγής περίληψης, το οποίο περιλαμβάνει την επιλογή ενός περιορισμένου αριθμού χαρακτηριστικών πλάνων. Για το σκοπό αυτό πραγματοποιείται αρχικά ανίχνευση αλλαγής πλάνων, χρονική κατάτμηση της ακολουθίας σε ένα αριθμό από διαδοχικά πλάνα καθώς και χαρακτηρισμός των πλάνων με τον προσδιορισμό ενός συνολικού διανύσματος περιγραφής για κάθε πλάνο. Η επιλογή χαρακτηριστικών πλάνων υλοποιείται με τη ομαδοποίηση πλάνων με παρόμοιο οπτικό περιεχόμενο σε ένα προκαθορισμένο αριθμό ομάδων και την επιλογή ενός αντιπροσωπευτικού πλάνου από κάθε ομάδα. Οι συγκρίσεις πραγματοποιούνται στο χώρο των διανυσμάτων περιγραφής, ενώ για τη βέλτιστη ομαδοποίηση χρησιμοποιείται ο γενικευμένος αλγόριθμος Lloyd-Max ή K-means. Η τεχνική αξιολογείται σε πραγματικές ακολουθίες βίντεο από δελτία ειδήσεων, ενώ παρουσιάζονται και πειραματικά αποτελέσματα στερεοσκοπικών ακολουθιών· στην περίπτωση αυτή αξιολογείται η βελτίωση στην ποιότητα της περίληψης σε σχέση με το απαιτούμενο επιπρόσθετο υπολογιστικό κόστος.

5.1 ΑΝΙΧΝΕΥΣΗ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΠΛΑΝΩΝ

Το πρώτο στάδιο για την αποτελεσματική εξαγωγή περίληψης μιας ακολουθίας βίντεο με σκοπό την προεπισκόπησης ή την αναζήτηση με βάση το περιεχόμενο είναι η ανάλυση της σε πλάνα και η επιλογή ενός συνόλου από χαρακτηριστικά πλάνα. Ένα πλάνο (shot) είναι ένα χρονικό τμήμα μιας ακολουθίας το οποίο αντιστοιχεί σε συνεχή λειτουργία μίας και μόνο κάμερας. Επειδή κάθε ακολουθία αποτελείται από ένα μεγάλο αριθμό πλάνων και οι χρονικές στιγμές αρχής και τέλους του κάθε πλάνου δεν είναι γνωστές, απαιτείται αρχικά *ανίχνευση αλλαγής πλάνων* (shot cut detection). Η διαδικασία αυτή έχει σκοπό τον εντοπισμό των χρονικών στιγμών, δηλαδή των καρέ, κατά τις οποίες τελειώνει ένα πλάνο και αρχίζει το επόμενο. Ο εντοπισμός αυτός πρέπει να γίνεται με αυτόματο τρόπο, με βάση το οπτικό περιεχόμενο της ακολουθίας. Επειδή οι αλλαγές πλάνων μπορεί να είναι απότομες (shot cut) ή πολλές φορές και συνεχείς, με ομοιόμορφη μεταβολή του περιεχομένου (*shot fading* ή *dissolve*), η ανίχνευση αλλαγής πλάνων δεν είναι πάντα εύκολη.

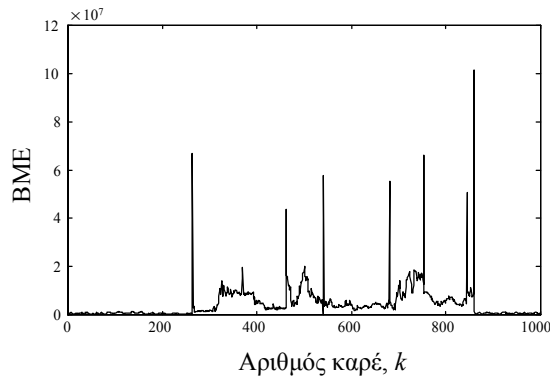
Έχουν προταθεί αρκετοί αλγόριθμοι στη βιβλιογραφία για αυτόματη ανίχνευση αλλαγής πλάνων, είτε σε ασυμπίεστες είτε απευθείας σε συμπίεσμένες ακολουθίες [97,149]. Οι αλγόριθμοι αυτοί αντιμετωπίζουν συνήθως απότομες αλλαγές (cuts), η ανίχνευση των οποίων θεωρείται πιο εύκολη, ή και αλλαγές τύπου fading ή dissolve. Στην περίπτωση των στερεοσκοπικών ακολουθιών, επειδή οι αλλαγές πλάνων συμβαίνουν την ίδια χρονική στιγμή και στα δύο κα-

νάλια, η ανίχνευση αλλαγής πλάνου στο ένα μόνο κανάλι είναι επαρκής και δεν απαιτείται ειδικός αλγόριθμος.

Στο πλαίσιο των πειραμάτων της διατριβής η ανίχνευση αλλαγής πλάνων βασίζεται στο σφάλμα εκτίμησης κίνησης (block motion estimation error, BME), το οποίο είναι διαθέσιμο στις ακολουθίες MPEG, ή προκύπτει ως ενδιάμεσο αποτέλεσμα της διαδικασίας εκτίμησης κίνησης. Το άθροισμα του σφάλματος εκτίμησης κίνησης σε όλα τα blocks ενός καρέ δίνει ένα μέτρο απόστασης του καρέ αυτού από το προηγούμενό του και μάλιστα δεν επηρεάζεται από τις κινήσεις των αντικειμένων, σε αντίθεση με τις τεχνικές που βασίζονται σε διαφορές pixel προς pixel. Η προτεινόμενη τεχνική ανιχνεύει μόνο απότομες αλλαγές πλάνων, ενώ το υπολογιστικό της κόστος είναι πολύ μικρό αν έχει ήδη προηγηθεί εκτίμηση κίνησης. Σε ακολουθίες MPEG οι αλλαγές πλάνων μπορούν επίσης να ανιχνευθούν με βάση την κατανομή των bits σε καρέ τύπου P και B [32].



Σχήμα 5-1. Ακολουθία τηλεοπτικών ειδήσεων μήκους 1000 καρέ, που περιέχει συνολικά οκτώ πλάνα.



Σχήμα 5-2. Σφάλμα εκτίμησης κίνησης συναρτήσει του χρόνου (αριθμού καρέ) για την ακολουθία του Σχήματος 5-1.



Σχήμα 5-3. Τα οκτώ πλάνα της ακολουθίας του Σχήματος 5-1. Παρουσιάζεται το πρώτο καρέ του κάθε πλάνου.

Ένα παράδειγμα ανίχνευσης πλάνων με την προτεινόμενη μέθοδο δίνεται για μια ακολουθία από πρόγραμμα τηλεοπτικών ειδήσεων συνολικής διάρκειας 40 δευτερολέπτων (1000 καρέ με ρυθμό 25 frames/sec). Η ακολουθία παρουσιάζεται στο Σχήμα 5-1, όπου εικονίζεται ένα καρέ κάθε 40. Αποτελείται συνολικά από οκτώ πλάνα: ένα πλάνο του παρουσιαστή στο

στούντιο, στη συνέχεια έξι εξωτερικά πλάνα αθλητικού ρεπορτάζ από αγώνα μπάσκετ και τέλος ακόμη ένα πλάνο με τον παρουσιαστή.

Το γράφημα του σφάλματος εκτίμησης κίνησης συναρτήσει του χρόνου για την ακολουθία αυτή φαίνεται στο Σχήμα 5-2. Οι χρονικές στιγμές αλλαγής πλάνου εντοπίζονται με χρήση κατωφλίσωσης στο γράφημα του σφάλματος εκτίμησης κίνησης. Μάλιστα το τελευταίο καρέ του αθλητικού ρεπορτάζ έχει πολύ μικρή διάρκεια (14 καρέ) και δεν διακρίνεται στο Σχήμα 5-1. Επίσης είναι εμφανής από το γράφημα η εντονότερη δράση (μεγαλύτερη κίνηση) στα εξωτερικά πλάνα σε σχέση με το στούντιο. Τα οκτώ πλάνα της ακολουθίας παρουσιάζονται στο Σχήμα 5-3, όπου εικονίζεται το πρώτο καρέ του κάθε πλάνου. Σημειώνεται ότι η ανίχνευση αλλαγής πλάνων είναι επιτυχής παρά τις γρήγορες κινήσεις της κάμερας στα πλάνα του αθλητικού ρεπορτάζ.

Το επόμενο βήμα μετά την ανίχνευση των πλάνων είναι ο *χαρακτηρισμός* τους, δηλαδή η κατασκευή ενός διανύσματος περιγραφής για την αναπαράσταση κάθε πλάνου. Για το σκοπό αυτό όλα τα καρέ του κάθε πλάνου αναλύονται σε αντικείμενα με βάση το χρώμα και την κίνηση (καθώς και το πεδίο βάθους για στερεοσκοπικές ακολουθίες), και στη συνέχεια παράγονται διανύσματα περιγραφής των καρέ σύμφωνα με τη διαδικασία του Κεφαλαίου 4. Το διάνυσμα περιγραφής ενός πλάνου ορίζεται ως η μέση τιμή των διανυσμάτων περιγραφής όλων των καρέ του πλάνου. Εναλλακτικά, μπορεί να γίνει πρώτα επιλογή χαρακτηριστικών καρέ και η μέση τιμή να υπολογισθεί για τα χαρακτηριστικά καρέ μόνο. Με τον τρόπο αυτό επιτυγχάνεται καλύτερη αναπαράσταση του οπτικού περιεχομένου των πλάνων.

5.2 ΟΜΑΔΟΠΟΙΗΣΗ ΠΛΑΝΩΝ

Με δεδομένα τα διανύσματα περιγραφής των πλάνων μιας ακολουθίας πραγματοποιείται βέλτιστη εξαγωγή των πιο χαρακτηριστικών (αντιπροσωπευτικών) πλάνων. Κάτι τέτοιο επιτυγχάνεται με την *ομαδοποίηση* (clustering) των πλάνων παρόμοιου οπτικού περιεχομένου σε ένα περιορισμένο αριθμό *ομάδων* (clusters) και την επιλογή ενός αντιπροσωπευτικού πλάνου από κάθε ομάδα. Η διαδικασία ομαδοποίησης υλοποιείται με τον αλγόριθμο Lloyd-Max ή K-means. Η εξαγωγή περίληψης μέσω ομαδοποίησης έχει ξαναχρησιμοποιηθεί [3,74], αλλά η επιλογή χαρακτηριστικών πλάνων μέσω ομαδοποίησης με τον αλγόριθμο Lloyd-Max [46] είναι πρωτότυπη· προτάθηκε αρχικά στην εργασία [14] και χρησιμοποιήθηκε επίσης στις [31,33,36,39].

Έστω ότι \mathbf{h}_i , $i = 1, 2, \dots, N_S$, είναι το διάνυσμα περιγραφής του i -οστού πλάνου, μήκους $Q^L + Q^M$, όπου N_S είναι ο συνολικός αριθμός πλάνων της ακολουθίας. Έτσι ορίζεται το σύνολο $E = \{\mathbf{h}_i, i = 1, 2, \dots, N_S\}$ όλων των διανυσμάτων περιγραφής πλάνων. Έστω ότι το πλήθος χαρακτηριστικών πλάνων που πρέπει να επιλεγούν είναι K_S , ενώ με \mathbf{q}_i , $i = 1, 2, \dots, K_S$, συμβολίζονται K_S αυθαίρετα διανύσματα του χώρου \mathcal{R}^M . Για κάθε \mathbf{q}_i σχηματίζεται μία *ζώνη επιρροής* (influence zone), έστω Z_i , η οποία περιέχει όλα τα διανύσματα περιγραφής πλάνων $\mathbf{h} \in E$ που είναι πλησιέστερα στο \mathbf{q}_i :

$$Z_i = \{\mathbf{h} \in E: \|\mathbf{h} - \mathbf{q}_i\| < \|\mathbf{h} - \mathbf{q}_j\| \ \forall j \neq i\}, \quad i = 1, 2, \dots, K_S \quad (5-1)$$

όπου $\|\cdot\|$ είναι η Ευκλείδεια απόσταση δύο διανυσμάτων. Με χρήση των ζωνών επιρροής επιτυγχάνεται *ομαδοποίηση* των διανυσμάτων περιγραφής, αφού σε κάθε διάνυσμα \mathbf{q}_i αντιστοιχεί

ένα σύνολο (ομάδα) διανυσμάτων περιγραφής. Για κάθε ομάδα το αντίστοιχο διάνυσμα \mathbf{q}_i ονομάζεται *κέντρο ομάδας* (cluster center). Τότε η μέση παραμόρφωση (distortion) των διανυσμάτων \mathbf{q}_i , $i = 1, 2, \dots, K_S$, ορίζεται ως το άθροισμα αποστάσεων των διανυσμάτων περιγραφής από το κέντρο της ομάδας τους, για όλες τις ομάδες:

$$R_S(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{K_S}) = \sum_{i=1}^{K_S} \sum_{\mathbf{h} \in Z_i} \|\mathbf{h} - \mathbf{q}_i\| \quad (5-2)$$

Η μέση παραμόρφωση είναι ένα μέτρο αξιολόγησης της αναπαράστασης των διανυσμάτων περιγραφής πλάνων από τα κέντρα ομάδων \mathbf{q}_i . Επομένως τα βέλτιστα διανύσματα $\hat{\mathbf{q}}_i$ υπολογίζονται με ελαχιστοποίηση της παραμόρφωσης R_S :

$$(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_{K_S}) = \arg \min_{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{K_S} \in \mathfrak{R}^M} R_S(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{K_S}) \quad (5-3)$$

Με άλλα λόγια, τα βέλτιστα κέντρα ομάδων θα πρέπει να ορίζονται με τέτοιο τρόπο, ώστε να έχουν την ελάχιστη απόσταση (και επομένως τη μεγαλύτερη ομοιότητα) με τα διανύσματα περιγραφής των ομάδων τους. Όμως η ελαχιστοποίηση του μέτρου παραμόρφωσης είναι μία πολύπλοκη διαδικασία γιατί οι άγνωστες παράμετροι \mathbf{q}_i , $i = 1, 2, \dots, K_S$, εμπλέκονται και στις αποστάσεις $\|\cdot\|$ αλλά και στις ζώνες επιρροής Z_i της Εξίσωσης (5-2). Για το λόγο αυτό η ελαχιστοποίηση πραγματοποιείται αναδρομικά με χρήση του γενικευμένου αλγορίθμου Lloyd-Max, ή *K-means* [46]. Συγκεκριμένα, ξεκινώντας από αυθαίρετα αρχικά κέντρα ομάδων $\mathbf{q}_i(0)$, $i = 1, 2, \dots, K_S$, τα επόμενα κέντρα υπολογίζονται με τις παρακάτω Εξισώσεις για την επανάληψη $n \geq 0$:

$$Z_i(n) = \{\mathbf{h} \in E: \|\mathbf{h} - \mathbf{q}_i(n)\| < \|\mathbf{h} - \mathbf{q}_j(n)\| \ \forall j \neq i\}, \quad i = 1, 2, \dots, K_S \quad (5-4)$$

$$\mathbf{q}_i(n+1) = \text{cent}(Z_i(n)), \quad i = 1, 2, \dots, K_S \quad (5-5)$$

όπου $\mathbf{q}_i(n)$ είναι το i -οστό κέντρο ομάδας στην n -οστή επανάληψη, και $Z_i(n)$ είναι η ζώνη επιρροής του. Το κέντρο, $\text{cent}(\cdot)$, της ζώνης $Z_i(n)$, υπολογίζεται ως η μέση τιμή των διανυσμάτων περιγραφής που ανήκουν στη ζώνη:

$$\text{cent}(Z_i(n)) = \frac{1}{|Z_i(n)|} \sum_{\mathbf{h} \in Z_i(n)} \mathbf{h} \quad (5-6)$$

όπου $|Z_i(n)|$ είναι το πλήθος στοιχείων της ζώνης $Z_i(n)$. Επαναλαμβάνοντας τη αναδρομική διαδικασία των Εξισώσεων (5-4) και (5-5), ο αλγόριθμος συγκλίνει στη λύση $(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_{K_S})$ έπειτα από μικρό αριθμό επαναλήψεων. Τελικά, επειδή δεν υπάρχει εγγύηση ότι τα βέλτιστα κέντρα ομάδων $(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_{K_S})$ αντιστοιχούν σε πραγματικά διανύσματα περιγραφής πλάνων, τα K_S χαρακτηριστικά πλάνα της ακολουθίας επιλέγονται ως εκείνα τα πλάνα των οποίων τα διανύσματα περιγραφής είναι πλησιέστερα στα $(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_{K_S})$.

Πρέπει να σημειωθεί ότι όσο το πλήθος χαρακτηριστικών πλάνων K_S αυξάνεται, η μέση παραμόρφωση R_S μειώνεται, αφού τα διανύσματα περιγραφής πλάνων είναι πλησιέστερα στα κέντρα των ομάδων πλάνων. Ειδικότερα στην οριακή περίπτωση $K_S = N_S$, όπου διατηρούνται όλα τα πλάνα της ακολουθίας ως χαρακτηριστικά, κάθε πλάνο παριστά μία ομάδα και επομένως η παραμόρφωση R_S μηδενίζεται. Αντίθετα, μικρές τιμές του K_S είναι συνήθως επιθυμητές για τη μείωση των απαιτήσεων αποθήκευσης και την εξαγωγή αποδοτικής περίληψης. Οι μικρές τιμές του K_S όμως οδηγούν σε μεγάλη παραμόρφωση και επομένως σε απώλεια πληρο-

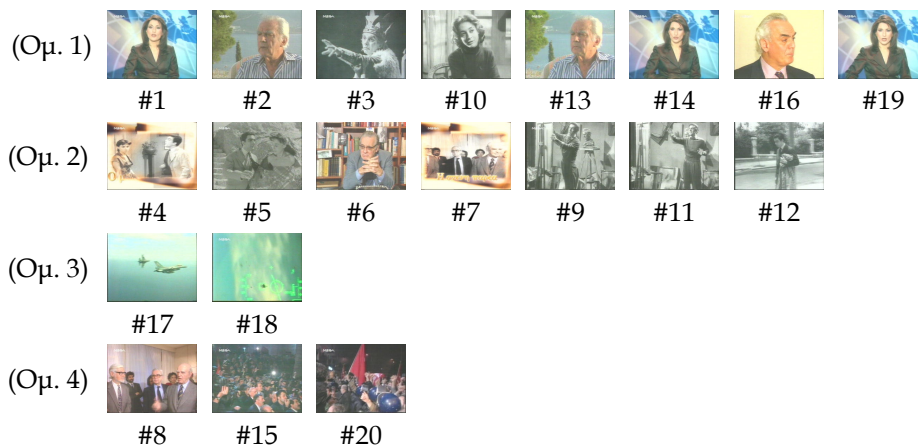
φορίας, αφού μικρός αριθμός πλάνων δε μπορεί να περιγράψει ικανοποιητικά το οπτικό περιεχόμενο της ακολουθίας. Η βέλτιστη τιμή του K_S εκτιμάται με χρήση ενός κριτηρίου που βασίζεται στη θεωρία πληροφορίας, και συγκεκριμένα το *minimum description length* (MDL) [144].

5.3 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Η προτεινόμενη μέθοδος επιλογής αντιπροσωπευτικών πλάνων δοκιμάστηκε σε μία ακολουθία συνολικής διάρκειας 2.5 λεπτών (3750 καρέ) που προέρχεται από τηλεοπτικό δελτίο ειδήσεων. Με την ανίχνευση αλλαγής πλάνων η ακολουθία αναλύθηκε αρχικά σε πλάνα και στη συνέχεια μέσω κατάτμησης χρώματος και κίνησης υπολογίσθηκαν τα διανύσματα περιγραφής των όλων των καρέ του κάθε πλάνου, καθώς και τα συνολικά διανύσματα περιγραφής των πλάνων. Για την κατασκευή διανυσμάτων περιγραφής, το πεδίο τιμών κάθε χαρακτηριστικού διαμερίστηκε σε $Q = 2$ κλάσεις μέσω τριών τριγωνικών συναρτήσεων συμμετοχής με 50% επικάλυψη. Έτσι το συνολικό μήκος των διανυσμάτων περιγραφής είναι $Q^{L^c} + Q^{L^m} = 96$ για $L_c = 6$ και $L_m = 5$.



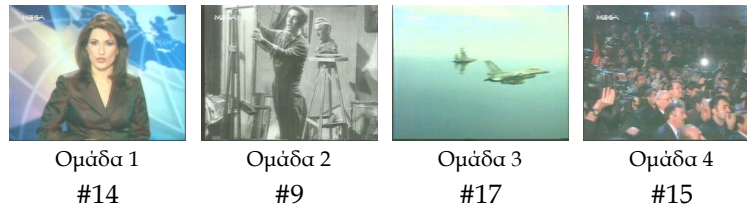
Σχήμα 5-4. Ακολουθία τηλεοπτικών ειδήσεων, διάρκειας 2.5 λεπτών (3750 καρέ), που περιέχει συνολικά $N_S = 20$ πλάνα. Παρουσιάζεται ένα καρέ για κάθε πλάνο.



Σχήμα 5-5. Ομάδες πλάνων για την ακολουθία του Σχήματος 5-4.

Η ακολουθία αποτελείται από $N_S = 20$ πλάνα, τα οποία φαίνονται στο Σχήμα 5-4. Με στόχο την αποτελεσματική παρουσίαση, για κάθε πλάνο εικονίζεται ένα καρέ, του οποίου το διάνυσμα περιγραφής είναι το πλησιέστερο στο διάνυσμα περιγραφής του όλου πλάνου, επομένως το καρέ αυτό αντιπροσωπεύει καλύτερα το οπτικό περιεχόμενο του πλάνου. Από το κρι-

τήριο MDL το βέλτιστο πλήθος χαρακτηριστικών πλάνων προκύπτει $K_S = 4$. Έτσι, τα 20 πλάνα ομαδοποιούνται με τον προτεινόμενο αλγόριθμο σε τέσσερις ομάδες, οι οποίες απεικονίζονται στο Σχήμα 5-5. Κάθε ομάδα περιέχει τα πλάνα των οποίων τα διανύσματα περιγραφής είναι πλησιέστερα στο κέντρο της ομάδας.



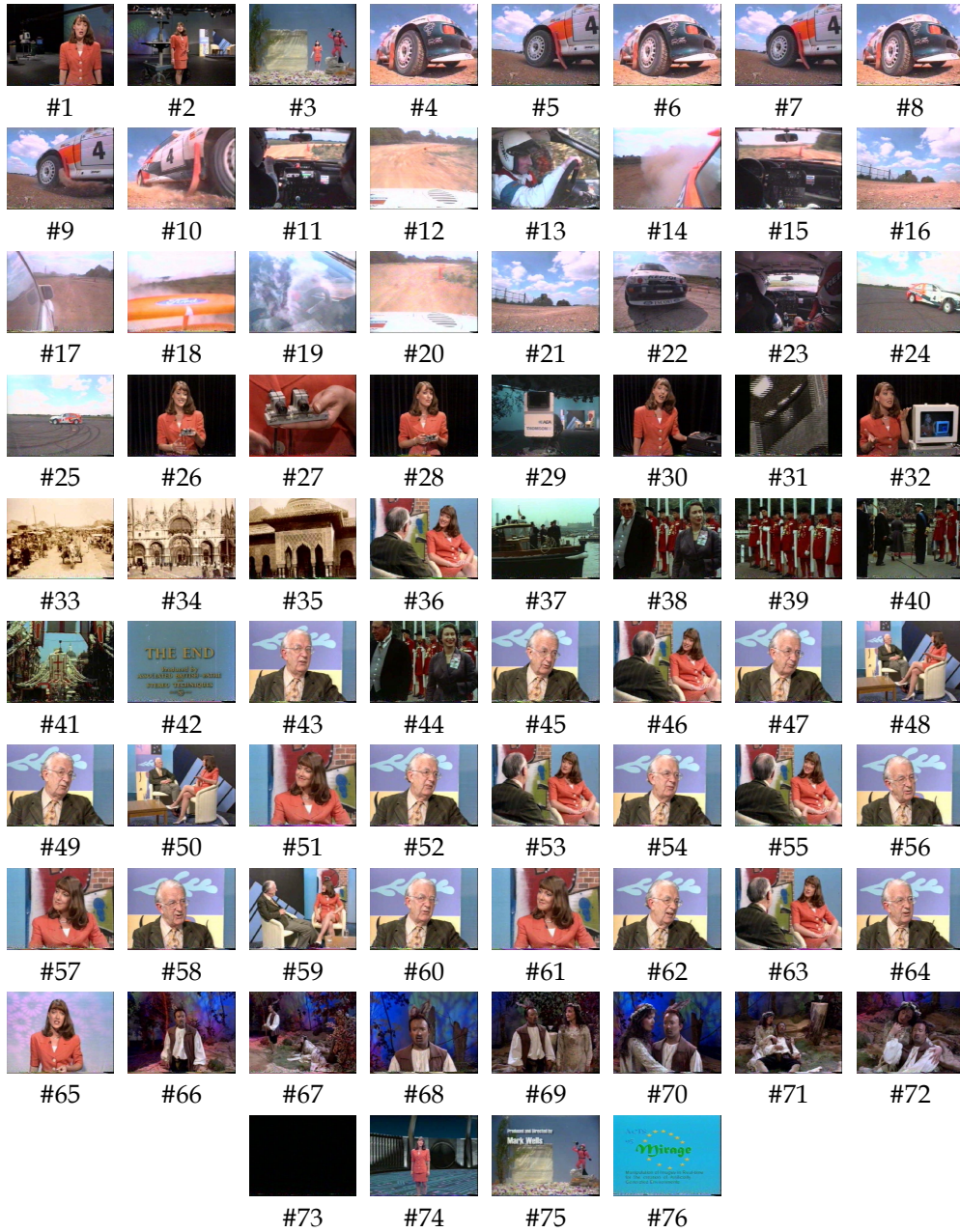
Σχήμα 5-6. Χαρακτηριστικά πλάνα για την ακολουθία του Σχήματος 5-4.

Τα τέσσερα επιλεγμένα χαρακτηριστικά πλάνα εικονίζονται στο Σχήμα 5-6. Πρόκειται για τα πλάνα εκείνα των οποίων τα διανύσματα περιγραφής είναι πλησιέστερα στα βέλτιστα κέντρα ομάδων που προκύπτουν από τον αλγόριθμο ομαδοποίησης. Είναι εμφανές ότι τα τέσσερα επιλεγμένα πλάνα δίνουν ικανοποιητική αναπαράσταση του οπτικού περιεχομένου της ακολουθίας χωρίς να περιέχουν περιττή πληροφορία. Παρατηρείται επίσης ότι κάθε ομάδα περιέχει πλάνα με παρόμοιες ιδιότητες, όπως το πλήθος και η πολυπλοκότητα των αντικειμένων.

5.4 ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΕ ΣΤΕΡΕΟΣΚΟΠΙΚΕΣ ΑΚΟΛΟΥΘΙΕΣ

Για την αξιολόγηση της προτεινόμενης μεθόδου εξαγωγής περίληψης σε στερεοσκοπικές ακολουθίες χρησιμοποιήθηκε το στερεοσκοπικό τηλεοπτικό πρόγραμμα “Eye to Eye” [118], συνολικής διάρκειας 25 λεπτών (12739 καρέ με ρυθμό 10 frames/sec). Η παραγωγή του προγράμματος έγινε στο πλαίσιο του έργου ACTS MIRAGE [47] σε συνεργασία με την AEA Technology και το Independent Television Commission (ITC). Οι λήψεις των εσωτερικών πλάνων πραγματοποιήθηκαν με τη στερεοσκοπική μονάδα στούντιο που αναπτύχθηκε από την AEA Technology από κοινού με την Thomson Multimedia στο πλαίσιο του παλαιότερου έργου RACE DISTIMA [153], ενώ οι λήψεις των εξωτερικών πλάνων με ειδική στερεοσκοπική κάμερα ελαφριάς και γερής κατασκευής που αναπτύχθηκε για το ITC από την AEA Technology.

Στην ακολουθία εφαρμόζεται αρχικά στερεοσκοπική ανάλυση και για κάθε ζεύγος εικόνων (καρέ) υπολογίζεται ένα πεδίο βάθους. Στη συνέχεια πραγματοποιείται κατάτμηση με βάση το χρώμα του αριστερού καναλιού και το πεδίο βάθους, με χρήση του αλγορίθμου M-RSST. Τα τμήματα χρώματος και βάθους συγχωνεύονται για την εξαγωγή αντικειμένων και καταγράφονται τα χαρακτηριστικά του κάθε αντικειμένου, συμπεριλαμβανομένων του μεγέθους, της θέσης, του χρώματος, του βάθους και της κίνησης. Για την κατασκευή διανυσμάτων περιγραφής των καρέ, το πεδίο τιμών κάθε χαρακτηριστικού διαμερίζεται σε $Q = 3$ κλάσεις μέσω τριών τριγωνικών συναρτήσεων συμμετοχής με 50% επικάλυψη. Έτσι το συνολικό μήκος των διανυσμάτων περιγραφής είναι $Q^L = 2,187$ καθώς στη συγκεκριμένη περίπτωση χρησιμοποιείται το σύνθετο διάνυσμα περιγραφής της Εξίσωσης (4-6) και επομένως $L = 7$. Στη συνέχεια ανιχνεύονται οι αλλαγές πλάνων και υπολογίζονται διανύσματα περιγραφής πλάνων. Όλες οι παραπάνω πληροφορίες χαρακτηρισμού της ακολουθίας αποθηκεύονται σε βάση δεδομένων και είναι διαθέσιμες για τη διεξαγωγή πειραμάτων επιλογής αντιπροσωπευτικών πλάνων και καρέ.

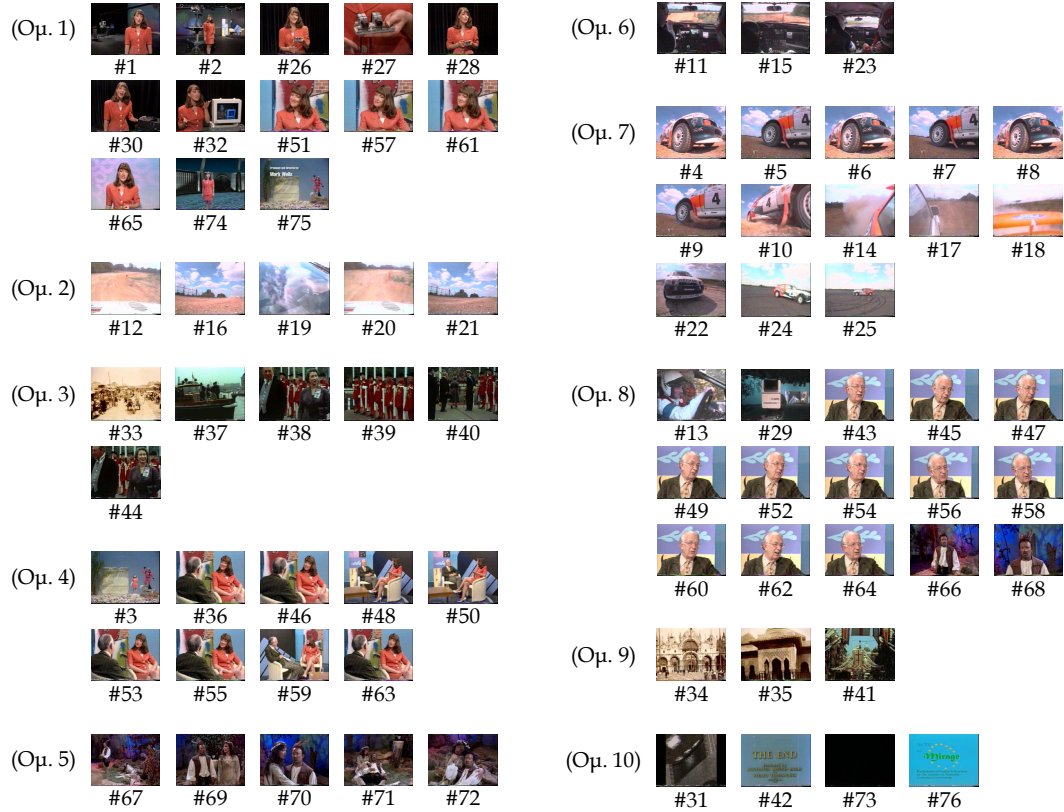


Σχήμα 5-7. Η στερεοσκοπική ακολουθία “Eye to Eye”, διάρκειας 25 λεπτών, η οποία περιέχει $N_S = 76$ πλάνα. Παρουσιάζεται ένα καρέ για κάθε πλάνο.

Η ακολουθία “Eye to Eye” αποτελείται συνολικά από $N_S = 76$ πλάνα τα οποία φαίνονται στο Σχήμα 5-7. Όπως και στο Σχήμα 5-4, για κάθε πλάνο εικονίζεται ένα καρέ, του οποίου το διάνυσμα περιγραφής είναι το πλησιέστερο στο διάνυσμα περιγραφής του όλου πλάνου, επομένως το καρέ αυτό αντιπροσωπεύει καλύτερα το οπτικό περιεχόμενο του πλάνου. Με χρήση του κριτηρίου MDL, το βέλτιστο πλήθος χαρακτηριστικών πλάνων προκύπτει $K_S = 10$. Με τον προτεινόμενο αλγόριθμο τα 76 πλάνα ομαδοποιούνται σε 10 ομάδες, οι οποίες απεικονίζονται στο Σχήμα 5-8.

Όπως παρατηρείται, η πλειοψηφία των πλάνων που περιέχουν παρόμοιο οπτικό περιεχόμενο, όσον αφορά στο πλήθος και στην πολυπλοκότητα των αντικειμένων, ανήκουν στην ίδια ομάδα πλάνων. Επίσης, στο Σχήμα 5-9 παρουσιάζονται τα δέκα αντιπροσωπευτικά πλάνα των ομάδων, δηλαδή εκείνα τα πλάνα των οποίων τα διανύσματα περιγραφής είναι πλη-

σιέστερα στα κέντρα των ομάδων. Είναι εμφανές ότι τα δέκα επιλεγμένα πλάνα αποτελούν ικανοποιητική περίληψη του οπτικού περιεχομένου της όλης ακολουθίας, συνολικής διάρκειας 25 λεπτών. Το σύνολο των χαρακτηριστικών πλάνων δίνει έτσι τη δυνατότητα παραγωγής περιληπτικών βίντεο κλιπ μικρής διάρκειας και χαμηλής ανάλυσης για σκοπούς προεπισκόπησης.



Σχήμα 5-8. Ομάδες πλάνων για την ακολουθία "Eye to Eye", με συγχώνευση τμημάτων χρώματος και βάθους.



Σχήμα 5-9. Χαρακτηριστικά πλάνα για την ακολουθία "Eye to Eye", με συγχώνευση τμημάτων χρώματος και βάθους.

Για να αξιολογηθεί το επιπρόσθετο όφελος από τη στερεοσκοπική ανάλυση, την εκτίμηση βάθους και τη συγχώνευση αντικειμένων στη διαδικασία εξαγωγής περίληψης, πραγματοποιείται ένα δεύτερο πείραμα στο οποίο χρησιμοποιείται μόνο η πληροφορία χρώματος του ενός καναλιού. Συγκεκριμένα, πραγματοποιείται μόνο κατάτμηση με βάση το χρώμα, με είσοδο την εικόνα αριστερού καναλιού για κάθε καρέ, χωρίς εκτίμηση βάθους ή συγχώνευση αντικειμένων. Τα διανύσματα περιγραφής που παράγονται από τα νέα τμήματα χρώματος χρησιμοποιούνται εκ νέου για την ομαδοποίηση των πλάνων σε $K_S = 10$ ομάδες. Τα αποτελέσματα του πειράματος αυτού παρουσιάζονται στο Σχήμα 5-10.

Είναι ξεκάθαρο ότι ο διαχωρισμός πλάνων ως προς το οπτικό περιεχόμενο δεν είναι το ίδιο επιτυχημένος όπως στην περίπτωση της στερεοσκοπικής ανάλυσης, όπου χρησιμοποιείται η πληροφορία του βάθους για πιο αξιόπιστη εξαγωγή αντικειμένων. Για παράδειγμα, όλα τα

πλάνα που απεικονίζουν την παρουσιάστρια του προγράμματος ταξινομούνται στη ίδια ομάδα πλάνων (ομάδα 1) με χρήση του πεδίου βάθους. Αντίθετα, με κατάτμηση χρώματος μόνο, τα πλάνα αυτά ταξινομούνται σε περισσότερες από μία ομάδες (κυρίως στις ομάδες 7, 8 και 9). Με όμοιο τρόπο, τα αντιπροσωπευτικά πλάνα των ομάδων με κατάτμηση χρώματος μόνο παρουσιάζονται στο Σχήμα 5-11. Παρατηρείται ότι τα επιλεγμένα χαρακτηριστικά πλάνα περιέχουν πολλαπλές εμφανίσεις του ίδιου οπτικού περιεχομένου, ενώ αντίθετα ορισμένα χαρακτηριστικά πλάνα της ακολουθίας και μάλιστα μεγάλης διάρκειας δεν έχουν ένα αντι-στοιχο αντιπρόσωπο ομάδας.



Σχήμα 5-10. Ομάδες πλάνων για την ακολουθία “Eye to Eye”, με κατάτμηση χρώματος μόνο.



Σχήμα 5-11. Χαρακτηριστικά πλάνα για την ακολουθία “Eye to Eye”, με κατάτμηση χρώματος μόνο.

Πρέπει επίσης να σημειωθεί ότι το επιπλέον υπολογιστικό κόστος για την εκτίμηση της απόκλισης και την κατάτμηση με βάση το βάθος δεν αποτελεί σοβαρό περιοριστικό παράγοντα. Για παράδειγμα, η επεξεργασία με κατάτμηση χρώματος μόνο απαιτεί 4.37 sec/frame σε υπολογιστή Sun Ultra 10 (333 MHz) για ανάλυση καρέ 352×264 pixels. Αντίθετα η επεξεργασία με κατάτμηση βάθους και συγχώνευση απαιτεί 9.91 sec/frame, συμπεριλαμβανομένης και της κατάτμησης χρώματος. Οι παραπάνω χρόνοι επεξεργασίας είναι οι μέσες τιμές για όλα τα καρέ της ακολουθίας.

Για επιτάχυνση της στερεοσκοπικής ανάλυσης πραγματοποιήθηκε στα παραπάνω πειράματα υποδειγματοληψία των εικόνων των δύο καναλιών για εκτίμηση και κατάτμηση βάθους (χρησιμοποιήθηκαν δηλαδή blocks διαστάσεων 2×2 pixels). Σημειώνεται ότι με επιπλέον υποδειγματοληψία (δηλαδή 4×4 blocks) ο συνολικός χρόνος επεξεργασίας μειώνεται στα 5.85 sec/frame, χρόνος συγκρίσιμος με την επεξεργασία ενός καναλιού με κατάτμηση χρώματος μόνο. Η υποδειγματοληψία δεν επηρεάζει σημαντικά την εξαγωγή αντικειμένων, αφού τα ακριβή περιγράμματα των αντικειμένων παράγονται ουσιαστικά από την κατάτμηση χρώματος. Επιπλέον μείωση του υπολογιστικού κόστους για την εκτίμηση του πεδίου απόκλισης μπορεί να επιτευχθεί με τον αλγόριθμο που προτάθηκε πρόσφατα στην εργασία [129]. Έτσι, συμπεραίνεται ότι όφελος από τη χρήση της πληροφορίας βάθους δικαιολογεί το απαιτούμενο υπολογιστικό κόστος.

Σε κάποιες περιπτώσεις βέβαια οι διαφορές απόκλισης μεταξύ αντικειμένων στη χρησιμοποιούμενη ανάλυση εικόνας είναι μικρές, έχοντας ως αποτέλεσμα λανθασμένη εκτίμηση βάθους. Αυτό συμβαίνει ιδιαίτερα στα μακρινά πλάνα, όπου οι διαφορές βάθους είναι πολύ μικρές για να εκτιμηθούν με ακρίβεια. Κάτι τέτοιο όμως δεν είναι πολύ συχνό, ειδικά στις στερεοσκοπικές ακολουθίες, στην παραγωγή των οποίων συνήθως δίνεται έμφαση στην πληροφορία βάθους. Συγκεκριμένα, στην ακολουθία “Eye to Eye”, 532 μόνο από τα 12,739 καρέ ανιχνεύονται ως καρέ χωρίς σημαντική πληροφορία βάθους, δηλαδή 4.18% της συνολικής ακολουθίας. Τα καρέ αυτά ανιχνεύονται εύκολα γιατί παράγουν ένα μόνο τμήμα βάθους, ενώ η επεξεργασία τους περιλαμβάνει κατάτμηση χρώματος μόνο, δηλαδή η πληροφορία βάθους αγνοείται.

Επιλογή Χαρακτηριστικών Καρέ με Βάση τη Χρονική Μεταβολή

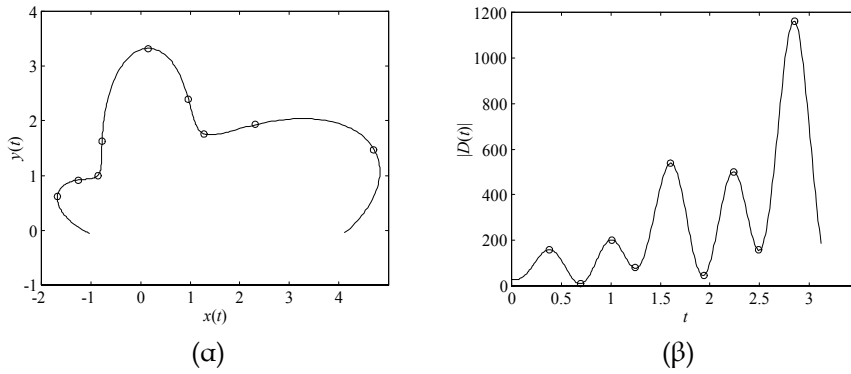
Το δεύτερο στάδιο εξαγωγής περίληψης από μία ακολουθία βίντεο είναι η επιλογή ενός περιορισμένου συνόλου χαρακτηριστικών καρέ τα οποία να δίνουν ικανοποιητική περιγραφή του οπτικού περιεχομένου της ακολουθίας. Ουσιαστικά η επιλογή χαρακτηριστικών καρέ αποτελεί χρονική δειγματοληψία της ακολουθίας, η οποία όμως είναι μη ομοιόμορφη και βασίζεται στο περιεχόμενο. Η επιλογή χαρακτηριστικών καρέ μπορεί να γίνει είτε στο σύνολο των χαρακτηριστικών πλάνων είτε σε ολόκληρη την ακολουθία· η πρώτη τεχνική είναι περισσότερο κατάλληλη για προεπισκόπηση ενώ η δεύτερη για ανάκληση με βάση το περιεχόμενο. Στο πλαίσιο της διατριβής προτείνονται δύο μέθοδοι επιλογής χαρακτηριστικών καρέ. Η πρώτη βασίζεται στη χρονική μεταβολή του διάνυσματος περιγραφής και παρουσιάζεται στο παρόν Κεφάλαιο ενώ η δεύτερη βασίζεται στην ελαχιστοποίηση ενός κριτηρίου συσχέτισης και παρουσιάζεται στο Κεφάλαιο 7. Πιο συγκεκριμένα, η μέθοδος χρονικής μεταβολής βασίζεται στην παρατήρηση ότι το διάνυσμα περιγραφής διαδοχικών καρέ ως συνάρτηση του χρόνου σχηματίζει μία τροχιά σε ένα πολυδιάστατο χώρο· επομένως η επιλογή χαρακτηριστικών καρέ είναι ισοδύναμη με την επιλογή κατάλληλων σημείων της τροχιάς που να περιγράφουν το σχήμα της. Ως τέτοια επιλέγονται τα ακρότατα σημεία της καμπύλης της τροχιάς, ενώ η επιλογή υλοποιείται με τη μεγιστοποίηση ενός μέτρου καμπυλότητας της τροχιάς, αφού πρώτα η τροχιά εξομαλυνθεί για την αφαίρεση του θορύβου που οφείλεται σε σφάλματα κατάτμησης. Η λειτουργία της μεθόδου χρονικής μεταβολής επιδεικνύεται στη συνθετική ακολουθία του Κεφαλαίου 4 καθώς και σε πραγματικές ακολουθίες.

6.1 ΤΡΟΧΙΑ ΔΙΑΝΥΣΜΑΤΟΣ ΠΕΡΙΓΡΑΦΗΣ

Αφού κάθε καρέ σε ένα πλάνο αντιστοιχεί σε μία συγκεκριμένη χρονική στιγμή και χαρακτηρίζεται από ένα συγκεκριμένο διάνυσμα περιγραφής, τα διανύσματα περιγραφής όλων των καρέ σχηματίζουν μία τροχιά σε ένα πολυδιάστατο χώρο η οποία εκφράζει τη χρονική μεταβολή του οπτικού περιεχομένου του πλάνου. Επομένως, η επιλογή των *χαρακτηριστικών καρέ* (key frames) μέσα σε ένα πλάνο είναι ισοδύναμη με την επιλογή ενός μικρού συνόλου σημείων της τροχιάς τα οποία είναι σε θέση να περιγράψουν τη μορφή της. Τα επιλεγμένα σημεία θα πρέπει να παρέχουν αρκετή πληροφορία σχετικά με το σχήμα της τροχιάς, έτσι ώστε αυτή να μπορεί να αναπαραχθεί από τα σημεία μέσω κάποιου είδους παρεμβολής. Αυτό μπορεί να επιτευχθεί με την εξαγωγή των χρονικών στιγμών, δηλαδή των αριθμών των καρέ, που βρίσκονται σε ακραίες θέσεις της τροχιάς. Για την εξαγωγή αυτών των χρονικών στιγμών προτείνεται η χρήση ενός *μέτρου καμπυλότητας* της τροχιάς, και συγκεκριμένα του *μέτρου της δεύτερης παραγώγου* του διανύσματος περιγραφής ως προς το χρόνο.

Η δεύτερη παράγωγος ουσιαστικά εκφράζει το μέγεθος και την κατεύθυνση της επιτάχυνσης ή επιβράδυνσης ενός αντικειμένου που κινείται κατά μήκος της τροχιάς. Τα τοπικά μέγιστα του μέτρου της παραγώγου αντιστοιχούν σε χρονικές στιγμές μέγιστης μεταβολής της τα-

χύτητας του αντικειμένου, δηλαδή σε ακραία σημεία της τροχιάς. Όμοια τα τοπικά ελάχιστα αντιστοιχούν σε χρονικές στιγμές όπου η ταχύτητα είναι σχεδόν σταθερή, δηλαδή σε σχεδόν ευθύγραμμα τμήματα της τροχιάς που βρίσκονται μεταξύ των ακραίων σημείων. Έτσι η επιλογή των χαρακτηριστικών καρέ γίνεται ακριβώς στα παραπάνω σημεία των τοπικών μεγίστων και ελαχίστων. Για παράδειγμα, έστω ότι ένα διδιάστατο διάνυσμα περιγραφής μεταβάλλεται χρονικά σχηματίζοντας τη συνεχή καμπύλη $\mathbf{r}(t) = (x(t), y(t))$ όπως φαίνεται στο Σχήμα 6-1. Τότε τα σημεία που αντιστοιχούν στα τοπικά μέγιστα και ελάχιστα του μέτρου της δεύτερης παραγώγου, $D(t)$, και που απεικονίζονται στα Σχήματα 6-1α,β ως κύκλοι, παρέχουν επαρκή πληροφορία για το σχήμα της καμπύλης $\mathbf{r}(t)$, αφού όπως φαίνεται η τελευταία μπορεί να αναπαραχθεί με ικανοποιητική ακρίβεια από τα επιλεγμένα σημεία με χρήση κάποιου είδους παρεμβολής. Η παραπάνω διαδικασία εξαγωγής χαρακτηριστικών καρέ ονομάζεται *μέθοδος χρονικής μεταβολής*. Στη συνέχεια δίνεται αναλυτική περιγραφή της μεθόδου, χρησιμοποιώντας διακριτό χρόνο για την αναπαράσταση των καρέ.



Σχήμα 6-1. (α) Μία διδιάστατη συνεχής καμπύλη $\mathbf{r}(t) = (x(t), y(t))$. (β) Το μέτρο της δεύτερης παραγώγου της καμπύλης, $D(t)$, συναρτήσει του χρόνου t .

Η ιδέα της εξαγωγής χαρακτηριστικών καρέ έχει προταθεί παλαιότερα στην εργασία [5], ενώ η επιλογή τους με βάση τις ιδιότητες της πολυδιάστατης καμπύλης τροχιάς του διανύσματος περιγραφής έχει προταθεί στην [30]. Όμως η χρήση του μέτρου καμπυλότητας της τροχιάς για το σκοπό αυτό στο πλαίσιο της διατριβής είναι πρωτότυπη. Η τεχνική αυτή δημοσιεύθηκε αρχικά στην εργασία [32], ενώ χρησιμοποιήθηκε και στις [8,14].

6.2 ΜΕΘΟΔΟΣ ΧΡΟΝΙΚΗΣ ΜΕΤΑΒΟΛΗΣ

Ας υποθέσουμε ότι ένα πλάνο μιας ακολουθίας βίντεο αποτελείται από N_F καρέ, και έστω ότι $\mathbf{f}(k)$, $k = 0, 1, \dots, N_F - 1$, είναι το διάνυσμα περιγραφής του k -οστού καρέ, όπως ορίζεται στην Εξίσωση (4-11). Στο διακριτό χρόνο που χρησιμοποιείται για την αναπαράσταση των καρέ, η παράγωγος του $\mathbf{f}(k)$ ως προς τον αύξοντα αριθμό καρέ k εκτιμάται ως η διαφορά διανυσμάτων περιγραφής δύο διαδοχικών καρέ:

$$\mathbf{d}_1(k) = \mathbf{f}(k+1) - \mathbf{f}(k), \quad k \in \{0, 1, \dots, N_F - 2\} \quad (6-1)$$

Όμως η παραπάνω διαφορά είναι ιδιαίτερα ευαίσθητη στο θόρυβο, αφού γενικά η διαδικασία παραγωγίσης ενισχύει τις υψηλές συχνότητες. Έτσι, για να εξαλειφθεί η επίδραση του θορύβου, υπολογίζεται ένας μέσος όρος με βάρη της διαφοράς $\mathbf{d}_1(k)$ σε ένα παράθυρο προκαθορισμένου μήκους $2N_W + 1$ με κέντρο το καρέ k :

$$\mathbf{d}_1^w(k) = \sum_{l=a_1(k)}^{l=\beta_1(k)} w_{l-k} \mathbf{d}_1(l) = \sum_{l=a_1(k)}^{l=\beta_1(k)} w_{l-k} (\mathbf{f}(l+1) - \mathbf{f}(l)), \quad k = 0, 1, \dots, N_F - 2 \quad (6-2)$$

όπου $a_1(k) = \max(0, k - N_W)$ και $\beta_1(k) = \min(N_F - 2, k + N_W)$. Τα βάρη w_l ορίζονται για $l \in \{-N_W, N_W\}$, ενώ στην απλή περίπτωση ενός τετραγωνικού παραθύρου είναι όλα ίσα με $1 / (2N_W + 1)$. Όπως φαίνεται στην Εξίσωση (6-2), το μήκος του παραθύρου μειώνεται γραμμικά στα όρια (στην αρχή και το τέλος) του πλάνου. Με όμοιο τρόπο ορίζεται ένας μέσος όρος με βάρη για τη δεύτερη παράγωγο:

$$\mathbf{d}_2^w(k) = \sum_{l=a_2(k)}^{l=\beta_2(k)} w_{l-k} (\mathbf{d}_1^w(l+1) - \mathbf{d}_1^w(l)), \quad k = 0, 1, \dots, N_F - 3 \quad (6-3)$$

όπου, όπως προηγουμένως, $a_2(k) = \max(0, k - N_W)$, $\beta_2(k) = \min(N_F - 3, k + N_W)$, και τα βάρη w_l , $l \in \{-N_W, N_W\}$, είναι ίσα με τα βάρη της Εξίσωσης (6-2) για την πρώτη παράγωγο, με την υπόθεση ότι χρησιμοποιείται το ίδιο είδος παραθύρου για τις δύο παραγώγους.

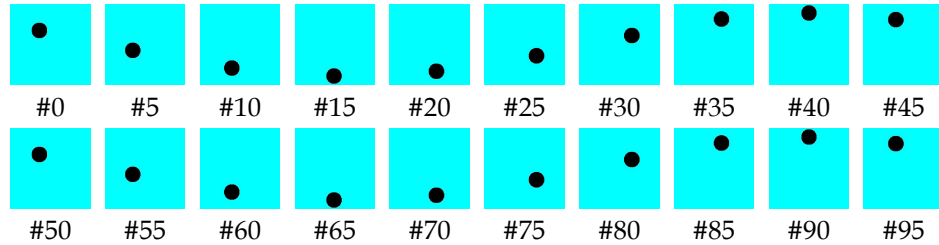
Τα στοιχεία της δεύτερης παραγώγου, $\mathbf{d}_2^w(k)$, εκφράζουν τη χρονική μεταβολή των επιμέρους στοιχείων του διανύσματος περιγραφής $\mathbf{f}(k)$, ενώ το διάνυσμα $\mathbf{d}_2^w(k)$ εκφράζει συνολικά την κατεύθυνση της επιτάχυνσης του $\mathbf{f}(k)$. Για να ληφθεί υπόψη η χρονική μεταβολή όλων των στοιχείων του $\mathbf{f}(k)$ υπολογίζεται το μέτρο της δεύτερης παραγώγου, $D(k) = |\mathbf{d}_2^w(k)|$, το οποίο εκφράζει το μέγεθος της επιτάχυνσης του $\mathbf{f}(k)$. Στη συνέχεια, οι τιμές του $k \in \{0, 1, \dots, N_F\}$ που αντιστοιχούν σε τοπικά μέγιστα ή ελάχιστα του $D(k)$ επιλέγονται ως χαρακτηριστικά καρέ. Πρέπει να σημειωθεί ότι το μέτρο $D(k)$ είναι μία ακολουθία διακριτού χρόνου, σε αντίθεση με το $D(t)$ του Σχήματος 6-1β που είναι μία συνεχής καμπύλη.

6.3 ΠΑΡΑΔΕΙΓΜΑ ΣΥΝΘΕΤΙΚΗΣ ΑΚΟΛΟΥΘΙΑΣ

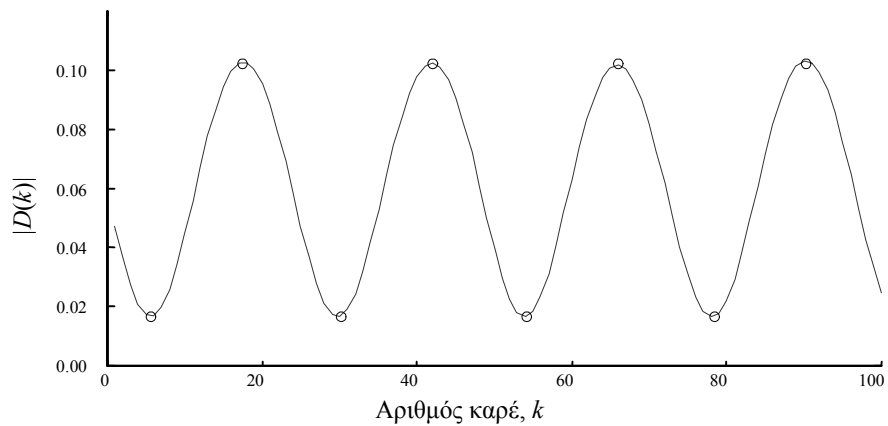
Για να γίνει περισσότερο κατανοητή η μέθοδος, παρουσιάζεται στη συνέχεια ένα παράδειγμα επιλογής χαρακτηριστικών καρέ από μία συνθετική ακολουθία βίντεο. Πρόκειται ουσιαστικά για την ίδια ακολουθία με εκείνη της Ενότητας 4.5, η οποία όμως επαναλαμβάνεται δύο φορές, έχει επομένως μήκος $N_F = 100$ καρέ, ενώ ο μαύρος κύκλος διαγράφει την ίδια ελλειπτική τροχιά δύο συνεχόμενες φορές. Πρόκειται για ακολουθία με ένα και μοναδικό πλάνο, επομένως δεν υφίσταται στην περίπτωση αυτή ανίχνευση πλάνων ή εξαγωγή χαρακτηριστικών πλάνων. Στο Σχήμα 6-2 φαίνονται 20 καρέ της ακολουθίας, των οποίων οι αριθμοί είναι ομοιόμορφα κατανεμημένοι μεταξύ 0 και 99 (#0, #5, #10, ..., #95). Όπως και στην Ενότητα 4.5, το διάνυσμα περιγραφής περιέχει χαρακτηριστικά χρώματος μόνο, δηλαδή $\mathbf{f}(k) = \mathbf{f}^c(k)$, $k = 0, 1, \dots, 99$, ενώ χρησιμοποιούνται και πάλι $Q = 2$ τριγωνικές συναρτήσεις συμμετοχής με 50% επικάλυψη για κάθε χαρακτηριστικό. Έτσι το μήκος του διανύσματος περιγραφής είναι και πάλι ίσο με 64.

Το Σχήμα 6-3 απεικονίζει το μέτρο της δεύτερης παραγώγου, $D(k)$, για όλα τα καρέ της ακολουθίας, $k = 0, 1, \dots, N_F - 1 = 99$. Τέσσερα τοπικά μέγιστα και τέσσερα τοπικά ελάχιστα ανιχνεύονται συνολικά και παρουσιάζονται στο σχήμα ως μικροί κύκλοι. Τα τοπικά μέγιστα αντιστοιχούν σε καρέ (ή και χρονικές στιγμές) της ακολουθίας όπου ο μαύρος κύκλος προσεγγίζει τις ακρότατες θέσεις της ελλειπτικής του τροχιάς κατά την κατακόρυφη κατεύθυνση, δηλαδή το πάνω και κάτω άκρο της εικόνας. Αντίστοιχα τα τοπικά ελάχιστα αντιστοιχούν σε

χρονικές στιγμές όπου ο κύκλος προσεγγίζει τις ακρότατες θέσεις του κατά την οριζόντια κατεύθυνση. Αφού η τροχιά διαγράφεται δύο συνεχόμενες φορές, τα δύο πρώτα τοπικά μέγιστα / ελάχιστα ανήκουν στη πρώτη περίοδο και τα δύο τελευταία στη δεύτερη. Τα χαρακτηριστικά καρέ της ακολουθίας επιλέγονται στις χρονικές στιγμές των τεσσάρων τοπικών μεγίστων και ελαχίστων. Πρόκειται συγκεκριμένα για τα καρέ #6, #18, #31, #43, #56, #68, #81 και #93 όπως φαίνεται στο Σχήμα 6-4.

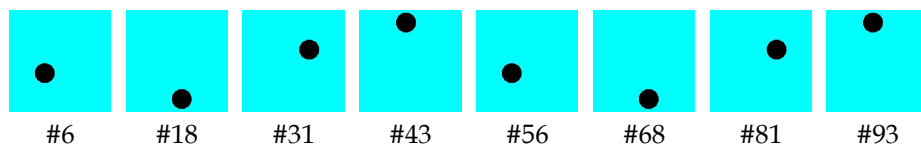


Σχήμα 6-2. Συνθετική ακολουθία βίντεο μήκους 100 καρέ για την επίδειξη της μεθόδου χρονικής μεταβολής.



Σχήμα 6-3. Μέτρο δεύτερης παραγώγου του διανύσματος περιγραφής για την ακολουθία του Σχήματος 6-2.

Οι χρονικές στιγμές των χαρακτηριστικών καρέ απεικονίζονται επίσης ως μικροί κύκλοι στα γραφήματα των Σχημάτων 4-5 και 4-6, τα οποία όπως έχει ήδη αναφερθεί αναπαριστούν προβολές της τροχιάς του διανύσματος περιγραφής σε διδιάστατους υπόχωρους που ορίζονται από ζεύγη στοιχείων του διανύσματος. Επιβεβαιώνεται ότι σε όλες τις περιπτώσεις τα επιλεγμένα χαρακτηριστικά καρέ βρίσκονται σε ακραία σημεία της τροχιάς. Παρατηρείται επίσης ότι μόνο οκτώ από τα 100 καρέ της αρχικής ακολουθίας απαιτούνται για την αναπαράσταση του οπτικού της περιεχομένου, οδηγώντας σε μείωση κατά 92% των απαιτήσεων αποθήκευσης. Παρά το γεγονός ότι η πραγματική πληροφορία μειώνεται κατά το ίδιο ποσοστό, το οπτικό περιεχόμενο της ακολουθίας διατηρείται.



Σχήμα 6-4. Χαρακτηριστικά καρέ με τη μέθοδο χρονικής μεταβολής για την ακολουθία του Σχήματος 6-2.

Η εξαγωγή χαρακτηριστικών καρέ με τη μέθοδο της χρονικής μεταβολής είναι μία πολύ απλή και γρήγορη διαδικασία, αφού στο διακριτό χρόνο η δεύτερη παράγωγος υλοποιείται με εξισώσεις διαφορών. Επιπλέον, το πλήθος των χαρακτηριστικών καρέ δεν είναι γνωστό εκ

των προτέρων. Αντίθετα, η μέθοδος εκτιμά ταυτόχρονα το πλήθος και τις θέσεις των χαρακτηριστικών καρέ. Βέβαια, σε περιπτώσεις όπου η μεταβολή του διανύσματος περιγραφής είναι σταθερή, η παραπάνω μέθοδος μπορεί να μη δώσει ικανοποιητικά αποτελέσματα. Τέτοιες περιπτώσεις όμως είναι πολύ σπάνιες και μπορεί να εμφανιστούν μόνο σε συνθετικές ακολουθίες βίντεο· σε πραγματικές ακολουθίες τα στοιχεία του διανύσματος περιγραφής δεν ακολουθούν κάποιο μαθηματικό ή φυσικό νόμο. Η συμπεριφορά του αλγορίθμου χρονικής μεταβολής μπορεί να βελτιωθεί σε αυτές τις περιπτώσεις με τη χρήση περισσότερων ασαφών συναρτήσεων συμμετοχής. Επίσης η χρήση μη γραμμικών συναρτήσεων συμμετοχής (π.χ. σιγμοειδών ή Γκαουσιανή) μπορεί να βοηθήσει σε αυτή την κατεύθυνση, αφού οι τριγωνικές συναρτήσεις είναι επίσης μη γραμμικές αλλά αποτελούνται από γραμμικά τμήματα.

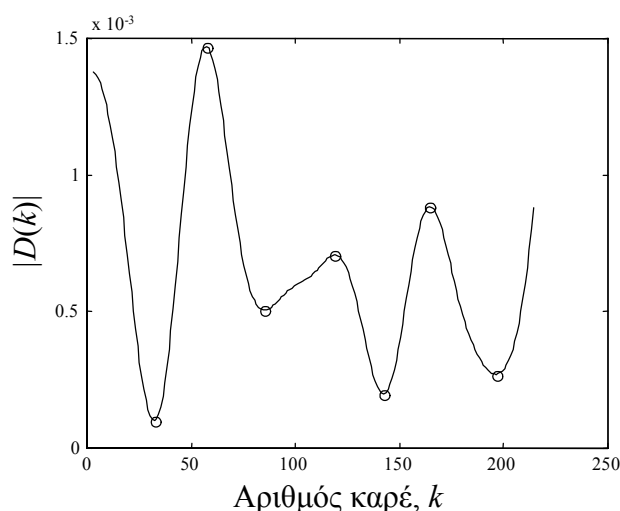
6.4 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την εκτίμηση της επίδοσης των προτεινόμενων αλγορίθμων επιλογής χαρακτηριστικών καρέ χρησιμοποιήθηκε μία βάση δεδομένων με ακολουθίες MPEG. Το οπτικοακουστικό υλικό της βάσης προέρχεται από πραγματικές ακολουθίες βίντεο συνολικής διάρκειας 3.5 περίπου ωρών και περιέχει πλάνα από τηλεοπτικά προγράμματα και συγκεκριμένα δελτία ειδήσεων, ενημερωτικές και αθλητικές εκπομπές, κινηματογραφικές ταινίες και διαφημιστικά σποτ. Οι διαδικασίες ανίχνευσης πλάνων και εξαγωγής διανυσμάτων περιγραφής εφαρμόστηκαν σε όλες τις ακολουθίες και οι πληροφορίες που σχετίζονται με τις χρονικές στιγμές αλλαγής πλάνου και τα διανύσματα περιγραφής κάθε καρέ αποθηκεύθηκαν στη βάση ώστε να είναι διαθέσιμες για τα πειράματα επιλογής χαρακτηριστικών καρέ.

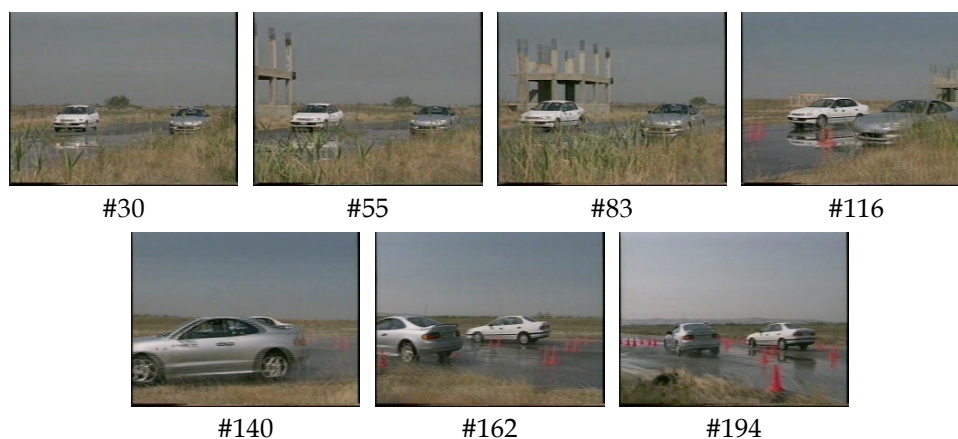


Σχήμα 6-5. Ακολουθία δοκιμαστικής οδήγησης αυτοκινήτων, μήκους 223 καρέ (22 δευτερολέπτων).

Οι προτεινόμενοι αλγόριθμοι επιλογής χαρακτηριστικών καρέ εφαρμόστηκαν ανεξάρτητα σε κάθε πλάνο του οπτικοακουστικού υλικού, χρησιμοποιώντας απευθείας τα διανύσματα περιγραφής όλων των καρέ του κάθε πλάνου. Στην Ενότητα αυτή παρουσιάζεται ένα παράδειγμα των πειραμάτων της μεθόδου χρονικής μεταβολής, ενώ στην Ενότητα 7.4 παρουσιάζονται αποτελέσματα για τις τρεις προτεινόμενες μεθόδους βέλτιστης επιλογής. Για την εξαγωγή διανυσμάτων περιγραφής, τα πεδία τιμών των χαρακτηριστικών χρώματος και κίνησης διαμερίστηκαν σε $Q = 3$ κλάσεις με χρήση τριών συναρτήσεων συμμετοχής με 50% επικάλυψη. Έτσι το συνολικό μήκος του διανύσματος περιγραφής είναι $Q^{L_c} + Q^{L_m} = 972$ για $L_c = 6$ και $L_m = 5$, όπως αναφέρεται στην Ενότητα 4.4.



Σχήμα 6-6. Μέτρο της δεύτερης παραγώγου του διανύσματος περιγραφής, $|D(k)|$, συναρτήσει του αριθμού καρέ, k , για την ακολουθία του Σχήματος 6-5.



Σχήμα 6-7. Επτά χαρακτηριστικά καρέ του πλάνου του Σχήματος 6-5, με τη μέθοδο χρονικής μεταβολής.

Ένα πλάνο από το περιεχόμενο της βάσης χρησιμοποιείται στην Ενότητα αυτή για επίδειξη της μεθόδου χρονικής μεταβολής. Το πλάνο προέρχεται από ένα πρόγραμμα δοκιμαστικής οδήγησης αυτοκινήτων, έχει διάρκεια 22 περίπου δευτερόλεπτα ($N_F = 223$ καρέ με ρυθμό 10 frames/sec) και περιέχει γρήγορα κινούμενα αντικείμενα. Το περιεχόμενο του πλάνου φαίνεται στο Σχήμα 6-5, όπου για λόγους παρουσίασης απεικονίζεται ένα κάθε 10 καρέ. Τα αποτελέσματα της μεθόδου χρονικής μεταβολής παρουσιάζονται στα Σχήματα 6-6 και 6-7. Συγκεκριμένα, το Σχήμα δείχνει το μέτρο της δεύτερης παραγώγου του διανύσματος περιγραφής, $|D(k)|$, συναρτήσει του αριθμού καρέ, k . Η μεταβολή του $|D(k)|$ φανερώνει ότι, λόγω του περιεχομένου του πλάνου, η τροχιά του διανύσματος περιγραφής είναι πιο πολύπλοκη από εκείνη του παραδείγματος της συνθετικής ακολουθίας του Σχήματος 6-2. Παρά το γεγονός αυτό, η καμπύλη του $|D(k)|$ διατηρείται λεία λόγω της διαδικασίας φιλτραρίσματος που παρουσιάσθηκε στην Ενότητα 6.2. Έτσι εξασφαλίζεται ότι τα τοπικά ακρότατα του $|D(k)|$, που φαίνονται ως μικροί κύκλοι στο Σχήμα 6-6, αντιστοιχούν σε πραγματικές μεταβολές του οπτικού περιεχομένου του πλάνου και όχι σε θόρυβο που προέρχεται από τη διαδικασία κατάτμησης.

Τα επτά επιλεγμένα χαρακτηριστικά καρτέ του πλάνου, που αντιστοιχούν στις χρονικές στιγμές των τοπικών ακρότατων του $|D(k)|$, παρουσιάζονται στο Σχήμα 6-7. Παρατηρείται ότι αυτά τα καρτέ παρέχουν ικανοποιητική πληροφορία για προεπισκόπηση του πλάνου το οποίο αποτελείται από 223 καρτέ συνολικά. Επίσης, παρά το γεγονός ότι η σημαντική δράση του πλάνου περιέχεται σε ένα τμήμα του με μικρή διάρκεια, η μέθοδος επιτυγχάνει την εξαγωγή της δράσης. Αντίθετα, η επιλογή επτά καρτέ σε ίσα χρονικά διαστήματα μεταξύ τους δεν θα είχε το ίδιο αποτέλεσμα. Όμως κάποια από τα χαρακτηριστικά καρτέ είναι παρόμοια μεταξύ τους, όπως για παράδειγμα τα καρτέ #30 και #55. Έτσι φαίνεται ότι ενώ η μέθοδος χρονικής μεταβολής έχει τη δυνατότητα να ανιχνεύει περιοδικές επαναλήψεις του οπτικού περιεχομένου, έχει την τάση να διατηρεί περιττή πληροφορία. Πιο σύντομη περίληψη του πλάνου μπορεί να επιτευχθεί με βέλτιστο τρόπο εξετάζοντας την ομοιότητα των καρτέ, και αυτό είναι το αντικείμενο της μεθόδου συσχέτισης που παρουσιάζεται στο επόμενο Κεφάλαιο.

Βέλτιστη Επιλογή Χαρακτηριστικών Καρέ

Η δεύτερη μέθοδος επιλογής χαρακτηριστικών καρέ βασίζεται στην ελαχιστοποίηση ενός κριτηρίου συσχέτισης μεταξύ των διανυσμάτων περιγραφής, έτσι ώστε τα επιλεγμένα καρέ να εμφανίζουν όσο το δυνατόν μικρότερες ομοιότητες μεταξύ τους. Έτσι ένας μικρός αριθμός κατάλληλα επιλεγμένων καρέ είναι δυνατόν να περιγράψει ικανοποιητικά το οπτικό περιεχόμενο μιας ακολουθίας. Όμως το πρόβλημα βελτιστοποίησης ανάγεται σε ένα συνδυαστικό πρόβλημα του οποίου η πολυπλοκότητα είναι τέτοια ώστε η εξαντλητική έρευνα για τη βέλτιστη λύση να είναι ανέφικτη. Για το λόγο αυτό προτείνονται δύο επιμέρους τεχνικές υλοποίησης της μεθόδου συσχέτισης. Η πρώτη προέρχεται σαν ιδέα από την εκτίμηση κίνησης σε ακολουθίες βίντεο και εκτελεί λογαριθμική αναζήτηση στον πολυδιάστατο χώρο των διανυσμάτων περιγραφής. Η δεύτερη χρησιμοποιεί γενετικό αλγόριθμο και έχει τη δυνατότητα παράλληλης αναζήτησης με στοχαστικό τρόπο. Οι δύο τεχνικές αναζήτησης, καθώς και μία στοχαστική εκδοχή του αλγορίθμου λογαριθμικής αναζήτησης αξιολογούνται σε ένα μεγάλο σύνολο πραγματικών ακολουθιών και δίνονται συγκριτικές μετρήσεις που αφορούν το βέλτιστο μέτρο συσχέτισης και τον απαιτούμενο χρόνο εκτέλεσης. Στα πειράματα η μέθοδος συσχέτισης συνδυάζεται με τη μέθοδο χρονικής μεταβολής, λόγω της ικανότητας της τελευταίας να ανιχνεύει αυτόματα τις μεταβολές του οπτικού περιεχομένου και συνεπώς τον απαιτούμενο αριθμό χαρακτηριστικών καρέ σε κάθε πλάνο της ακολουθίας.

