Εθνικό Μετσόβιο Πολυτεχνείο

Σχολη Ηλεκτρολογων Μηχανικων
Και Μηχανικων Υπολογιστων

Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Εργαστηριο Εικονων, Βιντεο και Πολυμεσικων Συστηματων

Χρήση υπολογιστικών μοντέλων οπτικής προσοχής για την ανάλυση στατικών και κινούμενων εικόνων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΚΩΝΣΤΑΝΤΙΝΟΥ Ε. ΡΑΠΑΝΤΖΙΚΟΥ

Διπλωματούχου Ηλεκτρονικού Μηχανικού &
Μηχανικού Υπολογιστών Πολυτεχνείου Κρήτης

Αθήνα, Απρίλιος 2008

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων & Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης
και Υπολογιστων
Εργαστηριο Εικονων, Βιντεο
και Πολυμεσικων Συστηματων

Χρήση υπολογιστικών μοντέλων οπτικής προσοχής για την ανάλυση στατικών και κινούμενων εικόνων

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΚΩΝΣΤΑΝΤΙΝΟΥ Ε. ΡΑΠΑΝΤΖΙΚΟΥ

Διπλωματούχου Ηλεκτρονικού Μηχανικού &
Μηχανικού Υπολογιστών Πολυτεχνείου Κρήτης

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
Πέτρος Μαραγκός
Ανδρέας–Γεώργιος Σταφυλοπάτης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 04η 19η Ιουλίου 2006.

...	...	...
Σ. Κόλλιας	Π. Μαραγκός	Α.–Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.
...	...	...
Π. Τσανάκας	Κ. Κοντογιάννης	Μ. Ζερβάκης
Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής Ε.Μ.Π.	Καθηγητής Π.Κ.
	...	
	Ν. Τσαπατσούλης	
	Επ. Καθηγητής Τ.Π.Κ.	

Αθήνα, Απρίλιος 2008

. . . . . .

ΚΩΝΣΤΑΝΤΙΝΟΣ Ε. ΡΑΠΑΝΤΖΙΚΟΣ
Ηλεκτρονικός Μηχανικός & Μηχανικός Υπολογιστών Πολυτεχνείου Κρήτης

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Στους γονείς μου.
Η εμπιστοσύνη τους είναι η δύναμη μου.

ΠΡΟΛΟΓΟΣ

Η παρούσα διδακτορική διατριβή έχει σαν κύριο θέμα την μελέτη, ανάπτυξη και εφαρμογή μοντέλων οπτικής προσοχής σε στατικές και κινούμενες εικόνες. Καταγράφει τους ερευνητικούς στόχους και τις συνεισφορές στον χώρο, όπως αυτές προέκυψαν μέσα από προσωπική μελέτη και δημιουργικές συνεργασίες. Κίνητρο αποτέλεσε η βιολογική σημασία του μηχανισμού οπτικής προσοχής στον άνθρωπο και ο καθοριστικός του ρόλος στην κατανόηση του οπτικού ερεθίσματος. Στόχος μας ήταν η μεταφορά ανάλογων μηχανισμών στην μηχανική όραση και η αξιολόγηση τους στην βελτίωση καθιερωμένων τεχνικών και στην εισαγωγή νέων μεθόδων επεξεργασίας.

Η έννοια της οπτικής προσοχής σχετίζεται με την ικανότητα του ανθρώπου να εστιάζει σχεδόν αυτόματα σε περιοχές που θεωρεί ενδιαφέρουσες αγνοώντας καποιες άλλες. Η διαδικασία αυτή είναι ουσιώδης για την κατανόηση του οπτικού ερεθίσματος και δρομολογείται μέσα από ένα καθοδικό (top-down) και ένα ανοδικό (bottom-up) κανάλι επεξεργασίας του ανθρώπινου οπτικού συστήματος. Η διατριβή στο σύνολο της στηρίζεται σε έναν φαινομενικά απλό ισχυρισμό, ο οποίος βρίσκεται σε αναλογία με την προηγούμενη ικανότητα του ανθρώπου: Η πληροφορία που εξάγεται από κατάλληλα επιλεγμένες περιοχές της εικόνας ή ακολουθίας είναι αρκετή για να περιγράψει το περιεχόμενο της ή να την χαρακτηρίσει συνολικά. Ταυτόχρονα η έρευνα μας προχώρησε λαμβάνοντας υπόψη το ότι ένα υπολογιστικό μοντέλο με βιολογικό ανάλογο πρέπει να κρίνεται με βάση δύο κριτήρια, όπως υποστηρίζουν οι Tsotsos *et al.* [137]: "Το πρώτο κριτήριο είναι η εγγύτητα του υπολογιστικού και βιολογικού μοντέλου, που αποδεικνύεται με ψυχοφυσικά πειράματα και το δεύτερο είναι η επιτυχία του υπολογιστικού μοντέλου να επιλύσει κοινά προβλήματα μηχανικής όρασης". Μπορεί επομένως ένα υπολογιστικό μοντέλο να μιμείται ικανοποιητικά τον ανθρώπινο τρόπο εστίασης προσοχής, αλλά να μην παρέχει την σωστότερη λύση σε ένα πρόβλημα μηχανικής όρασης κάτω από συγκεκριμένες συνθήκες και απαιτήσεις.

Στο πρώτο μέρος της διατριβής ασχολούμαστε με την μελέτη και επέκταση καθιερωμένων τεχνικών οπτικής προσοχής σε εικόνες. Περιγράφουμε αναλυτικά τις ιδιότητες τους και τα αξιολογούμε σε διαφορετικές εφαρμογές. Στο δεύτερο μέρος προτείνουμε νέα μοντέλα υπολογισμού τιμών ενδιαφέροντος για χωροχρονικές ακολουθίες. Περιγράφουμε τα βασικά στοιχεία τους, εισάγουμε νέους τελεστές και προτείνουμε νέα υπολογιστικά μοντέλα οπτικής προσοχής, τα οποία στηρίζονται σε λειτουργίες βιολογικών μοντέλων. Η διατριβή θα έχει πετύχει τον στόχο της αν αποτελέσει κίνητρο για μελλοντική ενασχόληση με εφαρμογή παρόμοιων μεθόδων στην ανάλυση και επεξεργασία οπτικής πληροφορίας. Το πρόσφατα αυξανόμενο ενδιαφέρον της ερευνητικής κοινότητας για βασική έρευνα και εφαρμογές γύρω από μοντέλα οπτική προσοχής είναι έκδηλο και φαίνεται να επιβεβαιώνει τα κίνητρα και τους στόχους της έρευνας μας.

ABSTRACT

Although human vision appears to be easy and unconscious there exist complex neural mechanisms in primary visual cortex that form the preattentive component of the Human Visual System (HVS) and lead to visual awareness. Considerable research has been carried out into the attention mechanisms of the HVS and computational models have been developed and employed to common computer vision problems. Most of the models simulate the bottom-up mechanism of the HVS and their major goal is to filter out redundant visual information and detect/enhance the most salient parts of the input. The Human Visual System (HVS) has the ability to fixate quickly on the most informative (salient) regions of a scene and reduce therefore the inherent visual uncertainty. Computational visual attention (VA) schemes have been proposed to account for this important characteristic of the HVS. The dissertation studies and expands the field of computational visual attention methods, proposes novel models both for spatial (images) and spatiotemporal (video sequences) analysis and evaluates both qualitatively and quantitavely in a variety of relevant applications.

Ευχαριστίες

Η εκπόνηση της παρούσας διδακτορικής διατριβής αποτέλεσε σημαντικό μέρος της ζωής μου. Οι διαπροσωπικές επαφές που ανέπτυξα με συνεργάτες και φίλους αυτά τα χρόνια ίσως είναι πιο σημαντικές από τις τεχνικές γνώσεις που αποκτήθηκαν.

Ιδιαίτερη ευγνωμοσύνη οφείλω στον κ. Στέφανο Κόλλια, ο οποίος μου έδωσε την δυνατότητα να εκπονήσω την παρούσα διατριβή κάτω από την επίβλεψη του στο εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων. Όλα αυτά τα χρόνια ο κ. Κόλλιας φροντίζει προσωπικά, ώστε το εργαστήριο να αποτελεί έναν δημιουργικό και συνεχώς εξελισσόμενο χώρο. Η εμπιστοσύνη του, οι ευκαιρίες που μου έδωσε και οι εύστοχες συμβουλές του την κατάλληλη στιγμή έκαναν εφικτή την παρούσα εργασία. Η σταθερά αισιόδοξη στάση και η θετική του σκέψη αποτέλεσαν μόνιμο κίνητρο για περισσότερη προσπάθεια με λιγότερο άγχος.

Στην διάρκεια του διδακτορικού συνεργάστηκα στενά με τον Νίκο Τσαπατσούλη και τον Γιάννη Αβρίθη. Ο Νίκος είναι ο ηθικός αυτουργός της ενασχόλησης μου με τον χώρο της υπολογιστικής οπτικής προσοχής σε εικόνες. Με στήριξε στα πρώτα μου βήματα με την μεθοδικότητα, την επιμονή και την υπομονή του. Θεωρώ ότι το σημαντικότερο που μου έμαθε είναι να μένω πιστός στον στόχο μου ακόμη και όταν αυτό φαίνεται δύσκολο. Οι συζητήσεις μας ήταν συχνά μακροσκελείς, πάντα ενδιαφέρουσες και συχνά με κοινωνικές προεκτάσεις.

Η συνεργασία μου με τον Γιάννη Αβρίθη οδήγησε στην μελέτη τεχνικών στο χώρο των κινούμενων εικόνων που αποτελεί και το μεγαλύτερο μέρος της διατριβής. Είναι δύσκολο να συνοψίσω την συνεισφορά του όλα αυτά τα χρόνια, αλλά θα σταθώ κυρίως στο ακούραστο πνεύμα του, στις δημιουργικές ιδέες και στην συνεχή υποστήριξη του. Με την εργατικότητα και το καθαρό του μυαλό αποτέλεσε πρότυπο για εμένα, ιδιαίτερα στις δύσκολες περιόδους του διδακτορικού που λόγω πίεσης ή συνθηκών έχανα το κίνητρο μου. Είναι από τους βασικούς υπεύθυνους της χαράς που μου δίνει η ολοκλήρωση της παρούσας διατριβής. Οι συζητήσεις μας, οι -συχνά πολύωρες- συναντήσεις μας και οι συμβουλές του μου άφηναν πάντα έναν ενθουσιασμό ικανό να με κάνει να προχωρήσω ακόμη ένα βήμα.

Στην αρχή του διδακτορικού είχα την τύχη να γνωρίσω τον κ. Πέτρο Μαραγκό, μέλος της συμβουλευτικής μου επιτροπής, μέσα από τις διαλέξεις του για μη-γραμμικά συστήματα, χάος και φράκταλς. Αν και συχνά απογοητευόμουν από την δυσκολία αφομοίωσης του συνόλου των διαλέξεων, η εμπεριστατωμένη άποψη και ο ενθουσιασμός του για το επιστημονικό του πεδίο κράτησαν αμείωτο το ενδιαφέρον μου. Οφείλω να τον ευχαριστήσω τόσο για τα κίνητρα για βασική έρευνα που μου έδωσε στην αρχή όσο και για τις μετέπειτα συζητήσεις και την ερευνητική μας συνεργασία.

Η ενασχόληση μου με τον χώρο της ψηφιακής επεξεργασίας εικόνας και σήματος ξεκίνησε στο δεύτερο έτος των προπτυχιακών μου σπουδών, όταν οι διαλέξεις του κ. Μιχάλη Ζερβάκη κίνησαν το ενδιαφέρον μου για τον χώρο. Ήταν η αρχή μίας πολύχρονης συνεργασίας που οδήγησε στην εκπόνηση της διπλωματικής και μεταπτυχιακής μου διατριβής. Του οφείλω μεγάλο μέρος του ενδιαφέροντος μου για τον τομέα, το οποίο παραμένει αμείωτο όλα αυτά τα χρόνια. Τις ευχαριστίες μου οφείλω να εκφράσω και στον κ. Κωνσταντίνο Μπάλα. Η συνεργασία μας με ωφέλησε σε προσωπικό και επαγγελματικό επίπεδο και μου άνοιξε νέους ορίζοντες στον χώρο της επεξεργασίας εικόνας.

Ευτυχώς που είχα γύρω μου καλούς φίλους -εντός και εκτός του εργαστηρίου- για να μου θυμίζουν ότι η ζωή έχει πολλές πλευρές, οι οποίες δεν σχετίζονται μόνο με εργασία και μελέτη. Η ολοκλήρωση του διδακτορικού δεν θα ήταν εφικτή χωρίς αυτούς. Η ενασχόληση με διαφορετικά επιστημονικά πεδία, η ποικιλία απόψεων, οι κοινωνικές ευαισθησίες και η δημιουργικά ανάλαφρη διάθεση κάποιων μελών του εργαστηρίου δημιουργούσαν μία μόνιμα ευχάριστη ατμόσφαιρα. Η παρέα από Σαλαμίνα, ο Νίκος, ο Κώστας και οι λίγοι αλλά διαλεχτοί στενοί μου φίλοι από τα πρώτα φοιτητικά χρόνια με βοήθησαν πολύ με την εμψυχωτική τους στάση και την ικανότητα τους να με στηρίζουν τόσο στη χαρά όσο και στη στενοχώρια.

Στην οικογένεια μου οφείλω τα πάντα. Η κατανόηση του στενού μου οικογενειακού κύκλου σε κάθε -εκ των υστέρων- σωστή ή λανθασμένη απόφαση μου ήταν η ίδια. Οι γονείς μου, στους οποίους αφιερώνεται η παρούσα διατριβή, με περιέβαλαν πάντα με αγάπη, υπομονή και απόλυτη εμπιστοσύνη. Κάθε επιτυχημένο ακαδημαϊκό ή επαγγελματικό μου ξεκίνημα οφείλεται κυρίως στην υποστήριξη που μου παρείχαν.

Κατάλογος συντμήσεων

1. **GOP** : Group Of Pictures
2. **DWT** : Discrete Wavelet Transform
3. **DTCWT** : Dual Tree Complex Wavelet Transform
4. **WTA** : Winner-Take-All
5. **IOR** : Inhibition-Of-Return
6. **pLSA** : probabilistic Latent Semantic analysis

Κατάλογος Απόδοσης Όρων

1. **χαρακτηρισμός/σχολιασμός** : annotation
2. **συρρίκνωση** : shrinkage
3. **ανοδικό κανάλι** : bottom-up channel
4. **καθοδικό κανάλι** : top-down channel
5. **ενδιάμεσος χάρτης ενδιαφέροντος** : conspicuity map
6. **δομοστοιχείο** : component
7. **ακρίβεια** : precision
8. **επανάκληση** : recall
9. **αναστολής της επιστροφής** : Inhibition-Of-Return/IOR
10. **σάκος λέξεων** : bag-of-words
11. **εξίσωση περιορισμού οπτικής ροής** : optical flow constraint equation
12. **περιορισμός χωρικής συνάφειας** : spatial coherence constraint
13. **μεταβλητος δυφιορρυθμός** : variable bit rate
14. **δυφιακός ρυθμός**: bit rate
15. **οπτικό περιβάλλον** : visual context
16. **αντιληπτική ανίχνευση** : perceptual detection
17. **όλα-για-το-νικητή** : winner-take-all
18. **χάρτης ενδιαφέροντος, κύριος χάρτης** : saliency map
19. **μοντέλο οπτικού συντονισμού** : selective tuning model
20. **εμπροσθόδοτος** : feed-forward
21. **φίλτρο κέντρου-περιφέρειας** : center-surround filter
22. **θεωρία διπλής χρωματικής αντίθεσης** : color double-opponent theory
23. **αντίπαλη διαδικασία** : opponent process

24. **καθοδηγήσιμο φίλτρο** : steerable filter
25. **Ανάλυση Κύριων Συνιστωσών** : Principal Components Analysis
26. **επικύρωση** : cross-validation
27. **Μορφή** : Gestalt
28. **ογκοστοιχείο** : voxel
29. **co-occurrence matrix** : πίνακας σύγχυσης
30. **ασθενής επισηματοθέτηση** : soft labeling
31. **ενδο-χαρακτηριστικός** : inter-feature
32. **δια-χαρακτηριστικός** : inter-feature
33. **δια-κλιμακωτός** : inter-scale
34. **με-μία-παράλειψη** : leave-one-out
35. **κατάβαση κλίσης** : gradient descent
36. **μέγιστη κατάβαση κλίσης** : steepest gradient descent
37. **κώδικας λέξεων** : codebook
38. **ενδείκτης ενδιαφέροντος** : saliency indicator function
39. **ιδιοχώρος** : eigenspace
40. **εξειδικευτικότητα** : specificity
41. **ευαισθησία** : sensitivity
42. **διαχωρίσιμος** : separable

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο

Οι περισσότερες εντυπώσεις και μνήμες μας για τον κόσμο βασίζονται στην όραση. Και όμως οι μηχανισμοί που παρεμβαίνουν στην όραση δεν είναι προφανείς ούτε για εκείνον που βλέπει ούτε για τον ερευνητή της όρασης. Πώς αντιλαμβανόμαστε την μορφή; Πώς αντιλαμβανόμαστε την κίνηση των αντικειμένων στο χώρο; Πώς αντιλαμβανόμαστε το χρώμα; Τα πρωτεύοντα θηλαστικά, συμπεριλαμβανόμενου του ανθρώπου, χρησιμοποιούν εστιασμένη οπτική προσοχή και γρήγορες κινήσεις των ματιών για να αναλύσουν πολύπλοκα οπτικά ερεθίσματα με έναν τρόπο που εξαρτάται από τις τρέχουσες συνθήκες ή από τον επιθυμητό στόχο. Η διαδικασία της οπτικής προσοχής θα μπορούσε επομένως να χαρακτηριστεί αφαιρετικά ως ο μηχανισμός που μας επιτρέπει να κατευθύνουμε το βλέμμα μας σε αντικείμενα ενδιαφέροντος στο περιβάλλον. Από την άποψη της εξέλιξης, η ικανότητα αυτή του ανθρώπου να κατευθύνει γρήγορα την προσοχή του ήταν και είναι σημαντική για να αντιμετωπίζει έγκαιρα πιθανούς κινδύνους από εισβολείς ή άλλα εμπόδια που παρουσιάζονται στο οπτικό του πεδίο.

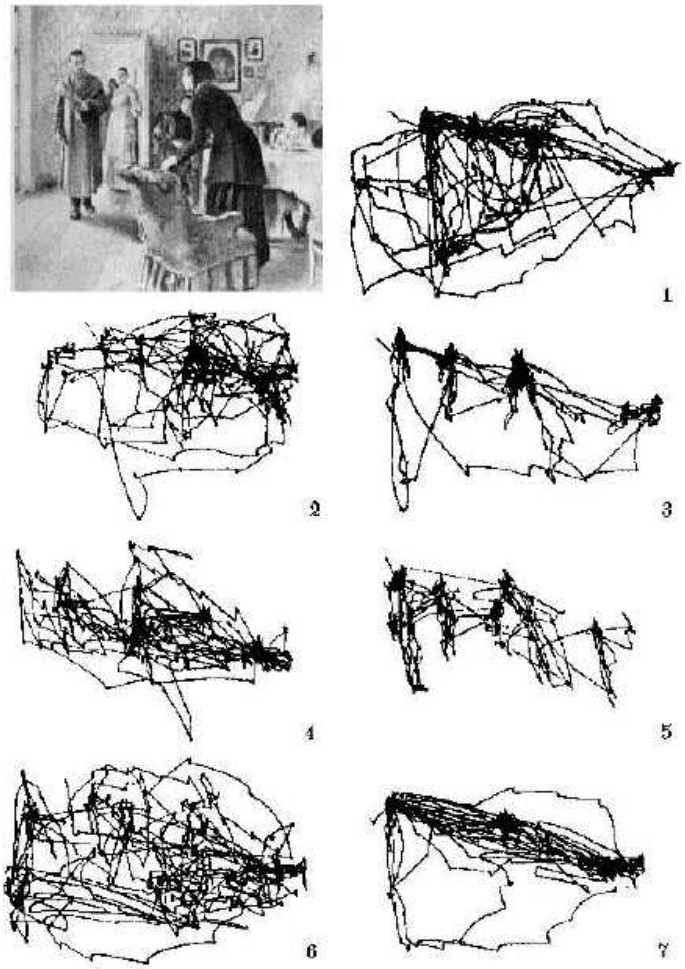
Σημαντική στην κατανόηση αυτής της λειτουργίας του ανθρώπινου οπτικού συστήματος ήταν η συνεισφορά του Yarbus [150], ο οποίος χρησιμοποίησε μία συσκευή παρακολούθησης σε ανθρώπους για να καταγράψει τις κινήσεις των ματιών τους στην διάρκεια παρακολούθησης μίας συγκεκριμένης φωτογραφίας υπό διαφορετικές οδηγίες. Το Σχήμα 1.1 δείχνει τα αποτελέσματα του πειράματος, τα οποία κάνουν προφανή την επίδραση της γνώσης για τον επιθυμητό στόχου στον τρόπο εστίασης και κατανόησης της σκηνής. Παρόμοια αποτελέσματα παρουσίασαν και οι Tanenhaus *et al.* [125]. Σε αυτό το πείραμα δόθηκε περισσότερη έμφαση στην επίδραση που έχει το οπτικό περιβάλλον[15] στην εστίαση της προσοχής. Η κατανόηση της λειτουργίας του ανθρώπινου συστήματος οπτικής προσοχής μπορεί να αποδειχθεί χρήσιμη στην υπολογιστική όραση τόσο σε περιπτώσεις μερικής ή ολικής έλλειψης περιορισμών στην σκηνή όσο και σε περιπτώσεις αλγορίθμων με μεγάλη υπολογιστική πολυπλοκότητα. Στην πρώτη περίπτωση οι διαδοχικές εστιάσεις στις σημαντικές περιοχές της εικόνας μπορούν να δημιουργήσουν επιθυμητούς περιορισμούς που είναι κρίσιμοι για την αποτελεσματικότητα του αλγόριθμου ανάλυσης που ακολουθεί. Στην δεύτερη περίπτωση το πρόβλημα επεξεργασίας της συνολικής οπτικής σκηνής μπορεί να αναλυθεί σε μικρότερα σειριακά προβλήματα επεξεργασίας περιοχών που προκύπτουν από τις διαδοχικές εστιάσεις. Οι ποικίλες εφαρμογές της υπολογιστικής όρασης που απαιτούν μεθόδους ανίχνευσης περιοχών σύμφωνες με την ανθρώπινη αντίληψη, όπως π.χ. η

επίβλεψη σκηνής, ο αυτόματος εντοπισμός στόχων ενδιαφέροντος, η καθοδήγηση μηχανών σε άγνωστα/γνωστά περιβάλλοντα θα μπορούσαν να επωφεληθούν από την ανάπτυξη υπολογιστικών μοντέλων οπτικής προσοχής.

Όταν ξεκίνησε η εκπόνηση της παρούσας διατριβής, οι εφαρμογές των καθιερωμένων υπολογιστικών μοντέλων ήταν περιορισμένες, αλλά ενδεικτικές για τον βελτιωτικό ρόλο τους σε συγκεκριμένες εφαρμογές μηχανικής όρασης. Ένα από τα γνωστότερα μοντέλα μέχρι σήμερα είναι αυτό των Itti *et al.*, το οποίο πλέον έχει εφαρμοστεί σε ποικίλες εφαρμογές από τους ίδιους συγγραφείς αλλά και πολλούς άλλους ερευνητές του χώρου [48]. Επομένως, εκτός από την θεωρητικά αναμενόμενη βελτίωση μεθόδων μηχανικής όρασης με την χρήση οπτικής προσοχής υπήρχαν και πειραματικές ενδείξεις -οι οποίες επιβεβαιώθηκαν με την πάροδο του χρόνου- που μας ώθησαν να μελετήσουμε την περιοχή. Σύντομα -μετά την πρώτη εμπεριστατωμένη μελέτη της βιβλιογραφίας του χώρου- προέκυψε ένα επιπλέον κίνητρο. Ήταν φανερή η έλλειψη κάποιου μοντέλου οπτικής προσοχής, το οποίο να λαμβάνει υπόψη τόσο την χωρική όσο και την χρονική εξέλιξη σε μία ακολουθία, καθώς η μέχρι τότε προσέγγιση ήταν η επεξεργασία ανά καρέ. Αυτό το κίνητρο άνοιξε νέους δρόμους στην έρευνα μας και οδήγησε στην ανάπτυξη και υλοποίηση μίας σειράς μοντέλων για χωροχρονική ανάλυση ακολουθιών. θα ήταν παράλειψη να μην αναφερθούμε και στον ρόλο που έπαιξε η σύντομη, αλλά ουσιαστική, ενασχόληση μας με την εισαγωγή νέων περιορισμών με τη μορφή τιμών εμπιστοσύνης σε μεθόδους ενεργών περιγραμμάτων. Συγκεκριμένα, στα πλαίσια ερευνητικής συνεργασίας προέκυψε ένας τελεστής εμπιστοσύνης, ο οποίος βασίζεται στην ιστορία της κίνησης των αντικειμένων, και ο οποίος χρησιμοποιείται στην πόλωση ενεργών περιγραμμάτων για κατάτμηση περιοχών σε ακολουθίες με χειρωνακτική αρχικοποίηση [141] [142]. Η αναζήτηση μας για εναλλακτικούς τρόπους υπολογισμού εμπιστοσύνης για μεθόδους ενεργών περιγραμμάτων μας έφερε πιο κοντά στην μελέτη τεχνικών υπολογισμού ενδιαφέροντος με οπτική προσοχή.

1.2 Στόχοι

Όπως αναφέραμε στον Πρόλογο, αν και οι μηχανισμοί οπτικής προσοχής έχουν το βιολογικό τους ανάλογο στο ανθρώπινο οπτικό σύστημα, ένα υπολογιστικό μοντέλο πρέπει να κρίνεται με βάση δύο κριτήρια: "Το πρώτο κριτήριο είναι η εγγύτητα του υπολογιστικού και βιολογικού μοντέλου, που αποδεικνύεται με ψυχοφυσικά πειράματα και το δεύτερο είναι η επιτυχία του υπολογιστικού μοντέλου να επιλύσει κοινά προβλήματα μηχανικής όρασης". Μπορεί επομένως ένα υπολογιστικό μοντέλο να μιμείται ικανοποιητικά τον ανθρώπινο τρόπο εστίασης προσοχής, αλλά να μην παρέχει την σωστότερη λύση σε ένα πρόβλημα μηχανικής όρασης κάτω από συγκεκριμένες συνθήκες και απαιτήσεις! Η ικανότητα προσαρμογής του ανθρώπινου οπτικού συστήματος είναι προς το παρόν αδύνατον να μοντελοποιηθεί υπολογιστικά γιατί μέρη του συστήματος παραμένουν άγνωστα ακόμη και για ερευνητές από τον χώρο της νευροεπιστήμης και φυσιολογίας. Στόχος μας λοιπόν δεν ήταν η -κατά το δυνατόν- πιστή μίμηση κάποιου βιολογικού μοντέλου οπτικής προσοχής, αλλά η ανάπτυξη αποδοτικών υπολογιστικών μοντέλων με εφαρμογή σε πραγματικά προβλήματα της μηχανικής όρασης. Σε αυτό το σημείο πρέπει να διευκρινίσουμε και την ορθότητα χρήσης κάποιων όρων: ο όρος "οπτική προσοχή" σχετίζεται περισσότερο με τα βιολογικά μοντέλα, καθώς αυτά είναι που περιγράφουν τον ανθρώπινο τρόπο εστίασης προσοχής. Στα υπολογιστικά μοντέλα ο όρος χρησιμοποιείται καταχρηστικά για να



Σχήμα 1.1: *Το οπτικό ερέθισμα και οι καταγεγραμμένες κινήσεις των ματιών υπό 7 διαφορετικές λεκτικές οδηγίες: 1) ελεύθερη παρατήρηση της εικόνας, 2) σκεφτείτε ένα πιθανό σενάριο, 3) αναφέρετε τις ηλικίες των ατόμων, 4) σκεφτείτε τι έκανε η οικογένεια πριν την άφιξη του απρόσμενου επισκέπτη, 5) απομνημονεύστε τα ρούχα των ατόμων στην σκηνή, 6) απομνημονεύστε την θέση των ατόμων και των αντικειμένων στη σκηνή, 7) υπολογίστε πόσο χρόνο ο απρόσμενος επισκέπτης ήταν μακριά από την οικογένεια. (από [150])*

δείξει τον συσχετισμό. Στα κεφάλαια που ακολουθούν χρησιμοποιούμε τον όρο αυτόν κυρίως για τα χωρικά μοντέλα που προτείνουμε, ενώ για τα χωροχρονικά προτιμούμε τον όρο "υπολογισμός τιμών ενδιαφέροντος" ή "υπολογισμός σημαντικότητας".

Συνοπτικά, οι βασικοί άξονες της έρευνας που καταγράφεται στην παρούσα διατριβή, οι οποίοι σχετίζονται και με τους στόχους της, είναι οι εξής:

- **Μελέτη:** Μελέτη της βιβλιογραφίας που σχετίζεται με τα βιολογικά και τα υπολογιστικά μοντέλα οπτικής προσοχής. Στόχος είναι η κατανόηση των αρχών των καθιερωμένων βιολογικών μοντέλων και η πλήρης ενημέρωση για τα υπολογιστικά μοντέλα που έχουν προταθεί. Η δυσκολία έγκειται κυρίως στον εντοπισμό κατάλληλων πηγών για τα βιολογικά μοντέλα, καθώς δεν είμαστε ειδικοί στον χώρο.

- **Συνεισφορά:** Προτάσεις για επέκταση καθιερωμένων μεθόδων και για καινοτόμα μοντέλα οπτικής προσοχής ή υπολογισμού τιμών ενδιαφέροντος. Επεκτάσεις επιλεγμένων μεθόδων χωρικής οπτικής προσοχής και κυρίως προτάσεις για

νέα μοντέλα χωροχρονικής ανάλυσης πληροφορίας. Όπως αναφέραμε νωρίτερα και σύμφωνα με την μέχρι σήμερα ενημέρωση μας, δεν υπάρχουν υπολογιστικά μοντέλα οπτικής προσοχής για πραγματικά χωροχρονική πληροφορία (συνήθως είναι ανά καρέ).

- **Ανάπτυξη:** Ανάπτυξη καθιερωμένων/δημοσιευμένων τεχνικών και πειραματισμός, ώστε να κατανοηθεί σε βάθος η λειτουργία τους. Στόχος η ανάδειξη των πλεονεκτημάτων/μειονεκτήματων τους, ώστε να οδηγηθούμε στην ανάπτυξη άρτιων και αποδοτικών νέων μεθόδων και μοντέλων. Αν και δεν αποτέλεσε πρωταρχικό στόχο της έρευνας μας, δόθηκε βαρύτητα και στην ανάπτυξη των μεθόδων με όσο το δυνατόν χαμηλή υπολογιστική πολυπλοκότητα για να είναι εφικτή η εφαρμογή τους σε μεγάλο όγκο δεδομένων.

- **Αξιολόγηση:** Αξιολόγηση όλων των μεθόδων που περιγράφονται σε ένα μεγάλο εύρος εφαρμογών, ώστε να αναδειχθεί η αξία της χρήσης οπτικής προσοχής σε τομείς της μηχανικής όρασης. Δυσκολίες αντιμετωπίσαμε λόγω έλλειψης χαρακτηρισμού[1] δεδομένων που να σχετίζεται με την χωρική ή χωροχρονική σημασία τους. Η υποκειμενικότητα του πεδίου είναι μεγάλη, καθώς το αποτέλεσμα της οπτικής προσοχή διαφέρει από άνθρωπο σε άνθρωπο.

1.3 Διάρθρωση της διατριβής

Στο **Κεφάλαιο 2** παρουσιάζουμε αναλυτική ανασκόπηση βιολογικών και υπολογιστικών μοντέλων οπτικής προσοχής. Παραθέτουμε επίσης την ανασκόπηση των πεδίων εφαρμογής όλων των μοντέλων που περιγράφουμε ή προτείνουμε στα κεφάλαια που ακολουθούν.

Το **Κεφάλαιο 3** περιλαμβάνει τις επεκτάσεις που προτείνουμε σε καθιερωμένα μοντέλα χωρικής οπτικής προσοχής. Περιγράφονται αναλυτικά οι αλλαγές/επεκτάσεις και παρουσιάζεται αναλυτική αξιολόγηση τους με πειραματικά αποτελέσματα σε αποθορυβοποίηση εικόνων με χρήση τιμών ενδιαφέροντος, σε ανίχνευση προσώπων υπό περίπλοκες συνθήκες και σε αποτελεσματική κωδικοποίηση ακολουθιών με χρήση περιοχών ενδιαφέροντος.

Στο **Κεφάλαιο 4** παρουσιάζουμε το πρώτο μοντέλο υπολογισμού τιμών ενδιαφέροντος με χρήση χωροχρονικής πληροφορίας της ακολουθίας. Πρόκειται για ένα μοντέλο που βασίζεται στον υπολογισμό χωροχρονικής κατευθυντικότητας στον χώρο των 3Δ κυματιδίων. Δίνεται αναλυτική περιγραφή των δυνατοτήτων του χώρου 3Δ κυματιδίων στην αναπαράσταση των κατευθύνσεων της ακολουθίας και παρουσιάζονται αποτελέσματα στο πεδίο αναγνώρισης ανθρώπινων οπτικών δραστηριοτήτων.

Το **Κεφάλαιο 5** περιγράφει και αναλύει ένα ολοκληρωμένο μοντέλο υπολογισμού χωροχρονικών τιμών ενδιαφέροντος, το οποίο περιλαμβάνει περισσότερα χαρακτηριστικά (φωτεινότητα, χρώμα, κατευθυντικότητα) και κινείται γύρω από την λογική του μοντέλου των Itti *et al.*. Η ποσοτική αξιολόγηση του μοντέλου γίνεται σε δύο εφαρμογές: επίβλεψη εσωτερικών χώρων και αναγνώριση σκηνής.

Στο **Κεφάλαιο 6** παρουσιάζουμε ένα διαφορετικό μοντέλο χωροχρονικής προσοχής, το οποίο βασίζεται στον δυναμικό ανταγωνισμό μεταξύ χαρακτηριστικών για τον τελικό υπολογισμό των τιμών ενδιαφέροντος. Στα πειραματικά αποτελέσματα περιλαμβάνεται μία συνολική αξιολόγηση όλων των μοντέλων που προτείναμε και

χρησιμοποιήσαμε σε αναγνώριση σκηνής, καθώς και η ξεχωριστή αξιολόγηση του τελευταίου μοντέλου στον χώρο της ανίχνευσης οπτικών δραστηριοτήτων και στην δημιουργία περιλήψεων ακολουθιών που σχετίζονται με την ανθρώπινη αντίληψη.

□

Κεφάλαιο 2

Επισκόπηση μοντέλων οπτικής προσοχής

□

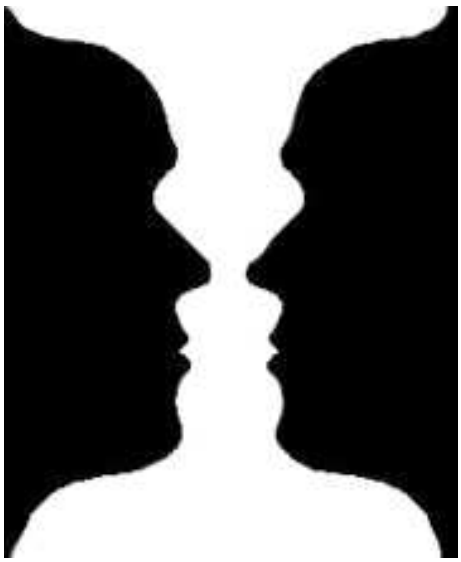
2.1 Εισαγωγή

Μελέτες από τον χώρο της νευροεπιστήμης αποδεικνύουν ότι οι νευρικές πληροφορίες που έχουν σχέση με την μορφή, την κίνηση και το χρώμα δεν μεταφέρονται από μία μόνο ιεραρχική οδό (ή ρεύμα επεξεργασίας) αλλά από τρεις τουλάχιστον (και πιθανόν περισσότερες) παράλληλες και διαπλεκόμενες οδούς στον εγκέφαλο. Η ύπαρξη παράλληλων οδών επεξεργασίας εγείρει ένα δεύτερο πρόβλημα, το πρόβλημα της σύνδεσης. Πώς ενώνονται σε μία μόνο εικόνα οι πληροφορίες που μεταφέρονται με τρεις χωριστές οδούς; Αντιμετωπίζοντας το πρόβλημα σύνδεσης, οι ερευνητές του χώρου εστιάζουν το ενδιαφέρον τους σε ένα από τα κεντρικά ερωτήματα της γνωστικής λειτουργίας: Πώς ο εγκέφαλος συγκροτεί τον κόσμο, ο οποίος γίνεται αντιληπτός μέσω των αισθητικών πληροφοριών, πώς τον φέρνει στην συνείδηση και ποιος είναι ο ρόλος της οπτικής προσοχής σε αυτήν τη διαδικασία; Σε αυτό το κεφάλαιο αναφερόμαστε στις βασικές ανατομικές λειτουργίες του εγκεφάλου, περιγράφουμε καθιερωμένα υπολογιστικά μοντέλα και επιχειρούμε επισκόπηση των εφαρμογών τους.

2.2 Βιολογικά μοντέλα οπτικής προσοχής

2.2.1 Σχηματισμός της οπτικής αντίληψης

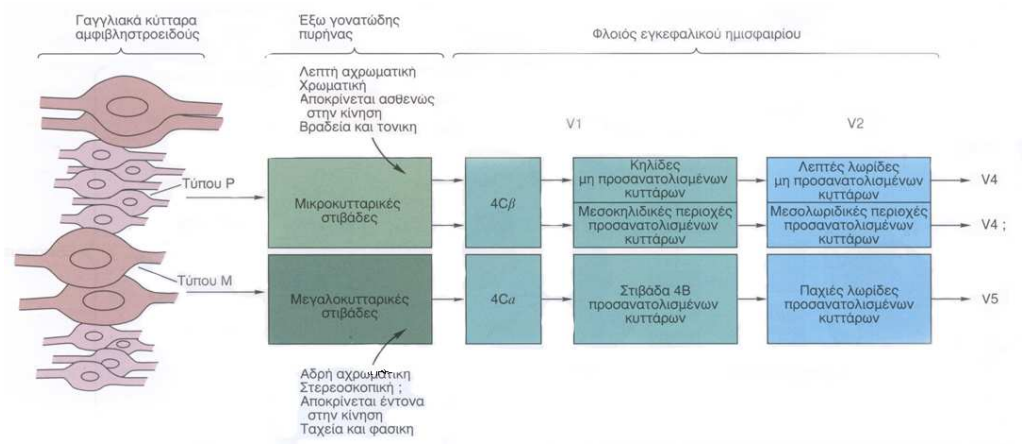
Το ανθρώπινο οπτικό σύστημα δημιουργεί μία τρισδιάστατη αντίληψη του κόσμου η οποία διαφέρει από την απλή δυσδιάστατη απεικόνιση του οπτικού ερεθίσματος στον αμφιβληστροειδή χιτώνα του ματιού. Αυτή η αντίληψη μας επιτρέπει να αναγνωρίζουμε ένα αντικείμενο ακόμη και όταν η πραγματική του εικόνα στον αμφιβληστροειδή διαφέρει σημαντικά λόγω αλλαγών φωτισμού ή προοπτικής. Ένα κόκκινο αντικείμενο π.χ. γίνεται αντιληπτό είτε σε ένα δωμάτιο με χαμηλό φωτισμό, είτε σε κάποιο εξωτερικό χώρο υπό το φως του ήλιου. Αντίστοιχα όταν ένας άνθρωπος περπατά προς το μέρος μας αντιλαμβανόμαστε ότι μας πλησιάζει και όχι ότι μεγαλώνει, όπως θα υποδήλωνε η απεικόνιση του αμφιβληστροειδούς. Η ικανότητα μας αυτή να



Σχήμα 2.1: *Σε αυτήν την οπτική απάτη βλέπουμε είτε δύο σκούρα πρόσωπα σε προφίλ είτε ένα άσπρο βάζο σε σκούρο φόντο (στρατηγική όλα-για-τον-νικητή)*

αντιλαμβανόμαστε το μέγεθος ή το χρώμα ενός αντικειμένου ως σταθερό ανεξαρτήτως συνθηκών δείχνει σαφώς ότι κάποια οπτικά χαρακτηριστικά είναι σημαντικά και μας βοηθούν να κατανοούμε το περιβάλλον.

Η σύγχρονη γνωστική άποψη ότι η αντίληψη είναι μία μετασχηματιστική και δημιουργική διεργασία, η οποία δεν αποτελείται απλά και μόνο από την πρόσληψη αισθητικών πληροφοριών, διατυπώθηκε για πρώτη φορά στις αρχές του 20ού αιώνα από τους Γερμανούς ψυχολόγους Max Wertheimer, Kurt Koffka και Wolfgang Kuhler, οι οποίοι ίδρυσαν την σχολή της Gestalt ψυχολογίας. Η γερμανική λέξη Gestalt σημαίνει μορφή και η κεντρική άποψη των ψυχολόγων αυτής της σχολής είναι ότι η διεργασία της αντίληψης διαμορφώνει την πλήρη μορφή από τις επιμέρους λεπτομέρειες του ερεθίσματος. Μία εικόνα επομένως δεν είναι άθροισμα των αντιληπτικών στοιχείων της, όπως πίστευαν οι εμπειριστές φιλόσοφοι του 17ου και 18ου αιώνα, αλλά τα στοιχεία της είναι επιλεκτικά οργανωμένα από τον εγκέφαλο μας ώστε να δημιουργούν μία μορφή η οποία υπερβαίνει το άθροισμα των τμημάτων της. Οι ψυχολόγοι της μορφής αρέσκονταν να συγκρίνουν την αντίληψη οπτικών μορφών με την αντίληψη μιας μελωδίας. Εκείνο που αναγνωρίζουμε σε μία μελωδία δεν είναι η αλληλουχία από συγκεκριμένες νότες αλλά η αμοιβαία σχέση τους. Μια μελωδία που εκτελείται σε διαφορετικούς τόνους εξακολουθεί να αναγνωρίζεται ως η ίδια μελωδία, γιατί η σχέση ανάμεσα στις νότες παραμένει η ίδια. Αντίστοιχα λοιπόν μπορούμε και αναγνωρίζουμε σύνθετες εικόνες ή αντικείμενα υπό διαφορετικές οπτικές συνθήκες, γιατί διατηρούνται οι σχέσεις μεταξύ των δομικών τους στοιχείων. Ο εγκέφαλος επιτυγχάνει την σύνθετη διαδικασία της αναγνώρισης κάνοντας ορισμένες υποθέσεις, οι οποίες φαίνεται ότι προέρχονται εν μέρει από την εμπειρία και εν μέρει από την έμφυτη συνδεσμολογία του συστήματος όρασης. Το οπτικό σύστημα οργανώνει τις αντιληπτικές εργασίες ακολουθώντας ορισμένους έμφυτους νόμους που διέπουν το σχήμα, το χρώμα, την απόσταση και την κίνηση των αντικειμένων στο οπτικό πεδίο. Η αντιληπτική οργάνωση είναι συνεχής και δυναμική, όπως αποδεικνύεται από σύνθετα πειράματα, αλλά και από γνωστές οπτικές απάτες. Στην γνωστή εικόνα που παρουσίασε πρώτος ο ψυχολόγος Edgar Rubin βλέπουμε είτε δύο σκούρα πρόσωπα σε προφίλ είτε ένα άσπρο βάζο σε σκούρο φόντο, αλλά είναι σχεδόν αδύνατον να



Σχήμα 2.2: *Τρεις κύριες παράλληλες οδοί για την οπτική αντίληψη αρχίζουν από τον έξω γονατώδη πυρήνα. Κάθε οδός έχει σχέση με ένα είδος οπτικής πληροφορίας (από [55])*

δούμε τις δύο εικόνες ταυτόχρονα (Σχήμα 2.1). Αυτή η στρατηγική ονομάζεται όλα για τον νικητή (WTA) και έχει χρησιμοποιηθεί ευρέως τόσο από καλλιτέχνες όσο και από ερευνητές στον χώρο των υπολογιστικών μοντέλων όρασης. Τα μάτια μας είναι συνηθισμένα να παρατηρούν συγκεκριμένα αντικείμενα και όταν συμβαίνει αυτό τα πάντα τριγύρω περιορίζονται σε φόντο. Το μάτι και ο νους του ανθρώπου δεν μπορούν να ασχολούνται ταυτόχρονα με δύο πράγματα. Πρέπει να υπάρχει μια γρήγορη και συνεχής μεταπήδηση από την μία περιοχή του οπτικού πεδίου στην άλλη, όπως υποστηρίζει ο Maurits Escher. Η διαφοροποίηση περιοχής ενδιαφέροντος και φόντου μας εισάγει σε μία βασική αρχή της οπτικής αντίληψης: μόνο κάποιο μέρος της εικόνας επιλέγεται ως εστία προσοχής, ενώ το υπόλοιπο "χάνεται" στο φόντο.

Στις σύγχρονες μελέτες της όρασης η ψυχολογία συναντά την νευροφυσιολογία και η απάντηση στο ερώτημα της δημιουργίας της οπτικής αντίληψης δίνεται μέσα από την προσεχτική μελέτη των ανατομικών οδών του εγκεφάλου με ψυχοφυσικά πειράματα που σχετίζονται με την επεξεργασία της μορφής, της κίνησης και του χρώματος από τρεις παράλληλες οδούς, οι οποίες εκτείνονται από τον αμφιβληστροειδή στον έξω γονατώδη πυρήνα και από εκεί στον εγκεφαλικό φλοιό. Στην συνέχεια, αφού περιγράψουμε συνοπτικά την φυσιολογία των επίμαχων συστημάτων του εγκεφάλου θα εστιάσουμε στην οπτική προσοχή, η οποία μπορεί να ενώσει την πληροφορία των παράλληλων οδών επεξεργασίας σε μία ενιαία συνειδητή εικόνα.

2.2.2 Ανατομία του ανθρώπινου οπτικού συστήματος

Οι νευρικοί άξονες των γαγγλιακών κυττάρων του αμφιβληστροειδούς σχηματίζουν το οπτικό νεύρο, το οποίο προβάλλει στον έξω γονατώδη πυρήνα του θαλάμου. Στη συνέχεια ο έξω γονατώδης πυρήνας προβάλλει στον ομόπλευρο πρωτοταγή οπτικό φλοιό ή V1, στο πεδίο Brodmann 17 που είναι γνωστό ως ταινιωτός φλοιός. Πέρα από τον ταινιωτό φλοιό βρίσκονται οι εξωταινιωτές περιοχές, οι οποίες περιέχουν 32 διαφορετικές προβολές της εικόνας του αμφιβληστροειδούς. Το σύνολο αυτών των περιοχών καταλαμβάνει περισσότερο από το μισό της συνολικής έκτασης του φλοιού! Οι περιοχές αυτές διαφέρουν ως προς την ευαισθησία των κυττάρων σε διαφορετικά χαρακτηριστικά της εισόδου. Π.χ. η περιοχή V5 (MT) έχει σχέση κυρίως με την κίνηση στο οπτικό πεδίο, ενώ η V4 έχει μεγαλύτερη σχέση με το χρώμα και με τον προσανατολισμό των ακμών. Επομένως κάθε περιοχή κωδικοποιεί ένα διαφορετικό

χαρακτηριστικό της οπτικής εισόδου.

Η πληροφορία για τις διαφορετικές αναπαραστάσεις του οπτικού ερεθίσματος φτάνει σε αυτές τις περιοχές μέσω τριών οδών. Γίνεται ένας πρώτος διαχωρισμός της πληροφορίας στον αμφιβληστροειδή χιτώνα από τα μεγάλα κύτταρα τύπου M που καταλήγουν στις μεγαλοκυτταρικές στοιβάδες και τα μικρά κύτταρα τύπου P που καταλήγουν στις μικροκυτταρικές στοιβάδες. Από τις δύο αυτές στοιβάδες ξεκινούν οι τρεις κύριες οδοί που εκτείνονται από τον έξω γονατώδη πυρήνα στην περιοχή V1 και από εκεί στην V2 και σε διάφορες άλλες περιοχές του εξωταινιωτού φλοιού (Σχήμα 2.2). Η πρώτη οδός έχει κυρίως σχέση με την αντίληψη των χρωμάτων και καταλήγει στον κάτω κροταφικό φλοιό, ο οποίος έχει σχέση με την αντίληψη των χρωμάτων και της μορφής. Η δεύτερη οδός έχει κυρίως σχέση με την αντίληψη των σχημάτων και καταλήγει επίσης στον κατωκροταφικό φλοιό. Οι νευρώνες αυτού του συστήματος είναι ευαίσθητοι στο περίγραμμα και στον προσανατολισμό των εικόνων και έχουν μεγάλη διακριτική ικανότητα, η οποία είναι σημαντική για να βλέπουμε τις λεπτομέρειες των ακίνητων αντικειμένων. Η τρίτη οδός είναι εξειδικευμένη στον έλεγχο της κίνησης και των χωρικών σχέσεων και συμβάλλει στην αντίληψη του βάθους. Η οδός αυτή έχει σχέση με το που βρίσκονται τα αντικείμενα παρά με το τι είναι. Οι νευρώνες αυτού του συστήματος δεν είναι πολύ ευαίσθητοι στα χρώματα και έχουν μικρή ικανότητα ανάλυσης ακίνητων αντικειμένων. Οι τρεις εξειδικευμένες οδοί αλληλεπιδρούν σε πολλά επίπεδα (Σχήμα 2.3).

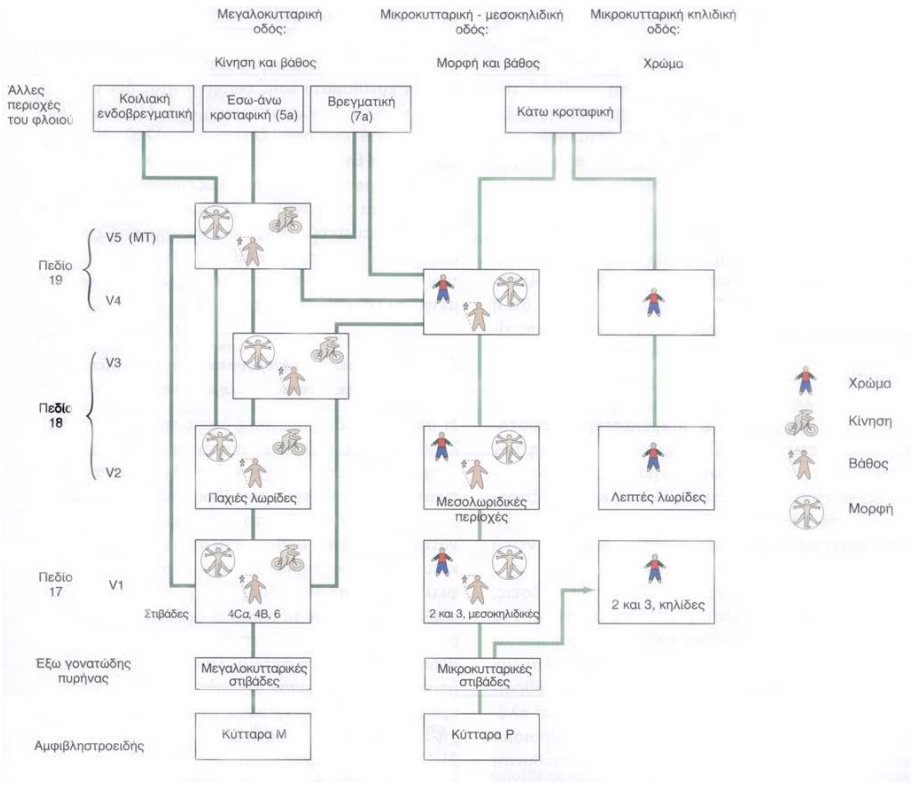
Αν και υπάρχει συμφωνία απόψεων στο ότι το οπτικό σύστημα χρησιμοποιεί παράλληλη επεξεργασία, οι ερευνητές διαφωνούν ως προς το πόσο ξεκάθαρη είναι η κατανομή των λειτουργιών στις τρεις οδούς (υπάρχει ακόμη και διαφωνία ως προς τον αριθμό τους). Η κύρια διαφωνία βρίσκεται στην έκταση της αλληλεπίδρασης των οδών. Ένας αριθμός ερευνητών υποστηρίζει ότι ανεξάρτητα του πόσο εξειδικευμένη είναι μία οδός σε συγκεκριμένη επεξεργασία της οπτικής εισόδου, οι άλλες οδοί συμμετέχουν ενεργά σε αυτήν την επεξεργασία.

2.2.3 Οπτική προσοχή και συντονισμός οπτικών οδών

Η οπτική αντίληψη δημιουργείται, όπως είδαμε, από την πληροφορία διαφορετικών παράλληλων οδών, οι οποίες επεξεργάζονται διαφορετικά στοιχεία (χρώμα, κίνηση, βάθος κτλ). Για να συνδυαστούν οι οδοί πρέπει να συνδεθούν παροδικά ανεξάρτητες ομάδες κυττάρων με διαφορετικές λειτουργίες. Ο εγκέφαλος επομένως χρειάζεται έναν μηχανισμό σύνδεσης των διαφορετικών στοιχείων που διαθέτει για να κατανοήσει αυτό που βλέπει. Αν και ο μηχανισμός αυτός παραμένει ακόμη αδιευκρίνιστος, αποτελεί μάλλον μέρος του μηχανισμού οπτικής προσοχής.

Η Ann Treisman, καθώς και ο Bella Julesz εργαζόμενοι ανεξάρτητα, απέδειξαν ότι ο σχηματισμός αυτών των συνδέσεων απαιτεί κάποιο είδος προσοχής. Προσπάθησαν να κατανοήσουν τον τρόπο εστίασης της προσοχής σε ένα αντικείμενο, όπως και το ποια χαρακτηριστικά του αντικειμένου το κάνουν να ξεχωρίζει από το περιβάλλον του. Διαπίστωσαν τελικά ότι στοιχειώδη χαρακτηριστικά, όπως η φωτεινότητα, το χρώμα και η κατεύθυνση των γραμμών δημιουργούν σαφή όρια μεταξύ αντικειμένων στο οπτικό πεδίο (Σχήμα 2.4). Παρατήρησαν επίσης ότι όταν τα όρια αυτά αποτελούνται από στοιχεία τα οποία διαφέρουν σαφώς από την γειτονιά τους, τα όρια προβάλλουν σχεδόν αυτόματα μέσα σε 50ms.

Τα συμπεράσματα αυτά ήταν σύμφωνα με τις πρώιμες παρατηρήσεις του James [51] και οδήγησαν τους δύο ερευνητές στην θεωρία της οπτικής προσοχής δύο επιπέδων:

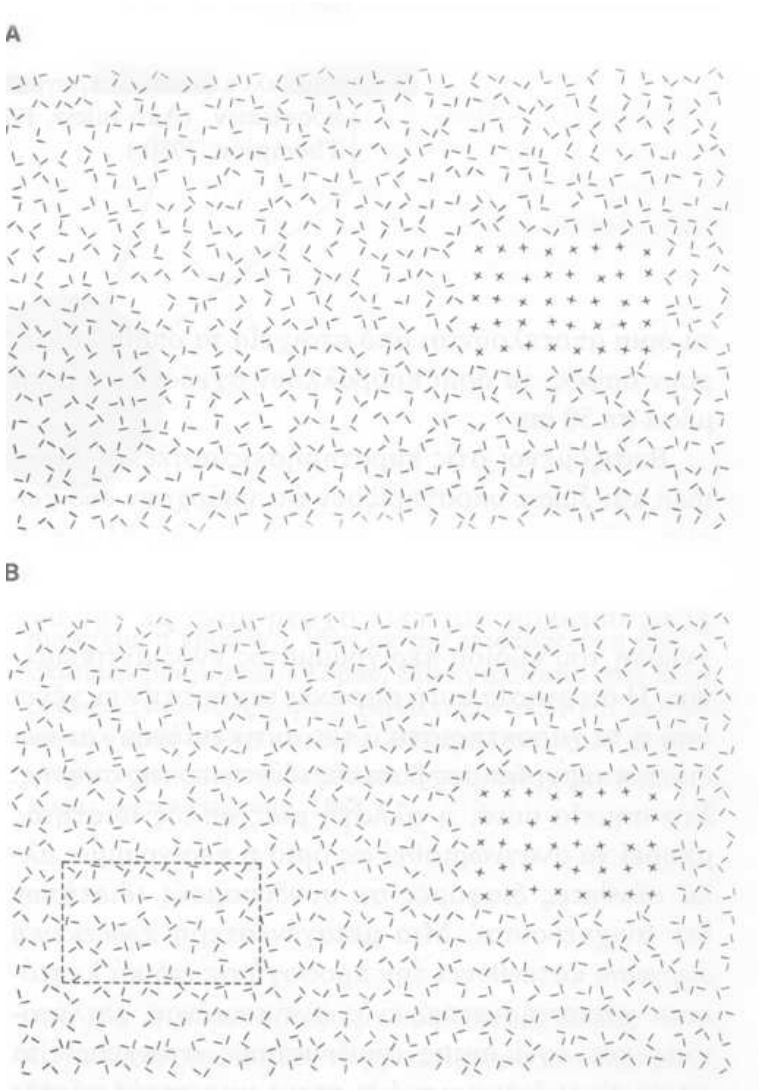


Σχήμα 2.3: *Καθεμία από τις τρεις κύριες παράλληλες οδούς του οπτικού συστήματος μεταφέρει ένα είδος οπτικής πληροφορίας. Η οργάνωση της εικόνας βασίζεται σε μελέτες που έγιναν σε πίθηκο μακάκο (από [55]).*

της προ-προσεχτικής και της προσεχτικής. Στο πρώτο επίπεδο η διαδικασία είναι ταχύτατη και σαρώνει τα χαρακτηριστικά του αντικειμένου και κωδικοποιεί τις βασικές ιδιότητες της σκηνής. Η αλλαγή μίας απλής ιδιότητας μπορεί σχεδόν αυτόματα να αναγνωριστεί σαν όριο ή περίγραμμα, αλλά σύνθετες διαφορές σε συνδυασμούς ιδιοτήτων δεν ανιχνεύονται (Σχήμα 2.5). Η προσεχτική διεργασία που ακολουθεί επικεντρώνει την προσοχή σε λεπτομέρειες του αντικειμένου επιλέγοντας και τονίζοντας συνδυασμούς στοιχείων τα οποία ξεχωρίζουν στον αντίστοιχο χάρτη. Η προσεχτική διεργασία κατευθύνεται κυρίως από την εμπειρία και την γνώση μας για τα οπτικά ερεθίσματα που μας περιβάλλουν. Πολλοί ερευνητές του χώρου αποδέχτηκαν και υποστήριξαν αυτό το μοντέλο [135, 131, 4, 80, 12, 11, 43, 51].

2.2.4 Ανοδική οπτική προσοχή

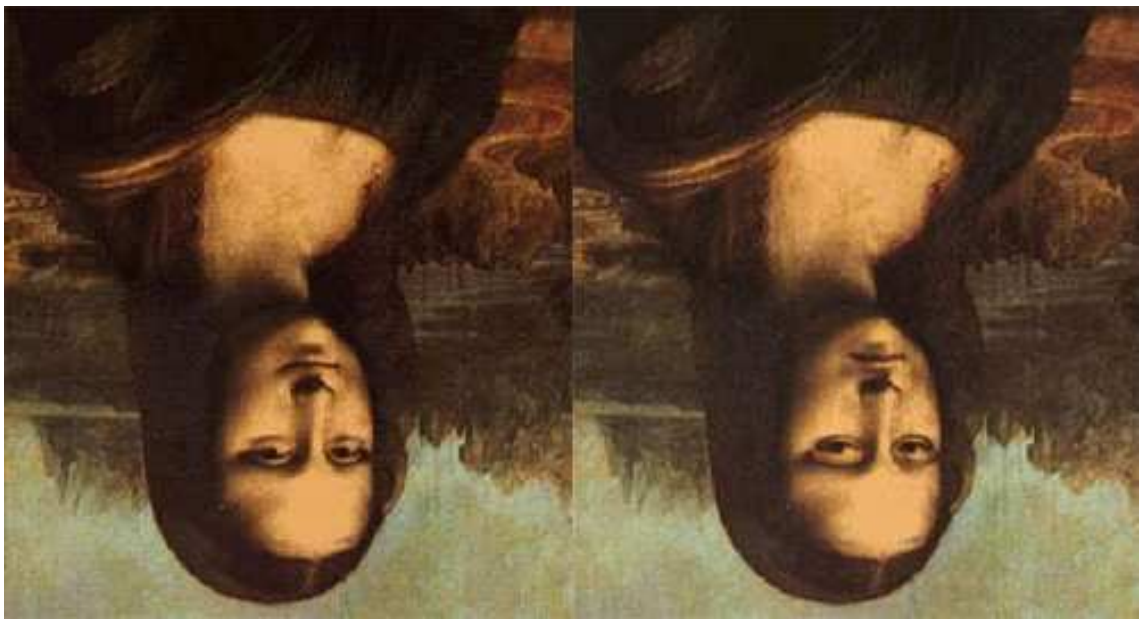
Η Treisman υποστήριξε ότι η προ-προσεχτικής διαδικασίας αποτελεί μία ανοδική[3] διαδικασία στην διάρκεια της οποίας οι διαφορετικές ιδιότητες κωδικοποιούνται σε διαφορετικούς χάρτες χαρακτηριστικών. Για να λύσει το πρόβλημα της σύνδεσης υπέθεσε επίσης ότι μάλλον υπάρχει ένας κύριος χάρτης που περιέχει τελικά τις πιο σημαντικές περιοχές της εικόνας. Πρόκειται για έναν διδιάστατο χάρτη, ο οποίος δέχεται πληροφορίες από όλους τους χάρτες χαρακτηριστικών και συγκρατεί μόνο εκείνες τις περιοχές που διαφοροποιούν το αντικείμενο προσοχής από το περιβάλλον του. Από την στιγμή που υπάρχει μία κεντρική αντιπροσώπευση των κύριων χαρακτηριστικών των αντικειμένων, οι επιμέρους λεπτομέρειες τους μπορούν να ανακτηθούν με αναφορά στους αντίστοιχους χάρτες χαρακτηριστικών που έχουν συνεισφέρει περισσότερο.



Σχήμα 2.4: *Ορισμένες αντιλήψεις παράγονται με προ-προσεχτική σάρωση, ενώ άλλες απαιτούν εστιακή προσοχή. Στο Α είναι εύκολο να διαφοροποιήσουμε άμεσα τη μικρή περιοχή που αποτελείται από σταυρούς. Η εικόνα περιλαμβάνει όμως και μία περιοχή που αποτελείται από Τ. Για να την βρούμε πρέπει να εστιάσουμε την προσοχή μας σε κάθε περιοχή της εικόνας (η περιοχή των Τ είναι περιγεγραμμένη στο Β) (από [4])*

Ο ανταγωνισμός μεταξύ των περιοχών του κυρίου χάρτη που θα οδηγήσει στην επιλογή ενός νικητή, ο οποίος θα αποτελέσει αυτόματα τον επόμενο στόχο της προσοχής μας, γίνεται με την στρατηγική του όλα-για-τον-νικητή. Σύμφωνα με αυτήν την στρατηγική όταν σε κάποια δεδομένη στιγμή ορισμένες περιοχές τονίζονται και τραβούν την προσοχή, οι υπόλοιπες αγνοούνται. Αφού γίνει η επεξεργασία αυτών των περιοχών, το οπτικό μας σύστημα επικεντρώνεται στην αμέσως λιγότερο σημαντική περιοχή κοκ. Ο κύριος χάρτης επομένως λαμβάνει είσοδο από τα πρώιμα στάδια οπτικής επεξεργασίας (εξαγωγή χαρακτηριστικών) και δημιουργεί ένα μονοπάτι για τις διαδοχικές εστιάσεις με σειρά μειούμενου ενδιαφέροντος.

Το υποθετικό μοντέλο που προέκυψε από τα πειράματα θεωρεί ότι ορισμένες βασικές ιδιότητες της εικόνας (π.χ. ένταση φωτεινότητας, χρώμα, προσανατολισμός, μέγεθος, απόσταση κτλ.) κωδικοποιούνται σε χωριστές παράλληλες οδούς, καθεμία από τις οποίες δημιουργεί τον αντίστοιχο χάρτη χαρακτηριστικών. Στην συνέχεια,



Σχήμα 2.5: *Οι δύο φωτογραφίες φαίνονται αρχικά ίδιες. Όταν τις δούμε τοποθετημένες σωστά, αποκαλύπτονται οι πραγματικές λεπτομέρειες των δύο προσώπων (από [55])*

επιλεγμένες περιοχές των χαρτών επεξεργάζονται και ανταγωνίζονται μεταξύ τους, ώστε τελικά να συνδεθούν και να δημιουργήσουν τον κύριο χάρτη[18]. Υπάρχουν τρεις διαφορετικοί μηχανισμοί για την σύνδεση των διαφορετικών χαρτών: (α) η επιλογή περιοχών των χαρτών με χρήση κάποιου χωρικού παραθύρου, (β) η καταστολή των περιοχών που δεν περιέχουν επιθυμητά χαρακτηριστικά και (γ) η συνειδητή ενεργοποίηση περιοχών που περιέχουν ένα γνωστό αντικείμενο.

Το μοντέλο σύνδεσης χαρακτηριστικών έχει χρησιμοποιηθεί ευρέως και έχει επιβεβαιωθεί πειραματικά από πολλές έρευνες της δημιουργού του, αλλά και πολλών ερευνητών του χώρου [23, 149, 135, 130, 131, 132, 133]. Αν και τις τελευταίες δεκαετίες το θεωρητικό μοντέλο έχει δεχτεί αλλαγές και προσθήκες, όπως π.χ. η παραλληλοποίηση κάποιων διαδικασιών, συνεχίζει να θέτει ένα γενικό πλαίσιο για την κατανόηση του συστήματος της οπτικής προσοχής. Ακολουθώντας την θεωρία αυτή ένας μεγάλος αριθμός υπολογιστικών μοντέλων έχει αναπτυχθεί στην ερευνητική κοινότητα. Οι κύριες διαφορές εντοπίζονται στους διαφορετικούς τρόπους δημιουργίας και σύνδεσης των επιμέρους χαρτών και στον τρόπο ελέγχου των μηχανισμών της μετατόπισης της οπτικής προσοχής από ένα σημείο ενδιαφέροντος στο άλλο.

Αν και το μοντέλο του κύριου χάρτη είναι αποδεκτό από πολλούς ερευνητές του χώρου, δεν αποτελεί το μοναδικό για την επεξήγηση του μηχανισμού εστίασης της προσοχής. Οι Desimone και Duncan υποστηρίζουν ότι δεν υπάρχει ξεκάθαρο βιολογικό ανάλογο του κύριου χάρτη και ότι η επικέντρωση της προσοχής γίνεται αποκλειστικά μετά από συνειδητή εστίαση σε περιοχές των χαρτών χαρακτηριστικών που σχετίζονται με το αντικείμενο που ψάχνουμε [23]. Το μοντέλο αυτό έχει ομοιότητες με αυτό της καθοδηγούμενης αναζήτησης (Guided Search) του Wolfe [149] και αποτελείται από δύο διαδικασίες: α) την προ-προσεχτική που στηρίζεται στον συνδυασμό ασυνείδητων και συνειδητών εντολών για τον υπολογισμό των χαρακτηριστικών και β) την διαδικασία ενεργοποίησης που κατευθύνει την οπτική προσοχή σε διαφορετικές περιοχές ενδιαφέροντος με σειριακό τρόπο. Το βασικό μειονέκτημα που έχουν αυτά τα μοντέλα σε πρακτικές υπολογιστικές εφαρμογές είναι το ότι πρέπει να

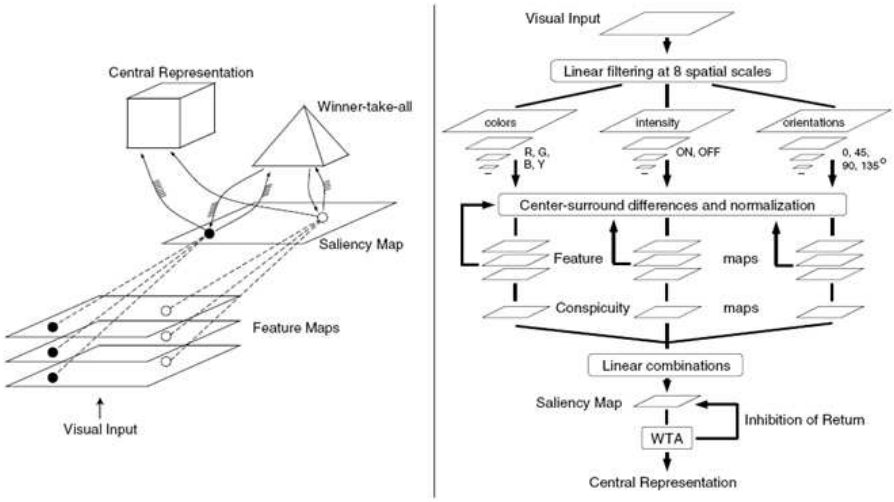
είναι γνωστό εκ των προτέρων το αντικείμενο ή η περιοχή που αναζητούμε.

