



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Στοχαστική οπτική παρακολούθηση αντικειμένου

Διπλωματική Εργασία

του

Ζαμπογιάννη Κωνσταντίνου

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2014



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Στοχαστική οπτική παρακολούθηση αντικειμένου

Διπλωματική Εργασία

του

Ζαμπογιάννη Κωνσταντίνου

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 25^η Σεπτεμβρίου 2014.

.....
Πέτρος Μαραγκός
Καθηγητής
Ε.Μ.Π.

.....
Κωνσταντίνος Τζαφέστας
Επίκουρος Καθηγητής
Ε.Μ.Π.

.....
Γεράσιμος Ποταμιάνος
Αναπληρωτής Καθηγητής
Πανεπιστήμιο Θεσσαλίας

Αθήνα, Σεπτέμβριος 2014

.....

Κωνσταντίνος Ζαμπογιάννης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Κωνσταντίνος Ζαμπογιάννης, 2014.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η οπτική παρακολούθηση αποτελεί ένα από τα μεγαλύτερου πρακτικού ενδιαφέροντος προβλήματα του επιστημονικού πεδίου της Όρασης Υπολογιστών και αντικείμενο εντατικής έρευνας. Στόχος της παρακολούθησης είναι ο χωρικός εντοπισμός, σε κάθε χρονική στιγμή μιας ακολουθίας βίντεο, ενός ή περισσότερων κινούμενων αντικειμένων, είτε στον τρισδιάστατο χώρο της ορατής σκηνής είτε στο επίπεδο της εικόνας. Συστήματα που υλοποιούν αλγορίθμους οπτικής παρακολούθησης χρησιμοποιούνται ευρέως σε εφαρμογές επιτήρησης, αναγνώρισης ενεργειών, αυτόματης πλοήγησης και διεπαφών ανθρώπου-μηχανής (HCI).

Στην παρούσα εργασία, κατόπιν επισκόπησης μέρους της σχετικής βιβλιογραφίας, αναπτύσσουμε μια ακολουθιακή μέθοδο δισδιάστατης παρακολούθησης ενός αντικειμένου, η οποία εντάσσεται στο θεωρητικό πλαίσιο του Bayesian filtering και χρησιμοποιεί χαρακτηριστικά χρώματος και κίνησης της ακολουθίας εικόνων εισόδου. Για το λόγο αυτό, σημαντικό μέρος της εργασίας αφιερώνεται στην επισκόπηση του θεωρητικού υποβάθρου για κάθε ένα από αυτά τα επιμέρους θέματα (στατιστική μοντελοποίηση χρώματος και κατάτμηση εικόνων με βάση αυτό, εκτίμηση κίνησης και οπτική ροή, Μπεϋζιανή παρακολούθηση). Η μέθοδος που προτείνουμε υλοποιεί ένα SIR Particle Filter, το μοντέλο μετάβασης του οποίου έχει μια προκαθορισμένη συναρτησιακή μορφή και οι παράμετροί του εκτιμούνται στο χρόνο εκτέλεσης (online), από δεδομένα χρώματος και κίνησης. Η θεώρηση ενός προσαρμοστικού στα δεδομένα εικόνας δυναμικού μοντέλου καθοδηγεί αποδοτικά την αναζήτηση στο χώρο των καταστάσεων, καθώς εξετάζονται κατά κύριο λόγο υποθέσεις που θεωρούνται εκ των προτέρων περισσότερο πιθανές. Το γεγονός αυτό έχει ως συνέπειες τη σύγκλιση στην υπό εκτίμηση ύστερη πυκνότητα με μικρότερο πλήθος σωματίων και την μη απαίτηση ενός ιδιαίτερα εκλεπτυσμένου μοντέλου παρατήρησης. Επιπλέον, ο αλγόριθμός μας ανιχνεύει αυτόματα, στο χρόνο εκτέλεσης, περιπτώσεις όπου η παρακολούθηση μπορεί να γίνει με ασφάλεια με βάση μόνο τη χρωματική κατάτμηση του τρέχοντος πλαισίου και προσαρμόζει κατάλληλα το μοντέλο μετάβασης. Για λόγους σύγκρισης, υλοποιείται και ένας αντίστοιχος αλγόριθμος που βασίζεται στο φίλτρο Kalman και χρησιμοποιεί τα ίδια χαρακτηριστικά εικόνας.

Η εφαρμογή της μεθόδου μας προϋποθέτει χειροκίνητη αρχικοποίηση της διαδικασίας και γνώση ενός χρωματικού μοντέλου για το υπό παρακολούθηση αντικείμενο, ενώ δεν πραγματοποιούνται περιοριστικές υποθέσεις για το είδος και την κίνηση του τελευταίου. Η πειραματική αξιολόγηση του αλγορίθμου μας πάνω σε πολλαπλών επιπέδων δυσκολίας ακολουθίες βίντεο επιβεβαιώνει ότι η μέθοδός μας είναι αρκετά γενική και μπορεί να εφαρμοστεί με επιτυχία σε ένα εύρος εφαρμογών. Επιπλέον, η παρακολούθηση με Particle Filter δίνει συστηματικά καλύτερα αποτελέσματα από την εφαρμογή του φίλτρου Kalman, για όλα τα θεωρούμενα μετρικά αξιολόγησης, ακόμα και για μικρές τιμές του πλήθους των σωματίων που χρησιμοποιεί.

Λέξεις-κλειδιά: όραση υπολογιστών, παρακολούθηση αντικειμένου, Bayesian filtering, φίλτρο σωματίων, φίλτρο Kalman, στατιστικά μοντέλα χρώματος, κατάτμηση εικόνας, εκτίμηση κίνησης, οπτική ροή.

Abstract

Visual tracking is a problem of major practical interest within the area of Computer Vision and a field of active study. The goal of tracking is the spatial localization of one or more moving objects, at any given time in a video sequence, in either the 3D space of the visible scene or the image plane. Systems that implement visual tracking algorithms are widely used in surveillance, action recognition, automatic navigation and human-computer interface (HCI) applications.

In this work, after a brief review of the tracking literature, we develop a sequential, two dimensional, single object tracking method, which is formulated in the Bayesian filtering framework and relies on color and motion features of the input video sequence. For completeness purposes, we dedicate a significant portion of this work to the theoretical background review of each of these individual subjects (statistical color modeling and color-based image segmentation, motion estimation and optical flow, Bayesian tracking). Our proposed method implements an SIR Particle Filter with a transition model of a predetermined functional form, whose parameters are estimated online, at runtime, from color and motion image data. The adoption of a partly data-driven dynamical model results in a more efficient state space search, as most of the generated hypotheses are considered to be more probable a priori. This enables the convergence of the particle set approximation toward the true posterior using less particles and renders the adoption of a carefully crafted observation model unnecessary. Our algorithm is also able to automatically detect, at runtime, situations where tracking can be safely performed using only the object color segmentation of the current frame, in which case the transition model is adapted accordingly. For comparison reasons, we also implement a Kalman filter tracker that uses the same type of image features.

Our method assumes manual initialization of the tracking process and the prior knowledge of a color model for the tracked object, while making no restricting assumptions about the allowed motion types of the latter. Our experimental evaluation on video sequences of varying tracking difficulty levels confirms that our method is general enough in its assumptions and can be successfully applied to a range of scenarios. In addition, Particle Filter tracking is shown to consistently produce more accurate results compared to our reference Kalman filter implementation, as suggested by all the evaluation metrics we consider, even when a small number of particles is used.

Keywords: computer vision, object tracking, Bayesian filtering, particle filter, Kalman filter, statistical color model, image segmentation, motion estimation, optical flow.

Ευχαριστίες

Ολοκληρώνοντας τη διπλωματική μου εργασία και κλείνοντας, έτσι, τον κύκλο των προπτυχιακών μου σπουδών στο Ε.Μ.Π., θα ήθελα να ευχαριστήσω όλους τους ανθρώπους της Σχολής μου που συνέβαλαν στην πορεία μου. Οφείλω πολλά στους καθηγητές μου, οι οποίοι, όλα αυτά τα χρόνια, αποτέλεσαν πρότυπα και κατεύθυναν τον τρόπο σκέψης και τις αποφάσεις μου, τόσο στα πλαίσια των σπουδών μου όσο και γενικότερα, με τρόπο για τον οποίο νιώθω ευγνώμων.

Ιδιαίτερα, θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου και επιβλέποντα της παρούσας εργασίας, κ. Πέτρο Μαραγκό. Η διδασκαλία του στα προπτυχιακά μαθήματα της Ψηφιακής Επεξεργασίας Σημάτων, της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων αποτέλεσε πηγή έμπνευσης για μένα με το βάθος και την πληρότητά της, πράγμα που λειτούργησε σαν καταλύτης για την απόφασή μου να μετακινήσω προς αυτή την κατεύθυνση το επίκεντρο των σπουδών μου και να εργαστώ στο ερευνητικό πεδίο της Όρασης Υπολογιστών. Του είμαι βαθύτατα ευγνώμων για την εμπιστοσύνη που μου έδειξε προτείνοντάς μου το θέμα της παρούσας εργασίας, για τις συμβουλές και την καθοδήγησή του καθ' όλη τη διάρκεια της εργασίας και τη διαρκή και άνευ όρων υποστήριξή του μέχρι την περάτωσή της. Επίσης, τον ευχαριστώ θερμά για την ανεκτίμητη βοήθεια και τις συστάσεις του κατά τη διάρκεια των αιτήσεών μου για μεταπτυχιακές σπουδές στο εξωτερικό. Του οφείλω σημαντικό μέρος της μέχρι τώρα πορείας μου και εύχομαι να καταφέρω να φανώ άξιος της εμπιστοσύνης του.

Θα ήθελα, επιπλέον, να ευχαριστήσω τους καθηγητές κ. Κωνσταντίνο Τζαφέστα και κ. Γεράσιμο Ποταμιάνο για το χρόνο που αφιέρωσαν με τη συμμετοχή τους στην επιτροπή εξέτασης της παρούσας διπλωματικής.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους στενούς μου φίλους για την ποικιλότητα υποστήριξη που μου παρείχαν σε όλη τη διάρκεια των σπουδών μου.

Κώστας Ζαμπογιάννης

Περιεχόμενα

1	Εισαγωγή	1
1.1	Όραση Υπολογιστών	1
1.2	Οπτική παρακολούθηση	2
1.3	Προτεινόμενη προσέγγιση	4
1.4	Διάρθρωση της εργασίας	5
2	Σχετικές εργασίες	9
2.1	Επιμέρους θέματα	9
2.1.1	Αναπαράσταση του αντικειμένου	9
2.1.2	Επιλογή χαρακτηριστικών για την παρακολούθηση	12
2.2	Παρακολούθηση αντικειμένου	13
2.2.1	Παρακολούθηση σημείων	14
2.2.2	Παρακολούθηση πυρήνα	15
2.2.3	Παρακολούθηση περιγράμματος/σιλουέτας	17
3	Κατάτμηση εικόνας με βάση το χρώμα	19
3.1	Εικόνες και χρώμα	19
3.1.1	Χώροι χρώματος	21
3.1.2	Κανονικοποίηση χρωματικών τιμών	25
3.1.3	Ψηφιακές εικόνες	27
3.2	Στατιστικά μοντέλα χρώματος	28
3.2.1	Ιστογράμματα χρώματος	29
3.2.2	Παραμετρικά χρωματικά μοντέλα	30
3.3	Κατάτμηση εικόνας σε αντικείμενα με βάση το χρώμα	33
3.3.1	Η κατάτμηση εικόνας σαν πρόβλημα Μπεϋζιανής απόφασης	34
3.3.2	Κατάτμηση αντικειμένου με καταφλίωση χρωματικών τιμών	38
3.3.3	Παραδείγματα εφαρμογών	39
4	Εκτίμηση κίνησης: οπτική ροή	45
4.1	Οπτική ροή	45
4.2	Μέθοδοι ταιριάσματος μπλοκ	47
4.2.1	Απλό ταίριασμα μπλοκ	47
4.2.2	Αφινικό ταίριασμα μπλοκ	48
4.3	Διαφορικές μέθοδοι εκτίμησης οπτικής ροής	49
4.3.1	Εξίσωση οπτικής ροής	49
4.3.2	Το πρόβλημα του “ανοίγματος” (aperture problem)	50
4.3.3	Η μέθοδος Horn-Schunck	51

4.3.4	Η μέθοδος Lucas-Kanade	54
4.4	Εφαρμογή: κατάτμηση με βάση την κίνηση	57
5	Bayesian filtering σε ακολουθιακά δεδομένα	61
5.1	Περιγραφή στο χώρο καταστάσεων	61
5.2	Συμβολισμοί και παραδοχές	62
5.3	Είδη συμπερασμού σε ακολουθιακά δεδομένα	62
5.4	Bayesian filtering	64
5.5	Το φίλτρο Kalman	66
5.6	Το επεκτεταμένο φίλτρο Kalman (EKF)	67
5.7	Particle Filtering	67
5.7.1	Δειγματοληψία σημαντικότητας (importance sampling)	68
5.7.2	SIS Particle Filter	69
5.7.3	Εκφυλισμός του συνόλου αναπαράστασης	70
5.7.4	Επιλογή της κατανομής σημαντικότητας	71
5.7.5	Επαναδειγματοληψία (resampling)	72
5.7.6	Γενικός αλγόριθμος Particle Filter	74
5.7.7	SIR Particle Filter	75
6	Θεωρητική περιγραφή της μεθόδου οπτικής παρακολούθησης	77
6.1	Ορισμός χώρου καταστάσεων	78
6.2	Χρήσιμα ενδιάμεσα μεγέθη και συμβολισμοί	78
6.3	Παρακολούθηση με Particle Filter	83
6.3.1	Ορισμός δυναμικού μοντέλου	85
6.3.2	Ορισμός μοντέλου παρατήρησης	88
6.4	Παρακολούθηση με Kalman Filter	91
6.4.1	Ορισμός δυναμικού μοντέλου	92
6.4.2	Ορισμός μοντέλου παρατήρησης	94
6.5	Παρακολούθηση χωρίς αβεβαιότητα	97
6.6	Παρατηρήσεις που αφορούν τις υλοποιήσεις	101
7	Πειράματα	103
7.1	Ακολουθίες βίντεο εισόδου	103
7.2	Περιγραφή της μεθοδολογίας αξιολόγησης	104
7.2.1	Ορισμός των μετρικών αξιολόγησης	105
7.2.2	Πειραματική διαδικασία	107
7.3	Πειραματικά αποτελέσματα	107
7.4	Συμπεράσματα	110
8	Πορίσματα της εργασίας	113
8.1	Συμβολή της εργασίας	113
8.2	Μελλοντικές κατευθύνσεις	114
A	Ταξινόμηση k-means	117
B	Ο αλγόριθμος EM για μείγματα Γκαουσιανών κατανομών	119
B.1	Μείγματα Γκαουσιανών κατανομών	119
B.2	Εκτίμηση παραμέτρων μέγιστης πιθανοφάνειας	120

C	Ενδεικτικά ενδιάμεσα αποτελέσματα και μεγέθη	125
C.1	Μεγέθη χρώματος και κίνησης	125
C.2	Κάποια ενδιάμεσα αποτελέσματα	127
	Βιβλιογραφία	131

List of Algorithms

1	Skin segmentation using explicit color space regions	39
2	Horn-Schunck optical flow estimation	54
3	Lucas-Kanade optical flow estimation	57
4	Kalman Filter	66
5	Extended Kalman Filter	68
6	SIS Particle Filter	70
7	Systematic Resampling	74
8	Generic Particle Filter	75
9	SIR Particle Filter	76
10	Particle Filter tracker	84
11	Kalman Filter tracker	92
12	k -means clustering (Lloyd's algorithm)	118
13	EM algorithm for Gaussian mixtures	122

Κατάλογος σχημάτων

2.1	Είδη αναπαράστασης αντικειμένου: (a) Κεντροειδές σημείο, (b) Πολλαπλά σημεία, (c) Ορθογώνια περιοχή, (d) Ελλειπτική περιοχή, (e) Περιοχές με βάση τα μέρη (part-based), (f) Σκελετός αντικειμένου, (g) Σημεία ελέγχου της καμπύλης-περιγράμματος, (h) Πλήρες περίγραμμα (καμπύλη), (i) Σιλουέτα (δείκτρια δυαδική εικόνα). Πηγή: [1].	10
2.2	Είδη παρακολούθησης: (a) Σημείων (αντιστοίχιση πολλαπλών σημείων), (b) Πυρήνα (παραμετρικό μοντέλο κίνησης ορθογώνιας περιοχής), (c,d) Περιγράμματος/σιλουέτας (εξέλιξη καμπύλης). Πηγή: [1].	14
3.1	Το ορατό φάσμα (μήκη κύματος σε nm).	20
3.2	Οι (κανονικοποιημένες) φασματικές αποκρίσεις (ευαισθησίες) των τριών ειδών κωνίων του ανθρώπινου οφθαλμού.	21
3.3	Απεικόνιση του χώρου RGB στο μοναδιαίο τριδιάστατο κύβο. Οι άξονες του καρτεσιανού συστήματος συντεταγμένων αντιστοιχούν σε διαφορετικές εντάσεις των (μονοχρωματικών) πρωταρχικών χρωμάτων (R, G, B).	22
3.4	Το χρωματικό διάγραμμα CIE xy .	23
3.5	Ο χώρος χρώματος HSV.	24
3.6	Αποτελέσματα κανονικοποίησης χρωματικών τιμών στο χώρο RGB.	27
3.7	Εικόνες δοκιμής.	34
3.8	Κατανομές χρωμάτων των εικόνων 3.7a και 3.7b.	35
3.9	Εφαρμογή ημιαυτόματης κατάτμησης	40
3.10	Χρωματικά μοντέλα των αντικειμένων/κλάσεων της εικόνας 3.9a. Πάνω σειρά: μοντελοποίηση του χρώματος του αντικειμένου (ω_1). Κάτω σειρά: μοντελοποίηση του χρώματος του παρασκηνίου της εικόνας (ω_2).	41
3.11	Αποτέλεσμα κατάτμησης του αντικειμένου της 3.9a.	41
3.12	Αρχική κατάτμηση και δεδομένα εκπαίδευσης μοντέλων χρώματος για 4 πλαίσια της ακολουθίας που χρησιμοποιήθηκε για την εφαρμογή κατάτμησης δέρματος. Πρώτη στήλη: εικόνα από την ακολουθία της δοκιμής. Δεύτερη στήλη: αρχική (πρόχειρη) κατάτμηση του αντικειμένου με βάση τον αλγόριθμο 1. Τρίτη στήλη: μάσκα ορισμού δεδομένων εκπαίδευσης για το μοντέλο χρώματος του αντικειμένου (μορφολογικό άνοιγμα της μάσκας της δεύτερης στήλης ως προς ένα κυκλικό δίσκο ακτίνας 5). Τέταρτη στήλη: μάσκα ορισμού δεδομένων εκπαίδευσης για το μοντέλο χρώματος του παρασκηνίου των εικόνων (μορφολογικό άνοιγμα του συμπληρώματος της μάσκας της δεύτερης στήλης ως προς ένα κυκλικό δίσκο ακτίνας 5).	42

3.13	Χρωματικά μοντέλα των αντικειμένων/κλάσεων της ακολουθίας εικόνων που χρησιμοποιήθηκε για την εφαρμογή κατάτμησης δέρματος. Πάνω σειρά: μοντελοποίηση του χρώματος του αντικειμένου (ω_1). Κάτω σειρά: μοντελοποίηση του χρώματος του παρασκηνίου της ακολουθίας (ω_2).	43
3.14	Τελικό αποτέλεσμα κατάτμησης του αντικειμένου για τα 4 πλαίσια της 3.12.	43
4.1	Το πρόβλημα του “ανοίγματος” στην τοπική εκτίμηση διδιάστατης κίνησης. Το έντονα σχεδιασμένο διάνυσμα συμβολίζει τη συνιστώσα της κάθετης ροής f_{\perp} . Η διακεκομμένη ευθεία αναπαριστά τον περιορισμό που θέτει η εξίσωση οπτικής ροής 4.10: οποιοδήποτε διάνυσμα ταχύτητας f την ικανοποιεί αποτελεί μια τοπική λύση.	51
4.2	Εικόνες πρώτης εφαρμογής.	58
4.3	59
4.4	Εικόνες δεύτερης εφαρμογής.	59
4.5	59
5.1	Είδη online συμπερασμού σε ακολουθιακά δεδομένα. Η σκιασμένη περιοχή αντιστοιχεί στο χρονικό διάστημα για το οποίο υπάρχουν διαθέσιμα δεδομένα, τα βέλη δείχνουν τη χρονική στιγμή στην οποία επιθυμούμε να εκτιμήσουμε την κατάσταση του συστήματος και t είναι η τρέχουσα χρονική στιγμή.	63
6.1	Από αριστερά προς τα δεξιά: η μάσκα $\partial B(x_k)$ (περιορισμένη στην έκταση της $B(x_k)$), ο μετασχηματισμός (Ευκλείδειας) απόστασης της $\partial B(x_k)$ (τιμές της $d(i, \partial(B(x_k)))$ για κάθε $i \in B(x_k)$) και η συνάρτηση στάθμησης $w_{x_k}(i)$, για $i \in B(x_k)$. Για τις ανάγκες της απεικόνισης, οι τιμές των $d(i, \partial(B(x_k)))$ και $w_{x_k}(i)$ έχουν κανονικοποιηθεί στο διάστημα $[0, 1]$	90
7.1	Ενδεικτικά πλαίσια από τις ακολουθίες εισόδου	104
7.2	Ακολουθία GSL10_5 (πείραμα 1).	108
7.3	Ακολουθία GSL10_6 (πείραμα 1).	108
7.4	Ακολουθία ASL (πείραμα 1).	108
7.5	Ακολουθία GSL10_5 (πείραμα 2).	109
7.6	Ακολουθία GSL10_6 (πείραμα 2).	109
7.7	Ακολουθία ASL (πείραμα 2).	109
C.1	Διαδοχικά πλαίσια της ακολουθίας τις στιγμές $k = 48$ και $k = 49$	125
C.2	Χρωματικές μάσκες των I_{48} και I_{49}	126
C.3	Πεδία οπτικής ροής που συνδέουν τις I_{48} και I_{49} (δύο εκδοχές ανά κατεύθυνση). Για τις ανάγκες της απεικόνισης, έχει πραγματοποιηθεί υποδειγματοληψία στο 15% του αρχικού μεγέθους.	126
C.4	Κατατμήσεις με βάση την κίνηση σε $N_c = 4$ κλάσεις για κάθε εκδοχή και κατεύθυνση πεδίου οπτικής ροής. Θυμίζουμε από την ενότητα 6.2 ότι ταξινομούνται μόνο τα διανύσματα κίνησης των οποίων η αφετηρία ανήκει στη χρωματική μάσκα B_{48}^c για την πάνω σειρά (προς τα εμπρός κίνηση) και τη χρωματική μάσκα B_{49}^c για την κάτω σειρά (προς τα πίσω κίνηση).	127

- C.5 Πρώτη σειρά: γραφική απεικόνιση του συνόλου σταθμισμένων δειγμάτων (particle set) τις χρονικές στιγμές $k = 48$ και $k = 49$. Για κάθε δείγμα/κατάσταση, σχεδιάζεται το πλαίσιο οριοθέτησης (bounding box) και το διάνυσμα ταχύτητας (με αφητηρία το κέντρο του πλαισίου οριοθέτησης) που του αντιστοιχούν. Τα πάχη των γραμμών της σχεδίασης είναι ενδεικτικά (ανάλογα) των βαρών στάθμισης. Δεύτερη σειρά: τιμές των βαρών στάθμισης w_k^i των δειγμάτων, για $i = 1, \dots, 10$ και για τις δύο χρονικές στιγμές. Τρίτη σειρά: εκτίμηση (προσδοκώμενη τιμή) της κατάστασης για τις δύο χρονικές στιγμές από τα αντίστοιχα particle sets (MMSE εκτιμήσεις, εξίσωση 6.9). 128
- C.6 Δυαδικές εικόνες που αντιστοιχούν στις κύριες συνιστώσες κίνησης στην κατάσταση/δείγμα (particle) x_{48}^{10} με βάση τις κατατμήσεις της προς τα εμπρός κίνησης $S_{49}^{f,\alpha}$ του Σχήματος C.4, για $\alpha = 1, 2$. Θυμίζουμε ότι είναι $m_{49}^{f,\alpha}(x_{48}^{10}) \equiv PMC(S_{49}^{f,\alpha}, B(x_{48}^{10}))$ (εξίσωση 6.12), ενώ χρησιμοποιούμε την αναπαράσταση 6.4 για τις $S_{49}^{f,\alpha}$ (για $\alpha = 1, 2$). 129
- C.7 Δυαδικές εικόνες που αντιστοιχούν στις κύριες συνιστώσες κίνησης στην κατάσταση/δείγμα (particle) x_{49}^{10} με βάση τις κατατμήσεις της προς τα πίσω κίνησης $S_{49}^{b,\alpha}$ του Σχήματος C.4, για $\alpha = 1, 2$. Θυμίζουμε ότι είναι $m_{49}^{b,\alpha}(x_{49}^{10}) \equiv PMC(S_{49}^{b,\alpha}, B(x_{49}^{10}))$ (εξίσωση 6.28), ενώ χρησιμοποιούμε την αναπαράσταση 6.4 για τις $S_{49}^{b,\alpha}$ (για $\alpha = 1, 2$). 129

Κεφάλαιο 1

Εισαγωγή

1.1 Όραση Υπολογιστών

Η παρούσα εργασία εντάσσεται στο επιστημονικό πεδίο της Όρασης Υπολογιστών. Αντικείμενο του ερευνητικού αυτού πεδίου αποτελεί η ανάπτυξη τεχνικών για την αυτόματη εξαγωγή πληροφορίας από σύνολα εικόνων για το μέρος του πραγματικού κόσμου που απεικονίζεται σε αυτές. Οι εικόνες εισόδου είναι, τυπικά, αριθμητικά δεδομένα που αντιστοιχούν σε δισδιάστατα σήματα, ενώ η εξαγόμενη πληροφορία είναι *συμβολική*, σε όρους περιγραφής των (τρισδιάστατων) αντικειμένων της εικονιζόμενης σκηνής. Τα βασικά ερωτήματα που καλείται να απαντήσει ένα σύστημα Όρασης Υπολογιστών περιλαμβάνουν τον προσδιορισμό της *ταυτότητας* και της *εμφάνισης* των αντικειμένων της σκηνής (σχήμα, μέγεθος, χρώμα, υφή), της *θέσης* και της *ταχύτητάς/κίνησής* τους και, τέλος, των *γεωμετρικών, σημασιολογικών σχέσεων και σχέσεων ομοιότητας* που τα συνδέουν.

Η Όραση Υπολογιστών αποτελεί μία από τις, στενά συνδεδεμένες μεταξύ τους και συχνά επικαλυπτόμενες, *επιστήμες εικόνων*. Σε αυτές περιλαμβάνονται, επίσης, η Απεικόνιση (Imaging, φυσικός κόσμος \mapsto εικόνες), που ασχολείται με την αναπαράσταση σε εικονική μορφή τμήματος του φυσικού κόσμου, η Επεξεργασία Εικόνας (Image Processing, εικόνες \mapsto εικόνες), με αντικείμενο το *μετασχηματισμό* μιας εικόνας σε μια άλλη ή την *εξαγωγή* ενός συνόλου *χαρακτηριστικών* ή παραμέτρων από αυτή και η Γραφική Υπολογιστών (Computer Graphics, σύμβολα \mapsto εικόνες), που ασχολείται με την ανάπτυξη μεθόδων *σύνθεσης* εικονικής φύσης περιεχομένου από συμβολική πληροφορία. Η Όραση Υπολογιστών (Computer Vision, εικόνες \mapsto σύμβολα) έχει ως αντικείμενο την επίλυση *αντίστροφων προβλημάτων*, όπου στόχος είναι η *ανάλυση και κατανόηση* των, εν γένει, δισδιάστατων εικόνων εισόδου, προκειμένου να εξαχθεί από αυτές *συμβολική πληροφορία* (π.χ., ύπαρξη και σχέσεις ορατών αντικειμένων) ή να ανακτηθεί χαμένη πληροφορία της εικονιζόμενης σκηνής (π.χ., βάθος). Τρία θεμελιώδη υψηλού επιπέδου προβλήματα που καλείται να επιλύσει η Όραση Υπολογιστών είναι η *ανακατασκευή* (ανάκτηση τρισδιάστατης δομής της εικονιζόμενης σκηνής), η *κατάτμηση* (διαμέριση εικόνων σε ομογενείς, ως προς κάποιο κριτήριο, περιοχές) και η *αναγνώριση* (κατηγοριοποίηση των ορατών αντικειμένων).

Η Όραση Υπολογιστών ξεκίνησε την ανάπτυξή της τη δεκαετία του 1960, από τη συμβολή των επιστημονικών πεδίων της Τεχνητής Νοημοσύνης, της Επεξεργασίας Σημάτων και της Αναγνώρισης Προτύπων. Τις δεκαετίες που ακολούθησαν, εξελίχθηκε ραγδαία σε μια ευρεία διεπιστημονική ερευνητική περιοχή, στην οποία συμβάλλουν και άλλα πεδία, όπως Μηχανική Μάθηση και Στατιστική, Ρομποτική, Εφαρμοσμένα Μαθηματικά, Νευροβιολογία και Ψυχολογία. Ο όρος Μηχανική Όραση (Machine Vision) χρησιμοποιείται συχνά ως συνώνυμος του Όραση Υπολογιστών και αναφέρεται στον εφοδιασμό μηχανών (υπολογιστών, ρομπότ) με κατάλληλους αισθητήρες και

ικανότητες συμπερασμού παρόμοιων δυνατοτήτων, ιδανικά, με το ανθρώπινο σύστημα όρασης. Προς αυτή την κατεύθυνση, για τη δημιουργία αισθητήρων και κατάλληλων υπολογιστικών μοντέλων, σημαντικό μέρος της έρευνας εντός του πεδίου της Όρασης Υπολογιστών επικεντρώνεται στη μελέτη και κατανόηση της Βιολογικής Όρασης και της Μοντελοποίησης Εγχεφάλου.

Συστήματα που υλοποιούν τεχνικές Όρασης Υπολογιστών χρησιμοποιούνται με επιτυχία σε ένα ιδιαίτερα ευρύ φάσμα εφαρμογών. Μια μη εξαντλητική λίστα των τελευταίων δίνεται ακολούθως:

- Κατασκευή 3D μοντέλων (φωτογραμμετρία): αυτόματη κατασκευή τρισδιάστατων μοντέλων αντικειμένων (π.χ., κτιρίων) από σύνολα δισδιάστατων εικόνων.
- Βιομηχανικός έλεγχος ποιότητας: αυτόματη ανίχνευση ελαττωμάτων σε εξαρτήματα κατά την παραγωγή τους.
- Οπτική αναγνώριση χαρακτήρων (Optical Character Recognition, OCR): αυτόματη μετατροπή χειρόγραφου κειμένου σε ψηφιακή μορφή, αναγνώριση αριθμού πινακίδων κυκλοφορίας.
- Βάσεις δεδομένων με εικόνες: ερωτήματα (queries) και ανάκτηση εικόνων με βάση το περιεχόμενο (content-based retrieval).
- Βιοϊατρική και ιατρική απεικόνιση: υποβοήθηση χειρουργικών επεμβάσεων, αυτόματη διάγνωση και ανίχνευση παθολογικών καταστάσεων.
- Αναγνώριση βιομετρικών στοιχείων: ταυτοποίηση δακτυλικών αποτυπωμάτων ή σάρωσης αμφιβληστροειδούς (retinal scan) για πιστοποίηση ταυτότητας.
- Επιτήρηση (surveillance): παρακολούθηση κυκλοφορίας οχημάτων σε αυτοκινητόδρομους, αυτόματη ανίχνευση κρίσιμων καταστάσεων.
- Καταγραφή κίνησης (motion capture): καταγραφή της τρισδιάστατης κίνησης ηθοποιών (με χρήση πολλών καμερών και παρακολούθηση σημείων-κλειδιών πάνω σε αυτούς) και χρήση αυτής σε εφαρμογές εμπύχωσης χαρακτήρων (animation) της Γραφικής Υπολογιστών.
- Αλληλεπίδραση ανθρώπου-μηχανής (Human-Computer Interaction, HCI): σχεδίαση διεπαφών χρήστη που βασίζονται στην οπτική αναγνώριση χειρονομιών, ηλεκτρονικά παιχνίδια (π.χ., με χρήση του αισθητήρα Microsoft Kinect).
- Εφαρμογές επαυξημένης πραγματικότητας (augmented reality), μοντελοποίηση και απόδοση με βάση την εικόνα (image-based modeling and rendering).

Η εισαγωγική αυτή ενότητα βασίστηκε κατά κύριο λόγο στο [2]. Για εκτενέστερη περιγραφή του επιστημονικού πεδίου της Όρασης Υπολογιστών, παραπέμπουμε στα [2, 3, 4, 5].

1.2 Οπτική παρακολούθηση

Με τον όρο *οπτική παρακολούθηση* (visual tracking) αναφερόμαστε στο πρόβλημα του χωρικού εντοπισμού ενός ορατού αντικειμένου σε κάθε πλαίσιο (frame) μιας ακολουθίας βίντεο. Η ραγδαία εξέλιξη της υπολογιστικής ισχύος των ηλεκτρονικών υπολογιστών, σε συνδυασμό με την ευρεία, πλέον, διαθεσιμότητα προσιτών βιντεοκαμερών, έδωσε σημαντική ώθηση στην ανάπτυξη αλγορίθμων παρακολούθησης και την εφαρμογή τους σε πλήθος σεναρίων αυτόματης ανάλυσης

βίντεο. Επιτυχημένοι αλγόριθμοι παρακολούθησης χρησιμοποιούνται σε μεγάλο εύρος εφαρμογών, στις οποίες περιλαμβάνονται, ενδεικτικά, οι οπτικές διεπαφές ανθρώπου-μηχανής (HCI), σενάρια αναγνώρισης με βάση την κίνηση (ταυτοποίηση αντικειμένου με βάση τον τρόπο κίνησής του, αναγνώριση ενεργειών), συστήματα επιτήρησης (παρακολούθηση κυκλοφορίας οχημάτων, ανίχνευση ασυνήθιστων δραστηριοτήτων), συστήματα αυτόματης πλοήγησης οχημάτων (σχεδιασμός τροχιάς και αποφυγή εμποδίων) και η καταγραφή κίνησης (motion capture).

Ένα ολοκληρωμένο σύστημα οπτικής παρακολούθησης στην πράξη καλείται να λύσει τα εξής τρία υποπροβλήματα:

1. *Ανίχνευση* του κινούμενου αντικειμένου προς παρακολούθηση: το στάδιο αυτό είτε χρησιμεύει στην αρχικοποίηση του συστήματος είτε καλείται συχνά κατά τη διάρκεια της παρακολούθησης (π.χ., σε σενάρια παρακολούθησης πολλών αντικειμένων ή αν το αντικείμενο πάει, κατά διαστήματα, να είναι ορατό).
2. *Παρακολούθηση* του αντικειμένου πλαίσιο προς πλαίσιο του βίντεο εισόδου: το υπολογιστικό αυτό στάδιο υλοποιεί τη βασική λειτουργία της παρακολούθησης. Λόγω απαιτήσεων για online (και συχνά πραγματικού χρόνου) υπολογισμό της θέσης του αντικειμένου από την πλειψηφία των εφαρμογών, μια συνήθης πρακτική είναι η υιοθέτηση αιτιατών, ακολουθιακών μεθόδων, όπου ο εντοπισμός σε κάθε πλαίσιο πραγματοποιείται με βάση προηγούμενα αποτελέσματα και τις νέες μετρήσεις που προκύπτουν από ανάλυση αυτού.
3. *Ανάλυση της τροχιάς* του αντικειμένου και εξαγωγή συμπερασμάτων για τη συμπεριφορά του: το στάδιο αυτό διαφοροποιείται ανάλογα με την εφαρμογή στην οποία αναφέρεται το σύστημα παρακολούθησης. Για παράδειγμα, μπορεί συνίσταται στην αναγνώριση μιας χειρονομίας ή μιας υψηλότερου επιπέδου, σύνθετης δραστηριότητας. Σε γενικού σκοπού συστήματα παρακολούθησης, όπου το ζητούμενο είναι μόνο ο χωρικός εντοπισμός του αντικειμένου, αυτό το στάδιο συμπερασμού μπορεί να απουσιάζει.

Στο πλαίσιο της εργασίας αυτής, θα μας απασχολήσει μόνο το δεύτερο υποπρόβλημα, που αποτελεί την καρδιά ενός συστήματος παρακολούθησης, ενώ θα επικεντρώσουμε το ενδιαφέρον μας στη *δισδιάστατη παρακολούθηση* (2D tracking), που αναφέρεται στον εντοπισμό του ορατού αντικειμένου στο επίπεδο της εικόνας.

Η δυσκολία του προβλήματος της οπτικής παρακολούθησης, στην πλήρη γενικότητά του, προέρχεται από έναν αριθμό παραγόντων:

- Απώλεια πληροφορίας από την προβολή της τρισδιάστατης κινούμενης σκηνής στην ακολουθία των δισδιάστατων εικόνων του βίντεο.
- Μερική ή ολική απόκρυψη (occlusion) του υπό παρακολούθηση αντικειμένου είτε από άλλα αντικείμενα της ορατής σκηνής είτε από τμήματα του εαυτού του (self-occlusion). Τα φαινόμενα απόκρυψης καθιστούν προβληματικό το ταίριασμα τοπικών χαρακτηριστικών (feature matching) στο οποίο στηρίζονται πολλές μέθοδοι εκτίμησης κίνησης.
- Αρθρωτή (articulated) ή, γενικά, παραμορφώσιμη (non-rigid) φύση του υπό παρακολούθηση αντικειμένου. Στις περιπτώσεις όπου το αντικείμενο δεν είναι άκαμπτο (rigid), απαιτείται η θεώρηση πολύπλοκων μοντέλων για την περιγραφή της σύνθετης κίνησής του και παρατηρούνται φαινόμενα αυτο-απόκρυψης (self-occlusion).
- Σύνθετη ή με απότομες αλλαγές ταχύτητας κίνηση του αντικειμένου. Η σύνθετη κίνηση παραμορφώσιμων αντικειμένων, όπως αναφέραμε, προϋποθέτει τη θεώρηση κατάλληλων κινηματικών μοντέλων. Ακόμα, όμως, και στην περίπτωση άκαμπτων αντικειμένων, η γρήγορη

ή με απότομες μεταβολές στην ταχύτητα κίνηση συχνά καθιστά δύσκολο το ταίριασμα χαρακτηριστικών και καθιστά ανεπαρκή τη χρήση απλών κινηματικών μοντέλων (π.χ., μοντέλο σταθερής ταχύτητας).

- Ομοιότητα της εμφάνισης (appearance) του αντικειμένου με το παρασκήνιο (background) της εικόνας. Σε πολλές περιπτώσεις, κάποια χαρακτηριστικά εμφάνισης του υπό παρακολούθηση αντικειμένου, όπως το χρώμα και η υφή, είναι παρόμοια με αυτά άλλων αντικειμένων της ορατής σκηνής και παρουσιάζεται η ανάγκη θεώρησης πιο εκλεπτυσμένων μοντέλων εμφάνισης για το αντικείμενο, που ενσωματώνουν περισσότερη πρότερη γνώση για αυτό, ή άλλου τύπου χαρακτηριστικών. Σαν παράδειγμα, αναφέρουμε την εφαρμογή παρακολούθησης του δεξιού χεριού (παλάμης) του νοηματιστή σε ένα βίντεο νοηματικής γλώσσας. Αν χρησιμοποιήσουμε μόνο χαρακτηριστικά χρώματος για την παρακολούθηση, δεν θα υπάρχει η δυνατότητα διαχωρισμού ανάμεσα στα δύο χέρια και το πρόσωπο αν αυτά επικαλύπτονται. Η κατάσταση είναι ακόμη πιο δύσκολη αν, στο ίδιο σενάριο, ο νοηματιστής φοράει κοντά μανίκια, καθώς, με βάση μόνο το χρώμα, η παλάμη δε θα διακρίνεται από το υπόλοιπο χέρι.
- Αλλαγές στις συνθήκες φωτισμού της ορατής σκηνής κατά τη διάρκεια της κίνησης, που, αν δε ληφθούν υπόψη, δυσκολεύουν την εξαγωγή χρήσιμων για την παρακολούθηση χαρακτηριστικών (π.χ., χρωματική κατάτμηση, εκτίμηση κίνησης).
- Έπαρξη θορύβου στην ακολουθία εικόνων, που επίσης δυσχεραίνει την εξαγωγή εύρωστων (robust) χαρακτηριστικών για την παρακολούθηση.
- Απαίτηση για αυτόματη αρχικοποίηση (εντοπισμό του αντικειμένου) της διαδικασίας παρακολούθησης. Η ικανοποίηση αυτής της απαίτησης προϋποθέτει μηχανισμούς αυτόματου εντοπισμού της προβολής του αντικειμένου στο επίπεδο της εικόνας, στην περίπτωση της διδιάστατης παρακολούθησης, ή την επίλυση πιθανώς σύνθετων προβλημάτων βελτιστοποίησης για τον προσδιορισμό της θέσης του στον τρισδιάστατο χώρο, σε σενάρια τρισδιάστατης παρακολούθησης.
- Απαίτηση για πραγματικού χρόνου (real-time) παρακολούθηση από την εφαρμογή, που καθιστά αναγκαία την ανάπτυξη ιδιαίτερα αποδοτικών αλγορίθμων.

Το πρόβλημα της παρακολούθησης μπορεί να απλοποιηθεί σημαντικά αν πραγματοποιήσουμε κάποιο σύνολο περιοριστικών υποθέσεων για την εμφάνιση και την κίνηση του αντικειμένου. Οι περιορισμοί αυτοί προκύπτουν συνήθως από την ενσωμάτωση πρότερης γνώσης για την εμφάνιση (π.χ., χρωματικό μοντέλο) και την κίνηση (π.χ., μοντέλο σταθερής επιτάχυνσης) του αντικειμένου ή της κλάσης αντικειμένων της εφαρμογής.

Η ενότητα αυτή βασίστηκε στα [1, 6]. Για πληρέστερη επισκόπηση του προβλήματος της οπτικής παρακολούθησης, παραπέμπουμε τον αναγνώστη στα [7, 3, 1, 6], ενώ μια μερική παρουσίαση του πεδίου πραγματοποιείται στο Κεφάλαιο 2 της παρούσας εργασίας.

1.3 Προτεινόμενη προσέγγιση

Στην εργασία αυτή, προτείνουμε μια μέθοδο διδιάστατης παρακολούθησης ενός αντικειμένου που εντάσσεται στο θεωρητικό πλαίσιο του Bayesian Filtering. Συγκεκριμένα, η μεθοδός μας υλοποιεί ένα Particle Filter τύπου SIR, του οποίου το δυναμικό μοντέλο και το μοντέλο παρατήρησης ορίζονται με βάση ένα σύνολο μετρήσιμων μεγεθών που προκύπτουν από την ανάλυση

του χρώματος και της διαδιάστατης κίνησης (οπτική ροή) που χαρακτηρίζουν την ακολουθία εικόνων εισόδου. Για λόγους σύγκρισης (benchmarking), υλοποιήθηκε, επιπλέον, ένας “παρόμοιος” αλγόριθμος παρακολούθησης με φίλτρο Kalman.

Ο εντοπισμός του αντικειμένου σε κάθε πλαίσιο της ακολουθίας εισόδου γίνεται *ακολουθιακά*, με βάση το αποτέλεσμα του εντοπισμού στο προηγούμενο χρονικό βήμα (πλαίσιο) και ένα σύνολο χαρακτηριστικών χρώματος και κίνησης για το προηγούμενο και το τρέχον πλαίσιο. Μία από τις σημαντικότερες ιδιότητες της μεθόδου μας (Particle Filter) είναι η θεώρηση ενός *προσαρμοστικού* δυναμικού μοντέλου για το υπό παρακολούθηση σύστημα. Σε κάθε χρονικό βήμα, η μέθοδός μας *μαθαίνει* ένα σύνολο παραμέτρων για το θεωρούμενης μορφής μοντέλο μετάβασης από το θεωρούμενο σύνολο μεγεθών χρώματος και κίνησης. Η θεώρηση ενός “πληροφορημένου” δυναμικού μοντέλου μπορεί να διευκολύνει σημαντικά την αναζήτηση στο χώρο των καταστάσεων, καθώς οδηγεί την τελευταία περιορίζοντάς την στο σύνολο των “έγκυρων” (περισσότερο πιθανών) καταστάσεων, και καθιστά ένα εκλεπτυσμένο μοντέλο παρατήρησης λιγότερο αναγκαίο.

Οι υλοποιήσεις μας δέχονται ως είσοδο μια ακολουθία βίντεο, ένα στατιστικό χρωματικό μοντέλο για το υπό παρακολούθηση αντικείμενο και τη θέση αυτού στο αρχικό πλαίσιο της ακολουθίας, μέσω χειροκίνητης αρχικοποίησης της διαδικασίας. Σε θεωρητικό επίπεδο, οι υποθέσεις που πραγματοποιεί η μέθοδος μας και οι οποίες περιορίζουν το πρόβλημα της παρακολούθησης, οριοθετώντας το σύνολο των πιθανών εφαρμογών, είναι δύο:

- Απαιτείται, όπως αναφέραμε, η εκ των προτέρων γνώση της στατιστικής κατανομής του χρώματος του υπό παρακολούθηση αντικειμένου και, προαιρετικά, του παρασκήνιου των εικόνων.
- Υποτίθεται ότι η κίνηση του αντικειμένου είναι *κατά κύριο λόγο* μεταφορική. Η έννοια αυτή, με τρόπο που θα αποσαφηνιστεί όταν περιγράψουμε τυπικά τη μέθοδό μας, αναφέρεται στην υπόθεση ότι η κύρια συνιστώσα διαδιάστατης κίνησης στην έκταση της προβολής του αντικειμένου, που θα οριστεί παρακάτω, είναι αντιπροσωπευτική της συνολικής κίνησης (ακριβέστερα, της κίνησης του κέντρου) της προβολής του αντικειμένου στο επίπεδο της εικόνας.

Σημειώνουμε ότι η δεύτερη υπόθεσή μας δεν επιδρά με ιδιαίτερα περιοριστικό τρόπο στο σύνολο των τύπων κίνησης του αντικειμένου για τους οποίους η μέθοδός μας αναμένεται να λειτουργεί σωστά. Συγκεκριμένα, δεν αποκλείονται, επί της αρχής, σύνθετες κινήσεις μη άκαμπτων αντικειμένων, ενώ, επιπλέον, ο πιθανοτικός χαρακτήρας της μεθόδου της παρέχει τη δυνατότητα καλής συμπεριφοράς ή και ανάνηψης από φαινόμενα ολίσθησης (drifting), ακόμα και όταν η υπόθεση αυτή δεν ικανοποιείται.

1.4 Διάρθρωση της εργασίας

Στην ενότητα αυτή, περιγράφουμε την οργάνωση της παρούσας εργασίας, παρέχοντας μια σύνοψη των περιεχομένων κάθε κεφαλαίου. Το βασικό προϊόν της εργασίας είναι η ανάπτυξη μιας μεθόδου οπτικής παρακολούθησης ενός αντικειμένου σε ένα βίντεο, την οποία εισάγουμε στο Κεφάλαιο 6 και η οποία στηρίζεται στο θεωρητικό υπόβαθρο που παρέχεται στα Κεφάλαια 3 (χρώμα), 4 (κίνηση) και 5 (Bayesian Filtering). Αναλυτικότερα:

- Στο Κεφάλαιο 2, πραγματοποιούμε μια συνοπτική επισκόπηση ενός ενδεικτικού συνόλου προσεγγίσεων και εργασιών πάνω στο πρόβλημα της οπτικής παρακολούθησης.
- Στο Κεφάλαιο 3, κάνουμε μια εισαγωγή σε θέματα αναπαράστασης του χρώματος σε ψηφιακές εικόνες, παρουσιάζουμε τις πιο συνήθεις τεχνικές στατιστικής μοντελοποίησης του

χρώματος ενός αντικειμένου και περιγράφουμε κάποιες απλές μεθόδους κατάτμησης μιας εικόνας σε αντικείμενα με βάση το χρώμα. Μία από αυτές θα αποτελέσει συστατικό των αλγορίθμων οπτικής παρακολούθησης που αναπτύσσουμε στο Κεφάλαιο 6.

- Στο Κεφάλαιο 4, κάνουμε μια εισαγωγή στην οπτική ροή (optical flow) και παρουσιάζουμε μερικές, κλασικές πλέον, μεθόδους εκτίμησής της. Η μέθοδος οπτικής παρακολούθησης που αναπτύσσουμε στο Κεφάλαιο 6 στηρίζεται σε σημαντικό βαθμό στην οπτική ροή και σε κάποια παράγωγα μεγέθη αυτής για την εκτίμηση της κίνησης και τον προσδιορισμό των μοντέλων που περιγράφουν τη χρονική εξέλιξη του συστήματος υπό παρακολούθηση.
- Στο Κεφάλαιο 5, κάνουμε μια εισαγωγή στο γενικό πρόβλημα της ακολουθιακής εκτίμησης της κατάστασης ενός δυναμικού συστήματος στο Μπεϋζιανό θεωρητικό πλαίσιο (Bayesian filtering/tracking) και παρουσιάζουμε κάποιες από τις πιο διαδεδομένες προσεγγίσεις του. Το πρόβλημα της οπτικής παρακολούθησης μπορεί εύκολα να διατυπωθεί στο θεωρητικό αυτό πλαίσιο: οι αλγόριθμοι που αναπτύσσουμε στο Κεφάλαιο 6 υλοποιούν ένα SIR Particle Filter και ένα Kalman Filter, η λειτουργία και κάποιες ιδιότητες των οποίων αναλύονται στο κεφάλαιο αυτό.
- Στο Κεφάλαιο 6, περιγράφουμε δύο αλγορίθμους οπτικής παρακολούθησης που εντάσσονται στο γενικό θεωρητικό πλαίσιο του Κεφαλαίου 5 και χρησιμοποιούν έννοιες και μεθόδους που παρουσιάσαμε στα Κεφάλαια 3 και 4. Ορίζοντας έναν κατάλληλο χώρο καταστάσεων, διατυπώνουμε το πρόβλημα της οπτικής παρακολούθησης σαν ένα πρόβλημα Bayesian Filtering και το προσεγγίζουμε με ένα Particle Filter, που αποτελεί το βασικό μας αλγόριθμο, και ένα Kalman Filter, που υλοποιήθηκε ως μέθοδος αναφοράς. Τα κατά περίπτωση θεωρούμενα δυναμικά μοντέλα ορίζονται με βάση ένα σύνολο μεγεθών που προκύπτουν από ανάλυση του χρώματος και της κίνησης που χαρακτηρίζει τις εικόνες της ακολουθίας βίντεο που παρέχεται ως είσοδος. Οι μέθοδοι που ορίζουμε στο κεφάλαιο αυτό αποτελούν το βασικό προϊόν της παρούσας εργασίας.
- Στο Κεφάλαιο 7, εφαρμόζουμε τις μεθόδους που αναπτύξαμε σε πραγματικές ακολουθίες βίντεο. Επικεντρώσαμε το ενδιαφέρον μας σε βίντεο νοηματικής γλώσσας διαφορετικών επιπέδων δυσκολίας για την παρακολούθηση ενός χεριού του νοηματιστή. Στη συνέχεια, αφού περιγράφουμε τα δεδομένα εισόδου μας και ορίσουμε κατάλληλα μέτρα σφάλματος στα οποία θα στηριχθεί η μεθοδολογία αξιολόγησής μας, παραθέτουμε ένα σύνολο πειραματικών αποτελεσμάτων.
- Στο Κεφάλαιο 8, πραγματοποιείται μια γενική αξιολόγηση της μεθόδου παρακολούθησης που αναπτύξαμε και προτείνεται ένα σύνολο κατευθύνσεων για πιθανές μελλοντικές επεκτάσεις και βελτιώσεις αυτής.
- Στο Παράρτημα A, παρουσιάζεται η μέθοδος ταξινόμησης k -means, την οποία χρησιμοποιούμε για τον υπολογισμό ενός είδους κατάτμησης με βάση την κίνηση (ενότητες 4.4 και 6.2).
- Στο Παράρτημα B, γίνεται μια εισαγωγή στα μοντέλα μείγματος Γκαουσιανών κατανομών και παρουσιάζεται ο αλγόριθμος EM για την εκπαίδευσή τους. Χρησιμοποιούμε κατανομές αυτής της μορφής τόσο για την αναπαράσταση των χρωματικών μας μοντέλων (ενότητα 3.2), όσο και για την πιθανοτική περιγραφή του δυναμικού μοντέλου στο οποίο στηρίζεται η βασική μας μέθοδος παρακολούθησης (ενότητα 6.3.1).

- Στο Παράρτημα C, απεικονίζουμε ένα σύνολο βοηθητικών μεγεθών και ενδιάμεσων αποτελεσμάτων που χρησιμοποιούνται από τους αλγορίθμους παρακολούθησης που αναπτύξαμε για μια πραγματική ακολουθία βίντεο, προκειμένου να παρέχουμε μια καλύτερη διαίσθηση για τη λειτουργία τους. Οι απεικονίσεις βασίζονται σε μια πραγματική ακολουθία βίντεο και δίνουν τη μορφή που λαμβάνουν στην πράξη κάποια από τα βοηθητικά μεγέθη και ενδιάμεσα αποτελέσματα που περιγράφουμε θεωρητικά στο Κεφάλαιο 6.

Κεφάλαιο 2

Σχετικές εργασίες

Στο κεφάλαιο αυτό, επιχειρούμε μια συνοπτική επισκόπηση των προσεγγίσεων που έχουν προταθεί για το πρόβλημα της οπτικής παρακολούθησης. Αρχικά, αναφερόμαστε σε κάποια επιμέρους θέματα που πρέπει να απαντηθούν από ένα σύστημα παρακολούθησης για τα διάφορα στάδια επεξεργασίας που αυτό εμπεριέχει και περιγράφουμε τα πιο διαδεδομένα είδη προσέγγισής τους. Στη συνέχεια, παρουσιάζουμε συνοπτικά ένα ενδεικτικό σύνολο μεθόδων δισδιάστατης παρακολούθησης, επιχειρώντας μια αδρή κατηγοριοποίησή τους με βάση το είδος αναπαράστασης που χρησιμοποιούν για το υπό παρακολούθηση αντικείμενο.

2.1 Επιμέρους θέματα

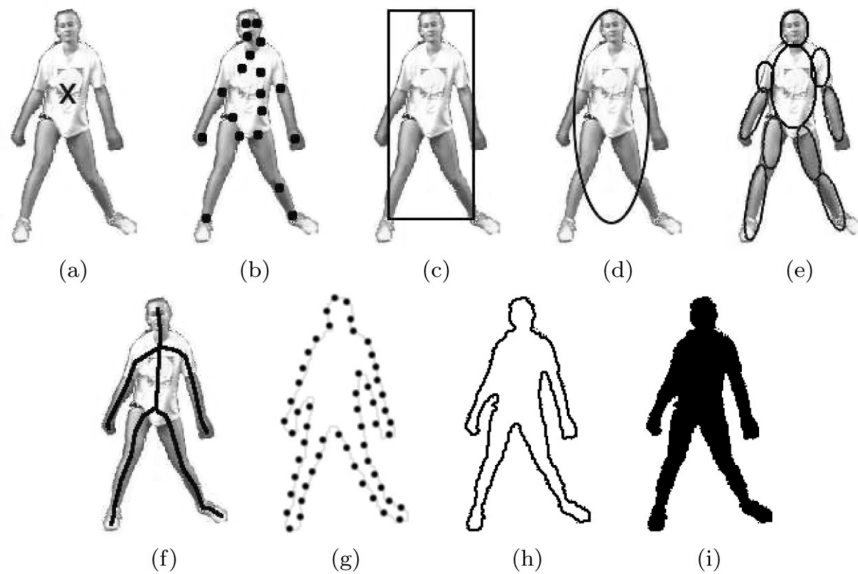
Κάθε μέθοδος οπτικής παρακολούθησης καλείται να πραγματοποιήσει έναν αριθμό σχεδιαστικών επιλογών για τα διάφορα στάδια επεξεργασίας που περιλαμβάνει. Τα κυριότερα επιμέρους ερωτήματα που πρέπει να απαντηθούν αφορούν τον τρόπο αναπαράστασης του κινούμενου αντικειμένου και την επιλογή κατάλληλων για την παρακολούθηση χαρακτηριστικών των εικόνων εισόδου, τα οποία θα αποτελέσουν την είσοδο του αλγορίθμου παρακολούθησης. Οι προτεινόμενες στη βιβλιογραφία μέθοδοι διαφοροποιούνται ως προς τις προσεγγίσεις τους στα παραπάνω θέματα, ανάλογα με τις απαιτήσεις και τους περιορισμούς του εκάστοτε πεδίου εφαρμογής. Συνεπώς, κάθε ένα από τα δύο αυτά θέματα αυτά μπορεί να χρησιμεύσει σαν ένα κριτήριο κατηγοριοποίησης των υπαρχόντων μεθόδων. Στις παραγράφους που ακολουθούν, παρουσιάζουμε συνοπτικά τις πιο διαδεδομένες προσεγγίσεις ανά περίπτωση. Το περιεχόμενο της ενότητας αυτής βασίζεται εν μέρει στην επισκόπηση του προβλήματος που πραγματοποιείται στο [1].

2.1.1 Αναπαράσταση του αντικειμένου

Για τις ανάγκες της οπτικής παρακολούθησης, το υπό παρακολούθηση αντικείμενο μπορεί να αναπαρασταθεί με βάση μια περιγραφή του σχήματος και της εμφάνισής του. Το σχήμα του αντικειμένου καθορίζεται από τη χωρική μορφή της προβολής του αντικειμένου στο επίπεδο της εικόνας, ενώ χαρακτηριστικά εμφάνισης προκύπτουν συνήθως από ανάλυση του χρώματος και της υφής του.

Οι πιο κοινές μορφές περιγραφής χωρικής πληροφορίας (θέση και σχήμα) που χρησιμοποιούνται από αλγορίθμους παρακολούθησης απεικονίζονται στο Σχήμα 2.1 και περιγράφονται με συντομία στη συνέχεια:

- **Σημεία.** Ένα αντικείμενο μπορεί να αναπαρασταθεί σαν ένα μεμονωμένο σημείο που αντιστοιχεί στο κεντροειδές του ή σαν ένα διακριτό σύνολο σημείων χωρίς ιδιαίτερους γεωμετρι-



Σχήμα 2.1: Είδη αναπαράστασης αντικειμένου: (a) Κεντροειδές σημείο, (b) Πολλαπλά σημεία, (c) Ορθογώνια περιοχή, (d) Ελλειπτική περιοχή, (e) Περιοχές με βάση τα μέρη (part-based), (f) Σκελετός αντικειμένου, (g) Σημεία ελέγχου της καμπύλης-περιγράμματος, (h) Πλήρες περίγραμμα (καμπύλη), (i) Σιλουέτα (δείκτρια δυαδική εικόνα). Πηγή: [1].

κούς περιορισμούς για τη χωρική κατανομή τους [8, 9, 10]. Οι αναπαραστάσεις αντικειμένου με σημεία είναι περισσότερο κατάλληλες στις περιπτώσεις όπου η προβολή του αντικειμένου καταλαμβάνει μικρή έκταση στο επίπεδο της εικόνας.

- **Αρχέτυπα γεωμετρικά σχήματα.** Η προβολή του αντικειμένου μπορεί να οριοθετηθεί από απλά, θεμελιώδη γεωμετρικά σχήματα, όπως παραλληλόγραμμα πλαίσια οριοθέτησης [11] και ελλείψεις [12, 13, 14, 15]. Για αυτού του είδους τις αναπαραστάσεις, η κίνηση του αντικειμένου μοντελοποιείται, συνήθως, από έναν αφινικό μετασχηματισμό ή από απλή μετατόπιση. Παρότι είναι φαινομενικά καταλληλότερα για απλά, άκαμπτα αντικείμενα, τα αρχέτυπα γεωμετρικά σχήματα χρησιμοποιούνται και για την παρακολούθηση παραμορφώσιμων αντικειμένων που εκτελούν σύνθετη κίνηση.
- **Αρθρωτά μοντέλα σχήματος και σκελετικά μοντέλα.** Τα αρθρωτά μοντέλα σχήματος συντίθενται από έναν αριθμό μερών, κάθε ένα από τα οποία αναπαρίσταται, συνήθως, από κάποιο βασικό γεωμετρικό σχήμα και τα οποία συνδέονται μεταξύ τους με συνδέσμους. Η σχέση ανάμεσα στα μέρη καθορίζεται από κατάλληλα κινηματικά μοντέλα που μπορεί να περιγράφουν, για παράδειγμα, τις πιθανοτικές κατανομές των γωνιών των συνδέσμων. Παράδειγμα αντικειμένου για το οποίο μια τέτοια αναπαράσταση είναι κατάλληλη είναι το ανθρώπινο σώμα [16].
- **Περίγραμμα και σιλουέτα του αντικειμένου.** Η σιλουέτα του αντικειμένου ορίζεται ως το χωρίο του επιπέδου της εικόνας που καταλαμβάνεται από την προβολή του αντικειμένου, ενώ το περίγραμμά του είναι η καμπύλη που αντιστοιχεί το σύνορο του χωρίου αυτού. Μια καμπύλη στο επίπεδο μπορεί να αναπαρασταθεί με πολλούς τρόπους. Από τη σκοπιά της οπτικής παρακολούθησης, έχουν εφαρμοστεί με επιτυχία παραμετρικές αναπαραστάσεις [17, 18].

- **Ενεργά μοντέλα σχήματος.** Τα ενεργά μοντέλα σχήματος (Active Shape Models, ASM [19]) είναι στατιστικά μοντέλα που βασίζονται σε κατανομές σημείων-οροσθίων (landmarks) του αντικειμένου και επιτρέπουν μεταβολές του σχήματος μόνο με “γνωστούς” από τα επισημειωμένα δεδομένα εκπαίδευσής τους τρόπους. Μια μέθοδος οπτικής που στηρίζεται σε ενεργά μοντέλα σχήματος μπορεί να βρεθεί στο [20].