7.1 ΜΕΘΟΔΟΣ ΣΥΣΧΕΤΙΣΗΣ

Όπως έγινε φανερό από το παράδειγμα της Ενότητας 6.4, η μέθοδος της χρονικής μεταβολής έχει την ικανότητα να ανιχνεύει παραπάνω από μία επαναλήψεις του ίδιου οπτικού περιεχομένου. Αυτό μπορεί να είναι χρήσιμο για την κατανόηση της χρονικής εξέλιξης της δράσης ενός πλάνου. Σε περιπτώσεις όμως όπου η χρονική εξέλιξη ενός πλάνου δεν είναι το ζητούμενο, η μέθοδος αυτή δεν παρέχει συμπαγή αναπαράσταση του οπτικού περιεχομένου, αφού διατηρεί περιττές πληροφορίες. Σε τέτοιες περιπτώσεις, είναι επιθυμητή η επιλογή ενός όσο το δυνατόν μικρού συνόλου καρέ που να χαρακτηρίζουν επαρκώς το οπτικό περιεχόμενο αλλά ταυτόχρονα να είναι όσο το δυνατόν ανόμοια μεταξύ τους. Γι' αυτό το λόγο προτείνεται στη συνέχεια ένας αλγόριθμος βέλτιστης επιλογής χαρακτηριστικών καρέ (optimal key-frame selection) που βασίζεται σε μια μέθοδο βελτιστοποίησης, η οποία εντοπίζει το σύνολο των πλέον ασυσχέτιστων διανυσμάτων περιγραφής σε ένα πλάνο. Η μέθοδος ονομάζεται μέθοδος συσχέτισης και επιτυγχάνει το στόχο της με την ελαχιστοποίηση ενός κριτηρίου ετεροσυσχέτισης μεταξύ των διανυσμάτων περιγραφής των καρέ σε ένα δεδομένο πλάνο μιας ακολουθίας.

Υπενθυμίζεται ότι $\mathbf{f}(k)$ είναι το διάνυσμα περιγραφής του k -οστού καρέ του δεδομένου πλάνου, όπου $k \in V$, και

$$V = \{0, 1, \dots, N_F - 1\} \quad (7-1)$$

είναι το σύνολο των καρτέ που περιέχονται στο πλάνο και N_F είναι το πλήθος των καρτέ του πλάνου. Έστω επίσης ότι K_F είναι το πλήθος των καρτέ που πρέπει να επιλεγούν ως αντιπροσωπευτικά του πλάνου. Σε αντίθεση με την προηγούμενη μέθοδο, ο αριθμός αυτός πρέπει να είναι εκ των προτέρων γνωστός. Ένας τρόπος εκτίμησής του είναι η χρήση της μεθόδου χρονικής μεταβολής. Συγκεκριμένα, τα πειράματα δείχνουν ότι στις περισσότερες περιπτώσεις ο αριθμός K_F πρέπει να είναι περίπου ο μισός από τον αριθμό χαρακτηριστικών καρτέ που εξάγονται με τη μέθοδο της χρονικής μεταβολής. Ως μέτρο ομοιότητας μεταξύ δύο καρτέ $k, l \in V$ χρησιμοποιείται ο *συντελεστής συσχέτισης* $\rho_{k,l}$ μεταξύ των αντίστοιχων διανυσμάτων περιγραφής $\mathbf{f}(k), \mathbf{f}(l)$

$$\rho_{k,l} = C_{k,l} / (\sigma_k \sigma_l), \quad k, l \in V \quad (7-2)$$

όπου $C_{k,l}$ είναι η *συμμεταβλητότητα* των διανυσμάτων $\mathbf{f}(k), \mathbf{f}(l)$

$$C_{k,l} = (\mathbf{f}(k) - \mathbf{m})^T (\mathbf{f}(l) - \mathbf{m}), \quad k, l \in V \quad (7-3)$$

ενώ \mathbf{m} είναι η μέση τιμή των διανυσμάτων περιγραφής για όλη τη διάρκεια του πλάνου

$$\mathbf{m} = \frac{1}{N_F} \sum_{i=0}^{N_F-1} \mathbf{f}(i) \quad (7-4)$$

και τέλος $\sigma_k^2 = C_{k,k}$, είναι η διασπορά του διανύσματος $\mathbf{f}(k)$, $k \in V$. Με βάση τους συντελεστές συσχέτισης μεταξύ ζευγών από διανύσματα περιγραφής, μπορεί να ορισθεί ένα μέτρο συσχέτισης ενός *συνόλου* από K_F διανύσματα. Για το σκοπό αυτό ορίζεται αρχικά ένα *διάνυσμα δεικτών*:

$$\mathbf{x} = (x_1, \dots, x_{K_F}) \in W \subset V^{K_F} \quad (7-5)$$

όπου

$$W = \{(x_1, \dots, x_{K_F}) \in V^{K_F} : x_1 < \dots < x_{K_F}\} \quad (7-6)$$

είναι το υποσύνολο του V^{K_F} που περιέχει όλα διανύσματα δεικτών \mathbf{x} τα οποία αντιστοιχούν σε σύνολα καρτέ *ταξινομημένα* κατά αύξουσα σειρά αριθμού εμφάνισης στο πλάνο. Έτσι, το *μέτρο συσχέτισης* $R(\mathbf{x})$ του συνόλου των διανυσμάτων περιγραφής $\mathbf{f}(k)$, $k = x_1, \dots, x_{K_F}$ ορίζεται ως

$$R(\mathbf{x}) = R(x_1, \dots, x_{K_F}) = \frac{2}{K_F(K_F-1)} \sum_{i=1}^{K_F-1} \sum_{j=i+1}^{K_F} (\rho_{x_i, x_j})^2 \quad (7-7)$$

και παίρνει τιμές στο κλειστό διάστημα $[0,1]$. Με βάση τον παραπάνω ορισμό είναι εμφανές ότι η αναζήτηση ενός συνόλου από K_F ελάχιστα συσχετισμένων διανυσμάτων περιγραφής είναι ισοδύναμη με την αναζήτηση ενός διανύσματος δεικτών $\mathbf{x} \in W$ το οποίο να ελαχιστοποιεί το μέτρο $R(\mathbf{x})$. Η αναζήτηση περιορίζεται στο υποσύνολο W , αφού τα διανύσματα δεικτών χρησιμοποιούνται για την κατασκευή *συνόλων* από διανύσματα περιγραφής, και επομένως αν ένα διάνυσμα δεικτών \mathbf{x} αντιστοιχεί σε ένα δεδομένο σύνολο διανυσμάτων περιγραφής, τότε οποιαδήποτε αναδιάταξη των στοιχείων του \mathbf{x} αντιστοιχεί στο ίδιο ακριβώς σύνολο. Είναι επιπλέον φανερό ότι το μέτρο συσχέτισης $R(\mathbf{x})$ όπως ορίζεται στην Εξίσωση (7-7) είναι ανεξάρτητο από τη διάταξη των στοιχείων του διανύσματος \mathbf{x} . Έτσι, το σύνολο των K_F ελάχιστα συσχετισμένων διανυσμάτων περιγραφής, που αντιστοιχεί στο σύνολο των K_F χαρακτηριστικών καρτέ του δεδομένου πλάνου, ορίζεται ως

$$\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_{K_F}) = \arg \min_{\mathbf{x} \in W} R(\mathbf{x}) \quad (7-8)$$

Δυστυχώς, το πλήθος δυνατών συνδυασμών καρέ και επομένως η πολυπλοκότητα της εξαντλητικής αναζήτησης της ελάχιστης τιμής του μέτρου $R(\mathbf{x})$ είναι τέτοια ώστε η απ' ευθείας αναζήτηση είναι πρακτικά ανέφικτη. Για παράδειγμα, για την επιλογή πέντε χαρακτηριστικών καρέ από ένα πλάνο που περιέχει συνολικά μόνο 128 καρέ, απαιτείται αναζήτηση σε ένα σύνολο 264 εκατομμυρίων συνδυασμών καρέ, κάθε ένας από τους οποίους απαιτεί και υπολογισμούς για την εκτίμηση του $R(\mathbf{x})$. Γι' αυτό το λόγο προτείνονται στη συνέχεια δύο εναλλακτικές τεχνικές υλοποίησης της μεθόδου συσχέτισης: η *λογαριθμική αναζήτηση* και η *αναζήτηση με γενετικό αλγόριθμο*.

Η τεχνική λογαριθμικής αναζήτησης για την υλοποίηση της μεθόδου συσχέτισης είναι πρωτότυπη, δημοσιεύθηκε για πρώτη φορά στην εργασία [36] και ξαναχρησιμοποιήθηκε στην [37], αλλά και στην [8] όπου γίνεται εκτεταμένη σύγκριση με την τεχνική του γενετικού αλγορίθμου. Η τελευταία είναι επίσης πρωτότυπη: η πρώτη της δημοσίευση έγινε στην εργασία [33] και στη συνέχεια λόγω των σημαντικών επιδόσεων της χρησιμοποιήθηκε σε πολλές ακόμη εργασίες [8,31,34,38,39].

7.2 ΛΟΓΑΡΙΘΜΙΚΗ ΑΝΑΖΗΤΗΣΗ

Η πρώτη λύση που προτείνεται για την πρακτική υλοποίηση της μεθόδου συσχέτισης για την επιλογή χαρακτηριστικών καρέ ονομάζεται *λογαριθμική αναζήτηση* και βασίζεται σε μία τεχνική παρόμοια με εκείνη που χρησιμοποιείται στα πρότυπα MPEG για την *εκτίμηση κίνησης* (block motion estimation) [127]. Η βασική διαφορά είναι ότι η εκτίμηση κίνησης γίνεται στο διδιάστατο χώρο των εικόνων ενώ η αναζήτηση χαρακτηριστικών καρέ γίνεται στον πολυδιάστατο χώρο W . Συγκεκριμένα, αντί να εκτελεσθεί εξαντλητική αναζήτηση σε όλα τα διανύσματα δεικτών του χώρου W , ακολουθείται ένα απλό *μονοπάτι* σημείων του χώρου, που έχει ως σημείο εκκίνησης ένα κατάλληλα επιλεγμένο σημείο του W . Σε κάθε σημείο του μονοπατιού εξετάζεται μόνο το σύνολο των *γειτονικών* του σημείων και το επόμενο σημείο επιλέγεται ως εκείνο το γειτονικό σημείο του συνόλου, που αντιστοιχεί στο ελάχιστο μέτρο συσχέτισης. Σε κάθε βήμα του αλγορίθμου η περιοχή που καλύπτει το σύνολο γειτονικών σημείων μειώνεται εκθετικά. Όταν η περιοχή αυτή φτάσει σε μέγεθος το ένα σημείο, το σημείο αυτό επιλέγεται ως λύση του προβλήματος βελτιστοποίησης. Παρά το γεγονός ότι εξετάζεται μόνο ένα πολύ μικρό υποσύνολο του χώρου αναζήτησης W , ο αλγόριθμος παρουσιάζει ικανοποιητική επίδοση. Αυτό είναι αναμενόμενο διότι τα καρέ που έχουν μικρή απόσταση μεταξύ τους (στο χρόνο) έχουν συνήθως παρόμοια διανύσματα περιγραφής, και επομένως τα διανύσματα δεικτών που έχουν μικρή απόσταση μεταξύ τους (στο χώρο W) έχουν συνήθως παρόμοιο μέτρο συσχέτισης.

Χωρίς βλάβη της γενικότητας γίνεται παρακάτω η υπόθεση ότι το πλήθος των καρέ είναι δύναμη του δύο, δηλαδή $N_F = 2^M$ όπου M ακέραιος. Στην περίπτωση που το πλήθος των καρέ δεν ικανοποιεί αυτό τον περιορισμό γίνεται επέκτασή του προς την αμέσως μεγαλύτερη δύναμη του δύο με προσθήκη «εικονικών» καρέ, των οποίων οι συντελεστές συσχέτισης παίρνουν μεγάλη τιμή ώστε να απορρίπτονται από τη διαδικασία επιλογής χαρακτηριστικών καρέ.

Στην αρχικοποίηση του αλγορίθμου πρέπει να επιλεγεί ένα αρχικό διάνυσμα δεικτών, έστω $\mathbf{x}(0)$. Μία προφανής επιλογή είναι το *κεντρικό σημείο* $\tilde{\mathbf{x}} = (\mu, \dots, \mu)$ όπου $\mu = 2^{M-1} - 1$ είναι η κεντρική χρονική στιγμή του πλάνου, μήκους $N_F = 2^M$. Όμως το σημείο αυτό δεν ανήκει στο χώρο W αφού τα στοιχεία του δεν ικανοποιούν τις συνθήκες ανισότητας της Εξίσωσης (7-6). Έτσι το $\mathbf{x}(0)$ επιλέγεται ως το σημείο του χώρου W που είναι το πλησιέστερο στο κεντρικό σημείο $\tilde{\mathbf{x}}$, δηλαδή

$$\mathbf{x}(0) = (\mu - \lfloor K_F / 2 \rfloor, \dots, \mu - 1, \mu + 1, \dots, \mu + \lfloor K_F / 2 \rfloor) \quad (7-9)$$

αν το K_F είναι άρτιο, και

$$\mathbf{x}(0) = (\mu - \lfloor K_F / 2 \rfloor, \dots, \mu - 1, \mu, \mu + 1, \dots, \mu + \lfloor K_F / 2 \rfloor) \quad (7-10)$$

αν το K_F είναι περιττό. Ας υποθέσουμε τώρα ότι στη n -οστή επανάληψη του αλγορίθμου το μονοπάτι περνά από το σημείο $\mathbf{x}(n)$. Τότε για την επιλογή του επόμενου σημείου του μονοπατιού, $\mathbf{x}(n+1)$, υπολογίζεται το μέτρο συσχέτισης για όλα τα γειτονικά σημεία του $\mathbf{x}(n)$ στη *γειτονική περιοχή* $N(\mathbf{x}(n), \delta(n))$ που ορίζεται γενικά ως

$$N(\mathbf{x}, \delta) = \{\mathbf{y} \in W : \mathbf{y} = \mathbf{x} + \delta \mathbf{p}, \mathbf{p} \in G^{K_F}\}, \quad \mathbf{x} \in W \quad (7-11)$$

όπου $G = \{-1, 0, 1\}$ και $\delta(n)$ είναι ένας ακέραιος που καθορίζει το μέγεθος της γειτονικής περιοχής στη n -οστή επανάληψη. Η παραπάνω εξίσωση υποδεικνύει ότι τα γειτονικά σημεία του $\mathbf{x}(n)$ βρίσκονται σε ένα πλέγμα που προκύπτει από το G^{K_F} με επέκταση κατά $\delta(n)$ και μετατόπιση κατά $\mathbf{x}(n)$. Το μέγεθος γειτονικής περιοχής αρχικοποιείται ως $\delta(0) = 2^{M-2}$ έτσι ώστε ο αλγόριθμος να μπορεί να καλύψει όλα τα δυνατά σημεία του χώρου W . Με βάση τα παραπάνω, οι ενέργειες που επαναλαμβάνονται σε κάθε βήμα του αλγορίθμου είναι (α) επιλογή του γειτονικού σημείου του $\mathbf{x}(n)$ με το ελάχιστο μέτρο συσχέτισης ως το επόμενο σημείο του μονοπατιού, $\mathbf{x}(n+1)$, και (β) μείωση στο μισό του μεγέθους γειτονικής περιοχής. Ισοδύναμα,

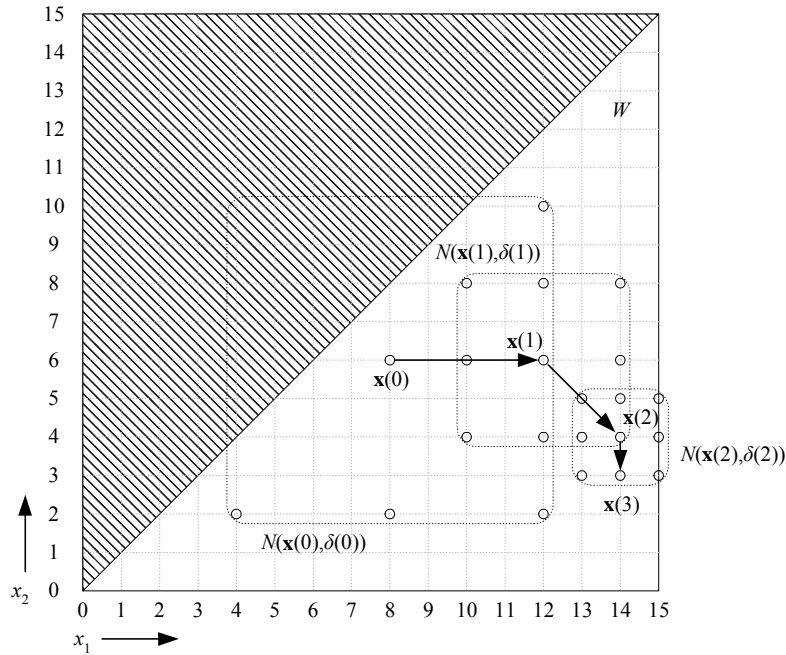
$$\mathbf{x}(n+1) = \arg \min_{\mathbf{x} \in N(\mathbf{x}(n), \delta(n))} R(\mathbf{x}) \quad (7-12)$$

$$\delta(n+1) = \delta(n) / 2 \quad (7-13)$$

Μετά την αρχικοποίηση του αλγορίθμου, δηλαδή τον υπολογισμό των $\mathbf{x}(0)$ και $\delta(0)$, τα παραπάνω βήματα επαναλαμβάνονται για $n = 0, 1, \dots, M-2$, δηλαδή μέχρι $\delta(n) = 1$. Έτσι, ο αλγόριθμος σταματά μετά από $M-1$ επαναλήψεις και το τελικό αποτέλεσμα είναι $\hat{\mathbf{x}} = \mathbf{x}(M-1)$. Αυτό σημαίνει ότι οι χρονικές στιγμές των χαρακτηριστικών καρέ επιλέγονται ως τα στοιχεία του διανύσματος $\mathbf{x}(M-1)$.

Το Σχήμα 7-1 παρουσιάζει μία γραφική αναπαράσταση του αλγορίθμου για $N_F = 16$ ($M = 4$) και $K_F = 2$, όπου ο οριζόντιος και κατακόρυφος άξονας αντιστοιχούν στα δύο στοιχεία, x_1 και x_2 , του διανύσματος δεικτών \mathbf{x} . Η τριγωνική διαγραμματισμένη περιοχή περιλαμβάνει τα διανύσματα δεικτών που δεν ανήκουν στο χώρο W , δηλαδή δεν αποτελούν λύσεις του προβλήματος ελαχιστοποίησης. Στο ίδιο Σχήμα, τα βέλη δείχνουν την πορεία του μονοπατιού αναζήτησης $\mathbf{x}(0) \rightarrow \mathbf{x}(1) \rightarrow \mathbf{x}(2) \rightarrow \mathbf{x}(3) = \hat{\mathbf{x}}$. Για κάθε σημείο του μονοπατιού, η γειτονική περιοχή αναπαριστάται με ένα διακεκομμένο τετράγωνο, ενώ τα γειτονικά σημεία που βρίσκονται εντός του χώρου W φαίνονται ως μικροί κύκλοι. Είναι εμφανές ότι το μέγεθος της γειτονικής περιοχής μειώνεται στο μισό σε κάθε βήμα του αλγορίθμου, ώστε να εντοπίζεται πρώτα η ευρύτερη περιοχή της βέλτιστης λύσης και στη συνέχεια το ακριβές σημείο της. Η όλη διαδικασία στηρίζεται στην υπόθεση ότι όσο πιο κοντά βρίσκονται δύο σημεία στο χώρο W , τόσο μικρότερη είναι και η διαφορά των μέτρων συσχέτισής τους. Η υπόθεση αυτή επιβεβαιώνεται

στην πράξη στις περισσότερες περιπτώσεις, αφού οι περισσότερες ακολουθίες βίντεο παρουσιάζουν «συνεχή» μεταβολή ενώ οι απότομες μεταβολές εκλαμβάνονται ως αλλαγές πλάνου.



Σχήμα 7-1. Γραφική αναπαράσταση του αλγορίθμου λογαριθμικής αναζήτησης για $N_F = 16$ ($M = 4$) και $K_F = 2$.

Η προτεινόμενη τεχνική της λογαριθμικής αναζήτησης μειώνει δραστικά την απαιτούμενη πολυπλοκότητα και καθιστά την εξαγωγή χαρακτηριστικών καρέ με τη μέθοδο συσχέτισης όχι μόνο πρακτικά εφικτή αλλά και πολύ γρήγορη. Όμως εξετάζει ένα μόνο συγκεκριμένο μονοπάτι του χώρου W που αποτελεί ένα πολύ μικρό υποσύνολο του χώρου και αγνοεί πλήρως τον υπόλοιπο χώρο. Έτσι πολύ συχνά «εγκλωβίζεται» σε ένα τοπικό ελάχιστο του $R(\mathbf{x})$ και το τελικό της αποτέλεσμα πολλές φορές απέχει σημαντικά από το βέλτιστο. Γι' αυτό το λόγο έχει προταθεί μία στοχαστική εκδοχή του αλγορίθμου [8] η οποία εξετάζει περισσότερα από ένα μονοπάτια, η επιλογή των οποίων γίνεται με τυχαίο τρόπο. Η τεχνική αυτή δίνει πάντα καλύτερα αποτελέσματα, αλλά με αυξημένο υπολογιστικό κόστος.

7.3 ΑΝΑΖΗΤΗΣΗ ΜΕ ΓΕΝΕΤΙΚΟ ΑΛΓΟΡΙΘΜΟ

Ο αλγόριθμος λογαριθμικής αναζήτησης που παρουσιάστηκε παραπάνω παρέχει πολύ γρήγορη σύγκλιση σε μία υπο-βέλτιστη λύση του προβλήματος ελαχιστοποίησης του μέτρου συσχέτισης $R(\mathbf{x})$, αλλά με σημαντική πιθανότητα σύγκλισης σε τοπικά ελάχιστα του $R(\mathbf{x})$. Η πιθανότητα αυτή μειώνεται σε μεγάλο βαθμό με τη χρήση της στοχαστικής εκδοχής του αλγορίθμου, η οποία βέβαια έχει αυξημένη υπολογιστική πολυπλοκότητα. Η στοχαστική εκδοχή βασίζεται στην ιδέα της καθοδηγούμενης τυχαίας αναζήτησης (guided random search), αφού ένας αριθμός από τυχαία μονοπάτια ακολουθούνται παράλληλα, ενώ η πιθανότητα αλλαγής πορείας ενός μονοπατιού προς μία νέα κατεύθυνση εξαρτάται από το μέτρο συσχέτισης του γειτονικού σημείου στην κατεύθυνση αυτή. Η ιδέα αυτή μπορεί να επεκταθεί περισσότερο με τη χρήση ενός εξελικτικού προγράμματος (evolution program, EP) [88].

Γενετικοί Αλγόριθμοι

Σε αντίθεση με τις απαριθμητικές τεχνικές αναζήτησης, όπως ο *δυναμικός προγραμματισμός*, οι οποίες μπορεί να αποτύχουν σε πολύπλοκα προβλήματα, τα εξελικτικά προγράμματα παρέχουν μοναδική ευελιξία και ευρωστία σε τέτοια προβλήματα. Γι' αυτό το λόγο υιοθετείται στη συνέχεια η προσέγγιση ενός *γενετικού αλγόριθμου* (genetic algorithm, GA) [48]. Οι γενετικοί αλγόριθμοι αποτελούν ειδική περίπτωση των εξελικτικών προγραμμάτων και χρησιμοποιούνται κυρίως σε διακριτά προβλήματα βελτιστοποίησης. Η προσέγγιση αυτή φαίνεται πολύ αποτελεσματική για το συγκεκριμένο πρόβλημα βελτιστοποίησης, δεδομένων των διαστάσεων του χώρου αναζήτησης και της φύσης του κριτηρίου ελαχιστοποίησης, το οποίο παρουσιάζει μεγάλο αριθμό τοπικών ελαχίστων. Το γεγονός αυτό γίνεται εμφανές στην επόμενη Ενότητα όπου παρουσιάζονται συγκριτικά πειραματικά αποτελέσματα.

Σε έναν γενετικό αλγόριθμο, οι δυνατές λύσεις του προβλήματος βελτιστοποίησης αναπαρίστανται από *χρωμοσώματα* (chromosomes), των οποίων το γενετικό υλικό αποτελείται από έναν αριθμό *γονιδίων* (genes). Τα γονίδια ενός χρωμοσώματος περιέχουν πληροφορία που σχετίζεται με την αντίστοιχη λύση του προβλήματος και η οποία αποθηκεύεται με ένα συγκεκριμένο *σχέδιο κωδικοποίησης* (encoding scheme). Στην αρχικοποίηση του γενετικού αλγορίθμου παράγεται ένας αρχικός *πληθυσμός* (population) χρωμοσωμάτων. Στη συνέχεια, σε κάθε επανάληψη, επιλέγεται ένα υποσύνολο του πληθυσμού σύμφωνα με ένα κριτήριο *ικανότητας* (fitness). Το υποσύνολο αυτό περιέχει τα *γονικά* χρωμοσώματα (parents) στα οποία εφαρμόζεται μία σειρά από γενετικές διαδικασίες για τη δημιουργία *απογόνων* (offspring). Ο πληθυσμός της επόμενης γενιάς, και της επόμενης επανάληψης του αλγορίθμου, σχηματίζεται με το συνδυασμό των απογόνων με ένα ποσοστό χρωμοσωμάτων της προηγούμενης γενιάς, σύμφωνα με μία *στρατηγική αντικατάστασης* (replacement strategy). Η παραπάνω διαδικασία επαναλαμβάνεται μέχρι ο αλγόριθμος να συγκλίνει σε μία βέλτιστη λύση, δηλαδή μέχρι ο πληθυσμός να σταματήσει να εξελίσσεται.

Με την απομίμηση της διαδικασίας της *φυσικής επιλογής* που οφείλεται στη βιολογική εξέλιξη, ο γενετικός αλγόριθμος επιτυγχάνει παράλληλη αναζήτηση σε διαφορετικές περιοχές του χώρου αναζήτησης, ενώ ταυτόχρονα εντείνει την αναζήτηση σε περιοχές με αυξημένη πιθανότητα εύρεσης βέλτιστης λύσης. Έτσι αποφεύγει τον εγκλωβισμό σε τοπικά ελάχιστα και παρέχει λύσεις πολύ κοντά στη βέλτιστη με πολύ μικρό υπολογιστικό κόστος, το οποίο είναι πολλές φορές μικρότερο ακόμη και σε σχέση με τον αλγόριθμο λογαριθμικής αναζήτησης.

Κωδικοποίηση Γονιδίων

Στο συγκεκριμένο πρόβλημα βελτιστοποίησης, οι δυνατές λύσεις είναι σύνολα από καρέ ενός πλάνου μιας ακολουθίας βίντεο, τα οποία αναπαρίστανται από τα διανύσματα δεικτών $\mathbf{x} = (x_1, \dots, x_{K_F}) \in W$, ενώ τα στοιχεία x_i , $i = 1, 2, \dots, K_F$, κάθε διανύσματος είναι οι αριθμοί (δείκτες ή χρονικές στιγμές) των καρέ. Έτσι τα διανύσματα δεικτών \mathbf{x} χρησιμοποιούνται ως χρωμοσώματα ενώ τα στοιχεία τους x_i ως γονίδια των χρωμοσωμάτων. Τα γονίδια κωδικοποιούνται ως ακέραιοι αριθμοί (integer encoding scheme). Ο λόγος για την επιλογή αυτή, σε αντίθεση με μία εναλλακτική δυαδική κωδικοποίηση (binary encoding scheme) είναι ότι οι γενετικές διαδικασίες πρέπει να εφαρμόζονται απευθείας στα γονίδια x_i και όχι σε αυθαίρετα bits της δυαδικής τους αναπαράστασης.

Ο αρχικός πληθυσμός (initial population), $\mathbf{X}(0) = \{\mathbf{x}_1, \dots, \mathbf{x}_P\}$, αποτελούμενος από P χρωμοσώματα, παράγεται με την επιλογή P συνόλων από καρέ των οποίων τα διανύσματα περιγραφής βρίσκονται σε ακραίες θέσεις της τροχιάς του διανύσματος περιγραφής, σύμφωνα με τη μέθοδο χρονικής μεταβολής, όπως περιγράφεται στην Ενότητα 6.2. Συνήθως η επιλογή αρχικού πληθυσμού σε ένα γενετικό αλγόριθμο γίνεται με τυχαίο τρόπο. Αφού όμως η μέθοδος χρονικής μεταβολής δίνει κάποια αρχική πληροφορία σχετικά με την κατανομή των τοπικών ελαχίστων, η παραπάνω προσέγγιση εκμεταλλεύεται το χρονικό συσχετισμό των διανυσμάτων περιγραφής και αυξάνει την πιθανότητα εντοπισμού συνόλων διανυσμάτων περιγραφής με μικρή συσχέτιση από τα πρώτα κιάλας βήματα του γενετικού αλγορίθμου.

Εκτίμηση Ικανότητας

Για την εκτίμηση της επίδοσης όλων των χρωμοσωμάτων \mathbf{x}_i , $i = 1, \dots, P$, ενός δεδομένου πληθυσμού $\mathbf{X}(n)$, $n \geq 0$, χρησιμοποιείται το μέτρο συσχέτισης $R(\mathbf{x})$ ως *αντικειμενική συνάρτηση* (objective function). Όμως στη συνέχεια χρησιμοποιείται μία *συνάρτηση ικανότητας* (fitness function) για την απεικόνιση αντικειμενικών τιμών σε τιμές ικανότητας. Η απεικόνιση αυτή γίνεται με τη μέθοδο της *κανονικοποίησης ταξινόμησης* (rank-based normalization scheme). Συγκεκριμένα, τα χρωμοσώματα \mathbf{x}_i , $i = 1, \dots, P$, ταξινομούνται κατά αύξουσα σειρά του μέτρου συσχέτισης $R(\mathbf{x}_i)$, αφού η αντικειμενική συνάρτηση πρέπει να ελαχιστοποιηθεί. Έστω ότι $\text{rank}(\mathbf{x}_i) \in \{1, \dots, P\}$ είναι η σειρά ταξινόμησης του χρωμοσώματος \mathbf{x}_i ($\text{rank} = 1$ αντιστοιχεί στο καλύτερο χρωμόσωμα και $\text{rank} = P$ στο χειρότερο). Ορίζοντας αυθαίρετα την τιμή ικανότητας του καλύτερου χρωμοσώματος ως F_B , η ικανότητα $F(\mathbf{x}_i)$ του χρωμοσώματος \mathbf{x}_i δίνεται από τη γραμμική συνάρτηση

$$F(\mathbf{x}_i) = F_B - (\text{rank}(\mathbf{x}_i) - 1) D_R, \quad i = 1, \dots, P \quad (7-14)$$

όπου D_R είναι ο *ρυθμός μείωσης* (decrement rate) της τιμής ικανότητας. Το βασικό πλεονέκτημα της μεθόδου κανονικοποίησης ταξινόμησης είναι ότι, αφού οι τιμές ικανότητας είναι ομοιόμορφα κατανομημένες, εμποδίζει τη δημιουργία *υπερ-χρωμοσωμάτων* (super chromosomes), αποφεύγοντας έτσι την πρόωρη σύγκλιση σε τοπικά ελάχιστα. Επιπλέον, με τη ρύθμιση των δύο παραμέτρων F_B και D_R , μπορεί κανείς να ελέγξει την *τάση επιλογής* (selective pressure) του αλγορίθμου, επηρεάζοντας ουσιαστικά την ταχύτητα σύγκλισης.

Επιλογή Γονέων

Αφού υπολογισθεί η συνάρτηση ικανότητας $F(\mathbf{x}_i)$, $i = 1, \dots, P$, για όλα τα μέλη του πληθυσμού, πραγματοποιείται *επιλογή γονέων* (parent selection) με τρόπο τέτοιο ώστε ένα ικανότερο χρωμόσωμα να δώσει μεγαλύτερο αριθμό απογόνων και επομένως να έχει μεγαλύτερη πιθανότητα επιβίωσης στην επόμενη γενιά. Συγκεκριμένα, για την επιλογή γονέων χρησιμοποιείται η *μέθοδος της ρουλέτας* (roulette wheel selection), η οποία δίνει σε κάθε χρωμόσωμα \mathbf{x}_i , $i = 1, \dots, P$, πιθανότητα επιλογής p_i ανάλογη με την τιμή ικανότητάς του:

$$p_i = \frac{F(\mathbf{x}_i)}{\sum_{j=1}^P F(\mathbf{x}_j)}, \quad i = 1, \dots, P \quad (7-15)$$

Για την επιλογή ενός υποσυνόλου Q γονέων ($2 \leq Q \leq P$) από το σύνολο του πληθυσμού $\mathbf{X}(n) = \{\mathbf{x}_1, \dots, \mathbf{x}_P\}$, $n \geq 0$, κατασκευάζονται αρχικά οι αθροιστικές πιθανότητες

$$q_i = \sum_{j=1}^i p_j, \quad i = 1, \dots, P \quad (7-16)$$

με $q_0 = 0$, και στη συνέχεια παράγεται ένα σύνολο τυχαίων αριθμών r_k , $k = 1, \dots, Q$, με ομοιόμορφη κατανομή πιθανότητας στο διάστημα $[0,1]$. Τότε οι γονείς που επιλέγονται είναι

$$\mathbf{p}_k = \mathbf{x}_i \in \mathbf{X}(n) : q_{i-1} < r_k \leq q_i, \quad k = 1, \dots, Q \quad (7-17)$$

Σημειώνεται ότι με τη μέθοδο αυτή ένα χρωμόσωμα με μεγάλη ικανότητα μπορεί να επιλεγεί παραπάνω από μία φορές ως γονέας και έτσι να δώσει μεγαλύτερο αριθμό απογόνων σε σχέση με άλλα πιο αδύναμα χρωμοσώματα. Η μέθοδος της ρουλέτας είναι μία από τις πλέον δημοφιλείς μεθόδους επιλογής γονέων, διότι εγγυάται ότι κάθε χρωμόσωμα έχει ρυθμό ανάπτυξης ανάλογο με την ικανότητά του. Παρατηρείται επίσης ότι, λόγω της κανονικοποίησης ταξινομήσης, οι πιθανότητες επιλογής καθώς και οι αθροιστικές πιθανότητες p_i , q_i , $i = 1, \dots, P$, παραμένουν σταθερές σε κάθε γενιά.

Ανταλλαγή Γονιδίων

Στη συνέχεια πραγματοποιείται ζευγάρωμα των επιλεγμένων γονέων και παραγωγή ενός συνόλου απογόνων (offspring) με τη διαδικασία *ανταλλαγής γονιδίων* (crossover) μεταξύ ζευγών γονέων. Το γενετικό υλικό των γονέων συνδυάζεται με ένα τυχαίο τρόπο και παράγει το γενετικό υλικό των απογόνων. Για παράδειγμα, θεωρώντας ένα απλό σημείο ανταλλαγής γονιδίων (single point crossover), δύο γονείς

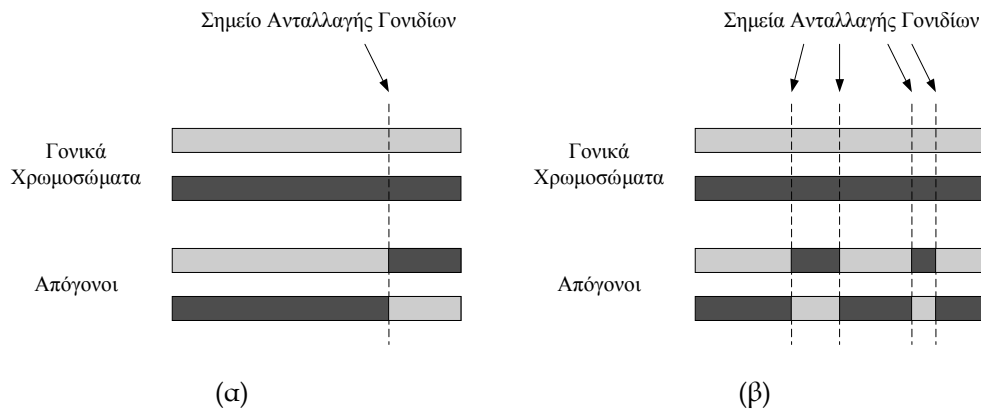
$$\mathbf{a} = (a_1, a_2, \dots, a_c, a_{c+1}, \dots, a_{K_F})$$

$$\mathbf{b} = (b_1, b_2, \dots, b_c, b_{c+1}, \dots, b_{K_F})$$

δημιουργούν τους απογόνους

$$\mathbf{a}' = (a_1, a_2, \dots, a_c, b_{c+1}, \dots, b_{K_F})$$

$$\mathbf{b}' = (b_1, b_2, \dots, b_c, a_{c+1}, \dots, a_{K_F})$$



Σχήμα 7-2. Ανταλλαγή γονιδίων. (α) Απλό σημείο ανταλλαγής. (β) Πολλαπλά σημεία ανταλλαγής.

Η παραπάνω διαδικασία μπορεί να γίνει και με *πολλαπλά σημεία ανταλλαγής* (multipoint crossover). Ένα παράδειγμα απλού σημείου ανταλλαγής γονιδίων παρουσιάζεται στο Σχήμα 7-2α, ενώ στο Σχήμα 7-2β δίνεται αντίστοιχα ένα παράδειγμα πολλαπλών σημείων. Μία πιο γενικευμένη τεχνική, που χρησιμοποιείται και στα πειράματα, είναι η *ομοιόμορφη ανταλλαγή γονιδίων* (uniform crossover), στην οποία κάθε γονίδιο θεωρείται ως πιθανό σημείο ανταλλαγής. Αυτό σημαίνει ότι δύο γονείς

$$\mathbf{a}_0 = (a_1^0, a_2^0, \dots, a_K^0)$$

$$\mathbf{a}_1 = (a_1^1, a_2^1, \dots, a_K^1)$$

δημιουργούν τους απογόνους

$$\mathbf{a}'_0 = (a_1^{s_1}, a_2^{s_2}, \dots, a_K^{s_K})$$

$$\mathbf{a}'_1 = (a_1^{1-s_1}, a_2^{1-s_2}, \dots, a_K^{1-s_K})$$

όπου το σύμβολο K_F αντικαταστάθηκε με K για απλότητα συμβολισμών, και οι αριθμοί s_i , $i = 1, \dots, K_F$ είναι τυχαίοι αριθμοί που παίρνουν τιμή 0 ή 1 με ίση πιθανότητα, έτσι ώστε κάθε γονίδιο να προέρχεται από τον πρώτο ή το δεύτερο γονέα με ίση πιθανότητα. Παρά το γεγονός ότι η ανταλλαγή απλού σημείου θεωρείται κατώτερη από άλλες τεχνικές, δεν έχουν αναφερθεί αποτελέσματα υπέρ της ομοιόμορφης ανταλλαγής, της ανταλλαγής πολλαπλών σημείων ή άλλων τεχνικών (όπως arithmetical, segmented ή shuffle crossover) [88]. Αντίθετα, η επιλογή εξαρτάται σε μεγάλο βαθμό από τη φύση του συγκεκριμένου προβλήματος βελτιστοποίησης, και στην περίπτωση της εξαγωγής χαρακτηριστικών καρέ η ομοιόμορφη ανταλλαγή έχει ελαφρώς καλύτερες επιδόσεις όσον αφορά στην ταχύτητα σύγκλισης.

Μετάλλαξη

Το επόμενο βήμα είναι η εφαρμογή *μετάλλαξης* (mutation) στα νέα χρωμοσώματα του πληθυσμού. Η διαδικασία της μετάλλαξης εισάγει ένα μικρό ποσοστό τυχαίων γονιδιακών μεταβολών σε κάθε χρωμόσωμα. Οι μεταβολές αυτές είναι χρήσιμες για την αποκατάσταση χαμένου γενετικού υλικού, αλλά κυρίως για την παραγωγή νέου υλικού που αντιστοιχεί σε νέες περιοχές αναζήτησης. Η *ομοιόμορφη μετάλλαξη* (uniform mutation) είναι το πιο συνηθισμένο είδος μετάλλαξης και επιλέγεται για το συγκεκριμένο πρόβλημα βελτιστοποίησης. Συγκεκριμένα, κάθε γονίδιο x_i , $i = 1, 2, \dots, K_F$, κάθε απογόνου, αντικαθίσταται από ένα τυχαίο γονίδιο $x'_i \in V = \{0, 1, \dots, N_F - 1\}$, με πιθανότητα p_m . Αυτό σημαίνει ότι παράγεται ένας τυχαίος αριθμός r με ομοιόμορφη κατανομή πιθανότητας στο διάστημα $[0,1]$, και η αντικατάσταση λαμβάνει χώρα αν $r < p_m$, αλλιώς το γονίδιο παραμένει ανέπαφο. Συνηθισμένες τιμές για την πιθανότητα μετάλλαξης p_m είναι μεταξύ 0.02 και 0.07. Άλλες εναλλακτικές τεχνικές είναι επίσης δυνατές, όπως οι nonuniform, boundary και swap mutation. Η *μη ομοιόμορφη* (nonuniform) μετάλλαξη προτιμάται γενικά σε αριθμητικά προβλήματα βελτιστοποίησης όσον αφορά στην ακρίβεια και στην ταχύτητα σύγκλισης, αλλά δεν επιτυγχάνει καλύτερη επίδοση στο συγκεκριμένο πρόβλημα της επιλογής χαρακτηριστικών καρέ.

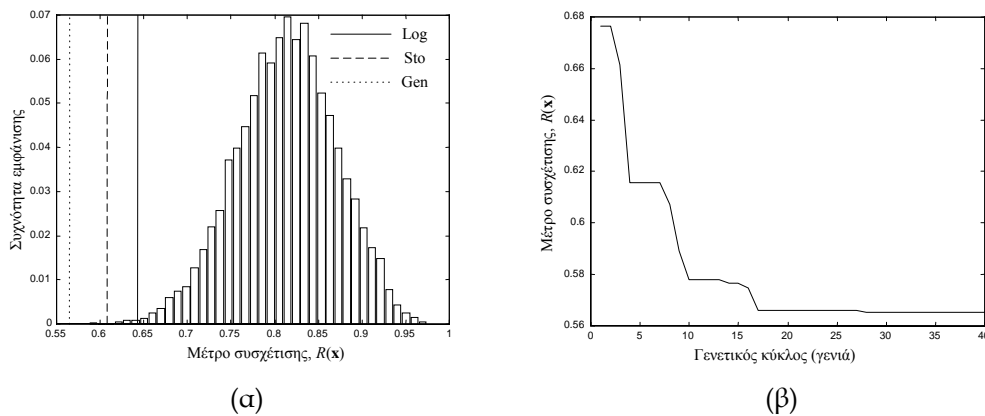
Αντικατάσταση Πληθυσμού

Αφού παραχθούν τα νέα χρωμοσώματα – απόγονοι για ένα δεδομένο πληθυσμό $\mathbf{X}(n)$, $n \geq 0$, ο πληθυσμός της επόμενης γενιάς, $\mathbf{X}(n+1)$, σχηματίζεται με την εισαγωγή αυτών των νέων χρωμοσωμάτων στο σύνολο $\mathbf{X}(n)$ και τη διαγραφή ενός κατάλληλου αριθμού παλαιών, έτσι ώστε ο πληθυσμός σε κάθε γενιά να αποτελείται από ακριβώς P χρωμοσώματα. Ο ακριβής αριθμός χρωμοσωμάτων που αντικαθίστανται από νέα καθορίζει τη *στρατηγική αντικατάστασης* (replacement strategy) του γενετικού αλγορίθμου. Στα πειράματα εξαγωγής χαρακτηριστικών καρέ χρησιμοποιήθηκε η *στρατηγική επίλεκτων* (elitist strategy), σύμφωνα με την οποία ένα μικρό ποσοστό των πιο ικανών χρωμοσωμάτων αντιγράφεται στον πληθυσμό της επόμενης γενιάς, μαζί με τους απογόνους τους. Η στρατηγική αυτή βελτιώνει σε μεγάλο βαθμό την ταχύ-

τητα σύγκλισης του αλγορίθμου [48]. Ένας αριθμός από *γενετικές περιόδους* (GA cycles) λαμβάνει χώρα με την επανάληψη των διαδικασιών εκτίμησης ικανότητας, επιλογής γονέων, ανταλλαγής γονιδίων, μετάλλαξης και αντικατάστασης πληθυσμού, μέχρι ο πληθυσμός να συγκλίνει σε μία λύση του προβλήματος αναζήτησης. Ο γενετικός αλγόριθμος τερματίζει όταν η ικανότητα του καλύτερου χρωμοσώματος παραμένει σταθερή για ένα μεγάλο αριθμό γενεών, υποδεικνύοντας ότι περαιτέρω βελτίωση είναι απίθανη.

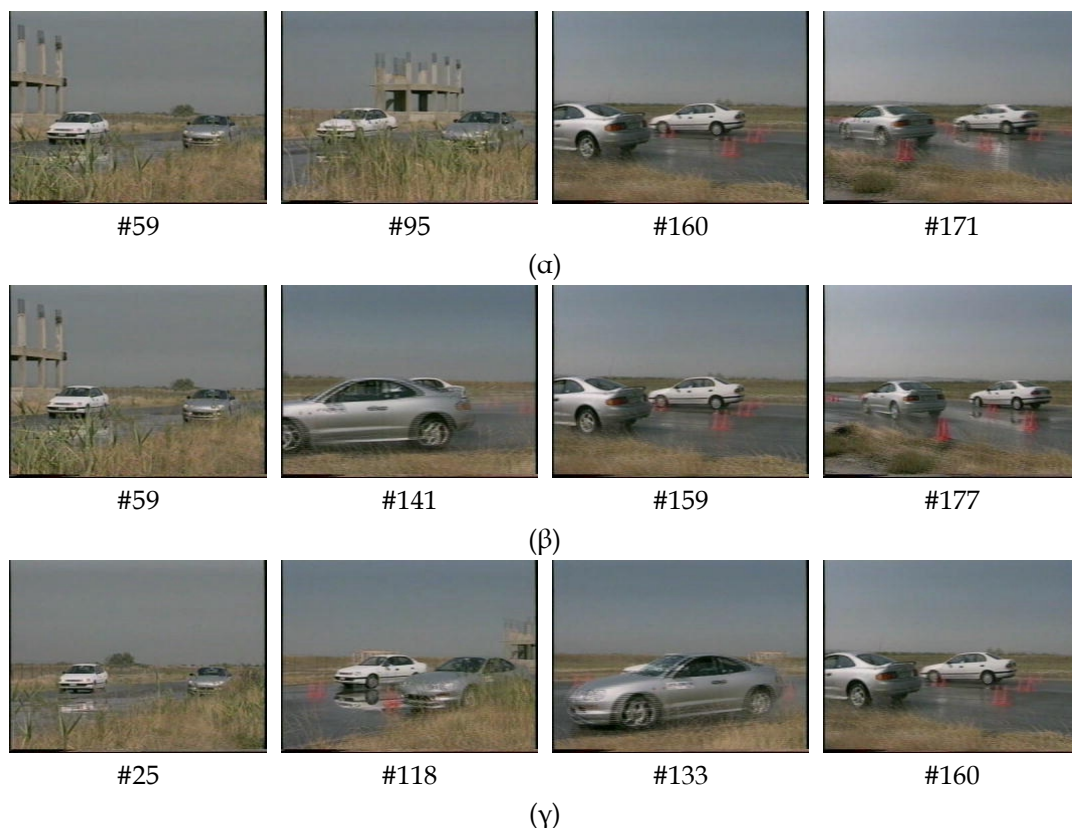
7.4 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για την εκτίμηση της επίδοσης των προτεινόμενων αλγορίθμων χρησιμοποιήθηκε η βάση δεδομένων που παρουσιάστηκε στην Ενότητα 6.4. Και πάλι οι χρονικές στιγμές αλλαγής πλάνου και τα διανύσματα περιγραφής κάθε καρέ είναι ήδη αποθηκευμένα στη βάση και διαθέσιμα για τα πειράματα επιλογής χαρακτηριστικών καρέ. Για την επίδειξη της μεθόδου συσχέτισης χρησιμοποιείται το πλάνο δοκιμαστικής οδήγησης αυτοκινήτων, διάρκειας 22 περίπου δευτερολέπτων ($N_F = 223$), που παρουσιάστηκε στη Σχήμα 6-5. Για το συγκεκριμένο πλάνο επιλέχθηκαν επτά χαρακτηριστικά καρέ με τη μέθοδο χρονικής μεταβολής, επομένως για τη μέθοδο συσχέτισης επιλέχθηκε $K_F = 4$, όπως εξηγείται στην Ενότητα 7.1.



Σχήμα 7-3. Μέθοδος συσχέτισης για την ακολουθία του Σχήματος 6-5. (α) Ιστογράμμο του μέτρου συσχέτισης $R(\mathbf{x})$, μαζί με τις βέλτιστες τιμές των τριών αλγορίθμων. Log: λογαριθμική αναζήτηση. Sto: στοχαστική αναζήτηση. Gen: γενετικός αλγόριθμος. (β) Ελάχιστη τιμή του $R(\mathbf{x})$, συναρτήσει της γενετικής περιόδου (γενιάς) για το γενετικό αλγόριθμο.

Παρατηρείται ότι όλοι οι προτεινόμενοι αλγόριθμοι υλοποίησης της μεθόδου συσχέτισης δίνουν εκτίμηση μόνο της ελάχιστης τιμής του μέτρου συσχέτισης $R(\mathbf{x})$, ενώ η πραγματική ελάχιστη τιμή δεν είναι γνωστή, αφού η πολυπλοκότητα υπολογισμού της είναι απαγορευτική. Σημειώνεται ότι στο συγκεκριμένο παράδειγμα το σύνολο των δυνατών λύσεων είναι περίπου 1.1×10^{10} . Έτσι, για να εκτιμηθεί η επίδοση των αλγορίθμων ως προς τον υπολογισμό του μέτρου συσχέτισης πραγματοποιήθηκε ένα πείραμα υπολογισμού του $R(\mathbf{x})$ για 100.000 τυχαία διανύσματα δεικτών \mathbf{x} και κατασκευής ενός ιστογράμματος του $R(\mathbf{x})$, όπως φαίνεται στο Σχήμα 7-3α. Το ιστογράμμο δίνει μία καλή εκτίμηση για την κατανομή του μέτρου συσχέτισης $R(\mathbf{x})$ και επομένως για το πόσο κοντά είναι μια τιμή στην πραγματική ελάχιστη. Στο Σχήμα 7-3α, μαζί με το ιστογράμμο του $R(\mathbf{x})$, φαίνονται ως κατακόρυφες γραμμές και οι εκτιμήσεις των τριών αλγορίθμων – λογαριθμική αναζήτηση, στοχαστική λογαριθμική αναζήτηση και γενετικός αλγόριθμος.



Σχήμα 7-4. Χαρακτηριστικά καρέ του πλάνου του Σχήματος 6-5, με τη μέθοδο συσχέτισης ($K_F = 4$).
(α) Λογαριθμική αναζήτηση. (β) Στοχαστική λογαριθμική αναζήτηση. (γ) Γενετικός αλγόριθμος.

Παρατηρείται αρχικά ότι εκτιμήσεις και των τριών αλγορίθμων είναι κοντά στο ελάχιστο της κατανομής του $R(x)$. Όμως ο αλγόριθμος λογαριθμικής αναζήτησης καθώς και η στοχαστική εκδοχή του δίνουν εκτίμηση μεγαλύτερη από το ελάχιστο του ιστογράμματος. Αντίθετα η εκτίμηση του γενετικού αλγορίθμου είναι όχι μόνο μικρότερη από τις άλλες δύο αλλά και από το ελάχιστο του πειράματος τυχαίων διανυσμάτων, παρά το γεγονός ότι το πείραμα αυτό απαιτεί περίπου 100 φορές μεγαλύτερο χρόνο εκτέλεσης. Επιβεβαιώνεται έτσι η ανωτερότητα του γενετικού αλγορίθμου ως προς την εκτίμηση της βέλτιστης επιλογής χαρακτηριστικών καρέ αλλά και ως προς την ταχύτητα.

Όσον αφορά στο γενετικό αλγόριθμο, στο Σχήμα 7-3β παρουσιάζεται η ελάχιστη τιμή, σε όλο τον πληθυσμό, του μέτρου συσχέτισης $R(x)$ συναρτήσει της γενετικής περιόδου, δηλαδή της γενιάς ή της επανάληψης του αλγορίθμου. Όπως αναμένεται, το $R(x)$ ελαττώνεται όσο η γενιά αυξάνεται, μέχρι να φτάσει μια ελάχιστη τιμή στη γενιά 40, μετά την οποία παραμένει σταθερό. Στο συγκεκριμένο πείραμα η στρατηγική αντικατάστασης είναι τέτοια ώστε μόνο τα μισά χρωμοσώματα του πληθυσμού αντικαθίστανται από νέα σε κάθε γενιά. Έτσι, στις περιπτώσεις όπου όλοι οι απόγονοι έχουν μικρότερη τιμή ικανότητας από τους γονείς, το μέτρο συσχέτισης παραμένει σταθερό. Αυτό εξηγεί την «κλιμακωτή» μορφή της καμπύλης του Σχήματος 7-3β. Σημειώνεται επίσης ότι το πλάτος κάθε βήματος της καμπύλης αυξάνεται με την αύξηση της γενιάς, αφού το πλάτος αυτό συνδέεται άμεσα με την πιθανότητα περαιτέρω βελτιστοποίησης.

Τα τέσσερα επιλεγμένα χαρακτηριστικά καρέ για το πλάνο του παραδείγματος παρουσιάζονται στο Σχήμα 7-4 για τους τρεις προτεινόμενους αλγορίθμους. Παρά το γεγονός ότι επιλέγεται ένα πολύ μικρό ποσοστό των συνολικών καρέ του πλάνου, είναι εμφανές ότι σε ό-

λες τις περιπτώσεις μπορεί κάποιος να κατανοήσει το οπτικό περιεχόμενο του πλάνου μόνο από τα επιλεγμένα καρέ. Έτσι τα χαρακτηριστικά καρέ δίνουν μια ικανοποιητική αναπαράσταση του περιεχομένου με τη μικρότερη δυνατή ποσότητα πληροφορίας. Μολονότι μία σύγκριση των τριών αλγορίθμων με βάση το περιεχόμενο των χαρακτηριστικών καρέ θα ήταν μάλλον υποκειμενική, μπορεί κανείς να ισχυρισθεί ότι τα καρέ του γενετικού αλγόριθμου είναι πιο αντιπροσωπευτικά, αφού για παράδειγμα τα καρέ #59 και #95, καθώς και τα #160 και #171, της λογαριθμικής αναζήτησης μοιάζουν αρκετά μεταξύ τους. Το ίδιο συμβαίνει και με τα καρέ #159 και #171 της στοχαστικής λογαριθμικής αναζήτησης.

Αλγόριθμος	\bar{R}	\bar{T} (sec)
Ομοιόμορφη επιλογή	0.85	–
Τυχαία αναζήτηση	0.52	55.38
Λογαριθμική αναζήτηση	0.63	1.92
Στοχαστική αναζήτηση	0.59	12.43
Γενετικός αλγόριθμος	0.44	0.54

Πίνακας 7-1. Αποτελέσματα μεθόδου συσχέτισης με τους τρεις προτεινόμενους αλγορίθμους για το σύνολο των πλάνων της βάσης: μέση τιμή βέλτιστου μέτρου συσχέτισης, \bar{R} , και μέσος χρόνος εκτέλεσης, \bar{T} (sec).

Τα παραπάνω πειράματα επαναλήφθηκαν για το σύνολο των ακολουθιών της βάσης δεδομένων, έτσι ώστε η σύγκριση μεταξύ των αλγορίθμων να είναι αξιόπιστη. Σε κάθε πλάνο εφαρμόστηκε πρώτα η μέθοδος χρονικής μεταβολής για την εκτίμηση του πλήθους χαρακτηριστικών καρέ, καθώς και του αρχικού πληθυσμού για την περίπτωση του γενετικού αλγόριθμου. Στη συνέχεια ακολούθησε η μέθοδος συσχέτισης, χρησιμοποιώντας ως K_F το μισό από το πλήθος καρέ που ανιχνεύθηκαν με τη μέθοδο χρονικής μεταβολής. Η μέση τιμή, \bar{R} , του βέλτιστου μέτρου συσχέτισης, παρουσιάζεται στον Πίνακα 7-1, μαζί με το μέσο χρόνο εκτέλεσης, \bar{T} , για κάθε αλγόριθμο. Στον Πίνακα παρουσιάζεται επίσης για σκοπούς σύγκρισης και η τυχαία αναζήτηση με χρήση 100,000 τυχαίων διανυσμάτων για κάθε πλάνο, καθώς και η περίπτωση επιλογής χαρακτηριστικών καρέ σε ομοιόμορφα χρονικά διαστήματα· η τελευταία έχει σχεδόν μηδενικό χρόνο εκτέλεσης, αλλά αποτυγχάνει να ανιχνεύσει ασυσχέτιστα καρέ.

Συμπεραίνεται ότι ο γενετικός αλγόριθμος έχει καλύτερη απόδοση όσον αφορά στην ακρίβεια των αποτελεσμάτων αλλά και την ταχύτητα. Σημειώνεται όμως ότι η μέθοδος χρονικής μεταβολής εξακολουθεί να είναι χρήσιμη σε περιπτώσεις περιοδικής επανάληψης του οπτικού περιεχομένου, αλλά και ως εργαλείο προεπεξεργασίας για την εκτίμηση της πολυπλοκότητας κάθε πλάνου. Πρέπει τέλος να τονισθεί ότι η εξαγωγή χαρακτηριστικών καρέ με τη μέθοδο συσχέτισης βασίζεται σε ένα αντικειμενικό, αριθμητικό κριτήριο, δηλαδή στην ελαχιστοποίηση του μέτρου συσχέτισης μεταξύ των διανυσμάτων περιγραφής των καρέ. Για την εκτίμηση όμως της απόδοσης των προτεινόμενων μεθόδων, θα μπορούσαν να χρησιμοποιηθούν υποκειμενικά κριτήρια, τα οποία να λαμβάνουν υπόψη την ανθρώπινη αντίληψη. Σε αυτή την περίπτωση, τα επιλεγμένα χαρακτηριστικά καρέ θα μπορούσαν να συγκριθούν με εκείνα που επιλέγονται από μία ομάδα ανθρώπων, ώστε να υπάρχει ένδειξη για το ποια μέθοδος δίνει αποτελέσματα πιο κοντά στην ανθρώπινη κρίση.

Αναζήτηση με Βάση το Περιεχόμενο

Η προτεινόμενη αναπαράσταση του οπτικού περιεχομένου μέσω διανυσμάτων περιγραφής επιτρέπει την ταχεία και αξιόπιστη σύγκριση μεταξύ διαφορετικών εικόνων, καρτέ ή πλάνων μιας ακολουθίας, αφού παρέχει ερμηνεία της οπτικής πληροφορίας που είναι κοντά στην ανθρώπινη αντίληψη. Έτσι με τη δεικτοδότηση του οπτικοακουστικού υλικού με διανύσματα περιγραφής και τον προσδιορισμό ενός μέτρου ομοιότητας διανυσμάτων είναι δυνατή η αναζήτηση και ανάκληση με βάση το περιεχόμενο. Επιπλέον, περιορίζοντας την αναζήτηση στα επιλεγμένα χαρακτηριστικά πλάνα και καρτέ, η αναζήτηση από ακολουθίες βίντεο ουσιαστικά ανάγεται σε αναζήτηση ακίνητων εικόνων και μάλιστα είναι πολύ πιο αποδοτική σε σχέση με την αναζήτηση από το σύνολο του διαθέσιμου υλικού. Προτείνονται δύο εναλλακτικοί μηχανισμοί αναζήτησης: αναζήτηση μέσω παραδείγματος και αναζήτηση μέσω σχεδιαγράμματος. Στον πρώτο ο χρήστης υποβάλλει μία διαθέσιμη εικόνα ως είσοδο στο σύστημα ανάκλησης, ενώ στο δεύτερο σχεδιάζει ένα προσεγγιστικό σκίτσο της εικόνας, τονίζοντας τα χαρακτηριστικά που τον ενδιαφέρουν. Επιπλέον, και στις δύο περιπτώσεις υιοθετείται μία τεχνική ανάδρασης από το χρήστη, η οποία καθιστά την αναζήτηση μία διαδικασία αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστικού συστήματος. Με την τεχνική αυτή ο χρήστης έχει τη δυνατότητα να επιλέξει ένα υποσύνολο των αποτελεσμάτων αναζήτησης που χαρακτηρίζει ως πλέον «κατάλληλα». Η πληροφορία αυτή ανατροφοδοτείται στο σύστημα για την αναπροσαρμογή του μέτρου ομοιότητας· έτσι ο χρήστης απαλλάσσεται από την επιλογή παραμέτρων ή την κατανόηση χαρακτηριστικών χαμηλού επιπέδου.

8.1 ΜΗΧΑΝΙΣΜΟΣ ΑΝΑΖΗΤΗΣΗΣ

Ο χαρακτηρισμός κάθε καρτέ και κάθε πλάνου μιας ακολουθίας βίντεο με διανύσματα περιγραφής και η επιλογή χαρακτηριστικών καρτέ και πλάνων, έτσι ώστε να απορριπτεται κάθε περιττή πληροφορία, επιτρέπει τη δημιουργία μιας αποδοτικής αναπαράστασης της ακολουθίας. Αν υπάρχει διαθέσιμο ένα μεγάλο σύνολο ακολουθιών σε μία βάση δεδομένων, η αναπαράσταση αυτή περιλαμβάνει τα διανύσματα περιγραφής των επιλεγμένων καρτέ / πλάνων κάθε ακολουθίας. Η αποθήκευση αυτής της πληροφορίας στη βάση επιτρέπει την αναζήτηση (query ή search) και ανάκληση με βάση το περιεχόμενο (content-based retrieval) για την εύρεση εικόνων ή πλάνων με συγκεκριμένες ιδιότητες, όπως για παράδειγμα καρτέ με σκοτεινά χρώματα, καρτέ με έντονη κίνηση κλπ.

Στο σημείο αυτό, το πρόβλημα της αναζήτησης με βάση το περιεχόμενο από μία βάση με ακολουθίες βίντεο έχει ουσιαστικά αναχθεί σε αναζήτηση ακίνητων εικόνων [13,60,62,72], αφού η αναζήτηση περιορίζεται στα διανύσματα περιγραφής των επιλεγμένων χαρακτηριστικών καρτέ / πλάνων και δεν εκτελείται σε όλο το διαθέσιμο οπτικοακουστικό υλικό. Στο πλαίσιο της διατριβής διερευνώνται δύο περιπτώσεις αναζήτησης: αναζήτηση μέσω παραδείγματος (query-by-example) και αναζήτηση μέσω σχεδιαγράμματος (query-by-sketch). Στην πρώτη περι-

πτώση ο χρήστης παρουσιάζει μια εικόνα ως είσοδο στο σύστημα αναζήτησης, ενώ στη δεύτερη σχεδιάζει ένα σκίτσο της επιθυμητής εικόνας.

Και στις δύο περιπτώσεις το σύστημα αναζήτησης βελτιώνεται με τη χρήση ενός μηχανισμού *ανάδρασης* (relevance feedback) ο οποίος με την αλληλεπίδραση με το χρήστη παρέχει ένα τρόπο απεικόνισης των υποκειμενικών κριτηρίων αναζήτησης σε χαμηλού επιπέδου χαρακτηριστικά και αυτόματης ενημέρωσης των παραμέτρων απόστασης για προσαρμογή του μηχανισμού αναζήτησης στις ανάγκες του χρήστη. Σημειώνεται ότι η ανάκληση μέσω παραδείγματος και σχεδιαγράμματος, καθώς και η τεχνική της ανάδρασης είναι γνωστές [106,107]. Όμως η προτεινόμενη στο πλαίσιο της διατριβής τεχνική ενημέρωσης παραμέτρων προτάθηκε για πρώτη φορά στην εργασία [13] και χρησιμοποιήθηκε επίσης και στην εργασία [35].

Αναζήτηση Μέσω Παραδείγματος

Σε ένα περιβάλλον *αναζήτησης μέσω παραδείγματος* ο χρήστης υποβάλλει *ερωτήματα* (queries) με τη μορφή στατικών εικόνων (καρέ) ή μικρών ακολουθιών (πλάνων), τα οποία αναλύονται από το σύστημα με τον ίδιο ακριβώς τρόπο όπως και το περιεχόμενο της βάσης, και στη συνέχεια τα παραγόμενα διανύσματα περιγραφής συγκρίνονται με τα διανύσματα που περιέχονται στη βάση. Ο χώρος των διανυσμάτων περιγραφής είναι ιδανικός για τέτοιες συγκρίσεις, καθώς περιέχει όλα τα ουσιαστικά χαρακτηριστικά των εικόνων, ενώ η διάστασή του είναι πολύ μικρότερη από εκείνη του αντίστοιχου χώρου των ίδιων των εικόνων. Επιπλέον, επιτυγχάνεται δραματική μείωση στον αριθμό των καρέ που απαιτούνται για αναζήτηση. Αντί να εξετάζεται κάθε καρέ μιας ακολουθίας, η αναζήτηση εκτελείται στο μικρό σύνολο των χαρακτηριστικών καρέ / πλάνων τα οποία παρέχουν αξιόπιστη αναπαράσταση του οπτικού περιεχομένου της ακολουθίας.

Για κάθε ακίνητη εικόνα ή πλάνο (μικρό σύνολο από διαδοχικά καρέ) που δίνεται από το χρήστη ως είσοδος, εξάγεται η πληροφορία χρώματος και κίνησης με τον αλγόριθμο κατάτμησης M-RSST. Στη συνέχεια τα χαρακτηριστικά των τμημάτων συγκεντρώνονται σε ένα διάνυσμα περιγραφής με τη μέθοδο του ασαφούς ιστογράμματος και ενεργοποιείται ο μηχανισμός αναζήτησης. Συγκεκριμένα, το σύνολο του υλικού που είναι διαθέσιμο στη βάση ερευνάται για καρέ ή πλάνα με ιδιότητες παρόμοιες με το παράδειγμα του χρήστη· τα M καλύτερα καρέ / πλάνα επιλέγονται και επιστρέφονται στο χρήστη ως αποτέλεσμα της αναζήτησης. Η έρευνα υλοποιείται υπολογίζοντας το διάνυσμα περιγραφής x του καρέ / πλάνου εισόδου και εκτελώντας συγκρίσεις ανάμεσα στο x και στα διανύσματα περιγραφής y όλων των καρέ / πλάνων της βάσης.

Όπως και στις παραδοσιακές εφαρμογές ανάκλησης ακίνητων εικόνων, χρησιμοποιείται μία *απόσταση* (distance) ή *μέτρο ομοιότητας* (similarity measure) για την εύρεση του συνόλου που ταιριάζουν καλύτερα με την εικόνα-παράδειγμα του χρήστη. Όμως, επειδή μπορεί να απαιτούνται από το χρήστη διαφορετικά κριτήρια δεικτοδότησης ή αναζήτησης, όπως αναζήτηση με βάση το χρώμα, την κίνηση, το σχήμα ή την υφή, ορισμένα στοιχεία των διανυσμάτων περιγραφής πρέπει να λαμβάνονται υπόψη σε μικρότερο ή μεγαλύτερο βαθμό κατά την εκτίμηση του μέτρου ομοιότητας.

Γι' αυτό το λόγο υιοθετείται μία *παραμετρική απόσταση* (parametric distance) ή *απόσταση με βάρη* (weighted distance), αυξάνοντας έτσι την ευελιξία του μηχανισμού ανάκλησης. Αυτό σημαίνει ότι ανατίθεται ένα σύνολο από *βάρη* (weights) στο μέτρο ομοιότητας και ο χρήστης

έχει τη δυνατότητα να προσδιορίσει τα βάρη ανάλογα με τις συγκεκριμένες ανάγκες για αναζήτηση πληροφορίας. Πιο συγκεκριμένα, η παραμετρική απόσταση ανάμεσα στο διάνυσμα περιγραφής \mathbf{x} της εισόδου και στα διανύσματα περιγραφής \mathbf{y} όλων χαρακτηριστικών καρτέ / πλάνων που είναι διαθέσιμα στη βάση, ορίζεται ως

$$d_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^N w_j (x_j - y_j)^2 = \sum_{j=1}^N w_j e_j^2 \quad (8-1)$$

όπου \mathbf{w} είναι το *διάνυσμα βαρών* (weight vector) ή *διάνυσμα παραμέτρων* (parameter vector), διαστάσεων $N \times 1$, N είναι το μήκος κάθε διανύσματος περιγραφής, $\mathbf{e} = \mathbf{x} - \mathbf{y}$ είναι ένα *διάνυσμα σφάλματος* (error vector) και x_j , y_j , w_j και e_j είναι τα στοιχεία των διανυσμάτων \mathbf{x} , \mathbf{y} , \mathbf{w} και \mathbf{e} αντίστοιχα. Το σύνολο των M χαρακτηριστικών καρτέ / πλάνων που αντιστοιχούν στα διανύσματα περιγραφής \mathbf{y}_i , $i = 1, \dots, M$, με την ελάχιστη απόσταση $d_{\mathbf{w}}(\mathbf{x}, \mathbf{y}_i)$ επιστρέφεται στο χρήστη ως αποτέλεσμα της αναζήτησης.

Αναζήτηση Μέσω Σχεδιαγράμματος

Σε ένα περιβάλλον *αναζήτησης μέσω σχεδιαγράμματος* ο χρήστης σχεδιάζει το σκίτσο μιας εικόνας και το υποβάλλει σαν ερώτημα στο σύστημα αναζήτησης. Το σκίτσο μπορεί να περιέχει μικρό αριθμό στοιχειωδών αντικειμένων με απλό γεωμετρικό σχήμα (όπως ορθογώνια, πολύγωνα, κύκλοι ή ελλείψεις) και συγκεκριμένες ιδιότητες όπως θέση, χρώμα, κίνηση και μέγεθος. Εναλλακτικά το σκίτσο μπορεί να περιλαμβάνει απλώς το περίγραμμα ενός αντικειμένου και η αναζήτηση να εκτελεσθεί με βάση το σχήμα του περιγράμματος· η περίπτωση αυτή εξετάζεται στο Κεφάλαιο 11.

Στο σενάριο αυτό τα αντικείμενα που σχεδιάζονται από το χρήστη μπορούν θεωρηθούν τμήματα εικόνας που προκύπτουν από κατάτμηση μιας εικόνας με βάση το χρώμα ή την κίνηση, ακριβώς όπως στην περίπτωση της αναζήτησης μέσω παραδείγματος. Αυτό σημαίνει ότι οι ιδιότητες των αντικειμένων μπορούν να χρησιμοποιηθούν άμεσα για την κατασκευή ενός πολυδιάστατου ασαφούς ιστογράμματος που περιγράφει το σκίτσο του χρήστη, σύμφωνα με τη μεθοδολογία του Κεφαλαίου 4. Το ιστόγραμμα είναι στην πράξη το διάνυσμα περιγραφής της εισόδου και έχει ακριβώς την ίδια μορφή με τα διανύσματα περιγραφής των χαρακτηριστικών καρτέ / πλάνων της βάσης. Συνεπώς, η αναζήτηση ανάγεται στην προηγούμενη περίπτωση και η άμεση σύγκριση του υπολογιζόμενου διανύσματος περιγραφής με τα διαθέσιμα διανύσματα της βάσης μπορεί να χρησιμοποιηθεί για σκοπούς ανάκλησης· η παρούσα περίπτωση μάλιστα είναι απλούστερη αφού παραλείπεται το στάδιο της κατάτμησης.

Όπως στην περίπτωση αναζήτησης μέσω παραδείγματος, οι συγκρίσεις πραγματοποιούνται στο χώρο των διανυσμάτων περιγραφής βάσει του μέτρου απόστασης της Εξίσωσης (8-1), ενώ η έρευνα περιορίζεται και πάλι στο σύνολο των επιλεγμένων χαρακτηριστικών καρτέ / πλάνων, επιτρέποντας ταχεία υλοποίηση της διαδικασίας αναζήτησης. Οι βασικές διαφορές των δύο προσεγγίσεων είναι ότι

- (α) δεν απαιτείται από το χρήστη η εισαγωγή μιας υπάρχουσας εικόνας ή πλάνου
- (β) η αναζήτηση περιορίζεται στα αντικείμενα που σχεδιάζονται και επομένως ενδιαφέρουν το χρήστη, και δεν επηρεάζεται από την ύπαρξη άλλων αντικειμένων που μπορεί να εμφανιστούν σε μια αναζήτηση μέσω παραδείγματος

Συνεπώς η αναζήτηση μέσω σχεδιαγράμματος τείνει να εκφράζει με μεγαλύτερη ακρίβεια τις πληροφοριακές ανάγκες του χρήστη. Επιπλέον, αυτό το είδος μπορεί σχετικά εύκολα να αναχθεί σε *αναζήτηση μέσω κειμένου* (textual query) χωρίς ουσιαστικές τροποποιήσεις στην υπολογιστική υποδομή. Για παράδειγμα, ένας χρήστης μπορεί να αναζητήσει ένα «σκούρο αντικείμενο που κινείται προς τα αριστερά» ή ένα «μεγάλο μπλε αντικείμενο στο κάτω άκρο της εικόνας». Στην περίπτωση αυτή με τη *συντακτική ανάλυση* (parsing) των ερωτημάτων μπορούν να προσδιορισθούν οι ιδιότητες των αντικειμένων που διαφορετικά θα υπολογίζονταν από το σκίτσο.

8.2 ΑΝΑΔΡΑΣΗ

Από τον ορισμό του διανύσματος περιγραφής ως πολυδιάστατο ιστόγραμμα και από τα παραδείγματα του Κεφαλαίου 4 προκύπτει ότι η πληροφορία για κάθε χαρακτηριστικό (θέση, χρώμα, κίνηση κλπ.) αποθηκεύεται αυτόνομα σε κάθε διάνυσμα· συνεπώς είναι δυνατή η αναζήτηση κάθε είδους πληροφορίας με κατάλληλη επιλογή βαρών στο μέτρο ομοιότητας (8-1). Αυτή η προσέγγιση όμως έχει το μειονέκτημα ότι ο χρήστης δεν έχει πάντα τη δυνατότητα να εκφράσει ένα ερώτημα βάσει χαρακτηριστικών χαμηλού επιπέδου, επομένως η διαδικασία επιλογής βαρών είναι μία επιπρόσθετη επιβάρυνση για το χρήστη και συνήθως οδηγεί σε κατώτερη απόδοση ανάκλησης. Γι' αυτό υιοθετείται μία προσέγγιση *ανάδρασης καταλληλότητας, σχετικότητας* ή απλώς *ανάδρασης* (relevance feedback) στην προτεινόμενη εφαρμογή ανάκλησης με βάση το περιεχόμενο, η οποία καθιστά την αναζήτηση μία διαδικασία *αλληλεπίδρασης* (interaction) μεταξύ ανθρώπου και υπολογιστικού συστήματος [106].