2.3 Υπολογιστικά μοντέλα οπτικής προσοχής

2.3.1 Επισκόπηση

Ο δρόμος για την ανάπτυξη υπολογιστικών μοντέλων οπτικής προσοχής άνοιξε με την θεωρία των Treisman and Gelade [135], οι οποίοι υποστήριξαν μεταξύ άλλων ότι μόνο πολύ απλά οπτικά χαρακτηριστικά εξάγονται από την είσοδο και υπολογίζονται με παράλληλο τρόπο κατά την προ-προσεχτική διαδικασία. Η οπτική προσοχή είναι μετά απαραίτητη για να συνδέσει αυτά τα χαρακτηριστικά με αναπαραστάσεις συγκεκριμένων αντικειμένων για περαιτέρω επεξεργασία. Η ιδέα αυτή επεκτάθηκε και αναπτύχθηκε στην συνέχεια με την χρήση νευρωνικής αρχιτεκτονικής από τους Koch και Ullman [60] και αργότερα από τους Itti *et al.* [48, 49]. Τα μοντέλα αυτά βασίζονται στην θεωρία του κύριου χάρτη σύμφωνα με την οποία, η εικόνα αναλύεται αρχικά σε ένα σύνολο χαρτών χαρακτηριστικών, οι οποίοι επεξεργάζονται με τελεστές εμπνευσμένους από την λειτουργία του ανθρώπινου εγκεφάλου. Οι τελεστές αποσκοπούν στον εντοπισμό σημαντικών περιοχών που ξεχωρίζουν από τη γειτονιά τους και είναι άξιες προσοχής. Η συνολική επεξεργασία είναι αυστηρά μη συνειδητή και καταλήγει σε ένα κύριο χάρτη. Ένα όλα-για-το-νικητή[17] δίκτυο αναλαμβάνει την σειριακή, ιεραρχική (πιο ενεργή → λιγότερο ενεργή) επιλογή των σημαντικών περιοχών της εισόδου. Το Σχήμα 2.6 αναπαριστά οπτικά τις δύο αρχιτεκτονικές. Η αρχιτεκτονική των Itti *et al.* θα αναλυθεί εκτενώς στην Ενότητα 2.3.2.

Πολλά υπολογιστικά μοντέλα οπτικής προσοχής αξιοποιούν την ιδέα του κύριου χάρτη. Οι διαφορές εντοπίζονται στην στρατηγική που ακολουθείται για την εξαγωγή των προ-προσεχτικών χαρακτηριστικών και στην σύνδεση τους. Η στρατηγική του Wolfe [149], όπως είδαμε στην Ενότητα 2.2.4, βασίζεται στην χρήση γνώσης για την συνειδητή εστίαση της προσοχής, η οποία θα διαφοροποιήσει τον επιθυμητό στόχο από το περιβάλλον του: π.χ. αν ψάχνουμε για ένα κόκκινο αντικείμενο τότε πρέπει να τονιστεί η συνεισφορά του χρώματος στον κύριο χάρτη και να υποβαθμιστεί η συνεισφορά της κατευθυντικότητας. Το μοντέλο FeatureGate των Cave *et al.* παρέχει μία ολοκληρωμένη υλοποίηση ενός συστήματος που συνδυάζει την προηγούμενη ιδέα με μη συνειδητούς μηχανισμούς οπτικής προσοχής, έτσι ώστε ένας στόχος να ελκύει την προσοχή βάσει συνειδητής και μη συνειδητής αναζήτησης [15]. Διατηρώντας αυτήν τη λογική, οι Rao *et al.* [97] πρότειναν τον υπολογισμό της σημαντικότητας[18] με την εύρεση της ευκλιδειας απόστασης μεταξύ ενός διανύσματος χαρακτηριστικών του επιθυμητού στόχου (π.χ. χρώμα, κατευθυντικότητα κτλ) και όλων των αντίστοιχων διανυσμάτων σε κάθε σημείο της οπτικής εισόδου. Επεκτείνοντας αυτές τις ιδέες κάτω από ένα Bayesian πλαίσιο, οι Torralba *et al.* [127, 128] υποστηρίζουν ότι μία μη λεπτομερής ολική ανάλυση της σκηνής παρέχει χρήσιμες πληροφορίες για την αναγνώριση αντικειμένων ή περιοχών: π.χ. αν ψάχνουμε για αυτοκίνητα και μπορούμε γρήγορα να προσδιορίσουμε την περιοχή της ασφάλτου στην εικόνα, τότε μπορούμε να περιορίσουμε σημαντικά την αναζήτηση μας. Οι Hamker *et al.* [38, 40] προτείνουν επίσης ένα υπολογιστικό μοντέλο που συνδυάζει την μη συνειδητή ανάλυση με την αναζήτηση συγκεκριμένου στόχου. Η κύρια διαφορά με το πρότυπο μοντέλο των Itti *et al.* είναι ότι τα διαφορετικά στάδια εξαγωγής και επεξεργασίας των χαρακτηριστικών αλληλεπιδρούν μεταξύ τους συνεχώς πριν καταλήξουν στο τελικό



Σχήμα 2.6: *α) Η αρχιτεκτονική των Koch και Ullman για τον υπολογισμό της οπτικής προσοχής; β) Η αρχιτεκτονική των Itti et al., η οποία βασίζεται σε αυτή των Koch και Ullman. (από [60])*

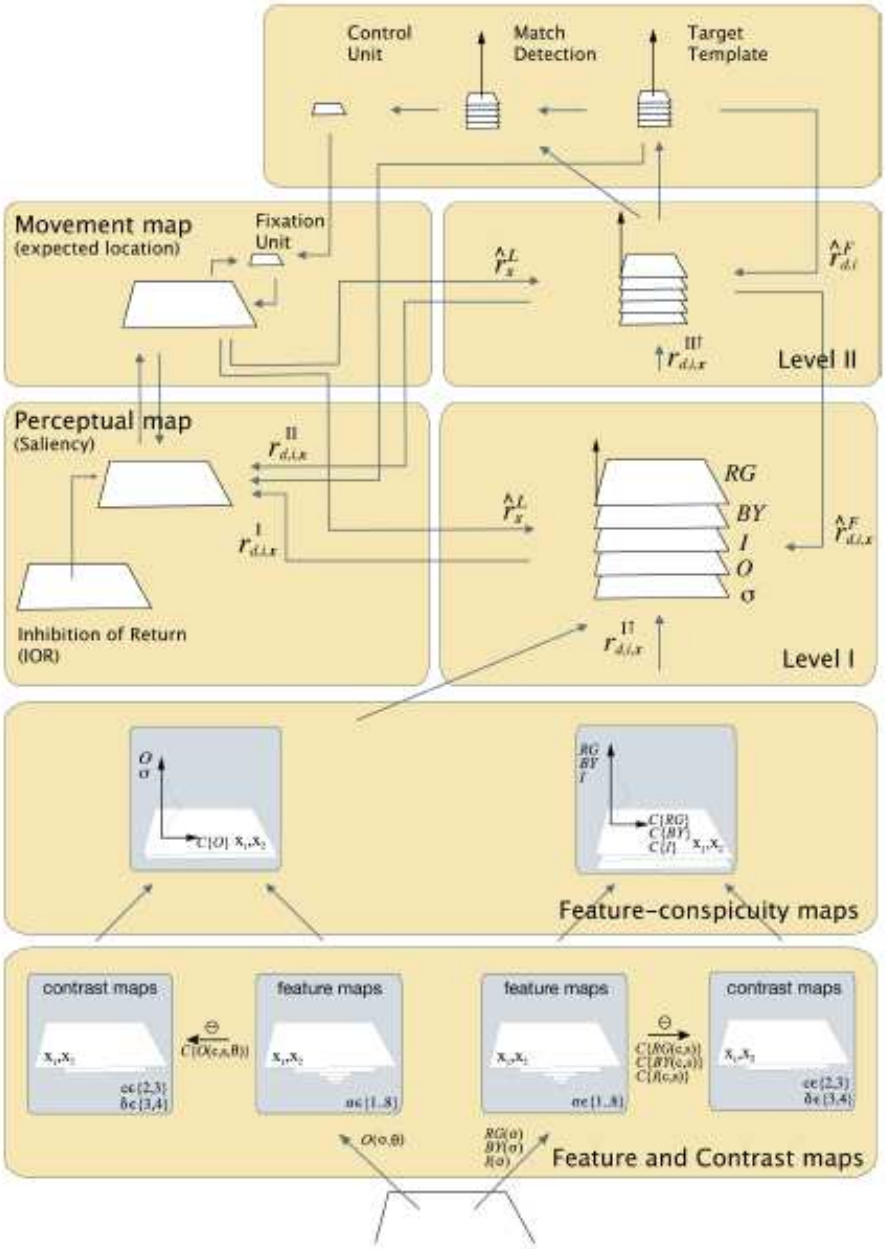
στάδιο συνδυασμού τους για την δημιουργία του κύριου χάρτη.

Μελετώντας μια αρχιτεκτονική με εκτενέστερες αλληλεπιδράσεις μεταξύ του κυρίου χάρτη και της ανάλυσης χαμηλού επιπέδου οι Tsotsos *et al.* [137] πρότειναν ένα μοντέλο οπτικής προσοχής, το οποίο αποτελείται από ένα συνδυασμό εμπροσθόδοτης[20] ανοδικής εξαγωγής χαρακτηριστικών και έναν μηχανισμό επιλεκτικού συντονισμού[19] αυτών των χαρακτηριστικών με ανάδραση. Η επεξεργασία είναι ιεραρχική και μετατρέπει την εικόνα εισόδου σε αναπαραστάσεις μειούμενης χωρικής ανάλυσης. Κάθε στοιχείο ενός επιπέδου της πυραμίδας που δημιουργείται συνδέεται με τα αντίστοιχα στοιχεία του πάνω και κάτω επιπέδου. Σε αυτό το μοντέλο, ο στόχος της οπτικής προσοχής επιλέγεται στην κορυφή της ιεραρχικής επεξεργασίας (κάτι ανάλογο με τον κυρίου χάρτη) και βασίζεται σε εμπροσθόδοτη ενεργοποίηση και πιθανή πόλωση μέσω γνώσης συγκεκριμένων στοιχείων της εισόδου. Μετά από την επιλογή της πιο σημαντικής τοποθεσίας ακολουθεί ένα στάδιο ανάδρασης που στοχεύει στην μεγαλύτερη ανάδειξη του στόχου σε σχέση με άλλα λιγότερο σημαντικά οπτικά ερεθίσματα. Σε κάθε επίπεδο απενεργοποιούνται εκείνα τα στοιχεία που δεν συνεισφέρουν ενεργά στην ανάδειξη του επιλεγμένου στόχου.

2.3.2 Μοντέλο κύριου χάρτη

Ο βασικός εκπρόσωπος της κατηγορίας μοντέλων οπτικής προσοχής που καταλήγουν σε κύριο χάρτη είναι το μοντέλο των Itti και Koch [48], το οποίο περιορίζεται στον μη συνειδητό ή ανοδικό[3] έλεγχο της εστίασης ενδιαφέροντος. Συνεπώς το βασικό μοντέλο καταλήγει στον εντοπισμό του οπτικού ερεθίσματος που είναι σημαντικό για το ανθρώπινο μάτι και όχι στην σημασιολογική αναγνώριση του. Οι κύριες υποθέσεις είναι οι εξής:

- Αρχικά, το οπτικό ερέθισμα αναλύεται σε οπτικούς χάρτες που αντιστοιχούν σε διαφορετικά χαρακτηριστικά (π.χ. φωτεινότητα, χρώμα, κατευθυντικότητα κτλ). Αυτοί οι χάρτες επεξεργάζονται με χρήση φίλτρων κέντρου-περιφέρειας ώστε να παραμείνουν ενεργές μόνο οι περιοχές που ξεχωρίζουν από το υπόβαθρο.



Σχήμα 2.7: *Το υπολογιστικό μοντέλο οπτικής προσοχής του Hamker. Η σημαντικότερη διαφορά με το μοντέλο των Itti et al. προσδιορίζεται στα επίπεδα I και II, όπου υπάρχει συνεχής αλληλεπίδραση μεταξύ των διαφορετικών χαρτών χαρακτηριστικών. (από [38])*

- Η πληροφορία των ενδιάμεσων χαρτών[5] ενδιαφέροντος συνδυάζεται σε ένα τελικό χάρτη ενδιαφέροντος ή κύριο χάρτη.

- Η μέγιστη τιμή του χάρτη ενδιαφέροντος είναι η πιο σημαντική τοποθεσία για μία συγκεκριμένη χρονική στιγμή και από αυτήν ξεκινούν οι διαδοχικές εστιάσεις του μοντέλου.

Στην περιγραφή του μοντέλου των Itti *et al.* που ακολουθεί διατηρήθηκε ο συμβολισμός του πρωτότυπου κειμένου [48].

2.3.3 Εξαγωγή χαρακτηριστικών

Το πρώτο στάδιο επεξεργασίας αφορά στην εξαγωγή χαρακτηριστικών χαμηλού επιπέδου από την αρχική έγχρωμη εικόνα. Στο βασικό μοντέλο των Itti et al χρησιμοποιούνται τέσσερα χρωματικά κανάλια (κόκκινο, πράσινο, μπλε, κίτρινο) και δύο κανάλια που κωδικοποιούν πληροφορία για κατευθυντικότητα και φωτεινότητα της εικόνας εισόδου. Τα χαρακτηριστικά αυτά εξάγονται σε πολλαπλές χωρικές κλίμακες με την χρήση Gaussian πυραμίδων [14], οι οποίες προκύπτουν με προοδευτική εξομάλυνση και υποδειγματοληψία της αρχικής εικόνας. Στην υλοποίηση τους οι Itti et al χρησιμοποίησαν πυραμίδες με βάθος εννέα επιπέδων με το πρώτο επίπεδο (επίπεδο 0) να αντιστοιχεί στην αρχική εικόνα και το τελευταίο (επίπεδο 8) να αντιστοιχεί σε 1:256 του μεγέθους εισόδου. Κάθε επίπεδο διαφέρει χωρικά από το προηγούμενο του κατά μία δύναμη του δύο ως προς τις δύο διαστάσεις.

Στη συνέχεια κάθε χαρακτηριστικό τροφοδοτεί την διαδικασία επεξεργασίας με φίλτρα κέντρου-περιφέρειας[21], τα οποία έχουν σαν βιολογικό ανάλογο στο σύστημα ανθρώπινης όρασης τα γαγγλιακά κύτταρα. Το υποδεκτικό πεδιο αυτών των κυττάρων αποκρίνεται έντονα στην αντίθεση φωτισμού και διαιρείται σε ένα κέντρο και μία περιφέρεια. Τα κύτταρα φωτεινού κέντρου διεγείρονται από φως το οποίο προσπίπτει στο κέντρο του υποδεκτικού τους πεδίου και υφίστανται αναστολή όταν το φως προσπίπτει στην περιφέρεια. Τα κύτταρα σκοτεινού κέντρου αποκρίνονται αντίστροφα. Αυτού του είδους η επεξεργασία καθιστά το σύστημα ευαίσθητο σε τοπικές χωρικές αντιθέσεις και όχι στην απόλυτη τιμή του εκάστοτε χαρακτηριστικού στον αντίστοιχο χάρτη. Η επεξεργασία κέντρου-περιφέρειας υλοποιείται στο μοντέλο σαν την διαφορά μεταξύ ενός αδρομερούς και λεπτομερούς επιπέδου στην πυραμίδα του κάθε χαρακτηριστικού: το κέντρο του υποδεκτικού πεδίου αντιστοιχεί σε ένα εικονοστοιχείο ενός λεπτομερούς επιπέδου c της πυραμίδας και η περιφέρεια στο αντίστοιχο εικονοστοιχείο ενός αδρομερούς επιπέδου s. Επομένως δημιουργούνται έξι χάρτες για κάθε είδος χαρακτηριστικού (στα επίπεδα $2-5, 2-6, 3-6, 3-7, 4-7, 4-8$). Στο μοντέλο χρησιμοποιούνται επτά είδη χαρακτηριστικών που υποστηρίζονται από βιολογικά και νευροφυσιολογικά πειράματα: ένα για την κωδικοποίηση αντίθεσης φωτεινότητας [65, 30], δύο για τα κανάλια χρωματικής αντίθεσης κόκκινο:πράσινο και μπλε:κίτρινο [70] και τέσσερα για την κωδικοποίηση της αντίθεσης σε τοπική κατευθυντικότητα [22, 129].

Οι έξι χάρτες που προκύπτουν από την ένταση της εισόδου κωδικοποιούν την αντίθεσης φωτεινότητας, η οποία υπολογίζεται σαν την απόλυτη τιμή της διαφοράς μεταξύ της έντασης του pixel στο κέντρο (σε ένα από τα τρία επίπεδα του c) και της έντασης στην περιφέρεια (σε ένα από τα έξι επίπεδα του s). Η απόλυτη τιμή εξασφαλίζει ότι τόσο οι σκοτεινές περιοχές σε φωτεινό υπόβαθρο, όσο και οι φωτεινές σε σκοτεινό υπόβαθρο θα αναδειχθούν. Δημιουργείται επομένως ένα σύνολο έξι χαρτών ως εξής:

$$I(c, s) = |I(c) - I(s)| \tag{2.1}$$

Ένα δεύτερο σύνολο χαρτών δημιουργείται για τα χρωματικά κανάλια ακολουθώντας παρόμοια διαδικασία και μετασχηματίζοντας τα σύμφωνα με την θεωρία διπλής χρωματικής αντίθεσης[22]. Κάθε ένας από τους 6 χάρτες κόκκινου:πράσινου και μπλε:κίτρινου υπολογίζονται ως

$$RG(c, s) = |(R(c) - G(c)) + (G(s) - R(s))| \tag{2.2}$$

$$RG(c, s) = |(B(c) - Y(c)) + (Y(s) - B(s))| \tag{2.3}$$

Η τοπική κατευθυντικότητα υπολογίζεται με την δημιουργία πυραμίδων Gabor από την μονόχρωμη αρχική εικόνα [36]. Τα Gabor φίλτρα συντονίζονται σε τέσσερις διαφορετικές κατευθύνσεις $(0^o, 45^o, 90^o, 135^o)$ και οι αντίστοιχοι χάρτες χαρακτηριστικών υπολογίζονται όπως και πριν:

$$O(c, s, \theta) = |O(c, \theta) - O(s, \theta)| \tag{2.4}$$

Συνολικά, υπολογίζονται 42 χάρτες χαρακτηριστικών: 6 για την φωτεινότητα, 12 για το χρώμα και 24 για την κατευθυντικότητα.

2.3.4 Δημιουργία κύριου χάρτη

Ο σκοπός του κύριου χάρτη είναι να αναπαραστήσει την σημασία κάθε περιοχής του οπτικού πεδίου με μία τιμή και να κατευθύνει την ιεραρχική επιλογή των περιοχών. Ο κύριος χάρτης θα δημιουργηθεί από τον συνδυασμό των ενδιάμεσων χαρτών που έχουν προκύψει από το προηγούμενο στάδιο. Αυτός ο συνδυασμός όμως δεν είναι εύκολος καθώς οι χάρτες έχουν διαφορετικό δυναμικό εύρος και προέρχονται από ποικίλους μηχανισμούς εξαγωγής. Ένα άλλο σημαντικό πρόβλημα είναι το ότι κάποιες σημαντικές περιοχές της εισόδου θα ενεργοποιούν μικρό αριθμό χαρτών, οπότε είναι πιθανό να καλυφθούν από θόρυβο ή από λιγότερο σημαντικές περιοχές που έχουν μικρότερη ενεργοποίηση αλλά συνεισφέρουν σε περισσότερους χάρτες.

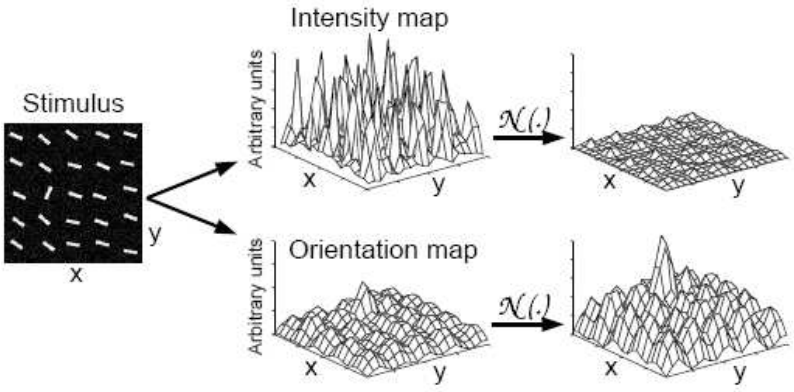
Οι Itti και Koch προτείνουν έναν απλό τελεστή κανονικοποίησης των χαρτών, $N(\cdot)$, ο οποίος αναδεικνύει εκείνους του χάρτες που έχουν λίγες αλλά σημαντικές περιοχές και υποβιβάζει τους υπόλοιπους που έχουν πολλές εξίσου σημαντικές περιοχές. Ο τελεστής N αποτελείται από τα εξής βήματα: 1) Κανονικοποίηση των τιμών του χάρτη σε συγκεκριμένο έυρος $[0...M]$, ώστε να υποβαθμιστούν οι διαφορές που οφείλονται στο διαφορετικό δυναμικό εύρος, 2) εύρεση του συνολικού μέγιστου M του χάρτη και υπολογισμός της μέσης τιμής των υπόλοιπων τοπικών μεγίστων $\overline{m}$, 3) πολλαπλασιασμός του χάρτη με την τιμή $(M - \overline{m})^2$. Ο τελεστής αυτός παρέχει έναν απλό τρόπο σύγκρισης της πιο ενεργής περιοχής του χάρτη με τις υπόλοιπες. Όταν η διαφορά αυτή είναι μεγάλη ο συγκεκριμένος χάρτης αποκτά μεγαλύτερη βαρύτητα, ενώ στην αντίθετη περίπτωση, στην οποία δεν υπάρχει κάποια σημαντική προεξέχουσα περιοχή, ο χάρτης υποβαθμίζεται. Σύμφωνα με τους συγγραφείς ο τελεστής βασίζεται σε ανάλογο βιολογικό υποσύστημα του εγκεφάλου. Οι ενδιάμεσοι χάρτες φωτεινότητας και χρώματος προκύπτουν ως εξής:

$$\overline{I} = \bigotimes_{c=2}^{\sigma_1} \bigotimes_{s=c+1}^{\sigma_1} N(I(c, s)) \tag{2.5}$$

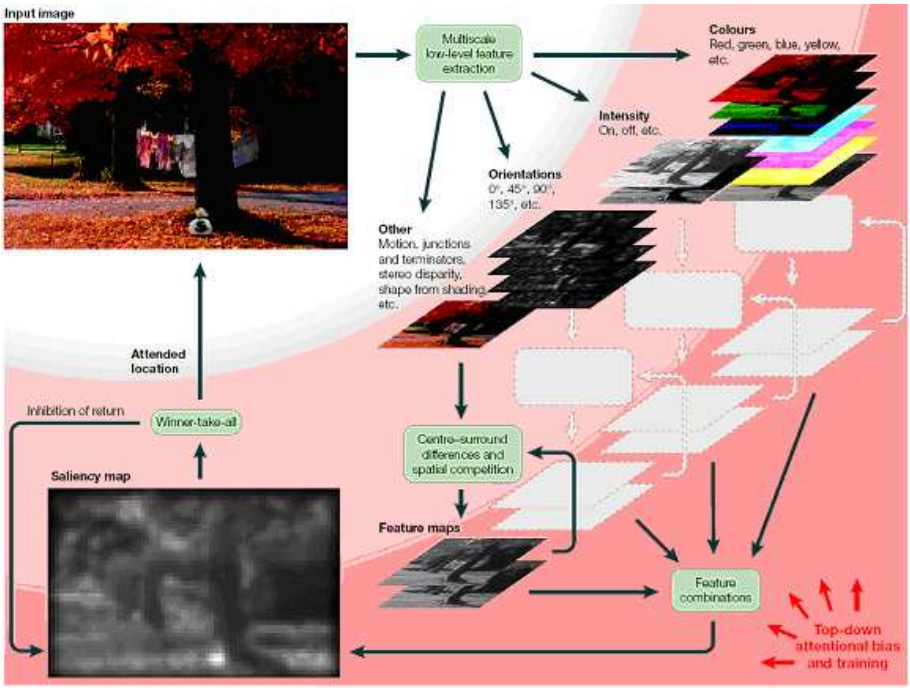
$$\overline{C} = \bigotimes_{c=2}^{\sigma_1} \bigotimes_{s=c+1}^{\sigma_1} [N(RG(c, s)) + N(BY(c, s)) \tag{2.6}$$

Για την κατευθυντικότητα δημιουργούνται αρχικά 4 ενδιάμεσοι χάρτες (ένας για κάθε θ), οι οποίοι μετά αθροίζονται κατά κλίμακα για να παράγουν τον ενδιάμεσο χάρτη κατευθυντικότητας:

$$\overline{O} = \sum_{\theta in A} \bigotimes_{c=2}^{\sigma_1} \bigotimes_{s=c+1}^{\sigma_1} N(O(c, s, \theta)) \tag{2.7}$$



Σχήμα 2.8: *Η ενδεικτική λειτουργία του τελεστή κανονικοποίησης N και το αποτέλεσμα για μια περιοχή που διαφέρει τοπικά από τις γύρω της στο κανάλι κατευθυντικότητας. (από [48])*



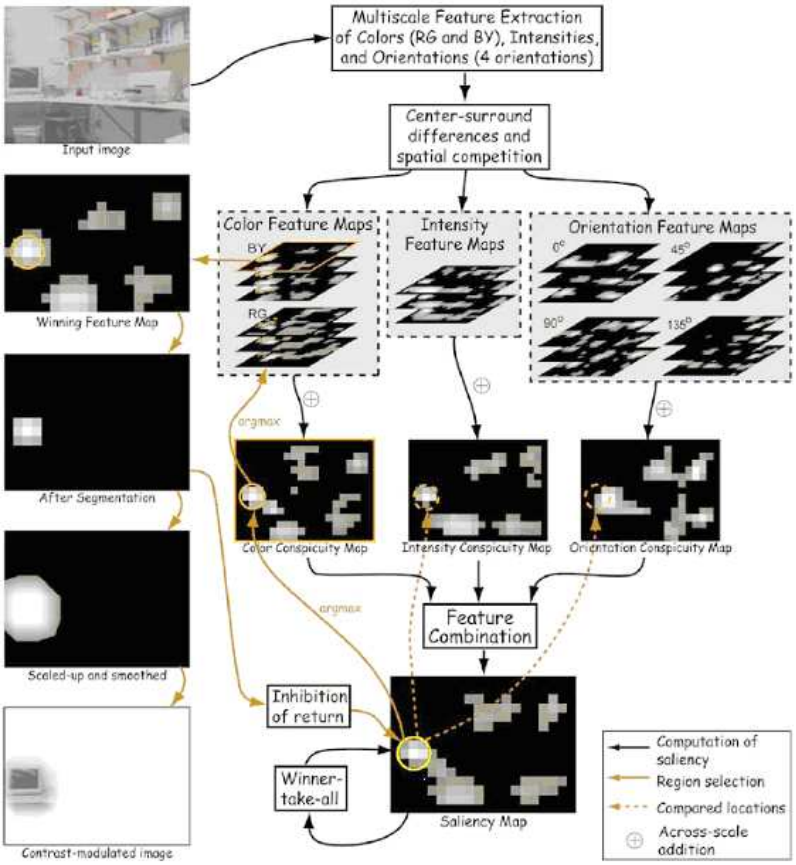
Σχήμα 2.9: *Σχηματική αναπαράσταση του μοντέλου των Itti et al. (από [50])*

Η υπόθεση που κρύβεται πίσω από την όλη διαδικασία δημιουργίας των τριών χαρτών μετά από επιμέρους κανονικοποίηση τους είναι το ότι παρόμοια χαρακτηριστικά (π.χ. κατευθυντικότητα σε διαφορετικές γωνίες) ανταγωνίζονται έντονα μεταξύ τους για να γίνουν σημαντικά, ενώ διαφορετικά χαρακτηριστικά συνεισφέρουν αυτόνομα στον τελικό κύριο χάρτη, ο οποίος δημιουργείται από τον μέσο όρο των ενδιάμεσων χαρτών:

$$S = \frac{1}{3}(N(\overline{I}) + N(\overline{C}) + N(\overline{O})) \tag{2.8}$$

2.4 Επισκόπηση εφαρμογών

Σε αυτήν την Ενότητα θα παρουσιάσουμε επισκόπηση εφαρμογών των μεθόδων υπολογισμού σημαντικότητας. Η επισκόπηση που επιχειρούμε αφορά κυρίως στους



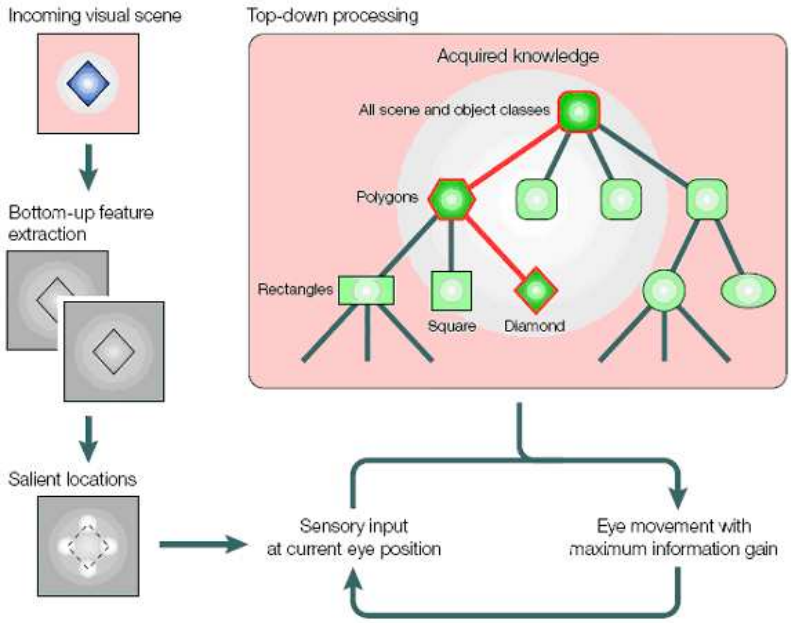
Σχήμα 2.10: *Ενδεικτικά αποτελέσματα για κάθε επίπεδο της αρχιτεκτονικής των Rutishauer et al.. Στο πρώτο επίπεδο γίνεται εξαγωγή χαρακτηριστικών σε πολλαπλές κλίμακες και επεξεργασία τους με φίλτρα κέντρου-περιφέρειας. Ακολουθεί ο συνδυασμός τους στον κύριο χάρτη και η επιλογή της πιο σημαντικής περιοχής από ένα δίκτυο όλα-για-το-νικητή. Στην συνέχεια εντοπίζεται ο χάρτης χαρακτηριστικών που έχει συνεισφέρει περισσότερο σε αυτήν την περιοχή και γίνεται η τμηματοποίηση του για να καταλήξει σε μία μάσκα για κάθε αντικείμενο ενδιαφέροντος. (από [110])*

τομείς εφαρμογής στους οποίους εστιάσαμε για την αξιολόγηση των προτεινόμενων μοντέλων που θα αναλύσουμε διεξοδικά στα επόμενα Κεφάλαια.

2.4.1 Ανίχνευση αντικειμένων

Οι Rutishauser *et al.* βασιζόμενοι στο βιολογικό μοντέλο του κύριου χάρτη παρουσίασαν μια πρώτη προσέγγιση αναγνώρισης αντικειμένων με καθοδικό κανάλι[4] οπτικής προσοχής (συνειδητή εστίαση προσοχής) [110]. Η τεχνική τους είναι αρκετά απλή και τα πειράματα περιορισμένα. Ωστόσο οι ενδείξεις για γρήγορη και αξιόπιστη αναγνώριση αντικειμένων είναι αρκετές για να κεντρίσουν το ερευνητικό ενδιαφέρον. Το Σχήμα 2.10 δείχνει το υπολογιστικό μοντέλο και ενδεικτικά αποτελέσματα για κάθε επίπεδο. Το αποτέλεσμα είναι μία μάσκα για κάθε πιθανό αντικείμενο στην σκηνή, η οποία χρησιμοποιείται για την εξαγωγή χαρακτηριστικών σημείων που τροφοδοτούν το υποσύστημα ανίχνευσης.

Στην ίδια κατεύθυνση βρίσκεται και ο συνδυασμός συνειδητής και μη συνειδητής πληροφορίας για την καθοδήγηση της οπτικής προσοχής, όπως περιγράφεται από τους Schill *et al.* [116]. Το μοντέλο που προτείνουν χρησιμοποιείται για αναγνώριση σκηνής



Σχήμα 2.11: *Ο συνδυασμός οπτικής προσοχής και αναγνώρισης αντικειμένου των Schill et al.. Με την βοήθεια της οπτικής προσοχής συλλέγεται όσο το δυνατόν περισσότερη πληροφορία για περιοχές της εικόνας, η οποία θα βοηθήσει στον αποκλεισμό υποθέσεων αναγνώρισης που έχει μάθει το σύστημα. Κάθε εστίαση του "ματιού" αντιστοιχεί σε ένα μονοπάτι του γράφου. (από [116])*

ή αντικειμένου και χρησιμοποιεί την οπτική προσοχή για να εστιάσει σε εκείνες τις περιοχές που οδηγούν στην προοδευτική προσέγγιση της ταυτότητας της σκηνής ή του αντικειμένου. Επομένως, δημιουργείται σταδιακά ένα ιεραρχικό δέντρο γνώσης μέσω εκπαίδευσης. Τα φύλλα αντιπροσωπεύουν ταυτοποιημένα αντικείμενα, ενώ οι ενδιάμεσοι κόμβοι αντιπροσωπεύουν γενικότερες κατηγορίες αντικειμένων. Οι σύνδεσμοι μεταξύ των κόμβων περιέχουν πληροφορία που βοηθάει τον διαχωρισμό μεταξύ των πιθανών αντικειμένων (αναμενόμενη απόκριση σε επιλεγμένα χαρακτηριστικά για συγκεκριμένα σημεία του αντικειμένου). Κατά την επαναληπτική διαδικασία αναγνώρισης το σύστημα προγραμματίζει την επόμενη εστίαση στην κατεύθυνση που μεγιστοποιεί την πληροφορία για ένα αντικείμενο που υπάρχει στο δέντρο. Το Σχήμα 2.11 απεικονίζει την διαδικασία που περιγράψαμε.

Οι Riesenhuber και Poggio [108, 109] έχουν εφαρμόσει τεχνικές οπτικής προσοχής σε ταυτοποίηση και αναγνώριση αντικειμένων υπό διαφορετικές γωνίες. Τα πειράματα τους έγιναν σε απλές εικόνες και ήταν ενδεικτικά της χρησιμότητας της οπτικής προσοχής στην επιτυχή αναγνώριση επιθυμητών αντικειμένων. Όπως αναφέρουν και οι συγγραφείς, το επόμενο βήμα είναι να ενσωματώσουν συνειδητή πληροφορία στο μοντέλο.

2.4.2 Αποθορυβοποίηση

Η αποθορυβοποίηση εικόνας παραμένει ένα σημαντικό πρόβλημα και συνεχίζει να θέτει προκλήσεις στους ερευνητές του χώρου. Ο θόρυβος στις εικόνες παρουσιάζεται συνήθως εξαιτίας των συσκευών λήψης, της μεταφοράς δεδομένων, της κβαντοποίησης κτλ και παρά τον μεγάλο αριθμό δημοσιευμένων τεχνικών καθεμία έχει τις δικές της προϋποθέσεις και τα δικά της μειονεκτήματα/πλεονεκτήματα και περιορισμούς.

Πρακτικά, η επιτυχημένη αποθορυβοποίηση εξαρτάται από την εφαρμογή.

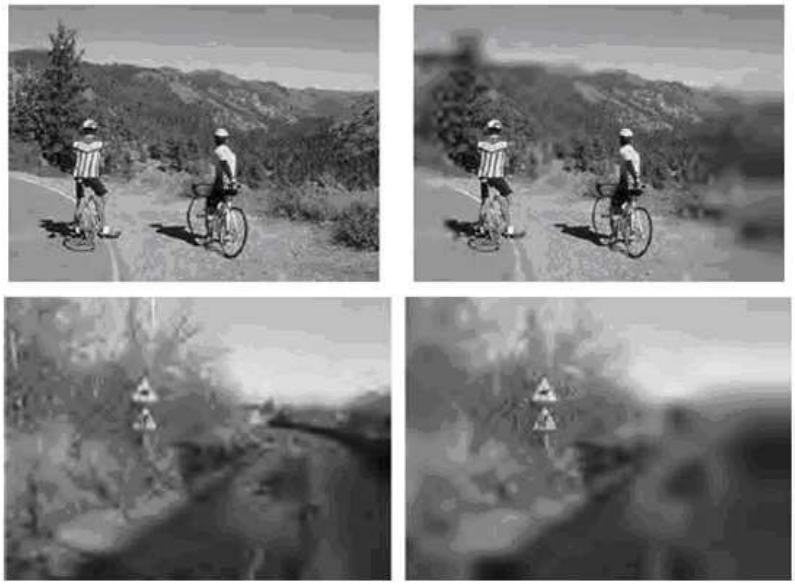
Ο μετασχηματισμος κυματιδίων έχει χρησιμοποιηθεί ευρέως για την καταστολή του θορύβου σε εικόνες, καθώς ιδιότητες όπως η σποραδικότητα και η αποσύνθεση σε πολλαπλές κλίμακες βοηθούν την αποθορυβοποίηση. Η σποραδικότητα είναι η ιδιότητα του μετασχηματισμού κυματιδίων να συγκεντρώνουν την ενέργεια του καθαρού σήματος σε μικρό αριθμό συντελεστών με μεγάλη ένταση επιτρέποντας έτσι την μείωση του θορύβου με την εφαρμογή κατάλληλου κατωφλίου [26]. Η συρρίκνωση στον χώρο των κυματιδίων, μία γνωστή μη-γραμμική μέθοδος κατωφλίωσης βασίζεται σε αυτήν την ιδιότητα.

Οι πρώτες τεχνικές συρρίκνωσης βασίζονταν είτε στην εφαρμογή ενός γενικού κατωφλίου, όπως η *VisuShrink* [26], ή στην εφαρμογή ενός κατωφλίου προσαρμοζόμενου στην ζώνη μετασχηματισμού, όπως οι SureShrink [27] και BayesShrink [20]. Επεκτάσεις αυτών των τεχνικών, οι οποίες εκμεταλλεύονται την εξάρτηση μεταξυ συντελεστών της ίδιας ή και διαφορετικής κλίμακας, έχουν προταθεί και παρουσιάζουν καλά αποτελέσματα. Η βιβλιογραφία είναι πλούσια σε μεθόδους που προβλέπουν την συνεισφορά ενός συντελεστή στο καθαρό σήμα βασιζόμενες και σε ενδό- και σε διαζωνική εξάρτηση [87, 18, 88, 33, 118, 119, 120]. Αυτές οι μέθοδοι είτε χρησιμοποιούν πληροφορία από την τοπική γειτονιά των συντελεστών είτε υποθέτουν ένα στατιστικό μοντέλο για την εξάρτηση γειτονικών συντελεστών. Ανεξάρτητα του αν υποθέτουν κάποιο μοντέλο ή όχι, απαιτούν τον υπολογισμό κατάλληλων στατιστικών ιδιοτήτων στο επίπεδο του κάθε συντελεστή ή/και της κάθε ζώνης του μετασχηματισμού. Οι Chen *et al.* πρότειναν την μέθοδο *NeighShrink*, η οποία βασίζεται σε κατωφλίωση ανάλογα με την τοπική γειτονιά [18]. Οι Sendur και Selesnick υπέθεσαν μη-κανονικές bivariate κατανομές εξάρτησης και πρότειναν μη γραμμικές συναρτήσεις συρρίκνωσης με χρήση Bayesian θεωρίας εκτίμησης [118, 119].

2.4.3 Κωδικοποίηση

Αν και σε πρώιμο στάδιο ακόμη, υπάρχει εξελισσόμενη έρευνα στην εφαρμογή των μοντέλων οπτικής προσοχής στην συμπίεση εικόνας και βίντεο. Απώτερος στόχος είναι η συμπίεση της πληροφορίας με βάση την ανθρώπινη οπτική αντίληψη, έτσι ώστε οι περιοχές που ελκύουν περισσότερο την προσοχή μας να υποστούν λιγότερη συμπίεση από τις υπόλοιπες [13, 90]. Η λογική πίσω από την συμπίεση με περιοχές ενδιαφέροντος στηρίζεται στην μη ομοιόμορφη κατανομή των φωτοδεκτών του ανθρώπινου αμφιβληστροειδούς, σύμφωνα με την οποία μόνο ένα μικρό μέρος του οπτικού πεδίου γύρω από το κέντρο προσοχής καταγράφεται με μεγάλη ανάλυση [145]. Επομένως δεν είναι αναγκαία η κωδικοποίηση κάθε καρέ με ομοιόμορφη ποιότητα, καθώς ο ανθρώπινος παρατηρητής θα παρατηρήσει προσεχτικά μόνο μέρος του.

Μία από τις πρώτες ολοκληρωμένες δουλειές στον χώρο είναι η προσπάθεια των Bradley και Stentiford [13] να ενσωματώσουν την πληροφορία εστίασης οπτικής προσοχής σε πειράματα που αφορούν στην συμπίεση εικόνων με το πρότυπο JPEG 2000 (JP2K). Τα πειράματα έδειξαν ότι σε συνθήκες χαμηλού ρυθμού μετάδοσης η χρήση περιοχών ενδιαφέροντος μπορεί να οδηγήσει σε μικρή βελτίωση της ποιότητας συμπίεσης (Σχήμα 2.12). Οι Itti *et al.* παρούσιασαν μία αναλυτική μελέτη και αξιολόγηση του μοντέλου κύριου χάρτη στην συμπίεση ακολουθιών κατά MPEG-1 και MPEG-4 [46]. Τα αποτελέσματα τους αποδεικνύουν την αντικειμενική βελτίωση σε συμπίεση που επιτυγχάνει η χρήση σημαντικότητας. Στο Κεφάλαιο 3 περιγράφουμε την εφαρμογή του προτεινόμενου μοντέλου σε κωδικοποίηση ακολουθιών. Η



Σχήμα 2.12: *Σύγκριση μεταξύ JP2K (αριστερά) και JP2K με χρήση οπτικής προσοχής (δεξιά) για ρυθμό μετάδοσης 1 bpp και 0.125 bpp.*

μεθοδολογία είναι παρόμοια με αυτή των Itti *et al.* και παρουσιάζονται πλήθος ποιοτικών και ποσοτικών αποτελεσμάτων.

2.4.4 Χωροχρονική ανάλυση ακολουθιών

Η επεξεργασία/ανάλυση ακολουθιών γίνεται συνήθως καρέ-καρέ χρησιμοποιώντας πληροφορία από έναν μικρό αριθμό (τυπικά 2) γειτονικών καρέ. Στην συνέχεια τα ενδιάμεσα αποτελέσματα συνδέονται μεταξύ τους για την εξαγωγή συμπερασμάτων για την συνολική ακολουθία. Το αποτέλεσμα αυτής της διαδικασίας είναι να μην λαμβάνεται υπόψη η πραγματική μακροπρόθεσμη χωροχρονική εξέλιξη των περιοχών καθώς κατά την επεξεργασία γειτονικών καρέ γίνονται πολλές υποθέσεις (παραμετρικά μοντέλα, περιορισμοί εξομάλυνσης κτλ.). Επιπρόσθετα, αυτές οι τεχνικές είναι συχνά ευαίσθητες σε θόρυβο και γίνονται υπολογιστικά ακριβές όταν χρησιμοποιούν π.χ. τεχνικές για τον υπολογισμό οπτικής ροής [44, 3].

Οι Adelson και Bergen ήταν ανάμεσα στους πρώτους που χρησιμοποίησαν χωροχρονικούς τελεστές για την *επεξεργασία ογκομετρικών δεδομένων* και έκαναν ενδιαφέροντα πειράματα που αποδεικνύουν την ανωτερότητα αυτών των τελεστών σε πολλά προβλήματα ανάλυσης [2]. Οι Bolles και Baker χρησιμοποίησαν επίσης χωροχρονική αναπαράσταση της ακολουθίας (χωροχρονικοί όγκοι με την τρίτη διάσταση να είναι ο χρόνος) για την εξαγωγή παραμέτρων κίνησης της κάμερας [7, 8]. Πιο πρόσφατα η χωροχρονική επεξεργασία δεδομένων χρησιμοποιήθηκε για ανάλυση περιοδικότητας [68], ανάλυση της κίνησης της κάμερας [53], εφαρμογές επίβλεψης χώρων [52], υπολογισμό κίνησης και κατάτμηση [81, 89, 111]. Ερευνητές του χώρου έχουν επίσης χρησιμοποιήσει παρόμοιους τελεστές σε ανιχνευτές χωροχρονικά σημαντικών σημείων για την βελτίωση της απόδοσης αναγνώρισης αντικειμένων ή ανάκτησης περιοχών [76, 74]. Στην πλειοψηφία όμως των προηγούμενων τεχνικών η επεξεργασία γίνεται βάσει χωροχρονικών τομών και δεν χρησιμοποιείται η συνολική πληροφορία του όγκου.

Σημαντικός τομέας στην χωροχρονική ανάλυση ακολουθιών είναι η *ανίχνευση*

οπτικών δραστηριοτήτων. Οι περισσότερες μέθοδοι στο συγκεκριμένο χώρο βασίζονται είτε σε κάποιο μοντέλο [99, 113, 64], επιχειρώντας να υπολογίσουν τις παραμέτρους του μοντέλου από τα δεδομένα, είτε απευθείας στην οπτική πληροφορία, επιχειρώντας έτσι να αναπαραστήσουν την ζητούμενη πληροφορία με περιγραφείς χαμηλού επιπέδου [9, 152, 6, 56]. Για να επιτευχθεί η σημασιολογική ερμηνεία ενός γεγονότος δεν είναι απαραίτητο να γίνει επεξεργασία του συνόλου της διαθέσιμης οπτικής πληροφορίας. Συγκεκριμένα μέρη της σκηνής είναι συνήθως αντιπροσωπευτικά του περιεχομένου της, έτσι ώστε περιορίζοντας την επεξεργασία σε αυτά πετυχαίνουμε την σωστή περιγραφή με μικρότερο υπολογιστικό κόστος. Αυτή είναι και η κεντρική ιδέα των τεχνικών που εξάγουν περιοχές ή σημεία ενδιαφέροντος που συνήθως εντοπίζονται γύρω από "γωνίες", ακμές ή έντονη υφή. Αυτές οι περιοχές ενδιαφέροντος μπορούν να εντοπιστούν είτε απευθείας είτε μέσω εξαγωγής και ομαδοποίησης σημείων ενδιαφέροντος.

Η πλειοψηφία των *ανιχνευτών σημαντικών σημείων* βασίζεται στον υπολογισμό πινάκων που περιγράφουν την κατανομή της παραγώγου στην τοπική γειτονιά του σημείου. Οι ιδιοτιμές αυτών των πινάκων αναπαριστούν τις κύριες κατευθύνσεις των οποίων οι τιμές χρησιμοποιούνται για τον υπολογισμό της σημαντικότητας του σημείου. Σε παρόμοια λογική βασίζεται και ο ανιχνευτής σημείων του Lindeberg [66], ο οποίος χρησιμοποιεί τον Εσσιανό πίνακα. Οι δεύτερες παράγωγοι του πίνακα παρουσιάζουν έντονη απόκριση σε σημεία που ξεχωρίζουν από την γειτονιά τους. Αυτή η μέθοδος γίνεται αμετάβλητη σε κλίμακα μετά την επιλογή της χαρακτηριστικής κλίμακας του σημείου για την οποία μια δεδομένη συνάρτηση (π.χ. Λαπλασιανή) γίνεται μέγιστη στο σύνολο των κλιμάκων [67]. Δεδομένων των σημείων και των χαρακτηριστικών κλιμάκων τους, ένας επαναληπτικός υπολογισμός ελλιπτικής περιοχής γύρω από το σημείο καταλήγει στην επιθυμητή περιοχή ενδιαφέροντος. Παρόμοιες μέθοδοι έχουν χρησιμοποιηθεί με επιτυχία σε πολλά πεδία της μηχανικής όρασης, όπως μεγάλης-κλίμακας ανάκτηση εικόνων [115, 144], ανάκτηση αντικειμένων από ακολουθίες [123, 122], αναγνώριση υφής [64], κατηγοριοποίηση αντικειμένων [21, 84] και υπολογισμό/ανίχνευση συμμετρίας [143]. Έξι μέθοδοι για την ανάκτηση σημείων/περιοχών ενδιαφέροντος περιγράφονται και αξιολογούνται στην πρόσφατη δουλειά των Mikolajczyk *et al.* [75].

Αν και η έρευνα γύρω από την ανίχνευση χωρικών σημείων ενδιαφέροντος σε εικόνες έχει προχωρήσει αρκετά τα τελευταία χρόνια δεν συμβαίνει το αντίστοιχο στην *ανίχνευση χωροχρονικά σημαντικών σημείων ή περιοχών σε ακολουθίες.* Οι Bobick *et al.*, σε μία από τις πρώτες σχετικές εργασίες, υπολογίζουν εικόνες ιστορικού της κίνησης (Motion-History-Images) για την αναπαράσταση ανθρώπινων δραστηριοτήτων και χρησιμοποιούν ροπές για την περιγραφή τους [9]. Τα αποτελέσματα που παρουσιάζουν είναι ικανοποιητικά, αλλά η μέθοδος προϋποθέτει τον πολύ καλό διαχωρισμό μεταξύ του ανθρώπου και του υποβάθρου. Οι περισσότερες πρόσφατες μέθοδοι είναι επεκτάσεις των αντίστοιχων τεχνικών για χωρικά σημεία ενδιαφέροντος που αναφέρθηκαν. Οι Laptev *et al.* [62, 63] και οι Schuldt *et al.* [114] επεκτείνουν την ιδέα των Harris *et al.* [40] και προτείνουν ανίχνευση σημείων γύρω από περιοχές στις οποίες η κίνηση αλλάζει απότομα κατεύθυνση. Προτείνουν ένα πλαίσιο αναγνώρισης οπτικών γεγονότων και εφαρμόζουν με επιτυχία τον ανιχνευτή σε αναγνώριση ανθρώπινων δραστηριοτήτων. Οι Ke *et al.* εξάγουν ογκομετρικά χαρακτηριστικά από χωροχρονικές γειτονιές και προτείνουν έναν ανιχνευτή περίπλοκων γεγονότων με υψηλά ποσοστά επιτυχίας [56]. Οι Boiman *et al.* [10] και οι Zelnik-Manor *et al.* [152] χρησιμοποίησαν επικαλυπτόμενες ογκομετρικές γειτονιές για την περιγραφή

δυναμικών δραστηριοτήτων και εφάρμοσαν τις μεθόδους τους στην ανίχνευση ανθρώπινων δραστηριοτήτων.

Πλήθος πρόσφατων τεχνικών ανίχνευσης αντικειμένων και χωροχρονικών γεγονότων βασίζεται σε μεθόδους "σάκου λέξεων" ("bag-of-words"). Σύμφωνα με αυτές τις μεθόδους, η οπτική πληροφορία μπορεί να αναπαρασταθεί επιτυχώς από ένα σύνολο "οπτικών λέξεων", οι οποίες αντιστοιχούν σε περιοχές στην στατική ή κινούμενη εικόνα. Οι Wang *et al.* προτείνουν μία μέθοδο αναγνώρισης ανθρώπινης δραστηριότητας από στατικές εικόνες χρησιμοποιώντας οπτικές λέξειςπου χαρακτηρίζουν το σχήμα των ανθρώπων όταν εκτελούν την εκάστοτε δραστηριότητα [146]. Οι Bosch *et al.* προτείνουν παρόμοια αναπαράσταση για την ανίχνευση αντικειμένων σε εικόνες με χρήση του πιθανοτικής λανθάνουσας σημασιολογικής ανάλυσης (probabilistic Latent Semantic analysis-pLSA). Οι Niebles *et al.* επεκτείνουν την τεχνική σε πραγματική χωροχρονική ανάλυση ακολουθιών και παρουσιάζουν αποτελέσματα σε κινούμενες εικόνες. Αναπαριστούν τις ακολουθίες με ένα σύνολο χωροχρονικών οπτικών λέξεων, των οποίων η πιθανότητα κατανομής προκύπτει με pLSA. Τα στατιστικά αποτελέσματα που παρουσιάζουν σε δύο σύνολα δεδομένων είναι ενδεικτικά των δυνατοτήτων της μεθόδου να εντοπίζει επιτυχώς οπτικές δραστηριότητες. Παρόμοια λογική ακολουθούν και οι Dóllar *et al.*, οι οποίοι προτείνουν έναν ανιχνευτή χωροχρονικών σημείων που βασίζεται στην περιοδικότητα, όπως αυτή μοντελοποιείται από φίλτρα Gabor [28]. Αποτελέσματα σε δημόσια διαθέσιμα σύνολα δεδομένων αποδεικνύει την αξιόπιστη απόδοση του μοντέλου τους ακόμη και με την χρήση του απλού ανιχνευτή περιοδικότητας για την ανίχνευση χωροχρονικών σημείων ενδιαφέροντος.

Μία άλλη κατηγορία μεθόδων που σχετίζεται με την έρευνα μας είναι η γενικευμένη αναγνώριση σημαντικών γεγονότων σε ακολουθίες εικόνων. Συνήθως υπολογίζεται μία καμπύλη προσοχής, η οποία συνοψίζει το συνολικό ενδιαφέρον των σκηνών και επομένως επιτρέπει την απόφαση για το ποια σημεία της ακολουθίας είναι τα πιο σημαντικά. Αυτή η διαδικασία παρουσιάζει ομοιότητες με δύο πεδία: α) ανίχνευση ασυνήθιστων γεγονότων [39] [47] [153], β) έρευνα στον χώρο της περίληψης ακολουθιών με χρήση καμπύλης οπτικής προσοχής [71] [72]. Υπάρχουν δύο ορισμοί των ασυνήθιστων γεγονότων: α) γεγονότα που δεν μοιάζουν με κάποια από ένα σύνολο γνωστών και β) σπάνια γεγονότα που δεν μοιάζουν με αυτά που συναντώνται συχνά στην ακολουθία ("συνηθισμένα"). Οι περισσότερες τεχνικές που ανήκουν στην πρώτη κατηγορία προκαθορίζουν ένα σύνολο συνηθισμένων δραστηριοτήτων και στην συνέχεια μέσω συγκρίσεων ανιχνεύουν τα λιγότερο συνηθισμένα [10] [124]. Στην δεύτερη κατηγορία ανήκουν οι τεχνικές που εντοπίζουν σπάνια γεγονότα εξάγοντας πρώτα με στατιστικό τρόπο τα "συνηθισμένα". Οι Zhong *et al.* χωρίζουν την ακολουθία σε ίσα μέρη, εξάγουν χαρακτηριστικά, σχηματίζουν τον πίνακα σύγχυσης και θεωρούν ότι ασυνήθιστα γεγονότα συμβαίνουν σε περιοχές χαμηλής συσσώρευσης στοιχείων [153]. Παρόμοια λογική, αλλά κάτω από διαφορετικό πλαίσιο και σε διαφορετικές εφαρμογές εφαρμόζουν οι Adam *et al.* [1] σε πρόσφατη δουλειά τους. Πηγαίνοντας ένα βήμα παραπέρα οι Ma *et al.* [71] [72] προτείνουν ένα πλαίσιο για ανίχνευση σημαντικών γεγονότων σε ακολουθίες και χρήση τους για δημιουργία περιλήψεων. Το μοντέλο τους βασίζεται σε οπτική προσοχή που σχετίζεται με πρόσωπα, κίνηση της κάμερας, ήχο, μουσική και φωνή

□

Κεφάλαιο 3

Χωρικά μοντέλα οπτικής προσοχής

3.1 Εισαγωγή

Σε αυτό το κεφάλαιο περιγράφουμε τις πρώτες απόπειρες επέκτασης καθιερωμένων τεχνικών οπτικής προσοχής και την εφαρμογή τους σε πραγματικές εφαρμογές. Οι προτεινόμενες επεκτάσεις γίνονται πάνω στα μοντέλα επιλεκτικού συντονισμού των Tsotsos *et al.* [137] και κύριου χάρτη των Itti *et al.* [48], τα οποία σχολιάσαμε στο Κεφάλαιο 2. Χρησιμοποιούμε ένα απλοποιημένο μοντέλο επιλεκτικού συντονισμού για να ανιχνεύσουμε περιοχές ενδιαφέροντος, δηλαδή περιοχές με έντονη παρουσία καθαρού σήματος, σε θορυβώδεις εικόνες και εξάγοντας στατιστικά από αυτές επιτυγχάνουμε αποθορυβοποίηση με πολύ καλά ποιοτικά και στατιστικά αποτελέσματα. Οι επεκτάσεις στο μοντέλο του κύριου χάρτη σχετίζονται τόσο με την εισαγωγή νέων χαρτών χαρακτηριστικών όσο και με τον αποδοτικότερο συνδυασμό τους. Επιλέγουμε την ανίχνευση προσώπων σε περίπλοκα περιβάλλοντα και την κωδικοποίηση ακολουθιών για να αξιολογήσουμε τις προτάσεις μας. Ειδικότερα στην κωδικοποίηση διεξάγουμε πλήθος ποιοτικών και ποσοτικών πειραμάτων.

3.2 Απλοποιημένο μοντέλο επιλεκτικού συντονισμού

Στα πλαίσια της μελέτης του μοντέλου επιλεκτικού συντονισμού των Tsotsos *et al.* [137] αναπτύξαμε την μέθοδο *salienShrink* για αποθορυβοποίηση εικόνων, η οποία βασίζεται στον υπολογισμό ενός χάρτη σημαντικών συντελεστών του μετασχηματισμού κυματιδίων. Σαν σημαντικούς θεωρούμε τους συντελεστές που αντιστοιχούν στο καθαρό σήμα και κατά συνέπεια πρέπει να διατηρηθούν ανέπαφοι κατά την αποθορυβοποίηση. Προτείνουμε μία υπολογιστικά αποδοτική μέθοδο για την ανίχνευση των σημαντικών περιοχών των ζωνών του μετασχηματισμού σε πολλές κλίμακες. Από αυτές τις περιοχές εξάγουμε μία ακριβή ακριβή τιμή για το επιπέδου θορύβου, την οποία χρησιμοποιούμε για να βελτιώσουμε την απόδοση τεχνικών αποθορυβοποίησης που βασίζονται στην συρρίκνωση[2]. Παρουσιάζονται αναλυτικά αποτελέσματα που αποδεικνύουν τόσο την οπτική βελτίωση που επιτυγχάνεται όσο και την βελτίωση του σηματοθορυβικού λόγου. Μία εισαγωγή στις επικρατέστερες τεχνικές του χώρου δίνεται στην Ενότητα 2.4.2.

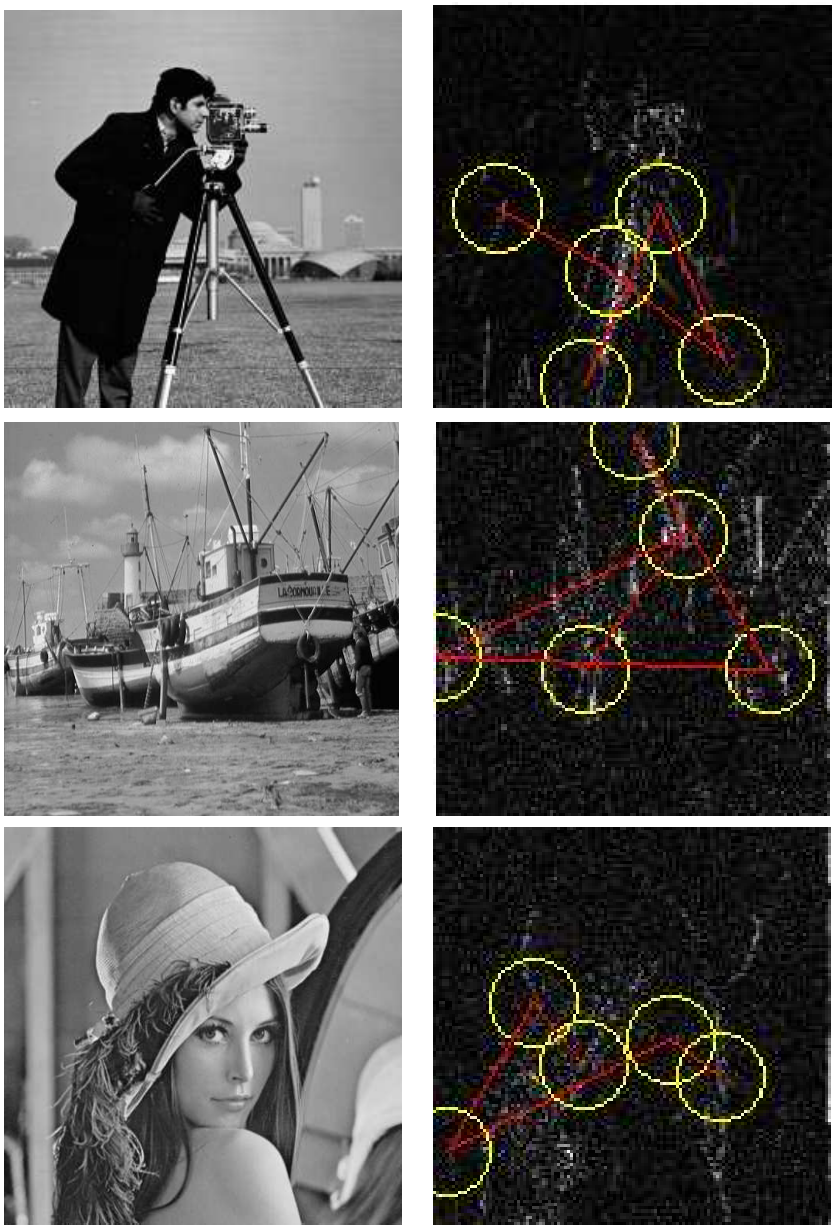
Στην προτεινόμενη μέθοδο, χρησιμοποιούμε την έννοια της σημαντικότητας στον χώρο του μετασχηματισμού κυματιδίων και προτείνουμε έναν απλό τρόπο αποθορυβοποίησης που παράγει εικόνες υψηλού PSNR και υψηλής οπτικής ποιότητας. Η βασική μας υπόθεση είναι ότι οι περιοχές που εντοπίζουμε και θεωρούμε σημαντικές

αντιστοιχούν σε περιοχές όπου το σήμα είναι έντονο και ο θόρυβος χαμηλός (π.χ. περιοχές πολλών ακμών). Η είσοδος αποσυντίθεται σε ζώνες συχνοτήτων με τον κλασσικό μετασχηματισμό κυματιδίων (DWT) ή τον Μιγαδικό μετασχηματισμό κυματιδίων Διπλού-Δένδρου (DTCWT) και η αναζήτηση σημαντικών συντελεστών γίνεται σε κάθε ζώνη. Η αναζήτηση υλοποιείται μέσω ενός δια-κλιμακωτού δικτύου όλα-για-τον-νικητή (WTA), ο οποίος εντοπίζει τους πιο σημαντικούς συντελεστές βάσει του μέτρου τους. Η μέθοδος καταλήγει σε έναν τελικό κύριο χάρτη S για κάθε υποζώνη. Στην Ενότητα 3.2.3 παρουσιάζονται πειραματικά αποτελέσματα και συγκρίσεις με δημόσια διαθέσιμες υλοποιήσεις γνωστών αλγορίθμων του χώρου. Ο κύριος χάρτης που προκύπτει χρησιμοποιείται για την βελτίωση των αποτελεσμάτων της μεθόδου *BiShrink* των Sendur και Selesnick [118] τόσο στο πεδίο του DWT όσο και στο πεδίο του DTCWT.

3.2.1 Υπολογισμός περιοχών ενδιαφέροντος στο πεδίο των κυματιδίων

Στον μετασχηματισμό κυματιδίων εμπλέκεται ένα βαθυπερατό $h_\phi(\cdot)$ και ένα υψιπερατό φίλτρο $h_\psi(\cdot)$, τα οποία εφαρμόζονται στην οριζόντια και κάθετη κατεύθυνση. Το αποτέλεσμα των φίλτρων υποδειγματοληπτείται κατά δύο και δημιουργείται ένα σύνολο $D_{DWT} = \{D_i\}$ με $i = \{1, 2, 3\}$, από τρεις ζώνες υψηλών συχνοτήτων, οι οποίες αντιστοιχούν σε οριζόντιες, κάθετες και διαγώνιες λεπτομέρειες της εισόδου και μία ζώνη χαμηλών συχνοτήτων A, η οποία αντιστοιχεί στους συντελεστές προσέγγισης. Αν και ο DWT έχει εφαρμοστεί επιτυχώς σε πλήθος εφαρμογών πάσχει σε επιλεκτικότητα σε κατεύθυνση και μεταβλητότητα σε μετατόπιση. Πρόσφατα ο Kingsbury πρότεινε τον μιγαδικό μετασχηματισμό διπλού-δένδρου ($DTCWT$), ο οποίος παρουσιάζει καλή επιλεκτικότητα σε κατεύθυνση και η απόκριση των ζωνών του είναι προσεγγιστικά μη μεταβλητή σε μετατόπιση [58]. Ο $DTCWT$ εφαρμόζει κυματίδια σε έξι διαφορετικές κατευθύνσεις και επομένως παράγει ένα σύνολο από έξι ζώνες υψηλών συχνοτήτων $D_{DTCWT} = \{D_i\}$ με $i = \{1, ..., 6\}$. Δύο κυματίδια σχετίζονται με κάθε κατεύθυνση, ένα για το πραγματικό και ένα για το φανταστικό μέρος. Οι Sendur και Selesnick χρησιμοποίησαν τον $DTCWT$ στην αποθορυβοποίηση εικόνων με πολύ καλά αποτελέσματα. Περισσότερες λεπτομέρειες για αυτήν την τεχνική δίνονται στην Ενότητα 3.2.2.

Για την επιλογή των πιο σημαντικών συντελεστών κυματιδίων προτείνουμε μια μέθοδο ανίχνευσης περιοχών ενδιαφέροντος καθοδηγούμενη από ένα μοντέλο οπτικής προσοχής. Αξιοποιούμε την αποσύνθεση σε πολλές κλίμακες του μετασχηματισμού DWT και εφαρμόζουμε μία απλοποιημένη μορφή του μοντέλου επιλεκτικού συντονισμού στις ζώνες υψηλών συχνοτήτων. Αρχικά επιλέγουμε τον ισχυρότερο συντελεστή στην κορυφή της πυραμίδας του μετασχηματισμού (αδρότερη κλίμακα του DWT) με μία διεργασία όλα-για-το-νικητή και διατηρούμε την θέση του "νικητή" σε όλες τις άλλες κλίμακες. Ακολουθούμε αυτήν τη λογική επειδή οι σημαντικοί συντελεστές (αυτοί με το μεγαλύτερο μέτρο) παραμένουν άθικτοι στις περισσότερες κλίμακες. Η διαδικασία αυτή αποτελείται και από έναν μηχανισμό αναστολής-με-επιστροφή[9] (IOR), ο οποίος εξασφαλίζει την σάρωση κάθε ζώνης με σειρά μειούμενης σημαντικότητας και δημιουργεί επομένως μία αλληλουχία εστιάσεων για κάθε ζώνη. Το κριτήριο τερματισμού ή -εναλλακτικά- ο αριθμός των εστιάσεων θα συζητηθεί στην Ενότητα 3.2.2. Τελικά υπολογίζεται ένας κύριος χάρτης S_i^ℓ για κάθε ζώνη i και σε κάθε κλίμακα ℓ του μετασχηματισμού.