Στη συνέχεια αναφερόμαστε στις πιο διαδεδομένες προσεγγίσεις μοντελοποίησης της εμφάνισης του υπό παρακολούθηση αντικειμένου. Σημειώνουμε ότι κάποιες από αυτές μοντελοποιούν από κοινού χαρακτηριστικά εμφάνισης και χωρική πληροφορία (σχήμα).

- **Στατιστικές κατανομές χαρακτηριστικών εμφάνισης.** Οι εκτιμήσεις των πιθανοτικών κατανομών μπορεί να είναι παραμετρικές, συνήθως με Γκαουσιανές κατανομές ή μείγματα αυτών, ή μη παραμετρικές, με πυρήνες (παράθυρα Parzen-Rosenblatt) ή, συνήθως, υπό τη μορφή ιστογραμμάτων. Ο υπολογισμός της υπό εκτίμηση πυκνότητας των θεωρούμενων χαρακτηριστικών εμφάνισης, τα οποία συνήθως αφορούν χρώμα ή υφή, πραγματοποιείται στις περιοχές της εικόνας που υποδεικνύονται από το χρησιμοποιούμενο μοντέλο σχήματος του αντικειμένου (π.χ., στο εσωτερικό της καμπύλης του περιγράμματος). Οι εργασίες [12, 13, 14, 15] χρησιμοποιούν απλά γεωμετρικά σχήματα για την αναπαράσταση του σχήματος του αντικειμένου, ενώ μοντελοποιούν την εμφάνισή του με μη παραμετρικές εκτιμήσεις κατανομών χρωματικών τιμών (ιστογράμματα).
- **Εικόνες-τεμπλέτες.** Οι τεμπλέτες σχηματίζονται συνήθως από σιλουέτες αντικειμένου που οριοθετούνται από απλά γεωμετρικά σχήματα. Παρότι περιέχουν από κοινού χωρική πληροφορία και χαρακτηριστικά εμφάνισης, οι τεμπλέτες κωδικοποιούν μόνο μία όψη του αντικειμένου, γεγονός που περιορίζει σημαντικά τους αποδεκτούς τρόπους κίνησης του αντικειμένου. Για παράδειγμα, στο [8] υποτίθεται αφινικό μοντέλο δισδιάστατης κίνησης. Οι μέθοδοι που βασίζονται σε τεμπλέτες είναι περισσότερο κατάλληλες για εφαρμογές όπου η πόζα του αντικειμένου δεν μεταβάλλεται σημαντικά κατά τη διάρκεια της παρακολούθησης.
- **Ενεργά μοντέλα εμφάνισης.** Τα ενεργά μοντέλα εμφάνισης (Active Appearance Models, AAM [21]) μπορούν να θεωρηθούν επέκταση των ενεργών μοντέλων σχήματος και έχουν εφαρμοστεί με επιτυχία στην παρακολούθηση προσώπου. Κωδικοποιούν από κοινού το σχήμα και την εμφάνιση του αντικειμένου, αποθηκεύοντας για κάθε σημείο-ορόσημο ένα σύνολο χαρακτηριστικών χρώματος και υφής που έχουν προκύψει από τη διαδικασία εκπαίδευσης του μοντέλου. Ένα ενδιαμέσο στα AAM και ASM είδος μοντέλου, που περιέχει μόνο τοπική πληροφορία εμφάνισης, χρησιμοποιείται στο [22] για την παρακολούθηση παραμορφώσιμων αντικειμένων.
- **Μοντέλα εμφάνισης πολλαπλών όψεων.** Τα μοντέλα αυτά κωδικοποιούν πληροφορία από πολλαπλές όψεις του αντικειμένου. Ένας τρόπος αναπαράστασης του σχήματος και της εμφάνισης που αντιστοιχεί σε πολλαπλές προκύπτει από τον υπολογισμό ενός κατάλληλου διανυσματικού ιδιόχωρου από τις δεδομένες όψεις [23, 24]. Μια άλλη προσέγγιση είναι η μάθηση των διαφορετικών όψεων μέσω κατάλληλης εκπαίδευσης ταξινομητών. Στο [25], η παρακολούθηση υποβοηθείται από παράλληλη ανίχνευση του αντικειμένου, όπου ο ταξινομητής ενημερώνεται online με βάση τις νέες όψεις του αντικειμένου που γίνονται γνωστές κατά τη διάρκεια της κίνησής του.

Η επιλογή του είδους αναπαράστασης του αντικειμένου εξαρτάται σε μεγάλο βαθμό από το πεδίο εφαρμογών της εκάστοτε μεθόδου. Έτσι, η αναπαράσταση με σημεία είναι περισσότερο κατάλληλη για αντικείμενα που καλύπτουν μικρό μέρος του επιπέδου της εικόνας, ενώ τα αρχέτυπα

γεωμετρικά σχήματα υιοθετούνται όταν είτε προσεγγίζουν ικανοποιητικά το σχήμα του αντικειμένου είτε δεν υπάρχει ανάγκη για ιδιαίτερα ακριβή οριοθέτηση αυτού (π.χ., αν μας ενδιαφέρει μόνο η θέση, η κλίμακα (μέγεθος) και, ενδεχομένως, ο γενικός προσανατολισμός του). Για την παρακολούθηση με ακρίβεια πολύπλοκου σχήματος αντικειμένων, προτιμούνται αναπαραστάσεις περιγράμματος. Αξίζει να σημειωθεί ότι, όπως θα δούμε και στη συνέχεια, υπάρχει στενή σχέση ανάμεσα στην αναπαράσταση του αντικειμένου, το είδος του αλγορίθμου παρακολούθησης και τα χαρακτηριστικά εικόνας που αυτός χρησιμοποιεί.

2.1.2 Επιλογή χαρακτηριστικών για την παρακολούθηση

Στην υποενότητα αυτή αναφερόμαστε στα πιο συνήθη είδη χαρακτηριστικών που εξάγονται από τις εικόνες της ακολουθίας βίντεο για να χρησιμοποιηθούν ως είσοδος του αλγορίθμου παρακολούθησης. Στο πλαίσιο της οπτικής παρακολούθησης, είναι φυσικό να προτιμούνται χαρακτηριστικά που να καθιστούν ευδιάκριτο το υπό παρακολούθηση αντικείμενο στο χώρο που αυτά ορίζουν (feature space). Τα περισσότερο συνηθισμένα χαρακτηριστικά εικόνων που χρησιμοποιούνται στο πλαίσιο της παρακολούθησης περιγράφονται στη συνέχεια:

- **Χρώμα.** Το χρώμα είναι το πιο ευρέως χρησιμοποιούμενο από τα χαρακτηριστικά εικόνων για παρακολούθηση [12, 13, 14, 15]. Μια εισαγωγή σε θέματα που αφορούν τους τρόπους αναπαράστασης και μοντελοποίησης του χρώματος δίνεται στο Κεφάλαιο 3. Σημειώνουμε ότι στη βιβλιογραφία χρησιμοποιείται ένας αριθμός από διαφορετικούς χώρους χρώματος, χωρίς να είναι ξεκάθαρη η υπεροχή κάποιου από αυτούς στην πράξη.
- **Οπτική ροή.** Η κίνηση των αντικειμένων της ορατής σκηνής ως προς την κάμερα έχει ως αποτέλεσμα την καταγραφή ενός χρονικά μεταβαλλόμενου (κινούμενου) προτύπου φωτεινής έντασης από τον αισθητήρα της. Η οπτική ροή είναι ένα δισδιάστατο διανυσματικό πεδίο που ορίζεται από τις προβολές των τρισδιάστατων διανυσμάτων κίνησης (ταχύτητα/μετατόπιση) των σημείων του τρισδιάστατου ορατού κόσμου πάνω στο επίπεδο της εικόνας και το οποίο, συνεπώς, αντιστοιχεί στις δισδιάστατες ταχύτητες των προβολών των ορατών αντικειμένων πάνω σε αυτό. Μια περιγραφή των πιο κλασικών μεθόδων εκτίμησης οπτικής ροής, με αναφορές σε πιο σύγχρονες εργασίες, γίνεται στο Κεφάλαιο 4.
- **Ακμές.** Τα περιγράμματα των αντικειμένων συνήθως χαρακτηρίζονται από απότομες μεταβολές στη συνάρτηση έντασης της εικόνας και, συνεπώς, υποδεικνύονται από τις ακμές αυτής. Τα χαρακτηριστικά ακμών έχουν τα πλεονεκτήματα ότι δεν είναι ιδιαίτερα ευαίσθητα σε αλλαγές των συνθηκών φωτισμού και μπορούν να υπολογιστούν αποδοτικά και με ακρίβεια (π.χ., ανιχνευτής ακμών Canny [26]). Χρησιμοποιούνται συνήθως από αλγορίθμους παρακολούθησης του περιγράμματος του αντικειμένου.
- **Υφή.** Η ανάλυση υφής έχει στόχο την εξαγωγή πληροφορίας για τον τρόπο με τον οποίο μεταβάλλεται η φωτεινή ένταση στην έκταση μιας επιφάνειας εξαιτίας της φύσης (εμφάνισης) της ίδιας της επιφάνειας και ανεξαρτήτως του περιβάλλοντος φωτισμού. Μέσω της ανάλυσης υφής, ποσοτικοποιούνται ιδιότητες της επιφάνειας που σχετίζονται με το πόσο ομαλή είναι και με το αν παρουσιάζει κάποια μορφή κανονικότητα, ντετερμινιστική ή στοχαστική (π.χ., εμφάνιση επαναλαμβανόμενων προτύπων). Μια επισκόπηση των μεθόδων ανάλυσης και αναπαράστασης της υφής μπορεί να βρεθεί στα Κεφάλαια 13 και 14 του [2]. Όπως και οι ακμές, τα χαρακτηριστικά υφής δεν είναι ιδιαίτερα ευαίσθητα σε μεταβολές των συνθηκών φωτισμού.

- **Τοπικά χαρακτηριστικά.** Περιγραφές (descriptors) τοπικών χαρακτηριστικών της εικόνας μπορούν να χρησιμοποιηθούν σαν είσοδος ενός αλγορίθμου παρακολούθησης. Αυτά μπορεί να αφορούν σημεία (π.χ., σημεία ενδιαφέροντος (γωνίες) τύπου Shi-Tomasi (KLT) [8] ή Harris-Stephens, χαρακτηριστικά SIFT [27]) ή περιοχές της εικόνας (π.χ., χαρακτηριστικά HOG [28]). Στα [9] και [29] χρησιμοποιούνται, αντίστοιχα, KLT και SIFT χαρακτηριστικά, ενώ στο [30] παρακολουθούνται HOG περιγραφές περιοχών για την αναγνώριση χειρονομιών.

Η επιλογή των χαρακτηριστικών είναι στενά συνδεδεμένη με το είδος της αναπαράστασης του αντικειμένου που χρησιμοποιείται. Για παράδειγμα, το χρώμα χρησιμοποιείται για αναπαραστάσεις της εμφάνισης του αντικειμένου που βασίζονται σε χρωματικά ιστογράμματα, ενώ χαρακτηριστικά ακμών χρησιμοποιούνται συνήθως για αναπαραστάσεις σχήματος με περιγράμματα. Πολλοί αλγόριθμοι παρακολούθησης λειτουργούν συνδυάζοντας περισσότερα του ενός είδη χαρακτηριστικών. Κάποιοι από αυτούς λειτουργούν στο πλαίσιο της παρακολούθησης μέσω ανίχνευσης του αντικειμένου (tracking via detection) και πραγματοποιούν αυτόματη επιλογή χαρακτηριστικών μέσω τεχνικών boosting [31]. Ενδεικτικά παραδείγματα μεθόδων αυτής της κατηγορίας μπορούν να βρεθούν στα [32] και [33].

2.2 Παρακολούθηση αντικειμένου

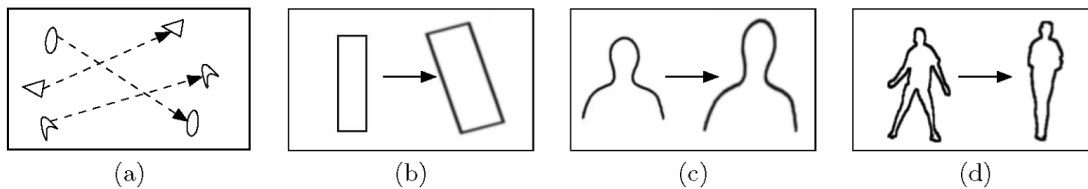
Η δισδιάστατη οπτική παρακολούθηση ενός αντικειμένου έγκειται στον εντοπισμό του σε κάθε πλαίσιο της ακολουθίας βίντεο και την κατάλληλη ενημέρωση της περιγραφής που χρησιμοποιείται για την αναπαράστασή του. Στην ενότητα αυτή, θα εστιάσουμε σε ένα ενδεικτικό σύνολο ακολουθιακών μεθόδων παρακολούθησης, όπου ο εντοπισμός στο τρέχον πλαίσιο γίνεται με βάση προηγούμενα αποτελέσματα και τα μέχρι τώρα χαρακτηριστικά που έχουν εξαχθεί από την ακολουθία εικόνων. Στην πιο συνηθισμένη περίπτωση, η εκτίμηση στο τρέχον πλαίσιο γίνεται με βάση το αποτέλεσμα της παρακολούθησης στο αμέσως προηγούμενο πλαίσιο και το σύνολο των χαρακτηριστικών (παρατηρήσεων) που εξάγονται από το τρέχον.

Η διαδικασία του εντοπισμού του αντικειμένου και της αντιστοίχισης των εμφανίσεών του που έχουν ανιχνευτεί σε διαδοχικά πλαίσια μπορεί να είναι πλήρως αποσυζευγμένες ή να πραγματοποιούνται από κοινού. Στην πρώτη περίπτωση, γίνεται χρήση κάποιου αλγορίθμου ανίχνευσης αντικειμένου, ο οποίος δέχεται σαν είσοδο ένα σύνολο χαρακτηριστικών εικόνας του τρέχοντος πλαισίου και επιστρέφει ένα σύνολο πιθανών εμφανίσεων του αντικειμένου σε αυτό. Στη συνέχεια, η μέθοδος παρακολούθησης καλείται να αντιστοιχίσει στιγμιότυπα του αντικειμένου από προηγούμενα πλαίσια με αυτά που ανιχνεύθηκαν στο τρέχον. Αυτός ο τρόπος λειτουργίας είναι γνωστός ως παρακολούθηση μέσω ανίχνευσης (tracking via detection) και ρίχνει σημαντικό βάρος της παρακολούθησης σε κάποιον αλγόριθμο εντοπισμού. Στη δεύτερη, γενικότερη και συνηθέστερη περίπτωση, ο εντοπισμός στο τρέχον πλαίσιο έγκειται στην κατάλληλη ενημέρωση του αποτελέσματος από το προηγούμενο πλαίσιο με βάση τις νέες παρατηρήσεις από το τρέχον και, ενδεχομένως, τη θεώρηση κάποιου εκ των προτέρων υποτιθέμενου δυναμικού μοντέλου που περιγράφει την κίνηση του αντικειμένου. Κάποιοι αλγόριθμοι παρακολούθησης χρησιμοποιούν και τις δύο προσεγγίσεις [25].

Η παρακολούθηση μπορεί να γίνεται είτε *ντετερμινιστικά*, πιθανώς με την υιοθέτηση κάποιων ευριστικών (heuristic) κανόνων που περιορίζουν τις θεωρούμενες δυνατές κινήσεις του αντικειμένου, είτε *στοχαστικά*, λαμβάνοντας ρητά υπόψη την αβεβαιότητα της παρακολούθησης λόγω του θορύβου τόσο του υποτιθέμενου δυναμικού μοντέλου όσο και των παρατηρήσεων (χαρακτηριστικών) που εξάγονται από τα σήματα εικόνας. Στη δεύτερη περίπτωση, υιοθετείται συνήθως μια περιγραφή του αντικειμένου ως δυναμικού συστήματος σε έναν κατάλληλο χώρο καταστάσεων

και εφαρμόζονται τεχνικές Bayesian filtering, με συνηθέστερες προσεγγίσεις το φίλτρο Kalman και τα Particle Filters. Μια σύντομη επισκόπηση των τεχνικών εκτίμησης σε αυτό το θεωρητικό πλαίσιο πραγματοποιείται στο Κεφάλαιο 5.

Όπως αναφέραμε στην προηγούμενη ενότητα, το θεωρούμενο είδος της αναπαράστασης του αντικειμένου περιορίζει τους δυνατούς τρόπους κίνησής του. Για παράδειγμα, μια αναπαράσταση μεμονωμένου σημείου περιορίζει το ενδιαφέρον μας στις μεταφορικές κινήσεις, στην περίπτωση χρήσης απλών γεωμετρικών σχημάτων θεωρούνται συχνά μοντέλα κίνησης που ορίζονται από αφινικούς μετασχηματισμούς, ενώ στις αναπαραστάσεις περιγράμματος/σιλουέτας χρησιμοποιούνται συνήθως πιο σύνθετα, παραμετρικά ή μη, μοντέλα κίνησης. Όπως είναι αναμενόμενο, το είδος της αναπαράστασης είναι στενά συνδεδεμένο με τον αλγόριθμο παρακολούθησης που εφαρμόζεται. Χρησιμοποιώντας το θεωρούμενο είδος αναπαράστασης ως κριτήριο ταξινόμησης των αλγορίθμων, διακρίνουμε τρεις γενικές κατηγορίες παρακολούθησης: *σημείων*, *πυρήνα* και *περιγράμματος/σιλουέτας*.



Σχήμα 2.2: Είδη παρακολούθησης: (a) Σημείων (αντιστοίχιση πολλαπλών σημείων), (b) Πυρήνα (παραμετρικό μοντέλο κίνησης ορθογώνιας περιοχής), (c,d) Περιγράμματος/σιλουέτας (εξέλιξη καμπύλης). Πηγή: [1].

Στο Σχήμα 2.2, απεικονίζονται ενδεικτικά παραδείγματα για κάθε ένα από τα τρία γενικά είδη παρακολούθησης. Στις υποενότητες που ακολουθούν, κάνουμε μια σύντομη αναφορά σε ενδεικτικές εργασίες από κάθε κατηγορία.

2.2.1 Παρακολούθηση σημείων

Η αναπαράσταση του αντικειμένου ως ενός μεμονωμένου σημείου μπορεί να είναι χρήσιμη όταν η προβολή του αντικειμένου καταλαμβάνει μικρή έκταση στο επίπεδο της εικόνας. Τυπικά, η παρακολούθηση μεμονωμένων σημείων συνίσταται στην αντιστοίχιση σημείων ενδιαφέροντος (τοπικών χαρακτηριστικών εικόνας) ανάμεσα στο προηγούμενο και το τρέχον πλαίσιο της ακολουθίας βίντεο. Χαρακτηριστική εφαρμογή αποτελεί η χρήση της μεθόδου KLT [8] για την παρακολούθηση γωνιών τύπου Shi-Tomasi. Σημειώνουμε ότι, όπως θα δούμε και στη συνέχεια, η μέθοδος αυτή πραγματοποιεί ουσιαστικά παρακολούθηση πυρήνα, αλλά η εφαρμογή της σε μικρής έκτασης περιοχές γύρω από κατάλληλα σημεία ενδιαφέροντος αποτελεί συνήθη πρακτική για την παρακολούθηση των τελευταίων.

Η αναπαράσταση μέσω ενός διακριτού συνόλου σημείων αποδεικνύεται χρήσιμη όταν το υπό παρακολούθηση αντικείμενο υφίσταται έντονες παραμορφώσεις κατά τη διάρκεια της κίνησής του. Στις περιπτώσεις αυτές, η θεώρηση περισσότερο περιγραφικών αναπαραστάσεων σχήματος, όπως το περίγραμμα, συχνά καθιστά αναγκαία την υιοθέτηση περίπλοκων μοντέλων για το αντικείμενο και την κίνησή του, και ξεφεύγει, ουσιαστικά, από το πλαίσιο της διδιάστατης παρακολούθησης. Μια τέτοια αναπαράσταση θα ήταν χρήσιμη, για παράδειγμα, σε σενάρια παρακολούθησης χεριού όπου μας ενδιαφέρει κυρίως η θέση του αντικειμένου. Στη συνέχεια, αναφερόμαστε σε δύο εργασίες αυτής της κατηγορίας.

Οι Kölsch και Turk [9] χρησιμοποιούν την έννοια του *σμήνους* τοπικών χαρακτηριστικών

για την αναπαράσταση του αντικειμένου. Η προτεινόμενη μέθοδος χρησιμοποιεί χαρακτηριστικά χρώματος και κίνησης (οπτική ροή τύπου Lucas-Kanade) για την παρακολούθηση, ενώ δεν θέτει περιορισμούς για τον τρόπο κίνησης και παραμόρφωσης του αντικειμένου. Η συμπεριφορά ενός συνόλου σημείων ως σμήνους ορίζεται από δύο κατώφλια: κάθε μέλος του συνόλου πρέπει να απέχει το πολύ μια προκαθορισμένη απόσταση από το κεντροειδές (διάμεσο) του συνόλου, ενώ παράλληλα περιορίζεται να βρίσκεται ακριβώς μακριά από οποιοδήποτε άλλο σημείο του συνόλου. Κάθε σημείο του συνόλου αρχικοποιείται ως η θέση ενός σημείου ενδιαφέροντος (γωνίας) Shi-Tomasi, ενώ η παρακολούθησή του γίνεται μέσω του αλγορίθμου KLT. Αν κάποια από τις δύο συνθήκες της συμπεριφοράς σμήνους παραβιάζεται από κάποιο σημείο κατά τη διάρκεια της παρακολούθησης, τότε αυτό επανατοποθετείται σε μία νέα γωνία, η οποία είναι συμβατή με αυτές και ανήκει στη χρωματική μάσκα του αντικειμένου.

Ο Hoey [10] χρησιμοποιεί την ίδια, ουσιαστικά, έννοια σμήνους για την αναπαράσταση του αντικειμένου, αλλά σε ένα στοχαστικό πλαίσιο: η παρακολούθηση διατυπώνεται σαν πρόβλημα Bayesian filtering και προσεγγίζεται με την εφαρμογή ενός Particle Filter. Το διάλυμα κατάστασης περιγράφει ακριβώς ένα σμήνος σημείων και περιλαμβάνει τη θέση και ένα τοπικό χρωματικό μοντέλο για κάθε σημείο/μέλος, τη μέση ταχύτητα του σμήνους, ένα “μέσο” (καθολικό) χρωματικό μοντέλο για το σύνολο του σμήνους και, τέλος, τις παραμέτρους που καθορίζουν τη συμπεριφορά του σμήνους (που μπορεί να είναι απλώς δύο κατώφλια απόστασης, όπως στην προηγούμενη εργασία). Όλα τα θεωρούμενα χρωματικά μοντέλα προσεγγίζονται από Γκαουσιανές κατανομές. Σύμφωνα με το προτεινόμενο δυναμικό μοντέλο, κάθε σημείο κινείται με ταχύτητα ίση με την τρέχουσα μέση ταχύτητα του σμήνους, στην οποία έχει προστεθεί Γκαουσιανός θόρυβος. Οι περιορισμοί της συμπεριφοράς σμήνους επιβάλλονται μέσω της ενσωμάτωσης στην πυκνότητα του δυναμικού μοντέλου δύο επιπλέον παραγόντων (κατανομές Gibbs), οι οποίοι θέτουν ποινή στα σμήνη με αποκλίνουσα συμπεριφορά, σύμφωνα με τις παραμέτρους σμήνους (π.χ., κατώφλια απόστασης) που περιέχονται στο διάλυμα κατάστασης. Η θεωρούμενη κατανομή σημαντικότητας του φίλτρου ορίζεται από ένα γραμμικό συνδυασμό της πυκνότητας του μοντέλου μετάβασης και ενός όρου που βασίζεται στα παρατηρούμενα δεδομένα. Τέλος, η συνάρτηση πιθανοφάνειας του μοντέλου παρατήρησης ορίζεται ανεξάρτητα για κάθε σημείο του σμήνους και είναι ενδεικτική της συμφωνίας της παρατηρούμενης εικόνας στη θέση του σημείου τόσο με το τοπικό όσο και με το καθολικό χρωματικό μοντέλο (γινόμενο δύο Γκαουσιανών πυκνοτήτων).

2.2.2 Παρακολούθηση πυρήνα

Με τον όρο “πυρήνας” αναφερόμαστε από κοινού στο σχήμα και την εμφάνιση του αντικειμένου. Παραδείγματα πυρήνων αποτελούν οι εικόνες-τεμπλέτες και οι περιγραφές σχήματος με αρχέτυπα γεωμετρικά σχήματα, εφοδιασμένες με κατανομές χαρακτηριστικών εμφάνισης στην έκταση που ορίζουν (π.χ., ιστογράμματα χρώματος). Η παρακολούθηση πυρήνα συνίσταται, συνήθως, στην εκτίμηση της κίνησης του αντικειμένου σε διαδοχικά πλαίσια του βίντεο, η οποία μπορεί να περιγράφεται παραμετρικά (π.χ., ως απλή μετατόπιση ή από έναν αφινικό μετασχηματισμό) ή να προκύπτει από το πυκνό πεδίο οπτικής ροής που συνδέει τα δύο πλαίσια. Για μια συνοπτική εισαγωγή σε θέματα εκτίμησης κίνησης, παραπέμπουμε τον αναγνώστη στο Κεφάλαιο 4.

Στην πιο απλή εκδοχή της, η παρακολούθηση τεμπλέτας πραγματοποιείται εφαρμόζοντας μεθόδους *ταιριάσματος μπλοκ* (block matching), στις οποίες αναφερόμαστε στην ενότητα 4.2. Οι μέθοδοι αυτές υπολογίζουν τις τιμές των παραμέτρων του υποτιθέμενου μοντέλου κίνησης για τις οποίες μεγιστοποιείται κάποιο κριτήριο ομοιότητας ανάμεσα στην τεμπλέτα που αναπαριστά το αντικείμενο στο προηγούμενο πλαίσιο της ακολουθίας και την αντίστοιχη, μέσω της κίνησης, περιοχή της εικόνας στο τρέχον πλαίσιο. Το θεωρούμενο μοντέλο κίνησης μπορεί να περιγράφει μια απλή μετατόπιση (ενότητα 4.2.1), οπότε αρκεί ο υπολογισμός ενός διανύσματος μετατόπι-

σης, ή έναν γενικότερο αφινικό μετασχηματισμό (ενότητα 4.2.2), οπότε απαιτείται η εκτίμηση 6 παραμέτρων (εξίσωση 4.5). Τυπικά παραδείγματα κριτηρίων ομοιότητας τεμπλετών που χρησιμοποιούνται στην πράξη είναι το άθροισμα (τετραγωνικών ή απόλυτων) μεταξύ τους διαφορών και η (κανονικοποιημένη) ετεροσυσχέτισή τους. Η παρακολούθηση τεμπλέτας μέσω ταιριάσματος μπλοκ στηρίζεται σε εξαντλητική αναζήτηση στο επίπεδο της εικόνας. Για τη μείωση του υπολογιστικού κόστους της αναζήτησης, η τελευταία συχνά περιορίζεται σε μια γειτονιά της προηγούμενης γνωστής θέσης του αντικειμένου.

Για την παρακολούθηση μιας περιοχής της εικόνας που ορίζεται από ένα αρχέτυπο γεωμετρικό σχήμα, μπορούν να χρησιμοποιηθούν και διαφορικές μέθοδοι εκτίμησης οπτικής ροής, μια εισαγωγή στις οποίες μπορεί να βρεθεί στην ενότητα 4.3. Οι μέθοδοι αυτής της κατηγορίας βασίζονται στον περιορισμό της σταθερότητας της έντασης, που εκφράζεται από τις εξισώσεις 4.8-4.10. Οι Shi και Tomasi [8] προτείνουν τον αλγόριθμο KLT για την παρακολούθηση ορθογώνιων περιοχών που έχουν κέντρο κάποιο κατάλληλα επιλεγμένο σημείο ενδιαφέροντος της εικόνας. Ως τοπικά χαρακτηριστικά προς παρακολούθηση επιλέγονται γωνίες τύπου Shi-Tomasi, για τις οποίες το πρόβλημα της τοπικής εκτίμησης κίνησης είναι, εκ κατασκευής, καλώς ορισμένο. Για την παρακολούθηση, υποτίθεται ένα μεταφορικό μοντέλο κίνησης. Η μέθοδος υπολογίζει επαναληπτικά το διάλυμα μετατόπισης της ορθογώνιας περιοχής (π.χ., παράθυρο 25×25) γύρω από κάθε σημείο που παρακολουθείται, υποθέτοντας σταθερή κίνηση στην έκταση αυτής. Ο υπολογισμός στηρίζεται στην εξίσωση 4.27 της μεθόδου τοπικής εκτίμησης οπτικής ροής των Lucas-Kanade [34], την εξαγωγή της οποίας περιγράφουμε στην ενότητα 4.3.4. Ο αλγόριθμος KLT περιλαμβάνει και ένα στάδιο αξιολόγησης της ποιότητας της παρακολούθησης για κάθε σημείο που έχει επιλεγεί, με βάση ένα αφινικό μοντέλο κίνησης. Σε κάθε βήμα, γίνεται εκτίμηση ενός αφινικού μετασχηματισμού που συνδέει αρχική τεμπλέτα, τη στιγμή της επιλογής του σημείου για παρακολούθηση, με το τρέχον αποτέλεσμα της παρακολούθησης για αυτό (εξίσωση 4.5). Αν το άθροισμα των τετραγωνικών διαφορών ανάμεσα στην τρέχουσα τεμπλέτα και την προβολή, μέσω του εκτιμηθέντος αφινικού μετασχηματισμού, της αρχικής είναι αρκετά μεγάλος, η παρακολούθηση του σημείου αυτού σταματά.

Οι Comaniciu et al. [12, 13] χρησιμοποιούν χωρικά σταθμισμένα ιστογράμματα χρώματος, υπολογισμένα σε μια ελλειπτικού σχήματος περιοχή, για την αναπαράσταση του αντικειμένου. Η χωρική στάθμιση πραγματοποιείται μέσω μιας ισοτροπικής, κυρτής και μονότονα φθίνουσας συνάρτησης πυρήνα, η οποία αναθέτει χαμηλότερα βάρη στις θέσεις που βρίσκονται μακρύτερα από το κέντρο του υποτιθέμενου μοντέλου. Το γεγονός αυτό καθιστά τη διαδικασία παρακολούθησης πιο εύρωστη και, επιπλέον, επιτρέπει τον ορισμό χωρικά ομαλών συναρτήσεων ομοιότητας ιστογραμμάτων, πράγμα που καθιστά εφικτή την εφαρμογή μεθόδων βελτιστοποίησης που βασίζονται στην κλίση (gradient-based). Η παρακολούθηση πραγματοποιείται μέσω του υπολογισμού της ελλειπτικής περιοχής της εικόνας, της οποίας το ιστόγραμμα παρουσιάζει τη μεγαλύτερη ομοιότητα με το γνωστό ιστόγραμμα του αντικειμένου. Ως συνάρτηση ομοιότητας ιστογραμμάτων, χρησιμοποιείται ο συντελεστής Bhattacharya, του οποίου η μεγιστοποίηση γίνεται επαναληπτικά, μέσω της διαδικασίας mean-shift [35, 36]. Σε κάθε επανάληψη, η εκτίμηση της θέσης του αντικειμένου μετατοπίζεται κατά το διάλυμα mean-shift, το οποίο αποτελεί εκτίμηση της κλίσης της (προσεγγισμένης με πυρήνες) υπό μεγιστοποίηση ποσότητας. Η προτεινόμενη μέθοδος συγκλίνει σε μικρό αριθμό επαναλήψεων, πράγμα που την καθιστά σημαντικά αποδοτικότερη από λύσεις που βασίζονται σε εξαντλητική αναζήτηση.

Οι Nummiaro et al. [15] χρησιμοποιούν ακριβώς την ίδια αναπαράσταση για το αντικείμενο, του οποίου το σχήμα περιγράφεται από μια έλλειψη και η εμφάνιση από ένα χωρικά σταθμισμένο ιστόγραμμα χρώματος. Η παρακολούθηση πραγματοποιείται στοχαστικά, με την εφαρμογή ενός SIR Particle Filter (ενότητα 5.7.7). Το διάλυμα της κατάστασης περιλαμβάνει τη θέση και την

ταχύτητα του κέντρου της έλλειψης, τις διαστάσεις των αξόνων της και ένα συντελεστή κλίμακας για το μέγεθός της. Το θεωρούμενο δυναμικό μοντέλο, που ορίζει και την κατανομή σημαντικότητας του SIR φίλτρου, είναι γραμμικό, με προσθετικό Γκαουσιανό θόρυβο, και περιγράφει μια συμπεριφορά κίνησης με σταθερή ταχύτητα και συντελεστή κλιμάκωσης μεγέθους. Η συνάρτηση πιθανοφάνειας των παρατηρήσεων, που ορίζει το μοντέλο παρατήρησης της μεθόδου, λαμβάνεται ως μια Γκαουσιανή κατανομή, μηδενικής μέσης τιμής, με όρισμα την απόσταση Bhattacharya ανάμεσα στο ιστόγραμμα-μοντέλο του αντικειμένου και αυτό που αντιστοιχεί στη δεδομένη υπόθεση (σωμάτιο). Για τη βελτίωση της ευρωστίας της μεθόδου, γίνεται ενημέρωση του χρωματικού μοντέλου του αντικειμένου κατά τη διάρκεια της εκτέλεσης της διαδικασίας, χρησιμοποιώντας μόνο παρατηρήσεις από διαστήματα επιτυχούς παρακολούθησης. Επιπλέον, προτείνονται τεχνικές αυτόματης αρχικοποίησης της διαδικασίας. Συνολικά, η προτεινόμενη μέθοδος παρακολούθησης μπορεί να θεωρηθεί ως το στοχαστικό ανάλογο της μεθόδου των Comaniciu et al. [12, 13]. Η εξέταση πολλαπλών υποθέσεων με βάση ένα υποτιθέμενο δυναμικό μοντέλο κατά την εφαρμογή του φίλτρου, σε συνδυασμό με τη δυναμική ενημέρωση του χρωματικού μοντέλου του αντικειμένου, καθιστούν την προτεινόμενη μέθοδο περισσότερο εύρωστη, τόσο σε περιπτώσεις επικαλύψεων (occlusions), όσο και σε μεταβολές των συνθηκών φωτισμού.

2.2.3 Παρακολούθηση περιγράμματος/σιλουέτας

Το υπό παρακολούθηση αντικείμενο έχει συχνά περίπλοκο σχήμα, το οποίο δεν μπορεί να περιγραφεί με ακρίβεια από απλά γεωμετρικά σχήματα. Οι μέθοδοι παρακολούθησης περιγράμματος/σιλουέτας έχουν ως στόχο τον ακριβή προσδιορισμό της περιοχής του επιπέδου της εικόνας που ορίζει η προβολή του υπό παρακολούθηση αντικειμένου.

Μια συνήθης πρακτική για την παρακολούθηση περιγράμματος είναι η θεώρηση κάποιου μοντέλου *ενεργών καμπυλών* (active contours) [37]. Οι ενεργές καμπύλες είναι καμπύλες του επιπέδου που μπορούν να κινούνται και να παραμορφώνονται κατά τρόπο που ελαχιστοποιεί την ενέργειά τους. Η τελευταία ορίζεται από συναρτήσεις δυναμικού που μοντελοποιούν αφενός *εσωτερικές* δυνάμεις της καμπύλης, ελέγχοντας ιδιότητες όπως η ελαστικότητα, η ομαλότητα και η συνέχεια αυτής, και αφετέρου *εξωτερικές* δυνάμεις που ασκούνται στην καμπύλη από την εικόνα, οι οποίες μπορεί, για παράδειγμα, να έλκουν την καμπύλη πάνω σε συγκεκριμένα χαρακτηριστικά εικόνας, όπως ακμές. Στην περίπτωση της κινούμενης εικόνας, το συναρτησιακό ενέργειας της καμπύλης είναι χρονικά μεταβαλλόμενο. Η παρακολούθηση μπορεί να πραγματοποιηθεί σε μια πλαίσιο προς πλαίσιο βάση, όπου κατά την τρέχουσα χρονική στιγμή θεωρείται γνωστή η θέση της ενεργού καμπύλης για το προηγούμενο πλαίσιο, η οποία βρίσκεται σε κατάσταση ισορροπίας για τη συνάρτηση δυναμικού την προηγούμενη χρονική στιγμή. Η τρέχουσα συνάρτηση δυναμικού ορίζει μια νέα κατάσταση ισορροπίας και, μέσω των δυνάμεων που ορίζει, έλκει την καμπύλη προς αυτήν. Η εξέλιξη της καμπύλης από την προηγούμενη θέση ισορροπίας της στην τρέχουσα μπορεί να μοντελοποιηθεί σαν μια δυναμική διαδικασία, η οποία περιγράφεται από μερικές διαφορικές εξισώσεις που προκύπτουν από εφαρμογή των αρχών της Λαγκρανζιανής μηχανικής. Σε αυτό το πλαίσιο, οι Terzopoulos και Szeliski [18] ορίζουν το διάνυσμα κατάστασης για την αναπαράσταση του περιγράμματος του αντικειμένου με βάση τις παραμέτρους που περιγράφουν τη θέση και τη δυναμική συμπεριφορά των σημείων ελέγχου της ενεργού καμπύλης. Η δυναμική συμπεριφορά των σημείων αυτών περιγράφεται από ένα μοντέλο ελατηρίου, το οποίο τα έλκει πάνω σε χαρακτηριστικά της εικόνας με δύναμη ανάλογη της σταθεράς σκληρότητάς του. Ορίζοντας κατάλληλες κατανομές Gibbs, το πρόβλημα ελαχιστοποίησης της ενέργειας της καμπύλης στο τρέχον βήμα διατυπώνεται ως υπολογισμός και μεγιστοποίηση της ύστερης πιθανότητας της καμπύλης, δεδομένων των μέχρι τώρα παρατηρήσεων, και επιλύεται με εφαρμογή του φίλτρου Kalman.

Μία άλλη προσέγγιση παρακολούθησης περιγράμματος με αβεβαιότητα δίνεται από τον αλγό-

ρυθμο CONDENSATION (Conditional Density Propagation) των Isard και Blake [17]. Το περιγράμμα του αντικειμένου αναπαρίσταται από B-splines και η παρακολούθησή του γίνεται με ένα SIR Particle Filter. Όπως και στο [18], το θεωρούμενο δυναμικό μοντέλο, που ορίζει και την κατανομή σημαντικότητας του φίλτρου, περιγράφει μια δεύτερης τάξης διαδικασία και, κατόπιν κατάλληλης επαύξησης του διανύσματος κατάστασης, περιγράφεται από ένα σύνολο γραμμικών εξισώσεων. Οι παράμετροι του μοντέλου εκτιμούνται από ένα στάδιο εκπαίδευσης. Για τον ορισμό της συνάρτησης πιθανοφάνειας των παρατηρήσεων, και, κατ' επέκταση, του μοντέλου παρατήρησης, χρησιμοποιούνται σημεία της εικόνας που ορίζονται από την τομή κάθετων στη δεδομένη καμπύλη (υπό εξέταση υπόθεση) ευθειών με τις ακμές της παρατηρούμενης εικόνας. Για κάθε τέτοια κάθετη ευθεία, επιλέγεται εκείνο το σημείο τομής της με ακμή που απέχει την ελάχιστη απόσταση από το σημείο τομής της ευθείας με τη δεδομένη καμπύλη. Το φίλτρο αναθέτει υψηλότερα βάρη στάθμισης στις υποθέσεις (καμπύλες) για τις οποίες το άθροισμα των τετραγώνων αυτών των αποστάσεων είναι χαμηλότερο. Ο προτεινόμενος αλγόριθμος αποτελεί την πρώτη εφαρμογή ενός Particle Filter στο πρόβλημα της οπτικής παρακολούθησης.

Η παρακολούθηση σιλουέτας είναι στενά συνδεδεμένη με το πρόβλημα της κατάτμησης βίντεο. Το ζητούμενο στην τελευταία είναι η διαμέριση του χωρο-χρονικού όγκου που αποτελεί το πεδίο ορισμού της κινούμενης εικόνας σε ομογενείς, ως προς κάποιο κριτήριο, περιοχές. Αν κάθε τρισδιάστατη περιοχή μιας τέτοιας κατάτμησης απαιτηθεί να αντιστοιχεί ακριβώς σε ένα αντικείμενο της ορατής σκηνής, τότε η επίλυση αυτού του προβλήματος παρέχει αυτόματα μια λύση και στο πρόβλημα της παρακολούθησης: η σιλουέτα του αντικειμένου που παρακολουθούμε και η χρονική εξέλιξή της περιγράφονται ακριβώς από μια περιοχή της κατάτμησης του βίντεο. Στο γενικό πλαίσιο αυτής της παρατήρησης, οι Brox και Malik [38] προτείνουν μια μέθοδο κατάτμησης των κινούμενων αντικειμένων σε μια ακολουθία βίντεο, μέσω της μη εποπτευόμενης ταξινόμησης (unsupervised clustering) των τροχιών που προκύπτουν από τη μακροπρόθεσμη παρακολούθηση ενός πυκνού συνόλου σημείων. Οι τροχιές μπορεί να είναι αυθαίρετα πυκνές χωρικά, είναι ασύγχρονες (καλύπτουν, εν γένει, διαφορετικά χρονικά διαστήματα) και υπολογίζονται από τη μέθοδο που υλοποιείται στο [39], η οποία στηρίζεται στη μέθοδο εκτίμησης οπτικής ροής που αναπτύσσεται στο [40]. Στο σύνολο των τροχιών, ορίζεται μια διμελής σχέση συγγένειας που συνδέει κάθε ζεύγος τροχιών που έχουν τουλάχιστον ένα κοινό πλαίσιο στο πεδίο ορισμού τους. Η απόσταση ανάμεσα στα μέλη κάθε τέτοιου ζεύγους τροχιών, σε μια δεδομένη χρονική στιγμή, δίνεται από ένα κανονικοποιημένο μέτρο χωρικής απόστασης των σημείων των τροχιών, οι θέσεις των οποίων υπολογίζονται στα άκρα ενός μικρού χρονικού παραθύρου γύρω από τη θεωρούμενη χρονική στιγμή, πολλαπλασιασμένο με τη μέση (χωρική) Ευκλείδεια απόσταση των δύο τροχιών, υπολογισμένη σε όλο το κοινό παράθυρο ορισμού τους. Η (χρονικά ανεξάρτητη) απόσταση ανάμεσα σε δύο σχετιζόμενες τροχιές, ορίζεται ως η μέγιστη, επί του κοινού διαστήματος ορισμού, απόσταση μεταξύ τους. Μετατρέποντας τις αποστάσεις των ζευγών σε τιμές ομοιότητας, κατασκευάζεται ο τετραγωνικός πίνακας συγγένειας (affinity matrix) του συνόλου των τροχιών και τροφοδοτείται σε έναν αλγόριθμο φασματικής ταξινόμησης (spectral clustering), ο οποίος επιβάλλει έναν περιορισμό χωρικής κανονικότητας (spatial regularity) για την κατάτμηση, ενώ, επιπλέον, επιτρέπει την αυτόματη επιλογή μοντέλου (πλήθος κλάσεων της κατάτμησης). Χάρη στη μεταβατικότητα του γράφου συγγένειας, η προτεινόμενη μέθοδος χειρίζεται με επιτυχία καταστάσεις μερικής απόκρυψης ενός αντικειμένου, χωρίς να απαιτείται κάποια διαδικασία μετεπεξεργασίας των τροχιών: δύο μη συγγενικές τροχιές (με κενή τομή πεδίων ορισμού) μπορεί πάλι να ανατεθούν στην ίδια ομάδα της ταξινόμησης, εφόσον υπάρχει μονοπάτι που τις συνδέει στο γράφο της σχέσης συγγένειας.

Κεφάλαιο 3

Κατάτμηση εικόνας με βάση το χρώμα

Στο κεφάλαιο αυτό, παρουσιάζουμε μια απλή μέθοδο κατάτμησης μιας εικόνας σε αντικείμενα με βάση το χρώμα. Η κατάτμηση μιας εικόνας στα αντικείμενα που περιέχει διατυπώνεται ευθέως σαν ένα πρόβλημα ταξινόμησης (classification, clustering) των εικονοστοιχείων της: μια έγκυρη κατάτμηση ορίζει μια διαμέριση των εικονοστοιχείων της εικόνας, όπου κάθε ομάδα της διαμέρισης αντιστοιχεί, ιδανικά, σε ένα αντικείμενο της εικόνας. Στόχος μας δεν είναι να καλύψουμε το εύρος των προσεγγίσεων που έχουν εφαρμοστεί στο γενικό πρόβλημα της κατάτμησης εικόνων. Αντί αυτού, περιγράφουμε τη μέθοδο που χρησιμοποιήθηκε στην εφαρμογή οπτικής παρακολούθησης (Κεφάλαιο 6) που αναπτύξαμε. Δεδομένου ενός στατιστικού μοντέλου για το χρώμα κάθε αντικειμένου σε μια εικόνα, διατυπώνουμε το πρόβλημα ταξινόμησης των εικονοστοιχείων αυτής (κατάτμηση) σαν ένα πρόβλημα Μπεϋζιανής θεωρίας απόφασης.

Αρχικά, παρουσιάζουμε συνοπτικά κάποια εισαγωγικά θέματα που αφορούν τις εικόνες, το σχηματισμό τους, τη μαθηματική περιγραφή τους, τη χρωματική πληροφορία που περιέχουν και τον τρόπο αναπαράστασής τους σε ψηφιακά μέσα. Μια αναλυτική και ολοκληρωμένη παρουσίαση αυτών των θεμάτων μπορεί να βρεθεί στις εργασίες [2] και [3]. Στη συνέχεια, περιγράφουμε τα βασικά στατιστικά εργαλεία που χρησιμοποιούνται για τη μοντελοποίηση κατανομών χρώματος και παρουσιάζουμε μια απλή μέθοδο κατάτμησης εικόνας με βάση το χρώμα σε επίπεδο εικονοστοιχείου, δεδομένων των στατιστικών κατανομών των χρωμάτων των αντικειμένων που απεικονίζονται.

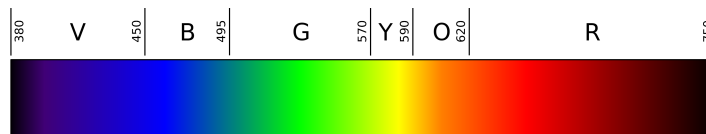
3.1 Εικόνες και χρώμα

Η εικόνα μιας σκηνής προκύπτει από τη συλλογή/συσσώρευση σε κάποιο αισθητήρα ενέργειας από την ηλεκτρομαγνητική ακτινοβολία που εκπέμπεται ή ανακλάται από τα αντικείμενα της σκηνής. Στο ανθρώπινο σύστημα όρασης, οι εικόνες σχηματίζονται στον αμφιβληστροειδή χιτώνα του οφθαλμού και, μετά από επεξεργασία/ερμηνεία από τον εγκέφαλο, δίνουν στον άνθρωπο τη δυνατότητα της οπτικής αντίληψης του χώρου γύρω του. Εκτός από τους βιολογικούς οπτικούς αισθητήρες, έχουν αναπτυχθεί και τεχνητοί, οι οποίοι επιτρέπουν την καταγραφή, αποθήκευση και μελλοντική ανάκτηση εικόνων. Μάλιστα, παρότι το ανθρώπινο σύστημα όρασης επιτρέπει την οπτική αντίληψη μιας στενής ζώνης του ηλεκτρομαγνητικού φάσματος (μήκη κύματος μεταξύ 400nm και 700nm), οι τεχνητοί αισθητήρες παρέχουν επιπλέον τη δυνατότητα καταγραφής μηκών κύματος εκτός του ορατού φάσματος. Τα σήματα αυτά, με κατάλληλο μετασχηματισμό, μπορούν να μετατραπούν σε ορατές εικόνες και να μελετηθούν εποπτικά από τον άνθρωπο, γε-

γονός που βρίσκει πολλές εφαρμογές σε ιατρικές και δορυφορικές απεικονίσεις, στη μελέτη του διαστήματος, κλπ.

Σε επίπεδο μαθηματικής αναπαράστασης, οι έγχρωμες εικόνες περιγράφονται από μαθηματικές συναρτήσεις, οι οποίες κωδικοποιούν την κατανομή της φωτεινής ενέργειας για τα διάφορα σημεία της εικόνας στα διάφορα μήκη κύματος του ορατού φάσματος. Συγκεκριμένα, μια έγχρωμη εικόνα τυπικά περιγράφεται από μια συνάρτηση δύο μεταβλητών θέσης με τιμές που κωδικοποιούν (σε κάθε θέση) το φασματικό περιεχόμενο της προσπίπτουσας ηλεκτρομαγνητικής ακτινοβολίας. Οι συνεχείς συναρτήσεις, με τα αντίστοιχα μαθηματικά εργαλεία (ανάλυση, διαφορικές εξισώσεις), παίζουν σημαντικό ρόλο στη θεωρητική ανάλυση και επεξεργασία των εικόνων, καθώς και στην εξαγωγή σημαντικών χαρακτηριστικών από αυτές. Παρόλα αυτά, στην πράξη εργαζόμαστε με διακριτές συναρτήσεις, καθώς μας ενδιαφέρει η ψηφιακή επεξεργασία τους με τη χρήση ηλεκτρονικών υπολογιστών.

Με τον όρο *χρώμα* αναφερόμαστε σε ένα συγκεκριμένο μήκος κύματος ηλεκτρομαγνητικής ακτινοβολίας εντός του ορατού φάσματος. Το ορατό φάσμα για τον άνθρωπο οριοθετείται από τη ζώνη μεταξύ μήκους κύματος $400nm$ και $700nm$ (Σχήμα 3.1). Συγκεκριμένα, τα $400nm$ αντιστοιχούν στο ιώδες χρώμα και τα $700nm$ στο ερυθρό. Μήκη κύματος μικρότερα από τα $400nm$ αντιστοιχούν στην υπεριώδη και μεγαλύτερα των $700nm$ στην υπέρυθη ακτινοβολία. Εάν το φασματικό περιεχόμενο της ακτινοβολίας αποτελείται μόνο μία συχνότητα (οπότε η φασματική κατανομή είναι ουσιαστικά μια συνάρτηση Dirac σε κάποιο μήκος κύματος), τότε έχουμε *μονοχρωματική* ακτινοβολία, το χρώμα της οποίας χαρακτηρίζεται ως καθαρό. Αν συνυπάρχουν πολλές συχνότητες στο φάσμα της ακτινοβολίας, τότε αυτή χαρακτηρίζεται *πολυχρωματική* και το αντίστοιχο χρώμα θεωρείται σύνθετο.

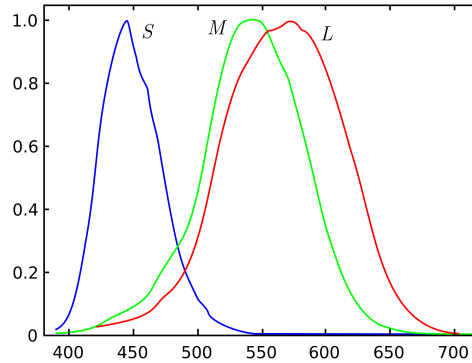


Σχήμα 3.1: Το ορατό φάσμα (μήκη κύματος σε nm).

Ο αισθητήρας του ανθρώπινου συστήματος όρασης, ο αμφιβληστροειδής χιτώνας (ρέτινα), βρίσκεται στο πίσω μέρος του ανθρώπινου οφθαλμού και διαθέτει τέσσερα είδη φωτοαισθητήρων, τα ραβδία και τρία είδη κωνίων, οι οποίοι είναι διάσπαρτοι πάνω στον αμφιβληστροειδή. Τα ραβδία είναι υπεύθυνα για την αντίληψη του αμυδρού φωτός (σκοτοπική όραση), η οποία στερείται χρώματος. Τα κωνία είναι υπεύθυνα για την έγχρωμη όραση και το καθένα χαρακτηρίζεται από μια πολυχρωματική (ζωνοπερατή) φασματική απόκριση (Σχήμα 3.2). Παρότι οι φασματικές αποκρίσεις των τριών ειδών κωνίων είναι σε ένα βαθμό επικαλυπτόμενες, είναι αρκούντως διαφορετικές:

- Τα *S*-κωνία είναι ευαίσθητα σε φωτόνια “βραχέως” (Short) μήκους κύματος και παρουσιάζουν μέγιστη απόκριση σε μήκος κύματος περίπου ($420nm$): μέγιστη ευαισθησία στο μπλε φως.
- Τα *M*-κωνία είναι ευαίσθητα σε φωτόνια “μεσαίου” (Medium) μήκους κύματος και παρουσιάζουν μέγιστη απόκριση σε μήκος κύματος περίπου ($530nm$): μέγιστη ευαισθησία στο πράσινο φως.
- Τα *L*-κωνία είναι ευαίσθητα σε φωτόνια “μακρού” (Long) μήκους κύματος και παρουσιάζουν μέγιστη απόκριση σε μήκος κύματος περίπου ($560nm$): μέγιστη ευαισθησία στο κόκκινο φως.

Η διαφορετική απόκριση που επάγεται από την αλληλεπίδραση του φωτός με τους τρεις διαφορετικούς τύπους αισθητήρων οδηγεί ουσιαστικά στην ανάλυση του προσπίπτοντος φωτός σε τρεις συνιστώσες, με κάθε μια να αντιστοιχεί σε ένα είδος κωνίου (χρώμα). Η λειτουργία αυτή του αισθητήρα του ανθρώπινου συστήματος όρασης είναι η αιτία της έγχρωμης όρασης.



Σχήμα 3.2: Οι (κανονικοποιημένες) φασματικές αποκρίσεις (ευαισθησίες) των τριών ειδών κωνίων του ανθρώπινου οφθαλμού.

3.1.1 Χώροι χρώματος

Η δυνατότητα ακριβούς και συνεπούς περιγραφής και αναπαράστασης των χρωμάτων κρίνεται αναγκαία σε πληθώρα σεναρίων. Η δυνατότητα αναπαραγωγής ενός χρώματος με ακρίβεια είναι μεγάλης εμπορικής σημασίας, καθώς πολλά προϊόντα είναι στενά συνδεδεμένα με συγκεκριμένα χρώματα. Επίσης, η συνέπεια του χρώματος είναι αναγκαία σε περιβάλλοντα διεπαφής χρήστη (user interfaces) και η συμβατότητα των χρωμάτων είναι απαραίτητη ανάμεσα σε διαφορετικές συσκευές απεικόνισης, όπως ανάμεσα σε οθόνες και εκτυπωτές. Η επιστήμη μέτρησης και κατανόησης του χρώματος ονομάζεται *χρωματομετρία* (colorimetry).

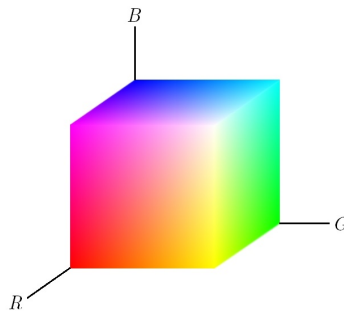
Ένα χρώμα αναπαρίσταται σαν ένα διάνυσμα, συνήθως τριδιάστατο, σε κάποιο κατάλληλο διανυσματικό χώρο χρώματος. Οι χώροι αυτοί κατηγοριοποιούνται σε γραμμικούς και μη γραμμικούς. Στους γραμμικούς χώρους χρώματος, η μείξη χρωμάτων μοντελοποιείται από τη γραμμική υπέρθεση των αντίστοιχων διανυσμάτων αναπαράστασης. Σε κάθε μία από τις δύο κατηγορίες, η βάση των χώρων ποικίλλει (ποιες συνιστώσες του χρώματος θεωρούνται βασικές). Πολλοί χώροι χρώματος έχουν προταθεί και μελετηθεί για τις ιδιότητές τους, ενώ πολλοί έχουν εμπνευστεί και επηρεαστεί από τις θεωρίες αντίληψης χρώματος. Για παράδειγμα, σύμφωνα με τη θεωρία της τριχρωμίας (trichromacy), που βασίζεται σε όσα αναφέραμε παραπάνω (Σχήμα 3.2), κάθε άγνωστο χρώμα μπορεί να παρασταθεί ως γραμμικός συνδυασμός τριών πρωταρχικών χρωμάτων (primary colors) που είναι γραμμικά ανεξάρτητα μεταξύ τους. Η ιδέα αυτή αναφέρεται ως η *θεωρία των τριών ερεθισμάτων* (tristimulus theory) [2].

Παρότι οι γραμμικοί χώροι χρώματος είναι πιο εύχρηστοι και, από μαθηματικής άποψης, λιγότερο πολύπλοκοι από τους μη γραμμικούς, έχουν ένα σημαντικό μειονέκτημα. Συγκεκριμένα, σε αυτούς τους χώρους χρώματος, η απόσταση δύο χρωμάτων, όπως αυτή καθορίζεται από τις απεικονίσεις τους στο χώρο χρώματος (Ευκλείδεια απόσταση των διανυσμάτων αναπαράστασης), δεν ανταποκρίνεται στη διαφορά των χρωμάτων που αντιλαμβάνεται ο άνθρωπος. Δηλαδή, δύο χρώματα που ο άνθρωπος θεωρεί κοντινά μπορεί να έχουν τρισδιάστατες απεικονίσεις δυσανάλογα διαφορετικές, και αντίστροφα. Για το λόγο αυτό, έχει γίνει εκτεταμένη έρευνα για τον ορισμό τέτοιων χώρων που ανταποκρίνονται όσο το δυνατόν καλύτερα στην ανθρώπινη αντίληψη περί χρωματικής διαφοράς.

Γραμμικοί χώροι χρώματος

Τα αναγκαία standards για τα διάφορα χρωματικά συστήματα αναπτύσσονται και τυποποιούνται από τη διεθνή επιτροπή CIE (Commission Internationale de l'Éclairage, International Commission on Illumination). Οι χώροι CIE RGB και CIE XYZ είναι οι πιο διαδεδομένοι γραμμικοί χώροι χρώματος και ο ένας προκύπτει από τον άλλο μέσω γραμμικού μετασχηματισμού.

Ο χώρος RGB προκύπτει χρησιμοποιώντας μονοχρωματικά πρωταρχικά χρώματα, συγκεκριμένα κόκκινο R στα $645.16nm$, πράσινο G στα $526.32nm$ και μπλε B στα $444.44nm$. Έτσι, σε αυτόν το χώρο κάθε χρώμα αναλύεται σύμφωνα με την περιεκτικότητά του όσον αφορά τρεις συνιστώσες. Αξίζει να σημειωθεί ότι έχει αποδειχθεί πειραματικά πως, για να ταιριάζει οπτικά κάποιο πραγματικό χρώμα με μια υπέρθεση κόκκινου, πράσινου και μπλε, ενδέχεται να απαιτηθεί η χρήση αρνητικών συντελεστών στο μείγμα: οι αντίστοιχες συναρτήσεις χρωματικού ταιριάσματος (Color Matching Functions, CMFs) λαμβάνουν και αρνητικές τιμές [2]. Αυτό συμβαίνει γιατί τα πρωταρχικά χρώματα του χώρου RGB αντιστοιχούν σε πραγματικά χρώματα. Θεωρώντας τις συνιστώσες (R, G, B) ενός χρώματος σαν καρτεσιανές συνιστώσες σε έναν Ευκλείδειο τριδιάστατο χώρο, ορίζεται ένας τριδιάστατος όγκος σε αυτό το γεωμετρικό χώρο. Στο μοντέλο RGB, ο όγκος αυτός αντιστοιχεί στον (κατά σύμβαση) μοναδιαίο κύβο, όπου κάθε συνιστώσα λαμβάνει μια τιμή από το διάστημα $[0, 1]$. Το μαύρο αντιστοιχεί στο σημείο $(0, 0, 0)$, κορυφή του κύβου που συμπίπτει με την αρχή των αξόνων, και το μέγιστης έντασης (αχρωματικό) λευκό στην κορυφή $(1, 1, 1)$. Η γεωμετρική αυτή απεικόνιση του χώρου RGB φαίνεται στο Σχήμα 3.3.



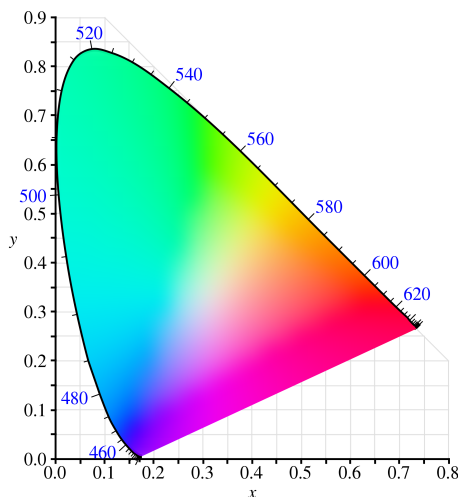
Σχήμα 3.3: Απεικόνιση του χώρου RGB στο μοναδιαίο τριδιάστατο κύβο. Οι άξονες του καρτεσιανού συστήματος συντεταγμένων αντιστοιχούν σε διαφορετικές εντάσεις των (μονοχρωματικών) πρωταρχικών χρωμάτων (R, G, B) .