Σύμφωνα με την τεχνική της ανάδρασης ο χρήστης μετά από μία ανάκληση έχει επιπλέον τη δυνατότητα να επιλέξει ένα υποσύνολο των αντικειμένων που ανακλήθηκαν, το οποίο κρίνει ότι ταιριάζει περισσότερο στο ερώτημα που έθεσε. Έστω λοιπόν ότι ο χρήστης επιλέγει και χαρακτηρίζει ως «*κατάλληλα*» ή «*σχετικά*» (relevant) τα m καρέ / πλάνα από τα M που επιστράφηκαν από το σύστημα ως αποτελέσματα αναζήτησης. Τότε τα υπόλοιπα $M - m$ καρέ / πλάνα θεωρούνται λανθασμένα ή «*ακατάλληλα*» ή «*μη σχετικά*» (irrelevant). Σε ένα γενικότερο σενάριο, ο χρήστης θα μπορούσε να ταξινομήσει τα αποτελέσματα δίνοντας διαφορετικούς βαθμούς *καταλληλότητας* ή *σχετικότητας* (relevance). Έτσι η πληροφορία καταλληλότητας θα ήταν πιο ακριβής, επιβαρύνοντας βέβαια το χρήστη με την εισαγωγή περισσότερης πληροφορίας.

Η πληροφορία καταλληλότητας ανατροφοδοτείται στο σύστημα και χρησιμοποιείται για την αυτόματη ενημέρωση ή αναπροσαρμογή των βαρών w του μέτρου ομοιότητας, έτσι ώστε στην αμέσως επόμενη απόπειρα αναζήτησης τα αποτελέσματα να προσεγγίσουν καλύτερα τις αρχικές ανάγκες πληροφόρησης [18]. Με την τροποποίηση των βαρών άλλα στοιχεία του διανύσματος περιγραφής υπολογίζονται σε μεγαλύτερο και άλλα σε μικρότερο βαθμό, έτσι ώστε η νέα απόσταση του διανύσματος περιγραφής x της εισόδου από τα m κατάλληλα καρέ / πλάνα να είναι μικρότερη, ενώ η αντίστοιχη νέα απόσταση από τα $M - m$ ακατάλληλα καρέ / πλάνα να είναι μεγαλύτερη.

Συνεπώς, στην αμέσως επόμενη φάση αναζήτησης επιλέγονται περισσότερα καρέ / πλάνα που σχετίζονται με το ερώτημα του χρήστη, αφού οι παράμετροι του μέτρου ομοιότητας έχουν προσαρμοσθεί στις ιδιότητες των καρέ / πλάνων που χαρακτηρίζονται από το χρήστη ως κατάλληλα. Πρακτικά με αυτό τον τρόπο ο χρήστης απαλλάσσεται από το βάρος της επιλογής

παραμέτρων ή της κατανόησης χαρακτηριστικών και αναπαραστάσεων χαμηλού επιπέδου· παράλληλα το υπολογιστικό σύστημα εφοδιάζεται με ένα εργαλείο για την αυτόματη αντιστοίχιση των χαρακτηριστικών χαμηλού επιπέδου σε υψηλού επιπέδου προδιαγραφές αναζήτησης, καθώς και για την αντιμετώπιση της υποκειμενικότητας με την οποία ο χρήστης καθορίζει τις προδιαγραφές αυτές.

Στις περισσότερες περιπτώσεις, η *απλή ανάδραση* (single relevance feedback), δηλαδή η μία μόνο επανάληψη της διαδικασίας ανάκλησης / ενημέρωσης παραμέτρων / νέας ανάκλησης, είναι επαρκής για την ανάκληση του επιθυμητού υλικού. Σε μια πρακτική εφαρμογή όμως, μία απλή ενημέρωση των παραμέτρων του μέτρου ομοιότητας δεν αποδίδει πάντα βέλτιστα αποτελέσματα αναζήτησης σύμφωνα με τις προδιαγραφές του χρήστη. Σε τέτοιες περιπτώσεις ο χρήστης πρέπει να έχει τη δυνατότητα να αναπροσαρμόσει και πάλι το μηχανισμό αναζήτησης, επιλέγοντας ένα δεύτερο σύνολο κατάλληλων αποτελεσμάτων. Τότε απαιτείται μία δεύτερη φάση ενεργοποίησης του μηχανισμού ενημέρωσης παραμέτρων καθώς και μία ακόμη φάση ανάκλησης. Η επανάληψη της παραπάνω διαδικασίας οδηγεί τελικά στην *πολλαπλή ανάδραση* (multiple relevance feedback). Στην επόμενη Ενότητα προτείνεται ένας αναδρομικός αλγόριθμος για την ενημέρωση των παραμέτρων που διακρίνεται ανάλογα με τον τύπο της ανάδρασης.

8.3 ΕΝΗΜΕΡΩΣΗ ΠΑΡΑΜΕΤΡΩΝ

Ο στόχος του μηχανισμού ενημέρωσης παραμέτρων είναι να μειωθεί η απόσταση του διανύσματος περιγραφής \mathbf{x} της εισόδου από τα διανύσματα που χαρακτηρίζονται «κατάλληλα» και ταυτόχρονα να αυξηθεί η απόσταση του από τα διανύσματα που χαρακτηρίζονται «ακατάλληλα». Ο προτεινόμενος αλγόριθμος ενημέρωσης για την απλή ανάδραση περιγράφεται αμέσως παρακάτω, ενώ για την περίπτωση πολλαπλής ανάδρασης ο αντίστοιχος αλγόριθμος μελετάται στη συνέχεια.

Απλή Ανάδραση

Έστω ότι $\mathbf{y}_i, i = 1, \dots, M$, είναι τα διανύσματα περιγραφής των καρτέ ή πλάνων που ανακαλούνται από τη βάση κατά την πρώτη αναζήτηση. Χωρίς βλάβη της γενικότητας, έστω επίσης ο χρήστης επιλέγει ως «κατάλληλα» τα διανύσματα $\mathbf{y}_i, i = 1, \dots, m$, όπου $m < M$. Τότε η απόσταση ανάμεσα στο διάνυσμα περιγραφής \mathbf{x} της εισόδου και τα διανύσματα $\mathbf{y}_i, i = 1, \dots, m$, πρέπει να ελαχιστοποιηθεί, ενώ η απόσταση του \mathbf{x} από τα $\mathbf{y}_i, i = m+1, \dots, M$, πρέπει να μεγιστοποιηθεί στην επόμενη αναζήτηση. Συνεπώς η *συνάρτηση κόστους* (cost function) που ορίζεται ως

$$J(\mathbf{w}) = \sum_{i=1}^m d_{\mathbf{w}}(\mathbf{x}, \mathbf{y}_i) - \sum_{i=m+1}^M d_{\mathbf{w}}(\mathbf{x}, \mathbf{y}_i) \quad (8-2)$$

πρέπει να ελαχιστοποιηθεί ως προς το διάνυσμα παραμέτρων \mathbf{w} , υπό τον περιορισμό ότι το μέτρο του διανύσματος \mathbf{w} παραμένει σταθερό. Ο περιορισμός είναι απαραίτητος διότι το μέτρο απόστασης χρησιμοποιείται μόνο σε συγκρίσεις, οπότε διανύσματα παραμέτρων που διαφέρουν μόνο ως προς το μέτρο τους δίνουν ακριβώς τα ίδια αποτελέσματα αναζήτησης. Χωρίς βλάβη της γενικότητας, έστω $\|\mathbf{w}\| = 1$:

$$\hat{\mathbf{w}} = \arg \min_{\|\mathbf{w}\|=1} J(\mathbf{w}) \quad (8-3)$$

Η βέλτιστη λύση της (8-3) μπορεί να προκύψει αναλυτικά σε κλειστή μορφή θέτοντας $\partial J(\mathbf{w})/\partial w_k = 0$ για $k = 1, \dots, N$. Η λύση είναι [13]:

$$\hat{w}_k = A_k \left(\sum_{l=1}^N A_l^2 \right)^{-1/2}, \quad k = 1, \dots, N \quad (8-4)$$

όπου οι όροι $A_k, k = 1, \dots, N$, ορίζονται ως

$$A_k = \sum_{i=1}^m (x_k - y_k^{(i)})^2 - \sum_{i=m+1}^M (x_k - y_k^{(i)})^2, \quad k = 1, \dots, N \quad (8-5)$$

και $y_k^{(i)}, k = 1, \dots, N$, είναι τα στοιχεία του διανύσματος \mathbf{y}_i . Στην παραπάνω ανάλυση έγινε επιπλέον η υπόθεση ότι ο χρήστης θα επιλέξει τουλάχιστον μία εικόνα ως σχετική με την αναζήτησή του, έτσι ώστε να υπάρχει τρόπος εντοπισμού των στοιχείων εκείνων του διανύσματος περιγραφής που αντιστοιχούν στο κριτήριο αναζήτησης (για παράδειγμα, χρώμα, μέγεθος ή κίνηση αντικειμένων). Η υπόθεση αυτή είναι λογική καθώς στην αντίθετη περίπτωση ο χρήστης θα ξεκινούσε νέα αναζήτηση αφού τα αποτελέσματα της πρώτης αναζήτησης θα ήταν όλα ακατάλληλα.

Πολλαπλή Ανάδραση

Με την παραπάνω διαδικασία ελαχιστοποίησης ουσιαστικά υπολογίζονται τα βέλτιστα βάρη \mathbf{w} για μία συγκεκριμένη εικόνα εισόδου (καρέ ή πλάνο) με διάνυσμα περιγραφής \mathbf{x} . Στην πράξη όμως, μία απλή ενημέρωση του διανύσματος παραμέτρων \mathbf{w} δεν αποδίδει πάντα βέλτιστα αποτελέσματα αναζήτησης σύμφωνα με τις προδιαγραφές του χρήστη. Τότε είναι απαραίτητη η *πολλαπλή ανάδραση* η οποία υλοποιείται μέσω πολλαπλών, διαδοχικών φάσεων αναζήτησης ανάμεσα στις οποίες μεσολαβεί η κατάλληλη ενημέρωση του \mathbf{w} .

Σε αυτή τη γενικότερη περίπτωση, τα διανύσματα εισόδου, \mathbf{x} , και εξόδου, $\mathbf{y}_i, i = 1, \dots, M$, μπορούν να θεωρηθούν ακολουθίες διακριτού χρόνου $\mathbf{x}(n)$ και $\mathbf{y}_i(n), i = 1, \dots, M$, αντίστοιχα. Συνεπώς ο αλγόριθμος ενημέρωσης παραμέτρων της περίπτωσης απλής ανάδρασης θα μπορούσε να εφαρμοσθεί αναδρομικά και στην παρούσα περίπτωση, παράγοντας μία νέα εκτίμηση του διανύσματος παραμέτρων σε κάθε επανάληψη. Όμως τα αποτελέσματα των προηγούμενων αναζητήσεων καθώς και οι αντίστοιχες τροποποιήσεις του διανύσματος παραμέτρων θα έπρεπε να ληφθούν επίσης υπόψη σε μικρό βαθμό κατά τη νέα ενημέρωση. Για το σκοπό αυτό εισάγουμε έναν *παράγοντα μνήμης* (memory factor) λ , όπου $0 < \lambda < 1$ με τον οποίο πολλαπλασιάζουμε τα αποτελέσματα των προηγούμενων σταδίων βελτιστοποίησης. Έτσι οι Εξισώσεις (8-4), (8-5) τροποποιούνται ως εξής [13]:

$$\hat{w}_k(n) = B_k(n) \left(\sum_{l=1}^N B_l^2(n) \right)^{-1/2}, \quad k = 1, \dots, N \quad (8-6)$$

όπου

$$B_k(n) = \sum_{j=0}^{\infty} \lambda^j A_k(n-j), \quad k = 1, \dots, N \quad (8-7)$$

$$A_k(n) = \sum_{i=1}^m (x_k(n) - y_k^{(i)}(n))^2 - \sum_{i=m+1}^M (x_k(n) - y_k^{(i)}(n))^2, \quad k = 1, \dots, N \quad (8-8)$$

Επιπλέον, προκύπτει ότι ο υπολογισμός των παραγόντων $B_k(n)$ ανάγεται τελικά στην παρακάτω αναδρομική σχέση:

$$B_k(n) = A_k(n) + \frac{1}{\lambda} B_k(n-1), \quad k = 1, \dots, N \quad (8-9)$$

Αυτός ο αλγόριθμος αναδρομικής υλοποίησης της ενημέρωσης του διανύσματος παραμέτρων οδηγεί σε σημαντική μείωση του υπολογιστικού κόστους ενημέρωσης. Επιπρόσθετα, η αναδρομική υλοποίηση λαμβάνει υπόψη της τη γνώση για τις προηγούμενες τροποποιήσεις, συνεπώς κατά τη λειτουργία του συστήματος ανάκλησης σε πραγματικές συνθήκες οι παράμετροι απόστασης δεν ενημερώνονται από την αρχή σε κάθε απόπειρα αναζήτησης.

8.4 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Η προτεινόμενη τεχνική αναζήτησης, ανάκλησης και ανάδρασης από το χρήστη δοκιμάστηκε και αξιολογήθηκε σε μία βάση δεδομένων που περιέχει οπτικοακουστικό υλικό κυρίως από τηλεοπτικά δελτία ειδήσεων. Το σύνολο των διαθέσιμων ακολουθιών και ακίνητων εικόνων χαρακτηρίστηκε υπολογίζοντας τα διανύσματα περιγραφής για κάθε καρέ (ή εικόνα) και επιλέγοντας χαρακτηριστικά καρέ και πλάνα, όσον αφορά στις ακολουθίες. Η πληροφορία χαρακτηρισμού αποθηκεύθηκε στη βάση και χρησιμοποιήθηκε για συγκρίσεις στα πειράματα αναζήτησης. Η διαδικασία αναζήτησης γίνεται σε πραγματικό χρόνο, αφού κατά τη διαδικασία αυτή αναλύεται μόνο η είσοδος του χρήστη, ενώ οι συγκρίσεις με το υλικό της βάσης πραγματοποιούνται στο χώρο των διανυσμάτων περιγραφής ενός πολύ μικρού ποσοστού των διαθέσιμων πλάνων, καρέ ή ακίνητων εικόνων. Αντίθετα η διαδικασία χαρακτηρισμού απαιτεί πολύ υπολογιστικό χρόνο και γι' αυτό πραγματοποιείται πριν από οποιοδήποτε πείραμα ανάκλησης (off-line).



Σχήμα 8-1. Αναζήτηση μέσω παραδείγματος από ακολουθίες δελτίων ειδήσεων. (α) Εικόνα εισόδου. (β) Αρχικά αποτελέσματα αναζήτησης για $M = 10$ εικόνες. Από αυτές, επιλέγεται ένα υποσύνολο από $m = 4$ εικόνες που χαρακτηρίζονται από το χρήστη ως «κατάλληλες» (διακρίνονται σε μαύρο πλαίσιο).

Στο Σχήμα 8-1 παρουσιάζεται μία περίπτωση αναζήτησης μέσω παραδείγματος από ακολουθίες τηλεοπτικών δελτίων ειδήσεων. Ο χρήστης δίνει στο σύστημα μία εικόνα εισόδου που εμφανίζεται στο 8-1α και περιέχει ένα δημοσιογράφο (ανταποκριτή) σε εξωτερική λήψη και σε σχετικά κοντινό πλάνο. Η εικόνα αυτή στην πραγματικότητα προέρχεται από ένα ρεπορτάζ ενός δελτίου ειδήσεων· το ίδιο το πλάνο του ρεπορτάζ όμως δεν περιέχεται στη βάση. Η εικόνα εισόδου αναλύεται με τον ίδιο ακριβώς τρόπο όπως και το υλικό της βάσης: πραγματοποιείται κατάτμηση με βάση το χρώμα και από τις ιδιότητες των τμημάτων παράγεται με ασαφή ταξινόμηση το διάνυσμα περιγραφής· πληροφορία κίνησης δεν υπάρχει αφού πρόκειται για στατική εικόνα. Το διάνυσμα περιγραφής της εισόδου συγκρίνεται στη συνέχεια με τα διανύσματα που είναι διαθέσιμα στη βάση με χρήση του μέτρου ομοιότητας (8-1), όπου όλα τα βάρη αρχικοποιούνται λαμβάνοντας ίσες τιμές.



Σχήμα 8-2. Αναζήτηση μέσω παραδείγματος, με ανάδραση. Αποτελέσματα αναζήτησης για την εικόνα εισόδου του Σχήματος 8-1α μετά από ένα στάδιο απλής ανάδρασης. Η ενημέρωση παραμέτρων γίνεται με βάση τις τέσσερις εικόνες που επιλέγονται στο Σχήμα 8-1β.

Τα αρχικά αποτελέσματα αναζήτησης εικονίζονται στο Σχήμα 8-1β. Τα $M = 10$ καρέ που ανακαλούνται αντιστοιχούν στα δέκα διανύσματα περιγραφής με τη μικρότερη απόσταση από το διάνυσμα της εικόνας εισόδου. Η τιμή του M καθορίζεται από το χρήστη· ο αυτόματος υπολογισμός της δεν θα ήταν ιδιαίτερα χρήσιμος, καθώς σε πολλές περιπτώσεις μπορεί να ανακαλούνται χιλιάδες εικόνες. Παρατηρείται ότι εμφανίζονται αρκετές ομοιότητες ανάμεσα στην εικόνα εισόδου και στις ανακαλούμενες εικόνες, ως προς η σύνθεσή τους από αντικείμενα, το χρώμα, τη θέση και το μέγεθος των αντικειμένων.

Όμως ο πραγματικός στόχος της έρευνας ήταν ο εντοπισμός ενός ανθρώπου σε σχετικά κοντινό πλάνο, χωρίς να υπάρχει ενδιαφέρον για το φόντο της εικόνας. Από τις ανακαλούμενες εικόνες μόνο τέσσερις ($m = 4$) θεωρείται ότι ικανοποιούν την προδιαγραφή αυτή· οι εικόνες αυτές χαρακτηρίζονται από το χρήστη ως «κατάλληλες» και διακρίνονται με ένα μαύρο πλαίσιο στο Σχήμα 8-1β. Οι υπόλοιπες έξι εξακολουθούν να παρουσιάζουν ομοιότητες με την εικόνα εισόδου, αλλά κυρίως ως προς τις ιδιότητες του φόντου· γι' αυτό χαρακτηρίζονται «ακατάλληλες».



(α)

Σχήμα 8-3. Αναζήτηση μέσω σχεδιαγράμματος από στατικές εικόνες τοπίων. (α) Σχεδιάγραμμα χρήστη. (β) Αρχικά αποτελέσματα αναζήτησης για $M = 8$ εικόνες. Από το χρήστη επιλέγονται $m = 2$ εικόνες που χαρακτηρίζονται ως «κατάλληλες» (διακρίνονται σε μαύρο πλαίσιο).

Η πληροφορία καταλληλότητας των εικόνων ανατροφοδοτείται στο σύστημα αναζήτησης και ενεργοποιείται ο μηχανισμός ενημέρωσης παραμέτρων. Τα βάρη του μέτρου ομοιότητας της (8-1) επανεκτιμούνται έτσι ώστε το ζητούμενο «άνθρωπος σε κοντινό πλάνο» να εκφραστεί σε χαμηλού επιπέδου χαρακτηριστικά και να εντοπισθούν τα στοιχεία εκείνα του διανύσματος περιγραφής βάσει των οποίων η εικόνα εισόδου ταιριάζει με τις τέσσερις «κατάλληλες». Τα νέα αποτελέσματα αναζήτησης μετά από ένα στάδιο απλής αναζήτησης παρουσιάζονται στο Σχήμα 8-2. Παρατηρείται ότι με την ελαχιστοποίηση της απόστασης μεταξύ της εικόνας εισόδου και των τεσσάρων εικόνων που επιλέγονται από το χρήστη επιτυγχάνεται η ανάκληση εικόνων με περισσότερη έμφαση στο προσκήνιο των εικόνων παρά στο φόντο, και μά-

λιστα χωρίς ο χρήστης να γνωρίζει τον τρόπο με τον οποίο τα διανύσματα περιγραφής αναπαριστούν την οπτική πληροφορία.

Στη συνέχεια δίνεται ένα παράδειγμα *αναζήτησης μέσω σχεδιαγράμματος* από μία συλλογή στατικών εικόνων με τοπία. Το Σχήμα 8-3α απεικονίζει την είσοδο του χρήστη η οποία αυτή τη φορά δεν είναι μία διαθέσιμη εικόνα, αλλά μία εικόνα που σχεδιάζει ο ίδιος ο χρήστης συνδυάζοντας απλά γεωμετρικά σχήματα. Συγκεκριμένα το σχέδιο περιλαμβάνει ένα μεγάλο ορθογώνιο αντικείμενο κυανού χρώματος στο πάνω μέρος του, που αναπαριστά τον ουρανό, καθώς και ένα επίσης ορθογώνιο αντικείμενο, γαλαζοπράσινου χρώματος, που καλύπτει το κάτω μέρος του και αναπαριστά τη θάλασσα. Οι ιδιότητες των δύο αντικειμένων (χρώμα, θέση, μέγεθος κλπ.) χρησιμοποιούνται απευθείας για την κατασκευή ενός διανύσματος περιγραφής με τη μέθοδο του ασαφούς ιστογράμματος· η ανάλυση λοιπόν της εισόδου είναι όμοια με εκείνη της αναζήτησης μέσω παραδείγματος, με τη διαφορά ότι δεν μεσολαβεί κατάτμηση.

Το διάνυσμα περιγραφής συγκρίνεται στη συνέχεια με όλα τα διαθέσιμα διανύσματα σε μία βάση με στατικές εικόνες τοπίων, με σκοπό να ανακληθούν εικόνες που περιέχουν θάλασσα και ουρανό. Ο χρήστης αυτή τη φορά επιλέγει να ανακληθούν οκτώ εικόνες από τη βάση ($M=8$). Και πάλι χρησιμοποιείται το μέτρο ομοιότητας (8-1) όπου όλα τα βάρη αρχικοποιούνται σε ίσες μεταξύ τους τιμές. Τα αποτελέσματα της αναζήτησης για $M=8$ εμφανίζονται στο Σχήμα 8-3β. Παρατηρείται ότι οι περισσότερες από τις οκτώ ανακαλούμενες εικόνες περιέχουν τμήματα ουρανού ή και θάλασσας και έχουν παρόμοια χρωματική σύνθεση με το σχεδιάγραμμα εισόδου. Όμως ο χρήστης κρίνει ότι δύο μόνο εικόνες ικανοποιούν τις ανάγκες του και τις χαρακτηρίζει «κατάλληλες». Οι δύο αυτές εικόνες σημειώνονται με μαύρο πλαίσιο στο Σχήμα 8-3β.



Σχήμα 8-4. Αναζήτηση μέσω σχεδιαγράμματος, με ανάδραση. Αποτελέσματα αναζήτησης για την εικόνα εισόδου του Σχήματος 8-3α μετά από ένα στάδιο απλής ανάδρασης. Η ενημέρωση παραμέτρων γίνεται με βάση τις δύο εικόνες που επιλέγονται στο Σχήμα 8-3β.

Με χρήση ενός σταδίου απλής ανάδρασης, η πληροφορία της αξιολόγησης του χρήστη ανατροφοδοτείται στο σύστημα και τα βάρη ανανεώνονται δυναμικά με σκοπό να δοθεί έμφαση στα χαρακτηριστικά των επιλεγμένων εικόνων. Τα νέα αποτελέσματα αναζήτησης εμφανίζονται στο Σχήμα 8-4, όπου είναι εμφανές ότι οι ανακαλούμενες εικόνες μετά την ανάδραση παρουσιάζουν μεγαλύτερη ομοιότητα με το αρχικό σχεδιάγραμμα. Είναι τέλος αξιοσημείωτο ότι στα περισσότερα πειράματα ανάκλησης, είτε βάσει παραδείγματος είτε βάσει σχεδιαγράμματος, η απλή ανάδραση είναι επαρκής. Ακόμη και στις περιπτώσεις όπου χρησιμοποιείται πολλαπλή ανάδραση, η μεγαλύτερη βελτίωση επιτυγχάνεται κατά την πρώτη επανάληψη. Αυτό το συμπέρασμα είναι ενθαρρυντικό καθώς ο τελικός χρήστης αναμένει να εντοπίσει κατάλληλα αποτελέσματα όσο το δυνατόν συντομότερα.

ΠΕΡΙΓΡΑΦΗ ΣΧΗΜΑΤΟΣ ΑΝΤΙΚΕΙΜΕΝΩΝ

Στο τρίτο Μέρος της διατριβής εξετάζεται η χρήση του σχήματος των αντικειμένων που παράγονται από τη διαδικασία κατάρτισης για σκοπούς ταξινόμησης και αναζήτησης. Η αναπαράσταση, ανάλυση και επεξεργασία του περιγράμματος αντικειμένων αποτελεί μία ερευνητική περιοχή με αυξημένο ενδιαφέρον, κυρίως λόγω της εμφάνισης τεχνικών αναζήτησης μέσω σχεδιαγράμματος, όπως παρουσιάστηκε στο Κεφάλαιο 8, αλλά και της ενσωμάτωσης της πληροφορίας σχήματος στις σύγχρονες τεχνικές κωδικοποίησης. Στο πλαίσιο της διατριβής προτείνεται μία μέθοδος *κανονικοποίησης* η οποία παρέχει περιγραφή *αναλλοίωτη σε affine μετασχηματισμούς* καθώς και μετασχηματισμούς *παραμέτρου*, χωρίς ταυτόχρονα να παρουσιάζει καμία απώλεια πληροφορίας.

Στο Κεφάλαιο 9 πραγματοποιείται μία βιβλιογραφική ανασκόπηση στις υφιστάμενες μεθόδους αναπαράστασης, ταξινόμησης, ταιριάσματος και αναγνώρισης περιγραμμάτων. Στη συνέχεια δίνεται η γενική περιγραφή της μεθόδου κανονικοποίησης, των ιδιοτήτων, προϋποθέσεων και περιορισμών της, και παρουσιάζεται η μοντελοποίηση καμπυλών με *B-Splines* που χρησιμοποιείται για την αναπαράσταση του σχήματος των περιγραμμάτων.

Στο Κεφάλαιο 10 δίνεται η αναλυτική περιγραφή και θεωρητική ανάλυση της προτεινόμενης μεθόδου κανονικοποίησης περιγραμμάτων. Η διαδικασία περιλαμβάνει σταδιακή κανονικοποίηση ως προς τους μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας, κλίσης (*skew*), σημείου αναφοράς, περιστροφής, κατοπτρισμού και φοράς διαγραφής. Η κανονικοποίηση είναι ένας γραμμικός μετασχηματισμός που πραγματοποιείται με χρήση πληροφορίας όπως οι ροπές μέχρι δευτέρου βαθμού και ο μετασχηματισμός Fourier κάθε καμπύλης.

Τέλος στο Κεφάλαιο 11 εξετάζεται το πρόβλημα της *ανάκλησης* αντικειμένων με βάση το σχήμα τους, η οποία βασίζεται σε τρία απλά μέτρα ομοιότητας περιγραμμάτων, ενώ εξετάζεται και το συναφές πρόβλημα της *ταξινόμησης* σχημάτων σε προκαθορισμένες κατηγορίες, το οποίο αντιμετωπίζεται με χρήση νευρωνικού δικτύου. Στη συνέχεια δίνονται παραδείγματα, πειραματικά αποτελέσματα και αλγοριθμικές οδηγίες τόσο για την κανονικοποίηση περιγραμμάτων όσο και για την ταξινόμηση, τα χρησιμοποιούμενα μέτρα ομοιότητας και την αναζήτηση με βάση το σχήμα.

Αναπαράσταση Περιγραμμάτων

Η αναπαράσταση, ανάλυση και επεξεργασία του περιγράμματος των αντικειμένων αποτελεί μία ερευνητική περιοχή με αυξημένο ενδιαφέρον, κυρίως λόγω της εμφάνισης τεχνικών αναζήτησης μέσω σχεδιαγράμματος, όπως παρουσιάστηκε στο Κεφάλαιο 8, αλλά και της ενσωμάτωσης της πληροφορίας σχήματος στις σύγχρονες τεχνικές κωδικοποίησης. Στο Κεφάλαιο αυτό παρουσιάζεται μία βιβλιογραφική ανασκόπηση στις διαθέσιμες μεθόδους αναπαράστασης, ταξινόμησης, ταιριάσματος και αναγνώρισης περιγραμμάτων, από την οποία προκύπτει η ανάγκη για περιγραφή περιγραμμάτων που να είναι αναλλοίωτη σε ένα σύνολο μετασχηματισμών όπως μετατόπιση, περιστροφή, αλλαγή κλίμακας κλπ. Στο πλαίσιο της διατριβής προτείνεται μία μέθοδος κανονικοποίησης η οποία παρέχει περιγραφή αναλλοίωτη σε affine μετασχηματισμούς καθώς και μετασχηματισμούς παραμέτρου (π.χ. κυκλική ολίσθηση, φορά διαγραφής) χωρίς ταυτόχρονα να παρουσιάζει καμία απώλεια πληροφορίας, αφού το σχήμα του περιγράμματος διατηρείται στο ακέραιο. Έτσι, η προκύπτουσα αναπαράσταση μπορεί να χρησιμοποιηθεί σε συνδυασμό με οποιαδήποτε μέθοδο ταξινόμησης, αναγνώρισης ή ταιριάσματος καμπυλών. Στη συνέχεια του Κεφαλαίου δίνεται η γενική περιγραφή της μεθόδου κανονικοποίησης, των ιδιοτήτων, προϋποθέσεων και περιορισμών της, και παρουσιάζεται η μοντελοποίηση καμπυλών με B-Splines που χρησιμοποιείται για την απλοποίηση του σχήματος των περιγραμμάτων, τη μείωση του θορύβου κατάτμησης και την ομοιόμορφη δειγματοληψία ως προς το μήκος τόξου. Η αναλυτική περιγραφή της μεθόδου κανονικοποίησης δίνεται στο επόμενο Κεφάλαιο.

9.1 ΕΙΣΑΓΩΓΗ

Πληροφορία Σχήματος

Η οπτική πληροφορία σε συστήματα αναζήτησης με βάση το περιεχόμενο μοντελοποιείται συνήθως με τη χρήση χαρακτηριστικών χαμηλού επιπέδου, όπως η σύνθεση χρώματος και υφής [14], το πεδίο κίνησης ή βάθους [39], καθώς και το σχήμα των αντικειμένων [145]. Υψηλότερου επιπέδου αναπαράσταση είναι εφικτή μέσω του συνδυασμού χαρακτηριστικών χαμηλού επιπέδου, κυρίως στο πλαίσιο ειδικευμένων εφαρμογών. Όπως προτείνεται στην εργασία [100], αν η βασική πληροφορία για την περιγραφή, αναγνώριση ή ταξινόμηση ενός αντικειμένου μπορεί να βρεθεί στο *σχήμα του περιγράμματός του* (contour shape), είναι φυσικό να διατηρήσει κανείς μόνο το σχήμα του για περαιτέρω επεξεργασία. Τέτοιες περιπτώσεις παρουσιάζονται για παράδειγμα στην αναγνώριση και ταξινόμηση αεροσκαφών και δορυφόρων με βάση το περίγραμμά τους, στην αναγνώριση χαρακτήρων και στην επεξεργασία εγγράφων [70]. Όμως η μελέτη και η χρήση του σχήματος για την περιγραφή, αναγνώριση ή αναζήτηση αντικειμένων σε εφαρμογές γενικού περιεχομένου, είτε από μόνο του είτε σε συνδυασμό με άλλα χαρακτηριστικά, είναι μία ενεργή περιοχή της τρέχουσας έρευνας [78,125,148].

Υπάρχουν δύο κύριοι λόγοι για το αυξημένο ενδιαφέρον στην ανάλυση και επεξεργασία του σχήματος αντικειμένων. Πρώτον, διότι παρέχει ένα ισχυρό εργαλείο για αναζήτηση και ανάκληση εικόνων με βάση το περιεχόμενο, χρησιμοποιώντας ένα μηχανισμό *αναζήτησης μέσω σχεδιαγράμματος / σκίτσου* (query-by-sketch) [19]. Σε μια τέτοια εφαρμογή ο τελικός χρήστης *σχεδιάζει* το σκίτσο ενός αντικειμένου και το σύστημα αναζήτησης ταιριάζει το σκίτσο αυτό με το σχήμα των αντικειμένων σε μια βάση δεδομένων με εικόνες ή βίντεο. Το ταίριασμα πραγματοποιείται με μέτρηση της ομοιότητας μεταξύ των προτύπων σχημάτων της βάσης και του σκίτσου του επιθυμητού αντικειμένου.

Δεύτερον, λειτουργίες που σχετίζονται με την περιγραφή περιεχομένου ενσωματώνονται στα πρόσφατα πρότυπα κωδικοποίησης οπτικοακουστικού υλικού [65]. Για το σκοπό αυτό έχουν προταθεί τεχνικές κωδικοποίησης *δεύτερης γενιάς* [128], στις οποίες η κωδικοποίηση βασίζεται στην κατάτμηση σε αντικείμενα και επιτρέπει τη *διαχείριση* (manipulation) περιεχομένου των αντικειμένων [110]. Έτσι η πληροφορία σχήματος ενσωματώνεται στα *επίπεδα αντικειμένων* (video object planes, VOPs) με τη μορφή δυαδικής εικόνας (μάσκας) και μπορεί να χρησιμοποιηθεί για την πρόβλεψη ή χρονική παρεμβολή στα τμήματων μιας ακολουθίας [84].

Υπάρχουσες Τεχνικές

Έχουν προταθεί πολυάριθμες μέθοδοι στη βιβλιογραφία για την ανάλυση, τη μοντελοποίηση και την αναπαράσταση του σχήματος. Οι μέθοδοι αυτές περιλαμβάνουν μεταξύ άλλων την *κωδικοποίηση αλυσίδας* (chain coding) [42], την *προσέγγιση πολυγώνων* (polygonal approximation) [98], το *μετασχηματισμό σκελετού* (medial axis / skeleton transform) [20], τους *περιγραφείς Fourier* (Fourier descriptors) [100], τις *ροπές καμπυλών* (curve moments) [55], τα *B-Splines* [28], τους *χώρους κλίμακας κυρτότητας* (curvature scale spaces) [90], τα *σημεία ενδιαφέροντος* (interest points) [141], τον *ημιτονοειδή μετασχηματισμό* (sinusoidal transform) [101], τους *περιγραφείς Legendre* (Legendre descriptors), καθώς και τις *ροπές Zernike* (Zernike moments) [73].

Οι περισσότερες τεχνικές εκμεταλλεύονται γεωμετρικά χαρακτηριστικά των καμπυλών, είτε *ολικά* (global) είτε *τοπικά* (local) για να επιτύχουν αναγνώριση, ταξινόμηση ή ταίριασμα καμπυλών. Τα ολικά χαρακτηριστικά προκύπτουν από την ανάλυση της καμπύλης ως σύνολο και παραδείγματά τους αποτελούν οι *ροπές* (moments), το *μήκος*, οι *κύριοι άξονες* (principal axes), η *επιμήκυνση* (elongation) και το *συμπαγές* (compactness). Αντίθετα τα τοπικά χαρακτηριστικά προκύπτουν από τμηματική ανάλυση της καμπύλης και παραδείγματά τους είναι τα *σημεία ενδιαφέροντος* (interest points), τα *μέτρα κυρτότητας* (curvature measures) και τα *έμμεσα πολυώνυμα* (implicit polynomials).

Ανεξάρτητα από το είδος της εφαρμογής, όλες οι μέθοδοι ανάλυσης σχήματος αντιμετωπίζουν ένα κοινό πρόβλημα: το σχήμα των αντικειμένων μεταβάλλεται δραστικά ανάλογα με την άποψη του αντικειμένου στον 3-Δ χώρο, λόγω του *μετασχηματισμού προοπτικής προβολής* (perspective transformation). Στις περισσότερες εργασίες η μεταβολή αυτή προσεγγίζεται με έναν *affine μετασχηματισμό* (affine transformation). Ο λόγος είναι ότι, σε αντίθεση με το μετασχηματισμό προοπτικής προβολής, ο affine μετασχηματισμός είναι γραμμικός, γεγονός που απλοποιεί σημαντικά τη διαδικασία ανάλυσης. Επιπλέον πρόκειται για πολύ καλή προσέγγιση, ιδιαίτερα όταν τα αντικείμενα είναι μακριά από την κάμερα, αφού η ελαφριά αλλοίωση του σχήματος που οφείλεται στην προοπτική μπορεί να θεωρηθεί τμήμα παραμόρφωσης, η οποία αντιμετωπίζεται με άλλες τεχνικές. Για να αποφευχθεί η αποθήκευση ή το ταίριασμα με

ένα μεγάλο αριθμό προτύπων σχημάτων που αντιστοιχούν σε διαφορετικούς affine μετασχηματισμούς (π.χ. διαφορετική περιστροφή, μετατόπιση ή αλλαγή κλίμακας), πρέπει να ορισθούν ποσότητες (ή μετασχηματισμοί) που είναι *αναλλοίωτες* (invariant) στους affine μετασχηματισμούς. Οι ποσότητες αυτές ονομάζονται *affine invariants*.

Μία ιδιότητα κοινή στις περισσότερες affine invariant τεχνικές στη βιβλιογραφία είναι ότι η απομάκρυνση της επίδρασης των affine μετασχηματισμών είναι «ενσωματωμένη» στη διαδικασία της αναγνώρισης, του ταιριάσματος καμπυλών ή της εκτίμησης μέτρων ομοιότητας. Για παράδειγμα, ένα μέτρο ομοιότητας αναλλοίωτο στην περιστροφή, τη μετατόπιση και την αλλαγή κλίμακας που βασίζεται σε *συναρτήσεις στροφής* (turning functions) για τη σύγκριση πολυγώνων έχει προταθεί στην εργασία [4], ενώ ένα ανάλογο μέτρο βασισμένο σε *τροποποιημένους περιγραφείς Fourier* (modified Fourier descriptors, MFD) έχει παρουσιασθεί στην εργασία [108]. Για τον ίδιο σκοπό έχουν χρησιμοποιηθεί αναλλοίωτες ποσότητες βασισμένες στις ροπές (moment invariants) [17]. Για την affine invariant αναγνώριση αντικειμένων έχουν χρησιμοποιηθεί *κανονικοποιημένοι περιγραφείς Fourier* (normalized Fourier descriptors, NFD) σε συνδυασμό με νευρωνικά δίκτυα [140], καθώς και γενετικοί αλγόριθμοι [130].

Τεχνικές βασισμένες σε τοπικά χαρακτηριστικά καμπυλών περιλαμβάνουν την αυτόματη ανίχνευση σημείων ενδιαφέροντος για ανάκληση εικόνων [113], affine invariant ποσότητες βασισμένες στο *κυρτό περίβλημα* (convex hull) για το ταίριασμα (registration) εικόνων [148], καθώς και τοπικά χαρακτηριστικά αναλλοίωτα στην παραμόρφωση για αναγνώριση καμπυλών με χρήση έμμεσων πολυνόμων [103]. Μία εναλλακτική προσέγγιση είναι το ταίριασμα δύο γνωστών καμπυλών με βέλτιστη εκτίμηση των παραμέτρων ενός affine μετασχηματισμού που μεγιστοποιεί ένα μέτρο ομοιότητας μεταξύ των δύο καμπυλών. Η βελτιστοποίηση βασίζεται, για παράδειγμα, στις ροπές των καμπυλών [56] ή στους περιγραφείς Fourier [100].

Τεχνικές Κανονικοποίησης

Το κύριο μειονέκτημα της πρώτης προσέγγισης – δηλαδή της ενσωμάτωσης της affine invariant περιγραφής στη διαδικασία αναγνώρισης ή ταιριάσματος – είναι ότι σχεδόν πάντα χάνεται μεγάλο μέρος πληροφορίας για το σχήμα της αρχικής καμπύλης. Από την άλλη μεριά, η δεύτερη προσέγγιση – η εκτίμηση των παραμέτρων του affine μετασχηματισμού μεταξύ δύο καμπυλών – απαιτεί εκ των προτέρων γνώση των δύο καμπυλών. Επομένως μπορεί να χρησιμοποιηθεί μόνο για ταίριασμα σε ένα συγκεκριμένο ζεύγος καμπυλών και όχι, για παράδειγμα, για αναγνώριση μέσω ενός νευρωνικού δικτύου ή οποιασδήποτε τεχνικής ταξινόμησης. Επιπλέον η υλοποίηση της δεύτερης προσέγγισης απαιτεί συνήθως υψηλό υπολογιστικό κόστος.

Για το λόγο αυτό έχει προταθεί η μέθοδος της *κανονικοποίησης* (normalization) ως εναλλακτική λύση για περιγραφή καμπυλών που να είναι αναλλοίωτη στους affine μετασχηματισμούς. Μία καμπύλη ή μία εικόνα μπορεί να κανονικοποιηθεί σε μία «πρότυπη» θέση, η οποία ορίζεται με τέτοιο τρόπο ώστε όλοι οι affine μετασχηματισμοί του ίδιου αντικειμένου να κανονικοποιούνται πάντα στην ίδια θέση. Με εξαίρεση τις παραμέτρους του affine μετασχηματισμού, στον οποίο η κανονικοποίηση είναι αναλλοίωτη, δεν χάνεται άλλη πληροφορία: η διαδικασία κανονικοποίησης αποτελείται στην ουσία από έναν affine (γραμμικό) μετασχηματισμό και το σχήμα της αρχικής καμπύλης παραμένει αμετάβλητο. Μία γενικευμένη μέθοδος κανονικοποίησης για τον προσδιορισμό αναλλοίωτων ποσοτήτων δίνεται στην εργασία [104],

η κανονικοποίηση εικόνων αντιμετωπίζεται στην εργασία [115], ενώ η κανονικοποίηση σχημάτων που έχουν υποστεί affine μετασχηματισμό μελετάται στην εργασία [126].

Για το ταίριασμα καμπυλών που έχουν υποστεί αυθαίρετη παραμόρφωση έχει προταθεί ένας αριθμός τεχνικών οι οποίες βασίζονται στα *παραμορφούμενα πρότυπα* (deformable templates) [19]. Τα παραμορφούμενα πρότυπα παράγονται επιβάλλοντας παραμετρικούς μετασχηματισμούς σε μία πρωτότυπη καμπύλη, ενώ η διακύμανση των προτύπων επιτυγχάνεται με στοχαστικό τρόπο [69]. Τα *μοντέλα ενεργών περιγραμμάτων* (active contour models ή snakes) χρησιμοποιούνται επίσης για τον ίδιο σκοπό [77]. Παρά το γεγονός ότι τα παραμορφούμενα πρότυπα αντιμετωπίζουν επιτυχώς το θόρυβο αλλά και τις τοπικές παραμορφώσεις των περιγραμμάτων (που οφείλονται σε τοπικές ανομοιοότητες ή και επικάλυψη αντικειμένων), οι τοπικές παραμορφώσεις εμφανίζονται συνήθως σε συνδυασμό με ολικές μεταβολές σχήματος που εκφράζονται με affine μετασχηματισμούς. Σε αυτές τις περιπτώσεις η επίδοσή τους είναι περιορισμένη. Όμως πολύ καλά αποτελέσματα προκύπτουν από τη χρήση των παραμορφούμενων προτύπων σε δεύτερο στάδιο, μετά την κανονικοποίηση, όπως αναφέρεται και στην εργασία [61].

9.2 ΠΡΟΤΕΙΝΟΜΕΝΗ ΑΝΑΠΑΡΑΣΤΑΣΗ

Μοντελοποίηση – Κανονικοποίηση

Στο πλαίσιο της παρούσας διατριβής προτείνεται μία πρωτότυπη μέθοδος κανονικοποίησης διδιάστατων καμπυλών ως προς τους affine μετασχηματισμούς. Η μέθοδος καθιστά την αναπαράσταση των καμπυλών αναλλοίωτη ως προς τους affine μετασχηματισμούς χωρίς καμία απώλεια πληροφορίας, ενώ στην περίπτωση κλειστών περιγραμμάτων η αναπαράσταση είναι επίσης αναλλοίωτη ως προς το αρχικό σημείο αναφοράς και τη φορά διαγραφής των καμπυλών. Συγκεκριμένα, η 2-Δ κλειστή καμπύλη που αναπαριστά το σχήμα του περιγράμματος ενός αντικειμένου μοντελοποιείται αρχικά μέσω κυβικών B-Splines. Έτσι απλοποιείται το σχήμα της καμπύλης και ο ελαττώνεται ο τυχόν θόρυβος που μπορεί να οφείλεται στην αυτόματη εξαγωγή του περιγράμματος (π.χ. μέσω κατάτμησης). Επίσης επιτυγχάνεται δειγματοληψία ομοιόμορφη ως προς το *μήκος τόξου* (arc length) της καμπύλης. Στη συνέχεια η καμπύλη κανονικοποιείται σε μια σειρά από βήματα στα οποία αφαιρείται σταδιακά η επίδραση της *μετατόπισης* (translation), της *αλλαγής κλίμακας* (scaling), της *κλίσης* (skew), του *αρχικού σημείου αναφοράς* (starting point), της *περιστροφής* (rotation), του *κατοπτρισμού* (reflection) και της *φοράς διαγραφής* (orientation).

Ο αλγόριθμος κανονικοποίησης βασίζεται σε ένα συνδυασμό χαρακτηριστικών της καμπύλης, συμπεριλαμβανομένων των ροπών (έως δεύτερου βαθμού) και ενός συνόλου περιγραφών Fourier. Η εκτίμηση όλων των χαρακτηριστικών πραγματοποιείται ολικά, ενώ δεν χρησιμοποιείται τοπική πληροφορία. Η συμμετοχή δηλαδή όλων των δειγμάτων της καμπύλης στον υπολογισμό των χαρακτηριστικών είναι ισοδύναμη. Το υπολογιστικό κόστος της προτεινόμενης μεθόδου κανονικοποίησης είναι αμελητέο, σε σχέση τουλάχιστον με το κόστος άλλων σταδίων επεξεργασίας, όπως π.χ. της κατάτμησης. Έτσι είναι δυνατή η ενσωμάτωση της μεθόδου σε συστήματα πραγματικού χρόνου για ανάκληση εικόνων ή ακόμη και για κωδικοποίηση βίντεο. Τα βασικά σημεία της μεθόδου έχουν παρουσιασθεί συνοπτικά στην εργασία [16].

Η ιδέα της κανονικοποίησης για την εξαγωγή αναλλοίωτων ποσοτήτων έχει παρουσιασθεί στην εργασία [104] και αξιοποιηθεί για την κανονικοποίηση εικόνων στην [115]. Όμως η προτεινόμενη τεχνική κανονικοποίησης για τη δημιουργία πλήρως αναλλοίωτης αναπαράστασης 2-Δ κλειστών καμπυλών σε affine μετασχηματισμούς καθώς και μετασχηματισμούς παραμέτρου χωρίς απώλεια πληροφορίας είναι εξολοκλήρου πρωτότυπη. Η αρχική ιδέα της κανονικοποίησης δημοσιεύθηκε για πρώτη φορά στην εργασία [11], και χρησιμοποιήθηκε ακόμμη στην [10]. Οι ιδιότητές της και οι λεπτομέρειες υλοποίησής της παρουσιάστηκαν στην [16] όπου χρησιμοποιήθηκε και για ανάκληση με βάση το σχήμα· αναλυτικότερη περιγραφή καθώς και αποδείξεις των σχετικών προτάσεων αναμένεται να δημοσιευθούν στην εργασία [12]. Η ταξινόμηση 2-Δ καμπυλών με χρήση B-Splines και νευρωνικού δικτύου παρουσιάστηκε στις εργασίες [10,11,145].

Ιδιότητες

Αποδεικνύεται ότι κάθε κανονικοποιημένη καμπύλη αντιστοιχεί *μοναδικά* (uniquely) σε ένα σύνολο καμπυλών που σχετίζονται μεταξύ τους μέσω αυθαίρετων affine μετασχηματισμών. Επιπλέον, η κανονικοποιημένη αναπαράσταση, μαζί με τις παραμέτρους του affine μετασχηματισμού που συνδέει την αρχική καμπύλη με την κανονικοποιημένη (οι οποίες εκτιμώνται κατά τη διαδικασία της κανονικοποίησης) περιγράφουν *πλήρως* (completely) την αρχική καμπύλη. Η αρχική καμπύλη μπορεί δηλαδή να ανακατασκευασθεί με ακρίβεια από την κανονικοποιημένη με χρήση των εκτιμώμενων παραμέτρων. Με βάση τις παραπάνω ιδιότητες *μοναδικότητας* (uniqueness) και *πληρότητας* (completeness) επιτυγχάνεται *αποσύνθεση* (decomposition) των καμπυλών σε *ολική* (global) «θέση» σχετιζόμενη με τους affine μετασχηματισμούς και σε *τοπική* (local) πληροφορία σχήματος.

Επομένως η προτεινόμενη μέθοδος μπορεί να εφαρμοσθεί ως ένα στάδιο προεπεξεργασίας σε οποιαδήποτε τεχνική αναπαράστασης, ταξινόμησης, αναγνώρισης ή ανάκλησης με βάση το σχήμα, αφού ουσιαστικά αποσυσχετίζει το πρόβλημα της affine-invariant περιγραφής από την *εξαγωγή χαρακτηριστικών* (feature extraction) και το *ταίριασμα με πρότυπα* (pattern matching). Ένας αριθμός από *μέτρα ομοιότητας* (similarity measures) χρησιμοποιούνται στα πειράματα για να επιδείξουν την ικανότητα της προτεινόμενης αναπαράστασης να διατηρεί όλη την πληροφορία της καμπύλης εκτός από τους αυθαίρετους affine μετασχηματισμούς, στους οποίους είναι αναλλοίωτη. Επίσης, σε όλες τις περιπτώσεις, τα πειράματα δείχνουν ότι η αναπαράσταση είναι ανθεκτική στο θόρυβο και στις παραμορφώσεις των σχημάτων.

9.3 ΠΡΟΫΠΟΘΕΣΕΙΣ – ΠΕΡΙΟΡΙΣΜΟΙ

Προέλευση Περιγραμμάτων

Η παρακάτω ανάλυση προϋποθέτει ότι το σχήμα του περιγράμματος ενός αντικειμένου είναι διαθέσιμο ως ένα διατεταγμένο σύνολο 2-Δ σημείων τα οποία σχηματίζουν μία 2-Δ επίπεδη και κλειστή καμπύλη. Το σύνολο αυτό μπορεί να προέρχεται από δειγματοληψία δεδομένων μιας εικόνας με χρήση αυτόματης, ημιαυτόματης ή και χειρονακτικής κατάτμησης. Στην πράξη μπορεί να εφαρμοσθεί οποιοσδήποτε αλγόριθμος κατάτμησης, βασιζόμενος για παράδειγμα στην *ομοιογένεια* (homogeneity) χρώματος ή κίνησης, στην *ανίχνευση ακμών* (edge detection) ή σε μορφολογικά εργαλεία [111]. Στα πειράματα χρησιμοποιήθηκε ο αλγόριθμος M-RSST, όπως περιγράφεται στην Ενότητα 1.2, για κατάτμηση με βάση το χρώμα.

Στην περίπτωση που ένα μεγάλο σύνολο ακολουθιών βίντεο είναι διαθέσιμο σε μία βάση δεδομένων, κάθε ακολουθία διαχωρίζεται αρχικά σε πλάνα που αντιστοιχούν σε χρονικά τμήματα συνεχούς λειτουργίας μίας κάμερας, και στη συνέχεια εφαρμόζεται κατάτμηση στα καρέ του κάθε πλάνου. Στην περίπτωση αυτή η ποιότητα της κατάτμησης μπορεί να βελτιωθεί με την εκμετάλλευση της πληροφορίας κίνησης. Συγκεκριμένα, για την κατάτμηση με βάση την κίνηση μπορούν να χρησιμοποιηθούν 2-Δ παραμετρικά μοντέλα κίνησης [96,127], ή η ανίχνευση των *κύριων κινούμενων αντικειμένων* (main mobile objects) [146]. Παρά το γεγονός ότι η κανονικοποίηση σημείου αναφοράς που παρουσιάζεται στην Ενότητα 10-5 εφαρμόζεται μόνο σε κλειστές καμπύλες περιγραμμάτων, τα υπόλοιπα βήματα κανονικοποίησης εξακολουθούν να ισχύουν και για ανοικτές καμπύλες.

Επικαλύψεις

Γίνεται επίσης η υπόθεση ότι οι καμπύλες αντιστοιχούν σε περιγράμματα αντικείμενων που δεν επικαλύπτονται μεταξύ τους, καθώς και ότι είναι εξ' ολοκλήρου γνωστές. Προβλήματα που σχετίζονται με την *επικάλυψη* (occlusion) αντικειμένων μπορούν να αντιμετωπισθούν μόνο με χρήση τοπικής πληροφορίας, ενώ στην προτεινόμενη μέθοδο χρησιμοποιούνται μόνο ολικά χαρακτηριστικά. Δυστυχώς, δεν είναι γνωστή καμία μέθοδος κανονικοποίησης στη βιβλιογραφία η οποία να μπορεί να αντιμετωπίσει επιτυχώς τα ζητήματα επικάλυψης χωρίς ταυτόχρονη απώλεια πληροφορίας, αν και έχουν καταβληθεί σημαντικές προσπάθειες [56]. Όμως, στην περίπτωση που ο βαθμός επικάλυψης είναι μικρός, η επικάλυψη μπορεί να θεωρηθεί ως μικρή τοπική παραμόρφωση και να αντιμετωπισθεί στη διαδικασία ταιριάσματος.

Μετασχηματισμοί

Όσον αφορά στους δυνατούς μετασχηματισμούς στους οποίους η προτεινόμενη αναπαράσταση είναι αναλλοίωτη, γίνεται η υπόθεση ότι κάθε καμπύλη εισόδου έχει υποστεί δύο ειδών μετασχηματισμούς: *μετασχηματισμούς παραμέτρου* (parameter transformations) και *μετασχηματισμούς συντεταγμένων* (coordinate transformations). Οι μετασχηματισμοί παραμέτρου οφείλονται στο γεγονός ότι οι καμπύλες προέρχονται από κατάτμηση εικόνων, στην οποία εμπλέκεται ουσιαστικά μια μορφή δειγματοληψίας που οδηγεί στη διακριτή αναπαράσταση συνεχών καμπυλών. Έτσι προκύπτει όχι μόνο «θόρυβος» κατάτμησης αλλά και δειγματοληψία μη ομοιόμορφη ως προς το *μήκος τόξου* (arc length). Επίσης, αφού τα περιγράμματα των αντικειμένων περιγράφονται από κλειστές καμπύλες, ένα σημείο που κινείται κατά μήκος του περιγράμματος παράγει συντεταγμένες που είναι περιοδικές συναρτήσεις του μήκους τόξου. Έτσι μπορεί να επιλεγεί αυθαίρετα οποιοδήποτε σημείο ως αρχικό σημείο αναφοράς για την περιγραφή μιας περιόδου. Επιπλέον η φορά διαγραφής της καμπύλης μπορεί να είναι είτε αριστερόστροφη είτε δεξιόστροφη.

Από την άλλη μεριά, οι μετασχηματισμοί συντεταγμένων οφείλονται στο γεγονός ότι οι εικόνες προέρχονται από την προβολή 3-Δ αντικειμένων σε ένα 2-Δ επίπεδο, προκαλώντας μη γραμμική προοπτική παραμόρφωση. Με την προϋπόθεση ότι ένα αντικείμενο είναι αρκετά μακριά από την κάμερα (σε σχέση βέβαια με το μέγεθος του αντικείμενου), η προοπτική παραμόρφωση μπορεί να προσεγγισθεί από ένα γραμμικό affine μετασχηματισμό. Το πρόβλημα λοιπόν είναι η κανονικοποίηση μιας καμπύλης και η εξαγωγή μιας αναπαράστασης η οποία, χωρίς απώλεια πληροφορίας σχήματος, να είναι αναλλοίωτη στους μετασχηματισμούς παραμέτρου αλλά και τους μετασχηματισμούς συντεταγμένων.

Βήματα Κανονικοποίησης

Οι μετασχηματισμοί παραμέτρου που οφείλονται στο θόρυβο κατάτμησης και στη μη ομοιόμορφη δειγματοληψία αντιμετωπίζονται στην προτεινόμενη μέθοδο με χρήση μοντέλων B-Splines, όπως περιγράφεται στη συνέχεια. Στην πραγματικότητα μια ποσότητα πληροφορίας χάνεται κατά τη διαδικασία της μοντελοποίησης, αλλά κάτι τέτοιο είναι απαραίτητο για την απομάκρυνση του θορύβου και την απλοποίηση της καμπύλης, ώστε να αφαιρούνται οι λεπτομέρειες και να διατηρείται το συνολικό σχήμα. Τα υπόλοιπα όμως βήματα κανονικοποίησης είναι πλήρως αναστρέψιμα.

Οι μετασχηματισμοί συντεταγμένων (affine) αναλύονται σε μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας, κλίσης, περιστροφής και κατοπτρισμού. Η απαλοιφή των τριών πρώτων (μετατόπιση, αλλαγής κλίμακας και κλίσης) επιτυγχάνεται μέσω μιας διαδικασίας *ορθοκανονικοποίησης* (orthogonalization) που βασίζεται στις ροπές μέχρι δεύτερου βαθμού. Στη συνέχεια οι δύο τελευταίες (περιστροφή και κατοπτρισμός), καθώς και το σημείο αναφοράς και η φορά διαγραφής, απαλείφονται με μία διαδικασία κανονικοποίησης που βασίζεται στο μετασχηματισμό Fourier της καμπύλης.

9.4 B-SPLINES

Μοντελοποίηση

Τα B-Splines έχουν χρησιμοποιηθεί ευρέως για την ανάλυση και τη μοντελοποίηση σχημάτων διότι έχουν έναν αριθμό από σημαντικές ιδιότητες, όπως *ομαλότητα* (smoothness), *συνέχεια* (continuity), και δυνατότητα *τοπικού ελέγχου* (local controllability). Είναι επίσης *φραγμένες* συναρτήσεις (built-in boundedness), και έχουν δυνατότητα αναλλοίωτης περιγραφής ως προς affine μετασχηματισμούς [28]. Στην παρούσα εργασία τα B-Splines χρησιμοποιούνται για την ομαλή και συνεχή αναπαράσταση των καμπυλών, οι οποίες είναι διαθέσιμες ως διατεταγμένα σύνολα 2-Δ σημείων. Παρακάτω δίνεται μία συνοπτική περιγραφή του αναγκαίου υπόβαθρου για τη μοντελοποίηση καμπυλών με B-Splines και την προσαρμογή των B-Splines στο σύνολο σημείων που περιγράφουν την κάθε καμπύλη. Στη συνέχεια δίνεται μία μέθοδος ανακατανομής των σημείων ώστε να επιτυγχάνεται ομοιόμορφη δειγματοληψία ως προς το μήκος τόξου.

Τα *κυβικά B-Splines* (cubic B-Splines) είναι σύνθετες καμπύλες που αποτελούνται από μεγάλο γενικά αριθμό *τμημάτων* (segments), τα οποία με τη σειρά τους περιγράφονται με μία υπέρθεση πολυωνύμων τρίτου βαθμού, ενώ στα *σημεία σύνδεσης* (connection points) των τμημάτων οι καμπύλες είναι C^2 *συνεχείς* (C^2 continuous). Δηλαδή όλες οι παράγωγοι ενός κυβικού B-Spline έως και δεύτερου βαθμού είναι συνεχείς. Γενικότερα, ένα B-Spline βαθμού k αποτελείται από τμήματα που εκφράζονται ως υπέρθεση πολυωνύμων βαθμού k και είναι C^{k-1} συνεχές στα σημεία σύνδεσης. Στη συνέχεια η ανάλυση περιορίζεται σε κυβικά B-Splines, αλλά μπορεί εύκολα να γενικευθεί για B-Splines αυθαίρετου βαθμού. Γίνεται επίσης η υπόθεση ότι είναι κλειστές καμπύλες, εφόσον περιγράφουν το κλειστό περίγραμμα αντικειμένων.

Έστω ότι ένα κλειστό κυβικό B-Spline \mathbf{r} αποτελείται από N συνδεδεμένα τμήματα \mathbf{r}_i , $i = 0, \dots, N-1$. Κάθε τέτοιο τμήμα παριστάνει ένα 2-Δ σημείο συναρτήσεως μιας παραμέτρου $u \in [0,1]$, δηλαδή $\mathbf{r}_i(u) = (x_i(u), y_i(u))$. Κάθε τμήμα \mathbf{r}_i εκφράζεται ως γραμμικός συνδυασμός τεσσάρων κυβικών πολυωνύμων:

$$\mathbf{r}_i(u) = \mathbf{C}_{i-1}Q_0(u) + \mathbf{C}_iQ_1(u) + \mathbf{C}_{i+1}Q_2(u) + \mathbf{C}_{i+2}Q_3(u), \quad i = 0, \dots, N-1 \quad (9-1)$$

όπου

$$Q_k(u) = a_{k,0}u^3 + a_{k,1}u^2 + a_{k,2}u + a_{k,3}, \quad k = 0, 1, 2, 3 \quad (9-2)$$

Τα κυβικά πολυώνυμα Q_k είναι γνωστά ως *συναρτήσεις βάσης* (basis functions). Επειδή πρόκειται για κλειστή καμπύλη, φαίνεται από την Εξίσωση (9-1) ότι ολόκληρο το B-Spline \mathbf{r} , που αποτελείται από N τμήματα, περιγράφεται από τα N διδιάστατα σημεία \mathbf{C}_i , $i = 0, \dots, N-1$. Τα σημεία αυτά ονομάζονται *σημεία ελέγχου* (control points).

Από τους περιορισμούς C^0 , C^1 , και C^2 , συνέχειας στα σημεία σύνδεσης των τμημάτων, προκύπτουν 15 εξισώσεις των αγνώστων παραμέτρων a_k που εμφανίζονται στην Εξίσωση (9-2). Για παράδειγμα, από τη C^0 συνέχεια, ή συνέχεια θέσης, προκύπτει $\mathbf{r}_i(1) = \mathbf{r}_{i+1}(0)$, $i = 0, \dots, N-1$. Ένας ακόμη περιορισμός προκύπτει από το γεγονός ότι η καμπύλη \mathbf{r} είναι αναλλοίωτη ως προς τους μετασχηματισμούς παραμέτρου:

$$\sum_{k=0}^3 Q_k(u) = 1, \quad u \in [0,1] \quad (9-3)$$

Έτσι σχηματίζονται 16 συνολικά εξισώσεις που επιτρέπουν τον υπολογισμό των 16 παραμέτρων a_k και επομένως τον πλήρη προσδιορισμό των συναρτήσεων βάσης $Q_k(u)$.

Αφού προσδιορισθούν οι συναρτήσεις βάσης, είναι απαραίτητη η παραμετροποίηση ολόκληρης της καμπύλης με μία συνεχή παράμετρο για την περιγραφή του B-Spline. Για το λόγο αυτό χρησιμοποιείται η παράμετρος u' που ορίζεται στο διάστημα $[0, N]$. Τότε για κάθε τμήμα i της καμπύλης ισχύει $u' = u + i$, όπου $u \in [0,1]$ είναι η παράμετρος του τμήματος i . Έτσι η καμπύλη του B-Spline μπορεί να εκφραστεί ως γραμμικός συνδυασμός των τμημάτων $\mathbf{r}_i(u)$ ως εξής:

$$\mathbf{r}(u') \equiv \sum_{i=0}^{N-1} \mathbf{r}_i(u) = \sum_{i=0}^{N-1} \mathbf{r}_i(u' - i) \quad (9-4)$$

όπου το $\mathbf{r}_i(u' - i)$ είναι μη μηδενικό για $u \in [0,1]$, ή ισοδύναμα $u' \in [i, i+1]$. Στις υπόλοιπες περιοχές του διαστήματος $[0, N]$, δηλαδή για $u' \in [0, i) \cup (i+1, N]$, ισχύει $\mathbf{r}_i(u' - i) = 0$. Αντικαθιστώντας τώρα την Εξίσωση (9-1) στην (9-4) προκύπτει

$$\mathbf{r}(u') = \sum_{i=0}^{N+2} \mathbf{C}_i B_i(u') \quad (9-5)$$

όπου τα \mathbf{C}_i ορίζονται αρχικά για $i = 0, \dots, N-1$, και στη συνέχεια πραγματοποιείται περιοδική επέκταση έτσι ώστε $\mathbf{C}_{-1} = \mathbf{C}_{N-1}$, $\mathbf{C}_N = \mathbf{C}_0$, $\mathbf{C}_{N+1} = \mathbf{C}_1$, και $\mathbf{C}_{N+2} = \mathbf{C}_2$. Οι συναρτήσεις $B_i(u')$ ονομάζονται *συναρτήσεις ανάμειξης* (blending functions) [143] και ορίζονται ως εξής:

$$B_i(u') = \begin{cases} Q_3(u' - i + 3), & i - 3 \leq u' < i - 2 \\ Q_2(u' - i + 2), & i - 2 \leq u' < i - 1 \\ Q_1(u' - i + 1), & i - 1 \leq u' < i \\ Q_0(u' - i), & i \leq u' < i + 1 \\ 0, & \text{αλλιώς} \end{cases} \quad (9-6)$$

Με δεδομένα τα σημεία ελέγχου \mathbf{C}_i μπορούν να ορισθούν τα *κομβικά σημεία* (knot points) \mathbf{p}_i , $i = 0, \dots, N-1$, τα οποία ταυτίζονται με τα σημεία σύνδεσης των τμημάτων της καμπύλης. Γενικά δηλαδή ισχύει $\mathbf{p}_i = \mathbf{r}_i(0) = \mathbf{r}_{i-1}(1)$. Από τις Εξισώσεις (9-1) και (9-2), και με αντικατάσταση

των πραγματικών τιμών των συντελεστών a_k , προκύπτει ότι τα κομβικά σημεία δίνονται από την ακόλουθη σχέση:

$$\mathbf{p}_i = \frac{1}{6}\mathbf{C}_{i-1} + \frac{2}{3}\mathbf{C}_i + \frac{1}{6}\mathbf{C}_{i+1}, \quad i = 0, \dots, N-1 \quad (9-7)$$

Από την παραπάνω ανάλυση προκύπτει ότι διαφορετικά ζεύγη σημείων ελέγχου και κομβικών σημείων μπορούν να περιγράψουν την ίδια καμπύλη B-Spline.

Εκτίμηση Σημείων Ελέγχου

Όταν ένα περίγραμμα είναι διαθέσιμο ως διατεταγμένο σύνολο 2-Δ δεδομένων σημείων (data points), η μοντελοποίησή του με B-Splines πραγματοποιείται με την προσαρμογή (fitting) μιας καμπύλης B-Spline στα δεδομένα σημεία, δηλαδή με τον προσδιορισμό των σημείων ελέγχου από τα δεδομένα σημεία. Έστω ότι είναι διαθέσιμα M δεδομένα σημεία $\mathbf{s}_j, j = 0, 1, \dots, M-1$. Η εκτίμηση των σημείων ελέγχου γίνεται με τέτοιο τρόπο ώστε να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα μεταξύ των δεδομένων σημείων και της καμπύλης B-Spline:

$$E^2 = \sum_{j=0}^{M-1} \|\mathbf{s}_j - \mathbf{r}(u'_j)\|^2 \quad (9-8)$$

όπου $\|\cdot\|$ είναι η Ευκλείδεια νόρμα και τα $u'_j, j = 0, \dots, M-1$, είναι κατάλληλα επιλεγμένες τιμές της παραμέτρου u' που ορίζονται παρακάτω. Αποδεικνύεται ότι τα σημεία ελέγχου που δίνουν το ελάχιστο μέσο τετραγωνικό σφάλμα (minimum mean square error, MMSE) δίνονται σε μορφή πίνακα από τη σχέση

$$\mathbf{C}_f = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{f} \quad (9-9)$$

όπου \mathbf{C}_f και \mathbf{f} είναι πίνακες διαστάσεων $N \times 2$ και $M \times 2$ αντίστοιχα, οι οποίοι περιέχουν τα σημεία ελέγχου \mathbf{C}_i και τα δεδομένα σημεία \mathbf{s}_j αντίστοιχα. Ο πίνακας \mathbf{P} , διαστάσεων $M \times N$, περιέχει τιμές των συναρτήσεων ανάμειξης στις επιλεγμένες τιμές u'_j της παραμέτρου u' [12]:

$$\mathbf{P} = \begin{bmatrix} B_0(u'_0) + B_N(u'_0) & B_0(u'_1) + B_N(u'_1) & \dots & B_0(u'_{M-1}) + B_N(u'_{M-1}) \\ B_1(u'_0) + B_{N+1}(u'_0) & B_1(u'_1) + B_{N+1}(u'_1) & \dots & B_1(u'_{M-1}) + B_{N+1}(u'_{M-1}) \\ B_2(u'_0) + B_{N+2}(u'_0) & B_2(u'_1) + B_{N+2}(u'_1) & \dots & B_2(u'_{M-1}) + B_{N+2}(u'_{M-1}) \\ B_3(u'_0) & B_3(u'_1) & \dots & B_3(u'_{M-1}) \\ \vdots & \vdots & \ddots & \vdots \\ B_{N-1}(u'_0) & B_{N-1}(u'_1) & \dots & B_{N-1}(u'_{M-1}) \end{bmatrix}^T \quad (9-10)$$

Επανερχόμενοι στις τιμές $u'_j, j = 0, \dots, M-1$, της παραμέτρου u' , η εκτίμησή τους γίνεται με τη μέθοδο *Chord Length* (μήκος χορδής) ή CL. Σύμφωνα με τη μέθοδο αυτή, κάθε τιμή u'_j αντιστοιχίζεται σε ένα δεδομένο σημείο $\mathbf{s}_j, j = 0, 1, \dots, M-1$ με τέτοιο τρόπο ώστε η απόσταση μεταξύ των δύο σημείων της καμπύλης $\mathbf{r}(u'_i), \mathbf{r}(u'_j)$ να προσεγγίζει την απόσταση μεταξύ των δύο αντίστοιχων δεδομένων σημείων $\mathbf{s}_i, \mathbf{s}_j$, για κάθε $i, j = 0, 1, \dots, M-1$. Πιο συγκεκριμένα, ορίζεται αρχικά $u'_0 = 0$ και $u'_{\max} = N-3$. Στη συνέχεια, η τιμή u'_j που αντιστοιχεί στο σημείο \mathbf{s}_j , υπολογίζεται για $j > 0$ από την αναδρομική σχέση

$$u'_j = u'_{j-1} + u'_{\max} \frac{\|\mathbf{s}_j - \mathbf{s}_{j-1}\|}{\sum_{l=2}^m \|\mathbf{s}_l - \mathbf{s}_{l-1}\|}, \quad j = 1, \dots, M-1 \quad (9-11)$$

Η μέθοδος βασίζεται στο γεγονός ότι το μήκος χορδής μεταξύ δύο σημείων της καμπύλης είναι πολύ καλή προσέγγιση του αντίστοιχου μήκους τόξου όταν τα σημεία είναι κοντά μεταξύ τους. Είναι ανθεκτική στον ομοιόμορφα κατανεμημένο θόρυβο αλλά δεν δίνει ικανοποιητικά αποτελέσματα όταν η δειγματοληψία των σημείων είναι ανομοιόμορφη, ή όταν υπάρχει ανομοιόμορφα κατανεμημένος θόρυβος. Εναλλακτικά μπορεί να χρησιμοποιηθεί η μέθοδος *Inverse Chord Length* (ICL) [56], η οποία δίνει καλύτερα αποτελέσματα στις περιπτώσεις αυτές.

Ανακατανομή Κομβικών Σημείων

Από την παραπάνω ανάλυση προκύπτει ότι το ταίριασμα δύο καμπυλών B-Splines δεν μπορεί να πραγματοποιηθεί με βάση τα σημεία ελέγχου, διότι διαφορετικά σύνολα σημείων ελέγχου μπορούν να περιγράψουν την ίδια ακριβώς καμπύλη. Για το λόγο αυτό χρησιμοποιούνται τα κομβικά σημεία $\mathbf{p}_i, i = 0, \dots, N-1$, τα οποία ανήκουν στην καμπύλη και μπορούν να υπολογισθούν από τα σημεία ελέγχου με χρήση της Εξίσωσης (9-7). Εναλλακτικά, τα κομβικά σημεία δίνονται σε μορφή πίνακα από τη σχέση

$$\mathbf{p}_f = \mathbf{A} \mathbf{C}_f \quad (9-12)$$

όπου ο \mathbf{p}_f είναι πίνακας διαστάσεων $N \times 2$ που περιέχει τα κομβικά σημεία, ενώ ο $N \times N$ πίνακας \mathbf{A} ορίζεται ως

$$\mathbf{A} = \begin{bmatrix} 2/3 & 1/6 & 0 & 0 & \dots & 0 & 1/6 \\ 1/6 & 2/3 & 1/6 & 0 & \dots & 0 & 0 \\ \vdots & & \vdots & & & \vdots & \\ 1/6 & 0 & 0 & 0 & \dots & 1/6 & 2/3 \end{bmatrix} \quad (9-13)$$

Πρέπει να σημειωθεί ότι, παρά το γεγονός ότι τα κομβικά σημεία ανήκουν στο B-Spline, δεν υπάρχει εγγύηση ότι δύο σύνολα κομβικών σημείων της ίδιας ή παρόμοιων καμπυλών B-Splines αντιστοιχούν μεταξύ τους, ακόμη και αν περιέχουν το ίδιο πλήθος σημείων. Για να υπάρχει μια τέτοια αντιστοιχία πρέπει να πραγματοποιηθεί *ανακατανομή* (reallocation) των κομβικών σημείων με τέτοιο τρόπο ώστε να αντιστοιχούν σε ισαπέχουσες τιμές της παραμέτρου u' σε κάθε καμπύλη. Το πρώτο κομβικό σημείο είναι το *αρχικό σημείο* ή *σημείο αναφοράς* και η επιλογή του είναι αρχικά αυθαίρετη. Στη συνέχεια, αφού πραγματοποιηθεί κανονικοποίηση του σημείου αναφοράς σύμφωνα με την προτεινόμενη μέθοδο που περιγράφεται στην Ενότητα 10.2, εκτελείται εκ νέου ανακατανομή κομβικών σημείων. Με την παραπάνω διαδικασία επιτυγχάνεται ομοιόμορφη δειγματοληψία ως προς το μήκος τόξου και σωστή αντιστοιχία μεταξύ των κομβικών σημείων.