Σχήμα 3.1: *Αρχικές εικόνες και 5 εστιάσεις της διαδικασίας όλα-για-το-νικητή για την διαγώνια ζώνη του μετασχηματισμού κυματιδίων των εικόνων (α) "cameraman", (β) "boat", (γ) "lena"*

3.2.2 Αποθορυβοποίηση

Η τεχνική αποθορυβοποίησης *BiShrink* που χρησιμοποιούμε στα πειράματα μας βασίζεται σε μη-Γκαουσσιανές κατανομές για την μοντελοποίηση των ενδο- και δια- ζωνικών εξαρτήσεων. Αν ο συντελεστής w_{2k} είναι στην ίδια θέση με τον k^{th} συντελεστή w_{1k}, αλλά σε αδρότερη κλίμακα, τότε η θορυβώδης παρατήρηση των w_{1k} και του "γονιού" του w_{2k} μπορεί να μοντελοποιηθεί ως $y_{1k} = w_{1k} + n_{1k}$ και $y_{2k} = w_{2k} + n_{2k}$ αντίστοιχα. Στην συνέχεια υπολογίζεται μία μη-γραμμική bivariate συνάρτηση συρρίκνωσης με την χρήση εκτιμήτριας μεγίστης εκ των υστέρων

πιθανοφάνειας και η εκτίμηση του w_{1k} υπολογίζεται ως

$$\hat{w_{1k}} = \frac{\left(\sqrt{(y_{1k}^2 + y_{2k}^2)} - \frac{\sqrt{3}\sigma_n^2}{\sigma}\right)_+}{\sqrt{(y_{1k}^2 + y_{2k}^2)}} \cdot y_{1k} \tag{3.1}$$

όπου

$$(g)_+ = \begin{cases} 0 & \text{if } g < 0 \\ g & \text{otherwise} \end{cases}$$

Η εκτιμήτρια απαιτεί γνώση της μεταβλητότητας του θορύβου σ_n^2 και της οριακής μεταβλητότητας σ^2 κάθε συντελεστή του μετασχηματισμού. Η μεταβλητότητα του θορύβου υπολογίζεται συνήθως με την χρήση ενός εκτιμητή ενδιάμεσης τιμής στην υψηλότερη κλίμακα της πυραμίδας στην ζώνη που ενισχύει τις διαγώνιες ακμές [119, 18].

$$\hat{\sigma_n^2} = \frac{median(|D_3|)}{0.6745} \tag{3.2}$$

Η οριακή μεταβλητότητα υπολογίζεται ως

$$\hat{\sigma_y^2} = \frac{1}{|N_{y_i}|} \cdot \sum_{y_i \in N_{y_i}} y_i^2 \tag{3.3}$$

όπου N_{y_i} είναι οι συντελεστές σε μία γειτονιά γύρω από τον συντελεστή y_i. Τέλος, η μεταβλητότητα σ υπολογίζεται ως

$$\hat{\sigma} = \sqrt{(\hat{\sigma_y^2} - \hat{\sigma_n^2})_+} \tag{3.4}$$

Για τον προσδιορισμό του κριτηρίου τερματισμού της διεργασίας όλα-για-το-νικητή εκμεταλλευόμαστε την (3.3) αφήνοντας το μοντέλο να εστιάζει σε διαφορετικές περιοχές όσο ο λόγος μεταξύ της $\hat{\sigma}$ μεταξύ δύο διαδοχικών εστιάσεων παραμένει σε ένα δεδομένο εύρος $[\epsilon_1, \epsilon_2]$. Η λογική που ακολουθούμε είναι η εξής: Για τις σημαντικές περιοχές χαμηλώνουμε την εκτίμηση της μεταβλητότητας του θορύβου, καθώς στις εστιασμένες περιοχές ο θόρυβος είναι χαμηλός, για να αποφύγουμε την αφαίρεση συντελεστών καθαρού σήματος. Οι χάρτες S_i^ℓ, όπως υπολογίζονται από τον μηχανισμό IOR, έχουν χαμηλές τιμές στις εστιασμένες περιοχές (καθαρό σήμα) και μεγάλες τιμές αλλού. Τελικά, πολλαπλασιάζουμε κάθε κύριο χάρτη με την εκτιμώμενη τιμή της μεταβλητότητας θορύβου μετά από φιλτράρισμα του με μία Γκαουσσιανή και κανονικοποίηση του στο διάστημα $[0, 1]$

$$\hat{\sigma} = \sqrt{(\sigma_y^2 - S_i^\ell \cdot \hat{\sigma_n^2})} \tag{3.5}$$

Οι πιο γνωστές τεχνικές συρρίκνωσης βασίζονται στην εκτίμηση της μεταβλητότητας του σήματος και την χρήση της για τον υπολογισμό ενός κατάλληλου κατωφλίου [26, 27, 20]. Επομένως ο προτεινόμενος τρόπος για την ακριβέστερη εκτίμηση της μεταβλητότητας του σήματος μπορεί να συνδυαστεί με οποιαδήποτε από αυτές τις τεχνικές. Στην Ενότητα 3.2.3 παρουσιάζουμε αποτελέσματα για την μέθοδο *BiShrink*.

3.2.3 Πειραματικά αποτελέσματα

Για την εκτίμηση των αποτελεσμάτων της προτεινόμενης μεθόδου χρησιμοποιήσαμε εικόνες που είναι κοινές στην σχετική βιβλιογραφία. Όλες οι εικόνες είναι

Πίνακας 3.1: *PSNR αποτελέσματα για την εικόνα "house" (σ dB)*

	15	20	25	30	35
DWT-salienShrink	31.29	29.75	28.45	27.38	26.48
DWT-BiShrink	30.43	28.58	27.15	26.13	25.31
BayeShrink	29.2	26.58	26.04	25.59	24.80
Donoho	28.40	25.32	22.93	20.98	19.42
DTCWT-salienShrink	32.71	**31.32**	**30.21**	**29.32**	**28.61**
DTCWT-BiShrink	**32.74**	31.27	30.12	29.14	28.32

ασπρόμαυρες και έχουν μέγεθος 256×256. Για συγκρίσεις επιλέγουμε τις τεχνικές *BiShrink*, την *BayeShrink* και την μέθοδο αποθορυβοποίησης των Donoho *et al.*, η οποία περιλαμβάνεται στο Matlab 5.1. Χρησιμοποιούμε την δημόσια διαθέσιμη υλοποίηση της *BiShrink* [119] και την δική μας υλοποίηση για την *BayeShrink*. Για τα αποτελέσματα που παρουσιάζουμε θέτουμε το εύρος $[\epsilon_1, \epsilon_2] = [0.8, 1.2]$ και το μέγεθος της IOR να είναι μέρος της ελάχιστης διάστασης της εικόνας ($\frac{1}{6} \cdot min\{rows, cols\}$). Η απόδοση των αλγορίθμων αποθορυβοποίησης μετράται με τον σηματοθορυβικό λόγο κορυφής PSNR, ο οποίος υπολογίζεται ως

$$PSNR = 20 \cdot \log_{10} \left(\frac{255}{\sqrt{(v-r)^2}} \right) \tag{3.6}$$

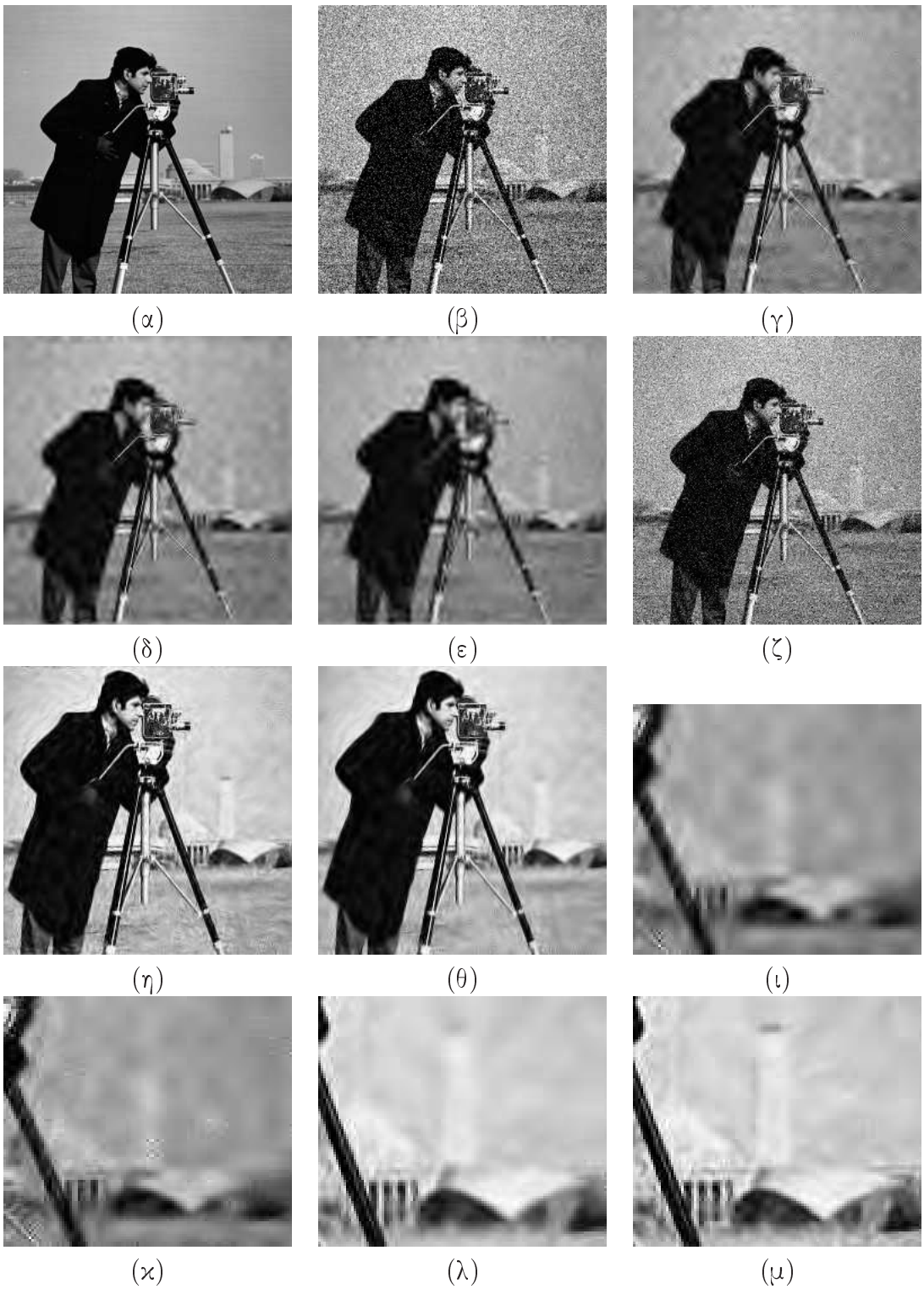
όπου v είναι η αρχική και r είναι η αποθορυβοποιημένη εικόνα αντίστοιχα.

Συνολικά ο προτεινόμενος αλγόριθμος *DWT-salienShrink* εμφανίζει μέσο κέρδος σε $PSNR$ 1.2dB σε σχέση με τον *DWT-BiShrink* και 0.58dB σε σχέση με τον *DTCWT-BiShrink*. Στις περισσότερες περιπτώσεις η βελτίωση είναι πιο εμφανής στα υψηλότερα επίπεδα θορύβου όπως φαίνεται στους πίνακες αποτελεσμάτων. Παράλληλα η οπτική βελτίωση είναι εμφανής σχεδόν σε όλες τις περιπτώσεις. Τα Σχήματα 3.2, 3.3 και 3.4 παρουσιάζουν συγκριτικά αποτελέσματα για όλες τις τεχνικές, καθώς και μεγεθυσμένες λεπτομέρειες των αποθορυβοποιημένων εικόνων. Οι Πίνακες 3.1, 3.2, 3.3 και 3.4 περιέχουν τις τιμές $PSNR$ σε dB των αλγορίθμων για ενδεικτικές εικόνες και για διάφορες τιμές θορύβου. Κάθε νούμερο εξάγεται ως ο μέσος όρος πέντε εφαρμογών της εκάστοτε μεθόδου. Ακόμη κ όταν η διαφορά μεταξύ της προτεινόμενης τεχνικής και της *BiShrink* σε $PSNR$ είναι μικρή (π.χ. "house", Πίνακας 3.1), η *salienShrink* έχει σαν αποτέλεσμα πιο καθαρές ακμές, όπως φαίνεται από τις μεγεθυσμένες λεπτομέρειες των εικόνων στα αντίστοιχα Σχήματα.

3.3 Επεκτεταμένο μοντέλο κύριου χάρτη

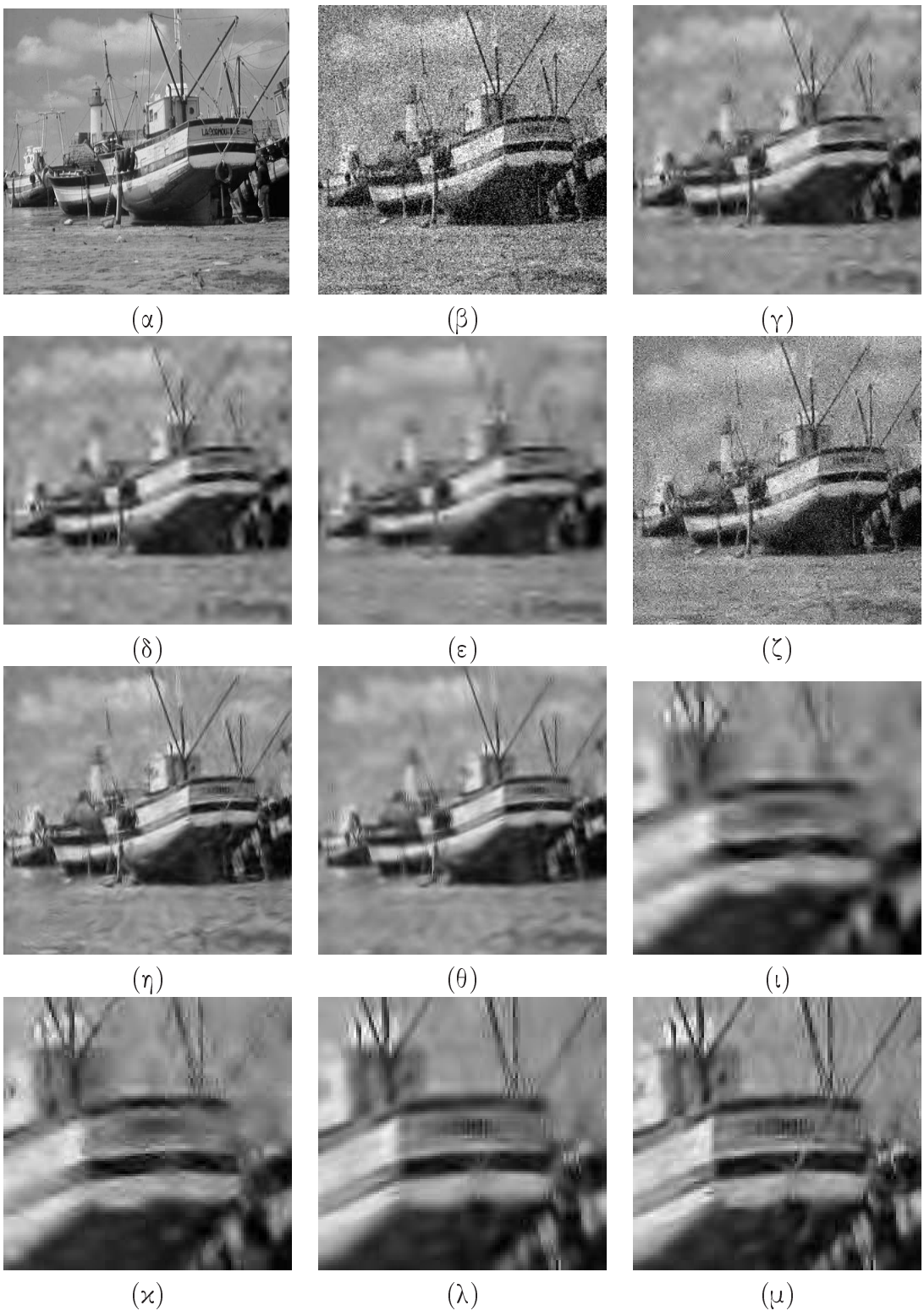
Στην πρώτη μας απόπειρα να ασχοληθούμε ενεργά με την μελέτη, υλοποίηση και ανάπτυξη του μοντέλου κύριου χάρτη, όπως αυτό παρουσιάστηκε για πρώτη φορά από τους Itti *et al.* [48], προτείναμε την επέκταση του μοντέλου με περισσότερους χάρτες χαρακτηριστικών (ανοδική διεργασία) και με μία διαδικασία ευφυούς συνδυασμού τους βάσει πρότερης γνώσης για το περιβάλλον της σκηνής (καθοδική διεργασία) [100, 101].

Η επεξεργασία στο ανοδικό κανάλι ακολουθεί τις αρχές που έθεσαν οι Itti *et al.*, αλλά χρησιμοποιεί διαφορετικό τρόπο υπολογισμού της κατευθυντικότητας. Πιο συγκεκριμένα, χρησιμοποιήθηκε η αποσύνθεση στον χώρο των κυματιδίων εξαιτίας της



Σχήμα 3.2: *(α) αρχική εικόνα, (β) εικόνα με προσθήκη Γκαουσσιανού θορύβου (σ = 35), αποτελέσματα για (γ) DWT-salienShrink, (δ) DWT-BiShrink, (ε) BayeShrink, (ζ) Donoho, (η) DTCWT-salienShrink, (θ) DTCWT-BiShrink και μεγεθυμένη περιοχή για (ι) DWT-BiShrink, (κ) DWT-salienShrink, (λ) DTCWT-BiShrink, (μ) DTCWT-salienShrink*

υπολογιστικής απλότητας και της ομοιότητας με την διεργασία κέντρου-περιφέρειας των Itti *et al.* (κάθε κλίμακα του μετασχηματισμού προκύπτει από διαφορές γειτονικών κλιμάκων). Τελικά οι παραγόμενοι ενδιάμεσοι χάρτες σημαντικότητας



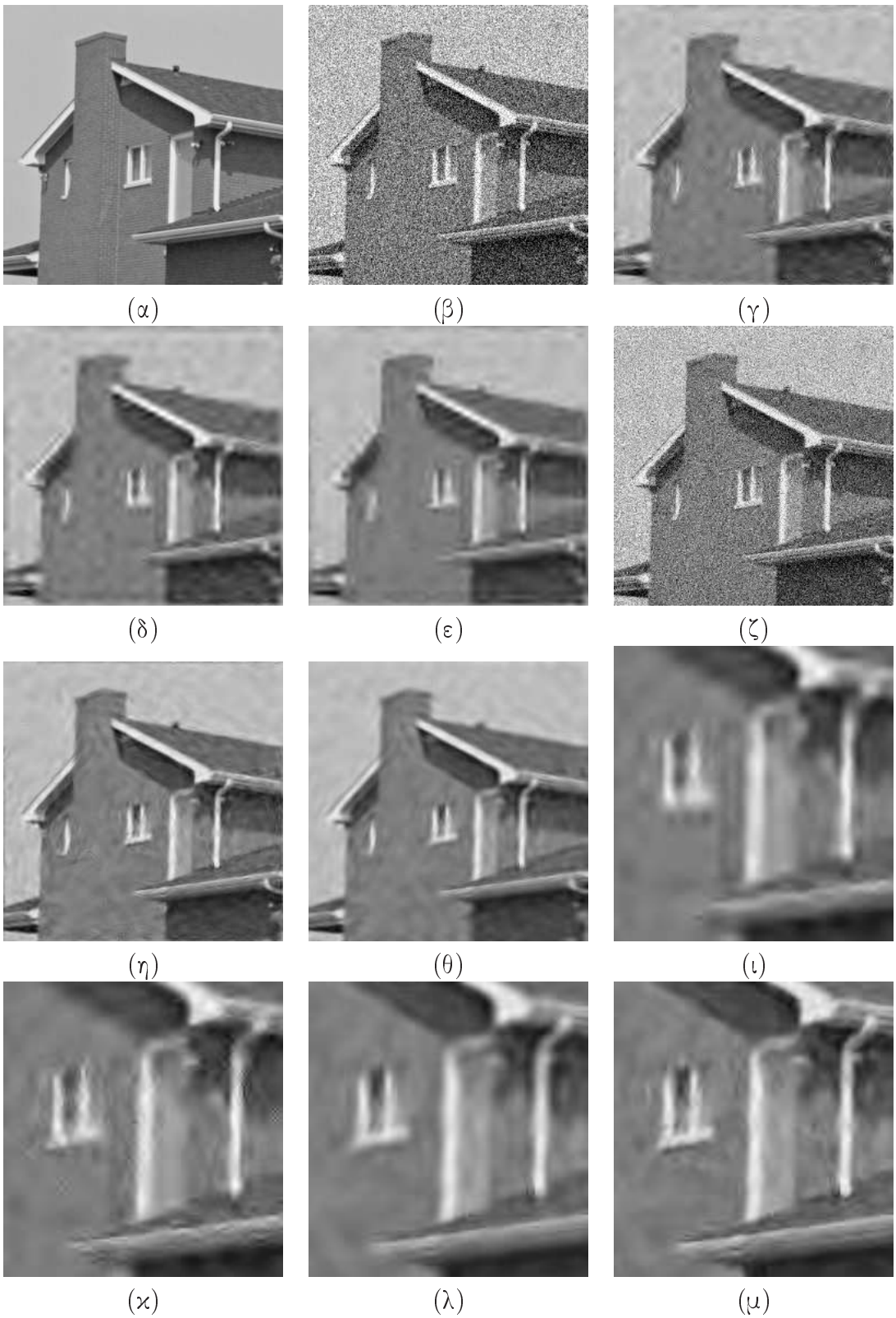
Σχήμα 3.3: *(α) αρχική εικόνα, (β) εικόνα με προσθήκη Γκαουσσιανού θορύβου (σ = 35), αποτελέσματα για (γ) DWT-salienShrink, (δ) DWT-BiShrink, (ε) BayeShrink, (ζ) Donoho, (η) DTCWT-salienShrink, (θ) DTCWT-BiShrink και μεγεθυμένη περιοχή για (ι) DWT-BiShrink, (κ) DWT-salienShrink, (λ) DTCWT-BiShrink, (μ) DTCWT-salienShrink*

συνδυάζονται για να προκύψει ο κύριος χάρτης. Η προτεινόμενη μέθοδος, με μικρές παραλλαγές από εφαρμογή σε εφαρμογή, χρησιμοποιήθηκε για ανίχνευση προσώπων σε περίπλοκα περιβάλλοντα (ανομοιόμορφος φωτισμός, θόρυβος κτλ) και για την

Σχήμα 3.4: *(α) αρχική εικόνα, (β) εικόνα με προσθήκη Γκαουσσιανού θορύβου (σ = 35), αποτελέσματα για (γ) DWT-salienShrink, (δ) DWT-BiShrink, (ε) BayeShrink, (ζ) Donoho, (η) DTCWT-salienShrink, (θ) DTCWT-BiShrink και μεγεθυμένη περιοχή για (ι) DWT-BiShrink, (κ) DWT-salienShrink, (λ) DTCWT-BiShrink, (μ) DTCWT-salienShrink*

βελτιωμένη συμπίεση βίντεο σε χαμηλούς ρυθμούς μετάδοσης. Η Ενότητα 3.3.4 περιέχει αναλυτικά τα αποτελέσματα της εφαρμογής του μοντέλου.

Η προτεινόμενη αρχιτεκτονική φαίνεται στο Σχήμα 3.6. Ο υπολογισμός του



Σχήμα 3.5: *(α) αρχική εικόνα, (β) εικόνα με προσθήκη Γκαουσσιανού θορύβου (σ = 35), αποτελέσματα για (γ) DWT-salienShrink, (δ) DWT-BiShrink, (ε) BayeShrink, (ζ) Donoho, (η) DTCWT-salienShrink, (θ) DTCWT-BiShrink και μεγεθυμένη περιοχή για (ι) DWT-BiShrink, (κ) DWT-salienShrink, (λ) DTCWT-BiShrink, (μ) DTCWT-salienShrink*

κύριου χάρτη βασίζεται τόσο στην ανοδική όσο και στην καθοδική διεργασία. Η

Πίνακας 3.2: *PSNR αποτελέσματα για την εικόνα "boat" (σ dB)*

	15	20	25	30	35
DWT-salienShrink	27.49	25.95	24.79	23.99	23.30
DWT-BiShrink	26.27	24.73	23.65	22.97	22.42
BayeShrink	24.54	23.08	22.46	21.97	21.50
Donoho	25.65	25.16	24.17	22.93	21.61
DTCWT-salienShrink	**28.62**	**27.23**	**26.33**	**25.44**	**24.76**
DTCWT-BiShrink	27.81	26.37	25.40	24.51	23.82

Πίνακας 3.3: *PSNR αποτελέσματα για την εικόνα "lena" (σ dB)*

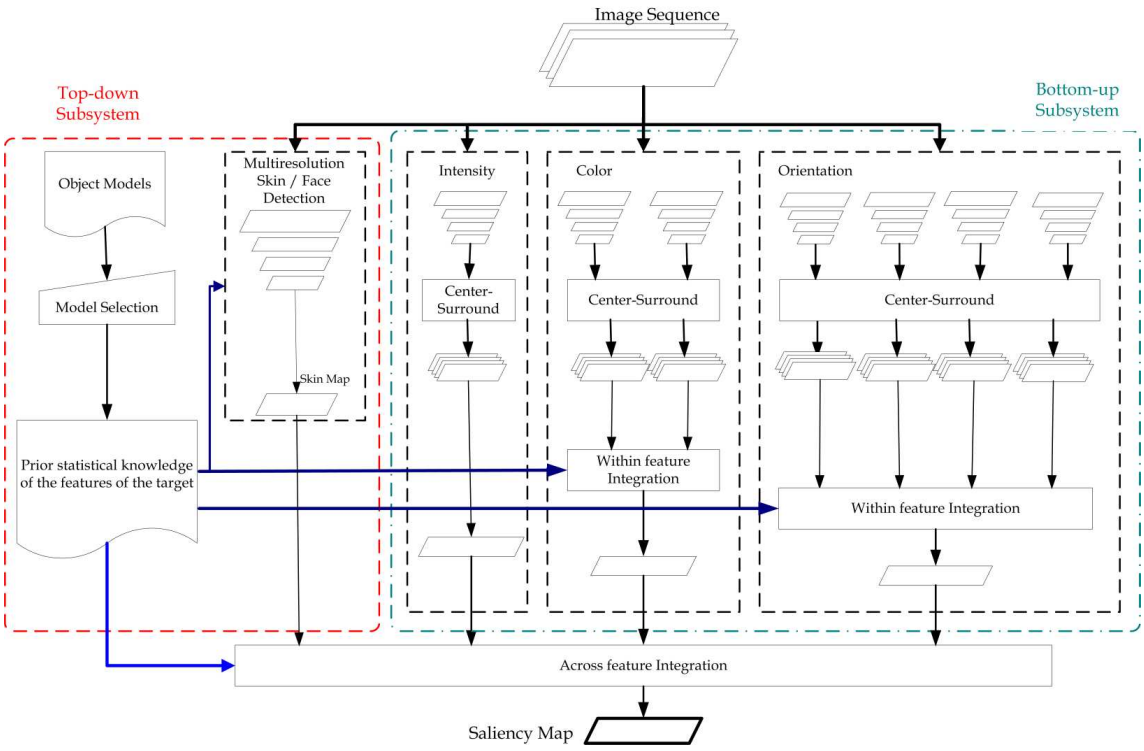
	15	20	25	30	35
DWT-salienShrink	29.19	27.42	26.31	25.30	24.47
DWT-BiShrink	27.85	26.04	24.92	24.10	23.45
BayeShrink	25.95	24.79	24.11	23.28	22.85
Donoho	27.77	26.29	24.47	22.68	21.08
DTCWT-salienShrink	**30.78**	**29.35**	**28.25**	**27.40**	**26.71**
DTCWT-BiShrink	30.14	28.65	27.49	26.54	25.83

Πίνακας 3.4: *PSNR αποτελέσματα για την εικόνα "cameraman" (σ dB)*

	15	20	25	30	35
DWT-salienShrink	28.96	27.15	25.73	24.54	23.52
DWT-BiShrink	28.05	26.28	24.80	23.59	22.52
BayeShrink	25.53	25.67	22.05	22.15	21.06
Donoho	27.58	25.54	23.66	22.02	20.61
DTCWT-salienShrink	**30.01**	**28.39**	**27.12**	**26.02**	**25.06**
DTCWT-BiShrink	29.43	27.84	26.56	25.48	24.50

υπόθεση μας είναι ότι η ακολουθία εισόδου αποτελείται από περιοχές ενδιαφέροντος και ασήμαντες περιοχές που μπορούν να θεωρηθούν ως υπόβαθρο. Ο ρόλος του καθοδικού καναλιού που φαίνεται αριστερά στο Σχήμα 3.6 είναι να επηρεάσει το μοντέλο προς περιοχές ενδιαφέροντος, οι οποίες μπορούν να μοντελοποιηθούν με στατιστικό τρόπο. Κάθε πληροφορία που μπορεί να αναπαρασταθεί με στατιστικό τρόπο μπορεί να αποτελέσει μέρος αυτού του δομοστοιχείου[6] της προτεινόμενης αρχιτεκτονικής.

Για την επίδειξη της προτεινόμενης μεθόδου χρησιμοποιούμε ένα στατιστικό μοντέλο αναπαράστασης ανθρώπινου δέρματος [29], το οποίο έχει εφαρμοστεί σε εφαρμογή εντοπισμού προσώπων. Εισάγουμε έναν επιπλέον ενδιάμεσο χάρτη σημαντικότητας, ο οποίος σχετίζεται με ανθρώπινο δέρμα και υπολογίζεται με βάση την χρωματική ομοιότητα των αντικειμένων με το ανθρώπινο δέρμα. Ο χάρτης δέρματος διαμορφώνεται μέσω πολλαπλασιασμού με έναν χάρτη υφής, ώστε να δοθεί έμφαση σε περιοχές που παρουσιάζουν κάποιου είδους δομή και ομοιότητα με δέρμα



Σχήμα 3.6: *Αρχιτεκτονική της προτεινόμενης επέκτασης του μοντέλου κύριου χάρτη*

αυξάνοντας έτσι την πιθανότητα να αντιστοιχούν σε πρόσωπο. Ο χάρτης υφής προκύπτει με φιλτράρισμα εύρους του χάρτη φωτεινότητας. Η αλληλεπίδραση μεταξύ των καθοδικών και ανοδικών υποσυστημάτων γίνεται τόσο στο επίπεδο των χαρτών (επίπεδο χαρακτηριστικών) όσο και στο επίπεδο της ενοποίησης των αποτελεσμάτων (Σχήμα 3.6, *within-feature* και *across-feature integration*).

Στην περίπτωση ενός σεναρίου εικονοτηλεφωνίας, το οποίο περιέχει ένα ή περισσότερα πρόσωπα, το προτεινόμενο μοντέλο ενεργοποιείται ως εξής: (α) Το μοντέλο ανθρώπινου δέρματος επιλέγεται ως το πιο κατάλληλο για αυτήν την εφαρμογή και η ακολουθία εισόδου αποσυντίθεται σε χάρτες χαρακτηριστικών (β) Κάθε χάρτης χαρακτηριστικών μετασχηματίζεται στον χώρο των κυματιδίων και εφαρμόζονται φίλτρα κέντρου-περιφέρειας. Τα φίλτρα αυτά εφαρμόζονται μεταξύ δύο γειτονικών κλιμάκων του μετασχηματισμού με σκοπό να τονιστούν οι περιοχές που διαφοροποιούνται από την γειτονιά τους. (γ) Οι ενδιάμεσοι χάρτες σημαντικότητας διαμορφώνονται από τα βάρη της καθοδικής διεργασίας σύμφωνα με το προ-επιλεγμένο μοντέλο και τελικά συνδυάζονται για να παραχθεί ο κύριος χάρτης.

3.3.1 Ανοδική διεργασία

Για την αναπαράσταση των χαρακτηριστικών στην ανοδική διεργασία χρησιμοποιήθηκε ο χώρος $YCbCr$ και η ανάλυση σε πολλές κλίμακες στον χώρο των κυματιδίων. Ο συγκεκριμένος χρωματικός χώρος χρησιμοποιήθηκε για να ταιριάζει η ανάλυση με το μοντέλο του ανθρώπινου δέρματος, αλλά και για να χρησιμοποιήσουμε το κανάλι Y για τον υπολογισμό του χάρτη φωτεινότητας και κατευθυντικότητας και τα κανάλια Cr και Cb για τον υπολογισμό των ενδιάμεσων χαρτών ενδιαφέροντος που σχετίζονται με το χρώμα. Χρησιμοποιήθηκε επίσης ένας χάρτης κίνησης που προκύπτει με την τεχνική υπολογισμού οπτικής ροής των Black και Anandan [5], η οποία βασίζεται

σε στιβαρή στατιστική με χρήση πολλαπλών κλιμάκων. Συνοπτικά, η οπτική ροή υπολογίζεται με ελαχιστοποίηση της εξίσωσης περιορισμού οπτικής ροής[11] υπό τον περιορισμό της χωρικής συνάφειας[12]. Πειράματα με αυτήν την τεχνική έχουν αποδείξει την ιδιότητα της να διατηρεί τις κινούμενες ακμές αποφεύγοντας φαινόμενα υπερεξομάλυνσης [106].

Υποθέτουμε μία έγχρωμη εικόνα f μετασχηματισμένη στον χώρο $YCbCr$. Στην προτεινόμενη υλοποίηση σημαντικές περιοχές με κριτήρια φωτεινότητας, κατευθυντικότητας και χρώματος υπολογίζονται σε διαφορετικές κλίμακες. Με αυτόν τον τρόπο εντοπίζονται εξέχουσες περιοχές διαφορετικών μεγεθών. Ο συνδυασμός των παραγόμενων χαρτών για κάθε χαρακτηριστικό παράγει τους ενδιάμεσους χάρτες σημαντικότητας για την φωτεινότητα (C_I), την κατευθυντικότητα (C_O), το χρώμα (C_C) και την κίνηση C_M. Για τον μετασχηματισμό στον χώρο των κυματιδίων εφαρμόζεται ένα βαθυπερατό h_ϕ και ένα υψιπερατό φίλτρο h_ψ σε κάθε κανάλι εισόδου προς την κάθετη και οριζόντια κατεύθυνση. Στην συνέχεια το αποτέλεσμα υποδειγματοληπτείται κατά δύο και τελικά παράγονται μία βαθυπερατή υποζώνη A (ζώνη προσέγγισης) και τρεις υψιπερατές ζώνες (ζώνες λεπτομέρειας), οι οποίες αναδεικνύουν τις οριζόντιες (H), κάθετες (V) και διαγώνιες ακμές (D) αντίστοιχα. Οι επόμενες κλίμακες του μετασχηματισμού προκύπτουν με επανάληψη της διαδικασίας στην υποζώνη A. Οι επόμενες εξισώσεις περιγράφουν την διαδικασία αποσύνθεσης του Y καναλιού. Με τον ίδιο τρόπο αποσυντίθενται και τα υπόλοιπα κανάλια.

$$\begin{aligned}
Y_A^{-(j+1)}(m,n) &= \{h_\phi(-m) * \{Y_A^{-j}(m,n) * h_\phi(-n)\} \downarrow^{2n}\} \downarrow^{2m} \\
Y_H^{-(j+1)}(m,n) &= \{h_\psi(-m) * \{Y_H^{-j}(m,n) * h_\phi(-n)\} \downarrow^{2n}\} \downarrow^{2m} \\
Y_V^{-(j+1)}(m,n) &= \{h_\phi(-m) * \{Y_V^{-j}(m,n) * h_\psi(-n)\} \downarrow^{2n}\} \downarrow^{2m} \\
Y_D^{-(j+1)}(m,n) &= \{h_\psi(-m) * \{Y_D^{-j}(m,n) * h_\psi(-n)\} \downarrow^{2n}\} \downarrow^{2m}
\end{aligned} \tag{3.7}$$

όπου $*$ είναι η συνέλιξη, $Y_A^{-j}(m,n)$ είναι η προσέγγιση του Y καναλιού στην κλίμακα j ($Y_A^{-0}(m,n) = Y$) και τα σύμβολα $\downarrow^{2n}$ και $\downarrow^{2m}$ υποδηλώνουν υποδειγματοληψία κατά δύο σε γραμμές και στήλες αντίστοιχα.

Μετά την αποσύνθεση κάθε καναλιού χρησιμοποιούμε διαφορές κέντρου-περιφέρειας για να ενισχύσουμε τις περιοχές που ξεχωρίζουν τοπικά από την γειτονιά τους. Οι διαφορές αυτές υπολογίζονται σε μία συγκεκριμένη κλίμακα j της ζώνης προσέγγισης ως η σημείο-προς-σημείο διαφορά από την αμέσως πιο αδρή κλίμακα $j+1$ της ίδιας ζώνης. Η διαφορά γίνεται εφικτή αφού πρώτα η αδρότερη κλίμακα $j+1$ παρεμβληθεί στο μέγεθος της πιο λεπτομερούς j. Οι εξισώσεις που ακολουθούν περιγράφουν αυτήν την διαδικασία για όλα τα εμπλεκόμενα κανάλια

$$\begin{aligned}
I^{-j} &= |Y_A^{-j}(m,n) - ((Y_A^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)| \\
C^{-j} &= w_{C_r} C_r^{-j} + w_{C_b} C_b^{-j} \\
O^{-j} &= w_{Y_D} |Y_D^{-j} - \hat{Y}_D^{-j}| + w_{Y_V} |Y_V^{-j} - \hat{Y}_V^{-j}| + w_{Y_H} |Y_H^{-j} - \hat{Y}_H^{-j}|
\end{aligned} \tag{3.8}$$

όπου

$$\begin{aligned}
C_r^{-j} &= |C_{r_A}^{-j}(m,n) - ((C_{r_A}^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)| \\
C_b^{-j} &= |C_{b_A}^{-j}(m,n) - ((C_{b_A}^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)| \\
\hat{Y}_D^{-j} &= ((Y_D^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)| \\
\hat{Y}_V^{-j} &= ((Y_V^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)| \\
\hat{Y}_H^{-j} &= ((Y_H^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n)|
\end{aligned} \tag{3.9}$$

Στις προηγούμενες εξισώσεις τα σύμβολα I^{-j}, O^{-j}, C^{-j}, αντιστοιχούν στους χάρτες χαρακτηριστικών που υπολογίζονται στην κλίμακα j, τα $C_{r_A}^{-j}$, $C_{b_A}^{-j}$ είναι οι ζώνες προσέγγισης των χρωματικών καναλιών Cr καιCb στην κλίμακα j, τα $\uparrow^{2m}$ και $\uparrow^{2n}$ συμβολίζουν πλειοδειγματοληψία σε γραμμές και στήλες, ενώ τα $\hat{Y}_A^{-j}$, $\hat{Y}_D^{-j}$, $\hat{Y}_H^{-j}$, $\hat{Y}_V^{-j}$ είναι οι πλειοδειγματοληπτημένες προσεγγίσεις των $Y_A^{-(j+1)}$, $Y_D^{-(j+1)}$, $Y_V^{-(j+1)}$ και $Y_H^{-(j+1)}$. Τα βάρη $\{w_{C_r}, w_{C_b}, w_{Y_A}, w_{Y_D}, w_{Y_V}, w_{Y_D}\}$ προκύπτουν από την καθοδική διεργασία. Στην περίπτωση απουσίας αυτής της διεργασίας τα βάρη μπορούν να είναι ίσα, αλλά θετικά και με άθροισμα ίσο με την μονάδα

Οι ενδιάμεσοι χάρτες ενδιαφέροντος προκύπτουν συνδυάζοντας τους χάρτες σε ποικίλες κλίμακες, έτσι ώστε να ανιχνευθούν σημαντικές περιοχές που συνολικά εξέχουν από την γειτονιά τους. Ο συνδυασμός γίνεται με παρεμβολή όλων των ενδιάμεσων χαρτών στην λεπτομερέστερη κλίμακα του μετασχηματισμού κυματιδίων, σημείου-προς-σημείου πρόσθεση και εφαρμογή συνάρτησης κορεσμού (σιγμοειδής) στο τελικό αποτέλεσμα. Η διαδικασία για τον ενδιάμεσο χάρτη σημαντικότητας C_I περιγράφεται στις εξισώσεις που ακολουθούν.

$$\begin{aligned} C_I &= \frac{2}{1+e^{-(\sum_{j=-J_{max}}^{-1} C_I^j)}} - 1 \\ C_I^{-j} &= I^{-j}(m,n) - ((C_I^{-(j+1)}(m,n) \uparrow^{2m}) * h_\phi(m)) \uparrow^{2n} * h_\phi(n) \\ C_I^{-J_{max}} &= I^{-J_{max}} \end{aligned} \tag{3.10}$$

Το μέγιστο βάθος αποσύνθεσης J_{max} υπολογίζεται ως

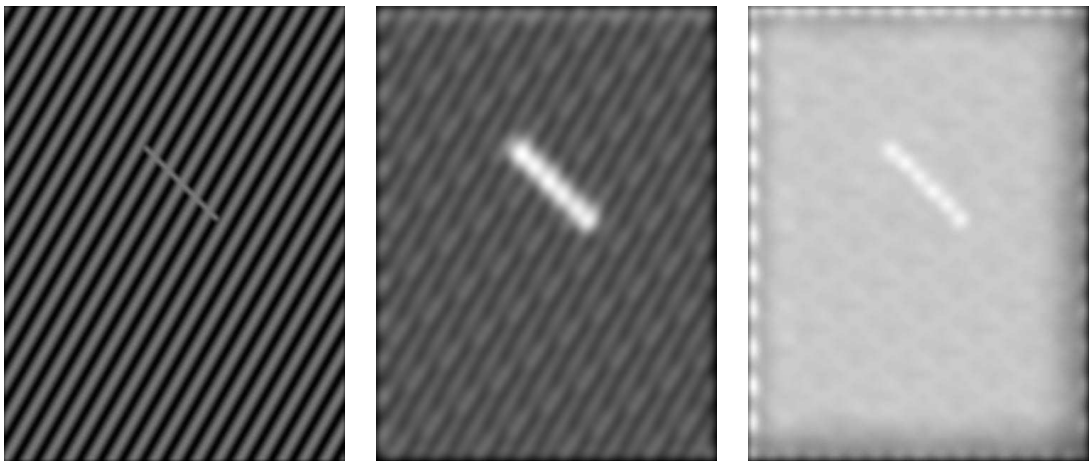
$$J_{max} = \lfloor \frac{log_2 N}{2} \rfloor, \quad N = min(R, C) \tag{3.11}$$

όπου στην συνάρτηση $y = \lfloor x \rfloor$ η y είναι η μεγαλύτερη ακέραια τιμή για την οποία $x \geq y$, και R, C είναι ο αριθμός των γραμμών και στηλών της εικόνας εισόδου αντίστοιχα.

Για να αναδείξουμε την σημασία του καναλιού κατευθυντικότητας δημιουργήσαμε μία σύνθετη εικόνα στην οποία απεικονίζεται ένα αντικείμενο που ξεχωρίζει από την γειτονιά του με βάση την κατευθυντικότητα (Σχήμα 3.7α). Ο ενδιάμεσος χάρτης κατευθυντικότητας (Σχήμα 3.7β) ενισχύει το αντικείμενο σε αντίθεση με τον ενδιάμεσο χάρτη φωτεινότητας (Σχήμα 3.7γ), ο οποίος είναι θορυβώδης μια και δεν υπάρχουν αντικείμενα στην εικόνα που να ξεχωρίζουν με βάση την φωτεινότητα.

3.3.2 Καθοδική διεργασία

Ο επιτυχημένος συνδυασμός των χαρτών χαρακτηριστικών είναι σημαντικός για την ανάδειξη του αντικείμενου ενδιαφέροντος (π.χ. πρόσωπα). Οι χάρτες είναι ανομοιογενείς και δεν είναι εύκολο να συνδυαστούν. Προτείνουμε την χρήση ενός committee machine για την αποτελεσματική αντιμετώπιση του προβλήματος, ο οποίος απεικονίζεται στο Σχήμα 3.8 [101, 102]. Η είσοδος αποτελείται από τους χάρτες χαρακτηριστικών,μία τιμή εμπιστοσύνης, της οποίας τον υπολογισμό θα αναλύσουμε παρακάτω, και την μοντελοποίηση του περίγυρου, όπως φαίνεται στο Σχήμα 3.8. Η είσοδος gating μοντελοποιεί τον περίγυρο με τον εξής τρόπο: Στις περιπτώσεις που είναι εκ των προτέρων γνωστή η ύπαρξη ανθρώπων στη σκηνή, όπως π.χ. σε ακολουθίες με πρόσωπα, τηλεοπτικές ειδήσεις κτλ, αυτή η πληροφορία οδηγεί το σύστημα στο να δώσει προτεραιότητα στα κανάλια εντοπισμού δέρματος και κίνησης. Ομοίως σε περίπτωση στατικού περιβάλλοντος, στα οποία



(α) (β) (γ)

Σχήμα 3.7: *Η σημασία του καναλιού κατευθυντικότητας. (α) Αρχική εικόνα, (β) Ενδιάμεσος χάρτης κατευθυντικότητας, (γ) Ενδιάμεσος χάρτης φωτεινότητας*

ίσως δεν υπάρχουν άνθρωποι, το committee machine δίνει προτεραιότητα στα στατικά κανάλια (φωτεινότητα, χρώμα, κατευθυντικότητα). Τέλος, στην περίπτωση που δεν υπάρχει εκ των προτέρων γνώση η είσοδος gating αποτελείται από έναν αριθμό προ-υπολογισμένων βαρών για κάθε χάρτη.

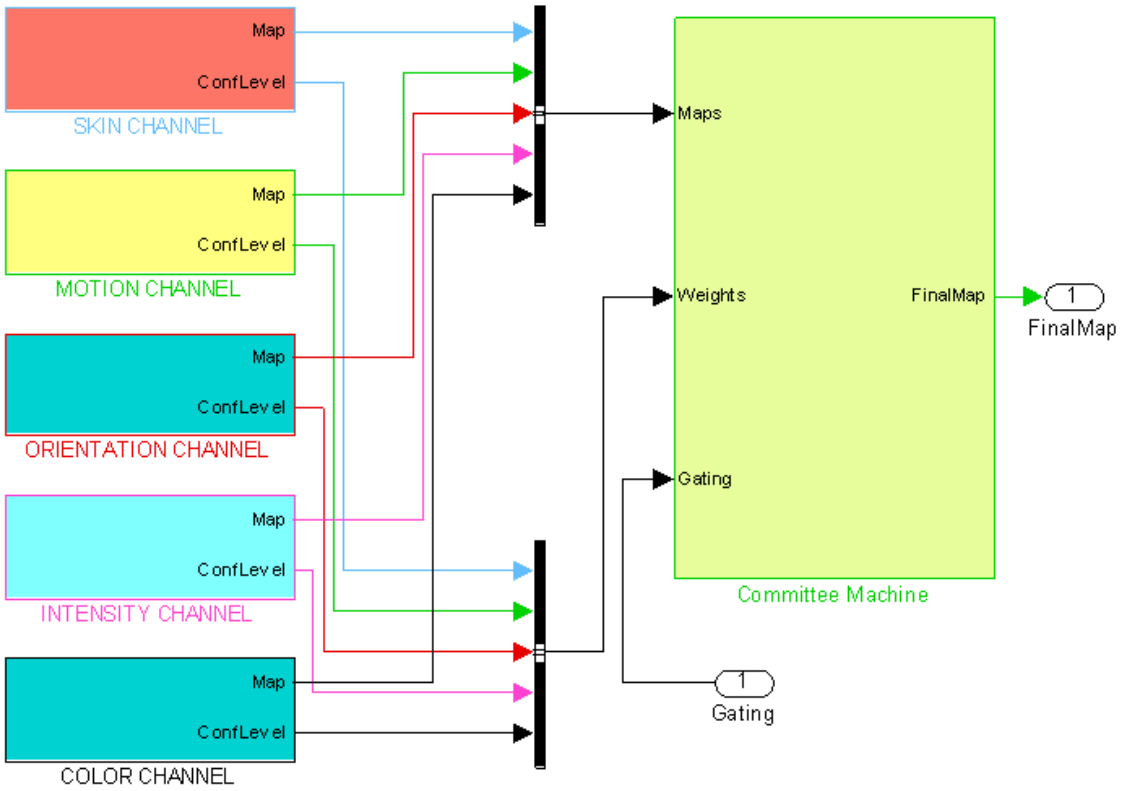
Τα επίπεδα εμπιστοσύνης είναι αυτά που καθορίζουν το ποιοι χάρτες θα συνεισφέρουν τελικά στην λειτουργία της μηχανής και με ποιο βάρος. Ακολουθώντας αυτήν την λογική, η διαδικασία είναι η εξής: Αν I είναι το τρέχον καρέ της ακολουθίας, g_i είναι η είσοδος του gating για τον i χάρτη, $y_i = f_i(I)$ είναι ο i χάρτης και $c_i(I)$ είναι το αντίστοιχο επίπεδο εμπιστοσύνης, τότε ο κύριος χάρτης υπολογίζεται ως:

$$F(I) = \frac{\sum_i^{N_c} g_i \cdot c_i(f) \cdot f_i(I)}{\sum_i^{N_c} g_i \cdot c_i(I)} \tag{3.12}$$

όπου N_c είναι ο αριθμός των καναλιών της αρχιτεκτονικής (στην περίπτωση μας έχουμε 5 κανάλια). Από την προηγούμενη εξίσωση προκύπτει ότι ο αριθμός των χαρτών που συνεισφέρει σε κάθε χρονική στιγμή (καρέ) αλλάζει ακολουθώντας πιθανές αλλαγές της ακολουθίας (π.χ. αλλαγές φωτισμού, απότομες κινήσεις). Σύμφωνα με την (3.12) τα επίπεδα εμπιστοσύνης των αρχικών χαρτών πολλαπλασιάζονται με τα gating βάρη των αντίστοιχων χαρτών. Τα επίπεδα εμπιστοσύνης υπολογίζονται για κάθε κανάλι ξεχωριστά και βασίζονται στην αυτοσυσχέτιση των χαρτών. Αυτό αποσκοπεί στην αντιστάθμιση γρήγορων ή απότομων αλλαγών στην είσοδο (συνθήκες φωτισμού, απότομη κίνηση, θόρυβος λόγω συμπίεσης) από τον προηγούμενο στον τρέχον χάρτη χαρακτηριστικών. Ο συντελεστής συσχέτισης μεταξύ δύο χρονικά γειτονικών καρέ υπολογίζεται και λειτουργεί ως ένας τρόπος καθορισμού της συνεισφοράς κάθε χάρτη. Αν, π.χ. η κίνηση είναι σταθερή μεταξύ των καρέ ο συντελεστής συσχέτισης θα είναι υψηλός, αν όμως αλλάξει ξαφνικά ο συντελεστής θα είναι χαμηλός και θα μειωθεί σημαντικά η συνεισφορά του χάρτη κίνησης. Ο συντελεστής $c_i(I)$ υπολογίζεται ως:

$$c_i(I) = \frac{\sum(f_i(I-1) \cap f_i(I))}{\sum(f_i(I-1) \cap + f_i(I) - f_i(I-1) \cap f_i(I))} \tag{3.13}$$

όπου $\cap$ είναι ο τελεστής AND (minimum) και το άθροισμα υπολογίζεται για όλες τις τιμές του i χάρτη.



Σχήμα 3.8: *Σχεδιάγραμμα του committee machine που χρησιμοποιήθηκε.*

3.3.3 Συνδυασμός ενδιάμεσων χαρτών ενδιαφέροντος

Ο τελικός κύριος χάρτης προκύπτει με συνδυασμό των ενδιάμεσων χαρτών ως αποτέλεσμα της καθοδικής και της ανοδικής διεργασίας του μοντέλου. Όπως και στην Ενότητα 3.3.1, χρησιμοποιούμε μία σιγμοειδή συνάρτηση για τον συνδυασμό των χαρτών. Η απλή κανονικοποίηση και πρόσθεση, όπως προτείνουν οι Itti *et al.* [48], παράγει συνήθως θορυβώδη και ανακριβή αποτελέσματα σε περιπτώσεις που μία περιοχή είναι σημαντική σε έναν και μόνο χάρτη. Για παράδειγμα, στο Σχήμα 3.7 περιμένουμε ότι μόνο το κανάλι κατευθυντικότητας θα παράγει μία σημαντική περιοχή. Αν απλά υπολογίσουμε τον μέσο όρο όλων των ενδιάμεσων χαρτών, τότε θα μειωθεί κατά πολύ η σημασία της ουσιαστικής περιοχής ενδιαφέροντος. Στο προτεινόμενο μοντέλο χρησιμοποιούμε την συνάρτηση κορεσμού για να διατηρήσουμε την συνεισφορά κάθε καναλιού ανέπαφη:

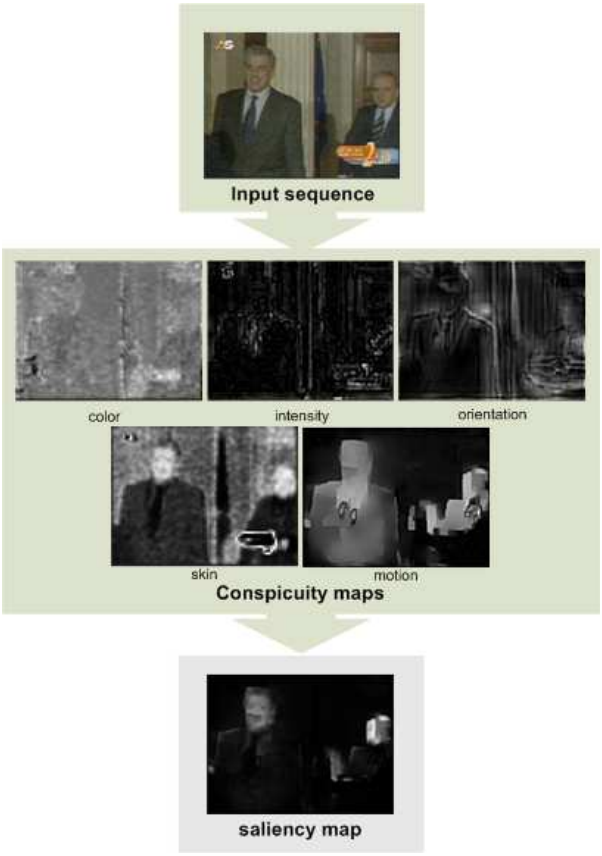
$$S = \frac{2}{1 + e^{-(C_I + C_O + C_C + C_F)}} - 1 \tag{3.14}$$

όπου C_F είναι ο ενδιάμεσος χάρτης δέρματος, ενώ S είναι ο τελικός κύριος χάρτης.

3.3.4 Εφαρμογές

3.3.4.1 Ανίχνευση προσώπων σε περίπλοκες συνθήκες

Ο εντοπισμός προσώπων σε δύσκολες συνθήκες είναι η πρώτη εφαρμογή που επιλέχτηκε για την αξιολόγηση του προτεινόμενου μοντέλου. Χρησιμοποιήθηκαν ακολουθίες από τηλεόραση και από ιδιωτικές λήψεις. Η έλλειψη εικόνων επαλήθευσης

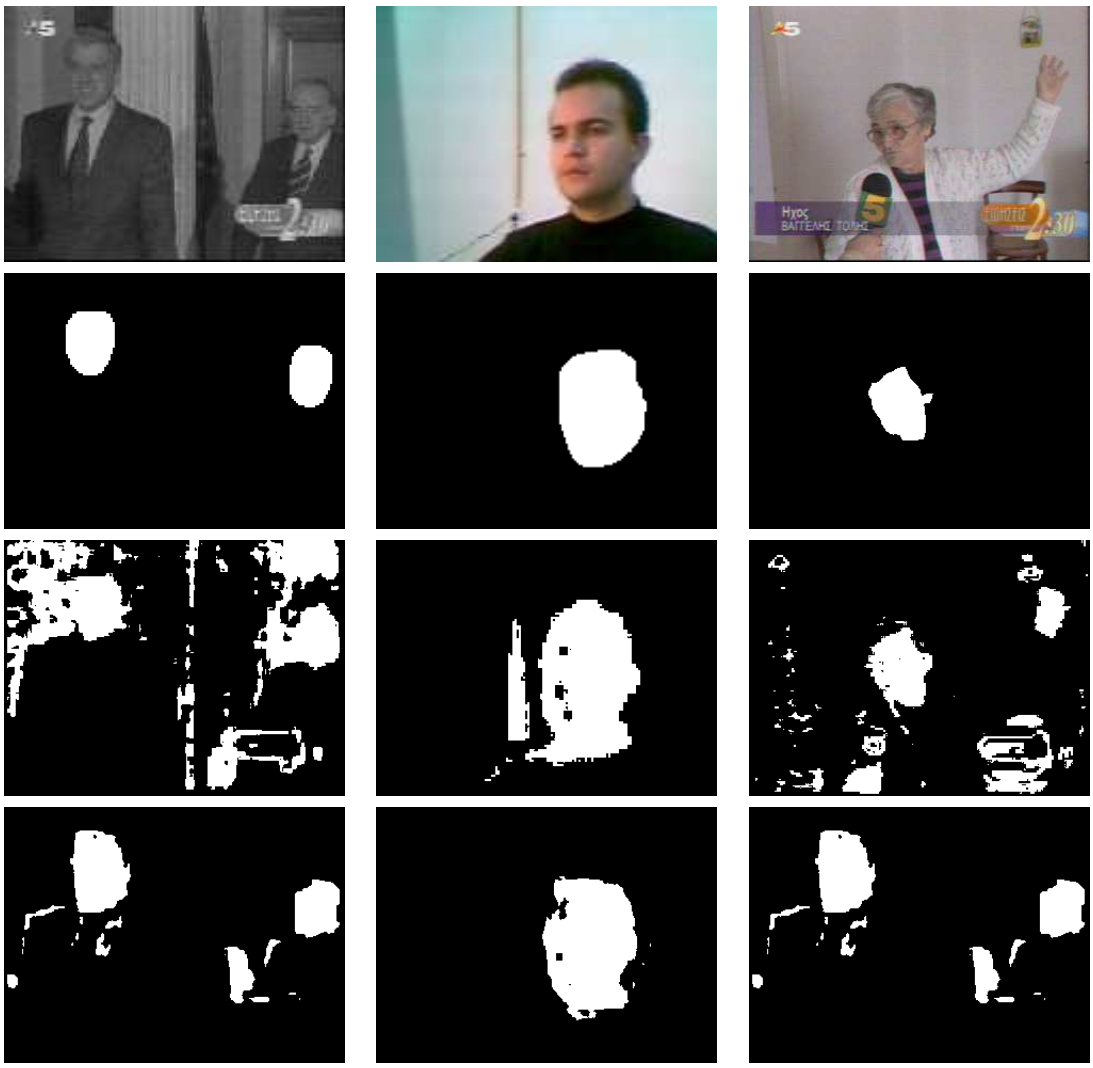


Σχήμα 3.9: *Παράδειγμα εξαγωγής κύριου χάρτη με χρήση της προτεινόμενης τεχνικής.*

σε επίπεδο δυαδικής μάσκας για κάθε πρόσωπο σε ένα σύνολο ακολουθιών μας απέτρεψε από την διεξαγωγή πειραμάτων μεγάλης κλίμακας. Με χειρωνακτικό τρόπο δημιουργήσαμε μάσκες που αντιστοιχούν σε όλα τα πρόσωπα για κάθε καρέ.

Το Σχήμα 3.9 απεικονίζει ένα ενδεικτικό παράδειγμα των χαρτών που προκύπτουν όταν το προτεινόμενο μοντέλο εφαρμόζεται σε μία τηλεοπτική ακολουθία που ονομάζουμε "twoFaces". Το Σχήμα 3.10 δείχνει περισσότερα αποτελέσματα της προτεινόμενης τεχνικής σε τρεις ακολουθίες, τις αντίστοιχες μάσκες επαλήθευσης, καθώς και τα αποτελέσματα κατωφλίωσης με μία κλασσική τεχνική του χώρου (κατωφλίωση ελάχιστης διασποράς [85]). Το κατώφλι υπολογίζεται στον κύριο χάρτη και στον χάρτη εντοπισμού δέρματος αντίστοιχα. Με βάση τις δυαδικές μάσκες που προκύπτουν υπολογίζουμε τις τιμές ακρίβειας[7] και επανάκλησης[8] για όλη την ακολουθία. Ο Πίνακας 3.5 παρουσιάζει αποτελέσματα για τρεις ενδεικτικές ακολουθίες.

Η ακολουθία "twoFaces" έχει μαγνητοσκοπηθεί με ακίνητη κάμερα. Οι κινήσεις των δυο ανθρώπων είναι μικρές, αλλά η ποιότητα είναι χαμηλή. Η ύπαρξη περιοχών που μοιάζουν χρωματικά με δέρμα αποτελεί πρόβλημα για μία απλή τεχνική εντοπισμού δέρματος. Επιπρόσθετα, ο χαμηλός ρυθμός καρέ δημιουργεί αναπόφευκτα προβλήματα στον αλγόριθμο υπολογισμού οπτικής ροής. Η προτεινόμενη τεχνική αντισταθμίζει τα προβλήματα που εντοπίζονται στους αντίστοιχους χάρτες και μειώνει κυρίως την συνεισφορά του χάρτη κίνησης λόγω της χαμηλής συσχέτισης που παρουσιάζει από καρέ σε καρέ. Η δεύτερη ακολουθία, "myFace", μαγνητοσκοπήθηκε με στατική κάμερα και δείχνει ένα πρόσωπο που κινείται κάτω από συνθήκες έντονα ανομοιόμορφου φωτισμού. Η τεχνική που βασίζεται στην οπτική προσοχή έχει σαφώς



Σχήμα 3.10: *Ενδεικτικά αποτελέσματα της προτεινόμενης τεχνικής. Ανά σειρά: Αρχικές εικόνες, μάσκες επαλήθευσης, αποτέλεσμα κατάτμησης χάρτη εντοπισμού δέρματος, αποτέλεσμα κατάτμησης αποτελεσμάτων της προτεινόμενης τεχνικής.*

Πίνακας 3.5: *Στατιστικές μετρήσεις για τις LISBON ακολουθίες*

Video Seq.	Method	Mean Precision (%)	Mean Recall (%)
twoFaces	VA	78.9	80.5
	Simple Skin	71.6	73.4
myFace	VA	80.4	79.8
	Simple Skin	73.3	73.0
grandma	VA	55.9	56.3
	Simple Skin	49.5	51.5

καλύτερη απόδοση και δεν επηρεάζεται από τον φωτισμό και τα αντικείμενα που έχουν χρωματική ομοιότητα με δέρμα, όπως η σωλήνωση που διακρίνεται στο βάθος. Μία δύσκολη περίπτωση αποτελεί η ακολουθία "grandma", η οποία χαρακτηρίζεται από κινούμενη κάμερα και πολύ θόρυβο. Σε αυτήν την περίπτωση οι τιμές της ακρίβειας και επανάκλησης είναι παρόμοιες για τις δύο συγκρινόμενες τεχνικές.

3.3.4.2 Κωδικοποίηση περιοχών ενδιαφέροντος με χρήση του κύριου χάρτη

Οι οπτικά ενδιαφέρουσες περιοχές, όπως προκύπτουν από το προτεινόμενο μοντέλο, μπορούν να χρησιμοποιηθούν για την αποτελεσματικότερη κωδικοποίηση ακολουθιών. Για αυτόν τον σκοπό θεωρούμε σαν περιοχές ενδιαφέροντος το αποτέλεσμα της κατωφλίωσης του κύριου χάρτη για κάθε καρέ. Οι περιοχές που δεν θεωρούνται σημαντικές εξομαλύνονται με ένα κατάλληλο φίλτρο με σκοπό να επιτευχθεί μεγαλύτερη συμπίεση σε αυτές λόγω της αραιής δομής που περιέχουν.

Η υπόθεση πίσω από την εξομάλυνση των μη σημαντικών περιοχών σχετίζεται και με το πείραμα που εκτελέσαμε, στο οποίο δείχνουμε τις συμπιεσμένες εικόνες σε ένα σύνολο παρατηρητών μετρώντας την ικανοποίηση τους. Ο παρατηρητής, στον περιορισμένο χρόνο που διαθέτει βλέποντας ένα καρέ, σχεδόν αυτόματα εστιάζει στις σημαντικές περιοχές δίνοντας μικρότερη σημασία στις υπόλοιπες "εξομαλύνοντας" τις. Η εξομάλυνση των περιοχών δεν είναι ιδανική σε σχέση με το βαθμό συμπίεσης, αλλά είναι συμβατή με υπάρχοντες κωδικοποιητές. Η ποιότητα της κωδικοποίησης ακολουθιών με την προτεινόμενη μέθοδο, η οποία στα σχήματα φαίνεται ως *VA-ROI* (Visual Attention-based ROI), αξιολογήθηκε με ένα σύνολο οπτικών δοκιμών για δέκα ακολουθίες (*fashion, eye_witness, grandma, justice, news_cast1, news_cast2, lecturer, night_interview, old_man, soldier*). Στους Πίνακες 3.8 και 3.9 παρουσιάζονται τα συνολικά αποτελέσματα για κωδικοποίηση κατά *MPEG-1*και *MPEG-4* αντίστοιχα. Συγκρίνουμε επίσης την προτεινόμενη τεχνική με αυτή των Itti *et al.*, η οποία καλείται *Itti-ROI*, και την συνολική κωδικοποιήση (χωρίς περιοχές ενδιαφέροντος) κατά *MPEG-1* και *MPEG-4*.

Στην περίπτωση της συμπίεσης κατά *MPEG-1* χρησιμοποιήσαμε κωδικοποίηση μεταβλητού δυφιακού ρυθμού[13] (VBR) με μέγεθος καρέ 288×352, ρυθμό καρέ 25fps και δομή GOP: *IBBPBBPBBPBB*. Στην περίπτωση της συμπίεσης κατά *MPEG-4* χρησιμοποιήθηκε επίσης VBR, με μέγεθος καρέ 144×176 και ρυθμό καρέ 15fps. Η κωδικοποίηση των ακολουθιών έγινε με τον κωδικοποιητή *ImTOO MPEG Encoder* (*http://www.imtoo.com/*). Η πρώτη αντιστοιχεί στην προτεινόμενη μέθοδο (*VA-ROI*), η δεύτερη στην *Itti-ROI* και η τρίτη στην *Std MPEG-1* ή *Std MPEG-4* αντίστοιχα. Στις δύο περιπτώσεις των μεθόδων με χρήση οπτικής προσοχής, οι περιοχές που επιλέγονται ως μη ενδιαφέρουσες εξομαλύνονται πριν κωδικοποιηθούν.