Προς αποφυγή των αρνητικών τιμών στις συναρτήσεις χρωματικού ταιριάσματος του χώρου RGB, δημιουργήθηκε ο χώρος XYZ, με γραμμικό μετασχηματισμό του χώρου RGB. Τα πρωταρχικά χρώματα του χώρου αυτού είναι φανταστικά, δηλαδή το φάσμα τους εμφανίζει αρνητικές τιμές για κάποια μήκη κύματος [3]. Παρόλα αυτά, έτσι εξασφαλίζεται το γεγονός ότι θα μπορεί πάντα ο άνθρωπος να αντιστοιχίζει κάποιο πραγματικό χρώμα που παρατηρεί σε ένα γραμμικό συνδυασμό στη βάση του χώρου XYZ, στον οποίο οι συντελεστές θα είναι πάντα θετικοί. Οι συντεταγμένες (X, Y, Z) ενός μετρηθέντος χρώματος στο χώρο XYZ ονομάζονται tristimulus τιμές. Για την εποπτεία του χώρου σε δύο διαστάσεις, οι tristimulus τιμές κανονικοποιούνται στις:

$$(x, y, z) = \left(\frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z} \right)$$

Οι κανονικοποιημένες συντεταγμένες ανήκουν σε ένα επίπεδο, αφού $x + y + z = 1$. Η παράσταση του χρώματος στο επίπεδο των ανεξάρτητων κανονικοποιημένων χρωματικών συνιστωσών (x, y) είναι γνωστή ως το CIE xy χρωματικό διάγραμμα (chromaticity diagram) (Σχήμα 3.4). Στο

διάγραμμα αυτό απεικονίζονται όλα τα χρώματα του φάσματος. Τα σημεία της περιφέρειας αντιστοιχούν στα μονοχρωματικά χρώματα, ενώ το (αχρωματικό) λευκό σημείο είναι στο κέντρο του διαγράμματος: ο κορεσμός των χρωμάτων αυξάνει με την απόσταση από το κέντρο. Σημειώνουμε ότι ο χώρος XYZ, όπως και ο χώρος RGB, παρά την ευρεία χρήση και μελέτη του στην υπολογιστική όραση και τη γραφική, θεωρείται ξεπερασμένος από επαγγελματίες χρωματομέτρους, μιας και δεν μπορεί να αποφύγει τα γενικά μειονεκτήματα των γραμμικών χώρων.



Σχήμα 3.4: Το χρωματικό διάγραμμα CIE xy .

Τέλος, ένας συμπληρωματικός χώρος του RGB είναι ο CMY. Η βάση του περιλαμβάνει τα χρώματα κυανό C , magenta M και κίτρινο Y , τα οποία ορίζονται ως τα συμπληρωματικά του κόκκινου, πράσινου και μπλε αντίστοιχα:

$$(C, M, Y) = (1 - R, 1 - G, 1 - B)$$

Τα χρώματα βάσης C , M , Y , μαζί με τα συμπληρωματικά τους, R , G , B , το λευκό και το μαύρο αποτελούν τις κορυφές του μοναδιαίου κύβου χρωμάτων του Σχήματος 3.3. Ο χώρος αυτός χρησιμοποιείται σε συσκευές έγχρωμης εκτύπωσης (εκτυπωτές, φωτοτυπικά). Ο λόγος είναι ότι το χρώμα των χρωστικών ουσιών που αντιλαμβάνομαστε ορίζεται από τις συνιστώσες του λευκού φωτός που αυτές αντανακλούν, ή, ισοδύναμα, από αυτές που απορροφούν. Για παράδειγμα, μια κόκκινη χρωστική αντανακλά το κόκκινο φως, ενώ απορροφά το πράσινο και το μπλε. Η μείξη χρωμάτων που επιτυγχάνεται με την επίθεση χρωστικών ουσιών πάνω σε χαρτί είναι αφαιρετική (subtractive). Καθώς με μείξη χρωστικών ουσιών δεν είναι εύκολο να επιτευχθεί καλή ποιότητα μαύρου, συνήθως χρησιμοποιείται το μαύρο σαν τέταρτο χρώμα και η ενισχυμένη αυτή βάση ορίζει το χώρο CMYK.

Μη γραμμικοί χώροι χρώματος

Ο άνθρωπος, όσον αφορά το χρώμα, αντιλαμβάνεται κυρίως τρεις παραμέτρους: την απόχρωση (hue), τον κορεσμό (saturation) και την φωτεινότητα/ένταση (brightness/intensity). Η απόχρωση σχετίζεται με το κυρίαρχο μήκος κύματος και είναι η ιδιότητα που μεταβάλλεται καθώς μετακινούμαστε από κάποιο κύριο χρώμα προς ένα άλλο. Ο κορεσμός έχει να κάνει με το βαθμό καθαρότητας του χρώματος: το μονοχρωματικό φως θεωρείται πλήρως καθαρό, σε αντίθεση με το λευκό φως, που θεωρείται ότι έχει μηδενική καθαρότητα. Τέλος, η φωτεινότητα, που είναι το κύριο οπτικό

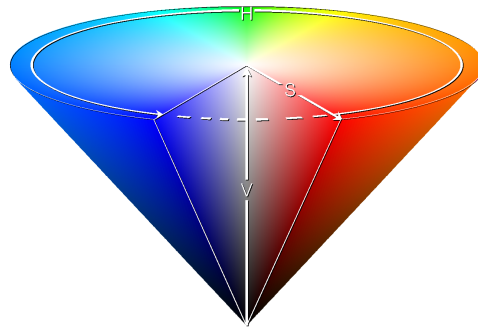
ερέθισμα, αυξάνει καθώς πηγαίνουμε από το μαύρο στο λευκό, και σχετίζεται με την ένταση της φωτεινής ακτινοβολίας. Με βάση αυτές τις διαπιστώσεις έχουν οριστεί οι μη γραμμικοί χώροι HSI (Hue, Saturation, Intensity) και HSV (Hue, Saturation, Value) μέσω μη γραμμικών μετασχηματισμών του χώρου RGB [2, 3]. Για παράδειγμα, αν $m = \min\{R, G, B\}$ και $M = \max\{R, G, B\}$, ο χώρος HSV ορίζεται με βάση τον ακόλουθο μετασχηματισμό του μοναδιαίου κύβου RGB σε έναν κώνο:

$$H = \begin{cases} 60^\circ \left(\frac{G-B}{M-m} \bmod 6 \right) & \text{αν } R = M \\ 60^\circ \left(\frac{B-R}{M-m} + 2 \right) & \text{αν } G = M \\ 60^\circ \left(\frac{R-G}{M-m} + 4 \right) & \text{αν } B = M \end{cases}$$

$$S = \begin{cases} 0 & \text{αν } M - m = 0 \\ \frac{M-m}{M} & \text{διαφορετικά} \end{cases}$$

$$V = M$$

Το αποτέλεσμα του μετασχηματισμού φαίνεται στο Σχήμα 3.5. Ο κατακόρυφος άξονας αντιστοιχεί στη συνιστώσα V που ανταποκρίνεται στη φωτεινότητα/ένταση, η γωνία γύρω από τον κατακόρυφο άξονα στη συνιστώσα H (απόχρωση) και η κάθετη απόσταση από το τρέχον σημείο στον κατακόρυφο άξονα στη συνιστώσα S (κορεσμός).



Σχήμα 3.5: Ο χώρος χρώματος HSV.

Όπως αναφέραμε παραπάνω, το βασικό μειονέκτημα των γραμμικών χώρων χρώματος έχει να κάνει με την ασυμφωνία ανάμεσα στις χρωματικές (Ευκλείδειες) αποστάσεις στους χώρους αυτούς και την αντίληψη της χρωματικής διαφοράς από τον άνθρωπο. Για τη σύγκριση χρωμάτων με τη χρήση της Ευκλείδειας απόστασης, υπάρχει η ανάγκη κατασκευής *ομοιόμορφων* (uniform) χώρων χρώματος, στους οποίους η συνήθης απόσταση μεταξύ δύο σημείων αποτελεί πιστή ένδειξη της ανθρώπινης σημασίας της διαφοράς μεταξύ των δύο χρωμάτων. Γύρω από κάθε σημείο ενός χώρου χρώματος μπορούμε να ορίσουμε μια περιοχή χρωμάτων που δεν είναι διακριτά από τον άνθρωπο από το χρώμα αναφοράς. Μια τέτοια περιοχή ορίζεται από όλα τα χρώματα που απέχουν από το σημείο αναφοράς απόσταση μικρότερη από αυτήν που αντιστοιχεί στο κατώφλι της *οριακά διακρινόμενης διαφοράς* (just noticeable difference). Ένα επιθυμητό χαρακτηριστικό ενός ομοιόμορφου χώρου είναι η ισοτροπία όσον αφορά τις περιοχές αυτές: το σχήμα και το μέγεθος μιας τέτοιας περιοχής πρέπει να είναι, ιδανικά, ανεξάρτητο της θέσης του σημείου-χρώματος αναφοράς.

Οι πιο διαδεδομένοι ομοιόμορφοι χώροι είναι οι CIE LAB, CIE LUV και CIE UVW. Ο

πρώτος είναι σήμερα ο πιο συχνά χρησιμοποιούμενος ομοιόμορφος χώρος αναπαράστασης χρώματος. Αν (X, Y, Z) η αναπαράσταση ενός χρώματος στο χώρο CIE XYZ (tristimulus τιμές) και (X_n, Y_n, Z_n) οι tristimulus τιμές που αντιστοιχούν στο λευκό χρώμα αναφοράς, τότε οι συνιστώσες L^* , a^* , b^* του χρώματος στο χώρο CIE LAB δίνονται από τον ακόλουθο μη γραμμικό μετασχηματισμό:

$$\begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16, \\ a^* &= 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right], \\ b^* &= 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right], \end{aligned}$$

όπου:

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{αν } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{διαφορετικά} \end{cases}$$

Η συνιστώσα L^* αντιστοιχεί στη φωτεινότητα (lightness), η συνιστώσα a^* στη διαφορά μεταξύ κόκκινης και πράσινης συνιστώσας και η συνιστώσα b^* στη διαφορά μεταξύ κίτρινης και μπλε συνιστώσας. Σημειώνουμε ότι ο χώρος CIE LAB, όπως έχει αποδειχθεί από εμπειρικά πειράματα, εμφανίζει με επιτυχία τα επιθυμητά χαρακτηριστικά ομοιομορφίας ως προς την ανθρώπινη αντίληψη. Ομοιόμορφες μεταβολές στις συνιστώσες L^* , a^* και b^* αντιστοιχούν σε ομοιόμορφες μεταβολές στο χρώμα που αντιλαμβάνεται ο άνθρωπος. Έτσι, η σύγκριση χρωμάτων, θεωρώντας τα ως σημεία ενός τριδιάστατου Ευκλείδειου χώρου με συντεταγμένες L^* , a^* και b^* , μέσω της Ευκλείδειας απόστασης, προσεγγίζει με επιτυχία τη διαφορά χρωμάτων που αντιλαμβάνεται ο άνθρωπος.

3.1.2 Κανονικοποίηση χρωματικών τιμών

Οι χρωματικές τιμές που χρησιμοποιούνται για την κατασκευή στατιστικών μοντέλων χρώματος μπορεί να είναι απλώς τριδιάστατα διανύσματα σε κάποιο χώρο χρώματος που περιγράψαμε παραπάνω. Μια συνήθης πρακτική που εφαρμόζεται σε σενάρια κατασκευής μοντέλων χρωματικών κατανομών και, γενικά, σε εργασίες που περιλαμβάνουν σύγκριση χρωμάτων, είναι η προβολή των χρωματικών τιμών σε ένα χώρο δύο διαστάσεων. Η προβολή αυτή ονομάζεται *κανονικοποίηση χρώματος* (color normalization) και αποσκοπεί στην αφαίρεση ενός βαθμού ελευθερίας από αυτούς που γεννούν το σύνολο των δυνατών χρωμάτων στο χώρο αναπαράστασης.

Όπως είναι φανερό από τη συζήτηση που προηγήθηκε, το χρώμα ενός αντικειμένου σε μια εικόνα, πλήρως περιγραφόμενο από μια κατανομή τριδιάστατων διανυσμάτων σε κάποιο χώρο χρώματος, εξαρτάται από τις συνθήκες φωτισμού του αντικειμένου. Συγκεκριμένα, η κατανομή των χρωματικών τιμών που μετρούνται σε μια εικόνα ενός αντικειμένου εξαρτάται από το φωτισμό της σκηνής (illumination) και από τη σκίαση (shading) του αντικειμένου λόγω της τοποθέτησής του τόσο ως προς τις πηγές φωτός όσο και ως προς άλλα αντικείμενα της σκηνής. Σε πολλά σενάρια αναγνώρισης με βάση το χρώμα, είναι επιθυμητή η εξάλειψη της εξάρτησης της περιγραφής του χρώματος από τις συνθήκες φωτισμού. Στόχος της κανονικοποίησης χρώματος είναι η αφαίρεση της πληροφορίας της έντασης/φωτεινότητας από μια χρωματική τιμή, με στόχο μια περιγραφή του χρώματος αναλλοίωτη στις συνθήκες φωτισμού.

Η αντίληψη του χρώματος ενός αντικειμένου από το ανθρώπινο σύστημα όρασης παραμένει σε μεγάλο βαθμό αναλλοίωτη για σχετικά μεγάλο εύρος διαφορετικών συνθηκών φωτισμού. Για

παράδειγμα, το χρώμα ενός πράσινου μήλου, υπό συνθήκες φυσικού φωτισμού, γίνεται αντιληπτό σαν πράσινο τόσο στο μέσο της ημέρας, όπου ο φωτισμός είναι ουσιαστικά λευκό φως, όσο και κατά τη διάρκεια του ηλιοβασιλέματος, όπου ο φωτισμός του περιβάλλοντος είναι κατά κύριο λόγο κόκκινος. Αυτή η προσαρμοστική αντίληψη του χρώματος είναι γνωστή σαν *χρωματική σταθερότητα* (color constancy). Η προσομοίωση αυτής της ικανότητας από ηλεκτρονικούς υπολογιστές αποτελεί ένα πρόβλημα από μόνη της και έχουν αναπτυχθεί αρκετές μέθοδοι για την προσέγγισή της [41, 42, 43]. Στόχος των μεθόδων αυτών είναι η εκτίμηση/μάθηση των συνθηκών φωτισμού και η αφαίρεση της επίδρασής τους από το χρώμα που μετράται, με απώτερο σκοπό την κατασκευή μιας “κανονικής”, ανεξάρτητης από συνθήκες φωτισμού, περιγραφής/αναπαράστασης του χρώματος μιας εικόνας. Οι τεχνικές κανονικοποίησης χρώματος δίνουν μια απλή (μερική) λύση στο πρόβλημα της προσομοίωσης της χρωματικής σταθερότητας [44].

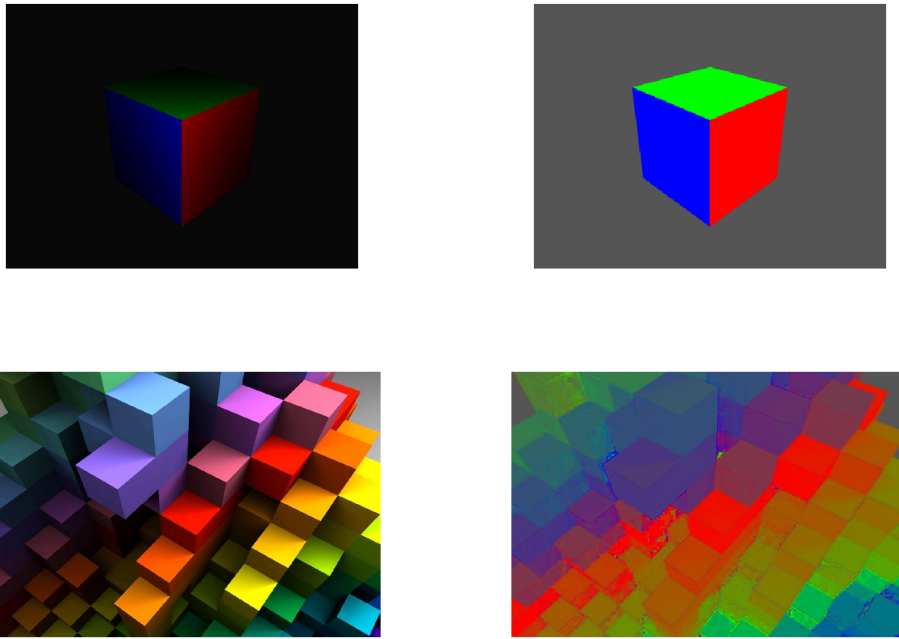
Όπως είδαμε στην προηγούμενη υποενότητα, σε πολλούς από τους μη γραμμικούς χώρους που περιγράψαμε, η μία από τις συνιστώσες αναπαράστασης κωδικοποιεί ακριβώς την ένταση (φωτεινότητα) της μετρηθείσας φωτεινής ακτινοβολίας. Μια περιγραφή χρώματος, αναλλοίωτη ως προς τις συνθήκες φωτισμού, στους χώρους αυτούς μπορεί να επιτευχθεί αν διατηρήσουμε μόνο τις συνιστώσες της χρωματικής πληροφορίας, αγνοώντας τη συνιστώσα της έντασης/φωτεινότητας. Για παράδειγμα, στο χώρο CIE LAB, αυτό μπορεί να πραγματοποιηθεί απλώς αγνοώντας τη συνιστώσα L^* (lightness), στο χώρο HSV το κανάλι V (value), κλπ.

Αν (R, G, B) οι συντεταγμένες ενός χρώματος στο χώρο RGB, για την απαλοιφή της πληροφορίας της έντασης χρησιμοποιείται συνήθως η κανονικοποίηση του χρώματος σε μοναδιαία L^1 νόρμα:

$$(r, g, b) = \left(\frac{R}{R+G+B}, \frac{G}{R+G+B}, \frac{B}{R+G+B} \right) \quad (3.1)$$

Καθώς οι κανονικοποιημένες συντεταγμένες ικανοποιούν πάντα $r + g + b = 1$, τα διανύσματα στη νέα αναπαράσταση ανήκουν σε ένα επίπεδο και η τελευταία είναι πλήρως καθορισμένη από ένα διδιάστατο διάνυσμα, για παράδειγμα το (r, g) . Ο διδιάστατος χώρος που ορίζεται έτσι είναι γνωστός σαν χώρος rg χρωματικότητας (rg chromaticity space). Σημειώνουμε ότι η L^1 νόρμα ενός διανύσματος στο χώρο RGB μοντελοποιεί την ένταση (intensity), πράγμα που δικαιολογεί την κανονικοποίηση 3.1. Μια επιπλέον χρήσιμη παρατήρηση είναι ότι, υποθέτοντας ένα γραμμικό μοντέλο μέτρησης για τον αισθητήρα λήψης της εικόνας, μια κλιμάκωση της έντασης του φωτισμού κατά έναν παράγοντα s προκαλεί την ίδια ακριβώς κλιμάκωση στις χρωματικές τιμές που μετρούμε: η τιμή (R, G, B) πριν την κλιμάκωση του φωτισμού γίνεται (sR, sG, sB) μετά από αυτήν. Η αναπαράσταση που ορίζει η 3.1 είναι προφανώς ανεξάρτητη της έντασης του φωτισμού σε αυτό το απλό φυσικό μοντέλο. Η κανονικοποίηση αυτή χρησιμοποιείται ευρέως και οδηγεί με επιτυχία σε χρωματικές αναπαραστάσεις ανεξάρτητες της έντασης του φωτισμού. Επιπλέον, έχειδειχθεί ότι η γεωμετρία του φωτισμού (θέση και προσανατολισμός της πηγής φωτός) επηρεάζει ουσιαστικά μόνο το μέγεθος της έντασης που σχετίζεται με μια μέτρηση (R, G, B) , και συνεπώς η παραπάνω αναπαράσταση εμφανίζει καλή συμπεριφορά σε σημαντικό εύρος συνθηκών [45]. Δύο παραδείγματα της επίδρασης της κανονικοποίησης 3.1 σε συνθετικές, φωτοσχιασμένες εικόνες απεικονίζονται στο Σχήμα 3.6. Οι κανονικοποιημένες τιμές απεικονίζονται σαν κανονικές συνιστώσες στο χώρο RGB. Και στις δύο περιπτώσεις, στις μετασχηματισμένες συνιστώσες έχουν εξαλειφθεί σχεδόν πλήρως όλα τα φαινόμενα σκίασης λόγω του φωτισμού της σκηνής.

Η αποσύμπλεξη της έντασης/φωτεινότητας από την αναπαράσταση ενός χρώματος είναι από μόνη της επιθυμητή σε ένα σενάριο κατασκευής ενός στατιστικού μοντέλου χρώματος ενός αντικειμένου, καθώς επιτρέπει αυτόματα μια άμεση γενίκευση της πληροφορίας που διαθέτουμε από τα δεδομένα εκπαίδευσης. Τα δεδομένα εκπαίδευσης, τυπικά, είναι ένα σύνολο τριδιάστατων διανυσμάτων που αντιστοιχούν σε χρωματικές τιμές, σε κάποιο χώρο χρώματος, που εμφανίζονται



Σχήμα 3.6: Αποτελέσματα κανονικοποίησης χρωματικών τιμών στο χρώρο RGB.

στα εικονοστοιχεία που απεικονίζουν το αντικείμενο στις επισημειωμένες εικόνες από τις οποίες αυτά αντλήθηκαν. Με την εφαρμογή κάποιας τεχνικής κανονικοποίησης, ένα τέτοιο τριδιάστατο διάλυσμα ταυτίζεται με την κλάση ισοδυναμίας του (σύνολο τριδιάστατων χρωματικών τιμών στις οποίες αλλάζει μόνο η συνιστώσα έντασης/φωτεινότητας). Το γεγονός αυτό επιτρέπει την κατασκευή στατιστικών μοντέλων με μικρότερες απαιτήσεις σε δεδομένα εκπαίδευσης. Σημειώνουμε ότι, παρότι μια τέτοια γενίκευση επιτυγχάνεται με την εφαρμογή οποιασδήποτε τεχνικής μείωσης διάστασης (dimensionality reduction) στα δεδομένα, στην περίπτωσή μας είναι ωφέλιμη, καθώς η μείωση της διάστασης συνίσταται ακριβώς στην απαλοιφή μιας (εν γένει) ανεπιθύμητης συνιστώσας.

Με τη χρήση κανονικοποιημένων αναπαραστάσεων χρώματος, τα στατιστικά μοντέλα που μας ενδιαφέρουν αφορούν κατανομές διδιάστατων (αντί τριδιάστατων) μεταβλητών. Μαζί με τις μικρότερες απαιτήσεις σε δεδομένα εκπαίδευσης, το γεγονός αυτό συνεπάγεται σημαντικά υπολογιστικά οφέλη για την εκπαίδευση και την αναπαράσταση/αποθήκευση των μοντέλων. Ενδεικτικά, για την αποθήκευση ενός ιστογράμματος τριδιάστατων τιμών απαιτείται χώρος μνήμης μιας τάξης μεγέθους περισσότερο από ό,τι ενός ιστογράμματος διδιάστατων τιμών.

3.1.3 Ψηφιακές εικόνες

Καθώς ασχολούμαστε σχεδόν αποκλειστικά με την ανάλυση και επεξεργασία εικόνων από ηλεκτρονικά μέσα, επικεντρώνουμε το ενδιαφέρον μας στις ψηφιακές εικόνες. Μια ψηφιακή εικόνα χαρακτηρίζεται από δύο επίπεδα διακριτοποίησης.

Καταρχήν, έχουμε διακριτοποίηση στο πεδίο ορισμού της εικόνας (έκταση στο επίπεδο της εικόνας). Η διακριτοποίηση αυτή, συνήθως, λαμβάνει τη μορφή της απλούστερης μορφής δειγματοληψίας ενός διδιάστατου σήματος και συνίσταται στον ορισμό, πάνω στο επίπεδο της εικόνας, ενός ομοιόμορφου, ορθογώνιου πλέγματος (rectangular grid). Με αυτόν τον τρόπο, αντί να αποθηκεύεται πληροφορία για κάθε ένα από τα άπειρα σημεία του επιπέδου της εικόνας, αποθη-

κεύονται οι τιμές της εικόνας μόνο για τα πεπερασμένου πλήθους παράθυρα που ορίζει το πλέγμα. Κάθε τέτοιο παράθυρο του πλέγματος ονομάζεται pixel (εικονοστοιχείο). Το pixel ορίζει τη μέγιστη δυνατή ευκρίνεια που έχουμε στη διάθεσή μας, αφού μία μοναδική τιμή ανατίθεται σε αυτό και, κατά συνέπεια, σε όλο το εμβαδόν της επιφάνειας του (συνεχούς) επιπέδου της εικόνας που αυτό αντιπροσωπεύει. Συνεπώς, αν δύο εικόνες απαθανατίζουν την ίδια ακριβώς σκηνή, αυτή που αποθηκεύεται σε περισσότερα pixels έχει μεγαλύτερη *ανάλυση*, και, συνεπώς, περιλαμβάνει περισσότερες λεπτομέρειες της σκηνής.

Το δεύτερο επίπεδο διακριτοποίησης έχει να κάνει με το πεδίο τιμών της εικόνας και αφορά την κβαντοποίηση της πληροφορίας που αποθηκεύεται σε κάθε εικονοστοιχείο. Για τις γκριζες (gray) εικόνες, οι οποίες δεν αποθηκεύουν πληροφορία χρώματος, κάθε pixel χαρακτηρίζεται από μια βαθμωτή τιμή που χαρακτηρίζει τη (μέση) φωτεινότητά του. Αυτός ο αριθμός, για να αποθηκευτεί, διακριτοποιείται σε ακρίβεια ενός πλήθους δυαδικών ψηφίων (bits). Όσο περισσότερα bits διατίθενται για την αποθήκευση των τιμών της φωτεινότητας, τόσο περισσότερες είναι οι διαφορετικές στάθμες κβαντοποίησής τους και τα διαφορετικά επίπεδα φωτεινότητας που μπορούν να αναπαρασταθούν. Συνήθως χρησιμοποιούνται 8bits, και συνεπώς υπάρχουν $2^8 = 256$ διαφορετικές στάθμες φωτεινότητας. Για τις έγχρωμες εικόνες, το μόνο που αλλάζει είναι ότι έχουμε 3 τιμές ανά pixel, κάθε μία από τις οποίες αντιστοιχεί σε ένα κανάλι (συνιστώσα) του χρησιμοποιούμενου χρωματικού χώρου. Κάθε μια από τις 3 συνιστώσες χρώματος κβαντοποιείται ξεχωριστά. Έτσι, για 8bits ανά συνιστώσα, έχουμε συνολικά 24bits ανά pixel.

3.2 Στατιστικά μοντέλα χρώματος

Στην ενότητα αυτή, επικεντρώνουμε το ενδιαφέρον μας σε τεχνικές μοντελοποίησης στατιστικών κατανομών χρώματος. Τα στατιστικά μοντέλα χρώματος χρησιμοποιούνται κατά κόρον τόσο σε εφαρμογές βελτίωσης και επεξεργασίας εικόνας, όσο και σε σενάρια αναγνώρισης η εντοπισμού ενός αντικειμένου με γνωστό μοντέλο χρώματος σε ένα σύνολο εικόνων. Όσον αφορά τη δεύτερη κατηγορία εφαρμογών, μας ενδιαφέρει συνήθως ο προσδιορισμός της πιθανοτικής κατανομής των χρωμάτων που εμφανίζονται σε διάφορες απεικονίσεις (εικόνες) του αντικειμένου ενδιαφέροντος. Αν ω το αντικείμενο (γενικότερα, κλάση) του οποίου το χρώμα μας ενδιαφέρει να μοντελοποιήσουμε και x μια τιμή χρώματος που μετρήθηκε από κάποια εικόνα του αντικειμένου, στόχος μας είναι ο προσδιορισμός της δεσμευμένης κατανομής πιθανότητας $p(x|\omega)$ των χρωματικών τιμών από κάθε πιθανή απεικόνιση του αντικειμένου ω .

Παρακάτω παρουσιάζουμε τις πιο συνηθισμένες μεθόδους *εκτίμησης πυκνότητας* (density estimation) που μπορούν να χρησιμοποιηθούν για τον προσδιορισμό και την αναπαράσταση μιας κατανομής χρωματικών τιμών. Στην περίπτωση κατασκευής στατιστικού μοντέλου χρώματος ενός αντικειμένου, υποθέτουμε ότι έχουμε στη διάθεσή μας έναν αριθμό εικόνων του αντικειμένου αυτού, από τις οποίες έχουν απομονωθεί/επισημειωθεί οι χρωματικές τιμές που ανήκουν στο αντικείμενο (οι τιμές χρώματος των εικονοστοιχείων που ανήκουν σε αυτό). Σε κάθε περίπτωση, θα αναφερόμαστε στις χρωματικές τιμές που θα χρησιμοποιήσουμε με τον όρο *δεδομένα εκπαίδευσης* (training data) του μοντέλου. Τα περισσότερα από όσα αναφερθούν αποτελούν βασικές μεθόδους κατασκευής στατιστικών μοντέλων από επισημειωμένα δεδομένα και δεν περιορίζονται σε κατανομές χρωματικών τιμών. Ολοκληρωμένες συγκριτικές περιγραφές των μεθόδων που χρησιμοποιούνται στη μοντελοποίηση χρώματος, με έμφαση στο χρώμα του ανθρώπινου δέρματος, μπορούν να βρεθούν στις εργασίες [46, 47, 48].

3.2.1 Ιστογράμματα χρώματος

Η πιο απλή μη παραμετρική αναπαράσταση μιας πυκνότητας πιθανότητας είναι από ένα ιστόγραμμα χρωματικών τιμών, το οποίο κατασκευάζεται από τα δεδομένα εκπαίδευσης. Το ιστόγραμμα για μια συνεχή τυχαία μεταβλητή $x \in \mathbb{R}^d$ ορίζεται με βάση μια διαμέριση του συνόλου των δυνατών τιμών της x σε περιοχές που ονομάζονται “κάδοι” (bins, buckets). Η κατασκευή ενός ιστογράμματος σε d διαστάσεις αποτελεί ευθεία γενίκευση της γνωστής περίπτωσης της κατασκευής ιστογράμματος για μονοδιάστατες τυχαίες μεταβλητές. Στην πιο απλή περίπτωση, έχουμε μια ομοιόμορφη διαμέριση του διαστήματος τιμών σε κάθε μια από τις d διαστάσεις της x και κάθε κάδος ορίζεται μοναδικά σαν τομή d ζωνών του \mathbb{R}^d , κάθε μια από τις οποίες αντιστοιχεί σε ένα από τα διαστήματα της διαμέρισης κάποιας από τις d διαστάσεις. Στην περίπτωση αυτή, οι κάδοι του ιστογράμματος είναι απλώς οι περιοχές ενός ορθογώνιου πλέγματος που καλύπτει/διαμερίζει το σύνολο τιμών της x και ορίζεται από τις επιμέρους διαμερίσεις σε κάθε διάσταση.

Ένα ιστόγραμμα σε d διαστάσεις, με σύνολο κάδων διαμέρισης \mathcal{B} , ορίζεται πλήρως από δύο συναρτήσεις:

- Μια συνάρτηση χβάντισης $b : \mathbb{R}^d \rightarrow \mathcal{B}$ που απεικονίζει μια μετρηθείσα τιμή (τυχαίο δείγμα) σε έναν κάδο του ιστογράμματος. Η συνάρτηση αυτή ουσιαστικά ορίζει τη γεωμετρία των κάδων διαμέρισης.
- Μια συνάρτηση “καταμέτρησης” $h : \mathcal{B} \rightarrow \mathbb{N}$ που επιστρέφει το πλήθος των δειγμάτων που έχουν τοποθετηθεί στον κάδο-όρισμα. Ο ορισμός αυτής της συνάρτησης είναι το αποτέλεσμα της κατασκευής του ιστογράμματος.

Με γνωστά τα παραπάνω, ένα ιστόγραμμα μπορεί να χρησιμοποιηθεί για τον ορισμό μιας τμηματικά σταθερής συνάρτησης πυκνότητας πιθανότητας σύμφωνα με τη σχέση:

$$p(x) = \frac{h(b(x))}{N \text{vol}(b(x))}, \quad (3.2)$$

όπου $N = \sum_{i \in \mathcal{B}} h(i)$ είναι το (γνωστό) πλήθος των δειγμάτων εκπαίδευσης που χρησιμοποιήθηκαν για την κατασκευή του ιστογράμματος (πλήθος δειγμάτων που έχουν τοποθετηθεί συνολικά σε όλους τους κάδους) και $\text{vol}(b(x))$ είναι ο όγκος του κάδου (στον \mathbb{R}^d) στον οποίο αντιστοιχίζεται η μέτρηση x . Καθώς η διαμέριση του πεδίου τιμών της x σε κάδους ορίζει αυτόματα μια διακριτοποίηση των τιμών της μέσω της συνάρτησης b , είναι πιο φυσιολογικό ένα ιστόγραμμα να χρησιμοποιηθεί για τον ορισμό μιας διακριτής κατανομής πιθανότητας επί των διακριτών αυτών τιμών (σύνολο κάδων). Αυτή είναι η ταχτική που εφαρμόζεται στην πράξη και η τιμή της διακριτής κατανομής για τον κάδο (διακριτή τιμή) $i \in \mathcal{B}$ δίνεται απλώς από την:

$$P(i) = \frac{h(i)}{N} \quad (3.3)$$

Τα βασικά ζητήματα που σχετίζονται με την ποιότητα της προσέγγισης της πραγματικής κατανομής της x από την 3.2/3.3 αφορούν το πλήθος των διαθέσιμων δεδομένων εκπαίδευσης και το μέγεθος των κάδων χβάντισης σε σχέση με το τελευταίο. Για μεγάλο μέγεθος κάδων, η εκτίμηση της πυκνότητας θα πάσχει από χαμηλή ανάλυση, ενώ για μικρό θα εμφανίζει μεγάλη στατιστική μεταβλητότητα. Έχουν προταθεί αρκετοί ευριστικοί κανόνες επιλογή του, υπό κάποια έννοια, βέλτιστου μεγέθους/πλήθους κάδους, αλλά το τελευταίο εξαρτάται γενικά από την πραγματική κατανομή της x και το πλήθος των διαθέσιμων δειγμάτων. Επίσης, τα ιστογράμματα είναι ιδιαίτερα ακατάλληλα για δεδομένα υψηλής διάστασης (curse of dimensionality): το πλήθος

των κάδων αυξάνεται εκθετικά με τη διάσταση d ενώ παράλληλα απαιτούνται αναλόγως πολλά δείγματα εκπαίδευσης για να μην είναι οι περισσότεροι κάδοι άδειοι.

Σημειώνουμε ότι αρκετά δημοφιλείς μη παραμετρικές μέθοδοι είναι οι μέθοδοι *εκτίμησης πυκνότητας με πυρήνες* (kernel density estimation), στις οποίες περιλαμβάνονται η μέθοδος παραθύρων Parzen-Rosenblatt και η εκτίμηση με βάση τους k κοντινότερους γείτονες (k NN, k nearest neighbor estimation) [49]. Οι εκτιμήσεις με βάση τις μεθόδους αυτές, σε γενικές γραμμές, αποτελούν παραλλαγές με, υπό συνθήκες, καλύτερα χαρακτηριστικά του απλού ιστογράμματος, ενώ ουσιαστικά δεν βελτιώνουν κάποιες από τις αδυναμίες του. Επιπλέον, σε αντίθεση με την περίπτωση του ιστογράμματος, η αποτίμηση της εκτίμησης της πυκνότητας για ένα νέο σημείο έχει δυνητικά μεγάλο υπολογιστικό κόστος, καθώς ενδέχεται να απαιτείται η προσπέλαση και χρήση όλων των δειγμάτων εκπαίδευσης. Το κόστος αυτό ελαττώνεται υπό συνθήκες και με τη χρήση περίτεχνων δομών δεδομένων.

Καθώς ένα χρώμα αναπαρίσταται, σχεδόν πάντοτε, σαν ένα σημείο ενός διδιάστατου ή τριδιάστατου χώρου, η διάσταση δεν αποτελεί εμπόδιο και τα ιστογράμματα χρώματος αποτελούν ένα ευρέως χρησιμοποιούμενο μέσο μοντελοποίησης χρωματικών κατανομών. Είναι ιδιαίτερα εύκολα στην κατασκευή και τη χρήση και μπορούν, επιπλέον, να χρησιμοποιηθούν σαν μέσο εποπτείας για τον εμπειρικό προσδιορισμό του τύπου ή παραμέτρων κάποιου πιθανού παραμετρικού μοντέλου.

3.2.2 Παραμετρικά χρωματικά μοντέλα

Στις μη παραμετρικές μεθόδους που αναφέραμε, η εκτίμηση της πυκνότητας γίνεται εξολοκλήρου με βάση τα δεδομένα εκπαίδευσης (data driven). Αντίθετα, στις παραμετρικές μεθόδους, υποτίθεται γνωστή η συναρτησιακή μορφή της άγνωστης πυκνότητας $p(x)$ και το ζητούμενο είναι η *εκτίμηση των παραμέτρων* (parameter estimation) της συνάρτησης που, σύμφωνα με κάποιο κριτήριο, ταιριάζουν καλύτερα στα δεδομένα εκπαίδευσης. Τα δύο βασικά θεωρητικά πλαίσια εκτίμησης παραμέτρων είναι η *εκτίμηση μέγιστης πιθανοφάνειας* (Maximum Likelihood Estimation) και η *Μπεϊζιανή εκτίμηση παραμέτρων* (Bayesian parameter estimation). Παρακάτω θα ασχοληθούμε με την πρώτη προσέγγιση.

Έστω $\theta \in \mathbb{R}^M$ το διάνυσμα των παραμέτρων που καθορίζουν πλήρως μια συνάρτηση πυκνότητας της υποτιθέμενης συναρτησιακής μορφής. Με $p(x; \theta)$ συμβολίζουμε την πυκνότητα αυτή για μια δεδομένη τιμή του διανύσματος θ . Προφανώς, η τιμή της $p(x; \theta)$ είναι συνάρτηση τόσο της τιμής x της τυχαίας μεταβλητής όσο και των παραμέτρων θ . Έστω $X = \{x_1, \dots, x_N\}$ το σύνολο των δεδομένων εκπαίδευσης, το οποίο υποθέτουμε ότι αποτελείται από ανεξάρτητα δείγματα (i.i.d., independent, identically distributed) της $p(x; \theta)$. Λόγω της ανεξαρτησίας που υποθέσαμε για τα δείγματα εκπαίδευσης, η τιμή της πυκνότητας της από κοινού κατανομής του συνόλου των δεδομένων είναι ίση με το γινόμενο:

$$p(X; \theta) = \prod_{i=1}^N p(x_i; \theta) \quad (3.4)$$

Η $p(X; \theta)$ ονομάζεται *συνάρτηση πιθανοφάνειας* (likelihood function) και εκφράζει την πιθανοφάνεια των παραμέτρων θ για το σύνολο δεδομένων X . Στο πλαίσιο της εκτίμησης παραμέτρων μέγιστης πιθανοφάνειας, αναζητούμε τις τιμές των παραμέτρων $\hat{\theta}$ που μεγιστοποιούν συνάρτηση πιθανοφάνειας:

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbb{R}^M} p(X; \theta) \quad (3.5)$$

Οι τιμές αυτές ορίζουν το στατιστικό μοντέλο της υποτιθέμενης μορφής που ταιριάζει καλύτερα με το σύνολο των παρατηρήσεων X . Οι τιμές $\hat{\theta}$ ονομάζονται *εκτιμήτριες μέγιστης πιθανοφάνειας* (Maximum Likelihood Estimates). Ανάλογα με την υποτιθέμενη συναρτησιακή μορφή της

$p(x; \theta)$, για τη μεγιστοποίηση 3.5 συνήθως χρησιμοποιείται, για αναλυτικούς λόγους, η συνάρτηση λογαριθμικής πιθανοφάνειας, που είναι ο λογάριθμος της 3.4:

$$L(X, \theta) = \ln p(X; \theta) = \sum_{i=1}^N \ln p(x_i; \theta) \quad (3.6)$$

Καθώς ο λογάριθμος είναι μονότονη συνάρτηση, το πρόβλημα μεγιστοποίησης 3.5 είναι ισοδύναμο με το:

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \mathbb{R}^M} L(X, \theta) = \operatorname{argmax}_{\theta \in \mathbb{R}^M} \sum_{i=1}^N \ln p(x_i; \theta) \quad (3.7)$$

Συχνά, για τη μεγιστοποίηση 3.7, χρησιμοποιούνται αναλυτικές μέθοδοι και οι τιμές $\hat{\theta}$ προσδιορίζονται σαν ρίζες της εξίσωσης:

$$\nabla_{\theta} L(X, \theta) = 0 \Leftrightarrow \nabla_{\theta} \left(\sum_{i=1}^N \ln p(x_i; \theta) \right) = 0 \Leftrightarrow \sum_{i=1}^N (\nabla_{\theta} \ln p(x_i; \theta)) = 0 \quad (3.8)$$

Παρακάτω παρουσιάζουμε δύο από τα πιο δημοφιλή παραμετρικά μοντέλα, τα οποία εφαρμόζονται ευρέως στη μοντελοποίηση κατανομών χρώματος.

Γκαουσιανές κατανομές

Το πιο κλασικό σενάριο εκτίμησης παραμέτρων περιλαμβάνει την εκτίμηση των παραμέτρων μιας Γκαουσιανής (κανονικής) κατανομής. Μια d -διάστατη Γκαουσιανή κατανομή παραμετροποιείται πλήρως από τη μέση τιμή της, $\mu \in \mathbb{R}^d$, και το συμμετρικό, θετικά ημιορισμένο $d \times d$ πίνακα συμμεταβλητότητας Σ , ο οποίος, επιπρόσθετα, υποτίθεται αντιστρέψιμος. Η συνάρτηση πυκνότητας πιθανότητάς της δίνεται από:

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}, \quad (3.9)$$

όπου $|\Sigma|$ η ορίζουσα του πίνακα Σ .

Σύμφωνα με τα παραπάνω, η συνάρτηση λογαριθμικής πιθανοφάνειας για το σύνολο X γράφεται:

$$L(X, \mu, \Sigma) = -\frac{Nd}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^N (x_i - \mu)^T \Sigma^{-1} (x_i - \mu), \quad (3.10)$$

Η εκτίμηση μέγιστης πιθανοφάνειας $\hat{\mu}$ για τη μέση τιμή πρέπει να ικανοποιεί την 3.8:

$$\nabla_{\mu} L = 0 \Leftrightarrow \sum_{i=1}^N (x_i - \hat{\mu}) = 0 \Leftrightarrow \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (3.11)$$

όπου έγινε χρήση του γεγονότος ότι, λόγω συμμετρικότητας του Σ , ο Σ^{-1} είναι επίσης συμμετρικός. Για την εύρεση της εκτιμήτριας μέγιστης πιθανοφάνειας για τον πίνακα συμμεταβλητότητας

Σ , ξαναγράφουμε την 3.10 ως:

$$\begin{aligned} L(X, \mu, \Sigma) &= -\frac{Nd}{2} \ln(2\pi) + \frac{N}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^N (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \\ &= -\frac{Nd}{2} \ln(2\pi) + \frac{N}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^N \text{tr} [\Sigma^{-1} (x_i - \mu)(x_i - \mu)^T] \\ &= -\frac{Nd}{2} \ln(2\pi) + \frac{N}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr} \left[\Sigma^{-1} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T \right], \end{aligned} \quad (3.12)$$

όπου χρησιμοποιήσαμε τη γραμμικότητα του τελεστή του ίχνους, την ταυτότητα $\text{tr} [AB] = \text{tr} [BA]$ και ότι $|\Sigma^{-1}| = |\Sigma|^{-1}$. Χρησιμοποιώντας, επιπλέον, τις ταυτότητες:

$$\nabla_A \ln |A| = A^{-T} \text{ και } \nabla_A \text{tr} [AB] = B^T,$$

βρίσκουμε τη βέλτιστη τιμή ως προς Σ^{-1} εφαρμόζοντας την 3.8 στην 3.12:

$$\nabla_{\Sigma^{-1}} L = 0 \Leftrightarrow N \hat{\Sigma} - \sum_{i=1}^N (x_i - \hat{\mu})(x_i - \hat{\mu})^T = 0 \Leftrightarrow \hat{\Sigma} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})(x_i - \hat{\mu})^T \quad (3.13)$$

Οι κλειστοί τύποι 3.11 και 3.13 για τις εκτιμήτριες μέγιστης πιθανοφάνειας μιας Γκαουσιανής κατανομής καθιστούν το εν λόγω μοντέλο ιδιαίτερα απλό στην εφαρμογή. Η μοντελοποίηση χρωματικών κατανομών με Γκαουσιανές είναι επιτυχής σε περιπτώσεις όπου η πραγματική κατανομή είναι “κοντά” σε κανονική. Για παράδειγμα, το μοντέλο αυτό είναι κατάλληλο για την περιγραφή του χρώματος ενός αντικειμένου, όταν αυτό αποτελείται από σχετικά μικρές μεταβολές ενός βασικού χρώματος. Παρόλα αυτά, μια Γκαουσιανή δεν επαρκεί για την μοντελοποίηση μιας πολυκόρυφης (multimodal) πραγματικής κατανομής.

Μείγματα Γκαουσιανών κατανομών

Τα μοντέλα μειγμάτων Γκαουσιανών κατανομών (Gaussian Mixture Models, GMMs) είναι συνεχείς κατανομές πιθανότητας, οι συναρτήσεις πυκνότητας των οποίων είναι κυρτοί συνδυασμοί Γκαουσιανών πυκνοτήτων. Ο φορμαλισμός των μοντέλων αυτών και η εκτίμηση μέγιστης πιθανοφάνειας για τις παραμέτρους τους περιγράφονται αναλυτικά στο Παράρτημα Β. Τα μείγματα Γκαουσιανών ορίζουν μια κλάση ιδιαίτερα εκφραστικών μοντέλων, καθώς, με την ενσωμάτωση ικανού αριθμού συνιστωσών σε ένα τέτοιο μείγμα, σχεδόν οποιαδήποτε συνεχής συνάρτηση πυκνότητας μπορεί να προσεγγισθεί αυθαίρετα καλά.

Μια άποψη του προβλήματος της μοντελοποίησης μιας πιθανοτικής κατανομής με μείγματα Γκαουσιανών που δεν καλύπτεται στο Παράρτημα Β είναι η επιλογή του πλήθους k των συνιστωσών του μείγματος. Η επιλογή αυτή μπορεί να γίνει εμπειρικά, αν υπάρχει κάποια πρότερη γνώση του φυσικού μοντέλου που παρήγαγε το σύνολο των δεδομένων ή αν υπάρχει η δυνατότητα κάποιας εποπτείας της κατανομής που ακολουθεί το σύνολο δεδομένων, για παράδειγμα, μέσω ενός ιστογράμματος. Είναι γεγονός ότι, ενσωματώνοντας οσοδήποτε πολλές συνιστώσες, η συνάρτηση πιθανοφάνειας μπορεί να λάβει αυθαίρετα μεγάλη τιμή. Παρόλα αυτά, μια τέτοια πρακτική οδηγεί στο φαινόμενο του overfitting του μοντέλου στα δεδομένα: ένα τέτοιο μοντέλο μπορεί να ερμηνεύει (αυθαίρετα) καλά τις δεδομένες παρατηρήσεις αλλά έχει μικρή αξία, καθώς, εν γένει, αποτυγχάνει να ερμηνεύσει νέες (μη γενικευσιμότητα). Μια γενικότερη, πιο εμπειριστατωμένη προσέγγιση για την επιλογή του πλήθους k των συνιστωσών περιλαμβάνει τη χρήση κριτηρίων

επιλογής μοντέλου (model selection) από τη θεωρία πληροφορίας, όπως το κριτήριο πληροφορίας του Akaike (Akaike Information Criterion, AIC) και το Μπεϋζιανό κριτήριο πληροφορίας του Schwarz (Bayesian Information Criterion, BIC). Δεδομένου ενός συνόλου παραμέτρων που εκτιμήθηκαν για το σύνολο των παρατηρήσεων, τα κριτήρια αυτά ορίζονται σαν συναρτήσεις της τιμής της συνάρτησης λογαριθμικής πιθανοφάνειας, του πλήθους των παραμέτρων του υποτιθέμενου μοντέλου και του πλήθους των παρατηρήσεων. Μπορούν να ερμηνευθούν σαν συναρτήσεις κόστους, που ευνοούν το καλό ταίριασμα του μοντέλου στις παρατηρήσεις (υψηλές τιμές πιθανοφάνειας), θέτοντας, όμως, ποινή για το πλήθος των παραμέτρων αυτού. Η τιμή του κριτηρίου BIC δίνεται από την έκφραση:

$$BIC = -2L(X, \hat{\theta}) + M \ln(N), \quad (3.14)$$

όπου L η συνάρτηση λογαριθμικής πιθανοφάνειας 3.6 των παραμέτρων που εκτιμήθηκαν για το σύνολο των παρατηρήσεων, M το πλήθος των προς εκτίμηση παραμέτρων του μοντέλου και N το πλήθος των παρατηρήσεων. Σημειώνουμε ότι σε ένα μείγμα k Γκαουσιανών συνιστωσών σε d διαστάσεις, υποθέτοντας γενικής μορφής πίνακες συμμεταβλητότητας (full covariance matrices), το πλήθος των προς εκτίμηση παραμέτρων είναι ίσο με:

$$M = k + k \left(d + \frac{d(d+1)}{2} \right) = k \frac{(d+1)(d+2)}{2}$$

Μια συγκριτική παρουσίαση των μεθόδων επιλογής μοντέλου για μείγματα Γκαουσιανών μπορεί να βρεθεί στην εργασία [50].

Χάρη στην πλούσια εκφραστική τους δύναμη, τα μείγματα Γκαουσιανών χρησιμοποιούνται ευρέως στη μοντελοποίηση χρωματικών κατανομών, οποιαδήποτε μορφή και αν αυτές έχουν. Ενδεικτικά, αναφέρουμε ότι εφαρμόζονται συχνά για την κατασκευή χρωματικών μοντέλων για το ανθρώπινο δέρμα [51] και αποτελούν το βασικό εργαλείο περιγραφής χρώματος σε πολλές μεθόδους κατάτμησης εικόνας, όπως, για παράδειγμα, ο δημοφιλής αλγόριθμος GrabCut [52]. Βασικό τους μειονέκτημα είναι η ανάγκη για τη χρήση ενός επαναληπτικού σχήματος (αλγόριθμος 13) για την εκτίμηση των παραμέτρων τους, το οποίο συγκλίνει σε κάποιο τοπικό ελάχιστο της συνάρτησης πιθανοφάνειας.

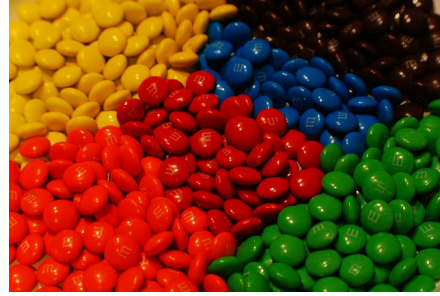
Κλείνουμε την ενότητα αυτή με δύο απλές εφαρμογές. Για κάθε μια από τις φυσικές εικόνες 3.7a και 3.7b, εκπαιδεύσαμε δύο στατιστικά μοντέλα για την κατανομή των χρωμάτων τους: ένα ιστόγραμμα και ένα μείγμα Γκαουσιανών πυκνοτήτων. Για την εικόνα 3.7a, χρησιμοποιήσαμε τον κανονικοποιημένο χώρο RGB (χώρος rg χρωματικότητας), ενώ για την 3.7b το χώρο CIE LAB, αγνοώντας το κανάλι L^* (διατηρήθηκαν μόνο οι συνιστώσες a^* και b^*). Τα ιστογράμματα αναπαριστούν τη διακριτή κατανομή πιθανότητας της εξίσωσης 3.3 σε ένα σύνολο 64×64 κάδων (64 σε κάθε διάσταση), ενώ ο αριθμός των συνιστωσών για τα μείγματα Γκαουσιανών εκτιμήθηκε εμπειρικά ανά περίπτωση, με βάση τα αντίστοιχα ιστογράμματα. Οι στατιστικές κατανομές απεικονίζονται στα Σχήματα 3.8a και 3.8b.

3.3 Κατάτμηση εικόνας σε αντικείμενα με βάση το χρώμα

Στην ενότητα αυτή, παρουσιάζουμε κάποιες από τις πιο συνηθισμένες μεθόδους κατάτμησης μιας εικόνας στα αντικείμενα (γενικότερα, κλάσεις/κατηγορίες) που απεικονίζει, οι οποίες βασίζονται αποκλειστικά στο χρώμα και λειτουργούν σε επίπεδο εικονοστοιχείου. Συγκεκριμένα, αν x το χρώμα ενός εικονοστοιχείου μιας εικόνας σε κάποιο (κανονικοποιημένο ή μη) χώρο χρώματος, αναζητούμε στρατηγικές ανάθεσης (ταξινόμησης) της μέτρησης x σε κάποια από τις (γνωστές εκ των προτέρων) κλάσεις που περιέχονται στην εικόνα.



(a)



(b)

Σχήμα 3.7: Εικόνες δοκιμής.

3.3.1 Η κατάτμηση εικόνας σαν πρόβλημα Μπεϋζιανής απόφασης

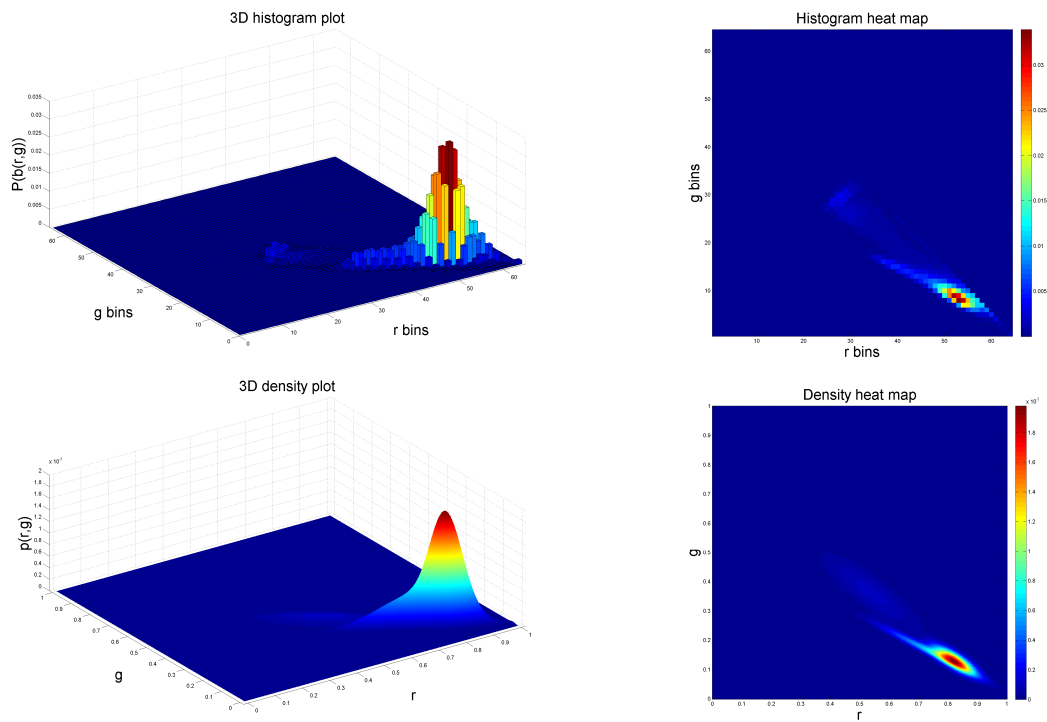
Η παρουσίαση που ακολουθεί βασίζεται εν μέρει στο Κεφάλαιο 2 του [49]. Έστω $\mathcal{C} = \{\omega_1, \dots, \omega_c\}$ το σύνολο των κατηγοριών που εμφανίζονται σε μια εικόνα και στις οποίες επιθυμούμε να ταξινομήσουμε τα εικονοστοιχεία της. Μια κλάση μπορεί να αντιστοιχεί σε ένα συγκεκριμένο αντικείμενο ή σε κάποια γενικότερη σημαντική έννοια, όπως για παράδειγμα το παρασκήνιο (φόντο) της εικόνας. Οι κατηγορίες αυτές είναι εξαρχής γνωστές και υποτίθενται αμοιβαία αποκλειόμενες (ανά δύο ξένες): κάθε εικονοστοιχείο ανήκει (ανατίθεται) σε ακριβώς μία κλάση. Υποθέτουμε, επιπλέον, ότι τόσο τα στατιστικά μοντέλα που περιγράφουν την κατανομή χρώματος κάθε κλάσης (δεσμευμένες πυκνότητες $p(x|\omega_j)$) όσο και οι πρότερες πιθανότητες εμφάνισης κάθε κλάσης, $P(\omega_j)$, είναι γνωστά για $j = 1, \dots, c$.

Σημειώνουμε ότι η εκτίμηση της υπό συνθήκη πυκνότητας $p(x|\omega_j)$ της κλάσης ω_j μπορεί να έχει γίνει με κάποια από τις τεχνικές της προηγούμενης ενότητας, με χρήση κατάλληλα επισημειωμένων εικόνων εκπαίδευσης. Μέσω της ίδιας διαδικασίας μπορεί να εκτιμηθεί και η πρότερη πιθανότητα $P(\omega_j)$: ένας απλός και συχνά χρησιμοποιούμενος τρόπος υπολογισμού της είναι ως ο λόγος του πλήθους των εικονοστοιχείων που είναι επισημειωμένα ότι ανήκουν στην κατηγορία ω_j προς το συνολικό πλήθος των εικονοστοιχείων των εικόνων εκπαίδευσης που χρησιμοποιήθηκε για την εκτίμηση της $p(x|\omega_j)$. Αν για την εκτίμηση των υπό συνθήκη πυκνοτήτων $p(x|\omega_j)$ δε χρησιμοποιήθηκαν εικόνες εκπαίδευσης που περιέχουν ακριβώς το ίδιο σύνολο κλάσεων, απαιτείται κανονικοποίηση των εκτιμήσεων των πρότερων πιθανοτήτων, ώστε $\sum_{j=1, \dots, c} P(\omega_j) = 1$.

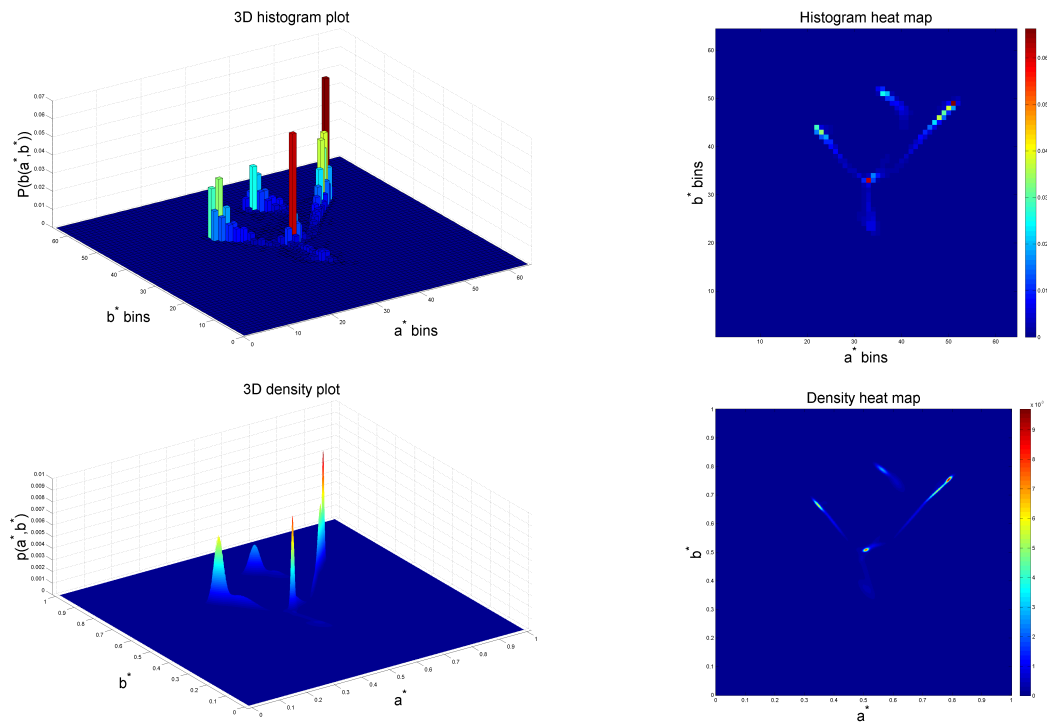
Η πρότερη πιθανότητα $P(\omega_j)$ αντανακλά την εκ των προτέρων γνώση μας του πόσο πιθανό είναι ένα εικονοστοιχείο να ανήκει στην κλάση ω_j , προτού γίνει γνωστό το χρώμα αυτού x . Η τιμή της ύστερης πυκνότητας $p(x|\omega_j)$ ονομάζεται *πιθανοφάνεια* (likelihood) της κλάσης ω_j με βάση το χρώμα x και εκφράζει πόσο “πιθανοφανής” (likely) είναι η κατηγορία ω_j για την παρατήρηση x . Ο κανόνας του Bayes μας επιτρέπει, με δεδομένη την παρατήρηση της τιμής x , να μετατρέψουμε την πρότερη πιθανότητα $P(\omega_j)$ στην ύστερη πιθανότητα $P(\omega_j|x)$:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)} = \frac{p(x|\omega_j)P(\omega_j)}{\sum_{i=1}^c p(x|\omega_i)P(\omega_i)} \quad (3.15)$$

Η $P(\omega_j|x)$ είναι ακριβώς η πιθανότητα μια παρατήρηση να προέρχεται από την κατηγορία ω_j , δεδομένου ότι η τιμή της είναι ίση με x . Συνεπώς, φαίνεται πολύ λογικό να αναθέτουμε μια δεδομένη μέτρηση x στην κλάση της οποίας η ύστερη πιθανότητα είναι μεγαλύτερη. Χρησιμοποιώντας το συμβολισμό $\alpha(x)$ για τον κανόνα απόφασής μας, όπου η συνάρτηση $\alpha: \mathbb{R}^d \rightarrow \mathcal{C}$ αντιστοιχίζει



(a) Κατανομή χρωματικών τιμών της εικόνας 3.7a στο χώρο rg χρωματικότητας (κανονικοποιημένος RGB). Πάνω σειρά: ιστόγραμμα χρωματικών τιμών με 64×64 κάδους. Κάτω σειρά: μείγμα Γκαουσιανών με $k = 3$ συνιστώσες για την προσέγγιση της κατανομής χρωματικών τιμών της εικόνας.



(b) Κατανομή χρωματικών τιμών της εικόνας 3.7b στο χώρο CIE LAB (συνιστώσες a^* και b^* μόνο). Πάνω σειρά: ιστόγραμμα χρωματικών τιμών με 64×64 κάδους. Κάτω σειρά: μείγμα Γκαουσιανών με $k = 16$ συνιστώσες για την προσέγγιση της κατανομής χρωματικών τιμών της εικόνας.

Σχήμα 3.8: Κατανομές χρωμάτων των εικόνων 3.7a και 3.7b.