Κανονικοποίηση Περιγραμμάτων

Στο Κεφάλαιο αυτό περιγράφεται αναλυτικά η προτεινόμενη μέθοδος κανονικοποίησης των περιγραμμάτων που έχουν ήδη μοντελοποιηθεί με χρήση B-Splines. Στο πρώτο στάδιο, με χρήση ρολών μέχρι δεύτερου βαθμού πραγματοποιείται ορθοκανονικοποίηση, δηλαδή κανονικοποίηση ως προς τους μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας και κλίσης (skew). Με την ορθοκανονικοποίηση οποιοσδήποτε affine μετασχηματισμός ανάγεται σε ορθογώνιο, δηλαδή σε μετασχηματισμό που περιλαμβάνει μόνο περιστροφή ή / και κατοπτρισμό. Στη συνέχεια με χρήση της φάσης κατάλληλων συντελεστών του μετασχηματισμού Fourier της καμπύλης (στην αναπαράστασή της με μιγαδικό διάνυσμα) κανονικοποιείται το σημείο αναφοράς της καμπύλης, το οποίο εν γένει μπορεί να διαφέρει λόγω μετασχηματισμού κυκλικής ολίσθησης. Στο τελευταίο στάδιο κανονικοποιείται η περιστροφή και ο κατοπτρισμός και πάλι με χρήση πληροφορίας του μετασχηματισμού Fourier. Σε όλα τα στάδια περιλαμβάνεται θεωρητική ανάλυση από την οποία προκύπτει ότι η αναπαράσταση των κανονικοποιημένων περιγραμμάτων είναι αναλλοίωτη στους affine μετασχηματισμούς και δεν προκαλεί απώλεια πληροφορίας αφού η διαδικασία κανονικοποίησης είναι αντιστρέψιμη. Παραδείγματα και πειραματικά αποτελέσματα της προτεινόμενης μεθόδου δίνονται στο Κεφάλαιο 11, μαζί με αντίστοιχα πειράματα ταξινόμησης και αναζήτησης με βάση το σχήμα αντικειμένων.

10.1 ΟΡΘΟΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Με τη χρήση των B-Splines για τη μοντελοποίηση των καμπυλών που προέρχονται από τα κλειστά περιγράμματα αντικειμένων, επιτυγχάνεται απλοποίηση του σχήματος των αντικειμένων, μείωση του θορύβου κατάτμησης, καθώς και ομοιόμορφη δειγματοληψία ως προς το μήκος τόξου. Στη συνέχεια ακολουθεί η κανονικοποίηση της κάθε καμπύλης, η οποία πλέον περιγράφεται από τα κομβικά της σημεία. Το πρώτο στάδιο της κανονικοποίησης είναι μία διαδικασία ορθοκανονικοποίησης (orthogonalization), η οποία κανονικοποιεί την καμπύλη ως προς τους μετασχηματισμούς μετατόπισης (translation), αλλαγής κλίμακας (scaling) και κλίσης (skew). Έτσι κάθε affine μετασχηματισμός ανάγεται ουσιαστικά σε ορθογώνιο (orthogonal) μετασχηματισμό, δηλαδή σε μετασχηματισμό που περιλαμβάνει μόνο περιστροφή (rotation) ή και κατοπτρισμό (reflection). Παρακάτω παρουσιάζεται αρχικά ο προτεινόμενος αλγόριθμος ορθοκανονικοποίησης, ενώ στη συνέχεια αποδεικνύεται ότι η αναπαράσταση της καμπύλης που προκύπτει είναι αναλλοίωτη στη μετατόπιση, την αλλαγή κλίμακας και την κλίση.

Αλγόριθμος Ορθοκανονικοποίησης

Έστω ότι $\mathbf{s}_i = [x_i \ y_i]^T$, $i = 0, \dots, N-1$, είναι τα N διδιάστατα σημεία που περιγράφουν μία καμπύλη και προκύπτουν από τη μοντελοποίηση με B-Splines, δηλαδή τα N κομβικά σημεία της

καμπύλης. Στο εξής θα χρησιμοποιείται ο $2 \times N$ πίνακας $\mathbf{s} = [\mathbf{s}_0 \mathbf{s}_1 \dots \mathbf{s}_{N-1}]$ για την περιγραφή της καμπύλης, ο οποίος επιτρέπει εύκολο συμβολισμό των affine μετασχηματισμών που εμπλέκουν πολλαπλασιασμό με 2×2 πίνακες. Με όμοιο τρόπο η οριζόντια και κατακόρυφη συντεταγμένη των σημείων θα συμβολίζεται με το διάνυσμα $\mathbf{x} = [x_0 x_1 \dots x_{N-1}]$ και $\mathbf{y} = [y_0 y_1 \dots y_{N-1}]$ αντίστοιχα, διαστάσεων $1 \times N$. Η ορθοκανονικοποίηση της καμπύλης βασίζεται στις ροπές έως και δευτέρου βαθμού, οι οποίες ορίζονται ως

$$m_{p,q}(\mathbf{s}) = \frac{1}{N} \sum_{i=0}^{N-1} x_i^p y_i^q \quad (10-1)$$

όπου $m_{p,q}(\mathbf{s})$ είναι η ροπή βαθμού (p,q) της καμπύλης \mathbf{s} . Χωρίς βλάβη της γενικότητας γίνεται στη συνέχεια η υπόθεση ότι τα σημεία \mathbf{s}_i , $i = 0, \dots, N-1$ δεν είναι συγγραμμικά, έτσι ώστε $m_{2,0} \neq 0$ και $m_{0,2} \neq 0$. Σε μια τέτοια περίπτωση, τα βήματα κανονικοποίησης που εμπλέκουν διαίρεση με το μηδέν παραλείπονται.

Η διαδικασία ορθοκανονικοποίησης αποτελείται από μια σειρά γραμμικών μετασχηματισμών (μετατόπιση, αλλαγή κλίμακας και περιστροφή), οι οποίοι είναι ανεξάρτητοι από το σημείο αναφοράς και γενικότερα από τη σειρά των σημείων στην καμπύλη. Για απλότητα συμβολισμών, η πρόσθεση ή αφαίρεση ενός βαθμωτού από ένα διάνυσμα θα συμβολίζει στην παρακάτω ανάλυση την πρόσθεση ή αφαίρεση του βαθμωτού από όλα τα στοιχεία του διανύσματος. Συγκεκριμένα πραγματοποιούνται τα παρακάτω βήματα κανονικοποίησης:

- (α) Το κέντρο βάρους της καμπύλης κανονικοποιείται ώστε να συμπίπτει με την αρχή των αξόνων:

$$\mathbf{x}_1 = \mathbf{x} - \mu_x, \quad \mathbf{y}_1 = \mathbf{y} - \mu_y \quad (10-2\alpha)$$

όπου $\mu_x = m_{1,0}(\mathbf{s})$, $\mu_y = m_{0,1}(\mathbf{s})$.

- (β) Η κλίμακα της καμπύλης μεταβάλλεται οριζόντια και κατακόρυφα ώστε οι ροπές δευτέρου βαθμού να κανονικοποιηθούν στη μονάδα:

$$\mathbf{x}_2 = \sigma_x \mathbf{x}_1, \quad \mathbf{y}_2 = \sigma_y \mathbf{y}_1 \quad (10-2\beta)$$

όπου $\sigma_x = 1/\sqrt{m_{2,0}(\mathbf{s}_1)}$, $\sigma_y = 1/\sqrt{m_{0,2}(\mathbf{s}_1)}$.

- (γ) Η καμπύλη περιστρέφεται αριστερόστροφα κατά γωνία $\theta_0 = \pi/4$:

$$\mathbf{s}_3 = \mathbf{R}_{\pi/4} \mathbf{s}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{x}_2 - \mathbf{y}_2 \\ \mathbf{x}_2 + \mathbf{y}_2 \end{bmatrix} \quad (10-2\gamma)$$

όπου \mathbf{R}_θ είναι ένας 2×2 πίνακας που αντιστοιχεί σε αριστερόστροφη περιστροφή κατά θ ακτίνια.

- (δ) Τέλος η κλίμακα της καμπύλης μεταβάλλεται πάλι, ακριβώς όπως στο βήμα (β):

$$\mathbf{x}_4 = \tau_x \mathbf{x}_3, \quad \mathbf{y}_4 = \tau_y \mathbf{y}_3 \quad (10-2\delta)$$

όπου $\tau_x = 1/\sqrt{m_{2,0}(\mathbf{s}_3)}$, $\tau_y = 1/\sqrt{m_{0,2}(\mathbf{s}_3)}$.

Η ορθοκανονικοποιημένη καμπύλη ορίζεται ως $n_a(\mathbf{s}) \equiv \mathbf{s}_4$, ενώ τα παραπάνω βήματα κανονικοποίησης συνοψίζονται σε μορφή πινάκων ως εξής:

$$n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s}) (\mathbf{s} - \boldsymbol{\mu}(\mathbf{s})) = \frac{1}{\sqrt{2}} \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix} \quad (10-3)$$

όπου $\mathbf{\mu}(s) = [m_{1,0}(s) \ m_{0,1}(s)]^T = [\mu_x \ \mu_y]^T$ και ο 2×2 πίνακας $\mathbf{N}(s)$ ονομάζεται *πίνακας ορθοκανονικοποίησης* της καμπύλης s . Παρά το γεγονός ότι η εξάρτηση των παραμέτρων $\mu_x, \mu_y, \sigma_x, \sigma_y, \tau_x, \tau_y$, από την s παραλήφθηκε για απλότητα συμβολισμών, ο πίνακας ορθοκανονικοποίησης εξακολουθεί να είναι συνάρτηση της s . Από τις Εξισώσεις 10-2α,β προκύπτει ότι η καμπύλη s_2 που δημιουργείται στο βήμα κανονικοποίησης (β) ικανοποιεί τις συνθήκες $m_{1,0}(s_2) = m_{0,1}(s_2) = 0$ και $m_{2,0}(s_2) = m_{0,2}(s_2) = 1$. Τα επιπλέον βήματα κανονικοποίησης (γ), (δ) απαιτούνται ώστε να ισχύει και η επιπρόσθετη συνθήκη $m_{1,1}(n_a(s)) = 0$. Η παρακάτω Πρόταση συνοψίζει τις αναγκαίες και ικανές συνθήκες για να είναι μια καμπύλη ορθοκανονικοποιημένη:

ΠΡΟΤΑΣΗ 10-1. Για κάθε αρχική καμπύλη s , η ορθοκανονικοποιημένη καμπύλη $n_a(s)$ που ορίζεται στην Εξίσωση (10-3) ικανοποιεί τις παρακάτω συνθήκες:

$$m_{1,0}(n_a(s)) = m_{0,1}(n_a(s)) = m_{1,1}(n_a(s)) = 0 \quad (10-4\alpha)$$

$$m_{2,0}(n_a(s)) = m_{0,2}(n_a(s)) = 1 \quad (10-4\beta)$$

Επιπλέον, οι παραπάνω συνθήκες ικανοποιούνται αν και μόνο αν η γωνία θ_0 που χρησιμοποιείται στο βήμα (γ) της κανονικοποίησης είναι ίση με $k\pi/2 + \pi/4, k \in \mathbb{Z}$.

Η απόδειξη της Πρότασης δίνεται στην Ενότητα 10.5. Οι συνθήκες (10-4) πρακτικά σημαίνουν ότι ο $2 \times N$ πίνακας που αναπαριστά την καμπύλη $n_a(s)$ είναι *ορθογώνιος*, δηλαδή $n_a(s)(n_a(s))^T = \mathbf{I}_2$, όπου \mathbf{I}_2 είναι ο 2×2 μοναδιαίος πίνακας. Σαν συνέπεια, αποδεικνύεται παρακάτω ότι αυτού του είδους η κανονικοποίηση ανάγει έναν αυθαίρετο affine μετασχηματισμό σε ορθογώνιο, απαλείφοντας έτσι τη μετατόπιση, την αλλαγή κλίμακας και την κλίση (skew). Επιπλέον, η διαδικασία ορθοκανονικοποίησης είναι και η ίδια ένας affine μετασχηματισμός, όπως φαίνεται στην (10-2). Έτσι δεν υφίσταται απώλεια πληροφορίας: η αρχική καμπύλη s μπορεί να ανακατασκευασθεί με την εφαρμογή του αντίστροφου μετασχηματισμού.

Κανονικοποίηση Μετατόπισης, Κλίμακας και Κλίσης

Ας θεωρήσουμε τώρα δύο καμπύλες s, s' που σχετίζονται μέσω ενός affine μετασχηματισμού:

$$\mathbf{s}' = \mathbf{A}\mathbf{s} + \mathbf{t} = \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (10-5)$$

όπου ο πίνακας \mathbf{A} είναι ομαλός. Στην αντίθετη περίπτωση όπου $\det \mathbf{A} = 0$, κάθε 2-Δ σημείο απεικονίζεται σε μία ευθεία γραμμή και η κανονικοποίηση είναι αδύνατη. Παρατηρούμε ότι $\mathbf{\mu}(s') = \mathbf{A}\mathbf{\mu}(s) + \mathbf{t}$, έτσι ώστε $\mathbf{s}'_1 = \mathbf{s}' - \mathbf{\mu}(s') = \mathbf{A}(\mathbf{s} - \mathbf{\mu}(s)) = \mathbf{A}\mathbf{s}_1$. Δηλαδή μετά το βήμα (α) της κανονικοποίησης μετατόπισης, οι καμπύλες \mathbf{s}_1 και \mathbf{s}'_1 σχετίζονται μέσω ενός πολλαπλασιασμού με τον 2×2 πίνακα \mathbf{A} . Προκύπτει τότε ότι η σχέση μεταξύ των ροπών των δύο καμπυλών δίνεται από τις Εξισώσεις

$$m_{2,0}(\mathbf{s}'_1) = a^2 m_{2,0}(\mathbf{s}_1) + b^2 m_{0,2}(\mathbf{s}_1) + 2ab m_{1,1}(\mathbf{s}_1) \quad (10-6\alpha)$$

$$m_{0,2}(\mathbf{s}'_1) = c^2 m_{2,0}(\mathbf{s}_1) + d^2 m_{0,2}(\mathbf{s}_1) + 2cd m_{1,1}(\mathbf{s}_1) \quad (10-6\beta)$$

$$m_{1,1}(\mathbf{s}'_1) = ac m_{2,0}(\mathbf{s}_1) + bd m_{0,2}(\mathbf{s}_1) + (ad + bc) m_{1,1}(\mathbf{s}_1) \quad (10-6\gamma)$$

Αν η \mathbf{s}_1 είναι ορθοκανονικοποιημένη, δηλαδή ικανοποιεί τις συνθήκες (10-4), τότε οι παραπάνω Εξισώσεις καταλήγουν στις

$$m_{2,0}(\mathbf{s}'_1) = a^2 + b^2 \quad (10-7\alpha)$$

$$m_{0,2}(\mathbf{s}'_1) = c^2 + d^2 \quad (10-7\beta)$$

$$m_{1,1}(\mathbf{s}'_1) = ac + bd \quad (10-7\gamma)$$

Αυτό σημαίνει, για παράδειγμα, ότι όταν μία ορθοκανονικοποιημένη καμπύλη περιστρέφεται ή κατοπτρίζεται, στην οποία περίπτωση ο \mathbf{A} είναι ορθογώνιος, τότε η καμπύλη παραμένει ορθοκανονικοποιημένη. Αντίστροφα, αν και οι δύο καμπύλες είναι ορθοκανονικοποιημένες, τότε ο \mathbf{A} πρέπει να είναι ορθογώνιος. Με αυτή την έννοια, η Πρόταση 10-2, που αποδεικνύεται στην Ενότητα 10.5, εκφράζει τη σχέση μεταξύ δύο ορθοκανονικοποιημένων καμπυλών όταν είναι γνωστή η σχέση μεταξύ των αντίστοιχων αρχικών καμπυλών:

ΠΡΟΤΑΣΗ 10-2. Αν δύο καμπύλες \mathbf{s}, \mathbf{s}' σχετίζονται μέσω ενός affine μετασχηματισμού, υπάρχει ένας ορθογώνιος 2×2 πίνακας \mathbf{Q} τέτοιος ώστε

$$n_a(\mathbf{s}') = \mathbf{Q} n_a(\mathbf{s}) \quad (10-8)$$

Επιπλέον, η ίδια σχέση ισχύει μεταξύ της $n_a(\mathbf{s})$ και μιας κανονικοποιημένης καμπύλης $\tilde{n}_a(\mathbf{s})$ που παράγεται από οποιαδήποτε διαδικασία κανονικοποίησης η οποία βασίζεται στις ιδιότητες (10-4).

Έτσι η προτεινόμενη διαδικασία ορθοκανονικοποίησης ανάγει τους affine μετασχηματισμούς σε ορθογώνιους μετασχηματισμούς, οι οποίοι περιέχουν μόνο περιστροφή ή / και κατοπτρισμό (ανάλογα με το αν $\det \mathbf{Q} = 1$ ή $\det \mathbf{Q} = -1$). Οι ορθοκανονικοποιημένες καμπύλες είναι λοιπόν αναλλοίωτες στους μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας και κλίσης (skew).

Είναι αξιοσημείωτο ότι η διαδικασία ορθοκανονικοποίησης πραγματοποιείται *χωρίς γνώση* των παραμέτρων του affine μετασχηματισμού και *χωρίς ταίριασμα* μεταξύ των δύο καμπυλών \mathbf{s}, \mathbf{s}' , δηλαδή η \mathbf{s} κανονικοποιείται χωρίς γνώση της \mathbf{s}' και αντίστροφα. Σημειώνεται επίσης ότι οι παράμετροι μετασχηματισμού $\{\mu_x, \mu_y, \sigma_x, \sigma_y, \tau_x, \tau_y\}$ υπολογίζονται άμεσα από τις ροπές της καμπύλης μέχρι και δευτέρου βαθμού. Επιπλέον, αυτό το σύνολο παραμέτρων μαζί με την $n_a(\mathbf{s})$ περιέχουν όλη την πληροφορία της αρχικής καμπύλης \mathbf{s} : παρά το γεγονός ότι η $n_a(\mathbf{s})$ ικανοποιεί πέντε μόνο περιοριστικές συνθήκες (10-4), οι παράμετροι είναι έξι διότι οι τ_x και τ_y συνδέονται με τη σχέση $1/\tau_x^2 + 1/\tau_y^2 = 1$. Ο μετασχηματισμός (10-3) είναι αντιστρέψιμος και επομένως αν οι παράμετροί του είναι γνωστοί, η αρχική καμπύλη \mathbf{s} μπορεί να ανακατασκευασθεί από την $n_a(\mathbf{s})$ με την εφαρμογή του αντίστροφου μετασχηματισμού.

Έτσι επιβεβαιώνεται η *μοναδικότητα* και η *πληρότητα* της $n_a(\mathbf{s})$, αφού περιέχει όλη την πληροφορία σχήματος της \mathbf{s} , εκτός από το μετασχηματισμό στον οποίο είναι αναλλοίωτη. Οι ιδιότητες αυτές είναι πολύτιμες για την ακριβή σύγκριση καμπυλών, με εφαρμογές στην ταξινόμηση, την αναγνώριση και την αναζήτηση με βάση το σχήμα. Επιπλέον, αφού η διαδικασία κανονικοποίησης δεν βασίζεται στο ταίριασμα καμπυλών, μπορεί να χρησιμοποιηθεί οποιαδήποτε μέθοδος ταξινόμησης, συμπεριλαμβανομένων για παράδειγμα των νευρωνικών δικτύων. Τέλος, όπως επιδεικνύεται στα πειράματα, η προτεινόμενη μέθοδος κανονικοποίησης είναι ιδιαίτερα ανθεκτική στο θόρυβο (ο οποίος μπορεί να προέρχεται από την κατάτμηση ή τη δειγματοληψία) και στις ανομοιογένειες ή παραμορφώσεις καμπυλών. Τούτο είναι αναμενόμενο αφού η μέθοδος βασίζεται σε ολικά (global) χαρακτηριστικά, όπως είναι οι ροπές των καμπυλών.

10.2 ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ ΣΗΜΕΙΟΥ ΑΝΑΦΟΡΑΣ

Στο σημείο αυτό έχει επιτευχθεί η αναγωγή των affine μετασχηματισμών σε ορθογώνιους· θα ήταν επαρκές λοιπόν να βρεθεί ένας μετασχηματισμός που να είναι αναλλοίωτος στην περιστροφή και τον κατοπτρισμό. Η συνολική κανονικοποίηση (δηλαδή ορθοκανονικοποίηση και κανονικοποίηση περιστροφής και κατοπτρισμού) θα ήταν αναλλοίωτη σε οποιοδήποτε affine μετασχηματισμό. Όμως, σε αντίθεση με την ορθοκανονικοποίηση, η διαδικασία κανονικοποίησης περιστροφής και κατοπτρισμού εμπλέκει ποσότητες που εξαρτώνται από το επιλεγμένο σημείο αναφοράς της κλειστής καμπύλης, και γενικότερα από τη σειρά εμφάνισης των σημείων στην καμπύλη. Για παράδειγμα, η εκτίμηση περιστροφής εξαρτάται από το σημείο αναφοράς ενώ η εκτίμηση κατοπτρισμού εξαρτάται από την περιστροφή. Αυτό συμβαίνει διότι η περιστροφή και το αρχικό σημείο αναφοράς είναι αλληλένδετες έννοιες· στην ειδική περίπτωση μάλιστα ενός κυκλικού σχήματος ταυτίζονται απόλυτα. Για τον παραπάνω λόγο πραγματοποιείται πρώτα κανονικοποίηση του σημείου αναφοράς, ενώ η κανονικοποίηση περιστροφής και κατοπτρισμού έπεται στη συνέχεια.

Στην ανάλυση που ακολουθεί, η κανονικοποίηση βασίζεται στο διακριτό μετασχηματισμό Fourier των σημείων της καμπύλης. Εξ' ορισμού ο μετασχηματισμός αυτός εξαρτάται από το σημείο αναφοράς σε αντίθεση με τις ροπές οι οποίες είναι αναλλοίωτες. Για το σκοπό αυτό δεν χρησιμοποιείται πλέον συμβολισμός πινάκων για τις καμπύλες, αλλά μιγαδικών διάνυσμάτων. Συγκεκριμένα, η οριζόντια και κατακόρυφη συντεταγμένη των σημείων της καμπύλης εξακολουθεί να συμβολίζεται με το διάνυσμα $\mathbf{x} = [x_0 \ x_1 \ \dots \ x_{N-1}]$ και $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_{N-1}]$, διαστάσεων $N \times 1$, αντίστοιχα· όμως η καμπύλη ως σύνολο συμβολίζεται με το μιγαδικό διάνυσμα $\mathbf{z} = \mathbf{x} + j\mathbf{y} = [z_0 \ \dots \ z_{N-1}]^T$, διαστάσεων επίσης $N \times 1$, όπου το $z_i = x_i + jy_i$, $i = 0, 1, \dots, N-1$, υποδηλώνει το i -οστό σημείο της καμπύλης.

Η διαδικασία κανονικοποίησης σημείου αναφοράς βασίζεται στο διακριτό μετασχηματισμό Fourier. Για μια καμπύλη N σημείων που αναπαρίστανται από το μιγαδικό διάνυσμα \mathbf{z} , ο μετασχηματισμός αυτός ορίζεται ως

$$\mathbf{u} = \mathfrak{F}(\mathbf{z}) = \mathbf{W} \mathbf{z} = \begin{bmatrix} w^0 & w^0 & \dots & w^0 \\ w^0 & w^1 & \dots & w^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ w^0 & w^{N-1} & \dots & w^{(N-1)^2} \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{N-1} \end{bmatrix} \quad (10-9)$$

όπου $w = e^{j2\pi/N}$, έτσι ώστε $w^{kN} = 1$, $k \in \mathbb{Z}$. Ισοδύναμα,

$$u_k = \sum_{i=0}^{N-1} z_i w^{-ki}, \quad k = 0, \dots, N-1 \quad (10-10)$$

Για κάθε στοιχείο u_k , $k = 0, \dots, N-1$, του μετασχηματισμού Fourier ορίζεται το *πρωτεύον όρισμα* (primary argument), ή *φάση* (phase), $a_k = \text{Arg } u_k = \theta \in [0, 2\pi)$: $u_k = r e^{j\theta}$, όπου $r \in \mathbb{R}^+$. Σχηματίζεται έτσι ένα *διάνυσμα φάσης* (phase vector) $\mathbf{a} = \text{Arg } \mathbf{u} = [a_0 \ \dots \ a_{N-1}]$. Θεωρούμε τώρα μία δεύτερη καμπύλη $\mathbf{z}' = [z'_0 \ \dots \ z'_{N-1}]$ η οποία είναι ίδια με τη \mathbf{z} εκτός από το σημείο αναφοράς, δηλαδή το διάνυσμα \mathbf{z}' έχει υποστεί *κυκλική ολίσθηση* (circular shift) σε σχέση με το \mathbf{z} κατά m στοιχεία, όπου $m \in \{0, \dots, N-1\}$:

$$\mathbf{z}' = S_m(\mathbf{z}): z'_i = z_{(i+m) \bmod N}, \quad i = 0, \dots, N-1 \quad (10-11)$$

Τότε, αν ορίσουμε με όμοιο τρόπο $\mathbf{u}' = \mathfrak{Z}(\mathbf{z}')$ και $\mathbf{a}' = \text{Arg } \mathbf{u}'$, αποδεικνύεται ότι $u'_k = w^{km} u_k$, $k = 0, \dots, N-1$, ή ισοδύναμα

$$a'_k = (a_k + 2\pi km/N) \bmod 2\pi, \quad k = 0, \dots, N-1 \quad (10-12)$$

Φαίνεται δηλαδή ότι η κυκλική ολίσθηση μεταξύ των \mathbf{z} και \mathbf{z}' ανάγεται στην πολύ απλή γραμμική σχέση (10-12) μεταξύ των αντίστοιχων διανυσμάτων φάσης \mathbf{a} και \mathbf{a}' . Αν και οι δύο καμπύλες \mathbf{z} και \mathbf{z}' ήταν ταυτόχρονα γνωστές, το ταίριασμά τους θα ήταν εφικτό με την εκτίμηση του m από την Εξίσωση (10-12) και την εφαρμογή της αντίθετης κυκλικής ολίσθησης στην \mathbf{z}' . Τούτο όμως δεν είναι δυνατό για μία μέθοδο κανονικοποίησης, στην οποία μία μόνο καμπύλη είναι γνωστή κάθε φορά. Ακόμη και αν είναι γνωστές και οι δύο καμπύλες, το m μπορεί να εκτιμηθεί μόνο αν οι καμπύλες αυτές προέρχονται από την ίδια ακριβώς καμπύλη με εφαρμογή διαφορετικής κυκλικής ολίσθησης. Σε κάθε άλλη περίπτωση, τα διανύσματα φάσης είναι τόσο θορυβώδη που η (10-12) δεν μπορεί να δώσει καμία χρήσιμη πληροφορία. Για τους παραπάνω λόγους προτείνεται η εκτίμηση μιας *πρότυπης* κυκλικής ολίσθησης για κάθε καμπύλη, η οποία βασίζεται στη διαφορά μεταξύ του πρώτου και του τελευταίου στοιχείου του διανύσματος φάσης:

$$p(\mathbf{z}) = \left\lfloor \frac{N}{4\pi} (a_1 - a_{N-1}) \right\rfloor \bmod N/2 \quad (10-13)$$

και στη συνέχεια η εφαρμογή της αντίθετης ολίσθησης για την κανονικοποίηση της καμπύλης:

$$n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z}) \quad (10-14)$$

Είναι εμφανές από την (10-12) ότι η διαφορά δύο διαδοχικών στοιχείων του \mathbf{a}' , για παράδειγμα $a'_2 - a'_1$, σχετίζεται με την αντίστοιχη διαφορά $a_2 - a_1$ μέσω μιας απλής άθροισης με την σταθερά $2\pi m/N$, modulo 2π . Έτσι, αν $a_2 - a_1 = 0$, μπορούμε να εκτιμήσουμε άμεσα το m από τη διαφορά $a'_2 - a'_1$, διαιρώντας δια $2\pi/N$ και στη συνέχεια υπολογίζοντας το modulo N του αποτελέσματος. Για την επιλογή της πρώτης και της τελευταίας φάσης όμως, η οποία δικαιολογείται παρακάτω, τα στοιχεία a_1 και a_{N-1} δεν είναι διαδοχικά· απέχουν κατά δύο θέσεις αφού οι συντελεστές Fourier είναι περιοδικοί με περίοδο N . Έτσι στην περίπτωση αυτή απαιτείται επιπλέον διαίρεση δια 2. Οι παραπάνω συλλογισμοί εξηγούν τον ορισμό του $p(\mathbf{z})$, όπου το ακέραιο μέρος είναι απαραίτητο αφού η κυκλική ολίσθηση γίνεται πάντα κατά ακέραιο αριθμό στοιχείων (αλλιώς απαιτείται νέα δειγματοληψία της καμπύλης).

Αποδεικνύεται ότι αν το N είναι άρτιο, η κανονικοποίηση (10-14) είναι αναλλοίωτη στο σημείο αναφοράς, με εξαίρεση μία αβεβαιότητα στην πρότυπη κυκλική ολίσθηση, που μπορεί να προκαλέσει μία επιπλέον ολίσθηση κατά $N/2$ δείγματα. Αν το N είναι περιττό, μία επιπρόσθετη ολίσθηση κατά ± 1 δείγματα είναι επίσης πιθανή μεταξύ $n_p(\mathbf{z})$ και $n_p(\mathbf{z}')$. Τα συμπεράσματα αυτά συνοψίζονται στην ακόλουθη Πρόταση, η οποία αποδεικνύεται στην Ενότητα 10.5:

ΠΡΟΤΑΣΗ 10-3. Αν δύο καμπύλες \mathbf{z}, \mathbf{z}' σχετίζονται μέσω ενός μετασχηματισμού κυκλικής ολίσθησης κατά m δείγματα, δηλαδή $\mathbf{z}' = S_m(\mathbf{z})$, τότε

$$p(\mathbf{z}') = (p(\mathbf{z}) + m) \bmod N/2 \quad (10-15\alpha)$$

$$p(n_p(\mathbf{z}')) = p(n_p(\mathbf{z})) = 0 \quad (10-15\beta)$$

$$n_p(\mathbf{z}') = \begin{cases} n_p(\mathbf{z}), & 0 \leq p(\mathbf{z}) + m < N/2 \\ S_{N/2}(n_p(\mathbf{z})), & N/2 \leq p(\mathbf{z}) + m < N \end{cases} \quad (10-15\gamma)$$

όπου οι $p(\mathbf{z})$ και $n_p(\mathbf{z})$ ορίζονται στις (10-13) και (10-14) αντίστοιχα.

Παρατηρείται ότι η αθροιστική σχέση (10-15α) οδηγεί στην κανονικοποίηση σημείου αναφοράς αν επιβάλλουμε – μέσω της κυκλικής ολίσθησης (10-14) – στις κανονικοποιημένες καμπύλες να ικανοποιούν τη συνθήκη (10-15β), ακριβώς όπως οι καμπύλες ορθοκανονικοποιούνται όταν ικανοποιούν τις ανάλογες συνθήκες (10-4). Παρατηρείται επίσης ότι η συνθήκη (10-15β) είναι ισοδύναμη με το να επιβάλλουμε στους συντελεστές Fourier u_1 και u_{N-1} να έχουν την ίδια φάση. Θα φανεί παρακάτω ότι αν επιπλέον η φάση αυτή είναι μηδενική, ώστε τα u_1 και u_{N-1} να είναι πραγματικοί και θετικοί αριθμοί, τότε επιτυγχάνεται ταυτόχρονα και κανονικοποίηση περιστροφής.

Επειδή η (10-15α) ισχύει για modulo $N/2$ και όχι για modulo N , απόλυτη κανονικοποίηση σημείου αναφοράς επιτυγχάνεται μόνο για $0 \leq p(\mathbf{z}) + m \leq N/2$. Στην αντίθετη περίπτωση, προκύπτει μία επιπλέον κυκλική ολίσθηση κατά $N/2$ δείγματα στην (10-15γ). Η αβεβαιότητα αυτή οφείλεται στην επιλογή των συντελεστών u_1 και u_{N-1} και μπορεί να αρθεί μόνο έπειτα από την κανονικοποίηση περιστροφής, όπως εξηγείται παρακάτω.

Παρά το γεγονός ότι η επιλογή δύο διαδοχικών συντελεστών, όπως για παράδειγμα των a_1 και a_2 , θα είχε ως αποτέλεσμα την απόλυτη κανονικοποίηση σημείου αναφοράς, χωρίς καμία αβεβαιότητα, η προτεινόμενη επιλογή (10-13) έχει δύο σημαντικά πλεονεκτήματα. Πρώτον, επιτρέπει την επιτυχή ανίχνευση *συμμετριών κατοπτρισμού* (reflectional symmetries), και δεύτερον (και σημαντικότερο) καθιστά την κανονικοποίηση ανθεκτική στο θόρυβο και στις παραμορφώσεις σχήματος, όπως επιδεικνύεται και στα πειράματα. Για παράδειγμα, η επιλογή των a_1 και a_2 δίνει τελείως διαφορετική, και στην ουσία *τυχαία*, εκτίμηση του σημείου αναφοράς για δύο καμπύλες που διαφέρουν σε μικρό βαθμό μεταξύ τους.

Η προτεινόμενη επιλογή των a_1 και a_{N-1} έχει επίσης χρησιμοποιηθεί στην εργασία [100] για την ανίχνευση του σημείου αναφοράς σε καμπύλες που περιγράφουν γραμμικά πρότυπα σχημάτων, όπως π.χ. περιγράμματα χαρακτήρων. Η χρήση άλλων συντελεστών Fourier είναι επίσης δυνατή [84]. Ο *αυτόματος προσδιορισμός* κατάλληλων συντελεστών βασίζεται συνήθως και στο μέτρο τους εκτός από τη φάση και μπορεί να αντιμετωπίσει καλύτερα περιπτώσεις *κυκλικής συμμετρίας* (rotational symmetry) [81,115] με αυξημένο βέβαια υπολογιστικό κόστος. Όμως, τέτοιες περιπτώσεις είναι σπάνιες σε πραγματικές εφαρμογές και η απόδοση της προτεινόμενης κανονικοποίησης είναι ικανοποιητική για φυσικές εικόνες, όσον αφορά στην ταχύτητα και την ποιότητα των αποτελεσμάτων.

10.3 ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ ΠΕΡΙΣΤΡΟΦΗΣ ΚΑΙ ΚΑΤΟΠΤΡΙΣΜΟΥ

Ας υποθέσουμε ότι δύο αρχικές καμπύλες έχουν ορθοκανονικοποιηθεί και στη συνέχεια κανονικοποιηθεί ως προς το σημείο αναφοράς. Έστω λοιπόν ότι οι κανονικοποιημένες καμπύλες

συμβολίζονται \mathbf{s} και \mathbf{s}' . Σύμφωνα με την Πρόταση 10-2, οι κανονικοποιημένες καμπύλες θα ικανοποιούν τη σχέση $\mathbf{s} = \mathbf{Q}\mathbf{s}'$, όπου \mathbf{Q} είναι ένας ορθογώνιος πίνακας διαστάσεων 2×2 . Τότε ο πίνακας \mathbf{Q} μπορεί με μοναδικό τρόπο να αναλυθεί σε γινόμενο ενός πίνακα περιστροφής (rotation matrix) και ενός πίνακα κατοπτρισμού (reflection matrix):

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \quad (10-16)$$

όπου $\theta \in [0, \pi)$, $s_x = \pm 1$, και $s_y = \pm 1$, έτσι ώστε να υπάρχει ένα-προς-ένα απεικόνιση μεταξύ των παραμέτρων αυτών και των στοιχείων του πίνακα \mathbf{Q} . Με άλλα λόγια, ο πίνακας \mathbf{Q} αντιστοιχεί με μοναδικό τρόπο σε μία περιστροφή μεταξύ 0 και π ακτινίων, καθώς και σε ένα πιθανό οριζόντιο και / ή κατακόρυφο κατοπτρισμό. Για απλότητα υιοθετείται και πάλι ο συμβολισμός μιγαδικών διανυσμάτων \mathbf{z} , \mathbf{z}' για τις καμπύλες \mathbf{s} και \mathbf{s}' αντίστοιχα. Έτσι, με χρήση της Εξίσωσης (10-16), η σχέση $\mathbf{s} = \mathbf{Q}\mathbf{s}'$ γράφεται για τα μιγαδικά διανύσματα ως εξής:

$$\mathbf{z}' = (s_x \mathbf{x} + j s_y \mathbf{y}) e^{j\theta} \quad (10-17)$$

όπου $\mathbf{z} = \mathbf{x} + j \mathbf{y}$. Η παραπάνω σχέση οδηγεί άμεσα στην κανονικοποίηση περιστροφής και κατοπτρισμού, η οποία αποτελείται από δύο βήματα. Στο πρώτο βήμα πραγματοποιείται η κανονικοποίηση περιστροφής και στο δεύτερο η κανονικοποίηση κατοπτρισμού. Η διαδικασία ορίζεται παρακάτω για την καμπύλη \mathbf{z} , και μπορεί με όμοιο τρόπο να εφαρμοσθεί και στην \mathbf{z}' . Όπως και στην περίπτωση του σημείου αναφοράς, η κανονικοποίηση περιστροφής βασίζεται στο πρώτο και τελευταίο στοιχείο, a_1 και a_{N-1} , του διανύσματος φάσης $\mathbf{a} = \text{Arg}\mathfrak{Z}(\mathbf{z}) = [a_0 \dots a_{N-1}]$ της \mathbf{z} . Αντίθετα η κανονικοποίηση κατοπτρισμού βασίζεται στις ροπές τρίτου βαθμού. Αναλυτικότερα, τα βήματα κανονικοποίησης είναι τα εξής:

- (α) Η περιστροφή της καμπύλης \mathbf{z} κανονικοποιούνται σύμφωνα με τη μέση τιμή των στοιχείων a_1 και a_{N-1} :

$$r(\mathbf{z}) = \left(\frac{1}{2} (a_1 + a_{N-1}) \right) \bmod \pi \quad (10-18\alpha)$$

$$\mathbf{z}_1 = \mathbf{z} e^{-jr(\mathbf{z})} \quad (10-18\beta)$$

- (β) Ο οριζόντιος και κατακόρυφος κατοπτρισμός κανονικοποιούνται σύμφωνα με τις ακόλουθες ροπές τρίτου βαθμού της καμπύλης \mathbf{z}_1 :

$$v(\mathbf{z}_1) = v_x(\mathbf{z}_1) + j v_y(\mathbf{z}_1) = \text{sgn} m_{1,2}(\mathbf{z}_1) + j \text{sgn} m_{2,1}(\mathbf{z}_1) \quad (10-19\alpha)$$

$$n_r(\mathbf{z}) = \mathbf{z}_2 = v_x(\mathbf{z}_1) \mathbf{x}_1 + j v_y(\mathbf{z}_1) \mathbf{y}_1 \quad (10-19\beta)$$

όπου $\text{sgn}(\cdot)$ υποδηλώνει τη συνάρτηση προσήμου (signum function).

Παρατηρείται ότι ακριβώς όπως στην περίπτωση του σημείου αναφοράς, πραγματοποιείται και εδώ η εκτίμηση μίας πρότυπης περιστροφής $r(\mathbf{z})$ και κατοπτρισμού $v(\mathbf{z}_1)$ και στη συνέχεια εφαρμόζεται ο αντίστροφος μετασχηματισμός.

Σημειώνεται ότι ισοδύναμα θα ήταν επαρκής ο ορισμός $r(\mathbf{z}) = a_1$. Ο ορισμός (10-18α) όμως έχει την επιπρόσθετη ιδιότητα ότι απαλείφει την επίδραση του σημείου αναφοράς που δίνεται στην Εξίσωση (10-12), έτσι ώστε η διαδικασία κανονικοποίησης περιστροφής να είναι αναλλοίωτη στο σημείο αναφοράς, όπως εξηγείται παρακάτω. Όμοια με τη διαδικασία ορθοκανονικοποίησης, μπορεί να αποδειχθεί η μοναδικότητα και η πληρότητα της $n_r(\mathbf{z})$, δηλαδή ότι η $n_r(\mathbf{z})$ είναι αναλλοίωτη στους μετασχηματισμούς περιστροφής και κατοπτρισμού διατηρώντας

παράλληλα όλη την πληροφορία σχήματος εκτός από τους μετασχηματισμούς στους οποίους είναι αναλλοίωτη. Συγκεκριμένα, στην Ενότητα 10.5 αποδεικνύεται η ακόλουθη Πρόταση:

ΠΡΟΤΑΣΗ 10-4. Αν δύο καμπύλες \mathbf{s}, \mathbf{s}' σχετίζονται μέσω ενός ορθογώνιου μετασχηματισμού, δηλαδή $\mathbf{s} = \mathbf{Q} \mathbf{s}'$, όπου \mathbf{Q} είναι ένας ορθογώνιος πίνακας διαστάσεων 2×2 που αναλύεται σε γινόμενο ενός πίνακα περιστροφής και ενός πίνακα κατοπτρισμού σύμφωνα με την (10-16), τότε για τις αναπαραστάσεις μιγαδικών διανυσμάτων z, z' των καμπυλών ισχύουν τα ακόλουθα:

$$r(\mathbf{z}') = (\lambda r(\mathbf{z}) + \theta) \bmod \pi \quad (10-20\alpha)$$

$$r(n_r(\mathbf{z}')) = r(n_r(\mathbf{z})) = 0 \quad (10-20\beta)$$

$$v_x(n_r(\mathbf{z}')) = v_y(n_r(\mathbf{z}')) = v_x(n_r(\mathbf{z})) = v_y(n_r(\mathbf{z})) = 1 \quad (10-20\gamma)$$

$$n_r(\mathbf{z}') = n_r(\mathbf{z}) \quad (10-20\delta)$$

όπου $\lambda = s_x s_y = \pm 1$ και οι ποσότητες θ, s_x, s_y ορίζονται στην (10-16).

Σημειώνεται ότι όπως και η ορθοκανονικοποίηση και η κανονικοποίηση σημείου αναφοράς πραγματοποιούνται βάσει των συνθηκών (10-4) και (10-15β) αντίστοιχα, έτσι και οι μετασχηματισμοί περιστροφής και κατοπτρισμού απαλείφονται με τέτοιο τρόπο ώστε οι κανονικοποιημένες καμπύλες να ικανοποιούν τις συνθήκες (10-20β) και (10-20γ) αντίστοιχα. Επιπρόσθετα, όπως και στην ορθοκανονικοποίηση, το σύνολο των παραμέτρων $\{r(\mathbf{z}), v_x(\mathbf{z}), v_y(\mathbf{z})\}$ μαζί με την κανονικοποιημένη καμπύλη $n_r(\mathbf{z})$ περιέχουν όλη την πληροφορία της αρχικής καμπύλης \mathbf{z} , έτσι ώστε η τελευταία να μπορεί να ανακατασκευασθεί πλήρως με την εφαρμογή του αντίστροφου μετασχηματισμού.

Συμπεραίνεται επίσης στην απόδειξη της Πρότασης 10-4 ότι η παραπάνω διαδικασία κανονικοποίησης περιστροφής / κατοπτρισμού δεν εξαρτάται από το επιλεγμένο σημείο αναφοράς, δηλαδή $n_r(S_m(\mathbf{z})) = S_m(n_r(\mathbf{z}))$ για κάθε $m \in \{0, \dots, N-1\}$, ακριβώς όπως $n_a(S_m(\mathbf{s})) = S_m(n_a(\mathbf{s}))$. Έτσι η κανονικοποίηση περιστροφής και κατοπτρισμού θα μπορούσε να εφαρμοσθεί ακόμη και πριν από την κανονικοποίηση αρχικού σημείου αναφοράς. Σε κάθε περίπτωση όμως ο κατοπτρισμός κανονικοποιείται μετά την περιστροφή.

10.4 ΤΕΛΙΚΑ ΣΤΑΔΙΑ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗΣ

Είναι αξιοσημείωτο ότι συνδυάζοντας τα αποτελέσματα των Προτάσεων 10-2 και 10-4, συμπεραίνεται ότι μία ορθοκανονικοποιημένη καμπύλη η οποία κανονικοποιείται επίσης ως προς περιστροφή και κατοπτρισμό, δηλαδή η $n_r(n_a(\mathbf{z}))$, είναι *αναλλοίωτη σε οποιοδήποτε affine μετασχηματισμό*. Η ανθεκτικότητα της κανονικοποίησης περιστροφής και κατοπτρισμού στο θόρυβο και την παραμόρφωση των καμπυλών επιδεικνύεται στα πειράματα της Ενότητας 11.3. Σημειώνεται επίσης ότι παρόμοιες προσπάθειες για κανονικοποίηση περιστροφής έχουν πραγματοποιηθεί με χρήση υψηλόβαθμων ροπών [126].

Δύο ακόμη βήματα κανονικοποίησης είναι απαραίτητα μετά την κανονικοποίηση περιστροφής και κατοπτρισμού. Πρώτον, υπενθυμίζεται ότι μετά την κανονικοποίηση σημείου αναφοράς παραμένει μία αβεβαιότητα $N/2$ δειγμάτων. Παρατηρείται όμως ότι μετά την κανονικοποίηση περιστροφής το αρχικό σημείο αναφοράς (x_0, y_0) βρίσκεται πολύ κοντά στον άξονα x ($y_0 \equiv 0$) και μακριά από τον άξονα y , σε οποιαδήποτε πλευρά του ($x_0 > 0$ ή $x_0 < 0$). Επομένως η αβεβαιότητα του σημείου αναφοράς αίρεται εφαρμόζοντας μία επιπρόσθετη κυκλική ολίσθηση $N/2$ δειγμάτων αν $x_0 > 0$, αλλιώς η καμπύλη παραμένει ανέπαφη. Μετά τη

διαδικασία αυτή ισχύει πάντα $x_0 < 0$, $y_0 \equiv 0$. Δεύτερον, η φορά διαγραφής της καμπύλης κανονικοποιείται ώστε να είναι δεξιόστροφη. Η διαδικασία αυτή πραγματοποιείται οπωσδήποτε μετά την κανονικοποίηση κατοπτρισμού, γιατί ο κατοπτρισμός της καμπύλης αντιστρέφει και τη φορά διαγραφής.

Στο σημείο αυτό η συνολική διαδικασία κανονικοποίησης είναι πλήρης. Η προκύπτουσα περιγραφή σχήματος μέσω της κανονικοποιημένης καμπύλης είναι αναλλοίωτη στους affine μετασχηματισμούς (μετατόπιση, αλλαγή κλίμακας, κλίση, περιστροφή και κατοπτρισμός) καθώς και στο αρχικό σημείο αναφοράς και τη φορά διαγραφής. Οι δύο τελευταίοι μετασχηματισμοί αφορούν μόνο κλειστές καμπύλες: στην περίπτωση των ανοικτών καμπυλών η παραπάνω ανάλυση εξακολουθεί να ισχύει και μάλιστα η κανονικοποίηση είναι απλούστερη γιατί δεν περιλαμβάνει το σημείο αναφοράς και τη φορά διαγραφής.

10.5 ΑΠΟΔΕΙΞΕΙΣ

Στην παρούσα Ενότητα δίνονται συνολικά οι αποδείξεις των Προτάσεων 10-1 έως 10-4. Οι αποδείξεις παραλήφθηκαν από την παραπάνω ανάπτυξη της διαδικασίας κανονικοποίησης ώστε να είναι το κείμενο πιο ευανάγνωστο.

Απόδειξη Πρότασης 10-1

Ο περιορισμός ότι η \mathbf{s} δεν αναπαριστά ευθύγραμμο τμήμα απαιτείται έτσι ώστε μετά από οποιαδήποτε μετατόπιση, περιστροφή ή μη μηδενική αλλαγή κλίμακας η \mathbf{s} να μην κείται επί του άξονα x ή y . Έτσι εξασφαλίζεται ότι $m_{2,0}(\mathbf{s}) \neq 0$, $m_{0,2}(\mathbf{s}) \neq 0$, οπότε όλες οι ποσότητες που εμπλέκονται στις Εξισώσεις (10-2α-δ) μπορούν να ορισθούν.

Από τις (10-2α) και (10-1) στο βήμα κανονικοποίησης (α) προκύπτει άμεσα ότι $m_{1,0}(\mathbf{s}_1) = m_{0,1}(\mathbf{s}_1) = 0$. Η ιδιότητα αυτή διατηρείται και για την $n_a(\mathbf{s}) = \mathbf{s}_4$, αφού τα υπόλοιπα βήματα κανονικοποίησης (β)-(δ) εμπλέκουν μόνο αλλαγή κλίμακας και περιστροφή, και όχι μετατόπιση. Έτσι $m_{1,0}(n_a(\mathbf{s})) = m_{0,1}(n_a(\mathbf{s})) = 0$. Είναι επίσης εμφανές ότι για την καμπύλη \mathbf{s}_2 του βήματος (β) ισχύει εξ' ορισμού $m_{2,0}(\mathbf{s}_2) = m_{0,2}(\mathbf{s}_2) = 1$. Το ίδιο ισχύει και για την καμπύλη \mathbf{s}_4 του βήματος (δ), δηλαδή για την $n_a(\mathbf{s})$: $m_{2,0}(n_a(\mathbf{s})) = m_{0,2}(n_a(\mathbf{s})) = 1$. Παρατηρείται επίσης ότι

$$m_{1,1}(\mathbf{s}_3) = \mathbf{x}_3 \mathbf{y}_3^T = \frac{1}{2}(\mathbf{x}_2 - \mathbf{y}_2)(\mathbf{x}_2 + \mathbf{y}_2)^T = \frac{1}{2}(m_{2,0}(\mathbf{s}_2) - m_{0,2}(\mathbf{s}_2)) = 0 \quad (10-21)$$

έτσι ώστε $m_{1,1}(n_a(\mathbf{s})) = m_{1,1}(\mathbf{s}_3) / (m_{2,0}(\mathbf{s}_3) m_{0,2}(\mathbf{s}_3))^{1/2} = 0$. Επιβεβαιώνεται λοιπόν ότι η $n_a(\mathbf{s})$ ικανοποιεί όλες τις συνθήκες (10-4).

Αν τώρα η γωνία $\theta_0 = \pi/4$ αντικατασταθεί από μία αυθαίρετη γωνία θ στην (10-2γ), τότε το βήμα κανονικοποίησης (γ) γίνεται

$$\mathbf{s}_3 = \mathbf{R}_\theta \mathbf{s}_2 = \begin{bmatrix} \mathbf{x}_2 \cos \theta - \mathbf{y}_2 \sin \theta \\ \mathbf{x}_2 \sin \theta + \mathbf{y}_2 \cos \theta \end{bmatrix} \quad (10-22)$$

και η (10-21) ξαναγράφεται ως εξής:

$$m_{1,1}(\mathbf{s}_3) = \mathbf{x}_3 \mathbf{y}_3^T = (\sin \theta \cos \theta)(\mathbf{x}_2 \mathbf{x}_2^T - \mathbf{y}_2 \mathbf{y}_2^T) + (\cos^2 \theta - \sin^2 \theta)(\mathbf{x}_2 \mathbf{y}_2^T) \quad (10-23)$$

ή ισοδύναμα:

$$m_{1,1}(\mathbf{s}_3) = (\cos 2\theta)(\mathbf{x}_2 \mathbf{y}_2^T) \quad (10-24)$$

Έτσι, για να είναι η ροπή $m_{1,1}(\mathbf{s}_3)$ πάντα μηδενική ανεξάρτητα από την αρχική καμπύλη, πρέπει να ισχύει $\cos 2\theta = 0$ και επομένως $\theta = k\pi/2 + \pi/4$, $k \in \mathbb{Z}$. \square

Απόδειξη Πρότασης 10-2

Από την (10-3) έχουμε $n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s}) (\mathbf{s} - \boldsymbol{\mu}(\mathbf{s})) = \mathbf{N}(\mathbf{s}) \mathbf{s}_1$ και $n_a(\mathbf{s}') = \mathbf{N}(\mathbf{s}') \mathbf{s}'_1 = \mathbf{N}(\mathbf{s}') \mathbf{A} \mathbf{s}_1$. Αφού οι \mathbf{s} και \mathbf{s}' δεν είναι ευθύγραμμα τμήματα, ισχύει $\det \mathbf{N}(\mathbf{s}) = \sigma_x \sigma_y \tau_x \tau_y / \sqrt{2} \neq 0$ και όμοια $\det \mathbf{N}(\mathbf{s}') \neq 0$. Οι δύο πίνακες ορθοκανονικοποίησης είναι επομένως αντιστρέψιμοι και μπορούμε να ορίσουμε τον πίνακα

$$\mathbf{Q} = \mathbf{N}(\mathbf{s}') \mathbf{A} (\mathbf{N}(\mathbf{s}))^{-1} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \quad (10-25)$$

έτσι ώστε $n_a(\mathbf{s}') = \mathbf{Q} n_a(\mathbf{s})$. Επειδή οι $n_a(\mathbf{s})$ και $n_a(\mathbf{s}')$ είναι ορθοκανονικοποιημένες, οι Εξισώσεις (10-6α,β,γ) γίνονται

$$m_{2,0}(n_a(\mathbf{s}')) = q_{11}^2 + q_{12}^2 = 1 \quad (10-26\alpha)$$

$$m_{0,2}(n_a(\mathbf{s}')) = q_{21}^2 + q_{22}^2 = 1 \quad (10-26\beta)$$

$$m_{1,1}(n_a(\mathbf{s}')) = q_{11} q_{21} + q_{12} q_{22} = 0 \quad (10-26\gamma)$$

που σημαίνει ότι $\mathbf{Q} \mathbf{Q}^T = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_2$, δηλαδή ο \mathbf{Q} είναι ορθογώνιος (με $\det \mathbf{Q} = \pm 1$). Άρα οι $n_a(\mathbf{s})$ και $n_a(\mathbf{s}')$ διαφέρουν μόνο κατά ένα μετασχηματισμό περιστροφής (αν $\det \mathbf{Q} = 1$), και πιθανόν κατά ένα επιπρόσθετο μετασχηματισμό κατοπτρισμού (αν $\det \mathbf{Q} = -1$).

Έστω τώρα ότι υπάρχει μία εναλλακτική μέθοδος κανονικοποίησης η οποία δίνει για την καμπύλη \mathbf{s} πίνακα κανονικοποίησης $\tilde{\mathbf{N}}(\mathbf{s})$ και κανονικοποιημένη καμπύλη $\tilde{n}_a(\mathbf{s}) = \tilde{\mathbf{N}}(\mathbf{s}) \mathbf{s}_1$. Στην περίπτωση αυτή μπορούμε όμοια να ορίσουμε $\tilde{\mathbf{Q}} = \tilde{\mathbf{N}}(\mathbf{s}) [\mathbf{N}(\mathbf{s})]^{-1}$ έτσι ώστε $\tilde{n}_a(\mathbf{s}) = \tilde{\mathbf{Q}} n_a(\mathbf{s})$. Αφού και πάλι οι $n_a(\mathbf{s})$ και $\tilde{n}_a(\mathbf{s})$ είναι κανονικοποιημένες, συμπεραίνουμε με όμοιο τρόπο ότι ο πίνακας $\tilde{\mathbf{Q}}$ είναι ορθογώνιος. Επομένως όλες οι δυνατές μέθοδοι κανονικοποίησης που βασίζονται στις ιδιότητες (10-4) καταλήγουν σε μία κανονικοποιημένη καμπύλη που σχετίζεται με την προτεινόμενη $n_a(\mathbf{s})$ μέσω μίας απλής περιστροφής (ή και κατοπτρισμού). \square

Απόδειξη Πρότασης 10-3

Αν $\mathbf{z}' = S_m(\mathbf{z})$, τότε ισχύει η (10-12) και μπορούμε να αντικαταστήσουμε τις ποσότητες $a'_1 = (a_1 + 2\pi m/N) \bmod 2\pi$ και $a'_{N-1} = (a_{N-1} - 2\pi m/N) \bmod 2\pi$ στη σχέση

$$p(\mathbf{z}') = \left\lfloor \frac{N}{4\pi} (a'_1 - a'_{N-1}) \right\rfloor \bmod N/2 \quad (10-27)$$

που προκύπτει από τον ορισμό (10-13) για την \mathbf{s}' . Επιπλέον, αν το N είναι άρτιο, ο αριθμός $N/2$ είναι ακέραιος και οι τελεστές $\bmod N/2$ και $\lfloor \cdot \rfloor$ (ακέραιο μέρος) μπορούν να αντιμετωπιστούν, δίνοντας

$$p(\mathbf{z}') = \left(\left\lfloor \frac{N}{4\pi} (a_1 - a_{N-1}) \right\rfloor \bmod N/2 + m \right) \bmod N/2 \quad (10-28)$$

η οποία οδηγεί στην (10-15α). Λαμβάνοντας τώρα υπόψη ότι η $n_p(\mathbf{z})$ προκύπτει από την \mathbf{z} μέσω μιας κυκλικής ολίσθησης κατά $N - p(\mathbf{z})$ δείγματα, μπορούμε να χρησιμοποιήσουμε την (10-15α) για τις καμπύλες \mathbf{z} και $n_p(\mathbf{z})$:

$$p(n_p(\mathbf{z})) = p(S_{N-p(\mathbf{z})}(\mathbf{z})) = [p(\mathbf{z}) + (N - p(\mathbf{z}))] \bmod N/2 = 0 \quad (10-29)$$

και με ένα παρόμοιο επιχείρημα για τις καμπύλες \mathbf{z}' και $n_p(\mathbf{z}')$ προκύπτει η (10-15β). Στη συνέχεια, συσχετίζοντας τις καμπύλες $n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z})$ και $n_p(\mathbf{z}') = S_{-p(\mathbf{z}')}(\mathbf{z}') = S_{m-p(\mathbf{z}')}(\mathbf{z})$, λαμβάνουμε $n_p(\mathbf{z}') = S_{p(\mathbf{z})+m-p(\mathbf{z}')} (n_p(\mathbf{z}))$. Αλλά τότε μπορούμε να αντικαταστήσουμε την ποσότητα

$$p(\mathbf{z}') = (p(\mathbf{z}) + m) \bmod N/2 = \begin{cases} p(\mathbf{z}) + m, & 0 \leq p(\mathbf{z}) + m < N/2 \\ p(\mathbf{z}) + m - N/2, & N/2 \leq p(\mathbf{z}) + m < N \end{cases} \quad (10-30)$$

στην προηγούμενη σχέση, η οποία τότε ανάγεται άμεσα στην (10-15γ). Τέλος, αν το N είναι περιττό, η αντιμετάθεση των τελεστών $\bmod N/2$ και $\lfloor \cdot \rfloor$ στην (10-27) μπορεί να οδηγήσει σε σφάλμα κατά ± 1 δείγμα στην εκτίμηση του $p(\mathbf{z}')$ από το $p(\mathbf{z})$, πιθανώς προκαλώντας μία επιπρόσθετη ολίσθηση κατά ± 1 δείγμα ανάμεσα στις $n_p(\mathbf{z})$ και $n_p(\mathbf{z}')$. \square

Απόδειξη Πρότασης 10-4

Αν $\mathbf{s} = \mathbf{Q}\mathbf{s}'$ και ο πίνακας \mathbf{Q} αναλύεται σύμφωνα με την (10-16), τότε η (10-17) ισχύει και επομένως μπορούμε να εκφράσουμε τη σχέση μεταξύ των διακριτών μετασχηματισμών Fourier $\mathbf{u} = \mathcal{Z}(\mathbf{z})$ και $\mathbf{u}' = \mathcal{Z}(\mathbf{z}')$ ως εξής:

$$u'_k = \sum_{i=0}^{N-1} z'_i w^{-ki} = s_x e^{j\theta} \sum_{i=0}^{N-1} (x_i + jy_i) w^{-ki} = \begin{cases} s_x e^{j\theta} u_k, & \lambda = 1 \\ s_x e^{j\theta} u_{N-k}^*, & \lambda = -1 \end{cases} \quad (10-31)$$

για $k = 0, \dots, N-1$, όπου $\lambda = s_x s_y = \pm 1$. Αν τώρα υπολογίσουμε τα διανύσματα φάσης $\mathbf{a} = \text{Arg } \mathbf{u}$ και $\mathbf{a}' = \text{Arg } \mathbf{u}'$, η σχέση (10-31) δίνει για $k = 1$ και $k = N-1$:

$$a'_1 = \begin{cases} a_1 + \theta + \text{Arg } s_x, & \lambda = 1 \\ -a_{N-1} + \theta + \text{Arg } s_x, & \lambda = -1 \end{cases} \quad (10-32\alpha)$$

$$a'_{N-1} = \begin{cases} a_{N-1} + \theta + \text{Arg } s_x, & \lambda = 1 \\ -a_1 + \theta + \text{Arg } s_x, & \lambda = -1 \end{cases} \quad (10-32\beta)$$

Προσθέτοντας κατά μέλη τις Εξισώσεις (10-32α,β) λαμβάνουμε $(a'_1 + a'_{N-1}) \bmod 2\pi = (\lambda(a_1 + a_{N-1}) + 2\theta) \bmod 2\pi$, αφού $s_x = \pm 1$ και επομένως $(2\text{Arg } s_x) \bmod 2\pi = 0$. Από τον ορισμό (10-18α) του $r(\mathbf{z})$, η (10-20α) προκύπτει άμεσα.

Ο υπολογισμός τώρα της \mathbf{z}'_1 από την (10-18β) δίνει $\mathbf{z}'_1 = \mathbf{z}' e^{-jr(\mathbf{z}')} = s_x (\mathbf{x} + jy) e^{j(\theta - r(\mathbf{z}'))}$. Αν λοιπόν ορίσουμε την ποσότητα

$$l(\mathbf{z}) = \begin{cases} 1, & \lambda r(\mathbf{z}) + \theta \in [0, \pi) \\ -1, & \lambda r(\mathbf{z}) + \theta \in [\pi, 2\pi) \end{cases} \quad (10-33)$$

τότε από την (10-20α) προκύπτει ότι

$$\mathbf{z}'_1 = s_x (\mathbf{x} + jy) e^{j(\theta - ((\lambda r(\mathbf{z}) + \theta) \bmod \pi))} = s_x l(\mathbf{z}) (\mathbf{x} + jy) e^{-j\lambda r(\mathbf{z})} = \begin{cases} s_x l(\mathbf{z}) \mathbf{z}_1, & \lambda = 1 \\ s_x l(\mathbf{z}) \mathbf{z}_1^*, & \lambda = -1 \end{cases} \quad (10-34)$$

έτσι ώστε $\mathbf{x}'_1 = s_x l(\mathbf{z}) \mathbf{x}_1$ και $\mathbf{y}'_1 = \lambda s_x l(\mathbf{z}) \mathbf{y}_1 = s_y l(\mathbf{z}) \mathbf{y}_1$. Συμπεραίνεται λοιπόν ότι τα διανύσματα \mathbf{x}'_1 και \mathbf{y}'_1 διαφέρουν από τα \mathbf{x}_1 και \mathbf{y}_1 μόνο ως προς το πρόσημο και ο υπολογισμός των ρο-

πών τρίτου βαθμού στην (10-19α) δίνει $v_x(\mathbf{z}'_1) = s_x l(\mathbf{z}) v_x(\mathbf{z}_1)$ και $v_y(\mathbf{z}'_1) = s_y l(\mathbf{z}) v_y(\mathbf{z}_1)$. Τότε, η κανονικοποίηση (10-19β) δίνει $\mathbf{x}'_2 = (s_x l(\mathbf{z}))^2 v_x(\mathbf{z}_1) \mathbf{x}_1 = \mathbf{x}_2$ και όμοια $\mathbf{y}'_2 = \mathbf{y}_2$, αποδεικνύοντας την (10-20δ). Τέλος, οι συνθήκες (10-20β) και (10-20γ) μπορούν να επιβεβαιωθούν άμεσα χρησιμοποιώντας τις (10-19α,β) και (10-20α). \square

Αναζήτηση με Βάση το Σχήμα - Αποτελέσματα

Η αναλλοίωτη αναπαράσταση που προκύπτει από την προτεινόμενη κανονικοποίηση περιγραμμάτων μπορεί να είναι εξαιρετικά χρήσιμη σε εφαρμογές όπως η σύγκριση, ταίριασμα, ταξινόμηση, αναγνώριση σχημάτων καθώς και η αναζήτηση και ανάκληση αντικειμένων με βάση το σχήμα. Στο παρόν Κεφάλαιο παρουσιάζονται αρχικά και συγκρίνονται με την προτεινόμενη μέθοδο δύο εναλλακτικές δημοφιλείς μέθοδοι αναλλοίωτης αναπαράστασης και ταιριάσματος καμπυλών και συγκεκριμένα οι κανονικοποιημένοι περιγραφείς Fourier και το ταίριασμα με χρήση ρολών. Στη συνέχεια προτείνεται μία μέθοδος αναζήτησης και ανάκλησης αντικειμένων με βάση το σχήμα τους, η οποία βασίζεται σε τρία απλά μέτρα ομοιότητας περιγραμμάτων, ενώ εξετάζεται και το συναφές πρόβλημα της ταξινόμησης σχημάτων σε προκαθορισμένες κατηγορίες, το οποίο αντιμετωπίζεται με χρήση νευρωνικού δικτύου. Τέλος δίνονται παραδείγματα, πειραματικά αποτελέσματα και αλγοριθμικές οδηγίες τόσο για την κανονικοποίηση περιγραμμάτων όσο και για την ταξινόμηση, τα χρησιμοποιούμενα μέτρα ομοιότητας και την αναζήτηση με βάση το σχήμα. Τα περιγράμματα που χρησιμοποιούνται στα πειράματα προέρχονται από μία συλλογή ακίνητων εικόνων αντικειμένων με αυτόματη κατάτμηση, η οποία όμως σε κάποιες περιπτώσεις απαιτεί ανθρώπινη παρέμβαση.

11.1 ΕΝΑΛΛΑΚΤΙΚΕΣ ΜΕΘΟΔΟΙ

Για την αναζήτηση και ανάκληση εικόνων από μία βάση δεδομένων με βάση το σχήμα των αντικειμένων, παρουσιάζονται δύο προβλήματα: (α) η σύγκριση ή το ταίριασμα περιγραμμάτων πρέπει να είναι μια διαδικασία αναλλοίωτη στους affine μετασχηματισμούς, και (β) πρέπει να παρέχεται ένας τρόπος γρήγορης αρχικής ταξινόμησης αφού συνήθως δεν είναι εφικτή η σύγκριση ενός δεδομένου περιγράμματος εισόδου με κάθε περίγραμμα που είναι διαθέσιμο στη βάση. Για το πρώτο πρόβλημα μπορεί να χρησιμοποιηθεί η προτεινόμενη μέθοδος κανονικοποίησης περιγραμμάτων, αλλά παρουσιάζονται παρακάτω για λόγους πληρότητας δύο εναλλακτικές και δημοφιλείς μέθοδοι αναπαράστασης / σύγκρισης. Τα ιδιαίτερα χαρακτηριστικά και οι περιορισμοί των δύο μεθόδων αναδεικνύουν μάλιστα τα πλεονεκτήματα της προτεινόμενης μεθόδου. Για το δεύτερο πρόβλημα προτείνεται μία τεχνική προκαταρκτικής ταξινόμησης περιγραμμάτων σε ένα προκαθορισμένο σύνολο κατηγοριών η οποία μπορεί να εφαρμοσθεί πριν από το ταίριασμα. Η ταξινόμηση πραγματοποιείται με χρήση νευρωνικού δικτύου και περιγράφεται στη συνέχεια.

Δύο από τις δημοφιλέστερες μεθόδους αναπαράστασης ή ταιριάσματος καμπυλών που είναι αναλλοίωτες στους affine μετασχηματισμούς είναι οι κανονικοποιημένοι περιγραφείς Fourier (normalized Fourier descriptors, NFD) [28] και το ταίριασμα με χρήση ρολών (moments) [56]. Το τελευταίο δίνει ικανοποιητικά αποτελέσματα ταιριάσματος και σύγκρισης περιγραμμάτων αλλά απαιτεί εκ των προτέρων γνώση των δύο περιγραμμάτων που προορίζονται για σύγκρι-

ση και έχει μεγάλο υπολογιστικό κόστος, αφού πρόκειται για μία διαδικασία βελτιστοποίησης. Από την άλλη μεριά οι Fourier descriptors έχουν μικρό υπολογιστικό κόστος αλλά εισάγουν σημαντική απώλεια πληροφορίας, επομένως δε μπορούν να αποτελέσουν γενικευμένη περιγραφή 2-Δ καμπυλών. Οι παρατηρήσεις αυτές γίνονται εμφανείς από την ακόλουθη περιγραφή των μεθόδων.

Κανονικοποιημένοι Περιγραφείς Fourier

Έστω ότι το περίγραμμα ενός αντικειμένου περιγράφεται από ένα σύνολο N σημείων· τα σημεία αυτά μπορούν να είναι είτε τα σημεία αρχικής δειγματοληψίας είτε τα κομβικά σημεία που προκύπτουν από τη μοντελοποίηση με B-Splines, όπως περιγράφεται στην Ενότητα 9.4. Για κάθε σημείο $\mathbf{s}_i = [x_i \ y_i]^T$, $i = 0, \dots, N-1$, της καμπύλης, χρησιμοποιείται η μιγαδική αναπαράσταση $z_i = x_i + jy_i$. Τότε σύμφωνα με τη (10-10) ο μετασχηματισμός Fourier της ακολουθίας z_i , $i = 0, \dots, N-1$, δίνεται από τη σχέση

$$u_k = \sum_{i=0}^{N-1} z_i \exp\left(-j \frac{2\pi i k}{N}\right), \quad k = 0, \dots, N-1 \quad (11-1)$$

Έστω επίσης ότι η μιγαδική ακολουθία z'_i , $i = 0, \dots, N-1$, αντιστοιχεί σε μια καμπύλη που προκύπτει από τη z_i με εφαρμογή μετατόπισης, αλλαγής κλίμακας, περιστροφής και κυκλικής ολίσθησης, δηλαδή

$$z'_i = a z_{(i+m) \bmod N} e^{j\theta} + t, \quad i = 0, \dots, N-1 \quad (11-2)$$

όπου $a \in \mathbb{R}$ είναι ο συντελεστής αλλαγής κλίμακας, t είναι ο μιγαδικός που ορίζει τη μετατόπιση, θ είναι η γωνία περιστροφής σε ακτίνια και m είναι το πλήθος στοιχείων ολίσθησης. Υπενθυμίζεται ότι ο παραπάνω μετασχηματισμός δεν περιλαμβάνει κλίση (skew) όπως ένας affine μετασχηματισμός. Τότε ο μετασχηματισμός Fourier της z'_i δίνεται από

$$u'_k = a u_k \exp\left(j \frac{\theta - 2\pi k m}{N}\right) + t \delta(k), \quad k = 0, \dots, N-1 \quad (11-3)$$

όπου $\delta(k)$ είναι η συνάρτηση Dirac. Με βάση την προηγούμενη σχέση, οι κανονικοποιημένοι περιγραφείς Fourier (NFD) ορίζονται ως το σύνολο των $N-2$ πραγματικών αριθμών

$$v_k = \frac{|u_k|}{|u_1|}, \quad k = 2, \dots, N-1 \quad (11-4)$$

Από τις (11-3) και (11-4) προκύπτει ότι $v'_k = v_k$ για $k = 2, \dots, N-1$.

Με άλλα λόγια οι κανονικοποιημένοι περιγραφείς Fourier είναι αναλλοίωτοι στους μετασχηματισμούς μετατόπισης, αλλαγής κλίμακας, περιστροφής και κυκλικής ολίσθησης (δηλ. αλλαγής σημείου αναφοράς). Επιπλέον αποτελούν ανεξάρτητη ποσοτική αναπαράσταση κάθε περιγράμματος, και μπορούν να χρησιμοποιηθούν σε συνδυασμό με οποιαδήποτε μέθοδο ταξινόμησης. Όμως η απαλοιφή της φάσης του μετασχηματισμού Fourier στην (11-4) προκαλεί σημαντική απώλεια πληροφορίας για το σχήμα του περιγράμματος, αφού είναι γνωστό ότι το μεγαλύτερο μέρος της πληροφορίας ενός σήματος περιέχεται στη φάση του μετασχηματισμού Fourier και όχι στο μέτρο του. Έτσι οι κανονικοποιημένοι περιγραφείς Fourier οδηγούν πολλές φορές σε λανθασμένο ταίριασμα ανόμοιων αντικειμένων.

Ταίριασμα με Χρήση Ροπών

Μία εναλλακτική μέθοδος ταίριασματος, επίσης αναλλοίωτη στους affine μετασχηματισμούς, η οποία όμως δεν πάσχει από απώλεια πληροφορίας, είναι το ταίριασμα με χρήση ροπών [56]. Στην περίπτωση αυτή, ένα περίγραμμα αναπαρίσταται από μία συνεχή καμπύλη B-Spline $\mathbf{r}(s) = (x(s), y(s))$ όπου s είναι το μήκος τόξου. Τότε ορίζεται η ροπή βαθμού (p, q) με βάρη, ως εξής:

$$m_{p,q}(\mathbf{r}) = \int_{s=0}^S x^p(s) y^q(s) w(x(s), y(s)) ds \quad (11-5)$$

όπου S είναι το συνολικό μήκος της καμπύλης και $w(x, y)$ είναι η συνάρτηση πυρήνα (kernel) της ροπής. Αν τώρα δύο δεδομένες καμπύλες $\mathbf{r}(s)$, $\mathbf{r}'(s)$ σχετίζονται μέσω ενός affine μετασχηματισμού, δηλαδή $\mathbf{r}'(s) = \mathbf{A}\mathbf{r}(s) + \mathbf{t}$, όπως στην Εξίσωση (10-5), τότε αποδεικνύεται ότι με κατάλληλη επιλογή των συναρτήσεων πυρήνα είναι δυνατή η πλήρης εκτίμηση των πινάκων \mathbf{A} , \mathbf{t} του affine μετασχηματισμού από τις ροπές μέχρι δευτέρου βαθμού [56]. Επιπλέον, αν οι δύο καμπύλες αντιστοιχούν σε δύο παρόμοια περιγράμματα, η διαδικασία αυτή επιτυγχάνει την «ευθυγράμμιση» ή ταίριασμα των δύο σχημάτων.