Πίνακας 3.6: *Συνολικές προτιμήσεις (ανεξαρτήτως ακολουθίας) για κωδικοποίηση MPEG-1*

Encoding method	**Preferences**	**Average Bit Rate** (Kbps)
VA-ROI	91	1125
Itti-ROI	16	1081
Standard MPEG-1	93	1527

Πίνακας 3.7: *Συνολικές προτιμήσεις (ανεξαρτήτως ακολουθίας) για κωδικοποίηση MPEG-4*

Encoding method	**Preferences**	**Average Bit Rate** (Kbps)
VA-ROI	79	197.5
Itti-ROI	42	194.0
Standard MPEG-4	79	224.6

Ανάλυση οπτικών δοκιμών και αποτελέσματα κωδικοποίησης Για την αξιολόγηση της οπτικής ποιότητας της κωδικοποίησης με τα μοντέλα οπτικής προσοχής που συγκρίνουμε διεξήχθη ένα πείραμα οπτικών δοκιμών. Οι περιοχές ενδιαφέροντος για την μέθοδο *VA-ROI* ανιχνεύθηκαν με το προτεινόμενο μοντέλο, ενώ για την μέθοδο *Itti-ROI* με το Neuromorphic Vision Toolkit (*http://ilab.usc.edu/toolkit/*). Και στις δύο περιπτώσεις οι κύριοι χάρτες κατωφλιώθηκαν με την τεχνική του Otsu [85] για να προκύψουν οι δυαδικές μάσκες των περιοχών. Η πειραματική διαδικασία που ακολουθήσαμε ήταν η εξής: ο κάθε παρατηρητής βλέπει τρεις διαφορετικές εκδόσεις της ακολουθίας (*VA-ROI, Itti-ROI, MPEG-based*) και επιλέγει αυτήν που προτιμά. Το πείραμα έγινε με δέκα φοιτητές και καθεμία από τις δέκα ακολουθίες προβλήθηκε δύο φορές ($10 \times 10 \times 2 = 200$ συγκρίσεις). Οι ακολουθίες προβλήθηκαν στην οθόνη ενός Smartphone στην περίπτωση της MPEG-4 κωδικοποίησης και σε μία τυπική οθόνη PC στην περίπτωση της κωδικοποίησης κατά MPEG-1.

Ο Πίνακας 3.6 περιέχει τις προτιμήσεις και τον μέσο δυφιακό ρυθμό[14] για την περίπτωση κωδικοποίησης κατά MPEG-1. Μία ελαφριά προτίμηση (46.5%) προς την κλασσική κωδικοποίηση MPEG-1 προκύπτει από τα αποτελέσματα. Επομένως, η διαφορά της VA-ROI σε οπτική ποιότητα δεν είναι σημαντική, αλλά το αντίθετο συμβαίνει με το κέρδος σε δυφιακό ρυθμό (36% κατά μέσο όρο). Οι ακολουθίες που προέκυψαν με την Itti-ROI επιλέχθηκαν μόνο στο 8% των περιπτώσεων. Ο ελαφρά καλύτερος μέσος όρος δυφιακού ρυθμού δεν είναι αρκετός για να δικαιολογήσει τις χαμηλές προτιμήσεις. Στον Πίνακα 3.7 παρουσιάζονται τα αντίστοιχα αποτελέσματα για την περίπτωση κωδικοποίησης κατά MPEG-4. Γενικά, το μέσο κέρδος σε δυφιακό ρυθμό είναι παρόμοιο για τις δύο μεθόδους, αλλά η προτίμηση στις ακολουθίες με VA-ROI είναι σαφής. Αναλυτικά αποτελέσματα για τις δέκα ακολουθίες περιέχονται στους Πίνακες 3.8 και 3.9.

3.4 Συμπεράσματα

Σε αυτό το κεφάλαιο περιγράψαμε και αξιολογήσαμε τις προτάσεις μας για χρήση και επέκταση δύο σημαντικών μοντέλων στον χώρο, το μοντέλο επιλεκτικού συντονισμού των Tsotsos *et al.* [137] και των Itti και Koch [48]. Χρησιμοποιήσαμε ένα απλοποιημένο μοντέλο επιλεκτικού συντονισμού για να βελτιώσουμε καθιερωμένες τεχνικές αποθορυβοποίησης που βασίζονται στον μετασχηματισμό κυματιδίων. Ένας κύριος χάρτης υπολογίζεται για κάθε υποζώνη του μετασχηματισμού και ένα δίκτυο όλα-για-το-νικητή επιλέγει τις πιο σημαντικές περιοχές. Τα στατιστικά αυτών των περιοχών βοηθούν στον ακριβέστερο υπολογισμό της μεταβλητότητας του θορύβου και κατά συνέπεια οδηγούν σε βελτιωμένα ποιοτικά και ποσοτικά αποτελέσματα σε αυτήν την εφαρμογή.

Στο δεύτερο μέρος του κεφαλαίου προτείναμε επεκτάσεις του μοντέλου των Itti *et al.* στο ανοδικό κανάλι και εισαγάγαμε ένα νέο καθοδικό κανάλι. Συγκεκριμένα, επεκτείναμε το βασικό τους μοντέλο με έναν χάρτη κίνησης, έναν χάρτη ανθρώπινου δέρματος, ο οποίος ταιριάζει με την εφαρμογή που επιλέξαμε, και με έναν τρόπο ελέγχου της συνεισφοράς κάθε χάρτη ανάλογα με τον περιβάλλον. Επιλέξαμε δύο εφαρμογές, την ανίχνευση προσώπων σε περίπλοκα περιβάλλοντα και την αντιληπτική κωδικοποίηση ακολουθιών, και παρουσιάσαμε αναλυτικά αποτελέσματα που δείχνουν την ποσοτική και ποιοτική βελτίωση που επιτυγχάνουμε.

Στην ανίχνευση προσώπων σε περίπλοκα περιβάλλοντα φαίνεται η σημασία του committee machine στην επιτυχία της μεθόδου, η οποία στην πλειοψηφία

Πίνακας 3.8: *Σύγκριση μεταξύ των μεθόδων VA-ROI, IttiROI και MPEG-1 σε 10 ακολουθίες*

Video Clip	**Encoding method**	**Bit Rate** (Kbps)	**Bit Rate Gain**
Eye_witness	VA-ROI	1610	30.5%
	Itti-ROI	1585	32.6%
	Standard MPEG-1	2101	
fashion	VA-ROI	1200	31.1%
	Itti-ROI	1188	32.4%
	Standard MPEG-1	1573	
grandma	VA-ROI	1507	15.2%
	Itti-ROI	1300	33.6%
	Standard MPEG-1	1737	
justice	VA-ROI	1468	30.5%
	Itti-ROI	1606	19.3%
	Standard MPEG-1	1916	
lecturer	VA-ROI	950	57.3%
	Itti-ROI	848	76.3%
	Standard MPEG-1	1495	
news_cast1	VA-ROI	991	39.7%
	Itti-ROI	999	38.5%
	Standard MPEG-1	1384	
news_cast2	VA-ROI	930	41.9%
	Itti-ROI	836	57.8%
	Standard MPEG-1	1319	
night_interview	VA-ROI	790	50.8%
	Itti-ROI	750	59.0%
	Standard MPEG-1	1192	
old_man	VA-ROI	1307	32.9%
	Itti-ROI	1085	60.0%
	Standard MPEG-1	1737	
soldier	VA-ROI	500	63.9%
	Itti-ROI	614	33.3%
	Standard MPEG-1	819	
Average	VA-ROI	1125	35.7%
	Itti-ROI	1081	41.2%
	Standard MPEG-1	1527	

των περιπτώσεων έχει πολύ καλά αποτελέσματα. Στην κωδικοποίηση ακολουθιών αξιολογήσαμε την δυνατότητα του προτεινόμενου μοντέλου να παράγει αντιληπτικά αποδεκτές περιοχών ενδιαφέροντος, έτσι ώστε η ισορροπία μεταξύ του κέρδους σε συμπίεση και οπτικής ποιότητας να είναι ικανοποιητική. Η απόδοση αξιολογήθηκε ποσοτικά και ποιοτικά. Τα αποτελέσματα που παρουσιάστηκαν για την προτεινόμενη μέθοδο αποδεικνύουν ότι: (α) μπορεί να επιτευχθεί σημαντικό κέρδος σε δυφιακό ρυθμό σε σχέση με τους *MPEG-1* και *MPEG-4* αλγορίθμους, (β) οι περιοχές ενδιαφέροντος είναι συμβατές με αυτές που θεωρούν σημαντικές οι ανθρώπινοι παρατηρητές, (γ) η οπτική ποιότητα που επιτυγχάνεται είναι αρκετά καλύτερη από την αντίστοιχη της μεθόδου *Itti-ROI*, αλλά το μέσο κέρδος σε συμπίεση είναι ελαφρά

Πίνακας 3.9: *Σύγκριση μεταξύ των μεθόδων VA-ROI, IttiROI και MPEG-4 σε 10 ακολουθίες*

Video Clip	**Encoding method**	**Bit Rate** (Kbps)	**Bit Rate Gain**
Eye_witness	VA-ROI	392	12.2%
	Itti-ROI	381	15.3%
	Standard MPEG-4	439	
fashion	VA-ROI	288	10.4%
	Itti-ROI	285	11.7%
	Standard MPEG-4	318	
grandma	VA-ROI	264	11.9%
	Itti-ROI	247	19.5%
	Standard MPEG-4	296	
justice	VA-ROI	227	11.5%
	Itti-ROI	236	6.9%
	Standard MPEG-4	253	
lecturer	VA-ROI	107	28.3%
	Itti-ROI	110	25.3%
	Standard MPEG-4	138	
news_cast1	VA-ROI	164	13.5%
	Itti-ROI	170	9.6%
	Standard MPEG-4	186	
news_cast2	VA-ROI	118	14.9%
	Itti-ROI	115	17.8%
	Standard MPEG-4	136	
night_interview	VA-ROI	127	15.7%
	Itti-ROI	128	14.8%
	Standard MPEG-4	147	
old_man	VA-ROI	217	13.3%
	Itti-ROI	189	30.3%
	Standard MPEG-4	246	
soldier	VA-ROI	71	22.5%
	Itti-ROI	79	10.4%
	Standard MPEG-4	87	
Average	VA-ROI	197.5	13.7%
	Itti-ROI	194.0	15.7%
	Standard MPEG-4	224.6	

μικρότερο.

□

Κεφάλαιο 4

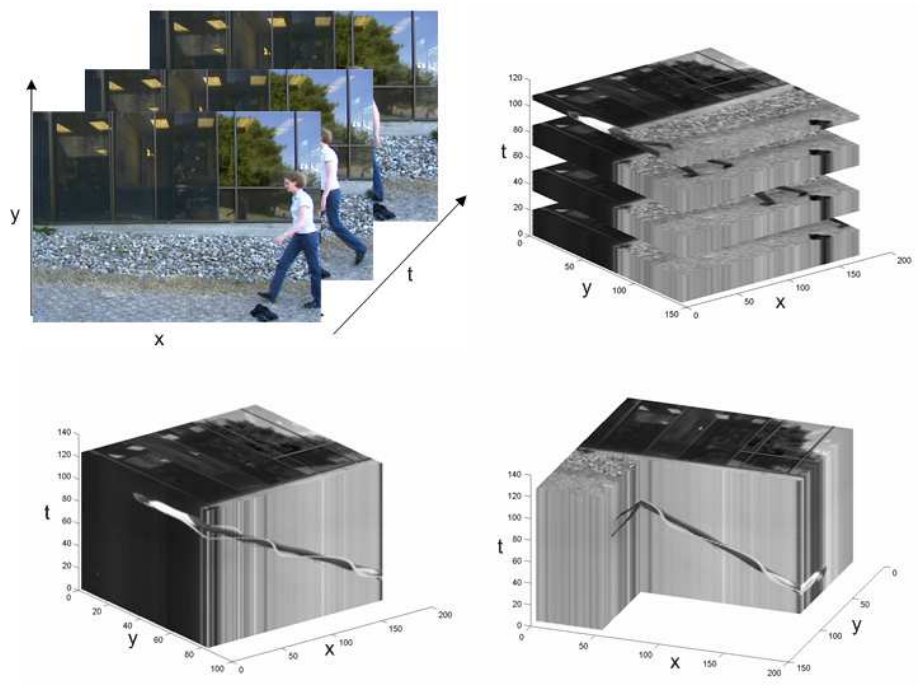
Χωροχρονικό μοντέλο ενδιαφέροντος στον χώρο του 3Δ μετασχηματισμού κυματιδίων

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο κάνουμε το πρώτο βήμα στην ανάλυση ακολουθιών ξεφεύγοντας από την συνηθισμένη επεξεργασία ανά καρέ και υιοθετώντας μία χωροχρονική αναπαράσταση της εισόδου. Αυτή η αναπαράσταση παρέχει την δυνατότητα ταυτόχρονης επεξεργασίας στον χώρο (επίπεδο καρέ) και στον χρόνο (χρονική εξέλιξη) και μάλιστα σε μεγάλο εύρος της ακολουθίας. Αναπτύσσουμε το μοντέλο χρησιμοποιώντας τον 3Δ μετασχηματισμό κυματιδίων και το αξιολογούμε σε αναγνώριση οπτικών δραστηριοτήτων. Η ανίχνευση και η αναγνώριση οπτικών γεγονότων παραμένει ένα από τα πιο ενεργά πεδία στον χώρο της μηχανικής όρασης, καθώς η πολυπλοκότητα των γεγονότων δημιουργεί την ανάγκη υπολογιστικά αποδοτικών λύσεων. Προτείνουμε ένα πλαίσιο υπολογισμού κύριου όγκου -σε αναλογία με τον διδιάστατο χάρτη-, το οποίο βασίζεται στις επιλεκτικά κατευθυντικές ζώνες του 3Δ μετασχηματισμού κυματιδίων και αναπαριστούμε τα οπτικά γεγονότα χρησιμοποιώντας απλά χαρακτηριστικά των σημαντικών περιοχών. Παρουσιάζουμε ποιοτική και ποσοτική ανάλυση της προτεινόμενης μεθόδου σε μία εφαρμογή αναγνώρισης ανθρώπινων κινήσεων από μία δημόσια διαθέσιμη βάση.

Αναλυτική επισκόπηση στον χώρο της χωροχρονικής ανάλυσης και της περιγραφής, ανίχνευσης και αναγνώρισης δυναμικών δραστηριοτήτων σε ακολουθίες παρουσιάζεται στην Ενότητα 2.4.4. Η πλειοψηφία των τεχνικών βασίζεται στην ανίχνευση σημείων ή περιοχών ενδιαφέροντος, τα οποία χρησιμοποιούνται για να αναπαραστήσουν την επιθυμητή δραστηριότητα. Κάποιες στηρίζονται στην ογκομετρική ανάλυση αναπαριστώντας την ακολουθία σαν έναν όγκο στον χώροχρόνο.

Στο προτεινόμενο πλαίσιο μία ακολουθία αναπαρίσταται σαν ένα στερεό στον τριδιάστατο Ευκλειδιανό χώρο με τον χρόνο να είναι η τρίτη διάσταση. Εφαρμόζουμε τον 3Δ μετασχηματισμό κυματιδίων, ο οποίος αποσυνθέτει τον αρχικό όγκο σε επιλεκτικά κατευθυντικές ζώνες και χρησιμοποιούμε τους συντελεστές του μετασχηματισμού για τον υπολογισμό της σημαντικότητας. Η υπόθεση μας είναι κοινή με αυτήν των περισσότερων ερευνητών στον χώρο: τα σημεία που εντοπίζονται σε γειτονιές έντονης κίνησης είναι τα πιο σημαντικά για την περιγραφή δυναμικών



Σχήμα 4.1: *Σχηματισμός χωροχρονικού όγκου και 3 διαφορετικές οπτικές γωνίες του*

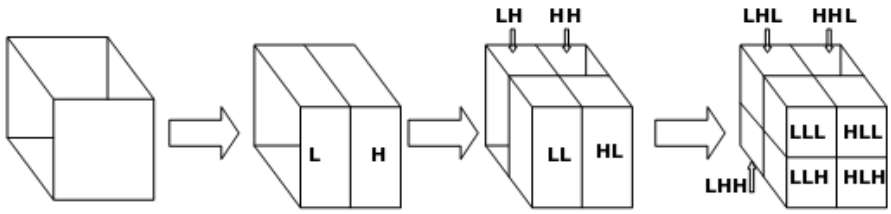
δραστηριοτήτων. Για την βελτίωση της μεθόδου ενσωματώνουμε στην περιγραφή του γεγονότος και γεωμετρικούς περιορισμούς. Ο κύριος στόχος της προτεινόμενης τεχνικής είναι η διερεύνηση των δυνατοτήτων του 3Δ μετασχηματισμού κυματιδίων να περιγράψει επιτυχώς οπτικά γεγονότα με τον περιορισμό της χαμηλής υπολογιστικής πολυπλοκότητας.

4.2 Μοντέλο υπολογισμού όγκου ενδιαφέροντος

4.2.1 Αποσύνθεση ακολουθίας βασισμένη στον 3Δ μετασχηματισμό κυματιδίων

Από κάθε ακολουθία προκύπτει ένας τριδιάστατος όγκος που αποτελείται από ένα σύνολο σημείων Q με το $q = (x, y, t)$ να είναι ένα σημείο στον χώρο-χρόνο. Η τρίτη διάσταση αντιστοιχεί στην χρονική εξέλιξη των καρέ. Με αυτήν την αναπαράσταση, ένα κινούμενο αντικείμενο καταλαμβάνει μία 3Δ περιοχή στον διακριτό χώρο που ορίσαμε πριν. Το Σχήμα 4.1 απεικονίζει τον όγκο που προκύπτει από μία ακολουθία ενός ανθρώπου που περπατά κατά μήκος του δρόμου. Οι διαφορετικές οπτικές γωνίες και τομές του όγκου που δημιουργείται δείχνουν την χωροχρονική εξέλιξη των περιοχών.

Ο 3Δ μετασχηματισμός κυματιδίων είναι διαχωρίσιμος[42] και υπολογίζεται με την εφαρμογή τριών ξεχωριστών 1Δ μετασχηματισμών. Το σήμα εισόδου συνελίσσεται στις τρεις διαστάσεις του με ένα βαθυπερατό φίλτρο L και ένα υψιπερατό φίλτρο H και υποδειγματοληπτείται κατά δύο. Αν ορίσουμε τον όγκο V να αντιστοιχεί σε μία ακολουθία εισόδου, τότε ο όγκος V φιλτράρεται αρχικά στην x διάσταση και παράγεται ένα βαθυπερατός όγκος L και ένας υψιπερατός H. Αυτοί οι όγκοι υποδειγματοληπτούνται κατά δύο και φιλτράρονται στην y διάσταση με αποτέλεσμα να δημιουργηθούν οι όγκοι LL, LH, HL, HH. Στην συνέχεια ακολουθούν πάλι



Σχήμα 4.2: *Ενδιάμεσα βήματα υπολογισμού του 3Δ μετασχηματισμού κυματιδίων*

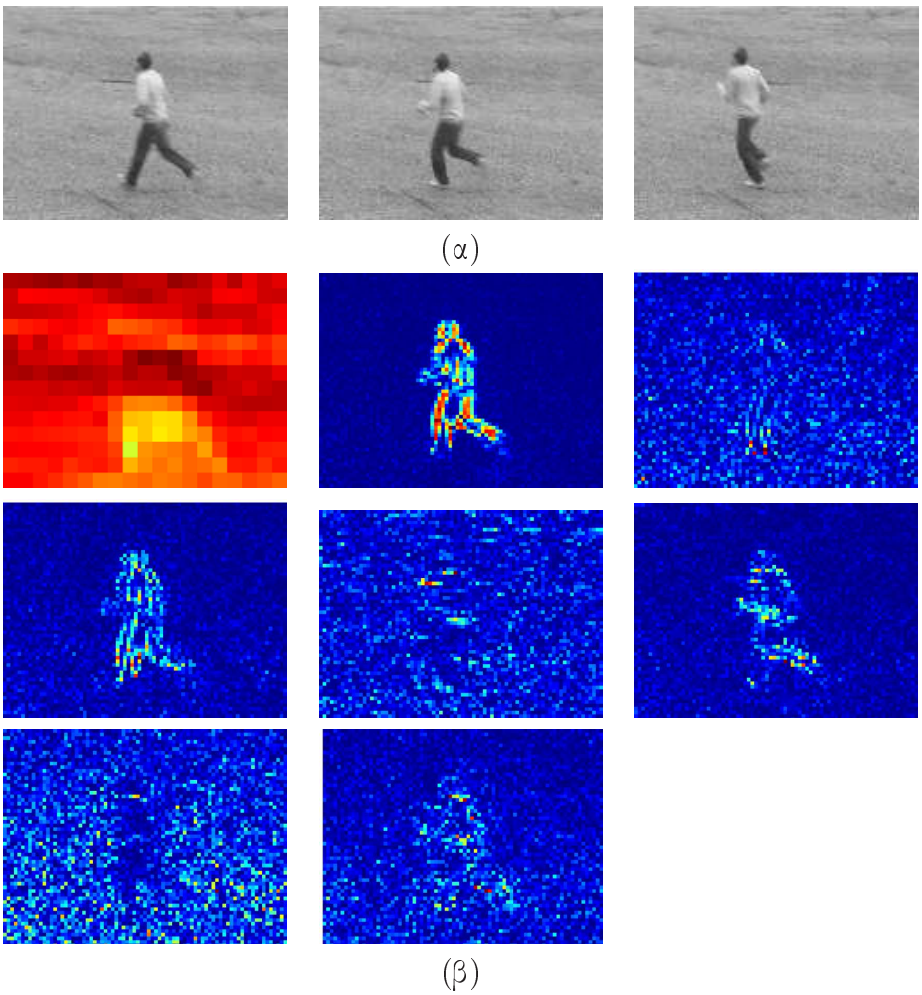
δειγματοληψία και φιλτράρισμα στην t διάσταση και το τελικό αποτέλεσμα είναι μία σειρά οχτώ όγκων, δηλαδή $LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH$. Το αποτέλεσμα επομένως της αποσύνθεσης είναι μία πολλαπλής κλίμακας αναπαράσταση της ακολουθίας που αποτελείται από τον όγκο προσέγγισης (approximation) και μία σειρά όγκων λεπτομέρειας (detail) $w_{i,\ell}$, όπου $\ell = 1, ..., L$ με L ναι είναι ο μέγιστος αριθμός κλιμάκων και $i = 1, ..., 7$ ο δείκτης των όγκων λεπτομέρειας του μετασχηματισμού. Το σχήμα 4.2 απεικονίζει την διαδικασία αποσύνθεσης ενός όγκου ακολουθίας.

Οι κατευθύνσεις με τις οποίες σχετίζονται οι ζώνες του μετασχηματισμού σχετίζονται με το εύρος των συχνοτήτων που περιλαμβάνουν. Για παράδειγμα, η ζώνη LLL αντιστοιχεί σε αργές αλλαγές ομαλών περιοχών της εισόδου, η LLH σε γρήγορα κινούμενες ομαλές περιοχές, η LHL σε αργά κινούμενες οριζόντιες περιοχές, η HLL σε αργά κινούμενες κάθετες περιοχές, η HHH σε γρήγορα κινούμενες διαγώνιες περιοχές κοκ. Το Σχήμα 4.3 απεικονίζει τομές για κάθε ζώνη του μετασχηματισμού για μία ακολουθία τρεξίματος, ενώ το Σχήμα 4.4 δείχνει την ογκομετρική αναπαράσταση κάποιων ζωνών του μετασχηματισμού.Οι φωτεινές κίτρινες περιοχές αντιστοιχούν στους συντελεστές με υψηλή τιμή και δείχνουν τα διαφορετικά μέρη της ακολουθίας που αναδεικνύονται σε κάθε ζώνη.

4.2.2 Υπολογισμός όγκου ενδιαφέροντος

Όπως αναφέρθηκε, ο μετασχηματισμός κυματιδίων παρέχει πληροφορία για τις χωροχρονικές συχνότητες του σήματος, καθώς και την θέση τους στον χώρο, χρόνο και κλίμακα. Οι συχνότητες αντιστοιχούν σε διαφορετικές χωροχρονικές κατευθύνσεις και επομένως σχετίζονται με την κίνηση στην ακολουθία. Περιμένουμε ότι οι περιοχές ενδιαφέροντος θα προκύψουν στο μετασχηματισμένο πεδίο σαν περιοχές υψηλής συγκέντρωσης ενέργειας σε μία ή περισσότερες ζώνες. Η ταυτόχρονα υψηλή συγκέντρωση σε περισσότερες από μία ζώνες σχετίζεται με σημαντικές αλλαγές κατεύθυνσης στην κίνηση και επομένως σημαίνει και αύξηση της τιμής ενδιαφέροντος για την συγκεκριμένη περιοχή.

Ακολουθώντας αυτήν τη λογική, υποθέτουμε ότι το δυναμικό περιεχόμενο της ακολουθίας μπορεί να χαρακτηριστεί υπολογίζοντας την κατανομή της χωροχρονικής ενέργειας στις ζώνες του 3Δ μετασχηματισμού. Ορίζουμε ένα ογκοστοιχείο ώς σημαντικό αν παρουσιάζει υψηλή ενδο- και διά- ζωνική συγκέντρωση ενέργειας σε όλες τις κλίμακες. Η ενέργεια κάθε ογκοστοιχείου υπολογίζεται σαν ο συνδυασμός της ενέργειας σε μία μικρή γειτονιά στην ίδια ζώνη (ενδο-ζωνική) και της ενέργειας των στοιχείου των υπόλοιπων ζωνών στην ίδια θέση (δια-ζωνική). Στον χώρο των κυματιδίων αυτού του είδους ο συνδυασμός γίνεται αρκετά απλός καθώς οι ισχυροί συντελεστές του σήματος ξεχωρίζουν καθαρά από τους πιο αδύναμους και ο θόρυβος διαχέεται σε όλες τις ζώνες. Η δυνατότητα της πυραμιδοειδούς αναπαράστασης της



Σχήμα 4.3: *(α) Τρία καρέ από μία ακολουθία τρεξίματος, (β) Η ζώνη χαμηλής συχνότητας (LLL) και οι 7 ζώνες υψηλής συχνότητας του 3Δ μετασχηματισμού κυματιδίων για το μεσαίο καρέ στο (α). Το κόκκινο χρώμα αντιστοιχεί σε υψηλές τιμές και το μπλε σε χαμηλές.*

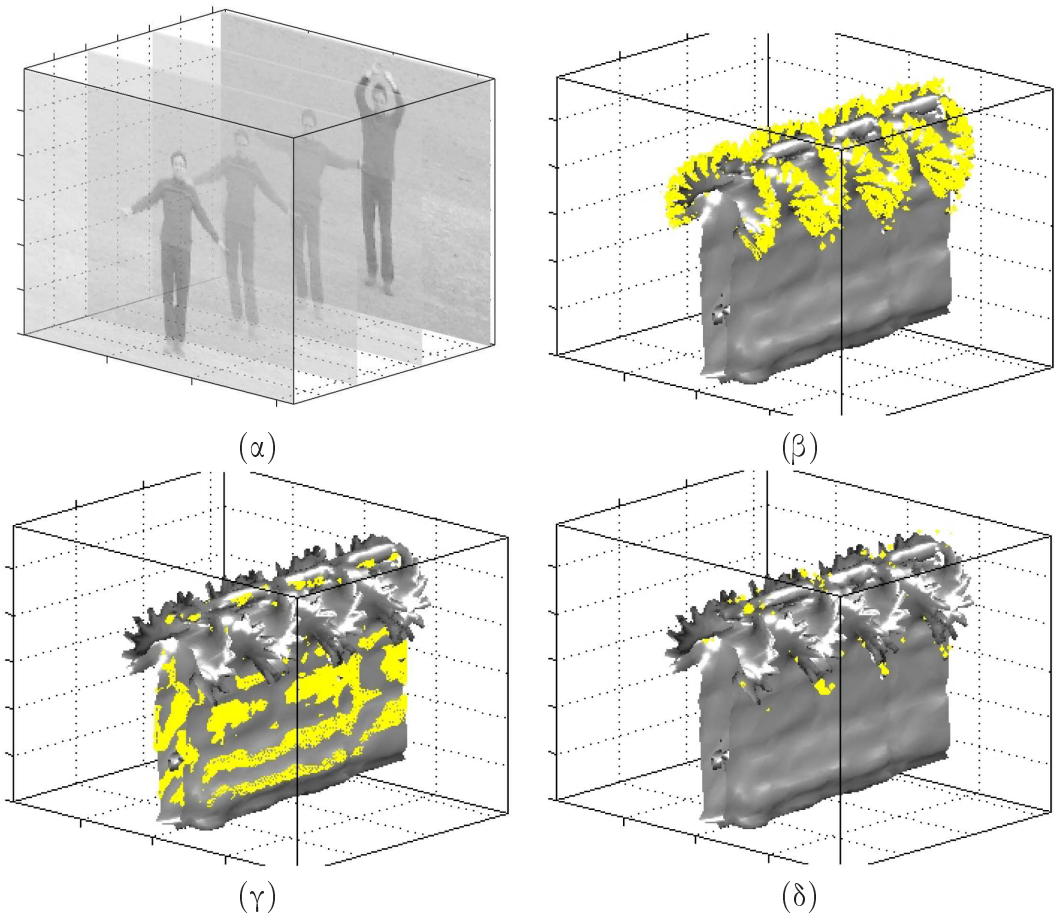
εισόδου να αναπαριστά σύντομα και εκτενή γεγονότα σε διαφορετικές κλίμακες μας οδήγησε στην χρήση μίας "ελεύθερης" παραμέτρου $\beta \in [0.1, 0.5]$, η οποία θα καθορίζει το επιθυμητό επίπεδο λεπτομέρειας δίνοντας διαφορετική βαρύτητα στις κλίμακες του μετασχηματισμού. Συσχετίζουμε την παράμετρο αυτήν με το σχήμα μιας θετικά ορισμένης Λαπλασιανής συνάρτησης, η οποία ορίζεται ως:

$$\Lambda(x|\mu, \beta) = \frac{1}{2\beta} \cdot \exp\left(-\frac{|x-\mu|}{\beta}\right) \tag{4.1}$$

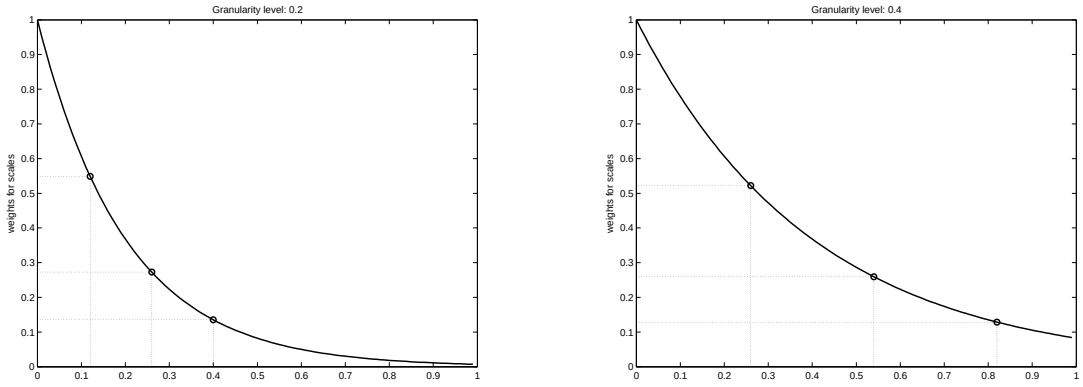
όπου η παράμετρος β καθορίζει το σχήμα της και μ είναι η μέση τιμή της. Τα βάρη κάθε κλίμακας ℓ υπολογίζονται ως

$$\nu(\ell) = \Lambda(\ell \log(2)|\mu = 0, \beta),\ \ell = 1, ..., L,\ \beta \in [0.1, 0.5] \tag{4.2}$$

Το Σχήμα 4.6 δείχνει το οπτικό αποτέλεσμα της μεταβολής της παραμέτρου σε μία ακολουθία. Όσο μεγαλύτερη είναι η τιμή τόσο πιο πιο συνεκτικές είναι οι περιοχές που εντοπίζονται.



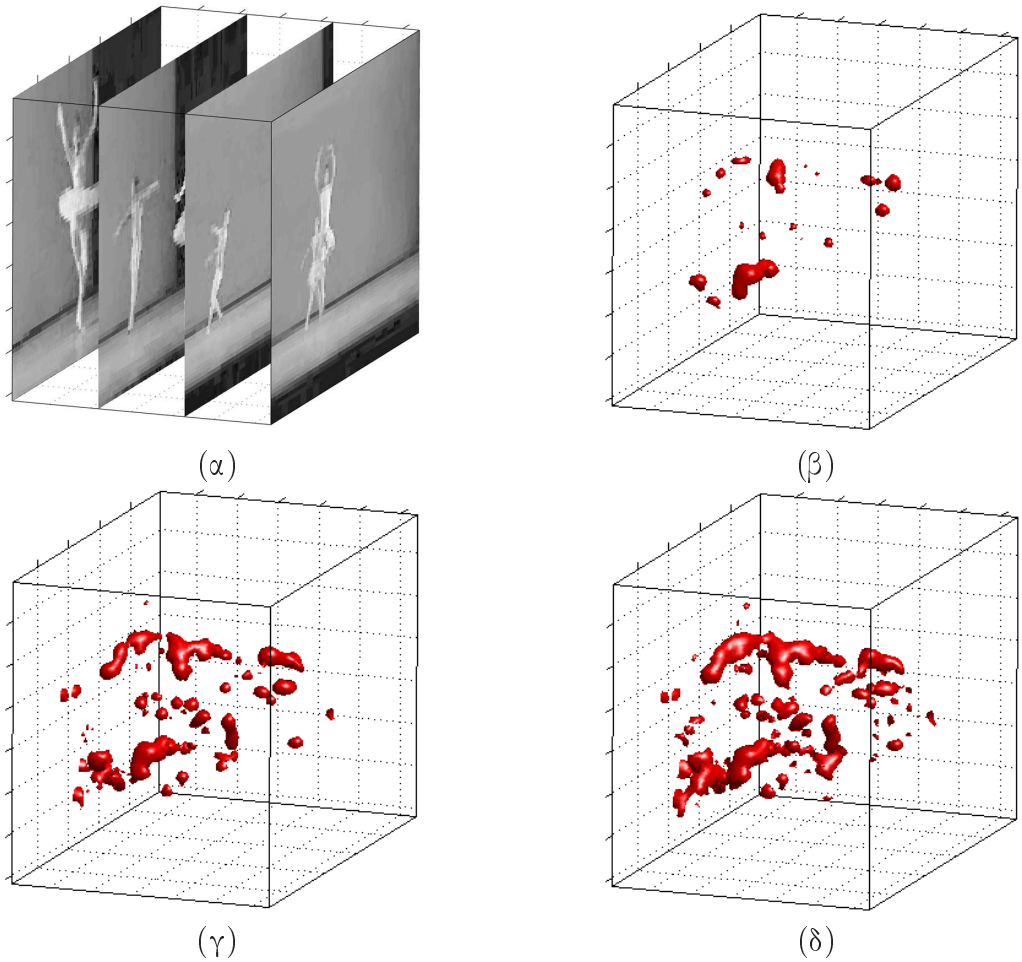
Σχήμα 4.4: *(α) Η αρχική ακολουθία χαιρετισμού και οι ογκομετρικές αναπαραστάσεις των ζωνών (β) LLH (γ) LHL και (δ) HHH του 3Δ μετασχηματισμού κυματιδίων. Απεικονίζονται οι ISO επιφάνειες των όγκων μετά από κάτάλληλη κατωφλίωση (οι φωτεινές κίτρινες περιοχές αντιστοιχούν σε υψηλές τιμές)*



Σχήμα 4.5: *Τα βάρη $\nu(\ell)$ (y-άξονας) των διαφορετικών κλιμάκων του μετασχηματισμού για (α) $\beta = 0.2$ και (β) $\beta = 0.4$. Στο συγκεκριμένο παράδειγμα η αποσύνθεση έχει γίνει σε τρεις κλίμακες*

Η χωροχρονική τοπική ενέργεια κάθε ζώνης υπολογίζεται ως

$$E_{i,\ell}(q) = \frac{1}{|N_q|}\left(\sum_{r\in N_q} |w_{i,\ell}(r)|^2\right)^{\nu(\ell)} \tag{4.3}$$

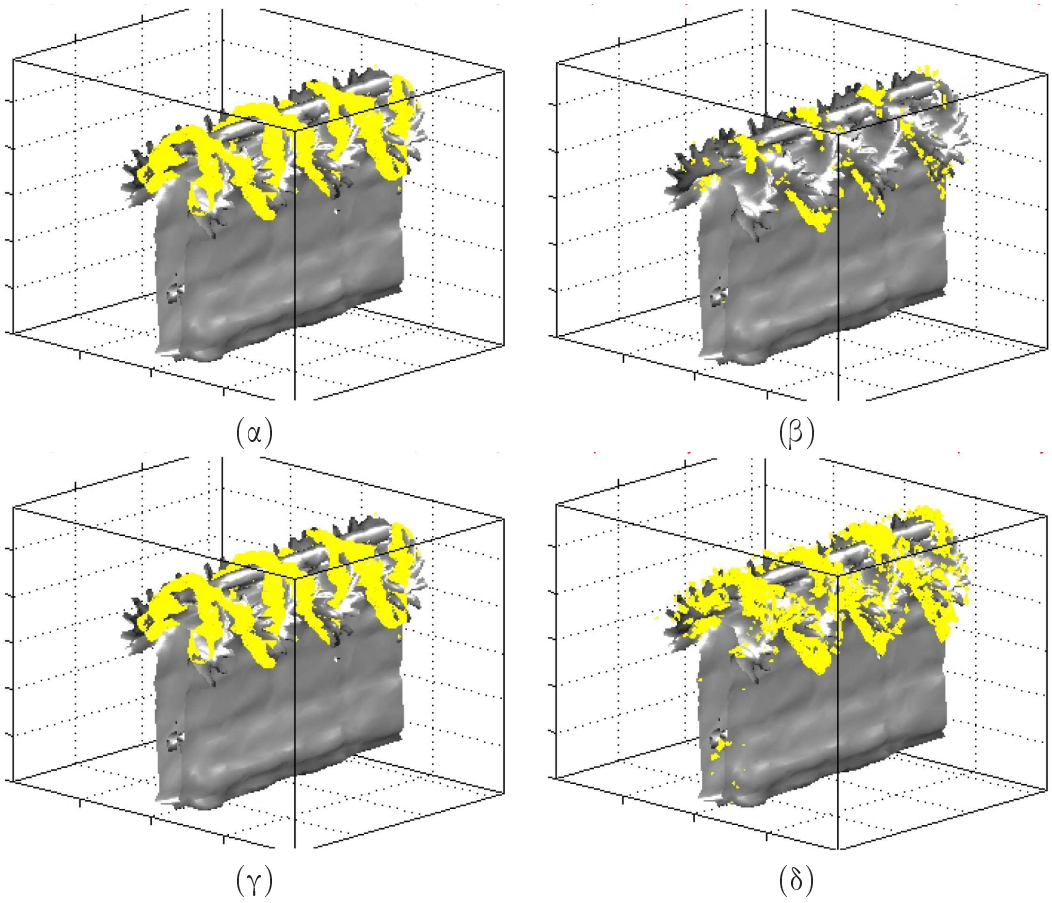


Σχήμα 4.6: *Επίδραση της παραμέτρου β στον κύριο χάρτη της ακολουθίας που απεικονίζεται στο (α). Οι τρεις εικόνες αντιστοιχούν σε (β) $\beta = 0.2$, (γ) $\beta = 0.4$ και (δ) $\beta = 0.7$. Όσο αυξάνεται η τιμή του β τόσο αναδεικνύονται αδρομερείς λεπτομέρειες της ακολουθίας*

όπου N_q είναι το σύνολο των 26 γειτόνων στην χωροχρονική τοπική γειτονιά του ογκοστοιχείου q. Οι τελικές ενέργειες για τις έξι ανθρώπινες δραστηριότητες για γειτονιά $3 \times 3 \times 3$ φαίνονται στα Σχήματα 4.8 και 4.9. Ο κύριος όγκος για την ακολουθία υπολογίζεται ως

$$S(q) = \sum_{\ell} \sum_{i} E_{i,\ell}(q) \tag{4.4}$$

Το Σχήμα 4.7 απεικονίζει τον κύριο όγκο μίας ακολουθίας χαιρετισμού για τρεις διαφορετικές τιμές της παραμέτρου β. Όσο αυξάνει η τιμή της τόσο πιο αδρομερείς περιοχές εντοπίζονται ως σημαντικές. Τα Σχήματα δείχνουν τις ISO επιφάνειες μετά από κατάλληλη κατωφλίωση για να φανούν καλύτερα οι λιγότερο και οι περισσότερο σημαντικές περιοχές, οι οποίες φαίνονται ως γκρι και κίτρινες αντιστοιχα. Όπως φαίνεται στην (4.4), χρησιμοποιείται μόνο η πρώτη κλίμακα του μετασχηματισμού. Πιο εκλεπτυσμένα αποτελέσματα θα μπορούσαν να προκύψουν με την χρήση περισσότερων κλιμάκων, αλλά δεν κρίνεται απαραίτητο στην εφαρμογή που εξετάζουμε. Μία πιο διεισδυτική ματιά για την φύση και των ζωνών που συνδυάζονται και την σημαντικότητα που προκύπτει δίνεται στην επόμενη ενότητα.

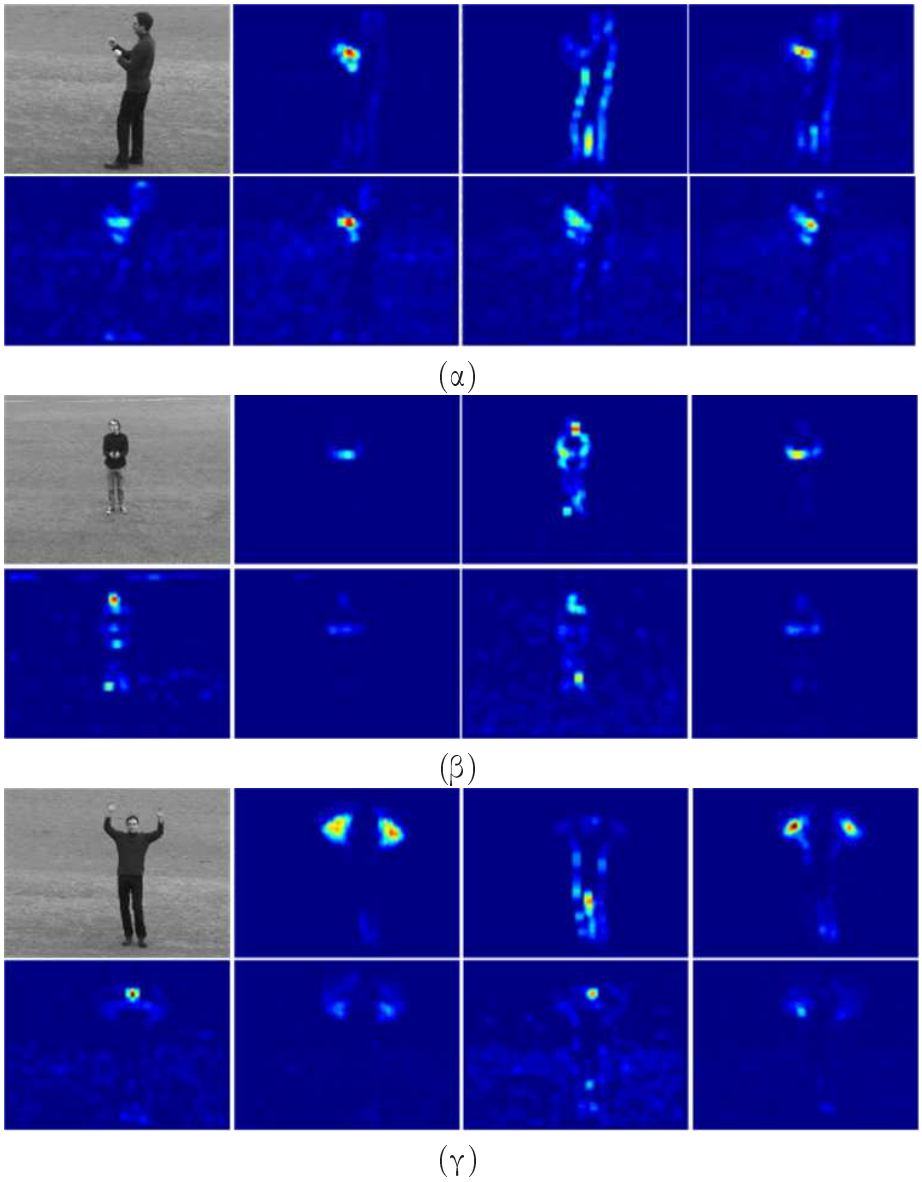


Σχήμα 4.7: *Εναλλακτική αναπαράσταση του κύριου όγκου μίας ακολουθίας χαιρετισμού, η οποία απεικονίζεται στο Σχήμα 4.4α. Οι σκούρες περιοχές είναι η ISO επιφάνεια που αντιστοιχεί στις λιγότερο σημαντικές περιοχές (κορμός του ανθρώπου), ενώ οι φωτεινές αντιστοιχούν στις πιο σημαντικές όπως αυτές εντοπίζονται με (β)* $\beta = 0.1$*, (γ)* $\beta = 0.3$ *και (δ)* $\beta = 0.4$

4.2.3 Ανίχνευση ανθρώπινων δραστηριοτήτων

Όπως ισχυριστήκαμε στην εισαγωγή, ο στόχος της προτεινόμενης δουλειάς είναι η δημιουργία ενός υπολογιστικά αποδοτικού αλλά και επιτυχημένου στην αναπαράσταση και ανίχνευση χωροχρονικών γεγονότων μοντέλου. Η ιδέα της χρήσης του 3Δ μετασχηματισμού κυματιδίων προέκυψε μετά από λεπτομερή οπτική εξέταση των διαφορετικών ζωνών του μετασχηματισμού και την ιδιότητα τους να αποσυνθέτουν μία περίπλοκη κίνηση σε απλούστερες. Μετά από την αποσύνθεση μερικών ακολουθιών που απεικονίζουν ανθρώπινες δραστηριότητες γίνεται προφανές ότι η τοπική ενεργοποίηση κάποιων υποσυνόλων ζωνών αντιστοιχεί σε συγκεκριμένες δραστηριότητες.

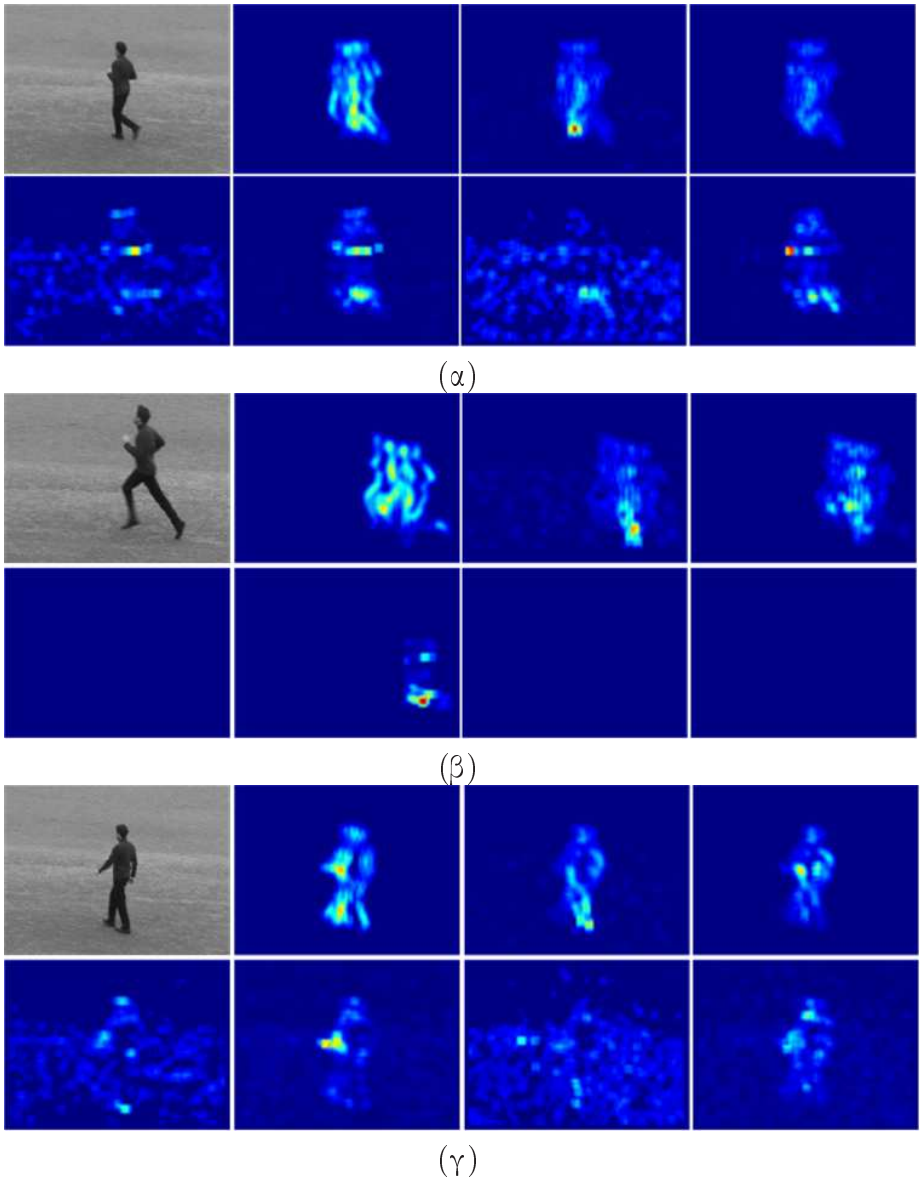
Στα Σχήματα 4.8 και 4.9 απεικονίζεται παραδείγματα καρέ από τις έξι ανθρώπινες δραστηριότητες που μελετήθηκαν μαζί με τις αντίστοιχες τομές των ζωνών λεπτομέρειας που αντιστοιχούν σε αυτές. Οι ακολουθίες των έξι δραστηριοτήτων που χρησιμοποιήθηκαν είναι δημόσια διαθέσιμες και αντιστοιχούν σε πυγμαχία, χειροκρότημα, χαιρετισμό, τζόκινγκ, τρέξιμο και περπάτημα. Περισσότερες πληροφορίες δίνονται στην ενότητα 4.3. Μία απλή ματιά στις εικόνες αποκαλύπτει την διαφορά σε ενεργοποίηση υπο-ζωνών μετά την αποσύνθεση της ακολουθίας και τον υπολογισμό της σημαντικότητας. Π.χ. η LLH ζώνη της πυγμαχίας, του χειροκροτήματος και του χαιρετισμού είναι διαφορετική τόσο σε ένταση όσο και στην σχετική θέση των



Σχήμα 4.8: *Τομές των 7 ζωνών του 3Δ μετασχηματισμού κυματιδίων για ακολουθίες (α) πυγμαχίας, (β) χειροκροτήματος και (γ) χαιρετισμού*

περιοχών υψηλής ενεργοποίησης. Το ίδιο ισχύει και για την LHH ζώνη. Παρόμοια συμπεράσματα μπορούν να εξαχθούν και για τις εικόνες στο σχήμα 4.9, το οποίο περιέχει δραστηριότητες με περισσότερη κίνηση.

Καθώς ο στόχος μας ήταν να μελετήσουμε την δυναμική του μετασχηματισμού στην αναπαράσταση οπτικών γεγονότων πειραματιστήκαμε με διαφορετικούς συνδυασμούς ζωνών για τον υπολογισμό του κύριου όγκου. Το Σχήμα 4.10 δείχνει αποτελέσματα για μία ακολουθία χαιρετισμού μετά τον συνδυασμό των συνόλων $b_1 = LLH, LHH, HLH, HHH$ και $b_2 = LHL, HLL, HHL$. Το σύνολο b_1 σχετίζεται περισσότερο με τις γοργά μεταβαλλόμενες περιοχές, οι οποίες αντιστοιχούν συνήθως στο προσκήνιο, ενώ το σύνολο b_2 με τις αργά μεταβαλλόμενες περιοχές του υπόβαθρου. Αυτή η διαφοροποίηση είναι περισσότερο αισθητή στις δραστηριότητες χαμηλής κινητικότητας, όπως η πυγμαχία, το χειροκρότημα και ο χαιρετισμός. Στο Σχήμα 4.10β απεικονίζεται το αποτέλεσμα του συνδυασμού των ζωνών στο b_1 και στο Σχήμα 4.10γ το αποτέλεσμα του συνδυασμού των ζωνών στο b_2. Στα Πειράματα

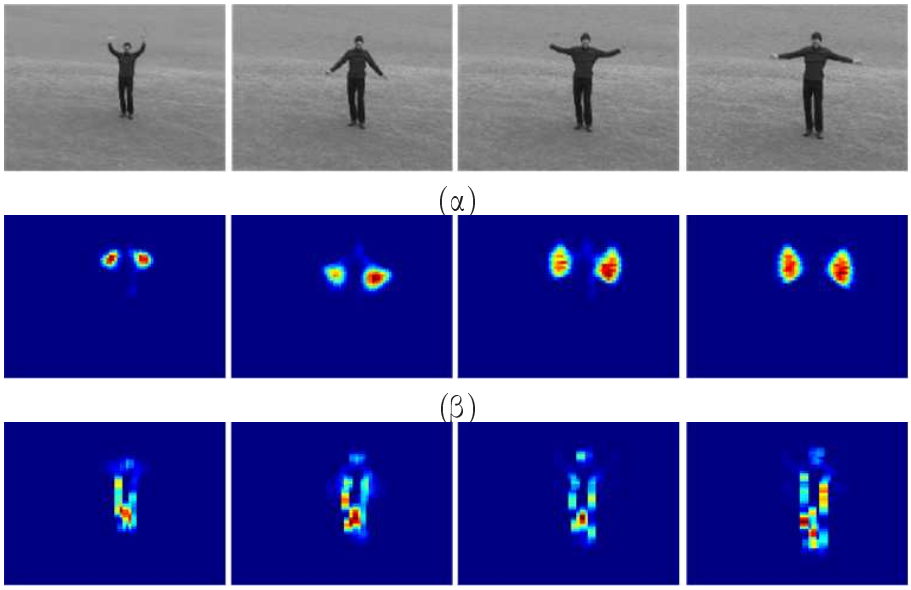


Σχήμα 4.9: *Τομές των 7 ζωνών του 3Δ μετασχηματισμού κυματιδίων για ακολουθίες (α) τζοκινγκ, (β) τρεξίματος και (γ) περπατήματος*

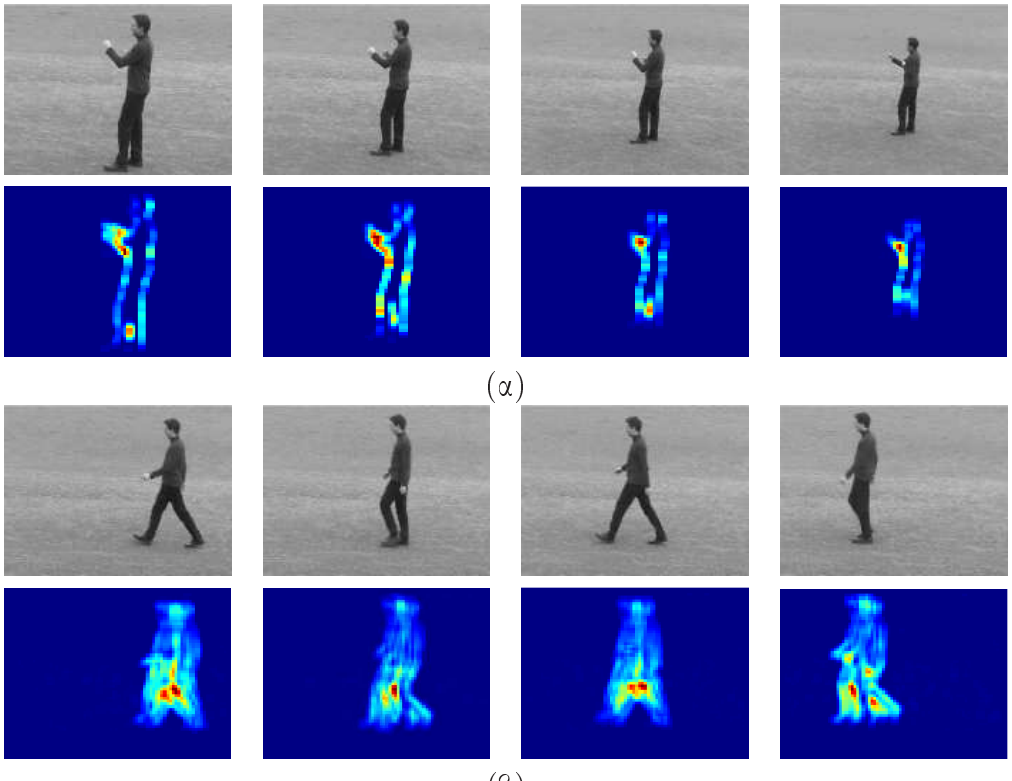
που παρουσιάζουμε στην Ενότητα 4.3 χρησιμοποιούμε ίδια βάρη για τα δύο σύνολα. Στο Σχήμα 4.11 απεικονίζονται παραδείγματα τομών κύριου χάρτη από ακολουθίες με αλλαγές βάθους της κάμερας. Ο κύριος χάρτης φαίνεται ότι δεν επηρεάζεται από τέτοιου είδους αλλαγές.

Η γεωμετρική μορφή των δραστηριοτήτων παίζει σημαντικό ρόλο στην κατανόηση τους. Σε μια προσπάθεια να εισάγουμε γεωμετρικά χαρακτηριστικά για την αναγνώριση δραστηριοτήτων, χωρίς να αυξηθούν σημαντικά οι απαιτήσεις σε υπολογιστική ισχύ, εντοπίζουμε έναν αριθμό ομάδων με την χρήση του k-means και επιλέγουμε τα p πολυπληθέστερα. Με αυτόν τον τρόπο επιλέγουμε p κέντρα, τα οποία αντιστοιχούν στις πιο σημαντικές περιοχές στο χωροχρονικό πεδίο που καταλαμβάνει η δραστηριότητα. Τέτοια κέντρα εντοπίζονται π.χ. στα δύο χέρια στις ακολουθίες χαιρετισμού, στα πόδια στις ακολουθίες περπατήματος και στις περιοχές του λαιμού και των ποδιών στις ακολουθίες τρεξίματος.

Το τελικό διάνυσμα χαρακτηριστικών F προκύπτει από τον συνδυασμό ιστογραμ-



Σχήμα 4.10: *Γειτονικά καρέ από μία ακολουθία χαιρετισμού υπό αλλαγές βάθους της κάμερας (zoom in/out) και οι αντίστοιχες τομές του κύριου όγκου χρησιμοποιώντας τις ζώνες του συνόλου (α)* b_1 *και (β)* b_2



Σχήμα 4.11: *Γειτονικά καρέ από μία ακολουθία (α) πυγμαχίας και (β) περπατήματος υπό αλλαγές βάθους της κάμερας (zoom in/out) και οι αντίστοιχες τομές του κύριου όγκου χρησιμοποιώντας όλες τις ζώνες*

μάτων γύρω από τα εντοπισμένα κέντρα:

$$F = \{H_p^\ell, C_p^\ell\} \tag{4.5}$$

όπου H_p^ℓ είναι τα ιστογράμματα σε μία γειτονιά γύρω από τα p κέντρα C_p^ℓ στο επίπεδο ℓ.

Για την αξιολόγηση της απόδοσης της μεθόδου υιοθετήθηκε η μέθοδος με-μία-παράλειψη[34] σε μία εφαρμογή ανάκτησης με βάση την ομοιότητα: κάθε ακολουθία συγκρίνεται με όλες τις υπόλοιπες και το αποτέλεσμα ελέγχεται για το αν ταυτίζεται με την πραγματική ταυτότητα της ακολουθίας. Για τον υπολογισμό της ομοιότητας χρησιμοποιήθηκε η Ευκλείδια απόσταση, η οποία για δύο ακολουθίες k_1 και k_2 ορίζεται ως

$$d_E(F_{k_1}, F_{k_2}) = \sqrt{\sum_j \left(F_{k_1}(j) - F_{k_2}(j)\right)^2} \tag{4.6}$$

όπου j είναι ο δείκτης στο διάνυσμα χαρακτηριστικών.

4.3 Πειραματικά αποτελέσματα

4.3.1 Πειραματικό πλαίσιο και μεθοδολογία

Για την αξιολόγηση της προτεινόμενης μεθόδου επιλέγουμε την αναγνώριση ανθρώπινων δραστηριοτήτων από μία δημόσια διαθέσιμη βάση ακολουθιών [45] και την συγκρίνουμε με την μέθοδο των Laptev *et al.*, η οποία και αυτή είναι δημόσια διαθέσιμη [62]. Όπως αναφέρθηκε η βάση αποτελείται από ένα σύνολο ακολουθιών σχετικές με έξι ανθρώπινες δραστηριότητες που έχουν εκτελεστεί από 25 ανθρώπους σε τέσσερα διαφορετικά σενάρια: εξωτερικού χώρου s_1, εξωτερικού χώρου υπό αλλαγή κλίμακας s_2, εξωτερικού χώρου με διαφορετικό ρουχισμό s_3 και εσωτερικού χώρου s_4. Όλες οι ακολουθίες έχουν βιντεοσκοπηθεί με σταθερή κάμερα και 25 καρέ/s, τα οποία έχουν μέγεθος 160×120. Οι έξι δρατηριότητες είναι: πυγμαχία (Box), χειροκρότημα (Hclp), χαιρετισμός (Hwav), τζόκινγκ (Jog), τρέξιμο (Run) και περπάτημα (Walk), με τις συντομογραφίες στις παρενθέσεις να υποδηλώνουν τον συμβολισμό που χρησιμοποιείται στους πίνακες αποτελεσμάτων.

Οι Laptev *et al.* ανιχνεύουν τοπικά χωροχρονικά σημεία, τα προσαρμόζουν σε θέση, μέγεθος και στην ταχύτητα της κίνησης και εξάγουν χωροχρονικά jets γύρω από αυτά [62]. Η αρχική τους υλοποίηση επιστρέφει όλα τα σημεία χωρίς κάποια επιλογή, αλλά σε μία μεταγενέστερη δουλειά τους προτείνουν την χρήση του k-means για την επιλογή των τεσσάρων πιο αντιπροσωπευτικών σημείων από τις πυκνότερες ομάδες. Στα πειράματα μας χρησιμοποιήθηκε και αυτό το επιπλέον βήμα. Το διάνυσμα χαρακτηριστικών που προκύπτει για αυτήν τη μέθοδο έχει μήκος 136, καθώς επιλέγονται 4 σημεία και 34 τοπικά Harris jets υπολογίζονται γύρω από το καθένα. Το διάνυσμα χαρακτηριστικών της προτεινόμενης μεθόδου είναι μεγαλύτερο καθώς χρησιμοποιούμε $n = 7$, $p = 3$ και $\ell = 2$ στην εξίσωση (4.5). Το τελικό διάνυσμα έχει μήκος 224.

4.3.2 Αποτελέσματα

Σε αυτήν την ενότητα παρουσιάζονται στατιστικά αποτελέσματα για την ανάκτηση ανθρώπινων δραστηριοτήτων από την συλλογή ακολουθιών που χρησιμοποιούμε. Οι Πίνακες 4.1 και 4.2 περιέχουν τα αποτελέσματα της προτεινόμενης μεθόδου για το σενάριο s_1 χωρίς και με γεωμετρικούς περιορισμούς αντίστοιχα. Άξια παρατήρησης είναι η βελτιωμένη διαφοροποίηση μεταξύ των ακολουθιών χαιρετισμού και πυγμαχίας

Πίνακας 4.1: *Αποτελέσματα προτεινόμενης μεθόδου στο σενάριο s_1*

	Box	Hclp	Hwav	Jog	Run	Walk
Box	**64**	10	17	0	0	8
Hclp	18	**59**	20	0	0	3
Hwav	15	12	**72**	0	0	1
Jog	0	0	1	**74**	10	15
Run	0	0	0	22	**75**	3
Walk	12	6	3	14	3	**62**
prec	0.587	0.678	0.637	0.673	0.852	0.674
rec	0.646	0.590	0.720	0.740	0.750	0.620

Πίνακας 4.2: *Αποτελέσματα προτεινόμενης μεθόδου με γεωμετρικούς περιορισμούς στο σενάριο s_1*

	Box	Hclp	Hwav	Jog	Run	Walk
Box	**73**	14	11	0	0	1
Hclp	15	**70**	12	0	0	3
Hwav	5	6	**85**	0	0	4
Jog	0	1	0	**74**	15	10
Run	0	0	0	22	**75**	3
Walk	5	5	3	14	1	**72**
prec	0.745	0.729	0.766	0.673	0.824	0.7749
rec	0.737	0.700	0.850	0.740	0.750	0.720

καθώς και μεταξύ χειροκροτήματος/περπατήματος και πυγμαχίας. Η βελτίωση οφείλεται κυρίως στην διαφορετική γεωμετρία των κινήσεων, όπως αυτή εκφράζεται μέσα από τα εντοπισμένα κέντρα της προτεινόμενης μεθόδου. Ο Πίνακας 4.3 περιέχει τα αποτελέσματα της μεθόδου των Laptev *et al.*. Τα ποσοστά είναι γενικά χαμηλότερα από αυτά της προτεινόμενης μεθόδου με το ζευγάρι τρέξιμο-τζόκινγκ να παρουσιάζουν το μεγαλύτερο πρόβλημα, όπως υποστηρίζουν και οι συγγραφείς σε προηγούμενη δουλειά τους [114]. Τέλος στον Πίνακα 4.4 παρουσιάζουμε αποτελέσματα της προτεινόμενης μεθόδου για όλα τα σενάρια που αναφέραμε. Τα ποσοστά είναι ενθαρρυντικά και δείχνουν ότι με μικρές προσαρμογές της μεθόδου θα ανέβουν περισσότερο.

Η υπολογιστική απόδοση ήταν ένας από τους αρχικούς στόχους της μεθόδου. Ενδεικτικά, για έναν Pentium IV, 2.4GHz με 512MB RAM, ο μέσος χρόνος επεξεργασίας για τις διαθέσιμες υλοποιήσεις σε MATLAB για έναν όγκο 32 καρέ είναι 19s, ενώ για την ανίχνευση των σημείων με την μέθοδο των Laptev *et al.* και την προσαρμογή τους είναι 319s. Ο μέγιστος αριθμός επαναλήψεων για την δεύτερη μέθοδο ορίστηκε στις 20, όπως και στον δημόσια διαθέσιμο κώδικα. Και οι δύο υλοποιήσεις παρουσιάζουν μεγαλύτερη καθυστέρηση στις πιο έντονες δραστηριότητες (τρέξιμο, περπάτημα, τζόκινγκ) εξαιτίας των περισσότερων σημαντικών σημείων που προκύπτουν και της επακόλουθης καθυστέρησης του k-means.

Πίνακας 4.3: *Αποτελέσματα της μεθόδου των Laptev et al. με προσαρμογή στο σενάριο* s_1

	Box	Hclp	Hwav	Jog	Run	Walk
Box	45	15	7	9	15	8
Hclp	15	44	11	3	13	14
Hwav	5	8	57	4	14	12
Jog	1	1	9	32	33	24
Run	1	6	1	11	75	6
Walk	1	7	8	21	19	44
prec	0.662	0.543	0.613	0.400	0.444	0.407
rec	0.455	0.440	0.570	0.320	0.750	0.440

Πίνακας 4.4: *Αποτελέσματα της προτεινόμενης μεθόδου με γεωμετρικούς περιορισμούς στα σενάρια* s_1, s_2, s_3

	Box	Hclp	Hwav	Jog	Run	Walk
Box	194	47	32	7	3	16
Hclp	70	154	36	11	8	17
Hwav	37	39	197	7	1	17
Jog	7	9	3	168	70	43
Run	5	7	7	80	180	21
Walk	22	23	27	46	21	161
prec	0.579	0.552	0.652	0.527	0.636	0.585
rec	0.649	0.520	0.661	0.560	0.600	0.537

4.4 Συμπεράσματα

Σε αυτό το κεφάλαιο προτείναμε μία μέθοδο για τον υπολογισμό χωροχρονικής σημαντικότητας στον χώρο των 3Δ κυματιδίων και την εφαρμόσαμε σε αναγνώριση ανθρώπινων δραστηριοτήτων. Χρησιμοποιήσαμε ένα σύνολο περιγραφέων, οι οποίοι προκύπτουν από ιστογράμματα γύρω από περιοχές έντονης ένδο- και διά- ζωνικής ενεργοποίησης των επιλεκτικών σε κατεύθυνση ζωνών του μετασχηματισμού και την γεωμετρία των σημαντικών περιοχών. Η δυνατότητα της μεθόδου να αναπαριστά και να εντοπίζει επιθυμητές δραστηριότητες παρουσιάζεται μέσα από ένα σύνολο σχετικών πειραμάτων σε μία δημόσια βάση με πολλές ακολουθίες και μέσω της σύγκρισης με μία εδραιωμένη στην βιβλιογραφία τεχνική.