μια παρατήρηση x σε μια κλάση, η στρατηγική αυτή διατυπώνεται ως:

$$\alpha(x) = \omega_i, \text{ όπου } i = \underset{j=1, \dots, c}{\operatorname{argmax}} P(\omega_j|x) = \underset{j=1, \dots, c}{\operatorname{argmax}} p(x|\omega_j)P(\omega_j) \quad (3.16)$$

Είναι εύκολο να δούμε ότι ο κανόνας απόφασης 3.16 ελαχιστοποιεί τη μέση πιθανότητα σφάλματος (λανθασμένης απόφασης) (error rate). Αν $P(\text{error}|x)$ η πιθανότητα λανθασμένης απόφασης με δεδομένη την παρατήρηση x , η μέση πιθανότητα σφάλματος δίνεται από την:

$$P(\text{error}) = \int P(\text{error}, x) dx = \int P(\text{error}|x)p(x) dx, \quad (3.17)$$

όπου η ολοκλήρωση λαμβάνει χώρα στον χώρο χρώματος όπου ανήκει η x . Αν επιλέξουμε την κλάση ω_j για την παρατήρηση x , η πιθανότητα $P(\text{error}|x)$ ισούται με $1 - P(\omega_j|x)$, για $j = 1, \dots, c$. Με με τον κανόνα 3.16, η $P(\text{error}|x)$ γίνεται ίση με:

$$P(\text{error}|x) = \min\{1 - P(\omega_1|x), \dots, 1 - P(\omega_c|x)\},$$

καθώς ο παρονομαστής $p(x)$ της 3.15 δεν επηρεάζει την απόφαση, και γίνεται όσο είναι δυνατόν μικρότερη για κάθε x . Το γεγονός αυτό εξασφαλίζει την ελαχιστότητα της μέσης πιθανότητας σφάλματος 3.17.

Σε ένα γενικότερο πλαίσιο, συχνά μας ενδιαφέρει η βελτιστότητα του κανόνα απόφασης ως προς κάποιο διαφορετικό κριτήριο από τη μέση πιθανότητα σφάλματος 3.17. Για παράδειγμα, ενδέχεται να είναι επιθυμητό η απώλεια (“κόστος”) που συνδέεται με μια απόφαση να είναι συνάρτηση τόσο της απόφασης που λαμβάνεται όσο και της πραγματικής κλάσης. Για τον ορισμό μιας γενικής μορφής κριτηρίου απόφασης, εισάγουμε τη συνάρτηση απώλειας $\lambda : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}$. Για $i, j = 1, \dots, c$, η τιμή $\lambda(\omega_i, \omega_j) = \lambda_{ij}$ είναι η απώλεια που επιφέρει η απόφαση ω_i , όταν η πραγματική κατηγορία είναι ω_j . Οι τιμές λ_{ij} κωδικοποιούν τόσο το κόστος μιας σωστής απόφασης, όσο και κάθε πιθανής περίπτωσης λανθασμένης. Στο πλαίσιο της κατάτμησης μιας εικόνας, αυτό μεταφράζεται σε όρους του πόσο σημαντική είναι η απόφαση υπέρ ενός συγκεκριμένου αντικειμένου, καθώς και η σχετική σημασία καθενός ως προς κάθε άλλο. Με δεδομένη μια παρατήρηση x , η αναμενόμενη τιμή της υπό συνθήκη απώλειας (υπό συνθήκη ρίσκο) λόγω της απόφασης $\alpha(x)$ είναι ίση με το σταθμισμένο άθροισμα:

$$R(\alpha(x)|x) = \sum_{j=1}^c \lambda(\alpha(x), \omega_j)P(\omega_j|x)$$

Συνολικά, η αναμενόμενη τιμή της απώλειας (συνολικό ρίσκο) είναι:

$$R = \int R(\alpha(x)|x)p(x) dx \quad (3.18)$$

Για την ελαχιστοποίηση της 3.18, αρκεί ο κανόνας απόφασης α να ελαχιστοποιεί το υπό συνθήκη ρίσκο $R(\alpha(x)|x)$ για κάθε x . Συνεπώς, στη γενική περίπτωση, ο βέλτιστος κανόνας απόφασης για τη συνάρτηση απώλειας δίνεται από την:

$$\begin{aligned} \alpha(x) &= \omega_i, \text{ όπου } i = \underset{j=1, \dots, c}{\operatorname{argmin}} R(\omega_j|x) = \underset{j=1, \dots, c}{\operatorname{argmin}} \sum_{k=1}^c \lambda(\omega_j, \omega_k)P(\omega_k|x) \\ &= \underset{j=1, \dots, c}{\operatorname{argmin}} \sum_{k=1}^c \lambda_{jk}p(x|\omega_k)P(\omega_k) \end{aligned} \quad (3.19)$$

Ο κανόνας απόφασης 3.19 ελαχιστοποιεί το συνολικό ρίσκο 3.18 και ονομάζεται Μπεϋζιανός κανόνας απόφασης (Bayesian decision rule).

Μια συνήθης επιλογή για τη συνάρτηση απώλειας είναι, για $i, j = 1, \dots, c$:

$$\lambda_{ij} = 1 - \delta_{ij} = \begin{cases} 0 & \text{αν } i = j \\ 1 & \text{αν } i \neq j \end{cases}$$

Η συνάρτηση αυτή ονομάζεται *συμμετρική* ή “0-1” (zero-one) συνάρτηση απώλειας και αναθέτει μηδενική τιμή απώλειας στις σωστές αποφάσεις και ίση, μοναδιαία απώλεια για όλες τις λανθασμένες. Το υπό συνθήκη κόστος για την απόφαση υπέρ της κλάσης ω_i , σε αυτή την περίπτωση, γράφεται για $i = 1, \dots, c$:

$$R(\omega_i|x) = \sum_{j=1}^c \lambda(\omega_i, \omega_j)P(\omega_j|x) = \sum_{j=1}^c \lambda_{ij}P(\omega_j|x) = \sum_{j \neq i} P(\omega_j|x) = 1 - P(\omega_i|x) \quad (3.20)$$

Το υπό συνθήκη ρίσκο 3.20 είναι ακριβώς η πιθανότητα λανθασμένης απόφασης αν επιλέξουμε την κλάση ω_i . Συνεπώς, το συνολικό ρίσκο που συνδέεται με αυτή τη συνάρτηση απώλειας είναι ακριβώς η μέση πιθανότητα σφάλματος 3.17 και, σύμφωνα με τα παραπάνω, ελαχιστοποιείται από τον κανόνα απόφασης 3.16. Ο Μπεϋζιανός κανόνας απόφασης σε αυτή την περίπτωση, όπως είδαμε, συνίσταται στην επιλογή της κλάσης με τη μεγαλύτερη ύστερη πιθανότητα $P(\omega_i|x)$. Σημειώνουμε ότι η μέση πιθανότητα σφάλματος 3.17 αναφέρεται και ως *ρυθμός σφάλματος* (error rate), ενώ ο κανόνας 3.16 ως *ταξινόμηση ελάχιστου ρυθμού σφάλματος* (minimum error rate classification).

Ένα ιδιαίτερα σύνθητες πρόβλημα κατάτμησης εικόνας αφορά τον εντοπισμό ή κατάτμηση ενός μόνο συγκεκριμένου αντικειμένου από μια εικόνα. Το πρόβλημα αυτό μπορεί να διατυπωθεί ευθέως στο πλαίσιο της Μπεϋζιανής θεωρίας απόφασης που περιγράψαμε σαν ένα πρόβλημα ταξινόμησης σε δύο κλάσεις, ω_1 και ω_2 . Σε ένα τέτοιο σενάριο δυαδικής ταξινόμησης (binary classification), η κλάση ω_1 μπορεί να θεωρηθεί ότι αντιστοιχεί στο αντικείμενο που μας ενδιαφέρει, ενώ η ω_2 στο υπόλοιπο της εικόνας (παρασκήνιο/φόντο). Για μια δεδομένη συνάρτηση απώλειας λ_{ij} , $i, j = 1, 2$, τα υπό συνθήκη ρίσκα των δύο δυνατών αποφάσεων γράφονται:

$$\begin{aligned} R(\omega_1|x) &= \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) \\ R(\omega_2|x) &= \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x) \end{aligned}$$

Οι τιμές λ_{ij} , στο πλαίσιο της δυαδικής απόφασης, μπορούν να χρησιμοποιηθούν για τον έλεγχο της σχέσης ανάμεσα στις συχνότητες εμφάνισης των δύο πιθανών ειδών σφαλμάτων: των ψευδώς αληθών αποφάσεων (λανθασμένη απόφαση υπέρ της ω_1 , false positives) και των ψευδώς αρνητικών αποφάσεων (λανθασμένη απόφαση υπέρ της ω_2 , false negatives). Ο Μπεϋζιανός κανόνας απόφασης 3.19, για αυτή την περίπτωση, αποφασίζοντας υπέρ της ω_2 στην περίπτωση ίσων ρίσκων, μπορεί να εκφραστεί ως:

$$\begin{aligned} \alpha(x) = \omega_1 &\Leftrightarrow R(\omega_1|x) < R(\omega_2|x) \\ &\Leftrightarrow (\lambda_{21} - \lambda_{11})P(\omega_1|x) > (\lambda_{12} - \lambda_{22})P(\omega_2|x) \\ &\Leftrightarrow (\lambda_{21} - \lambda_{11})p(x|\omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})p(x|\omega_2)P(\omega_2) \\ &\Leftrightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)} \end{aligned} \quad (3.21)$$

Σημειώνουμε ότι, καθώς η απώλεια λόγω ενός σφάλματος αναμένεται να είναι μεγαλύτερη από την απώλεια μιας σωστής απόφασης, οι παράγοντες $\lambda_{12} - \lambda_{22}$ και $\lambda_{21} - \lambda_{11}$ είναι θετικοί. Η γραφή

του κανόνα απόφασης όπως στην τελευταία γραμμή της 3.21 έχει μια χαρακτηριστική μορφή: το αριστερό μέλος της ανισότητας είναι μια συνάρτηση της παρατήρησης x , ενώ το δεξί μια ποσότητα ανεξάρτητη της x . Καθώς η $p(x|\omega_i)$, σαν συνάρτηση της ω_i , καλείται πιθανοφάνεια της ω_i , το αριστερό μέλος της 3.21 ονομάζεται *λόγος πιθανοφανειών* (likelihood ratio) [49, 35]. Ο κανόνας απόφασης 3.21 συνίσταται, συνεπώς, στη σύγκριση του λόγου των πιθανοφανειών $\frac{p(x|\omega_1)}{p(x|\omega_2)}$ με μια ποσότητα σταθερή ως προς x (κατώφλι). Ο έλεγχος αυτός για το λόγο πιθανοφανειών (likelihood ratio test) μπορεί να γραφεί στη γενικότερη μορφή:

$$\alpha(x) = \omega_1 \Leftrightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} > \tau, \quad (3.22)$$

όπου το κατώφλι τ μπορεί να είτε να τεθεί ίσο με το δεξί μέλος της 3.21, στο Μπεύζιανό θεωρητικό πλαίσιο, ή να προσδιοριστεί εμπειρικά, πράγμα που αποτελεί μια αρκετά συνηθισμένη πρακτική [48] σε εφαρμογές κατάτμησης αντικειμένου.

3.3.2 Κατάτμηση αντικειμένου με κατωφλίωση χρωματικών τιμών

Οι πιθανές τιμές του χρώματος ενός αντικειμένου είναι γενικά συγκεντρωμένες σε μια φραγμένη περιοχή του χώρου αναπαράστασής τους. Σε στατιστικούς όρους, το σύνολο σχεδόν της μάζας πιθανότητας της κατανομής χρωματικών τιμών του αντικειμένου ω , $p(x|\omega)$, είναι συγκεντρωμένο σε ένα φραγμένο χωρίο (“cluster”) του χώρου χρώματος που χρησιμοποιείται. Σε αρκετές περιπτώσεις, το σχήμα του χωρίου αυτού είναι τέτοιο ώστε να επιτρέπει μια απλή περιγραφή του συνόρου του. Το γεγονός αυτό καθιστά εύκολο τον έλεγχο για το αν ένα χρώμα (σημείο του χώρου χρώματος) ανήκει στην περιοχή αυτή και μπορεί να αξιοποιηθεί για τον ορισμό απλών μεθόδων κατάτμησης και εντοπισμού του αντικειμένου ενδιαφέροντος σε μια εικόνα.

Σύμφωνα με τις μεθόδους αυτές, που βασίζονται στη “ρητή” (explicit) περιγραφή του χωρίου που συγκεντρώνει τη μάζα της κατανομής $p(x|\omega)$, μια τιμή χρώματος x θεωρείται ότι ανήκει στην κλάση ω ακριβώς όταν το x ανήκει στο εν λόγω χωρίο. Η ιδέα αυτή είναι ισοδύναμη με απλή κατωφλίωση της πιθανοφάνειας $p(x|\omega)$ και το σύνολο απόφασης, που ταυτίζεται με το σύνολο του χωρίου, ορίζει μια ισοϋψή καμπύλη/επιφάνεια της $p(x|\omega)$. Παρότι μια τέτοια προσέγγιση έχει σαφή θεωρητικά μειονεκτήματα, καθώς δε λαμβάνει καθόλου υπόψη την πιθανοφάνεια άλλων κλάσεων/αντικειμένων στην εικόνα, βρίσκει εφαρμογή σε περιπτώσεις όπου η περιγραφή ή προσέγγιση του συνόρου απόφασης είναι ιδιαίτερα απλή, με αποτέλεσμα τον ορισμό πολύ γρήγορων υπολογιστικά μεθόδων κατάτμησης. Στις περισσότερες σχετικές εφαρμογές, η απόφαση ταξινόμησης λαμβάνεται με βάση μια σειρά κατωφλιώσεων των χρωματικών συνιστωσών ή απλών συναρτήσεών τους, οι οποίες ορίζουν μια προσέγγιση του συνόρου απόφασης.

Ο εντοπισμός και η κατάτμηση του ανθρώπινου δέρματος με απλές μεθόδους που λειτουργούν σε επίπεδο εικονοστοιχείου έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας [46, 47, 48] και το χωρίο των χρωμάτων δέρματος (“skin cluster”) έχει προσεγγιστεί με σχετικά απλό τρόπο σε αρκετούς χώρους χρώματος [53, 47, 54]. Οι προσεγγίσεις αυτές έχουν προκύψει τόσο μέσω εμπειρικής εκτίμησης όσο και με την εφαρμογή τεχνικών μηχανικής μάθησης [47]. Οι σχετικοί κανόνες απόφασης για κάποιους χώρους χρώματος δίνονται στον αλγόριθμο 1. Σημειώνουμε ότι μια γενική αδυναμία της εφαρμογής κάποιας τέτοιας μεθόδου κατωφλίωσης για την κατάτμηση δέρματος είναι ο έλεγχος των ψευδώς θετικών αποφάσεων (false positives), που είναι αποτέλεσμα της (ενδεχομένως υπερβολικά) απλουστευμένης προσέγγισης του συνόρου απόφασης. Παρόλα αυτά, οι απλές αυτές μέθοδοι δίνουν συχνά ικανοποιητικά αποτελέσματα, συγκρίσιμα με τις άλλες μεθόδους κατάτμησης δέρματος με βάση το χρώμα που λειτουργούν σε επίπεδο εικονοστοιχείου.

Algorithm 1 Skin segmentation using explicit color space regions

- RGB color space

Classify (R, G, B) as skin if and only if:

$$\begin{aligned} R > \frac{95}{255} \quad \wedge \quad G > \frac{40}{255} \quad \wedge \quad B > \frac{20}{255} \\ \wedge \quad \max\{R, G, B\} - \min\{R, G, B\} < \frac{15}{255} \\ \wedge \quad |R - G| > \frac{15}{255} \quad \wedge \quad R > \max\{G, B\} \end{aligned}$$

- rg chromaticity space (equation 3.1)

Classify (r, g) as skin if and only if:

$$\frac{r}{g} > 1.185 \quad \wedge \quad r(1 - r - g) > 0.107 \quad \wedge \quad rg > 0.112$$

- HSV color space

Classify (H, S, V) as skin if and only if [54]:

$$(0^\circ \leq H \leq 25^\circ \quad \vee \quad 335^\circ \leq H \leq 360^\circ) \quad \wedge \quad 0.2 \leq S \leq 0.6 \quad \wedge \quad V \geq 0.4$$

- YCbCr color space

Classify (Y, C_B, C_R) as skin if and only if:

$$\frac{77}{255} \leq C_B \leq \frac{127}{255} \quad \wedge \quad \frac{133}{255} \leq C_R \leq \frac{173}{255}$$

3.3.3 Παραδείγματα εφαρμογών

Στην ενότητα αυτή παρουσιάζουμε δύο απλές εφαρμογές κατάτμησης ενός αντικειμένου σε εικόνες με βάση το χρώμα. Και στα δύο παραδείγματα που ακολουθούν, συμβολίζουμε με ω_1 την κατηγορία που αντιστοιχεί στο αντικείμενο ενδιαφέροντος (“προσκήνιο”) και με ω_2 το “παρασκήνιο” της εικόνας. Η κατάτμηση πραγματοποιείται με βάση τον κανόνα απόφασης 3.21/3.22, για κατάλληλη επιλογή συνάρτησης απώλειας/τιμής κατωφλίου.

Εφαρμογή 1: Ημιαυτόματη κατάτμηση αντικειμένου

Το σενάριο αυτό περιλαμβάνει ένα αρχικό στάδιο πρόχειρης, χειροκίνητης “κατάτμησης” μιας εικόνας. Συγκεκριμένα, γίνεται ένας αδρός χειροκίνητος προσδιορισμός (μέρους του) αντικειμένου που μας ενδιαφέρει και (μέρους του) παρασκηνίου της εικόνας. Οι περιοχές που ορίζονται με αυτό τον τρόπο καθορίζουν τα δεδομένα εκπαίδευσης που θα χρησιμοποιηθούν στη συνέχεια για τις εκτιμήσεις των συναρτήσεων πιθανοφάνειας $p(x|\omega_1)$ και $p(x|\omega_2)$. Προφανώς, όσο ακριβέστερος είναι ο χειροκίνητος αυτός προσδιορισμός της περιοχής που αντιστοιχεί στην κλάση ω_i , $i = 1, 2$, τόσο περισσότερο η εκτίμηση της $p(x|\omega_i)$ θα ανταποκρίνεται στην πραγματικότητα. Σημειώνουμε ότι με τον τρόπο αυτό, εκτός από τις συναρτήσεις πιθανοφάνειας, μπορεί να γίνει και μια πρόχειρη

εκτίμηση των πρότερων πιθανοτήτων των κλάσεων $P(\omega_1)$ και $P(\omega_2)$. Καθώς, όμως, οι τιμές αυτές επηρεάζουν μόνο την τιμή του κατωφλίου τ του κανόνα απόφασης 3.21/3.22, η οποία θα προσδιοριστεί ούτως ή άλλως εμπειρικά, μπορούμε να τις αγνοήσουμε.

Στις εικόνες 3.9b και 3.9c φαίνονται αντίστοιχα μια χονδροειδής επισημείωση του αντικειμένου της εικόνας 3.9a που μας ενδιαφέρει και του παρασκήνιου της. Οι δύο αυτές μάσκες ορίζουν αντίστοιχα ποια δεδομένα της εικόνας θα χρησιμοποιηθούν για την εκτίμηση της $p(x|\omega_1)$ και της $p(x|\omega_2)$. Εργαζόμενοι στο χώρο χρώματος CIE LAB (a^* και b^* συνιστώσες μόνο), στα Σχήματα 3.10a και 3.10c απεικονίζουμε τα ιστογράμματα των χρωματικών τιμών στο επισημειωμένο αντικείμενο και παρασκήνιο αντίστοιχα. Με βάση την εποπτεία της κατανομής που μας παρέχουν τα ιστογράμματα, προσεγγίζουμε τις $p(x|\omega_1)$ και $p(x|\omega_2)$ σαν μείγματα Γκαουσιανών κατανομών με 2 και 3 συνιστώσες αντίστοιχα (Σχήματα 3.10b και 3.10d). Το αποτέλεσμα της κατάτμησης (μάσκα αντικειμένου) με βάση τον κανόνα 3.22 απεικονίζεται στο Σχήμα 3.11.

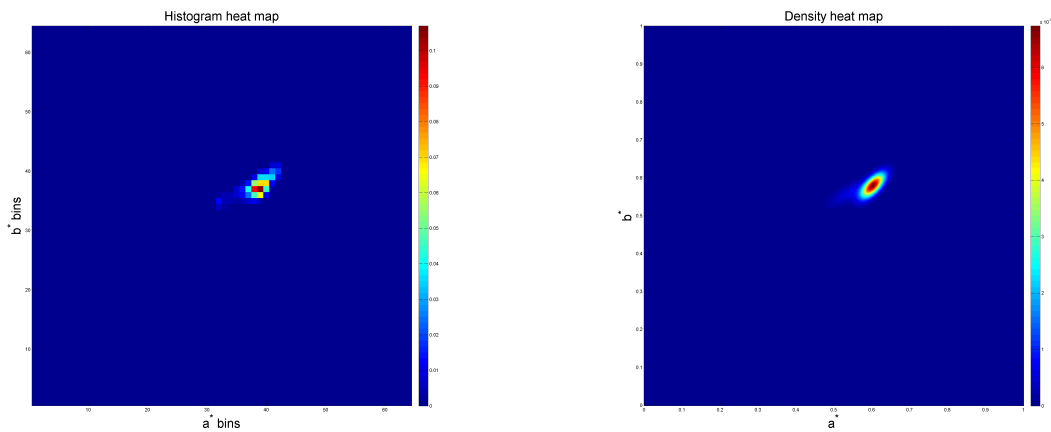


Σχήμα 3.9: Εφαρμογή ημιαυτόματης κατάτμησης

Εφαρμογή 2: Κατάτμηση δέρματος

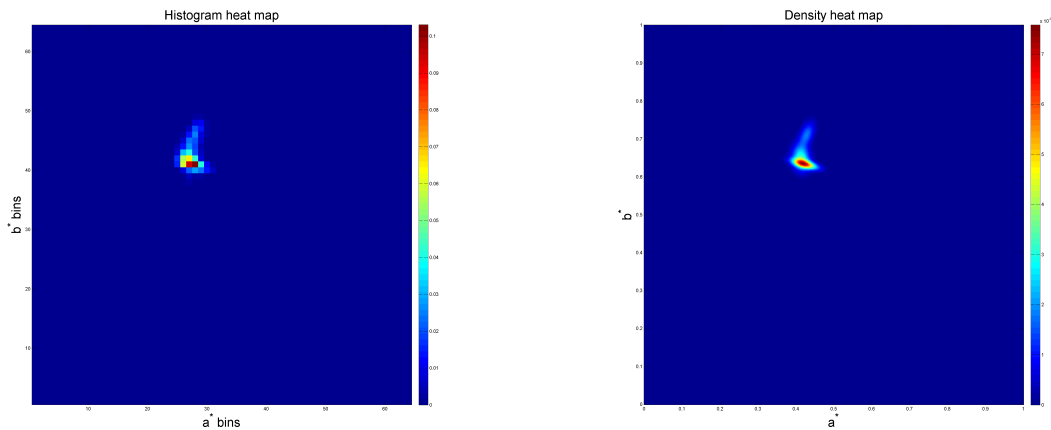
Στο παράδειγμα αυτό θα πραγματοποιήσουμε κατάτμηση του δέρματος του ομιλητή σε μια ακολουθία εικόνων νοηματικής γλώσσας. Θα ακολουθήσουμε το ίδιο σχήμα εκπαίδευσης μοντέλου χρώματος αντικειμένου και μοντέλου χρώματος παρασκήνιου όπως στην προηγούμενη εφαρμογή, χρησιμοποιώντας μια αρχική, πρόχειρη επισημείωση της περιοχής που αντιστοιχεί σε κάθε κλάση. Για την αρχική αυτή επισημείωση, χρησιμοποιούμε τον αλγόριθμο 1, ο οποίος επιστρέφει μια μάσκα για τις περιοχές της εικόνας που απεικονίζουν ανθρώπινο δέρμα. Η μάσκα αυτή μπορεί να χρησιμοποιηθεί απευθείας για τον ορισμό των δεδομένων εκπαίδευσης για την εκτίμηση της $p(x|\omega_1)$ και το συμπλήρωμά της, αντίστοιχα, για την $p(x|\omega_2)$. Καθώς έχουμε στη διάθεσή μας μια ολόκληρη ακολουθία εικόνων με τις κατηγορίες που θέλουμε να μοντελοποιήσουμε, μπορούμε να χρησιμοποιήσουμε οποιοδήποτε υποσύνολο εικόνων της για την εκπαίδευση του μοντέλου χρώματος κάθε κλάσης.

Σημειώνουμε ότι μάσκα του δέρματος που επιστρέφει ο αλγόριθμος 1 συνήθως παρουσιάζει μικρές έκτασης τεχνουργήματα (artifacts) επί του συνόρου της περιοχής του αντικειμένου, τα οποία εκτείνονται και σε μέρος του παρασκήνιου της εικόνας. Για την αποφυγή της ενσωμάτωσης αυτών των περιοχών στα δεδομένα εκπαίδευσης για την εκτίμηση της $p(x|\omega_1)$, μπορούμε να χρησιμοποιήσουμε μια πιο “συντηρητική” εκδοχή της μάσκας δέρματος που επιστρέφει ο αλγόριθμος 1, λαμβάνοντας το μορφολογικό άνοιγμα (morphological opening) της τελευταίας ως προς ένα δομικό στοιχείο (structuring element) κυκλικού δίσκου μικρής ακτίνας και χρησιμοποιώντας το αποτέλεσμα για τον ορισμό των δεδομένων εκπαίδευσης της $p(x|\omega_1)$. Συμμετρικά, για τον ορισμό των δεδομένων εκπαίδευσης της $p(x|\omega_2)$ (μοντέλο χρώματος παρασκήνιου), μπορούμε να χρησιμοποιήσουμε το μορφολογικό άνοιγμα του συμπληρώματος της μάσκας που επιστρέφει ο αλ-



(a) Ιστόγραμμα χρωματικών τιμών αντικειμένου.

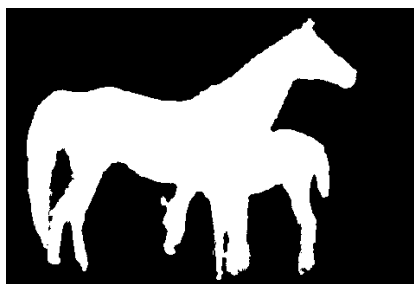
(b) Μείγμα Γκαουσιανών με 2 συνιστώσες για την προσέγγιση της $p(x|\omega_1)$.



(c) Ιστόγραμμα χρωματικών τιμών του παρασκηνίου της εικόνας.

(d) Μείγμα Γκαουσιανών με 3 συνιστώσες για την προσέγγιση της $p(x|\omega_2)$.

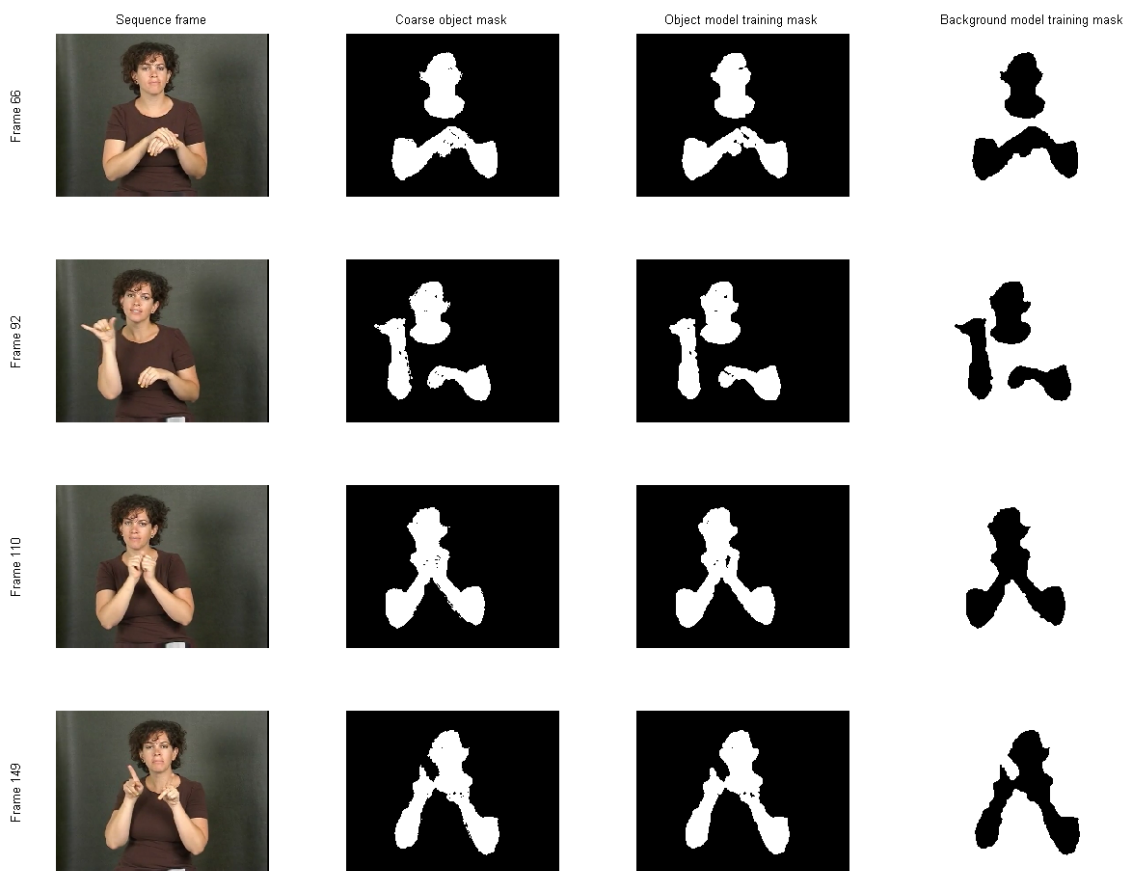
Σχήμα 3.10: Χρωματικά μοντέλα των αντικειμένων/κλάσεων της εικόνας 3.9a. Πάνω σειρά: μοντελοποίηση του χρώματος του αντικειμένου (ω_1). Κάτω σειρά: μοντελοποίηση του χρώματος του παρασκηνίου της εικόνας (ω_2).



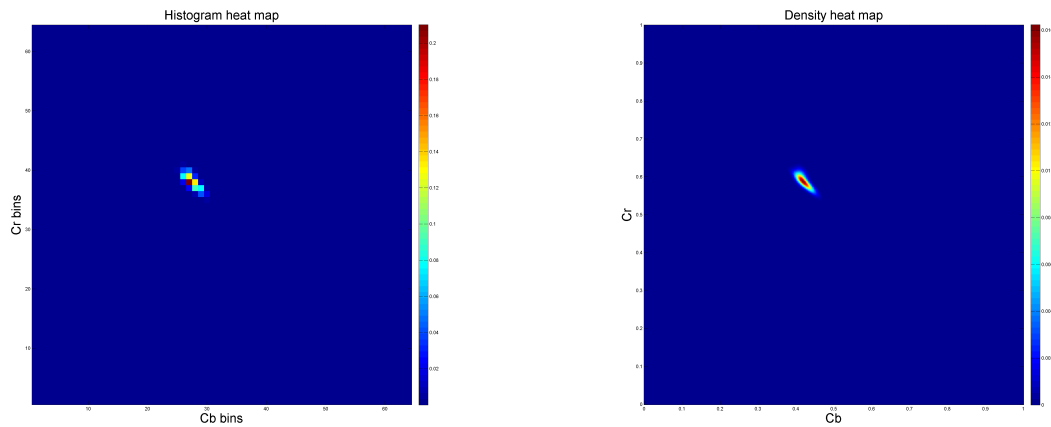
Σχήμα 3.11: Αποτέλεσμα κατάτμησης του αντικειμένου της 3.9a.

γόριθμος 1 ως προς κάποιο δομικό στοιχείο κυκλικού δίσκου μικρής ακτίνας. Με τον τρόπο αυτό βελτιώνουμε τη διαχωρισιμότητα των μετρήσεων που προέρχονται από τις δύο κλάσεις στο χώρο χρώματος. Μια εκτενής αναφορά στους μορφολογικούς τελεστές σε δυαδικές εικόνες μπορεί να βρεθεί στο Κεφάλαιο 7 του [2].

Στο Σχήμα 3.12 παρουσιάζονται οι μάσκες εκπαίδευσης για το αντικείμενο και το παρασκήνιο των εικόνων για 4 ενδεικτικά πλαίσια της ακολουθίας. Οι μάσκες προκύπτουν όπως περιγράψαμε παραπάνω, με την εφαρμογή μορφολογικών φίλτρων στο αποτέλεσμα του αλγορίθμου 1. Για την εκτίμηση της $p(x|\omega_1)$, χρησιμοποιήθηκαν τα πλαίσια 50-150 της ακολουθίας. Το ιστόγραμμα χρωματικών τιμών του αντικειμένου, με βάση αυτά τα πλαίσια, απεικονίζεται στο Σχήμα 3.13a, ενώ η $p(x|\omega_1)$ προσεγγίζεται τελικά σαν ένα μείγμα Γκαουσιανών με 2 συστατώσες στο Σχήμα 3.13b. Για την εκτίμηση της $p(x|\omega_2)$, χρησιμοποιήθηκαν τα 5 πρώτα πλαίσια της ακολουθίας. Το ιστόγραμμα των χρωματικών τιμών του παρασκήνιου απεικονίζεται στο Σχήμα 3.13c, ενώ η $p(x|\omega_2)$ προσεγγίζεται επίσης σαν ένα μείγμα Γκαουσιανών με 2 συστατώσες στο Σχήμα 3.13d. Τα τελικά αποτελέσματα της κατάτμησης του δέρματος του ομιλητή στα 4 πλαίσια της ακολουθίας του Σχήματος 3.12 με βάση τον κανόνα 3.22, για εμπειρική επιλογή του κατωφλίου τ , απεικονίζονται στο Σχήμα 3.14.

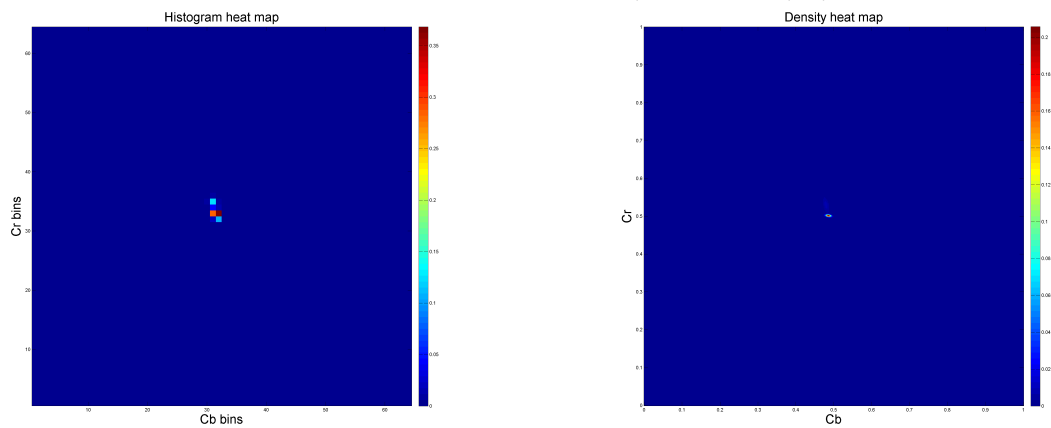


Σχήμα 3.12: Αρχική κατάτμηση και δεδομένα εκπαίδευσης μοντέλων χρώματος για 4 πλαίσια της ακολουθίας που χρησιμοποιήθηκε για την εφαρμογή κατάτμησης δέρματος. Πρώτη στήλη: εικόνα από την ακολουθία της δοκιμής. Δεύτερη στήλη: αρχική (πρόχειρη) κατάτμηση του αντικειμένου με βάση τον αλγόριθμο 1. Τρίτη στήλη: μάσκα ορισμού δεδομένων εκπαίδευσης για το μοντέλο χρώματος του αντικειμένου (μορφολογικό άνοιγμα της μάσκας της δεύτερης στήλης ως προς ένα κυκλικό δίσκο ακτίνας 5). Τέταρτη στήλη: μάσκα ορισμού δεδομένων εκπαίδευσης για το μοντέλο χρώματος του παρασκήνιου των εικόνων (μορφολογικό άνοιγμα του συμπληρώματος της μάσκας της δεύτερης στήλης ως προς ένα κυκλικό δίσκο ακτίνας 5).



(α) Ιστόγραμμα χρωματικών τιμών αντικειμένου.

(β) Μείγμα Γκαουσιανών με 2 συνιστώσες για την προσέγγιση της $p(x|\omega_1)$.



(α) Ιστόγραμμα χρωματικών τιμών του παρασκηνίου της ακολουθίας.

(β) Μείγμα Γκαουσιανών με 2 συνιστώσες για την προσέγγιση της $p(x|\omega_2)$.

Σχήμα 3.13: Χρωματικά μοντέλα των αντικειμένων/κλάσεων της ακολουθίας εικόνων που χρησιμοποιήθηκε για την εφαρμογή κατάτμησης δέρματος. Πάνω σειρά: μοντελοποίηση του χρώματος του αντικειμένου (ω_1). Κάτω σειρά: μοντελοποίηση του χρώματος του παρασκηνίου της ακολουθίας (ω_2).



Σχήμα 3.14: Τελικό αποτέλεσμα κατάτμησης του αντικειμένου για τα 4 πλαίσια της 3.12.

Κεφάλαιο 4

Εκτίμηση κίνησης: οπτική ροή

Η ανίχνευση και εκτίμηση της *οπτικής κίνησης* (κίνηση ορατών αντικειμένων στον τριδιάστατο χώρο) αποτελεί κρίσιμο συστατικό σχεδόν οποιουδήποτε συστήματος που βασίζεται στην οπτική αντίληψη για την ερμηνεία και κατανόηση του περιβάλλοντος κόσμου του. Η κίνηση σε μια ακολουθία εικόνων αποκαλύπτει μέρος της δομής της ορατής σκηνής, συσχετίζοντας χωρικά χαρακτηριστικά γνωρίσματα (spatial features) της εικόνας με χρονικές μεταβολές.

Στο κεφάλαιο αυτό θα αναφερθούμε στις πιο κλασικές μεθόδους εκτίμησης διδιάστατης κίνησης (κίνηση των προβολών των αντικειμένων στο επίπεδο της εικόνας) σε μια ακολουθία εικόνων. Όπως θα γίνει σαφές στο Κεφάλαιο 6, η εκτίμηση κίνησης αποτελεί μια ιδιαίτερα σημαντική λειτουργική μονάδα της μεθόδου οπτικής παρακολούθησης που υλοποιήσαμε.

4.1 Οπτική ροή

Η σχετική κίνηση της κάμερας ως προς το ορατό περιβάλλον έχει ως αποτέλεσμα την καταγραφή ενός μεταβαλλόμενου (κινούμενου) προτύπου φωτεινής έντασης από τον αισθητήρα της. Με τον τρόπο αυτό, ορίζεται ένα διδιάστατο διανυσματικό πεδίο που αντιστοιχεί στην ταχύτητα των προβολών των ορατών αντικειμένων στο επίπεδο της εικόνας και το οποίο καλείται *οπτική ροή* (optical flow). Το διανυσματικό πεδίο της οπτικής ροής, όντας η προβολή της τριδιάστατης κίνησης (ταχύτητας) των σημείων πάνω στο επίπεδο της εικόνας, χρησιμοποιείται σαν είσοδος ή αρχικό στάδιο επεξεργασίας σε ένα ευρύ φάσμα εφαρμογών, που περιλαμβάνουν σενάρια ανίχνευσης κίνησης, οπτικής παρακολούθησης και επίβλεψης, εκτίμησης κίνησης της κάμερας και τριδιάστατης ανακατασκευής της ορατής σκηνής, αυτόνομης πλοήγησης και ελέγχου της κίνησης ρομπότ και μεθόδους συμπίεσης/αναπαράστασης βίντεο.

Οι περισσότερες ευρέως χρησιμοποιούμενες μέθοδοι εκτίμησης της διδιάστατης κίνησης στο επίπεδο της εικόνας μπορούν να ενταχθούν σε μια από τις ακόλουθες γενικές κατηγορίες [2, 55, 7]:

- Μέθοδοι ταιριάσματος μπλοκ (block matching): οι μέθοδοι αυτές βασίζονται στην ελαχιστοποίηση του αθροίσματος των τετραγωνικών (ή απόλυτων) διαφορών σε μικρές χωρικές περιοχές (blocks) ανάμεσα σε δύο εικόνες για τον προσδιορισμό των διανυσμάτων μετατόπισης (displacement vectors). Οι μέθοδοι της κατηγορίας αυτής είναι συνδεδεμένες με τη μεγιστοποίηση της (κανονικοποιημένης) ετεροσυσχέτισης ανάμεσα στα μπλοκ των δύο εικόνων.
- Διαφορικές μέθοδοι ή μέθοδοι που βασίζονται στην κλίση (gradient-based) για την εκτίμηση της οπτικής ροής: οι μέθοδοι αυτές βασίζονται στη σχέση που συνδέει θεωρητικά τις

χωρικές και χρονικές παραγώγους της εικόνας (εξίσωση οπτικής ροής) και την επιβολή κάποιων επιπλέον κριτηρίων βελτιστότητας ή περιορισμών (regularization term or constraint) για το υπό αναζήτηση διανυσματικό πεδίο της ροής. Στην κατηγορία αυτή εντάσσονται οι (κλασικές πλέον) μέθοδοι:

- Horn-Schunck [56], η οποία διατυπώνεται ως η βελτιστοποίηση μιας αντικειμενικής συνάρτησης που επιβάλλει τετραγωνική ποινή τόσο στα “υπόλοιπα” της εξίσωσης οπτικής ροής όσο και, ολικά (globally), στις πρώτες παραγώγους του πεδίου ροής (μέσω του regularization/smoothness term που περιλαμβάνει) και
- Lucas-Kanade [34], η οποία βασίζεται στην υπόθεση της τοπικής σταθερότητας του πεδίου ροής στη γειτονιά του υπό εξέταση εικονοστοιχείου και λύνει το σύστημα των εξισώσεων οπτικής ροής που προκύπτουν από αυτή με τη μέθοδο των ελαχίστων τετραγώνων.

Σημειώνουμε ότι έχει προταθεί μεγάλος αριθμός μεθόδων μεταβολών (variational methods) που μπορούν να θεωρηθούν σαν παραλλαγές ή επεκτάσεις της Horn-Schunck και εισάγουν διαφορετικές συναρτήσεις ποινής τόσο για την ικανοποίηση της εξίσωσης οπτικής ροής, όσο και για την επιβολή κάποιας μορφής ομαλότητας στο υπό αναζήτηση διανυσματικό πεδίο (regularization/smoothness term). Ενδεικτικά, παραπέμπουμε στις εργασίες [57, 58, 59], οποίες χρησιμοποιούν κάποιον εύρωστο (robust) ορισμό της αντικειμενικής συνάρτησης που βασίζεται στην L^1 νόρμα ή κάποια διαφορίσιμη προσέγγισή της αντί για τη συνάρτηση τετραγωνικού σφάλματος της μεθόδου Horn-Schunck.

- Μέθοδοι που βασίζονται στην ανάλυση Fourier. Ενδεικτικά αναφέρουμε τις μεθόδους που βασίζονται στη φάση (phase-based), οι οποίες στηρίζονται στην εκτίμηση της τοπικής φάσης της χωροχρονικής συνάρτησης που αναπαριστά την κινούμενη εικόνα με βάση μια κατάλληλη συστοιχία φίλτρων. Η χρήση της πληροφορίας που περιέχεται στη φάση δικαιολογείται από την παρατήρηση ότι η φάση ενός σήματος που διέρχεται από ένα ζωνοπερατό φίλτρο είναι λιγότερο ευαίσθητη σε μεταβολές της φωτεινής έντασης από ό,τι το πλάτος του σήματος [55]. Οι τεχνικές αυτές μπορούν να θεωρηθούν σαν διαφορικές μέθοδοι που εφαρμόζονται στη φάση της χωροχρονικής συνάρτησης αντί απευθείας στην ένταση [55, 60].

Μια ποσοτική αξιολόγηση των αποτελεσμάτων που δίνουν οι κλασικές μέθοδοι εκτίμησης οπτικής ροής σε καθιερωμένες εικόνες εισόδου αναφοράς (benchmarks) μπορεί να βρεθεί στην εργασία [61]. Το ίδιο πραγματοποιείται και στην πιο πρόσφατη εργασία [62], όπου δίνεται έμφαση στη χρήση σύγχρονων τεχνικών βελτιστοποίησης και πρακτικών υλοποίησης για παραλλαγές των κλασικών μεθόδων, με εξαιρετικά αποτελέσματα.

Στις ενότητες που ακολουθούν, συμβολίζουμε με $I(x, y, t)$ τη χωροχρονική συνάρτηση φωτεινής έντασης που περιγράφει το σήμα της κινούμενης εικόνας, όπου οι συντεταγμένες (x, y) ορίζουν μια θέση στο επίπεδο της εικόνας και t είναι η συντεταγμένη χρόνου. Για λόγους απλότητας, περιορίζουμε την ανάλυσή μας σε γκριζες εικόνες. Συμβολίζουμε με (u, v) την τιμή του πεδίου διδιάστατης κίνησης (ταχύτητας/μετατόπισης) σε κάποια θέση του επιπέδου της εικόνας για κάποια χρονική στιγμή, όπου u και v είναι οι ταχύτητες/μετατοπίσεις στις κατευθύνσεις x και y αντίστοιχα. Σημειώνουμε ότι ο όρος μετατόπιση (displacement) είναι περισσότερο δόκιμος για την περίπτωση ακολουθιών εικόνων διακριτού χρόνου.

4.2 Μέθοδοι ταιριάσματος μπλοκ

Στην ενότητα αυτή θα αναφερθούμε σύντομα σε δύο μέλη της πρώτης κατηγορίας εκτίμησης διδιάστατης κίνησης που αναφέραμε παραπάνω, βασιζόμενοι κυρίως στα [2] και [4].

4.2.1 Απλό ταιρίασμα μπλοκ

Οι μέθοδοι “απλού ταιριάσματος μπλοκ” είναι γνωστές και ως μέθοδοι ευθυγράμμισης με βάση την μετατόπιση (translational alignment, [4]) και βασίζονται στην απλή ιδέα της προσέγγισης της σχετικής κίνησης ανάμεσα σε δύο εικόνες (ή περιοχές εικόνων) ως της παράλληλης μετατόπισης της μιας ως προς την άλλη, για την οποία η ομοιότητα των εικόνων/περιοχών μεγιστοποιείται με βάση κάποιο κριτήριο.

Έστω $I(x, y, t_1)$ και $I(x, y, t_2)$ δύο στιγμιότυπα της κινούμενης εικόνας τις χρονικές στιγμές t_1 και t_2 και R η γειτονιά (μπλοκ) της θέσης (x, y) . Η περιοχή R , ως υποσύνολο του επιπέδου της εικόνας, είναι συνάρτηση της θέσης (x, y) και συνήθως υποτίθεται ότι έχει συγκεκριμένο σχήμα και έκταση, ανεξαρτήτως της θέσης. Στην πράξη, μια συνήθης επιλογή είναι η R να είναι μια τετράγωνη (π.χ., 7×7) περιοχή με κέντρο το εικονοστοιχείο (x, y) . Για κάθε θέση (x, y) του επιπέδου της εικόνας, αναζητούμε κατάλληλο διάνυσμα μετατόπισης (u, v) , έτσι ώστε, στη γειτονιά της R (δηλαδή, για κάθε θέση $(x', y') \in R$), οι εικόνες $I(x', y', t_1)$ και $I(x' + u, y' + v, t_2)$ να είναι όσο γίνεται πιο “όμοιες”.

Η έννοια της τοπικής αυτής ομοιότητας της εικόνας τη χρονική στιγμή t_1 με την “παραμορφωμένη” (warped) εκδοχή της εικόνας τη στιγμή t_2 σύμφωνα με το πεδίο των (u, v) μπορεί να οριστεί με πολλούς τρόπους. Κάθε κριτήριο ομοιότητας ορίζει ένα αντίστοιχο βέλτιστο διανυσματικό πεδίο μετατοπίσεων. Οι πιο συνηθισμένοι στην πράξη ορισμοί βελτιστότητας της μετατόπισης (u, v) στη θέση (x, y) που βασίζονται σε κάποια μορφή απευθείας σύγκρισης των συναρτήσεων φωτεινής έντασης των δύο εικόνων συνοψίζονται στους ακόλουθους:

- Ελαχιστοποίηση του αθροίσματος των τετραγωνικών διαφορών (SSD, Sum of Squared Differences):

$$(u, v) = \operatorname{argmin}_{(\hat{u}, \hat{v})} \sum_{(x', y') \in R} [I(x', y', t_1) - I(x' + \hat{u}, y' + \hat{v}, t_2)]^2 \quad (4.1)$$

- Ελαχιστοποίηση του αθροίσματος των απόλυτων διαφορών (SAD, Sum of Absolute Differences):

$$(u, v) = \operatorname{argmin}_{(\hat{u}, \hat{v})} \sum_{(x', y') \in R} |I(x', y', t_1) - I(x' + \hat{u}, y' + \hat{v}, t_2)| \quad (4.2)$$

- Μεγιστοποίηση της ετεροσυσχέτισης (cross-correlation):

$$(u, v) = \operatorname{argmax}_{(\hat{u}, \hat{v})} \sum_{(x', y') \in R} I(x', y', t_1) I(x' + \hat{u}, y' + \hat{v}, t_2) \quad (4.3)$$

- Μεγιστοποίηση της κανονικοποιημένης ετεροσυσχέτισης (normalized cross-correlation):

$$(u, v) = \operatorname{argmax}_{(\hat{u}, \hat{v})} \sum_{(x', y') \in R} \frac{(I(x', y', t_1) - I_1)(I(x' + \hat{u}, y' + \hat{v}, t_2) - I_2)}{\sigma_1 \sigma_2}, \quad (4.4)$$

όπου I_1, I_2 είναι αντίστοιχα οι μέσες τιμές των $I(x', y', t_1), I(x' + \hat{u}, y' + \hat{v}, t_2)$ στην R και σ_1, σ_2 οι αντίστοιχες τυπικές αποκλίσεις.

Πέρα από το αυξημένο υπολογιστικό κόστος για την επίλυση των προβλημάτων βελτιστοποίησης που περιλαμβάνουν, οι μέθοδοι αυτής της κατηγορίας έχουν δύο σημαντικές θεωρητικές αδυναμίες [2]. Καταρχήν, δε λαμβάνεται υπόψη η προβολική παραμόρφωση (perspective distortion/foreshortening) μιας περιοχής R_1 (π.χ., της προβολής ενός κινούμενου αντικειμένου) της εικόνας τη χρονική στιγμή t_1 στην αντίστοιχη της R_2 στην εικόνα τη χρονική στιγμή t_2 . Επιπλέον, πέραν του ότι οι R_1 και R_2 έχουν γενικά διαφορετικό σχήμα, οι περιορισμοί των δύο εικόνων σε αυτές έχουν καταγραφεί υπό διαφορετικές συνθήκες φωτισμού, λόγω (και) της διαφορετικής γεωμετρίας τις δύο χρονικές στιγμές.

4.2.2 Αφινικό ταίριασμα μπλοκ

Μια λύση για τη μετρίαση των μειονεκτημάτων των μεθόδων της προηγούμενης υποενότητας δίνεται από τη χρήση του *αφινικού μοντέλου ταιριάσματος εικόνων* [63]. Οι μέθοδοι που περιγράψαμε υποθέτουν ένα απλό μοντέλο μετατόπισης (translational motion model) για την κίνηση που συνδέει τις δύο εικόνες. Αν R_1 και R_2 οι περιοχές που ορίζονται από την προβολή ενός κινούμενου αντικειμένου στις εικόνες για $t = t_1$ και $t = t_2$ αντίστοιχα, κερδίζουμε σημαντικά σε γενικότητα επιτρέποντας αφενός τις R_1 και R_2 να συνδέονται μέσω ενός αφινικού γεωμετρικού μετασχηματισμού (αντί μιας απλής μετατόπισης) και αφετέρου τους περιορισμούς των συναρτήσεων έντασης των δύο εικόνων στις περιοχές αυτές να συνδέονται επίσης μέσω ενός αφινικού μετασχηματισμού τους πλάτους τους (αντί να υποθέτουμε διατήρηση των τιμών της έντασης).

Ένας αφινικός μετασχηματισμός μπορεί να οριστεί σαν η σύνθεση ενός γραμμικού μετασχηματισμού και μιας μετατόπισης. Στις 2 διαστάσεις, μια τέτοια απεικόνιση παραμετροποιείται από ένα 2×2 πίνακα και ένα διδιάστατο διάνυσμα. Μια θέση (x, y) του επιπέδου, μέσω ενός αφινικού γεωμετρικού μετασχηματισμού, απεικονίζεται στην:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix} \Leftrightarrow (x, y) \mapsto (ax + by + u, cx + dy + v) \quad (4.5)$$

Ο 2×2 πίνακας της απεικόνισης περιγράφει μετασχηματισμούς περιστροφής (rotation) και (ανισοτροπικής) κλιμάκωσης (scaling), ενώ το διάνυσμα (προσθετικός όρος) της απεικόνισης περιγράφει τη χωρική μετατόπιση. Η κλάση των αφινικών γεωμετρικών μετασχηματισμών, παρότι όχι τόσο γενική όσο αυτή των προβολικών, είναι επαρκής για την επιτυχή μοντελοποίηση παραμορφώσεων λόγω κινήσεων υπολογίσιμου εύρους, ενώ έχειδειχθεί επιπλέον ότι περιγράφει ικανοποιητικά και παραμορφώσεις λόγω κίνησης παραμορφώσιμων (non-rigid) αντικειμένων [63]. Μια τιμή έντασης I , μέσω ενός αφινικού μετασχηματισμού πλάτους, απεικονίζεται στην:

$$I \mapsto eI + f,$$

όπου οι παράμετροι e , f ρυθμίζουν αντίστοιχα το δυναμικό εύρος του πλάτους και την DC συνιστώσα (offset) της έντασης. Στόχος αυτού του μετασχηματισμού είναι η σύνδεση των διαφορετικών συναρτήσεων έντασης στις δύο εικόνες λόγω διαφορετικών γεωμετρικών φωτισμού και προβολής σε καθεμία. Προφανώς, το απλό ταίριασμα μπλοκ που περιγράψαμε παραπάνω είναι ειδική περίπτωση του αφινικού μοντέλου, για $a = d = 1$, $b = c = 0$, $e = 1$ και $f = 0$.

Όπως και στην προηγούμενη υποενότητα, έστω $I(x, y, t_1)$ και $I(x, y, t_2)$ οι εικόνες για $t = t_1$ και $t = t_2$ αντίστοιχα και R μια δεδομένη γειτονιά (μπλοκ) της θέσης (x, y) . Στο πλαίσιο του αφινικού μοντέλου ταιριάσματος, αναζητούμε το σύνολο των παραμέτρων των δύο αφινικών μετασχηματισμών (γεωμετρικού και πλάτους), $\theta = (a, b, c, d, u, v, e, f)$, για τις οποίες, στη γειτονιά R μιας θέσης (x, y) (δηλαδή, για κάθε θέση $(x', y') \in R$), η εικόνα για $t = t_1$, $I(x', y', t_1)$, και η μετασχηματισμένη εικόνα για $t = t_2$, $eI(ax' + by' + u, cx' + dy' + v, t_2) + f$, να είναι όσο γίνεται

πιο “όμοιες”. Το κριτήριο της ελαχιστοποίησης του αθροίσματος των τετραγωνικών διαφορών (SSD), στο πλαίσιο του αφινικού ταιριάσματος, ορίζει το ακόλουθο πρόβλημα βελτιστοποίησης:

$$\theta = \operatorname{argmin}_{\hat{\theta}} \sum_{(x',y') \in R} \left[I(x', y', t_1) - \hat{e}I(\hat{a}x' + \hat{b}y' + \hat{u}, \hat{c}x' + \hat{d}y' + \hat{v}, t_2) - \hat{f} \right]^2, \quad (4.6)$$

όπου $\hat{\theta} = (\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{u}, \hat{v}, \hat{e}, \hat{f})$. Οι βέλτιστες τιμές των παραμέτρων \hat{u} και \hat{v} της συνιστώσας μετατόπισης του γεωμετρικού αφινικού μετασχηματισμού μας δίνουν τα ζητούμενα διανύσματα κίνησης (u, v) . Η 4.6 ορίζει ένα μη γραμμικό πρόβλημα βελτιστοποίησης σε ένα χώρο λύσεων 8 διαστάσεων και προσεγγιστικοί αλγόριθμοι για την επίλυσή του δίνονται στα [2, 63].

4.3 Διαφορικές μέθοδοι εκτίμησης οπτικής ροής

Στην ενότητα αυτή παρουσιάζουμε τις δύο πλέον κλασικές διαφορικές μεθόδους εκτίμησης οπτικής ροής. Οι μέθοδοι αυτής της κατηγορίας βασίζονται στη σχέση που συνδέει θεωρητικά τις χωρικές και χρονικές παραγωγούς της εικόνας, υποθέτοντας διατήρηση της έντασης για μικρές κινήσεις. Όπως θα δούμε, η εξίσωση που προκύπτει από αυτή την υπόθεση δεν αρκεί για τον πλήρη προσδιορισμό του διανύσματος ταχύτητας και απαιτείται η επιβολή κάποιων επιπλέον κριτηρίων βελτιστότητας ή περιορισμών για το υπό αναζήτηση διανυσματικό πεδίο της ροής. Οι επιπλέον αυτοί περιορισμοί/υποθέσεις, καθώς και η μορφή της συνάρτησης ποινής για την μη ικανοποίηση της εξίσωσης οπτικής ροής, αποτελούν τα βασικά σημεία διαφοροποίησης ανάμεσα στις μεθόδους αυτής της κατηγορίας.

4.3.1 Εξίσωση οπτικής ροής

Έστω $I(x, y, t)$ η χωροχρονική συνάρτηση που περιγράφει το καταγραφόμενο πρότυπο φωτεινής έντασης στο επίπεδο της εικόνας. Όπως αναφέραμε και στην εισαγωγική ενότητα, η κίνηση στον τριδιάστατο χώρο των σημείων της ορατής σκηνής ορίζει ένα διδιάστατο διανυσματικό πεδίο ταχυτήτων των προβολών τους στο επίπεδο της εικόνας, στο οποίο αναφερόμαστε με τον όρο οπτική ροή.

Έστω P ένα κινούμενο σημείο της τριδιάστατης σκηνής που τη χρονική στιγμή t προβάλλεται στη θέση (x, y) του επιπέδου της εικόνας. Μετά από χρόνο Δt , η προβολή του ίδιου τριδιάστατου σημείου P θα έχει μετακινηθεί στη θέση $(x + \Delta x, y + \Delta y)$. Υποθέτοντας ότι η συνάρτηση της κινούμενης εικόνας είναι ομαλή, οι τιμή της στο χωροχρονικό σημείο $(x + \Delta x, y + \Delta y, t + \Delta t)$ μπορεί να υπολογιστεί από το ανάπτυγμα Taylor της συνάρτησης γύρω από το σημείο (x, y, t) :

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + h, \quad (4.7)$$

όπου h είναι το άθροισμα των όρων ανώτερης (δεύτερης και άνω) τάξης. Καθώς στα σημεία (x, y, t) και $(x + \Delta x, y + \Delta y, t + \Delta t)$ είναι καταγεγραμμένη η προβολή του ίδιου σημείου P , είναι εύλογο να υποθέσουμε ότι η ένταση στα σημεία αυτά έχει την ίδια τιμή:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) \quad (4.8)$$

Η παραπάνω εξίσωση είναι γνωστή ως *περιορισμός σταθερότητας της έντασης* (brightness constancy constraint). Η 4.7, λόγω της 4.8 και αγνοώντας τους όρους ανώτερης τάξης, δίνει την εξίσωση:

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (4.9)$$

Θεωρώντας το όριο καθώς $\Delta t \rightarrow 0$, ορίζοντας τις συνιστώσες της οπτικής ροής ως:

$$u \equiv \frac{dx}{dt} \quad \text{και} \quad v \equiv \frac{dy}{dt}$$

και χρησιμοποιώντας το συνήθη συμβολισμό με δείκτες για τις μερικές παραγώγους, η εξίσωση 4.9 γράφεται:

$$I_x u + I_y v + I_t = 0 \quad (4.10)$$

Η παραπάνω σχέση είναι γνωστή ως εξίσωση οπτικής ροής και αποτελεί ουσιαστικά μια άλλη έκφραση του περιορισμού σταθερότητας της έντασης 4.8. Συγκεκριμένα, η ολική παράγωγος της I ως προς t γράφεται:

$$\frac{dI}{dt} = \frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} \frac{dt}{dt} = I_x u + I_y v + I_t$$

και, συνεπώς, η 4.10 εκφράζει ακριβώς την απαίτηση $\frac{dI}{dt} = 0$.

4.3.2 Το πρόβλημα του “ανοίγματος” (aperture problem)

Παρατηρούμε ότι ο περιορισμός οπτικής ροής 4.10 είναι μια εξίσωση ευθείας, την οποία το διάνυσμα ταχυτήτων $f = (u, v)$ στη θέση (x, y) απαιτείται να ικανοποιεί με βάση τις υποθέσεις μας. Καθώς ο περιορισμός αυτός εκφράζεται σαν μια εξίσωση με δύο αγνώστους, υπάρχει μια μονοπαραμετρική απειρία λύσεων που τον ικανοποιούν. Συνεπώς, η εκτίμηση της διδιάστατης ταχύτητας μόνο από την 4.10 αποτελεί ένα κακώς ορισμένο (ill-posed) πρόβλημα. Η αδυναμία αυτή είναι γνωστή ως το πρόβλημα του “ανοίγματος” (aperture problem) στην εκτίμηση διδιάστατης κίνησης.