Όμως η μέθοδος των ροπών απαιτεί την εκ των προτέρων γνώση των δύο καμπυλών που προορίζονται για ταίριασμα. Επομένως δεν μπορεί να χρησιμοποιηθεί για γενικευμένη αναπαράσταση περιγραμμάτων, ούτε για ταξινόμηση ή αναγνώριση π.χ. με νευρωνικό δίκτυο, αφού στην περίπτωση αυτή είναι γνωστό ένα μόνο περίγραμμα κάθε φορά. Επίσης το υπολογιστικό κόστος της μεθόδου είναι ιδιαίτερα αυξημένο. Αντίθετα η προτεινόμενη μέθοδος κανονικοποίησης παρέχει μία αναπαράσταση που είναι επίσης αναλλοίωτη στους affine μετασχηματισμούς αλλά δεν βασίζεται σε ταίριασμα οπότε μπορεί να χρησιμοποιηθεί σε συνδυασμό με οποιοδήποτε μηχανισμό ταξινόμησης. Επιπλέον το υπολογιστικό κόστος της είναι αμελητέο και δεν πάσχει από απώλεια πληροφορίας όπως οι κανονικοποιημένοι περιγραφείς Fourier.

11.2 ΤΑΞΙΝΟΜΗΣΗ ΚΑΙ ΑΝΑΖΗΤΗΣΗ

Αναζήτηση και Ανάκληση με Βάση το Σχήμα

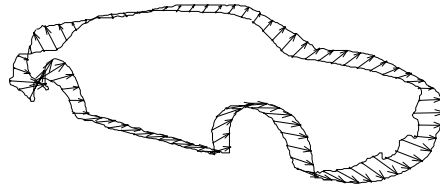
Έστω ότι ένα σύνολο εικόνων ή βίντεο είναι διαθέσιμο σε μία βάση δεδομένων. Τότε η αναζήτηση και ανάκληση με βάση το σχήμα είναι δυνατή αν έχει προηγηθεί ανάλυση των εικόνων σε αντικείμενα, ανίχνευση των περιγραμμάτων των αντικειμένων και κανονικοποίηση των περιγραμμάτων ώστε να είναι αναλλοίωτα στους affine μετασχηματισμούς. Η αναζήτηση μπορεί να πραγματοποιηθεί είτε μέσω παραδείγματος (query by example) είτε μέσω σχεδιαγράμματος / σκίτσου (query by sketch). Στην πρώτη περίπτωση ο χρήστης υποβάλλει στο σύστημα αναζήτησης μία εικόνα ενός αντικειμένου. Η εικόνα αναλύεται για την εξαγωγή του περιγράμματος του αντικειμένου· στη συνέχεια το περίγραμμα κανονικοποιείται και συγκρίνεται με τα διαθέσιμα για την ανάκληση των εικόνων που μοιάζουν περισσότερο με το παράδειγμα. Στη δεύτερη περίπτωση ο χρήστης σχεδιάζει το περίγραμμα ενός αντικειμένου, το οποίο αφού κανονικοποιηθεί χρησιμοποιείται απευθείας για σύγκριση.

Σε κάθε περίπτωση η κανονικοποίηση εξασφαλίζει ότι σχήματα με παρόμοιο οπτικό περιεχόμενο «ευθυγραμμίζονται» όπως απεικονίζεται στο Σχήμα 11-1. Στο Σχήμα αυτό φαίνεται η αντιστοιχία των κομβικών σημείων των κανονικοποιημένων περιγραμμάτων δύο αυτοκινή-

των. Είναι εμφανές ότι (α) τα κομβικά σημεία είναι ομοιόμορφα κατανομημένα ως προς το μήκος τόξου και στα δύο περιγράμματα, και (β) τα περιγράμματα είναι έτσι ευθυγραμμισμένα ώστε η συνολική απόσταση των κομβικών σημείων να ελαχιστοποιείται. Έτσι η σύγκριση ή το ταίριασμα των περιγραμμάτων μπορεί να πραγματοποιηθεί με χρήση ακόμη και ενός απλού μέτρου απόστασης όπως είναι η Ευκλείδεια απόσταση. Συγκεκριμένα, αν $\mathbf{p}_i, \mathbf{p}'_i, i = 0, \dots, N-1$, είναι τα κομβικά σημεία δύο κανονικοποιημένων καμπυλών, τότε η απόσταση

$$d_s^2 = \sum_{i=0}^{N-1} \|\mathbf{p}_i - \mathbf{p}'_i\|^2 \quad (11-6)$$

μπορεί να χρησιμοποιηθεί ως μέτρο ομοιότητας των καμπυλών. Οι εικόνες λοιπόν που ανακαλούνται από το σύστημα ως αποτέλεσμα της αναζήτησης είναι εκείνες που περιέχουν αντικείμενα των οποίων το περίγραμμα έχει τη μικρότερη απόσταση από το περίγραμμα του παραδείγματος. Οι εικόνες επιστρέφονται κατά φθίνουσα σειρά ομοιότητας, δηλαδή κατά αύξουσα σειρά απόστασης. Στα πειράματα της Ενότητας 11.4 εκτός από το παραπάνω μέτρο απόστασης χρησιμοποιείται για ταίριασμα περιγραμμάτων η αντίστοιχη Ευκλείδεια απόσταση των κανονικοποιημένων Fourier descriptors. Επίσης εξετάζονται για λόγους σύγκρισης οι τροποποιημένοι Fourier descriptors (modified Fourier descriptors, MFD) που προτείνονται στην εργασία [108].



Σχήμα 11-1. Ταίριασμα των κομβικών σημείων των κανονικοποιημένων περιγραμμάτων δύο αυτοκινήτων.

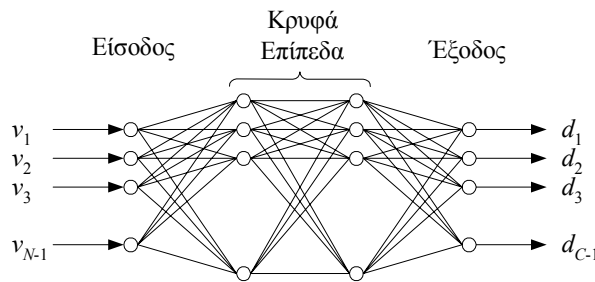
Ταξινόμηση Σχημάτων

Ένα πρόβλημα συναφές με την αναζήτηση και ανάκληση εικόνων είναι εκείνο της ταξινόμησης αντικειμένων με βάση το σχήμα τους. Στην περίπτωση αυτή, ομαδοποιώντας ένα σύνολο πρότυπων περιγραμμάτων ορίζονται κατηγορίες ή κλάσεις αντικειμένων (π.χ. αεροσκάφη, αυτοκίνητα κλπ.) οι οποίες μπορούν να οργανωθούν σε μία βάση δεδομένων. Επομένως το πρόβλημα της ταξινόμησης ενός δεδομένου περιγράμματος σε μία από τις προκαθορισμένες κατηγορίες ανάγεται στο εντοπισμό του πρότυπου περιγράμματος που ταιριάζει καλύτερα με το περίγραμμα εισόδου.

Όμως, παρά το γεγονός ότι η αναπαράσταση των περιγραμμάτων είναι αναλλοίωτη στους affine μετασχηματισμούς, είναι απαραίτητο για αξιόπιστη ταξινόμηση να υπάρχει για κάθε κατηγορία διαθέσιμος μεγάλος αριθμός προτύπων περιγραμμάτων που αντιστοιχούν σε διαφορετικές παραλλαγές των αντικειμένων ή σε διαφορετικά επίπεδα λεπτομέρειας. Συνεπώς σε ένα πρακτικό σύστημα απαιτείται συνολικά πολύ μεγάλος αριθμός προτύπων περιγραμμάτων, καθιστώντας το απευθείας ταίριασμα περιγραμμάτων εξαιρετικά χρονοβόρο. Το ίδιο ισχύει και στην περίπτωση αναζήτησης εικόνων με βάση το σχήμα.

Γι' αυτό το λόγο προτείνεται μία προσέγγιση αρχικής ταξινόμησης με χρήση *νευρωνικού δικτύου* (neural network, NN) [57]. Έτσι η διαδικασία αναζήτησης περιορίζεται σε ένα μικρό υποσύνολο των διαθέσιμων κατηγοριών αντικειμένων. Συγκεκριμένα, η αναπαράσταση των περιγραμμάτων, δηλαδή είτε τα κανονικοποιημένα περιγράμματα (κομβικά σημεία) είτε οι

κανονικοποιημένοι Fourier descriptors, χρησιμοποιούνται ως είσοδος σε ένα feedforward neural network, ενώ κάθε κατηγορία αντικειμένων αντιστοιχίζεται σε μία έξοδο του δικτύου. Το νευρωνικό δίκτυο πραγματοποιεί αντιστοίχιση του διανύσματος εισόδου $\mathbf{v} = [v_0 \dots v_{N-1}]$ σε ένα επιθυμητό διάνυσμα εξόδου $\mathbf{d} = [d_0 \dots d_{C-1}]$. Η χρησιμοποιούμενη αρχιτεκτονική δικτύου απεικονίζεται στο Σχήμα 11-2. Περιλαμβάνει δύο κρυφά επίπεδα (hidden layers), με N νευρώνες (neurons) εισόδου, N_1 και N_2 νευρώνες στο πρώτο και δεύτερο κρυφό επίπεδο, αντίστοιχα, και C νευρώνες εξόδου.



Σχήμα 11-2. Αρχιτεκτονική νευρωνικού δικτύου για ταξινόμηση περιγραμμάτων.

Οι νευρώνες διαδοχικών επιπέδων διασυνδέονται μέσω βαρών, έτσι ώστε η είσοδος κάθε νευρώνα s να είναι $n_s = \sum_i a_i w_i$, όπου a_i είναι η έξοδος του i -οστού νευρώνα του προηγούμενου επιπέδου, w_i είναι το βάρος διασύνδεσης του νευρώνα αυτού με το νευρώνα s , και η άθροιση περιλαμβάνει όλους τους νευρώνες του προηγούμενου επιπέδου. Η είσοδος n_s στη συνέχεια μετασχηματίζεται με χρήση της σιγμοειδούς συνάρτησης ενεργοποίησης (activation function)

$$o_s = f(n_s) = \frac{1}{1 + e^{-\lambda n_s}} \quad (11-7)$$

όπου o_s είναι η έξοδος του νευρώνα s και λ είναι μία παράμετρος κέρδους (gain) [57]. Στο στάδιο εκπαίδευσης του δικτύου, η αναπαράσταση $\mathbf{v}^{(p)}$, $p = 0, \dots, M-1$, ενός συνόλου από M πρότυπα περιγράμματα δίνεται ως είσοδος στο δίκτυο· τα επιθυμητά διανύσματα εξόδου $\mathbf{d}^{(p)}$, $p = 0, \dots, M-1$, καθορίζονται θέτοντας το στοιχείο του $\mathbf{d}^{(p)}$ που αντιστοιχεί στη σωστή κατηγορία του $\mathbf{v}^{(p)}$ ίσο με ένα και τα υπόλοιπα στοιχεία ίσα με μηδέν. Για την εκπαίδευση χρησιμοποιείται ο αλγόριθμος Levenberg-Marquardt, ο οποίος ελαχιστοποιεί το τετραγωνικό σφάλμα

$$E^2 = \sum_{p=0}^{M-1} \|\mathbf{d}^{(p)} - \mathbf{o}^{(p)}\|^2 = \sum_{p=0}^{M-1} \sum_{i=0}^{C-1} |d_i^{(p)} - o_i^{(p)}|^2 \quad (11-8)$$

ανάμεσα στο επιθυμητό και πραγματικό διάνυσμα εξόδου, $\mathbf{d}^{(p)}$ και $\mathbf{o}^{(p)}$, αντίστοιχα. Η διαδικασία ελαχιστοποίησης υλοποιείται με την ανανέωση βαρών που συνδέουν νευρώνες διαδοχικών επιπέδων και τον εκ νέου υπολογισμό του διανύσματος εξόδου και του τετραγωνικού σφάλματος, με αναδρομικό τρόπο.

Στο στάδιο ταξινόμησης, η αναπαράσταση $\mathbf{v} = [v_0 \dots v_{N-1}]$ ενός περιγράμματος εισάγεται στο νευρωνικό δίκτυο ως διάνυσμα εισόδου. Αφού σε κάθε κατηγορία αντικειμένων αντιστοιχεί μία έξοδος του δικτύου, τυπικά το περίγραμμα εισόδου ταξινομείται στην κατηγορία εκείνη που αντιστοιχεί στη μέγιστη έξοδο του δικτύου. Στην πράξη όμως, επειδή οι υπόλοιπες έ-

ξοδοι δεν είναι όλες μηδενικές (όπως στην περίπτωση των επιθυμητών εξόδων κατά τη φάση της εκπαίδευσης), σε κάποιες περιπτώσεις μπορεί να συμβεί λανθασμένη ταξινόμηση αν οι έξοδοι του νευρωνικού δικτύου έχουν παρόμοιες τιμές.

Γι' αυτό το λόγο επιλέγεται ένας μικρός αριθμός κατηγοριών, έστω R , για κάθε περίγραμμα εισόδου, οι οποίες αντιστοιχούν στις εξόδους με τις R μεγαλύτερες τιμές. Ο αριθμός αυτός αντιπροσωπεύει ένα μικρό ποσοστό του συνολικού αριθμού κατηγοριών, C . Αυτό το σύνολο κατηγοριών χρησιμοποιείται στη συνέχεια για ταίριασμα με το περίγραμμα εισόδου, ώστε να εντοπισθεί η κατηγορία εκείνη που ταιριάζει καλύτερα. Στο στάδιο αυτό πραγματοποιείται σύγκριση μεταξύ του περιγράμματος εισόδου και όλων των προτύπων περιγραμμάτων που ανήκουν στις R επιλεγμένες κατηγορίες. Εναλλακτικά, αν όλες οι κατηγορίες περιέχουν πολύ μεγάλο αριθμό περιγραμμάτων, η σύγκριση μπορεί να περιορισθεί σε ένα μικρότερο υποσύνολο αντιπροσωπευτικών προτύπων από κάθε κατηγορία. Με την παραπάνω τεχνική επιτυγχάνεται ταχεία επιλογή πιθανών κατηγοριών στο πρώτο στάδιο με το νευρωνικό δίκτυο και ταυτόχρονα αξιόπιστη ταξινόμηση στο δεύτερο στάδιο με το ταίριασμα περιγραμμάτων.

11.3 ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗΣ

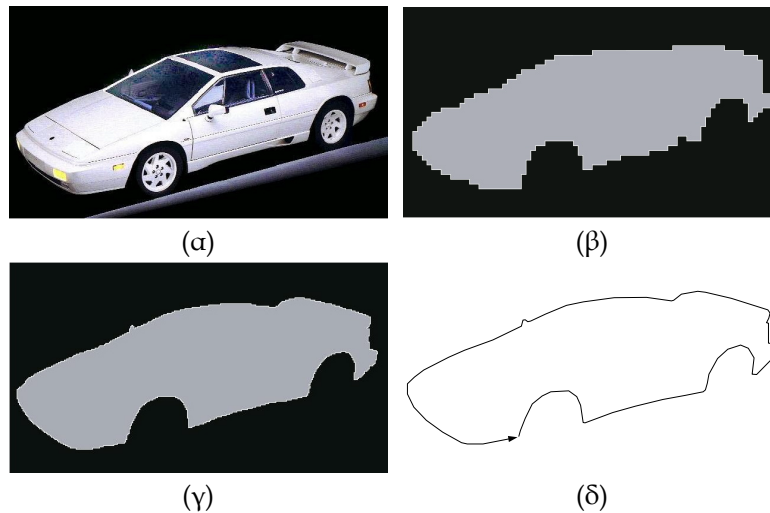
Στην Ενότητα αυτή εξετάζεται η επίδοση της προτεινόμενης μεθόδου κανονικοποίησης καμπυλών και περιλαμβάνονται κατάλληλες αλγοριθμικές οδηγίες, ενώ στην επόμενη Ενότητα παρουσιάζονται πειραματικά αποτελέσματα που αφορούν στην εφαρμογή της μεθόδου στην ταξινόμηση, αναζήτηση και ανάκληση με βάση το σχήμα. Αρχικά δίνονται ενδεικτικά αποτελέσματα κατάτμησης με βάση το χρώμα για την εξαγωγή περιγραμμάτων από εικόνες. Στη συνέχεια παρουσιάζονται αποτελέσματα για όλα τα βήματα κανονικοποίησης, δηλαδή μετατόπιση, αλλαγή κλίμακας, κλίση, σημείο αναφοράς, περιστροφή, και κατοπτρισμό. Η αποτελεσματικότητα της μεθόδου εξετάζεται εφαρμόζοντας affine μετασχηματισμούς (α) στο ίδιο αντικείμενο με ίδια δειγματοληψία, (β) στο ίδιο αντικείμενο με διαφορετική, μη ομοιόμορφη δειγματοληψία, (γ) σε παρόμοια αντικείμενα και (δ) σε ανόμοια αντικείμενα. Σε κάθε περίπτωση χρησιμοποιείται αναπαράσταση με B-Splines για την αντιμετώπιση του θορύβου και των προβλημάτων δειγματοληψίας.

Εξαγωγή Περιγραμμάτων

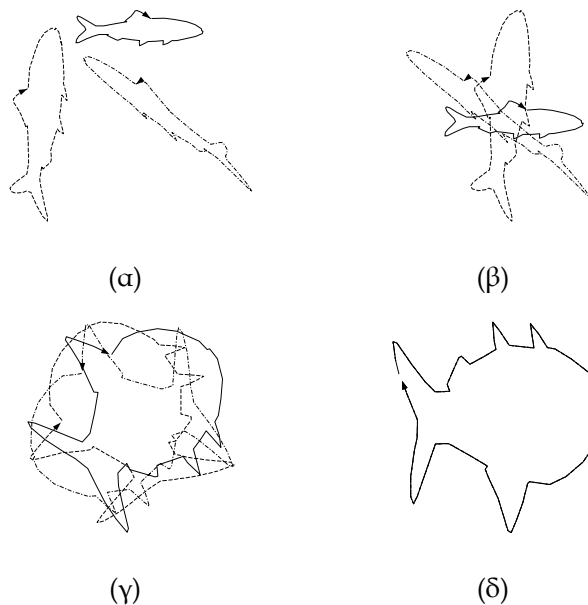
Όπως αναφέρεται στην Ενότητα 9.2, η ποιότητα της κατάτμησης επηρεάζει σε μεγάλο βαθμό την επίδοση της προτεινόμενης μεθόδου κανονικοποίησης. Αυτό αληθεύει γενικά για οποιαδήποτε τεχνική μοντελοποίησης και ταυρίσματος περιγραμμάτων που προέρχονται από φυσικές εικόνες. Στο πλαίσιο της διατριβής χρησιμοποιήθηκε ο αλγόριθμος M-RSST για την κατάτμηση με βάση το χρώμα σε εικόνες υψηλής ανάλυσης, ενώ η κατάτμηση με βάση την κίνηση σε ακολουθίες εικόνων πραγματοποιήθηκε βάσει 2-Δ παραμετρικών μοντέλων κίνησης. Ενδεικτικά αποτελέσματα για μία στατική εικόνα ενός αυτοκινήτου παρουσιάζονται στο Σχήμα 11-3. Συγκεκριμένα, στο Σχήμα 11-3α απεικονίζεται η αρχική εικόνα, ενώ στα Σχήματα 11-3β,γ το πρώτο και το τελικό στάδιο κατάτμησης του αλγορίθμου M-RSST. Το περίγραμμα του αυτοκινήτου φαίνεται στο Σχήμα 11-3δ.

Τα εξαγόμενα περιγράμματα αντικειμένων από τη διαδικασία κατάτμησης χρησιμοποιούνται στη συνέχεια για κανονικοποίηση με τον προτεινόμενο αλγόριθμο. Πρέπει να σημειωθεί ότι η λειτουργία του αλγορίθμου είναι πλήρως αυτόματη, με την έννοια ότι από τη

στιγμή που τα περιγράμματα είναι διαθέσιμα ως καμπόλες δεν απαιτείται ανθρώπινη παρέμβαση ούτε ρύθμιση παραμέτρων για κάθε καμπόλη.



Σχήμα 11-3. Ενδεικτικά αποτελέσματα κατάρτησης με βάση το χρώμα για την εικόνα ενός αυτοκινήτου. (α) Αρχική εικόνα. (β) Κατάρτηση στο χαμηλότερο επίπεδο ανάλυσης. (γ) Τελική κατάρτηση. (δ) Εξαγόμενο περίγραμμα.



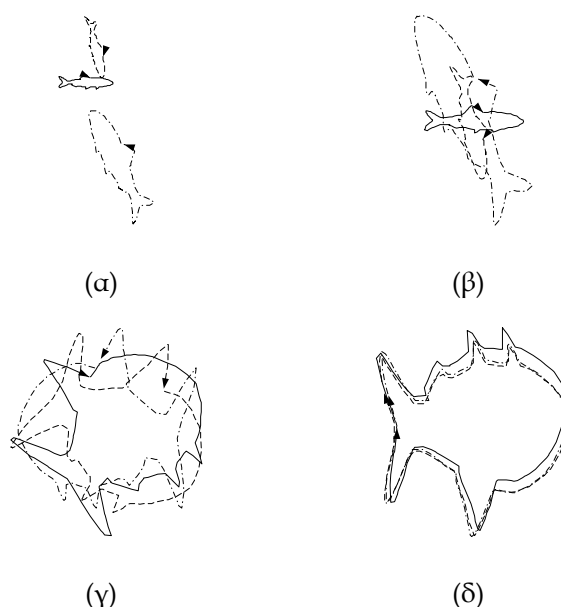
Σχήμα 11-4. Στάδια κανονικοποίησης με απευθείας χρήση δειγμάτων, χωρίς μοντελοποίηση με B-Splines ή νέα δειγματοληψία. (α) Τρία περιγράμματα ενός ψαριού, που προέρχονται από την ίδια καμπόλη με εφαρμογή διαφορετικών affine μετασχηματισμών. (β) Κανονικοποίηση μετατόπισης. (γ) Κανονικοποίηση κλίμακας και κλίσης (ορθοκανονικοποίηση). (δ) Κανονικοποίηση σημείου αναφοράς, περιστροφής και κατοπτρισμού (τελικό στάδιο κανονικοποίησης).

Ίδιο Αντικείμενο, Ίδια Δειγματοληψία

Το πρώτο πείραμα αφορά στην απευθείας χρήση των δειγμάτων της καμπόλης, δηλαδή χωρίς μοντελοποίηση με B-Splines ή νέα δειγματοληψία. Τα αποτελέσματα του πειράματος αυτού παρουσιάζονται στο Σχήμα 11-4. Το περίγραμμα ενός ψαριού, που αποτελείται από 100 σημεία και προέρχεται από κατάρτηση στατικής εικόνας, φαίνεται στην πάνω δεξιά πλευρά του Σχήματος 11-4α. Στο ίδιο Σχήμα περιλαμβάνονται άλλα δύο περιγράμματα που αποτελούνται επίσης από 100 σημεία και παράγονται από το αρχικό περίγραμμα με άμεση εφαρμογή

τυχαίων affine μετασχηματισμών. Σύμφωνα με τη μεθοδολογία που παρουσιάστηκε στο Κεφάλαιο 10, οι τρεις καμπύλες υποβάλλονται σε κανονικοποίηση μετατόπισης, κλίμακας, κλίσης, σημείου αναφοράς, περιστροφής και κατοπτρισμού.

Στο Σχήμα 11-4β απεικονίζονται οι καμπύλες μετά την κανονικοποίηση μετατόπισης. Αντίστοιχα το Σχήμα 11-4γ αφορά στην κανονικοποίηση κλίμακας και κλίσης (ορθοκανονικοποίηση), ενώ το Σχήμα 11-4δ στην κανονικοποίηση σημείου αναφοράς, περιστροφής και κατοπτρισμού (τελικό στάδιο κανονικοποίησης). Όμοια μορφή παρουσίασης ακολουθείται και για τα υπόλοιπα Σχήματα της παρούσας Ενότητας. Σε όλες τις περιπτώσεις το σημείο αναφοράς και η φορά διαγραφής προσδιορίζονται με ένα βέλος το οποίο είναι τοποθετημένο στο τελευταίο σημείο της καμπύλης (ένα σημείο πριν το σημείο αναφοράς). Παρατηρείται ότι οι τελικές καμπύλες μετά την κανονικοποίηση συμπίπτουν απόλυτα όταν οι αρχικές προκύπτουν με affine μετασχηματισμούς της ίδιας καμπύλης, επιβεβαιώνοντας έτσι τα θεωρητικά αποτελέσματα του Κεφαλαίου 10.



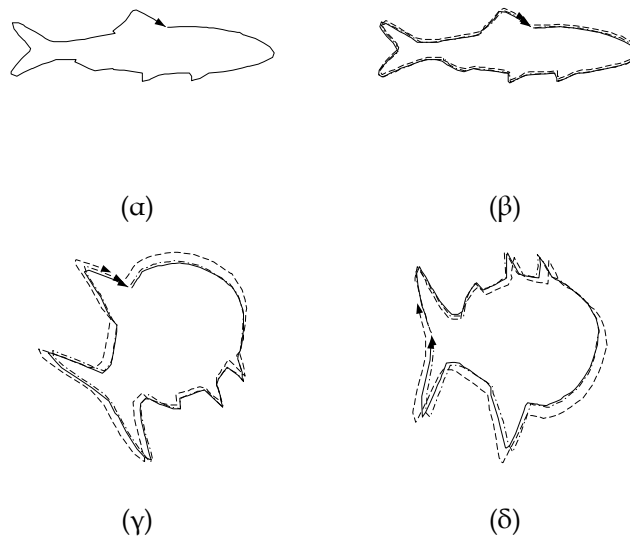
Σχήμα 11-5. Κανονικοποίηση περιγραμμάτων με διαφορετική, μη ομοιόμορφη δειγματοληψία και τυχαίους affine μετασχηματισμούς. (α) Αρχικές καμπύλες. (β,γ,δ) Στάδια κανονικοποίησης, όπως στο Σχήμα 11-4.

Διαφορετική, Μη Ομοιόμορφη Δειγματοληψία

Στην πραγματικότητα, οι κανονικοποιημένες καμπύλες είναι ταυτόσημες μόνο όταν προέρχονται από affine μετασχηματισμούς της ίδιας καμπύλης, και μάλιστα χωρίς νέα δειγματοληψία. Για να γίνει αυτό περισσότερο κατανοητό, τα Σχήματα 11-5 και 11-6 επιδεικνύουν τις ατέλειες στο ταίριασμα ορθοκανονικοποιημένων περιγραμμάτων που προέρχονται από την ίδια καμπύλη, αλλά με εφαρμογή διαφορετικής και μη ομοιόμορφης δειγματοληψίας. Συγκεκριμένα, το Σχήμα 11-5α απεικονίζει το περίγραμμα ενός ψαριού μαζί με άλλα δύο περιγράμματα που προέρχονται από το πρώτο με εφαρμογή τυχαίων affine μετασχηματισμών. Και οι τρεις καμπύλες έχουν υποστεί διαφορετική και μη ομοιόμορφη δειγματοληψία. Τα βήματα κανονικοποίησης των περιγραμμάτων παρουσιάζονται στα Σχήματα 11-5β-δ, όπου ακολουθείται η ίδια λογική παρουσίασης όπως και στο Σχήμα 11-4.

Τα αποτελέσματα δεν απέχουν πολύ από τα επιθυμητά, με την έννοια ότι και πάλι οι κανονικοποιημένες καμπύλες ταιριάζουν αρκετά μεταξύ τους. Αυτό ισχύει παρά το γεγονός ότι

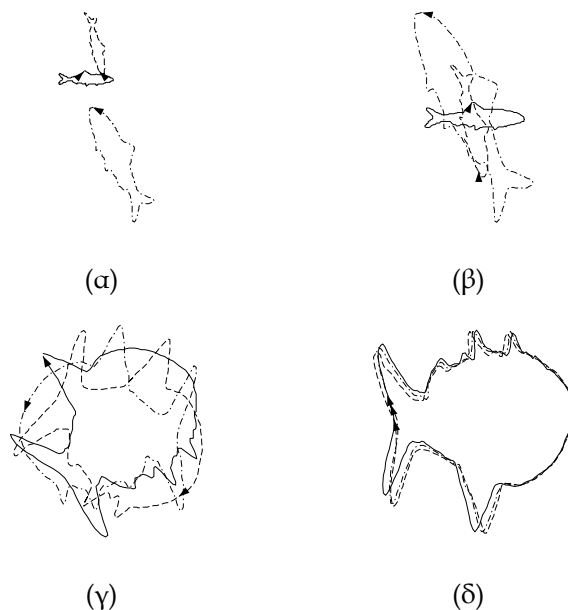
οι διαφορές στη διαδικασία δειγματοληψίας των τριών περιγραμμάτων είναι σημαντικές. Όμως οι παρατηρούμενες ανομοιότητες των κανονικοποιημένων περιγραμμάτων μπορούν να είναι ανεπιθύμητες όταν απαιτείται μεγάλη ακρίβεια στο ταίριασμα καμπύλων σε μια συγκεκριμένη εφαρμογή. Είναι αξιοσημείωτο ότι ανομοιότητες προκύπτουν λόγω διαφορετικής δειγματοληψίας ακόμη και απουσία μετασχηματισμών. Αυτό είναι εμφανές στο Σχήμα 11-6, όπου το πείραμα επαναλαμβάνεται χωρίς την εφαρμογή affine μετασχηματισμών στις τρεις καμπύλες. Παρόλο που στην περίπτωση αυτή οι αρχικές καμπύλες ταυτίζονται, οπτικά τουλάχιστον, οι κανονικοποιημένες διαφέρουν, σε μικρό βέβαια βαθμό.



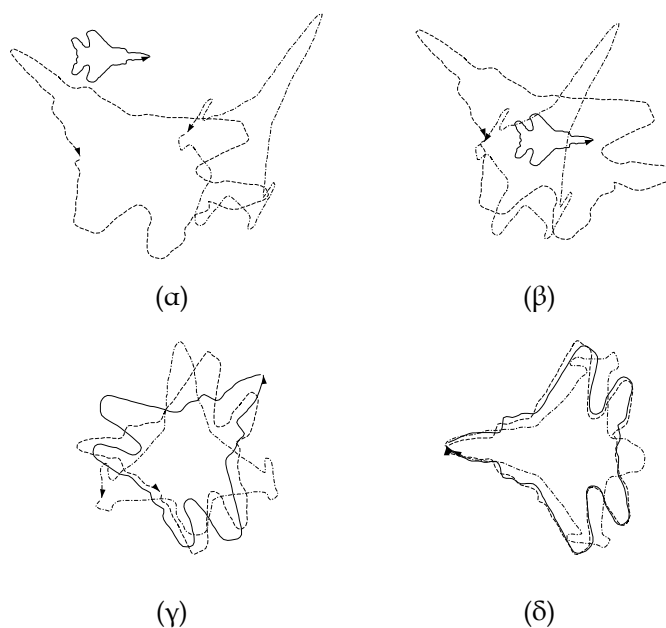
Σχήμα 11-6. Κανονικοποίηση περιγραμμάτων με διαφορετική, μη ομοιόμορφη δειγματοληψία αλλά χωρίς μετασχηματισμούς. (α) Αρχικές καμπύλες, οπτικά ταυτόσημες. (β, γ, δ) Στάδια κανονικοποίησης, όπως στο Σχήμα 11-4.

Διαφορετική, Ομοιόμορφη Δειγματοληψία

Τα αποτελέσματα βελτιώνονται σημαντικά με την εφαρμογή νέα δειγματοληψίας πριν τη διαδικασία κανονικοποίησης. Αυτό επιτυγχάνεται μοντελοποιώντας τα περιγράμματα με B-Splines και ανακατανέμοντας τα κομβικά σημεία ώστε η δειγματοληψία να είναι ομοιόμορφη ως προς το μήκος τόξου, όπως περιγράφεται στην Ενότητα 9.4. Τα ανακατανομημένα κομβικά σημεία διατηρούνται πλέον ως δείγματα της καμπύλης στην περίπτωση αυτή, και σε αυτά εφαρμόζεται η διαδικασία κανονικοποίησης. Το Σχήμα 11-7 απεικονίζει τη βελτίωση στο ταίριασμα των κανονικοποιημένων περιγραμμάτων πραγματοποιώντας δύο φορές τη μοντελοποίηση με B-Splines και την ανακατανομή κομβικών σημείων: μία πριν την κανονικοποίηση μετατόπισης και μία πριν την κανονικοποίηση σημείου αναφοράς και περιστροφής. Ένα τελικό στάδιο δειγματοληψίας πραγματοποιείται στα πλήρως κανονικοποιημένα περιγράμματα, ώστε να επιτυγχάνεται όσο το δυνατόν καλύτερο ταίριασμα. Το Σχήμα 11-7 είναι άμεσα συγκρίσιμο με το 11-5, αποκαλύπτοντας σημαντική βελτίωση στο ταίριασμα των τριών περιγραμμάτων. Στα όλα τα πειράματα που ακολουθούν εφαρμόζεται η παραπάνω διαδικασία μοντελοποίησης με B-Splines και ανακατανομής κομβικών σημείων.



Σχήμα 11-7. Επανάληψη του πειράματος του Σχήματος 11-5 με εφαρμογή μοντελοποίησης με B-Splines και ανακατανομής κομβικών σημείων. (α) Αρχικές καμπύλες. (β,γ,δ) Στάδια κανονικοποίησης, όπως στο Σχήμα 11-4.



Σχήμα 11-8. Ευθυγράμμιση περιγραμμάτων με παρόμοιο οπτικό περιεχόμενο. (α) Αρχικά περιγράμματα τριών διαφορετικών αεροσκαφών. (β,γ,δ) Στάδια κανονικοποίησης, όπως στο Σχήμα 11-4.

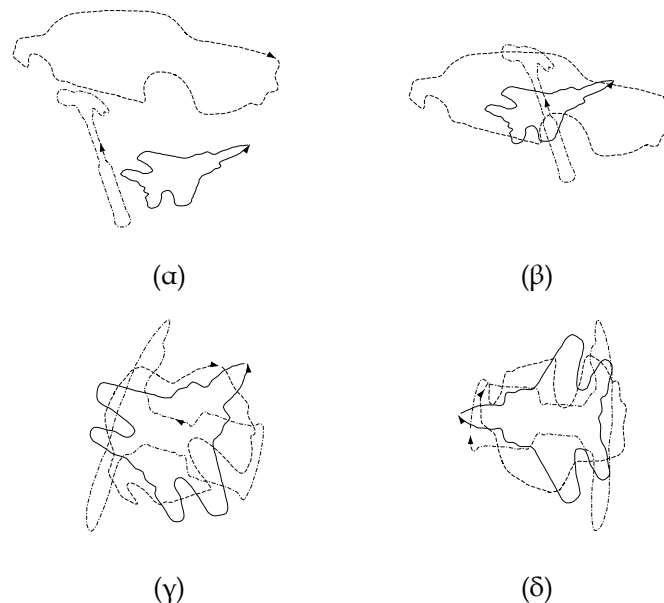
Παρόμοια Αντικείμενα

Το ως η σπουδαιότερη ιδιότητα της προτεινόμενης μεθόδου κανονικοποίησης είναι η ικανότητά της να «ευθυγραμμίζει» σχήματα που έχουν παρόμοιο οπτικό περιεχόμενο. Η ιδιότητα αυτή εξετάστηκε με χρήση μεγάλου αριθμού περιγραμμάτων τα οποία ανήκουν σε διαφορετικές κατηγορίες αντικειμένων, όπως για παράδειγμα αεροσκάφη, αυτοκίνητα, ψάρια, ποτήρια και σφυριά. Το Σχήμα 11-8 παρουσιάζει τα αποτελέσματα κανονικοποίησης για τρία διαφορετικά περιγράμματα αεροσκαφών που προέρχονται από κατάτμηση στατικών εικόνων, και επομένως η δειγματοληψία τους διαφέρει σε μεγάλο βαθμό. Η μέθοδος κανονικοποίησης, σε συν-

δυσασμό με τη μοντελοποίηση B-Splines, ευθυγραμμίζει σωστά τα περιγράμματα και μάλιστα χωρίς να πραγματοποιεί ταίριασμα μεταξύ τους.

Ανόμοια Αντικείμενα

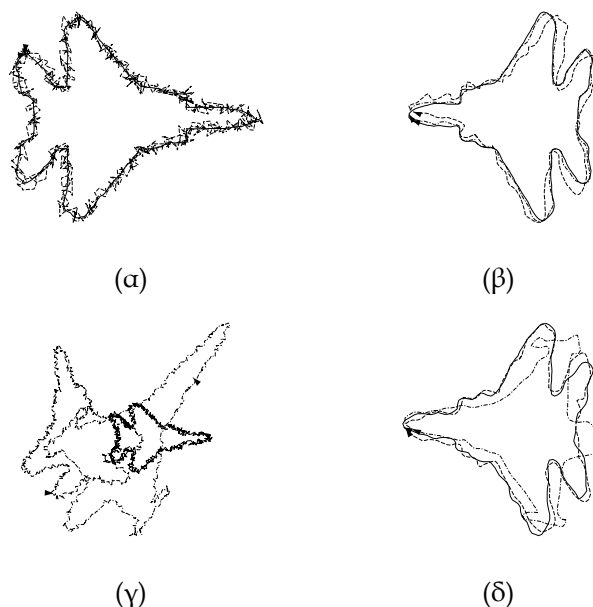
Τα αποτελέσματα κανονικοποίησης του Σχήματος 11-8 θα πρέπει να συγκριθούν με εκείνα του Σχήματος 11-9, όπου τα τρία αρχικά περιγράμματα αντιστοιχούν σε διαφορετικές κατηγορίες αντικειμένων, και επομένως έχουν διαφορετικό οπτικό περιεχόμενο. Συγκεκριμένα εξετάζεται το περίγραμμα ενός αεροσκάφους, ενός αυτοκινήτου και ενός σφυριού. Είναι εμφανές ότι με οποιαδήποτε τεχνική ταιριάσματος, τα περιγράμματα του Σχήματος 11-8 θα έδιναν μεγαλύτερο βαθμό ομοιότητας από εκείνα του Σχήματος 11-9. Συμπεραίνεται λοιπόν ότι η προτεινόμενη μέθοδος κανονικοποίησης μπορεί να χρησιμοποιηθεί ως στάδιο προεπεξεργασίας σε οποιαδήποτε τεχνική ταιριάσματος, ταξινόμησης ή αναγνώρισης σχημάτων βελτιώνοντας τα αποτελέσματά τους.



Σχήμα 11-9. Ευθυγράμμιση περιγραμμάτων με διαφορετικό οπτικό περιεχόμενο. (α) Αρχικά περιγράμματα ενός αεροσκάφους, ενός αυτοκινήτου και ενός σφυριού. (β,γ,δ) Στάδια κανονικοποίησης, όπως στο Σχήμα 11-4.

Επίδραση Θορύβου

Στη συνέχεια εξετάζεται η επίδραση του θορύβου στην κανονικοποίηση περιγραμμάτων. Ο θόρυβος στα περιγράμματα αντικειμένων συνήθως προέρχεται από τα σφάλματα στη διαδικασία κατάτμησης εικόνων με βάση το χρώμα ή την κίνηση. Για τη μοντελοποίηση των αποτελεσμάτων του θορύβου εισάγεται τεχνητά τυχαίος θόρυβος κανονικής κατανομής και σχετικά μεγάλης διασποράς στα διαθέσιμα αρχικά περιγράμματα και εξετάζεται η επίδρασή του στα αντίστοιχα κανονικοποιημένα περιγράμματα. Όπως φαίνεται από τα πειράματα, η απόδοση του αλγορίθμου κανονικοποίησης είναι ικανοποιητική, αρκεί να μη δημιουργείται σημαντική παραμόρφωση σχήματος λόγω του θορύβου. Ενδεικτικά πειραματικά αποτελέσματα απεικονίζονται στο Σχήμα 11-10.



Σχήμα 11-10. Κανονικοποίηση περιγραμμάτων παρουσία τυχαίου θορύβου. (α,β) Κανονικοποίηση τριών περιγραμμάτων του ίδιου αεροσκάφους με διαφορετική ισχύ θορύβου. (γ,δ) Κανονικοποίηση τριών περιγραμμάτων με παρόμοιο οπτικό περιεχόμενο, παρουσία θορύβου και τυχαίων affine μετασχηματισμών.











Συγκεκριμένα, στο Σχήμα 11-10α παρουσιάζεται το αρχικό περίγραμμα ενός αεροσκάφους (χωρίς θόρυβο), μαζί με άλλα δύο που προέρχονται από το αρχικό με πρόσθεση θορύβου. Στο ένα από τα δύο θορυβώδη περιγράμματα ο σηματοθορυβικός λόγος είναι $SNR = 30dB$ και στο άλλο $20dB$. Τα τρία αντίστοιχα κανονικοποιημένα περιγράμματα απεικονίζονται στο Σχήμα 11-10β. Πρέπει να σημειωθεί ότι ο θόρυβος εισάγεται στο αρχικό περίγραμμα, πριν τη μοντελοποίηση με B-Splines. Έτσι φαίνεται στο Σχήμα 11-10β πώς η μοντελοποίηση αυτή «απλοποιεί» τα περιγράμματα και ουσιαστικά αφαιρεί μεγάλο τμήμα του θορύβου. Φαίνεται επίσης ότι η ισχύς του θορύβου επηρεάζει τη διαδικασία κανονικοποίησης, αλλά σε μικρό μόνο βαθμό. Στα Σχήματα 11-10γ,δ το πείραμα επαναλαμβάνεται για τρία διαφορετικά περιγράμματα αεροσκαφών τα οποία έχουν υποστεί προσθήκη θορύβου και τυχαίους affine μετασχηματισμούς. Η σύγκριση του Σχήματος 11-10δ με το 11-8δ δείχνει ότι ιδιότητα «ευθυγράμμισης» σχημάτων με παρόμοιο οπτικό περιεχόμενο δεν επηρεάζεται σημαντικά από την παρουσία του θορύβου.

11.4 ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΑΞΙΝΟΜΗΣΗΣ ΚΑΙ ΑΝΑΖΗΤΗΣΗΣ

Απόσταση Περιγραμμάτων

Η αναγνώριση και ταξινόμηση αντικειμένων βάσει της πληροφορίας περιγράμματος έχει αντιμετωπισθεί σε μία πληθώρα εργασιών, όπως αναφέρεται στην Εισαγωγή της Ενότητας 9.1, και δεν αποτελεί βασικό αντικείμενο ανάλυσης στην παρούσα εργασία. Ενδεικτικά πειραματικά αποτελέσματα ταιριάσματος περιγραμμάτων παρουσιάζονται όμως στον Πίνακα 11-1. Στον Πίνακα αυτό εμφανίζονται πέντε ζεύγη περιγραμμάτων που αντιστοιχούν σε περιπτώσεις ίδιων αντικειμένων (περιγράμματα του ίδιου αεροσκάφους με διαφορετικό affine μετασχηματισμό), παρόμοιων (περιγράμματα δύο διαφορετικών αεροσκαφών) ή και διαφορετικών (περιγράμματα αντικειμένων που ανήκουν σε διαφορετικές κατηγορίες).

Σε όλες τις περιπτώσεις πραγματοποιούνται συγκρίσεις των περιγραμμάτων 1 και 2 του Πίνακα 11-1 με χρήση τριών μετρικών απόστασης, και συγκεκριμένα (α) Ευκλείδεια απόσταση κομβικών σημείων, (β) Ευκλείδεια απόσταση κανονικοποιημένων Fourier descriptors (NFD), και (γ) τροποποιημένοι Fourier descriptors (MFD) [108]. Στην περίπτωση της απόστασης κομβικών σημείων, πραγματοποιούνται μετρήσεις πριν και μετά την κανονικοποίηση. Όλες οι μετρικές είναι κανονικοποιημένες έτσι ώστε οι τιμές 0 και 1 να υποδηλώνουν απόλυτη ταύτιση και πλήρη ανομοιότητα, αντίστοιχα. Όπως είναι αναμενόμενο, η απόσταση που υπολογίζεται απευθείας στα κομβικά σημεία αποτυγχάνει για τα αρχικά περιγράμματα που δεν έχουν υποστεί κανονικοποίηση, όπως φαίνεται χαρακτηριστικά στο πείραμα #1. Αντίθετα, όπως είναι εμφανές από τον Πίνακα 11-1, όλες οι υπόλοιπες μετρικές απόστασης δίνουν αξιόπιστα αποτελέσματα. Είναι αξιοσημείωτο ότι ένα τόσο απλό μέτρο όπως η Ευκλείδεια απόσταση των δειγμάτων μπορεί να δώσει ικανοποιητική σύγκριση περιγραμμάτων, και μάλιστα καλύτερη σε σχέση με τις πιο εξειδικευμένες τεχνικές (NFD και MFD) σε όλα τα πειράματα εκτός από το #2.

α/α	Περιγράμμα 1	Περιγράμμα 2	ΑΚΣ (αρχ)	ΑΚΣ (καν)	NFD	MFD
#1			0.69	0.01	0.02	0.01
#2			0.34	0.19	0.12	0.11
#3			0.81	0.75	0.41	0.57
#4			0.93	0.89	0.62	0.65
#5			0.84	0.76	0.25	0.32

Πίνακας 11-1. Ενδεικτικές αποστάσεις μεταξύ των περιγραμμάτων 1 και 2 για το ίδιο αντικείμενο με διαφορετικό μετασχηματισμό, παρόμοια, ή και διαφορετικά αντικείμενα. ΑΚΣ: Απόσταση κομβικών σημείων (αρχικά / κανονικοποιημένα περιγράμματα). FD: Απόσταση κανονικοποιημένων Fourier Descriptors. MFD: Απόσταση Modified Fourier Descriptors [108].

Ταξινόμηση Περιγραμμάτων

Στη συνέχεια παρουσιάζονται αποτελέσματα ταξινόμησης περιγραμμάτων με την προτεινόμενη τεχνική νευρωνικού δικτύου. Για το σκοπό αυτό χρησιμοποιήθηκε μία μικρή βάση δεδομένων η οποία περιέχει 100 στατικές εικόνες αντικειμένων. Τα αντικείμενα διακρίνονται σε πέντε κατηγορίες ($C=5$), και συγκεκριμένα αεροσκάφη, αυτοκίνητα, ψάρια, ποτήρια και σφυριά· στη βάση υπάρχουν 20 εικόνες αντικειμένων από κάθε κατηγορία. Στο στάδιο εκπαίδευσης του δικτύου χρησιμοποιήθηκαν 10 εικόνες από κάθε κατηγορία, ενώ οι υπόλοιπες 10 χρησιμοποιήθηκαν για πειράματα ταξινόμησης. Έτσι το πειραματικό σύνολο περιλαμβάνει συνολικά 50 περιγράμματα αντικειμένων. Σημειώνεται ότι κάθε περίγραμμα κανονικοποιείται πριν εισαχθεί για ταξινόμηση στο νευρωνικό δίκτυο· επειδή η κανονικοποιημένη αναπαράσταση είναι αναλλοίωτη στους affine μετασχηματισμούς, καμπύλες που έχουν υποστεί οποιοδήποτε affine μετασχηματισμό δίνουν ακριβώς το ίδιο αποτέλεσμα ταξινόμησης.

Έτσι η ταξινόμηση περιγραμμάτων που ανήκουν στο σύνολο εκπαίδευσης είναι 100% επιτυχής.

Κατηγορία	NN (R = 1)	NN (R = 2)	Ταίριασμα
Αεροσκάφη	9/10	10/10	9/10
Αυτοκίνητα	8/10	10/10	10/10
Ψάρια	8/10	9/10	8/9
Ποτήρια	6/10	9/10	9/9
Σφυριά	10/10	10/10	9/10
Σύνολο	41/50	48/50	45/48

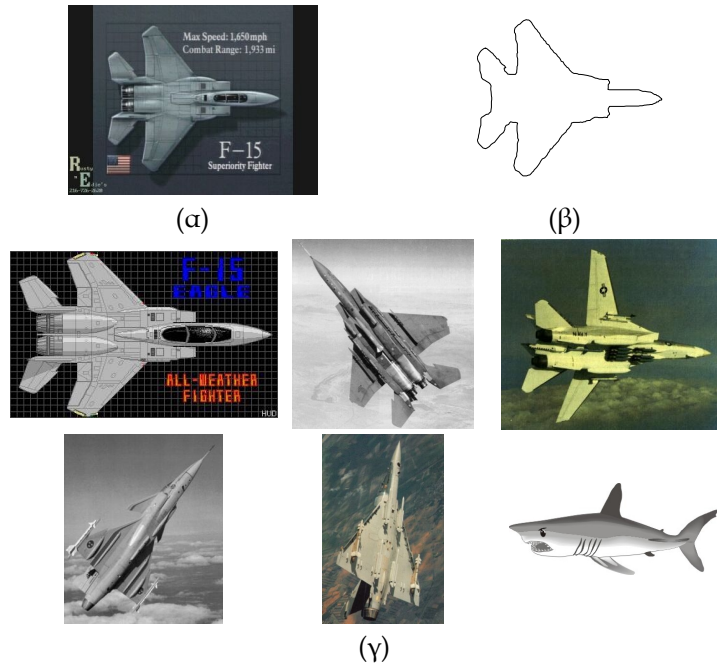
Πίνακας 11-2. Αποτελέσματα ταξινόμησης περιγραμμάτων με χρήση νευρωνικού δικτύου μόνο (R = 1) και με συνδυασμό νευρωνικού δικτύου (με R = 2) και ταιριάσματος περιγραμμάτων στις επιλεγμένες κατηγορίες.

Αφού ολοκληρωθεί η εκπαίδευση του νευρωνικού δικτύου με χρήση του συνόλου εκπαίδευσης που αποτελείται από 50 περιγράμματα, πραγματοποιούνται δοκιμές ταξινόμησης με χρήση των 50 περιγραμμάτων του πειραματικού συνόλου. Αρχικά πραγματοποιείται ταξινόμηση με το νευρωνικό δίκτυο μόνο, χωρίς ταίριασμα, επιλέγεται δηλαδή η μέγιστη έξοδος του νευρωνικού δικτύου (R = 1). Τα αποτελέσματα ταξινόμησης παρουσιάζονται στην πρώτη στήλη του Πίνακα 11-2. Από τα 50 περιγράμματα τα 41 ταξινομούνται σωστά, δίνοντας ποσοστό επιτυχίας 82%. Στη συνέχεια διεξάγεται ένα δεύτερο πείραμα στο οποίο για κάθε περίγραμμα εισόδου επιλέγονται οι δύο κατηγορίες (R = 2) που αντιστοιχούν στις δύο μεγαλύτερες εξόδους του δικτύου. Στην περίπτωση αυτή το περίγραμμα εισόδου συγκρίνεται με το σύνολο των 10 προτύπων περιγραμμάτων εκπαίδευσης κάθε μιας από τις δύο επιλεγμένες κατηγορίες. Έτσι υπολογίζονται συνολικά 20 αποστάσεις και το περίγραμμα εισόδου ταξινομείται στην κατηγορία εκείνη που παρουσιάζει την ελάχιστη απόσταση.

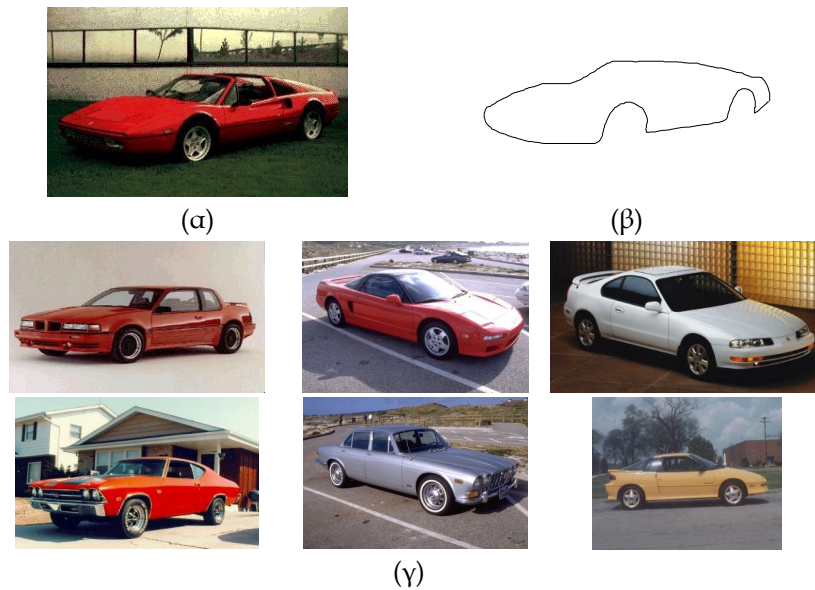
Όπως φαίνεται στη δεύτερη και τρίτη στήλη του Πίνακα 11-2, στα 48 από τα 50 περιγράμματα εισόδου το νευρωνικό δίκτυο πραγματοποιεί σωστή ταξινόμηση, δηλαδή μία από τις δύο επιλεγόμενες κατηγορίες είναι η σωστή. Από τα 48 αυτά περιγράμματα, τα 45 ταξινομούνται σωστά στο τελικό στάδιο ταιριάσματος, δίνοντας ποσοστό επιτυχίας 90%. Συμπεραίνεται ότι η προτεινόμενη τεχνική ταξινόμησης σε δύο στάδια με συνδυασμό νευρωνικού δικτύου και ταιριάσματος είναι πιο αξιόπιστη από την ταξινόμηση με νευρωνικό δίκτυο μόνο· από την άλλη μεριά η χρήση του νευρωνικού δικτύου είναι απαραίτητη για την επιτάχυνση της διαδικασίας ταξινόμησης στην περίπτωση που ο αριθμός προτύπων περιγραμμάτων είναι πολύ μεγάλος.

Αναζήτηση – Ανάκληση με Βάση το Σχήμα

Τέλος, η προτεινόμενη μέθοδος εξετάστηκε για τους σκοπούς αναζήτησης και ανάκλησης με βάση το σχήμα στην ίδια βάση δεδομένων που περιέχει 100 στατικές εικόνες αντικειμένων από πέντε διαφορετικές κατηγορίες. Η ανάκληση βασίστηκε σε μετρήσεις ομοιότητας των κανονικοποιημένων περιγραμμάτων των αντικειμένων, και συγκεκριμένα στην Ευκλείδεια απόσταση κομβικών σημείων που δίνεται στην Εξίσωση (11-6). Ιδιαίτερη έμφαση δόθηκε στην ακριβή εξαγωγή περιγραμμάτων από τις εικόνες· γι' αυτό το λόγο απαιτήθηκε παρέμβαση στη διαδικασία κατάτμησης σε ορισμένες περιπτώσεις όπου η πλήρως αυτοματοποιημένη κατάτμηση δεν έδινε τα σωστά περιγράμματα. Τα αποτελέσματα μιας αναζήτησης αεροσκάφους παρουσιάζονται στο Σχήμα 11-11.



Σχήμα 11-11. Αναζήτηση μέσω παραδείγματος (query-by-example) με βάση το σχήμα για μία εικόνα αεροσκάφους. (α) Εικόνα εισόδου. (β) Εξαγόμενο περίγραμμα. (γ) Αποτελέσματα αναζήτησης.



Σχήμα 11-12. Αναζήτηση με βάση το σχήμα για μία εικόνα αυτοκινήτου. (α) Εικόνα εισόδου. (β) Εξαγόμενο περίγραμμα. (γ) Αποτελέσματα αναζήτησης.

Πρόκειται για μια περίπτωση αναζήτησης μέσω παραδείγματος (query-by-example). Συγκεκριμένα, η εικόνα ενός αεροσκάφους που φαίνεται στο Σχήμα 11-11α παρουσιάζεται ως είσοδος στο σύστημα αναζήτησης. Στην εικόνα αυτή πραγματοποιείται κατάτμηση για την εξαγωγή του περιγράμματος του κύριου αντικειμένου, που εικονίζεται στο Σχήμα 11-11β. Στη συνέχεια υπολογίζεται η απόσταση μεταξύ του περιγράμματος εισόδου και όλων των περιγραμμάτων της βάσης, και οι εικόνες που αντιστοιχούν στα περιγράμματα εκείνα με τη μεγαλύτερη ομοιότητα παρουσιάζονται από το σύστημα ως αποτελέσματα ανάκλησης. Ως μέτρο σύγκρισης χρησιμοποιείται η Ευκλείδεια απόσταση των κομβικών σημείων των κανονικοποιημένων περιγραμμάτων. Οι έξι πρώτες εικόνες εμφανίζονται κατά φθίνουσα σειρά ομοιότητας στο Σχήμα 11-11γ. Σημειώνεται ότι η αναζήτηση θα μπορούσε εξίσου να πραγματοποιηθεί

ποιηθεί μέσω *σχεδιαγράμματος ή σκίτσου* (query-by-sketch). Η περίπτωση αυτή είναι απλούστερη αφού δεν είναι απαραίτητη η εξαγωγή περιγράμματος, αλλά μόνο η κανονικοποίησή του.

Ανάλογα αποτελέσματα ανάκλησης παρουσιάζονται στο Σχήμα 11-12 για μια περίπτωση αναζήτησης ενός αυτοκινήτου. Παρατηρείται ότι ενώ στην περίπτωση αυτή όλες οι ανακαλούμενες εικόνες ανήκουν στη σωστή κατηγορία, πρόκειται δηλαδή για εικόνες αυτοκινήτων, στην περίπτωση των αεροσκαφών έχουμε μία λανθασμένη ανάκληση μιας εικόνας ψαριού. Το συγκεκριμένο περίγραμμα ψαριού δίνει μεγαλύτερο βαθμό ομοιότητας με το περίγραμμα εισόδου από ότι το περίγραμμα άλλων αεροσκαφών της βάσης. Κάτι τέτοιο είναι αναμενόμενο αφού το χρησιμοποιούμενο μέτρο σύγκρισης είναι ιδιαίτερα απλοϊκό. Πιο αξιόπιστη σύγκριση μπορεί να επιτευχθεί με τεχνικές όπως τα *παραμορφούμενα πρότυπα* (deformable templates) [19] ή τα *μοντέλα ενεργών περιγραμμάτων* (active contour models) [77]. Πρέπει να σημειωθεί ότι είναι δυνατή η χρήση οποιοσδήποτε τεχνικής σύγκρισης αφού η κανονικοποίηση δεν εισάγει καμία απώλεια πληροφορίας. Άλλα χαρακτηριστικά των αντικειμένων, όπως για παράδειγμα το χρώμα ή η υφή, θα μπορούσαν επίσης να χρησιμοποιηθούν σε συνδυασμό με το σχήμα σε ένα ολοκληρωμένο σύστημα αναζήτησης με βάση το περιεχόμενο.

ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΔΕΛΤΙΩΝ ΕΙΔΗΣΕΩΝ

Το αντικείμενο του τέταρτου και τελευταίου Μέρους της διατριβής είναι η εφαρμογή τεχνικών αυτόματης ανάλυσης βίντεο στην ειδική περίπτωση των τηλεοπτικών δελτίων ειδήσεων. Η κατηγορία αυτή οπτικοακουστικού υλικού παρουσιάζει αυξημένο ενδιαφέρον για δύο κυρίως λόγους: πρώτον διότι η τυπική οργάνωσή τους σε θεματικές ενότητες επιτρέπει την εύκολη χρονική κατάτμηση και συντακτική ανάλυση, και δεύτερον διότι μέχρι σήμερα για το χαρακτηρισμό τους χρησιμοποιούνται χειρονακτικές μέθοδοι με υψηλό κόστος.

Επειδή το επικρατέστερο τμήμα περιεχομένου των δελτίων ειδήσεων σχετίζεται με ανθρώπινες δραστηριότητες, οι εικόνες ανθρώπων και ιδιαίτερα προσώπων παίζουν σημαντικότερο ρόλο στην ανάλυση της δομής των δελτίων· έτσι η προτεινόμενη τεχνική στηρίζεται στην *ανίχνευση προσώπων*. Στο Κεφάλαιο 12 παρουσιάζεται αρχικά μία πλήρως αυτοματοποιημένη διαδικασία ανίχνευσης και εντοπισμού προσώπων. Η διαδικασία αποτελείται από μία σειρά διαδοχικών σταδίων στα οποία ανιχνεύεται αρχικά το χρώμα του ανθρώπινου δέρματος και στη συνέχεια πραγματοποιείται κατάτμηση με βάση το χρώμα, συγχώνευση των τμημάτων δέρματος, ανάλυση του σχήματος των τμημάτων και επαλήθευση με χρήση μιας τεχνικής ταιριάσματος με πρότυπο πρόσωπου.

Ο αξιόπιστος εντοπισμός ανθρώπινων προσώπων σε συνδυασμό με άλλα χαρακτηριστικά χρώματος και κίνησης σε μία ακολουθία δελτίων ειδήσεων χρησιμοποιείται για την ταξινόμηση των πλάνων της ακολουθίας σε προκαθορισμένες κατηγορίες όπως «παρουσιαστής», «ρεπορτάζ / σύνεντευξη», «στατική εικόνα» και «εξωτερικό πλάνο». Στο Κεφάλαιο 13 περιγράφεται η προτεινόμενη τεχνική ταξινόμησης με χρήση ενός αριθμού απλών κανόνων. Με δεδομένο ότι στην τυπική οργάνωση των τηλεοπτικών ειδήσεων περιλαμβάνονται επαναλαμβανόμενες εμφανίσεις του παρουσιαστή εναλλάξ με άλλα είδη πλάνων, η ταξινόμηση αυτή οδηγεί άμεσα στη *συντακτική ανάλυση*, δηλαδή στη χρονική κατάτμηση των δελτίων σε στοιχειώδεις ενότητες ειδήσεων.

Τέλος, στο Κεφάλαιο 14 εξετάζεται η εφαρμογή των προτεινόμενων μεθόδων χαρακτηρισμού, δεικτοδότησης, εξαγωγής περίληψης, προεπισκόπησης και αναζήτησης με βάση το περιεχόμενο στην ειδική περίπτωση των ακολουθιών δελτίων ειδήσεων. Επιπλέον, για την περίπτωση της αναζήτησης και ανάκλησης εξετάζεται ένας αριθμός από νέα σενάρια αναζήτησης που σχετίζονται με την ύπαρξη ανθρώπινων προσώπων, και συγκεκριμένα η αναζήτηση με βάση τα χαρακτηριστικά χρώματος, το μέγεθος και το πλήθος των προσώπων.

Εντοπισμός Προσώπων

Μια ειδική κατηγορία οπτικοακουστικού υλικού με ιδιαίτερο ενδιαφέρον είναι το υλικό που προέρχεται από τηλεοπτικά δελτία ειδήσεων, για δύο κυρίως λόγους: πρώτον διότι η τυπική οργάνωσή τους σε θεματικές ενότητες επιτρέπει την εύκολη χρονική κατάτμηση και συντακτική ανάλυση, και δεύτερον διότι μέχρι σήμερα για το χαρακτηρισμό τους χρησιμοποιούνται χειρονακτικές μέθοδοι με υψηλό κόστος. Στην ανάλυση ακολουθιών βίντεο από δελτία ειδήσεων πρωταρχικό ρόλο παίζει η εμφάνιση ανθρώπινων προσώπων· για το λόγο αυτό εξετάζεται στο παρόν Κεφάλαιο το πρόβλημα της αυτόματης ανίχνευσης και εντοπισμού προσώπων με βάση ένα πλήθος χαρακτηριστικών. Συγκεκριμένα, το πρώτο στάδιο της προτεινόμενης μεθόδου βασίζεται σε μία ταχεία και αξιόπιστη τεχνική μοντελοποίησης και ανίχνευσης του χρώματος δέρματος. Στη συνέχεια πραγματοποιείται κατάτμηση με βάση το χρώμα και συγχώνευση τμημάτων δέρματος ώστε να προκύψει ένα τμήμα εικόνας για κάθε πρόσωπο, αναλύεται το σχήμα των τμημάτων και υπολογίζονται απλά χαρακτηριστικά σχήματος για την απόρριψη τμημάτων με ακανόνιστο σχήμα. Τέλος, στο στάδιο επαλήθευσης, τμήματα με υψηλή πιθανοφάνεια προσώπου εξετάζονται με μία τεχνική ταιριάσματος με πρότυπο πρόσωπο, λαμβάνοντας έτσι υπόψη και την υφή των τμημάτων. Η συνολική μέθοδος παρουσιάζει υψηλή ακρίβεια ανίχνευσης έχοντας ταυτόχρονα χαμηλό υπολογιστικό κόστος· η πλήρως αυτοματοποιημένη διαδικασία εντοπισμού προσώπων αποτελεί ισχυρό εργαλείο για την περαιτέρω ανάλυση δελτίων ειδήσεων. Περιλαμβάνονται πειραματικά αποτελέσματα ανίχνευσης δέρματος και εντοπισμού προσώπων σε ένα μεγάλο σύνολο πραγματικών εικόνων, τα οποία αξιολογούνται με μετρήσεις ακρίβειας και ανάκλησης.

12.1 ΕΙΣΑΓΩΓΗ

Έχει διαπιστωθεί ότι τα υπάρχοντα συστήματα αναζήτησης με βάση το περιεχόμενο αδυνατούν να εντοπίσουν σημασιολογική πληροφορία από οπτικοακουστικό υλικό, γεγονός που οφείλεται στο ότι η δυνατότητα αυτή απαιτεί πάντα εκ των προτέρων γνώση και μπορεί να επιτευχθεί μόνο στο πλαίσιο εξειδικευμένων εφαρμογών. Ένα σημαντικό παράδειγμα τέτοιας εφαρμογής είναι η αναζήτηση εικόνων που περιέχουν ανθρώπινα πρόσωπα, καθώς και ο εντοπισμός ή η αναγνώριση των προσώπων αυτών. Επειδή οι εικόνες και οι ακολουθίες εικόνων που περιέχουν ανθρώπινα πρόσωπα παίζουν σημαντικό ρόλο στην περιγραφή οπτικοακουστικού υλικού, η *ανίχνευση προσώπων* (face detection) έχει καθιερωθεί ως ένα σημαντικό εργαλείο στο πλαίσιο εφαρμογών όπως η συντακτική ανάλυση, η δεικτοδότηση, η αναζήτηση με βάση το περιεχόμενο και η εξαγωγή περίληψης από δελτία ειδήσεων [9].

Στο παρελθόν η ανίχνευση προσώπων ήταν συνδεδεμένη με την *αναγνώριση προσώπων* (face recognition)· το γεγονός αυτό είχε σημαντικό αντίκτυπο στους σχετικούς αλγορίθμους. Για να επιτευχθεί η απαιτούμενη ακρίβεια ανίχνευσης επιβάλλονταν αυστηροί περιορισμοί στο περιβάλλον των εικόνων [112], ενώ η πλειοψηφία των αλγορίθμων βασιζόταν σε ασπρό-

μαυρες εικόνες και χρησιμοποιούσε ταίριασμα με πρότυπα προσώπων ή χαμηλού επιπέδου χαρακτηριστικά για την ανίχνευση στοιχείων του προσώπου όπως τα μάτια, η μύτη και το στόμα [147,151]. Οι σύγχρονες εφαρμογές όμως βασίζονται σε έγχρωμες εικόνες και απαιτούν γρήγορη υλοποίηση με ικανοποιητική ακρίβεια, και όχι εξαντλητική έρευνα με πολύ υψηλή ακρίβεια. Ως αποτέλεσμα, αλγόριθμοι που έχουν καθιερωθεί και χρησιμοποιηθεί με επιτυχία για αναγνώριση προσώπων δεν είναι πλέον κατάλληλες ή απαιτούν τροποποιήσεις, ενώ παράλληλα έχουν εμφανισθεί νέες τεχνικές που ουσιαστικά αποσυνδέουν την ανίχνευση από την αναγνώριση προσώπων.

Συγκεκριμένα, η εργασία [139] έχει προκαλέσει το ενδιαφέρον και έχει εμπνεύσει πολλούς ερευνητές για την ανάπτυξη αλγορίθμων ανίχνευσης προσώπων από έγχρωμες εικόνες, καθώς συνδυάζει πολύ γρήγορη υλοποίηση με ικανοποιητικά αποτελέσματα. Η βασική της ιδέα είναι η μοντελοποίηση του χρώματος του ανθρώπινου δέρματος μέσω μιας *κατανομής χρώματος δέρματος* (skin color distribution) που βασίζεται στις χρωματικές συνιστώσες του χώρου $YCrCb$. Οι περισσότερες μελέτες που βασίζονται στην ιδέα αυτή αποκαλύπτουν ότι απαιτούνται πολλά στάδια μετα-επεξεργασίας, ενώ παρά το γεγονός ότι ο χώρος του χρώματος δέρματος πράγματι καλύπτει μικρή επιφάνεια του χρωματικού επιπέδου $Cr-Cb$, δε μπορεί να μοντελοποιηθεί με τόσο γενικό τρόπο ώστε αποδίδει για όλες τις εικόνες προσώπων [15,43]. Η γενίκευση του μοντέλου συνήθως αυξάνει το ποσοστό *λανθασμένης ανίχνευσης* (false alarms), ενώ η εξειδίκευσή του αυξάνει το ποσοστό *απωλειών* (dismissals). Επιπλέον, η επίδραση του καναλιού φωτεινότητας, Y , δεν είναι πάντα αμελητέα.

Για τους παραπάνω λόγους, έχει προταθεί στην εργασία [102] μία τεχνική για τη δυναμική ενημέρωση του χρωματικού μοντέλου όταν παρουσιάζονται αλλαγές λόγω συνθηκών φωτισμού, με χρήση ενός *συνδυασμού Γκαουσιανών μοντέλων* (Gaussian mixture model) για τις κατανομές πιθανότητας του χρώματος δέρματος, ρουχισμού και φόντου της εικόνας. Επίσης, στην εργασία [116] χρησιμοποιείται ένα μοντέλο Markov για την πρόβλεψη της εξέλιξης ενός ιστογράμματος του χρώματος δέρματος. Το ιστογράμμο περιλαμβάνει και τις τρεις χρωματικές συνιστώσες του χώρου HSV , σε αντίθεση με τις τεχνικές που βασίζονται στο χώρο $YCrCb$ όπου χρησιμοποιείται μόνο το επίπεδο $Cr-Cb$. Μία ανάλογη εκτίμηση ιστογράμματος του χρώματος δέρματος έχει προταθεί στην εργασία [71], όπου χρησιμοποιείται ο χώρος RGB απευθείας: από τα πειράματα μάλιστα σε ένα μεγάλο σύνολο εικόνων συμπεραίνεται ότι η επίδοση του συστήματος αυτού είναι ανώτερη από το συνδυασμό Γκαουσιανών μοντέλων.

Όλες οι παραπάνω εργασίες αντιμετωπίζουν το πρόβλημα της ανίχνευσης δέρματος, ενώ στην εργασία [124] τα χαρακτηριστικά του χρώματος δέρματος συνδυάζονται με πληροφορίες τοπικής συμμετρίας και περιγράμματος για τον εντοπισμό πιθανών σημείων προσώπου, με στόχο την ανίχνευση και εντοπισμό ανθρώπινων προσώπων. Ο συνδυασμός της κατάτμησης με βάση το χρώμα με ένα Γκαουσιανό μοντέλο χρώματος δέρματος και επεξεργασία σχήματος έχει επίσης προταθεί στην εργασία [131] για ανίχνευση προσώπων.

Στο πλαίσιο της διατριβής τα χαρακτηριστικά χρώματος και σχήματος που χρησιμοποιούνται εργασία [131] συνδυάζονται περαιτέρω με μία τεχνική ταίριασματος με πρότυπο για την ανίχνευση και εντοπισμό προσώπων. Συγκεκριμένα, προτείνεται ένα *διδιάστατο Γκαουσιανό μοντέλο* για την κατανομή πιθανότητας του χρώματος δέρματος, του οποίου οι παράμετροι επανεκτιμούνται δυναμικά. Σε κάθε εικόνα πραγματοποιείται *κατάτμηση με βάση το χρώμα*, η οποία ακολουθείται από ένα επιπλέον στάδιο συγχώνευσης των τμημάτων που ταίριαζουν με το μοντέλο του χρώματος δέρματος. Στα τμήματα αυτά υπολογίζονται κατάλληλα

χαρακτηριστικά σχήματος και απορρίπτονται εκείνα των οποίων το σχήμα παρουσιάζει μεγάλη απόκλιση από το σχήμα του προσώπου. Τα εναπομείναντα τμήματα εξετάζονται για ύπαρξη προσώπου στο στάδιο επαλήθευσης, το οποίο βασίζεται σε *ταίριασμα με πρότυπο* προσώπου (template matching). Η αναζήτηση προτύπων ολόκληρου προσώπου δίνει καλύτερα αποτελέσματα σε σχέση με την αναζήτηση τοπικών χαρακτηριστικών προσώπου, ιδιαίτερα σε εικόνες χαμηλής ανάλυσης ή για πρόσωπα σε μικρή κλίμακα.

Ο συνδυασμός ανίχνευσης χρώματος δέρματος, προσαρμογής του χρωματικού μοντέλου, κατάτμησης με βάση το χρώμα, συγχώνευσης τμημάτων δέρματος και ανάλυσης σχήματος για εντοπισμό ανθρώπινων προσώπων είναι πρωτότυπος, αν και οι επιμέρους τεχνικές έχουν ξαναχρησιμοποιηθεί. Δημοσιεύθηκε αρχικά στην εργασία [131], ενώ ο επιπλέον συνδυασμός με χαρακτηριστικά υψής στο στάδιο επαλήθευσης (ταίριασμα με πρότυπο προσώπου) για μεγαλύτερη ακρίβεια ανίχνευσης παρουσιάστηκε στην εργασία [132].

Η προτεινόμενη τεχνική βελτιώνει την επίδοση σε σχέση με υπάρχουσες τεχνικές με δύο τρόπους. Πρώτον, η ικανότητα γενίκευσης του μοντέλου χρώματος δέρματος βελτιώνεται με την επανεκτίμηση των παραμέτρων της κατανομής, και δεύτερον, το στάδιο επαλήθευσης είναι πολύ πιο αξιόπιστο, αφού λαμβάνει υπόψη και την υφή του προσώπου, πέρα από τα χαρακτηριστικά χρώματος και σχήματος. Αυτό το όφελος όσον αφορά στην αξιοπιστία απαιτεί βέβαια επιπρόσθετο υπολογιστικό κόστος, το οποίο όμως είναι περιορισμένο αφού το ταίριασμα με πρότυπο εφαρμόζεται μόνο σε τμήματα της εικόνας όπου έχει ανιχνευθεί πρόσωπο.