Αν και τα αποτελέσματα είναι πολύ ενθαρρυντικά, η προτεινόμενη δουλειά θα πρέπει να χαρακτηριστεί ως μία προσπάθεια να αναδειχθούν οι δυνατότητες του μετασχηματισμού κυματιδίων στην αναπαράσταση δυναμικών οπτικών γεγονότων παράλληλα με την υπολογιστική του απλότητα. Υπάρχουν λογικές επεκτάσεις της μεθόδου, οι οποίες αναμένουμε ότι θα βελτιώσουν τα ποσοστά αναγνώρισης. Τα σημαντικότερα μειονεκτήματα είναι η μεταβλητότητα σε μετατόπιση και η μειωμένη επιλεκτικότητα σε κατεύθυνση που παρουσιάζει ο απλός μετασχηματισμός. Στην κατεύθυνση της

βελτίωσης αυτών των προβλημάτων έχει προταθεί ο πραγματικός και ο μιγαδικός 3Δ μετασχηματισμός κυματιδίων, καθώς και ο αντίστοιχος πραγματικός και μιγαδικός μετασχηματισμός Διπλού-Δένδρου που είδαμε στην Ενότητα 3.2.1 [57] [58]. Αυτοί οι μετασχηματισμοί παρέχουν μεγαλύτερη επιλεκτικότητα σε κατεύθυνση αλλά με αυξημένο υπολογιστικό κόστος. Επιπροσθέτως, θα μπορούσαμε να αυξήσουμε τις διαφοροποιήσεις μεταξύ των γεγονότων που θέλουμε να ταυτοποιήσουμε προτείνοντας έναν τρόπο αυτόματου υπολογισμού των βαρών των συνόλων b_1 και b_2, ο οποίος να βασίζεται π.χ. στην μέση ενεργοποίηση του αντίστοιχου συνόλου για τις συγκεκριμένες δραστηριότητες.

□

Κεφάλαιο 5

Χωροχρονικό μοντέλο υπολογισμού κύριου όγκου

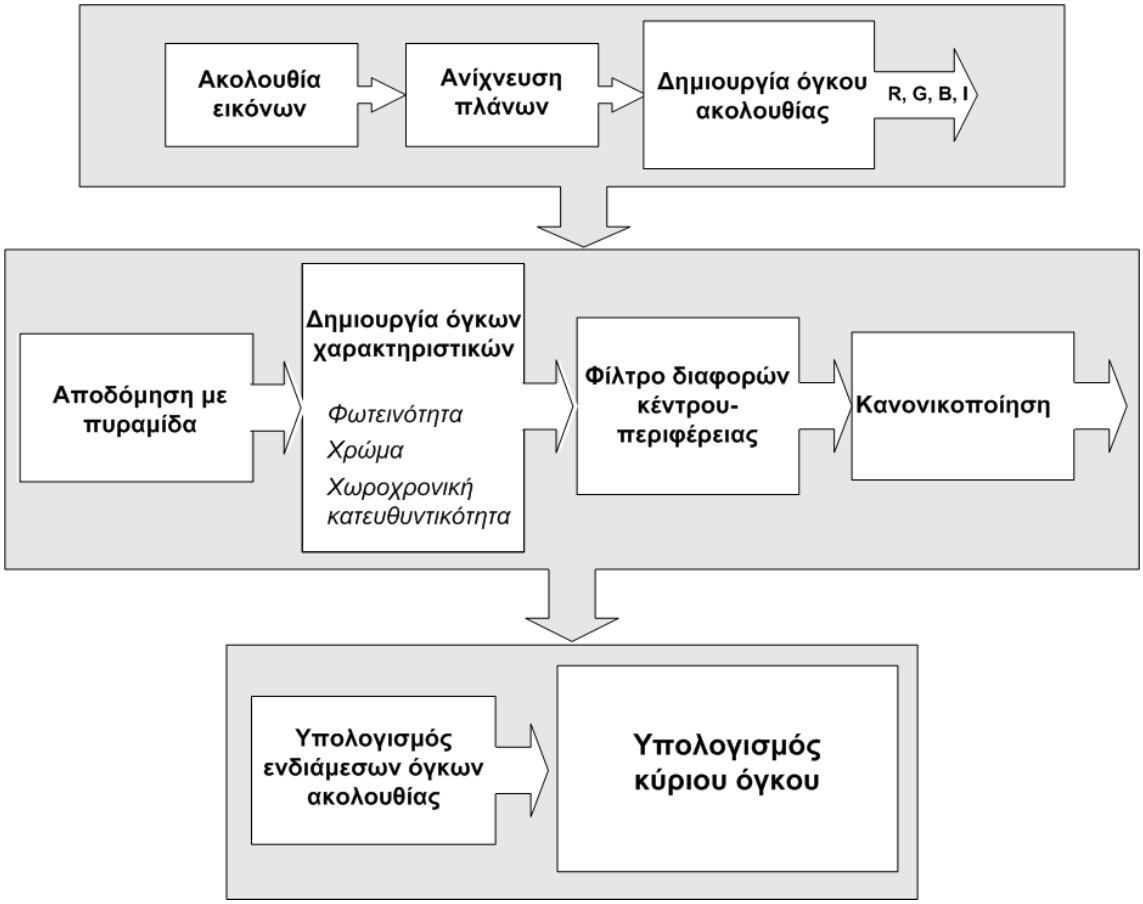
5.1 Εισαγωγή

Σε αυτό το κεφάλαιο προτείνουμε ένα υπολογιστικό μοντέλο οπτικής προσοχής για ανάλυση ακολουθιών, το οποίο βασίζεται στην χωροχρονική επέκταση του υπολογισμού κύριου όγκου ενδιαφέροντος των Itti *et al.* [48]. Συνυπολογίζεται δηλαδή τόσο η χωρική έκταση όσο και η δυναμική εξέλιξη των περιοχών στην διάρκεια του χρόνου. Η χωροχρονική επεξεργασία μπορεί να λύσει πολλά από τα συνήθη προβλήματα της ανάλυσης ακολουθιών, αλλά η ποσότητα των δεδομένων και η συνεπαγόμενη υπολογιστική πολυπλοκότητα γίνονται συχνά εμπόδιο. Είναι απαραίτητος ένας μηχανισμός που θα επιλέγει τα πιο σημαντικά μέρη της οπτικής εισόδου εξάγοντας απλά χαρακτηριστικά με γρήγορο τρόπο και θα περιορίζει την επεξεργασία που ακολουθεί μόνο σε συγκεκριμένες περιοχές. Αυτή η ανάγκη ταιριάζει απόλυτα με τους μηχανισμούς οπτικής προσοχής, που αναλύσαμε στα προηγούμενα κεφάλαια.

Επεκτείνουμε το μοντέλο χωρικής επιλεκτικής οπτικής προσοχής με χρήση κύριου χάρτη της Ενότητας 2.3.2 και αντιμετωπίζουμε την ακολουθία εικόνων σαν ένα όγκο τριών διαστάσεων στον χώρο με την τρίτη διάσταση να αντιστοιχεί στον χρόνο, όπως και στο Κεφάλαιο 4. Ισχυριζόμαστε ότι η οπτική προσοχή σε χωροχρονική πληροφορία περιγράφει αποδοτικά το περιεχόμενο της ακολουθίας με μορφή χωροχρονικών περιοχών ενδιαφέροντος και μπορεί να περιορίσει την ανάλυση που ακολουθεί μόνο σε αυτές. Αξιολογούμε ποσοτικά και ποιοτικά το προτεινόμενο μοντέλο μέσα από εφαρμογές κατηγοριοποίησης/αναγνώρισης σκηνής και ανίχνευσης περιοχών ενδιαφέροντος [104, 103, 95, 96, 91].

5.2 Υπολογισμός όγκου ενδιαφέροντος

Στο Σχήμα 5.1 συνοψίζεται η συνολική διαδικασία που ακολουθείται στο προτεινόμενο μοντέλο. Η ακολουθία εισόδου αναλύεται σε όγκους χαρακτηριστικών (φωτεινότητα, χρώμα, χωροχρονική κατευθυντικότητα), οι οποίοι μετά από επεξεργασία με 3Δ τελεστές κέντρου-περιφέρειας συνδυάζονται για την δημιουργία του κύριο όγκου. Η συνολική προσέγγιση μπορεί να χωριστεί στα στάδια εξαγωγής χαρακτηριστικών, δημιουργίας ενδιάμεσων όγκων ενδιαφέροντος και στην δημιουργία του τελικού



Σχήμα 5.1: *Αρχιτεκτονική του προτεινόμενου μοντέλου χωροχρονικής προσοχής*

κύριου όγκου.

5.2.1 Εξαγωγή Χαρακτηριστικών

Η παρακολούθηση γεγονότων έχει νόημα μόνο όταν η εξέλιξη τους είναι σχετικά σταθερή στην διάρκεια της ακολουθίας. Απότομες μεταβολές λόγω αλλαγής σκηνής ή γενικότερα αλλαγή οπτικής γωνίας της κάμερας μπορεί να δημιουργήσουν συγκεχυμένα χωροχρονικά μοτίβα. Επομένως, το πρώτο βήμα επεξεργασίας είναι η κατάτμηση της ακολουθίας εισόδου σε πλάνα με την χρήση μίας κοινής τεχνικής ανίχνευσης πλάνου [86]. Ο αριθμός των καρέ που θα επεξεργαστούμε με το προτεινόμενο μοντέλο μπορεί να είναι ίσος με τον αριθμό των καρέ της σκηνής ή μικρότερος, αλλά αρκετά μεγάλος ώστε να αναπαριστά ικανοποιητικά τα χωροχρονικά γεγονότα της σκηνής. Κάθε ένα λοιπόν από τα πλάνα αντιστοιχεί σε έναν τρισδιάστατο όγκο, ο οποίος αποτελείται από ένα σύνολο σημείων Q στον 3Δ χώρο με κάθε σημείο $q(x, y, t)$ να ορίζεται από τις καρτεσιανές συντεταγμένες του στον χώρο-χρόνο. Η τρίτη διάσταση αντιστοιχεί στην χρονική εξέλιξη των καρέ. Με αυτήν την αναπαράσταση, ένα κινούμενο αντικείμενο καταλαμβάνει μία 3Δ περιοχή στον διακριτό χώρο που ορίσαμε πριν.

Ακολουθώντας την παραπάνω αναπαράσταση η ακολουθία αποσυντίθεται σε μία σειρά χαρακτηριστικών, όπως φωτεινότητα, χρώμα και χωροχρονική κατευθυντικότητα. Για τα κανάλια φωτεινότητας και χρώματος υιοθετούμε την θεωρία διπλής χρωματικής αντίθεσης[22] σύμφωνα με την οποία υπάρχουν τρεις αντίπαλες

διαδικασίες[23] ή αντίπαλα κανάλια, τα οποία δημιουργούνται από τους τρεις τύπους κώνων (κόκκινος, πράσινος, μπλε) του ανθρώπινου αμφιβληστροειδούς [54]. Το μάτι δημιουργεί ένα αχρωματικό κανάλι ("μαύρο-λευκό" ή κανάλι φωτεινότητας), ένα "κόκκινο/πράσινο" με την διαφορά των κόκκινων και πράσινων κώνων και ένα "κίτρινο/μπλε" με το άθροισμα των κόκκινων-πράσινων (κίτρινο κανάλι) και την διαφορά από τους μπλε κώνους. Το κανάλι φωτεινότητας λαμβάνεται συνήθως με την μέση τιμή των αποκρίσεων των τριών διαφορετικών κώνων. Αυτή η επεξεργασία εξηγεί διάφορα οπτικά φαινόμενα, όπως το γιατί είμαστε σε θέση να αντιληφθούμε ταυτόχρονα τα κοκκινωπά μπλε και τα κοκκινωπά-πράσινα χρώματα. Ο βαθμός στον οποίο αυτά τα αντίπαλα κανάλια προσελκύουν την προσοχή των ανθρώπων έχει ερευνηθεί λεπτομερώς τόσο για τα βιολογικά [135] όσο και για τα υπολογιστικά πρότυπα της προσοχής [71]. Σύμφωνα με το σχήμα που περιγράφηκε αν τα r, g, b είναι οι όγκοι χρώματος, τότε ο όγκος φωτεινότητας στην μηδενική κλίμακα προκύπτει ως

$$F_1 = \frac{1}{3} \cdot (r + g + b) \tag{5.1}$$

και ο χρωματικός ως

$$F_2 = RG_0 + BY_0 \tag{5.2}$$

όπου

$$RG = R - G \tag{5.3}$$
$$BY = B - Y \tag{5.4}$$

και

$$\begin{aligned} R &= r - \frac{(g+b)}{2} \\ G &= g - \frac{(r+b)}{2} \\ B &= b - \frac{(r+g)}{2} \\ Y &= \frac{(r+g)}{2} - \frac{|r-g|}{2} - b \end{aligned} \tag{5.5}$$

5.2.1.1 Χωροχρονική κατευθυντικότητα

Σε μία προσπάθεια εύρεσης ενός τρόπου υπολογισμού της κατευθυντικότητας που να είναι όσο το δυνατόν πιό κοντά στην ευαισθησία του υποδεκτικού πεδίου των εκλεκτικών νευρώνων προσανατολισμού του ανθρώπινου οπτικού φλοιού, οι Itti et al χρησιμοποίησαν τα φίλτρα Gabor για να ανιχνεύσουν τις τοπικές πληροφορίες προσανατολισμού. Τα φίλτρα Gabor πρέπει να έχουν στενή ζώνη περατότητας στο πεδίο συχνοτήτων προκειμένου να έχουν ευκρίνεια στην ανίχνευση κατευθυντικότητας. Επομένως, σύμφωνα με την αρχή αβεβαιότητας το φίλτρο πρέπει να έχει μεγάλη κλίμακα στο χωρικό πεδίο. Στην περίπτωση της εστίασης οπτικής προσοχής η απόλυτη ακρίβεια στον εντοπισμό της κατευθυντικότητας δεν είναι απαραίτητη προϋπόθεση για την σωστή λειτουργία του μοντέλου. Το πραγματικό μειονέκτημα είναι η θετική ασυμμετρία κατανομής (skewness) στις αποκρίσεις των φίλτρων [37], η οποία γίνεται πιο έντονη στην τριδιάστατη έκδοσης τους. Η απόκριση ενός τρισδιάστατου φίλτρου Gabor που εφαρμόζεται σε μία ακολουθία μπορεί να συσχετιστεί με πληροφορία

κίνησης. Όταν μόνο μια κίνηση είναι παρούσα στην ακολουθία, το μέγιστο (στη σφαίρα προσανατολισμού) είναι καλά εντοπισμένο παρά την ασυμμετρία κατανομής. Αλλά αν υπάρχουν περισσότερες κινήσεις, η επικάλυψη των διαφορετικών αποκρίσεων θα επηρεάσει τις θέσεις των μέγιστων τιμών [151].

Το προτεινόμενο μοντέλο οπτικής προσοχής χρησιμοποιεί καθοδηγήσιμα φίλτρα[24] για τον υπολογισμό της κατευθυντικότητας, τα οποία προκύπτουν από τον γραμμικό συνδυασμό ενός συνόλου "φίλτρων βάσης" [34]. Στο προτεινόμενο μοντέλο χρησιμοποιήσαμε την δεύτερη παράγωγο τριδιάστατων Γκαουσσιανών φίλτρων G_2^θ για τον υπολογισμό της προσανατολισμένης ενέργειας [24].

$$E_v^\theta = [G_2^\theta * F_1]^2 + [H_2^\theta * F_1]^2 \tag{5.6}$$

όπου $\theta \equiv \alpha, \beta, \gamma$ και α, β, γ αντιστοιχούν στα συνημίτονα κατεύθυνσης στα οποία προσανατολίζονται τα φίλτρα. Οι χωροχρονικοί προσανατολισμοί σχετίζονται άμεσα με την ανάλυση κινήσεων και παρέχουν τη βάση για τους ενεργειακούς μηχανισμούς που εξάγουν και αναλύουν την κίνηση [34] [42]. Οι Freeman *et al.* παραμετροποίησαν τον προσανατολισμό του τριδιάστατου πυρήνα των φίλτρων με τα συνημίτονα κατεύθυνσης μεταξύ του άξονα προσανατολισμού και των κύριων αξόνων α, β, γ, όπως είδαμε πριν [34]. Οι τρεις κατευθυντικές γωνίες εκφράζονται σε σφαιρικές συντεταγμένες ως:

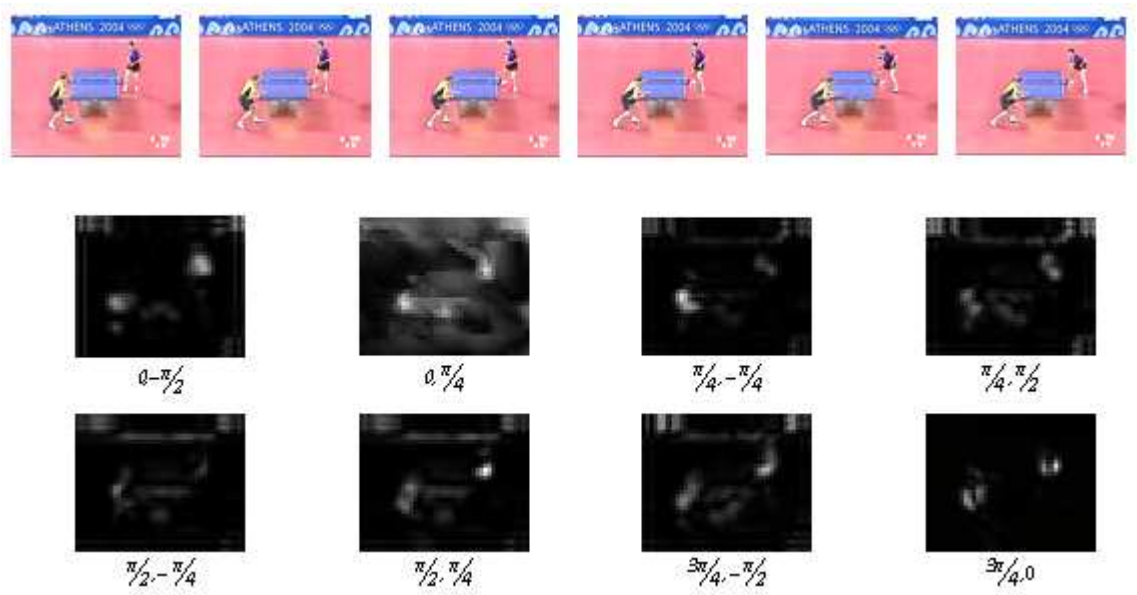
$$\begin{aligned} \alpha &= cos(\theta)sin(\phi) \\ \beta &= sin(\theta)sin(\phi) \\ \gamma &= cos(\phi) \end{aligned} \tag{5.7}$$

Με την χρήση των χωροχρονικών φίλτρων προσανατολισμού, ο εντοπισμός ενδιαφέροντων γεγονότων σε έναν μεγάλο αριθμό καρέ μπορεί να γίνει χωρίς την ανάγκη, π.χ., μιας υπολογιστικά ακριβής εκτίμησης οπτικής ροής που προϋποθέτει χωρική συνέχεια (π.χ. παράγωγοι εικόνας). Το Σχήμα 5.2 απεικονίζει αποτελέσματα για γειτονικά καρέ της ίδιας ακολουθίας, όπου τα αντικείμενα ενδιαφέροντος (παίχτες) κινούνται σε ποκίλες κατευθύνσεις. Κάθε εικόνα αντιστοιχεί σε έναν συγκεκριμένο χωροχρονικό προσανατολισμό. Αν και μέρος των προσανατολισμένων φίλτρων εντοπίζει ακριβώς τις μετακινήσεις στη σκηνή, το πρόβλημα της συγχώνευσης/κανονικοποίησης όλων των προσανατολισμών παραμένει. Με τις τιμές της γωνίας θ που επιλέχθηκαν παράγονται 20 όγκοι διαφορετικών χωροχρονικών προσανατολισμών. Όπως αναφέραμε, αυτοί οι όγκοι πρέπει να συνδυαστούν για να παράγουν έναν ενιαίο όγκο προσανατολισμού.

Η συγχώνευση των χωροχρονικών όγκων μπορεί να αποδειχτεί θορυβώδης εξαιτίας του μεγάλου αριθμού διαφορετικών προσανατολισμών. Προτείνουμε έναν τελεστή που βασίζεται στην ανάλυση κύριων συνιστωσών[25] (PCA), έναν μετασχηματισμό που συνδέεται συνήθως με την στατιστική πολλών μεταβλητών, και παράγουμε τον τελικό χωροχρονικό όγκο προσοχής προσανατολισμού ως

$$F_3 = \Phi_{PCA}[E_{3d}(\theta, \phi)] \tag{5.8}$$

όπου Φ_{PCA} είναι ο μετασχηματισμός PCA. Ο τελεστής ανάλυσης κύριων συνιστωσών βρίσκει τους ορθογώνιους γραμμικούς συνδυασμούς ενός συνόλου χαρακτηριστικών n που μεγιστοποιούν την μεταβλητότητα, με αποτέλεσμα να αναπαριστά το μεγαλύτερο μέρος της αρχικής μεταβλητότητας σε έναν ίσο ή μικρότερο αριθμό διαστάσεων. Στο



Σχήμα 5.2: *Γειτονικά καρέ που δείχνουν τις χωροχρονικές κινήσεις της ακολουθίας και η έξοδος των φίλτρων για συγκεκριμένες χωροχρονικές κατευθύνσεις.*

μετασχηματισμένο χώρο η πρώτη διάσταση περιέχει την μεγαλυτερη μεταβλητότητα των δεδομένων, η δεύτερη την αμέσως μικρότερη κοκ. Συνήθως μέρος των συνιστωσών που περιέχουν την μεγαλύτερη μεταβλητότητα χρησιμοποιούνται για την αντιπροσωπεύση της οπτικής εισόδου [112, 91]. Αρχικά, δημιουργείται ένας πίνακας S από το σύνολο των n block διανυσμάτων, τα οποία αντιστοιχούν στους $n(n = 20)$ προσανατολισμούς και υπολογίζεται το n-διάστατο διάνυσμα μέσης τιμής μ. Στη συνέχεια υπολογίζονται οι ιδιοτιμές λ_i και τα ιδιοδιανύσματα e_i με $i = \{1, ..., n\}$, τα οποία ταξινομούνται με βάση τις τιμές των αντίστοιχων ιδιοτιμών. Ο πίνακας προβολής W με διαστάσεις $n \times n'$ δημιουργείται έτσι ώστε να περιέχει n' ιδιοδιανύσματα $e_i, ..., e_{n'}$ που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές $\lambda_i, ..., \lambda_{n'}$. Επομένως, το σύνολο των αρχικών δεδομένων μετασχηματίζεται ως

$$S' = W^t(S - \mu) \tag{5.9}$$

έτσι ώστε τα αρχικά δεδομένα να είναι ασυσχέτιστα μετά τον μετασχηματισμό [29]. Τελικά, για την δημιουργία του συνολικού όγκου κατευθυντικότητας διατηρούμε τις δύο πρώτες κύριες συνιστώσες και υπολογίζουμε την μέση τιμή τους. Οι συνιστώσες που απορρίπτονται αντιστοιχούν συνήθως σε θόρυβο [35]. Ένα οπτικό παράδειγμα συνδυαμού με μέση τιμή και με PCA απεικονίζεται στο Σχήμα 5.3.

5.2.2 Δημιουργία κύριου όγκου

Μετά την αποσύνθεση της ακολουθίας στο σύνολο των χαρακτηριστικών, οι όγκοι που προκύπτουν χρησιμοποιούνται για την δημιουργία χωροχρονικών πυραμίδων. Κάθε κλίμακα αυτής της ογκομετρικής πυραμίδας προκύπτει με εξομάλυνση με Γκαουσσιανό φίλτρο και υποδειγματοληψία. Πρόκειται για γενίκευση της Γκαουσσιανής πυραμίδας για εικόνες στον χωροχρόνο [103, 16]. Με αυτήν την διαδικασία παράγεται ένα σύνολο όγκων $F_{i,\ell}$, $\ell \in \{1, 2, ..., L\}$, όπου L είναι η μέγιστη κλίμακα της πυραμίδας.

Σχήμα 5.3: *Παράδειγμα συνδυασμού χαρτών με PCA. Ανά γραμμή: αρχικά καρέ, συνδυασμός με μέση τιμή, συνδυασμός με PCA*

Στο Κεφάλαιο 3 αναφερθήκαμε στα πεδία κέντρου-περιφέρειας, τα οποία επιτελούν μία από τις βασικότερες λειτουργίες του ανθρώπινου οπτικού συστήματος: την διαφοροποίηση των έντονων οπτικών ερεθισμάτων από τα υπόλοιπα. Υπάρχουν διαφορετικοί τύποι υποδεκτικών τομέων με τον κέντρου-περιφέρειας να είναι ο σημαντικότερος. Αυτά τα υποδεκτικά πεδία χαρακτηρίζονται από την κυκλική συμμετρία και την παρουσία δύο ξεχωριστών, αμοιβαία-ανταγωνιστικών υπο-περιφερειών, του κέντρου και της περιφέρειας. Τα σχεδιαγράμματα ευαισθησίας τους είναι πιθανότατα Γκαουσσιανή, δηλαδή η απόκριση του κέντρου είναι μικρότερη κοντά στα όρια απ'ότι στο μεσαίο σημείο του. Αυτή η τοπολογία βοηθά στην ανίχνευση των θέσεων που ξεχωρίζουν τοπικά από την γειτονιά τους. Ο τελεστής κέντρου-περιφέρειας, ο οποίος συμβολίζεται ως $\ominus$, εφαρμόζεται στο μοντέλο ως η διαφορά μεταξύ των υψηλών και χαμηλών κλιμάκων της πυραμίδας. Το κέντρο είναι ένα στοιχείο του όγκου στην κλίμακα c και η περιφέρεια είναι το αντίστοιχο στοιχείο στην κλίμακα s. Δημιουργούνται τρεις όγκοι χαρακτηριστικών:

$$\begin{aligned} F_{1,c/s} &= |F_{1,c} \ominus F_{1,s}| \\ F_{2,c/s} &= |(R_c - G_c) \ominus (G_s - R_s)| + |(B_c - Y_c) \ominus (Y_s - B_s)| \\ F_{3,c/s} &= |F_{3,c} \ominus F_{3,s}| \end{aligned} \tag{5.10}$$

όπου c/s είναι η κλίμακα που προκύπτει από τον "συνδυασμό" των κλιμάκων c και s.

Μετά την επεξεργασία των χαρακτηριστικών με τους τελεστές κέντρου-περιφέρειας προκύπτει ένας αριθμός πυραμίδων από όγκους που κωδικοποιούν την σημαντικότητα κάθε χαρακτηριστικού σε διαφορετική κλίμακα. Ο επιθυμητός στόχος είναι η δημιουργία αντίστοιχων ενδιάμεσων όγκων ενδιαφέροντος, οι οποίοι θα συνδυαστούν για να προκύψει ο κύριος όγκος. Οι όγκοι αυτοί, C_1 για την ένταση, C_2 για το χρώμα και C_3 για την 3Δ κατευθυντικότητα, υπολογίζονται με χωροχρονική παρεμβολή σε μια ενδιάμεση κλίμακα της πυραμίδας και με πρόσθεση στοιχείου με στοιχείο [48, 49]:

$$\overline{C_i} = \bigotimes_{c=2}^{\sigma_1} \bigotimes_{s=c+1}^{\sigma_1} N(F_{i,c/s}) \tag{5.11}$$

όπου N είναι ο τελεστής κανονικοποίησης, ο οποίος υποβαθμίζει ασήμαντες περιοχές κάθε όγκου. Το κίνητρο για τη δημιουργία των χωριστών καναλιών και της μεμονωμένης κανονικοποίησής προκύπτει από την υπόθεση των Koch και Ullman [60] ότι ενώ τα παρόμοια χαρακτηριστικά ανταγωνίζονται έντονα για να την δημιουργία του κύριου όγκου, τα διαφορετικά συμβάλλουν ανεξάρτητα στον κύριο όγκο.

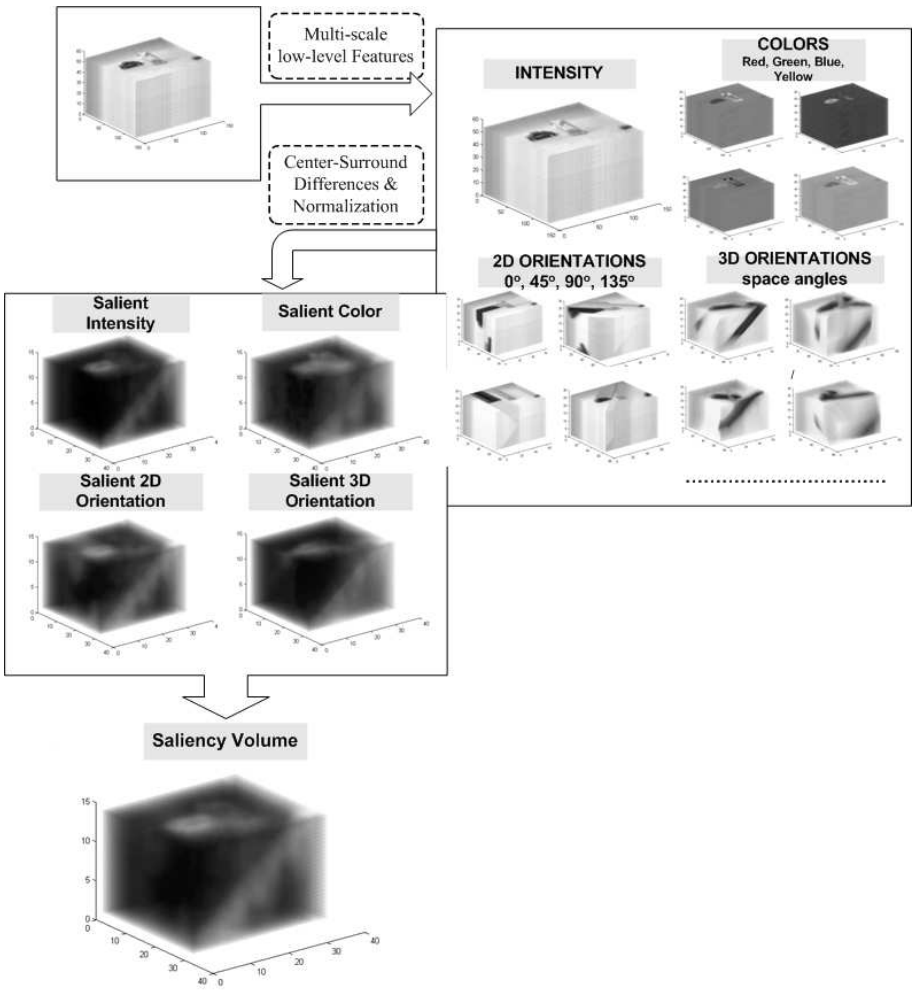
Για να μειωθεί ο αναπόφευκτος θόρυβος χρησιμοποιήσαμε μία χωροχρονική επέκταση της απλής διαδικασίας των Itti *et al.* [48]. Σύμφωνα με αυτήν την επέκταση ο τελεστής κανονικοποίησης N αποτελείται από τα εξής βήματα: 1) κανονικοποίηση των χωροχρονικών όγκων στην ίδια δυναμική περιοχή, προκειμένου να εξαλειφθούν οι διαφορές εύρους μεταξύ τους, 2) εύρεση του συνολικού μέγιστου M και το μέσο όρο $\overline{m}$ όλων των τοπικών μεγίστων, 3) πολλαπλασιασμός κάθε όγκου με την τιμή $(M - \overline{m})^2$. Για λόγους υπολογιστικής απλότητας χρησιμοποιήσαμε το μορφολογικό μετασχηματισμό grayscale top-hat $THT(V, S) = V - (V \circ S)$ για την εύρεση των τοπικών μεγίστων κάθε όγκου. Η σύγκριση της μέγιστης περιοχής δραστηριότητας με το μέσο όρο των υπόλοιπων μεγίστων μετρά πόσο σημαντική, σε ένταση, είναι η πιο ενεργός περιοχή. Όταν αυτή η διαφορά είναι μεγάλη, προάγουμε έντονα το χάρτη. Ο τελικός όγκος S προκύπτει με απλό άθροισμα των επιμέρους κανονικοποιημένων αποτελεσμάτων:

$$S = \frac{1}{3} \cdot \sum_{i=1}^{3} N(C_i) \tag{5.12}$$

Το Σχήμα 5.4 απεικονίζει σε μία πιο παραστατική έκδοση την διαδικασία που περιγράψαμε.

5.3 Εφαρμογές

Σε αυτήν την Ενότητα παρουσιάζουμε ποιοτικά και ποσοτικά αποτελέσματα του προτεινόμενου μοντέλου. Χρησιμοποιήσαμε γνωστές ακολουθίες για να δείξουμε την εφαρμογή του μοντέλου και να αναλύσουμε ποιοτικά τα αποτελέσματα του. Για να αξιολογήσουμε ποσοτικά την μέθοδο επιλέξαμε δύο διαφορετικά πεδία εφαρμογής, τα οποία θεωρούμε ότι μπορούν να επωφεληθούν από την χωροχρονική ανάλυση. Το πρώτο μας πείραμα σχετίζεται με την αξιόπιστη αναγνώριση σκηνής και το δεύτερο με την γρήγορη και αποτελεσματική ανίχνευση περιοχών σε εφαρμογές επίβλεψης χώρων.



Σχήμα 5.4: *Μία πιο παραστατική έκδοση της προτεινόμενης αρχιτεκτονικής*

5.3.1 Παραδείγματα

Στα Σχήματα 5.5, 5.6 απεικονίζονται αποτελέσματα της μεθόδου για διαφορετικές ακολουθίες. Κάθε αποτέλεσμα αντιστοιχεί στην τομή του όγκου που αναγράφεται. Στα Σχήματα 5.5, 5.6 η εκάστοτε τομή του κύριου χάρτη υπερτίθεται στο αρχικό καρέ, ώστε να φαίνονται μόνο οι σημαντικές περιοχές που προκύπτουν. Τα Σχήματα 5.5, 5.6 δείχνουν ενδεικτικά αποτελέσματα για τις ακολουθίες "coast guard" και "table tennis" που έχουν χρησιμοποιηθεί ευρέως στην βιβλιογραφία. Στην πρώτη ακολουθία οι περιοχές ενδιαφέροντος αντιστοιχούν κυρίως στα δύο πλοιάρια που διασχίζουν το ποτάμι. Συνολικά στην ακολουθία παρατηρούνται οι κινήσεις των δύο αντικειμένων και η συνολική κίνηση της κάμερας, η οποία αρχικά ακολουθεί το μικρό και μετά το μεγαλύτερο πλοιάριο. Το προτεινόμενο μοντέλο εστιάζει σωστά στην χωροχρονική περιοχή που αντιστοιχεί στα πλοιάρια και δεν επηρεάζεται από την κάμερα, όπως δείχνουν τα αποτελέσματα του Σχήματος 5.5.

Στην δεύτερη ακολουθία απεικονίζεται ένα απόσπασμα από επιτραπέζια αντισφαίριση, το οποίο περιέχει κινήσεις των παιχτών, της κάμερας (zoom in/out) και εμφάνιση νέων αντικειμένων στην διάρκεια της ακολουθίας (διαφημιστικό ποστερ στον τοίχο και ο δεύτερος παίχτης). Το Σχήμα 5.6 δείχνει την έξοδο του μοντέλου για κάποια ενδεικτικά καρέ. Στην δεύτερη γραμμή του ίδιου Σχήματος απεικονίζονται και τα αντίστοιχα αποτελέσματα της τεχνικής των Black και Anandan [5] για υπολογισμό

καρέ 74 καρέ 99 καρέ 149

Σχήμα 5.5: *Αποτελέσματα της χωροχρονικής οπτικής προσοχής για την ακολουθία "coast-guard". Η συνολική κίνηση της κάμερας δεν επηρεάζει την εστίαση στις περιοχές ενδιαφέροντος.*

οπτικής ροής. Ο σκοπός αυτής της σύγκρισης είναι να τονιστεί η ικανότητα της προτεινόμενης τεχνικής να μην επηρεάζεται από αλλαγές στην κίνηση της κάμερας και να παραμένει εστιασμένη στα αντικείμενα ενδιαφέροντος, όπως είναι οι παίχτες και οι διαφημίσεις στους τοίχους. Στο καρέ 55 η οπτική ροή επηρεάζεται σημαντικά από το zoom-out της κάμερας ενώ ο κύριος χάρτης συνεχίζει να αναδεικνύει τις σημαντικές περιοχές. Η αντίστοιχη παρατήρηση ισχύει και για τα επόμενα καρέ.

καρέ 55 καρέ 115 καρέ 122

Σχήμα 5.6: *Αποτελέσματα της χωροχρονικής οπτικής προσοχής για την ακολουθία "table-tennis". Κατά σειρές: αρχικά καρέ, κύριοι χάρτες, χάρτες οπτικής ροής.*

5.3.2 Κατηγοριοποίηση σκηνής

Η αξιολόγηση της απόδοσης ενός ανιχνευτή ενδιαφέροντος είναι δύσκολη λόγω της υποκειμενικότητας της έννοιας "ενδιαφέρον". Σε αυτήν την ενότητα επιχειρούμε να αξιολογήσουμε ποσοτικά το προτεινόμενο μοντέλο μετρώντας την βελτίωση που επιτυγχάνει στα ποσοστά αναγνώρισης σκηνής. Στην συνέχεια παραθέτουμε λεπτομέρειες για το σύνολο ακολουθιών που χρησιμοποιήσαμε και τα αποτελέσματα για την προτεινόμενη μέθοδο. Σε αυτήν την Ενότητα αξιολογούμε την μέθοδο συγκρίνοντας την με συνολική κατηγοριοποίση σκηνής (εξαγωγή χαρακτηριστικών από όλη την ακολουθία και όχι μόνο από περιοχές ενδιαφέροντος), ενώ στο Κεφάλαιο 6 (Ενότητα 6.6.1) παραθέτουμε τα αποτελέσματα όλων των μεθόδων που έχουμε προτείνει σε σχέση με κλασσικές τέχνικες ανίχνευσης ενδιαφέροντος.

Για την εφαρμογή σε αναγνώριση σκηνής επιλέξαμε ακολουθίες από το πεδίο των αθλημάτων. Ακολουθίες εφτά διαφορετικών αθλημάτων κατηγοριοποιήθηκαν χειρωνακτικά στις εξής κατηγορίες: ποδόσφαιρο (SO), κολύμβηση (SW), καλαθοσφαίριση (BA), πυγμαχία (BO), σνούκερ (SN), αντισφαίριση (TE) και επιτραπέζια αντισφαίριση (TB). Οι συντομογραφίες στις παρενθέσεις χρησιμοποιούνται στους πίνακες αποτελεσμάτων που ακολουθούν. Οι ακολουθίες είναι στο μεγαλύτερο ποσοστό από του Ολυμπιακούς αγώνες της Αθήνας το 2004. Κάθε κατηγορία περιλαμβάνει μακρινές και κοντινές λήψεις παικτών και γηπέδου, καθώς και καρέ που περιέχουν ταυτόχρονα περιοχές γηπέδου, παικτών και ακροατηρίου. Οι ακολουθίες διαρκούν περίπου 6-7 δευτερόλεπτα και έχουν ίδιο χωρικό μέγεθος.

Ο τελικός κύριος όγκος δημιουργήθηκε με την προτεινόμενη τεχνική. Για την κατηγοριοποίηση είναι απαραίτητη η κατάτμηση του κύριου όγκου και η εξαγωγή κατάλληλων χαρακτηριστικών, τα οποία θα τροφοδοτήσουν τον ταξινομητή. Το

Σχήμα 5.7 δείχνει ενδεικτικά αποτελέσματα για τις εφτά κατηγορίες αθλητικών ακολουθιών που χρησιμοποιήσαμε. Η πρώτη στήλη περιέχει τα αρχικά καρέ, η δεύτερη την τομή του κύριου όγκου που αντιστοιχεί στο αντίστοιχο καρέ (οι φωτεινές περιοχές αντιστοιχούν στις πιο σημαντικές) και η τρίτη το αποτέλεσμα της κατάτμησης (η μαύρη είναι η λιγότερο σημαντική περιοχή). Λεπτομέρειες για τις τεχνικές που χρησιμοποιούνται δίνονται στις επόμενες ενότητες.

5.3.2.1 Κατάτμηση και εξαγωγή χαρακτηριστικών

Η ακριβής κατάτμηση της εικόνας σε αντικείμενα δεν είναι απαραίτητη για την συνολική κατηγοριοποίηση των ακολουθιών. Η κατάτμηση που χρησιμοποιούμε βασίζεται στην ομαδοποίηση των στοιχείων με k-means. Το κριτήριο ομαδοποίησης είναι η τιμή ενδιαφέροντος κάθε στοιχείου και ο αριθμός των κλάσεων k που θα προκύψουν είναι συγκεκριμένος. Τελικά οι ομάδες στοιχείων που προκύπτουν ταξινομούνται με σειρά μειούμενης μέσης τιμής ενδιαφέροντος και απορρίπτεται αυτή με την χαμηλότερη τιμή, καθώς σχετίζεται συνήθως με περιοχές που υπάρχουν στην μεγαλύτερη διάρκεια της ακολουθίας, όπως μεγάλο μέρος του γηπέδου και του ακροατηρίου. Ο ισχυρισμός είναι απλός και σχετίζεται με την κύρια ιδέα της οπτικής προσοχής: οι λιγότερο σημαντικές περιοχές της εισόδου δεν αντιπροσωπεύουν επαρκώς την είσοδο και κατά συνέπεια τα χαρακτηριστικά που θα εξαχθούν από αυτές ίσως δημιουργήσουν ασάφεια στον ταξινομητή ικανή να οδηγήσει σε αυξημένο λάθος κατηγοριοποίησης. Στην τρίτη στήλη του Σχήματος 5.7 απεικονίζονται ενδεικτικά αποτελέσματα.

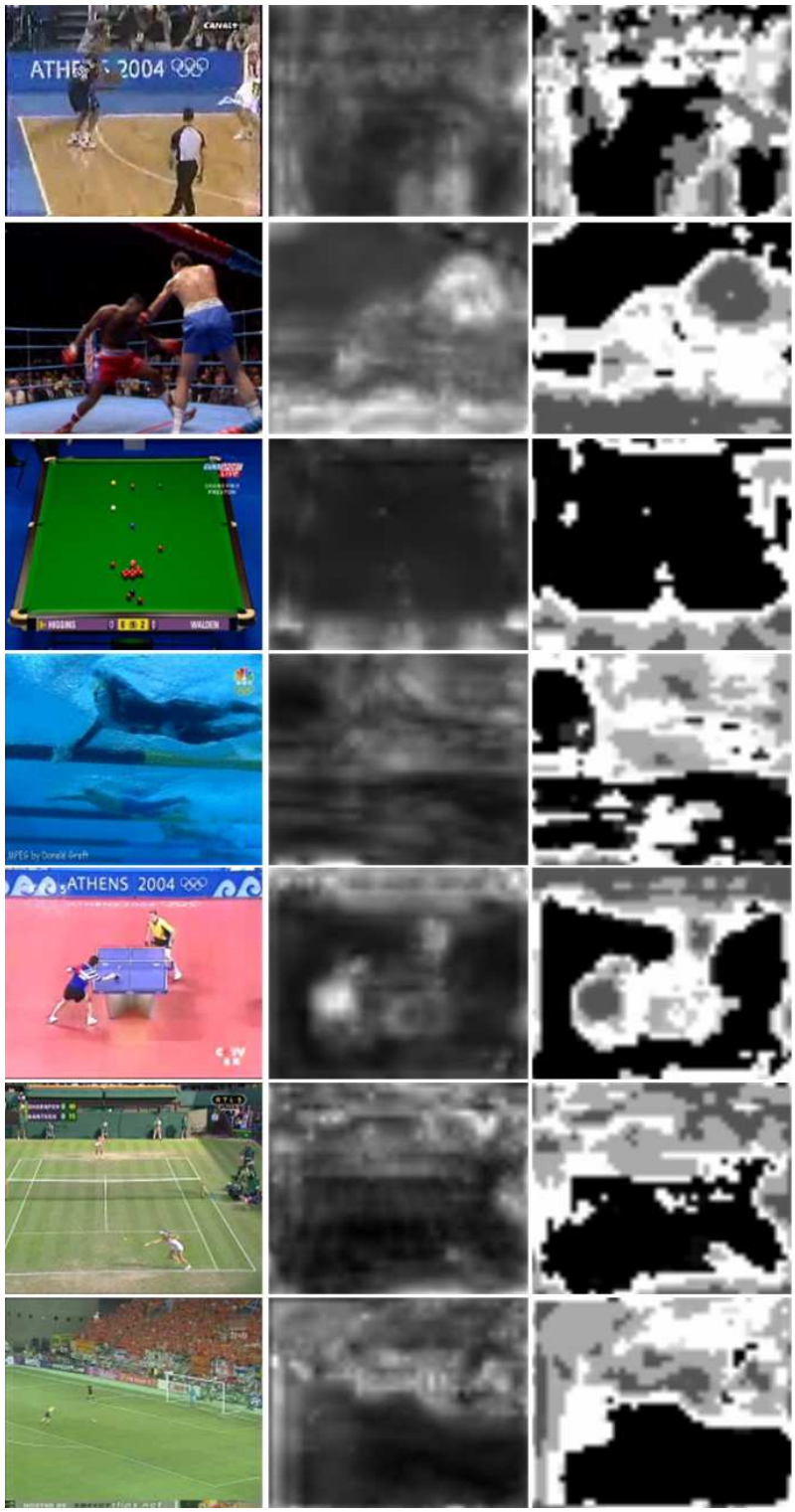
Τα χαρακτηριστικά που χρησιμοποιήσαμε είναι ιστογράμματα χρώματος και μετρήσεις υφής από τον πίνακα συνεμφάνισης. Για να διατηρήσουμε το μέγεθος του διανύσματος χαρακτηριστικών μικρό, κβαντίσαμε τα ιστογράμματα και χρησιμοποιήσαμε τέσσερις μετρήσεις υφής: εντροπία, ροπή, ενέργεια και ομοιογένεια. Ακολουθώντας την διαδικασία που περιγράψαμε πριν, κβαντίζουμε τα ιστογράμματα σε 8 στήλες για κάθε χρωματικό κανάλι (24 τιμές χαρακτηριστικών για κάθε περιοχή) και υμπεριλαμβάνουμε τις μετρήσεις υφής για τέσσερις διαφορετικές τομές κάθε υπο-όγκου ενδιαφέροντος (16 τιμές για κάθε περιοχή). Το συνολικό μέγεθος του διανύσματος είναι 120.

5.3.2.2 Κατηγοριοποίηση

Για την κατηγοριοποίηση των ακολουθιών χρησιμοποιούμε Support Vector Machines (SVM), τα οποία έχουν προταθεί για ταξινόμηση δεδομένων σε δύο κατηγορίες. Τα SVMs μεγιστοποιούν την απόσταση μεταξύ ενός υπερ-επιπέδου w και του κοντινότερου σημείου σε αυτό, με τον περιορισμό τα δείγματα των δύο κατηγοριών να βρίσκονται σε διαφορετικές κλάσεις του υπερ-επιπέδου. Τα πιο κοντινά σημεία ονομάζονται διανύσματα υποστήριξης (support vectors). Αν π.χ. έχουμε ένα σύνολο στιγμιότυπων (x_i, y_i), $i = \{1, ..., l\}$ όπου $x_i \in \Re_n$ και $y \in \{-1, 1\}$, τα SVMs απαιτούν την λύση στο εξής πρόβλημα βελτιστοποίησης:

$$min\|w\|^2 + P\sum_{i=1}^{l}\xi_i, \quad s.t.\ y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \tag{5.13}$$

όπου τα δεδομένα εκμάθησης απεικονίζονται σε έναν χώρο μεγαλύτερης διάστασης με χρήση της συνάρτησης ϕ και ο δεύτερος όρος της (5.13) είναι η συνάρτηση ποινής με παράμετρο P.



Σχήμα 5.7: *Ενδεικτικά αποτελέσματα του υπολογισμού και κατάτμησης κύριου χάρτη για τις αθλητικές ακολουθίες που χρησιμοποιήθηκαν*

Το πρόβλημα της κατηγοριοποίησης σε περισσότερες των δύο κατηγορίες λύνεται με την ανάλυση του σε πολλά μικρότερα προβλήματα δυαδικής κατηγοριοποίησης. Η ανάλυση ενός-ενάντιον-όλων χρησιμοποιείται συχνά σε τέτοιες περιπτώσεις. Σύμφωνα με αυτήν την τακτική η κατηγοριοποίηση σε k κατηγορίες επιτυγχάνεται με την εκπαίδευση k ταξινομητών να διαφοροποιούν τα παραδείγματα μίας κλάσης από όλες

Πίνακας 5.1: *Πίνακας σύγχυσης για τα δεδομένα δοκιμής χωρίς εξαγωγή ενδιαφέροντος (***συνολικό σφάλμα: 26.37%***)*

	SN	SW	BA	BO	SO	TE	TB
SN	35	0	5	0	20	0	0
SW	0	50	0	0	0	0	0
BA	0	0	70	5	5	10	10
BO	0	0	5	35	0	0	0
SO	10	0	5	0	60	5	0
TE	0	0	20	0	5	50	0
TB	0	0	15	0	0	0	35
Prec	**0,778**	**1,000**	**0,583**	**0,875**	**0,667**	**0,769**	**0,778**
Rec	**0,583**	**1,000**	**0,700**	**0,875**	**0,750**	**0,667**	**0,700**

τις υπόλοιπες. Επομένως, σε κάποια άγνωστη είσοδο ενεργοποιούνται όλοι οι k ταξινομητές και επιλέγεται αυτός με την μεγαλύτερη τιμή εξόδου.

Στα πειράματα μας, εκπαιδεύουμε τα SVMs με χρήση πυρήνα Radial-Basis-Function (RBF) μετά από κατάλληλη επιλογή παραμέτρων. Για αυτήν την επιλογή κάνουμε εξαντλητική αναζήτηση σε ένα πεδίο τιμών για την παράμετρο P ως $P = 2^0, 2^1, 2^2, 2^3, 2^4$ με 5-απλή επικύρωση[26]. Αφού καταλήξουμε στην τιμή που δίνει το χαμηλότερο λάθος, επαναλαμβάνουμε την αναζήτηση σε ένα εύρος τιμών γύρω από την επιλεγμένη και επιλέγουμε την τελική τιμή που είναι ίδια για όλους τους ταξινομητές.

5.3.2.3 Πειραματικά αποτελέσματα

Στα αρχικά πειράματα συγκρίναμε την κατηγοριοποίηση με τιμές ενδιαφέροντος με την συνολική κατηγοριοποίηση των ακολουθιών. Η διαφορά επομένως είναι η εξής: στην πρώτη περίπτωση εντοπίζουμε χωροχρονικές περιοχές ενδιαφέροντος και εξάγουμε χαρακτηριστικά μόνο από αυτές, ενώ στην δεύτερη περίπτωση τα χαρακτηριστικά εξάγονται από ολόκληρο τον όγκο της ακολουθίας. Αποτελέσματα σε μορφή πινάκων σύγχυσης[29] δίνονται στους Πίνακες 5.1 και 5.2. Κάθε γραμμή των πινάκων δείχνει τα αποτελέσματα της εκάστοτε κατηγορίας, ενώ οι δύο τελευταίες δίνουν τις τιμές ακρίβειας/επανάκλησης για κάθε κατηγορία. Π.χ. η πρώτη γραμμή του Πίνακα 5.1 μας δίνει την εξής πληροφορία: 20 ακολουθίες σνούκερ ταξινομούνται λανθασμένα σε ακολουθίες ποδοσφαίρου και 5 σε ακολουθίες καλαθοσφαίρισης.

Ο Πίνακας 5.1 περιέχει τα αποτελέσματα χωρίς την χρήση οπτικής προσοχής, ενώ ο Πίνακας 5.2 δείχνει τα αποτελέσματα με την χρήση του προτεινόμενου μοντέλου. Στην πρώτη περίπτωση το συνολικό σφάλμα ταξινόμησης είναι 26.37%, ενώ στην δεύτερη 15.38%. Αν και η απόλυτη διαφορά στο σφάλμα δεν είναι πολύ μεγάλη, υπάρχει ένα ενδιαφέρον αποτέλεσμα που υποστηρίζει τον ισχυρισμό μας ότι το μοντέλο οπτικής προσοχής ανιχνεύει περιοχές που αντιπροσωπεύουν καλύτερα την ακολουθία. Ζευγάρια κατηγοριών, όπως το ποδόσφαιρο-σνούκερ ή η καλαθοσφαίριση-επιτραπέζια αντισφαίριση έχουν παρόμοια γενικά χαρακτηριστικά εξαιτίας της ομοιότητας του χρώματος του γηπέδου και ίσως των διαφημίσεων των αγώνων της Αθήνας-2004 που είναι μπλε-άσπρες. Αυτός είναι και ο λόγος που τα στατιστικά του Πίνακα 5.1 που αντιστοιχούν σε αυτές τις κατηγορίες δεν είναι ικανοποιητικά. Στον Πίνακα 5.2 όμως

Πίνακας 5.2: *Πίνακας σύγχυσης για τα δεδομένα δοκιμής με εξαγωγή ενδιαφέροντος (συνολικό σφάλμα: 15.38%)*

	SN	SW	BA	BO	SO	TE	TB
SN	50	0	5	0	5	0	0
SW	0	50	0	0	0	0	0
BA	5	0	75	10	0	10	0
BO	0	0	0	35	5	0	0
SO	0	0	5	0	70	5	0
TE	0	0	5	0	0	70	0
TB	0	0	10	0	5	0	35
Prec	**0,909**	**1,000**	**0,750**	**0,778**	**0,824**	**0,824**	**1,000**
Rec	**0,833**	**1,000**	**0,750**	**0,875**	**0,875**	**0,933**	**0,700**

Πίνακας 5.3: *Συνολικά σφάλματα για τις δυαδικές ταξινομήσεις*

Μέθοδος	**Συνολική ταξινόμηση**	**Προτεινόμενο μοντέλο**
SN vs. SO	21.43%	3.57%
BA vs. TB	20.00%	8.57%

τα αποτελέσματα είναι πολύ καλύτερα. Για να αναδείξουμε αυτήν την παρατήρηση μας επιχειρήσαμε την δυαδική ταξινόμηση χρησιμοποιώντας μόνο ζευγάρια κατηγοριών. Ο ταξινομητής επιλέχτηκε και σε αυτήν την περίπτωση όπως πριν. Τα αποτελέσματα δείχνουν την ικανότητα διάκρισης μεταξύ των κλάσεων όταν χρησιμοποιηθεί το προτεινόμενο μοντέλο, καθώς το συνολικό σφάλμα είναι πολύ μικρότερο, όπως φαίνεται στον Πίνακα 5.3.

5.3.3 Ανίχνευση περιοχών ενδιαφέροντος

Η ιδιότητα των μοντέλων οπτικής προσοχής να εστιάζουν σε αντιληπτικά ενδιαφέρουσες περιοχές ταιριάζει με εφαρμογές επίβλεψης χώρων, στις οποίες το πρόβλημα είναι η γρήγορη ανίχνευση περιοχών που ξεχωρίζουν από την γειτονιά τους (π.χ. ένας άνθρωπος που ξαφνικά εισβάλει σε ένα δωμάτιο ή μία απότομη λάμψη σε έναν κλειστό χώρο). Για την εφαρμογή της προτεινόμενης μεθόδου επιλέξαμε ακολουθίες από την βάση CAVIAR, η οποία είναι δημόσια διαθέσιμη [17]. Σε αυτήν περιέχονται ακολουθίες που έχουν μαγνητοσκοπηθεί με ευρυγώνιους φακούς σε έναν διάδρομο εμπορικού κέντρου στην Λισσαβόνα (LISBON ακολουθίες) και στον χώρο υποδοχής των εργαστηρίων INRIA στην Grenoble (INRIA ακολουθίες). Οι μάσκες επαλήθευσης έχουν δημιουργηθεί χειρωνακτικά και έχουν τη μορφή ορθογωνίων που περιέχουν την επιθυμητή περιοχή (περαστικοί).

Τα Σχήματα 5.8,5.9,5.10 δείχνουν ενδεικτικά καρέ των δύο κατηγοριών, καθώς και τις αντίστοιχες μάσκες επαλήθευσης. Ο επιθυμητός στόχος είναι να υπολογίσουμε τον κύριο χάρτη με το μοντέλο χωροχρονικής οπτικής προσοχής, να τον κατατμήσουμε σε περιοχές ενδιάφεροντος και να παρατηρήσουμε το αν και πόσο γρήγορα μπορεί ο αλγόριθμος να εστιάσει στις πραγματικές περιοχές ενδιαφέροντος. Για την πληρότητα των αποτελεσμάτων επιχειρήσαμε και σύγκριση με την επεκτεταμένη με χάρτη κίνησης

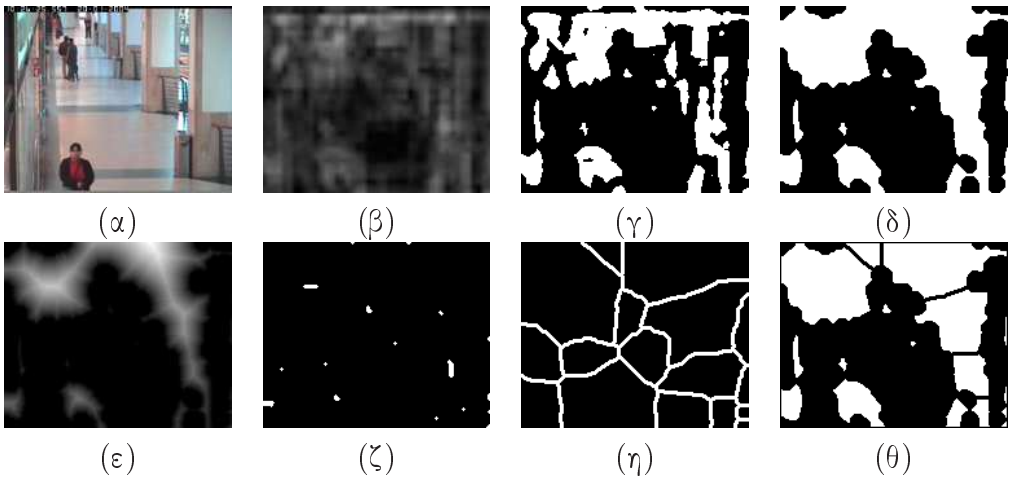
τεχνική των Itti *et al.* που παρουσιάσαμε στο Κεφάλαιο 2 [101].

5.3.3.1 Πειραματικά αποτελέσματα

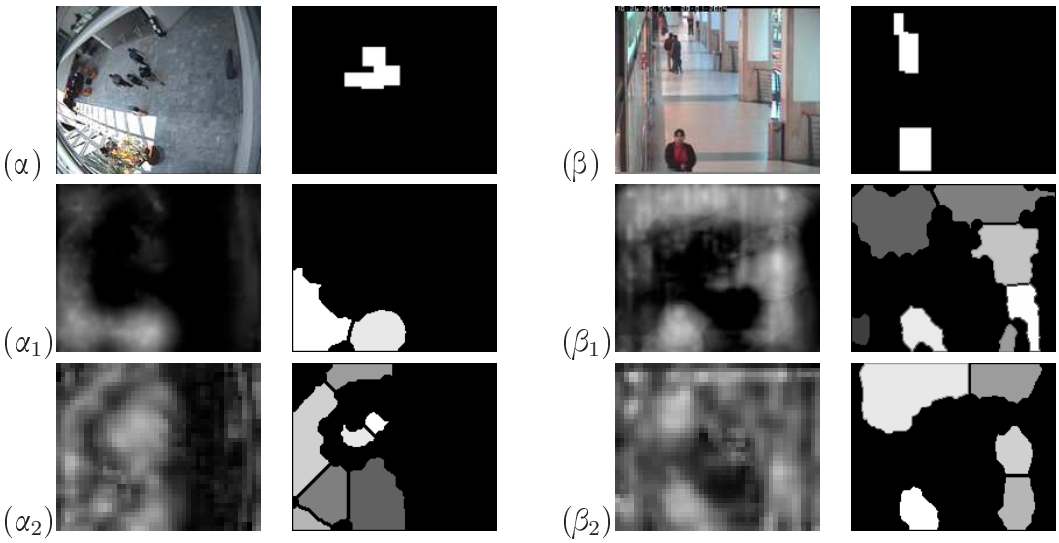
Για την ανίχνευση των περιοχών και την σύγκριση τους με τις μάσκες επαλήθευσης είναι απαραίτητη η κατάτμηση του κύριου όγκου σε περιοχές μειούμενου ενδιαφέροντος. Το στάδιο αυτό πρέπει να είναι γρήγορο, αλλά σχετικά αξιόπιστο. Δεν είναι απαραίτητη η ακριβής κατάτμηση σε περιοχές που αντιστοιχούν σε ένα και μοναδικό αντικείμενο, αλλά είναι σημαντικό οι περιοχές που θα προκύψουν να έχουν παρόμοια σημαντικότητα. Χρησιμοποιούμε μία τεχνική με χρήση μορφολογικών τελεστών, η οποία μοιάζει με την geodesic SKIZ (Skeleton Influence Zone) [117]. Αρχικά κάθε τομή του κύριου όγκου, κατωφλιοποιείται με την μέθοδο ελαχιστοποίησης διασποράς του Otsu [85] και επεξεργάζεται με απλούς μορφολογικούς τελεστές για να εξαλειφθούν προβληματικές περιοχές και να γεμίσουν τα κενά (Σχήμα 5.8γ). Ο κύριος χάρτης διαχωρίζει αρκετά καλά τις σημαντικές από τις λιγότερο σημαντικές περιοχές, οπότε η αυτόματη κατωφλιοποίηση δεν αλλοιώνει την αντιληπτική σημασία των δυαδικών περιοχών που προκύπτουν. Στην συνέχεια υπολογίζεται ένας marker για κάθε υποψήφια περιοχή με διαστολή των τοπικών μεγίστων του μετασχηματισμού απόστασης, όπως φαίνεται στα σχήματα 5.8εζ, και υπολογίζεται ο μετασχηματισμός watershed υπό τον περιορισμό των markers στην αρνητική έκδοση της εικόνας του μετασχηματισμού απόστασης. Η τελική μάσκα (Σχήμα 5.8θ) προκύπτει από την τομή της εικόνας στο Σχήμα 5.8γ και την αρνητική έκδοση της εικόνας στο Σχήμα 5.8η.

Ενδεικτικά αποτελέσματα για τις διαφορετικές ακολουθίες της βάσης CAVIAR παρουσιάζονται στις Εικόνες 5.9,5.10. Για κάθε καρέ διακρίνονται οι κύριοι χάρτες που προκύπτουν από το προτεινόμενο μοντέλο χωροχρονικής οπτικής προσοχής και το αντίστοιχο χωρικό των Itti *et al.*, το οποίο επεκτείναμε με χάρτη κίνησης (Κεφάλαιο 3, [101]), καθώς και οι αντίστοιχες μάσκες επαλήθευσης. Απεικονίζονται επίσης τα αποτελέσματα κατάτμησης του εκάστοτε κύριου χάρτη. Οι περιοχές που προκύπτουν μετά την κατάτμηση έχουν χαρακτηριστεί με διαφορετικά επίπεδα του γκρι ανάλογα με την μέση τιμή σημαντικότητας τους. Οι ανοιχτόχρωμες αντιστοιχούν σε περιοχές υψηλού ενδιαφέροντος. Για την ανάλυση των πειραμάτων προσδιορίζουμε δύο κριτήρια: τις τιμές ακρίβειας/επανάκλησης, όπως αυτές υπολογίζονται από την επικάλυψη των περιοχών ενδιαφέροντος με τις μάσκες επαλήθευσης, και τον αριθμό των εστιάσεων που απαιτείται για να επιτευχθούν οι προηγούμενες τιμές ή ισότιμα ο αριθμός των εστιάσεων που απαιτείται για να καλυφθεί ολόκληρος ή μέρος του επιθυμητού στόχου. Είναι προφανές, ότι οι υψηλές τιμές ακρίβειας είναι πιο δύσκολο να επιτευχθούν από τις τιμές επανάκλησης, καθώς οι μη συνειδητές τεχνικές οπτικής προσοχής, όπως αυτές που προτείνουμε, δεν είναι σχεδιασμένες να εντοπίζουν με ακρίβεια στόχους με συγκεκριμένα χαρακτηριστικά ή προδιαγεγραμμένη συμπεριφορά στον χρόνο.

Υπολογίζουμε επομένως το ποσοστό επικάλυψης των περιοχών ενδιαφέροντος και των μασκών επαλήθευσης για κάθε εστίαση του μοντέλου. Ως εστίαση ορίζεται το σχήμα της εκάστοτε περιοχής του κατατμημένου κύριου χάρτη, όπως φαίνεται στις εικόνες. Για παράδειγμα, αν παρατηρήσουμε την εικόνα κατάτμησης στο Σχήμα 5.9α_2, οι δύο πρώτες εστιάσεις θα είναι γύρω από το κέντρο (περιοχές ανοιχτού γκρι) και οι υπόλοιπες θα καλύψουν σειριακά το αριστερό μέρος του καρέ. Αυτή η διαδικασία μετρήσεων ακολουθείται για εκατοντάδες καρέ των ακολουθιών και υπολογίζονται οι μέσες τιμές ακρίβειας/επανάκλησης για τα δύο είδη ακολουθιών του CAVIAR: αυτών που δείχνουν ανθρώπους σε ένα εμπορικό κέντρο (π.χ. Σχήματα 5.9α, 5.10α)



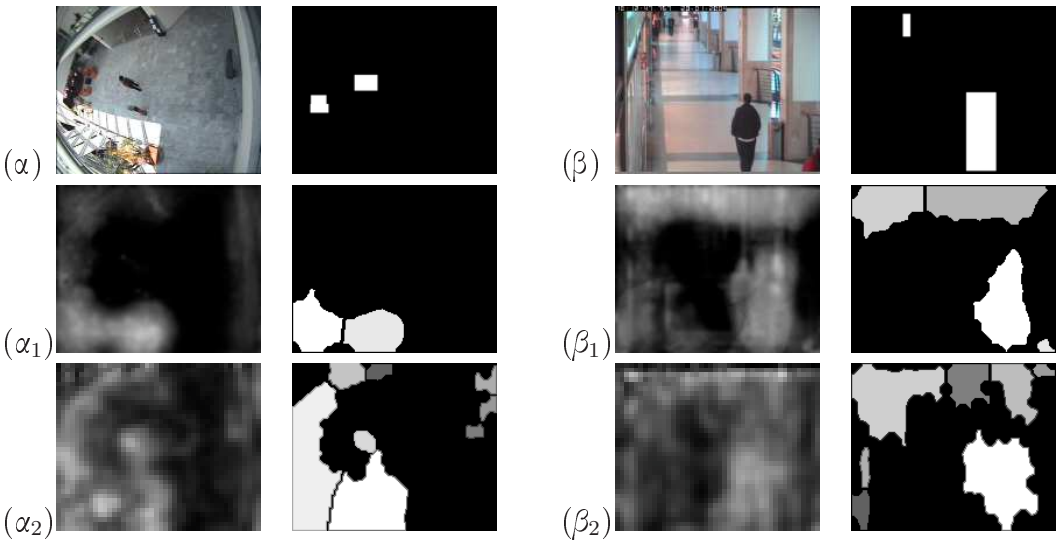
Σχήμα 5.8: *Τα βήματα της κατάτμησης του κύριου χάρτη: (α) αρχικό καρέ, (β) κύριος χάρτης, (γ) κατωφλίωση, (δ) μορφολογικό φιλτράρισμα, (ε) μετασχηματισμός απόστασης, (ζ) τοπικά μέγιστα, (η) μετασχηματισμός watershed με περιορισμό από την αρνητική εικόνα του μετασχηματισμού απόστασης και των τοπικών μεγίστων, (θ) τομή των (γ) και η αρνητική έκδοση της (η)*



Σχήμα 5.9: *(α)-(β) αρχικά καρέ και οι μάσκες επαλήθευσης για μία INRIA και μία LISBON ακολουθία αντίστοιχα, (α_1)-(β_1) οι κύριοι χάρτες της μεθόδου των Itti et al. με επέκταση κίνησης και η κατάτμηση τους, (α_2)-(β_2) οι κύριοι χάρτες της προτεινόμενης μεθόδου και η κατάτμηση τους*

και αυτών που δείχνουν επισκέπτες σε χώρο υποδοχής (π.χ. Σχήματα 5.9β, 5.10β). Συνοπτικά η διαδικασία που ακολουθείται είναι η εξής: 1) ένας χωροχρονικός κύριος όγκος υπολογίζεται για κάθε 64 καρέ της εισόδου, 2) κάθε τομή αυτού του όγκου κατατμείται με την παραπάνω διαδικασία, 3) οι κατατμημένες περιοχές ταξινομούνται με βάση την μέση τιμή ενδιαφέροντος, 4) υπολογίζονται οι τιμές ακρίβειας/επανάκλησης. Η ίδια διαδικασία ακολουθείται για την επεκτεταμένη τεχνική των Itti *et al.* για κάθε καρέ. Για να είναι πιο δίκαιη η σύγκριση φιλτράρουμε το αποτέλεσμα της δεύτερης τεχνικής με ένα 3Δ φίλτρο ενδιάμεσης τιμής για την εξάλειψη του θορύβου και την αύξηση της χωροχρονικής συνεκτικότητας.