Μια ισοδύναμη γραφή της εξίσωσης οπτικής ροής 4.10 δίνεται από την:

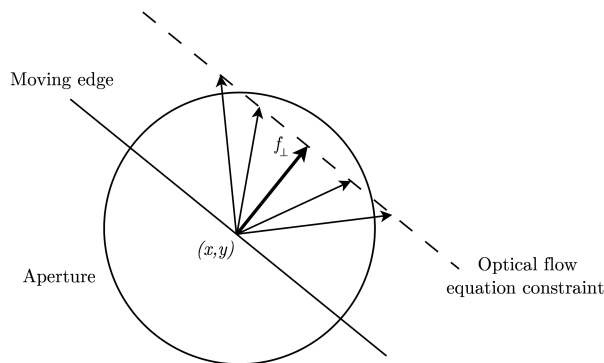
$$\nabla I \cdot (u, v) = -I_t \Leftrightarrow \nabla I \cdot f = -I_t, \quad (4.11)$$

όπου $\nabla I = (I_x, I_y)$ είναι η χωρική κλίση (spatial gradient) της εικόνας τη χρονική στιγμή t . Σύμφωνα με την 4.11, είναι γνωστή η προβολή του διανύσματος ταχύτητας στην κατεύθυνση της κλίσης της εικόνας, ή, ισοδύναμα, η συνιστώσα της ταχύτητας που είναι κάθετη στις ισοϋψείς καμπύλες της εικόνας τη χρονική στιγμή t . Αν συμβολίσουμε με f_{\perp} αυτή τη συνιστώσα της ταχύτητας, η οποία είναι γνωστή ως *κάθετη ροή* (normal flow), και με n το μοναδιαίο διάνυσμα στην κατεύθυνση της κλίσης, ισχύει:

$$f_{\perp} = -\frac{I_t}{\|\nabla I\|} n = -\frac{I_t}{\sqrt{I_x^2 + I_y^2}} n, \quad \text{όπου} \quad n = \frac{\nabla I}{\|\nabla I\|} \quad (4.12)$$

Σαν συνέπεια του προβλήματος του ανοίγματος, δεν είναι δυνατός ο τοπικός προσδιορισμός της συνιστώσας $f_{\parallel} = f - f_{\perp}$ της ταχύτητας που είναι παράλληλη στις ισοϋψείς καμπύλες της εικόνας τη στιγμή t , ή, ισοδύναμα, κάθετη στη χωρική της κλίση. Μια απεικόνιση του προβλήματος δίνεται στο Σχήμα 4.1.

Παρότι η εξίσωση της οπτικής ροής διατυπώνεται για μια μεμονωμένη θέση του επιπέδου της εικόνας, οποιαδήποτε διακριτή προσέγγιση των χωρικών και χρονικών παραγώγων της κινούμενης εικόνας απαιτεί τη θεώρηση μιας γειτονιάς της θέσης όπου γίνεται η εκτίμηση. Το γεγονός αυτό δημιουργεί το ερώτημα του πώς η αναζήτηση της ταχύτητας f σε κάποια θέση της εικόνας μπορεί να επωφεληθεί από τη θεώρηση μιας πεπερασμένης έκτασης γειτονιάς της θέσης



Σχήμα 4.1: Το πρόβλημα του “ανοίγματος” στην τοπική εκτίμηση διδιάστατης κίνησης. Το έντονα σχεδιασμένο διάνυσμα συμβολίζει τη συνιστώσα της κάθετης ροής f_{\perp} . Η διακεκομμένη ευθεία αναπαριστά τον περιορισμό που θέτει η εξίσωση οπτικής ροής 4.10: οποιοδήποτε διάνυσμα ταχύτητας f την ικανοποιεί αποτελεί μια τοπική λύση.

αυτής, αντί της τελευταίας μεμονωμένα. Συγκεκριμένα, η μεταβλητότητα της κατεύθυνσης του διανύσματος της χωρικής κλίσης της εικόνας μέσα σε μια τέτοια γειτονιά/άνοιγμα (aperture), υπό κατάλληλες υποθέσεις, μπορεί να χρησιμοποιηθεί για την κατάρτιση επιπλέον περιορισμών που επιτρέπουν τη μονοσήμαντη εκτίμηση της διδιάστατης ροής f . Το πρόβλημα, βεβαίως, εξακολουθεί να υπάρχει αν η κατεύθυνση της κλίσης της εικόνας είναι σταθερή στην έκταση του θεωρούμενου ανοίγματος. Έχει προταθεί ένας μεγάλος αριθμός μεθόδων που βασίζονται στην 4.10 για τον ορισμό της αντικειμενικής τους συνάρτησης, ενσωματώνοντας επιπλέον περιορισμούς που αφορούν μια πεπερασμένη περιοχή της εικόνας. Στις ακόλουθες υποενότητες, παρουσιάζουμε ενδεικτικά δύο κλασικές διαφορικές μεθόδους εκτίμησης οπτικής ροής: αυτήν των Horn-Schunck, η οποία επιβάλλει έναν ολικό (global) περιορισμό ομαλότητας για το πεδίο της ροής και αυτήν των Lucas-Kanade, η οποία βασίζεται στην υπόθεση της τοπικής (local) σταθερότητας της ροής σε μια περιοχή του υπό εξέταση εικονοστοιχείου.

4.3.3 Η μέθοδος Horn-Schunck

Στα περισσότερα πρακτικά σενάρια, η εξίσωση οπτικής ροής 4.10 δεν μπορεί να ικανοποιηθεί με ακρίβεια σε κάθε θέση του επιπέδου της εικόνας, λόγω είτε αποτυχίας της υπόθεσης σταθερότητας της έντασης 4.8 ως ακριβούς ισότητας είτε της ύπαρξης θορύβου στο σήμα της εικόνας. Έτσι, μια εύλογη απαίτηση για το πεδίο (u, v) είναι αυτό να ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα (“υπόλοιπο”) ικανοποίησης της εξίσωσης οπτικής ροής, ή, ισοδύναμα, τον όρο:

$$E_b = \iint (I_x u + I_y v + I_t)^2 \, dx dy, \quad (4.13)$$

όπου η ολοκλήρωση λαμβάνει χώρα σε όλη την έκταση του επιπέδου της εικόνας.

Όπως αναφέραμε στην προηγούμενη υποενότητα, ο υπολογισμός του διανύσματος ταχύτητας ανεξάρτητα σε κάθε θέση αποτελεί ένα κακώς ορισμένο πρόβλημα. Ένας επιπλέον περιορισμός για τις άγνωστες ταχύτητες μπορεί να προέλθει από την παρατήρηση ότι γειτονικά σημεία της προβολής ενός κινούμενου αντικείμενου αναμένεται να κινούνται με περίπου την ίδια ταχύτητα. Έτσι, το πεδίο των ταχυτήτων αναμένεται να μεταβάλλεται σχεδόν παντού ομαλά, με εξαίρεση τις περιοχές όπου ένα κινούμενο αντικείμενο αποκρύπτει κάποιο άλλο. Αυτές οι περιοχές ασυνέχειας του πεδίου των ταχυτήτων είναι γνωστές ως *σύνορα κίνησης/απόκρυψης* (motion/occlusion boundaries). Η πιο συνήθης έκφραση ενός τέτοιου περιορισμού ομαλότητας δίνεται από την απαι-

τηση το μέσο άθροισμα των τετραγώνων των μέτρων των κλίσεων των συνιστωσών u και v του πεδίου να είναι ελάχιστο [56]. Η προς ελαχιστοποίηση ποσότητα δίνεται από τον όρο:

$$E_s = \iint (\|\nabla u\|^2 + \|\nabla v\|^2) \, dx dy = \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) \, dx dy, \quad (4.14)$$

όπου η ολοκλήρωση λαμβάνει χώρα σε όλη την έκταση του επιπέδου της εικόνας.

Η εκτίμηση της οπτικής ροής με τη μέθοδο Horn-Schunck διατυπώνεται ως το πρόβλημα της ελαχιστοποίησης της αντικειμενικής συνάρτησης που ορίζεται από τον όρο E_b της 4.13 και τον όρο κανονικοποίησης (regularization term) E_s της 4.14. Ο όρος E_b εκφράζει τον περιορισμό της τοπικής διατήρησης της έντασης και συχνά καλείται ως όρος δεδομένων (data term). Ο όρος κανονικοποίησης E_s εκφράζει έναν ολικό περιορισμό ομαλότητας (global smoothness constraint) για το ζητούμενο πεδίο. Αν λ είναι ένας μη αρνητικός συντελεστής που καθορίζει το βάρος (σημασία) που έχει η απαίτηση ομαλότητας του πεδίου σε σχέση την απαίτηση ικανοποίησης του περιορισμού της οπτικής ροής στον ορισμό της αντικειμενικής συνάρτησης, τότε η τελευταία γράφεται:

$$E(u, v) = E_b + \lambda E_s = \iint (I_x u + I_y v + I_t)^2 \, dx dy + \lambda \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) \, dx dy \quad (4.15)$$

Η σχεδιαστική επιλογή της παραμέτρου κανονικοποίησης λ αποτελεί μέρος του ορισμού του προβλήματος.

Η ελαχιστοποίηση της 4.15 αποτελεί πρόβλημα του λογισμού μεταβολών. Η υπό ολοκλήρωση ποσότητα είναι μια συνεχής χωροχρονική συνάρτηση:

$$e(x, y, t) = (I_x u + I_y v + I_t)^2 + \lambda (u_x^2 + u_y^2 + v_x^2 + v_y^2) \quad (4.16)$$

Μια αναγκαία συνθήκη για την ελαχιστότητα της 4.15 είναι η στασιμότητα της e ως προς u και v , η οποία εξασφαλίζεται από την απαίτηση οι συναρτησιακές παράγωγοι (functional/variational derivatives) της e να είναι ίσες με 0:

$$\begin{aligned} \nabla_u e &= \frac{\partial e}{\partial u} - \frac{\partial}{\partial x} \frac{\partial e}{\partial u_x} - \frac{\partial}{\partial y} \frac{\partial e}{\partial u_y} = 2I_x (I_x u + I_y v + I_t) - 2\lambda(u_{xx} + u_{yy}) = 0 \\ \nabla_v e &= \frac{\partial e}{\partial v} - \frac{\partial}{\partial x} \frac{\partial e}{\partial v_x} - \frac{\partial}{\partial y} \frac{\partial e}{\partial v_y} = 2I_y (I_x u + I_y v + I_t) - 2\lambda(v_{xx} + v_{yy}) = 0 \end{aligned}$$

Οι παραπάνω είναι οι εξισώσεις Euler-Lagrange για το συγκεκριμένο πρόβλημα και ξαναγράφονται ως:

$$\begin{aligned} \lambda \nabla^2 u &= (I_x u + I_y v + I_t) I_x \\ \lambda \nabla^2 v &= (I_x u + I_y v + I_t) I_y \end{aligned} \quad (4.17)$$

Οι εξισώσεις 4.17 ορίζουν ένα συζευγμένο σύστημα (δευτεροβάθμιων) ελλειπτικών μερικών διαφορικών εξισώσεων.

Στη συνέχεια, θα παρουσιάσουμε μια αριθμητική μέθοδο επίλυσης των 4.17, η οποία εντάσσεται στην κατηγορία των μεθόδων πεπερασμένων διαφορών. Θεωρούμε δεδομένη μια διακριτοποιημένη συνάρτηση $I[i, j, k]$ για την κινούμενη εικόνα, όπου η τιμή $I[i, j, k]$ αντιστοιχεί στην $I(i\Delta x, j\Delta y, k\Delta t)$. Για λόγους απλότητας, υποθέτουμε ότι το πλέγμα της διακριτοποίησης είναι ομοιόμορφο και επιπλέον ότι $\Delta x = \Delta y = \Delta t = 1$. Το ζητούμενο είναι η εκτίμηση του πεδίου των διδιάστατων ταχυτήτων ($u[i, j, k], v[i, j, k]$) στο διακριτοποιημένο επίπεδο της εικόνας, για κάποια δεδομένη διακριτή χρονική τιμή k . Οι χωρικές και η χρονική μερικές παράγωγοι της κινούμενης εικόνας προσεγγίζονται από τις μέσες τιμές των προς τα εμπρός διαφορών (forward

differences) σε ένα χωροχρονικό παράθυρο (κύβο) διαστάσεων $2 \times 2 \times 2$. Για παράδειγμα, η τιμή της παραγώγου $I_x[i, j, k]$ λαμβάνεται ίση με:

$$I_x[i, j, k] = \frac{1}{4}(I[i+1, j, k] - I[i, j, k] + I[i+1, j+1, k] - I[i, j+1, k] + I[i+1, j, k+1] - I[i, j, k+1] + I[i+1, j+1, k+1] - I[i, j+1, k+1])/\Delta x \quad (4.18)$$

Για τη διακριτή εκδοχή του τελεστή Laplace, συνήθως χρησιμοποιείται μια κεντρική συμμετρική προσέγγιση της μορφής:

$$\nabla^2 w[i, j, k] = \kappa (\bar{w}[i, j, k] - w[i, j, k]), \quad (4.19)$$

όπου χρησιμοποιήσαμε καταχρηστικά το ίδιο σύμβολο για το διακριτό τελεστή Laplace, $\bar{w}[i, j, k]$ είναι ένας σταθμισμένος μέσος από τιμές σε ένα 3×3 χωρικό παράθυρο με κέντρο τη θέση (i, j) και κ είναι ένας παράγοντας που εξαρτάται από τον τύπο του σταθμισμένου μέσου. Με w συμβολίζουμε με συντομία μια από τις συνιστώσες u και v του ζητούμενου πεδίου. Οι Horn και Schunck στο [56], προτείνουν ένα σχήμα προσέγγισης που εκμεταλλεύεται όλα τα στοιχεία του συμμετρικού παραθύρου, όπου $\kappa = 3$ και:

$$\bar{w}[i, j, k] = \frac{1}{6} (w[i-1, j, k] + w[i+1, j, k] + w[i, j-1, k] + w[i, j+1, k]) + \frac{1}{12} (w[i-1, j-1, k] + w[i-1, j+1, k] + w[i+1, j-1, k] + w[i+1, j+1, k]) \quad (4.20)$$

Περισσότερο συνηθισμένη στη βιβλιογραφία είναι η προσέγγιση με $\kappa = 4$ και:

$$\bar{w}[i, j, k] = \frac{1}{4} (w[i-1, j, k] + w[i+1, j, k] + w[i, j-1, k] + w[i, j+1, k]) \quad (4.21)$$

Υιοθετώντας την τελευταία, η 4.19 γράφεται:

$$\nabla^2 w[i, j, k] = 4(\bar{w}[i, j, k] - w[i, j, k]), \quad (4.22)$$

όπου η τιμή $\bar{w}[i, j, k]$ δίνεται από την 4.21.

Αντικαθιστώντας την 4.22 στις 4.17 και θέτοντας $\alpha = 4\lambda$, λαμβάνουμε τις εξισώσεις:

$$\begin{aligned} (I_x^2 + \alpha)u + I_x I_y v &= \alpha \bar{u} - I_x I_t \\ I_x I_y u + (I_y^2 + \alpha)v &= \alpha \bar{v} - I_y I_t \end{aligned} \quad (4.23)$$

Σύμφωνα με το σχήμα διακριτοποίησης που ορίσαμε προηγουμένως, είναι εύκολο να επαληθεύσουμε ότι οι 4.23 ορίζουν ένα $2N \times 2N$ γραμμικό σύστημα, όπου N είναι το πλήθος των διακριτών χωρικών θέσεων (εικονοστοιχεία της εικόνας). Λόγω του μεγέθους του πίνακα του συστήματος, για την επίλυση των 4.23 προτιμούνται στην πράξη επαναληπτικές μέθοδοι. Μια απλή μέθοδος (τύπου Gauss-Seidel) μπορεί να οριστεί αν θεωρήσουμε τις μέσες τιμές \bar{u} και \bar{v} γνωστές (από την προηγούμενη επανάληψη) και λύσουμε τις 4.23 ως προς u και v . Χρησιμοποιώντας εκθέτες για το συμβολισμό του βήματος της επανάληψης και παραλείποντας τους δείκτες $[i, j, k]$, το βήμα της μεθόδου ορίζεται από τις εξισώσεις:

$$\begin{aligned} u^{n+1} &= \bar{u}^n - I_x \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{\alpha + I_x^2 + I_y^2} \\ v^{n+1} &= \bar{v}^n - I_y \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{\alpha + I_x^2 + I_y^2} \end{aligned} \quad (4.24)$$

Οι τιμές \bar{u}^n και \bar{v}^n υπολογίζονται από τη σχέση 4.21 όταν έχει ολοκληρωθεί ο υπολογισμός των u^n και v^n σε όλες τις θέσεις. Η αρχικοποίηση (ανάθεση τιμών στις u^0 και v^0) μπορεί να γίνει αυθαίρετα (για παράδειγμα, παντού 0). Ο έλεγχος της σύγκλισης μπορεί να γίνει με βάση το εάν η τιμή μιας διακριτοποιημένης εκδοχής της αντικειμενικής συνάρτησης 4.15 έχει πρακτικά σταματήσει να μειώνεται. Το αριθμητικό επαναληπτικό σχήμα που ορίσαμε για τη μέθοδο των Horn-Schunck δίνεται υπό τη μορφή ψευδοκώδικα στον Αλγόριθμο 2.

Σημειώνουμε ότι μπορούσαμε να έχουμε καταλήξει στην ίδια επαναληπτική μέθοδο ξεκινώντας από μια διακριτοποιημένη εκδοχή της 4.16 και εξάγοντας τις αναγκαίες συνθήκες βελτιστότητας (σε διακριτή μορφή) απευθείας από αυτή, έχοντας προσεγγίσει τις χωρικές μερικές παραγώγους των συνιστωσών u και v με προς τα εμπρός διαφορές [2, 64].

Algorithm 2 Horn-Schunck optical flow estimation

Input: Discretized image function $I[i, j, k]$, regularization parameter λ , initial estimates for velocity components u^0, v^0

Output: 2D velocity components u, v

```

1:  $n = 0$ 
2: repeat
3:   for all image pixels do
4:     Compute  $\bar{u}^n$  and  $\bar{v}^n$  according to 4.21 from  $u^n$  and  $v^n$ 
5:     /* Image partial derivatives are computed according to scheme 4.18: */
6:      $u^{n+1} = \bar{u}^n - I_x \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{4\lambda + I_x^2 + I_y^2}$ 
7:      $v^{n+1} = \bar{v}^n - I_y \frac{I_x \bar{u}^n + I_y \bar{v}^n + I_t}{4\lambda + I_x^2 + I_y^2}$ 
8:   end for
9:    $n = n + 1$ 
10: until convergence is achieved
  
```

Η ενσωμάτωση του όρου ομαλότητας στην αντικειμενική συνάρτηση 4.15 έχει ως αποτέλεσμα τη “διάδοση” του πεδίου των ταχυτήτων σε περιοχές της εικόνας με χαμηλή μεταβλητότητα της χωρικής της κλίσης (απουσία υψής), προσφέροντας μια κομψή λύση στο πρόβλημα του ανοίγματος. Παρόλα αυτά, η επίδραση του όρου ομαλότητας είναι ανεπιθύμητα υπαρκτή και στις περιοχές ασυνέχειας του πεδίου ταχυτήτων (σύνορα κίνησης/απόκρυψης). Για την αποφυγή του φαινομένου, έχουν προταθεί πιο σύνθετοι όροι ομαλότητας που λαμβάνουν υπόψη αυτή την επιπλέον απαίτηση [55].

4.3.4 Η μέθοδος Lucas-Kanade

Οι Lucas και Kanade [34] προτείνουν μια *τοπική εκτίμηση σταθμισμένων ελαχίστων τετραγώνων* (local weighted least squares) για την εξίσωση οπτικής ροής 4.10 σε μια δεδομένη θέση, υποθέτοντας ότι το πεδίο της οπτικής ροής (u, v) είναι σταθερό σε μια μικρή γειτονιά \mathcal{R} της θέσης υπολογισμού. Η συνάρτηση σφάλματος που πρέπει να ελαχιστοποιηθεί ορίζεται ως το άθροισμα των τετραγώνων των υπολοίπων της εξίσωσης οπτικής ροής, σταθμισμένων με βάση τις τιμές μιας χωρικής συνάρτησης στάθμισης $w(x, y)$, το στήριγμα (support) της οποίας καθορίζει τη γειτονιά \mathcal{R} της θεωρούμενης θέσης.

Η αρχική διατύπωση του προβλήματος [34] περιλάμβανε την ελαχιστοποίηση ενός σταθμισμένου διακριτού αθροίσματος τετραγώνων των υπολοίπων της εξίσωσης οπτικής ροής στην \mathcal{R} . Στα [55, 7] δίνεται μια μαθηματικά ισοδύναμη, συνεχής διατύπωση, όπου η συνάρτηση σφάλματος λαμβάνει τη μορφή ενός συνελκτικού ολοκληρώματος. Για μια δεδομένη θέση του επιπέδου της

εικόνας (x, y) (και χρονική στιγμή t), η αντικειμενική συνάρτηση ορίζεται από τη σχέση:

$$E(u, v) = \int_{-\infty}^{+\infty} w(x - x', y - y') (I_x(x', y', t)u + I_y(x', y', t)v + I_t(x', y', t))^2 dx' dy', \quad (4.25)$$

Στην πράξη, ως συνάρτηση στάθμισης επιλέγεται συχνά ένας Γκαουσιανός πυρήνας, καθώς δίνει μεγαλύτερη βαρύτητα στους περιορισμούς οπτικής ροής που αντιστοιχούν σε θέσεις κοντά στο κέντρο της γειτονιάς \mathcal{R} . Γενικά, η συνάρτηση w μπορεί να σταθμίζει κάθε θέση σύμφωνα με κάποιο μέτρο εμπιστοσύνης, όπως, για παράδειγμα, το μέτρο της κλίσης της εικόνας [55]. Οι τιμές των u και v που ελαχιστοποιούν την 4.25 αποτελούν την εκτίμηση σταθμισμένων ελαχίστων τετραγώνων για την οπτική ροή στη θέση (x, y) (τη χρονική στιγμή t).

Η συνάρτηση E της 4.25 γίνεται ελάχιστη για τιμές των u και v που ικανοποιούν τις αναγκαίες συνθήκες μηδενισμού των μερικών παραγώγων της ως προς u και v :

$$\begin{aligned} \frac{\partial E}{\partial u} = 0 &\Leftrightarrow \int_{-\infty}^{+\infty} w(x - x', y - y') (I_x u + I_y v + I_t) I_x dx' dy' = 0 \\ \frac{\partial E}{\partial v} = 0 &\Leftrightarrow \int_{-\infty}^{+\infty} w(x - x', y - y') (I_x u + I_y v + I_t) I_y dx' dy' = 0 \end{aligned}, \quad (4.26)$$

όπου όλες οι μερικές παράγωγοι της εικόνας υπολογίζονται στο σημείο (x', y', t) . Σύμφωνα με την υπόθεση σταθερότητας των u και v σε όλη την έκταση της γειτονιάς \mathcal{R} που σταθμίζει η w , είναι εύκολο να δούμε ότι οι εξισώσεις 4.26 ορίζουν ένα 2×2 σύστημα με αγνώστους τα u και v . Χρησιμοποιώντας τον ακόλουθο συμβολισμό για τα συνελικτικά ολοκληρώματα της μορφής:

$$\langle g \rangle_w \equiv \int_{-\infty}^{+\infty} w(x - x', y - y') g dx' dy',$$

οι 4.26 ξαναγράφονται στη μητρική μορφή:

$$\underbrace{\begin{bmatrix} \langle I_x^2 \rangle_w & \langle I_x I_y \rangle_w \\ \langle I_x I_y \rangle_w & \langle I_y^2 \rangle_w \end{bmatrix}}_A \underbrace{\begin{bmatrix} u \\ v \end{bmatrix}}_f = \underbrace{\begin{bmatrix} -\langle I_x I_t \rangle_w \\ -\langle I_y I_t \rangle_w \end{bmatrix}}_b \quad (4.27)$$

Εφόσον ο πίνακας A είναι αντιστρέψιμος, η λύση του συστήματος 4.27 που δίνεται από την:

$$f = A^{-1}b \quad (4.28)$$

ορίζει την εκτίμηση ελαχίστων τετραγώνων για τη ροή $(u, v) = f^T$ στη θέση (x, y) .

Αν η κατεύθυνση της κλίσης της εικόνας είναι σταθερή σε όλη την έκταση της \mathcal{R} , ο πίνακας του συστήματος 4.27 δεν είναι αντιστρέψιμος: η θεώρηση των περιορισμών από την εξίσωση οπτικής ροής 4.10 στο σύνολο της γειτονιάς \mathcal{R} δεν προσθέτει κάποια επιπλέον πληροφορία από αυτήν του σημείου υπολογισμού και το πρόβλημα του ανοίγματος παραμένει. Στην περίπτωση αυτή, μόνο η κάθετη ροή μπορεί να υπολογιστεί για τη θεωρούμενη θέση του επιπέδου της εικόνας, σύμφωνα με την εξίσωση 4.12.

Η εξίσωση 4.28 δίνει τη βέλτιστη λύση στο πρόβλημα της ελαχιστοποίησης της 4.25. Ο ορισμός της τελευταίας βασίζεται στην εξίσωση οπτικής ροής 4.10, η οποία εκφράζει τον περιορισμό

σταθερότητας της έντασης με βάση το ανάπτυγμα Taylor 4.7, από το οποίο λάβαμε υπόψη μόνο τους όρους πρώτης τάξης. Λόγω αυτής της προσέγγισης που υπάρχει εγγενώς στον περιορισμό οπτικής ροής 4.10, ο υπολογισμός της 4.28 είναι ακριβής μόνο για σχετικά μικρά διανύσματα μετατόπισης, όπου η επίδραση των όρων ανώτερης τάξης στην 4.7 είναι αμελητέα. Για το λόγο αυτό, η μέθοδος Lucas-Kanade εφαρμόζει μια επαναληπτική μέθοδο καθόδου κατά την κλίση (gradient descent), σε κάθε επανάληψη της οποίας υπολογίζεται μια διαφορική μετατόπιση που βελτιώνει την τρέχουσα εκτίμηση.

Θεωρώντας την I σαν μια διακριτού χρόνου ακολουθία εικόνων με μοναδιαία περίοδο χρονικής δειγματοληψίας, το ζητούμενο είναι ο υπολογισμός των μετατοπίσεων u και v που ελαχιστοποιούν τη συνάρτηση:

$$E(u, v) = \int_{-\infty}^{+\infty} w(x - x', y - y') (I(x' + u, y' + v, t + 1) - I(x', y', t))^2 dx' dy' \quad (4.29)$$

Η 4.29 αποτελεί μια εναλλακτική γραφή της 4.25 (αν αγνοήσουμε τους όρους ανώτερης τάξης). Στο πλαίσιο του επαναληπτικού σχήματος της μεθόδου Lucas-Kanade, έχουμε δεδομένες τις τρέχουσες/αρχικές εκτιμήσεις για τα διανύσματα μετατόπισης u^n και v^n (όπου ο εκθέτης είναι δείκτης της επανάληψης) και αναζητούμε τις βελτιώσεις δu , δv ώστε $u = u^n + \delta u$ και $v = v^n + \delta v$. Ο ορισμός των u^0 και v^0 (αρχικοποίηση) μπορεί να γίνει αυθαίρετα (για παράδειγμα, παντού 0). Το ανάπτυγμα Taylor πρώτης τάξης για την $I(x + u^n + \delta u, y + v^n + \delta v, t + 1)$ γύρω από το σημείο $(x + u^n, y + v^n, t + 1)$ γράφεται:

$$I(x + u, y + v, t + 1) \approx I(x + u^n, y + v^n, t + 1) + I_x \delta u + I_y \delta v, \quad (4.30)$$

όπου οι μερικές παράγωγοι I_x και I_y είναι υπολογισμένες στο σημείο $(x + u^n, y + v^n, t + 1)$. Αντικαθιστώντας την προσέγγιση 4.30 στην 4.29, η αντικειμενική συνάρτηση γράφεται:

$$E(\delta u, \delta v) = \int_{-\infty}^{+\infty} w(x - x', y - y') (J_x(x', y', t) \delta u + J_y(x', y', t) \delta v + J_t(x', y', t))^2 dx' dy', \quad (4.31)$$

όπου έχουμε ορίσει:

- $J_x(x, y, t) \equiv I_x(x + u^n, y + v^n, t + 1)$
- $J_y(x, y, t) \equiv I_y(x + u^n, y + v^n, t + 1)$
- $J_t(x, y, t) \equiv I(x + u^n, y + v^n, t + 1) - I(x, y, t)$

Η 4.31 είναι ακριβώς της ίδιας μορφής με την 4.25: το μόνο που αλλάζει είναι ότι χρησιμοποιούμε μια παραμορφωμένη (warped) εκδοχή της εικόνας σύμφωνα με το πεδίο (u^n, v^n) τη χρονική στιγμή $t + 1$. Συνεπώς, οι τιμές των δu και δv που ελαχιστοποιούν την 4.31 δίνονται από μια λύση της μορφής 4.28:

$$\begin{bmatrix} \delta u \\ \delta v \end{bmatrix} = \begin{bmatrix} \langle J_x^2 \rangle_w & \langle J_x J_y \rangle_w \\ \langle J_x J_y \rangle_w & \langle J_y^2 \rangle_w \end{bmatrix}^{-1} \begin{bmatrix} -\langle J_x J_t \rangle_w \\ -\langle J_y J_t \rangle_w \end{bmatrix}, \quad (4.32)$$

όπου οι συναρτήσεις J_x , J_y και J_t ορίζονται παραπάνω. Η λύση 4.32 χρησιμοποιείται για την ενημέρωση της τρέχουσας εκτίμησης για τις συνιστώσες της μετατόπισης: $u^{n+1} = u^n + \delta u$ και

Algorithm 3 Lucas-Kanade optical flow estimation

Input: Consecutive image sequence frames $I(x, y, t)$ and $I(x, y, t + 1)$, initial displacement estimates u^0 and v^0

Output: Displacements u, v

```

1:  $n = 0$ 
2: repeat
3:   /* Compute warped image derivatives */
4:    $J_x(x, y, t) \equiv I_x(x + u^n, y + v^n, t + 1)$ 
5:    $J_y(x, y, t) \equiv I_y(x + u^n, y + v^n, t + 1)$ 
6:    $J_t(x, y, t) \equiv I(x + u^n, y + v^n, t + 1) - I(x, y, t)$ 
7:   for all image pixels do
8:     /* Update local displacement estimates */
9:     
$$\begin{bmatrix} \delta u \\ \delta v \end{bmatrix} = \begin{bmatrix} \langle J_x^2 \rangle_w & \langle J_x J_y \rangle_w \\ \langle J_x J_y \rangle_w & \langle J_y^2 \rangle_w \end{bmatrix}^{-1} \begin{bmatrix} -\langle J_x J_t \rangle_w \\ -\langle J_y J_t \rangle_w \end{bmatrix}$$

10:     $u^{n+1} = u^n + \delta u$ 
11:     $v^{n+1} = v^n + \delta v$ 
12:   end for
13:    $n = n + 1$ 
14: until all values of  $\delta u$  and  $\delta v$  are sufficiently small

```

$v^{n+1} = v^n + \delta v$. Η μέθοδος τερματίζει όταν οι τιμές $\delta u, \delta v$ είναι αρκούντως μικρές. Μια περιγραφή της μεθόδου υπό τη μορφή ψευδοκώδικα δίνεται στον Αλγόριθμο 3.

Για καλύτερα αποτελέσματα (π.χ., δυνατότητα υπολογισμού κινήσεων μεγάλου εύρους), συνήθως χρησιμοποιείται κάποια πολυ-κλίμακωτή εκδοχή της μεθόδου Lucas-Kanade. Σε μια από τις απλούστερες μεθόδους αυτής της κατηγορίας, αρχικά οι εικόνες αναλύονται σε Γκαουσιανές πυραμίδες και το πεδίο της οπτικής ροής υπολογίζεται από τις πιο μικρές (αδρές) στις πιο μεγάλες (λεπτομερείς) κλίμακες, χρησιμοποιώντας τη λύση της μικρής κλίμακας (κατόπιν κατάλληλης κλιμάκωσης και υπερδειγματοληψίας/παρεμβολής) ως αρχική εκτίμηση για τη μεγάλη κλίμακα. Οι λύσεις από τις χαμηλής ανάλυσης εικόνες περιέχουν ουσιαστικά τις χαμηλής συχνότητας συνιστώσες της λύσης, οι οποίες σταδιακά βελτιώνονται αποκτώντας περισσότερη λεπτομέρεια καθώς η μέθοδος προχωράει στις πιο υψηλές αναλύσεις (coarse-to-fine refinement).

4.4 Εφαρμογή: κατάτμηση με βάση την κίνηση

Στην ενότητα αυτή παρουσιάζουμε μια ιδιαίτερα απλή μέθοδο κατάτμησης μιας εικόνας που ανήκει σε μια ακολουθία βίντεο σε περιοχές που χαρακτηρίζονται από ομοιόμορφη κίνηση. Ολόκληρα αντικείμενα (στην περίπτωση μεταφορικής κίνησης) ή περιοχές αντικειμένων (σε σύνθετη κίνηση πιθανώς παραμορφώσιμων αντικειμένων) που κινούνται χαρακτηρίζονται από πρακτικά την ίδια ταχύτητα. Το ίδιο ισχύει, σε ειδικές περιπτώσεις για ολόκληρα αντικείμενα και γενικά για αρκούντως μικρές περιοχές αντικειμένων, για τις καταγραφόμενες προβολές τους σε ένα βίντεο. Συνεπώς, η χρησιμότητα του προϊόντος μιας τέτοιας κατάτμησης είναι μεγάλη, καθώς οι περιοχές (περίπου) σταθερής διδιάστατης ταχύτητας που ορίζει αναμένεται να αντιστοιχούν σε κινούμενα αντικείμενα ή τμήματα κινούμενων αντικειμένων.

Έχοντας ως δεδομένα δύο διαδοχικά πλαίσια μιας ακολουθίας βίντεο, μπορούμε να υπολογίσουμε το (πυκνό) πεδίο οπτικής ροής που τις συνδέει. Μια κατάτμηση σε περιοχές ομοιόμορφης κίνησης μπορεί στη συνέχεια να προκύψει άμεσα με την εφαρμογή κάποιου αλγορίθμου ταξινόμη-

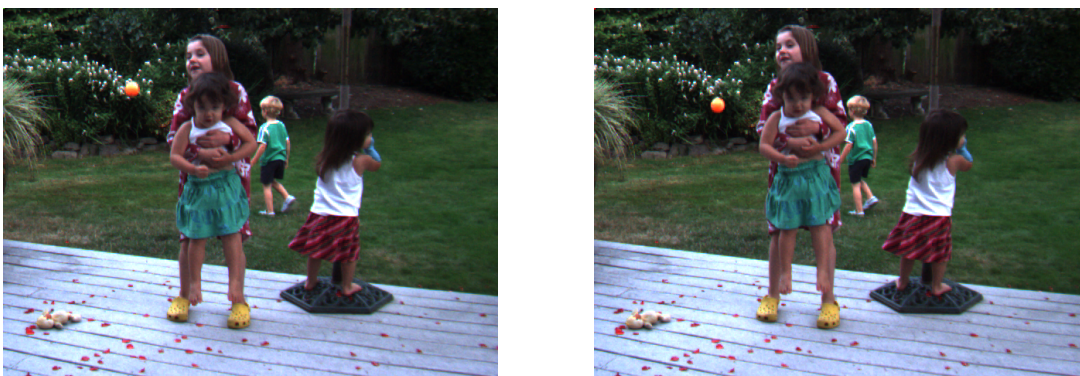
σης (clustering) στα διανύσματα κίνησης του πεδίου της οπτικής ροής. Ένας τέτοιος αλγόριθμος απαιτεί τον ορισμό κάποιου μετρικού απόστασης (ή μέτρου ομοιότητας) ανάμεσα σε δύο αντικείμενα προς ταξινόμηση – εν προκειμένω, διανύσματα κίνησης. Η απλούστερη, ίσως, προσέγγιση είναι χρήση της μεθόδου ταξινόμησης k -means, που, στη συνήθη μορφή της, υποθέτει τη θεωρία της Ευκλείδειας απόστασης δύο διανυσμάτων κίνησης $h = (u, v)$ και $h' = (u', v')$ ως μετρικού απόστασης:

$$d(h, h') = \sqrt{(u - u')^2 + (v - v')^2}$$

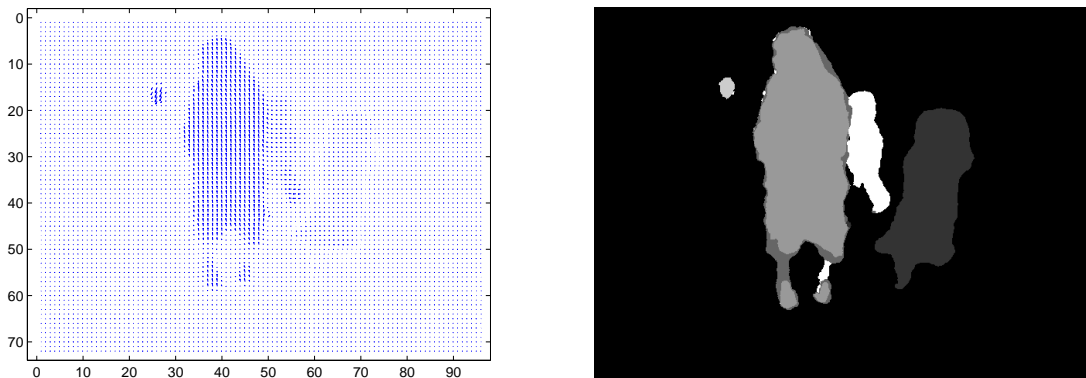
Μια περιγραφή της μεθόδου ομαδοποίησης k -means δίνεται στο Παράρτημα A. Η ζητούμενη κατάτμηση προκύπτει απλά από την ανάθεση σε κάθε εικονοστοιχείο της εικόνας του δείκτη της ομάδας (cluster label) που επιστρέφει ο αλγόριθμος ταξινόμησης για το διάνυσμα κίνησης στην ίδια θέση.

Η μέθοδος αυτή, παρότι εξαιρετικά απλή, απαιτεί κάποια πρότερη εκτίμηση για το πλήθος των περιοχών της κατάτμησης (απαιτούμενη παράμετρος για την ταξινόμηση k -means), είναι χρονικά τοπική, υπό την έννοια ότι ο υπολογισμός σε κάθε πλαίσιο γίνεται ανεξάρτητα με βάση μόνο το τρέχον και το επόμενο πλαίσιο, και βασίζεται σε ένα “αφελές” μετρικό απόστασης που δε λαμβάνει υπόψη, για παράδειγμα, τη χωρική εγγύτητα των σημείων στα οποία έχουν υπολογιστεί τα διανύσματα κίνησης. Η παρουσίασή της εδώ λειτουργεί συμπληρωματικά με την περιγραφή της μεθόδου οπτικής παρακολούθησης του Κεφαλαίου 6. Σημειώνουμε ότι η προσέγγιση αυτή μπορεί να θεωρηθεί ως μια (απλουστευμένη) ειδική περίπτωση της μεθόδου κατάτμησης σε κινούμενα αντικείμενα που περιγράφεται στο [38], όπου ορίζεται ένα μετρικό απόστασης ανάμεσα σε *τροχιές σημείων* (point trajectories), οι οποίες μπορεί να έχουν μήκος (διάρκεια) αρκετών πλαισίων, και η κατάτμηση προκύπτει σαν αποτέλεσμα της εφαρμογής μιας μεθόδου φασματικής ταξινόμησης (spectral clustering) [65] στις τροχιές αυτές.

Τέλος, παραθέτουμε δύο παραδείγματα εφαρμογής της μεθόδου που περιγράψαμε σε πραγματικά δεδομένα (Σχήματα 4.2 και 4.4). Οι εικόνες του Σχήματος 4.2 προέρχονται από το σύνολο εικόνων αναφοράς [66] για αλγόριθμους εκτίμησης οπτικής ροής, ενώ αυτές του Σχήματος 4.4 αποτελούν διαδοχικά πλαίσια μιας ακολουθίας βίντεο νοηματικής γλώσσας. Τα πεδία οπτικής ροής φαίνονται αντίστοιχα στα Σχήματα 4.3a και 4.5a, ενώ τα αποτελέσματα της κατάτμησης στα Σχήματα 4.3b και 4.5b. Για τον υπολογισμό της οπτικής ροής, βασιστήκαμε στην υλοποίηση του αλγορίθμου που παρέχεται από το [40].

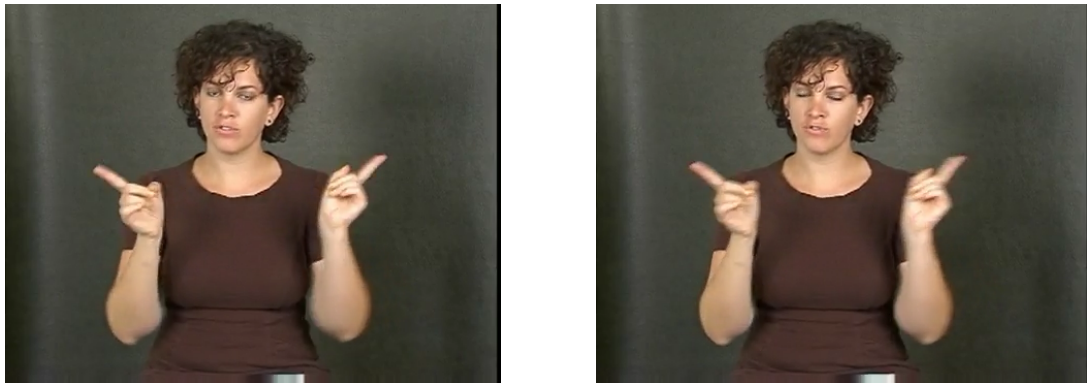


Σχήμα 4.2: Εικόνες πρώτης εφαρμογής.

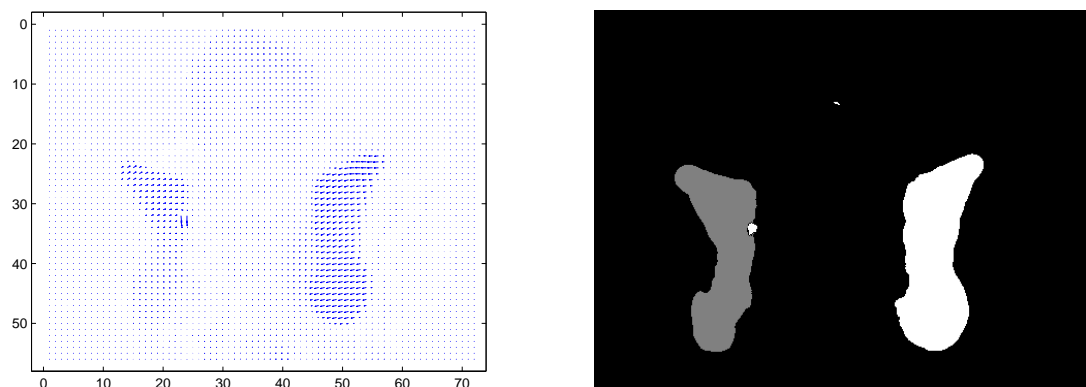


(a) Πεδίο οπτικής ροής (ζεύγος εικόνων 4.2) κα- (b) Αποτέλεσμα κατάτμησης σε $k = 6$ περιοχές με τόνιν υποδειγματοληψίας στο 15% του αρχικού μέγεθους. βάση την κίνηση (ζεύγος εικόνων 4.2).

Σχήμα 4.3



Σχήμα 4.4: Εικόνες δεύτερης εφαρμογής.



(a) Πεδίο οπτικής ροής (ζεύγος εικόνων 4.4) κα- (b) Αποτέλεσμα κατάτμησης σε $k = 3$ περιοχές με τόνιν υποδειγματοληψίας στο 15% του αρχικού μέγεθους. βάση την κίνηση (ζεύγος εικόνων 4.4).

Σχήμα 4.5

Κεφάλαιο 5

Bayesian filtering σε ακολουθιακά δεδομένα

Το πρόβλημα της οπτικής παρακολούθησης (visual tracking) συνίσταται στον εντοπισμό, κάθε χρονική στιγμή, του αντικειμένου ενδιαφέροντος σε μία ακολουθία εικόνων. Ο εντοπισμός μπορεί να γίνεται ντετερμινιστικά ή πιθανοτικά και μπορεί να εμπλέκει πληροφορία παρατηρήσεων από το παρελθόν, το παρόν ή/και το μέλλον του τρέχοντος πλαισίου (καρέ) της ακολουθίας. Στο Μπεϋζιανό θεωρητικό πλαίσιο, το ζητούμενο είναι ο υπολογισμός, κάθε χρονική στιγμή, της *πυκνότητας πιθανότητας* της θέσης (ή, γενικότερα, της κατάστασης, όπως θα ορίσουμε παρακάτω) του αντικειμένου (ή του συστήματος) με βάση την πληροφορία από τις διαθέσιμες παρατηρήσεις, θεωρώντας δεδομένα, σε πιθανοτική μορφή, το δυναμικό μοντέλο (dynamic/transition model) του συστήματος και το μοντέλο παρατήρησης (measurement/observation model). Η ζητούμενη κατανομή ενσωματώνει όλη τη διαθέσιμη στατιστική πληροφορία και από αυτή μπορεί στη συνέχεια να προκύψει ένας αριθμός εκτιμήσεων της θέσης/κατάστασης, ανάλογα με το κριτήριο που επιθυμούμε να βελτιστοποιήσουμε.

Στις ακόλουθες ενότητες, διατυπώνουμε περισσότερο φορμαλιστικά τον ορισμό του προβλήματος με τις σχετικές υποθέσεις/παραδοχές και στη συνέχεια παρουσιάζουμε τις δύο κυριότερες προσεγγίσεις στην εφαρμογή του Bayesian filtering στο πρόβλημα της οπτικής παρακολούθησης: το Kalman filter και την οικογένεια των Particle Filters.

5.1 Περιγραφή στο χώρο καταστάσεων

Ακολουθιακά δεδομένα (χρονοσειρές) εμφανίζονται σε έναν αριθμό εφαρμογών. Τα δεδομένα αυτά θεωρούμε ότι αποτελούν μετρήσεις (παρατηρήσεις) που λαμβάνονται καθώς κάποιο δυναμικό σύστημα εξελίσσεται. Συνήθη ζητούμενα είναι η πρόβλεψη μελλοντικών παρατηρήσεων και η εξαγωγή πληροφορίας για το ίδιο το σύστημα.

Η παρακάτω ανάλυση βασίζεται στην προσέγγιση μέσα από ένα χώρο (κρυφών) καταστάσεων. Θεωρούμε ότι, κάθε χρονική στιγμή, η κατάσταση του υπό εξέταση συστήματος περιγράφεται πλήρως από ένα σύνολο κρυφών (latent) μεταβλητών, οι οποίες συνιστούν το διάνυσμα κατάστασης (state vector). Το είδος της εφαρμογής καθορίζει το περιεχόμενο του τελευταίου: σε μια εφαρμογή οπτικής παρακολούθησης, οι μεταβλητές κατάστασης τυπικά είναι κινηματικά χαρακτηριστικά του στόχου. Στα προβλήματα συμπερασμού που αφορούν μοντέλα χώρου κατάστασης (state-space models), το ενδιαφέρον επικεντρώνεται ακριβώς στον προσδιορισμό του διανύσματος κατάστασης του συστήματος κάθε χρονική στιγμή.

Η προσέγγιση μέσω του χώρου καταστάσεων παρουσιάζει πλεονεκτήματα έναντι “κλασικών” μεθόδων ανάλυσης χρονοσειρών (μοντέλα τύπου ARMA, νευρωνικά δίκτυα, δένδρα αποφάσεων) [67, 68]. Συγκεκριμένα, δεν επιβάλλονται εξ αρχής από τη φύση του μοντέλου περιορισμοί που αφορούν τη Μαρκοβιανή τάξη της διαδικασίας που περιγράφει τη δυναμική εξέλιξη του συστήματος, το είδος (συνεχές/διακριτό, διάσταση) των “εισόδων” και τη δυνατότητα ενσωμάτωσης πρότερης γνώσης στο μοντέλο. Σε επίπεδο μοντελοποίησης, είναι συχνά επιθυμητή η εκτίμηση μεταβλητών που δεν είναι άμεσα παρατηρήσιμες, των οποίων η ενσωμάτωση στο διάνυσμα κατάστασης ως κρυφών μεταβλητών (latent variables) οδηγεί στον ορισμό ενός μοντέλου που βρίσκεται, ως προς τους μηχανισμούς αιτιατότητας που το διέπουν, πιο κοντά στην πραγματική υπό μοντελοποίηση οντότητα. Ακόμη, όμως, και αν μας ενδιαφέρουν μόνο παρατηρήσιμες μεταβλητές, η εισαγωγή “πλασματικών” κρυφών μεταβλητών, ενδεχομένως χωρίς καν φυσική ερμηνεία, συχνά οδηγεί στον ορισμό πολύ απλούστερων μοντέλων (χαμηλότερης διάστασης διάνυσμα κατάστασης από ό,τι οι παρατηρήσεις).

5.2 Συμβολισμοί και παραδοχές

Η κατάσταση του συστήματος είναι, εν γένει, διαφορετικής διάστασης από τις λαμβανόμενες από αυτό μετρήσεις. Στα επόμενα, θα συμβολίζουμε με $x_k \in \mathbb{R}^{n_x}$ την κατάσταση του συστήματος τη χρονική στιγμή k και με $y_k \in \mathbb{R}^{n_y}$ την παρατήρηση που λαμβάνουμε τη στιγμή αυτή. Θα περιοριστούμε σε συνεχείς χώρους καταστάσεων και σε συστήματα διακριτού χρόνου.

Με $x_{0:k}$ και $y_{1:k}$ συμβολίζουμε αντίστοιχα το ιστορικό των καταστάσεων $\{x_i, i = 0, \dots, k\}$ και το σύνολο των παρατηρήσεων $\{y_i, i = 1, \dots, k\}$ μέχρι τη στιγμή k . Στο Μπεϋζιανό θεωρητικό πλαίσιο, το ζητούμενο είναι ο υπολογισμός της πιθανότητας $p(x_k | y_{1:k})$. Στην περίπτωση του filtering, όπου εστιάζει η παρούσα εργασία, στόχος μας είναι να συμπεράνουμε την κατάσταση x_k υπολογίζοντας εκτιμήσεις αυτής με βάση την πυκνότητα $p(x_k | y_{1:k})$.

Στα επόμενα, θα θεωρείται δεδομένο ότι ισχύουν οι ακόλουθες υποθέσεις στατιστικής ανεξαρτησίας:

- Θεωρούμε ότι το μοντέλο του συστήματος περιγράφεται από μια Μαρκοβιανή διαδικασία πρώτης τάξης, δηλαδή ότι:

$$p(x_k | x_{0:k-1}) = p(x_k | x_{k-1})$$

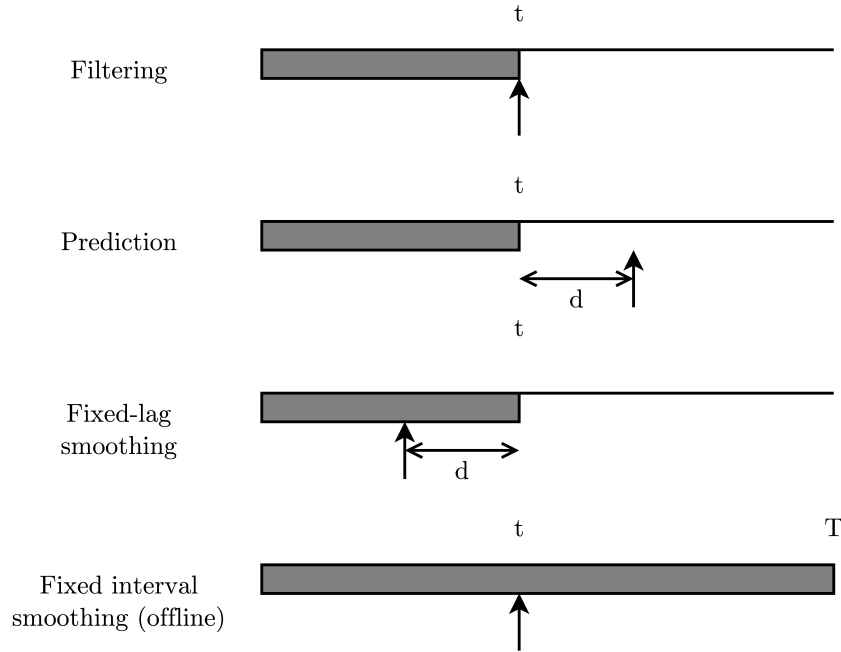
Αν το παραπάνω δεν ισχύει, μπορούμε να το επιβάλλουμε επαυξάνοντας κατάλληλα το διάνυσμα κατάστασης. Για παράδειγμα, αν η x_k είναι Μαρκοβιανή διαδικασία p -οστής τάξης, το επαυξημένο διάνυσμα $\tilde{x}_k = (x_k, \dots, x_{k-p+1})$ περιγράφει διαδικασία πρώτης τάξης.

- Υποθέτουμε, επίσης, ότι οι τρέχουσες παρατηρήσεις εξαρτώνται μόνο από την τρέχουσα κατάσταση. Η y_k είναι υπό συνθήκη ανεξάρτητη από όλες τις άλλες μετρήσεις, δεδομένης της x_k , δηλαδή:

$$p(y_k, y_1, \dots, y_{l_2} | x_k) = p(y_k | x_k) p(y_1, \dots, y_{l_2} | x_k)$$

5.3 Είδη συμπερασμού σε ακολουθιακά δεδομένα

Θεωρώντας ότι σε κάθε πρόβλημα online ανάλυσης το ζητούμενο είναι ο υπολογισμός της κατανομής $p(x_l | y_{1:k})$, διακρίνουμε τρεις βασικές κατηγορίες προβλημάτων συμπερασμού ανάλογα με τη σχέση που συνδέει τα l, k :



Σχήμα 5.1: Είδη online συμπερασμού σε ακολουθιακά δεδομένα. Η σκιασμένη περιοχή αντιστοιχεί στο χρονικό διάστημα για το οποίο υπάρχουν διαθέσιμα δεδομένα, τα βέλη δείχνουν τη χρονική στιγμή στην οποία επιθυμούμε να εκτιμήσουμε την κατάσταση του συστήματος και t είναι η τρέχουσα χρονική στιγμή.

- **Filtering** ($k = l$). Η εκτίμηση της κατάστασης τη χρονική στιγμή k με δεδομένες όλες τις μέχρι τότε παρατηρήσεις αποτελεί το πιο συνηθισμένο πρόβλημα συμπερασμού. Όπως θα δούμε αναλυτικότερα σε επόμενη ενότητα, ο υπολογισμός της $p(x_k|y_{1:k})$ γίνεται αναδρομικά, με χρήση του κανόνα του Bayes, και περιλαμβάνει ένα στάδιο πρόβλεψης (υπολογισμός της $p(x_k|y_{1:k-1})$) και ένα στάδιο ενημέρωσης (predict-update). Η υπολογισθείσα $p(x_k|y_{1:k})$ ενσωματώνει όλη τη στατιστική πληροφορία μέχρι τη χρονική στιγμή k και καθιστά την αποθήκευση των προηγούμενων εκτιμήσεων περιττή. Έτσι, η διαδικασία έχει σταθερό κόστος σε χώρο και χρόνο ανά χρονικό βήμα.
- **Smoothing** ($l < k$). Σε κάποιες περιπτώσεις απαιτείται η εκτίμηση μιας κατάστασης του παρελθόντος, με δεδομένες τις παρατηρήσεις μέχρι την τρέχουσα χρονική στιγμή, δηλαδή ο υπολογισμός της $p(x_{k-d}|y_{1:k})$, για κάποιο $d > 0$. Τυπική εφαρμογή που εντάσσεται στην online περίπτωση είναι το fixed-lag smoothing, όπου η καθυστέρηση d είναι σταθερή. Στις offline εφαρμογές, το ζητούμενο είναι συνήθως η $p(x_k|y_{1:T})$, για κάθε $1 \leq k \leq T$ (fixed-interval smoothing).
- **Prediction** ($l > k$). Εκτός από την εκτίμηση της τρέχουσας ή μιας περασμένης κατάστασης, ενίοτε ζητείται η εκτίμηση μιας μελλοντικής κατάστασης, δηλαδή ο υπολογισμός της $p(x_{k+d}|y_{1:k})$, για κάποιο $d > 0$. Η παραπάνω κατανομή μπορεί επίσης να δώσει μια πρόβλεψη για την παρατήρηση y_{k+d} , κάνοντας περιθωριοποίηση στην x_{k+d} :

$$p(y_{k+d}|y_{1:k}) = \int_{\mathbb{R}^{n_x}} p(y_{k+d}|x_{k+d}) p(x_{k+d}|y_{1:k}) dx_{k+d}$$

5.4 Bayesian filtering

Για τον ορισμό του προβλήματος του Bayesian filtering, απαιτείται η γνώση ενός μοντέλου που περιγράφει τη χρονική εξέλιξη της κατάστασης του συστήματος και ενός μοντέλου που περιγράφει τη σχέση των παρατηρήσεων με την πραγματική, κρυφή κατάσταση του συστήματος. Το πρώτο αποτελεί το *δυναμικό μοντέλο* (ή *μοντέλο μετάβασης*) και το δεύτερο το *μοντέλο παρατήρησης* της περιγραφής. Σε συστημικούς όρους και σε συμφωνία με τις υποθέσεις ανεξαρτησίας που περιγράψαμε πιο πάνω, υποθέτουμε ότι τα δυο αυτά μοντέλα περιγράφονται από τις παρακάτω γενικές εξισώσεις:

$$x_k = f_k(x_{k-1}, v_{k-1}) \quad (5.1)$$

$$y_k = h_k(x_k, w_k) \quad (5.2)$$

όπου:

- $k \in \mathbb{N}$ είναι ο δείκτης της (τρέχουσας) χρονικής στιγμής
- $\{v_{k-1}, k \in \mathbb{N}^*\}$ είναι μια ακολουθία τιμών, διάστασης n_v , ανεξάρτητων και πανομοιότυπα κατανομημένων (independent, identically distributed: i.i.d.), που μοντελοποιεί την αβεβαιότητα της διαδικασίας της κατάστασης (αποτελεί την πηγή του πιθανοτικού χαρακτήρα του δυναμικού μοντέλου)
- η $f_k: \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \mapsto \mathbb{R}^{n_x}$ περιγράφει τη χρονική εξέλιξη του συστήματος συναρτήσει της προηγούμενης κατάστασης (βλ. παραδοχές) και της αβεβαιότητας v_{k-1}
- $\{w_k, k \in \mathbb{N}\}$ είναι μια ακολουθία τιμών, διάστασης n_w , ανεξάρτητων και πανομοιότυπα κατανομημένων (independent, identically distributed: i.i.d.), που περιγράφει την αβεβαιότητα του μοντέλου παρατήρησης (μοντελοποιεί το θόρυβο των μετρήσεων/παρατηρήσεων)
- η $h_k: \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \mapsto \mathbb{R}^{n_y}$ ορίζει την πιθανοτική σχέση ανάμεσα στην κατάσταση του συστήματος τη χρονική στιγμή k και την τρέχουσα παρατήρηση y_k

Με βάση τα προηγούμενα, στο Μπεϋζιανό θεωρητικό πλαίσιο, το πρόβλημα του filtering διατυπώνεται ως ο υπολογισμός της πυκνότητας $p(x_k|y_{1:k})$, από την οποία μπορεί να εξαχθεί ένας αριθμός εκτιμήσεων της κατάστασης x_k . Θεωρώντας ότι η αρχική (πρότερη, prior) πυκνότητα $p(x_0|y_0) \equiv p(x_0)$ είναι γνωστή (το y_0 είναι ένα κενό σύνολο μετρήσεων), ο υπολογισμός της $p(x_k|y_{1:k})$ μπορεί να γίνει αναδρομικά με βάση τον κανόνα του Bayes, με δεδομένη την πυκνότητα $p(x_{k-1}|y_{1:k-1})$ της προηγούμενης χρονικής στιγμής $k-1$. Ο υπολογισμός περιλαμβάνει ένα στάδιο *πρόβλεψης* (prediction), όπου υπολογίζεται η $p(x_k|y_{1:k-1})$ χωρίς να ληφθεί υπόψη η παρατήρηση y_k , και ένα στάδιο *ενημέρωσης* (update), όπου η πρόβλεψη “διορθώνεται” με βάση τη μέτρηση y_k :

• Πρόβλεψη

Η *πρότερη* πυκνότητα $p(x_k|y_{1:k-1})$ υπολογίζεται ως η περιθώρια της $p(x_k, x_{k-1}|y_{1:k-1})$ ως προς τη μεταβλητή x_{k-1} :

$$\begin{aligned} p(x_k|y_{1:k-1}) &= \int_{\mathbb{R}^{n_x}} p(x_k, x_{k-1}|y_{1:k-1}) dx_{k-1} \\ &= \int_{\mathbb{R}^{n_x}} p(x_k|x_{k-1}, y_{1:k-1})p(x_{k-1}|y_{1:k-1}) dx_{k-1} \\ &= \int_{\mathbb{R}^{n_x}} p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1}) dx_{k-1} \end{aligned} \quad (5.3)$$

Παραπάνω, έγινε χρήση του γεγονότος ότι $p(x_k|x_{k-1}, y_{1:k-1}) = p(x_k|x_{k-1})$, καθώς η x είναι μαρκοβιανή διαδικασία πρώτης τάξης, σε συμφωνία με τις παραδοχές μας και την 5.1. Σημειώνουμε ότι η $p(x_k|x_{k-1})$ ορίζεται από την f_k και τη γνωστή κατανομή της v_{k-1} στην 5.1 και αποτελεί μια ισοδύναμη περιγραφή του δυναμικού μοντέλου του συστήματος. Σύμφωνα με τις υποθέσεις ανεξαρτησίας μας, $p(v_{k-1}|x_{k-1}) = p(v_{k-1})$ και:

$$\begin{aligned} p(x_k|x_{k-1}) &= \int_{\mathbb{R}^{n_v}} p(x_k|x_{k-1}, v_{k-1})p(v_{k-1}|x_{k-1}) dv_{k-1} \\ &= \int_{\mathbb{R}^{n_v}} \delta(x_k - f_k(x_{k-1}, v_{k-1})) p(v_{k-1}) dv_{k-1} \end{aligned}$$

Η πρότερη πυκνότητα $p(x_k|y_{1:k-1})$ ενσωματώνει όλη τη διαθέσιμη στατιστική πληροφορία για την x_k με δεδομένες τις παρατηρήσεις $y_{1:k-1}$ (ισοδύναμα: την $p(x_{k-1}|y_{1:k-1})$) και το δυναμικό μοντέλο του συστήματος.