Ο σημασιολογικός χαρακτηρισμός οπτικοακουστικών προγραμμάτων συνήθως απαιτεί χειρονακτική εργασία και έχει μεγάλο κόστος. Επειδή λοιπόν τα ανθρώπινα πρόσωπα είναι το πλέον κοινό είδος σημασιολογικού αντικειμένου σε προγράμματα όπως δελτία ειδήσεων, συζητήσεις, τηλεοπτικές σειρές ή κινηματογραφικές ταινίες, η προτεινόμενη τεχνική ανίχνευσης και εντοπισμού προσώπων μπορεί να αποτελέσει χρήσιμο εργαλείο για την αυτόματη ανάλυση και δεικτοδότηση σε βάσεις δεδομένων οπτικοακουστικού υλικού. Σε περιπτώσεις όπου απαιτείται υψηλή ακρίβεια και απόλυτη αξιοπιστία ή προσθήκη επιπρόσθετης πληροφορίας χαρακτηρισμού, η τεχνική μπορεί να χρησιμοποιηθεί ως ημι-αυτόματη διαδικασία που να αλληλεπιδρά με τον χειριστή.

12.2 ΜΟΝΤΕΛΟ ΠΙΘΑΝΟΤΗΤΑΣ ΧΡΩΜΑΤΟΣ ΔΕΡΜΑΤΟΣ

Στην παρούσα Ενότητα περιγράφεται η μέθοδος με την οποία ανιχνεύεται το χρώμα του δέρματος. Σε ορισμένες κλασσικές μελέτες [54,109] παρατηρείται ότι οι αποχρώσεις του ανθρώπινου δέρματος καλύπτουν μια μικρή μόνο περιοχή του χρωματικού επιπέδου $Cr-Cb$ του μοντέλου $YCrCb$. Με βάση την παρατήρηση αυτή, προτάθηκε στην εργασία [139] ένας γρήγορος αλγόριθμος ανίχνευσης δέρματος που ενέπνευσε πολλούς ερευνητές. Με ένα παρόμοιο τρόπο στο πλαίσιο της διατριβής η κατανομή του χρώματος δέρματος προσεγγίστηκε από μία διδιάστατη κανονική (Γκαουσιανή) συνάρτηση πυκνότητας πιθανότητας. Επιπλέον, επεκτείνοντας την τεχνική της εργασίας [139], χρησιμοποιείται ένα μοντέλο ανάδρασης για την επανεκτίμηση και την προσαρμογή των παραμέτρων της κατανομής από τα διαθέσιμα δεδομένα.

Παρά το γεγονός ότι σε πολλές περιπτώσεις ένας συνδυασμός Γκαουσιανών συναρτήσεων (mixture of Gaussians) προσεγγίζει καλύτερα την πραγματική κατανομή [102], τυχόν σφάλματα στην εκτίμηση της κατανομής αντισταθμίζονται με την προσαρμογή των παραμέτρων

της, ενώ το στάδιο επαλήθευσης μειώνει το ποσοστό λανθασμένης ανίχνευσης (false alarm) που μπορεί να προκύψει από ατέλειες του μοντέλου. Επιπλέον, η εκτίμηση τέτοιων συνδυασμών συναρτήσεων είναι εν γένει δυσκολότερη, ενώ έχει αναφερθεί ότι μέθοδοι που βασίζονται σε ιστογράμματα δέρματος επιτυγχάνουν μεγαλύτερο ποσοστό ανίχνευσης, αλλά η προσαρμογή τους έχει μεγαλύτερο υπολογιστικό κόστος [71,116].

Έστω ότι μ_0 και C είναι αντίστοιχα η μέση τιμή και ο πίνακας συµμεταβλητότητας των χρωματικών συνιστωσών Cr , Cb του ανθρώπινου δέρματος, διαστάσεων 2×1 και 2×2 αντίστοιχα. Η εκτίμηση των μ_0 και C πραγματοποιείται αρχικά από ένα σύνολο εικόνων εκπαίδευσης οι οποίες περιέχουν πρόσωπα διαφορετικών φυλών και με διαφορετικές συνθήκες φωτισμού. Τότε, σύμφωνα με το μοντέλο της κανονικής ή Γκαουσιανής κατανομής (normal / Gaussian distribution), το διάνυσμα $\mathbf{x} = [Cr \ Cb]^T$ των χρωματικών συνιστωσών ενός pixel ή μιας περιοχής (τμήματος) μιας εικόνας θεωρείται τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$P_s(\mathbf{x} | \mu_0, C) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_0)^T C^{-1}(\mathbf{x} - \mu_0)\right\}}{(2\pi)^{k/2} |C|^{1/2}} \quad (12-1)$$

όπου $k = 2$ είναι ο αριθμός χρωματικών συνιστωσών. Με δεδομένη την κανονική κατανομή και ένα pixel ή μία περιοχή εικόνας που περιγράφεται από το διάνυσμα \mathbf{x} , η πιθανοφάνεια (likelihood) να προέρχεται η περιοχή από ανθρώπινο πρόσωπο μπορεί να προσεγγισθεί από την ποσότητα

$$p_s(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_0)^T C^{-1}(\mathbf{x} - \mu_0)\right\} \quad (12-2)$$

Με άλλα λόγια, με κατάλληλη κατωφλίωση της συνάρτησης $p_s(\mathbf{x})$, μία περιοχή εικόνας μπορεί να ταξινομηθεί σε περιοχή δέρματος ή μη δέρματος. Παρά το γεγονός ότι το παραπάνω μοντέλο κανονικής κατανομής είναι επαρκές κάτω από ελεγχόμενες συνθήκες φωτισμού, καλύτερη επίδοση ταξινόμησης σε μεταβαλλόμενο περιβάλλον μπορεί να επιτευχθεί με την επανεκτίμηση (re-estimation) της μέσης τιμής μ_0 με βάση τα διαθέσιμα δεδομένα της εικόνας. Συγκεκριμένα, για την περίπτωση στατικών εικόνων, πραγματοποιείται αρχικά ένα πρώτο στάδιο ταξινόμησης με χρήση ενός κατωφλίου το οποίο υπολογίζεται από τα στατιστικά στοιχεία πιθανοφάνειας της παρούσας εικόνας, ώστε το μοντέλο να προσαρμόζεται στις μεταβαλλόμενες συνθήκες φωτισμού. Τότε οι περιοχές που ταξινομούνται ως περιοχές δέρματος χρησιμοποιούνται για την επανεκτίμηση της μ_0 σύμφωνα με τη σχέση

$$\mu_0 = (1 - m)\mu + m\mu_0 \quad (12-3)$$

όπου μ είναι το μέσο διάνυσμα χρωματικών συνιστωσών των περιοχών που ταξινομούνται ως περιοχές δέρματος και m είναι μία παράμετρος μνήμης που καθορίζει το μέγεθος αναπροσαρμογής του μοντέλου. Μία μεγάλη τιμή μνήμης δεν επιτρέπει στο μοντέλο να μεταβληθεί σημαντικά, ενώ αντίθετα μία μικρή τιμή αναπροσαρμόζει γρήγορα το μοντέλο στα διαθέσιμα δεδομένα. Η εκτίμηση της σταθεράς πραγματοποιείται με την ταξινόμηση μεγάλου αριθμού εικόνων και την καταγραφή του ποσοστού εσφαλμένων ανιχνεύσεων (false alarm) συναρτήσει της m . Τότε η m προσδιορίζεται με την κατάλληλη επιλογή ενός σημείου στην καμπύλη που σχηματίζεται (ROC curve) [116]. Τα πειράματα δείχνουν ότι μία κατάλληλη τιμή είναι $m = 0.7$. Σημειώνεται ότι η επανεκτίμηση του πίνακα συµμεταβλητότητας C είναι επίσης δυνατή· η επίδρασή του όμως στην επιτυχία της ταξινόμησης προκύπτει αμελητέα.

Στη συνέχεια ακολουθεί ένα *δεύτερο στάδιο ταξινόμησης* με χρήση του αναπροσαρμοσμένου μοντέλου σε συνδυασμό με ένα νέο κατώφλι που υπολογίζεται βάσει του κριτηρίου μέγιστης πιθανοφάνειας στο σύνολο εικόνων εκπαίδευσης. Στην περίπτωση που δεν ανιχνεύεται καμία περιοχή δέρματος, η αναπροσαρμογή της μ_0 θεωρείται άκυρη και η προηγούμενη τιμή της αποκαθίσταται. Η Εξίσωση (12-3) μπορεί να χρησιμοποιηθεί και σε πλήρως δυναμικό περιβάλλον. Συγκεκριμένα, στην περίπτωση των ακολουθιών εικόνων, το μοντέλο αναπροσαρμόζεται σε κάθε καρέ με βάση τα στατιστικά στοιχεία του προηγούμενου καρέ.

Η προτεινόμενη μέθοδος αναπροσαρμογής επιτυγχάνει την παρακολούθηση μεταβολών φωτεινότητας σε ακολουθίες βίντεο ενώ ταυτόχρονα αντιμετωπίζει την εμφάνιση προσώπων με νέα χαρακτηριστικά χρώματος, χωρίς απόκλιση από το γενικό μοντέλο του χρώματος δέρματος, διατηρώντας έτσι χαμηλό ποσοστό εσφαλμένων ανιχνεύσεων. Επιπλέον, σε ένα αλληλεπιδραστικό περιβάλλον αναζήτησης με βάση το περιεχόμενο, η επανεκτίμηση των παραμέτρων του μοντέλου μπορεί να βασισθεί στα τμήματα προσώπου που επιλέγονται από τον τελικό χρήστη [15], όπως περιγράφεται στην Ενότητα 14.3.

12.3 ΕΞΑΓΩΓΗ ΤΜΗΜΑΤΩΝ ΠΡΟΣΩΠΟΥ

Με χρήση του μοντέλου χρώματος δέρματος που παρουσιάστηκε, η ανίχνευση, εντοπισμός και εξαγωγή τμημάτων προσώπου από εικόνες πραγματοποιείται σε τρία στάδια. Αρχικά πραγματοποιείται κατάτμηση με βάση το χρώμα, ακολουθούμενη από συγχώνευση των τμημάτων που αντιστοιχούν σε δέρμα, και υπολογίζεται η πιθανοφάνεια δέρματος για κάθε τμήμα εικόνας. Στη συνέχεια το σχήμα κάθε τμήματος αναλύεται για να προσδιορισθεί κατά πόσο ταιριάζει με το ωοειδές σχήμα του προσώπου. Τέλος, στο στάδιο επαλήθευσης πραγματοποιείται ταίριασμα με πρότυπο προσώπου στο εσωτερικό των τμημάτων που έχουν ανιχνευθεί κατά τα δύο πρώτα στάδια.

Κατάτμηση με Βάση το Χρώμα

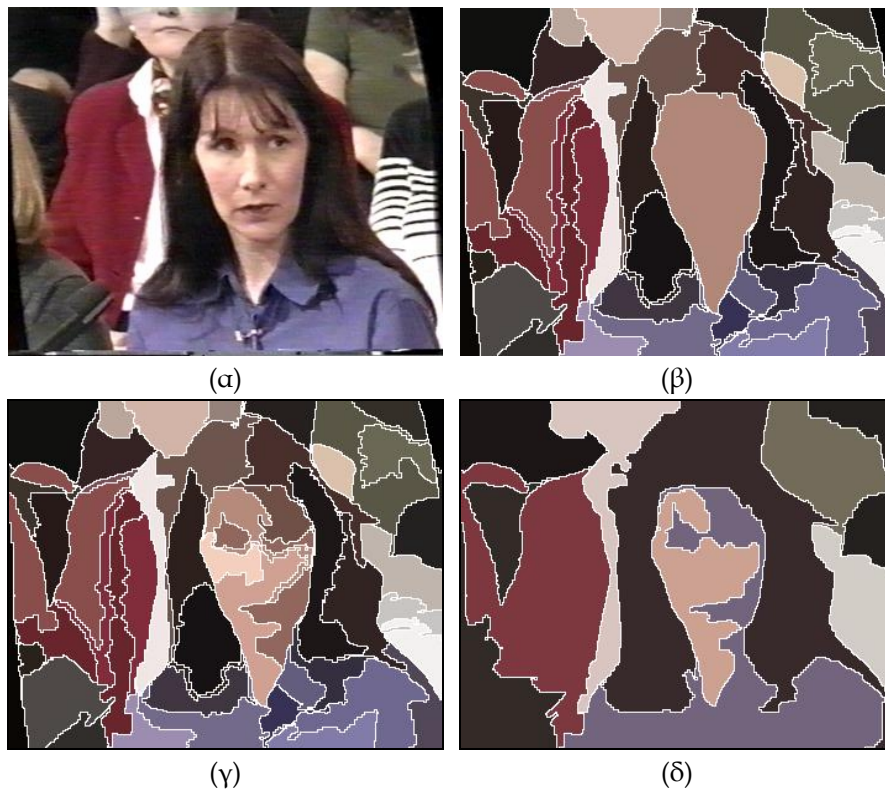
Για την κατάτμηση με βάση το χρώμα χρησιμοποιείται ο αλγόριθμος M-RSST που παρουσιάστηκε στην Ενότητα 1.2. Λόγω του ορισμού της απόστασης τμημάτων που δίνεται στην Εξίσωση (1-1) και της χαμηλής ανάλυσης της αρχικής εικόνας, τα μικρά τμήματα που αντιστοιχούν στις λεπτομέρειες ενός προσώπου συνήθως απαλείφονται και διατηρείται ένα μόνο τμήμα για ολόκληρη την επιφάνεια ενός προσώπου. Γι' αυτό το λόγο ο αλγόριθμος M-RSST έχει χρησιμοποιηθεί και αυτόνομα για την ανίχνευση προσώπων με βάση το χρώμα [15]. Όπως φανερώνεται στα πειράματα όμως, εξακολουθούν να υπάρχουν αρκετές περιπτώσεις – ιδιαίτερα για πρόσωπα σε μεγάλη κλίμακα – όπου ακόμη και η βέλτιστη επιλογή κατωφλίου για τον τερματισμό της κατάτμησης δεν μπορεί να δώσει ένα και μόνο τμήμα για όλη την επιφάνεια του προσώπου χωρίς να συγχωνεύσει το τμήμα αυτό με γειτονικές περιοχές.

Για τον παραπάνω λόγο η κατάτμηση συνδυάζεται με ένα επιπρόσθετο βήμα συγχώνευσης. Συγκεκριμένα, επιλέγεται αρχικά μία μικρή τιμή κατωφλίου τερματισμού, έτσι ώστε κάθε πρόσωπο να δίνει περισσότερα από ένα – τοπικά από δύο μέχρι έξι – τμήματα, αλλά να μη συγχωνεύεται με γειτονικές περιοχές άλλων αντικειμένων. Στη συνέχεια για κάθε τμήμα υπολογίζεται η *πιθανοφάνεια χρώματος δέρματος* ή *πιθανοφάνεια δέρματος* (skin-color likelihood) από την Εξίσωση (12-2), όπου ως x χρησιμοποιείται η μέση τιμή του διανύσματος χρωματικών συνιστωσών για όλη την επιφάνεια του τμήματος. Με βάση τις τιμές πιθανοφάνειας, πραγματο-

ποιείται μία δευτέρα φάση *συγχώνευσης τμημάτων δέρματος* (skin-color merging). Στη φάση αυτή το μέτρο απόστασης για τα βάρη συνδέσμων μεταξύ δύο τμημάτων X, Y της εικόνας ορίζεται ως

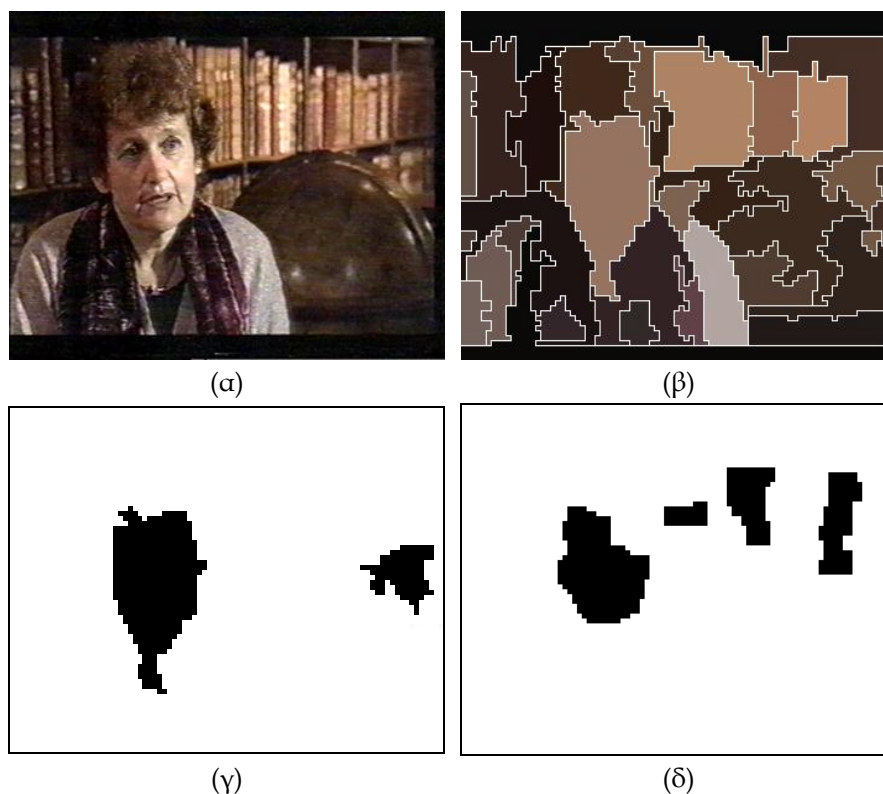
$$\delta_c(X, Y) = [\max(1 - p_s(X), 1 - p_s(Y))]^2 \quad (12-4)$$

όπου $p_s(X), p_s(Y)$ είναι η πιθανοφάνεια δέρματος των τμημάτων X και Y αντίστοιχα, δηλαδή αντί για το την (1-1) χρησιμοποιείται η (12-4) στον αλγόριθμο M-RSST. Έτσι, στη φάση αυτή η συγχώνευση τμημάτων δεν εξαρτάται από το χρώμα ή το μέγεθος των τμημάτων, αλλά μόνο από την πιθανοφάνεια να προέρχονται από περιοχή προσώπου. Συνεπώς όλα τα γειτονικά τμήματα προσώπου συγχωνεύονται σε ένα και μόνο τμήμα ενώ η υπόλοιπη κατάτμηση παραμένει ανεπηρέαστη. Τέλος για κάθε ένα από τα τελικά τμήματα υπολογίζεται εκ νέου η πιθανοφάνεια δέρματος.



Σχήμα 12-1. Κατάτμηση με βάση το χρώμα για ανίχνευση προσώπου. (α) Αρχική εικόνα. (β) Κατάτμηση με συγχώνευση τμημάτων δέρματος. (γ,δ) Κατάτμηση χωρίς συγχώνευση τμημάτων δέρματος, κατώφλι τερματισμού 2 και 22 αντίστοιχα.

Ενδεικτικά αποτελέσματα της μεθόδου εμφανίζονται στο Σχήμα 12-1. Στην αρχική εικόνα του Σχήματος 12-1α εφαρμόζεται η προτεινόμενη τεχνική κατάτμησης με βάση το χρώμα και συγχώνευσης τμημάτων δέρματος, δίνοντας την εικόνα 12-1β, όπου παρατηρείται ότι τα δύο πρόσωπα – ένα σε κοντινό πλάνο και ένα μικρό τμήμα στο φόντο – εξάγονται με μεγάλη ακρίβεια. Το αποτέλεσμα αυτό δεν θα ήταν εφικτό μέσω κατάτμησης μόνο. Συγκεκριμένα, μία μικρή κατωφλίου προκαλεί τη διάσπαση του κεντρικού τμήματος προσώπου σε πολλά τμήματα, όπως φαίνεται στο Σχήμα 12-1γ· στην πραγματικότητα η εικόνα του Σχήματος 12-1β υπολογίστηκε από την εικόνα του Σχήματος 12-1γ με συγχώνευση τμημάτων δέρματος. Από την άλλη μεριά, η προσπάθεια αύξησης του κατωφλίου προκαλεί τη συνένωση του δεξιού τμήματος του προσώπου με το υπόλοιπο σώμα, όπως φαίνεται στο Σχήμα 12-1δ.



Σχήμα 12-2. Σύγκριση προτεινόμενης μεθόδου κατάτμησης με την απευθείας κατωφλίωση. (α) Αρχική εικόνα. (β) Κατάτμηση με συγχώνευση τμημάτων δέρματος. (γ) Κατωφλίωση πιθανοφάνειας δέρματος των τμημάτων. (δ) Απευθείας κατωφλίωση των blocks της εικόνας με μορφολογικό closing.

Όπως προτείνεται στην εργασία [132], ανίχνευση προσώπου είναι δυνατή και χωρίς κατάτμηση, δηλαδή με απ' ευθείας κατωφλίωση στα blocks της εικόνας, σε συνδυασμό με κατάλληλη μορφολογική επεξεργασία στις δυαδικές μάσκες που προκύπτουν από την κατωφλίωση. Η προσέγγιση αυτή συγκρίνεται για μία ενδεικτική εικόνα με την προτεινόμενη τεχνική κατάτμησης στο Σχήμα 12-2. Για λόγους σύγκρισης, στο παράδειγμα αυτό η κατάτμηση και η συγχώνευση τμημάτων προσώπου εκτελούνται με υψηλότερο επίπεδο ανάλυσης ίσο με δύο, που αντιστοιχεί δηλαδή σε blocks μεγέθους 4×4 pixels. Το αποτέλεσμα της κατάτμησης για την αρχική εικόνα του Σχήματος 12-2α φαίνεται στο Σχήμα 12-2β, ενώ η δυαδική μάσκα που προκύπτει με κατωφλίωση της πιθανοφάνειας δέρματος φαίνεται στο Σχήμα 12-2γ. Αντίστοιχα, η απ' ευθείας κατωφλίωση σε 4×4 blocks της αρχικής εικόνας με μορφολογικό closing δίνει τη μάσκα του Σχήματος 12-2δ. Παρατηρείται ότι παρά το γεγονός ότι το τμήμα του προσώπου είναι παραμορφωμένο στην τελευταία αυτή εικόνα, άσχετα αντικείμενα όπως τμήματα της βιβλιοθήκης στο φόντο δεν μπορούν να αποφευχθούν.

Ανάλυση Σχήματος

Επειδή αντικείμενα ξένα προς τα ανθρώπινα πρόσωπα, των οποίων όμως οι χρωματικές συνιστώσες είναι παρόμοιες με εκείνες των προσώπων, εμφανίζονται συχνά σε εικόνες και ακολουθίες εικόνων, είναι απαραίτητο να ληφθούν υπόψη και άλλα χαρακτηριστικά πέρα από το χρώμα. Στο δεύτερο στάδιο λοιπόν της ανίχνευσης αναλύεται το σχήμα του κάθε αντικειμένου ώστε να απορριφθούν τμήματα των οποίων το σχήμα απέχει πολύ από το τυπικό ωοειδές ή ελλειψοειδές σχήμα του προσώπου. Ιδανικά, μπορεί να υπολογισθεί ο βαθμός ομοιότητας του σχήματος ενός τμήματος με μία έλλειψη. Το ταίριασμα σχημάτων που έχουν υποστεί αυ-

θαίρετες παραμορφώσεις είναι δυνατό με τη χρήση ενεργών περιγραμμάτων (active contours) ή παραμορφούμενων προτύπων (deformable templates) [19,69]. Επιπλέον, η ομοιόμορφη κίνηση αντικειμένων καθώς και οι affine μετασχηματισμοί όπως η μετατόπιση, η αλλαγή κλίμακας και η περιστροφή μπορούν να αφαιρεθούν από το περίγραμμα των αντικειμένων με χρήση κανονικοποίησης ή ποσοτήτων και μέτρων αναλλοίωτων στους μετασχηματισμούς αυτούς [16], όπως αναλύεται στο Μέρος III.

Στις περισσότερες πρακτικές περιπτώσεις όμως, τα περιγράμματα αντικειμένων που προκύπτουν από την κατάτμηση απέχουν πολύ από τα ιδανικά. Ακόμη και αν τα περιγράμματα απλοποιούνται ή εξομαλύνονται, για παράδειγμα με χρήση splines, το ταίριασμα περιγραμμάτων δεν δίνει ικανοποιητικά αποτελέσματα. Γι' αυτό το λόγο χρησιμοποιούνται μόνο δύο απλά χαρακτηριστικά του περιγράμματος. Συγκεκριμένα, για το περίγραμμα κάθε τμήματος X της εικόνας υπολογίζεται αρχικά το *συμπαγές* (compactness) $g(X)$, το οποίο ορίζεται συναρτήσει της περιμέτρου και της συνολικής επιφάνειας του τμήματος ως εξής:

$$g(X) = \frac{1}{4\pi} \frac{[r(X)]^2}{a(X)} \quad (12-5)$$

όπου το $r(X)$ συμβολίζει την περίμετρο (πλήθος σημείων περιγράμματος) και το $a(X)$ την επιφάνεια (πλήθος pixels) του τμήματος X . Σημειώνεται ότι μέγιστη τιμή συμπαγούς είναι ίση με 1 και επιτυγχάνεται για το σχήμα του κύκλου, συνεπώς το $g(X)$ είναι πάντα κανονικοποιημένο στο διάστημα $[0,1]$. Αυτό βέβαια ισχύει μόνο σε έναν Ευκλείδειο χώρο και ο διακριτός χώρος εικόνας στον οποίο υπολογίζονται οι ποσότητες $r(X)$ και $a(X)$ δεν είναι τέτοιος χώρος. Αυτό σημαίνει ότι το $g(X)$ όπως ορίζεται στην Εξίσωση (12-5) μπορεί σε κάποιες περιπτώσεις να πάρει τιμές μεγαλύτερες από 1, αλλά αυτό δεν αποτελεί πρόβλημα γιατί το συμπαγές κανονικοποιείται περαιτέρω με χρήση μιας μη γραμμικής συνάρτησης όπως περιγράφεται παρακάτω.

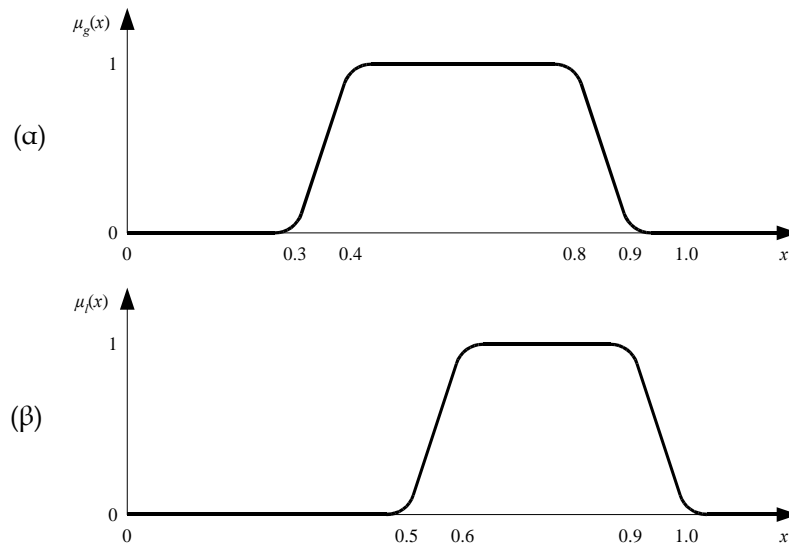
Στη συνέχεια υπολογίζεται η *επιμήκυνση* (elongation) με χρήση του μετασχηματισμού *Houghling* ή διακριτού μετασχηματισμού *Karhunen-Loeve*. Έστω τα διανύσματα \mathbf{x} και \mathbf{y} , διαστάσεων $N \times 1$, περιέχουν τις συντεταγμένες των N σημείων του κλειστού περιγράμματος του τμήματος X . Τότε ο πίνακας συμμεταβλητότητας \mathbf{C}_X των σημείων αυτών, διαστάσεων $N \times 1$, δίνεται από τη σχέση

$$\mathbf{C}(X) = \frac{1}{N} [\mathbf{x} - \mu_x \mathbf{e} \quad \mathbf{y} - \mu_y \mathbf{e}]^T [\mathbf{x} - \mu_x \mathbf{e} \quad \mathbf{y} - \mu_y \mathbf{e}] \quad (12-6)$$

όπου (μ_x, μ_y) είναι το κέντρο βάρους του τμήματος και \mathbf{e} είναι το $N \times 1$ διάνυσμα $[1 \ 1 \ \dots \ 1]^T$. Τα δύο *ιδιοδιανύσματα* (eigenvectors) του πίνακα συμμεταβλητότητας εκφράζουν τους *κύριους άξονες* (principal axes) του σχήματος, και συγκεκριμένα το *ιδιοδιάνυσμα* που αντιστοιχεί στη μέγιστη *ιδιοτιμή* (eigenvalue) εκφράζει τον *πρωτεύοντα* ή *μείζονα άξονα* (major axis), ενώ εκείνο που αντιστοιχεί στην ελάχιστη ιδιοτιμή εκφράζει τον *δευτερεύοντα* ή *ελάσσονα άξονα* (minor axis). Επιπλέον, ο λόγος των δύο ιδιοτιμών καθορίζει την επιμήκυνση του περιγράμματος του τμήματος X :

$$l(X) = \sqrt{\lambda_2(\mathbf{C}(X)) / \lambda_1(\mathbf{C}(X))} \quad (12-7)$$

όπου $\lambda_1(\mathbf{C}(X))$, $\lambda_2(\mathbf{C}(X))$ είναι αντίστοιχα η μέγιστη και ελάχιστη τιμή του $\mathbf{C}(X)$. Τα παραπάνω ολικά (global) χαρακτηριστικά περιγράμματος είναι ανθεκτικά στο θόρυβο και τις παραμορφώσεις, καθώς και *αναλλοίωτα* στη μετατόπιση, την αλλαγή κλίμακας και την περιστροφή.



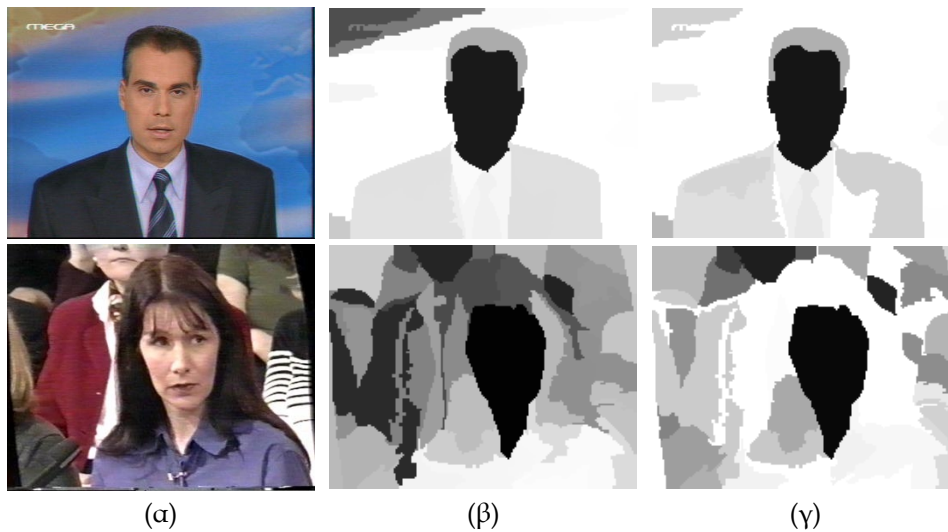
Σχήμα 12-3. Μη γραμμικές συναρτήσεις συμμετοχής για το μετασχηματισμό των χαρακτηριστικών σχήματος. (α) Συμπαγές (compactness). (β) Επιμήκυνση (elongation).

Τα πειράματα δείχνουν ότι οι τυπικές τιμές που αντιστοιχούν σε πραγματικά τμήματα προσώπου κυμαίνονται από 0.44 έως 0.79 για το συμπαγές και από 0.59 έως 0.91 για την επιμήκυνση. Συνεπώς, το ταίριασμα ενός περιγράμματος με περίγραμμα προσώπου επιτυγχάνεται με το μετασχηματισμό του συμπαγούς και της επιμήκυνσης μέσω κατάλληλων μη γραμμικών συναρτήσεων που παίρνουν τιμές στο διάστημα $[0,1]$, όμοια τις *ασαφείς συναρτήσεις συμμετοχής*: $g'(X) = \mu_g(g(X))$, $l'(X) = \mu_l(l(X))$. Οι μη γραμμικές συναρτήσεις που χρησιμοποιούνται στα πειράματα φαίνονται στο Σχήμα 12-3. Το μετασχηματισμένο συμπαγές / επιμήκυνση είναι λοιπόν πάντα κανονικοποιημένο στο διάστημα $[0,1]$ και ουσιαστικά εκφράζει την ομοιότητα με το συμπαγές / την επιμήκυνση ενός τυπικού περιγράμματος προσώπου.

Τέλος, το μετασχηματισμένο συμπαγές $g'(X)$ και επιμήκυνση $l'(X)$ του τμήματος X συνδυάζονται με την πιθανοφάνεια δέρματος $p_s(X)$ που προκύπτει από την Εξίσωση (12-2), μέσω ενός *γεωμετρικού μέσου με βάρη* (weighted geometric mean). Έτσι προκύπτει μία συνολική *πιθανοφάνεια προσώπου* (face likelihood) του τμήματος X , η οποία συμβολίζεται με $p_f(X)$. Στην πράξη, τα βάρη εκτιμώνται εμπειρικά με βάση τα πειράματα και εξαρτώνται από τις τιμές συμπαγούς και επιμήκυνσης. Συγκεκριμένα, αφού η πιθανοφάνεια δέρματος $p_s(X)$ είναι συνήθως πιο αξιόπιστη για την ανίχνευση προσώπου – το χρώμα του δέρματος χαρακτηρίζει πολύ περισσότερο ένα ανθρώπινο πρόσωπο σε σχέση με το σχήμα του – ανατίθεται σε αυτό μεγαλύτερο βάρος, εκτός αν τα $g'(X)$ και $l'(X)$ παίρνουν τιμές κοντά στο μηδέν. Αυτό σημαίνει ότι τα χαρακτηριστικά σχήματος χρησιμοποιούνται ουσιαστικά μόνο για την απόρριψη τμημάτων με πολύ ακανόνιστο σχήμα των οποίων το χρώμα ταιριάζει με το μοντέλο χρώματος δέρματος.

Ενδεικτικά αποτελέσματα της προτεινόμενης ανάλυσης παρουσιάζονται στα Σχήματα 12-4, 12-5 και 12-6 για μια ποικιλία ανάλυσης, ποιότητας και contrast εικόνας καθώς και πλήθους προσώπων και συνθηκών φωτισμού. Συγκεκριμένα, στο Σχήμα 12-4 παρουσιάζονται δύο εικόνες με ένα πρόσωπο σε κοντινό πλάνο. Η δεύτερη περιέχει επίσης ένα δεύτερο τμήμα προσώπου καθώς και ένα τμήμα χεριού. Στο Σχήμα 12-5 δύο εικόνες με δύο ή τρία πρόσωπα σε μακρινό πλάνο, δηλαδή σε μικρότερη κλίμακα από το Σχήμα 12-4. Η δεύτερη απ' αυτές έχει ιδιαίτερα έντονο φωτισμό. Τέλος, στο Σχήμα 12-6 εμφανίζονται δύο εικόνες με περισσότερα

από ένα πρόσωπο από τα οποία άλλα είναι σε μικρή και άλλα σε μεγάλη κλίμακα, ενώ οι γωνίες των προσώπων ποικίλουν επίσης. Οι τελευταίες εικόνες προέρχονται από εξωτερική λήψη και η πρώτη έχει πολύ χαμηλό contrast, ενώ η δεύτερη πολύ έντονο.

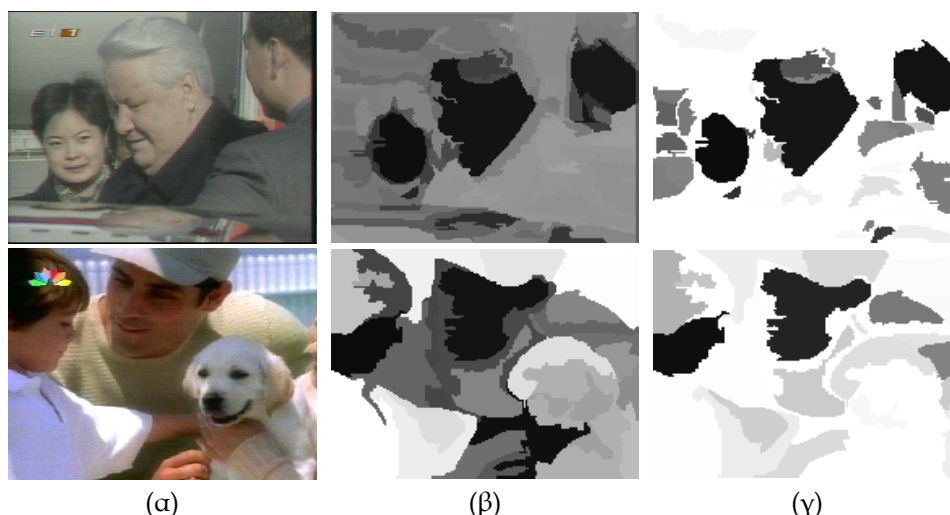


Σχήμα 12-4. Ανίχνευση προσώπου σε εικόνες με ένα πρόσωπο σε κοντινό πλάνο. (α) Αρχική εικόνα. (β) Εικόνα πιθανοφάνειας δέρματος. (γ) Εικόνα πιθανοφάνειας προσώπου.



Σχήμα 12-5. Ανίχνευση προσώπου σε εικόνες με περισσότερα από ένα πρόσωπα σε μακρινό πλάνο. (α) Αρχική εικόνα. (β) Εικόνα πιθανοφάνειας δέρματος. (γ) Εικόνα πιθανοφάνειας προσώπου.

Σε όλα τα παραπάνω Σχήματα, στη στήλη (α) φαίνεται η αρχική εικόνα, στη στήλη (β) η εικόνα πιθανοφάνειας δέρματος μετά την κατάτμηση και τη συγχώνευση τμημάτων προσώπου, και στη στήλη (γ) η εικόνα πιθανοφάνειας προσώπου, μετά δηλαδή την ανάλυση σχήματος. Οι εικόνες πιθανοφάνειας σχηματίζονται αναθέτοντας σε κάθε τμήμα εικόνας ένα επίπεδο γκρι έτσι ώστε η τιμή 0 να αντιστοιχεί σε λευκό χρώμα και η τιμή 1 σε μαύρο. Παρατηρείται ότι σε όλες τις περιπτώσεις η εικόνα πιθανοφάνειας προσώπου υπερέχει της εικόνας πιθανοφάνειας δέρματος όσον αφορά στην ανίχνευση προσώπων. Αυτό είναι αναμενόμενο καθώς τα τμήματα με ακανόνιστο σχήμα απορρίπτονται ακόμη και αν το χρώμα ταιριάζει με το μοντέλο χρώματος δέρματος. Παρατηρείται επίσης ότι λόγω της επανεκτίμησης και αναπροσαρμογής του μοντέλου, τα τμήματα δέρματος ανιχνεύονται σωστά παρά τις έντονες διαφορές στις συνθήκες φωτισμού.



Σχήμα 12-6. Ανίχνευση προσώπου σε εικόνες με περισσότερα από ένα πρόσωπα, σε διαφορετικές κλίμακες και διαφορετικές συνθήκες φωτισμού. (α) Αρχική εικόνα. (β) Εικόνα πιθανοφάνειας δέρματος. (γ) Εικόνα πιθανοφάνειας προσώπου.

Τα πρόσωπα που ανιχνεύονται από την εικόνα πιθανοφάνειας προσώπου ποικίλουν από ιδανικά έως οριακά αναγνωρίσιμα, ανάλογα με την ποιότητα και την πολυπλοκότητα της αρχικής εικόνας. Παρά το γεγονός ότι το σχήμα των τμημάτων προσώπου είναι σε πολλές περιπτώσεις έντονα παραμορφωμένο, σε όλα τα πρόσωπα αντιστοιχεί από ένα μόνο τμήμα εικόνας και οι απώλειες (dismissals) προσώπων είναι ελάχιστες. Παρατηρείται όμως ότι διατηρούνται πολλά τμήματα εικόνας, κυρίως τμήματα δέρματος, τα οποία δεν περιέχουν πρόσωπο. Για την αφαίρεση των τμημάτων αυτών τα χαρακτηριστικά χρώματος και σχήματος δεν επαρκούν, γι' αυτό εξετάζεται περαιτέρω και η υφή τους, όπως περιγράφεται παρακάτω.

Ταίριασμα με Πρότυπο

Σε κάθε τμήμα εικόνας υπολογίζεται η πιθανοφάνεια προσώπου με βάση το μοντέλο χρώματος δέρματος και την ανάλυση σχήματος· αν η τιμή της πιθανοφάνειας προσώπου ξεπερνά ένα κατάλληλα επιλεγμένο κατώφλι τότε το τμήμα σημειώνεται ως ένα πιθανό τμήμα προσώπου και εξάγεται η ορθογώνια περιοχή που περιέχει το τμήμα (*bounding rectangle*). Κατά τη φάση *επαλήθευσης*, στο εσωτερικό της περιοχής αυτής πραγματοποιείται *ταίριασμα με πρότυπο* (*template matching*). Σκοπός της τεχνικής αυτής είναι να προσδιορισθεί αν υπάρχει ανθρώπινο πρόσωπο στην περιοχή και αν ναι να υπολογισθεί η θέση, η κλίμακα και ο προσανατολισμός του. Περισσότερες λεπτομέρειες υλοποίησης της τεχνικής περιλαμβάνονται στην εργασία [132].

Πρόκειται για μία διαδικασία με υψηλό κόστος διότι περιλαμβάνει εξαντλητική έρευνα για όλους τους συνδυασμούς των παραμέτρων (θέσης, κλίμακας και προσανατολισμού). Όμως η επιβάρυνση στη συνολική τεχνική ανίχνευσης προσώπων δεν είναι μεγάλη αφού η έρευνα πραγματοποιείται μόνο στο εσωτερικό των περιοχών που έχουν εντοπισθεί στα προηγούμενα στάδια και που συνήθως αντιπροσωπεύουν μικρό ποσοστό της συνολικής εικόνας. Επιπλέον, στην περίπτωση των ακολουθιών εικόνων, η χρονική συσχέτιση των διαδοχικών καρέ λαμβάνεται υπόψη για την επιτάχυνση της διαδικασίας. Η πληροφορία για τη θέση, κλίμακα και προσανατολισμό των προσώπων που ανιχνεύονται σε κάθε καρέ χρησιμοποιούνται για τον προσδιορισμό του αρχικού σημείου έρευνας στο επόμενο καρέ· έτσι ο απαιτούμενος χρόνος επαλήθευσης σε βίντεο μειώνεται δραστικά.

Ενδεικτικά αποτελέσματα επαλήθευσης περιλαμβάνονται στο Σχήμα 12-7. Συγκεκριμένα, στο Σχήμα 12-7α φαίνεται μία εικόνα με πολλά πρόσωπα και τμήματα δέρματος που δεν προέρχονται από πρόσωπο. Τα εξωτερικά πλαίσια στο σχήμα αυτό δείχνουν τις περιοχές της εικόνας που ανιχνεύονται από το μοντέλο χρώματος δέρματος και την ανάλυση σχήματος, ενώ τα εσωτερικά δείχνουν την ακριβή θέση προσώπου που προσδιορίζεται μέσω του ταιριάσματος με πρότυπο. Παρατηρείται ότι από τα τρία πιθανά τμήματα προσώπου μόνο τα δύο διατηρούνται, ενώ το τμήμα χεριού απορρίπτεται. Παράλληλα η θέση του προσώπου προσδιορίζεται με μεγαλύτερη ακρίβεια. Όμοια, στο Σχήμα 12-7β φαίνεται μία εικόνα που περιέχει ένα μόνο πρόσωπο, το οποίο όμως καλύπτει μια μικρή περιοχή του αντίστοιχου τμήματος δέρματος. Και πάλι στο στάδιο επαλήθευσης εξάγεται η σωστή περιοχή του προσώπου.



Σχήμα 12-7. Ανίχνευση και επαλήθευση πολλαπλών προσώπων. Τα εξωτερικά πλαίσια προκύπτουν από την ανίχνευση δέρματος και την ανάλυση σχήματος· τα εσωτερικά από ταιρίασμα με πρότυπο προσώπου.
(α) Εικόνα με πολλά πρόσωπα και τμήματα δέρματος που δεν προέρχονται από πρόσωπο (β) Εικόνα με ένα πρόσωπο που καλύπτει μικρή έκταση του αντίστοιχου τμήματος δέρματος.

Σημειώνεται ότι η τεχνική ταιριάσματος με πρότυπο έχει χρησιμοποιηθεί και αυτόνομα για ανίχνευση και εντοπισμό προσώπων [133], και ιδιαίτερα σε ασπρόμαυρες εικόνες όπου δεν υπάρχει πληροφορία χρώματος και επομένως η τεχνική ανίχνευσης δέρματος δε μπορεί να εφαρμοσθεί. Το κύριο μειονέκτημά της είναι η υπολογιστική πολυπλοκότητα, αφού στην περίπτωση αυτή απαιτείται έρευνα σε όλη την επιφάνεια της εικόνας. Επιπλέον, για τον εντοπισμό περισσότερων από ένα πρόσωπα, το πλήθος των προσώπων πρέπει να είναι εκ των προτέρων γνωστό, αλλιώς απαιτείται αυθαίρετη ενός κατωφλίου που μπορεί να οδηγήσει σε αναξιόπιστα αποτελέσματα. Παρά τα μειονεκτήματα, η τεχνική αυτή εξακολουθεί να είναι χρήσιμη αφού υπάρχουν μεγάλες ποσότητες οπτικοακουστικού υλικού στο οποίο οι εικόνες είναι ασπρόμαυρες, ιδιαίτερα σε περιπτώσεις ιστορικών αρχείων.

12.4 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στην παρούσα Ενότητα δίνονται πειραματικά αποτελέσματα για την αξιολόγηση των προτεινόμενων τεχνικών. Συγκεκριμένα, περιλαμβάνονται δύο ομάδες πειραμάτων. Η πρώτη αφορά στην ανίχνευση δέρματος και περιλαμβάνει την κατάτμηση με βάση το χρώμα, τη συγχώνευση τμημάτων δέρματος και τον υπολογισμό και κατωφλίωση της πιθανοφάνειας δέρματος. Η δεύτερη αφορά στην ανίχνευση και στον εντοπισμό προσώπων και περιλαμβάνει επιπλέον την ανάλυση σχήματος και την επαλήθευση μέσω ταιριάσματος με πρότυπο.

Πειράματα Ανίχνευσης Δέρματος

Για την αξιολόγηση της ανίχνευσης δέρματος χρησιμοποιείται υποσύνολο της βάσης δεδομένων του Compaq Cambridge Research Laboratory (CRL), η οποία παρουσιάζεται στην εργασία [71]. Η βάση αποτελείται από 4670 *εικόνες δέρματος* (skin images) που περιέχουν πρόσωπα και τμήματα δέρματος, και 8964 *γενικές εικόνες* (non-skin images) που δεν περιέχουν κανένα τμήμα δέρματος. Όλες οι εικόνες προέρχονται από παράλληλη έρευνα στο διαδίκτυο και είναι διαθέσιμες σε ανάλυση που κυμαίνεται από πολύ χαμηλή (50×50 pixels) έως σχετικά υψηλή (800×600 pixels), ενώ η ανάλυση χρώματος είναι για όλες τις εικόνες είναι 24bpp. Στις εικόνες δέρματος, όλα τα τμήματα δέρματος έχουν εντοπισθεί χειρονακτικά, με τη βοήθεια ενός εργαλείου που επιτρέπει τον ανεξάρτητο έλεγχο και ρύθμιση ενός αλγορίθμου κατάτμησης για κάθε εικόνα. Οι *δυναδικές μάσκες* που περιέχουν τις θέσεις των τμημάτων δέρματος περιλαμβάνονται επίσης στη βάση.

Για τις ανάγκες των πειραμάτων επιλέχθηκαν τυχαία 213 εικόνες δέρματος μαζί με τις δυναδικές μάσκες τους, καθώς και 107 γενικές εικόνες. Το προτεινόμενο μοντέλο χρώματος δέρματος σε συνδυασμό με τη μέθοδο κατάτμησης και συγχώνευσης τμημάτων δέρματος εφαρμόστηκε στο σύνολο των 320 εικόνων. Με κατωφλίωση της πιθανοφάνειας δέρματος κάθε τμήματος προκύπτει μια δυναδική μάσκα για κάθε εικόνα. Θεωρώντας τις μάσκες της βάσης ως «αληθείς» (ground truth) αξιολογήσαμε τα αποτελέσματα μετρήσεις *ακρίβειας* (precision) και *ανάκλησης* (recall) που παρουσιάζονται στον Πίνακα 12-1.

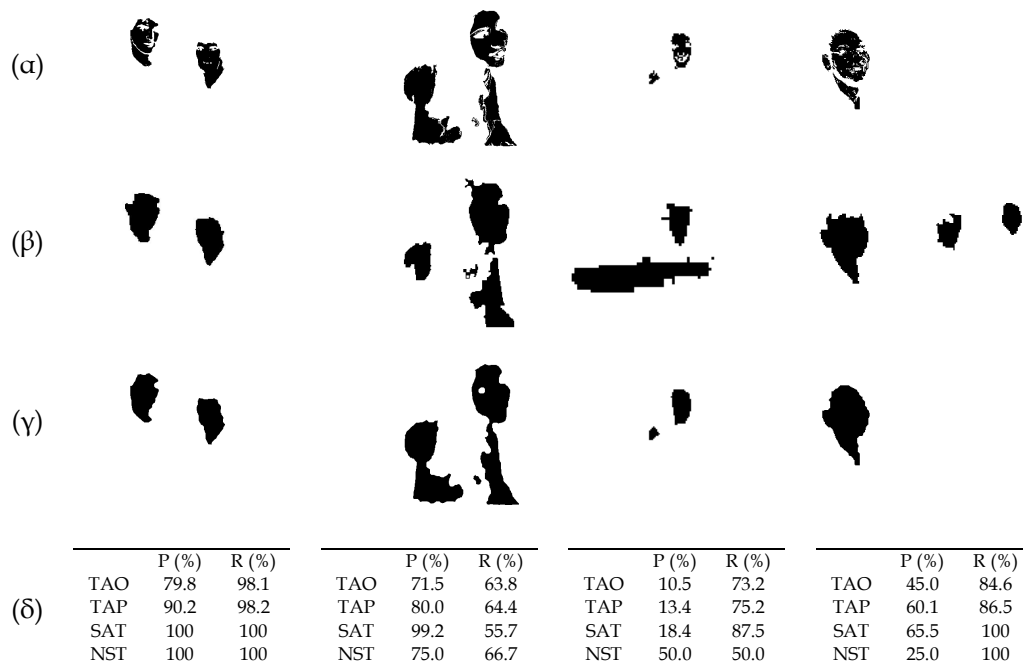
Μέτρηση	P (%)	R (%)
TAO	66.42	85.69
TAP	77.82	85.36
SAT	87.61	93.57
NST	85.77	88.46

Πίνακας 12-1. Αξιολόγηση ανίχνευσης δέρματος με βάση τις αληθείς μάσκες που περιλαμβάνονται στη βάση δεδομένων CRL. P: Precision (ακρίβεια). R: Recall (ανάκληση). Τα σύμβολα TAO, TAP, SAT και NST εξηγούνται στο κείμενο.

Στην πραγματικότητα δίνονται τέσσερις τύποι μετρήσεων στον Πίνακα αυτό. Στον πρώτο τύπο μετρήσεων, η ακρίβεια μετράται ως η μέση τιμή του λόγου της κοινής επιφάνειας που καλύπτεται και από τις δύο μάσκες (εκτιμώμενη και αληθής) προς την επιφάνεια της εκτιμώμενης μάσκας, ενώ η για την ανάκληση χρησιμοποιείται ο λόγος της κοινής επιφάνειας προς την επιφάνεια της αληθούς μάσκας. Η μέτρηση αυτή συμβολίζεται ως TAO (Total Area, Original) στον Πίνακα 12-1. Στον δεύτερο τύπο μετρήσεων, που συμβολίζεται ως TAP (Total Area, after Processing), οι αληθείς μάσκες υφίστανται πρώτα επεξεργασία μορφολογικού closing. Παρατηρείται ότι η ακρίβεια βελτιώνεται σημαντικά αφού οι αρχικές αληθείς μάσκες δεν περιέχουν λεπτομέρειες του προσώπου όπως τα μάτια, τα μαλλιά και το στόμα.

Στον τρίτο τύπο μετρήσεων, που συμβολίζεται ως SAT (Segment Area, with Thresholding), οι μάσκες χωρίζονται πρώτα σε ανεξάρτητα τμήματα δέρματος. Ξεκινώντας από μία εκτιμώμενη μάσκα, κάθε τμήμα της θεωρείται «σωστό» αν το 55% τουλάχιστον της επιφάνειάς του καλύπτεται από τμήματα δέρματος της αληθούς μάσκας. Η αντικατάσταση της κοινής επιφάνειας με τη συνολική επιφάνεια «σωστών» τμημάτων δίνει μια νέα μέτρηση ακρίβειας, ενώ η ανάκληση υπολογίζεται με μία ανάλογη διαδικασία ξεκινώντας από τις αληθείς μάσκες. Λόγω της χρησιμοποιούμενης κατωφλίωσης, η ακρίβεια και η ανάκληση είναι και οι δύο υψηλότερες. Επιπλέον, οι μετρήσεις αυτές είναι πιο αξιόπιστες αφού ένα τμήμα

δέρματος θα έπρεπε να θεωρείται σωστό ακόμη και όταν δεν ταιριάζει απόλυτα με το αληθές, και λανθασμένο ακόμη και όταν έχει μια μικρή επικάλυψη με το αληθές. Τέλος, η ακρίβεια και η ανάκληση μετρούνται με βάση τον αριθμό τμημάτων δέρματος αντί για τις αντίστοιχες επιφάνειες. Αυτός ο τύπος μετρήσεων συμβολίζεται ως NST (Number of Segments, with Thresholding) στον Πίνακα 12-1. Οι μετρήσεις σε αυτή την περίπτωση είναι ελαφρώς μικρότερες αφού η απώλεια ενός πολύ μικρού τμήματος είναι τώρα ισοδύναμη με την απώλεια ενός μεγάλου.



Σχήμα 12-8. Χαρακτηριστικά παραδείγματα ανίχνευσης δέρματος σε εικόνες της βάσης CRL. (α) Αληθείς μάσκες. (β) Εκτιμώμενες μάσκες. (γ) Αληθείς μάσκες μετά από μορφολογικό closing. (δ) Μετρήσεις ακρίβειας / ανάκλησης.

Στο Σχήμα 12-8 δίνονται κάποια χαρακτηριστικά παραδείγματα για να γίνουν οι παραπάνω μετρήσεις πιο κατανοητές. Στο Σχήμα περιλαμβάνονται μόνο μάσκες· για τις αρχικές εικόνες δεν υπάρχει άδεια δημοσίευσης λόγω ζητημάτων πνευματικής ιδιοκτησίας. Η πρώτη στήλη του Σχήματος 12-8 παρουσιάζει μία τυπική περίπτωση επιτυχίας. Παρόλο που αναμένεται υψηλή τιμή ακρίβειας, η μέτρηση TAO είναι μόνο 79.76% λόγω των περιοχών των ματιών και της μύτης που δεν περιλαμβάνονται στην αρχική αληθή μάσκα. Αντίθετα, μετά το μορφολογικό closing (μέτρηση TAP) η ακρίβεια αυξάνεται στο 90.23%. Για τις μετρήσεις SAT και NST παρουσιάζεται απόλυτη επιτυχία, αφού και τα δύο τμήματα προσώπου έχουν ανιχνευθεί.

Στη δεύτερη στήλη του Σχήματος 12-8 παρουσιάζεται μία ακραία περίπτωση χαμηλής ανάκλησης· σχεδόν ένα ολόκληρο τμήμα χεριού χάνεται λόγω των συνθηκών φωτισμού (χρωματισμένη πηγή). Η μέτρηση SAT για την ανάκληση είναι η χαμηλότερη, δείχνοντας ότι η διαδικασία κατωφλίωσης δεν ωφελεί πάντα τις μετρήσεις. Όμοια, στην τρίτη στήλη φαίνεται η χειρότερη περίπτωση όσον αφορά στην ακρίβεια. Ένα μεγάλο αντικείμενο ταξινομείται εσφαλμένα ως τμήμα δέρματος· τέτοιες περιπτώσεις είναι αναπόφευκτες και αναδεικνύουν τη σημασία του σταδίου επαλήθευσης. Τέλος, στην τέταρτη στήλη εμφανίζεται μια περίπτωση όπου η αληθής μάσκα δεν περιλαμβάνει τρία τμήματα δέρματος – δύο πρόσωπα και ένα χέρι

– που υπάρχουν στην αρχική εικόνα. Στην περίπτωση αυτή όλες οι μετρήσεις είναι προφανώς εσφαλμένες.

Πειράματα Ανίχνευσης Προσώπων

Για την αξιολόγηση της ανίχνευσης και εντοπισμού προσώπων, που περιλαμβάνει τα προηγούμενα στάδια ανάλυσης καθώς και την ανάλυση σχήματος και το στάδιο επαλήθευσης (ταίριασμα με πρότυπο), πραγματοποιήθηκαν δύο πειράματα: ένα σε ακίνητες εικόνες και ένα σε ακολουθίες εικόνων. Και στις δύο περιπτώσεις η αξιολόγηση έγινε με μετρήσεις ακρίβειας και ανάκλησης, όπως και για την ανίχνευση δέρματος. Σε όλες τις διαθέσιμες εικόνες (ή καρέ) τα «αληθή» (ground truth) στοιχεία των προσώπων είτε υπήρχαν στις βάσεις δεδομένων που χρησιμοποιήθηκαν είτε προσδιορίστηκαν χειρονακτικά. Έτσι η ακρίβεια υπολογίστηκε ως ο λόγος των προσώπων που εντοπίστηκαν σωστά προς το συνολικό αριθμό προσώπων που εντοπίστηκαν, ενώ η ανάκληση ως ο λόγος των προσώπων που εντοπίστηκαν σωστά προς το συνολικό αριθμό αληθών προσώπων. Παρατηρείται ότι η ακρίβεια είναι η αντίθετη έννοια του ποσοστού *λανθασμένης ανίχνευσης* (false alarm) και η ανάκληση η αντίθετη έννοια του ποσοστού *απόλειψης* (dismissal).

Μέτρηση	Εικόνες	Ακολουθίες
Συνολικές εικόνες / καρέ	420	701
Εικόνες με πρόσωπα	305	643
Συνολικά πρόσωπα	409	736
Εντοπισμένα πρόσωπα	379	681
Σωστά εντοπισμένα πρόσωπα	368	667
Ακρίβεια (%)	97.10	97.94
Ανάκληση (%)	89.98	90.63
MET (sec)	1.9	1.5

Πίνακας 12-2. Αξιολόγηση ανίχνευσης προσώπων σε έγχρωμες εικόνες και ακολουθίες εικόνων. MET: Mean Execution Time (μέσος χρόνος εκτέλεσης).

Το οπτικό υλικό του πειράματος στατικών εικόνων περιλαμβάνει 100 καρέ από τηλεοπτικά προγράμματα σε ανάλυση CIF, καθώς και τις 320 εικόνες της βάσης Compaq CRL που χρησιμοποιήθηκαν και στο πείραμα ανίχνευσης δέρματος. Τα αληθή στοιχεία προσώπων προσδιορίστηκαν χειρονακτικά για τις 100 εικόνες τηλεοπτικών προγραμμάτων, ενώ για τις υπόλοιπες 320 εικόνες απλώς επιλέχθηκαν οι περιοχές προσώπων από τις μάσκες δέρματος που είναι διαθέσιμες στη βάση CRL. Από το σύνολο των 420 εικόνων κάποιες δεν περιέχουν κανένα πρόσωπο ενώ άλλες περιέχουν ένα ή και περισσότερα πρόσωπα· συνολικά σε όλες τις εικόνες περιέχονται 409 πρόσωπα. Τα αποτελέσματα ανίχνευσης προσώπων παρουσιάζονται στην πρώτη γραμμή του Πίνακα 12-2: ανιχνεύονται 379 πρόσωπα από τα οποία 368 είναι σωστά. Η κύρια αιτία απωλειών είναι η αποτυχία ανίχνευσης όλων των πιθανών τμημάτων δέρματος, ενώ οι λανθασμένες ανιχνεύσεις (false alarms) οφείλονται κυρίως στο στάδιο επαλήθευσης. Παρατηρείται ότι λόγω του συνδυασμού του μοντέλου χρώματος δέρματος και του ταίριασματος με πρότυπο επιτυγχάνεται πολύ υψηλό ποσοστό ακρίβειας· η ανάκληση είναι χαμηλότερη λόγω ατελειών του χρωματικού μοντέλου.

Στο πείραμα ακολουθιών εικόνων το οπτικό υλικό αποτελείται κυρίως από 20 ακολουθίες μήκους 30 καρέ η κάθε μια, οι οποίες καταγράφηκαν από προγράμματα τηλεοπτικών ειδήσεων. Περιέχονται επίσης δύο ακολουθίες της βάσης Boston University IVC [116] από κινημα-

τογραφικές ταινίες, και συγκεκριμένα οι ακολουθίες #3 (μήκους 52 καρέ) και #12 (μήκους 49 καρέ). Και πάλι τα αληθή στοιχεία για τα τηλεοπτικά προγράμματα υπολογίσθηκαν χειρονακτικά, ενώ για τις ακολουθίες της βάσης IVC επιλέχθηκαν τα τμήματα προσώπου από τις μάσκες δέρματος που ήταν διαθέσιμες στη βάση. Τα αποτελέσματα ανίχνευσης για το σύνολο των 701 καρέ που περιέχουν 736 πρόσωπα παρουσιάζονται στη δεύτερη γραμμή του Πίνακα 12-2. Οι μετρήσεις ακρίβειας και ανάκλησης είναι κοντά σε εκείνες του πειράματος στατικών εικόνων. Ο μέσος χρόνος εκτέλεσης ανά καρέ όμως μειώνεται λόγω της εκμετάλλευσης της χρονικής συσχέτισης των διαδοχικών καρέ. Πλήρης έρευνα στο στάδιο επαλήθευσης πραγματοποιείται μόνο στο πρώτο καρέ κάθε ακολουθίας· στα υπόλοιπα η θέση κάθε προσώπου ενός καρέ χρησιμοποιείται για τον προσδιορισμό του αρχικού σημείου έρευνας του επόμενου καρέ.

Συντακτική Ανάλυση Δελτίων

Ο αξιόπιστος εντοπισμός ανθρώπινων προσώπων σε συνδυασμό με άλλα χαρακτηριστικά χρώματος και κίνησης σε μία ακολουθία δελτίων ειδήσεων είναι δυνατόν να οδηγήσει στην ταξινόμηση των πλάνων της ακολουθίας σε προκαθορισμένες κατηγορίες. Συγκεκριμένα, σε κάθε ακολουθία πραγματοποιείται αρχικά ανίχνευση αλλαγής πλάνων και στη συνέχεια τα πλάνα ταξινομούνται μέσω ενός αριθμού απλών κανόνων στις κατηγορίες «παρουσιαστής», «ρεπορτάζ / συνέντευξη», «στατική εικόνα» και «εξωτερικό πλάνο». Κάθε κατηγορία χαρακτηρίζεται από συγκεκριμένες ιδιότητες όπως την ύπαρξη ενός, δύο ή περισσότερων προσώπων σε κοντινό πλάνο ή στο φόντο της εικόνας, την κίνηση των προσώπων αυτών, τη μεταβολή του ιστογράμματος χρώματος της συνολικής εικόνας και την κίνηση που παρουσιάζεται στο φόντο. Με δεδομένο ότι στην τυπική οργάνωση των τηλεοπτικών ειδήσεων περιλαμβάνονται επαναλαμβανόμενες εμφανίσεις του παρουσιαστή εναλλάξ με άλλα είδη πλάνων, η ταξινόμηση αυτή οδηγεί άμεσα στη συντακτική ανάλυση, δηλαδή στη χρονική κατάτμηση των δελτίων σε στοιχειώδεις ενότητες ειδήσεων. Πρόκειται λοιπόν για ένα θετικό βήμα προς την κατεύθυνση της σημασιολογικής κατάτμησης σε θεματικές ενότητες, η οποία βεβαίως απαιτεί και άλλα είδη πληροφορίας εκτός από την οπτική, όπως για παράδειγμα την ακουστική και τη γλωσσική πληροφορία. Περιλαμβάνονται πειράματα ανάλυσης μεγάλου αριθμού ακολουθιών από δελτία ειδήσεων που αξιολογούνται, όπως και στην περίπτωση του εντοπισμού προσώπων, με μετρήσεις ακρίβειας και ανάκλησης· τα αποτελέσματα είναι ιδιαίτερα ενθαρρυντικά αν ληφθεί υπόψη η απλότητα των χρησιμοποιούμενων κριτηρίων ταξινόμησης.

13.1 ΑΝΑΛΥΣΗ ΔΕΛΤΙΩΝ ΕΙΔΗΣΕΩΝ

Τα ψηφιακά αρχεία τηλεοπτικών δελτίων ειδήσεων αποτελούν σημαντική πηγή οπτικοακουστικού υλικού για δύο κυρίως λόγους. Πρώτον, παρά το γεγονός ότι δεν έχουν κάποια συγκεκριμένη δομή όσον αφορά στο περιεχόμενο, όπως οποιοδήποτε οπτικοακουστικό υλικό [152], η τυπική οργάνωσή τους σε θεματικές ενότητες που περιλαμβάνουν επαναλαμβανόμενη εμφάνιση παρουσιαστών, ρεπορτάζ και εξωτερικών λήψεων, επιτρέπει εύκολη συντακτική ανάλυση (parsing). Με άλλα λόγια είναι δυνατή η χρονική τους κατάτμηση σε στοιχειώδεις ενότητες και η εξαγωγή πληροφορίας περιεχομένου από χαμηλού επιπέδου οπτικοακουστικά χαρακτηριστικά. Δεύτερον, συστήματα πλήρως αυτοματοποιημένης δεικτοδότησης (indexing) και χαρακτηρισμού (annotation) δελτίων ειδήσεων θα είναι πολύτιμα για τους αναλυτές δεδομένων σε πρακτορεία ειδήσεων, δημοσιογραφικούς οργανισμούς, κρατικές υπηρεσίες, τηλεοπτικούς παραγωγούς και υπηρεσίες παροχής πληροφοριών [87], καθώς για το σκοπό αυτό χρησιμοποιούνται παραδοσιακά χειρονακτικές μέθοδοι με υψηλό κόστος.

Για το σκοπό αυτό έχει αναπτυχθεί πρόσφατα σημαντικός αριθμός πρότυπων συστημάτων που επιτρέπουν την αυτόματη ή ημι-αυτόματη ανάλυση και χαρακτηρισμό δελτίων ειδήσεων [22,86,93,152]. Τέτοια συστήματα καθιστούν δυνατή την πλοήγηση ειδήσεων με

αλληλεπίδραση του χρήστη, την αναζήτηση με βάση το περιεχόμενο, καθώς και τη δημιουργία εφαρμογών *ειδήσεων κατ' απαίτηση* (news on demand, NoD). Τα περισσότερα απ' αυτά όμως βασίζονται σχεδόν αποκλειστικά σε *γλωσσική πληροφορία* (linguistic cues) και *πληροφορία κειμένου* (textual cues), όπως για παράδειγμα οι *υπότιτλοι* (closed-caption tokens) και οι *εγγραφές teletext* (teletext transcripts) που είναι διαθέσιμες από την παραγωγή των δελτίων. Είναι όμως κοινά αποδεκτό ότι η *οπτικοακουστική πληροφορία* (audiovisual cues) θα πρέπει να διαδραματίσει σημαντικότερο ρόλο στο μέλλον [134], ενώ για την ενσωμάτωση όλων των πληροφοριών περιεχομένου σε μια ενιαία περιγραφή για τη σημασιολογική κατάτμηση των ειδήσεων θα απαιτήσει μεθόδους τεχνητής νοημοσύνης.

Η ενσωμάτωση πληροφορίας έχει αντιμετωπισθεί με χρήση τεχνικών όπως ο *δυναμικός προγραμματισμός* (dynamic programming) [93], τα *finite state machines* [87] και τα *hidden Markov models* [40]. Από την άλλη μεριά όμως, ο ρόλος της οπτικοακουστικής πληροφορίας στα υπάρχοντα συστήματα συνήθως περιορίζεται σε πολύ απλά στάδια επεξεργασίας. Για παράδειγμα η ακουστική πληροφορία χρησιμοποιείται για την ανίχνευση περιόδων σιωπής, οι οποίες συνήθως λαμβάνουν χώρα κατά την αλλαγή θεματικής ενότητας σε ένα δελτίο, ενώ η οπτική πληροφορία για την ανίχνευση της μαύρης οθόνης η οποία συνήθως εμφανίζεται κατά τη διακοπή ενός δελτίου για διαφημιστικά μηνύματα [87]. Η περαιτέρω αξιοποίηση λοιπόν της οπτικοακουστικής πληροφορίας θα μπορούσε να συμβάλλει αποφασιστικά στη βελτίωση της επίδοσης αυτών των συστημάτων. Επιπλέον η απουσία πληροφορίας κειμένου από ένα σημαντικό αριθμό τηλεοπτικών προγραμμάτων ειδήσεων και κυρίως από ιστορικά αρχεία ειδήσεων δε μπορεί να υποτιμηθεί.

Για τους παραπάνω λόγους προτείνεται στο πλαίσιο της διατριβής μία τεχνική χρονικής κατάτμησης και συντακτικής ανάλυσης δελτίων ειδήσεων βασισμένη αποκλειστικά στην οπτική πληροφορία. Επειδή το επικρατέστερο τμήμα περιεχομένου των δελτίων ειδήσεων σχετίζεται με ανθρώπινες δραστηριότητες, οι εικόνες ανθρώπων και ιδιαίτερα προσώπων παίζουν σημαντικότερο ρόλο στην ανάλυση της δομής των δελτίων· έτσι η προτεινόμενη τεχνική στηρίζεται στην *ανίχνευση προσώπων*. Όπως περιγράφηκε στο προηγούμενο Κεφάλαιο, ο χρησιμοποιούμενος αλγόριθμος ανίχνευσης προσώπων περιλαμβάνει την κατάτμηση με βάση το χρώμα, την ανίχνευση του χρώματος δέρματος με χρήση ενός προσαρμοζόμενου μοντέλου κατανομής πιθανότητας, την ανάλυση σχήματος των τμημάτων δέρματος καθώς και το ταιριασμα με πρότυπα προσώπου στο στάδιο επαλήθευσης. Ο αλγόριθμος έχει παρουσιασθεί στις εργασίες [131,132].

Η ανίχνευση αλλαγής πλάνων που παρουσιάσθηκε στην Ενότητα 5.1 μαζί με ένα σύνολο απλών κανόνων για τον κατάλληλο συνδυασμό των χαρακτηριστικών των τμημάτων προσώπου (όπως η θέση, το μέγεθος, τα χαρακτηριστικά χρώματος, το σχήμα και η κίνηση) με τη δραστηριότητα που εμφανίζεται στο φόντο της εικόνας, επιτρέπει την αναγνώριση τυπικών σκηνών των δελτίων, όπως οι σκηνές του *παρουσιαστή* (anchorman shots), τα *ρεπορτάζ* (reports), τα *εξωτερικά πλάνα* (outdoor shots) και οι *στατικές εικόνες* (static images). Έτσι στην προτεινόμενη τεχνική ανάλυσης τα πλάνα ταξινομούνται σε μία από τις προηγούμενες κατηγορίες· η ταξινόμηση αυτή αποτελεί σημαντικό βήμα προς την κατεύθυνση της αναγνώρισης των θεματικών ενότητων ενός δελτίου, αφού οι θεματικές ενότητες συνήθως περιλαμβάνουν επαναλαμβανόμενη εμφάνιση του παρουσιαστή εναλλάξ με τα υπόλοιπα είδη πλάνων.

Η προτεινόμενη τεχνική συντακτικής ανάλυσης δελτίων ειδήσεων με χρήση οπτικής πληροφορίας είναι πρωτότυπη και δημοσιεύθηκε για πρώτη φορά στην εργασία [9]. Η τεχνική

μπορεί είτε να χρησιμοποιηθεί ως ανεξάρτητη εφαρμογή για την ανάλυση και χαρακτηρισμό δελτίων ειδήσεων στα οποία απουσιάζει η πληροφορία κειμένου, είτε να ενσωματωθεί σε υπάρχοντα συστήματα συνδυάζοντας την οπτικοακουστική με τη γλωσσική πληροφορία.

13.2 ΤΑΞΙΝΟΜΗΣΗ ΠΛΑΝΩΝ

Η ακολουθία βίντεο ενός δελτίου ειδήσεων αναλύεται αρχικά σε πλάνα με την τεχνική που παρουσιάστηκε στην Ενότητα 5.1. Στη συνέχεια σε κάθε καρέ της ακολουθίας πραγματοποιείται ανίχνευση προσώπων με τον αλγόριθμο του Κεφαλαίου 12. Η εικόνα πιθανοφάνειας προσώπου που παράγεται για κάθε καρέ χρησιμοποιείται για την ανίχνευση *κοντινών πλάνων* (close-ups) προσώπου, τα οποία αποτελούν μία καλή ένδειξη για την εμφάνιση ενός παρουσιαστή, ενός ρεπόρτερ / ανταποκριτή ή γενικότερα ενός ομιλητή σε συνέντευξη. Συγκεκριμένα, το μέγεθος των εξαγόμενων τμημάτων προσώπου χρησιμοποιείται για την απομόνωση κοντινών προσώπων, δηλαδή μεγάλων τμημάτων, και την απόρριψη μικρότερων, που ενδεχομένως αντιστοιχούν σε πρόσωπα στο φόντο της εικόνας, σε χέρια ή και άλλα αντικείμενα.