Τα στατιστικά του Σχήματος 5.11 είναι οι μέσες τιμές ακρίβειας/επανάκλησης για 3154 καρέ των INRIA ακολουθιών, ενώ αυτά του Σχήματος 5.12 έχουν προκύψει



Σχήμα 5.10: *(α)-(β) αρχικά καρέ και οι μάσκες επαλήθευσης για μία INRIA και μία LISBON ακολουθία αντίστοιχα, (α1)-(β1) οι κύριοι χάρτες της μεθόδου των Itti et al. με επέκταση κίνησης και η κατάτμηση τους, (α2)-(β2) οι κύριοι χάρτες της προτεινόμενης μεθόδου και η κατάτμηση τους*

Πίνακας 5.4: *Στατιστικές μετρήσεις για τις INRIA ακολουθίες*

	Prec.	# εστιάσ.	1η εστίαση	Rec.	# εστιάσ.	1η εστίαση
Προτεινόμενη μέθοδος	14.20	8	11.75	71.57	15	31.75
Itti *et al.* με κίνηση	8.47	1	8.47	20.84	14	11.96

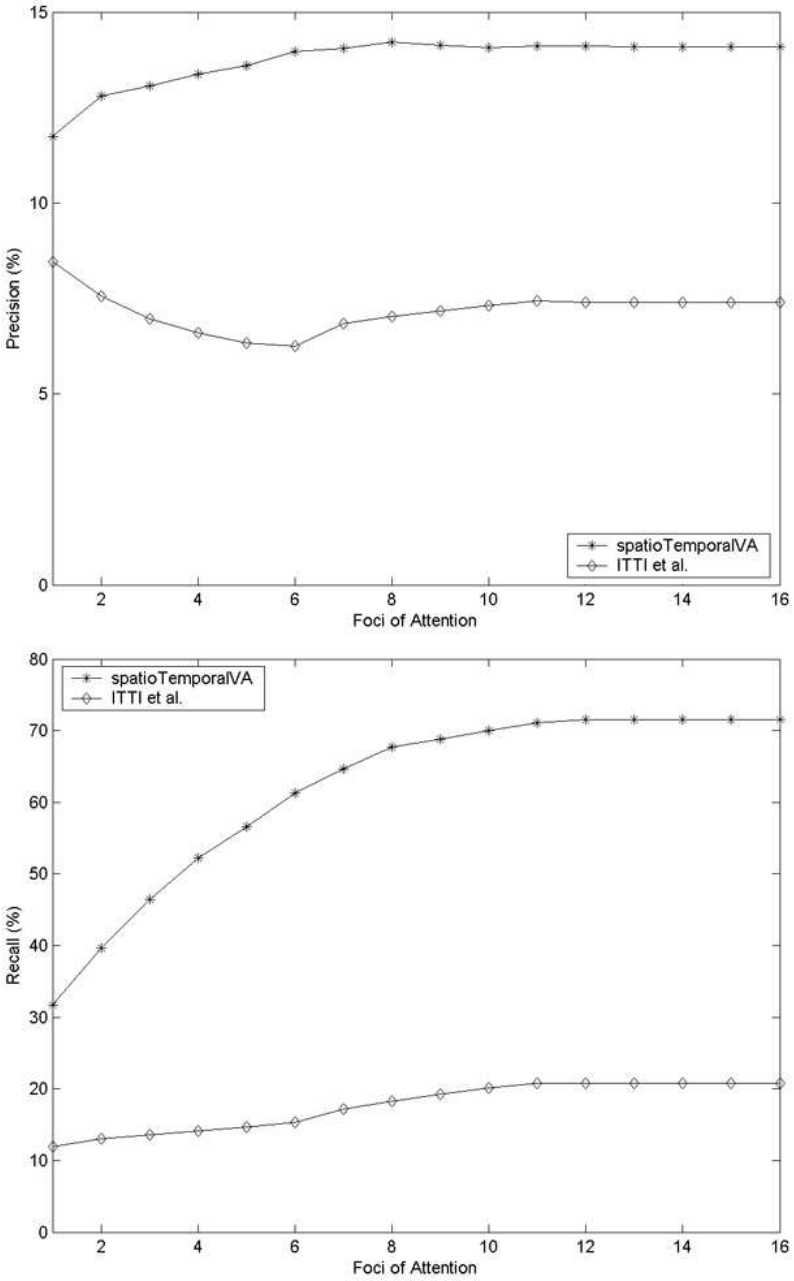
Πίνακας 5.5: *Στατιστικές μετρήσεις για τις LISBON ακολουθίες*

	Prec.	# εστιάσ.	1η εστίαση	Rec.	# εστιάσ.	1η εστίαση
Προτεινόμενη μέθοδος	11.94	2	11.55	61.52	15	35.90
Itti *et al.* με κίνηση	5.37	1	5.37	15.21	14	11.21

gg

από 3720 καρέ των LISBON ακολουθιών. Ο οριζόντιος άξονας αντιστοιχεί στον αριθμό των εστιάσεων και ο κάθετος στις τιμές των στατιστικών μετρήσεων. Οι Πίνακες 5.4 και 5.5 συνοψίζουν τα κύρια στατιστικά αποτελέσματα. Η προτεινόμενη τεχνική φτάνει στην μεγαλύτερη τιμή επανάκλησης (71.57%) στις INRIA ακολουθίες μετά από 15 εστιάσεις, ενώ στην πρώτη εστίαση η τιμή επανάκλησης είναι ήδη στο 31.75%. Παρόμοια η ακρίβεια φτάνει στο 14.20% μετά από 8 εστιάσεις.

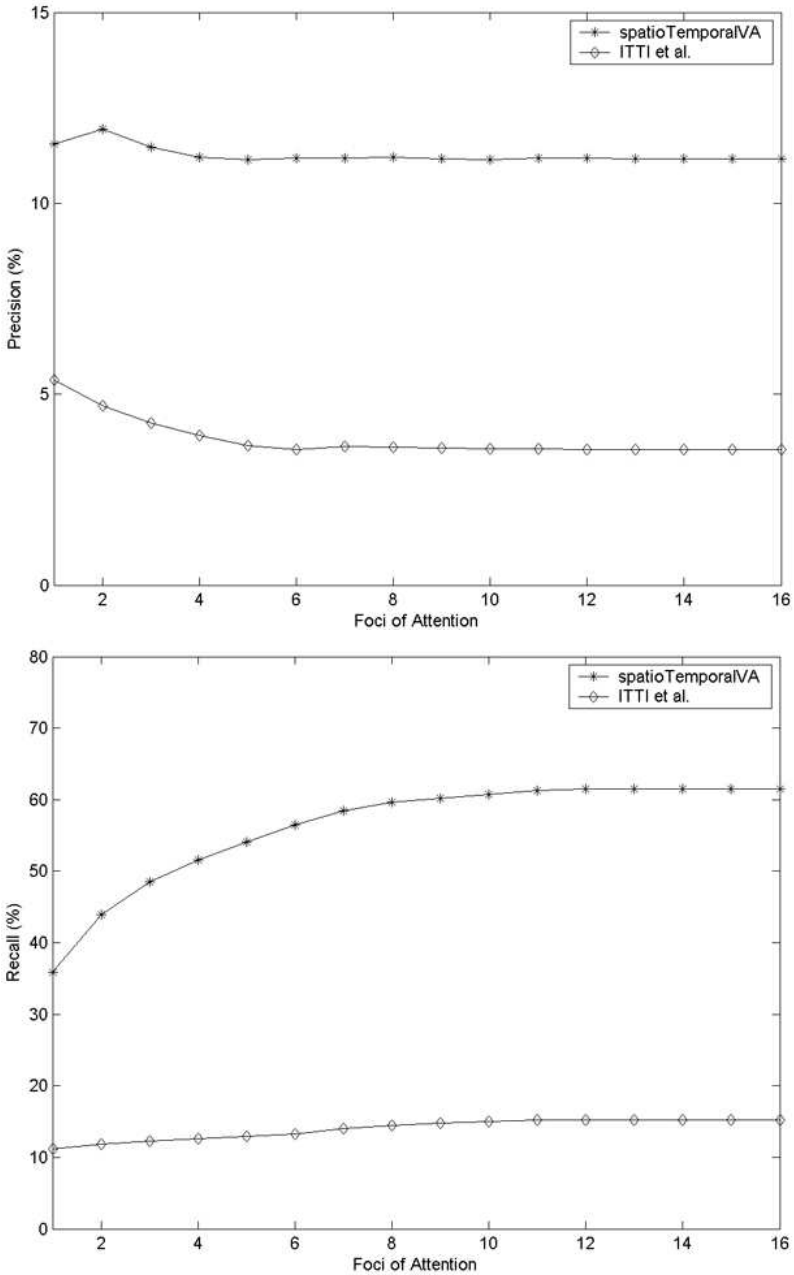
Οι μέσες τιμές ακρίβειας/επανάκλησης στους Πίνακες 5.4, 5.5 αναδεικνύουν το πλεονέκτημα του προτεινόμενου μοντέλου, καθώς γίνεται φανερό ότι από την πρώτη εστίαση έχει φτάσει σε τιμή μεγαλύτερη του μισού της μέγιστης. Πρέπει να τονιστεί ότι η διαφορά στην απόδοση των δύο τεχνικών οφείλεται καθαρά στην χωροχρονική αρχιτεκτονική του προτεινόμενου μοντέλου, καθώς τα αποτελέσματα της δεύτερης μεθόδου έχουν φιλτραριστεί χρονικά για την εξάλειψη προβλημάτων λόγω ασυνεχειών από καρέ σε καρέ. Η αρχιτεκτονική που αναπτύξαμε και οι προτεινόμενοι 3Δ τελεστές



(α) (β)

Σχήμα 5.11: *Αποτελέσματα των δύο μεθόδων για τις ακολουθίες INRIA. (α) ακρίβεια, (β) επανάκληση*

κανονικοποίησης και συνδυασμού εντοπίζουν χωροχρονικές περιοχές που διαφέρουν από την 3Δ γειτονιά τους και δεν είναι ευαίσθητοι σε ανεπαίσθητες αλλαγές στην σκηνή. Οι ακολουθίες LISBON περιέχουν μεγάλους στόχους σε σχετικά απλό περιβάλλον, ενώ οι ακολουθίες INRIA είναι πιο δύσκολες στην ανάλυση τους καθώς περιέχουν μικρούς στόχους σε σχετικά περίπλοκο περιβάλλον (πολύπλοκης υφής με πολλές αντανακλάσεις).



(α) (β)

Σχήμα 5.12: *Αποτελέσματα των δύο μεθόδων για τις ακολουθίες LISBON. (α) ακρίβεια, (β) επανάκληση*

5.4 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάσαμε ένα μοντέλο οπτικής προσοχής για ανάλυση ακολουθιών. Σκοπός ήταν η γρήγορη και αποδοτική ανίχνευση περιοχών ενδιαφέροντος, ώστε η επεξεργασία που ακολουθεί να εστιαστεί μόνο σε αυτές τις περιοχές. Η απόδοση της μεθόδου αξιολογείται ποιοτικά, αλλά και ποσοτικά με πειράματα αναγνώρισης σκηνής και ανίχνευση στόχου ενδιαφέροντος.

Το προτεινόμενο μοντέλο αποτελεί την πρώτη ολοκληρωμένη μας πρόταση για χωροχρονικό υπολογισμό τιμών ενδιαφέροντος, καθώς χρησιμοποιεί περισσότερα του ενός χαρακτηριστικά-σε αντίθεση με το μοντέλο στον χώρο των 3Δ κυματιδίων- και

βασίζεται στο υπάρχον μοντέλο χωρικής οπτικής προσοχής (Itti *et al.*, [48]), το οποίο έχει μελετηθεί και χρησιμοποιηθεί εκτενώς στην βιβλιογραφία.

□

Κεφάλαιο 6

Χωροχρονικό μοντέλο ενδιαφέροντος με ανταγωνισμό

6.1 Εισαγωγή

Στα μοντέλα που μελετήσαμε και προτείναμε μέχρι τώρα δεν λαμβάνεται υπόψη ο ανταγωνισμός μεταξύ οπτικών χαρακτηριστικών, όπως αυτός έχει υποστηριχτεί από πειράματα σχετικά με το ανθρώπινο οπτικό σύστημα [55]. Σύμφωνα με αυτά τα πειράματα, υπάρχει ένας εν μέρει άγνωστος -ακόμη- ανταγωνισμός μεταξύ των διαφορετικών οπτικών μονοπατιών, τα οποία σχετίζονται με κίνηση/βάθος (M pathway) και μορφή/βάθος/χρώμα (P pathway) αντίστοιχα (η Ενότητα 2.2.2 παρέχει περισσότερες πληροφορίες). Συνεχίζοντας την μελέτη και ανάπτυξη αλγορίθμων οπτικής προσοχής εισάγουμε την έννοια του ανταγωνισμού προτείνοντας ένα νέο πλαίσιο υπολογισμού τιμών ενδιαφέροντος.

Το προτεινόμενο χωροχρονικό μοντέλο επιτρέπει τον ανταγωνισμό διαφορετικών χαρακτηριστικών μέσω μιας διαδικασίας ελαχιστοποίησης ενός ενεργειακού με περιορισμούς που προέρχονται από προηγούμενα μοντέλα μας και από περιορισμούς που εμπνεύστηκαν από την θεωρία Μορφής (Gestalt). Ο μαθηματικός ορισμός του προβλήματος παρουσιάζει ομοιότητες με την δουλειά των Milanese et al., οι οποίοι κατέληξαν σε υπολογισμό τιμών ενδιαφέροντος σε εικόνες με την ελαχιστοποίηση ενός ενεργειακού [?]. Παρουσιάζουμε δύο μοντέλα οπτικής προσοχής, τα οποία μοιράζονται κοινές ιδέες, με το δεύτερο να αποτελεί εξέλιξη του πρώτου, και τα αξιολογούμε σε εφαρμογές αναγνώρισης σκηνής, αναγνώρισης οπτικών δραστηριοτήτων και δημιουργίας περιλήψεων ακολουθιών σύμφωνες προς την ανθρώπινη αντίληψη. Για την εφαρμογή αναγνώρισης σκηνής παρουσιάζουμε και σύγκριση με όλα τα μοντέλα χωρικής και χωροχρονικής ανάλυσης που έχουμε προτείνει ως τώρα.

Διατηρούμε την ογκομετρική αναπαράσταση της ακολουθίας και την αποσύνθεση της σε διάφορα χαρακτηριστικά. Η ενέργεια που προτείνουμε αποτελείται από έναν όρο παρατήρησης και έναν εξομάλυνσης. Αναλυτική επισκόπηση στον χώρο της χωροχρονικής ανάλυσης και της περιγραφής, ανίχνευσης και αναγνώρισης δυναμικών δραστηριοτήτων σε ακολουθίες παρουσιάζεται στην Ενότητα 2.4.4. Η πλειοψηφία των τεχνικών, όπως είδαμε και στο Κεφάλαιο 4, βασίζεται στην ανίχνευση σημείων ή περιοχών ενδιαφέροντος, τα οποία χρησιμοποιούνται για να αναπαραστήσουν την επιθυμητή δραστηριότητα.

6.2 Ορισμός του προβλήματος

Εμπνευσμένοι από τις θεωρίες Μορφής[27] και αντιληπτικής οργάνωσης προτείνουμε ένα νέο μοντέλο υπολογισμού κύριου χάρτη, το οποίο βασίζεται στην ογκομετρική αναπαράσταση της ακολουθίας και στον ανταγωνισμό διαφορετικών χαρακτηριστικών, ο οποίος μοντελοποιείται μέσω της ελαχιστοποίησης μίας ενέργειας με περιορισμούς. Οι περιορισμοί προκύπτουν από την εμπειρία των προηγούμενων μοντέλων που αναπτύξαμε (π.χ. φίλτρα κέντρου-περιφέρειας, πόλωση των χαρακτηριστικών με κίνηση) και από τους βασικούς νόμους της θεωρίας Μορφής που αναλύσαμε στην Ενότητα 2.2.1. Σε αυτό το κεφάλαιο παρουσιάζονται δύο μοντέλα που λειτουργούν κάτω από το ίδιο πλαίσιο, αλλά με διαφορετικούς περιορισμούς και μικρές διαφορές στον υπολογισμό του χωροχρονικού χάρτη κατευθυντικότητας. Ο διαχωρισμός υπάρχει λόγω της χρονικής εξέλιξης της ερευνάς. Στα Σχήματα 6.1 και 6.2 απεικονίζονται οι δύο προτεινόμενες αρχιτεκτονικές. Σε αυτήν την Ενότητα θα διατυπώσουμε το πρόβλημα, το οποίο είναι κοινό και για τις δύο.

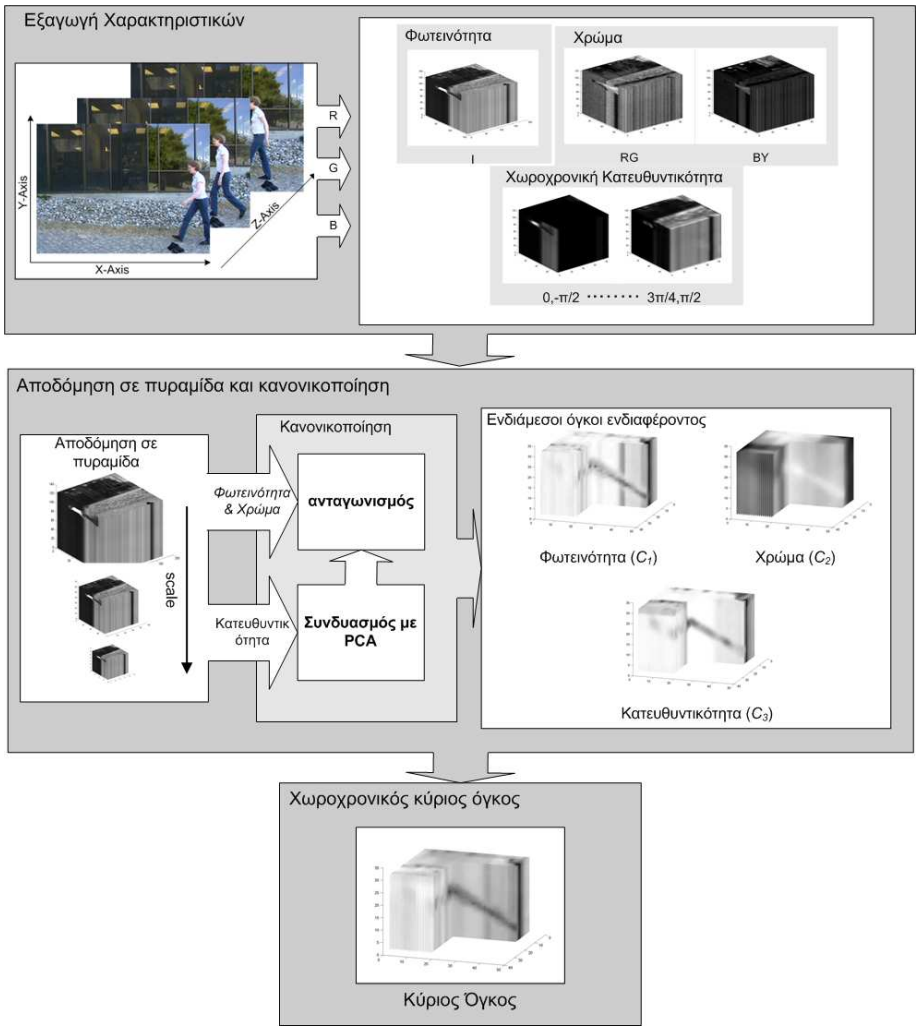
Η ακολουθία εισόδου μετασχηματίζεται σε έναν όγκο στον χώρο-χρόνο, όπως και στο Κεφάλαιο 5. Ο όγκος αυτός αποσυντίθεται σε ένα σύνολο χαρακτηριστικών, τα οποία σχηματίζουν χωροχρονικές πυραμίδες, οι οποίες κωδικοποιούν την χωροχρονική μεταβολή του εκάστοτε χαρακτηριστικού. Στην συνέχεια κανονικοποιείται κάθε πυραμίδα με έναν τελεστή αντίθεσης, ο οποίος βασίζεται σε γειτνίαση και ομοιότητα, έτσι ώστε να έχουμε κοινό σημείο αναφοράς πριν ξεκινήσει η διαδικασία βελτιστοποίησης.

Έστω V ο όγκος της ακολουθίας ορισμένος σε ένα σύνολο σημείων Q με το $q = (x, y, t)$ να είναι ένα σημείο στον χώρο-χρόνο. Τα σημεία $q \in Q$ δημιουργούν έναν 3Δ χώρο, ο οποίος ορίζεται από τις Καρτεσιανές τους συντεταγμένες. Σύμφωνα με αυτήν την αναπαράσταση το σημείο q είναι το αντίστοιχο ενός ογκοστοιχείου. Τα ογκοστοιχεία του V, τα οποία ανήκουν σε κάποιο κινούμενο αντικείμενο, καταλαμβάνουν μια χωροχρονική περιοχή του όγκου. Η τιμή του όγκου V στο σημείο q συμβολίζεται ως $V(q)$. Στην συνέχεια, ο όγκος V αποσυντίθεται σε ένα σύνολο χαρακτηριστικών F_i με $i = 1, 2, ..., M$. Αυτά τα χαρακτηριστικά είναι κοινά με τα προηγούμενα μοντέλα που αναπτύξαμε (φωτεινότητα, χρώμα, χωροχρονική κατευθυντικότητα).

Από κάθε όγκο χαρακτηριστικών σχηματίζουμε μία πυραμίδα και καταλήγουμε με ένα νέο σύνολο $\mathbf{F} = \{F_{i,\ell}\}$, το οποίο περιέχει τα χαρακτηριστικά σε όλες τις κλίμακες ℓ, όπου $\ell = 1, 2, ..., L$ και L είναι το μέγιστο βάθος της πυραμίδας. Η πυραμιδοειδής αναπαράσταση επιτυγχάνεται με επέκταση της κλασσικής Γκαουσσιανής πυραμίδας στον χώρο-χρόνο, όπως είδαμε και στην Ενότητα 6.2. Για να εξασφαλίσουμε κοινή κωδικοποίηση, το σύνολο $\mathbf{F}$ φιλτράρεται με έναν τελεστή αντίθεσης και δημιουργείται ένα νέο σύνολο ενδιάμεσων όγκων ενδιαφέροντος $\mathbf{C} = \{C_{i,\ell}\}$. Αυτός ο τελεστής αυξάνει την τιμή ενός ογκοστοιχείου όταν η τιμή του διαφέρει από την μέση τιμή της γειτονιάς του. Ορίζουμε τον υπολογισμό ενδιαφέροντος ως ένα πρόβλημα συνολικής βελτιστοποίησης και υπολογίζουμε μία τιμή ενδιαφέροντος για κάθε q μέσω του ανταγωνισμού των ενδιάμεσων όγκων ενδιαφέροντος $\mathbf{C}$. Διατυπώνουμε αυτό το πρόβλημα ως την βελτιστοποίηση μίας ενέργειας E, η οποία αποτελείται από έναν όρο παρατήρησης E_d και έναν όρο εξομάλυνσης E_s

$$E(\mathbf{C}) = \lambda_d \cdot E_d(\mathbf{C}) + \lambda_s \cdot E_s(\mathbf{C}) \tag{6.1}$$

Ο δεύτερος όρος της (6.1) σχετίζεται με εξομάλυνση, καθώς κανονικοποιεί (regularize) την τρέχουσα παρατήρηση σύμφωνα με τους επιλεγμένους περιορισμούς [126] [106].



Σχήμα 6.1: *Προτεινόμενη αρχιτεκτονική για το μοντέλο με περιορισμούς κίνησης*

Ο τελικός κύριος χάρτης S υπολογίζεται ως το άθροισμα των ενδιάμεσων όγκων ενδιαφέροντος που προκύπτουν από την ελαχιστοποίηση της (6.1).

6.3 Εξαγωγή Χαρακτηριστικών

Τα χαρακτηριστικά που χρησιμοποιήσαμε είναι ίδια με αυτά της Ενότητας 5.2.1. Η μόνη διαφοροποίηση εστιάζεται στον χάρτη κατευθυντικότητας. Για λόγους απλοποίησης περιορίσαμε τον αριθμό των κατευθύνσεων των φίλτρων καθοδήγησης και επιλέξαμε διαφορετικό τρόπο συνδυασμού τους.

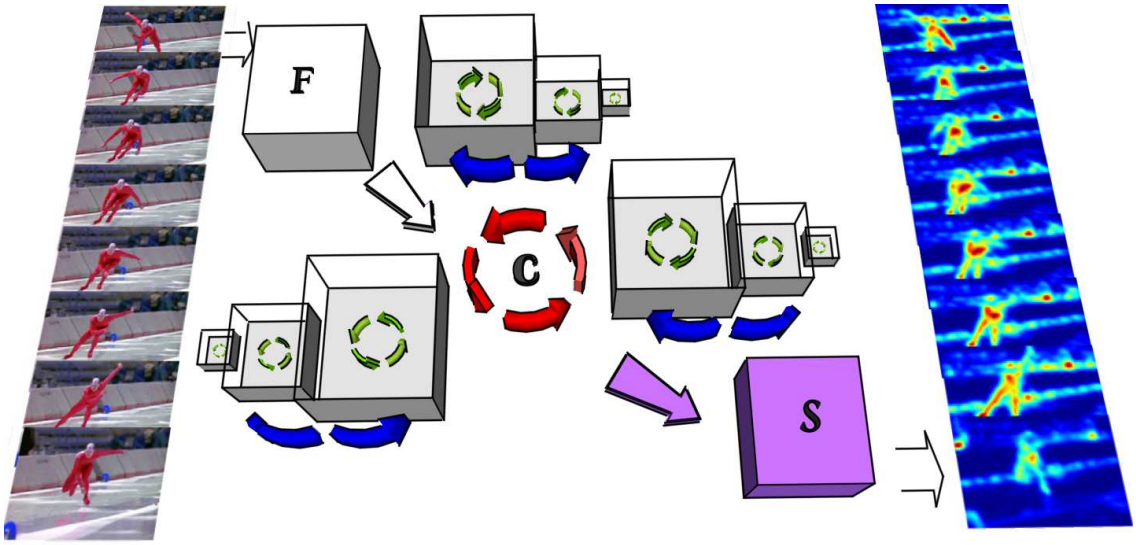
Τα χαρακτηριστικά φωτεινότητας και χρώματος προκύπτουν ως

$$F_1 = \frac{1}{3} \cdot (r + g + b) \tag{6.2}$$

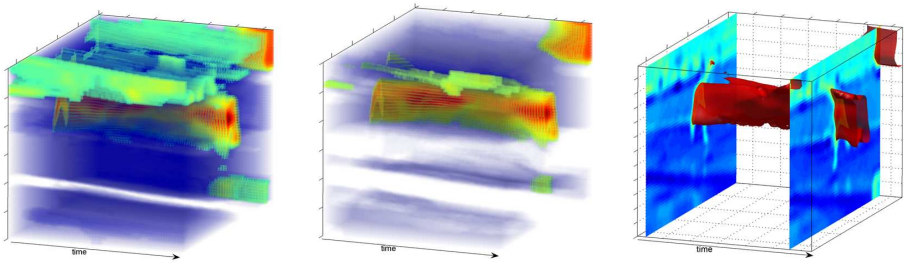
$$F_2 = RG + BY \tag{6.3}$$

όπου

$$RG = R - G \tag{6.4}$$
$$BY = B - Y \tag{6.5}$$



Σχήμα 6.2: *Προτεινόμενη αρχιτεκτονική για το μοντέλο με περιορισμούς Μορφής*



(a) (b) (b)

Σχήμα 6.3: *(α),(β) διαφανείς εκδόσεις του κύριου όγκου, (γ) Ισομετρική επιφάνεια που περιλαμβάνει τις πιο σημαντικές περιοχές του κύριου όγκου*

και

$$\begin{aligned} R &= r - \frac{(g+b)}{2} \\ G &= g - \frac{(r+b)}{2} \\ B &= b - \frac{(r+g)}{2} \\ Y &= \frac{(r+g)}{2} - \frac{|r-g|}{2} - b \end{aligned} \tag{6.6}$$

Για τον υπολογισμό της χωροχρονικής κατευθυντικότητας χρησιμοποιήσαμε φίλτρα καθοδήγησης, όπως και στα προηγούμενα κεφάλαια, αλλά ακολουθήσαμε διαφορετική λογική για τον συνδυασμό τους. Συγκεκριμένα, χρησιμοποιήσαμε φίλτρα σχετικά με κινήσεις προς τα αριστερά $(\frac{-1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, δεξιά $(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, πάνω $(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ και κάτω $(0, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, όπως προτείνουν και οι Wildes *et al.* [147]. Οι αριθμοί στις παρενθέσεις αντιστοιχούν στις τιμές των α, β, γ αντίστοιχα, οι οποίες αντιστοιχούν στα συνημίτονα κατεύθυνσης στα οποία προσανατολίζονται τα φίλτρα. Περισσότερες λεπτομέρειες δίνονται στην Ενότητα 5.2.1.1. Για να καταλήξουμε σε μία πιο ξεκάθαρη μέτρηση κατευθυντικότητας το αποτέλεσμα κάθε φίλτρου κανονικοποιείται με το

συνολικό άθροισμα των αποκρίσεων και τελικά ο ενδιάμεσος όγκος ενδιαφέροντος υπολογίζεται ως

$$F_3(q) = \frac{\sum_\theta E_v^\theta(q)}{\sum_r \sum_\theta E_v^\theta(r)} \tag{6.7}$$

όπου η E_v^θ ορίζεται στην (5.6).

Για να ενισχύσουμε την τιμή των ογκοστοιχείων που ξεχωρίζουν από την γειτονιά τους (με υψηλή τιμή ανάμεσα σε χαμηλές και το αντίστροφο) εφαρμόζουμε έναν τελεστή αντίθεσης, ο οποίος εξασφαλίζει κοινό εύρος τιμών ανάμεσα σε όλα τα εμπλεκόμενα χαρακτηριστικά έτσι ώστε να είναι συγκρίσιμα και να μπορούν να χρησιμοποιηθούν σαν αρχικές λύσεις στην διαδικασία ελαχιστοποίησης. Το νέο σύνολο ως

$$C_{i,\ell}(q)\} = \left| F_{i,\ell}(q) - \frac{1}{|N_q|} \sum_{r \in N_q} F_{i,\ell}(r) \right| \tag{6.8}$$

όπου N_q είναι το σύνολο των 26 γειτόνων του ογκοστοιχείου q στον 3Δ χώρο. Η 26-γειτονιά είναι η επέκταση σε 3Δ της τυπικής 8-γειτονιάς στον 2Δ χώρο της εικόνας.

6.4 Ανταγωνισμός χαρακτηριστικών με περιορισμό κίνησης

Στην πρώτη μας απόπειρα να υλοποιήσουμε την ιδέα του ανταγωνισμού εκμεταλλευτήκαμε την πρότερη γνώση μας από τα μοντέλα κύριου χάρτη που περιγράψαμε και προτείναμε στα Κεφάλαια 3 και **??**. Το Σχήμα 6.1 απεικονίζει το προτεινόμενο μοντέλο με όλα τα δομοστοιχεία του: εξαγωγή χαρακτηριστικών, αποσύνθεση με Γκαουσσιανές πυραμίδες, κανονικοποίηση, δημιουργία ενδιάμεσων όγκων ενδιαφέροντος και συνδυασμός για τον υπολογισμό του κύριου όγκου.

6.4.1 Ορισμός ενέργειας

Η ενέργεια εμπεριέχει πράξεις μεταξύ ογκοστοιχείων σε διαφορετικές κλίμακες της πυραμίδας. Συγκεκριμένα οι πράξεις γίνονται μεταξύ στοιχείων της κλίμακας κέντρου c και της κλίμακας περιφέρειας h. Προτείνουμε δύο όρους σχετικούς (α) με τον χωροχρονικό τελεστή κέντρου-περιφέρειας και (β) την επίδραση που ασκεί η κίνηση (χωροχρονική κατευθυντικότητα) στον υπολογισμό τιμών ενδιαφέροντος.

Ακολουθώντας την διατύπωση της (6.1), ο πρώτος όρος της ορίζεται ως

$$E_d(\mathbf{C}) = \sum_i \sum_\ell \sum_q \left(C_{i,c}(q) \cdot |C_{i,c}(q) - C_{i,h}(q)| \right) \tag{6.9}$$

και αντιστοιχεί στην ενέργεια κέντρου-περιφέρειας. Η διαφορά σε κάθε ογκοστοιχείο υπολογίζεται μετά από παρεμβολή του $F_{i,h}$ στο μέγεθος του $F_{i,c}$. Ο όρος αυτός καλείται να ενισχύσει περιοχές που ξεχωρίζουν από την χωροχρονική γειτονιά τους και επομένως ελκύουν την προσοχή μας. Η ενέργεια εξομάλυνσης E_s που προτείνουμε σχετίζεται με ανταγωνισμό ενός ογκοστοιχείου με τη γειτονιά του και την επίδραση του καναλιού κίνησης:

$$E_s(\mathbf{C}) = \sum_i \sum_\ell \sum_q \left(C_{i,c}(q) \cdot \frac{1}{|N_q|} \cdot \sum_r (C_{i,c}(r) + C_3(r)) \right) \tag{6.10}$$

Σύμφωνα με αυτόν τον περιορισμό, ένα ογκοστοιχείο αυξάνει την σημαντικότητα του μόνο όταν η ενεργοποίηση της γειτονιάς του είναι ψηλή και παρουσιάζει έντονη κίνηση.

6.4.2 Ελαχιστοποίηση ενέργειας

Η ελαχιστοποίηση της ενέργειας (6.1) μπορεί να γίνει με οποιαδήποτε μέθοδο κατάβασης κλίσης[35]. Υιοθετούμε έναν αλγόριθμο μέγιστης κατάβασης κλίσης[36] σύμφωνα με τον οποίο η τιμή κάθε ογκοστοιχείου αλλάζει προς την κατεύθυνση του εκτιμώμενου ελαχίστου.

$$C^{\tau}_{i,\ell}(q) = C^{\tau-1}_{i,\ell}(q) + \triangle C^{\tau-1}_{i,\ell}(q) \tag{6.11}$$

όπου

$$\triangle C^{\tau-1}_{i,\ell}(q) = -\gamma \cdot \frac{\partial E(\mathbf{C}^{\tau}))}{\partial C^{\tau}_{i,\ell}(q)} + \mu \cdot \triangle C^{\tau-1}_{i,\ell}(q) \tag{6.12}$$

όπου τ είναι ο αριθμός της επανάληψης, γ είναι ο ρυθμός εκμάθησης και μ είναι μία σταθερά ορμής (momentum) που βοηθά στην σταθεροποίηση του αλγόριθμου [107]. Οι δύο παράμετροι είναι σημαντικές τόσο για την ευστάθεια όσο και για την ταχύτητα σύγκλισης. Πρακτικά λίγες επαναλήψεις είναι αρκετές για να προσεγγίσουμε το επιθυμητό ελάχιστο. Για να κρατήσουμε τους συμβολισμούς απλούς παραλείπουμε το σύμβολο τ από τις εξισώσεις που ακολουθούν.

Για την ελαχιστοποίηση της (6.1) απαιτείται ο υπολογισμός των μερικών παραγώγων που εμφανίζονται στην (6.12):

$$\begin{aligned} \frac{\partial E(\mathbf{C})}{\partial C_{k,m}(s)} &= \lambda_d \cdot \frac{\partial E_d(\mathbf{C})}{\partial C_{k,m}(s)} + \lambda_d \cdot \frac{\partial E_s(\mathbf{C})}{\partial C_{k,m}(s)} \\ &= \lambda_d \cdot \frac{\partial E_d(\mathbf{C})}{\partial C_{k,m}(s)} + \lambda_s \cdot \sum_{c=1}^{3} \frac{\partial E_n(\mathbf{C})}{\partial C_{k,m}(s)} \end{aligned} \tag{6.13}$$

όπου E_n με $n = 1, 2, M$ είναι οι περιορισμοί του όρου εξομάλυνσης. Η μερική παράγωγος της E_d υπολογίζεται ως

$$\frac{\partial E_d}{\partial C_{k,m}(s)} = |C_{k,c}(s) - C_{k,h}(s)| + sign(C_{k,c}(s)) \cdot C_{k,h}(s) \tag{6.14}$$

και η μερική παράγωγος της E_s ως

$$\frac{\partial E_s}{\partial C_{k,m}(s)} = \frac{1}{|N_s|} \cdot \sum_{r \in N_s} (C_{k,c}(r) + C_{3,c}(r)) \tag{6.15}$$

όπου $C_{3,c}$ είναι ο ενδιάμεσος όγκος χωροχρονικής κατευθυντικότητας, ο οποίος κωδικοποιεί την κινητική δραστηριότητα της ακολουθίας και υπολογίζεται ως

$$C_{3,c} = T[\oplus(F_{3,c}, F_{3,h})] \tag{6.16}$$

όπου $\oplus$ είναι ο τελεστής σημαντικότητας, ο οποίος αποτελείται από μείωση των όγκων κάθε επιπέδου σε μία συγκεκριμένη κλίμακα της πυραμίδας και σημείο-προς-σημείο πρόσθεση τους, και T είναι ένας τελεστής αντίθεσης που ενισχύει ακόμη περισσότερο τις τοπικά ενεργές περιοχές. Στην υλοποίηση μας χρησιμοποιήσαμε τον μορφολογικό τελεστή *top-hat* με έναν 3Δ δομικό στοιχείο.

6.4.3 Υπολογισμός κύριου όγκου

Για τον υπολογισμό του κύριου όγκου απαιτείται πρώτα η δημιουργία των ενδιάμεσων όγκων ενδιαφέροντος, οι οποίοι κωδικοποιούν την σημασία κάθε ογκοστοιχείου ως προς το αντίστοιχο χαρακτηριστικό. Οι ενδιάμεσοι όγκοι παράγονται με τον τελεστή σημαντικότητας που είδαμε στην προηγούμενη ενότητα (6.4.2) εφαρμόζοντας τον στα αποτελέσματα της ελαχιστοποίησης $\hat{C}_{i,\ell}$. Ο ενδιάμεσος όγκος φωτεινότητας υπολογίζεται ως

$$I = \oplus(\hat{C}_{1,c}, \hat{C}_{1,h}) \tag{6.17}$$

και ο αντίστοιχος του χρώματος ως

$$C = \oplus(\hat{C}_{2,c}, \hat{C}_{2,h}) \tag{6.18}$$

Τελικά ο κύριος όγκος προκύπτει ως

$$S = \frac{1}{2} \cdot (I + C) \tag{6.19}$$

6.5 Ανταγωνισμός χαρακτηριστικών με περιορισμούς Μορφής (Gestalt)

Το δεύτερο μοντέλο που προτείνουμε βασίζεται στην ομαδοποίηση "όμοιων" περιοχών και χρησιμοποιεί ενέργειες εμπνευσμένες από τους νόμους της θεωρίας Μορφής. Αρκετά νωρίς οι ψυχολόγοι αυτής της σχολής περιέγραψαν τρόπους σύμφωνα με τους οποίους οργανώνονται αντικείμενα και δυναμικά γεγονότα, έτσι ώστε να γίνουν εύκολα και γρήγορα κατανοητά από τον ανθρώπινο εγκέφαλο. Στην Ενότητα 2.2.1 αναφερθήκαμε σε εκπροσώπους αυτής της σχολής. Μια απλή εξήγηση είναι ότι η θεωρία Μορφής αναφέρεται σε μια ενοποιημένη αναπαράσταση αντικειμένων/γεγονότων, η οποία έχει συγκεκριμένα χαρακτηριστικά που είναι σημαντικότερα από το άθροισμα των επιμέρους χαρακτηριστικών, τα οποία έχουν κοινές ιδιότητες (π.χ. γειτνίαση, ομοιότητα, απλότητα, κοινό προορισμό). Σε αντιστοιχία με το παράδειγμα της μελωδίας στην Ενότητα 2.2.1 παραθέτουμε το εξής: Όταν παρατηρούμε έναν άνθρωπο να περπατάει, αυτόματα αντιλαμβανόμαστε το γεγονός στο σύνολο του και όχι σαν ένα σύνολο από κινούμενα χέρια, πόδια κτλ. Κάθε επιμέρους κίνηση είναι ξεκάθαρα αυτόνομη, αλλά η συνολική ερμηνεία του γεγονότος εξαρτάται από την ταξινόμηση των υπο-γεγονότων σε μία συγκεκριμένη διάταξη (περπάτημα). Με αυτήν τη λογική, μία μεθοδολογία μηχανικής όρασης για ανίχνευση και αναγνώριση οπτικών γεγονότων πρέπει να παράγει συνεκτικές οπτικές μορφές, οι οποίες ανήκουν κατά το δυνατόν στο ίδιο γεγονός.

Όπως αναφέρθηκε, οι ενέργειες που προτείνουμε εμπνεύστηκαν από τους νόμους της Μορφής. Συγκεκριμένα οι ενέργειες σχετίζονται με τους εξής νόμους: (α) φόντου/υπόβαθρου (τα αντικείμενα ξεχωρίζουν από το υπόβαθρο τους), (β) γειτνίασης (γειτονικά στοιχεία στον χώρο ή στον χρόνο ομαδοποιούνται), (γ) κλεισίματος (μία μορφή είναι συμπαγής και επομένως οπτικά κενά καλύπτονται), (δ) ομοιότητας (στοιχεία όμοια ως προς σχήμα, χρώμα, μέγεθος ή φωτεινότητα ομαδοποιούνται) και (3) κοινού προορισμού (στοιχεία με παρόμοια κίνηση αντιλαμβάνονται ως σύνολο).

Στο Σχήμα 6.2 απεικονίζεται με πιο παραστατικό τρόπο το δεύτερο μοντέλο που περιγράφουμε σε αυτό το κεφάλαιο και διαφέρει από το προηγούμενο κυρίως στο είδος των ενεργειών που επιλέξαμε για να μοντελοποιήσουμε το πρόβλημα.

Για να κρατηθεί το σχήμα απλό περιλαμβάνονται σε αυτό μόνο οι πυραμίδες των όγκων ενδιαφέροντος. Τα βέλη στο σχήμα είναι ενδεικτικά των τριών τρόπων που αλληλεπιδρούν τα ογκοστοιχεία κάθε όγκου: (α) ενδο-χαρακτηριστικός[31] ανταγωνισμός μεταξύ ογκοστοιχείων του ιδίου χαρακτηριστικού, (β) δια-χαρακτηριστικός[32] ανταγωνισμός μεταξύ ογκοστοιχείων διαφορετικών χαρακτηριστικών και (γ) δια-κλιμακωτός[33] ανταγωνισμός μεταξύ ογκοστοιχείων του ίδιου χαρακτηριστικού αλλά διαφορετικής κλίμακας. Τα πράσινα βέλη αντιστοιχούν στον ενδο-χαρακτηριστικό, τα μπλε στον δια-κλιμακωτό και τα κόκκινα στον δια-χαρακτηριστικό ανταγωνισμό αντίστοιχα. Στο Σχήμα 6.2 φαίνονται επίσης ενδεικτικά καρέ εισόδου και εξόδου. Τα καρέ εξόδου είναι τομές του κύριου όγκου που προκύπτει με τις σημαντικές περιοχές να είναι κόκκινες και τις λιγότερο σημαντικές μπλε.

Ο πρώτος μας στόχος είναι η ανίχνευση σημαντικών διατάξεων ογκοστοιχείων που σχηματίζουν μία συνεκτική επιφάνεια στον χώρο-χρόνο και που ανήκει σε ένα ενιαίο γεγονός και ο δεύτερος είναι η αναπαράσταση και αναγνώριση του γεγονότος ή ο υπολογισμός της σημαντικότητας του και η σύγκριση με το υπόλοιπο της ακολουθίας. Αξιολογούμε το προτεινόμενο μοντέλο στην αναγνώριση δραστηριοτήτων και στην δημιουργία περιλήψεων ακολουθιών. Παρουσιάζουμε πειράματα με ακολουθίες δραστηριοτήτων από βάσεις δημόσιας χρήσης και συγκρίνουμε με καθιερωμένες τεχνικές του χώρου. Επιπροσθέτως, αξιολογούμε την δυνατότητα του μοντέλου να ανιχνεύει σημαντικά γεγονότα σε μία ακολουθία παράγοντας περιλήψεις ακολουθιών από μία πρόσφατη βάση [78], η οποία αποτελείται από ακολουθίες γνωστών ταινιών με σχολιασμό επί της σημασίας των μερών τους. Αξιολογούμε την ικανότητα του μοντέλου να δημιουργεί περιλήψεις που συμφωνούν με την ανθρώπινη αντίληψη συγκρίνοντας το αποτέλεσμα με τον σχολιασμό. Τα αποτελέσματα αποδεικνύουν την δυναμική του μοντέλου σε αυτές τις εφαρμογές.

6.5.1 Ορισμός ενέργειας

Οι ενέργειες που περιγράφουμε σε αυτήν την ενότητα προέκυψαν από τους νόμους της θεωρίας Μορφής, όπως είδαμε και στην Ενότητα 6.5. Ο στόχος είναι η αντιληπτική ομαδοποίηση περιοχών με τέτοιο τρόπο ώστε να αυξηθεί η σημασία τους. Οι ενέργειες που περιγράφονται παρακάτω σχετίζονται με τους νόμους Μορφής ως εξής: (α) Η γειτνίαση και η κλειστότητα είναι η αφορμή για το ενδο-χαρακτηριστικό περιορισμό ενέργειας, σύμφωνα με τον οποίο γειτονικά ογκοστοιχεία τείνουν να έχουν παρόμοιες τιμές και επομένως μικρά κενά μεταξύ τους καλύπτονται εξαιτίας της δυναμικής του συστήματος, (β) Η ομοιότητα σχετίζεται με όλους τους περιορισμούς, καθώς ογκοστοιχεία παρόμοια σε ενδο-,δια- χαρακτηριστική και δια-κλιμακωτή τιμή τείνουν να ομαδοποιούνται και (γ) ο νόμος του κοινού προορισμού είναι η αφορμή για την συνολική ελαχιστοποίηση του ενεργειακού με στόχο την δημιουργία συνεκτικών και ομογενών περιοχών.

Ο περιορισμός παρατήρησης E_d διατηρεί μία σχέση μεταξύ της τρέχουσας παρατήρησης και της αρχικής λύσης με σκοπό να αποφευχθεί η υπέρ-εξομάλυνση του αποτελέσματος. Για λόγους πληρότητας υπενθυμίζουμε ότι η αρχική μας λύση είναι το αποτέλεσμα του τελεστή αντίθεσης (Ενότητα 6.3). Ο περιορισμός υλοποιείται στο μοντέλο ως εξής:

$$E_d(C) = \sum_i \sum_l \sum_q (C_{i,\ell}(q) - C^0_{i,\ell}(q))^2 \tag{6.20}$$

Ο ενδο-χαρακτηριστικός περιορισμός σχετίζεται με τον ανταγωνισμό μεταξύ ογκο-

στοιχείων του ίδιου χαρακτηριστικού. Αυξάνει την τιμή των στοιχείων όταν αυτά συνορεύουν με άλλα στοιχεία που ξεχωρίζουν από την γειτονιά τους:

$$E_1(\mathbf{C}) = \sum_i \sum_\ell \sum_q \left(C_{i,\ell}(q) - \frac{1}{|N_q|} \sum_{r \in N_q} C_{i,\ell}(r) \right)^2 \tag{6.21}$$

Όπως προκύπτει από την προηγούμενη εξίσωση, η ενέργεια ελαχιστοποιείται όταν η τιμή του ογκοστοιχείου q γίνεται ίση με την μέση τιμή των γειτόνων του. Ιδανικά με αυτόν τον τρόπο θα κλείσουν μικρά "κενά" στους ενδιάμεσους όγκους ενδιαφέροντος και θα επιτραπεί η ομαδοποίηση όμοιων σε τιμή στοιχείων σε μεγαλύτερες περιοχές.

Ο δια-χαρακτηριστικός περιορισμός επιτρέπει τον ανταγωνισμό μεταξύ των διαφορετικών χαρακτηριστικών. Ογκοστοιχεία που ξεχωρίζουν από την γειτονιά τους σε όλους τους όγκους χαρακτηριστικών αυξάνουν την τιμή ενδιαφέροντος τους. Στο μοντέλο μας υλοποιείται ως μία ενέργεια που επηρεάζει την τιμή του q προς την μέση τιμή των ανταγωνιστών του στους υπόλοιπους όγκους:

$$E_2(\mathbf{C}) = \sum_i \sum_\ell \sum_q \left(C_{i,\ell}(q) - \frac{1}{M-1} \sum_{j \neq i} F_{j,\ell}(q) \right)^2 \tag{6.22}$$

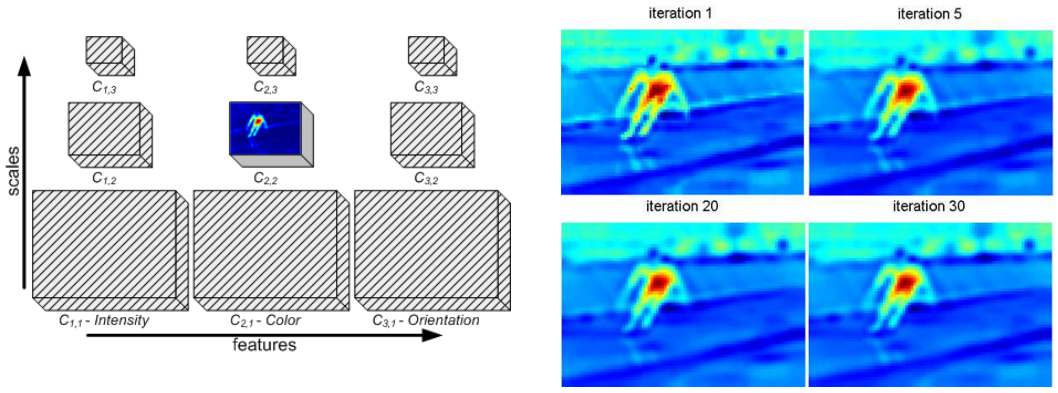
Ο δια-κλιμακωτός περιορισμός επενεργεί στις κλίμακες της πυραμίδας κάθε χαρακτηριστικού και σχετίζεται με ομοιογένεια σε όλες τις κλίμακες της αναπαράστασης. Αυτό σημαίνει ότι αν ένα ογκοστοιχείο έχει υψηλή τιμή σε όλες τις κλίμακες, τότε πρέπει να γίνει πιο σημαντικό:

$$E_3(\mathbf{C}) = \sum_i \sum_\ell \sum_q \left(C_{i,\ell}(q) - \frac{1}{L-1} \sum_{n \neq l} C_{i,n}(q) \right)^2 \tag{6.23}$$

Τα Σχήματα 6.4, 6.6, 6.5 και 6.7 απεικονίζουν την λειτουργία του προτεινόμενου μοντέλου όταν εμπλέκεται ο ενδο-χαρακτηριστικός, ο δια-χαρακτηριστικός, ο δια-κλιμακωτός περιορισμός και ο συνδυασμός τους αντίστοιχα. Η πρώτη στήλη των Σχημάτων δείχνει τις αλληλεπιδράσεις μεταξύ ογκοστοιχείων που ενεργοποιούνται σε κάθε περίπτωση. Ο οριζόντιος άξονας αντιστοιχεί στα χαρακτηριστικά, ενώ ο κατακόρυφος στις κλίμακες. Οι όγκοι που φαίνονται απενεργοποιημένοι είναι αυτοί που δεν συμβάλουν στην ελαχιστοποίηση σύμφωνα με τα επιλεγμένα κριτήρια. Η δεύτερη στήλη δείχνει ενδεικτικά αποτελέσματα για τον ενδιάμεσο όγκο ενδιαφέροντος $C_{2,2}$ μετά από 1, 5, 20 και 30 επαναλήψεις. Οι κόκκινες τιμές αντιστοιχούν στα πιο σημαντικά ογκοστοιχεία, ενώ οι μπλε στα πιο ασήμαντα. Τα αποτελέσματα αυτά είναι για την ακολουθία που χρησιμοποιήσαμε και στο Σχήμα 6.3, η οποία δείχνει έναν παγοδρόμο να γλυστράει στον πάγο κουνώντας και τα χέρια του. Η κάμερα κινείται ακολουθώντας τον παγοδρόμο. Η ακολουθία με κινούμενη κάμερα επιλέχτηκε σκόπιμα με σκοπό να φανεί η αξιοπιστία της προτεινόμενης μεθόδου σε περιπτώσεις κίνησης του υποβάθρου.

6.5.2 Ελαχιστοποίηση ενέργειας

Όπως αναφέραμε πριν, η ελαχιστοποίηση της (6.1) απαιτεί τον υπολογισμό μερικών παραγώγων των ενεργειών. Αναλυτικοί υπολογισμοί παρατίθενται στο Παράρτημα A.



(α) (β)

Σχήμα 6.4: *(α) Ενδιάμεσοι όγκοι ενδιαφέροντος που ενεργοποιούνται κατά την ελαχιστοποίηση με χρήση του ενδο-χαρακτηριστικού περιορισμού, (β) Αποτέλεσμα για τον $C_{2,2}$ μετά από 1, 5, 20 και 30 επαναλήψεις*

Συγκεκριμένα, η μερική παράγωγος της E_d υπολογίζεται ως

$$\frac{\partial E_d}{\partial C_{k,m}(s)} = 2 \cdot \sum_q \left(C_{k,m}(s) - C^0_{k,m}(s) \right) \tag{6.24}$$

Η μερική παράγωγος του ενδο-χαρακτηριστικού περιορισμού που δίνεται στην (6.21) υπολογίζεται ως

$$\frac{\partial E_1}{\partial C_{k,m}(s)} = 2 \cdot \left[C_{k,m}(s) - \frac{1}{|N_q|^2} \cdot \sum_{q \in N(s)} \left(2N \cdot C_{k,m}(q) - \sum_{r \in N_q} C_{i,\ell}(r) \right) \right] \tag{6.25}$$

Όπως φαίνεται στο Σχήμα 6.4 μόνο ογκοστοιχεία του ίδιου χαρακτηριστικού εμπλέκονται σε αυτό το κριτήριο. Δεν ενεργοποιούνται δια-κλιμακωτές ή δια-χαρακτηριστικές αλληλεπιδράσεις. Το αποτέλεσμα της χρήσης αυτού του κριτηρίου φαίνεται από την ενίσχυση της περιοχής του κορμού του αθλητή, καθώς είναι η πιο συνεκτική χωροχρονική περιοχή σε σχέση με φωτεινότητα, χρώμα και κατευθυντικότητα.

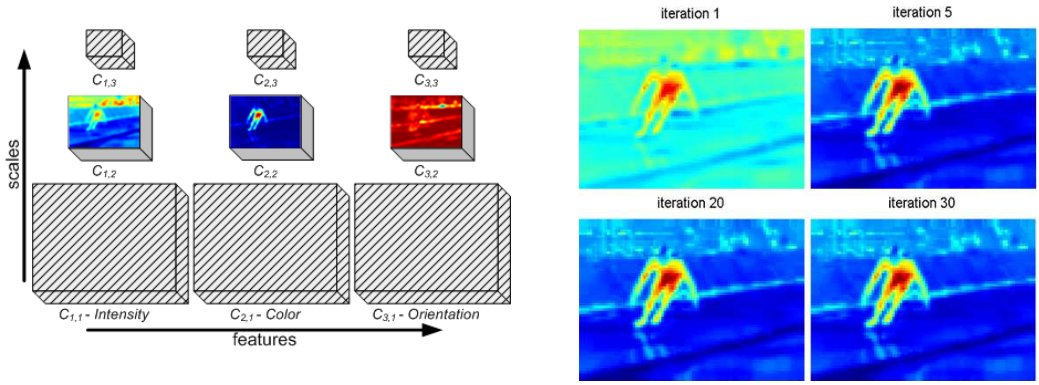
Η μερική παράγωγος του δια-ζωνικού περιορισμού, ο οποίος δίνεται στην (6.22), υπολογίζεται ως

$$\frac{\partial E_2}{\partial C_{k,m}(s)} = 2 \cdot \frac{M}{M-1} \cdot \left(C_{k,m}(s) - \frac{1}{M-1} \cdot \sum_{j \neq k} C_{j,m}(s) \right) \tag{6.26}$$

Όπως φαίνεται στο Σχήμα 6.5, μόνο ογκοστοιχεία διαφορετικών χαρακτηριστικών στην ίδια κλίμακα επιτρέπεται να ανταγωνιστούν όταν είναι ενεργοποιημένο το συγκεκριμένο κριτήριο. Άξιο παρατήρησης είναι το γεγονός ότι ο όγκος κίνησης δεν παρέχει χρήσιμη πληροφορία, καθώς η σχετική κίνηση της κάμερας και του αθλητή είναι μικρή. Παρόλ'αυτά οι περιοχές που ανήκουν στον αθλητή αναδεικνύονται εξαιτίας των υπολοίπων χαρακτηριστικών.

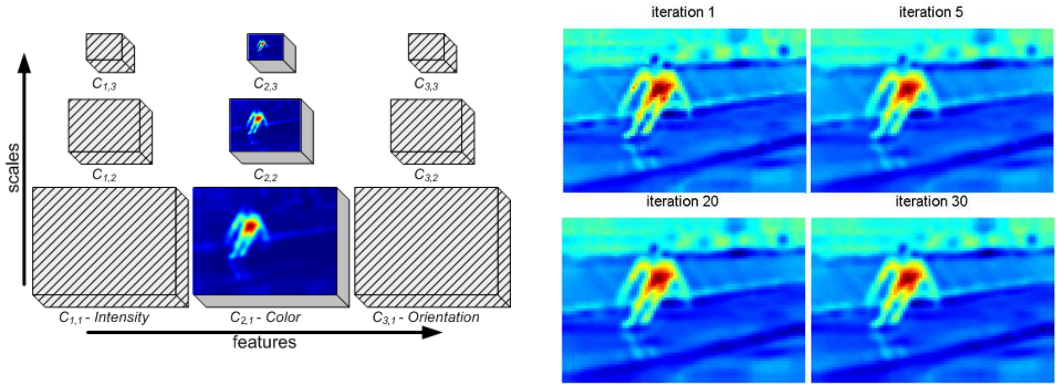
Η παράγωγος της τελευταίας ενέργειας που σχετίζεται με την δια-κλιμακωτή αλληλεπίδραση των χαρακτηριστικών και δίνεται στην (6.23) υπολογίζεται ως

$$\frac{\partial E_3}{\partial C_{k,m}(s)} = 2 \cdot \frac{L}{L-1} \cdot \left[C_{k,m}(s) - \frac{1}{L-1} \cdot \sum_{n \neq \ell} C_{k,n}(s) \right] \tag{6.27}$$



(α) (β)

Σχήμα 6.5: *(α) Ενδιάμεσοι όγκοι ενδιαφέροντος που ενεργοποιούνται κατά την ελαχιστοποίηση με χρήση του δια-χαρακτηριστικού περιορισμού, (β) Αποτέλεσμα για τον $C_{2,2}$ μετά από 1, 5, 20 και 30 επαναλήψεις*

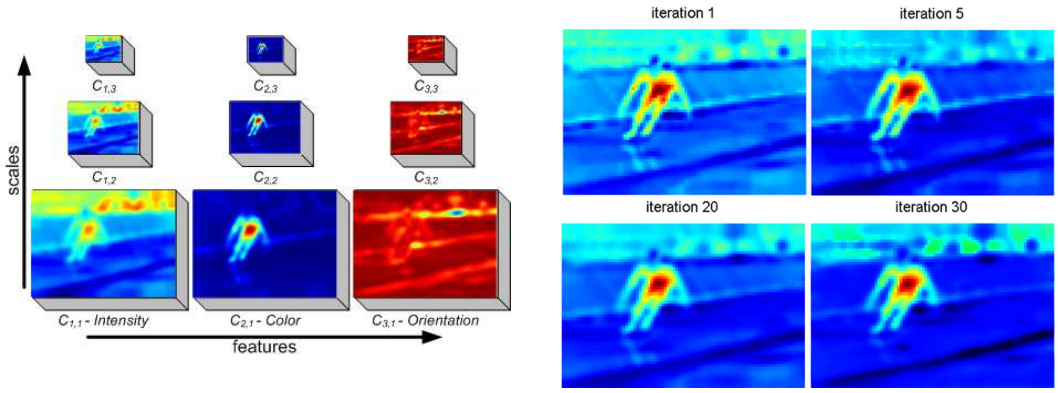


(α) (β)

Σχήμα 6.6: *(α) Ενδιάμεσοι όγκοι ενδιαφέροντος που ενεργοποιούνται κατά την ελαχιστοποίηση με χρήση του δια-κλιμακωτού περιορισμού, (β) Αποτέλεσμα για τον $C_{2,2}$ μετά από 1, 5, 20 και 30 επαναλήψεις*

Όπως φαίνεται στο Σχήμα 6.6, μόνο τα ογκοστοιχεία του ίδιου χαρακτηριστικού σε όλες τις κλίμακες επιτρέπεται να αλληλεπιδρούν. Τα ογκοστοιχεία που αντιστοιχούν στον "σκελετό" του παγοδρόμου είναι αυτά που γίνονται σημαντικά από επανάληψη σε επανάληψη. Εκτός από τον κορμό του αθλητή, τα πόδια και τα χέρια του φαίνονται πιο σημαντικά σε σχέση με το Σχήμα 6.4, στο οποίο η χωροχρονική εξομάλυνση είναι μεγαλύτερη λόγω του ενδο-χαρακτηριστικού περιορισμού. Το Σχήμα 6.7 δείχνει το τελικό αποτέλεσμα του μοντέλου όταν όλες οι ενέργειες εμπλέκονται στην ελαχιστοποίηση. Το υπόβαθρο γίνεται συνεχώς πιο ασήμαντο από επανάληψη σε επανάληψη, ενώ τα ογκοστοιχεία που ανήκουν στο αντικείμενο ενδιαφέροντος ομαδοποιούνται και ενισχύονται. Όπως φαίνεται από τα αποτελέσματα, η περιοχή του κορμού του αθλητή είναι αυτή που αναδεικνύεται εξαιτίας της έντονης συνεκτικότητας της σε όλη την διάρκεια της σκηνής.

Το Σχήμα 6.8 απεικονίζει με παραστατικό τρόπο τον κύριο όγκο για μία ακολουθία περπατήματος. Το Σχήμα 6.8α δείχνει ενδεικτικές τομές της αρχικής ακολουθίας, ενώ τα Σχήματα 6.8β και 6.8γ περιέχουν διαφορετικές απεικονίσεις του κύριου όγκου.Το Σχήμα 6.8δ δείχνει την εξέλιξη της ελαχιστοποίησης για μία σειρά επαναλήψεων. Είναι φανερή η ενίσχυση περιοχών που αντιστοιχούν στο αντικείμενο ενδιαφέροντος



(α) (β)

Σχήμα 6.7: *(α) Ενδιάμεσοι όγκοι ενδιαφέροντος που ενεργοποιούνται κατά την ελαχιστοποίηση με χρήση όλων των περιορισμών, (β) Αποτέλεσμα για τον $C_{2,2}$ μετά από 1, 5, 20 και 30 επαναλήψεις*

και η εξασθένηση αυτών που αντιστοιχούν στο υπόβαθρο.

Το αποτέλεσμα της ελαχιστοποίησης, αφού επιτευχθεί η επιθυμητή τιμή σφάλματος $\triangle C$, είναι ένα σύνολο μετασχηματισμένων ενδιάμεσων όγκων ενδιαφέροντος $\hat{\mathbf{C}} = \{\hat{C}_{i,\ell}\}$, οι οποίοι πρέπει να συνδυαστούν για να προκύψει ο κύριος όγκος, ως εξής:

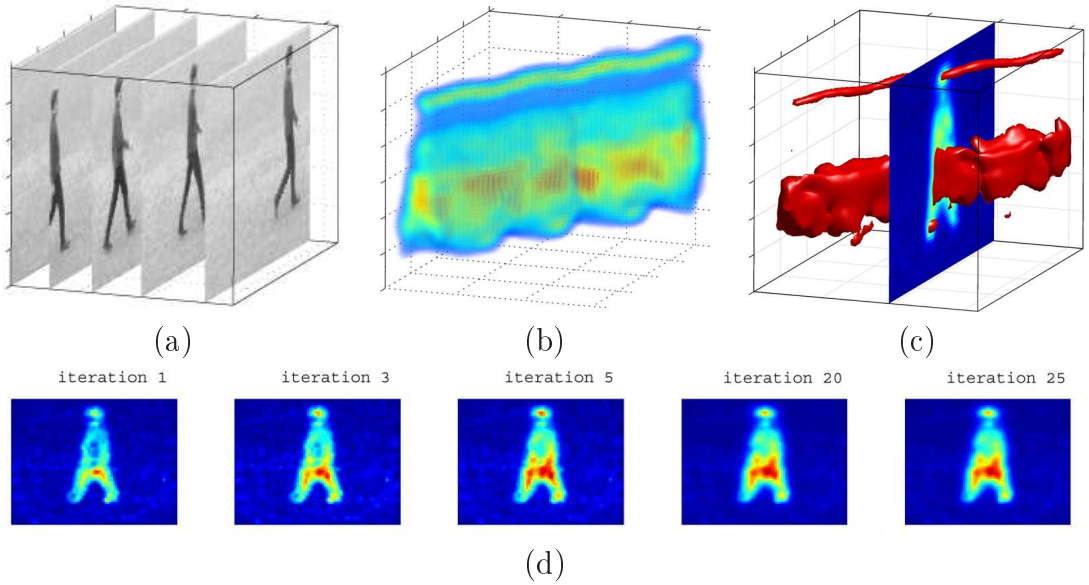
$$S = \frac{1}{3 \cdot L} \cdot \sum_{i=1}^{3} \sum_{\ell=1}^{L} \hat{C}_{i,\ell} \tag{6.28}$$

Παραδείγματα τομών κύριου όγκου απεικονίζονται στα Σχήματα 6.7, 6.8, 6.9 και στην ενότητα των πειραματικών αποτελεσμάτων. Το Σχήμα 6.9 δείχνει το αποτέλεσμα του μοντέλου όταν εφαρμόζεται σε μία κινούμενη οπτική απάτη, στην οποία ένα άλογο τρέχει υπό την έντονη παρουσία τυχαίου θορύβου [1]. Το άλογο γίνεται αντιληπτό εξαιτίας της τοπικής χωροχρονικής συνεκτικότητας των μερών του και όχι εξαιτίας της χωρικής του μορφής (είναι δύσκολο να γίνει αντιληπτό το άλογο μόνο παρατηρώντας μεμονωμένα καρέ). Το Σχήμα 6.9α δείχνει ένα αρχικό καρέ, ενώ τα υπόλοιπα δείχνουν την εξέλιξη της μεθόδου. Το άλογο γίνεται εύκολα αντιληπτό μετά από 25 επαναλήψεις (Σχήμα 6.9δ) λόγω της ενίσχυσης της χωροχρονικής συνεκτικότητας που επιβάλλει το μοντέλο.