• Ενημέρωση

Με γνωστή την παρατήρηση y_k , ο υπολογισμός της ύστερης πυκνότητας $p(x_k|y_{1:k})$ γίνεται με βάση τον κανόνα του Bayes:

$$\begin{aligned} p(x_k|y_{1:k}) &= \frac{p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \\ &= \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}, \end{aligned} \quad (5.4)$$

όπου η σταθερά κανονικοποίησης στον παρονομαστή ισούται με:

$$p(y_k|y_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(y_k|x_k)p(x_k|y_{1:k-1}) dx_k \quad (5.5)$$

Στις παραπάνω, έγινε χρήση της $p(y_k|x_k, y_{1:k-1}) = p(y_k|x_k)$ (δεύτερη υπόθεση στατιστικής ανεξαρτησίας: η y_k εξαρτάται μόνο από την x_k). Σημειώνουμε ότι η $p(y_k|x_k)$ ορίζεται από την h_k και τη γνωστή κατανομή της w_k στην 5.2 και αποτελεί ισοδύναμη περιγραφή του μοντέλου παρατήρησης του συστήματος. Σύμφωνα με τις υποθέσεις ανεξαρτησίας μας, $p(w_k|x_k) = p(w_k)$ και:

$$\begin{aligned} p(y_k|x_k) &= \int_{\mathbb{R}^{n_w}} p(y_k|x_k, w_k)p(w_k|x_k) dw_k \\ &= \int_{\mathbb{R}^{n_w}} \delta(y_k - h_k(x_k, w_k)) p(w_k) dw_k \end{aligned}$$

Η ύστερη πυκνότητα $p(x_k|y_{1:k})$ ενσωματώνει όλη τη διαθέσιμη πληροφορία για την x_k με δεδομένες τις μετρήσεις $y_{1:k}$ και προκύπτει σαν “διόρθωση” της πρότερης $p(x_k|y_{1:k-1})$ με βάση τη νέα μέτρηση y_k και το μοντέλο παρατήρησης του συστήματος.

Οι εξισώσεις 5.3 και 5.4 ορίζουν τη (βέλτιστη) λύση του προβλήματος του Bayesian filtering. Στη γενική περίπτωση (για αυθαίρετες συναρτήσεις f_k και h_k και κατανομές των v_{k-1} και w_k – ισοδύναμα: για αυθαίρετες $p(x_k|x_{k-1})$ και $p(y_k|x_k)$), δεν υπάρχουν αποδοτικές μεθόδους αναλυτικού (ακριβούς) προσδιορισμού της $p(x_k|y_{1:k})$.

5.5 Το φίλτρο Kalman

Το φίλτρο Kalman [69, 70, 71] είναι ένας αναδρομικός αλγόριθμος που δίνει ακριβή λύση στο Bayesian filtering στην ειδική περίπτωση όπου η ύστερη κατανομή σε κάθε βήμα είναι γκαουσιανή. Συγκεκριμένα, στην πιο συνηθισμένη περίπτωση (γραμμικό δυναμικό σύστημα και γκαουσιανοί θόρυβοι: linear-Gaussian assumption), υποθέτουμε ότι οι 5.1 και 5.2 έχουν τις ακόλουθες ειδικές μορφές:

$$x_k = A_k x_{k-1} + v_{k-1} \quad (5.6)$$

$$y_k = H_k x_k + w_k \quad (5.7)$$

όπου τα A_k και H_k είναι πίνακες διαστάσεων $n_x \times n_x$ και $n_y \times n_x$ αντίστοιχα και τα v_{k-1} , w_k ακολουθούν γκαουσιανές κατανομές με μηδενική μέση τιμή και πίνακες συμμεταβλητότητας Q_{k-1} και R_k αντίστοιχα: $v_{k-1} \sim \mathcal{N}(v_{k-1}; 0, Q_{k-1})$ και $w_k \sim \mathcal{N}(w_k; 0, R_k)$, όπου με $\mathcal{N}(x; \mu, \Sigma)$ συμβολίζουμε μια κανονική πυκνότητα με όρισμα x , μέση τιμή μ και πίνακα συμμεταβλητότητας Σ . Όπως προκύπτει από τη μορφή των εξισώσεων, σε αυτό το μοντέλο οι θόρυβοι v_{k-1} και w_k είναι ίδιας διάστασης με τα διανύσματα x_k και y_k αντίστοιχα.

Είναι σαφές από τη μορφή των 5.6 και 5.7 ότι, αν η πρότερη $p(x_0)$ και οι θόρυβοι v_{k-1} και w_k ακολουθούν γκαουσιανές κατανομές, το ίδιο θα ισχύει τόσο για την πρότερη $p(x_k|y_{1:k-1})$ όσο και για το μοντέλο παρατήρησης $p(y_k|x_k)$ (επαγωγικά) σε κάθε βήμα, αφού κάθε γραμμικός συνδυασμός τυχαίων μεταβλητών που ακολουθούν κανονικές κατανομές ακολουθεί επίσης κανονική κατανομή. Από την 5.4, το ίδιο συμπεραίνουμε και για την ύστερη πυκνότητα $p(x_k|y_{1:k})$ (“συμπλήρωση τετραγώνου”). Συνεπώς, κάθε μια από τις $p(x_k|y_{1:k-1})$ και $p(x_k|y_{1:k})$ είναι γκαουσιανή και παραμετροποιείται από μια μέση τιμή και ένα πίνακα συμμεταβλητότητας.

Έστω $m_{k|k-1}$ και $P_{k|k-1}$ οι παράμετροι της $p(x_k|y_{1:k-1})$ και $m_{k|k}$, $P_{k|k}$ αντίστοιχα της $p(x_k|y_{1:k})$ στο k -οστό βήμα. Οι τιμές $m_{k-1|k-1}$ και $P_{k-1|k-1}$ είναι γνωστές από το προηγούμενο βήμα του αναδρομικού υπολογισμού (η ύστερη $p(x_{k-1}|y_{1:k-1}) = \mathcal{N}(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1})$ του προηγούμενου βήματος είναι γνωστή). Με βάση τα παραπάνω, τα στάδια πρόβλεψης και ενημέρωσης συνοψίζονται στις εξισώσεις του αλγορίθμου 4.

Algorithm 4 Kalman Filter

- Πρόβλεψη

$$p(x_k|y_{1:k-1}) = \mathcal{N}(x_k; m_{k|k-1}, P_{k|k-1}) \quad (5.8)$$

$$m_{k|k-1} = A_k m_{k-1|k-1} \quad (5.9)$$

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^T + Q_{k-1} \quad (5.10)$$

- Ενημέρωση

$$p(x_k|y_{1:k}) = \mathcal{N}(x_k; m_{k|k}, P_{k|k}) \quad (5.11)$$

$$m_{k|k} = m_{k|k-1} + K_k (y_k - H_k m_{k|k-1}) \quad (5.12)$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (5.13)$$

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \quad (5.14)$$

Καθώς η $p(x_k|y_{1:k})$ είναι γκαουσιανή, η εκτίμηση μέγιστης ύστερης πιθανότητας (MAP estimation) για την x_k μετά το στάδιο της πρόβλεψης είναι αυτή που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα και είναι ίση με τη μέση τιμή της κατανομής: $\hat{x}_{k|k} = m_{k|k}$. Θα αναφερόμαστε στην τιμή αυτή ως (τελική) εκτίμηση και στην $\hat{x}_{k|k-1} = m_{k|k-1}$ ως πρόβλεψη της κατάστασης στο k -οστό βήμα.

Όπως υποδεικνύει η 5.12, η τελική εκτίμηση $\hat{x}_{k|k}$ προκύπτει σαν το άθροισμα της πρόβλεψης $\hat{x}_{k|k-1}$ και μιας γραμμικής συνάρτησης της διαφοράς της k -οστής παρατήρησης y_k και της “παρατήρησης” που αντιστοιχεί στην τιμή της πρόβλεψης με βάση το μοντέλο παρατήρησης. Η (διαστάσεων $n_x \times n_y$) σταθερά της γραμμικής συνάρτησης, K_k , ονομάζεται κέρδος Kalman και καθορίζει το συντελεστή βάρους ανάμεσα σε αυτές τις δύο ποσότητες. Ποιοτικά, για μικρή R_k (μικρή αβεβαιότητα για την παρατήρηση y_k), δίνεται ισχυρό βάρος στον όρο της διαφοράς (μεγάλη “εμπιστοσύνη” στην παρατήρηση), ενώ, αντίθετα, όταν η $P_{k|k-1}$ είναι μικρή (μικρή αβεβαιότητα στην πρόβλεψη/προβλεπόμενη “παρατήρηση”), ο όρος αυτός καταπιέζεται. Η τιμή του κέρδους K_k που δίνεται στην 5.14 ελαχιστοποιεί την αβεβαιότητα της τελικής εκτίμησης $P_{k|k}$.

Το φίλτρο Kalman είναι ένας αποδοτικός αλγόριθμος και δίνει ακριβή λύση στο πρόβλημα του Bayesian filtering υπό τις υποθέσεις του γραμμικού, γκαουσιανού μοντέλου που αναφέραμε.

5.6 Το επεκτεταμένο φίλτρο Kalman (EKF)

Η παραδοχή ενός γραμμικού μοντέλου μετάβασης με προσθετικό γκαουσιανό θόρυβο είναι ιδιαίτερα περιοριστική σε ένα μεγάλο εύρος εφαρμογών. Το γεγονός αυτό οδήγησε στην ανάπτυξη μιας “επεκτεταμένης” εκδοχής του αλγορίθμου (Extended Kalman Filter, EKF) που μπορεί να εφαρμοστεί σε μια ευρύτερη κλάση δυναμικών συστημάτων όπου οι συναρτήσεις f , h του μοντέλου (5.1, 5.2) είναι διαφορίσιμες (όχι απαραίτητα γραμμικές).

Το Extended Kalman Filter είναι ένας προσεγγιστικός αλγόριθμος και ο υπολογισμός σε κάθε βήμα γίνεται ουσιαστικά όπως στο κλασικό Kalman φίλτρο, αφού γίνει γραμμικοποίηση του συστήματος γύρω από τις τρέχουσες εκτιμήσεις. Η ύστερη κατανομή προσεγγίζεται τότε από μια γκαουσιανή με παραμέτρους που προκύπτουν από τις τρέχουσες εκτιμήσεις και το (τρέχον) γραμμικοποιημένο σύστημα. Με δεδομένη την προσέγγιση της ύστερης του προηγούμενου βήματος, $p(x_{k-1}|y_{1:k-1}) \approx \mathcal{N}(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1})$, η προσέγγιση της ύστερης στο k -οστό βήμα γίνεται με βάση τους υπολογισμούς του αλγορίθμου 5.

Ο αλγόριθμος είναι ίδιος με αυτόν του κλασικού φίλτρου Kalman, με τη διαφορά ότι σε κάθε χρονική στιγμή το σύστημα προσεγγίζεται από τους πρώτους (γραμμικούς) όρους των αναπτυγμάτων Taylor των συναρτήσεων f και h του μοντέλου γύρω από τις τρέχουσες εκτιμήσεις. Έχουν αναπτυχθεί εκδοχές του EKF που διατηρούν όρους ανώτερης τάξης, όμως η προστιθέμενη πολυπλοκότητα τις καθιστά μη πρακτικές [68].

5.7 Particle Filtering

Στις βασισμένες στο Kalman φίλτρο μεθόδους Bayesian filtering που είδαμε μέχρι τώρα, η ύστερη πυκνότητα $p(x_k|y_{1:k})$ είναι ή προσεγγίζεται από μια γκαουσιανή κατανομή. Το γεγονός αυτό περιορίζει τη μορφή των συναρτήσεων του μοντέλου και κατά συνέπεια την κλάση των δυναμικών συστημάτων στα οποία μια τέτοια προσέγγιση μπορεί να εφαρμοστεί με επιτυχία. Μια άλλη, εγγενής αδυναμία των μεθόδων αυτών είναι η απαίτηση για γνώση της αναλυτικής μορφής των μοντέλων μετάβασης και παρατήρησης. Στην ενότητα αυτή, παρουσιάζουμε μια οικογένεια γενικών μεθόδων οι οποίες δίνουν μια προσεγγιστική λύση στο πρόβλημα του Bayesian filtering, χωρίς την αναγκαιότητα κάποιας ιδιαίτερα περιοριστικής παραδοχής για τη μορφή του πιθανοτικού

Algorithm 5 Extended Kalman Filter

- Γραμμικοποίηση δυναμικού μοντέλου

$$\hat{A}_k \equiv \left. \frac{df_k(x)}{dx} \right|_{x=m_{k-1|k-1}} \quad (5.15)$$

- Πρόβλεψη

$$p(x_k|y_{1:k-1}) \approx \mathcal{N}(x_k; m_{k|k-1}, P_{k|k-1}) \quad (5.16)$$

$$m_{k|k-1} = f_k(m_{k-1|k-1}) \quad (5.17)$$

$$P_{k|k-1} = \hat{A}_k P_{k-1|k-1} \hat{A}_k^T + Q_{k-1} \quad (5.18)$$

- Γραμμικοποίηση μοντέλου παρατήρησης

$$\hat{H}_k \equiv \left. \frac{dh_k(x)}{dx} \right|_{x=m_{k|k-1}} \quad (5.19)$$

- Ενημέρωση

$$p(x_k|y_{1:k}) \approx \mathcal{N}(x_k; m_{k|k}, P_{k|k}) \quad (5.20)$$

$$m_{k|k} = m_{k|k-1} + K_k (y_k - h_k(m_{k|k-1})) \quad (5.21)$$

$$P_{k|k} = (I - K_k \hat{H}_k) P_{k|k-1} \quad (5.22)$$

$$K_k = P_{k|k-1} \hat{H}_k^T (\hat{H}_k P_{k|k-1} \hat{H}_k^T + R_k)^{-1} \quad (5.23)$$

μοντέλου.

Τα Particle Filters (“φίλτρα σωματίων”) (στιγμιότυπα ή εναλλακτικές ονομασίες των οποίων είναι το bootstrap filter και οι αλγόριθμοι CONDENSATION, interacting particle approximations και survival of the fittest) είναι μια κατηγορία ακολουθιακών Monte Carlo μεθόδων (Sequential Monte Carlo – SMC). Η βασική ιδέα είναι η αναπαράσταση/προσέγγιση της ζητούμενης ύστερης κατανομής με ένα πεπερασμένο σύνολο τυχαίων δειγμάτων, σε κάθε ένα από τα οποία αντιστοιχεί ένα βαθμωτό βάρος.

5.7.1 Δειγματοληψία σημαντικότητας (importance sampling)

Έστω $\{(x_{0:k}^i, w_k^i) : i = 1, \dots, N_s\}$ ένα τυχαίο μέτρο που περιγράφει την ύστερη κατανομή $p(x_{0:k}|y_{1:k})$. Σημειώνουμε ότι η $p(x_k|y_{1:k})$ την οποία αναζητούμε είναι περιθώρια κατανομή της “πλήρους” $p(x_{0:k}|y_{1:k})$. Για κάθε $i = 1, \dots, N_s$, η $x_{0:k}^i$ είναι μια ακολουθία καταστάσεων μέχρι και τη στιγμή k , στην οποία έχει αποδοθεί βάρος w_k^i και για κάθε k , είναι $\sum_{i=1}^{N_s} w_k^i = 1$. Με βάση

την περιγραφή αυτή, η $p(x_{0:k}|y_{1:k})$ προσεγγίζεται από το πεπερασμένο άθροισμα:

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_{0:k} - x_{0:k}^i), \quad (5.24)$$

όπου δ είναι το συναρτησιακό Dirac. Εάν οι $x_{0:k}^i$, $i = 1, \dots, N_s$, είναι τυχαία δείγματα από την $p(x_{0:k}|y_{1:k})$, τα βάρη λαμβάνονται όλα ίσα μεταξύ τους και ίσα με $w_k^i = \frac{1}{N_s}$. Γενικά, είναι συχνά αδύνατη ή ασύμφορη η δειγματοληψία απευθείας από την $p(x_{0:k}|y_{1:k})$. Στη γενική περίπτωση, τα βάρη επιλέγονται με βάση την αρχή/μέθοδο της *δειγματοληψίας σημαντικότητας* (importance sampling). Συγκεκριμένα, υποθέτουμε ότι τα δείγματα $x_{0:k}^i$, $i = 1, \dots, N_s$, ακολουθούν (έχουν αντληθεί από) τη γνωστή κατανομή $q(x_{0:k}|y_{1:k})$, την οποία ονομάζουμε *κατανομή σημαντικότητας* (importance/proposal/instrumental distribution), ενώ για την πραγματική ύστερη κατανομή μπορούμε μόνο να αποτιμήσουμε μια “ανάλογη” πυκνότητα $\pi(x_{0:k}|y_{1:k}) \propto p(x_{0:k}|y_{1:k})$ (γνωρίζουμε την τιμή $p(x_{0:k}|y_{1:k})$ μέχρι έναν σταθερό παράγοντα – up to proportionality). Προκειμένου η 5.24 να αποτελεί μια πιστή προσέγγιση, τα δείγματα σταθμίζονται σύμφωνα με τα βάρη:

$$w_k^i \propto \frac{p(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})} \quad (5.25)$$

$$\sum_{i=1}^{N_s} w_k^i = 1 \quad (5.26)$$

5.7.2 SIS Particle Filter

Επιστρέφοντας στο σενάριο της (ακολουθιακής) εκτίμησης της ύστερης $p(x_{0:k}|y_{1:k})$, υποθέτουμε ότι έχουμε μια προσέγγιση της $p(x_{0:k-1}|y_{1:k-1})$ από ένα σύνολο σταθμισμένων όπως παραπάνω δειγμάτων $\{(x_{0:k-1}^i, w_{k-1}^i) : i = 1, \dots, N_s\}$ από την οποία θα παράγουμε ένα νέο σύνολο σταθμισμένων δειγμάτων (προσέγγιση) της ζητούμενης ύστερης.

Υποθέτουμε ότι η κατανομή σημαντικότητας επιλέγεται έτσι ώστε η πυκνότητά της να παραγοντοποιείται ως:

$$q(x_{0:k}|y_{1:k}) = q(x_k|x_{0:k-1}, y_{1:k})q(x_{0:k-1}|y_{1:k-1}) \quad (5.27)$$

Στην περίπτωση αυτή, μπορούμε να παράγουμε δείγματα $x_{0:k}^i \sim q(x_{0:k}|y_{1:k})$ απλώς επαυξάνοντας κάθε ένα από τα υπάρχοντα δείγματα $x_{0:k-1}^i \sim q(x_{0:k-1}|y_{1:k-1})$ με τη νέα κατάσταση $x_k^i \sim q(x_k|x_{0:k-1}, y_{1:k})$. Για να καταλήξουμε στην εξίσωση ενημέρωσης των βαρών, εκφράζουμε την $p(x_{0:k}|y_{1:k})$ σαν συνάρτηση των $p(x_{0:k-1}|y_{1:k-1})$ (ύστερη την προηγούμενη χρονική στιγμή), $p(y_k|x_k)$ (μοντέλο παρατήρησης) και $p(x_k|x_{k-1})$ (μοντέλο μετάβασης):

$$\begin{aligned} p(x_{0:k}|y_{1:k}) &= \frac{p(x_{0:k}, y_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \\ &= \frac{p(y_k|x_{0:k}, y_{1:k-1})p(x_{0:k}|y_{1:k-1})}{p(y_k|y_{1:k-1})} \\ &= \frac{p(y_k|x_{0:k}, y_{1:k-1})p(x_{0:k-1}, x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \\ &= \frac{p(y_k|x_{0:k}, y_{1:k-1})p(x_k|x_{0:k-1}, y_{1:k-1})}{p(y_k|y_{1:k-1})} p(x_{0:k-1}|y_{1:k-1}) \\ &= \frac{p(y_k|x_k)p(x_k|x_{k-1})}{p(y_k|y_{1:k-1})} p(x_{0:k-1}|y_{1:k-1}) \\ &\propto p(y_k|x_k)p(x_k|x_{k-1})p(x_{0:k-1}|y_{1:k-1}), \end{aligned} \quad (5.28)$$

όπου εφαρμόσαμε τις υποθέσεις ανεξαρτησίας που διατυπώσαμε νωρίτερα. Σημειώνουμε ότι η περιθώρια ύστερη $p(x_k|y_{1:k})$ προκύπτει από ολοκλήρωση της 5.28 ως προς $x_{0:k-1}$. Αντικαθιστώντας τις 5.27 και 5.28 στην 5.25, λαμβάνουμε την ακόλουθη εξίσωση ενημέρωσης των βαρών:

$$\begin{aligned} w_k^i &\propto \frac{p(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})} \\ &= \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)p(x_{0:k-1}^i|y_{1:k-1})}{q(x_k^i|x_{0:k-1}^i, y_{1:k})q(x_{0:k-1}^i|y_{1:k-1})} \\ &= w_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{0:k-1}^i, y_{1:k})} \end{aligned} \quad (5.29)$$

Όπως πάντα, απαιτούμε τα βάρη τη χρονική στιγμή k να έχουν μοναδιαίο άθροισμα.

Υποθέτουμε, επιπλέον, ότι $q(x_k|x_{0:k-1}, y_{1:k}) = q(x_k|x_{k-1}, y_k)$, δηλαδή ότι η κατανομή σημαντικότητας εξαρτάται μόνο από την προηγούμενη κατάσταση x_{k-1} και την τρέχουσα παρατήρηση y_k . Η παραδοχή αυτή είναι ιδιαίτερα χρήσιμη στην περίπτωση όπου μας ενδιαφέρει μόνο η εκτίμηση της περιθώριας ύστερης $p(x_k|y_{1:k})$, καθώς μας επιτρέπει να αποθηκεύουμε μόνο την x_k^i , αγνοώντας την ακολουθία καταστάσεων $x_{0:k-1}^i$ και το ιστορικό των παρατηρήσεων $y_{1:k-1}$. Στο σενάριο αυτό, όπου θα επικεντρωθούμε, η εξίσωση ενημέρωσης των βαρών διαμορφώνεται ως εξής:

$$w_k^i \propto w_{k-1}^i \frac{p(y_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, y_k)} \quad (5.30)$$

Για τα παραπάνω βάρη, η ζητούμενη ύστερη $p(x_k|y_{1:k})$ προσεγγίζεται ως:

$$p(x_k|y_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \quad (5.31)$$

Μπορεί να δειχθεί ότι, για $N_s \rightarrow \infty$, η 5.31 προσεγγίζει την πραγματική ύστερη πυκνότητα $p(x_k|y_{1:k})$ [68].

Ο αλγόριθμος SIS (Sequential Importance Sampling – ακολουθιακή δειγματοληψία σημαντικότητας) συνίσταται στην ακολουθιακή “χρονική διάδοση” (propagation) των σημείων (particles) που αναπαριστούν την ύστερη κάθε χρονική στιγμή με βάση την κατανομή σημαντικότητας και την ενημέρωση των αντίστοιχων βαρών όπως περιγράφηκε παραπάνω. Μια περιγραφή υπό τη μορφή ψευδοκώδικα για ένα βήμα της διαδικασίας δίνεται στον αλγόριθμο 6.

Algorithm 6 SIS Particle Filter

Input: $\{(x_{k-1}^i, w_{k-1}^i) : i = 1, \dots, N_s\}$, y_k

Output: $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$

- 1: **for** $i = 1 : N_s$ **do**
 - 2: Draw particle $x_k^i \sim q(x_k|x_{k-1}^i, y_k)$
 - 3: Assign the particle a weight w_k^i according to 5.30
 - 4: **end for**
 - 5: Normalize weights so that $\sum_{i=1}^{N_s} w_k^i = 1$
-

5.7.3 Εκφυλισμός του συνόλου αναπαράστασης

Μία σημαντική αδυναμία του SIS αλγορίθμου είναι ο εκφυλισμός (degeneracy) του συνόλου των τυχαίων δειγμάτων (particle set) που αναπαριστούν την ύστερη πυκνότητα $p(x_k|y_{1:k})$: μετά από

έναν (μικρό στην πράξη) αριθμό επαναλήψεων, τα βάρη όλων των σημείων (particles) με εξαίρεση ένα αποκτούν πρακτικά μηδενική τιμή. Από πρακτική άποψη, συνέχιση του αλγορίθμου πάνω σε ένα εκφυλισμένο σύνολο έχει σαν συνέπεια την εκτέλεση μεγάλου αριθμού υπολογιστικών βημάτων για τη χρονική διάδοση (propagation) και ενημέρωση των βαρών για καταστάσεις (particles) που έχουν αμελητέα συνεισφορά στην αναπαράσταση/προσέγγιση της $p(x_k|y_{1:k})$. Η αποφυγή του φαινομένου αυτού είναι αδύνατη: έχειδειχθεί ότι η μεταβλητότητα των βαρών στάθμισης μπορεί μόνο να αυξηθεί (στοχαστικά) με το χρόνο [72].

Ένα ενδεικτικό μέτρο του εκφυλισμού του συνόλου αναπαράστασης είναι το ενεργό μέγεθος δείγματος (effective sample size), N_{eff} , το οποίο ορίζεται ως [68, 73]:

$$N_{eff} = \frac{N_s}{1 + \text{Var}(w_k^{*i})}, \quad (5.32)$$

όπου τα w_k^{*i} θεωρούνται ως τα “πραγματικά βάρη” της δειγματοληψίας σημαντικότητας και δίνονται από:

$$w_k^{*i} = \frac{p(x_k^i|y_{1:k})}{q(x_k^i|x_{k-1}^i, y_k)}$$

Καθώς δεν είναι δυνατή η ακριβής αποτίμηση της ποσότητας αυτής, στην πράξη χρησιμοποιείται η εκτίμηση \hat{N}_{eff} , που είναι συνάρτηση των διαθέσιμων (κανονικοποιημένων σε μοναδιαίο άθροισμα) βαρών (5.29, 5.30) και δίνεται από:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2} \quad (5.33)$$

Από τον ορισμό του, $N_{eff} \leq N_s$ (και το ίδιο ισχύει για την εκτίμηση \hat{N}_{eff} από την ανισότητα Cauchy–Schwarz). Μικρή τιμή του μέτρου N_{eff} είναι ένδειξη εκφυλισμού του συνόλου αναπαράστασης.

Ο εκφυλισμός του συνόλου των σταθμισμένων δειγμάτων είναι ένα αναπόφευκτο και ανεπιθύμητο φαινόμενο. Μια αφελής προσέγγιση για το μετριασμό του θα ήταν η χρήση μιας (πολύ) μεγάλης τιμής για το πλήθος N_s των σταθμισμένων δειγμάτων (particles) που αναπαριστούν την ύστερη πυκνότητα, αλλά κάτι τέτοιο γρήγορα καθίσταται μη πρακτικό, ιδιαίτερα σε περιπτώσεις όπου η διάσταση του χώρου κατάστασης είναι υψηλή. Δύο πιο πρακτικές προσεγγίσεις του προβλήματος παρουσιάζονται στις ενότητες που ακολουθούν.

5.7.4 Επιλογή της κατανομής σημαντικότητας

Ένας τρόπος αντιμετώπισης του εκφυλισμού του particle set είναι μέσω της βέλτιστης επιλογής της κατανομής σημαντικότητας από την οποία πραγματοποιείται η δειγματοληψία σημαντικότητας για τη χρονική διάδοση των δειγμάτων. Από την κατανομή σημαντικότητας απαιτείται η ελαχιστοποίηση του $\text{Var}(w_k^{*i})$, ώστε να μεγιστοποιηθεί το μέγεθος N_{eff} της 5.32.

Έχειδειχθεί [72, 73] ότι η πυκνότητα της βέλτιστης κατανομής σημαντικότητας, η οποία ελαχιστοποιεί τη μεταβλητότητα των πραγματικών βαρών δεδομένων των x_{k-1}^i και y_k , είναι η $q_{opt}(x_k|x_{k-1}^i, y_k) = p(x_k|x_{k-1}^i, y_k)$. Η βέλτιστη πυκνότητα, λαμβάνοντας υπόψη τις υποθέσεις

ανεξαρτησίας μας, γράφεται ως:

$$\begin{aligned}
 q_{opt}(x_k|x_{k-1}^i, y_k) &= p(x_k|x_{k-1}^i, y_k) \\
 &= \frac{p(x_k, y_k|x_{k-1}^i)}{p(y_k|x_{k-1}^i)} \\
 &= \frac{p(y_k|x_k, x_{k-1}^i)p(x_k|x_{k-1}^i)}{p(y_k|x_{k-1}^i)} \\
 &= \frac{p(y_k|x_k)p(x_k|x_{k-1}^i)}{p(y_k|x_{k-1}^i)} \tag{5.34}
 \end{aligned}$$

Αντικαθιστώντας την 5.34 στην 5.30, παίρνουμε την ακόλουθη εξίσωση ενημέρωσης των βαρών:

$$w_k^i \propto w_{k-1}^i p(y_k|x_{k-1}^i) = \int_{\mathbb{R}^{n_x}} p(y_k|x_k)p(x_k|x_{k-1}^i) dx_k \tag{5.35}$$

Όπως προκύπτει από την 5.35, για αυτή την επιλογή της κατανομής σημαντικότητας, το βάρος w_k^i παίρνει την ίδια τιμή δεδομένης της x_{k-1}^i , οποιοδήποτε δείγμα και αν ληφθεί από την $q_{opt}(x_k|x_{k-1}^i, y_k)$ για την x_k^i . Έτσι, δεδομένης της x_{k-1}^i , είναι $\text{Var}(w_k^{*i}) = 0$: δεν υπάρχει μεταβλητότητα στις τιμές των w_k^i για τα διάφορα πιθανά δείγματα x_k^i .

Η επιλογή της $q_{opt}(x_k|x_{k-1}^i, y_k)$ ως πυκνότητας σημαντικότητας δεν είναι πάντοτε εφικτή. Μία προφανής αντικειμενική δυσκολία έγκειται στο ότι δεν είναι πάντοτε δυνατή η δειγματοληψία από την $p(x_k|x_{k-1}^i, y_k)$. Επιπλέον, το ολοκλήρωμα της 5.35 πάνω στην x_k μπορεί να μην έχει αναλυτική μορφή και ο υπολογισμός του να μην μπορεί να γίνει αποδοτικά. Δύο περιπτώσεις όπου η επιλογή της βέλτιστης κατανομής σημαντικότητας είναι δυνατή είναι αυτή του πεπερασμένου χώρου καταστάσεων (οπότε το ολοκλήρωμα της 5.35 είναι πεπερασμένο άθροισμα) και αυτή όπου η $p(x_k|x_{k-1}^i, y_k)$ είναι γκαουσιανή. Το τελευταίο μπορεί να προκύψει για ένα σύστημα της μορφής 5.1-5.2, όπου οι θόρυβοι v_{k-1} και w_k είναι προσθετικοί, γκαουσιανοί θόρυβοι μηδενικής μέσης τιμής και η h_k είναι επιπλέον γραμμική ως προς x_k [68, 72]. Σημειώνουμε ότι είναι δυνατή η κατασκευή προσεγγίσεων της βέλτιστης πυκνότητας σημαντικότητας χρησιμοποιώντας τεχνικές τοπικής γραμμικοποίησης, στις οποίες η $p(x_k|x_{k-1}^i, y_k)$ προσεγγίζεται από μια γκαουσιανή πυκνότητα [72].

Μία συνήθης επιλογή για την πυκνότητα σημαντικότητας είναι να τη λάβουμε ίση με την πρότερη που ορίζει το δυναμικό μοντέλο του συστήματος:

$$q(x_k|x_{k-1}^i, y_k) = p(x_k|x_{k-1}^i) \tag{5.36}$$

Στην περίπτωση αυτή, η ενημέρωση των βαρών 5.30 παίρνει την απλή μορφή:

$$w_k^i \propto w_{k-1}^i p(y_k|x_k^i), \tag{5.37}$$

όπου το βάρος κάθε δείγματος ενημερώνεται απλά με βάση την πιθανοφάνεια (likelihood) της συντρέχουσας παρατήρησης. Η παραπάνω επιλογή της πυκνότητας σημαντικότητας είναι συνηθισμένη στην πράξη, διαισθητικά δικαιολογημένη και οδηγεί σε απλές υλοποιήσεις, ενώ δεν διαθέτει κάποιο θεωρητικό πλεονέκτημα.

5.7.5 Επαναδειγματοληψία (resampling)

Μία άλλη μέθοδος αντιμετώπισης του φαινομένου του εκφυλισμού του συνόλου των σταθμισμένων δειγμάτων είναι η εισαγωγή ενός σταδίου επαναδειγματοληψίας (resampling). Η βασική ιδέα

αυτής της διαδικασίας είναι η εξάλειψη των δειγμάτων (particles) με μικρό βάρος, που ούτως ή άλλως συνεισφέρουν αμελητέα στην αναπαράσταση της ύστερης $p(x_k|y_{1:k})$, και η επικέντρωση στα περισσότερα σημαντικά. Η επαναδειγματοληψία μπορεί να πραγματοποιείται είτε στο τέλος κάθε χρονικού βήματος είτε όταν το ενεργό μέγεθος δείγματος N_{eff} γίνεται μικρότερο από κάποιο κατώφλι N_T , συνθήκη που λαμβάνεται σαν ένδειξη σοβαρού εκφυλισμού.

Έστω $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$ ένα σύνολο σταθμισμένων δειγμάτων που ορίζει την προσέγγιση $p_{N_s}(x_k|y_{1:k})$ της $p(x_k|y_{1:k})$:

$$p(x_k|y_{1:k}) \approx p_{N_s}(x_k|y_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \quad (5.38)$$

Το αποτέλεσμα της επαναδειγματοληψίας είναι η παραγωγή ενός νέου συνόλου σταθμισμένων δειγμάτων $\{(x_k^{i*}, w_k^{i*}) : i = 1, \dots, N_s\}$, όπου τα x_k^{i*} έχουν προκύψει από δειγματοληψία (με επανατοποθέτηση) από την 5.38: $\Pr(x_k^{i*} = x_k^i) = w_k^i$ για $i = 1, \dots, N_s$. Καθώς τα δείγματα του νέου συνόλου λαμβάνονται ανεξάρτητα από την 5.38 (independent, identically distributed: i.i.d.), τα νέα βάρη στάθμισης τίθενται ίσα μεταξύ τους:

$$w_k^{i*} = \frac{1}{N_s}, \text{ για } i = 1, \dots, N_s \quad (5.39)$$

Η νέα προσέγγιση $\hat{p}_{N_s}(x_k|y_{1:k})$ της ύστερης πυκνότητας με βάση το σύνολο των σταθμισμένων δειγμάτων μετά την επαναδειγματοληψία, $\{(x_k^{i*}, w_k^{i*}) : i = 1, \dots, N_s\}$, γράφεται:

$$\hat{p}_{N_s}(x_k|y_{1:k}) = \sum_{i=1}^{N_s} \frac{1}{N_s} \delta(x_k - x_k^{i*}) \quad (5.40)$$

Οι αποδείξεις σύγκλισης της νέας προσέγγισης $\hat{p}_{N_s}(x_k|y_{1:k})$ στην πραγματική ύστερη $p(x_k|y_{1:k})$ βασίζονται στην υπόθεση ότι η $\hat{p}_{N_s}(x_k|y_{1:k})$ είναι “κοντά” στην αρχική προσέγγιση $p_{N_s}(x_k|y_{1:k})$, δηλαδή ότι για κάθε συνάρτηση g ισχύει [74, 75]:

$$\begin{aligned} d &= \mathbb{E} \left[\left(\int_{\mathbb{R}^{n_x}} g(x_k) p_{N_s}(x_k|y_{1:k}) dx_k - \int_{\mathbb{R}^{n_x}} g(x_k) \hat{p}_{N_s}(x_k|y_{1:k}) dx_k \right)^2 \right] \\ &= \mathbb{E} \left[\left(\sum_{i=1}^{N_s} (w_k^i g(x_k^i)) - \frac{1}{N_s} \sum_{i=1}^{N_s} g(x_k^{i*}) \right)^2 \right] \rightarrow 0, \text{ καθώς } N_s \rightarrow \infty \end{aligned} \quad (5.41)$$

Η “απόσταση” των ολοκληρωμάτων/αθροισμάτων της 5.41 είναι ενδεικτική της ποιότητας της επαναδειγματοληψίας [76].

Οι περισσότερες χρησιμοποιούμενες μέθοδοι επαναδειγματοληψίας είναι πιθανοτικοί αλγόριθμοι γραμμικού χρόνου και βασίζονται στη δειγματοληψία από την πολυωνυμική κατανομή (multinomial distribution) ή είναι απλώς στιγμιότυπα αυτής της διαδικασίας. Οι τέσσερις “βασικοί” αλγόριθμοι επαναδειγματοληψίας στη βιβλιογραφία είναι η πολυωνυμική επαναδειγματοληψία (multinomial resampling), η επαναδειγματοληψία κατά στρώματα (stratified resampling), η επαναδειγματοληψία υπολοίπου (residual resampling) και η συστηματική επαναδειγματοληψία (systematic resampling). Οι παραπάνω αλγόριθμοι έχουν ασυμπτωτική πολυπλοκότητα $\mathcal{O}(N_s)$, καθώς βασίζονται στην παραγωγή σε γραμμικό χρόνο N_s ταξινομημένων δειγμάτων από την $\mathcal{U}[0, 1]$ (ομοιόμορφη κατανομή στο $[0, 1]$), αλλά παρουσιάζουν διαφορετικούς χρόνους εκτέλεσης και Monte Carlo μεταβλητότητες [76, 77]. Η συστηματική επαναδειγματοληψία είναι απλή στην

Algorithm 7 Systematic Resampling

Input: $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$
Output: $\{(x_k^{i*}, w_k^{i*}) : i = 1, \dots, N_s\}$

- 1: /* Construct CDF */
- 2: $c_1 = 0$
- 3: **for** $i = 2 : N_s$ **do**
- 4: $c_i = c_{i-1} + w_k^i$
- 5: **end for**
- 6: /* Move along the CDF */
- 7: $i = 1$
- 8: Draw a random starting point $u \sim \mathcal{U}\left[0, \frac{1}{N_s}\right]$
- 9: **for** $j = 1 : N_s$ **do**
- 10: $u_j = u + \frac{j-1}{N_s}$
- 11: **while** $u_j > c_i$ **do**
- 12: $i = i + 1$
- 13: **end while**
- 14: Assign new sample $x_k^{j*} = x_k^i$
- 15: Assign new weight $w_k^{j*} = \frac{1}{N_s}$
- 16: **end for**

υλοποίηση, έχει πολυπλοκότητα $\mathcal{O}(N_s)$ και παρουσιάζει καλά χαρακτηριστικά ελαχιστοποίησης της Monte Carlo μεταβλητότητας [68]. Μια υλοποίησή της παρουσιάζεται υπό τη μορφή ψευδοκώδικα στον αλγόριθμο 7.

Η επαναδειγματοληψία μειώνει αποτελεσματικά το φαινόμενο του εκφυλισμού, αφού το νέο σύνολο δειγμάτων είναι περισσότερο συγκεντρωμένο σε περιοχές του χώρου των καταστάσεων με μεγαλύτερη ύστερη πιθανότητα. Παρόλα αυτά, η εφαρμογή της εισάγει κάποια πρακτικά προβλήματα. Καθώς τα δείγματα με υψηλά βάρη επιλέγονται στατιστικά πολλές φορές, το νέο σύνολο δειγμάτων (particle set) θα περιέχει πολλές επαναλαμβανόμενες καταστάσεις. Αυτό το φαινόμενο της απώλειας της διαφορετικότητας εντός του συνόλου ονομάζεται *πτωχοποίηση* του δείγματος (sample impoverishment) και είναι ιδιαίτερα έντονο στην περίπτωση όπου ο θόρυβος της διαδικασίας είναι μικρός. Στην πράξη, για πολύ μικρό θόρυβο διαδικασίας, το σύνολο των δειγμάτων καταλήγει να αποτελείται από πολλαπλά αντίγραφα μιας μόνο κατάστασης μετά από ένα μικρό αριθμό επαναλήψεων. Ένα άλλο, περισσότερο τεχνικό πρόβλημα που σχετίζεται με την εφαρμογή της επαναδειγματοληψίας είναι οι περιορισμένες δυνατότητες παράλληλης υλοποίησης, καθώς πρέπει ολόκληρο το σύνολο να ληφθεί υπόψη.

5.7.6 Γενικός αλγόριθμος Particle Filter

Μία γενίκευση του βήματος του SIS φίλτρου στο οποίο έχει εισαχθεί ένα στάδιο επαναδειγματοληψίας δίνεται στον αλγόριθμο 8. Το στάδιο της επαναδειγματοληψίας πραγματοποιείται εφόσον η εκτίμηση \tilde{N}_{eff} του ενεργού μεγέθους δείγματος γίνει μικρότερη από κάποιο προκαθορισμένο κατώφλι. Μία άλλη στρατηγική για την εφαρμογή της επαναδειγματοληψίας είναι στο τέλος κάθε χρονικού βήματος, όπως συμβαίνει με το SIR φίλτρο που περιγράφεται παρακάτω.

Algorithm 8 Generic Particle Filter**Input:** $\{(x_{k-1}^i, w_{k-1}^i) : i = 1, \dots, N_s\}, y_k$ **Output:** $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$

- 1: **for** $i = 1 : N_s$ **do**
- 2: Draw particle $x_k^i \sim q(x_k | x_{k-1}^i, y_k)$
- 3: Assign the particle a weight w_k^i according to 5.30
- 4: **end for**
- 5: Normalize weights so that $\sum_{i=1}^{N_s} w_k^i = 1$
- 6: Calculate \hat{N}_{eff} according to 5.33
- 7: **if** $\hat{N}_{eff} < N_T$ **then**
- 8: Resample using algorithm 7 and update particle set $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$
- 9: **end if**

5.7.7 SIR Particle Filter

Ο αλγόριθμος SIS και η γενίκευσή του αποτελούν τη βάση για τα περισσότερα μέλη της οικογένειας των particle filters που έχουν αναπτυχθεί. Οι διάφορες μέθοδοι σε αυτή την κατηγορία μπορεί να θεωρηθούν σαν ειδικές περιπτώσεις του αλγορίθμου αυτού, για διαφορετικές επιλογές της κατανομής σημαντικότητας ή/και της στρατηγικής/μεθόδου επαναδειγματοληψίας που εφαρμόζεται.

Ο αλγόριθμος SIR (Sampling Importance Resampling) είναι μια τέτοια μέθοδος, όπου ως πυκνότητα σημαντικότητας $q(x_k | x_{k-1}^i, y_k)$ επιλέγεται η πρότερη $p(x_k | x_{k-1}^i)$ και το στάδιο της επαναδειγματοληψίας πραγματοποιείται στο τέλος κάθε χρονικού βήματος. Οι προϋποθέσεις για την εφαρμογή του SIR φίλτρου είναι αρκετά χαλαρές και συνοψίζονται στη δυνατότητα δειγματοληψίας από την $p(x_k | x_{k-1}^i)$ και τη δυνατότητα υπολογισμού της τιμής της πυκνότητας πιθανοφάνειας $p(y_k | x_k^i)$ τουλάχιστον μέχρι έναν παράγοντα (up to proportionality).

Σε όρους συστημικής περιγραφής της μορφής 5.1-5.2, προϋποτίθεται γνώση των συναρτήσεων f_k και h_k του δυναμικού μοντέλου και του μοντέλου παρατήρησης, καθώς και η δυνατότητα δειγματοληψίας από την κατανομή του θορύβου της διαδικασίας, v_{k-1} . Σημειώνουμε ότι ένα τυχαίο δείγμα x_k^i από την $p(x_k | x_{k-1}^i)$ μπορεί να παραχθεί από ένα τυχαίο δείγμα της διαδικασίας θορύβου v_{k-1}^i , θέτοντας $x_k^i = f_k(x_{k-1}^i, v_{k-1}^i)$.

Όπως έχει αναφερθεί προηγουμένως, για αυτή την επιλογή της κατανομής σημαντικότητας, τα βάρη ενημερώνονται σύμφωνα με την 5.37. Καθώς η επαναδειγματοληψία πραγματοποιείται στο τέλος κάθε χρονικού βήματος, στο k -οστό βήμα είναι $w_{k-1}^i = \frac{1}{N_s}$, για κάθε $i = 1, \dots, N_s$, και η ενημέρωση των βαρών στάθμισης παίρνει την ακόμη απλούστερη μορφή:

$$w_k^i \propto p(y_k | x_k^i) \quad (5.42)$$

Όπως πάντα, τα βάρη κανονικοποιούνται στο μοναδιαίο άθροισμα προτού εφαρμοστεί η επαναδειγματοληψία. Μία περιγραφή υπό τη μορφή ψευδοκώδικα για ένα βήμα του φίλτρου SIR δίνεται στον αλγόριθμο 9.

Μία από τις αδυναμίες του SIR φίλτρου προέρχεται από το γεγονός ότι η πυκνότητα σημαντικότητας είναι ανεξάρτητη από την τρέχουσα παρατήρηση y_k . Καθώς η εξερεύνηση του χώρου των καταστάσεων γίνεται χωρίς γνώση των παρατηρήσεων, η αναζήτηση που εκτελεί το φίλτρο αυτό είναι εκ των πραγμάτων λιγότερο αποδοτική από ό,τι θα μπορούσε. Επίσης, το γεγονός ότι η επαναδειγματοληψία πραγματοποιείται κάθε χρονική στιγμή μπορεί να οδηγήσει σε απώλεια της διαφορετικότητας του συνόλου των δειγμάτων. Παρόλα αυτά, το SIR φίλτρο διαθέτει τα πλεονεκτήματα των χαλαρών προϋποθέσεων για την εφαρμογή του (είναι ευρέως εφαρμόσιμο) και της

Algorithm 9 SIR Particle Filter

Input: $\{(x_{k-1}^i, w_{k-1}^i) : i = 1, \dots, N_s\}, y_k$ **Output:** $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$

- 1: **for** $i = 1 : N_s$ **do**
 - 2: Draw particle $x_k^i \sim p(x_k | x_{k-1}^i)$
 - 3: Assign the particle a weight $w_k^i \propto p(y_k | x_k^i)$
 - 4: **end for**
 - 5: Normalize weights so that $\sum_{i=1}^{N_s} w_k^i = 1$
 - 6: Resample using algorithm 7 and overwrite particle set $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$
-

εύκολης υλοποίησης: η δειγματοληψία από την $p(x_k | x_{k-1}^i)$ γίνεται συνήθως εύκολα και το ίδιο ισχύει για τον υπολογισμό των βαρών στάθμισης σε κάθε βήμα.

Κεφάλαιο 6

Θεωρητική περιγραφή της μεθόδου οπτικής παρακολούθησης

Στο κεφάλαιο αυτό παρουσιάζουμε τη μέθοδο οπτικής παρακολούθησης ενός κινούμενου αντικειμένου που αποτελεί το βασικό προϊόν της παρούσας εργασίας. Η βασική μας μέθοδος εντάσσεται στο θεωρητικό πλαίσιο του Bayesian filtering που περιγράψαμε στο Κεφάλαιο 5 και βασίζεται σε ένα Particle Filter, το οποίο χρησιμοποιεί μετρήσεις που προκύπτουν το χρώμα και τη διδιάστατη κίνηση (οπτική ροή) που χαρακτηρίζουν την ακολουθία των εικόνων εισόδου. Για λόγους σύγκρισης, υλοποιήθηκε επίσης ένας “αντίστοιχος” αλγόριθμος οπτικής παρακολούθησης που βασίζεται στο φίλτρο Kalman. Στα Κεφάλαια 3 και 4 παρουσιάζονται, αντίστοιχα, όλες οι σχετικές με το χρώμα και τη διδιάστατη κίνηση έννοιες και τεχνικές που θα χρειαστούμε για την περιγραφή των μεθόδων μας.

Οι αλγόριθμοί μας δέχονται ως είσοδο μια ακολουθία βίντεο, ένα στατιστικό χρωματικό μοντέλο για το υπό παρακολούθηση αντικείμενο (ενότητα 3.2) και τη θέση αυτού στο αρχικό πλαίσιο της ακολουθίας, μέσω χειροκίνητης αρχικοποίησης της διαδικασίας. Ο εντοπισμός του αντικειμένου σε κάθε πλαίσιο γίνεται, στη συνέχεια, ακολουθιακά, με βάση το αποτέλεσμα του εντοπισμού στο προηγούμενο πλαίσιο και ένα σύνολο μετρήσεων που προκύπτουν από τη χρωματική ανάλυση του προηγούμενου και του τρέχοντος πλαισίου καθώς και της διδιάστατης κίνησης (οπτική ροή) που τα συνδέει. Ένα από τα σημαντικότερα χαρακτηριστικά της κύριας μεθόδου παρακολούθησής μας (Particle Filter), με τρόπο που θα αποσαφηνιστεί στα επόμενα, είναι η θεώρηση ενός *προσαρμοστικού* δυναμικού μοντέλου του υπό παρακολούθηση συστήματος. Σε κάθε χρονικό βήμα, η μέθοδός μας *μαθαίνει* ένα σύνολο παραμέτρων για το θεωρούμενης μορφής μοντέλο μετάβασης από ένα σύνολο μετρήσιμων μεγεθών που προκύπτουν από τη χρωματική ανάλυση και την ανάλυση της κίνησης που χαρακτηρίζει το προηγούμενο και το τρέχον πλαίσιο της ακολουθίας εισόδου. Η θεώρηση ενός “πληροφορημένου” δυναμικού μοντέλου μπορεί να διευκολύνει σημαντικά την αναζήτηση στο χώρο των καταστάσεων, καθώς οδηγεί την τελευταία περιορίζοντάς την στο σύνολο των “έγκυρων” (περισσότερο πιθανών) καταστάσεων, και καθιστά ένα εκλεπτυσμένο μοντέλο παρατήρησης λιγότερο αναγκαίο.

Οι περιοριστικές υποθέσεις της μεθόδου μας που οριοθετούν το πρόβλημα της παρακολούθησης και καθορίζουν το σύνολο των πιθανών εφαρμογών είναι δύο. Πρώτον, όπως αναφέραμε, απαιτείται εκ των προτέρων γνώση του χρώματος (στατιστικού μοντέλου) του υπό παρακολούθηση αντικειμένου και, προαιρετικά, του παρασκηνίου των εικόνων. Δεύτερον, υποτίθεται ότι η κίνηση του αντικειμένου είναι *κατά κύριο λόγο* μεταφορική. Η έννοια αυτή, με τρόπο που θα απο-

σαφηνιστεί στη συνέχεια, αναφέρεται στην υπόθεση ότι η κύρια συνιστώσα διδιάστατης κίνησης στην έκταση της προβολής του αντικειμένου, την οποία θα ορίσουμε παρακάτω, είναι αντιπροσωπευτική της συνολικής κίνησης (ακριβέστερα, της κίνησης του κέντρου) της προβολής του αντικειμένου στο επίπεδο της εικόνας. Σημειώνουμε ότι η υπόθεση αυτή δεν θέτει ιδιαίτερα ισχυρούς περιορισμούς στο σύνολο των τρόπων κίνησης για τους οποίους η μέθοδός μας αναμένεται να λειτουργεί σωστά, καθώς δεν αποκλείονται, επί της αρχής, σύνθετες κινήσεις μη άκαμπτων αντικειμένων.

Στις ενότητες που ακολουθούν, ορίζονται ο θεωρούμενος χώρος καταστάσεων, ένα σύνολο ενδιάμεσων αποτελεσμάτων που χρησιμοποιούνται από τους αλγορίθμους παρακολούθησης, τα γεωμετρικά σχήματα των δύο αλγορίθμων και, τέλος, τα βασικά συστατικά τους (δυναμικό μοντέλο και μοντέλο παρατηρήσεων, κατά περίπτωση). Στο Παράρτημα C, παραθέτουμε ένα σύνολο απεικονίσεων για κάποια από τα ενδιάμεσα μεγέθη και αποτελέσματα που θα περιγράψουμε στη συνέχεια για μια πραγματική ακολουθία βίντεο.

6.1 Ορισμός χώρου καταστάσεων

Το πρόβλημα της οπτικής παρακολούθησης με το οποίο ασχολούμαστε έχει ως ζητούμενο τον προσδιορισμό, σε κάθε πλαίσιο της κινούμενης εικόνας, της θέσης και της ταχύτητας του αντικειμένου ενδιαφέροντος. Για την περιγραφή της θέσης, χρησιμοποιούμε ένα πλαίσιο οριοθέτησης (bounding box) του αντικειμένου, το οποίο ορίζεται ως η μικρότερη έκταση ορθογώνια περιοχή του επιπέδου της εικόνας, με πλευρές παράλληλες στους άξονες, που περικλείει την προβολή του αντικειμένου στο τρέχον πλαίσιο.

Ένα διδιάστατο πλαίσιο οριοθέτησης έχει 4 βαθμούς ελευθερίας (θέση και μέγεθος) και ένα διάνυσμα διδιάστατης ταχύτητας 2. Συνεπώς, απαιτείται η θεώρηση ενός χώρου καταστάσεων 6 διαστάσεων. Ένα διάνυσμα κατάστασης έχει τη μορφή:

$$(x, y, w, h, x', y') \in \mathbb{R}^6, \quad (6.1)$$

όπου:

- $p = (x, y)$ είναι η θέση του κέντρου του πλαισίου οριοθέτησης στο επίπεδο της εικόνας
- $s = (w, h)$ είναι το μέγεθος του πλαισίου (μήκη στις x και y διαστάσεις)
- $v = (x', y')$ είναι η ταχύτητα του αντικειμένου/πλαίσιου

Στις 3 παραπάνω ομάδες των μεταβλητών κατάστασης θα αναφερόμαστε αντίστοιχα απλώς με τους όρους *θέση*, *μέγεθος (πλαίσιο)* και *ταχύτητα*.

Σημειώνουμε ότι η εκτίμηση της ταχύτητας δεν αποτελεί πρωταρχικό στόχο στο πρόβλημα της παρακολούθησης που μας απασχολεί και, έχοντας βοηθητικό χαρακτήρα, δε θα αξιολογηθεί ως πειραματικό αποτέλεσμα στο Κεφάλαιο 7. Η χρησιμότητα της συμπερίληψης της ταχύτητας στο διάνυσμα κατάστασης της μεθόδου παρακολούθησης θα αποσαφηνιστεί, κατά περίπτωση, στα επόμενα.

6.2 Χρήσιμα ενδιάμεσα μεγέθη και συμβολισμοί

Οι αλγόριθμοι παρακολούθησης που θα ορίσουμε λειτουργούν αναδρομικά: ο εντοπισμός του αντικειμένου στο πλαίσιο k γίνεται με δεδομένο το αποτέλεσμα του προηγούμενου βήματος στο πλαίσιο $k - 1$. Συμβολίζουμε με I_k το k -οστό πλαίσιο (καρέ) της ακολουθίας βίντεο εισόδου και,

διατηρώντας τις συμβάσεις του Κεφαλαίου 5, με x_k την εκτίμηση της κατάστασης του συστήματος τη χρονική στιγμή (πλαίσιο) k .

Υποθέτουμε ότι οι ακολουθίες βίντεο που παρέχονται ως είσοδοι στους αλγορίθμους μας έχουν καταγραφεί με μοναδιαία περίοδο χρονικής δειγματοληψίας ($\Delta t = 1$). Καθώς η χρονική διαφορά από το πλαίσιο I_{k-1} μέχρι το πλαίσιο I_k είναι ίση, με βάση αυτήν μας την υπόθεση, με τη μονάδα του χρόνου, οι τιμές των μετρήσεων της ταχύτητας και της μετατόπισης σημείων ή αντικειμένων, με βάση αυτά τα δύο πλαίσια, ταυτίζονται. Κατά συνέπεια, όπως θα δούμε στα επόμενα, δεν γίνεται πουθενά ρητή αναφορά στο χρόνο (χρήση της τιμής του k ή του $t = k\Delta t$) από τις εξισώσεις που περιγράφουν τα μοντέλα μας. Παρακάτω, θα χρησιμοποιούμε εναλλάξιμα τους όρους *πεδίο ταχυτήτων* και *πεδίο μετατοπίσεων* για κάποιο πεδίο οπτικής ροής που συνδέει δύο διαδοχικά πλαίσια.

Το k -οστό βήμα των αλγορίθμων παρακολούθησης λαμβάνει υπόψη τα πλαίσια I_{k-1} και I_k και περιλαμβάνει τον υπολογισμό ενός συνόλου βοηθητικών, ενδιάμεσων αποτελεσμάτων, τα οποία χρησιμεύουν στον ορισμό των κατά περίπτωση μοντέλων μετάβασης και μοντέλων παρατήρησης. Παρακάτω, ορίζουμε τα πιο σημαντικά από αυτά και εισάγουμε τους συμβολισμούς που θα χρησιμοποιήσουμε στις ακόλουθες ενότητες. Ο αναγνώστης μπορεί να ανατρέξει στο Παράρτημα C για παραδείγματα της μορφής των μεγεθών αυτών, καθώς και άλλων από τη συνέχεια του κεφαλαίου, στην πράξη, για μια πραγματική ακολουθία βίντεο.

Συμβολισμοί. Στα επόμενα, θα χρησιμοποιήσουμε το συμβολισμό $x_{[i]}$ για την i -οστή συνιστώσα του διανύσματος x . Με $x_{[i_1, \dots, i_N]}$ θα συμβολίζουμε το διάνυσμα διάστασης N , του οποίου η j συνιστώσα, για $j = 1, \dots, N$, είναι ίση με την i_j συνιστώσα του x .

Σύμφωνα με τον ορισμό του χώρου καταστάσεών μας, ένα διάνυσμα κατάστασης ορίζει ένα πλαίσιο οριοθέτησης. Αν x είναι ένα διάνυσμα κατάστασης, συμβολίζουμε με $B(x)$ τη *μάσκα κατάστασης*: η $B(x)$ είναι μια δυαδική εικόνα που υποδεικνύει ακριβώς τις θέσεις του επιπέδου της εικόνας που ανήκουν στο εσωτερικό ή πάνω στο σύνορο του πλαισίου οριοθέτησης που ορίζει η x . Η $B(x)$ ορίζει πάντα μια ορθογώνια, απλά συνεκτική περιοχή του επιπέδου της εικόνας, με πλευρές παράλληλες στους άξονες συντεταγμένων. Καθώς οι συνιστώσες του διανύσματος κατάστασης x είναι πραγματικοί αριθμοί, στις υλοποιήσεις μας πραγματοποιούνται οι απαραίτητες στρογγυλοποιήσεις στους εγγύτερους ακεραίους για τις θέσεις των πλευρών της $B(x)$ (ακριβεία εικονοστοιχείου).

Χρωματικές μάσκες. Για τα πλαίσια I_{k-1} και I_k υπολογίζουμε αντίστοιχα τις χρωματικές μάσκες B_{k-1}^c και B_k^c του αντικειμένου. Οι B_{k-1}^c και B_k^c είναι *δυαδικές εικόνες* που προκύπτουν από την κατάτμηση των I_{k-1} και I_k σε δύο κλάσεις με βάση το χρώμα: το αντικείμενο ενδιαφέροντος (προσκήνιο) και το παρασκήνιο της εικόνας.

Για τον υπολογισμό των χρωματικών μασκών, χρησιμοποιείται το χρωματικό μοντέλο του αντικειμένου που παρέχεται σαν παράμετρος εισόδου στους αλγορίθμους μας και, προαιρετικά, ένα χρωματικό μοντέλο για το παρασκήνιο. Τα θεωρούμενα χρωματικά μοντέλα κωδικοποιούν τις πιθανοφάνειες $p(c|\omega_i)$ και παρέχουν εκτιμήσεις των πρότερων πιθανοτήτων $P(\omega_i)$, όπου τα ω_i , $i = 1, 2$, αντιστοιχούν στο αντικείμενο και το παρασκήνιο και το c είναι μια χρωματική τιμή. Όταν δεν παρέχεται χρωματικό μοντέλο παρασκήνιου στην είσοδο, υποθέτουμε ότι η $p(c|\omega_2)$ είναι ομοιόμορφη και όλες οι χρωματικές τιμές θεωρούνται εξίσου πιθανές για το παρασκήνιο. Τα χρωματικά μοντέλα εισόδου μπορεί να είναι παραμετρικά (μείγματα Γκαουσιανών κατανομών) ή μη παραμετρικά (ιστογράμματα χρωματικών τιμών) και η εκπαίδευσή τους έχει γίνει offline, πριν την εφαρμογή των αλγορίθμων παρακολούθησης, με κάποια από τις τεχνικές που περιγράψαμε στην ενότητα 3.2.

Η κατάτμηση των I_{k-1} και I_k πραγματοποιείται με βάση ένα στατιστικό έλεγχο που βασίζεται στο λόγο των πιθανοφανειών της μορφής 3.22 (ενότητα 3.3.1):

$$\alpha(x) = \omega_1 \Leftrightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} > \tau,$$

όπου το κατώφλι τ προσδιορίζεται στο στάδιο εκπαίδευσης των χρωματικών μοντέλων και παρέχεται ως είσοδος σαν μέρος αυτών. Σημειώνουμε ότι, σύμφωνα με την ενότητα 3.3.1 (εξίσωση 3.21), η πιο ορθή επιλογή για το κατώφλι τ , δεδομένης της επιλογής της “0-1” συνάρτησης απώλειας, είναι η:

$$\tau = \frac{P(\omega_2)}{P(\omega_1)}$$

Καθώς οι εκτιμήσεις των πρότερων πιθανοτήτων $P(\omega_1)$ και $P(\omega_2)$ δεν είναι πάντα αξιόπιστες, συχνά απαιτείται η χειροκίνητη ρύθμιση του τ εμπειρικά (trial and error).