Επιπλέον, η έρευνα στην αναγνώριση προσώπων έχει δείξει ότι πρόσωπα με ανάλυση μικρότερη από 32×32 pixels είναι μη αναγνωρίσιμα. Βασιζόμενοι σε αυτό το συμπέρασμα επιβάλλουμε ένα όριο για την απόρριψη τμημάτων προσώπου τα οποία καταλαμβάνουν ποσοστό λιγότερο από 3% της συνολικής επιφάνειας της εικόνας. Με τον τρόπο αυτό διατηρούνται συνήθως ένα, δύο ή και τρία *επικρατέστερα πρόσωπα* (dominant faces), τα οποία θεωρείται ότι αντιστοιχούν σε κοντινά πλάνα προσώπου (close-ups). Στη συνέχεια χρησιμοποιούνται και πάλι ασαφείς συναρτήσεις συμμετοχής για τη θέση και το μέγεθος των επικρατέστερων προσώπων για την ταξινόμηση κάθε καρέ ανάλογα με το πλήθος κοντινών πλάνων προσώπου που περιέχουν. Η ταξινόμηση περιλαμβάνει τρεις κατηγορίες καρέ, δηλαδή τα καρέ που περιέχουν κανένα, ένα ή δύο κοντινά πλάνα προσώπου.

Η χρονική μεταβολή των χαρακτηριστικών των επικρατέστερων τμημάτων προσώπου, και κυρίως της θέσης και του μεγέθους τους λαμβάνεται υπόψη για την λεπτομερέστερη ταξινόμηση των καρέ. Για την ταξινόμηση αυτή τα παραπάνω χαρακτηριστικά συνδυάζονται με την πληροφορία κίνησης του *φόντου* (background) της εικόνας, η οποία αποτελεί ένδειξη για την δραστηριότητα που λαμβάνει χώρα στο πλάνο, και επομένως για το αν πρόκειται για εσωτερική ή εξωτερική λήψη. Συγκεκριμένα:

- (α) ένα ή δύο κοντινά πλάνα προσώπου σε στατικό φόντο ταξινομούνται ως *σκηνές ενός ή δύο παρουσιαστών* (single / double anchor)
- (β) ένα ή δύο κοντινά πλάνα προσώπου σε κινούμενο φόντο ταξινομούνται ως *ρεπορτάζ / συνέντευξη* (report / interview)
- (γ) εικόνες με στατικό φόντο που δεν περιέχουν πρόσωπα ταξινομούνται ως *στατικές εικόνες* (static images): τέτοιες εικόνες συνήθως εμφανίζονται σε οικονομικά ή αθλητικά ρεπορτάζ, ή σε δελτία καιρού
- (δ) τέλος, άλλες περιπτώσεις καρέ που παρουσιάζουν κίνηση και περιέχουν μικρά ή καθόλου πρόσωπα ταξινομούνται ως *εξωτερικά πλάνα* (outdoor shots)

Η κίνηση του φόντου εκτιμάται μέσω της απόλυτης διαφοράς εικόνας μεταξύ διαδοχικών καρέ: η αντιστάθμιση κίνησης δεν είναι απαραίτητη εδώ όπως στην ανίχνευση αλλαγής πλάνου,

αφού ο σκοπός είναι μόνο η ταξινόμηση του φόντου ως στατικού ή κινούμενου, επιτρέποντας τη διάκριση ανάμεσα σε εξωτερική και εσωτερική λήψη (στούντιο). Αφού ταξινομηθεί κάθε καρέ ενός πλάνου σε μία από τις παραπάνω κατηγορίες, είναι δυνατή η ταξινόμηση ολόκληρου του πλάνου· κάτι τέτοιο πραγματοποιείται με την εκτίμηση της κατηγορίας (παρουσιαστής, ρεπορτάζ, στατική εικόνα ή εξωτερικό πλάνο) που περιγράφει καλύτερα το πλάνο, δηλαδή την πλειοψηφία των καρέ. Σε περιπτώσεις μάλιστα που διαδοχικά τμήματα ενός πλάνου ταξινομούνται σε διαφορετικές κατηγορίες, είναι δυνατή η επιπλέον χρονική κατάτμηση του πλάνου.

Η χρονική κατάτμηση ενός δελτίου ειδήσεων είναι λοιπόν εφικτή μέσω της *ανίχνευσης αλλαγής πλάνου* (shot change detection) και της *ταξινόμησης πλάνων* (shot classification) με χρήση των παραπάνω κριτηρίων. Η ανίχνευση αλλαγής πλάνου, όπως αναλύεται στην Ενότητα 5.1, πραγματοποιείται με την κατωφλίωση των διαφορών καρέ μετά από αντιστάθμιση κίνησης· πρόκειται για μια απλή αλλά γρήγορη τεχνική, ιδιαίτερα όταν τα διανύσματα κίνησης και το σφάλμα εκτίμησης κίνησης είναι ήδη διαθέσιμα, όπως σε ακολουθίες MPEG. Με τον τρόπο αυτό ανιχνεύονται μόνο οι απότομες αλλαγές πλάνου, οι οποίες βέβαια έχουν και τη συντριπτική πλειοψηφία σε ακολουθίες βίντεο από δελτία ειδήσεων. Για την ανίχνευση ομαλής *μετάβασης* (transition) από ένα πλάνο στο επόμενο, όπως για παράδειγμα για μεταβάσεις τύπου *zoom*, *wipe* και *dissolve*, μπορούν να χρησιμοποιηθούν πιο εξειδικευμένες τεχνικές [21].

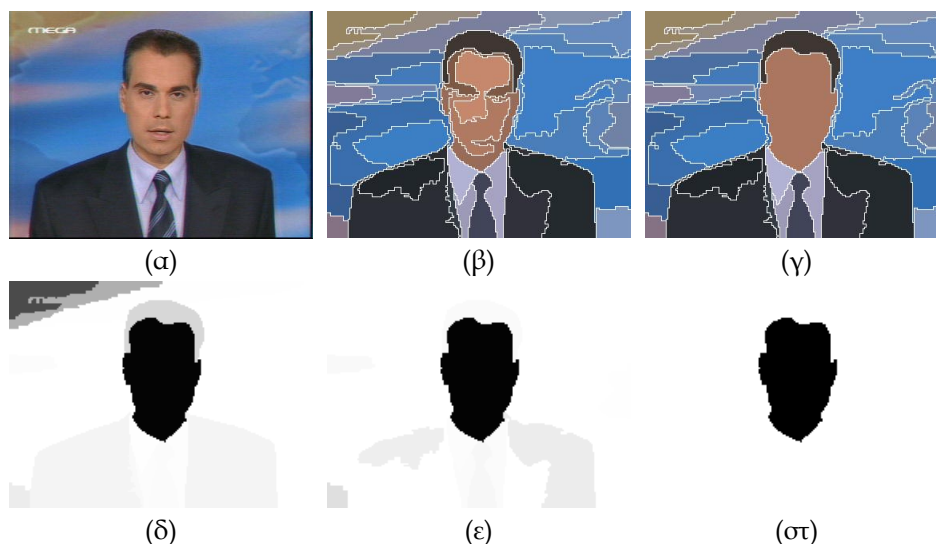
Επειδή κοντινά πλάνα προσώπου σε στατικό φόντο μπορούν να εμφανιστούν και σε περιπτώσεις διαφορετικές από εκείνη του κεντρικού παρουσιαστή (π.χ. στην περίπτωση ενός ανταποκριτή που μεταδίδει ρεπορτάζ από εσωτερικό χώρο), τα πλάνα παρουσιαστή φιλτράρονται περαιτέρω. Συγκεκριμένα, υλοποιείται *ομαδοποίηση* (clustering) σύμφωνα με το *ιστόγραμμα χρώματος* (color histogram) του φόντου με χρήση του αλγορίθμου *K-means*, όπως και στην περίπτωση επιλογής χαρακτηριστικών πλάνων στην Ενότητα 5.2. Στη συνέχεια επιλέγεται η ομάδα πλάνων με τα περισσότερα μέλη, η οποία αντιστοιχεί στο πλάνο παρουσιαστή με τις περισσότερες εμφανίσεις. Έτσι τα πλάνα παρουσιαστή περιορίζονται σημαντικά, ενώ τα υπόλοιπα πλάνα κατατάσσονται στην κατηγορία του ρεπορτάζ / συνέντευξης.

Τα πλάνα που παρεμβάλλονται ανάμεσα σε δύο διαδοχικές εμφανίσεις του κεντρικού παρουσιαστή ομαδοποιούνται σε *στοιχειώδεις ενότητες ειδήσεων* (elementary story units). Αυτό είναι ένα θετικό βήμα προς την κατεύθυνση της *σημασιολογικής κατάτμησης* (semantic segmentation) των δελτίων, δηλαδή της ανίχνευσης των πραγματικών *θεματικών ενότητων* (news topics). Οι θεματικές ενότητες αποτελούνται από μία ή περισσότερες στοιχειώδεις ενότητες· επομένως για τον ακριβή προσδιορισμό των θεματικών ενότητων απαιτείται περαιτέρω ομαδοποίηση. Φυσικά, κάτι τέτοιο είναι ανέφικτο με χρήση της οπτικής πληροφορίας μόνο. Αντίθετα είναι απαραίτητος ο συνδυασμός με την ακουστική πληροφορία ή και την πληροφορία κειμένου, αν βέβαια η τελευταία είναι διαθέσιμη. Για παράδειγμα, δύο διαδοχικά πλάνα στα οποία η ομιλία είναι συνεχόμενη εύκολα κατατάσσονται στην ίδια στοιχειώδη ενότητα ειδήσεων, αφού στην αλλαγή θεματικών ενότητων παρεμβάλλονται μικρά χρονικά διαστήματα σιωπής. Από την οπτική πληροφορία ένα τέτοιο συμπέρασμα είναι πολύ δύσκολο να εξαχθεί.

13.3 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Το οπτικό περιεχόμενο που χρησιμοποιήθηκε για την πειραματική αξιολόγηση της προτεινόμενης τεχνικής προέρχεται από μία βάση δεδομένων η οποία περιλαμβάνει εγγραφές δελτίων

ειδήσεων από τέσσερις Ελληνικούς τηλεοπτικούς σταθμούς, και συγκεκριμένα τους ET1, Mega, Antenna και Alter. Έξι τμήματα δελτίων ειδήσεων, διάρκειας 10 λεπτών περίπου το καθένα ψηφιοποιήθηκαν για το σκοπό αυτό με ρυθμό 10 frames/sec και σε ανάλυση $384 \times 288 \times 24$ brrp. Αρχικά δίνονται ενδεικτικά παραδείγματα ανίχνευσης του επικρατέστερου προσώπου σε ένα καρέ του κεντρικού παρουσιαστή και ταξινόμησης πλάνων σε ένα τμήμα δελτίου διάρκειας 100 δευτερολέπτων, αποτελούμενου από 15 πλάνα. Στη συνέχεια παρουσιάζονται στατιστικά αποτελέσματα αξιολόγησης της τεχνικής ταξινόμησης για ολόκληρο το περιεχόμενο της βάσης δεδομένων.



Σχήμα 13-1. Ανίχνευση επικρατέστερου τμήματος προσώπου. (α) Αρχικό καρέ παρουσιαστή. (β,γ) Κατάτμηση με βάση το χρώμα, πριν και μετά τη συγχώνευση τμημάτων δέρματος, αντίστοιχα. (δ) Πιθανοφάνεια δέρματος. (ε) Πιθανοφάνεια προσώπου. (στ) Επικρατέστερο τμήμα προσώπου.

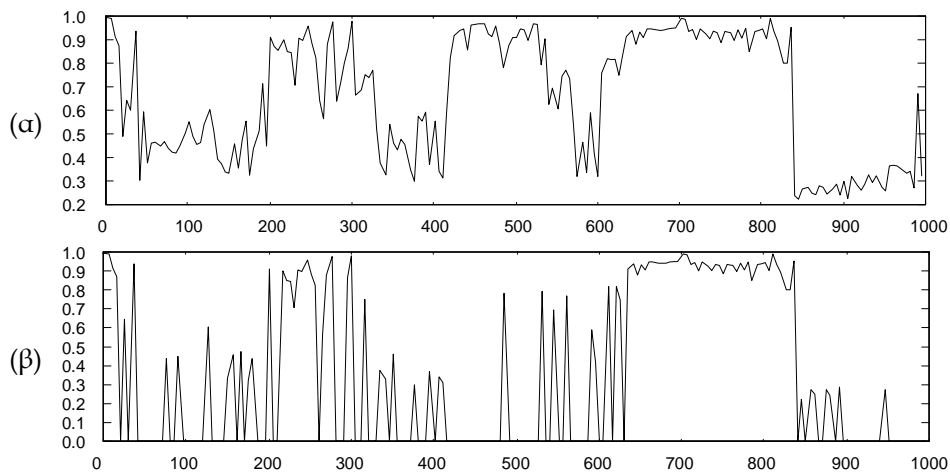
Ένα τυπικό παράδειγμα ανίχνευσης του επικρατέστερου τμήματος προσώπου απεικονίζεται στο Σχήμα 13-1. Η αρχική εικόνα ενός καρέ του κεντρικού παρουσιαστή φαίνεται στο Σχήμα 13-1α. Στην εικόνα εφαρμόζεται κατάτμηση με βάση το χρώμα και συγχώνευση τμημάτων δέρματος όπως περιγράφεται στην Ενότητα 12.3 και παρουσιάζεται στα Σχήματα 13-1β,γ αντίστοιχα. Στη συνέχεια παράγονται οι εικόνες πιθανοφάνειας δέρματος και προσώπου που φαίνονται στα Σχήματα 13-1δ,ε αντίστοιχα. Τέλος, με κατωφλίωση της πιθανοφάνειας προσώπου και του μεγέθους των τμημάτων εξάγεται το επικρατέστερο τμήμα προσώπου. Παρατηρείται ότι με το συνδυασμό της κατάτμησης, της ανίχνευσης χρώματος δέρματος, της επεξεργασίας σχήματος, του ταιριάσματος με πρότυπο προσώπου και της κατωφλίωσης μεγέθους, εξάγεται με μεγάλη ακρίβεια το τμήμα του προσώπου του παρουσιαστή.

Η χρονική κατάτμηση σε πλάνα και η ταξινόμηση πλάνων σε κατηγορίες επιδεικνύεται για την ενδεικτική ακολουθία που εμφανίζεται στο Σχήμα 13-2. Η ακολουθία αυτή είναι τμήμα δελτίου ειδήσεων του σταθμού Mega, αποτελείται από 1000 καρέ, έχει δηλαδή συνολική διάρκεια 100 δευτερόλεπτα, και περιέχει 15 πλάνα, ένα από τα οποία είναι πλάνο παρουσιαστή. Το Σχήμα 13-3α απεικονίζει τη γραφική παράσταση της *μέγιστης πιθανοφάνειας προσώπου* (maximum face likelihood) συναρτήσεως του χρόνου, δηλαδή του αριθμού καρέ, για όλη τη διάρκεια της ακολουθίας (0-1000). Σε κάθε καρέ, η μέγιστη πιθανοφάνεια υπολογίζεται ως η μέγιστη τιμή της πιθανοφάνειας όλων των τμημάτων, πριν τα τμήματα φιλτραριστούν με βάση το μέγεθος.



Σχήμα 13-2. Η ακολουθία Mega, διάρκειας 1000 καρέ (100 sec), που αποτελείται από 15 πλάνα, ένα από τα οποία (πλάνο #13) είναι πλάνο παρουσιαστή. Για κάθε πλάνο παρουσιάζεται το πρώτο καρέ.

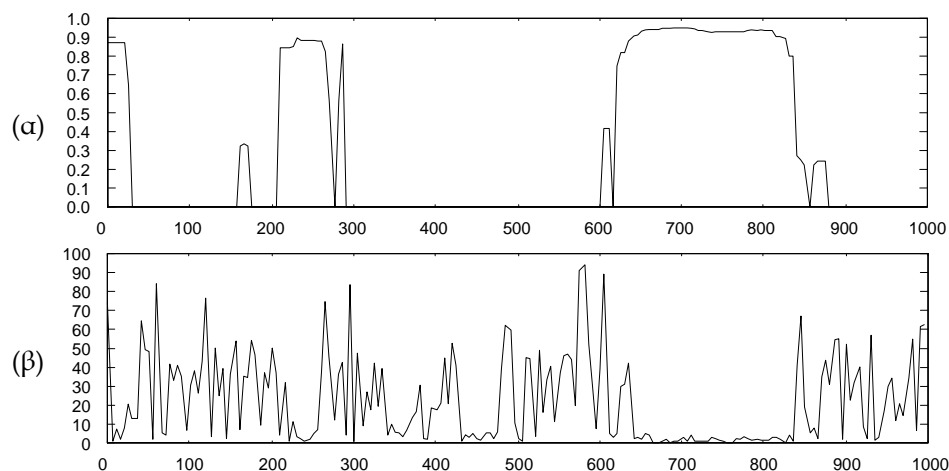
Αντίστοιχα στο Σχήμα 13-3β φαίνεται η ανάλογη γραφική παράσταση μετά την απόρριψη των μικρότερων τμημάτων, για τα οποία η πιθανοφάνεια τίθεται ίση με μηδέν. Πρόκειται δηλαδή για τη μέγιστη πιθανοφάνεια επικρατέστερου προσώπου (maximum dominant face likelihood), η τιμή της οποίας υπολογίζεται και πάλι για όλα τα τμήματα της εικόνας. Είναι εμφανές ότι ορισμένα καρέ τα οποία περιέχουν μικρά τμήματα με υψηλή πιθανοφάνεια προσώπου απορρίπτονται, ενώ διατηρούνται κυρίως τα καρέ που περιέχουν μεγάλα πρόσωπα σε κοντινό πλάνο – πρόσωπα δηλαδή που πιθανόν να αντιστοιχούν σε παρουσιαστή, ανταποκριτή ή συνομιλητή σε συνέντευξη.



Σχήμα 13-3. Μετρήσεις πιθανοφάνειας συναρτήσει του χρόνου (αριθμού καρέ, 0-1000) για την ακολουθία Mega. (α) Μέγιστη πιθανοφάνεια προσώπου. (β) Μέγιστη πιθανοφάνεια επικρατέστερου προσώπου, μετά την αφαίρεση μικρών τμημάτων.

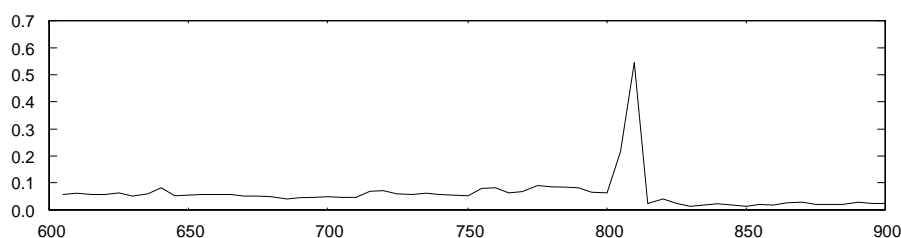
Παρατηρείται από το Σχήμα 13-3β ότι από τη μέγιστη πιθανοφάνεια επικρατέστερου προσώπου δεν μπορεί να εξαχθεί αξιόπιστο συμπέρασμα για την ταξινόμηση ενός ολόκληρου πλάνου, διότι η χρονική μεταβολή της πιθανοφάνειας από καρέ σε καρέ είναι ιδιαίτερα έντονη. Για το λόγο αυτό πραγματοποιείται φιλτράρισμα της παραπάνω καμπύλης ως προς το χρόνο με χρήση ενός φίλτρου median. Η φιλτραρισμένη καμπύλη της μέγιστης πιθανοφάνειας επικρατέστερου προσώπου απεικονίζεται στο Σχήμα 13-4α. Το φιλτράρισμα είναι απαραίτητο για να απορριφθούν πλάνα πολύ μικρής διάρκειας που περιέχουν πρόσωπα, αφού

για παράδειγμα ένα πλάνο παρουσιαστή σπάνια έχει διάρκεια μικρότερη από λίγα δευτερόλεπτα. Η επιλογή του μήκους παραθύρου που χρησιμοποιείται στο φιλτράρισμα μπορεί να προσαρμόζεται στην επιθυμητή ελάχιστη χρονική διάρκεια πλάνων.



Σχήμα 13-4. (α) Μέγιστη πιθανοφάνεια επικρατέστερου προσώπου για την ακολουθία Mega συναρτήσει του χρόνου (αριθμού καρέ), μετά από φιλτράρισμα με χρήση φίλτρου median. (β) Μέτρο διανύσματος κίνησης του επικρατέστερου προσώπου, συναρτήσει του χρόνου.

Με απλή παρατήρηση του Σχήματος 13-4α προκύπτει ότι υπάρχουν τρία πιθανά χρονικά τμήματα της ακολουθίας που περιέχουν κοντινό πλάνο προσώπου. Το πρώτο (πλάνο #1, καρέ 0-38) δεν αντιστοιχεί σε πραγματικό πλάνο προσώπου, έχει γίνει δηλαδή λανθασμένη ανίχνευση (false alarm). Το δεύτερο τμήμα αντιστοιχεί σε δύο πραγματικά πλάνα προσώπων, και αποτελείται από τα δύο διαδοχικά πλάνα #5 και #6 (καρέ 209-286). Τέλος, το τρίτο τμήμα αποτελείται επίσης από δύο διαδοχικά πλάνα προσώπων μεγαλύτερης διάρκειας, τα πλάνα #13 και #14 της ακολουθίας (καρέ 614-867). Σημειώνεται ότι η ανίχνευση προσώπου είναι επιτυχής στην ακολουθία αυτή παρά τις άσχημες συνθήκες φωτισμού που επικρατούν, αφού η ακολουθία περιέχει εξωτερικά πλάνα ρεπορτάζ και συνεντεύξεων κατά τη διάρκεια της νύχτας.



Σχήμα 13-5. Μέτρο της διαφοράς ιστογράμματος χρώματος διαδοχικών καρέ για την ακολουθία Mega, καρέ 600-900 (πλάνα #13 και #14) Η μέγιστη τιμή αντιστοιχεί στη χρονική στιγμή αλλαγής πλάνου.

Η λανθασμένη ανίχνευση προσώπου στο πλάνο #1 αντισταθμίζεται εύκολα λαμβάνοντας υπόψη την κίνηση του επικρατέστερου προσώπου, όπως φαίνεται στο Σχήμα 13-4β, αφού τα πραγματικά τμήματα προσώπου συνήθως εμφανίζουν μικρή κίνηση ενώ οι ταχείες ταλαντώσεις στην κίνηση του επικρατέστερου προσώπου οφείλονται ουσιαστικά σε εξωτερικά πλάνα με έντονη δραστηριότητα που ταξινομούνται λανθασμένα ως πλάνα προσώπου. Η κίνηση του επικρατέστερου προσώπου στο πλάνο #1 υποδηλώνει πως το πλάνο δεν περιέχει κοντινή λήψη προσώπου, παρά τη μεγάλη τιμή πιθανοφάνειας. Από την άλλη μεριά, το γεγονός ότι δύο ζεύγη διαδοχικών πλάνων (#5,#6 και #13,#14) εμφανίζονται στο Σχήμα 13-4α ως ενιαία χρονικά τμήματα δεν αποτελεί πρόβλημα, αφού τα πλάνα έχουν ήδη διαχωριστεί σε προηγούμε-

νο στάδιο· θα μπορούσαν βέβαια εύκολα να διαχωριστούν με χρήση κριτηρίων όπως η μεταβολή του ιστογράμματος χρώματος, όπως φαίνεται και στο Σχήμα 13-5.

13.4 ΑΞΙΟΛΟΓΗΣΗ

Η επίδοση της προτεινόμενης τεχνικής χρονικής κατάτμησης δελτίων ειδήσεων αξιολογήθηκε μέσω μετρήσεων *ακρίβειας* (precision) και *ανάκλησης* (recall). Οι έξι δεκάλεπτες ακολουθίες της βάσης δεδομένων χωρίστηκαν σε πλάνα και κάθε πλάνο ταξινομήθηκε και χαρακτηρίστηκε χειρωνακτικά σε μία από τις κατηγορίες *Παρουσιαστής*, *Ρεπορτάζ / Συνέντευξη*, *Στατική Εικόνα* και *Εξωτερικό Πλάνο*. Η ταξινόμηση αυτή θεωρείται *αληθής* (ground truth) ταξινόμηση. Τα αποτελέσματα λοιπόν της αυτόματης διαδικασίας ταξινόμησης μπορούν να συγκριθούν με την αληθή ταξινόμηση ώστε να προκύψουν στατιστικές μετρήσεις για την αξιολόγηση της τεχνικής.

Ακολουθία	Παρουσιαστής		Ρεπορτάζ / Συνέντευξη		Στατική Εικόνα		Εξωτερικό Πλάνο	
	P	R	P	R	P	R	P	R
A5 (α)	0.92	0.94	0.65	0.82	–	–	0.75	0.85
A5 (β)	0.95	1.00	0.83	0.94	0.50	1.00	0.73	0.87
ANT1	0.93	0.94	0.77	0.88	0.75	0.66	0.85	0.86
ET-1	1.00	1.00	0.71	0.88	0.66	1.00	0.81	0.93
MEGA (α)	0.93	0.93	0.76	0.86	0.75	0.75	0.67	0.86
MEGA (β)	0.96	1.00	0.84	0.91	–	–	0.74	0.81
Σύνολο	0.95	0.97	0.76	0.88	0.67	0.85	0.76	0.86

Πίνακας 13-1. Μετρήσεις ακρίβειας (precision) και ανάκλησης (recall) για την ταξινόμηση πλάνων στις έξι δοκιμαστικές ακολουθίες δελτίων ειδήσεων.

Όπως περιγράφεται στην εργασία [87], η *ακρίβεια* ορίζεται ως ο λόγος των σωστά ευθυγραμμισμένων γεγονότων (correctly aligned events) προς το συνολικό αριθμό γεγονότων που ανιχνεύονται, ενώ η *ανάκληση* ως ο λόγος ως ο λόγος των σωστά ευθυγραμμισμένων γεγονότων προς το συνολικό αριθμό αληθών γεγονότων. Ως *γεγονός* (event) ορίζεται κάθε μετάβαση ανάμεσα σε δύο διαφορετικές κατηγορίες πλάνων· ένα γεγονός είναι *σωστά ευθυγραμμισμένο* (correctly aligned) αν λαμβάνει χώρα σε ένα χρονικό διάστημα που απέχει το πολύ ± 2 καρέ από το αντίστοιχο αληθές γεγονός. Σημειώνεται ότι η ακρίβεια είναι η αντίθετη έννοια του *ρυθμού λανθασμένης ανίχνευσης* (false alarm rate), ενώ η ανάκληση είναι η αντίθετη έννοια του *ρυθμού απώλειας* (dismissal rate). Οι πρώτες χρησιμοποιούνται γενικά στο πλαίσιο της *αναζήτησης πληροφορίας* (information retrieval) ενώ οι δεύτερες στην *ανίχνευση και αναγνώριση προσώπων* αλλά και γενικότερα στην *εκτίμηση υποθέσεων* (hypothesis testing) και τη στατιστική.

Οι μετρήσεις ακρίβειας και ανάκλησης για κάθε μία από τις έξι δοκιμαστικές ακολουθίες της βάσης και για κάθε μία από τις τέσσερις κατηγορίες πλάνων δίνονται στον Πίνακα 13-1. Παρατηρείται αρχικά ότι τις καλύτερες μετρήσεις εμφανίζει η κατηγορία του παρουσιαστή· κάτι τέτοιο είναι αναμενόμενο καθώς η κατηγορία αυτή βασίζεται και στην ομαδοποίηση (clustering) πέρα από την ανίχνευση προσώπων. Οι μετρήσεις ταξινόμησης για την κατηγορία των ρεπορτάζ / συνεντεύξεων είναι ελαφρώς μικρότερες, κυρίως λόγω των ανεξέλεγκτων συνθηκών φωτισμού που γενικά επηρεάζουν τη διαδικασία ανίχνευσης χρώματος δέρματος. Συνήθως τα πλάνα ρεπορτάζ / συνεντεύξεων ταξινομούνται λανθασμένα στην κατηγορία των εξωτερικών πλάνων και αντίστροφα. Τέλος, λόγω του περιορισμένου αριθμού διαθέσιμων

στατικών εικόνων στη βάση δεν είναι δυνατό να εξαχθούν αξιόπιστα συμπεράσματα για την κατηγορία αυτή.

Γενικά παρατηρείται ότι οι μετρήσεις ακρίβειας και ανάκλησης είναι ιδιαίτερα ενθαρρυντικές παρά την απλότητα της προτεινόμενης τεχνικής. Εκτιμάται λοιπόν ότι ο συνδυασμός της με ακουστική και γλωσσική πληροφορία και η ενσωμάτωσή της σε υφιστάμενα συστήματα θα μπορούσε να δώσει πολύ ικανοποιητικά αποτελέσματα για την πλήρως αυτοματοποιημένη ανάλυση δελτίων ειδήσεων.

Προεπισκόπηση και Ανάκληση

Στο τελευταίο Κεφάλαιο της διατριβής εξετάζεται η εφαρμογή των προτεινόμενων μεθόδων χαρακτηρισμού, δεικτοδότησης, εξαγωγής περίληψης, προεπισκόπησης και αναζήτησης με βάση το περιεχόμενο στην ειδική περίπτωση των ακολουθιών που προέρχονται από δελτία ειδήσεων. Στην περίπτωση αυτή η συντακτική ανάλυση των δελτίων σε συνδυασμό με την επιλογή χαρακτηριστικών πλάνων και καρέ μπορεί να οδηγήσει στη δημιουργία ενός ισχυρού συστήματος παρουσίασης, πλοήγησης και αναζήτησης από δελτία ειδήσεων με βάση το οπτικοακουστικό περιεχόμενο, στο οποίο η αλληλεπίδραση με το χρήστη παίζει πρωταρχικό ρόλο. Επιπλέον, για την περίπτωση της αναζήτησης και ανάκλησης εξετάζεται ένας αριθμός από νέα σενάρια αναζήτησης που σχετίζονται με την ύπαρξη ανθρώπινων προσώπων, αφού η πληροφορία των τμημάτων προσώπου παράγεται από τον αυτόματο εντοπισμό τους και μπορεί να αποθηκευθεί για σκοπούς δεικτοδότησης. Συγκεκριμένα εξετάζεται η αναζήτηση με βάση τα χαρακτηριστικά χρώματος, το μέγεθος και το πλήθος των προσώπων. Τα κριτήρια αυτά μπορούν επίσης να συνδυαστούν με γενικότερα κριτήρια αναζήτησης όπως αυτά που παρουσιάστηκαν στο Κεφάλαιο 8 για την ανάκληση μέσω παραδείγματος ή σχεδιαγράμματος, εμπλουτίζοντας έτσι τις δυνατότητες υφιστάμενων συστημάτων διαχείρισης οπτικοακουστικού υλικού και αναζήτησης με βάση το περιεχόμενο. Παρουσιάζονται επίσης ενδεικτικά αποτελέσματα ανάκλησης με βάση το χρώμα και το μέγεθος των προσώπων, καθώς και ποσοτική αξιολόγηση της ανάκλησης με βάση το πλήθος προσώπων με μετρήσεις ακρίβειας και ανάκλησης σε ένα μεγάλο σύνολο ακολουθιών δελτίων ειδήσεων.

14.1 ΕΞΑΓΩΓΗ ΠΕΡΙΛΗΨΗΣ ΚΑΙ ΠΡΟΕΠΙΣΚΟΠΗΣΗ

Μία από τις σημαντικότερες εφαρμογές της συντακτικής ανάλυσης ακολουθιών δελτίων ειδήσεων είναι η χρήση της πληροφορίας χαρακτηρισμού των πλάνων για σκοπούς προεπισκόπησης της ακολουθίας. Με δεδομένο ότι η τυπική οργάνωση των δελτίων ειδήσεων περιλαμβάνει περιοδικές εμφανίσεις του κεντρικού παρουσιαστή ανάμεσα στις οποίες εμφανίζονται συνήθως ρεπορτάζ / συνεντεύξεις, εξωτερικά πλάνα ή στατικές εικόνες, η αυτόματη ταξινόμηση πλάνων σε μία από τις παραπάνω κατηγορίες επιτρέπει τον προσδιορισμό ενότητων πλάνων (shot units) που αντιστοιχούν σε στοιχειώδεις ενότητες ειδήσεων (elementary story units). Μία ενότητα πλάνων ορίζεται ως το σύνολο πλάνων που μεσολαβούν μεταξύ δύο διαδοχικών πλάνων του κεντρικού παρουσιαστή. Η προεπισκόπηση ενός δελτίου ειδήσεων περιλαμβάνει την απεικόνιση στον τελικό χρήστη των ενότητων πλάνων του δελτίου, καθώς και των πλάνων που περιλαμβάνονται σε κάθε ενότητα, με μικρά εικονίδια, και της πληροφορίας χαρακτηρισμού των πλάνων. Η πληροφορία αυτή συνίσταται στο αποτέλεσμα ταξινόμησης κάθε πλάνου σε μία από τις παραπάνω κατηγορίες. Επιπλέον ο χρήστης έχει τη δυνατότητα να επιλέξει τις κατηγορίες πλάνων που επιθυμεί για επισκόπηση, δηλαδή για παράδειγμα μόνο τα πλάνα παρουσιαστή, μόνο τα εξωτερικά πλάνα κλπ.



Σχήμα 14-1. Παράδειγμα προεπισκόπησης ακολουθίας δελτίου ειδήσεων. Διακρίνονται εννέα ενότητες πλάνων (E1-E9) σε κάθε μία από τις οποίες αντιστοιχούν από 2 έως 14 πλάνα. Η κατηγορία κάθε πλάνου εμφανίζεται δίπλα στον αριθμό της. (A): Anchor (Παρουσιαστής). (R): Report (Ρεπορτάζ/ Συνέντευξη). (O): Outdoor shot (Εξωτερικό πλάνο). Στατικές εικόνες δεν υπάρχουν.

Αυτός ο τύπος προεπισκόπησης μπορεί να συνδυαστεί με την εξαγωγή περίληψης που παρουσιάστηκε στο Μέρος II της διατριβής. Συγκεκριμένα, οι ενότητες πλάνων πολύ συχνά περιλαμβάνουν πολλά πλάνα με παρόμοιο οπτικό περιεχόμενο, και μάλιστα με περιοδική επανάληψη. Η ομαδοποίηση πλάνων λοιπόν με βάση το περιεχόμενο χρησιμεύει για την απει-

κόνιση στο χρήστη ενός περιορισμένου αριθμού πλάνων, δηλαδή των αντιπροσωπευτικών πλάνων κάθε ομάδας· ο χρήστης τότε μπορεί να επιλέξει ένα από αυτά για να δει και το σύνολο των πλάνων της ομάδας που τον ενδιαφέρει. Επιπλέον, η εξαγωγή χαρακτηριστικών καρέ από κάθε πλάνο επιτρέπει την επιλογή από το χρήστη ενός πλάνου και την επισκόπηση των περιεχομένων του με τη μορφή εικονιδίων. Δημιουργείται έτσι μία ιεραρχική δομή των δελτίων ειδήσεων που δίνει στο χρήστη τη δυνατότητα προεπισκόπησης και πλοήγησης στα περιεχόμενα των δελτίων με διαφορετικό επίπεδο λεπτομέρειας κάθε φορά.

Ένα χαρακτηριστικό παράδειγμα φαίνεται στο Σχήμα 14-1 για μία ακολουθία δελτίου ειδήσεων της ET1, διάρκειας 8.5 περίπου λεπτών (5115 καρέ με ρυθμό δειγματοληψίας 10 frames/sec). Η ακολουθία περιέχει συνολικά 68 πλάνα, από τα οποία 9 χαρακτηρίζονται ως πλάνα *παρουσιαστή* (Anchor, A), άλλα 9 ως *ρεπορτάζ / συνέντευξη* (Report, R) και 50 ως *εξωτερικά πλάνα* (Outdoor shot, O), ενώ πλάνα στατικών εικόνων δεν υπάρχουν στο συγκεκριμένο δελτίο. Με βάση τα πλάνα του παρουσιαστή, ορίζονται εννέα ομάδες πλάνων, κάθε μία από τις οποίες περιέχει από 2 έως 14 πλάνα. Η απεικόνιση των ομάδων πλάνων όπως στο Σχήμα 14-1 επιτρέπει στο χρήστη την άμεση επισκόπηση των στοιχειωδών ενότητων ειδήσεων. Φυσικά, οι πραγματικές *θεματικές ενότητες* (news topics) του δελτίου αποτελούνται από σύνολα στοιχειωδών ενότητων ειδήσεων, τα οποία όμως δεν είναι δυνατό να προσδιορισθούν μόνο με βάση την οπτική πληροφορία, όπως αναφέρεται στο Κεφάλαιο 13.

Παρατηρείται ότι πολλά από τα πλάνα παρουσιάζουν όμοιο ή παρόμοιο οπτικό περιεχόμενο, όπως για παράδειγμα τα #2/#7, #17/#18, #19/#20/#21, #24/#31, #37/#38, #42/#45/#46, #43/#44 και #54/#55/#56. Με την ομαδοποίηση πλάνων η παρουσίαση των πλάνων αυτών θα μπορούσε να γίνει με ένα εικονίδιο για κάθε ομάδα, που να αντιστοιχεί στο αντιπροσωπευτικό πλάνο της ομάδας. Παρατηρείται επίσης ότι πολλές φορές η ταξινόμηση πλάνων αποτυγχάνει. Χαρακτηριστικό παράδειγμα είναι το πλάνο #58 το οποίο είναι ασπρόμαυρο με αποτέλεσμα να μην ανιχνεύεται το πρόσωπο (dismissal). Έτσι το πλάνο θεωρείται εξωτερικό και όχι ρεπορτάζ / συνέντευξη. Πάντως τα πλάνα παρουσιαστή εντοπίζονται όλα σωστά, αφενός λόγω των κατάλληλων συνθηκών φωτισμού και του στατικού φόντου, και αφετέρου λόγω της διαδικασίας ομαδοποίησης πλάνων παρουσιαστή που περιγράφηκε στην Ενότητα 13.2.

14.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΡΟΣΩΠΩΝ

Η ανίχνευση και εντοπισμός προσώπων σε μια βάση δεδομένων με εικόνες και βίντεο δίνει τη δυνατότητα εκμετάλλευσης της εξαγόμενης πληροφορίας, όπως το πλήθος προσώπων, η κλίμακα και η μέση τιμή των χρωματικών συνιστωσών. Η πληροφορία αυτή μπορεί να αποθηκευτεί στη βάση και να χρησιμοποιηθεί για δεικτοδότηση (indexing) και *ανάκληση με βάση το περιεχόμενο* (content-based retrieval). Έτσι, σε ένα περιβάλλον *αναζήτησης μέσω παραδείγματος* (query-by-example), η πληροφορία προσώπων μπορεί να οδηγήσει σε νέα κριτήρια δεικτοδότησης. Στην περίπτωση αυτή οι εικόνες εισόδου αναλύονται σε πραγματικό χρόνο και τα χαρακτηριστικά των προσώπων που εντοπίζονται χρησιμοποιούνται για *ανάκληση με βάση τα πρόσωπα*. Τα προτεινόμενα σενάρια αναζήτησης με βάση την πληροφορία προσώπων αποτέλεσαν το αντικείμενο των εργασιών [15,132].

Για το σκοπό της ανάκλησης με βάση τα πρόσωπα, κάθε εικόνα (ή καρέ ακολουθίας) αναλύεται για την ανίχνευση προσώπων, και στη συνέχεια εξάγονται και αποθηκεύονται στη βάση τα ακόλουθα χαρακτηριστικά:

- (α) Το πλήθος τμημάτων προσώπου
- (β) Η μέση τιμή των χρωματικών συνιστωσών κάθε τμήματος προσώπου. Έστω ότι c_{Cr} και c_{Cb} είναι αντίστοιχα οι χρωματικές συνιστώσες Cr και Cb μιας εικόνας, και S_i είναι το i -οστό τμήμα προσώπου της εικόνας. Τότε οι μέσες τιμές των συνιστωσών Cr και Cb για το τμήμα S_i υπολογίζονται ως εξής:

$$\mu_{Cr}(S_i) = \frac{1}{\|S_i\|} \sum_{s \in S_i} c_{Cr}(s), \quad \mu_{Cb}(S_i) = \frac{1}{\|S_i\|} \sum_{s \in S_i} c_{Cb}(s) \quad (14-1)$$

- (γ) Το ποσοστό $a(S_i)$ της επιφάνειας της εικόνας που καλύπτεται από κάθε τμήμα προσώπου.

Επιπρόσθετα χαρακτηριστικά μπορούν επίσης να χρησιμοποιηθούν στη διαδικασία δεικτοδότησης, όπως η θέση κάθε προσώπου, ο προσανατολισμός του, χαρακτηριστικά σχήματος (π.χ. συμπαγές, επιμήκυνση) καθώς και χαρακτηριστικά υφής (π.χ. ροπές, συντελεστές FFT κλπ.). Επιπλέον, τα χαρακτηριστικά αυτά μπορούν να συνδυαστούν και με τα γενικότερα χαρακτηριστικά που χρησιμοποιούνται σε ένα περιβάλλον αναζήτησης, οδηγώντας σε μια ποικιλία από σενάρια αναζήτησης. Όμως τα τρία χαρακτηριστικά που αναφέρθηκαν παραπάνω είναι επαρκή για τα τρία κριτήρια αναζήτησης που χρησιμοποιούνται στην παρούσα εργασία και αναλύονται αμέσως παρακάτω.

14.3 ΑΝΑΚΛΗΣΗ ΜΕ ΒΑΣΗ ΤΑ ΠΡΟΣΩΠΑ

Παρουσιάζονται στη συνέχεια τρία κριτήρια ανάκλησης εικόνων με βάση τα πρόσωπα, τα οποία έχουν εξεταστεί πειραματικά και δίνουν αξιολογικά αποτελέσματα. Συγκεκριμένα τα κριτήρια αυτά είναι οι μέσες χρωματικές συνιστώσες των προσώπων, η κλίμακα και το πλήθος προσώπων σε μια εικόνα.

Μέσες Χρωματικές Συνιστώσες Προσώπων

Με βάση το κριτήριο αυτό ο χρήστης ενδιαφέρεται για την ανάκληση εικόνων ή καρέ από βίντεο που περιέχουν πρόσωπα με χαρακτηριστικά χρώματος παρόμοια με εκείνα μιας εικόνας που εισάγει στο σύστημα αναζήτησης. Στην περίπτωση αυτή ανιχνεύεται αυτόματα και χαρακτηρίζεται το επικρατέστερο τμήμα προσώπου από την εικόνα· αν υπάρχουν περισσότερα από ένα πρόσωπα ο χρήστης μπορεί να επιλέξει το συγκεκριμένο που τον ενδιαφέρει.

Έστω ότι F είναι το επιλεγμένο τμήμα προσώπου. Για το τμήμα αυτό υπολογίζονται οι μέσες χρωματικές συνιστώσες $\mu_{Cr}(F)$ και $\mu_{Cb}(F)$ αντίστοιχα, σύμφωνα με τον ορισμό (14-1). Οι τιμές αυτές χρησιμοποιούνται για την *επανεκτίμηση* του μέσου διανύσματος μ στο μοντέλο χρώματος δέρματος της σχέσης (12-1). Συγκεκριμένα, σύμφωνα με τη σχέση επανεκτίμησης (12-3),

$$\hat{\mu}_0 = (1 - m) \mu + m \mu_0 \quad (14-2)$$

όπου $\mu = [\mu_{Cr}(F) \mu_{Cb}(F)]^T$ και η παράμετρος μνήμης m λαμβάνει σχετικά μικρή τιμή (τυπικά 0.4) έτσι ώστε το μοντέλο να προσαρμόζεται καλά στο επιλεγμένο τμήμα προσώπου, διατηρών-

ντας ταυτόχρονα τα γενικά χαρακτηριστικά του χρώματος δέρματος. Έστω τώρα ότι με $S_i^{(k)}$ συμβολίζεται το i -οστό τμήμα προσώπου της k -οστής εικόνας στη βάση και $\mathbf{x}(S_i^{(k)}) = [\mu_{Cr}(S_i^{(k)}) \ \mu_{Cb}(S_i^{(k)})]^T$ είναι το αντίστοιχο μέσο διάνυσμα χρωματικών συνιστωσών. Τότε ο βαθμός ομοιότητας του τμήματος $S_i^{(k)}$ με το επιλεγμένο τμήμα F ορίζεται ως

$$s_c(S_i^{(k)}) = \exp\left\{-\frac{1}{2}(\mathbf{x}(S_i^{(k)}) - \hat{\mu}_0)^T \mathbf{C}^{-1}(\mathbf{x}(S_i^{(k)}) - \hat{\mu}_0)\right\} \quad (14-3)$$

και ο βαθμός ομοιότητας χρώματος ολόκληρης της k -οστής εικόνας με την εικόνα εισόδου είναι

$$s_c^{(k)} = \max_i \{s_c(S_i^{(k)})\} \quad (14-4)$$

Οι εικόνες της βάσης ταξινομούνται σύμφωνα με το παραπάνω μέτρο ομοιότητας και εκείνες με τη μεγαλύτερη τιμή ομοιότητας ανακαλούνται πρώτες και επιστρέφονται στο χρήστη. Σημειώνεται ότι θα μπορούσε να χρησιμοποιηθεί απευθείας σύγκριση ανάμεσα στις μέσες χρωματικές συνιστώσες του επιλεγμένου τμήματος προσώπου και των τμημάτων που περιέχονται στη βάση· όμως με την παραπάνω τεχνική τα τμήματα προσώπου των ανακαλούμενων εικόνων εξακολουθούν να ικανοποιούν το γενικό μοντέλο χρώματος δέρματος.

Κλίμακα Προσώπων

Το κριτήριο αυτό απευθύνεται σε χρήστες που ενδιαφέρονται για πρόσωπα σε κλίμακα παρόμοια με εκείνη ενός τμήματος προσώπου που παρουσιάζουν ως είσοδο στο σύστημα ανάκλησης. Η περίπτωση αυτή μπορεί να είναι χρήσιμη, για παράδειγμα, για την ανίχνευση κοντινών πλάνων προσώπων ή προσώπων μακριά στο φόντο της εικόνας. Έστω ότι $a(F)$ είναι το ποσοστό της συνολικής επιφάνειας της εικόνας που καλύπτεται από το επιλεγμένο τμήμα προσώπου F , κανονικοποιημένο στο διάστημα $[0,1]$. Τότε ο βαθμός ομοιότητας του τμήματος $S_i^{(k)}$ (του i -οστού τμήματος προσώπου της k -οστής εικόνας στη βάση, όπως στην προηγούμενη περίπτωση) με το επιλεγμένο τμήμα F ορίζεται ως

$$s_r(S_i^{(k)}) = 1 - |a(S_i^{(k)}) - a(F)| \quad (14-5)$$

και ο βαθμός ομοιότητας κλίμακας ολόκληρης της k -οστής εικόνας με την εικόνα εισόδου είναι

$$s_r^{(k)} = \max_i \{s_r(S_i^{(k)})\} \quad (14-6)$$

Και πάλι οι εικόνες της βάσης ταξινομούνται σύμφωνα με το παραπάνω μέτρο ομοιότητας, το οποίο είναι επίσης κανονικοποιημένο στο διάστημα $[0,1]$ ενώ η τιμή $s_r^{(k)} = 1$ αντιστοιχεί σε μέγιστη ομοιότητα, δηλαδή σε ίσες κλίμακες. Οι εικόνες που περιέχουν πρόσωπα με τη μέγιστη τιμή ομοιότητας ανακαλούνται πρώτες και επιστρέφονται στο χρήστη ως αποτέλεσμα της αναζήτησης.

Πλήθος Τμημάτων Προσώπου

Τέλος, ο χρήστης μπορεί να ενδιαφέρεται για εικόνες ή καρέ από ακολουθίες που περιέχουν ένα συγκεκριμένο αριθμό προσώπων. Στην περίπτωση αυτή οι εικόνες της βάσης που περιέχουν τον επιθυμητό αριθμό τμημάτων προσώπου ανακαλούνται από τη βάση, ενώ η σειρά ταξινόμησης των ανακαλούμενων εικόνων καθορίζεται είτε σύμφωνα με την ομοιότητα χρώματος των τμημάτων προσώπου με ένα παράδειγμα προσώπου που δίνει ο χρήστης σε μια εικόνα

να εισόδου, είτε σύμφωνα με την ομοιότητα κλίμακας, είτε τέλος με ένα συνδυασμό και των δύο μέτρων ομοιότητας.

14.4 ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Η ανάκληση με βάση τα πρόσωπα με τα τρία προτεινόμενα κριτήρια δεικτοδότησης και αναζήτησης αξιολογήθηκε σε μία βάση δεδομένων με ακολουθίες βίντεο. Το περιεχόμενο της βάσης δημιουργήθηκε χρησιμοποιώντας 200 συνολικά ακολουθίες μήκους 120-850 καρέ η κάθε μία, οι οποίες καταγράφηκαν από προγράμματα δελτίων ειδήσεων πέντε Ελληνικών καναλιών σε ανάλυση CIF και με ρυθμό 10 frames/sec. Από τα 10850 συνολικά καρέ, 28% περίπου δεν περιέχουν καθόλου πρόσωπα (ή περιέχουν πρόσωπα σε πολύ μικρή κλίμακα), 49% περιέχουν ακριβώς ένα πρόσωπο, ενώ 23% περιέχουν δύο ή περισσότερα πρόσωπα. Η συνολική διαδικασία εντοπισμού προσώπων, που περιλαμβάνει κατάτμηση με βάση το χρώμα, συγχώνευση τμημάτων δέρματος, ανάλυση σχήματος, υπολογισμό πιθανοφάνειας δέρματος και προσώπου και ταίριασμα με πρότυπο, εφαρμόστηκε σε όλο το υλικό της βάσης. Το χρώμα, η κλίμακα, η θέση, ο προσανατολισμός και τα χαρακτηριστικά σχήματος (συμπαγές και επιμήκυνση) αποθηκεύτηκαν στη συνέχεια ως πληροφορία δεικτοδότησης στη βάση.



Σχήμα 14-2. Ανάκληση με βάση το χρώμα προσώπων. (α) Εικόνα εισόδου. Σημειώνεται το επιλεγμένο τμήμα προσώπου. (β) Ανακαλούμενες εικόνες, κατά φθίνουσα σειρά ομοιότητας χρώματος (s_c : ομοιότητα χρώματος).

Παραδείγματα Ανάκλησης

Η ανάκληση αξιολογήθηκε σε ένα περιβάλλον αναζήτησης μέσω παραδείγματος (query-by-example). Οι εικόνες εισόδου που παρουσιάστηκαν στο σύστημα αναζήτησης αναλύθηκαν με τον ίδιο ακριβώς τρόπο για τον εντοπισμό προσώπων και τα χαρακτηριστικά των προσώπων που εντοπίστηκαν συγκρίθηκαν με εκείνα των προσώπων που υπάρχουν αποθηκευμένα στη βάση, σύμφωνα με τα τρία διαφορετικά κριτήρια δεικτοδότησης που προτείνονται. Για παράδειγμα, στο Σχήμα 14-2 εμφανίζονται ενδεικτικά αποτελέσματα αναζήτησης με βάση την ομοιότητα χρώματος. Το μοναδικό πρόσωπο της εικόνας εισόδου του Σχήματος 14-2α ανιχνεύεται σωστά και οι μέσες χρωματικές συνιστώσες του συγκρίνονται με τις διαθέσιμες στη βάση, ανακαλώντας τις οκτώ εικόνες που παρουσιάζονται στο Σχήμα 14-2β κατά φθίνουσα σειρά ομοιότητας χρώματος. Παρατηρείται ότι ανακαλούνται πρόσωπα παρόμοιου χρώματος, ανεξάρτητα από το μέγεθος ή το πλήθος προσώπων.

Το Σχήμα 14-3 παρουσιάζει την περίπτωση της ανάκλησης με βάση την κλίμακα προσώπου. Το κύριο τμήμα προσώπου εξάγεται και πάλι σωστά από την εικόνα εισόδου του Σχήματος 14-3α αλλά τώρα η ομοιότητα υπολογίζεται ως προς την κλίμακα του προσώπου, δηλαδή

ως προς το ποσοστό της συνολικής επιφάνειας εικόνας που καλύπτεται από το πρόσωπο. Όπως φαίνεται στο Σχήμα 14-3β, τα αποτελέσματα αναζήτησης είναι τώρα ανεξάρτητα από το χρώμα και ταξινομούνται κατά φθίνουσα σειρά ομοιότητας κλίμακας. Τέλος, το Σχήμα 14-4 παρουσιάζει αποτελέσματα για το κριτήριο του πλήθους τμημάτων προσώπου. Αφού ανιχνεύονται δύο τμήματα στην εικόνα εισόδου του Σχήματος 14-4α, η έρευνα στη βάση περιορίζεται στις εικόνες εκείνες που περιέχουν δύο ακριβώς πρόσωπα· όμως η ομοιότητα υπολογίζεται ως προς την κλίμακα στην περίπτωση αυτή.



Σχήμα 14-3. Ανάκληση με βάση την κλίμακα προσώπων. (α) Εικόνα εισόδου. Σημειώνεται το επιλεγμένο τμήμα προσώπου. (β) Ανακαλούμενες εικόνες, κατά φθίνουσα σειρά ομοιότητας κλίμακας (s_r : ομοιότητα κλίμακας).



Σχήμα 14-4. Ανάκληση με βάση το πλήθος προσώπων. (α) Εικόνα εισόδου. Σημειώνονται τα δύο επιλεγμένα τμήματα προσώπων. (β) Ανακαλούμενες εικόνες που περιέχουν δύο πρόσωπα, κατά φθίνουσα σειρά ομοιότητας κλίμακας (s_r : ομοιότητα κλίμακας).

Μπορεί να ισχυρισθεί κανείς ότι η ανάκληση στα παραπάνω παραδείγματα είναι επιτυχής· είναι επίσης σαφές ότι τέτοια αποτελέσματα αναζήτησης δεν θα μπορούσαν να επιτευχθούν με τη χρήση γενικών χαρακτηριστικών των εικόνων, όπως η χρωματική σύνθεση ή το σχήμα και η υφή των αντικειμένων, χωρίς γνώση της δομής του ανθρώπινου προσώπου. Βέβαια τα αποτελέσματα που παρουσιάστηκαν είναι ποιοτικά και μια ποσοτική αξιολόγηση δεν μπορεί να πραγματοποιηθεί άμεσα, κυρίως λόγω της υποκειμενικής φύσης του προβλήματος. Πιο αξιόπιστα πειράματα που λαμβάνουν υπόψη τους την ανθρώπινη αντίληψη θα μπορούσαν να διεξαχθούν αναθέτοντας σε μια ομάδα κριτών να αξιολογήσουν κατά πόσο οι ανακαλούμενες είναι σχετικές με την εικόνα εισόδου.

Ποσοτική Αξιολόγηση

Στην ειδική περίπτωση της αναζήτησης με βάση το πλήθος προσώπων πραγματοποιήθηκε ένα είδος ποσοτικής αξιολόγησης σε ένα υποσύνολο της βάσης το οποίο αποτελείται από 500 καρτέ. Τα καρτέ αυτά ταξινομήθηκαν χειρονακτικά σε τέσσερις κατηγορίες ανάλογα με το πλήθος

προσώπων που περιέχουν, όπως φαίνεται στον Πίνακα 14-1. Η ταξινόμηση αυτή θεωρείται *αληθής* (ground truth). Στη συνέχεια εκτελέσθηκε μία αναζήτηση για κάθε κατηγορία και οι ανακαλούμενες εικόνες χαρακτηρίστηκαν «σωστές» αν περιείχαν τον ίδιο ακριβώς αριθμό προσώπων με τον επιθυμητό, δηλαδή τον αριθμό προσώπων που προέκυψαν χειρονακτικά. Από τη σύγκριση αυτή των αποτελεσμάτων αναζήτησης με την αληθή ταξινόμηση υπολογίσθηκαν τιμές *ακρίβειας* (precision) και *ανάκλησης* (recall) για κάθε κατηγορία. Όπως και στις περιπτώσεις εντοπισμού προσώπων και ταξινόμησης πλάνων στα Κεφάλαια 12 και 13, η ακρίβεια μετρήθηκε ως ο λόγος σωστά ανακαλούμενων εικόνων προς τις συνολικά ανακαλούμενες εικόνες, ενώ η ανάκληση ως ο λόγος των σωστά ανακαλούμενων εικόνων προς τις συνολικές εικόνες.

Πλήθος προσώπων	Κανένα	1	2	3+
Συνολικές εικόνες	62	275	112	51
Ανακαλούμενες εικόνες	66	282	106	46
Σωστά ανακαλούμενες	59	273	101	42
Ακρίβεια (%)	89.34	96.81	93.57	91.30
Ανάκληση (%)	95.16	99.27	87.61	82.35

Πίνακας 14-1. Αριθμητική αξιολόγηση ανάκλησης με βάση το πλήθος προσώπων.

Παρατηρείται από τις μετρήσεις ακρίβειας και ανάκλησης του Πίνακα 14-1 ότι καλύτερα αποτελέσματα λαμβάνονται στην περίπτωση του ενός προσώπου· αυτό είναι αναμενόμενο αφού το ένα πρόσωπο είναι συνήθως σε κοντινό πλάνο και επομένως σε μεγαλύτερη ανάλυση, οπότε ανιχνεύεται ευκολότερα. Επιπλέον, πολλά καρέ με ένα πρόσωπο εμφανίζουν πρόσοψη παρουσιαστών ή ομιλητών σε εσωτερική λήψη όπου οι συνθήκες φωτισμού είναι ελεγχόμενες. Η μειωμένη ανάλυση είναι ο κυριότερος παράγοντας για τις χαμηλότερες τιμές ανάκλησης στις περιπτώσεις των δύο ή περισσότερων προσώπων, καθώς προκαλεί μεγαλύτερο αριθμό *απωλειών* (dismissals). Η ακρίβεια της κατηγορίας «κανένα πρόσωπο» είναι χαμηλότερη για τον ίδιο ακριβώς λόγο, αφού όταν τα πρόσωπα δεν ανιχνεύονται οι αντίστοιχες εικόνες κατατάσσονται στην κατηγορία αυτή. Σημειώνεται τέλος ότι μία παρόμοια ποσοτική αξιολόγηση για τις περιπτώσεις ανάκλησης με βάση το χρώμα και την κλίμακα προσώπου δεν είναι εύκολη διότι σε αυτές τις περιπτώσεις δε μπορεί να καθοριστεί αντικειμενικά το επιθυμητό αποτέλεσμα (ground truth).

Συμπεράσματα

Η παραδοσιακή ακολουθιακή ή γραμμική αναπαράσταση βίντεο με διαδοχικά καρέ έχει σημαντικούς περιορισμούς για τις μελλοντικές εφαρμογές πολυμέσων όπως η δεικτοδότηση, η πλοήγηση και η ανάκληση με βάση το περιεχόμενο, κυρίως λόγω του τεράστιου όγκου πληροφοριών που εμπλέκονται και των αντίστοιχων απαιτήσεων χωρητικότητας αποθήκευσης ή εύρους ζώνης μετάδοσης. Μία από τις σημαντικότερες συνεισφορές της παρούσας διατριβής είναι ότι επιτυγχάνει μη γραμμική αναπαράσταση ως προς το χρόνο, με βάση το οπτικό περιεχόμενο. Οι ακολουθίες διαμερίζονται αρχικά σε διαδοχικά πλάνα και στη συνέχεια πραγματοποιείται ουσιαστικά μη ομοιόμορφη δειγματοληψία με βάση το περιεχόμενο, απορρίπτοντας πλάνα ή καρέ με παρόμοιο οπτικό περιεχόμενο. Η προσέγγιση αυτή παρέχει αυτόματη εξαγωγή περιλήψης της οπτικής πληροφορίας με τρόπο παρόμοιο με εκείνο που χρησιμοποιείται σε συστήματα διαχείρισης εγγράφων και πληροφορίας κειμένου.

Η χρησιμοποιούμενη πολυδιακριτική υλοποίηση του αλγορίθμου RSST για την κατάτμηση με βάση το χρώμα, την κίνηση και το πεδίο βάθους, πέρα από τη δραματική επιτάχυνση της διαδικασίας κατάτμησης, επιτυγχάνει και μείωση του φαινομένου του oversegmentation. Το πεδίο βάθους υπολογίζεται στην περίπτωση των στερεοσκοπικών ακολουθιών από την απόκλιση μεταξύ των εικόνων των δύο καναλιών, ενώ οι επικαλυπτόμενες περιοχές ανιχνεύονται και αντισταθμίζονται αποτελεσματικά με κατάλληλες τιμές απόκλισης. Επιπλέον, με τη συγχώνευση της πληροφορίας χρώματος και βάθους επιτυγχάνεται ικανοποιητική προσέγγιση των σημασιολογικών αντικειμένων, και μάλιστα με μεγάλη ακρίβεια στα περιγράμματά τους. Πιο αξιόπιστη ανάλυση μπορεί να επιτευχθεί με την παρακολούθηση (tracking) των αντικειμένων, ώστε να μειωθεί η επίδραση του θορύβου κατάτμησης [83,142]. Τα πειράματα εξαγωγής περιλήψης δείχνουν ότι επιτυγχάνεται καλύτερη αναπαράσταση με τη χρήση της πληροφορίας βάθους, γεγονός που δικαιολογεί το επιπλέον υπολογιστικό κόστος εκτίμησης του πεδίου απόκλισης. Η όλη διαδικασία ανάλυσης είναι προσανατολισμένη στην εκμετάλλευση πληροφορίας που είναι διαθέσιμη σε ακολουθίες MPEG, με κριτήριο την ταχεία υλοποίηση. Στην περίπτωση αυτή, παράμετροι όπως οι μέσες χρωματικές συνιστώσες στα block των εικόνων και τα διανύσματα κίνησης είναι ήδη υπολογισμένα· έτσι απαιτείται μικρού μόνο βαθμού αποκωδικοποίηση και ταυτόχρονα υπάρχει σημαντικό όφελος στην ταχύτητα επεξεργασίας.

Με βάση την κατάτμηση ακολουθιών σε αντικείμενα, τα χαρακτηριστικά του κάθε αντικειμένου, όπως η θέση, το μέγεθος, το χρώμα, η κίνηση και το βάθος, χρησιμοποιούνται για την κατασκευή διανυσμάτων περιγραφής μέσω ασαφών πολυδιάστατων ιστογραμμάτων, μειώνοντας έτσι την επίδραση των ασυνεχειών και του θορύβου κατάτμησης. Η αναπαράσταση του οπτικού περιεχομένου με διανύσματα περιγραφής σταθερών διαστάσεων, στα οποία υπάρχει ένα προς ένα αντιστοιχία των επιμέρους στοιχείων, επιτρέπει ταχεία και αξιόπιστη σύγκριση εικόνων, καρέ ή πλάνων. Έτσι, με χρήση της Ευκλείδειας απόστασης διανυσμάτων περιγραφής, είναι η δυνατή η ομαδοποίηση των πλάνων παρόμοιου οπτικού περιεχομένου και η επιλογή ενός περιορισμένου αριθμού αντιπροσωπευτικών πλάνων με το γενικευμένο αλγόριθμο Lloyd-Max, ο οποίος έχει χαμηλό υπολογι-

οτικό κόστος. Σημειώνεται ότι η επιλογή γίνεται πάντα με βάση το περιεχόμενο· επιλογή πλάνων σε ισοπέχοντα χρονικά διαστήματα δε μπορεί να αποδώσει σωστά το οπτικό περιεχόμενο.

Για κάθε πλάνο, επιλέγεται ένας περιορισμένος αριθμός χαρακτηριστικών καρέ, με δύο εναλλακτικές μεθόδους: την επιλογή ακραίων σημείων της τροχιάς του διανύσματος περιγραφής ως συνάρτησης του χρόνου και την ελαχιστοποίηση ενός κριτηρίου συσχέτισης των διανυσμάτων περιγραφής των καρέ. Η πρώτη μέθοδος είναι πολύ γρήγορη και απλή στην υλοποίηση· επιπλέον, ανιχνεύει τη δράση ενός πλάνου, επιτρέποντας την εκτίμηση του βέλτιστου πλήθους χαρακτηριστικών καρέ, και λειτουργεί ακόμη και σε περιπτώσεις περιοδικής επανάληψης του οπτικού περιεχομένου. Αντίθετα, η δεύτερη μέθοδος επιλέγει χαρακτηριστικά καρέ με βέλτιστο τρόπο, βάσει της ομοιότητας. Επειδή η εξαντλητική αναζήτηση της βέλτιστης λύσης είναι πρακτικά ανέφικτη, προτείνεται ένας αλγόριθμος λογαριθμικής αναζήτησης, μία στοχαστική εκδοχή του αλγορίθμου αυτού, καθώς και ένας γενετικός αλγόριθμος για την υλοποίηση της μεθόδου συσχέτισης. Η πειραματική αξιολόγηση δείχνει ότι η επίδοση του γενετικού αλγορίθμου είναι ανώτερη όσον αφορά στην ταχύτητα αλλά και στην ακρίβεια των αποτελεσμάτων, ενώ η μέθοδος χρονικής μεταβολής αποτελεί ένα ισχυρό εργαλείο προ-επεξεργασίας για την εκτίμηση της πολυπλοκότητας και της δράσης ενός πλάνου.

Η χρήση της προτεινόμενης αναπαράστασης με διανύσματα περιγραφής των επιλεγμένων καρέ δίνει τη δυνατότητα ανάκλησης βάσει παραδείγματος ή σχεδιαγράμματος και ουσιαστικά ανάγει το πρόβλημα της ανάκλησης βίντεο σε ανάκληση ακίνητων εικόνων. Επιπλέον, ο μηχανισμός ανάδρασης, ακριβώς όπως και στα συστήματα αναζήτησης πληροφορίας εγγράφων με βάση το κείμενο, επιτρέπει την αλληλεπίδραση ανάμεσα στο χρήστη και το υπολογιστικό σύστημα και αυξάνει την ευελιξία ανάκλησης, αφού διαφορετικές εφαρμογές ή διαφορετικοί χρήστες μπορεί να απαιτούν διαφορετικούς τύπους αναζήτησης. Τα πειραματικά αποτελέσματα αναζήτησης από πραγματικές ακολουθίες βίντεο φανερώνουν ότι ένα απλό στάδιο ανάδρασης είναι στις περισσότερες περιπτώσεις επαρκές, ενώ στην περίπτωση που απαιτείται πολλαπλή ανάδραση το μεγαλύτερο ποσοστό βελτίωσης επιτυγχάνεται στην πρώτη επανάληψη.

Η προτεινόμενη τεχνική κανονικοποίησης περιγραμμάτων που παρουσιάστηκε στο τρίτο μέρος της διατριβής επιτρέπει την αναλλοίωτη περιγραφή ως προς τους affine μετασχηματισμούς χωρίς καμία απώλεια πληροφορίας σχήματος. Η τεχνική μπορεί να εφαρμοσθεί ως ένα στάδιο προ-επεξεργασίας σε συνδυασμό με οποιαδήποτε μέθοδο αναπαράστασης, ταξινόμησης, αναγνώρισης ή ανάκλησης, αφού ουσιαστικά αποσυνδέει το πρόβλημα της αναλλοίωτης περιγραφής από την εξαγωγή χαρακτηριστικών και το ταίριασμα. Το γεγονός αυτό επιβεβαιώνεται με ένα αριθμό πειραμάτων με διάφορα μέτρα απόστασης καμπυλών στο πλαίσιο της ταξινόμησης και ανάκλησης αντικειμένων με βάση το σχήμα από βάσεις εικόνων και βίντεο. Σε κάθε περίπτωση, τα πειράματα έδειξαν ότι η προτεινόμενη κανονικοποίηση είναι ανθεκτική στο θόρυβο και στις περιορισμένες παραμορφώσεις σχήματος. Επιπλέον, το υπολογιστικό κόστος της τεχνικής είναι πολύ μικρό και επομένως μπορεί να ενσωματωθεί σε οποιοδήποτε σύστημα πραγματικού χρόνου για ανάκληση με βάση το περιεχόμενο ή ακόμη και για κωδικοποίηση βίντεο.

Η βελτίωση της τεχνικής είναι δυνατή προς διάφορες κατευθύνσεις. Κατ' αρχάς, μπορεί να γενικευθεί ώστε να χειρίζεται σύνολα από ανοικτές ή κλειστές καμπύλες, είτε συνεκτικές είτε όχι, για το σκοπό της οπτικής αναγνώρισης χαρακτήρων ή τη δεικτοδότηση σε βάσεις δεδομένων με τεχνικά διαγράμματα. Επίσης είναι επιθυμητή η κανονικοποίηση διδιάστατων δεδομένων από εικόνες, πέρα από τα περιγράμματα αντικειμένων, ή ακόμη και η κανονικοποίηση τρισδιάστατων μοντέλων και πολυδιάστατων δεδομένων. Τέλος ένα πολύ ενεργό πεδίο έρευνας είναι η κανονικοποίηση ως προς άλλους, μη γραμμικούς μετασχηματισμούς, και κυρίως ως το μετασχηματισμό προοπτικής προβο-

λής. Τεχνικές υπολογισμού αναλλοίωτων ποσοτήτων ως προς το μετασχηματισμό αυτό υπάρχουν [103], αλλά δεν επιτρέπουν κανονικοποίηση και δεν έχουν την ιδιότητα διατήρησης της πληροφορίας σχήματος. Τα παραπάνω ζητήματα μπορούν να αποτελέσουν αντικείμενο μελλοντικής έρευνας.

Η τεχνική εντοπισμού ανθρώπινων προσώπων σε ακολουθίες βίντεο που εξετάσθηκε στο τελευταίο μέρος της διατριβής μπορεί να χρησιμοποιηθεί σε διάφορες εφαρμογές, όπως η χρονική κατάτμηση, η εξαγωγή περίληψης και η πλοήγηση οπτικοακουστικού υλικού, καθώς και η συντακτική ανάλυση και δεικτοδότηση τηλεοπτικών δελτίων ειδήσεων. Η χρήση των χαρακτηριστικών υφής και του ταιριάσματος με πρότυπο προσώπου βελτιώνει σημαντικά την ακρίβεια του αλγορίθμου, με κόστος την αυξημένη πολυπλοκότητα. Όμως η τελευταία είναι μειωμένη έως και 100 φορές σε σχέση με την εξαντλητική αναζήτηση σε ολόκληρη την εικόνα (όπως γίνεται στις ασπρόμαυρες εικόνες), αφού εξετάζονται μόνο οι περιοχές στις οποίες ανιχνεύεται δέρμα. Το μοντέλο χρώματος δέρματος προσαρμόζεται σε κάθε καρέ μιας ακολουθίας βίντεο, παρέχοντας την απαιτούμενη γενίκευση για την αποτελεσματική ανίχνευση σε μεταβαλλόμενες συνθήκες φωτισμού. Η τεχνική κατάτμησης με βάση το χρώμα σε συνδυασμό με τη συγχώνευση τμημάτων δέρματος επιτυγχάνει εξαγωγή ενός μόνο τμήματος για κάθε πρόσωπο και δίνει καλύτερα αποτελέσματα σε σχέση με την κατωφλίωση της πιθανοφάνειας δέρματος ως προς το ποσοστό ανάκλησης, ενώ η χρήση χαρακτηριστικών σχήματος βελτιώνει σημαντικά το ποσοστό ακρίβειας. Η συνολική πειραματική αξιολόγηση δείχνει ότι η μέθοδος αντιμετωπίζει το πρόβλημα της ακριβούς ανίχνευσης επιτυγχάνοντας ταυτόχρονα ταχεία υλοποίηση.

Η προτεινόμενη τεχνική χρονικής κατάτμησης, συντακτικής ανάλυσης και δεικτοδότησης δελτίων ειδήσεων με ανάλυση της οπτικής πληροφορίας δίνει ενθαρρυντικά αποτελέσματα, παρά το γεγονός ότι σε συνδυασμό με τον εντοπισμό προσώπων χρησιμοποιείται μόνο ένα μικρό σύνολο κανόνων που βασίζεται σε απλά οπτικά χαρακτηριστικά όπως τα ιστογράμματα χρώματος, οι απόλυτες διαφορές καρέ και η κίνηση. Αν και η σημασιολογική ανάλυση σε πραγματικές θεματικές ενότητες θα απαιτούσε ταυτόχρονα και ηχητική πληροφορία καθώς και πληροφορία κειμένου, η επίδοση της χρησιμοποιούμενης τεχνικής εντοπισμού προσώπων επιτρέπει αξιόπιστη ανάλυση ακόμη και όταν οι πληροφορίες αυτές δεν είναι διαθέσιμες. Η προτεινόμενη τεχνική ανάλυσης μπορεί είτε να χρησιμοποιηθεί σαν ανεξάρτητη εφαρμογή για δελτία ειδήσεων που δεν συνοδεύονται από υπότιτλους ή εγγραφές teletext, είτε και να ενσωματωθεί σε υφιστάμενα συστήματα που βασίζονται σε πληροφορία κειμένου και ήχου.

Επιπλέον, τα πειράματα δείχνουν ότι πέρα από την αυτοματοποίηση της διαδικασίας δεικτοδότησης ανθρώπινων προσώπων, η χρήση χαρακτηριστικών των προσώπων μπορεί επίσης να οδηγήσει σε νέα κριτήρια αναζήτησης σε εφαρμογές ανάκλησης μέσω παραδείγματος. Τα τρία απλά κριτήρια που προτείνονται, δηλαδή της αναζήτησης με βάση το χρώμα, το μέγεθος και το πλήθος προσώπων, μπορούν να συνδυαστούν και με γενικότερα κριτήρια που βασίζονται σε χαρακτηριστικά χρώματος, κίνησης, σχήματος ή υφής, επιτρέποντας τη γενίκευση των σεναρίων αναζήτησης. Έτσι παρέχεται ένα ισχυρό εργαλείο σημασιολογικής αναζήτησης που μπορεί να αποτελέσει τμήμα υφιστάμενων συστημάτων δεικτοδότησης και ανάκλησης.