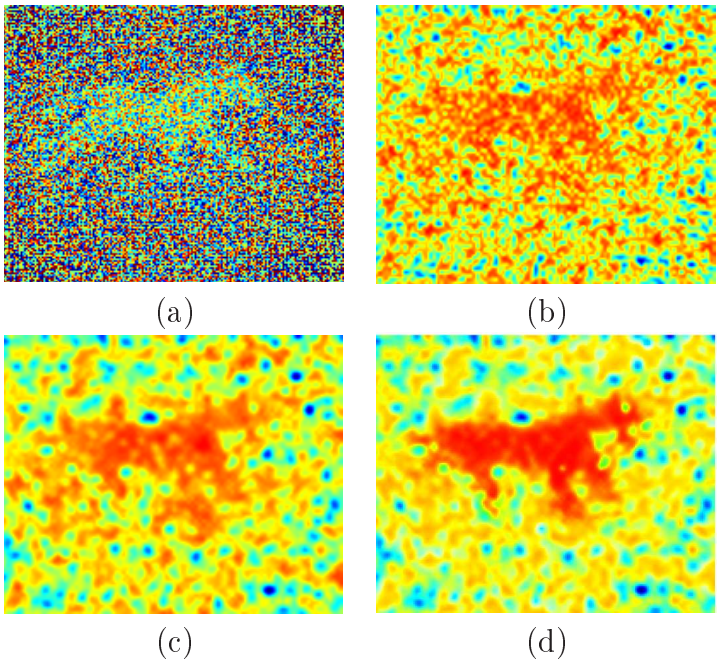
6.6 Εφαρμογές

Αξιολογούμε το πρώτο από τα δύο μοντέλα που προτείνουμε σε αυτό το κεφάλαιο σε αναγνώριση/κατηγοριοποίηση σκηνής και το δεύτερο σε αναγνώριση οπτικών δραστηριοτήτων σε ένα πλήθος διαφορετικών βάσεων ακολουθιών και στην δημιουργία περιλήψεων ταινιών από μία βάση με ανθρώπινο χαρακτηρισμό για ενδιαφέροντα γεγονότα. Στην περίπτωση της αναγνώρισης σκηνής συγκρίνουμε με καθιερωμένα μοντέλα οπτικής προσοχής, καθώς και με το μοντέλο που προτείναμε στο Κεφάλαιο 5. Για τις εφαρμογές του δεύτερου μοντέλου συγκρίνουμε με τεχνικές που σχετίζονται περισσότερο με αναγνώριση δραστηριοτήτων. Ο στόχος μας είναι να αποδείξουμε πειραματικά την δυνατότητα των μοντέλων να εντοπίζουν αντιπροσωπευτικές περιοχές της εισόδου και να διαχωρίζουν επιτυχώς τα σημαντικά γεγονότα από τα ασήμαντα.

[1] http://viperlib.york.ac.uk/, search for filename "horse2.mov"



Σχήμα 6.8: *(α) Τομές από μία ακολουθία περπατήματος, (β) Κύριος όγκος με μερική διαφάνεια για να φανούν οι σημαντικές περιοχές, (γ) Ισομετρική επιφάνεια, η οποία περιλαμβάνει τις περιοχές που το μοντέλο επέλεξε ως πιο σημαντικές, (δ) Αποτέλεσμα της ελαχιστοποίησης μετά από 1, 3, 5, 20 και 25 επαναλήψεις για την τομή που φαίνεται στο Σχήμα 6.8γ*



Σχήμα 6.9: *(α) Αρχικό καρέ της ακολουθίας οπτικής απάτης με το κινούμενο άλογο και η αντίστοιχη τομή του κύριου όγκου μετά από (β) 1, (γ) 10 και (δ) 25 επαναλήψεις. Το άλογο γίνεται αντιληπτό εξαιτίας της ενισχυμένης τοπικής χωροχρονικής συνεκτικότητας που εξασφαλίζει το προτεινόμενο μοντέλο.*

6.6.1 Κατηγοριοποίηση σκηνής

Στο Κεφάλαιο 5 περιγράψαμε την εφαρμογή του μοντέλου με επέκταση κίνησης σε αναγνώριση σκηνής. Σε αυτήν την ενότητα αναπτύσσουμε περισσότερο τις ιδέες μας γύρω από την χρήση οπτικής προσοχής σε κατηγοριοποίηση σκηνής,

εκτελούμε εκτενέστερα πειράματα που εξετάζουν διαφορετικές πτυχές της μεθόδου και συγκρίνουμε με τεχνικές του χώρου.

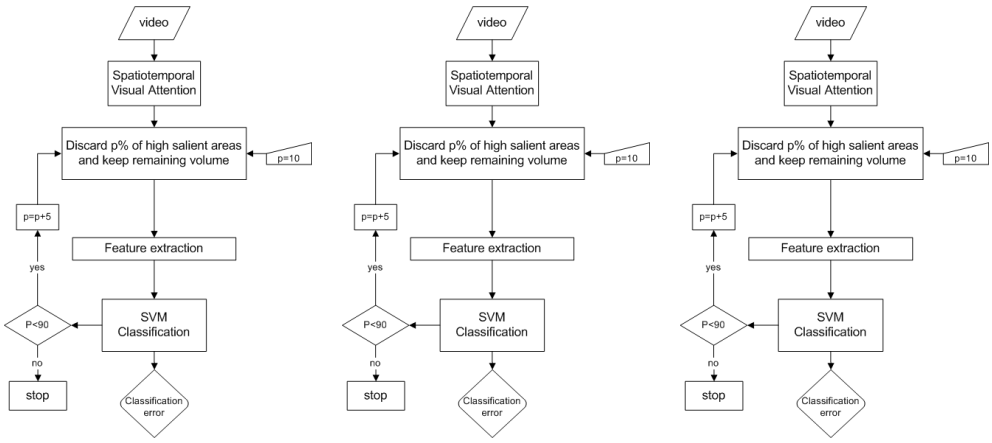
Θεωρούμε την αναγνώριση σκηνής ως μία διαδικασία αναγνώρισης που στηρίζεται σε περιοχές της ακολουθίας και όχι στο σύνολο της. Η επίτευξη κατάλληλης κατάτμησης που θα διαχωρίσει αυτές τις περιοχές δεν είναι εύκολη. Σε αυτό ακριβώς το σημείο εκμεταλλευόμαστε τον υπολογισμό σημαντικότητας ώστε να διαχωρίσουμε την οπτική είσοδο σε περιοχές χαμηλού/μέσου/υψηλού ενδιαφέροντος. Η κύρια ιδέα είναι να απορρίπτουμε σταδιακά περιοχές παρόμοιας σημαντικότητας ξεκινώντας από αυτές με ψηλή τιμή και να παρατηρούμε την απόδοση της κατηγοριοποίησης. Περιμένουμε ότι οι περιοχές χαμηλού ενδιαφέροντος θα σχετίζονται περισσότερο με το υπόβαθρο της σκηνής και επομένως θα την χαρακτηρίζουν καλύτερα (π.χ. το γήπεδο ποδοσφαίρου/καλαθοσφαίρισης και όχι οι παίκτες, οι λήψεις εντός του γηπέδου και όχι το κοινό κτλ). Για να υποστηρίξουμε αυτήν την ιδέα θέτουμε μία προϋπόθεση, η οποία συνήθως πληροίται: οι περιοχές του πραγματικού υποβάθρου (με την σημασιολογική έννοια) πρέπει να καλύπτουν μεγαλύτερη έκταση από αυτές του παρασκηνίου.

Για να σχηματίσουμε το διάνυσμα χαρακτηριστικών που θα τροφοδοτήσει τον κατηγοριοποιητή χρησιμοποιούμε ιστογράμματα των χαρακτηριστικών εισόδου, και συγκεκριμένα ιστογράμματα χρώματος και χωροχρονικής κατευθυντικότητας (κίνηση). Για να διατηρήσουμε το μήκος μικρό τα κβαντίζουμε και τελικά σχηματίζουμε το διάνυσμα χαρακτηρηστικών. Για την κατηγοριοποίηση χρησιμοποιούμε Support Vector Machines (SVMs), όπως και στην Ενότητα 5.3.2.2, η οποία παρέχει αναλυτικές πληροφορίες.

6.6.1.1 Κατηγοριοποίηση με περιοχές ενδιαφέροντος

Αξιολογούμε τα αποτελέσματα του προτεινόμενου μοντέλου συγκρίνοντας τα με μία ευρετική μέθοδο, δύο μεθόδους οπτικής προσοχής, το μοντέλο που παρουσιάσαμε στο Κεφάλαιο 5 και μία μέθοδο που βασίζεται στην Ανάλυση Κύριων Συνιστωσών (PCA), η οποία έχει χρησιμοποιηθεί για αφαίρεση υποβάθρου σε ακολουθίες [83, 77]. Ο λόγος που συμπεριλαμβάνουμε την τελευταία μέθοδο, αν και δεν σχετίζεται με κάποιο μοντέλο υπολογισμού ενδιαφέροντος, είναι για να επιβεβαιώσουμε την ορθότητα της υπόθεσης μας, το ότι δηλαδή το υπόβαθρο της σκηνής είναι σημαντικό για την αναγνώριση της (που σημαίνει ότι μία αξιόπιστη τεχνική αφαίρεσης υποβάθρου θα οδηγήσει σε χαμηλό σφάλμα κατηγοριοποίησης.

Η τεχνική που βασίζεται σε PCA αποτελείται από ένα βήμα υπολογισμού ιδιοτιμών και ένα βήμα απόρριψης των ιδιοδιανυσμάτων που αντιστοιχούν σε μικρές ιδιοτιμές. Η κύρια ιδέα είναι ότι τα κινούμενα αντικείμενα είναι συνήθως μικρότερα σε μέγεθος από το υπόβαθρο και επομένως δεν συνεισφέρουν σημαντικά στο μοντέλο ιδιοχώρου[39]. Πρακτικά, αυτό σημαίνει ότι τα μέρη της ακολουθίας που περιέχουν κινούμενα αντικείμενα δεν περιγράφονται ικανοποιητικά από το προηγούμενο μοντέλο και επομένως μπορούν να χαρακτηριστούν ως παρασκήνιο [83]. Το μοντέλο που προτείναμε στο Κεφάλαιο 5 μοιράζεται την ίδια έννοια χωροχρονικής σημαντικότητας, αλλά χωρίς το δομοστοιχείο ανταγωνισμού [91]. Στην Ενότητα 5.3.2 παρουσιάσαμε αποτελέσματα σε αναγνώριση σκηνής. Οι δύο μέθοδοι υπολογισμού ενδιαφέροντος είναι η χωρική μέθοδος των Itti *et al.* [48, 49, 50] και η επέκταση της με χάρτη κίνησης, όπως έχουμε προτείνει εμείς στο παρελθόν [100] αλλά και άλλοι ερευνητές του χώρου [148, 47]. Και οι δύο μέθοδοι καταλήγουν σε μία τιμή ενδιαφέροντος για κάθε εικονοστοιχείο. Η χωρική μέθοδος επεξεργάζεται την ακολουθία σε επίπεδο καρέ. Αφού υπολογιστεί ο κύριος χάρτης για κάθε καρέ, δημιουργούμε έναν όγκο



Σχήμα 6.10: *Αναγνώριση σκηνής με χρήση τιμών ενδιαφέροντος. (α) Διαχωρισμός παρασκηνίου/υπόβαθρου, (β) κατάτμηση σε > 1 περιοχών ενδιαφέροντος (γ) Επίπτωση του μεγέθους του συνόλου εκμάθησης*

από αυτά και εφαρμόζουμε ένα 3Δ φίλτρο ενδιάμεσης τιμής για να βελτιώσουμε την χρονική συνεκτικότητα και για να είμαστε δίκαιοι όταν συγκρίνουμε με χωροχρονικές τεχνικές. Ο χάρτης κίνησης υπολογίζεται με την τεχνική υπολογισμού οπτικής ροής των Black και Annandan, η οποία βασίζεται σε εύρωστη στατιστική [5]. Με την ίδια διαδικασία δημιουργούμε τον κύριο όγκο για την τεχνική που βασίζεται σε PCA. Για λόγους πληρότητας συμπεριλαμβάνουμε και μία ευρετική τεχνική, η οποία όμως σχετίζεται με την οπτική προσοχή ως εξής: Οι περισσότεροι δείχνουμε περισσότερο ενδιαφέρον για αυτά που συμβαίνουν στο κέντρο του οπτικού πεδίου [69] και επομένως αναμένουμε καλύτερα αποτελέσματα στην αναγνώριση όταν την περιορίζουμε στο αντίστοιχο μέρος της ακολουθίας. Στα πειράματα μας ο αρχικός όγκος μειώνεται κατά $p\%$ κάθε φορά και παρατηρούμε το σφάλμα αναγνώρισης. Η μείωση γίνεται χωρικά και ομοιόμορφα που σημαίνει ότι μειώνουμε τον όγκο από τις άκρες προς το κέντρο αφήνοντας την χρονική διάσταση ανέπαφη.

6.6.1.2 Πειραματική μεθοδολογία

Σε αυτή την ενότητα επιχειρούμε να δείξουμε τα οφέλη της χρήσης περιοχών ενδιαφέροντος που προκύπτουν από μεθόδους οπτικής προσοχής στην αναγνώριση σκηνής με τρία διαφορετικά πειράματα. Κάθε ένα από αυτά εφαρμόζεται στο ίδιο σύνολο ακολουθιών, αλλά "εξετάζει" την απόδοση με διαφορετικό τρόπο. Το πρώτο πείραμα φαίνεται στο Σχήμα 6.10α και αποτελείται από τρία βήματα: (1) απόρριψη ογκοστοιχείων με ψηλή τιμή ενδιαφέροντος μέχρι να παραμείνει το $p\%$ του αρχικού όγκου, (2) εξαγωγή ιστογραμμάτων των προϋπολογισμένων χαρακτηριστικών, (3) τροφοδοσία του κατηγοριοποιητή με τα διανύσματα χαρακτηριστικών και εξαγωγή τιμής σφάλματος για κάθε τιμή p. Ο κύριος όγκος κατατμείται σε κάθε επανάληψη του πειράματος σε δύο περιοχές, σε μία με μεγάλη και μία με χαμηλη τιμή μέσου ενδιαφέροντος. Η κατάτμηση γίνεται με κατωφλίωση ([85]) σύμφωνα με το ποσοστό των ογκοστοιχείων που πρέπει να διατηρηθεί. Πρακτικά, ο όγκος κατωφλιώνεται επαναληπτικά με ένα μικρό βήμα κατωφλίωσης ώσπου να επιτευχθεί προσεγγιστικά το επιθυμητό ποσοστό p. Σε κάθε επανάληψη παράγεται μία σημαντική και μία ασήμαντη περιοχή. Το διάνυσμα χαρακτηριστικών που προκύπτει από την λιγότερο σημαντική περιοχή (υπόβαθρο) έχει πάντα ίδιο μέγεθος και δημιουργείται από δύο ιστογράμματα:

(α) ιστόγραμμα χρώματος με 32 κελιά ανά κανάλι χρώματος (96 κελιά συνολικά) και (β) ιστόγραμμα κίνησης με 16 κελιά. Το τελικό μήκος του διανύσματος είναι επομένως 112.

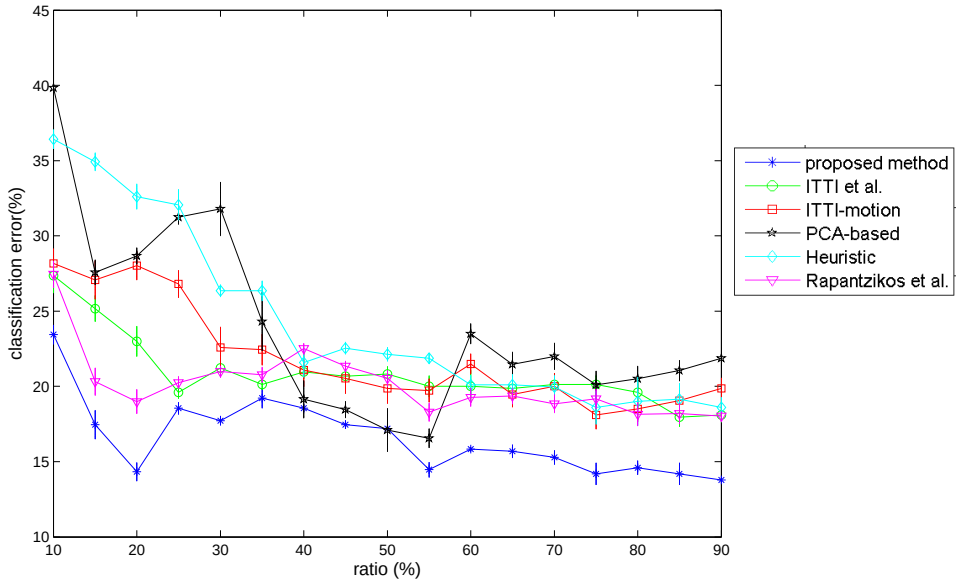
Διαισθητικά μπορούμε να ισχυριστούμε ότι υπάρχει ένας αριθμός περιοχών που αναπαριστά καλύτερα την σκηνή. Για παράδειγμα στην περίπτωση των σκηνών από αθλήματα, μία περιοχή μπορεί να περιέχει το γήπεδο, μία άλλη τους παίκτες, τις διαφημίσεις, το κοινό κτλ. Κάθε μία από αυτές τις περιοχές αντιστοιχεί σε ένα χαρακτηριστικό της σκηνής, αλλά δεν είναι όλες απαραίτητες για να καταλάβουμε το περιεχόμενο της. Ακολουθώντας την λογική του προηγούμενου πειράματος, περιμένουμε ότι αν επιλεχθούν οι κατάλληλες περιοχές το σφάλμα κατηγοριοποίησης θα μειωθεί. Στο δεύτερο πείραμα επομένως κατατμούμε τον εκάστοτε κύριο όγκο σε ποικίλο αριθμό περιοχών, όπως φαίνεται σχηματικά στο Σχήμα 6.10β. Ακολουθείται παρόμοια διαδικασία με το πρώτο πείραμα με την διαφορά ότι σε κάθε επανάληψη ο όγκος κατατμείται σε περισσότερες της μίας περιοχής. Μετά την κατάτμηση οι περιοχές ταξινομούνται με βάση την μέση τιμή ενδιαφέροντος τους και η πιο σημαντική απορρίπτεται. Αυτό το σενάριο έχει μία δυσκολία, καθώς αν ο αριθμός των περιοχών δεν είναι σταθερός για κάθε ακολουθία, το μήκος του διανύσματος χαρακτηριστικών δεν θα είναι σταθερό και επομένως δεν θα είναι εφικτή η απευθείας σύγκριση μεταξύ των διανυσμάτων. Για να ξεπεράσουμε αυτό το πρόβλημα, κατατμούμε τον κύριο όγκο σε έναν προκαθορισμένο αριθμό περιοχών με την μέθοδο k-means. Τα ογκοστοιχεία ομαδοποιούνται ανάλογα με την τιμή ενδιαφέροντος τους και καταλήγουμε με έναν προκαθορισμένο αριθμό περιοχών. Στην συνέχεια ταξινομούμε τις κλάσσεις και απορρίπτουμε την πιο σημαντική. Σε αυτό το σενάριο χρησιμοποιούμε 8 κελιά ανά χρωματικό κανάλι (24 στοιχεία ανά περιοχή) και 4 κελιά για την κίνηση με αποτέλεσμα το τελικό μήκος του διανύσματος χαρακτηριστικών να είναι 32.

Πόσο αξιόπιστες είναι οι συγκρινόμενες τεχνικές με μικρό μέγεθος συνόλου εκμάθησης; Στο τρίτο πείραμα μελετούμε την απόδοση των αλγορίθμων με διαφορετικά μεγέθη συνόλων εκμάθησης. Αντί λοιπόν να χρησιμοποιήσουμε ένα σταθερό σύνολο εκμάθησης, το οποίο συνήθως είναι μεγάλο σε σχέση με το σύνολο των διαθέσιμων δεδομένων, χρησιμοποιούμε διαφορετικά μεγέθη για να εξετάσουμε την ευαισθησία των τεχνικών. Αυτό το πείραμα σχετίζεται με τα πρώτα δύο ως εξής: Από τα αποτελέσματα του πρώτου πειράματος, επιλέγουμε τρεις τιμές για τις οποίες οι τεχνικές παρουσιάζουν προσεγγιστικά ελάχιστο (π.χ. στις τιμές 20%, 55% και 90%) και εκπαιδεύουμε τους κατηγοριοποιητές με διαφορετικά μεγέθη συνόλων εκμάθησης. Η ίδια διαδικασία ακολουθείται για τις θέσεις των ελαχίστων στο δεύτερο πείραμα. Τα βήματα συνοψίζονται στο Σχήμα 6.10γ.

6.6.1.3 Πειραματικά αποτελέσματα

Για τα πειράματα επεκτείναμε το σύνολο ακολουθιών που είδαμε στην Ενότητα 5.3.2. Αυξήθηκε ο αριθμός των ακολουθιών για τα εφτά διαφορετικά αθλήματα: ποδόσφαιρο (144), κολύμβηση (120), καλαθοσφαίριση (140), πυγμαχία (132), σνούκερ (128), αντισφαίριση (130) και επιτραπέζια αντισφαίριση (130). Οι αριθμοί στις παρενθέσεις δείχνουν τον αριθμό ακολουθιών για κάθε κατηγορία. Οι ακολουθίες είναι στο μεγαλύτερο ποσοστό από τους Ολυμπιακούς αγώνες της Αθήνας το 2004. Κάθε κατηγορία περιλαμβάνει μακρινές και κοντινές λήψεις παικτών και γηπέδου, καθώς και καρέ που περιέχουν ταυτόχρονα περιοχές γηπέδου, παικτών και ακροατηρίου. Οι ακολουθίες διαρκούν περίπου 6-7 δευτερόλεπτα και έχουν ίδιο χωρικό μέγεθος.

Στην συνέχεια παραθέτουμε στατιστικά για τα τρία πειράματα που θα μας



Σχήμα 6.11: *Πείραμα Ι, Αναγνώριση σκηνής με διαχωρισμό παρασκηνίου/υπόβαθρου βάσει τιμών ενδιαφέροντος*

βοηθήσουν να αξιολογήσουμε τα προτεινόμενα μοντέλα. Τα κύρια κριτήρια είναι η εξειδικευτικότητα (specificity) και η ευαισθησία (sensitivity). Υπολογίζονται επίσης διάφορα στατιστικά πάνω στο σφάλμα κατηγοριοποίησης $p(r)$, όπου r είναι η εξαρτημένη μεταβλητή.

$$
\begin{aligned}
Avg\ error\ of\ \nu-fold\ validation &: \mu_{err} = \frac{1}{\nu \cdot |r|} \sum_{\nu} \sum_{r} p(r) \\
Std\ deviation\ of\ \nu-fold\ validation &: \sigma_{err} = \frac{1}{|r|} \sum_{r} \left(p(r) - \frac{1}{|r| \sum_r p(r)} \right)^2 \\
Avg\ std\ error\ of\ \nu-fold\ validation &: \hat{\sigma}_{err} = \frac{1}{|r|\sqrt{\nu}} \sum_{r} \left(p(r) - \frac{1}{|r| \sum_r p(r)} \right)^2 \\
Min\ classification\ error &: MCE = arg\min r(p(r)) \\
sensitivity &: \frac{tp}{(tp+fn)} \\
specificity &: \frac{tn}{(fp+tn)}
\end{aligned}
\tag{6.29}
$$

όπου tp, fn, fp είναι οι *true-positive, false-negative* και *false-positive* ρυθμοί.

Αναγνώριση βασισμένη σε σημαντικές περιοχές Μία αυθόρμητη ερώτηση για τον ερευνητή του χώρου είναι το πόσο σημαντικές είναι τελικά οι περιοχές που προκύπτουν από τα μοντέλα οπτικής προσοχής. Ακολουθώντας την μεθοδολογία που περιγράψαμε στην Ενότητα 6.6.1.2 επιχειρούμε να δώσουμε απάντηση σε αυτό το ερώτημα. Το πρώτο πείραμα αντιστοιχεί στο Σχήμα 6.10α. Στο γράφημα 6.11 απεικονίζεται το σφάλμα κατηγοριοποίησης για όλες τις μεθόδους όταν αλλάζει το μέγεθος του μέρους της σκηνής που απορρίπτουμε. Κάθε σημείο στο γράφημα οδηγεί στο εξής συμπέρασμα: αν π.χ. απορρίψουμε 10% των σημαντικών ογκοστοιχείων

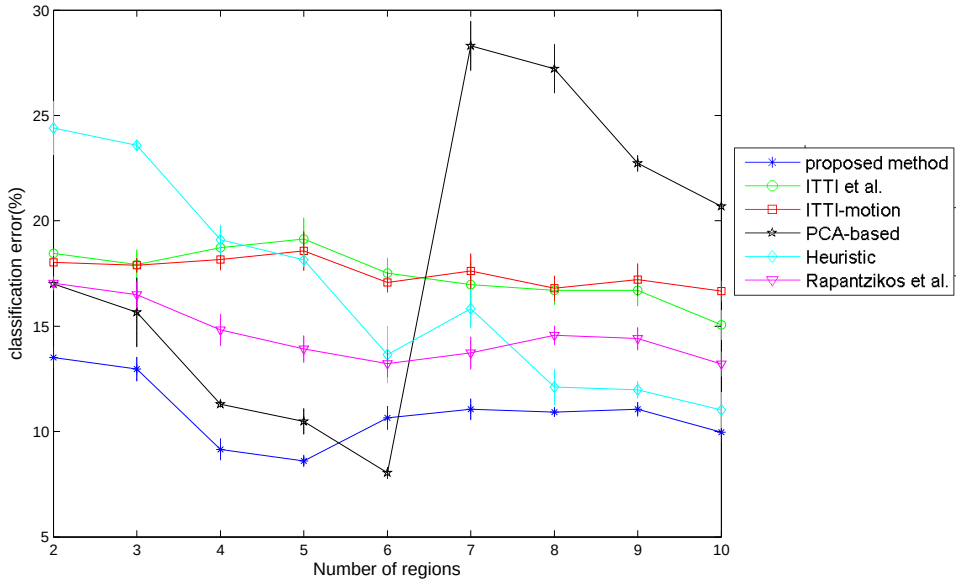
Πίνακας 6.1: *Στατιστικά αναγνώρισης σκηνής με διαχωρισμό παρασκηνίου/υπόβαθρου (%)*

	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	13.78±0. 14	16.58	0.48	2.51
Itti et al.	17.96±0.63	20.86	0.67	2.35
ITTI-motion	18.10±0.93	21.93	0.93	3.43
PCA-based	16.56±0.63	23.83	0.94	6.16
Heuristic	18.60±1.1	24.26	0.75	6.09
Rapantzikos et al.	18.04±0.3	20.14	0.58	2.27

(άξονας x), τότε επιτυγχάνουμε ένα μέσο σφάλμα (άξονας y) μετά από ν επαναλήψεις του πειράματος (τυπικό εύρος σφάλματος για κάθε σημείο). Στα πειράματα μας $\nu = 5$. Στην περίπτωση της ευρετικής μεθόδου το ποσοστό στον άξονα x αντιστοιχεί στο ποσοστό της περιοχής που απορρίπτεται ξεκινώντας από το κέντρο του αρχικού όγκου της ακολουθίας. Οι μέθοδοι Itti *et al.* και *ITTI-motion* έχουν παρόμοια αποτελέσματα σε απόλυτο σφάλμα για τιμές $p > 30\%$ με την *ITTI-motion* να έχει ψηλότερο $\hat{\sigma}_{err}$ και επομένως μεγαλύτερη αβεβαιότητα στην αναγνώριση. Η απόδοση της μεθόδου που προτείναμε στο Κεφάλαιο 5, η οποία συμβολίζεται στα γραφήματα και στους πίνακες ως Rapantzikos *et al.*, βρίσκεται κοντά στις δύο προηγούμενες τεχνικές. Η *PCA-based* μέθοδος παρουσιάζει αυξημένες διακυμάνσεις (απότομες αλλαγές από το 10% στο 20% και από το 30% στο 40%), αλλά επιτυγχάνει μικρότερο σφάλμα (στο 55%). Το σφάλμα της προτεινόμενης μεθόδου είναι συνεχώς μικρότερο από των υπολοίπων και με μικρότερη μέση διακύμανση για κάθε τιμή του p. Ο Πίνακας 6.1 δείχνει στατιστικά για όλες τις μεθόδους.

Το δεύτερο πείραμα εξετάζει την απόδοση των μεθόδων όταν τα χαρακτηριστικά εξάγονται από περισσότερες της μίας περιοχές, όπως εξηγήσαμε στην Ενότητα 6.6.1.2. Το πείραμα αυτό αντιστοιχεί στο Σχήμα 6.10β. Το γράφημα 6.12 απεικονίζει το σφάλμα κατηγοριοποίησης σε σχέση με τον αριθμό σημαντικών περιοχών που επιλέγονται. Οι δύο μέθοδοι που βασίζονται στην μέθοδο των Itti *et al.* έχουν παρόμοια απόδοση χωρίς ιδαίτερες διακυμάνσεις. Αν και η μέση τιμή σφάλματος είναι σχεδόν ίδια (Πίνακας 6.2), το μέσο τυπικό σφάλμα για την χωρική μέθοδο είναι ψηλότερο. Αυτό είναι αναμενόμενο, καθώς η επέκταση με κίνηση παράγει πιο συνεκτικές χωροχρονικές περιοχές. Η ευρετική μέθοδος καταλήγει σε χαμηλότερο σφάλμα μετά από μία απότομη διακύμανση (στις 6 περιοχές), αλλά έχει υψηλό μέσο τυπικό σφάλμα. Η μέθοδος *Rapantzikos et al.* αποδεικνύεται καλύτερη των προηγούμενων τεχνικών, ενώ η *PCA-based* και η προτεινόμενη αποδίδουν συνολικά καλύτερα.

Επίπτωση του μεγέθους του συνόλου εκμάθησης Ιδανικά, μία αξιόπιστη μέθοδος αναγνώρισης θα αποδίδει ακόμη και όταν τα διαθέσιμα δεδομένα εκμάθησης είναι λίγα. Υιοθετούμε την διαδικασία που περιγράφεται στο Σχήμα 6.10γ για να αξιολογήσουμε τις μεθόδους ως προς αυτήν την παράμετρο. Για το πρώτο μέρος του πειράματος επιλέγουμε τις τρεις τιμές για τις οποίες οι μέθοδοι παρουσιάζουν τα καλύτερα αποτελέσματα, συγκεκριμένα 20%, 55% και 90%, ενώ για το δεύτερο μέρος επιλέγουμε τις τιμές που δίνουν τα καλύτερα αποτελέσματα στο δεύτερο πείραμα



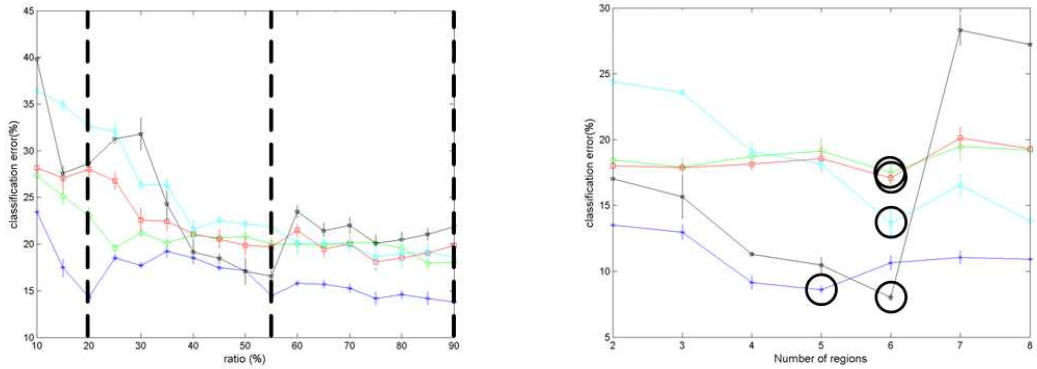
Σχήμα 6.12: *Πείραμα ΙΙ, Αναγνώριση σκηνής με κατάτμηση σε περιοχές ενδιαφέροντος*

Πίνακας 6.2: *Στατιστικά αναγνώρισης σκηνής με κατάτμηση σε περιοχές ενδιαφέροντος (%)*

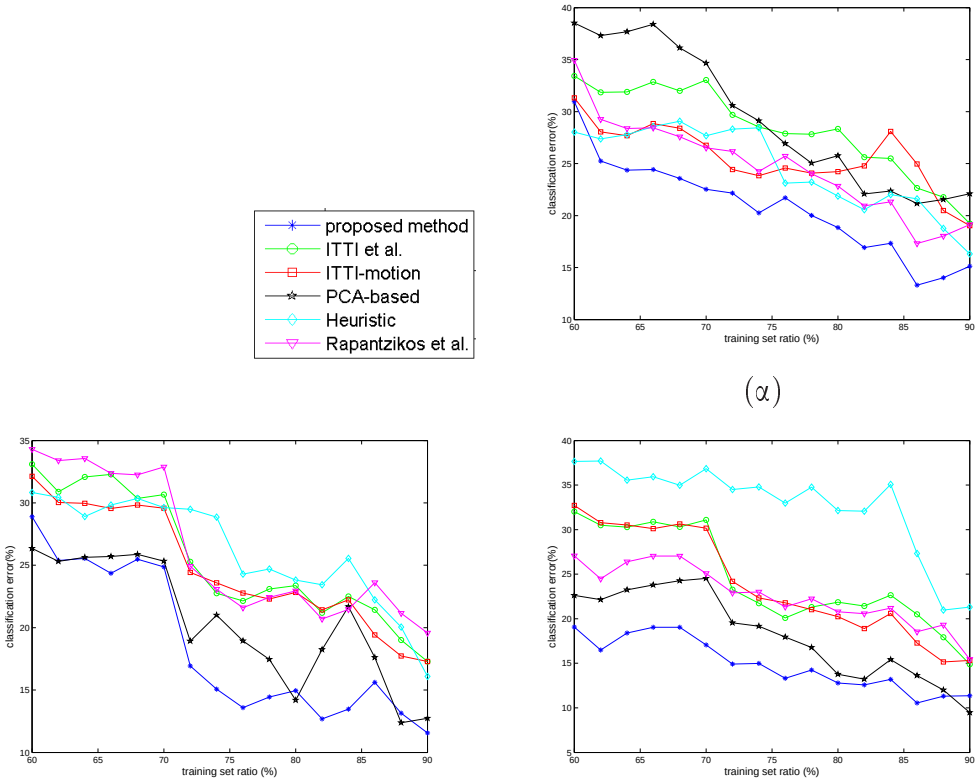
	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	8.61±0.27	**10.88**	**0.43**	1.59
Itti et al.	17.51±0.69	18.63	0.74	1.27
ITTI-motion	17.07±0.48	18.44	0.56	**0.66**
PCA-based	**8.03±0.27**	16.85	0.78	7.31
Heuristic	13.65±1.35	18.46	0.78	4.37
Rapantzikos et al.	13.21±0.58	14.60	0.59	1.35

(κατάτμηση σε περισσότερες της μίας σημαντικές περιοχές). Οι επιλεγμένες τιμές φαίνονται στο Σχήμα 6.13.

Όταν το σύνολο εκμάθησης είναι αρκετά μεγάλο, όλες οι μέθοδοι αποδίδουν καλά με την προτεινόμενη και την *PCA-based* να είναι οι καλύτερες. Τα Σχήματα 6.14 και 6.15 απεικονίζουν την απόδοση των μεθόδων σε σχέση με το μέγεθος του συνόλου εκμάθησης. Οι τιμές στον άξονα x αντιστοιχούν στο ποσοστό του συνόλου εκμάθησης ως προς το σύνολο των διαθέσιμων δεδομένων. Για κάθε μέγεθος του συνόλου, η κατηγοριοποίηση επαναλαμβάνεται δέκα φορές με τυχαία διχοτόμηση. Στο Σχήμα 6.14 φαίνεται η αναμενόμενη καθοδική τάση του σφάλματος όταν αυξάνει το μέγεθος του συνόλου. Το γράφημα δείχνει καθαρά ότι η προτεινόμενη μέθοδος αποδίδει καλύτερα, καθώς η συνολική μεταβλητότητα και το σφάλμα είναι χαμηλότερα των άλλων. Τα στατιστικά των Πινάκων 6.3, 6.4 και 6.5 επιβεβαιώνουν αυτήν την παρατήρηση. Οι τεχνικές που βασίζονται στην αρχιτεκτονική των Itti *et al.* έχουν παρόμοια συμπεριφορά, ενώ η *PCA-based* έχει χαμηλό σφάλμα και μέτρια μεταβλητότητα. Η ευρετική μέθοδος έχει συνολικά καλή απόδοση, αλλά δεν μπορεί να χαρακτηριστεί σταθερή, όπως φαίνεται και από την μέση τιμή σφάλματος.



Σχήμα 6.13: *Πείραμα III, (α) Επιλεγμένες τιμές από το Πείραμα I στα σημεία που οι μέθοδοι φτάνουν προσεγγιστικά σε ελάχιστο, (β) το αντίστοιχο για τις τιμές από το Πείραμα II*



Σχήμα 6.14: *Πείραμα III, Απόδοση των μεθόδων με μεταβλητό σύνολο εκμάθησης για τις τιμές (α) 20%, (β) 55% και (γ) 90% από το πείραμα I*

Στην περίπτωση της κατηγοριοποίησης με κατάτμηση σε σημαντικές περιοχές (Σχήμα 6.15), η *PCA-based* μέθοδος επιτυγχάνει το χαμηλότερο σφάλμα με σύνολο εκμάθησης 90% με πέντε σημαντικές περιοχές (Πίνακας 6.6). Το προτεινόμενο μοντέλο έχει παραπλήσιο σφάλμα με την προηγούμενη μέθοδο στην ίδια τιμή του συνόλου, αλλά αποδίδει καλύτερα και είναι πιο σταθερή σε όλες τις υπόλοιπες τιμές. Οι μέθοδοι που βασίζονται σο μοντέλο των Itti έχουν παρόμοια απόδοση, ενώ η ευρετική επιτυγχάνει απροσδόκητα χαμηλό σφάλμα.

Πίνακας 6.3: *Στατιστικά αναγνώρισης σκηνής με μεταβλητό μέγεθος συνόλου εκμάθησης για την τιμή 20% του πειράματος I*

	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	**13.31±0.64**	**20.67**	0.68	4.67
Itti et al.	19.24±0.27	28.25	0.68	4.34
ITTI-motion	19.02±0.72	25.59	0.79	**3.15**
PCA-based	21.15±0.15	29.33	**0.66**	6.8
Heuristic	16.30±0.88	24.54	0.98	4.08
Rapantzikos et al.	17.31±0.73	24.67	0.68	4.67

Πίνακας 6.4: *Στατιστικά αναγνώρισης σκηνής με μεταβλητό μέγεθος συνόλου εκμάθησης για την τιμή 55% του πειράματος I*

	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	**11.57±0.65**	**18.50**	0.57	6.00
Itti et al.	17.28±0.65	25.46	**0.53**	5.22
ITTI-motion	17.28±0.27	24.69	0.61	4.81
PCA-based	12.38±0.85	20.46	0.83	4.89
Heuristic	16.09±0.41	26.15	0.69	**4.34**
Rapantzikos et al.	19.57±0.65	26.26	0.55	5.64

Πίνακας 6.5: *Στατιστικά αναγνώρισης σκηνής με μεταβλητό μέγεθος συνόλου εκμάθησης για την τιμή 90% του πειράματος I*

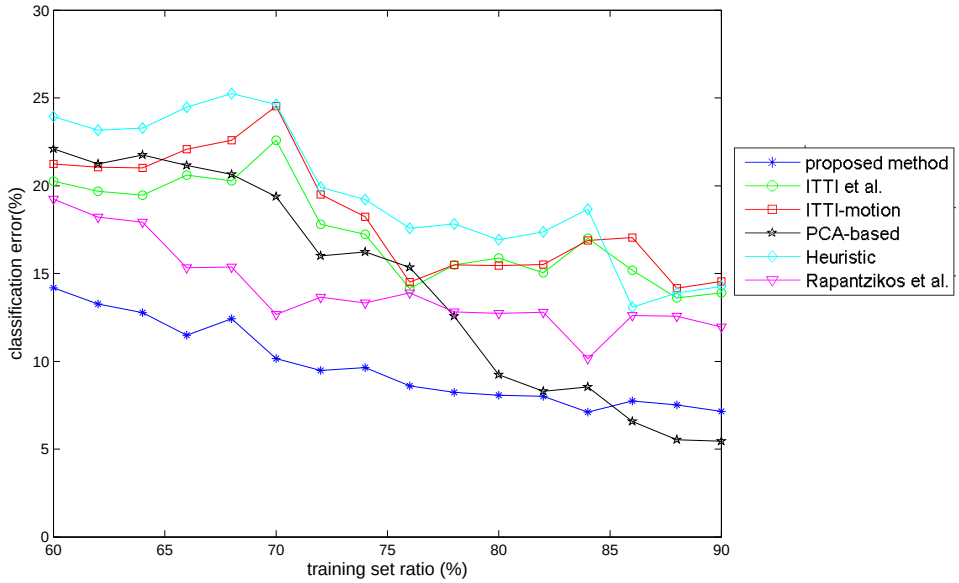
	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	**9.81±0.38**	**11.63**	0.63	**1.76**
Itti et al.	16.20±0.43	23.92	**0.61**	5.75
ITTI-motion	16.63±0.75	24.85	0.66	5.77
PCA-based	11.77±0.60	17.26	0.69	4.60
Heuristic	26.96±1.39	39.08	0.78	4.72
Rapantzikos et al.	17.81±0.74	19.63	0.62	1.82

6.6.2 Ανίχνευση οπτικών δραστηριοτήτων

6.6.2.1 Σύνολα δεδομένων και αλγόριθμοι για σύγκριση

Χρησιμοποιήσαμε τρία σύνολα ακολουθιών για να αξιολογήσουμε τη μέθοδο μας σε αναγνώριση οπτικών δυναμικών γεγονότων. Το πρώτο είναι το σύνολο ακολουθιών KTH, ένα από τα μεγαλύτερα που είναι και δημόσια διαθέσιμα, το οποίο αποτελείται από 6 διαφορετικές ανθρώπινες δραστηριότητες [114]. Πρόκειται για το ίδιο σύνολο που περιγράψαμε στην Ενότητα 4.3.1 όπου και παρέχονται αναλυτικές πληροφορίες. Το δεύτερο σύνολο ακολουθιών σχετίζεται με εκφράσεις προσώπου και είναι επίσης δημόσια διαθέσιμο[2] [28]. Περιέχει ακολουθίες από δύο ανθρώπους που εκτελούν

[2] http://vision.ucsd.edu/∼pdollar/



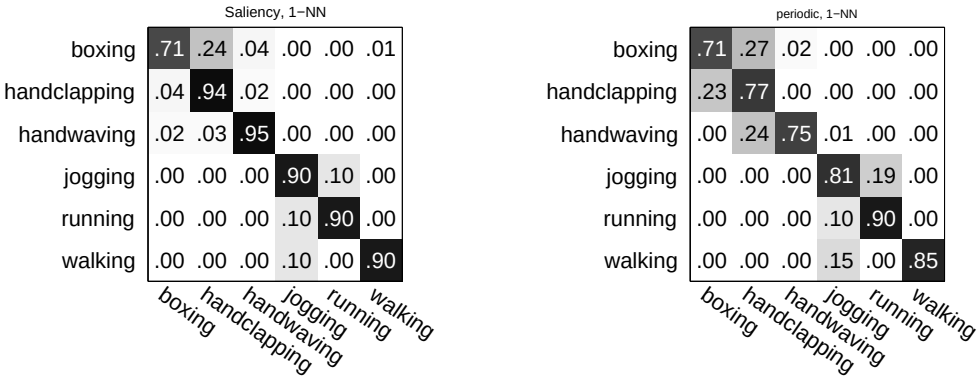
Σχήμα 6.15: *Πείραμα ΙΙΙ, Απόδοση μεθόδων με μεταβλητό σύνολο εκμάθησης για τις περιοχές που επιλέχθηκαν από το πείραμα ΙΙ*

Πίνακας 6.6: *Στατιστικά αναγνώρισης σκηνής με μεταβλητό μέγεθος συνόλου εκμάθησης για την τιμή 90% του πειράματος Ι*

	MCE	μ_{err}	$\hat{\sigma}_{err}$	σ_{err}
Proposed method	7.11±0.27	**9.74**	0.60	**2.36**
Itti et al.	13.61±0.45	17.38	0.58	2.79
ITTI-motion	14.16±0.18	18.36	0.47	3.36
PCA-based	**5.46±0.63**	14.38	**0.40**	6.32
Heuristic	13.08±0.57	19.59	0.80	4.09
Rapantzikos et al.	10.16±0.50	14.08	0.59	2.50

έξι διαφορετικές εκφράσεις προσώπου: θυμός, αποστροφή, φόβος, χαρά, θλίψη και έκπληξη. Κάθε άνθρωπος επαναλαμβάνει κάθε έκφραση οχτώ φορές. Το τρίτο σύνολο περιέχει ακολουθίες ανθρώπινων δραστηριοτήτων σε πολύπλοκα περιβάλλοντα, τα οποία χαρακτηρίζονται από κίνηση στο υπόβαθρο, πλήθος αντικειμένων, επικαλύψεις και αλλαγές βάθους/γωνίας λήψης της κάμερας.

Για τα πειράματα χρησιμοποιούμε το ίδιο πλαίσιο με αυτό που προτείνουν οι Dollár *et al.* [28] και συγκρίνουμε τα αποτελέσματα με δύο διαφορετικές τεχνικές ανίχνευσης χωροχρονικών σημείων ενδιαφέροντος, και συγκεκριμένα αυτές των Laptev *et al.* [62] και των Dollár *et al.* [28]. Οι Laptev *et al.* εντοπίζουν τοπικά χωροχρονικά σημεία επεκτείνοντας τον ανιχνευτή γωνίας του Harris [40]στο χωροχρονικό πεδίο. Ο ανιχνευτής βασίζεται στον πίνακα δεύτερων ροπών, ο οποίος περιγράφει την τοπική κατανομή κλίσεων σε πολλαπλές χωρικές και χρονικές κλίμακες. Τα τοπικά μέγιστα αυτής της κατανομής, η οποία αποτελεί το αντίστοιχο του προτεινόμενου κύριου όγκου, αντιστοιχούν στα σημεία ενδιαφέροντος. Η λογική τους είναι ότι χωροχρονικές γωνίες, δηλαδή σημεία στα οποία το διάνυσμα ταχύτητας αντιστρέφεται, ταιριάζουν στην αναπαράσταση ανθρώπινων δραστηριοτήτων. Οι Dollár *et al.*



(a) (b)

Σχήμα 6.16: *Αποτελέσματα αναγνώρισης δραστηριοτήτων για (α) την προτεινόμενο μέθοδο, (β) την μέθοδο ανίχνευσης περιοδικότητας και (γ) την μέθοδο ανίχνευσης χωροχρονικών γωνιών (stHarris)*

προτείνουν και αξιολογούν ενα πλαίσιο αναγνώρισης, το οποίο βασίζεται σε έναν ανιχνευτή περιοδικών σημείων. Συνοπτικά, ο ανιχνευτής υλοποιείται με χωρική εξομάλυνση του όγκου φωτεινότητας και με φιλτράρισμα στον χρόνο με ένα ζευγάρι 1Δ φίλτρων Gabor. Διαισθητικά, οι τιμές ενδιαφέροντος που προκύπτουν είναι μεγάλες στα σημεία που η φωτεινότητα αλλάζει χρονικά με έναν συγκεκριμένο ρυθμό, ο οποίος καθορίζεται από την συχνότητα και την κλίμακα των φίλτρων Gabor.

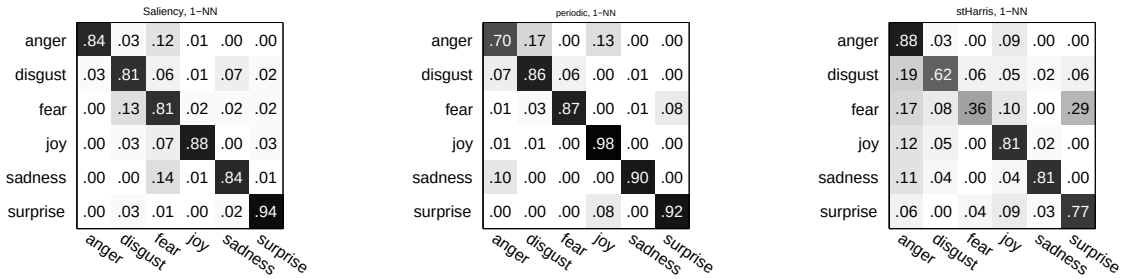
6.6.2.2 Αναγνώριση δραστηριοτήτων

Χρησιμοποιούμε το πειραματικό πλαίσιο των Dóllar *et al.* για να είμαστε σε θέση να παρέχουμε δίκαια αποτελέσματα και να απομονώσουμε την επίδραση του ανιχνευτή σημαντικών σημείων που βασίζεται στο προτεινόμενο μοντέλο. Τα σημαντικά σημεία ανιχνεύονται ως τα τοπικά μέγιστα του κύριου όγκου S. Δεν υπάρχει η ανάγκη καθορισμού ενός ελάχιστου κατωφλίου ενεργοποίησης, καθώς οι περιοχές χαμηλού ενδιαφέροντος, όπως προκύπτουν από το προτεινόμενο μοντέλο, είναι συνεκτικές και ομοιογενείς με αποτέλεσμα να μην έχουν έντονα τοπικά μέγιστα.

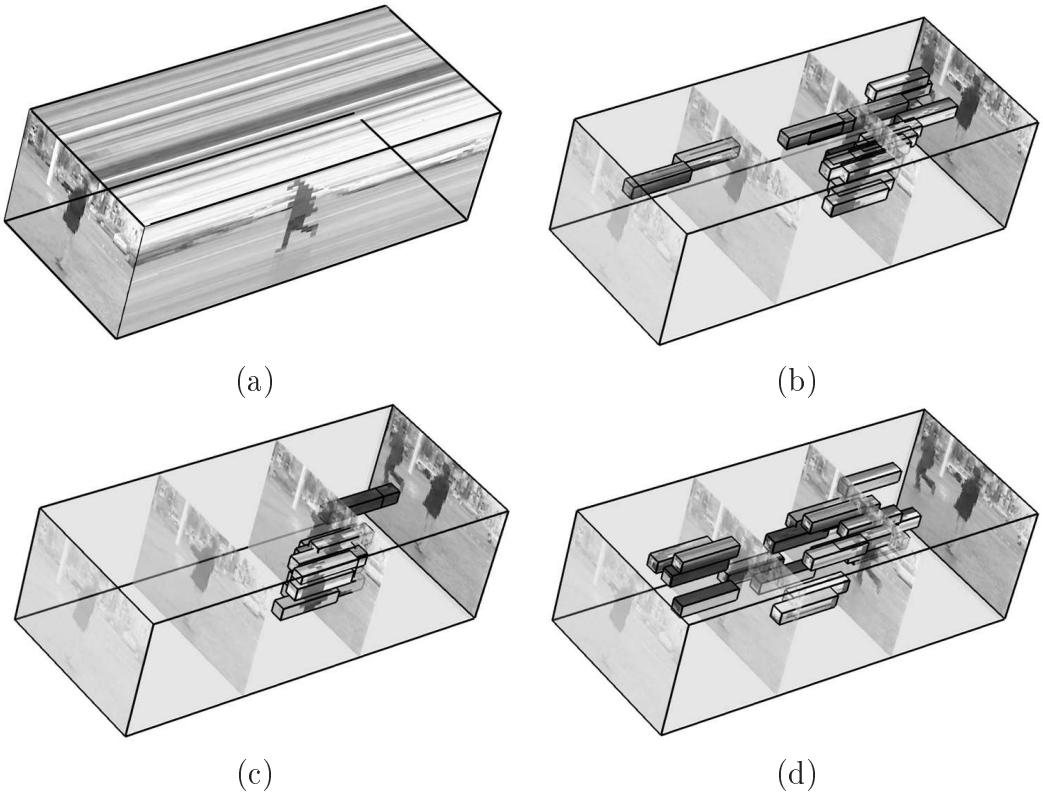
Γύρω από κάθε σημείο τοποθετείται ένας κύβος συγκεκριμένων διαστάσεων και ένας PCA-SIFT περιγραφέας χρησιμοποιείται για να τον περιγράψει. Ο περιγραφέας εξάγεται από όλους τους κύβους στο σύνολο εκμάθησης $D = d_m$ που αποτελείται από n ακολουθίες. Κάθε ακολουθία περιέχει ένα σύνολο από N_d λέξεις. Ένας κώδικας [37] $W = w_k, k \in 1, ..., K$ οπτικών λέξεων δημιουργείται με ομαδοποίηση k-means πάνω στις N_d λέξεις. Τελικά, το σύνολο εκμάθησης αναπαρίσταται από ένα σύνολο διανυσματικών περιγραφέων υπολογίζοντας τα ιστογράμματα του τύπου των κύβων που εμφανίζονται σε κάθε ακολουθία. Αυτή η περιγραφή είναι η αντίστοιχη ενός πίνακα σύγχυσης[29] c μεγέθους $K \times D$ με κάθε στοιχείο του να αντιστοιχεί στην συχνότητα εμφάνισης της λέξης w_k στην ακολουθία d_m.

Για τα σύνολα ακολουθιών KTH και εκφράσεις προσώπου χρησιμοποιούμε εκμάθηση με-μία-παράλειψη[34], όπως και στην [28]. Η εκμάθηση και η αξιολόγηση γίνεται ως εξής: η εκμάθηση γίνεται για όλες τις ακολουθίες εκτός από μία και η αξιολόγηση γίνεται σε αυτήν που έμεινε εκτός. Η διαδικασία αυτή επαναλαμβάνεται για όλες τις ακολουθίες. Οι κώδικες λέξεων δημιουργούνται με ένα πλήθος ομάδων k και η κατηγοριοποίηση γίνεται με τον 1-NN ταξινομητή.

Το σύνολο ακολουθιών σε περίπλοκα περιβάλλοντα είναι το πιο προκλητικό, καθώς

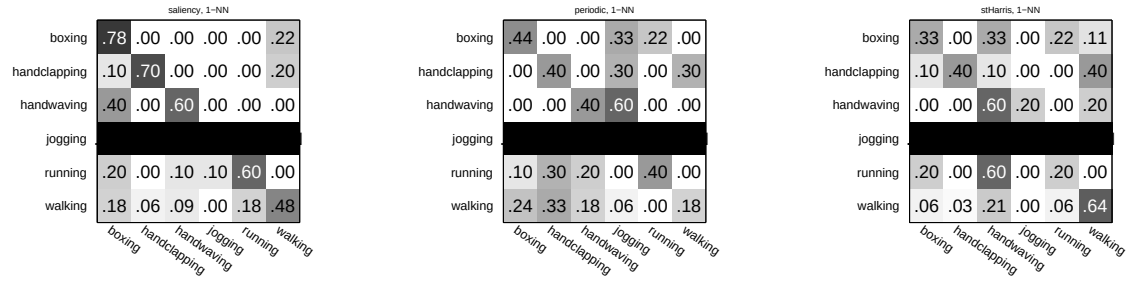


Σχήμα 6.17: *Αποτελέσματα αναγνώρισης ανθρώπινων εκφράσεων για (α) την προτεινόμενο μέθοδο, (β) την μέθοδο ανίχνευσης περιοδικότητας και (γ) την μέθοδο ανίχνευσης χωροχρονικών γωνιών (stHarris)*



Σχήμα 6.18: *Αρχικός όγκος για μία ακολουθία πυγμαχίας με σύνθετο περιβάλλον με τα 20 πιο σημαντικά σημεία ενδιαφέροντος για για (α) την προτεινόμενο μέθοδο, (β) την μέθοδο ανίχνευσης περιοδικότητας και (γ) την μέθοδο ανίχνευσης χωροχρονικών γωνιών*

προϋποθέτει επιτυχημένη ανίχνευση σημείων όσο και κατάλληλη ομαδοποίηση τους ώστε να διαχωριστούν τα οπτικά γεγονότα μεταξύ τους. Αν και η διαφοροποίηση οπτικών γεγονότων που συμβαίνουν ταυτόχρονα δεν μας απασχόλησε ερευνητικά, αξιολογούμε το προτεινόμενο μοντέλο σε αυτές τις ακολουθίες και χρησιμοποιούμε την μέθοδο pLSA ως πιο αξιόπιστη σε τέτοιες περιπτώσεις. Για να περιορίσουμε κάπως το προηγούμενο πρόβλημα διατηρούμε μόνο τα πιο σημαντικά σημεία κάθε μεθόδου, όπως αυτό προκύπτει από την τιμή ενδιαφέροντος τους. Η μέθοδος αναγνώρισης pLSA έχει χρησιμοποιηθεί στο παρελθόν για αναγνώριση χωρικών αντικειμένων [121] και χωροχρονικών γεγονότων [82]. Η κλασσική pLSA μέθοδος χρησιμοποιήθηκε αρχικά στην ανάλυση εγγράφων κειμένου. Στην περίπτωση μας, τα έγγραφα αντιστοιχούν στις ακολουθίες εικόνων και οι λέξεις στους κύβους γύρω από τα σημεία



(a) (b) (c)

Σχήμα 6.19: *Αποτελέσματα αναγνώρισης ανθρώπινων δραστηριοτήτων σε περίπλοκα περιβάλλοντα για (α) την προτεινόμενο μέθοδο, (β) την μέθοδο ανίχνευσης περιοδικότητας και (γ) την μέθοδο ανίχνευσης χωροχρονικών γωνιών*

ενδιαφέροντος. Εκτός από τον κώδικα λέξεων και τις οπτικές λέξεις, οι οποίες είναι παρατηρήσιμες και εξάγονται όπως είδαμε πριν, υπάρχει και μία μη-παρατηρήσιμη μεταβλητή z που περιγράφει τις Z θεματικές κλάσεις του συνόλου εκμάθησης και την σχέση τους με τις οπτικές λέξεις. Ο αλγόριθμος pLSA χρησιμοποιείται για την εκμάθηση της σχέσης $P(w|z)$ (οπτικές λέξεις που προκύπτουν από κάποια κλάση) και $P(z|d)$ (η κλάση στην οποία ανήκει μία ακολουθία) μέσω μεγιστοποίησης αναμονής (Expectation-Maximization) [25] πάνω στον πίνακα σύγχυσης.

Οι πίνακες σύγχυσης για την αναγνώριση σύνθετων ακολουθιών δίνονται στο Σχήμα 6.19 για τις τρεις μεθόδους που συγκρίνουμε. Συνολικά, τα στατιστικά δεν είναι ιδιαίτερα ψηλά εξαιτίας της έντονης ποικιλομορφίας του περιεχομένου και της έλλειψης κάποιου φιλτραρίσματος των σημείων που δεν σχετίζονται με κάποια από τις γνωστές δραστηριότητες. Παρόλ'αυτά τα αποτελέσματα είναι ενδεικτικά της δυνατότητας των μεθόδων να ανιχνεύουν επιτυχώς ή ανεπιτυχώς σημεία που χαρακτηρίζουν πολύ καλά την παρατηρούμενη δραστηριότητα. Το Σχήμα 6.18 απεικονίζει τα 20 πιο σημαντικά χωροχρονικά σημεία που ανίχνευσαν οι τρεις μέθοδοι σε μία σύνθετη ακολουθία. Η ακολουθία χαρακτηρίζεται ως "πυγμαχία" και περιέχει στο προσκήνιο μία δραστηριότητα πυγμαχίας και στο υπόβαθρο ένα κινούμενο αυτοκίνητο και έναν άνθρωπο να τρέχει. Ο ανιχνευτής περιοδικότητας εστιάζει στο υπόβαθρο, ενώ το προτεινόμενο μοντέλο και αυτό των Laptev *et al.* (stHarris) εντοπίζουν και σημεία πάνω στην δραστηριότητα ενδιαφέροντος διευκολύνοντας έτσι την αναγνώριση. Αυτή η παρατήρηση ισχύει για τις περισσότερες ακολουθίες του σύνθετου συνόλου. Για να είμαστε δίκαιο πρέπει να αναφέρουμε ότι τα στατιστικά της μεθόδου stHarris είναι χαμηλότερα από αυτά που δημοσιεύονται στην [61] επειδή οι μέθοδοι αναγνώρισης είναι διαφορετικές. Οι Laptev *et al.* δεν χρησιμοποιούν pLSA και -το σημαντικότερο- χρησιμοποιούν για την αναγνώριση μόνο ένα υποσύνολο σημείων που θεωρούν ως "όμοιο" με κάποια από τις δραστηριότητες που ψάχνουν [61].

6.6.3 Ανίχνευση σημαντικών γεγονότων

Η αξιολόγηση μεθόδων που σχετίζονται με οπτική αντίληψη αποτελεί πάντα πρόκληση για τους ερευνητές του χώρου, καθώς συνήθως ο ανθρώπινος χαρακτηρισμός δεν υπάρχει ή -αν υπάρχει- είναι πολύ υποκειμενικός. Για την επαλήθευση των αποτελεσμάτων χρησιμοποιούμε μία βάση με ανθρώπινο χαρακτηρισμό για σημαντικά γεγονότα [78], η οποία προέκυψε από το Δίκτυο Αριστείας MUSCLE-NoE [79]. Οι ακολουθίες

της βάσης αποτελούν υποσύνολο γνωστών ταινιών και χαρακτηρίστηκαν σύμφωνα με διάφορα κριτήρια, όπως ακουστική, οπτική και οπτικοακουστική σημαντικότητα του περιεχομένου τους. Αυτό σημαίνει ότι μέρη της ακολουθίας χαρακτηρίστηκαν ως σημαντικά, λιγότερο σημαντικά ή ασήμαντα ανάλογα με την εκτίμηση του κάθε παρατηρητή. Οι παρατηρητές κλήθηκαν να προσδιορίσουν ένα διακριτό επίπεδο ενδιαφέροντος σύμφωνα με χαλαρές οδηγίες, καθώς οι αυστηρές οδηγίες δεν μπορούν να τηρηθούν εξαιτίας της μεγάλης υποκειμενικότητας της διαδικασίας. Το αποτέλεσμα είναι χαρακτηρισμός μόνο για ακουστική, οπτική και οπτικοακουστική σημαντικότητα χωρίς όμως να λαμβάνεται η σημασιολογική ερμηνεία της ακολουθίας υπόψη. Αυτό σημαίνει ότι οι παρατηρητές κλήθηκαν να χαρακτηρίζουν την ακολουθία με βάση αυτό που βλέπουν εκείνη την στιγμή (π.χ. απότομες οπτικές αλλαγές, δυνατός θόρυβος κτλ) και όχι με βάση το σημασιολογικό περιεχόμενο (π.χ. ένα αυτοκίνητο στον δρόμο είναι πιο σημαντικό από τα άλλα επειδή το οδηγεί ο πρωταγωνιστής).

Στην πιο πρόσφατη έκδοση της η βάση αποτελείται από τρεις χαρακτηρισμένες ακολουθίες ($\simeq$ 10 min), τις "300", "Lord of the Rings I" (LOTRI) και "Cold Mountain" (CM). Εξαιτίας του μικρού τους μήκους σε σχέση με όλη την ταινία δεν μπορούν να χαρακτηριστούν αντιπροσωπευτικά του θεματικού περιεχομένου της ταινίας, αν και η επιλογή τους βασίστηκε εν μέρει και σε αυτό το κριτήριο. Η ακολουθία από την ταινία "300" περιέχει πολλές σκηνές δράσης και επομένως μπορεί να χαρακτηριστεί ως ακολουθία δράσης. Το επιλεγμένο μέρος του "Lord of the Rings I" είναι από την αρχή της ταινίας και επομένως δεν περιέχει σκηνές έντονης δράσης, όπως το υπόλοιπο της ταινίας. Αποτελείται κυρίως από σκηνές εσωτερικού χώρου με ανθρώπους να διαλέγονται και μερικές ήρεμες εξωτερικές σκηνές με αποτέλεσμα η ακολουθία να θεωρείται ατμοσφαιρική, φαντασίας με σύντομες σκηνές δράσης. Η ακολουθία από την ταινία "Cold Mountain" αποτελείται από εξωτερικές λήψεις, χωρίς δράση και μπορεί να χαρακτηριστεί ως δράμα.

6.6.3.1 Καμπύλη οπτικής προσοχής

Για να παράγουμε μία καμπύλη οπτικής προσοχής, η οποία θα κωδικοποιεί την σημασία κάθε τμήματος της ακολουθίας, ακολουθούμε παρόμοια λογική με αυτήν που είδαμε στην αναγνώριση οπτικών γεγονότων στην Ενότητα 6.6.2. Αναπαριστούμε κάθε τμήμα της ακολουθίας σαν ένα σύνολο από οπτικές λέξεις από έναν κώδικα λέξεων, ο οποίος προκύπτει με ομαδοποίηση ενός μεγάλου αριθμού περιγραφέων με χρήση του k-means.

Το προτεινόμενο μοντέλο αναλύει όλη την ακολουθία με βήμα 64 καρέ ($\simeq$ 2.5s-όλες οι ακολουθίες έχουν ρυθμό λήψης 25 καρέ/s). Μία τιμή ενδιαφέροντος προκύπτει για κάθε τμήμα ακολουθίας, ώστε να προκύψει η καμπύλη ενδιαφέροντος. Στις ταινίες η ερμηνεία της σημαντικότητας είναι ελαφρά διαφορετική από αυτήν που είδαμε μέχρι τώρα. Σε όλη την διάρκεια της ταινίας συναντώνται πολλά τυπικά/συνηθισμένα γεγονότα που εμφανίζονται συχνά και λίγα σποραδικά γεγονότα, τα οποία συχνά σχετίζονται με σημαντικές οπτικές εξελίξεις. Μοντελοποιούμε αυτού του είδους την σημαντικότητα με μία αναπαράσταση γεγονότων που θεωρεί τα σημαντικά γεγονότα ως αυτά που προεξέχουν στην καμπύλη προσοχής. Το ελάχιστο μήκος αυτών των γεγονότων είναι το μήκος του τμήματος της ακολουθίας, όπως ορίσαμε παραπάνω.

Η τιμή ενδιαφέροντος κάθε τμήματος υπολογίζεται με την Haussdorff απόσταση μεταξύ του συνόλου των προτύπων στο W και ενός συνόλου σημείων $U = u_1, ..., u_n$ που ανιχνεύονται σε κάθε τμήμα. Επομένως, η καμπύλη προσοχής A είναι ίση με την

Hausdorff απόσταση μεταξύ αυτών των δύο συνόλων και ορίζεται ως

$$A(W,U) = max(h(W,U), h(W,U)) \tag{6.30}$$

όπου

$$h(W,U) = \max_{w_k} \min_{u_n} \rho(w,u) \tag{6.31}$$

είναι η κατευθυνόμενη Hausdorff απόσταση από το W στο U. Η απόσταση ρ μεταξύ w και u είναι η Ευκλείδια και ο πίνακας αποστάσεων είναι μεγέθους $k \times n$.