Για αρκετές περιπτώσεις κατανομών $p(c|\omega_i)$, $i = 1, 2$, η προβολή του κινούμενου αντικειμένου στο επίπεδο της εικόνας προσδιορίζεται χωρικά απλώς από μια συνεκτική συνιστώσα της μάσκας B_k^c και το πρόβλημα του εντοπισμού διευκολύνεται δραματικά (ενότητα 6.5). Όπως θα δούμε παρακάτω, η μάσκα B_{k-1}^c χρησιμοποιείται για τον ορισμό του δυναμικού μοντέλου μετάβασης του συστήματος, ενώ η B_k^c για τον ορισμό του μοντέλου παρατήρησης τη χρονική στιγμή k .

Πεδία οπτικής ροής. Υπολογίζουμε την προς τα εμπρός ροή (forward flow) από το πλαίσιο I_{k-1} στο I_k και την προς τα πίσω ροή (backward flow) από το πλαίσιο I_k στο I_{k-1} και συμβολίζουμε τα διανυσματικά πεδία μετατοπίσεων, αντίστοιχα, με $f_k^{f,1}$ και $f_k^{b,1}$. Ο υπολογισμός της προς τα πίσω ροής (από την I_k στην I_{k-1}) αποτελεί συνήθη πρακτική για τον προσδιορισμό περιοχών απόκρυψης (occlusion), μέσω του ελέγχου της συνέπειας της προς τα εμπρός με την προς τα πίσω ροή [39]. Καθώς η εκτίμηση της οπτικής ροής από μια εικόνα σε μια άλλη μπορεί να είναι αναξιόπιστη για πληθώρα λόγων, εμείς θα χρησιμοποιήσουμε την προς τα πίσω ροή $f_k^{b,1}$ για τον ορισμό μιας δεύτερης εκδοχής προς τα εμπρός ροής, $f_k^{f,2}$, και, αντιστοίχως, την $f_k^{f,1}$ για τον ορισμό μιας δεύτερης εκδοχής προς τα πίσω ροής, $f_k^{b,2}$.

Η αλλαγή κατεύθυνσης ενός πεδίου οπτικής ροής βασίζεται στην παρατήρηση ότι, στις περιοχές όπου δεν έχουμε απόκρυψη (occlusion), ένα διάνυσμα προς τα πίσω ροής πρέπει να είναι αντίθετο από το διάνυσμα της προς τα εμπρός ροής στη θέση που δείχνει. Αν συμβολίσουμε με (u_f, v_f) και (u_b, v_b) αντίστοιχα τα πεδία της προς τα εμπρός και της προς τα πίσω οπτικής ροής ανάμεσα σε δύο εικόνες, η παραπάνω σχέση που τα συνδέει ποσοτικοποιείται από τις εξισώσεις:

$$\begin{aligned} u_b(x, y) &= -u_f(x + u_b(x, y), y + v_b(x, y)) \\ v_b(x, y) &= -v_f(x + u_b(x, y), y + v_b(x, y)) \end{aligned} \quad (6.2)$$

Με δεδομένο ένα πεδίο προς τα πίσω ροής, η 6.2 μπορεί να χρησιμοποιηθεί για τον ορισμό ενός πεδίου προς τα εμπρός ροής (u_f, v_f) , του οποίου οι τιμές είναι γνωστές σε όλες τις θέσεις $(x + u_b(x, y), y + v_b(x, y))$ του επιπέδου της εικόνας. Ο υπολογισμός του (u_f, v_f) στις θέσεις (x, y) μπορεί να γίνει μέσω παρεμβολής. Καθώς το σύνολο των σημείων $(x + u_b(x, y), y + v_b(x, y))$ όπου το (u_f, v_f) είναι γνωστό μπορεί να είναι διεσπαρμένο με αυθαίρετο τρόπο στο επίπεδο της εικόνας, ο υπολογισμός της παρεμβάλλουσας βασίζεται σε έναν Delaunay τριγωνισμό αυτού. Η αλλαγή κατεύθυνσης ενός πεδίου προς τα εμπρός ροής γίνεται με βάση την τελείως συμμετρική παρατήρηση ότι ένα διάνυσμα προς τα εμπρός ροής πρέπει να είναι αντίθετο από το διάνυσμα της προς τα πίσω ροής στη θέση που δείχνει:

$$\begin{aligned} u_f(x, y) &= -u_b(x + u_f(x, y), y + v_f(x, y)) \\ v_f(x, y) &= -v_b(x + u_f(x, y), y + v_f(x, y)) \end{aligned} \quad (6.3)$$

Με βάση τα παραπάνω, οι 6.2 χρησιμοποιούνται για τον υπολογισμό της δεύτερης εκδοχής προς τα εμπρός ροής, $f_k^{f,2}$, από την προς τα πίσω ροή $f_k^{b,1}$, ενώ οι (συμμετρικές τους) 6.3 για τον υπολογισμό της δεύτερης εκδοχής προς τα πίσω ροής, $f_k^{b,2}$, από την προς τα εμπρός ροή $f_k^{f,1}$.

Όπως θα δούμε στα επόμενα, οι δύο εκδοχές της προς τα εμπρός ροής, $f_k^{f,1}$ και $f_k^{f,2}$, θα χρησιμοποιηθούν για τον ορισμό του δυναμικού μοντέλου μετάβασης του συστήματος, ενώ τα πεδία της προς τα πίσω ροής, $f_k^{b,1}$ και $f_k^{b,2}$, για τον ορισμό του μοντέλου παρατήρησης τη χρονική στιγμή k .

Κατατμήσεις με βάση την κίνηση. Με βάση τα πεδία της προς τα εμπρός ροής $f_k^{f,1}$, $f_k^{f,2}$ και τη χρωματική μάσκα B_{k-1}^c , υπολογίζουμε δύο κατατμήσεις με βάση την κίνηση για την εικόνα I_{k-1} . Οι κατατμήσεις πραγματοποιούνται ακολουθώντας την απλή μέθοδο που περιγράψαμε στην ενότητα 4.4, με τη διαφορά ότι ταξινομούμε μόνο τα διανύσματα κίνησης των οποίων η αφετηρία ανήκει στη χρωματική μάσκα B_{k-1}^c , ορίζοντας, ουσιαστικά, μια κατάτμηση αυτής. Με βάση τα πεδία της προς τα πίσω ροής $f_k^{b,1}$, $f_k^{b,2}$ και τη χρωματική μάσκα B_k^c , κάνουμε το ίδιο για την I_k .

Η εφαρμογή της ταξινόμησης k -means (Παράρτημα Α) προϋποθέτει μια εκ των προτέρων εκτίμηση του πλήθους των κλάσεων της εκάστοτε κατάτμησης. Οι αλγόριθμοι παρακολούθησής μας χρησιμοποιούν την ίδια τιμή N_c για το μέγεθος αυτό για όλες τις κατατμήσεις με βάση την κίνηση που πραγματοποιούν και την οποία δέχονται ως είσοδο. Η τιμή της N_c εκτιμάται εμπειρικά και εξαρτάται από τα πρότυπα κίνησης που υπάρχουν στην ακολουθία εικόνων εισόδου. Το γεγονός ότι επικεντρώνουμε το ενδιαφέρον μας στην ταξινόμηση μόνο εκείνων των διανυσμάτων κίνησης που έχουν την αφετηρία τους στις θεωρούμενες χρωματικές μάσκες διευκολύνει σημαντικά την επιλογή της N_c , καθώς πρέπει να εκτιμηθεί μόνο το πλήθος των συνιστωσών κίνησης σε περιοχές της εικόνας που είναι χρωματικά όμοιες με το κινούμενο αντικείμενο που παρακολουθούμε. Για παράδειγμα, στο σενάριο παρακολούθησης του δεξιού χεριού σε μια ακολουθία βίντεο νοηματικής γλώσσας, όπου η χρωματική μάσκα κάθε πλαισίου υποδεικνύει τα εικονοστοιχεία που απεικονίζουν ανθρώπινο δέρμα, μια εύλογη επιλογή τιμής για την παράμετρο είναι $N_c = 4$ (χέρια, πρόσωπο, “παρασκήνιο”).

Η κατάτμηση μιας εικόνας με βάση ένα πεδίο οπτικής ροής και μια χρωματική μάσκα αντιστοιχίζει κάθε εικονοστοιχείο της εικόνας που ανήκει στη μάσκα σε μία από τις τιμές του συνόλου $\{1, \dots, N_c\}$. Προϊόν της κατάτμησης αποτελεί, επιπλέον, το κεντροειδές κάθε κλάσης, το οποίο θυμίζουμε ότι είναι η μέση τιμή των διανυσμάτων κίνησης που ανήκουν στην κλάση. Θα αναπαραστήσουμε μια κατάτμησης αυτής της μορφής ως το σύνολο:

$$S_k^\alpha \equiv \left\{ \left(B_k^{\alpha,i}, c_k^{\alpha,i} \right) : i = 1, \dots, N_c \right\}, \quad (6.4)$$

όπου $c_k^{\alpha,i}$ είναι το κεντροειδές της κλάσης i για αυτή την κατάτμηση και $B_k^{\alpha,i}$ είναι μια δυαδική εικόνα που υποδεικνύει ακριβώς τα εικονοστοιχεία της εικόνας που ανήκουν στην κλάση i . Η παράμετρος α εξυπηρετεί ανάγκες παραμετρικού συμβολισμού και η τιμή της υποδηλώνει τον τύπο της κατάτμησης: $S_k^{f,1}$ και $S_k^{f,2}$ είναι, αντίστοιχα, οι κατατμήσεις της I_{k-1} με βάση τα πεδία $f_k^{f,1}$ και $f_k^{f,2}$ για τη μάσκα B_{k-1}^c και $S_k^{b,1}$, $S_k^{b,2}$ είναι, αντίστοιχα, οι κατατμήσεις της I_k με βάση τα πεδία $f_k^{b,1}$ και $f_k^{b,2}$ για τη μάσκα B_k^c .

Καθώς ταξινομούμε μόνο διανύσματα κίνησης με αφετηρία σε μια δεδομένη χρωματική μάσκα, οι κατατμήσεις μας ουσιαστικά ορίζουν διαμερίσεις της θεωρούμενης μάσκας. Το γεγονός αυτό συνεπάγεται ότι οι δυαδικές εικόνες αναπαράστασης των κατατμήσεων ικανοποιούν:

$$\bigcup_{i=1, \dots, N_c} B_k^{f,\alpha,i} = B_{k-1}^c \quad \text{και} \quad \bigcup_{i=1, \dots, N_c} B_k^{b,\alpha,i} = B_k^c, \quad \text{για } \alpha = 1, 2 \quad (6.5)$$

Αναφέρουμε εκ των προτέρων ότι οι κατατμήσεις με βάση την προς τα εμπρός κίνηση, $S_k^{f,1}$ και $S_k^{f,2}$, θα χρησιμοποιηθούν για τον ορισμό του δυναμικού μοντέλου του συστήματος, ενώ οι κατατμήσεις με βάση την προς τα πίσω κίνηση, $S_k^{b,1}$ και $S_k^{b,2}$, για τον ορισμό του μοντέλου παρατήρησης τη χρονική στιγμή k .

Κύρια συνιστώσα κίνησης. Ορίζουμε την *κύρια συνιστώσα κίνησης* μιας κατάτμησης με βάση την κίνηση S_k^α ως προς μια μάσκα (δυναδική εικόνα) αναφοράς B , ως εκείνη την κλάση (cluster) της S_k^α που καλύπτει το μεγαλύτερο μέρος της μάσκας B . Καθώς μια δυναδική εικόνα ερμηνεύεται άμεσα ως μια αναπαράσταση ενός συνόλου εικονοστοιχείων, χρησιμοποιώντας την αναπαράσταση 6.4 της κατάτμησης S_k^α , η κύρια συνιστώσα κίνησης της S_k^α ως προς την B ορίζεται ως:

$$PMC(S_k^\alpha, B) \equiv \operatorname{argmax}_{i=1, \dots, N_c} |B_k^{\alpha, i} \cap B|, \quad (6.6)$$

όπου με $|X|$ συμβολίζουμε τον πληθάρημο της δυναδικής εικόνας/συνόλου X (πλήθος των εικονοστοιχείων που υποδεικνύει).

Σημειώνουμε ότι το μέγεθος αυτό θα χρησιμοποιηθεί, μεταξύ άλλων, για τον ορισμό του δυναμικού μοντέλου της μεθόδου παρακολούθησης με Particle Filter (ενότητα 6.3.1), όπου θα αποσαφηνιστεί η υπόθεση της *κατά κύριο λόγο* μεταφορικής κίνησης της μεθόδου.

Εκτίμηση αλλαγής μεγέθους. Τα μεγέθη που ορίσαμε παραπάνω, με τρόπο που θα αποσαφηνιστεί στα επόμενα, μας παρέχουν εργαλεία εκτίμησης των συνιστωσών θέσης και ταχύτητας της κατάστασης του συστήματος. Χρειαζόμαστε, επιπλέον, μία μέθοδο εκτίμησης της μεταβολής των διαστάσεων (μεγέθους) του πλαισίου οριοθέτησης που ορίζει η υπό εκτίμηση κατάσταση. Θα περιγράψουμε έναν απλό τρόπο εκτίμησης συντελεστών κλιμάκωσης για τις δύο διαστάσεις του πλαισίου οριοθέτησης, ο οποίος βασίζεται στον υπολογισμό λόγων αποστάσεων πριν και μετά την “εφαρμογή” ενός πεδίου προς τα εμπρός ροής.

Ο υπολογισμός των συντελεστών κλιμάκωσης πραγματοποιείται με βάση ένα πεδίο προς τα εμπρός ροής $f_k^{f, \alpha}$ (το οποίο μπορεί να είναι ένα από τα $f_k^{f, 1}$, $f_k^{f, 2}$ για $\alpha = 1, 2$) και μια δυναδική εικόνα B που ορίζει σε ποια περιοχή της εικόνας πρέπει να εκτιμηθεί η αλλαγή κλίμακας. Καθώς, στα επόμενα, στόχος μας θα είναι μια εκτίμηση για την αλλαγή μεγέθους του πλαισίου οριοθέτησης της κατάστασης x_{k-1} , τα εικονοστοιχεία που υποδεικνύει η B θα περικλείονται από το πλαίσιο αυτό: $B \subseteq B(x_{k-1})$.

Έστω $X_k(B) = (x_i : i = 1, \dots, |B|)$ και $Y_k(B) = (y_i : i = 1, \dots, |B|)$ δύο διανύσματα που περιέχουν, αντίστοιχα, τις x και y συνιστώσες θέσης όλων των εικονοστοιχείων που υποδεικνύει η μάσκα B . Αν $f_k^{f, \alpha} = (u^\alpha, v^\alpha)$, ορίζουμε, με το ίδιο σχήμα διανυσματοποίησης, τα διανύσματα $U_k^\alpha(B) = (u_i^\alpha : i = 1, \dots, |B|)$ και $V_k^\alpha(B) = (v_i^\alpha : i = 1, \dots, |B|)$ των x και y μετατοπίσεων, αντίστοιχα, όλων των σημείων της B σύμφωνα με το $f_k^{f, \alpha}$. Το γεγονός ότι χρησιμοποιούμε το ίδιο σχήμα διανυσματοποίησης για τις συνιστώσες θέσης και μετατόπισης, εξασφαλίζει ότι, για $i = 1, \dots, |B|$, τα x_i και u_i^α προήλθαν από την ίδια θέση της εικόνας/μάσκας και ότι το ίδιο ισχύει για τα y_i και v_i^α . Τα διανύσματα:

$$\begin{aligned} X_k^\alpha(B) &= X_k(B) + U_k^\alpha(B) = (x_i^\alpha = x_i + u_i^\alpha : i = 1, \dots, |B|) \\ Y_k^\alpha(B) &= Y_k(B) + V_k^\alpha(B) = (y_i^\alpha = y_i + v_i^\alpha : i = 1, \dots, |B|) \end{aligned} \quad (6.7)$$

περιέχουν τις συνιστώσες θέσης των εικονοστοιχείων που υποδεικνύει η B , μετατοπισμένων κατά το πεδίο $f_k^{f, \alpha}$. Με βάση τους προηγούμενους ορισμούς, σχηματίζουμε τα σύνολα των λόγων των

αποστάσεων, σε κάθε μία από τις x και y διευθύνσεις, όλων των ζευγών σημείων της B μετά την μετατόπισή τους κατά το $f_k^{f,\alpha}$ προς τις αρχικές τους αποστάσεις:

$$\begin{aligned} R_k^{x,\alpha}(B) &\equiv \left\{ \frac{|x_i^\alpha - x_j^\alpha|}{|x_i - x_j|} : i, j = 1, \dots, |B|, j < i \right\} \\ R_k^{y,\alpha}(B) &\equiv \left\{ \frac{|y_i^\alpha - y_j^\alpha|}{|y_i - y_j|} : i, j = 1, \dots, |B|, j < i \right\} \end{aligned} \quad (6.8)$$

Από τα $R_k^{x,\alpha}(B)$ και $R_k^{y,\alpha}(B)$ μπορούμε να αντλήσουμε στατιστικά συμπεράσματα για το πώς το πεδίο $f_k^{f,\alpha}$ μεταβάλλει τις αποστάσεις (στις διευθύνσεις των αξόνων) των σημείων της B . Για παράδειγμα, οι εκτιμήσεις $r_k^{x,\alpha}$ και $r_k^{y,\alpha}$ για τους συντελεστές κλιμάκωσης των x και y διαστάσεων του πλαισίου οριοθέτησης μπορούν να προκύψουν από τις μέσες τιμές ή τις διαμέσους (medians) των $R_k^{x,\alpha}(B)$ και $R_k^{y,\alpha}(B)$ αντίστοιχα, ενώ από τα ίδια σύνολα μπορούμε να εξάγουμε και εκτιμήσεις για τις μεταβλητότητες των συντελεστών αυτών. Καθώς έχουμε δύο εκδοχές προς τα εμπρός ροής ($\alpha \in \{1, 2\}$), έχουμε αντίστοιχα δύο εκτιμήσεις για την κλιμάκωση σε κάθε διάσταση. Οι εκτιμήσεις με βάση τα σύνολα 6.8 θα χρησιμεύσουν στον ορισμό του δυναμικού μοντέλου του συστήματος τη χρονική στιγμή k .

Κάθε ένα από τα σύνολα 6.8 έχει $\binom{|B|}{2} = \frac{|B|(|B|-1)}{2} = O(|B|^2)$ στοιχεία, όσα και τα ζεύγη σημείων της B . Καθώς ο παραπάνω υπολογισμός μπορεί να γίνει αρκετά απαιτητικός χρονικά για μεγάλα $|B|$, μπορούμε να τον πραγματοποιήσουμε για ένα υποσύνολο της B , που έχει προκύψει από χωρικά ομοιόμορφη δειγματοληψία αυτής.

6.3 Παρακολούθηση με Particle Filter

Στην ενότητα αυτή, παρουσιάζουμε τον κύριο αλγόριθμο οπτικής παρακολούθησης που αναπτύξαμε. Η μέθοδός μας υλοποιεί ένα SIR Particle Filter (ενότητα 5.7.7). Θυμίζουμε ότι, στην περίπτωση του SIR φίλτρου, ως πυκνότητα σημαντικότητας επιλέγεται η πρότερη πυκνότητα μετάβασης $p(x_k|x_{k-1})$, η οποία περιγράφει το δυναμικό μοντέλο του συστήματος. Στην υποενότητα 6.3.1, θα ορίσουμε την πυκνότητα σημαντικότητας που θα υιοθετήσουμε. Θεωρώντας ότι αυτή μοντελοποιεί ικανοποιητικά τη δυναμική συμπεριφορά του συστήματος, θα υποθέσουμε ότι ταυτίζεται με την $p(x_k|x_{k-1})$.

Αντίθετα από ό,τι στον Αλγόριθμο 9, όπου δίνεται μια γενική μορφή ενός SIR φίλτρου, η μέθοδός μας δεν κάνει επαναδειγματοληψία (resampling) στο τέλος κάθε επανάληψης: ακολουθήσαμε τη γενικότερη προσέγγιση του Αλγορίθμου 8, προσθέτοντας μια παράμετρο εισόδου N_T για το ελάχιστο ενεργό μέγεθος δείγματος. Καθώς δεν πραγματοποιείται πάντοτε επαναδειγματοληψία, η ενημέρωση των βαρών στάθμισης του φίλτρου γίνεται σύμφωνα με την κάπως γενικότερη εξίσωση 5.37 αντί της 5.42 που χρησιμοποιεί ο Αλγόριθμος 9 και η οποία υποθέτει πως όλα τα βάρη είναι ίσα.

Μια επισκόπηση της μεθόδου μας δίνεται στον Αλγόριθμο 10. Στην ενότητα 6.2, έχουν περιγραφεί τα μεγέθη εισόδου και τα ενδιάμεσα αποτελέσματα που χρησιμοποιούνται. Σημειώνουμε ότι, στην υλοποίησή μας, η αρχική κατάσταση x_1 (παράμετρος εισόδου) προσδιορίζεται απλώς μέσω του ορισμού ενός πλαισίου οριοθέτησης του αντικειμένου στο πρώτο πλαίσιο της ακολουθίας βίντεο (4 πρώτες διαστάσεις του διανύσματος κατάστασης), ενώ οι συνιστώσες ταχύτητας της x_1 τίθενται ίσες με 0.

Στο τέλος κάθε χρονικού βήματος, γίνεται η εκτίμηση της κατάστασης x_k με βάση το τρέχον

Algorithm 10 Particle Filter tracker

Input: Image sequence $\langle I_k : k = 1, \dots, N \rangle$, object (and background) color model(s), initial state x_1 , number of particles N_s , minimum effective sample size N_T , number of motion segmentation clusters N_c

Output: State estimate sequence $\langle x_k : k = 1, \dots, N \rangle$

```

1: /* Initialize particle set */
2: for  $i = 1 : N_s$  do
3:   Set  $x_1^i = x_1$ 
4:   Set  $w_1^i = \frac{1}{N_s}$ 
5: end for
6: for  $k = 2 : N$  do
7:   Compute object color masks  $B_{k-1}^c, B_k^c$ 
8:   Compute flow fields  $f_k^{f,1}, f_k^{f,2}, f_k^{b,1}, f_k^{b,2}$  as described in section 6.2
9:   Compute motion segmentations  $S_k^{f,1}, S_k^{f,2}, S_k^{b,1}, S_k^{b,2}$  as described in section 6.2
10:  /* Propagate particles and update particle weights */
11:  for  $i = 1 : N_s$  do
12:    Determine transition prior  $p(x_k|x_{k-1}^i)$  as in section 6.3.1 (equation 6.11), based on
       $B_{k-1}^c, f_k^{f,1}, f_k^{f,2}, S_k^{f,1}$  and  $S_k^{f,2}$ 
13:    Draw particle  $x_k^i \sim p(x_k|x_{k-1}^i)$ 
14:    Determine likelihood density  $p(y_k|x_k^i)$  as in section 6.3.2 (equation 6.20), based on
       $B_k^c, S_k^{b,1}$  and  $S_k^{b,2}$ 
15:    Assign the particle a weight  $w_k^i \propto w_{k-1}^i p(y_k|x_k^i)$ 
16:  end for
17:  Normalize weights so that  $\sum_{i=1}^{N_s} w_k^i = 1$ 
18:  /* Resample, if needed */
19:  Calculate effective sample size  $\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2}$ 
20:  if  $\hat{N}_{eff} < N_T$  then
21:    Resample using algorithm 7 and update particle set  $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$ 
22:  end if
23:  Estimate state  $x_k$  from particle set according to 6.9 or 6.10
24: end for

```

σύνολο σταθμισμένων δειγμάτων (particle set) $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$. Καθώς το τελευταίο αποτελεί, σύμφωνα με τις υποθέσεις που περιγράψαμε στο Κεφάλαιο 5, προσέγγιση της $p(x_k|y_{1:k})$, δύο συνήθεις τρόποι εκτίμησης της x_k από αυτό αναφέρονται ακολούθως.

- Εκτίμηση της x_k ως του σταθμισμένου μέσου των δειγμάτων (particles):

$$x_k = \sum_{i=1}^{N_s} w_k^i x_k^i \quad (6.9)$$

Η 6.9 ορίζει την εκτίμηση ελάχιστου μέσου τετραγωνικού σφάλματος (Minimum Mean Square Error – MMSE) με βάση το σύνολο των σταθμισμένων δειγμάτων, η οποία αντιστοιχεί στην προσδοκώμενη τιμή της τυχαίας μεταβλητής x_k υπό την $p(x_k|y_{1:k})$.

- Εκτίμηση της x_k ως του δείγματος (particle) με το μέγιστο βάρος:

$$x_k = x_k^m, \text{ όπου } m = \underset{i=1, \dots, N_s}{\operatorname{argmax}} w_k^i \quad (6.10)$$

Η 6.10 ορίζει την εκτίμηση μέγιστης ύστερης πιθανότητας (Maximum A Posteriori – MAP) με βάση το σύνολο των σταθμισμένων δειγμάτων, η οποία αντιστοιχεί στην πιο πιθανή τιμή της τυχαίας μεταβλητής x_k υπό τη θεωρούμενη προσέγγιση της $p(x_k|y_{1:k})$.

Η εκτίμηση της x_k με βάση κάποιο από τα παραπάνω κριτήρια αποτελεί απλώς μέρος της εξόδου του αλγορίθμου και δεν χρησιμοποιείται από τα επόμενα χρονικά βήματα.

Στις υποενότητες που ακολουθούν, ορίζουμε την πυκνότητα σημαντικότητας/δυναμικό μοντέλο $p(x_k|x_{k-1})$ και το μοντέλο παρατήρησης που περιγράφεται από τη συνάρτηση πιθανοφάνειας $p(y_k|x_k)$.

6.3.1 Ορισμός δυναμικού μοντέλου

Στην ενότητα αυτή θα ορίσουμε την πυκνότητα σημαντικότητας του φίλτρου που υλοποιεί ο Αλγόριθμος 10, η οποία, όπως αναφέραμε, υποθέτουμε ότι ταυτίζεται με το μοντέλο μετάβασης του συστήματος. Ο ορισμός του δυναμικού αυτού μοντέλου, $p(x_k|x_{k-1})$, θα γίνει με βάση τη χρωματική μάσκα B_{k-1}^c του πλαισίου I_{k-1} , τις εκδοχές της προς τα εμπρός οπτικής ροής που συνδέει τις I_{k-1} και I_k και τα παράγωγα αυτών, όπως τα ορίσαμε στην ενότητα 6.2. Σημειώνουμε ότι, κατά την εκτέλεση του k -οστού βήματος του Αλγορίθμου 10, υποστασιοποιούνται N_s (εν γένει) διαφορετικές κατανομές $p(x_k|x_{k-1}^i)$: μία για κάθε σωματίο $i = 1, \dots, N_s$. Η $p(x_k|x_{k-1}^i)$ θα είναι υπεύθυνη για τη χρονική διάδοση (propagation) του i -οστού δείγματος (particle).

Θυμίζουμε ότι οι δύο πρώτες συνιστώσες ενός διανύσματος κατάστασης κωδικοποιούν τη θέση του (κέντρου του) πλαισίου οριοθέτησης, η τρίτη και η τέταρτη συνιστώσα το μέγεθος (x και y διαστάσεις) αυτού, ενώ η δύο τελευταίες την ταχύτητά του. Δεδομένης της x_{k-1} , υποθέτουμε ότι οι συνιστώσες θέσης, μεγέθους και ταχύτητας της x_k είναι αμοιβαία ανεξάρτητες. Θα ορίσουμε την $p(x_k|x_{k-1})$ κατά τους παράγοντες που αυτή η υπόθεση ανεξαρτησίας επάγει:

$$p(x_k|x_{k-1}) = p(x_{k[1,2]}|x_{k-1})p(x_{k[3,4]}|x_{k-1})p(x_{k[5,6]}|x_{k-1}), \quad (6.11)$$

όπου έγινε χρήση του συμβολισμού που εισάγαμε στην ενότητα 6.2. Θα αναφερόμαστε στις επιμέρους πυκνότητες $p(x_{k[1,2]}|x_{k-1})$, $p(x_{k[3,4]}|x_{k-1})$ και $p(x_{k[5,6]}|x_{k-1})$ αντίστοιχα με τους όρους *πυκνότητα θέσης*, *πυκνότητα μεγέθους* και *πυκνότητα ταχύτητας*.

Θα μοντελοποιήσουμε κάθε μία από τις επιμέρους πυκνότητες σαν ένα μείγμα Γκαουσιανών κατανομών (ενότητα B.1) σε 2 διαστάσεις. Θυμίζουμε ότι η συνάρτηση πυκνότητας πιθανότητας ενός μείγματος c Γκαουσιανών κατανομών έχει τη μορφή:

$$p(x) = \sum_{i=1}^c \pi_i \mathcal{N}(x; \mu_i, \Sigma_i),$$

όπου με $\mathcal{N}(x; \mu_i, \Sigma_i)$ συμβολίζουμε μια Γκαουσιανή πυκνότητα με μέση τιμή μ_i και πίνακα συμμεταβλητότητας Σ_i και με π_i το συντελεστή μείξης της Γκαουσιανής συνιστώσας i . Οι συντελεστές μείξης είναι μη αρνητικοί και έχουν μοναδιαίο άθροισμα, για $i = 1, \dots, c$.

Σημειώνουμε ότι η υπόθεση ανεξαρτησίας 6.11 και η μοντελοποίηση των επιμέρους πυκνοτήτων ως μείγματα Γκαουσιανών καθιστά τη δειγματοληψία από την $p(x_k|x_{k-1})$ εύκολη: η λήψη ενός δείγματος της x_k γίνεται μέσω της λήψης ενός δείγματος, ανεξάρτητα, από κάθε μια από τις τρεις επιμέρους πυκνότητες. Η δειγματοληψία από ένα μείγμα Γκαουσιανών μπορεί να γίνει

απλά, πρώτα αντλώντας ένα δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[0, 1]$, το οποίο, με βάση τους συντελεστές μείξης, επιλέγει μια Γκαουσιανή συνιστώσα j , και στη συνέχεια αντλώντας ένα δείγμα από την j -οστή συνιστώσα $\mathcal{N}(x; \mu_j, \Sigma_j)$.

Πυκνότητα ταχύτητας. Θα μοντελοποιήσουμε την πυκνότητα ταχύτητας σαν ένα μείγμα από δύο Γκαουσιανές κατανομές, με ίσους συντελεστές μείξης, κάθε συνιστώσα του οποίου αντιστοιχεί στην κατάτμηση με βάση την κίνηση που προκύπτει από μία εκδοχή προς τα εμπρός οπτικής ροής. Συγκεκριμένα, θα χρησιμοποιήσουμε τα κεντροειδή (μέσες ταχύτητες των κλάσεων μιας κατάτμησης) των κύριων συνιστωσών κίνησης των κατατμήσεων $S_k^{f,1}$ και $S_k^{f,2}$ στη μάσκα $B(x_{k-1})$ της κατάστασης x_{k-1} ως μέσες τιμές για τις Γκαουσιανές συνιστώσες.

Χρησιμοποιώντας τον ορισμό 6.6, συμβολίζουμε το δείκτη της κύριας συνιστώσας κίνησης της κατάτμησης με βάση την κίνηση $S_k^{f,\alpha}$, για $\alpha = 1, 2$, στη μάσκα $B(x_{k-1})$ με:

$$m_k^{f,\alpha}(x_{k-1}) \equiv PMC \left(S_k^{f,\alpha}, B(x_{k-1}) \right) \quad (6.12)$$

Σύμφωνα με την αναπαράσταση 6.4, το κεντροειδές (μέση ταχύτητα) που αντιστοιχεί στην παραπάνω κύρια συνιστώσα κίνησης, για $\alpha = 1, 2$, είναι το $c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$.

Με βάση τα παραπάνω, η πυκνότητα ταχύτητας γράφεται στη μορφή:

$$p(x_{k[5,6]}|x_{k-1}) = 0.5\mathcal{N}(x_{k[5,6]}; \mu_k^{v,1}, \Sigma_k^{v,1}) + 0.5\mathcal{N}(x_{k[5,6]}; \mu_k^{v,2}, \Sigma_k^{v,2}), \quad (6.13)$$

όπου $\mu_k^{v,\alpha} = c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$, για $\alpha = 1, 2$. Οι 2×2 πίνακες συμμεταβλητότητας $\Sigma_k^{v,\alpha}$ προσδιορίζονται ως οι συμμεταβλητότητες δείγματος των (διδιάστατων) τιμών των διανυσμάτων κίνησης των πεδίων προς τα εμπρός ροής $f_k^{f,\alpha}$, για $\alpha = 1, 2$, που έχουν αφετηρία στη μάσκα $B(x_{k-1}) \cap B_{k-1}^c$, δηλαδή ανήκουν τόσο στη μάσκα κατάστασης όσο και στη χρωματική μάσκα του αντικειμένου. Περιορίζουμε το ενδιαφέρον εντός της χρωματικής μάσκας διότι η ορθογώνια μάσκα $B(x_{k-1})$ αναμένεται να περιέχει υπολογίσιμοι μεγέθους περιοχές που δεν ανήκουν καν στο αντικείμενο υπό παρακολούθηση και, συνεπώς, η πληροφορία οπτικής ροής που τις χαρακτηρίζει θα επηρεάσει αρνητικά την εκτίμηση των $\Sigma_k^{v,\alpha}$. Υπενθυμίζουμε ότι κάθε μία από τις κατατμήσεις $S_k^{f,\alpha}$ έχει γίνει ούτως ή άλλως με μάσκα αναφοράς την B_{k-1}^c (εξίσωση 6.5) και, επομένως, η κύρια συνιστώσα κίνησης 6.12 θα ανήκει εξορισμού σε αυτήν.

Πυκνότητα θέσης. Η πυκνότητα θέσης θα μοντελοποιηθεί σαν ένα μείγμα τριών Γκαουσιανών κατανομών. Η μία συνιστώσα ορίζεται με βάση την ταχύτητα της προηγούμενης κατάστασης, $x_{k-1[5,6]}$, μοντελοποιώντας μια συμπεριφορά σταθερής ταχύτητας, ενώ οι άλλες δύο, όπως και στον ορισμό της πυκνότητας ταχύτητας, αντιστοιχούν στις κύριες συνιστώσες κίνησης των κατατμήσεων $S_k^{f,1}$ και $S_k^{f,2}$ στη μάσκα $B(x_{k-1})$. Συγκεκριμένα, η μέση τιμή κάθε μιας από τις τρεις συνιστώσες ορίζεται ως το άθροισμα της θέσης στην προηγούμενη κατάσταση, $x_{k-1[1,2]}$, και της ταχύτητας στην προηγούμενη κατάσταση, $x_{k-1[5,6]}$, για την πρώτη συνιστώσα, ή του κεντροειδούς των κυρίων συνιστωσών κίνησης στην $B(x_{k-1})$ για τις άλλες δύο συνιστώσες. Όπως και πριν, συμβολίζουμε τον δείκτη της κύριας συνιστώσας κίνησης της κατάτμησης $S_k^{f,\alpha}$ στη μάσκα $B(x_{k-1})$ με $m_k^{f,\alpha}(x_{k-1})$ (εξίσωση 6.12), ενώ το αντίστοιχο κεντροειδές (μέση ταχύτητα της κύριας συνιστώσας κίνησης) είναι το $c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$, για $\alpha = 1, 2$.

Σύμφωνα με τα παραπάνω, η πυκνότητα θέσης γράφεται:

$$p(x_{k[1,2]}|x_{k-1}) = 0.4\mathcal{N}(x_{k[1,2]}; \mu_k^{p,0}, \Sigma_k^{p,0}) + 0.3\mathcal{N}(x_{k[1,2]}; \mu_k^{p,1}, \Sigma_k^{p,1}) + 0.3\mathcal{N}(x_{k[1,2]}; \mu_k^{p,2}, \Sigma_k^{p,2}), \quad (6.14)$$

όπου:

- $\mu_k^{p,0} = x_{k-1[1,2]} + x_{k-1[5,6]}$ και $\Sigma_k^{p,0} = \sigma_{p,0}^2 I_2$ (μοντέλο σταθερής ταχύτητας)
- $\mu_k^{p,\alpha} = x_{k-1[1,2]} + c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$ και $\Sigma_k^{p,\alpha} = \Sigma_k^{v,\alpha}$, για $\alpha = 1, 2$

Με I_2 συμβολίζουμε τον 2×2 μοναδιαίο πίνακα. Στην υλοποίησή μας, χρησιμοποιούμε $\sigma_{p,0} = 1$. Σημειώνουμε ότι οι πίνακες συμμεταβλητότητας $\Sigma_k^{p,\alpha}$ ($\alpha = 1, 2$) για τις συνιστώσες θέσης είναι ακριβώς ίδιοι με αυτούς των συνιστωσών ταχύτητας $\Sigma_k^{v,\alpha}$ για το ίδιο πεδίο ροής, των οποίων τον υπολογισμό περιγράψαμε στην προηγούμενη παράγραφο.

Πυκνότητα μεγέθους. Θα μοντελοποιήσουμε την πυκνότητα μεγέθους σαν ένα μείγμα τριών Γκαουσιανών κατανομών. Η μία συνιστώσα θα οριστεί με βάση το μέγεθος στην προηγούμενη κατάσταση, $x_{k-1[3,4]}$, και αντιστοιχεί στην υπόθεση σταθερότητας του μεγέθους. Οι άλλες δύο συνιστώσες αντιστοιχούν στις εκτιμήσεις αλλαγής μεγέθους με βάση τα πεδία $f_k^{f,1}$ και $f_k^{f,2}$ στη μάσκα $B(x_{k-1}) \cap B_{k-1}^c \equiv B^c(x_{k-1})$ (μέρος της χρωματικής μάσκας του αντικειμένου στο πλαίσιο $k-1$ που βρίσκεται μέσα στη μάσκα της προηγούμενης κατάστασης), σύμφωνα με τη μέθοδο που περιγράψαμε στην ενότητα 6.2. Συγκεκριμένα, από τα σύνολα λόγων αποστάσεων 6.8, μπορούμε να εξάγουμε εκτιμήσεις για τους συντελεστές κλιμάκωσης $r_k^{x,\alpha}$ και $r_k^{y,\alpha}$ στις x και y διαστάσεις του πλαισίου οριοθέτησης που ορίζει η κατάσταση (για $\alpha = 1, 2$). Οι εκτιμήσεις αυτές επιδρούν *πολλαπλασιαστικά* σε κάθε μια από τις διαστάσεις μεγέθους $x_{k-1[3]}$ και $x_{k-1[4]}$ του πλαισίου οριοθέτησης στην προηγούμενη κατάσταση. Έτσι, τα διανύσματα $[r_k^{x,\alpha} x_{k-1[3]} \quad r_k^{y,\alpha} x_{k-1[4]}]^T$, για $\alpha = 1, 2$, μπορούν να αποτελέσουν εκτιμήσεις για την $x_{k[3,4]}$.

Θεωρούμε τα σύνολα λόγων αποστάσεων 6.8, υπολογισμένων στην παραπάνω μάσκα $B^c(x_{k-1})$, και, για $\alpha = 1, 2$, συμβολίζουμε τις μέσες τιμές τους με:

$$r_k^{x,\alpha} \equiv \frac{1}{|B^c(x_{k-1})|} \sum_{r_i \in R_k^{x,\alpha}(B^c(x_{k-1}))} r_i \quad (6.15)$$

$$r_k^{y,\alpha} \equiv \frac{1}{|B^c(x_{k-1})|} \sum_{r_i \in R_k^{y,\alpha}(B^c(x_{k-1}))} r_i \quad (6.16)$$

και τις μεταβλητότητές τους με:

$$(s_k^{x,\alpha})^2 \equiv \frac{1}{|B^c(x_{k-1})|} \sum_{r_i \in R_k^{x,\alpha}(B^c(x_{k-1}))} (r_i - r_k^{x,\alpha})^2 \quad (6.17)$$

$$(s_k^{y,\alpha})^2 \equiv \frac{1}{|B^c(x_{k-1})|} \sum_{r_i \in R_k^{y,\alpha}(B^c(x_{k-1}))} (r_i - r_k^{y,\alpha})^2 \quad (6.18)$$

Σημειώνουμε ότι οι ποσότητες $r_k^{x,\alpha}$, $r_k^{y,\alpha}$, $(s_k^{x,\alpha})^2$ και $(s_k^{y,\alpha})^2$, για $\alpha = 1, 2$, είναι συναρτήσεις της κατάστασης x_{k-1} , γεγονός που δεν περιλαμβάνεται ρητά στο συμβολισμό μας για λόγους συνοπτικότητας αυτού.

Με βάση τα παραπάνω, η πυκνότητα μεγέθους γράφεται στη μορφή:

$$p(x_{k[3,4]} | x_{k-1}) = 0.3 \mathcal{N}(x_{k[3,4]}; \mu_k^{s,0}, \Sigma_k^{s,0}) + 0.35 \mathcal{N}(x_{k[3,4]}; \mu_k^{s,1}, \Sigma_k^{s,1}) + 0.35 \mathcal{N}(x_{k[3,4]}; \mu_k^{s,2}, \Sigma_k^{s,2}), \quad (6.19)$$

όπου ορίζουμε:

- $\mu_k^{s,0} = x_{k-1[3,4]}$ και $\Sigma_k^{s,0} = \sigma_{s,0}^2 I_2$ (υπόθεση σταθερού μεγέθους)
- $\mu_k^{s,\alpha} = [r_k^{x,\alpha} x_{k-1[3]} \quad r_k^{y,\alpha} x_{k-1[4]}]^T$ και $\Sigma_k^{s,\alpha} = \begin{bmatrix} (x_{k-1[3]})^2 (s_k^{x,\alpha})^2 & 0 \\ 0 & (x_{k-1[4]})^2 (s_k^{y,\alpha})^2 \end{bmatrix}$, για $\alpha = 1, 2$

Στην υλοποίησή μας, χρησιμοποιήσαμε την τιμή $\sigma_{s,0} = 1$.

6.3.2 Ορισμός μοντέλου παρατήρησης

Στην ενότητα αυτή θα ορίσουμε το μοντέλο παρατήρησης του φίλτρου που υλοποιεί ο Αλγόριθμος 10, το οποίο περιγράφεται από τη συνάρτηση πιθανοφάνειας $p(y_k|x_k)$ της κατάστασης x_k για την παρατήρηση y_k . Ο ορισμός θα βασιστεί στη χρωματική μάσκα B_k^c του πλαισίου I_k και τις κατατμήσεις με βάση την κίνηση $S_k^{b,1}$ και $S_k^{b,2}$ της I_k που βασιζονται στις εκδοχές της προς τα πίσω οπτικής ροής (από την I_k στην I_{k-1}), όπως αυτές ορίστηκαν στην ενότητα 6.2. Τα τρία αυτά μεγέθη θεωρούμε ότι αποτελούν την παρατήρηση y_k .

Υποθέτουμε την ακόλουθη μορφή για την $p(y_k|x_k)$:

$$p(y_k|x_k) = \pi_c p_c(y_k|x_k) + \pi_m p_m(y_k|x_k), \quad (6.20)$$

όπου $\pi_c, \pi_m \geq 0$ και $\pi_c + \pi_m = 1$. Ονομάζουμε τις $p_c(y_k|x_k)$ και $p_m(y_k|x_k)$, αντίστοιχα, πιθανοφάνεια χρώματος και πιθανοφάνεια κίνησης. Οι συντελεστές μείζης π_c και π_m ρυθμίζουν το βάρος που έχουν οι μετρήσεις (παρατηρήσεις) χρώματος και κίνησης, αντίστοιχα, στην τιμή της συνάρτησης πιθανοφάνειας. Η $p(y_k|x_k)$ χρησιμοποιείται μόνο για την ενημέρωση των βαρών στάθμισης, σύμφωνα με την εξίσωση 5.37:

$$w_k^i \propto w_{k-1}^i p(y_k|x_k^i)$$

Λόγω της μορφής της παραπάνω, αρκεί ο υπολογισμός της $p(y_k|x_k)$ μέχρι μια πολλαπλασιαστική σταθερά (up to proportionality). Καθώς οι συντελεστές μείζης μπορούν να οριστούν αυθαίρετα, ο υπολογισμός των επιμέρους συναρτήσεων $p_c(y_k|x_k)$ και $p_m(y_k|x_k)$ θα γίνει, χωρίς βλάβη της γενικότητας, μέχρι μια πολλαπλασιαστική σταθερά. Συγκεκριμένα, θα σχεδιάσουμε κανονικοποιημένες εκδοχές των $p_c(y_k|x_k)$ και $p_m(y_k|x_k)$ που έχουν πεδίο τιμών το διάστημα $(0, 1]$. Οι συντελεστές μείζης μπορούν να προσδιορισθούν εμπειρικά. Η υλοποίησή μας χρησιμοποιεί τις τιμές $\pi_c = \pi_m = 0.5$.

Πιθανοφάνεια χρώματος. Η πιθανοφάνεια με βάση το χρώμα ορίζεται με βάση το γινόμενο δύο παραγόντων, $g_{c,1}(x_k, y_k)$ και $g_{c,2}(x_k, y_k)$, οι οποίοι έχουν τιμές στο διάστημα $(0, 1]$ και κάθε ένας από τους οποίους εκφράζει την πιθανοφάνεια ενός διαφορετικού μέτρου κάλυψης της μάσκας κατάστασης $B(x_k)$ από τη χρωματική μάσκα του αντικειμένου B_k^c στο πλαίσιο I_k . Η $p_c(y_k|x_k)$, μέχρι μια πολλαπλασιαστική σταθερά, γράφεται:

$$p_c(y_k|x_k) \propto g_{c,1}(x_k, y_k) g_{c,2}(x_k, y_k) \quad (6.21)$$

Στο σημείο αυτό, εισάγουμε το συμβολισμό ∂B για το (εσωτερικό) σύνορο της μάσκας B . Το ∂B είναι μια δυαδική εικόνα που αποτελεί υποσύνολο της B και έχει την ιδιότητα ότι για κάθε στοιχείο της (θέση του διακριτού επιπέδου της εικόνας που αυτή υποδεικνύει) υπάρχει μια γειτονική θέση που δεν ανήκει στην B . Ο ορισμός της σχέσης γειννίας γίνεται με βάση κάποιο μοντέλο συνεκτικότητας (π.χ., 4-συνεκτικότητα). Ένας τυπικός ορισμός για το ∂B που εξυπηρετεί τις ανάγκες μας είναι ο:

$$\partial B \equiv \{i \in B \mid \exists j \in \mathcal{N}(i) : j \notin B\},$$

όπου με $\mathcal{N}(i)$ συμβολίζουμε το σύνολο των γειτονικών θέσεων του εικονοστοιχείου i , με βάση το μοντέλο 4-συνεκτικότητας. Συγκεκριμένα, μας ενδιαφέρει το σύνορο της μάσκας κατάστασης $\partial B(x_k)$, το οποίο θα χρησιμοποιηθεί για τους ορισμούς των $g_{c,1}(x_k, y_k)$ και $g_{c,2}(x_k, y_k)$ της 6.21. Σύμφωνα με τα παραπάνω, η δυαδική εικόνα $\partial B(x_k)$ απεικονίζει απλώς ένα ορθογώνιο, με πάχος περιγράμματος ίσο με ένα εικονοστοιχείο.

Ο παράγοντας $g_{c,1}(x_k, y_k)$ ορίζεται με βάση ένα μέτρο σταθμισμένης κάλυψης της $B(x_k)$ από τη χρωματική μάσκα B_k^c . Θεωρούμε την παρακάτω συνάρτηση στάθμισης, η οποία ορίζεται για τα εικονοστοιχεία $i \in B(x_k)$:

$$w_{x_k}(i) = 1 - e^{-\frac{1}{80}d^2(i, \partial(B(x_k)))}, \quad (6.22)$$

όπου με $d(i, \partial(B(x_k)))$ συμβολίζουμε την απόσταση του εικονοστοιχείου i από το σύνορο $\partial B(x_k)$. Η απόσταση ενός σημείου από ένα σύνολο (δυαδική εικόνα) ορίζεται απλώς ως η ελάχιστη απόσταση του σημείου από κάποιο στοιχείο του συνόλου:

$$d(i, B) \equiv \min_{j \in B} d(i, j),$$

όπου $d(i, j)$ είναι η συνήθης (Ευκλείδεια) απόσταση των θέσεων i και j . Ο υπολογισμός της $d(i, \partial(B(x_k)))$, για κάθε $i \in B(x_k)$, μπορεί να γίνει αποδοτικά μέσω κάποιου αλγορίθμου υπολογισμού του μετασχηματισμού απόστασης της $\partial B(x_k)$. Στο Σχήμα 6.1 απεικονίζεται η μορφή του μετασχηματισμού απόστασης και της συνάρτησης w_{x_k} , για κάποια αυθαίρετη κατάσταση x_k . Με βάση τη συνάρτηση στάθμισης 6.22, ορίζουμε το ακόλουθο μέτρο σταθμισμένης κάλυψης της $B(x_k)$ από τη χρωματική μάσκα B_k^c :

$$a_{c,1}(x_k, y_k) \equiv \frac{\sum_{i \in B(x_k) \cap B_k^c} w_{x_k}(i)}{\sum_{i \in B(x_k)} w_{x_k}(i)} \quad (6.23)$$

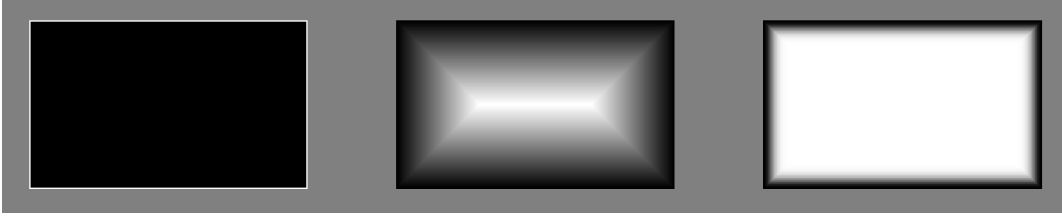
Η ποσότητα $a_{c,1}(x_k, y_k)$ λαμβάνει τιμές στο διάστημα $[0, 1]$ και, καθώς η w_{x_k} έχει χαμηλές τιμές κοντά στο σύνορο $\partial B(x_k)$, ενθαρρύνει (λαμβάνοντας μεγαλύτερη τιμή) καταστάσεις x_k που καλύπτουν τη χρωματική μάσκα B_k^c στο εσωτερικό τους. Με τον τρόπο αυτό, καταστάσεις με καλύτερο “κεντράρισμα” πάνω σε κάποιο τμήμα της χρωματικής μάσκας του αντικειμένου θα χαρακτηρίζονται από μεγαλύτερη τιμή για αυτό το κριτήριο σταθμισμένης κάλυψης. Καθώς οι καταστάσεις μας ορίζουν ορθογώνια πλαίσια οριοθέτησης, ενώ η χρωματική μάσκα του αντικειμένου μπορεί να έχει αυθαίρετη μορφή, η μέγιστη τιμή του μέτρου αυτού είναι, γενικά, μικρότερη του 1. Αν $\hat{a}_{c,1} \in (0, 1)$ είναι μια εκτίμηση της μέγιστης τιμής της $a_{c,1}(x_k, y_k)$, ο παράγοντας $g_{c,1}(x_k, y_k)$ ορίζεται τελικά ως:

$$g_{c,1}(x_k, y_k) = e^{-10(a_{c,1}(x_k, y_k) - \hat{a}_{c,1})^2} \quad (6.24)$$

Η εκτίμηση $\hat{a}_{c,1}$ του μέτρου της βέλτιστης σταθμισμένης κάλυψης εξαρτάται από τη μορφή του αντικειμένου υπό παρακολούθηση. Στην υλοποίησή μας, η $\hat{a}_{c,1}$ υποτίθεται σταθερή για κάθε χρονική στιγμή και αποτελεί παράμετρο προς ρύθμιση με προεπιλεγμένη την τιμή $\hat{a}_{c,1} = 0.9$.

Ο παράγοντας $g_{c,2}(x_k, y_k)$ της 6.21 θέτει ποινή στις καταστάσεις x_k για τις οποίες η μάσκα-σύνορο $\partial B(x_k)$ καλύπτεται σε μεγάλο βαθμό από τη χρωματική μάσκα B_k^c του αντικειμένου. Συγκεκριμένα, θεωρούμε το ποσοστό κάλυψης της $\partial B(x_k)$ από την B_k^c :

$$a_{c,2}(x_k, y_k) \equiv \frac{|\partial B(x_k) \cap B_k^c|}{|\partial B(x_k)|} \quad (6.25)$$



Σχήμα 6.1: Από αριστερά προς τα δεξιά: η μάσκα $\partial B(x_k)$ (περιορισμένη στην έκταση της $B(x_k)$), ο μετασχηματισμός (Ευκλείδειας) απόστασης της $\partial B(x_k)$ (τιμές της $d(i, \partial(B(x_k)))$) για κάθε $i \in B(x_k)$ και η συνάρτηση στάθμησης $w_{x_k}(i)$, για $i \in B(x_k)$. Για τις ανάγκες της απεικόνισης, οι τιμές των $d(i, \partial(B(x_k)))$ και $w_{x_k}(i)$ έχουν κανονικοποιηθεί στο διάστημα $[0, 1]$.

και ορίζουμε τη συνάρτηση $g_{c,2}(x_k, y_k)$ ως:

$$g_{c,2}(x_k, y_k) = e^{-5(a_{c,2}(x_k, y_k))^2} \quad (6.26)$$

Ιδανικά, εφόσον η χρωματική μάσκα B_k^c υποδεικνύει *μόνο* την προβολή του υπό παρακολούθηση αντικειμένου στο επίπεδο της εικόνας, θα υπήρχε κατάσταση x_k της οποίας το πλαίσιο οριοθέτησης θα περιέκλειε την B_k^c με κάλυψη συνόρου $a_{c,2}(x_k, y_k) = 0$. Μια μεγάλη τιμή (κοντά στη μονάδα) του $a_{c,2}(x_k, y_k)$ μπορεί να αποτελεί ένδειξη κακής τοποθέτησης του πλαισίου οριοθέτησης της κατάστασης x_k πάνω στη χρωματική μάσκα του αντικειμένου, τόσο λόγω της θέσης (κακό κεντράρισμα) όσο και λόγω των διαστάσεών του (κάλυψη μόνο μέρους του αντικειμένου λόγω μεγέθους). Η συμπερίληψη του $g_{c,2}(x_k, y_k)$ στη συνάρτηση πιθανοφάνειας με βάση το χρώμα “αποθαρρύνει” τις καταστάσεις με αυτά τα χαρακτηριστικά, ενώ, στην περίπτωση όπου η χρωματική μάσκα δεν περιορίζεται στην έκταση της προβολής του αντικειμένου, δεν έχει κάποια δραματική επίδραση.

Οι 6.24 και 6.26, μέσω της 6.21, ορίζουν την πιθανοφάνεια με βάση το χρώμα $p_c(y_k|x_k)$ μέχρι μια πολλαπλασιαστική σταθερά.

Πιθανοφάνεια κίνησης. Όπως και η πιθανοφάνεια χρώματος, η συνάρτηση πιθανοφάνειας με βάση την κίνηση ορίζεται με βάση το γινόμενο δύο παραγόντων, $g_{m,1}(x_k, y_k)$ και $g_{m,2}(x_k, y_k)$, οι οποίοι λαμβάνουν τιμές στο διάστημα $(0, 1]$ και ο ορισμός των οποίων στηρίζεται στις καταταμίσεις με βάση την κίνηση $S_k^{b,1}$ και $S_k^{b,2}$ από τα προς τα πίσω πεδία ροής. Η $p_m(y_k|x_k)$, μέχρι μια πολλαπλασιαστική σταθερά, γράφεται:

$$p_m(y_k|x_k) \propto g_{m,1}(x_k, y_k)g_{m,2}(x_k, y_k) \quad (6.27)$$

Χρησιμοποιώντας τον ορισμό 6.6, συμβολίζουμε το δείκτη της κύριας συνιστώσας κίνησης της κατάταξης με βάση την κίνηση $S_k^{b,\alpha}$, για $\alpha = 1, 2$, στη μάσκα $B(x_k)$ με:

$$m_k^{b,\alpha}(x_k) \equiv PMC \left(S_k^{b,\alpha}, B(x_k) \right) \quad (6.28)$$

Σύμφωνα με την αναπαράσταση 6.4, για $\alpha = 1, 2$, η δυαδική εικόνα που υποδεικνύει τα εικονοστοιχεία της I_k που ανήκουν στην παραπάνω κύρια συνιστώσα είναι η $B_k^{b,\alpha, m_k^{b,\alpha}(x_k)}$, ενώ το κεντροειδές (μέση ταχύτητα) που αντιστοιχεί σε αυτήν είναι το $c_k^{b,\alpha, m_k^{b,\alpha}(x_k)}$.

Ο παράγοντας $g_{m,1}(x_k, y_k)$ εκφράζει την πιθανοφάνεια της μέσης κάλυψης της μάσκας κατάστασης $B(x_k)$ από τις δυαδικές μάσκες $B_k^{b,\alpha, m_k^{b,\alpha}(x_k)}$, για $\alpha = 1, 2$, που αντιστοιχούν στις κύριες

συνιστώσες της προς τα πίσω κίνησης για την $B(x_k)$. Συγκεκριμένα, ο ορισμός του βασίζεται στη μέση κάλυψη:

$$a_{m,1}(x_k, y_k) = \frac{1}{2} \left(\frac{|B_k^{b,1,m_k^{b,1}(x_k)} \cap B(x_k)|}{|B(x_k)|} + \frac{|B_k^{b,2,m_k^{b,2}(x_k)} \cap B(x_k)|}{|B(x_k)|} \right) \quad (6.29)$$

Μεγαλύτερη τιμή της ποσότητας $a_{m,1}(x_k, y_k)$ υποδεικνύει ότι σημαντικό μέρος της $B(x_k)$ χαρακτηρίζεται από μία σχετικά ομοιόμορφη κίνηση, μια ιδιότητα που αναμένουμε ότι, γενικά, έχει η προβολή του αντικειμένου υπό παρακολούθηση. Όπως και στον ορισμό του πρώτου παράγοντα της πιθανοφάνειας με βάση το χρώμα, η μέγιστη τιμή του παραπάνω μέτρου κάλυψης αναμένεται να είναι μικρότερη του 1. Αν $\hat{a}_{m,1} \in (0, 1)$ είναι μια εκτίμηση της μέγιστης τιμής της $a_{m,1}(x_k, y_k)$, ο παράγοντας $g_{m,1}(x_k, y_k)$ ορίζεται τελικά ως:

$$g_{m,1}(x_k, y_k) = e^{-10(a_{m,1}(x_k, y_k) - \hat{a}_{m,1})^2} \quad (6.30)$$

Η εκτίμηση $\hat{a}_{m,1}$ του μέτρου της βέλτιστης μέσης κάλυψης της μάσκας κατάστασης από τις κύριες συνιστώσες κίνησης εξαρτάται από την εφαρμογή παρακολούθησης και μπορεί να γίνει εμπειρικά. Στην υλοποίησή μας, η $\hat{a}_{m,1}$ υποτίθεται σταθερή για κάθε χρονική στιγμή και αποτελεί παράμετρο προς ρύθμιση με προεπιλεγμένη την τιμή $\hat{a}_{m,1} = 0.85$.

Ο παράγοντας $g_{m,2}(x_k, y_k)$ της 6.27 εκφράζει τη μέση πιθανοφάνεια των διαφορών ανάμεσα στην ταχύτητα $x_{k[5,6]}$ στην κατάσταση x_k και τα αντίθετα των κεντροειδών (μέσων ταχυτήτων) των κύριων συνιστωσών της προς τα πίσω κίνησης στην $B(x_k)$. Κάθε ένα από τα $-c_k^{b,\alpha,m_k^{b,\alpha}(x_k)}$, για $\alpha = 1, 2$, εκφράζει μια επικρατούσα τιμή ταχύτητας στη μάσκα $B(x_k)$. Ο $g_{m,2}(x_k, y_k)$ ορίζεται ως:

$$g_{m,2}(x_k, y_k) = \frac{1}{2} \left(e^{-\frac{1}{7} \left(-c_k^{b,1,m_k^{b,1}(x_k)} - x_{k[5,6]} \right)^2} + e^{-\frac{1}{7} \left(-c_k^{b,2,m_k^{b,2}(x_k)} - x_{k[5,6]} \right)^2} \right) \quad (6.31)$$

και ενθαρρύνει τις καταστάσεις x_k για τις οποίες η ταχύτητα $x_{k[5,6]}$ είναι σε καλύτερη συμφωνία με τις (αντεστραμμένου πρόσημου) μέσες ταχύτητες των κύριων συνιστωσών της προς τα πίσω κίνησης στις αντίστοιχες μάσκες $B(x_k)$.

Οι 6.30 και 6.31, μέσω της 6.27, ορίζουν την πιθανοφάνεια με βάση την κίνηση $p_m(y_k|x_k)$ μέχρι μια πολλαπλασιαστική σταθερά.

6.4 Παρακολούθηση με Kalman Filter

Θυμίζουμε ότι η εφαρμογή του φίλτρου Kalman υποθέτει ένα γραμμικό δυναμικό σύστημα και Γκαουσιανούς θορύβους διαδικασίας και παρατήρησης. Η μορφή του δυναμικού μοντέλου και του μοντέλου παρατήρησης δίνεται από τις εξισώσεις 5.6 και 5.7, οι οποίες επαναλαμβάνονται εδώ:

$$x_k = A_k x_{k-1} + v_{k-1} \quad (6.32)$$

$$y_k = H_k x_k + w_k \quad (6.33)$$

Για το σύστημά μας, ο πίνακας A_k του δυναμικού μοντέλου και ο πίνακας H_k του μοντέλου παρατήρησης είναι διαστάσεων 6×6 και 4×6 αντίστοιχα (ο χώρος παρατηρήσεων θα οριστεί στην

ενότητα 6.4.2). Ο θόρυβος v_{k-1} της διαδικασίας και ο θόρυβος w_k της παρατήρησης ακολουθούν Γκαουσιανές κατανομές με μηδενική μέση τιμή και πίνακες συμμεταβλητότητας Q_{k-1} και R_k αντίστοιχα:

$$v_{k-1} \sim \mathcal{N}(v_{k-1}; 0, Q_{k-1}) \text{ και } w_k \sim \mathcal{N}(w_k; 0, R_k)$$

Σύμφωνα με την ενότητα 5.5, η ύστερη $p(x_k|y_{1:k})$, κάθε χρονική στιγμή k , είναι μια Γκαουσιανή πυκνότητα με μέση τιμή $m_{k|k}$ και πίνακα συμμεταβλητότητας $P_{k|k}$. Στόχος, σε κάθε βήμα, είναι η εκτίμηση των δύο αυτών παραμέτρων, χρησιμοποιώντας τις τιμές τους από το προηγούμενο βήμα. Χρησιμοποιώντας τον ίδιο συμβολισμό όπως στο Κεφάλαιο 5, θα ονομάσουμε $m_{k|k-1}$ και $P_{k|k-1}$ τις *προβλέψεις* για τις τιμές των παραμέτρων, πριν από τη διόρθωσή τους με βάση την παρατήρηση y_k . Η *τελική εκτίμηση* για την κατάσταση x_k δεν είναι άλλη από την $m_{k|k}$ στο τέλος του k -οστού βήματος, κατόπιν του σταδίου διόρθωσης. Μια επισκόπηση της μεθόδου παρακολούθησης με φίλτρο Kalman δίνεται στον Αλγόριθμο 11, ο οποίος βασίζεται ουσιαστικά στις εξισώσεις του Αλγορίθμου 4.