B ι β λ ι ο γ ρ α φ ι α

- [1] Alatan A., Onural L., Wollborn M., Mech R., Tuncel E., and Sikora T., "Image Sequence Analysis for Emerging Interactive Multimedia Services – The European Cost 211 Framework," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 7, pp. 802-813, Nov. 1998.
- [2] Androutsos D., Plataniotis K. N., and Venetsanopoulos A. N., "Extraction of Detailed Image Regions for Content-Based Image Retrieval," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Seattle, WA, May 1998.
- [3] Ariki Y. and Saito Y., "Extraction of TV News Articles Based on Scene Cut Detection Using DCT Clustering," in *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, Vol. 3, pp. 847-850, 1996.
- [4] Arkin E. M., Chew L. P., Huttenlocher D. P., Kadem K., and Mitchell J. S. B., "An Efficiently Computable Metric for Comparing Polygonal Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 3, pp. 209-216, 1991.
- [5] Arman F., Depommier R., Hsu A., and Chiu M-Y., "Content-based Browsing of Video Sequences," in *Proc. of ACM Multimedia Conf.*, pp. 97-103, San Francisco, CA, Oct. 1994.
- [6] Avrithis Y. S. and Kollias S. D., "Fuzzy Image Classification Using Multiresolution Neural Networks with Applications to Remote Sensing," in *Proc. of 13th Int. Conf. on Digital Signal Processing (DSP)*, Vol. 1, pp. 261-264, Santorini, Greece, July 1997.
- [7] Avrithis Y. S., Delopoulos A. N., and Papageorgiou G. C., "Ultrasonic Array Imaging Using CDMA Techniques," in *Proc. of IX European Signal Processing Conf. (EUSIPCO)*, Rhodes, Greece, September 1998.
- [8] Avrithis Y. S., Doulamis A. D., Doulamis N. D., and Kollias S. D., "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," *Computer Vision and Image Understanding*, Vol. 75, No. 1/2, pp. 3-24, July/August 1999.
- [9] Avrithis Y. S., Tsapatsoulis N. A., and Kollias S. D., "Broadcast News Parsing Using Visual Cues: A Robust Face Detection Approach," in *Proc. of IEEE Int. Conf. on Multimedia and Expo (ICME)*, Vol. 3, pp. 1469-1472, New York City, NY, July-August 2000.
- [10] Avrithis Y. S., Xirouhakis Y. S., and Kollias S. D., "Affine Invariant Representation and Classification of Object Contours for Image and Video Retrieval," in *Computational Intelligence and Applications*, N. Mastorakis (editor), World Scientific and Engineering Society Press, pp. 342-347, 1999.
- [11] Avrithis Y. S., Xirouhakis Y. S., and Kollias S. D., "Affine Invariant Representation and Classification of Object Contours for Image and Video Retrieval," in *Proc. of 3rd WSES/IEEE/IMACS World Multiconference on Circuits, Systems, Communications and Computers (CSCC)*, Athens, Greece, July 1999.
- [12] Avrithis Y. S., Xirouhakis Y. S., and Kollias S. D., "Affine-Invariant Curve Normalization for Object Shape Representation, Classification and Retrieval," *Machine Vision and Applications* (submitted for publication).
- [13] Avrithis Y., Doulamis A., Doulamis N., and Kollias S., "An Adaptive Approach to Video Indexing and Retrieval Using Fuzzy Classification," in *Proc. of Workshop on Very Low Bit Rate Video Coding (VLBV)*, Urbana-Champaign, IL, Oct. 1998.

- [14] Avrithis Y., Doulamis N., Doulamis A., and Kollias S., "Efficient Content Representation in MPEG Video Databases," in *Proc. of IEEE Workshop of Content-Based Access of Image and Video Libraries (CBAIVL)*, pp. 91-95, Santa Barbara, CA, June 1998.
- [15] Avrithis Y., Tsapatsoulis N., and Kollias S., "Color-Based Retrieval of Facial Images," in *Proc. of IX European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, September 2000.
- [16] Avrithis Y., Xirouhakis Y., and Kollias S., "Affine-Invariant Curve Normalization for Shape-Based Retrieval," in *Proc. of Int. Conf. on Pattern Recognition (ICPR)*, Barcelona, Spain, Sept. 2000.
- [17] Balslev I., "Noise Tolerance of Moment Invariants in Pattern Recognition," *Pattern Recognition Letters*, Vol. 19, pp. 1183-1189, 1998.
- [18] Bhanu B., Peng J., and Qing S., "Learning Feature Relevance and Similarity Metrics in Image Databases," in *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)*, Santa Barbara, CA, June 1998.
- [19] Bimbo A. D. and Pala P., "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 2, pp. 121-132, 1997.
- [20] Blum H., "A Transformation for Extracting New Descriptors of Shape," in *Models for the Perception of Speech and Visual Form*, W. Wathen-Dum (editor), MIT Press, Cambridge, Mass, 1967.
- [21] Bouthemmy P., Gelgon M., and Ganancia F., "A Unified Approach to Shot Change Detection and Camera Motion Characterization," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 9, No. 7, pp.1030-1044, Oct. 1999.
- [22] Brown M., Foote J., Jones G., Sparck-Jones K., and Young S., "Automatic Content-Based Retrieval of Broadcast News," in *Proc. of ACM Multimedia Conf.*, San Francisco, CA, Nov. 1995.
- [23] Castagno R., Ebrahimi T., and Kunt M., "Video Segmentation Based on Multiple Features for Interactive Multimedia Applications," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 562-571, Sept. 1998.
- [24] CCITT Recommendation H.261, "Video Codec for Audiovisual Data at p×64 kb/s," Geneva, 1990.
- [25] Chang S.-F., Chen W., Meng H. J., Sundaram H., and Zhong D., "A Fully Automated Content-Based Video Search Engine Supporting Spatiotemporal Queries," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, Sept. 1998.
- [26] Chang S.-F., Eleftheriadis A., and McClintock R., "Next-Generation Content Representation Creation, and Searching for New-Media Applications in Education," *Proceedings of the IEEE*, Vol. 86, No. 5., pp. 884-904, May 1998.
- [27] Chiariglione L., "MPEG and Multimedia Communications," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, pp. 5-18, Feb. 1997.
- [28] Cohen F. S., Huang Z., and Yang Z., "Invariant Matching and Identification of Curves using B-Splines Curve Representation," *IEEE Trans. Image Processing*, Vol. 4, No. 1, pp. 1-10, Jan. 1995.
- [29] Dell'Acqua F. and Gamba P., "Simplified Model analysis and Search for Reliable Shape Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 656-666, Sept. 1998.
- [30] DeMenthon D., Kobla V., and Doermann D., "Video Summarization by Curve Simplification," in *Proc. of ACM Multimedia Conf.*, pp. 211-218, Bristol, UK, 1998.
- [31] Doulamis A. D., Avrithis Y. S., Doulamis N. D., and Kollias S. D., "A Genetic Algorithm for Efficient Video Content Representation," in *Computational Intelligence in Systems and Control – Design and Applications*, S. G. Tzafestas (editor), Kluwer Academic Publishers, Jan. 2000.

- [32] Doulamis A. D., Avrithis Y. S., Doulamis N. D., and Kollias S. D., "Indexing and Retrieval of the Most Characteristic Frames / Scenes in Video Databases," in *Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 105-110, Louvain-la-Neuve, Belgium, June 1997.
- [33] Doulamis A. D., Avrithis Y. S., Doulamis N. D., and Kollias S. D., "A Genetic Algorithm for Efficient Video Content Representation," in *Proc. of IMACS/IFAC Int. Symp. on Soft Computing in Engineering Applications (SOFTCOM)*, Athens, Greece, June 1998.
- [34] Doulamis A. D., Doulamis N. D., Avrithis Y. S., and Kollias S. D., "A Fuzzy Video Content Representation for Video Summarization and Content-Based Retrieval," *Signal Processing Journal*, Vol. 80, No. 6, June 2000.
- [35] Doulamis A., Avrithis Y., Doulamis N., and Kollias S., "Interactive Content-Based Retrieval in Video Databases Using Fuzzy Classification and Relevance Feedback," in *Proc. of IEEE Int. Conf. on Multimedia Computing and Systems (ICMCS)*, Florence, Italy, June 1999.
- [36] Doulamis N. D., Doulamis A. D., Avrithis Y. S., and Kollias S. D., "Video Content Representation Using Optimal Extraction of Frames and Scenes," in *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, Vol. 1, pp. 875-879, Chicago, IL, October 1998.
- [37] Doulamis N. D., Doulamis A. D., Avrithis Y. S., Kollias S. D., "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," in *Proc. of IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, pp. 141-146, Copenhagen, Denmark, September 1999.
- [38] Doulamis N. D., Doulamis A. D., Avrithis Y. S., Ntalianis K. S., and Kollias S. D., "An Optimal Framework for Summarization of Stereoscopic Video Sequences," in *Proc. of Int. Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI)*, Santorini, Greece, September 1999.
- [39] Doulamis N., Doulamis A., Avrithis Y., Ntalianis K., and Kollias S., "Efficient Summarization of Stereoscopic Video Sequences," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, No. 4, pp. 501-517, June 2000.
- [40] Eickeler S., Kosmala A., and Rigoll G., "A New Approach to Content-Based Video Indexing Using Hidden Markov Models," in *Proc. of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS)*, Louvain-la-Neuve, Belgium, June 1997.
- [41] Flickner M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele D., and Yanker P., "Query by Image and Video Content: the QBIC System," *IEEE Computer Magazine*, pp. 23-32, Sept. 1995.
- [42] Freeman H., "Boundary Encoding and Processing," in *Picture Processing and Psychopictorics*, Lipkin and Rosenfeld (editors), Academic Press, New York, 1970.
- [43] Garcia C. and Tziritas G., "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," *IEEE Trans. Multimedia*, Vol. 1, No. 3, pp. 264-277, 1999.
- [44] Garrido L., Marques F., Pardas M., Salembier P., and Vilaplana V., "A Hierarchical Technique for Image Sequence Analysis," in *Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 13-20, Louvain-la-Neuve, Belgium, June 1997.
- [45] Gelgon M. and Bouthemy P., "A Hierarchical Motion-Based Segmentation and Tracking Technique for Video Storyboard-Like Representation and Content-Based Indexing," in *Proc. of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS)*, Louvain-la-Neuve, Belgium, June 1997.
- [46] Gersho A. and Gray R. M., *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1993.
- [47] Girdwood C. and Chiwy P., "MIRAGE: An ACTS Project in Virtual Production and Stereoscopy," *IBC Conference Publication*, No. 428, pp. 155-160, Sept. 1996.
- [48] Goldberg D. E., *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison Wesley, 1989.

- [49] Gonzalez R. C. and Woods R. E., *Digital Image Processing*, Addison-Wesley, 1992.
- [50] Grammalidis N. and Strintzis M. G., "Disparity and Occlusion Estimation in Multicocular Systems and Their Coding for the Communication of Multiview Image Sequences," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 3, pp. 328-344, June 1998.
- [51] Gu C. and Lee M.-C., "Semiautomatic Segmentation and Tracking of Semantic Video Objects," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 572-584, Sept. 1998.
- [52] Hamrapur A., Gupta A., Horowitz B., Shu C. F., Fuller C., Bach J., Gorkani M., and Jain R., "Virage Video Engine," in *SPIE Proc. Storage and Retrieval for Video and Image Databases V*, pp. 188-197, San Jose, CA, Feb. 1997.
- [53] Han K. J. and Tewfik A. H., "Eigen-Image Video Segmentation and Indexing," in *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, pp. 538-541, Santa Barbara, CA, Oct. 1997.
- [54] Harwood L. A., "A Chrominance Demodulator IC with Dynamic Flesh Correction," *IEEE Trans. Consumer Electronics*, Vol. 22, No. 2, pp. 111-117, 1976.
- [55] Hu M. K., "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Information Theory*, Vol. 8, pp. 179-187, 1962.
- [56] Huang Z. and Cohen F. S., "Affine-invariant B-Spline Moments for Curve Matching," *IEEE Trans. Image Processing*, Vol. 5, No. 10, pp. 1473-1480, 1996.
- [57] Hush D. R. and Horne B. G., "Progress in Supervised Neural Networks," *IEEE Signal Processing Magazine*, Jan. 1993.
- [58] Idris F. M. and Panchanathan S., "Spatio-Temporal Indexing of Vector Quantized Video Sequences," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 728-740, Oct. 1997.
- [59] *IEEE Workshop on Content Based Access of Image and Video Libraries (CBAIVL)*, Santa Barbara, 21 June, 1998.
- [60] Ioannou S., Moschovitis G., Ntalianis K., Karpouzis K., and Kollias S., "Effective Access to Large Audio-visual Assets Based on User Preferences," *ACM Multimedia*, Marina del Ray, CA, Nov. 2000.
- [61] Ip H. S. and Shen D., "An Affine-Invariant Active Contour Model (AI-Snake) for Model-Based Segmentation," *Image and Vision Computing*, Vol. 16, No. 2, pp. 135-146, 1998.
- [62] Irani M. and Anandan P., "Video Indexing Based on Mosaic Representation," *Proceedings of the IEEE*, Vol. 86, No. 5, pp. 805-921, May 1998.
- [63] ISO/CD 11172-2, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbps," March 1991.
- [64] ISO/IEC 13818-2, "Generic Coding of Moving Pictures and Associated Audio," H.262, Committee Draft, May 1994.
- [65] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Video Verification Model Version 11.0," Doc. N2172, 1998.
- [66] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Overview (v. 1.0)," Doc. N3158, Dec. 1999.
- [67] ISO/IEC JTC1/SC29/WG11, "MPEG-7: Context and Objectives (v.5)," Doc. N1920, Oct. 1997.
- [68] ITU-T SG 15 Experts Group for Very Low Bitrate Visual Telephony, Draft Recommendation H.263, Feb. 1995.
- [69] Jain A. K., Zhong Y., and Lakshmanan S., "Object Matching Using Deformable Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 3, pp. 267-278, 1996.
- [70] Jiang X., Bunke H., and Widmer-Kljajo D., "Skew Correction of Document Images by Focused Nearest-Neighbor Clustering," in *Proc. of 5th Int. Conf. on Document Analysis and Recognition (ICDAR'99)*, pp 629-632, Bangalore, 1999.

- [71] Jones M. J. and Rehg M. R., "Statistical Color Models with Application to Skin Detection," *Compaq Cambridge Research Lab Technical Report*, CRL 98/11, 1998.
- [72] Karpouzis K., Votsis G., Xirouhakis Y., Stamou G., and Kollias S., "An Intelligent Multimedia System for Efficient Image Browsing and Retrieval," in *Proc. of European Workshop on Content-Based Multimedia Indexing (CBMI)*, Toulouse, France, Nov. 1999.
- [73] Khotanzad A. and Hong Y. H., "Invariant Image Recognition by Zernike Moments," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, No. 5, pp. 489-497, 1990.
- [74] Kobla V., Doermann D., and Faloutsos C., "Video Trails: Representing and Visualizing Structure in Video Sequences," in *Proc. of ACM Multimedia Conf.*, pp. 335-346, Seattle, WA, Nov. 1997.
- [75] Kompatsiaris I., Tzovaras D., and Strintzis M. G., "3-D Model Based Segmentation of Videoconference Image Sequences," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 547-561, Sept. 1998.
- [76] Kosko B., *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, 1992.
- [77] Lai K. K. and Chin R. T., "Deformable Contours: Modeling and Extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, No. 11, pp. 1084-1090, 1995.
- [78] Lei Z., Chan Y., and Lopresti D., "Image Curvelet Feature Extraction and Matching," in *Proc. of Int. Conf. on Image Processing (ICIP)*, Santa Barbara, CA, 1997.
- [79] Li C.-S., Mahan R., and Smith J., "Multimedia Content Description in the Infopyramid," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3789-3792, Seattle, WA, May 1998.
- [80] Lin H.-C., Wang L.-L., and Yang S.-N., "Color Image Retrieval Based on Hidden Markov Models," *IEEE Trans. Image Processing*, Vol. 6, No. 2, pp. 332-339, Feb. 1997.
- [81] Llados J., Bunke H., and Marti E., "Finding rotational symmetries by cyclic string matching," *Pattern Recognition Letters*, Vol. 18, No. 14, pp. 1435-1442, 1997.
- [82] Ma W. Y. and Manjunath B. S., "Netra: A Toolbox for Navigating Large Image Databases," in *Proc. of Int. Conf. on Image Processing (ICIP)*, Santa Barbara, CA, Oct. 1997.
- [83] Malassiotis S. and Strintzis M., "Tracking Textured Deformable Objects Using a Finite-Element Mesh," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, pp. 756-774, Oct. 1998.
- [84] Marques F., Llorens B., and Gasull A., "Prediction of Image Partitions Using Fourier Descriptors: Application to Segmentation-Based Coding Schemes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, No. 4, pp. 529-542, 1998.
- [85] Meier T. and Ngan K. N., "Automatic Segmentation of Moving Objects for Video Object Plane Generation," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 525-538, Sept. 1998.
- [86] Merialdo B., "Automatic Indexing of TV News," in *Proc. of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS)*, Louvain-la-Neuve, Belgium, June 1997.
- [87] Merlino A., Morey D., and Maybury M., "Broadcast News Navigation Using Story Segments," in *Proc. of ACM Multimedia Conf.*, Seattle, WA, Nov. 1997.
- [88] Michalewicz Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Verlag, 1994.
- [89] Mills M., Cohen J., and Wong Y. Y., "A Magnifier Tool for Video Data," in *Proc. ACM Computer Human Interface (CHI)*, pp. 93-98, May 1992.
- [90] Mokhtarian F., "Silhouette-Based Isolated Object Recognition through Curvature Scale Space," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 17, No. 5, pp. 539-544, 1995.
- [91] Morris O. J., Lee M. J., and Constantinides A. G., "Graph Theory for Image Analysis: An Approach Based on the Shortest Spanning Tree," *IEE Proceedings*, Vol. 133, pp. 146-152, April 1986.

- [92] Mulroy P. J., "Video Content Extraction: Review of Current Automatic Segmentation Algorithms," in *Proc. of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS)*, Louvain-la-Neuve, Belgium, June 1997.
- [93] Nakamura Y. and Kanade T., "Semantic Analysis for Video Contents Extraction – Spotting by Association in News Video," in *Proc. of ACM Multimedia Conf.*, Seattle, WA, Nov. 1997.
- [94] Nam J. and Tewfik A., "Progressive Resolution Motion Indexing of Video Object," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Seattle, USA, May 1998.
- [95] Ngan K. N., Panchanathan S., Sikora T., and Sun M.-T., Guest Editorial, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, Sept. 1998.
- [96] Odobez J. M. and Bouthemy P., "Separation of Moving Regions from Background in an Image Sequence Acquired with a Mobile Camera," in *Video Data Compression for Multimedia Computing*, Kluwer Academic Publishers, pp. 238-311, 1997.
- [97] Patel N. V. and Sethi I. K., "Video Shot Detection and Characterization for Video Databases," *Pattern Recognition*, Vol. 30, No. 4, pp. 583-592, April 1997.
- [98] Pavlidis T. and Ali F., "Computer Recognition of Handwritten Numerals by Polygonal Approximation," *IEEE Trans. Systems, Man and Cybernetics*, Vol. 6, pp. 610-614, 1975.
- [99] Pentland A., Picard R. W., and Sclaroff S., "Photobook: Content-Based Manipulation of Image Databases," *Int. J. Computer Vision*, Vol. 18, No. 3, pp. 233-254, 1996.
- [100] Persoon E. and Fu K. S., (1986) "Shape Discrimination Using Fourier Descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, No. 3, pp. 388-397, 1986.
- [101] Pratt I., "Shape Representation using Fourier Coefficients of the Sinusoidal Transform," *Technical Report Series UMCS-96-7-1*, University of Manchester, 1996.
- [102] Raja Y., McKenna S. J., and Gong S., "Tracking and Segmenting People in Varying Lighting Conditions Using Color," in *Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, April 1998.
- [103] Rivlin E. and Weiss I., "Deformation Invariants in Object Recognition," *Computer Vision and Image Understanding*, Vol. 65, No. 1, pp. 95-108, 1999.
- [104] Rothe I., Susse H., and Voss K., "The Method of Normalization to Determine Invariants," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 4, pp. 366-375, 1996.
- [105] Rui Y., Huang T. S., and Chang S.-F., "Digital Image/Video Library and MPEG-7: Standardization and Research Issues," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3785-3788, Seattle, WA, May 1998.
- [106] Rui Y., Huang T. S., and Mehrotra S., "Content-Based Image Retrieval with Relevance Feedback in MARS," in *Proc. of Int. Conf. on Image Processing (ICIP)*, Santa Barbara, CA, Oct. 1997.
- [107] Rui Y., Huang T. S., Ortega M., and Mehrotra S., "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, Sept. 1998.
- [108] Rui Y., She A., and Huang T. S., "A Modified Fourier Descriptor for Shape Matching in MARS," in *Image Databases and Multimedia Search*, Series on Software Engineering and Knowledge Engineering 8, S. K. Chang (editor), World Scientific Publishing House in Singapore, pp. 165-180, 1998.
- [109] Rzeszewski T., "A Novel Automatic Hue Control System," *IEEE Trans. Consumer Electronics*, Vol. 21, No. 2, pp. 155-163, 1975.
- [110] Salembier P., Marques F., Pardas M., Morros R., Corset I., Jeannin S., Bouchard L., Meyer F., Marcotequi B., "Segmentation-based Video Coding System Allowing the Manipulation of Objects," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, No. 1, pp. 60-73, 1997.

- [111] Salembier P. and Pardas M., "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Processing*, Vol. 3, No. 5, pp. 639-651, Sept. 1994.
- [112] Samal A. and Iyengar P. A., "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, Vol. 25, No. 1, pp. 65-77, 1992.
- [113] Schmid C. and Mohr R., "Local Grayvalue Invariants for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 530-535, 1997.
- [114] Shahraray B., "Scene Change Detection and Content-Based Sampling of Video Sequences," in *Proc. of SPIE 2419: Digital Video Compression: Algorithms and Technologies*, pp. 2-13, Feb. 1995.
- [115] Shen D. and Ip H., "Generalized Affine Invariant Image Normalization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 431-440, 1997.
- [116] Sigal L., Sclaroff S., and Athitsos V., "Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, Hilton Head Island, SC, June 2000.
- [117] Sikora T., "The MPEG-4 Video Standard Verification Model," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, No. 1, pp. 19-31, February 1997.
- [118] Slater J., "Eye to Eye with Stereoscopic TV," *Image Technology*, p. 23, Nov./Dec. 1996.
- [119] Smith J. R. and Chang S.-F., "VisualSEEK: A Fully Automated Content-Based Image Query System," in *Proc. of ACM Multimedia Conf.*, pp. 87-98, Boston, MA, Nov. 1996.
- [120] Smoliar S. W. and Zhang H. J., "Content-Based Video Indexing and Retrieval," *IEEE Multimedia*, pp. 62-72, Summer 1994.
- [121] Special Issue on Content-Based Image Retrieval Systems, *IEEE Computer Magazine*, Venkat N. Gudivada and J. V. Raghavan (editors), Vol. 28, No. 9, 1995.
- [122] Special Issue on Segmentation, Description and Retrieval of Video Content, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, 1998.
- [123] Special Issue on Visual Information Management, *Communications of ACM*, Ramesh Jain (guest editor), Dec. 1997.
- [124] Sun Q. B., Huang W. M., and Wu J. K., "Face Detection Based on Color and Local Symmetry Information," in *Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, April 1998.
- [125] Swanson M. D. and Tewfik A. H., "Affine-Invariant Multiresolution Image Retrieval using B-Splines," in *Proc. of Int. Conf. on Image Processing (ICIP)*, Santa Barbara, CA, 1997.
- [126] Taubin G. and Cooper D., "Object Recognition Based on Moment (or Algebraic) Invariants," in *Geometric Invariance in Computer Vision*, J. L. Mundy and A. Zisserman (editors), pp. 375-397, 1992.
- [127] Tekalp A. M., *Digital Video Processing*, Prentice Hall, 1995.
- [128] Torres L. and Kunt M., *Video Coding: The Second Generation Approach*, Kluwer Academic Publishers, Boston, MA, 1996.
- [129] Tsai C.-J. and Katsaggelos A. K., "Dense Disparity Estimation with a Divide-and Conquer Disparity Space Image Technique," *IEEE Trans. Multimedia*, Vol. 1, No. 1, pp. 18-29, Mar. 1999.
- [130] Tsang P. W. M., "A Genetic Algorithm for Affine Invariant Recognition of Object Shapes from Broken Boundaries," *Pattern Recognition Letters*, Vol. 18, pp. 631-639, 1997.
- [131] Tsapatsoulis N. A., Avrithis Y. S., and Kollias S. D., "Efficient Face Detection for Multimedia Applications," in *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, Vancouver, BC, Canada, September 2000.
- [132] Tsapatsoulis N. A., Avrithis Y. S., and Kollias S. D., "Facial Image Indexing in Multimedia Databases," *Pattern Analysis and Applications*, Special Issue on Image Indexation, Springer-Verlag (accepted for publication).

- [133] Tsapatsoulis N., Avrithis Y., and Kollias S., "On the use of Radon Transform for Facial Expression Recognition," in *Proc. of Int. Conf. on Information Systems Analysis and Synthesis (ISAS)*, Orlando, FL, August 1999.
- [134] Tsekeridou S. and Pitas I., "Audio-Visual Content Analysis for Content-Based Video Indexing," in *Proc. of Int. Conf. on Multimedia Computing and Systems*, Florence, Italy, June 1999.
- [135] Tzovaras D., Grammalidis N., and Strintzis M. G., "Disparity Field and Depth Map Coding for Multiview 3D Image Generation," *Signal Processing: Image Communication*, No. 11, pp. 205-230, 1998.
- [136] Tzovaras D., Grammalidis N., and Strintzis M. G., "Object-Based Coding of Stereo Image Sequences Using Joint 3-D Motion/Disparity Compensation," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, No. 2, pp. 312-327, April 1997.
- [137] Vasconcelos N. and Lippman A., "A Spatiotemporal Motion Model for Video Summarization," in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 361-366, Santa Barbara, CA, June 1998.
- [138] Wan X. and Kuo C.-C. J., "A New Approach to Image Retrieval with Hierarchical Color Clustering," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 628-643, Sept. 1998.
- [139] Wang H. and Chang S.-F., "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, No. 4, pp. 615-628, 1997.
- [140] Wang J. and Cohen F. S., "Part II: 3-D Object Recognition and Shape Estimation from Image Contours Using B-Splines, Shape Invariant Matching, and Neural Network," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 13-23, 1994.
- [141] Wang J., Yang W.-J., and Acharya R., "Efficient Access to and Retrieval from a Shape Image Database," in *Proc. of IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)*, Santa Barbara, CA, June 1998.
- [142] Wang Y. and Lee O., "Active Mesh - A Feature Seeking and Tracking Image Sequence Representation Scheme," *IEEE Trans. Image Processing*, Vol. 3, pp. 610-624, Sept. 1994.
- [143] Watt A. and Watt M., *Advanced Animation and Rendering Techniques*, ACM Press, New York, 1992.
- [144] Wax M. and Kailath T., "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 33, No. 2, April 1985.
- [145] Xirouhakis Y., Avrithis Y., and Kollias S., "Image Retrieval and Classification Using Affine Invariant B-Spline Representation and Neural Networks," in *Proc. of IEE Colloquium on Neural Nets and Multimedia*, pp. 4/1-4/4, London, UK, Oct. 1998.
- [146] Xirouhakis Y., Mathioudakis V., and Delopoulos A., "An Efficient Algorithm for Mobile Object Localization in Video Sequences," in *Proc. of Visual Modeling and Visualization Workshop (VMV)*, Erlangen, Germany, 1999.
- [147] Yang G. and Huang T. S., "Human Face Detection in Complex Background," *Pattern Recognition*, Vol. 27, No. 1, pp. 55-63, 1994.
- [148] Yang Z. and Cohen F. S., "Image Registration and Object Recognition Using Affine Invariants and Convex Hulls," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, No. 7, pp. 934-946, 1999.
- [149] Yeo B.-L. and Liu B., "Rapid Scene Analysis on Compressed Videos," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 5, pp. 533-544, Dec. 1995.
- [150] Yeung M. M. and Yeo B.-L., "Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 7, No. 5, pp. 771-785, Oct. 1997.

- [151] Yow K. C. and Cipolla C., "Feature-based Human Face Detection in Complex Background," *Image and Vision Computing*, Vol. 15, No. 9, pp. 713-735, 1997.
- [152] Zhang H. J., Tan S. Y., Smoliar S., and Yihong G., "Automatic Parsing and Indexing of News Video," *Multimedia Systems*, Vol. 2, pp. 256-266, 1995.
- [153] Ziegler M., "Digital Stereoscopic Imaging and Applications: A Way Towards New Dimensions, the RACE II Project DISTIMA," in *Proc. of IEE Colloquium on Stereoscopic Television*, London, UK, 1992.

Ευρετήριο

Ελληνικοί όροι

A

αβεβαιότητα σημείου αναφοράς, 105, 107

ακεραίων αριθμών, κωδικοποίηση, 72

ακολουθιακή πρόσβαση, 1

ακολουθίες

MPEG, 19, 20, 46, 51, 148

δελτίων ειδήσεων, 3, 85

στερεοσκοπικές, 3, 6, 22

συνθετικές, 46, 62

ψευδοτυχαίες, 9

ακουστική πληροφορία, 148, 153

ακραία σημεία, 61

ακρίβεια, 141, 144, 152, 161

ακρότατα, τοπικά, 66

αλγόριθμος

color clustering, 14

K-means, 53, 148

Levenberg-Marquardt, 116

Lloyd-Max, 9, 52, 53

M-RSST, 6, 9, 16, 117, 133

pyramidal region growing, 14

RSST, 6, 9, 14

watershed, 14

ανίχνευσης προσώπων, 129, 146

γενετικός, 7, 10, 69, 72

λογαριθμικός, 69

ορθοκανονικοποίησης, 99

στοχαστικός, 71

συγχώνευσης, 6, 33

ταχείας ανίχνευσης δέρματος, 131

αληθείς μάσκες, 141, 143, 144

αληθής ταξινόμηση, 152, 161

αλλαγή κλίμακας, 92, 113, 117

αλλαγή πλάνου, 148

απότομη, 148

αλληλεπίδραση, 82, 146

αλυσίδας, κωδικοποίηση, 90

αναγνώριση

θεματικών ενότητων, 146

περιγραμμάτων, 123

προσώπων, 129, 147

χαρακτήρων, 89

ανάδραση, 7, 10, 80, 82, 131

απλή, 83

καταλληλότητας, 82

πολλαπλή, 83, 84, 87

αναζήτηση, 1

λογαριθμική, 7, 10, 69

με βάση τα πρόσωπα, 8

κλίμακα προσώπων, 158, 159

πλήθος προσώπων, 158, 160

χρώμα προσώπων, 157, 159

με βάση το περιεχόμενο, 7, 79, 129

με βάση το σχήμα, 5, 112, 115, 117

μέσω κειμένου, 82

μέσω παραδείγματος, 7, 79, 85, 114, 126, 156

μέσω σκίτσου, 114

μέσω σχεδιαγράμματος, 5, 7, 79, 81, 90, 114

παράλληλη, 72

περιγραμμάτων, 125

περιοχή, 26

πληροφορίας, 152

σενάρια, 157

τυχαία, 71

ανακατανομή κομβικών σημείων, 98, 120

ανακατασκευή εικόνων υπερήχων, 9

ανάκληση, 1, 141, 144, 152, 161

λανθασμένη, 127

με βάση τα πρόσωπα, 156

με βάση το περιεχόμενο, 7, 79, 156

με βάση το σχήμα, 8, 10, 93, 112, 117, 125

μέσω παραδείγματος, 80, 87, 159

μέσω σχεδιαγράμματος, 5, 127

αναλλοίωτα χαρακτηριστικά, 136

αναλλοίωτες ποσότητες, 5, 91

αναλλοίωτη αναπαράσταση, 114, 115

αναλλοίωτη περιγραφή, 7

σε affine μετασχηματισμούς, 107

αναλλοίωτοι μετασχηματισμοί, 91, 113

ανάλυση

δελτίων ειδήσεων, 5, 8, 13, 146

επίπεδο, 16

οπτικοακουστικού υλικού, 131

σε αντικείμενα, 114

στερεοσκοπική, 3, 9, 22

συντακτική, 3, 82

σχήματος προσώπου, 133, 135

αναλυτές δεδομένων, 145

ανάμειξης, συνάρτησης, 96

ανανέωση βαρών, 116

αναπαράσταση

αναλλοίωτη σε affine μετασχηματισμούς, 114,

115, 124

οπτικού περιεχομένου, 6

περιγραμμάτων, 113, 114, 116

αναπροσαρμογή μοντέλου, 132, 133

αναφοράς, σημείο, 92, 98, 103

ανίχνευση

ακμών, 93

αλλαγής πλάνων, 4, 50, 146, 148
 δέρματος, 130, 141
 επικάλυψης, 6, 28
 κύριων κινούμενων αντικειμένων, 94
 λανθασμένη, 130, 132
 μαύρης οθόνης, 146
 περιγραμμάτων, 114
 περιόδων σιωπής, 146
 προσώπων, 8, 13, 129, 131, 133, 140, 146
 σημείων ενδιαφέροντος, 5
 χρώματος δέρματος, 131
 ανταλλαγή γονιδίων, 74
 αντικατάστασης, στρατηγική, 72, 75
 αντικείμενα, 2
 αντικειμενική συνάρτηση, 73
 αντιπροσωπευτικά πλάνα, 7
 αντιστάθμιση
 επικάλυψης, 6, 29
 κίνησης, 147, 148
 λανθασμένης ανίχνευσης, 151
 αντιστρέψιμος πίνακας, 109
 απαλοιφή φάσης, 113
 απαριθμητικές τεχνικές, 72
 απλή ανάδραση, 83
 απλό σημείο ανταλλαγής γονιδίων, 74
 απλοποίηση περιγραμμάτων, 123
 απόγονοι, 72, 74
 απόκλιση, 6
 διάνυσμα, 25
 πεδίο, 25
 αποκωδικοποίηση, 19
 απόλυτη διαφορά, 147
 απόσταση
 διανυσμάτων περιγραφής, 80
 Ευκλείδεια, 124
 κανονικοποιημένων Fourier descriptors, 115, 124
 κομβικών σημείων, 124, 125
 με βάρη, 80
 παραμετρική, 80
 περιγραμμάτων, 115, 124
 αποσύνθεση, 93
 απότομες αλλαγές πλάνου, 148
 απώλεια, 130, 139, 143, 161
 πληροφορίας, 113, 114
 αρχικό σημείο αναφοράς, 92, 98
 αρχικός πληθυσμός, 73
 αρχιτεκτονική νευρωνικού δικτύου, 116
 ασαφείς συναρτήσεις συμμετοχής, 137, 147
 ασαφές ιστόγραμμα, 9, 41, 42, 44
 ασαφή συστήματα, 44
 ασαφής ταξινόμηση, 41
 αυτόματη ανάλυση δελτίων ειδήσεων, 145

B

βαθμός
 ομοιότητας
 κλίμακας, 158
 περιγραμμάτων, 122, 127
 χρώματος, 158
 συμμετοχής, 41, 43
 βάθος, 6

πεδίο, 25
 βάρη
 ανανέωση, 116
 απόσταση με, 80
 διάνυσμα, 81
 διασύνδεσης νευρώνων, 116
 συνδέσμων, 134
 βάρους εξομάλυνσης, συνάρτηση, 27
 βάσεις δεδομένων
 κατανεμημένες, 3
 πολυμέσων, 1
 βάσης, απόσταση, 23
 βάσης, συνάρτηση, 96
 βέλτιστη επιλογή χαρακτηριστικών καρτέ, 67
 βελτιστοποίηση, 69, 72, 84, 91, 113
 βήματα κανονικοποίησης, 117
 βήματα ορθοκανονικοποίησης, 100
 βιντεοτηλέφωνο, 13, 32

Γ

γεγονός, 152
 γειτονική περιοχή, 70
 γενετικός αλγόριθμος, 7, 10, 69, 72, 91
 γενίκευση μοντέλου χρώματος δέρματος, 131
 γεωμετρικός μέσος με βάρη, 137
 Γκαουσιανή κατανομή, 131, 132
 γλώσσα ορισμού περιγραφής, 3
 γλωσσική πληροφορία, 5, 146, 147, 153
 γονέων, επιλογή, 73
 γονίδιο, 72
 γονιδίων, ανταλλαγή, 74
 γονικά χρωμοσώματα, 72

Δ

δεδομένα σημεία, 97
 δειγματοληψία, 92
 με βάση το περιεχόμενο, 7
 μη ομοιόμορφη, 117, 119
 ομοιόμορφη, 120
 περιγραμμάτων, 117, 118
 δεικτοδότηση, 1, 129, 145, 156
 οπτικοακουστικού υλικού, 131
 δεικτών, διάνυσμα, 68
 δεικτών, συνάρτηση, 45
 δελτία ειδήσεων, 3, 8, 129, 131, 145
 δέρμα
 ανίχνευση, 130
 χρώμα, 130
 δευτερεύων άξονας, 136
 δεύτερη παράγωγος, 60
 διαγραφής, φορά, 92
 διαδίκτυο, 141
 διαίρεση τμημάτων, 16
 διακριτός μετασχηματισμός Fourier, 103
 διακριτός μετασχηματισμός Karhunen-Loeve, 136
 διαμέριση, 41, 43
 διάνυσμα
 απόκλισης, 25
 βαρών, 81
 δεικτών, 68
 εισόδου, 116
 εξόδου, 116

κίνησης, 20, 148
 μετατόπισης, 26
 μιγαδικό, 103, 106
 παραμέτρων, 81
 περιγραφής, 7, 14, 45, 80
 απόσταση, 80
 μέτρο ομοιότητας, 80
 σφάλματος, 81
 φάσης, 103, 110
 χαρακτηριστικών τμήματος, 42
 χρωματικών συνιστωσών, 132, 133
 διασπορά, 122
 διασύνδεση νευρώνων, 116
 διαχείριση, 2, 90
 διδιάστατη κανονική κατανομή, 131
 διοπτρικό σύστημα κάμερας, 23
 δυαδική κωδικοποίηση, 72
 δυαδική μάσκα, 135, 141
 δυναμικός προγραμματισμός, 72, 146

Ε

εγγραφές teletext, 146
 ειδήσεις, 129
 κατ' απαίτηση, 5, 146
 εικόνα
 γενική, 141
 δέρματος, 141
 πιθανοφάνειας
 δέρματος, 138
 προσώπου, 138
 υπερήχων, 9
 εκπαίδευση, 125
 νευρωνικού δικτύου, 116, 117
 εκτίμηση
 ικανότητας, 73
 κίνησης, 19, 69
 σφάλμα, 51
 φόντου, 147
 πεδίου απόκλισης, 25
 πεδίου βάθους, 25
 υποθέσεων, 152
 εκφράσεις, 9
 ελάσσων άξονας, 136
 ελάχιστο μέσο τετραγωνικό σφάλμα, 97, 116
 ελέγχου, σημεία, 96
 έμμεσα πολυώνυμα, 5
 ενεργά περιγράμματα, 92, 127, 136
 ενεργοποίησης, συνάρτηση, 116
 ενότητες
 ειδήσεων, 148
 πλάνων, 154
 εντοπισμός
 περιγράμματος, 115
 προσώπων, 10, 13, 129, 131, 133, 140
 εξαγωγή
 περιγραμμάτων, 114, 117, 125
 περίληψης, 1, 4, 129, 155
 τμημάτων προσώπου, 133
 χαρακτηριστικών, 93
 χαρακτηριστικών καρέ, 47
 εξελικτικό πρόγραμμα, 71
 εξομάλυνσης, συνάρτηση, 26

εξωτερικά πλάνα, 146, 147, 151, 152
 επαλήθευση, 131, 132, 139
 επανεκτίμηση
 μοντέλου χρώματος δέρματος, 139
 παραμέτρων κατανομής, 131, 132, 133
 χρωματικού μοντέλου, 157
 επεξεργασία
 εγγράφων, 89
 μορφολογική, 135
 σχήματος, 130
 επίδραση θορύβου, 122
 επιθυμητό διάλυσμα εξόδου, 116
 επικάλυψη, 6, 28, 94
 ανίχνευση, 28
 αντιστάθμιση, 29
 ποσοστό, 42
 τμημάτων, 142
 επικρατέστερα πρόσωπα, 147, 149
 μέγιστη πιθανοφάνεια, 150
 επίλεκτων, στρατηγική, 75
 επιλογή
 απλού καρέ, 4
 γονέων, 73
 καρέ, ομοιόμορφη, 4
 πολλαπλών καρέ, 4
 φυσική, 72
 χαρακτηριστικών καρέ, 76
 βέλτιστη, 67
 επιλογής, τάση, 73
 επιμήκυνση, 90, 136
 επίπεδα αντικειμένων βίντεο, 2
 επίπεδο
 Cr-Cb, 130, 131
 αντικειμένων, 90
 εικόνας, 24
 κρυφό, 116
 νευρώνων, 116
 επιρροής, ζώνη, 52
 επιτάχυνση, 62
 επιφάνεια
 περιγράμματος, 136
 προσώπου, 158
 ερώτημα, 80
 εστιακό
 μήκος, 23
 σημείο, 24
 εσφαλμένη ανίχνευση, 132, 133
 ευθυγράμμιση περιγραμμάτων, 114, 121
 ευθυγραμμισμένα γεγονότα, 152
 Ευκλείδεια απόσταση, 124
 εύρος ζώνης μετάδοσης, 1, 3
 εφαρμογές πολυμέσων, 1

Z

ζώνη επιρροής, 52

H

ημιτονοειδής μετασχηματισμός, 90

Θ

θεματικές ενότητες, 4, 146, 148, 156
 θόρυβος περιγραμμάτων, 122

I

ιδιοδιανύσματα, 136
 ιδιοτιμές, 136
 ιεραρχική ομαδοποίηση χρωμάτων, 5
 ικανότητα, 72
 εκτίμηση, 73
 συνάρτηση, 73
 ιστόγραμμα
 ασαφές, 42, 44
 πολυδιάστατο, 9, 40, 41
 χρώματος, 45, 148
 δέρματος, 130, 132

K

καθοδηγούμενη τυχαία αναζήτηση, 71
 καμπύλη
 μιγαδική αναπαράσταση, 103, 106, 113
 ομοιότητα, 115
 ορθοκανονικοποιημένη, 100, 102, 109
 περιγραφή, 113
 ταίριασμα, 102
 καμπυλότητας, μέτρο, 60
 κανάλι
 αριστερό/δεξιό, 26
 φωτεινότητας, 130
 κανονική κατανομή, 122, 131, 132
 κανονικοποιημένα περιγράμματα, 115
 κανονικοποιημένοι περιγραφείς Fourier, 91, 112, 113
 Ευκλείδεια απόσταση, 124
 κανονικοποίηση, 5, 7, 91
 affine μετασχηματισμού, 107
 αλλαγής κλίμακας, 99
 ανοικτών καμπυλών, 108
 βήματα, 117
 εικόνων, 92
 κατοπτρισμού, 103, 106, 119
 κλίμακας, 101, 119
 κλίσης, 99, 101, 119
 κυκλικής ολίσθησης, 104, 105
 μετατόπισης, 99, 101, 119
 περιγραμμάτων, 10, 93, 114, 117, 119, 136
 περιστροφής, 103, 106, 119
 σημείου αναφοράς, 103, 104, 105, 119
 ταξινόμησης, 73
 φοράς διαγραφής, 108
 καρτέ, 1
 ταξινόμηση, 147
 χαρακτηριστικά, 7, 60, 68, 76, 156
 Καρτεσιανές συντεταγμένες, 24
 καταλληλότητα, 82
 κατανομή
 κανονική, 122, 131
 κομβικών σημείων, 115
 προσαρμογή, 131
 χρώματος δέρματος, 130, 131
 κατάτμηση, 2, 6, 13, 126
 M-RSST, 117
 με βάση την κίνηση, 6, 19
 με βάση το πεδίο βάθους, 6, 13, 30
 με βάση το χρώμα, 6, 17, 117, 130, 133
 κατηγορίες περιγραμμάτων, 115, 117, 121, 124, 125

κατοπτρισμός, 92, 117
 κανονικοποίηση, 106, 119
 μετασχηματισμός, 109
 πίνακας, 106, 107
 πρότυπος, 106
 συμμετρία, 105
 κατώφλι
 επιλογής καρτέ, 4
 κατάτμησης, 16, 37, 133
 ταξινόμησης, 132
 κατωφλίωση, 142
 διαφοράς καρτέ, 148
 πιθανοφάνειας, 132
 πιθανοφάνειας δέρματος, απευθείας, 135
 κειμένου, πληροφορία, 146
 κέντρα ομάδων, 53
 βέλτιστα, 53
 κεντρικές ροπές, 46
 κεντρικό σημείο, 70
 κέντρο βάρους, 43
 κέρδος, 116
 κίνηση, 2
 διάνυσμα, 20
 εκτίμηση, 19, 69
 φόντου, 147
 κλάσεις, 40
 κλειδιά, 4
 κλίμακα
 αλλαγή, 92
 κανονικοποίηση, 119
 προσώπου, 158
 κλίση, 92, 113, 117
 κανονικοποίηση, 119
 κόλινος πυραμίδα εικόνων, 16
 κομβικά σημεία, 96, 114, 115
 ανακατανομή, 120
 Ευκλείδεια απόσταση, 124, 125
 κοντινά πλάνα, 147
 προσώπου, 147, 158
 κόστους, συνάρτηση, 26, 83
 κριτήριο
 αναζήτησης, 157
 μέγιστης πιθανοφάνειας, 133
 κρυφά επίπεδα, 116
 κυβικά B-Splines, 95
 κυκλική ολίσθηση, 103, 113
 κανονικοποίηση, 104, 105
 μετασχηματισμός, 105
 πρότυπη, 104
 κυκλική συμμετρία, 105
 κύρια κινούμενα αντικείμενα, 94
 κύριοι άξονες, 90, 136
 κυρτό περίβλημα, 5, 91
 κυρτότητας, μέτρο, 90
 κωδικοποίηση
 ακεραίων αριθμών, 72
 αλυσίδας, 90
 βίντεο, 2
 δεύτερης γενιάς, 5, 90
 δυναμική, 72
 σχέδιο, 72

Λ

λανθασμένη
 ανάκληση, 127
 ανίχνευση, 130, 143, 151
 ταξινόμηση, 117
 λέξεις-κλειδιά, 4
 λειπομέρειες προσώπου, 133
 λογαριθμική αναζήτηση, 7, 10, 69

Μ

μάσκες, 90, 135, 141
 αληθείς, 141, 144
 μέγιστη πιθανοφάνεια, 133
 προσώπου, 149
 μέθοδος
 ρουλέτας, 73
 συσχέτισης, 7, 67, 76
 χρονικής μεταβολής, 7, 61, 73
 μείζων άξονας, 136
 μείωσης, ρυθμός, 73
 μέσες χρωματικές συνιστώσες, 157
 μετάβαση πλάνου, 148
 μεταβολές φωτεινότητας, 133
 μετάλλαξη, 75
 μετασχηματισμός
 affine, 7, 90, 101, 102, 114, 117, 119, 124
 Fourier, 7, 95, 103
 φάση, 113
 Hotelling, 136
 Karhunen-Loeve, 136
 Radon, 8
 ημιτονοειδής, 90
 κατοπτρισμού, 109
 κυκλικής ολίσθησης, 105
 μετατόπισης, 24
 ορθογώνιος, 99, 102, 107
 παραμέτρου, 94
 περιστροφής, 24, 109
 προοπτικής προβολής, 24, 90
 σκελετού, 90
 συμπαγούς / επιμήκυνσης, 137
 συντεταγμένων, 94
 μετατόπιση, 92, 113, 117
 διάνυσμα, 26
 κανονικοποίηση, 119
 μετρήσεις ακρίβειας, 141
 μέτρο
 απόστασης
 βάρη, 134
 περιγραμμάτων, 115, 124
 συνδέσμων, 134
 δεύτερης παραγώγου, 60
 καμπυλότητας, 60
 κυρτότητας, 90
 ομοιότητας, 158
 διανυσμάτων περιγραφής, 80
 καμπυλών, 93, 115
 συσχέτισης, 68
 μη ομοιόμορφη δειγματοληψία, 117, 119
 μη ομοιόμορφη μετάλλαξη, 75
 μήκος
 παραθύρου, 151

περιγράμματος, 90

τόξου, 92, 94, 115, 120

χορδής, 97

μηχανισμός ταξινόμησης, 114

μιγαδικό διάνυσμα, 103, 106

μνήμη, 84, 132, 157

μοναδικότητα, 93, 102, 106

μονοδιάστατη ταξινόμηση, 41

μονοπάτι, 69

μοντέλο

Markov, 5, 130

YCrCb, 131

ανάδρασης, 131

Γκαουσιανό, 130

ενεργών περιγραμμάτων, 5, 92, 127

κανονικής κατανομής, 132

κίνησης, παραμετρικό, 117

μοντελοποίηση

B-Splines, 120

περιγραμμάτων, 117

χρώματος δέρματος, 130, 133

μορφολογική επεξεργασία, 135

μορφολογικό closing, 141

μωσαϊκό εικόνων, 4

N

νευρώνες, 116

νευρωνικό δίκτυο, 8, 9, 91, 112, 115, 124

αρχιτεκτονική, 116

εκπαίδευση, 116, 125

O

ολικά χαρακτηριστικά, 45, 90, 102, 136

ολίσθηση, κυκλική, 103

ομαδοποίηση, 152

περιγραμμάτων, 115

πλάνων, 52, 148, 155

χρωμάτων, ιεραρχική, 5

ομαλότητα, 95

ομογενείς συντεταγμένες, 24

ομοιογένεια

χρώματος, 93

ομοιόμορφη

ανταλλαγή γονιδίων, 74

δειγματοληψία, 120

επιλογή καρέ, 4

μετάλλαξη, 75

ομοιότητα

καμπυλών, 115

κλίμακας, 158

περιγραμμάτων, 122, 125, 127

χρώματος, 158

οπτική ροή, 19

οπτικοακουστική πληροφορία, 146

οργάνωση

δελτίων ειδήσεων, 145

περιγραμμάτων, 115

ορθογώνιος μετασχηματισμός, 99, 102, 107

ορθοκανονικοποιημένη καμπύλη, 109

ορθοκανονικοποίηση, 95, 99, 119

βήματα, 100, 108

πίνακας, 101, 109

συνθήκες, 101, 108
 όρισμα, πρωτεύον, 103

Π

παράθυρο, τετραγωνικό, 62
 παράλληλη αναζήτηση, 72
 παραμετρική απόσταση, 80
 παραμετρικό μοντέλο κίνησης, 117
 παράμετροι κατανομής, 131
 επανεκτίμηση, 131
 παραμορφούμενα πρότυπα, 5, 92, 127, 136
 παραμόρφωση, 53
 σχήματος, 122
 παρουσιαστές, 146, 147, 152
 πεδίο
 απόκλισης, 6, 25
 βάθους, 6, 25
 περιβάλλον αλληλεπίδρασης, 133
 περίβλημα, κυρτό, 5
 περιγράμματα, 5, 46, 81, 89
 αναλλοιώτα σε affine μετασχηματισμούς, 114
 αναπαράσταση, 113, 114
 ανίχνευση, 114
 απλοποίηση, 123
 απόσταση, 115
 ενεργά, 5, 92
 εξαγωγή, 117
 επιφάνεια, 136
 ευθυγράμμιση, 114, 121
 θόρυβος, 122
 κανονικοποίηση, 114, 115
 ομαδοποίηση, 115
 περίμετρος, 136
 πρότυπα, 116, 117, 125
 σύγκριση, 114, 115, 117
 σχέδιαση, 114
 ταξινόμηση, 124
 περιγραφείς, 3
 Fourier, 5, 90
 κανονικοποιημένοι, 91, 112
 τροποποιημένοι, 5, 91, 115
 Legendre, 90
 περιγραφή
 αναλλοιώτη, 7
 διάνυσμα, 7, 40, 45
 καμπυλών, 113
 υλικού πολυμέσων, 2
 περίληψη, 4, 129
 περίμετρος περιγράμματος, 136
 περιοχές, 14
 απόσταση, 14
 γειτονικές, 70
 δέρματος, 132
 συνοριακές, 19
 περιστροφή, 92, 113, 117
 κανονικοποίηση, 106, 119
 μετασχηματισμός, 109
 πίνακας, 106, 107
 πρότυπη, 106
 πιθανότητα μετάλλαξης, 75
 πιθανοφάνεια
 προσώπου, 137

 μέγιστη, 149
 χρώματος δέρματος, 132, 133, 137
 πίνακας
 affine μετασχηματισμού, 114
 αντιστρέψιμος, 109
 κατοπτρισμού, 106, 107
 ορθογώνιος, 101, 102, 107
 ορθοκανονικοποίησης, 101, 109
 περιστροφής, 106, 107
 προοπτικής προβολής, 24
 συμμεταβλητότητας, 132, 136
 πλάνο
 αλλαγή, 50
 αντιπροσωπευτικά, 156
 διάνυσμα περιγραφής, 52
 κοντινά, 147
 ομαδοποίηση, 52
 παρουσιαστή, 148
 ταξινόμηση, 148
 χαρακτηρισμός, 52
 χαρακτηριστικά, 7, 53
 πληθυσμός, 72
 αρχικός, 73
 πληρότητα, 93, 102, 106
 πληροφορία
 απώλεια, 113
 γλωσσική, 5, 146, 147
 κειμένου, 5, 146, 148
 οπτικοακουστική, 5, 146
 σημασιολογική, 129
 σχήματος, 5
 πλοήγηση, 1, 5, 7, 156
 ειδήσεων, 145
 πολλαπλά σημεία ανταλλαγής γονιδίων, 74
 πολλαπλή ανάδραση, 83, 84, 87
 πολλαπλή προσπέλαση διαίρεσης κώδικα, 9
 πολυγώνων, προσέγγιση, 90
 πολυδιακριτική υλοποίηση, 16
 πολυδιάστατη ταξινόμηση, 42
 πολυμέσα, 1
 πολυπλοκότητα
 M-RSST, 17
 RSST, 16
 ανίχνευσης προσώπων, 131
 ασαφούς ιστογράμματος, 45
 εκτίμησης απόκλισης, 58
 εξαντλητικής αναζήτησης, 69
 μεθόδου συσχέτισης, 7
 ταιριάσματος
 με πρότυπο, 139
 περιγραμμάτων, 114, 115
 υφής, 45
 πολώνυμα, έμμεσα, 5
 ποσοστό
 ανίχνευσης, 132
 επικάλυψης, 42
 πραγματικό διάνυσμα εξόδου, 116
 προβολής, συνάρτηση, 34
 προεπεξεργασία περιγραμμάτων, 122
 προεπισκόπηση, 1, 7, 50, 66, 154
 προοπτική προβολή, 24
 προσαρμογή

καμπύλης B-Spline, 97
 κατανομή, 131
 προσέγγιση πολυγώνων, 90
 προσήμου, συνάρτηση, 106
 πρόσωπα
 αναγνώριση, 129
 ανίχνευση, 129
 επικρατέστερα, 147, 149
 κοντινά, 147
 υφή, 131
 χαρακτηριστικά στοιχεία, 130
 πρότυπα
 παραμορφούμενα, 92
 περιγράμματα, 116, 117, 125
 προσώπου, 131
 πρότυπη
 θέση, 91
 κυκλική ολίσθηση, 104
 περιστροφή, 106
 πρότυπος κατοπτρισμός, 106
 πρωτεύον όρισμα, 103
 πρωτεύων άξονας, 136
 πυρήνας, 114

P

ρεπορτάζ, 146, 147, 152
 ροπές, 5, 7, 46, 90, 112, 114, 157
 Zernike, 90
 καμπυλών, 90, 100
 με βάρη, 114
 πυρήνας, 114
 ρουλέτας, μέθοδος, 73
 ρυθμός
 απώλειας, 152
 λανθασμένης ανίχνευσης, 152
 μείωσης, 73

Σ

σενάρια αναζήτησης, 157
 σημασιολογική
 κατάτμηση, 13, 148
 πληροφορία, 129
 σημασιολογικό περιεχόμενο, 13, 32
 σημασιολογικός χαρακτηρισμός, 131
 σηματοθορυβικός λόγος, 123
 σημεία
 δεδομένα, 97
 ελέγχου, 96
 ενδιαφέροντος, 90
 κομβικά, 96, 114, 115
 ανακατανομή, 98
 σύνδεσης, 95
 σημείο αναφοράς, 92, 98, 103, 113, 117
 αβεβαιότητα, 105, 107
 κανονικοποίηση, 103, 119
 σφάλμα, 110
 σκελετός, 90
 σκηνές
 εξωτερικών πλάνων, 146, 147
 παρουσιαστή, 146, 147
 ρεπορτάζ, 146, 147
 στατικών εικόνων, 146, 147

σκίτσο, 80, 81, 90, 114, 127
 στάδιο επαλήθευσης, 131, 132, 133
 στατικές εικόνες, 146, 147, 152
 στερεοσκοπικές ακολουθίες, 3, 6, 22
 στερεοσκοπική ανάλυση, 3, 9, 22, 23
 στοιχειώδεις ενότητες ειδήσεων, 148, 154
 στρατηγική αντικατάστασης, 75
 στροφής, συνάρτηση, 91
 σύγκριση περιγραμμάτων, 114, 115, 117
 συγχώνευση, 6, 14, 15, 16, 33
 τμημάτων δέρματος, 134
 τμημάτων προσώπου, 130, 133
 συμμεταβλητότητα, 68, 132, 136
 συμμετοχή
 βαθμός, 41, 43
 συνάρτηση, 41
 συμμετρία
 κατοπτρισμού, 105
 κυκλική, 105
 τοπική, 130
 συμπαγές, 90, 136
 συμπίεση βίντεο, 2
 συνάρτηση
 Dirac, 113
 ανάμειξης, 96
 αντικειμενική, 73
 βάρους εξομάλυνσης, 27
 βάσης, 96
 δεικτών, 45
 ενεργοποίησης, 116
 εξομάλυνσης, 26
 ικανότητας, 73
 κόστους, 26, 83
 προβολής, 34
 προσήμου, 106
 πυκνότητας πιθανότητας, 131, 132
 πυρήνα, 114
 στροφής, 91
 συμμετοχής, 41, 137, 147
 Γκαουσιανή, 64
 σιγμοειδής, 64
 τριγωνική, 41
 σφάλματος, 26
 φραγμένη, 95
 σύνδεσης, σημεία, 95
 σύνδεσμοι, 14
 βάρος, 14
 ταξινόμηση, 15, 16
 συνδυασμός Γκαουσιανών μοντέλων, 130, 131
 συνέντευξη, 147
 συνέχεια, 95
 συνθετική ακολουθία, 46, 62
 συνθήκες
 κανονικοποίησης
 κατοπτρισμού, 107, 111
 περιστροφής, 107, 111
 σημείου αναφοράς, 105
 ορθοκανονικοποίησης, 101, 108
 φωτισμού, 132, 138, 142, 151
 μεταβαλλόμενες, 132
 συντακτική ανάλυση, 3, 8, 10, 82, 129, 145, 146
 σύστημα αναφοράς, 24

συστήματα

αναζήτησης, 126

πρωτότυπα, 2

ασαφή, 44

συσχέτιση, 7

μέθοδος, 67, 76

μέτρο, 68

συντελεστής, 68

σφάλμα

διάνυσμα, 81

εκτίμησης κίνησης, 51, 148

σημείου αναφοράς, 110

συνάρτηση, 26

τετραγωνικό, 97, 116

σχεδιάγραμμα, 81

σχεδίαση περιγραμμάτων, 114

σχετικότητα, 82

σχήμα, 2, 5, 81, 89

επεξεργασία, 130

περιγράμματος, 89

περιγραφής, 3

προσώπου, 135

σωστά ευθυγραμμισμένα γεγονότα, 152

T

ταίριασμα

αναλλοίωτο σε affine μετασχηματισμούς, 112, 114

καμπυλών, 5, 91, 102

με πρότυπο, 8, 93, 130

προσώπου, 131, 133, 139

με χρήση ροπών, 112, 114

περιγραμμάτων, 114, 117, 119, 123

ταξινόμηση

αληθής, 152, 161

ασαφής, 41

εκφράσεων, 9

κανονικοποίηση, 73

καρέ, 147

λανθασμένη, 117

με βάση το σχήμα, 115

με νευρωνικό δίκτυο, 116, 124

μονοδιάστατη, 41

περιγραμμάτων, 112, 113, 117, 123, 124

περιοχών δέρματος, 132

πλάνων, 8, 146, 148, 154

πολυδιακριτική, 9

πολυδιάστατη, 42

συνδέσμων, 15, 16

σχημάτων, 8

τάση επιλογής, 73

τερματισμός κατάτμησης, 133

τετραγωνικό

παράθυρο, 62

σφάλμα, 116

τεχνητή νοημοσύνη, 6, 146

τηλεοπτικά δελτία ειδήσεων, 3, 8, 145

τμήματα, 14

B-Splines, 95

βάθους, 33, 43

δέρματος, 134

κίνησης, 33, 43

προσώπου, 133

χρώματος, 33, 43

τόξο, 94, 120

τοπικά ελάχιστα, 71

τοπικά χαρακτηριστικά, 90

τοπική συμμετρία, 130

τριγωνική συνάρτηση συμμετοχής, 41

τρισδιάστατες ακολουθίες, 3, 22

τροποποιημένοι περιγραφείς Fourier, 5, 91, 115, 124

τροχιά διανύσματος περιγραφής, 10, 61

τυχαία αναζήτηση, 76, 78

Υ

υπερήχων, εικόνες, 9

υπερ-χρωμοσώματα, 73

υποδειγματοληψία, 59

υποθέσεων, εκτίμηση, 152

υπότιτλοι, 146

υφή, 2, 18, 157

πολυπλοκότητα, 45

Φ

φάση, 103

απαλοιφή, 113

διάνυσμα, 103, 110

επαλήθευσης, 139

μετασχηματισμού Fourier, 113

φίλτρο median, 20, 150

φόντο, 147

φορά διαγραφής, 92

κανονικοποίηση, 108

φυσική επιλογή, 72

φωτεινότητα, 130

μεταβολές, 133

φωτισμός, 132, 138

Χ

χαρακτηρισμός, 1, 5, 14, 85, 145

με λέξεις-κλειδιά, 4

πλάνων, 52

σημαιολογικός, 131

χαρακτηριστικά

διάνυσμα, 40

εξαγωγή, 93

καρέ, 7, 47, 60, 68, 76

ολικά, 45, 90, 102

πλάνα, 7, 53

σχήματος, 131, 157

τοπικά, 90

υφής, 157

χρώματος, 157

χορδή, 97

χρονική μεταβολή, 7

μέθοδος, 61, 73

χρώμα, 2

δέρματος

ανίχνευση, 131

ιστόγραμμα, 130, 132

κατανομή, 131

μοντελοποίηση, 130

ιστόγραμμα, 148

χρωματικές συνιστώσες, 15, 157
 Cr-Cb, 130, 131, 132
 διάνυσμα, 132, 133
 χρωματισμένη πηγή φωτισμού, 142
 χρωμοσώματα, 72
 γονικά, 72
 χωρητικότητα αποθήκευσης, 1
 χώρος
 Ευκλείδειος, 136
 κλίμακας κυρτότητας, 90

Ξενόγλωσσοι όροι

A

activation function, 116
 active contours, 92, 127, 136
 affine
 invariant, 7, 91
 transformation, 90, 101
 aligned events, 152
 anchorperson shots, 146
 anchors, 147
 annotation, 1, 145
 keyword, 4
 approximation, polygonal, 90
 arc length, 94
 argument, primary, 103
 arithmetical crossover, 75
 audiovisual cues, 146

B

background, 147
 baseline distance, 23
 basis function, 96
 bin, 40
 binary encoding scheme, 72
 binocular camera system, 23
 blending function, 96
 block, 19, 46
 error function, 26
 matching, 25
 motion estimation, 69
 error, 51
 Boston University IVC database, 144
 boundedness, 95
 bounding rectangle, 139
 browsing, 1
 B-Splines, 7, 90, 95, 117, 120
 cubic, 95
 segments, 95

C

CDMA, 9
 chain coding, 90
 channel, left/right, 26
 chord length, 97
 chromosome, 72
 circular shift, 103
 CL, 97
 classification

χρωμάτων
 HSV, 130
 RGB, 15, 130
 YCrCb, 130
 δέρματος, 130

Ψ

ψευδοαντίστροφος πίνακας, 25
 ψευδοτυχαίες ακολουθίες, 9

fuzzy, 41
 shot, 148
 closed-caption tokens, 146
 close-ups, 147
 closing, 141
 cluster center, 53
 clustering, 52, 148, 152
 code division multiple access, 9
 color
 clustering, 14
 histogram, 148
 compactness, 90, 136
 Compaq research database, 141
 completeness, 93
 connection points, 95
 content-based
 retrieval, 1, 79, 156
 sampling, 7
 continuity, 95
 contour, 89
 active, 92
 contrast, 138
 control points, 96
 convex hull, 91
 coordinate transformation, 94
 correctly aligned events, 152
 cost function, 26, 83
 COST211ter, 14
 CRL, 141, 143
 crossover, 74
 cubic B-Splines, 95
 cues
 audiovisual, 146
 linguistic, 146
 textual, 146
 curvature
 measure, 90
 scale spaces, 90
 curve moments, 90

D

data points, 97
 decomposition, 93
 deformable templates, 92, 127, 136
 degree of membership, 41
 depth field, 25
 description
 definition language, 3

- multimedia content, 2
- schemes, 3
- descriptors, 3
 - Fourier, 90
 - modified, 91
 - normalized, 91, 112
 - Legendre, 90
- detection
 - face, 129
 - main mobile objects, 94
 - shot change, 148
- Dirac function, 113
- dismissal, 130, 139, 143, 156, 161
 - rate, 152
- disparity
 - field, 25
 - vector, 25
- displacement vector, 26
- dissolve, 50, 148
- distance
 - feature vectors, 80
 - parametric, 80
 - weighted, 80
- DISTIMA, 55
- distortion, 53
- distribution, normal / Gaussian, 132
- dominant face, 147
 - maximum likelihood, 150
- dynamic programming, 146

E

- eigenvalues, 136
- eigenvectors, 136
- elementary story units, 148, 154
- elitist strategy, 75
- elongation, 90, 136
- encoding scheme, 72
- EP, 71
- error
 - minimum mean square, 97
 - vector, 81
- event, 152
- evolution program, 71

F

- face
 - detection, 129
 - dominant, 147
 - likelihood, 137
 - maximum, 149
 - recognition, 129
- fading, 50
- false alarm, 130, 132, 143, 151
 - rate, 152
- feature
 - extraction, 93
 - vector, 40
- feedback, 82
- feedforward neural network, 116
- FFT, 157
- finite state machines, 146
- fitness, 72

- function, 73
- fitting, B-Spline curve, 97
- focal
 - length, 23
 - point, 24
- Fourier descriptors, 90
 - modified, 115
 - normalized, 112
- frame, 1
 - feature vector, 40
 - selection, 60
 - optimal, 67
- function
 - activation, 116
 - basis, 96
 - blending, 96
 - cost, 83
 - Dirac, 113
 - signum, 106
 - turning, 91
- fusion, 33
- fuzzy classification, 41

G

- gain, 116
- Gaussian
 - distribution, 132
 - mixture model, 130, 131
- gene, 72
- genetic algorithm, 72
- global features, 45, 90, 136
- ground truth, 141, 143, 152, 161
- guided random search, 71

H

- H.261/3, 2
- hidden layer, 116
- hidden Markov models, 146
- histogram
 - bin, 40
 - color, 148
- homogeneous coordinates, 24
- Hotelling transform, 136
- HSV, 130
- hypothesis testing, 152

I

- ICL, 98
- image
 - mosaic, 4
 - non-skin, 141
 - plane, 24
 - skin, 141
- indexing, 1, 145, 156
- influence zone, 52
- information retrieval, 152
- initial population, 73
- integer encoding scheme, 72
- interaction, 82
- interest points, 90
- inverse chord length, 98
- IVC, 144

K

Karhunen-Loeve transform, 136
 kernel, 114
 key frames, 7, 60
 selection, 60
 optimal, 67
 keyword annotation, 4
 K-means, 53, 148
 knot points, 96
 reallocation, 98

L

layer, hidden, 116
 Legendre descriptors, 90
 Levenberg-Marquardt, 116
 likelihood
 face, 137
 maximum, 149
 skin-color, 132, 133
 linguistic cues, 5, 146
 links, 14
 weight, 14
 Lloyd-Max, 9, 52, 53
 local controllability, 95
 local features, 90

M

main mobile objects, 94
 major axis, 136
 manipulation, 2, 90
 Markov models, 130
 hidden, 146
 MARS, 2
 matrix
 perspective projection, 24
 reflection, 106
 rotation, 106
 maximum face likelihood, 149
 MDL, 54
 measure
 curvature, 90
 similarity, 93
 medial axis transform, 90
 median filter, 20, 150
 membership function, 41
 memory factor, 84
 merging, 15
 skin-color, 134
 MFD, 91, 115, 124
 minimum description length, 54
 minimum mean square error, 97
 minor axis, 136
 MIRAGE, 55
 mixture of Gaussians, 130, 131
 MMSE, 97
 model
 active contour, 127
 Gaussian mixture, 130
 Markov, 130
 modified Fourier descriptors, 91, 115
 moments, 90, 112
 invariants, 91

Zernike, 90

mosaic, 4
 motion estimation, 19
 MPEG, 19, 20, 46, 51, 64, 69
 MPEG 1/2, 2
 MPEG-4, 2, 32
 MPEG-7, 2, 3, 32
 M-RSST, 6, 9, 16, 117, 133
 multimedia content description, 2
 multi-ocular camera system, 23
 multiple relevance feedback, 83
 multipoint crossover, 74
 multiresolution
 decomposition, 16
 implementation, 16
 RSST, 16
 mutation, 75

N

Netra, 2
 neural network, 115
 neurons, 116
 news
 on demand, 5, 146
 topics, 148, 156
 NFD, 112, 113, 124
 NN, 115
 NoD, 5, 146
 non-skin images, 141
 nonuniform mutation, 75
 normal distribution, 132
 normalization, 91
 scaling, 99
 skew, 99
 translation, 99
 normalized Fourier descriptors, 91, 112
 NST, 142

O

objective function, 73
 occlusion, 28, 94
 detection, 28, 29
 offspring, 72, 74
 optical flow, 19
 orientation, 92
 orthogonal transformation, 99
 orthogonalization, 95, 99
 outdoor shots, 146, 147
 oversegmentation, 18

P

parameter
 transformation, 94
 vector, 81
 parametric distance, 80
 parents, 72
 selection, 73
 parsing, 3, 82, 145
 pattern matching, 93
 PCA, 5
 perspective
 projection, 24

- matrix, 24
- transformation, 90
- phase, 103
 - vector, 103
- Photobook, 2
- points
 - connection, 95
 - controll, 96
 - data, 97
 - interest, 90
 - knot, 96
- polygonal approximation, 90
- population, 72
- precision, 141, 152, 161
- preview, 1
- primary argument, 103
- principal
 - axes, 90, 136
 - component analysis, 5
- projection function, 34
- pseudorandom sequences, 9
- pyramidal region growing, 14

Q

- QBIC, 2
- query, 79
 - by example, 79, 114, 126, 156, 159
 - by sketch, 79, 90, 114, 127
 - textual, 82

R

- Radon transform, 8
- rank-based normalization, 73
- reallocation, 98
- recall, 141, 152, 161
- re-estimation, 132
- reflection, 92
 - matrix, 106
- reflectional symmetry, 105
- regions, 14
 - distance, 14
- registration, 91
- relevance feedback, 80, 82
 - multiple, 83
 - single, 83
- replacement strategy, 72, 75
- reports, 146, 147
- representative shots, 7
- retrieval, 1
 - content-based, 156
- RGB, 15, 130
- ROC curve, 132
- rotation, 92
 - matrix, 106
- rotational symmetry, 105
- roulette wheel selection, 73
- RSST, 6, 9, 14

S

- sampling, content-based, 7
- SAT, 141
- scaling, 92

- search, 1, 79
 - area, 26
- segmentation, 13
- segmented crossover, 75
- segments, 14
 - B-Splines, 95
- selective pressure, 73
- semantic
 - meaning, 13
 - segmentation, 13, 148
- shape, 89
- shift, circular, 103
- shot
 - anchorperson, 146, 147
 - change detection, 148
 - classification, 148
 - clustering, 148
 - cut detection, 50
 - dissolve, 50
 - fading, 50
 - outdoor, 146, 147
 - report, 146, 147
 - representative, 7
 - static image, 146, 147
 - transition, 148
 - unit, 154
- shuffle crossover, 75
- signum function, 106
- similarity measure, 93
 - feature vector, 80
- single point crossover, 74
- single relevance feedback, 83
- sinusoidal transform, 90
- skeleton transform, 90
- sketch, 114, 127
- skew, 92, 113
- skin images, 141
- skin-color
 - likelihood, 133
 - merging, 134
- smoothing weight function, 27
- smoothness, 95
 - function, 26
- snake, 92
- splines, 90, 120, 136
- splitting, 16
- starting point, 92
- static images, 146, 147
- story units, 4, 148
- summarization, 1, 4
- super chromosomes, 73
- symmetry
 - reflectional, 105
 - rotational, 105

T

- TAO, 141
- TAP, 141
- teletext transcripts, 146
- template
 - deformable, 92
 - matching, 131, 139

textual

cues, 5, 146

query, 82

texture, 19

threshold, segmentation, 16

thresholding, 141

transform

Fourier, 103

Hotelling, 136

Karhunen-Loeve, 136

medial axis, 90

Radon, 8

sinusoidal, 90

skeleton, 90

transformation

affine, 90, 101

coordinate, 94

orthogonal, 99

parameter, 94

perspective, 90

transition, 148

translation, 92

truncated image pyramid, 16

turning function, 91

U

uniform crossover, 74

uniform mutation, 75

uniqueness, 93

V

vector

error, 81

parameter, 81

phase, 103

weight, 81

video object planes, 2, 90

VideoQ, 2

Virage, 2

VisualSeek, 2

VOP, 2, 90

W

watershed, 14

wavelets, 5

weight vector, 81

weighted

distance, 80

geometric mean, 137

wipe, 148

world coordinate system, 24

Y

YCrCb, 130

Z

Zernike moments, 90

zoom, 148