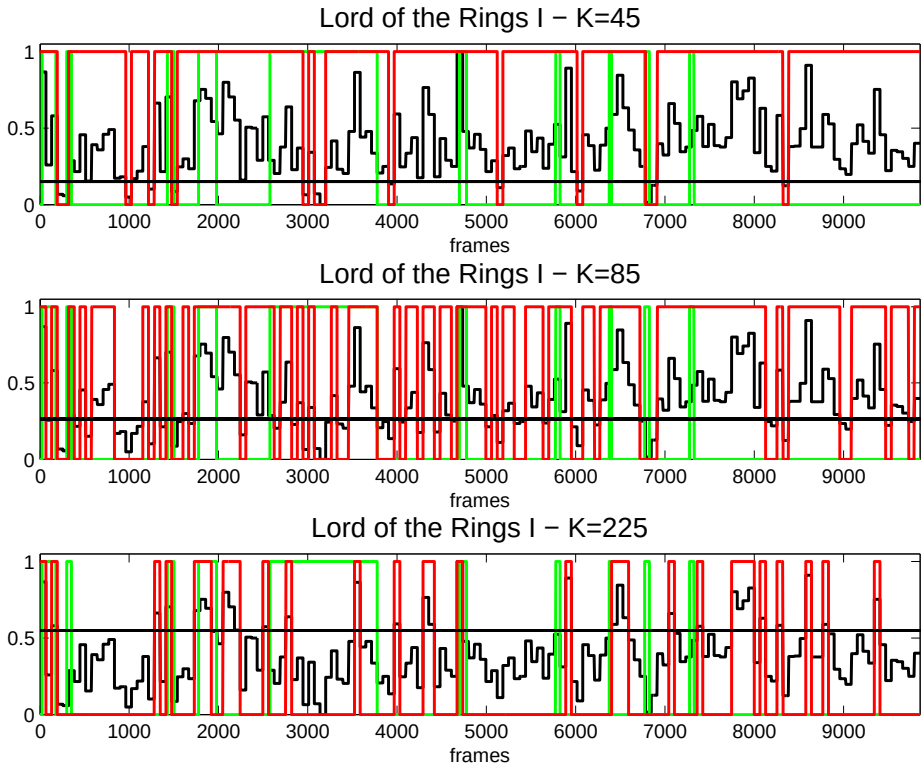
Η σύγκριση με τον υπάρχον χαρακτηρισμό δεν είναι μία προφανής διαδικασία. Κατά την διάρκεια του χαρακτηρισμού των ακολουθιών το ανθρώπινο οπτικό σύστημα είναι ικανό να εντοπίζει σχεδόν αυτόματα την σημαντική πληροφορία σε ολόκληρα τμήματα της ακολουθίας. Το αποτέλεσμα του χαρακτηρισμού είναι ένας ενδείκτης ενδιαφέροντος[38] I_{sal}. Διατυπώνουμε το πρόβλημα σύγκρισης μεταξύ της καμπύλης προσοχής και του ενδείκτη ενδιαφέροντος ως πρόβλημα συντονισμού δύο παραμέτρων, ειδικότερα τον αριθμό των οπτικών λέξεων K και το κατώφλι T, το οποίο χρησιμοποιείται για να την δυαδική απόφαση σημαντικό ή μη-σημαντικό για κάθε σημείο της καμπύλης. Τα πειράματα αποκάλυψαν μία λογική σχέση μεταξύ των δύο παραμέτρων. Ο αυξανόμενος αριθμός οπτικών λέξεων περιγράφει όλο και καλύτερα την υποκείμενη δομή της ακολουθίας και επομένως παράγει μία πιο οξεία (peaky) καμπύλη (τα σημαντικά γεγονότα γίνονται όλο και πιο διαφορετικά και σποραδικά από το "υπόβαθρο"), ενώ το αντίθετο θα παράγει μία πιο ομαλή καμπύλη. Μία οξεία καμπύλη πρέπει να συνοδεύεται από ένα υψηλό κατώφλι, ενώ μία ομαλή με χαμηλότερο. Για να απλοποιήσουμε την απόφαση συσχετίζουμε αυτές τις παραμέτρους με μία εκθετική συνάρτηση

$$T(K) = (1/B)\exp(-K/B) \tag{6.32}$$

όπου B είναι η κλίμακα. Για τα πειράματα θέτουμε $B = 0.5$ και κανονικοποιούμε την $T(K)$ ώστε $T(0) = 1$.

Η σχέση μεταξύ των δύο παραμέτρων φαίνεται στο Σχήμα 6.20, το οποίο απεικονίζει τρία διαφορετικά αποτελέσματα για μία ταινία που αντιστοιχούν σε τρεις διαφορετικές τιμές του K. Η καμπύλη προσοχής απεικονίζεται με μαύρο χρώμα, ο χαρακτηρισμός με πράσινο, τα τμήματα που ανιχνεύονται ως σημαντικά με κόκκινο και η τιμή του κατωφλίου ως μαύρη οριζόντια γραμμή. Από τα σχήματα γίνεται φανερή η συμπεριφορά που περιγράψαμε πριν: όσο αυξάνει ο αριθμός των οπτικών λέξεων τόσο πιο σποραδικά γίνονται τα σημαντικά γεγονότα που ανιχνεύει το προτεινόμενο μοντέλο. Ο προσδιορισμός του κατάλληλου μεγέθους των σημαντικών τμημάτων της ακολουθίας δεν είναι προφανής. Σε προηγούμενη δουλειά μας, και κάτω από ένα πολύ διαφορετικό πλαίσιο, εφαρμόσαμε ένα φίλτρο ενδιάμεσης τιμής μεταβλητού μήκους και κρατήσαμε τελικά αυτό που έδινε τα καλύτερα στατιστικά [98]. Σε αυτό το κεφάλαιο επιλέγουμε να μην εισάγουμε και άλλη παράμετρο στην μέθοδο και χρησιμοποιούμε απευθείας την έξοδο της κατωφλίωσης σαν το μήκος των σημαντικών γεγονότων.

Το Σχήμα 6.21 δείχνει τις καμπύλες ακρίβειας και επανάκλησης για τις τρεις ταινίες της βάσης. Οι καμπύλες υπολογίστηκαν με μεταβλητό αριθμό οπτικών λέξεων K που αποτελούν τον κώδικα λέξεων W. Όπως αναμένεται, οι τιμές recall είναι υψηλότερες στο μεγαλύτερο εύρος, καθώς υψηλή ακρίβεια σε μία τόσο υποκειμενική εφαρμογή είναι δύσκολο να επιτευχθεί. Παρόλ'αυτά όλες οι μέθοδοι πετυχαίνουν ένα ικανοποιητικό επίπεδο ακρίβειας για όλες τις ταινίες, με την προτεινόμενη τεχνική να έχει εμφανώς υψηλότερες τιμές για δύο από τις τρεις ταινίες,



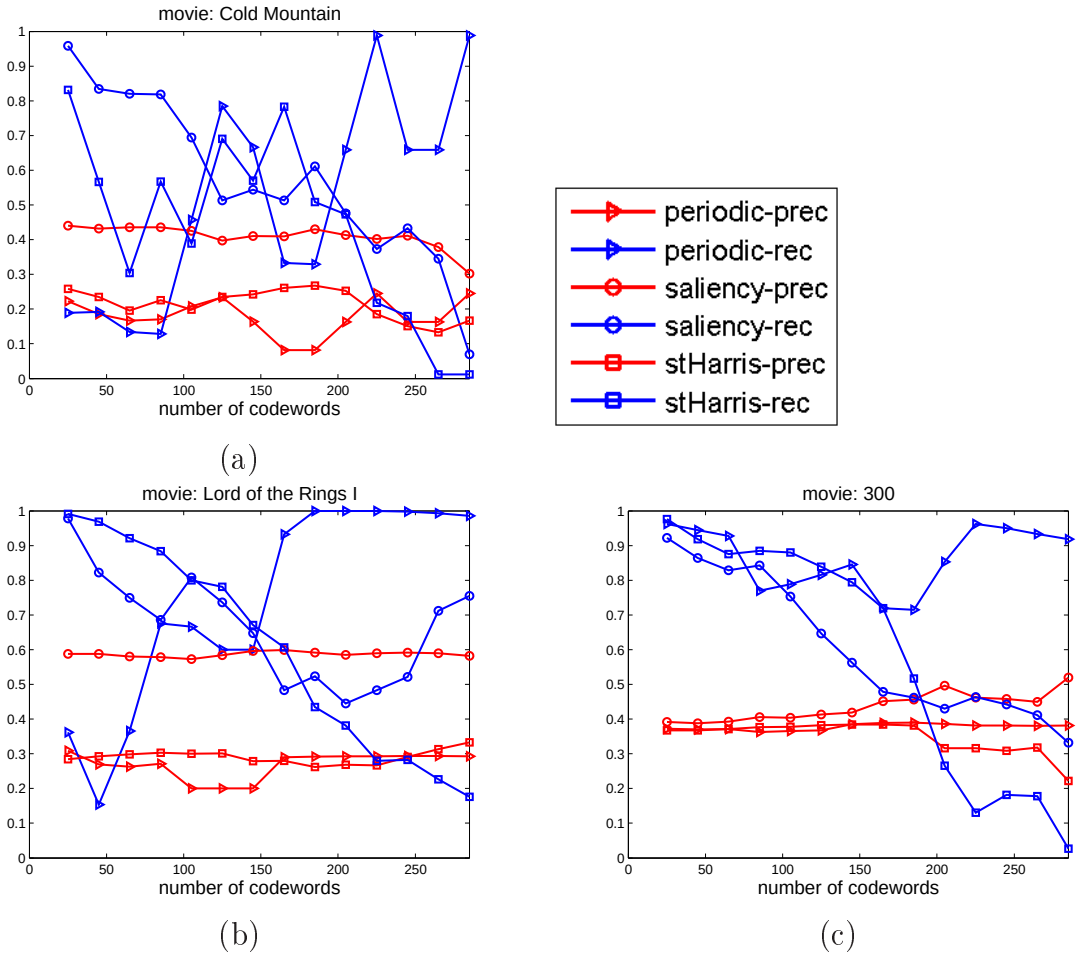
Σχήμα 6.20: *Ανίχνευση σημαντικών τμημάτων ακολουθίας με την προτεινόμενη μέθοδο. Κάθε διάγραμμα απεικονίζει τον ανθρώπινο χαρακτηρισμό (πράσινο), την καμπύλη προσοχής (μάυρο), τα επιλεγμένα σημαντικά τμήματα (κόκκινα) και το κατώφλι(οριζόντια μαύρη γραμμή)*

όπως φαίνεται από τα σχεδιαγράμματα. Παρουσιάζει ιδιαίτερο ενδιαφέρον το ότι για την ταινία "300" όλες οι μέθοδοι έχουν παρόμοια αποτελέσματα. Φαίνεται ότι το περιεχόμενο δράσης και οι ιδιομορφίες των ψηφιακών σκηνικών (έντονα χρώματα, χαρακτήρες που μοιάζουν με σκίτσα, εξωπραγματικές σκηνές δράσης κτλ) δίνουν αφορμή για παρόμοια χωροχρονικά σημεία για όλες τις τεχνικές. Για παράδειγμα, οι σκηνές δράσης στο πεδίο μάχης, όπως αυτές αποτυπώνονται στην ταινία, περιέχουν εξέχοντα γεγονότα εξαιτίας χρώματος/φωτεινότητας/κίνησης (προτεινόμενος ανιχνευτής), περιοδικές κινήσεις (ανιχνευτής περιοδικότητας) και απότομες αλλαγές στην κατεύθυνση κίνησης, οι οποίες αντιστοιχούν σε χωροχρονικές γωνίες (ανιχνευτής stHarris).

6.7 Συμπεράσματα

Σε αυτό το κεφάλαιο προτείναμε και υλοποιήσαμε δύο μοντέλα υπολογισμού τιμών εδιαφέροντος για ακολουθίες, τα οποία βασίζονται σε ανταγωνισμό χαρακτηριστικών. Ο ανταγωνισμός μοντελοποιείται με μία ενέργεια που αποτελείται από τον όρο παρατήρησης και τον όρο εξομάλυνσης. Τα δύο μοντέλα μοιράζονται κοινές ιδέες, αλλά χρησιμοποιούν διαφορετικούς περιορισμούς στην ελαχιστοποίηση της προηγούμενης ενέργειας.

Όπως είδαμε και στα προηγούμενα κεφάλαια, η επεξεργασία και ανάλυση ακολου-



Σχήμα 6.21: *Καμπύλες ακρίβειας (κόκκινο) και επανάκλησης (μπλε) για τις ακολουθίες (α) "300", (β) "Lord of the Rings I" και (γ) "Cold Mountain"*

θιών με βάση την σημαντικότητα συνεισφέρει σημαντικά στην επίλυση προβλημάτων στον χώρο της μηχανικής όρασης. Σε αυτό το κεφάλαιο μελετούμε εκτενέστερα το πρόβλημα αναγνώριση σκηνής με χρήση τιμών ενδιαφέροντος και παρουσιάζουμε αποτελέσματα σε αναγνώριση οπτικών δραστηριοτήτων και δημιουργία αντιληπτικών περιλήψεων ακολουθιών.

Στην αναγνώριση σκηνής παρουσιάζουμε εκτενή σύγκριση ανάμεσα στα τρία μοντέλα που έχουμε προτείνει συνολικά σε αυτήν τη διατριβή και σε τρεις καθιερωμένες μεθόδους οπτικής προσοχής που είδαμε στο Κεφάλαιο 2. Παρουσιάζουμε τρεις πειραματικές μεθοδολογίες που εξετάζουν διαφορετικές παραμέτρους της αναγνώρισης και βασίζονται στην κατάτμηση της αρχικής ακολουθίας σε σημαντικές και μη-σημαντικές περιοχές βάσει των τιμών ενδιαφέροντος που προκύπτουν από κάθε μέθοδο. Η πρώτη και η δεύτερη εξετάζουν την βελτίωση του σφάλματος όταν χρησιμοποιείται μέρος της σκηνής, ενώ το τρίτο μελετά την επίπτωση του μεγέθους του συνόλου εκμάθησης στο συνολικό σφάλμα.

Το δεύτερο μοντέλο που προτείνουμε το εφαρμόζουμε σε αναγνώριση χωροχρονικών δραστηριοτήτων και δημιουργία αντιληπτικών περιλήψεων συγκρίνοντας με δύο καθιερωμένες μεθόδους του χώρου.

□

Κεφάλαιο 7

Επίλογος

7.1 Συμπεράσματα

Σε αυτήν τη διατριβή προτείναμε μοντέλα και μεθόδους για υπολογισμό τιμών ενδιαφέροντος για εικόνες και ακολουθίες. Με κίνητρο την σημασία της λειτουργίας οπτικής προσοχής για τον άνθρωπο επιχειρήσαμε να μελετήσουμε και να προτείνουμε υπολογιστικά αποδοτικούς τρόπους για την μεταφορά ανάλογων λειτουργιών στην μηχανική όραση. Αρχικά, οι προσπάθειες άλλων ερευνητών ήταν περιορισμένες, αλλά αρκετά κατατοπιστικές ώστε να να αναδείξουν την αξία τέτοιων τεχνικών σε εφαρμογές επεξεργασίας και ανάλυσης. Η έρευνα μας σε χωρικά μοντέλα οπτικής προσοχής βασίστηκε κυρίως σε καθιερωμένα υπολογιστικά μοντέλα του χώρου. Το μοντέλο του κύριου χάρτη των Itti *et al.* [48] και η θεωρία των Koch και Ullman [60] στην οποία βασίστηκε ήταν καθοριστικές στην ανάπτυξη των προτεινόμενων χωρικών μοντέλων, αλλά και στις μεταγενέστερες εξελίξεις τους σε χωροχρονικά. Η συνολική ερευνητική δουλειά γύρω από την θεωρία του επιλεκτικού συντονισμού [137], όπως αυτή περιγράφεται σε μία σειρά δημοσιεύσεων, και η διεισδυτική ματιά των συγγραφέων σε θέματα οπτικής προσοχής συνεισέφεραν επίσης σημαντικά στις καινοτομίες της διατριβής.

Η διατριβή στο σύνολο της αποτελεί μία εμπεριστατωμένη μελέτη της θεωρίας και εφαρμογής μοντέλων οπτικής προσοχής. Αν και η έρευνα μας ξεκίνησε από τον χώρο της εικόνας (Κεφάλαια 2, 3), ήταν από νωρίς φανερό το ενδιαφέρον μας για την χωροχρονική ανάλυση ακολουθιών (Κεφάλαια 4, 5, 6), η οποία είχε μελετηθεί λιγότερο στην σχετική βιβλιογραφία. Περιγράψαμε αναλυτικά τις συνεισφορές μας στον χώρο τόσο στην μορφή επεκτάσεων αλλά και καινοτομιών και ασχοληθήκαμε κυρίως με εφαρμογές αποθορυβοποίησης εικόνων, κωδικοποίησης ακολουθιών, αναγνώριση σκηνής, επίβλεψης σκηνής, ανίχνευσης οπτικών δραστηριοτήτων και περίληψης ακολουθιών. Κάθε κεφάλαιο εισάγει τις ιδέες μας, παρουσιάζει λεπτομερώς τα μοντέλα που αναπτύξαμε και τελικά τα αξιολογεί ποσοτικά και ποιοτικά. Συμπεράσματα και μελλοντικές επεκτάσεις δίνονται στο τέλος κάθε κεφαλαίου. Η Ενότητα 7.3 συνοψίζει τις συνεισφορές μας στον χώρο.

Στο ξεκίνημα της διατριβής, το πεδίο υπολογιστικών μοντέλων οπτικής προσοχής φαίνονταν περιορισμένο, αλλά πολύ ενδιαφέρον λόγω των συσχετίσεων με τον ανθρώπινο τρόπο όρασης. Στην συνέχεια και αφού αναπτύξαμε τα πρώτα μοντέλα οπτικής προσοχής αντιμετωπίσαμε σοβαρές δυσκολίες στην αξιολόγηση τους εξαιτίας της έντονης υποκειμενικότητας των σχετικών εννοιών (ενδιαφέρον, σημαντικότητα κτλ). Στην πορεία όμως τόσο με την προσωπική μας έρευνα όσο και με την παράλληλη

έρευνα άλλων στο χώρο άρχισαν να φαίνονται τα οφέλη, να αυξάνεται το ενδιαφέρον για θεωρίες οπτικής προσοχής και τελικά να μεγαλώνει το πεδίο εφαρμογών. Το αποτέλεσμα ήταν να προκύψουν νέες ιδέες για υπολογιστικά μοντέλα και εφαρμογή τους σε πλήθος τομέων. Λογικό είναι επομένως και η παρούσα διατριβή να αποτελεί έμπνευση για μελλοντική ερευνητική δουλειά στον χώρο. Συγκεκριμένα, συνδέσαμε την έννοια της σημαντικότητας ή υπολογισμού ενδιαφέροντος με πολλούς τομείς της μηχανικής όρασης και ιδιαίτερα με τομείς σχετικούς με χωροχρονική ανάλυση ακολουθιών.

7.2 Μελλοντικές επεκτάσεις

Στην διατριβή συμπεριλάβαμε μόνο πειράματα με στατιστική αξία, δηλαδή πειράματα τα οποία έγιναν σε σχετικά μεγάλο όγκο δεδομένων. Εξετάστηκε όμως και η εφαρμογή των προτεινόμενων μοντέλων είτε επιφανειακά είτε σε μεγαλύτερο βάθος σε άλλους τομείς, όπως η κατάτμηση εικόνων και ακολουθιών, ο διαχωρισμός παρασκηνίου και υποβάθρου σε ακολουθίες και η επίδραση τιμών ενδιαφέροντος σε ενεργά περιγράμματα για την βελτιωμένη χρονική ανίχνευση αντικειμένων. Χωρίς να περιορίσουμε τις μελλοντικές εφαρμογές και επεκτάσεις των μεθόδων στις τρεις που αναφέραμε, αλλά με στόχο να γίνουμε πιο συγκεκριμένοι αναφέρουμε πιθανές μελλοντικές συνεισφορές μόνο σε αυτές:

- **Κατάτμηση εικόνων/ακολουθιών:** Μια απλοποιημένη, αλλά έγκυρη, ερμηνεία των μοντέλων οπτικής προσοχής είναι ότι αποσκοπούν στην ανίχνευση περιοχών, οι οποίες τραβούν το ενδιαφέρον του παρατηρητή και ταυτόχρονα είναι σχεδόν συνεκτικές όσον αφορά στα χαρακτηριστικά που χρησιμοποιούν. Αρχίζει να γίνεται προφανής επομένως η πιθανή συσχέτιση τους με τεχνικές τμηματοποίησης, οι οποίες ιδανικά παράγουν συνεκτικές περιοχές με βάση τα χαρακτηριστικά που υπολογίζουν. Θεωρούμε ότι ακόμη και κλασσικοί αλγόριθμοι του χώρου, όπως ο Watershed, θα επωφελούνταν π.χ. από ενδεικτικά όρια περιοχών που πιθανότατα συμπίπτουν με τα πραγματικά όρια των αντικειμένων ενδιαφέροντος. Ακόμη και απλές προσεγγίσεις, όπως π.χ. η κατάτμηση του χωρικού κύριου χάρτη στην Ενότητα 5.3.3.1 και η κατάτμηση του αντίστοιχου χωροχρονικού στην Ενότητα 6.6.1.2, είναι ενδεικτικές της αξίας της οπτικής προσοχής σε αυτές τις εφαρμογές.

- **Διαχωρισμός παρασκηνίου/υποβάθρου:** Ένα πλήθος τεχνικών αυτής της περιοχής βασίζεται στην μοντελοποίηση της σκηνής με κατανομές χρώματος/φωτεινότητας (π.χ. Gaussian Mixture Models). Σε ακολουθίες, ο στόχος είναι να προβλέπεται σε κάθε στιγμή σε ποια κατανομή ανήκουν τα εικονοστοιχεία του καρέ, ώστε να χαρακτηρίζονται ως μέρος του υποβάθρου ή παρασκηνίου. Οι τιμές ενδιαφέροντος συχνά σχετίζονται με κινούμενα αντικείμενα του παρασκηνίου ή με αντικείμενα του υποβάθρου που ξεχωρίζουν και επομένως για ένα μικρό χρονικό διάστημα μπορούν να χαρακτηριστούν μέρος του παρασκηνίου. Οι μηχανισμοί των μεθόδων που είναι υπεύθυνοι για την ενημέρωση της κατάστασης κάθε εικονοστοιχείου μπορούν να συμπεριλάβουν και την επίδραση του χάρτη ενδιαφέροντος για να παράγουν πιο "σωστές" - σε σχέση με την ανθρώπινη αντίληψη- περιοχές. Πειράματα μικρής κλίμακας με επέκταση της τεχνικής των Stauffer και Grimson [124] ενίσχυσαν τον προηγούμενο ισχυρισμό μας.

- **Εισαγωγή 3Δ μετασχηματισμό κυματιδίων στο μοντέλο ανταγωνισμού και επεκτάσεις:** Στο Κεφάλαιο 4 προτείναμε το πρώτο χωροχρονικό μοντέλο υπολογισμού ενδιαφέροντος και το εφαρμόσαμε σε αναγνώριση ανθρώπινων δραστηριοτήτων. Η αφορμή ήταν η χαμηλή υπολογιστική πολυπλοκότητα και οι εύκολα κατανοητές ιδιότητες των ζωνών του μετασχηματισμού (συσχέτιση με συγκεκριμένες ιδιότητες της κίνησης στην ακολουθία). Η πρώτη προφανής επέκταση επομένως είναι η εισαγωγή αυτού του πλαισίου στο μοντέλο ανταγωνισμού που περιγράψαμε στο Κεφάλαιο 6 και συγκεκριμένα προς αντικατάσταση του δομοστοιχείου κατευθυντικότητας. Συνολικά όμως ο 3Δ μετασχηματισμός μπορεί να χρησιμοποιηθεί τόσο για τον υπολογισμό του ενδιάμεσου όγκου κατευθυντικότητας όσο και για την αξιόπιστη περιγραφή περιοχών ενδιαφέροντος. Το ερευνητικό ενδιαφέρον βρίσκεται στην κατάλληλη επιλογή και χρήση του μετασχηματισμού, ώστε να αποφευχθούν κυρίως προβλήματα λόγω μεταβλητότητας σε μετατόπιση. Προτάσεις για χρήση απλών μιγαδικών η μιγαδικών Διπλού Δένδρου μετασχηματισμών κυματιδίων πρέπει να μελετηθούν και να αξιολογηθούν με βάση την αμεταβλητότητα σε μετατόπιση αλλά και την συνεπαγόμενη υπολογιστική πολυπλοκότητα [58] [57].

- **Ενεργά περιγράμματα:** Στην αρχή της έρευνας μας ασχοληθήκαμε με την εισαγωγή πληροφορίας οπτικής ροής σε ένα υπάρχον πλαίσιο υπολογισμού ενεργών περιγραμμάτων [141] [142]. Αν και η συγκεκριμένη δουλειά αποτέλεσε κίνητρο για την μελέτη του χώρου της οπτικής προσοχής, όπως αναφέραμε στην Ενότητα 1.1, δεν συνεχίστηκε εξαιτίας της εκτενούς ενασχόλησης μας με τα υπολογιστικά μοντέλα υπολογισμού ενδιαφέροντος. Θεωρούμε ότι τα προτεινόμενα μοντέλα μπορούν να συνδυαστούν με σύγχρονες πλέον τεχνικές ενεργών περιγραμμάτων ώστε να βελτιώσουν την παρακολούθηση αντικειμένων ή την ακριβέστερη κατάτμηση τους. Συγκεκριμένα, στις προαναφερθείσες εργασίες προτείναμε έναν τελεστή εμπιστοσύνης που βασίζεται στην ιστορία της κίνησης των αντικειμένων (με χειρωνακτική αρχικοποίηση του περιγράμματος). Με αντίστοιχο τρόπο θα μπορούσε να αξιοποιηθεί η σημαντικότητα τόσο στην εξέλιξη του περιγράμματος όσο και στην αυτόματη αρχικοποίηση του.

- **Εφαρμογές:** Οι μελλοντικές εφαρμογές των μοντέλων οπτικής προσοχής που αναπτύχθηκαν στην διατριβή είναι πολλές. Ενδεικτικά αναφέρουμε την γενικευμένη ανίχνευση οπτικών γεγονότων, την περίληψη ακολουθιών με βάση την ανθρώπινη αντίληψη και την κωδικοποίηση, οι οποίες μας απασχόλησαν εν μέρει και στην πορεία της έρευνας μας. Συγκεκριμένα, η ανίχνευση δραστηριοτήτων με έντονη ποικιλομορφία εμφάνισης, όπως “ανοίγω πόρτα”, “αφήνω τσάντα”, “πίνω”, “κάθομαι” κτλ, αποτελεί ερευνητικό στόχων πρόσφατων μεθόδων. Τα προτεινόμενα χωροχρονικά μοντέλα οπτικής προσοχής μπορούν να δώσουν λύση σε παρόμοια προβλήματα. Η περίληψη ακολουθιών εικόνων μπορεί να προκύψει από την μεθοδολογία που αναπτύξαμε στο Κεφάλαιο 6 για ανίχνευση σημαντικών γεγονότων. Τα σημαντικά τμήματα μίας ακολουθίας εικόνων μπορούν να συνδυαστούν για να σχηματίσουν μια περίληψη, η οποία θα είναι κοντά στην ανθρώπινη αντίληψη. Στο Κεφάλαιο 3 παρουσιάσαμε πειράματα για κωδικοποίηση με περιοχές ενδιαφέροντος που προέκυψαν από τα χωρικά μοντέλα που προτείναμε (ανά καρέ). Η αξιοποίηση των χωροχρονικών μοντέλων οπτικής προσοχής σε κωδικοποίηση ίσως ανοίξει τον δρόμο για πιο αποδοτική και ποιοτική κωδικοποίηση, καθώς θα είναι διαθέσιμη η χωροχρονικά συνεκτική

περιοχή ενδιαφέροντος.

Γενικότερα, θεωρούμε ότι οποιαδήποτε μέθοδος απαιτεί τιμές εμπιστοσύνης ή χωρικούς/χωροχρονικούς περιορισμούς για την οπτική πληροφορία μπορεί να επωφεληθεί από τις τεχνικές που περιγράφουμε στην διατριβή.

7.3 Συνεισφορές και δημοσιεύσεις

Στην διάρκεια της έρευνας μας εστιάσαμε στην επέκταση καθιερωμένων χωρικών μοντέλων οπτικής προσοχής και προτείναμε νέα χωροχρονικά μοντέλα υπολογισμού ενδιαφέροντος για ακολουθίες. Περιγράψαμε αναλυτικά τις προτάσεις μας και αξιολογήσαμε διεξοδικά τις νέες μεθόδους σε πλήθος εφαρμογών. Συνοπτικά, οι κύριες συνεισφορές μας ήταν οι εξής:

- Στο Κεφάλαιο 3 προτείναμε *επεκτάσεις καθιερωμένων μοντέλων οπτικής προσοχής*. Πιο συγκεκριμένα προτείναμε μία απλοποιημένη έκδοση του μοντέλου επιλεκτικού συντονισμού των Tsotsos *et al.* [137], η οποία βασίζεται στην χρήση δικτύων όλα-για-τον-νικητή και χρησιμοποιείται για την ανίχνευση περιοχών ενδιαφέροντος. Εφαρμόσαμε το μοντέλο σε αποθορυβοποίηση εικόνων. Ο συνδυασμός υπολογισμού ενδιαφέροντος με τεχνικές αποθορυβοποίησης αποτελεί συνεισφορά της έρευνας μας. Στη συνέχεια επεκτείναμε το βασικό μοντέλο των Itti *et al.* [48] με δύο τρόπους: α) με χρήση περισσότερων χαρακτηριστικών και β) με χρήση ενός καθοδικού (top-down) καναλιού επεξεργασίας. Εφαρμόσαμε το δεύτερο μοντέλο σε κωδικοποίηση ακολουθιών συγκρίνοντας το με άλλες μεθόδους. Το πρώτο μέρος αυτού του κεφαλαίου δημοσιεύτηκε στο συνέδριο ICIP'07 [92]. Το δεύτερο μέρος του κεφαλαίου, καθώς και βελτιώσεις/εξελίξεις που προέκυψαν σε όλη την διάρκεια της έρευνας μέσα από ερευνητικές συνεργασίες, δημοσιεύτηκαν στα συνέδρια ICIP'05 [101], ICIP'07 [138], MobiMedia'06 [139] και στο περιοδικό International Journal of Neural Systems [140].

- Στο Κεφάλαιο 4 μελετήσαμε τη χρήση του *3Δ μετασχηματισμού κυματιδίων για υπολογισμό του όγκου ενδιαφέροντος*. Προτείναμε τον δια-ζωνικό και ενδο-ζωνικό συνδυασμό συντελεστών του μετασχηματισμού και αξιολογήσαμε την μέθοδο σε ανίχνευση οπτικών δραστηριοτήτων. Η ογκομετρική αποσύνθεση της ακολουθίας και η μελέτη του 3Δ μετασχηματισμού κυματιδίων για εφαρμογές ανάλυσης κινούμενων εικόνων αποτελεί συνεισφορά της διατριβής. Μέρος της έρευνας που εκπονήθηκε στα πλαίσια αυτού του κεφαλαίου δημοσιεύτηκε στο συνέδριο CIVR [93].

- Στο Κεφάλαιο 5 προτείναμε ένα *ολοκληρωμένο μοντέλο υπολογισμού τιμών ενδιαφέροντος*, το οποίο βασίζεται σε ογκομετρική αναπαράσταση της εισόδου και σε πλήθος χαρακτηριστικών και το εφαρμόσαμε σε ανίχνευση σημαντικών περιοχών και σε αναγνώριση σκηνής. Οι κεντρικές ιδέες για την λειτουργία του μοντέλου προήλθαν από το μοντέλο κύριου χάρτη των Itti *et al.* [48], αλλά το συνολικό ογκομετρικό πλαίσιο, η επέκταση των 2Δ τελεστών σε 3Δ, καθώς και η εφαρμογή τεχνικών οπτικής προσοχής σε αναγνώριση σκηνής αποτελούν συνεισφορά της διατριβής. Το μοντέλο και τα αποτελέσματα αυτού του κεφαλαίου δημοσιεύτηκαν στα συνέδρια MMSP'04 [103], CBMI'05 [91],

VLBV'05 [95], FUZZ-IEEE'05 [96] και στο περιοδικό IET (πρώην IEE) Image Processing [104].

- Στο Κεφάλαιο 6 εισήχθηκε η έννοια του *ανταγωνισμού των χαρακτηριστικών για τον υπολογισμό τιμών ενδιαφέροντος* της εισόδου. Η κεντρική ιδέα προέκυψε από την εισαγωγική μελέτη μας στον χώρο των βιολογικών μοντέλων οπτικής προσοχής και συγκεκριμένα από τον ανταγωνισμό οπτικών μονοπατιών στο ανθρώπινο οπτικό σύστημα, ο οποίος έχει ανιχνευθεί πειραματικά, και από την θεωρία Μορφής (Gestalt). Η μοντελοποίηση υπολογισμού ενδιαφέροντος σε ακολουθίες με χρήση ανταγωνισμού αποτελεί συνεισφορά της διατριβής. Σε αυτό το κεφάλαιο παρουσιάσαμε συνολική σύγκριση όλων των προτεινόμενων μοντέλων σε εφαρμογή αναγνώρισης σκηνής. Επιπροσθέτως εφαρμόσαμε το μοντέλο ανταγωνισμού σε ανίχνευση οπτικών δραστηριοτήτων και σημαντικών γεονότων. Η θεωρία και τα αποτελέσματα αυτού του κεφαλαίου έχουν σταλεί σε δύο περιοδικά και είναι υπό-έγκριση. Πρόκειται για τα περιοδικά Signal Processing: Image Communication [105] (υπό αναθεώρηση) και IEEE Transactions on Pattern Analysis and Machine Intelligence [94]. Μέρος των εφαρμογών έχει δημοσιευτεί στο συνέδριο MMSP'07 [98], σε κεφάλαιο του βιβλίου "Multimodal Processing and Interaction" [31] και έχει σταλεί στο συνέδριο ICIP'08 [32].

□

Κεφάλαιο 8

Κατάλογος δημοσιεύσεων

8.1 Περιοδικά

8.1.1 Δημοσιευμένα

1. K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis and S. Kollias, "A Bottom-Up Spatiotemporal Visual Attention Model for Video Analysis", IET Image Processing, vol. 1, no. 2, pp. 237- 248, Jun 2007.

2. N. Tsapatsoulis, K. Rapantzikos, C. Pattichis, "An Embedded Saliency Map Estimator Scheme: Application to Video Encoding", International Journal of Neural Systems, vol. 17, No. 4, pp. 1-16, Aug 2007.

3. K. Rapantzikos, N. Tsapatsoulis, "A committee machine scheme for feature map fusion under uncertainty: the face detection case", IJISTA, special Issue on "Intelligent Image and Video Processing and Applications: The Role of Uncertainty", vol. 1, no. 3, pp. 346-358, 2006.

4. G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis and S. Kollias, "A Snake Model for Object Tracking in Natural Sequences", Elsevier, Signal Processing: Image Communication, vol. 19, no. 3, pp. 219-238, Mar 2004.

5. G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis, S. Kollias, "Rule-driven Object Tracking in Clutter and Partial Occlusion with Model-based Snakes", EURASIP Journal on Applied Signal Processing, vol. 2004, no. 6, pp. 841-860, Jun 2004.

6. G. Evangelopoulos, K. Rapantzikos, P. Maragos, Y. Avrithis, "Audio-Visual Attention Modeling and Salient Event Detection", in Multimodal Processing and Interaction, P. Maragos, A. Potamianos, Eds., Springer, to be published.

8.1.2 Υποβεβλημένα προς κρίση

7. K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis and S. Kollias, "Spatiotemporal Visual Attention for Video classification", Signal Processing: Image Communication, *under revision.*

8. K. Rapantzikos, Y. Avrithis and S. Kollias, "Volumetric saliency for event detection and activity representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, *submitted.*

8.2 Συνέδρια

8.2.1 Δημοσιευμένα

9. P. Kapsalas, K. Rapantzikos, A. Sofou, Y. Avrithis, "Regions Of Interest for Object Detection", 6th Int'l Workshop on Content-Based Multimedia Indexing (CBMI), Jun 18-20, London, UK, *accepted.*

10. K. Rapantzikos, G. Evangelopoulos, P. Maragos, Y. Avrithis, "An Audio-Visual Saliency Model for Movie Summarization", Proc. IEEE Int'l Workshop on Multimedia Signal Processing (MMSP07), Chania, Greece, Oct. 2007.

11. K. Rapantzikos, Y. Avrithis, S. Kollias, "SALIENShrink: Saliency-based image denoising", ICIP'07, vol. 1, pp. 305-308, San Antonio, Sep 2007.

12. K. Rapantzikos, Y.Avrithis, S. Kollias, "Spatiotemporal saliency for event detection and representation in the 3D Wavelet Domain: Potential in human action recognition", Proc. ACM Int'l Conference on Image and Video Retrieval (CIVR), pp. 294 - 301 , Jul 2007.

13. N. Tsapatsoulis, C. Pattichis, K. Rapantzikos, "Biologically inspired region of interest selection for low bit-rate video coding", Proc. of the IEEE Int'l Conference on Image Processing (ICIP'07), vol. 3, pp.333- 336, San Antonio, Sep 2007.

14. N. Tsapatsoulis, K. Rapantzikos and Y. Avrithis, "Priority Coding for Video-telephony Applications based on Visual Attention", 2nd International Mobile Multimedia Communications Conference (MobiMedia 2006), Alghero, Italy, September 18-20, 2006.

15. K. Rapantzikos, N. Tsapatsoulis, "Enhancing the robustness of skin-based face detection schemes through a visual attention architecture", Proc. of the IEEE Int'l Conference on Image Processing (ICIP), Genova, Italy, Sep 2005.

16. K. Rapantzikos, Y. Avrithis, "An enhanced spatiotemporal visual attention model for sports video analysis", Intern. Workshop on content-based Multimedia indexing (CBMI), Riga, Latvia, Jun 2005.

17. K. Rapantzikos, Y. Avrithis, S. Kollias, "Handling uncertainty in video analysis with spatiotemporal visual attention", Proc. of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '05), Reno, Nevada, May 22-25, 2005.

18. K. Rapantzikos, Y. Avrithis, S. Kollias, "On the use of spatiotemporal visual attention for video classification", Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV '05), Sardinia, Italy, September 15-16, 2005.

19. K. Rapantzikos, N. Tsapatsoulis, and Y. Avrithis, "Spatiotemporal Visual Attention Architecture for Video Analysis", Proc. of IEEE International Workshop On Multimedia Signal Processing (MMSP'04), pp. 83-86, 2004

8.2.2 Υποβεβλημένα προς κρίση

20. G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, Y. Avrithis, "Movie summarization based on audio-visual saliency detection", Proc. of the IEEE Int'l Conference on Image Processing (ICIP'08), *submitted.*

□

Βιβλιογραφία

[1] A. Adam, E. Rivlin, I. Shimshoni, D. Reinitz, "Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 555-560, Mar 2008

[2] Adelson E.H., Bergen J., "Spatiotemporal energy models for the perception of motion", J. Opt. Soc. Amer., vol. 2, no. 2, pp. 284-299, Feb. 1985.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Systems and experiment performance of optical flow techniques", Int'l Journal of Computer Vision, vol. 12, pp. 43-77, 1994

[4] Bergen J.R., Julesz B., "Parallel versus serial processing in rapid pattern discrimination", Nature, vol. 303, no. 5919,pp. 696-698, 1983

[5] M.J. Black, P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields". CVIU, vol. 63, no. 1, pp. 75-104, 1996.

[6] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes", IEEE Int'l Conference on Computer Vision (ICCV), Beijing, Oct 2005.

[7] C.R. Bolles, H.H. Baker, "Generalizing epipolar-plane image analysis on the spatio-temporal surface", Int'l J. Comput. Vision, vol. 3, pp. 33-49, 1989.

[8] C.R. Bolles, H.H. Baker, D.H. Mariont, "Epipolar-plane image analysis: An approach to determining structure from motion", Int'l J. Comput. Vision, vol. 1, pp. 7-55, 1987.

[9] A.F. Bobick, J.W. Davis, "The recognition of human movement using temporal templates", IEEE Trans. PAMI, vol. 23, pp. 257-267, 2001.

[10] O. Boiman and M. Irani, "Detecting irregularities in images and in video", IEEE Int'l Conference on Computer Vision (ICCV), Beijing, Oct 2005.

[11] Braun J., "Vision and attention: the role of training", Nature, vol. 393, pp. 424-425, 1998

[12] Braun J., Sagi D., "Vision outside the focus of attention", Percept. Psychophys., vol. 48, no. 1, pp. 45-58, 1990

[13] A.P. Bradley, F.W. Stentiford, "Visual attention for region of interest coding in JPEG 2000", Visual Communication and Image Representation, vol. 14, pp. 232-250, 2003.

[14] Burt, P., Adelson, E., "The laplacian pyramid as a compact image code", IEEE Transactions on Communications, vol. 31, pp. 532-540, 1983

[15] K. Cave, "The feature gate model of visual selection". Psychological Research, vol. 62, pp.182-194, 1999.

[16] Y. Caspi and M. Irani, Spatio-Temporal Alignment of Sequences. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 24, No. 11, pp. 1409-1424, Nov. 2002.

[17] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/, EC Funded CAVIAR project/IST-2001-37540.

[18] G.Y. Chen, T.D. Bui, and A. Krzyzak, "Image denoising using neighbouring wavelet coefficients", ICASSP'04, vol. 2, pp. 917-920, 2004.

[19] A. Chambolle, R.A. DeVore, N.-Y. Lee, B.J. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage", IEEE Trans. On Image Processing, vol. 7, no. 3, Mar. 1998.

[20] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression", IEEE Trans. Image Processing, vol. 9, pp. 1532-1546, Sep 2000.

[21] G. Csurka, C. Dance, C. Bray, L. Fan, "Visual categorization with bags of keypoints", in Proc. Statistical Learning in Computer Vision, 2004.

[22] DeValois, R. L., Albrecht, D. G., Thorell, L. G., "Spatial-frequency selectivity of cells in macaque visual cortex", Vision Research, 22, pp. 545-559, 1982.

[23] Desimone, R. Duncan, J. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, pp. 193-222, 1995.

[24] K.G. Derpanis, J.M. Gryn, "Three-dimensional n^th derivative of Gaussian separable steerable filters", Technical report CS-2004-05, York University, Nov 2004.

[25] A. Dempster, N. Laird, D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", J. Royal Statistical Soc., Ser. B, vol. 39, no. 1, pp. 1-38, 1977.

[26] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", Biometrika, vol. 81, pp. 425-455, 1994.

[27] D.L. Donoho and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", J. Amer. Statist. Assoc., vol. 90, no. 432, pp. 1200-1224, 1995.

[28] P. Dollár, V. Rabaud, G. Cottrell, S. Belongie, "Behavior Recognition via Sparse Spatio-Temporal Features", VS-PETS, pp. 65-72, Oct 2005.

[29] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification", John Wiley and Sons, New York 2001

[30] Engel, S., Zhang, X., Wandell, B., "Colour tuning in human visual cortex measured with functional magnetic resonance imaging", Nature, 388, pp. 68-71, 1997.

[31] G. Evangelopoulos, K. Rapantzikos, P. Maragos, Y. Avrithis, "Audio-Visual Attention Modeling and Salient Event Detection", in Multimodal Processing and Interaction, P. Maragos, A. Potamianos, Eds., Springer, to be published.

[32] G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, Y. Avrithis, "Movie summarization based on audio-visual saliency detection", Proc. of the IEEE Int'l Conference on Image Processing (ICIP'08), *submitted.*

[33] F. Faghih, and M. Smith, "Combining spatial and scale-space techniques for edge detection to provide a spatially adaptive wavelet-based noise filtering algorithm", IEEE Trans. Image Process., vol. 11, no. 9, pp. 1062-1071, Sep 2002.

[34] W.T. Freeman, E.H. Adelson, "The Design and Use of Steerable Filters", IEEE Transactions on Pattern Analysis and Machine Intelligence., 1991

[35] P. Groves, P. Bajcsy, "Methodology for hyperspectral band and classification model selection", IEEE Workshop on Advances in Techniques for Anlaysis of Remotely sensed data, pp. 120-128, Oct. 2003

[36] Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S., Anderson, C. H., "Overcomplete steerable pyramid filters and rotation invariance", In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Seattle, WA,pp. 222-228, Jun. 1994

[37] N.M. Grzywacz, A.L. Yuille, "A model for the estimate of local image velocity by cells in the visual cortex", Proc. Royal Society of London, B 239:129-161, 1990

[38] F.H. Hamker, "Modeling Attention: From Computational Neuroscience to Computer Vision", Neurobiology of Attention, L. Itti, G. Rees, J. Tsotsos (editors), Academic Press, 2005.

[39] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, G. Coleman, "Detection and explanation of anomalous activities: representing activities as bags of event n-grams", CVPR'05, vol. 1, pp. 1031-1038, Jun 2005.

[40] C. Harris, M. Stephens, "A combined corner and edge detector, Alvey Vision Conference, pp. 147-152, 1988.

[41] F.H. Hamker, J. Worcester, "Object detection in natural scenes by feedback", H.H. Bulthoff et al. (eds.), Biologically Motivated Computer Vision, Lecture Notes in Computer Science. Berlin, Heidelberg, New York: Springer Verlag, pp. 398-407, 2002.

[42] David J. Heeger. Optical flow using spatiotemporal filters. Int'l Journal of Computer Vision, pp. 279-302, 1988.

[43] Hikosaka O., Miyauchi S., Shimojo S., "Orienting a spatial attention - its reflexive, compensatory and voluntary mechanisms", Brain Res Cogn Brain Res, vol. 5, no. 1-2, pp. 1-9,1996

[44] B. Horn and B. Shunck, "Determining optical flow", Artificial Intelligence, no. 17, 185-203, 1981.

[45] Human action dataset, http://www.nada.kth.se/cvap/actions/

[46] L. Itti, "Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention", IEEE Transactions on Image Processing, Vol. 13, No. 10, pp. 1304-1318, Oct 2004.

[47] L. Itti, P. Baldi, "A Principled Approach to Detecting Surprising Events in Video", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 631-637, Jun 2005.

[48] Itti L., Koch C., Niebur E., "A model of saliency-based visual attention for rapid scene analysis", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol. 20, no. 11, pp. 1254-1259, 1998.

[49] Itti L., Koch C., "A saliency-based search mechanism for overt and covert shifts of visual attention", Vision Research, vol. 40, pp. 1489-1506, 2000.

[50] L. Itti, C. Koch, "Computational modelling of visual attention", Nature reviews, Neuroscience, vol. 2, no. 3, pp. 194-203, Mar 2001

[51] James W., "The principles of psychology", Cambridge, MA: Harvard UP, 1890/1981.

[52] Jaehne B., "Spatio-temporal image processing: Theory and scientific applications", New York: Springer-Verlag, 1993

[53] Joly P., Kim H.K., "Efficient automatic analysis of camera work and microsegmentation of video using spatiotemporal images", Signal Process.: Image Commun., no. 8, pp. 295-307, 1996

[54] P.K. Kaiser, R.M. Boynton, "Human Color Vision", Second Edition, Washington, DC, Optical Society of America, 1996

[55] E.R. Kandel, H.J. Schwartz, M.T. Jessell, "Essentials of Neural Science and Behavior", Appleton and Lange, 1995.

[56] Y. Ke, R. Sukthankar, M. Hebert, "Efficient Visual Event Detection using Volumetric Features", Intern. Conference on Computer Vision, vol. 1, pp. 166-173, Oct 2005.

[57] N.G. Kingsbury, "Shift invariant properties of the dual-tree complex wavelet transform", Proceedings of the Acoustics, Speech, and Signal Processing (ICASSP), vol. 3, pp. 1221-1224, 1999.

[58] N.G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals", Applied Computational Harmonic Anal., vol. 10, no. 3, pp. 234-253, May 2001

[59] E. Loupias, N. Sebe, S. Bres, J.-M. Jolion, "Wavelet-based salient points for image retrieval", in Proc.IEEE Int. Conf. Image Processing,Vancouver, BC, Canada,vol.2, pp.518-521, 2000.

[60] Koch C., Ullman S., "Shifts in selective visual attention: towards the underlying neural circuitry", Human Neurobiology, vol. 4, pp. 219-227, 1985.

[61] I. Laptev, B. Caputo, C. Schuldt, T. Lindeberg, "Local Velocity-Adapted Motion Events for Spatio-Temporal Recognition", Computer Vision and Image Understanding, vol. 108, pp. 207-229, 2007.

[62] I. Laptev and T. Lindeberg, "Space-Time Interest Points", in Proc. ICCV'03, Nice, France, pp. 432-443, 2003 (matlab implementation in http://www.nada.kth.se/ laptev/code.html).

[63] I. Laptev, T. Lindeberg , "Local Descriptors for Spatio-Temporal Recognition", ECCV Workshop "Spatial Coherence for Visual Motion Analysis", May 2004.

[64] S. Lazebnik, C. Schmid, J. Ponce,"Affine-invariant local descriptors and neighborhood statistics for texture recognition", in ICCV, vol. 1 , 649-655, 2003.

[65] Leventhal, A., "The neural basis of visual function", In Vision and visual dysfunction, vol. 4. Boca Raton, FL: CRC Press, 1991

[66] T. Lindeberg, "Feature detection with automatic scale selection", in Intern. Journal of Computer Vision, vol. 30, no.2, pp. 79-116, 1998.

[67] T. Lindeberg, J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure", in Image and Vision Computing, vol. 15, no.6, pp. 415-434, 1997.

[68] Liu F., Picard R.W., "Finding periodicity in space and time", in Proc. IEEE Int. Conf. On Computer Vision, pp. 376-383, 1998.

[69] F.-F. Li, R. VanRullen, C. Koch, P. Perona, "Rapid natural scene categorization in the near absence of attention", Proc. Natl. Acad. Sci. 99, pp. 8378 - 8383, 2002.

[70] Luschow, A., Nothdurft, H. C., "Pop-out of orientation but no pop-out of motion at isoluminance", Vision Research, 33, pp. 91-104, 1993.

[71] Y.-F. Ma, L. Lu, H.-J. Zhang, and M. Li, "A user attention model for video summarization," ACM Multimedia Conf., pp.533-542, 2002

[72] Y. Ma, L. Lu, H. Zhang, M. Li, "A generic framework of user attention model and its application in video summarization", IEEE Trans. on Multimedia, vol. 7, no. 5, pp. 907-919, 2005

[73] R. Milanese, S. Gil, T.Pun, "Attentive mechanisms for dynamic and static scene analysis", Optical Engineering, vol. 34, no. 8, pp. 2428-2434,Aug 1995.

[74] K. Mikolajczyk, C. Schmid, "Scale & Affine Invariant Interest Point Detectors", Int'l Journal of Computer Vision, vol. 60, no. 1, pp. 63-86, 2004

[75] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, "A comparison of affine region detectors", in Intern. Journal of Computer Vision, vol. 65, no. 1, pp. 43-72, 2006.

[76] N. Moenne-Loccoz, E. Bruno, S. Marchand-Maillet, "Knowledge-based Detection of Events in Video Streams from Salient Regions of Activity", Pattern Analysis and Applications, vol. 7, no. 4, pp. 422-429, Dec. 2004

[77] A. Monnet, A. Mittal, N. Paragios, V. Ramesh, "Background modelling and subtraction of dynamic scenes", IEEE International Conf. On Computer Vision (ICCV'03), vol. 2, pp. 1305-1313, 2003.

[78] MUSCLE WP5 Movie Dialogue DataBase v1.1, Aristotle University of Thessaloniki, AIIA Lab, 2007.

[79] EU FP6 Network of Excellence on Multimedia Understanding through Semantics, Computation and Learning (MUSCLE), FP6-507752.

[80] Nakayama K., Mackeben M., "Sustained and transient components of focal visual attention", Vis Res, vo. 29, no. 11, pp. 1631-1647, 1989

[81] Ngo C.-W., Pong T.-C., Zhang H.-J., "Motion analysis and segmentation through spatio-temporal slices processing", IEEE Trans. On Image Processing, vol. 12, no. 3, Mar 2003

[82] J.C. Niebles, H. Wang, Li Fei-Fei, "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words", British Machine Vision Conference (BMVC), Edinburgh, 2006.

[83] N.M. Oliver, B. Rosario, A.P. Pentland, "A Bayesian computer vision system for modelling human interactions", IEEE Trans. on Pattern Analysis and Mach. Intelligence, vol. 22, no. 8, Aug 2000.

[84] A. Opelt, M. Fussenegger, A. Pinz, P. Auer, "Weak hypotheses and boosting for generic object detection and recognition", in Proc. of European Conference on Computer Vision, Prague, Czech Republic, pp. 71-84, 2004.

[85] N. Otsu, "A threshold selection method from gray level histograms," IEEE Transactions on Systems, Man and Cybernetics, vol. 9, pp. 62-66, 1979

[86] Patel N.V., Sethi I.K., "Video Shot Detection and Characterization for Video Databases", Pattern Recognition, vol. 30, no. 4, pp. 583-592, April 1997

[87] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intra scale statistical model for Bayesian wavelet based image denoising", IEEE Trans. Image Process., vol. 11, no. 5, pp. 545-557, May 2002.

[88] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model", ICIP, 2001.

[89] Porikli F., Wang Y., "Automatic video object segmentation using volume growing and hierarchical clustering", EURASIP Journal on Applied Signal Processing (Object-based and Semantic Image & Video Analysis), vol. 2004, no. 6, pp. 814-832, Jun 2004.

[90] C. M. Privitera, L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22 (9), pp. 970-982, 2000.

[91] K. Rapantzikos, Y. Avrithis, "An enhanced spatiotemporal visual attention model for sports video analysis", Int'l Workshop on Content-based Multimedia indexing (CBMI'05), Riga, Latvia, Jun 2005

[92] K. Rapantzikos, Y.Avrithis, S. Kollias, "Spatiotemporal saliency for event detection and representation in the 3D Wavelet Domain: Potential in human action recognition", Proc. ACM Int'l Conference on Image and Video Retrieval (CIVR), pp. 294 - 301 , Jul 2007.

[93] K. Rapantzikos, Y. Avrithis, S. Kollias, "SALIENShrink: Saliency-based image denoising", ICIP'07, vol. 1, pp. 305-308, San Antonio, Sep 2007.

[94] K. Rapantzikos, Y. Avrithis and S. Kollias, "????Volumetric saliency for event detection and activity representation????", IEEE Trans. , *submitted.*

[95] K. Rapantzikos, Y. Avrithis, S. Kollias, "On the use of spatiotemporal visual attention for video classification", Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV '05, Sardinia, Italy, September 15-16, 2005.

[96] K. Rapantzikos, Y. Avrithis, S. Kollias, "Handling uncertainty in video analysis with spatiotemporal visual attention", Proc. of IEEE Int'l Conference on Fuzzy Systems (FUZZ-IEEE '05), Reno, Nevada, May 22-25, 2005.

[97] R.P.N. Rao, D.H. Ballard, "Probabilistic Models of Attention based on Iconic Representations and Predictive Coding", Neurobiology of Attention, L. Itti, G. Rees, and J. Tsotsos (ed-itors), Academic Press, 2004.

[98] K. Rapantzikos, G. Evangelopoulos, P. Maragos, Y. Avrithis, "An Audio-Visual Saliency Model for Movie Summarization", Proc. IEEE Int'l Workshop on Multimedia Signal Processing (MMSP07), Chania, Greece, Oct 2007.

[99] Ramanan, D., Forsyth, D. A. Using temporal coherence to build models of animals. In Proc. ICCV, vol. 1, pp. 338- 345, Oct 2003.

[100] K. Rapantzikos N. Tsapatsoulis, "On the implementation of visual attention architectures", Tales of the Disappearing Computer, Santorini, June 2003.

[101] K. Rapantzikos, N. Tsapatsoulis, "Enhancing the robustness of skin-based face detection schemes through a visual attention architecture", Proc. IEEE Int'l Conference on Image Processing (ICIP), Genova, Italy, Sep 2005.

[102] K. Rapantzikos, N. Tsapatsoulis, "A committee machine scheme for feature map fusion under uncertainty: the face detection case", IJISTA, special Issue on "Intelligent Image and Video Processing and Applications: The Role of Uncertainty", Int'l Journal of Intelligent Systems Technologies and Applications, vol. 1, no. 3/4, pp. 346-358, 2006.

[103] K. Rapantzikos, N. Tsapatsoulis, and Y. Avrithis, "Spatiotemporal visual attention architecture for video analysis", Proc. of IEEE Int'l Workshop On Multimedia Signal Processing (MMSP04), 2004.

[104] K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis, S. Kollias, "A bottom-up spatiotemporal visual attention model for video analysis", IET Image Processing, vol. 1, no. 2, pp. 237- 248, Jun 2007.

[105] K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis and S. Kollias, "Spatiotemporal Visual Attention for Video classification", Signal Processing: Image Communication, *under revision.*

[106] K. Rapantzikos. M. Zervakis, "Robust optical flow estimation in MPEG sequences", IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP), Mar 2005.

[107] M. Riedmiller, "Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms", Int. Journal of Computer Standards and Interfaces, special Issue on Neural Networks, vol. 16, pp. 265-278, 1994.

[108] Riesenhuber, M., T. Poggio. Models of Object Recognition, Nature Neuroscience, 3 Supp., 1199-1204, 2000.

[109] Riesenhuber, M. T. Poggio. Neural Mechanisms of Object Recognition, Current Opinion in Neurobiology, vol. 12, pp. 162-168, 2002.

[110] U. Rutishauser, D. Walther, C. Koch, P. Perona, "Is bottom-up attention useful for object recogntion?", CVPR'04, pp. 37-44, Jul 2004.

[111] Sarkar S., Majchrzak D., Korimilli K., "Perceptual organization based computational model for robust segmentation of moving objects", CVIU, no. 86, pp. 141-170, 2002.

[112] E. Sahouria, A. Zakhor, "Content analysis of video using principal components," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 9, No. 8, Dec 1999.

[113] C. Schmid, "Constructing models for content-based image retrieval", in Proc. CVPR, vol. 2, pp. 39-45, 2001.

[114] C. Schuldt, I. Laptev, B. Caputo , "Recognizing Human Actions: A Local SVM Approach", in Proc. ICPR'04, 2004.

[115] C. Schmid, R. Mohr, "Local gray value invariants for image retrieval", in. IEEE Trans. On Pattern Analysis and Mach. Intelligence, vol. 19, no. 5, pp. 530-535, 1997.

[116] Schill, K., Umkehrer, E., Beinlich, S., Krieger, G. , Zetzsche, C., "Scene analysis with saccadic eye movements: top-down and bottom-up modeling", J. Electronic Imaging,vol. 10, no. 1, pp. 152-160, 2001

[117] J. Serra, "Image Analysis and Mathematical Morphology", New York: Academic Press, 1982.

[118] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising, exploiting interscale dependencies", IEEE Trans. on Signal Processing, vol. 50, no. 11, pp. 2744-2756, Nov 2002.

[119] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation", IEEE Signal Processing Letters, vol. 9, no. 12, pp. 438-441, Dec 2002. http://taco.poly.edu/WaveletSoftware/denoise2.html

[120] E.P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain", Proc. SPIE , vol. 3813, pp. 188-195, 1999.

[121] J. Sivic, B. Russell, A.A. Efros, A. Zisserman, B. Freeman,"Discovering Objects and Their Location in Images", ICCV'05, Oct 2005.

[122] J. Sivic, F. Schaffalitzky, A. Zisserman, "Object level grouping for video shots", in Proc. of the 8th European Conf. on Computer Vis., pp. 724-734, 2004.

[123] J. Sivic, A. Zisserman, "Video google: A text retrieval approach to object matching in videos", in Proc. Int'l Conf. on Computer Vision (ICCV03), vol. 2, pp. 1470, 2003.

[124] C. Stauffer, W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking", Proc. Conf. Computer Vision and Pattern Recognition, vol.2, pp.246-252, Jun 1999.

[125] M.K. Tanemhaus, M.J. Spivey-Knowlton, K.M. Eberhard, J.C. Sedivy, "Integration of visual and linguistic information in spoken language comprehension", Science, vol. 268, no. 5217, pp. 1632-1634, 1995.

[126] A. Tikhonov, V. Arsenin, "Solutions of ill-posed problems", New York,Winston and Sons, 1977.

[127] A. Torralba, "Contextual Priming for Object Detection", Intern. Journal on Comp. Vis., vol. 53, no. 2, pp. 169-191, Jul 2003.

[128] A. Torralba, "Contextual Influences on Saliency", Neurobiology of Attention, Eds. L. Itti, G. Rees and J. Tsotsos, Academic Press / Elsevier, pp. 586-593., 2005.

[129] Tootell, R. B., Hamilton, S. L., Silverman, M. S., Switkes, E., "Functional anatomy of macaque striate cortex. i. ocular dominance, binocular interactions, and baseline conditions", Journal of Neuroscience, vol. 8, pp.1500-1530, 1988.

[130] A. Treisman, "Perceptual grouping and attention in visual search for features and for objects," J. Exp. Psychol: Hum. Percept. Perf., 8, pp. 194-214, 1982.

[131] A. Treisman, "Features and objects: the fourteenth Bartlett Memorial lecture", Q. J. Experimential Psychology, 40A, pp. 201-237, 1988.

[132] A. Treisman, "The perception of features and objects," In A. Baddeley and L. Weiskrantz (Eds.) Attention: Selection, Awareness and Control, Oxford: Uarendon Press, pp. 5-35, 1993.

[133] A. Treisman, "Feature binding, attention and object perception," Phil. Trans. R. Soc. Lond. B., 353, pp. 1295-1306, 1998.

[134] V. Tresp, "Committee Machines", Handbook for Neural Network Signal Processing, Yu Hen Hu and Jeng-Neng Hwang (eds.), CRC Press, 2001.

[135] A. Treisman and G. Gelade, "A feature integration theory of attention," Cognition Psychology, 12, pp. 97-136, 1980.

[136] N. Tsapatsoulis N, Y. Avrithis, and S. Kollias, "Facial Image Indexing in Multimedia Databases," Pattern Analysis and Applications: Special Issue on Image Indexation; vol. 4(2/3), pp. 93-107, 2001.

[137] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, N. Davis, F. Nuflo, "Modeling visual attention via selective tuning", Artificial Intelligence, vol. 78, pp. 507-545, 1995

[138] N. Tsapatsoulis, C. Pattichis, K. Rapantzikos, "Biologically inspired region of interest selection for low bit-rate video coding", Proc. of the IEEE Int'l Conference on Image Processing (ICIP'07), vol. 3, pp.333- 336, San Antonio, Sep 2007.

[139] N. Tsapatsoulis, K. Rapantzikos and Y. Avrithis, "Priority Coding for Video-telephony Applications based on Visual Attention", 2nd International Mobile Multimedia Communications Conference (MobiMedia 2006), Alghero, Italy, September 18-20, 2006.

[140] N. Tsapatsoulis, K. Rapantzikos, C. Pattichis, "An Embedded Saliency Map Estimator Scheme: Application to Video Encoding", International Journal of Neural Systems, vol. 17, No. 4, pp. 1-16, Aug 2007.

[141] G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis and S. Kollias, "A Snake Model for Object Tracking in Natural Sequences", Elsevier, Signal Processing: Image Communication, vol. 19, no. 3, pp. 219-238, Mar 2004.

[142] G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis, S. Kollias, "Rule-driven Object Tracking in Clutter and Partial Occlusion with Model-based Snakes", EURASIP Journal on Applied Signal Processing, vol. 2004, no. 6, pp. 841-860, Jun 2004.

[143] A. Turina, T. Tuytelaars, L. VanGool, "Efficient Grouping under perspective skew", in Proc. IEEE Conf. on Computer Vis. and Patt. Recogn. (CVPR01), 2001.

[144] T. Tuytelaars,L. VanGool, L. D'haene, R. Koch, "Matching of affinely invariant regions for visual servoing", in Proc. Int. Conf. Robotics and Automation, ICRA'99.

[145] B. A. Wandell, "Foundations of Vision", Sinauer Associates, Sunderland, MA 01375, 1995.

[146] Y. Wang, H. Jiang, H. Jiang, M.S. Drew, Z. Li, G. Mori, "Unsupervised Discovery of Action Classes". In Proc. of CVPR'06, vol. 2, pp. 17-22, 2006

[147] R.P. Wildes, J.R. Bergen, "Qualitative Spatiotemporal Analysis Using an Oriented Energy Representation", ECCV, vol. 2, pp. 768-784, 2000.

[148] T.J. Williams, B.A. Draper, "An Evaluation of Motion in Artificial Selective Attention", CVPR'05, pp., Jun 2005

[149] Wolfe, J. M. Visual search in continuous, naturalistic stimuli. Vision Res. vol. 34, pp. 1187-1195, 1994.

[150] A. Yarbus, "Eye movements and vision", Plenum Press, New York, 1967

[151] W.Yu, G. Sommer, K. Daniilidis, "Using skew Gabor filter in source signal separation and local spectral multi-orientation analysis", Image & Vis. Computing, vol. 4, no. 1 , pp. 377-392,Apr 2005

[152] L. Zelnik-Manor, M. Irani, "Statistical Analysis of Dynamic Actions", IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol. 28, no. 9, pp. 1530–1535, Sep 2006.

[153] H. Zhong, J. Shi, M. Visontai, "Detecting Unusual Activity in Video", CVPR'04, Washington, DC, vol. 2, pp. 819-826, Jun 2004.

□

Παράρτημα A

Παράρτημα

A.1 Υπολογισμός μερικών παραγώγων

Ο υπολογισμός της μερικής παραγώγου του ενδο-χαρακτηριστικού περιορισμού, όπως αυτός ορίζεται στην (6.21) υπολογίζεται ως εξής:

Δεδομένου του ότι

$$A_{k,m}(s) = C_{k,m}(s) - \frac{1}{|N_q|} \cdot \sum_{r \in N_q} C_{k,m}(r)$$

$$B_{k,m}(q,s) = \frac{\partial}{\partial C_{k,m}(s)} \cdot \left[\frac{1}{|N_q|} \cdot \sum_{r \in N_q} C_{k,m}(r) \right], \quad B_{k,m}(q,s) = \begin{cases} 1 & \text{if } s \in N_q \\ 0 & \text{if } s \neq N_q \end{cases}$$

$$\begin{aligned}
\frac{\partial E_1(\mathbf{C})}{\partial C_{k,m}(s)} &= \sum_s \frac{\partial}{\partial C_{k,m}(s)} \left(C_{k,m}(s) - \frac{1}{|N_q|} \sum_{r \in N_q} C_{k,m}(r) \right)^2 \\
&= 2 \cdot \sum_s A_{k,m}(s) \cdot \frac{\partial}{\partial C_{k,m}(s)} \left(A_{k,m}(s) \right) \\
&= 2 \cdot \left[\left(A_{k,m}(s) \cdot \frac{\partial}{\partial C_{k,m}(s)} (C_{k,m}(s)) \right) - \sum_s A_{k,m}(s) \cdot B_{k,m}(q,s) \right] \\
&= 2 \cdot \left[A_{k,m}(s) - \sum_{q \in N(s)} \frac{1}{|N_q|} A_{k,m}(q) \right] \\
&= 2 \cdot \left[C_{k,m}(s) - \frac{1}{N^2} \cdot \sum_{q \in N(s)} \left(2N \cdot C_{k,m}(q) - \sum_{r \in N_q} C_{k,m}(r) \right) \right]
\end{aligned}$$

Η μερική παράγωγος του δια-χαρακτηριστικού περιορισμού, όπως αυτός ορίζεται στην (6.22) υπολογίζεται ως:

$$\begin{aligned}
\frac{\partial E_2}{\partial C_{k,\ell}(s)} &= \sum_i \frac{\partial}{\partial C_{k,\ell}(s)} \left(C_{k,m}(s) - \frac{1}{M-1} \sum_{j \neq i} C_{j,\ell}(s) \right)^2 \\
&= 2 \cdot \sum_i C_{k,m}(s) - \frac{1}{M-1} \cdot \sum_{j \neq i} C_{j,\ell}(s) \cdot \frac{\partial}{\partial C_{k,\ell}(s)} \left(C_{k,m}(s) - \frac{1}{M-1} \sum_{j \neq i} C_{k,m}(s) \right) \\
&= 2 \cdot \left[\left(C_{k,\ell}(s) - \frac{1}{M-1} \sum_{j \neq k} C_{j,\ell}(s) \right) - \frac{1}{M-1} \cdot \sum_{i \neq k} \left(C_{k,m}(s) - \frac{1}{M-1} \cdot \sum_{j \neq i} C_{j,\ell}(s) \right) \right] \\
&= 2 \cdot \frac{M}{M-1} \cdot \left(C_{k,\ell}(s) - \frac{1}{M-1} \cdot \sum_{j \neq k} C_{j,\ell}(s) \right)
\end{aligned}$$

with

$$\frac{\partial}{\partial C_{k,\ell}(s)} \left(C_{k,m}(s) \right) = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$$

$$\frac{\partial}{\partial C_{k,\ell}(s)} \left(\sum_{j \neq i} C_{k,m}(s) \right) = \begin{cases} 1 & \text{if } i \neq k \\ 0 & \text{if } i = k \end{cases}$$

$$\sum_{i \neq k} \sum_{j \neq i} C_{j,\ell}(s) = (M-1) \cdot C_{k,\ell}(s) + (M-2) \cdot \sum_{j \neq k} C_j(s)$$

□