Algorithm 11 Kalman Filter tracker

Input: Image sequence $\langle I_k : k = 1, \dots, N \rangle$, object (and background) color model(s), initial state x_1 , number of motion segmentation clusters N_c

Output: State estimate sequence $\langle x_k : k = 1, \dots, N \rangle$

- 1: /* Initialize */
 - 2: $m_{1|1} = x_1$
 - 3: $P_1 = 0_{6,6}$
 - 4: **for** $k = 2 : N$ **do**
 - 5: Compute object color masks B_{k-1}^c, B_k^c
 - 6: Compute flow fields $f_k^{f,1}, f_k^{f,2}, f_k^{b,1}, f_k^{b,2}$ as described in section 6.2
 - 7: Compute motion segmentations $S_k^{f,1}, S_k^{f,2}$ as described in section 6.2
 - 8: Determine transition model parameters A_k and Q_{k-1} as described in section 6.4.1, based on $B_{k-1}^c, f_k^{f,1}, f_k^{f,2}$
 - 9: /* Prediction */
 - 10: $m_{k|k-1} = A_k m_{k-1|k-1}$
 - 11: $P_{k|k-1} = A_k P_{k-1|k-1} A_k^T + Q_{k-1}$
 - 12: /* Correction */
 - 13: Define y_k and determine observation model parameters H_k and R_k as described in section 6.4.2, based on $B_{k-1}^c, f_k^{f,1}, f_k^{f,2}, S_k^{f,1}, S_k^{f,2}, B_k^c, f_k^{b,1}, f_k^{b,2}$
 - 14: Compute Kalman gain: $K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}$
 - 15: $m_{k|k} = m_{k|k-1} + K_k (y_k - H_k m_{k|k-1})$
 - 16: $P_{k|k} = (I - K_k H_k) P_{k|k-1}$
 - 17: Estimate state as $x_k = m_{k|k}$
 - 18: **end for**
-

Στις ενότητες που ακολουθούν, περιγράφουμε τον τρόπο ορισμού του δυναμικού μοντέλου και του μοντέλου παρατήρησης για αυτή τη μέθοδο παρακολούθησης.

6.4.1 Ορισμός δυναμικού μοντέλου

Σύμφωνα με την εξίσωση 6.32, το μοντέλο μετάβασης του συστήματος ορίζεται πλήρως από τον πίνακα A_k και τον πίνακα συμμεταβλητότητας Q_{k-1} του θορύβου της διαδικασίας v_{k-1} . Με δεδομένη την εκτίμηση της κατάστασης x_{k-1} από το προηγούμενο βήμα, θα προσδιορίσουμε τα

μεγέθη αυτά με βάση τη χρωματική μάσκα του αντικειμένου B_{k-1}^c στο πλαίσιο I_{k-1} και τις δύο εκδοχές της προς τα εμπρός οπτικής ροής $f_k^{f,1}$ και $f_k^{f,2}$ από την εικόνα I_{k-1} στην I_k .

Για αυτή τη μέθοδο παρακολούθησης, θα θεωρήσουμε ότι το σύστημα ακολουθεί ένα μοντέλο σταθερής ταχύτητας. Σύμφωνα με αυτό το μοντέλο, με δεδομένη την κατάσταση $x_{k-1} \equiv m_{k-1|k-1}$, η ταχύτητα στην πρόβλεψη $m_{k|k-1}$ θα είναι ίση με την ταχύτητα στην $m_{k-1|k-1}$:

$$m_{k|k-1}[5,6] = m_{k-1|k-1}[5,6] \quad (6.34)$$

Η θέση στην πρόβλεψη $m_{k|k-1}$ θα προκύπτει απλά από την ενημέρωση της θέσης στην $m_{k-1|k-1}$ με βάση την ταχύτητα στην ίδια κατάσταση $m_{k-1|k-1}$. Λαμβάνοντας υπόψη την υπόθεσή μας για μοναδιαία περίοδο χρονικής δειγματοληψίας της ακολουθίας εικόνων ($\Delta t = 1$), η τιμή της ταχύτητας και της μετατόπισης που αντιστοιχεί σε αυτή για ένα χρονικό βήμα ταυτίζονται και η παραπάνω ενημέρωση λαμβάνει την απλή μορφή:

$$m_{k|k-1}[1,2] = m_{k-1|k-1}[1,2] + m_{k-1|k-1}[5,6] \quad (6.35)$$

Για τις τιμές των συνιστωσών μεγέθους του πλαισίου οριοθέτησης στην πρόβλεψη $m_{k|k-1}$, υπολογίζουμε εκτιμήσεις r_k^x και r_k^y για τους συντελεστές κλιμάκωσης για τις x και y διαστάσεις. Θυμίζουμε ότι, στον ορισμό του μοντέλου μετάβασης της μεθόδου παρακολούθησης με Particle Filter, έγινε χρήση δύο τέτοιων εκτιμήσεων για κάθε διάσταση, κάθε μία από τις οποίες προέκυψε από μια εκδοχή προς τα εμπρός οπτικής ροής. Με τον ίδιο ακριβώς τρόπο, υπολογίζουμε και εδώ τις εκτιμήσεις $r_k^{x,\alpha}$ και $r_k^{y,\alpha}$, για $\alpha = 1, 2$, ως τις μέσες τιμές των συνόλων των λόγων αποστάσεων 6.8 για τη μάσκα $B(x_{k-1}) \cap B_{k-1}^c \equiv B^c(x_{k-1})$ (όπου $x_{k-1} \equiv m_{k-1|k-1}$), σύμφωνα με τις εξισώσεις 6.15 και 6.16. Με βάση αυτές, ορίζουμε τις εκτιμήσεις των συντελεστών κλιμάκωσης r_k^x και r_k^y ως τις μέσες τιμές των δύο εκδοχών τους σε κάθε διάσταση:

$$\begin{aligned} r_k^x &= \frac{1}{2} (r_k^{x,1} + r_k^{x,2}) \\ r_k^y &= \frac{1}{2} (r_k^{y,1} + r_k^{y,2}) \end{aligned} \quad (6.36)$$

Σύμφωνα με τα παραπάνω, οι συνιστώσες μεγέθους στην πρόβλεψη $m_{k|k-1}$ δίνονται από την εξίσωση:

$$m_{k|k-1}[3,4] = \begin{bmatrix} r_k^x & 0 \\ 0 & r_k^y \end{bmatrix} m_{k-1|k-1}[3,4] \quad (6.37)$$

Οι εξισώσεις 6.34, 6.35 και 6.37 ορίζουν τον πίνακα A_k . Η εξίσωση πρόβλεψης γράφεται στην ενιαία μητρική μορφή:

$$m_{k|k-1} = A_k m_{k-1|k-1},$$

όπου:

$$A_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & r_k^x & 0 & 0 & 0 \\ 0 & 0 & 0 & r_k^y & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.38)$$

Όπως και όταν ορίσαμε το μοντέλο μετάβασης της μεθόδου παρακολούθησης με Particle Filter, θα υποθέσουμε ότι οι συνιστώσες θέσης, μεγέθους και ταχύτητας είναι αμοιβαία ανεξάρτητες. Στην περίπτωση αυτή, ο πίνακας συμμεταβλητότητας Q_{k-1} του θορύβου της διαδικασίας

θα είναι της μπλοκ μορφής:

$$Q_{k-1} = \begin{bmatrix} Q_{k-1}^p & 0_{2,2} & 0_{2,2} \\ 0_{2,2} & Q_{k-1}^s & 0_{2,2} \\ 0_{2,2} & 0_{2,2} & Q_{k-1}^v \end{bmatrix}, \quad (6.39)$$

όπου με $0_{m,n}$ συμβολίζουμε τον μηδενικό $m \times n$ πίνακα και τα Q_{k-1}^p , Q_{k-1}^s και Q_{k-1}^v είναι, αντίστοιχα, οι 2×2 πίνακες συμμεταβλητότητας των συνιστωσών θέσης, μεγέθους και ταχύτητας της κατάστασης. Για τους πίνακες συμμεταβλητότητας των συνιστωσών θέσης και ταχύτητας Q_{k-1}^p και Q_{k-1}^v , υποθέτουμε τις μορφές:

$$Q_{k-1}^p = \sigma_{p,0}^2 I_2 \text{ και } Q_{k-1}^v = \sigma_{v,0}^2 I_2 \quad (6.40)$$

Στην περίπτωση μας, είναι $Q_{k-1}^p = Q_{k-1}^v$ και, κατά συνέπεια, $\sigma_{p,0} = \sigma_{v,0}$. Στην υλοποίησή μας, χρησιμοποιήσαμε την τιμή $\sigma_{p,0} = \sigma_{v,0} = 2$. Για τον πίνακα συμμεταβλητότητας των συνιστωσών μεγέθους του πλαισίου, υπολογίζουμε τις μεταβλητότητες $(s_k^{x,\alpha})^2$ και $(s_k^{y,\alpha})^2$, για $\alpha = 1, 2$, των συνόλων των λόγων αποστάσεων 6.8 για τη μάσκα $B(x_{k-1}) \cap B_{k-1}^c \equiv B^c(x_{k-1})$ (όπου $x_{k-1} \equiv m_{k-1|k-1}$) σε κάθε μια από τις x και y διαστάσεις και για κάθε εκδοχή οπτικής ροής, σύμφωνα με τις εξισώσεις 6.17 και 6.18. Με βάση αυτές, ορίζουμε τις εκτιμήσεις μας για τις μεταβλητότητες των συντελεστών κλιμάκωσης $(s_k^x)^2$ και $(s_k^y)^2$ ως τις μέσες τιμές των δύο εκδοχών τους σε κάθε διάσταση:

$$\begin{aligned} (s_k^x)^2 &= \frac{1}{2} \left((s_k^{x,1})^2 + (s_k^{x,2})^2 \right) \\ (s_k^y)^2 &= \frac{1}{2} \left((s_k^{y,1})^2 + (s_k^{y,2})^2 \right) \end{aligned} \quad (6.41)$$

Ο πίνακας συμμεταβλητότητας των συνιστωσών μεγέθους λαμβάνει τελικά τη μορφή:

$$Q_{k-1}^s = \begin{bmatrix} (m_{k-1|k-1[3]})^2 (s_k^x)^2 & 0 \\ 0 & (m_{k-1|k-1[4]})^2 (s_k^y)^2 \end{bmatrix} \quad (6.42)$$

Οι εξισώσεις 6.40 και 6.42 ορίζουν πλήρως τον πίνακα συμμεταβλητότητας 6.39 και, κατά συνέπεια, το μηδενικής μέσης τιμής, Γκαουσιανό θόρυβο της διαδικασίας.

6.4.2 Ορισμός μοντέλου παρατήρησης

Σε αντίθεση με την βασική μας μέθοδο παρακολούθησης με Particle Filter, όπου ο ορισμός του χώρου παρατηρήσεων έγινε έμμεσα μέσω της συνάρτησης $p(y_k|x_k)$, στην περίπτωση του Kalman Filter απαιτείται ο ακριβής (ρητός) ορισμός του διανύσματος παρατήρησης. Θα θεωρήσουμε ένα χώρο παρατήρησης 4 διαστάσεων, παρόμοιο με το χώρο κατάστασης, αλλά χωρίς την πληροφορία για τις διαστάσεις ενός πλαισίου οριοθέτησης. Ένα διάνυσμα παρατήρησης περιέχει μόνο τη θέση του κέντρου ενός πλαισίου οριοθέτησης του υπό παρακολούθηση αντικειμένου και την ταχύτητά του και είναι της μορφής:

$$(x, y, x', y') \in \mathbb{R}^4$$

όπου με $p = (x, y)$ συμβολίζουμε, για τις ανάγκες του παραπάνω ορισμού και μόνο, τη θέση του κέντρου του πλαισίου οριοθέτησης του αντικειμένου στο επίπεδο της εικόνας και με $v = (x', y')$ την ταχύτητά του. Σημειώνουμε ότι, καθώς το διάνυσμα παρατήρησης δεν περιέχει πληροφορία για το μέγεθος του πλαισίου οριοθέτησης του αντικειμένου, το στάδιο της διόρθωσης του φίλτρου δεν παρέχει κάποια αναπληροφόρηση (feedback) για τις συνιστώσες μεγέθους της υπό εκτίμηση κατάστασης.

Με βάση τα παραπάνω, ο πίνακας H_k που απεικονίζει μια κατάσταση σε μια παρατήρηση ορίζεται απλά ως:

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.43)$$

Σύμφωνα με την εξίσωση 6.33, για τον πλήρη ορισμό του μοντέλου παρατήρησης, απαιτείται επιπλέον ο ορισμός του τρόπου εξαγωγής της παρατήρησης y_k στο k -οστό βήμα και του πίνακα συμμεταβλητότητας του θορύβου παρατήρησης R_k .

Η παρατήρηση y_k , όπως θα την ορίσουμε αμέσως, προκύπτει ουσιαστικά με βάση μια μετατόπιση της κατάστασης $x_{k-1} \equiv m_{k-1|k-1}$ σύμφωνα με τα πεδία της προς τα εμπρός κίνησης. Χρησιμοποιώντας τον ορισμό 6.6, συμβολίζουμε το δείκτη της κύριας συνιστώσας κίνησης της κατάτμησης με βάση την κίνηση $S_k^{f,\alpha}$, για $\alpha = 1, 2$, στη μάσκα $B(x_{k-1})$ όπως κάναμε προηγουμένως, στον ορισμό του δυναμικού μοντέλου της μεθόδου παρακολούθησης με Particle Filter (εξίσωση 6.12):

$$m_k^{f,\alpha}(x_{k-1}) \equiv PMC \left(S_k^{f,\alpha}, B(x_{k-1}) \right) \quad (6.44)$$

Σύμφωνα με την αναπαράσταση 6.4, το κεντροειδές (μέση ταχύτητα) και η δυαδική εικόνα που αντιστοιχούν στην παραπάνω κύρια συνιστώσα κίνησης, για $\alpha = 1, 2$, είναι αντίστοιχα τα $c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$ και $B_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$. Ορίζουμε τα βάρη στάθμισης για κάθε μια από τις δύο εκδοχές της κύριας συνιστώσας κίνησης ως:

$$w_k^\alpha(x_{k-1}) \equiv \frac{\left| B_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})} \cap B(x_{k-1}) \right|}{\left| B_k^{f,1, m_k^{f,1}(x_{k-1})} \cap B(x_{k-1}) \right| + \left| B_k^{f,2, m_k^{f,2}(x_{k-1})} \cap B(x_{k-1}) \right|}, \text{ για } \alpha = 1, 2 \quad (6.45)$$

Προφανώς, είναι $w_k^1(x_{k-1}) + w_k^2(x_{k-1}) = 1$ και, για $\alpha = 1, 2$, το βάρος στάθμισης $w_k^\alpha(x_{k-1})$ είναι ανάλογο της κάλυψης της μάσκας $B(x_{k-1})$ από την κύρια συνιστώσα κίνησης της κατάτμησης $S_k^{f,\alpha}$ στη μάσκα κατάστασης $B(x_{k-1})$. Θυμίζουμε ότι κάθε μία από τις κατατμήσεις $S_k^{f,\alpha}$ έχει γίνει ούτως ή άλλως με μάσκα αναφοράς την B_{k-1}^c (εξίσωση 6.5) και, επομένως, η κύρια συνιστώσα κίνησης 6.44 θα ανήκει εξορισμού σε αυτήν. Συνεπώς, τα παραπάνω βάρη στάθμισης είναι επίσης ανάλογα της κάλυψης της μάσκας $B(x_{k-1}) \cap B_{k-1}^c$ από την αντίστοιχη εκδοχή κύριας συνιστώσας κίνησης στην $B(x_{k-1})$. Ορίζουμε το κύριο διάνυσμα κίνησης στην $B(x_{k-1})$ ως το σταθμισμένο μέσο των κεντροειδών $c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})}$ από τα βάρη της 6.45, για $\alpha = 1, 2$:

$$c_k^f = w_k^1(x_{k-1})c_k^{f,1, m_k^{f,1}(x_{k-1})} + w_k^2(x_{k-1})c_k^{f,2, m_k^{f,2}(x_{k-1})} \quad (6.46)$$

Για λόγους συνοπτικότητας, παραλείπουμε από το συμβολισμό μας τη ρητή έκφραση της συναρτησιακής εξάρτησης της ποσότητας c_k^f από την κατάσταση $x_{k-1} \equiv m_{k-1|k-1}$. Η ταχύτητα/μετατόπιση c_k^f , όπως θα δούμε, αποτελεί μια υποψήφια τιμή για την ταχύτητα που αντιστοιχεί στην παρατήρηση y_k .

Ο προσδιορισμός της y_k θα γίνει με βάση μια βοηθητική κατάσταση \hat{x}_k , την οποία θα ορίσουμε αμέσως και η οποία αντιστοιχεί σε μία πιο πληροφορημένη πρόβλεψη για την x_k . Θεωρούμε την κατάσταση \hat{x}_k^0 , στην οποία η θέση προκύπτει από τη μετατόπιση της θέσης της x_{k-1} κατά το διάνυσμα c_k^f και οι διαστάσεις μεγέθους από αυτές της x_{k-1} , κλιμακωμένες κατά r_k^x και r_k^y στις

διαστάσεις των αξόνων, όπου οι συντελεστές κλιμάκωσης έχουν υπολογιστεί σύμφωνα με τις εξισώσεις 6.36 κατά τον προσδιορισμό του μοντέλου μετάβασης:

$$\hat{x}_k^0 = \left(x_{k-1[1]} + c_{k[1]}^f, x_{k-1[2]} + c_{k[2]}^f, r_k^x x_{k-1[3]}, r_k^y x_{k-1[4]}, c_{k[1]}^f, c_{k[2]}^f \right) \quad (6.47)$$

Σημειώνουμε ότι η ταχύτητα στην \hat{x}_k^0 , που έχει τεθεί ίση με c_k^f , δεν θα παίξει κάποιο ρόλο στον παρακάτω υπολογισμό. Η \hat{x}_k^0 δεν αποτελεί παρά μια εύλογη πρόβλεψη για την κατάσταση τη χρονική στιγμή k , η οποία βασίζεται στην $x_{k-1} \equiv m_{k-1|k-1}$, παράγωγα μεγέθη των πεδίων προς τα εμπρός κίνησης και, έμμεσα, μέσω των κατατιμήσεων με βάση την προς τα εμπρός κίνηση, τη χρωματική μάσκα B_{k-1}^c . Ένας απλός τρόπος αξιολόγησης της \hat{x}_k^0 δίνεται από τον έλεγχο του αν η χρωματική μάσκα του αντικειμένου B_k^c στο πλαίσιο I_k καλύπτει ένα ελάχιστο ποσοστό $\tau^c \in [0, 1]$ της μάσκας κατάστασης $B(\hat{x}_k^0)$:

$$\frac{|B(\hat{x}_k^0) \cap B_k^c|}{|B(\hat{x}_k^0)|} \stackrel{?}{\geq} \tau^c \quad (6.48)$$

Στην υλοποίησή μας, το ποσοστό-κατώφλι έχει τεθεί εμπειρικά στην τιμή $\tau^c = 0.6$ και υποτίθεται σταθερό. Η βοηθητική κατάσταση \hat{x}_k γίνεται με αφετηρία την \hat{x}_k^0 και τον έλεγχο της συνθήκης 6.48. Συγκεκριμένα, αν η παραπάνω είναι αληθής, οπότε η μάσκα της \hat{x}_k^0 καλύπτεται “επαρκώς” από τη χρωματική μάσκα B_k^c , η βοηθητική κατάσταση ορίζεται ως $\hat{x}_k \equiv \hat{x}_k^0$. Σε διαφορετική περίπτωση, η \hat{x}_k προκύπτει από μια τοπική αναζήτηση γύρω από την \hat{x}_k^0 , κατά την οποία εξετάζεται ένα σύνολο “μετατοπισμένων” εκδοχών της \hat{x}_k^0 , δηλαδή καταστάσεων που ορίζουν ένα πλαίσιο οριοθέτησης που προκύπτει από απλή μετατόπιση αυτού της \hat{x}_k^0 , και επιλέγεται αυτή που εμφανίζει τη μεγαλύτερη κάλυψη από τη χρωματική μάσκα B_k^c (αριστερό μέλος της 6.48). Προϊόν της αναζήτησης είναι το διάνυσμα μετατόπισης (d_x, d_y) που συνδέει την \hat{x}_k^0 με τη βέλτιστης κάλυψης, υπό την παραπάνω έννοια, βοηθητική κατάσταση \hat{x}_k :

$$(d_x, d_y) = \underset{(d_x, d_y) \in [-5, 5]^2}{\operatorname{argmax}} \frac{|B(\hat{x}_k^0 + (d_x', d_y', 0, 0, 0, 0)) \cap B_k^c|}{|B(\hat{x}_k^0 + (d_x', d_y', 0, 0, 0, 0))|} \quad (6.49)$$

Παρότι η αναζήτηση που περιγράψαμε έχει διατυπωθεί σαν ένα συνεχές πρόβλημα βελτιστοποίησης στην 6.49, στην υλοποίησή μας στρογγυλοποιούμε τις συντεταγμένες θέσης στον εγγύτερο ακέραιο (ακρίβεια εικονοστοιχείου) και πραγματοποιούμε την αναζήτηση σε ένα 11×11 παράθυρο με κέντρο την (στρογγυλοποιημένη) θέση στην \hat{x}_k^0 . Συνοψίζοντας, η βοηθητική κατάσταση \hat{x}_k , που θα χρησιμοποιήσουμε για τον ορισμό της παρατήρησης y_k , δίνεται από την:

$$\hat{x}_k = \begin{cases} \hat{x}_k^0, & \text{αν } \frac{|B(\hat{x}_k^0) \cap B_k^c|}{|B(\hat{x}_k^0)|} \geq \tau^c \text{ (εξίσωση 6.47)} \\ \hat{x}_k^0 + (d_x, d_y, 0, 0, 0, 0), & \text{διαφορετικά (εξισώσεις 6.47 και 6.49)} \end{cases} \quad (6.50)$$

Τέλος, ορίζουμε την παρατήρηση y_k στο k -οστό βήμα με βάση την \hat{x}_k και την προηγούμενη κατάσταση $x_{k-1} \equiv m_{k-1|k-1}$. Θεωρούμε ότι στην y_k η θέση είναι ίδια με αυτήν στη βοηθητική κατάσταση \hat{x}_k και θέτουμε την ταχύτητα/μετατόπιση ίση με τη μεταβολή της θέσης από την x_{k-1} στην \hat{x}_k :

$$y_k \equiv (\hat{x}_{k[1]}, \hat{x}_{k[2]}, \hat{x}_{k[1]} - x_{k-1[1]}, \hat{x}_{k[2]} - x_{k-1[2]}) \quad (6.51)$$

Οι εξισώσεις 6.51 και 6.50 ορίζουν το μέγεθος που θεωρούμε ως παρατήρηση στο k -οστό βήμα και περιγράφουν τον τρόπο υπολογισμού του.

Για τον πλήρη ορισμό του μοντέλου παρατήρησης του Kalman Filter (εξίσωση 6.33), μένει ο προσδιορισμός του πίνακα συμμεταβλητότητας R_k του θορύβου παρατήρησης. Υποθέτοντας,

όπως προηγουμένως, ανεξαρτησία των συνιστωσών θέσης και ταχύτητας, ο R_k γράφεται στην μπλοκ μορφή:

$$R_k = \begin{bmatrix} R_k^p & 0_{2,2} \\ 0_{2,2} & R_k^v \end{bmatrix}, \quad (6.52)$$

όπου R_k^p και R_k^v είναι, αντίστοιχα, οι 2×2 πίνακες συµµεταβλητότητας των συνιστωσών θέσης και ταχύτητας της παρατήρησης. Έστω $\Sigma_k^{v,\alpha}$, για $\alpha = 1, 2$, οι συµµεταβλητότητες δείγµατος των (διδιάστατων) τιµών των διανυσµάτων κίνησης των πεδίων προς τα πίσω ροής $f_k^{b,\alpha}$ που έχουν αφετηρία στη µάσκα $B(\hat{x}_k) \cap B_k^c$, δηλαδή ανήκουν τόσο στη µάσκα της βοηθητικής κατάστασης \hat{x}_k µε βάση την οποία ορίζεται η παρατήρηση y_k όσο και στη χρωµατική µάσκα του αντικειµένου στο πλαίσιο I_k . Ορίζουµε τον πίνακα συµµεταβλητότητας των συνιστωσών ταχύτητας της y_k ως τη μέση τιμή των δύο αυτών εκτιμήσεων:

$$R_k^v \equiv \frac{1}{2} \left(\Sigma_k^{v,1} + \Sigma_k^{v,2} \right) \quad (6.53)$$

Στην περίπτωση µας είναι $R_k^p = R_k^v$ και, συνεπώς, η 6.53 ορίζει πλήρως τον πίνακα συµµεταβλητότητας 6.52.

6.5 Παρακολούθηση χωρίς αβεβαιότητα

Οι μέθοδοι οπτικής παρακολούθησης που περιγράψαµε στις προηγούμενες ενότητες είναι αρκετά γενικές, καθώς προϋποθέτουν γνώση μόνο ενός χρωµατικού μοντέλου για το υπό παρακολούθηση αντικείμενο και βασίζονται μόνο στην υπόθεση ότι η κίνηση του τελευταίου είναι κατά κύριο λόγο μεταφορική. Οι μέθοδοί µας χρησιµοποιούν χαρακτηριστικά χρώµατος (χρωµατικές µάσκες που προκύπτουν από κατάτµηση του αντικειµένου µε βάση το χρώµα) και κίνησης (εκδοχές οπτικής ροής και παράγωγες κατατμήσεις µε βάση την κίνηση) µε ένα γενικό τρόπο, χωρίς να υποθέτουν *εκ των προτέρων* την ισχύ κάποιας ειδικής ιδιότητας αυτών. Στην ενότητα αυτή, θα περιγράψουµε µια προσθήκη/τροποποίηση των μεθόδων µας ώστε αυτές να ανιχνεύουν και να εκµεταλλεύονται αυτόµατα (on the fly) ειδικές µορφές της χρωµατικής µάσκας του αντικειµένου, για τις οποίες το πρόβληµα της παρακολούθησης προσεγγίζεται σηµαντικά πιο εύκολα και µε µεγαλύτερη (δυνάµει πλήρη) ακρίβεια.

Σε αρκετές εφαρµογές, η κατανοµή των χρωµατικών τιµών του αντικειµένου υπό παρακολούθηση είναι πλήρως διαχωρισµένη από αυτές των αντικειµένων του παρασκήνιου. Σε αυτή την ακραία περίπτωση, η προβολή του αντικειµένου στο επίπεδο της εικόνας ανά πάσα στιγµή υποδεικνύεται ακριβώς από τη χρωµατική του µάσκα και η οπτική παρακολούθηση ανάγεται ουσιαστικά στην κατάτµηση του αντικειµένου από τα πλαίσια της ακολουθίας βίντεο µε βάση το χρώµα. Σε άλλα σενάρια, αυτό µπορεί να μην ισχύει ποτέ. Σαν παράδειγµα, µπορούµε να θεωρήσουµε µια εφαρµογή παρακολούθησης του δεξιού χεριού (παλάµης) σε ένα βίντεο νοηµατικής γλώσσας, όπου ο νοηµατιστής φοράει κοντά µανίκια. Η χρωµατική µάσκα του χεριού θα περιλαµβάνει αναπόφευκτα τον πήχυ και το βραχίονα του χεριού, ολόκληρο το αριστερό χέρι και το πρόσωπο του νοηµατιστή. Το αντικείμενο υπό παρακολούθηση (δεξιά παλάμη) θα υποδεικνύεται, ανά πάσα στιγµή, από μέρος της χρωµατικής του µάσκας, το οποίο, µάλιστα, είναι συνδεδεµένο µε το υπόλοιπο αυτής. Φυσικά, υπάρχουν και “ενδιάµεσης δυσκολίας” εφαρµογές, όπου η χρωµατική µάσκα περιλαµβάνει µεν και τµήµατα αντικειµένων του παρασκήνιου, αλλά το µέρος της που αντιστοιχεί στο αντικείμενο υπό παρακολούθηση αποτελεί, τουλάχιστον σε κάποια χρονικά διαστήµατα του βίντεο, µια *αυτόνοµη συνεκτική συµιστώσα* αυτής. Σαν παράδειγµα µιας τέτοιας ενδιάµεσης περίπτωσης, µπορούµε να θεωρήσουµε και πάλι ένα σενάριο παρακολούθησης του δεξιού χεριού σε

ένα βίντεο νοηματικής γλώσσας, όπου, όμως, ο νοηματιστής φοράει μακριά μανίγια. Η χρωματική μάσκα, τότε, θα υποδεικνύει τις προβολές των δύο χεριών (παλάμες) και του προσώπου. Στα διαστήματα του βίντεο όπου τα αντικείμενα αυτά δεν επικαλύπτονται/αποκρύπτονται το ένα από το άλλο (occlusion), η χρωματική μάσκα αναμένεται να αποτελείται από τρεις συνεκτικές συνιστώσες, με μία από αυτές να αντιστοιχεί ακριβώς στη δεξιά παλάμη που παρακολουθούμε. Στα διαστήματα επικάλυψης/απόκρυψης, δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα αναλύοντας μόνο τη χρωματική μάσκα.

Η τροποποίηση των μεθόδων παρακολούθησης που θα περιγράψουμε έγκειται στην ενσωμάτωση του ελέγχου αν το πλαίσιο οριοθέτησης μιας βοηθητικής κατάστασης-πρόβλεψης, που προκύπτει από μετασχηματισμό της (γνωστής) κατάστασης στο προηγούμενο βήμα με βάση τις κύριες συνιστώσες κίνησης και τους συντελεστές κλιμάκωσης που υπολογίσαμε παραπάνω, καλύπτει σημαντικό μέρος μιας συνεκτικής συνιστώσας της χρωματικής μάσκας του τρέχοντος βήματος. Στην περίπτωση που η συνθήκη αυτή είναι αληθής, μπορούμε να θεωρήσουμε με ασφάλεια ότι το πλαίσιο οριοθέτησης της τρέχουσας κατάστασης είναι απλώς το (ελάχιστου μεγέθους) πλαίσιο οριοθέτησης που περικλείει τη συνεκτική αυτή συνιστώσα της χρωματικής μάσκας και η εκτίμηση στο τρέχον βήμα ορίζεται αναλόγως, όντας, ουσιαστικά, ντετερμινιστική. Στη συνέχεια, διατυπώνουμε τυπικά τον έλεγχο εφαρμοσιμότητας αυτής της τεχνικής στο k -οστό βήμα, χρησιμοποιώντας πάλι τα ενδιάμεσα μεγέθη που ορίσαμε προηγουμένως, και περιγράφουμε ακριβώς τις προσθήκες που απαιτούνται στις δύο μεθόδους μας. Όπως θα αποσαφηνιστεί αμέσως στα επόμενα, η ενσωμάτωση των επιπλέον ελέγχων δεν παραβιάζει τα θεωρητικά πλαίσια των μεθόδων. Για παράδειγμα, η μέθοδος παρακολούθησης με Particle Filter, κατόπιν της τροποποίησης, εξακολουθεί να υλοποιεί ένα Particle Filter με λίγο διαφορετική πυκνότητα σημαντικότητας (μοντέλο μετάβασης, στην SIR περίπτωση μας).

Έλεγχος εφαρμοσιμότητας. Στο k -οστό βήμα, υπολογίζουμε δύο προβλέψεις \hat{x}_k^α , για $\alpha = 1, 2$, που αντιστοιχούν σε δύο εκδοχές “διάδοσης” (propagation) της γνωστής κατάστασης x_{k-1} . Στην περίπτωση του Particle Filter, έχουμε N_s τέτοιες υποθέσεις και η διαδικασία που θα περιγράψουμε εφαρμόζεται ξεχωριστά για κάθε δείγμα (particle), ενώ στο Kalman Filter είναι $x_{k-1} \equiv m_{k-1|k-1}$. Από τον ορισμό 6.6, ο δείκτης της κύριας συνιστώσας προς τα εμπρός κίνησης στη μάσκα $B(x_{k-1})$ με βάση την κατάτμηση $S_k^{f,\alpha}$, για $\alpha = 1, 2$, γράφεται:

$$m_k^{f,\alpha}(x_{k-1}) \equiv PMC \left(S_k^{f,\alpha}, B(x_{k-1}) \right)$$

Σύμφωνα με την αναπαράσταση 6.4, το κεντροειδές (μέση ταχύτητα) που αντιστοιχεί στην παραπάνω κύρια συνιστώσα κίνησης, για $\alpha = 1, 2$, είναι το $c_k^{f,\alpha, m_k^{f,\alpha}(x_{k-1})} \equiv c_k^\alpha$ (για συντομία στο συμβολισμό). Σημειώνουμε ότι το μέγεθος αυτό υπολογίζεται ούτως ή άλλως από τις μεθόδους μας (εξισώσεις 6.12 και 6.44). Από τις εξισώσεις 6.15 και 6.16, έχουμε στη διάθεσή μας από δύο εκδοχές για τους συντελεστές κλιμάκωσης του πλαισίου οριοθέτησης στις διευθύνσεις των αξόνων, $r_k^{x,\alpha}$ και $r_k^{y,\alpha}$, για $\alpha = 1, 2$, οι οποίες επίσης χρησιμοποιούνται ήδη από τις μεθόδους μας. Με βάση τα παραπάνω μεγέθη και την x_{k-1} , υπολογίζουμε τις προβλέψεις, για $\alpha = 1, 2$:

$$\hat{x}_k^\alpha = \left(x_{k-1[1]} + c_{k[1]}^\alpha, x_{k-1[2]} + c_{k[2]}^\alpha, r_k^{x,\alpha} x_{k-1[3]}, r_k^{y,\alpha} x_{k-1[4]}, c_{k[1]}^\alpha, c_{k[2]}^\alpha \right) \quad (6.54)$$

Σημειώνουμε ότι η ταχύτητα στις \hat{x}_k^α , που έχει τεθεί ίση με c_k^α , για $\alpha = 1, 2$, δε θα παίζει κάποιο ρόλο στον έλεγχο που ακολουθεί.

Στη συνέχεια, θεωρούμε την ανάλυση της χρωματικής μάσκας B_k^c του πλαισίου I_k σε συνεκτικές συνιστώσες, όπου η σχέση γειτνίασης ανάμεσα σε δύο διακριτές θέσεις (εικονοστοιχεία)

ορίζεται από το μοντέλο της 4-συνεκτικότητας. Αν η B_k^c έχει N_k^c συνεκτικές συνιστώσες, η ανάλυση αυτή μπορεί να αναπαρασταθεί ως το σύνολο των δυαδικών εικόνων:

$$CC(B_k^c) = \{B_k^{c,i} : i = 1, \dots, N_k^c\} \quad (6.55)$$

με τις ακόλουθες ιδιότητες:

- $B_k^c = \bigcup_{i=1, \dots, N_k^c} B_k^{c,i}$
- Για κάθε $i = 1, \dots, N_k^c$, η $B_k^{c,i}$ είναι συνεκτική
- Για κάθε $i, j = 1, \dots, N_k^c$, με $i \neq j$, η $B_k^{c,i} \cup B_k^{c,j}$ δεν είναι συνεκτική

Η ανάλυση μιας δυαδικής εικόνας σε συνεκτικές συνιστώσες, και συνεπώς ο υπολογισμός του συνόλου $CC(B_k^c)$, μπορεί να γίνει αποδοτικά. Χρησιμοποιώντας αυτή την αναπαράσταση, συμβολίζουμε με $B_k^{c,p}(x)$ τη συνεκτική συνιστώσα (δυαδική εικόνα) της B_k^c που καλύπτει το μεγαλύτερο μέρος της μάσκας κατάστασης της x :

$$B_k^{c,p}(x) \equiv \operatorname{argmax}_{B \in CC(B_k^c)} \frac{|B \cap B(x)|}{|B(x)|} = \operatorname{argmax}_{B \in CC(B_k^c)} |B \cap B(x)| \quad (6.56)$$

Από τις καταστάσεις-προβλέψεις 6.54, επιλέγουμε την \hat{x}_k^β , $\beta \in \{1, 2\}$, της οποίας η μάσκα κατάστασης καλύπτει σε μεγαλύτερο ποσοστό την κυρίαρχη σε αυτήν (μέγιστης κάλυψης) συνεκτική συνιστώσα της B_k^c , όπως αυτή ορίζεται στην εξίσωση 6.56, όπου:

$$\beta = \operatorname{argmax}_{\alpha=1,2} \frac{|B_k^{c,p}(\hat{x}_k^\alpha) \cap B(\hat{x}_k^\alpha)|}{|B_k^{c,p}(\hat{x}_k^\alpha)|} \quad (6.57)$$

Η συνθήκη που καθορίζει αν η εκτίμηση στο k -οστό βήμα μπορεί να γίνει χωρίς αβεβαιότητα συνίσταται στον έλεγχο του αν η μάσκα κατάστασης της \hat{x}_k^β που επιλέξαμε καλύπτει ένα ελάχιστο ποσοστό $\tau^d \in [0, 1]$ της κυρίαρχης σε αυτήν συνεκτικής συνιστώσας $B_k^{c,p}(\hat{x}_k^\beta)$ της χρωματικής μάσκας B_k^c :

$$\frac{|B_k^{c,p}(\hat{x}_k^\beta) \cap B(\hat{x}_k^\beta)|}{|B_k^{c,p}(\hat{x}_k^\beta)|} \stackrel{?}{\geq} \tau^d \quad (6.58)$$

Για το ποσοστό-κατώφλι, στην υλοποίησή μας, χρησιμοποιήσαμε την τιμή $\tau^d = 0.8$.

Αν η συνθήκη 6.58 δεν είναι αληθής, η τεχνική που περιγράφουμε σε αυτή την ενότητα δεν μπορεί να εφαρμοστεί και οι μέθοδοι παρακολούθησής μας συμπεριφέρονται όπως στη γενική περίπτωση που περιγράψαμε στις αντίστοιχες ενότητες. Διαφορετικά, θεωρούμε ότι η συνεκτική συνιστώσα $B_k^{c,p}(\hat{x}_k^\beta)$ της χρωματικής μάσκας υποδεικνύει την προβολή του κινούμενου αντικειμένου στο επίπεδο της εικόνας. Ορίζουμε την τελική μας εκτίμηση \hat{x}_k με βάση το ελάχιστο πλαίσιο οριοθέτησης της μάσκας $B_k^{c,p}(\hat{x}_k^\beta)$ και την κατάσταση x_{k-1} ως ακολούθως:

- Οι συνιστώσες θέσης $\hat{x}_{k[1]}$ και $\hat{x}_{k[2]}$ είναι οι συντεταγμένες του κέντρου του πλαισίου οριοθέτησης της $B_k^{c,p}(\hat{x}_k^\beta)$.
- Οι συνιστώσες μεγέθους $\hat{x}_{k[3]}$ και $\hat{x}_{k[4]}$ είναι ίσες με τις διαστάσεις μεγέθους του πλαισίου οριοθέτησης της $B_k^{c,p}(\hat{x}_k^\beta)$.

- Οι συνιστώσες ταχύτητας/μετατόπισης $\hat{x}_{k[5]}$ και $\hat{x}_{k[6]}$ ορίζονται από τη μεταβολή της θέσης από την x_{k-1} στην \hat{x}_k : $\hat{x}_{k[5]} = \hat{x}_{k[1]} - x_{k-1[1]}$ και $\hat{x}_{k[6]} = \hat{x}_{k[2]} - x_{k-1[2]}$.

Στις επόμενες παραγράφους, περιγράφουμε τον τρόπο της “επιβολής” της \hat{x}_k ως εκτίμησης χωρίς αβεβαιότητα στο k -οστό βήμα των δύο αλγορίθμων παρακολούθησής μας, εφόσον η 6.58 είναι αληθής.

Τροποποίηση Particle Filter. Η ενσωμάτωση των παραπάνω στη μέθοδο παρακολούθησης με Particle Filter γίνεται με ιδιαίτερα κομψό τρόπο, μέσω μιας απλής γενίκευσης του θεωρούμενου δυναμικού μοντέλου $p(x_k|x_{k-1})$, το οποίο, στην περίπτωσή μας, ταυτίζεται με την πυκνότητα σημαντικότητας του φίλτρου. Συγκεκριμένα, αν η συνθήκη 6.58 είναι αληθής, θεωρούμε ότι το μοντέλο μετάβασης ορίζεται από την:

$$p(x_k|x_{k-1}) = \delta(x_k - \hat{x}_k), \quad (6.59)$$

όπου δ είναι το συναρτησιακό Dirac και η \hat{x}_k έχει υπολογισθεί με βάση την x_{k-1} , όπως ακριβώς περιγράψαμε παραπάνω. Η δειγματοληψία από αυτή την “πυκνότητα” δίνει, με πιθανότητα 1, τιμές ίσες με \hat{x}_k . Μια πρακτικά ισοδύναμη εναλλακτική, ενδεχομένως προτιμότερη για τεχνικούς λόγους, είναι να θεωρήσουμε ότι η $p(x_k|x_{k-1})$ είναι μια Γκαουσιανή πυκνότητα με μέση τιμή \hat{x}_k και πίνακα συμμεταβλητότητας $\sigma^2 I_6$:

$$p(x_k|x_{k-1}) = \mathcal{N}(x_k; \hat{x}_k, \sigma^2 I_6), \quad (6.60)$$

όπου σ μια πολύ μικρή θετική σταθερά, κοντά στην ακρίβεια της μηχανής. Αν η συνθήκη 6.58 δεν ισχύει, η $p(x_k|x_{k-1})$ ορίζεται ακριβώς όπως περιγράψαμε στην ενότητα 6.3.1.

Σημειώνουμε ότι όλα τα παραπάνω επαναλαμβάνονται ξεχωριστά για κάθε ένα από τα N_s δείγματα (particles). Κατά τη διάρκεια της εκτέλεσης της μεθόδου παρακολούθησης, όσα από αυτά πληρούν τις προϋποθέσεις που περιγράψαμε σε αυτή την ενότητα εξελίσσονται ντετερμινιστικά στο τρέχον βήμα, ενώ τα υπόλοιπα ακολουθούν το γενικό μοντέλο της ενότητας 6.3.1.

Τροποποίηση Kalman Filter. Η ενσωμάτωση των παραπάνω στη μέθοδο παρακολούθησης με Kalman Filter γίνεται μέσω του μοντέλου παρατήρησης του φίλτρου. Συγκεκριμένα, αν η συνθήκη 6.58 είναι αληθής, θεωρούμε ότι ο χώρος των παρατηρήσεων, για το τρέχον βήμα, ταυτίζεται με το χώρο καταστάσεων. Συγκεκριμένα, ως παρατήρηση στο k -οστό βήμα θεωρούμε ακριβώς την κατάσταση \hat{x}_k που υπολογίσαμε παραπάνω:

$$y_k = \hat{x}_k \quad (6.61)$$

Ο πίνακας H_k , που απεικονίζει καταστάσεις σε παρατηρήσεις, είναι τότε απλά ο ταυτοτικός 6×6 πίνακας:

$$H_k = I_6 \quad (6.62)$$

Στην περίπτωση αυτή, έχουμε αναπληροφόρηση (feedback) της εκτίμησης στο k -οστό βήμα από την αντίστοιχη παρατήρηση και για τις συνιστώσες μεγέθους, πράγμα που δεν συνέβαινε με το γενικό μοντέλο παρατήρησης της ενότητας 6.4.2. Η απόλυτη βεβαιότητα για την εκτίμηση \hat{x}_k , μπορεί να εκφραστεί από την ανάθεση του μηδενικού 6×6 πίνακα ως πίνακα συμμεταβλητότητας R_k του θορύβου παρατήρησης:

$$R_k = 0_{6,6} \quad (6.63)$$

Μια προτιμότερη, για πρακτικούς λόγους, εναλλακτική είναι να θέσουμε τον R_k στην τιμή:

$$R_k = \sigma^2 I_6, \quad (6.64)$$

όπου σ μια πολύ μικρή θετική σταθερά, κοντά στην ακρίβεια της μηχανής. Οι παραπάνω εξισώσεις δίνουν τα τρία μεγέθη που ορίζουν πλήρως το μοντέλο παρατήρησης της εξίσωσης 6.33. Αν η συνθήκη 6.58 δεν ισχύει, το μοντέλο παρατήρησης του φίλτρου ορίζεται όπως ακριβώς περιγράψαμε στην ενότητα 6.4.2.

6.6 Παρατηρήσεις που αφορούν τις υλοποιήσεις

Οι μέθοδοι παρακολούθησής μας υλοποιήθηκαν σε περιβάλλον MATLAB. Το σύνολο της λογικής των αλγορίθμων αποτελεί προϊόν δικής μας εργασίας. Για την εξαγωγή ορισμένων από τα ενδιάμεσα αποτελέσματα στα οποία βασιζόμαστε, τα οποία προκύπτουν από την επίλυση κλασικών προβλημάτων Όρασης Υπολογιστών (π.χ., εκτίμηση οπτικής ροής) και Αναγνώρισης Προτύπων (εκπαίδευση μοντέλων μειγμάτων Γκαουσιανών κατανομών, ταξινόμηση), χρησιμοποιήσαμε ήδη διαθέσιμες υλοποιήσεις:

- Για τον υπολογισμό των πεδίων οπτικής ροής, πειραματιστήκαμε με έναν αριθμό από ελεύθερα διαθέσιμες υλοποιήσεις, καθώς και μια απλή, δική μας υλοποίηση μιας πολυκλιμακωτής εκδοχής της μεθόδου Lucas-Kanade [34]. Από αυτές, κατόπιν εμπειρικής αξιολόγησης στις ακολουθίες που χρησιμοποιούμε για τα πειράματα του Κεφαλαίου 7, ξεχωρίσαμε αυτές που συνοδεύουν τα [62, 40, 78], θεωρώντας ότι δίνουν, σε γενικές γραμμές, εξίσου ακριβή αποτελέσματα. Ως “προεπιλεγμένη” μέθοδο εκτίμησης της οπτικής ροής στους αλγορίθμους μας, υιοθετήσαμε αυτήν που υλοποιείται από το [78], η οποία συνδυάζει ικανοποιητική ακρίβεια με, συγκριτικά, χαμηλούς χρόνους εκτέλεσης. Συγκεκριμένα, από την ελεύθερα διαθέσιμη βιβλιοθήκη σε C++ (Particle Video Library) που παρέχουν οι συγγραφείς του [78], δημιουργήσαμε μια δυναμική βιβλιοθήκη mex που υλοποιεί το κομμάτι εκτίμησης οπτικής ροής της γενικότερης εργασίας. Η βιβλιοθήκη υλοποιεί τη μέθοδο [58], που είναι προγενέστερη δουλειά της [40].
- Χρησιμοποιήσαμε την κλάση `gmdistribution` του Statistics Toolbox της MATLAB που υλοποιεί μείγματα Γκαουσιανών κατανομών. Για παράδειγμα, αναπαριστούμε ως αντικείμενα αυτής της κλάσης τα χρωματικά μοντέλα των αντικειμένων και τις επιμέρους πυκνότητες της 6.11 που ορίζει το δυναμικό μοντέλο της μεθόδου παρακολούθησης με Particle Filter. Για την εκπαίδευση των χρωματικών μοντέλων, χρησιμοποιήσαμε τη μέθοδο `fit` αυτής της κλάσης, η οποία υλοποιεί ένα στιγμιότυπο του αλγορίθμου EM (Παράρτημα B).
- Ως υλοποίηση της ταξινόμησης k -means (Παράρτημα A) για την απλή κατάτμηση με βάση την κίνηση που περιγράψαμε, χρησιμοποιήσαμε αυτήν που παρέχει η συνάρτηση `kmeans` του Statistics Toolbox της MATLAB.

Κεφάλαιο 7

Πειράματα

Στο κεφάλαιο αυτό, εφαρμόζουμε τις μεθόδους που αναπτύξαμε σε πραγματικές ακολουθίες βίντεο. Αρχικά, περιγράφουμε τις ακολουθίες βίντεο που θα χρησιμοποιήσουμε ως εισόδους για τους αλγορίθμους μας και στη συνέχεια, αφού περιγράψουμε τη μεθοδολογία της πειραματικής αξιολόγησης που θα ακολουθήσουμε, παραθέτουμε ένα σύνολο αποτελεσμάτων και προβαίνουμε σε ένα σύντομο σχολιασμό αυτών.

7.1 Ακολουθίες βίντεο εισόδου

Για τις δοκιμές μας, επικεντρώσαμε το ενδιαφέρον μας σε ακολουθίες βίντεο νοηματικής γλώσσας, όπου παρακολουθούμε το δεξί χέρι (παλάμη) του νοηματιστή. Δύο από αυτές αφορούν την ελληνική νοηματική γλώσσα και περιλαμβάνουν πολλαπλές επαναλήψεις της χειρονομίας που αντιστοιχεί σε ένα συγκεκριμένο νόημα. Η τρίτη αφορά την αμερικανική νοηματική και έχει ληφθεί απευθείας από το διαδίκτυο. Λεπτομερέστερα, σε αύξουσα σειρά δυσκολίας για το πρόβλημα της οπτικής παρακολούθησης, θα χρησιμοποιήσουμε τις ακολουθίες:

- **GSL10_5** (ελληνική νοηματική). Περιορίσαμε τα πειράματά μας στα πλαίσια της ακολουθίας που αντιστοιχούν σε μία επανάληψη της χειρονομίας. Η κίνηση του δεξιού χεριού είναι γρήγορη και με απότομες μεταβολές ταχύτητας, ενώ επιπλέον η προβολή του αλλάζει σημαντικά σχήμα και μέγεθος κατά τη διάρκεια της κίνησης. Ο νοηματιστής φοράει μακριά μανίκια και σχεδόν σε όλη τη διάρκεια της εκτέλεσης της χειρονομίας, δεν υπάρχει επικάλυψη ανάμεσα στις προβολές στο επίπεδο της εικόνας των χεριών ή των χεριών και του προσώπου. Χάρη στο γεγονός αυτό, η παρακολούθηση του χεριού αναμένεται να γίνει από τους αλγορίθμους μας με ακρίβεια στηριζόμενη το χρώμα (ενότητα 6.5).
- **GSL10_6** (ελληνική νοηματική). Έχουμε ακριβώς το ίδιο σενάριο με την προηγούμενη ακολουθία, για χειρονομία που αντιστοιχεί σε διαφορετικό νόημα. Η διαφορά εδώ είναι ότι η χειρονομία περιλαμβάνει κίνηση και των δύο χεριών και μάλιστα υπάρχει επικάλυψη των προβολών τους στο επίπεδο της εικόνας για σημαντικό μέρος της θεωρούμενης διάρκειας. Στο σενάριο αυτό, αναμένουμε εναλλαγές (στο επίπεδο των particles, για το Particle Filter) στη λειτουργία των αλγορίθμων ανάμεσα στη γενική μέθοδο που υλοποιούν και την τεχνική της ενότητας 6.5.
- **ASL** (αμερικανική νοηματική). Η ακολουθία αυτή είναι με διαφορά η δυσκολότερη για τους αλγορίθμους μας. Ο νοηματιστής φοράει κοντά μανίκια και, καθώς μας ενδιαφέρει η παρακολούθηση μόνο της παλάμης, η παρακολούθηση δεν μπορεί να γίνει με ακρίβεια μόνο

με βάση το χρώμα και, συνεπώς, η τεχνική της ενότητας 6.5 δεν εφαρμόζεται. Επιπλέον, υπάρχουν σημαντικές επικαλύψεις του δεξιού χεριού με το πρόσωπο, πράγμα που ρίχνει το μεγαλύτερο βάρος για την εκτίμηση στο στοιχείο της κίνησης. Καθώς το βίντεο δεν είναι ιδιαίτερα υψηλής ποιότητας, η τελευταία δεν μπορεί να εκτιμηθεί με ακρίβεια μέσω της οπτικής ροής, διότι οι περιοχές της εικόνας που χαρακτηρίζονται από γρήγορη κίνηση είναι θολές (motion blur).

Ενδεικτικά πλαίσια από τις παραπάνω ακολουθίες απεικονίζονται στο Σχήμα 7.1.



Σχήμα 7.1: Ενδεικτικά πλαίσια από τις ακολουθίες εισόδου

Η εκπαίδευση του χρωματικού μοντέλου για το αντικείμενο, το οποίο θυμίζουμε ότι αποτελεί μέρος της εισόδου των αλγορίθμων μας, έγινε με τις τεχνικές που περιγράψαμε στο Κεφάλαιο 3. Συγκεκριμένα, ακολουθήσαμε τη διαδικασία που περιγράψαμε στην ενότητα 3.3.3 (Εφαρμογή 2): σε ένα μικρό υποσύνολο εικόνων από την εκάστοτε ακολουθία, εφαρμόσαμε τον Αλγόριθμο 1, ο οποίος επιτυγχάνει μια πρόχειρη κατάτμηση των περιοχών της εικόνας που απεικονίζουν ανθρώπινο δέρμα. Με τον τρόπο που περιγράψαμε στην εφαρμογή, χρησιμοποιήσαμε τα αποτελέσματα της πρόχειρης αυτής κατάτμησης για την αυτοματοποιημένη εξαγωγή δεδομένων εκπαίδευσης για το χρωματικό μοντέλο του αντικειμένου. Για το τελευταίο, σε κάθε περίπτωση, εκπαίδευσουμε ένα μείγμα Γκαουσιανών κατανομών με 2 συνιστώσες σε 2 διαστάσεις εργαζόμενοι στο χώρο χρώματος YCbCr (συνιστώσες Cb και Cr).

7.2 Περιγραφή της μεθοδολογίας αξιολόγησης

Η πειραματική αξιολόγηση των αποτελεσμάτων της μεθόδου που αναπτύξαμε συνίσταται στη σύγκριση των εκτιμήσεων που αυτή παράγει με αποτελέσματα που έχουν προκύψει από χειροκίνητη, πλαίσιο προς πλαίσιο επισημείωση των ακολουθιών βίντεο εισόδου και τα οποία θεωρούμε ότι αντιπροσωπεύουν πραγματικά, μηδενικού σφάλματος αποτελέσματα παρακολούθησης (ground truth). Ακολούθως, περιγράφουμε τη διαδικασία και τα δεδομένα επισημείωσης και στη συνέχεια ορίζουμε τα μετρικά αξιολόγησης που θα χρησιμοποιήσουμε.

Χρησιμοποιώντας τη σημειογραφία του Κεφαλαίου 6, συμβολίζουμε με $\langle I_k : k = 1, \dots, N \rangle$ την ακολουθία εικόνων εισόδου και με $\langle x_k : k = 1, \dots, N \rangle$ την ακολουθία των εκτιμήσεων της κατάστασης του συστήματος που υπολογίζεται για αυτήν από κάποια από τις μεθόδους παρακολούθησης. Στη μέθοδο παρακολούθησης με Particle Filter, η x_k λαμβάνεται ως ο σταθμισμένος μέσος των δειγμάτων (particles) σύμφωνα με τα βάρη στάθμισης που τα συνοδεύουν. Αν $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$ είναι το σύνολο των σταθμισμένων δειγμάτων με τα οποία αναπαριστούμε την $p(x_k | y_{1:k})$ στο k -οστό χρονικό βήμα, η εκτίμηση x_k λαμβάνεται από την εξίσωση 6.9

(εκτίμηση MMSE):

$$x_k = \sum_{i=1}^{N_s} w_k^i x_k^i$$

και αντιστοιχεί στην προσδοκώμενη τιμή της τυχαίας μεταβλητής x_k υπό τη θεωρούμενη προσέγγιση της $p(x_k | y_{1:k})$. Θυμίζουμε ότι, στην περίπτωση του Kalman Filter, η x_k είναι απλώς η εκτίμηση $m_{k|k}$.

Η επισημείωση της ακολουθίας εισόδου γίνεται πλαίσιο προς πλαίσιο, μέσω του χειροκίνητου προσδιορισμού του ελάχιστου μεγέθους πλαισίου οριοθέτησης (bounding box) του προς παρακολούθηση αντικειμένου στο τρέχον καρέ. Από αυτήν ορίζουμε την ακολουθία των πραγματικών (ground truth) καταστάσεων $\langle \chi_k : k = 1, \dots, N \rangle$. Οι 4 πρώτες συνιστώσες της χ_k ορίζονται άμεσα από τις συντεταγμένες θέσης του κέντρου και τις διαστάσεις μεγέθους του πλαισίου οριοθέτησης στο k -οστό πλαίσιο της ακολουθίας (ενότητα 6.1). Καθώς οι συνιστώσες ταχύτητας της κατάστασης δε θα ληφθούν υπόψη από τα μετρικά της πειραματικής αξιολόγησης που θα ορίσουμε στη συνέχεια, οι συνιστώσες $\chi_{k[5]}$ και $\chi_{k[6]}$ θα αγνοηθούν και μπορεί να θεωρηθεί ότι τίθενται σε κάποια αυθαίρετη τιμή.

7.2.1 Ορισμός των μετρικών αξιολόγησης

Προτού περιγράψουμε τα μετρικά που θα χρησιμοποιήσουμε για την αξιολόγηση των πειραματικών μας αποτελεσμάτων, θα ορίσουμε το F-score, ένα μέτρο ακρίβειας του αποτελέσματος ενός πειράματος δυαδικής ταξινόμησης. Σε ένα σενάριο δυαδικής ταξινόμησης, το ζητούμενο είναι ο προσδιορισμός ενός υποσυνόλου των δεδομένων στοιχείων που διαθέτει κάποια ιδιότητα. Έστω G το πραγματικό (ground truth) υποσύνολο του οποίου όλα τα στοιχεία και μόνο αυτά διαθέτουν την ιδιότητα που μας ενδιαφέρει και H το υποσύνολο που επέστρεψε ως αποτέλεσμα ένας ταξινομητής. Το precision (ακρίβεια) του πειράματος ορίζεται ως το κλάσμα των στοιχείων που επέστρεψε ο ταξινομητής που πράγματι διαθέτουν διαθέτουν την εν λόγω ιδιότητα:

$$\mathcal{P}(H, G) = \frac{|H \cap G|}{|H|} \quad (7.1)$$

Η ποσότητα αυτή είναι ίση με την πιθανότητα ένα τυχαία επιλεγμένο (από την ομοιόμορφη κατανομή) στοιχείο που επέστρεψε ο ταξινομητής να διαθέτει πράγματι την ιδιότητα που μας ενδιαφέρει. Το recall (ευαισθησία) του πειράματος ορίζεται ως το κλάσμα των στοιχείων που πράγματι διαθέτουν την ιδιότητα και που επιπλέον επιστρέφονται από τον ταξινομητή:

$$\mathcal{R}(H, G) = \frac{|H \cap G|}{|G|} \quad (7.2)$$

Η ποσότητα αυτή είναι ίση με την πιθανότητα ένα τυχαία επιλεγμένο (από την ομοιόμορφη κατανομή) στοιχείο που πράγματι διαθέτει την ιδιότητα που μας ενδιαφέρει να επιστραφεί από τον ταξινομητή. Το F-score συνδυάζει τα δύο αυτά μεγέθη και ορίζεται ως ο αρμονικός μέσος (αντίστροφος του αριθμητικού μέσου των αντίστροφων τιμών) του precision και του recall:

$$\mathcal{F}(H, G) = 2 \frac{\mathcal{R}(H, G) \mathcal{P}(H, G)}{\mathcal{R}(H, G) + \mathcal{P}(H, G)} = 2 \frac{|H \cap G|}{|H| + |G|} \quad (7.3)$$

Όπως προκύπτει από την παραπάνω, το F-score αποτελεί ένα μέτρο ομοιότητας (similarity measure) των συνόλων H και G και λαμβάνει τιμές στο διάστημα $[0, 1]$.

Θυμίζουμε από την ενότητα 6.2 ότι, αν x είναι ένα διάνυσμα κατάστασης, συμβολίζουμε με $B(x)$ την αντίστοιχη μάσκα κατάστασης. Με βάση όλα τα παραπάνω, ορίζουμε τα ακόλουθα τρία μέτρα για την αξιολόγηση των πειραματικών μας αποτελεσμάτων:

- Το *σφάλμα θέσης* e_k^p τη χρονική στιγμή k δίνεται από την Ευκλείδεια απόσταση της θέσης στην εκτιμηθείσα κατάσταση x_k από αυτήν στην πραγματική κατάσταση χ_k :

$$e_k^p \equiv \sqrt{(x_{k[1]} - \chi_{k[1]})^2 + (x_{k[2]} - \chi_{k[2]})^2} \quad (7.4)$$

Το μέσο σφάλμα θέσης \bar{e}^p δίνεται από τη μέση τιμή του συνόλου τιμών της ακολουθίας $\langle e_k^p : k = 1, \dots, N \rangle$:

$$\bar{e}^p \equiv \frac{1}{N} \sum_{k=1}^N e_k^p \quad (7.5)$$

Το μέτρο αυτό δίνει τη (μέση) απόσταση σε εικονοστοιχεία του κέντρου του πλαισίου οριοθέτησης που υπολογίστηκε από αυτό του πραγματικού πλαισίου (ground truth).

- Το μέτρο *ομοιότητας μάσκας κατάστασης* F_k^b τη χρονική στιγμή k δίνεται από το F-score της μάσκας της εκτιμηθείσας κατάστασης $B(x_k)$ ως προς τη μάσκα της πραγματικής κατάστασης $B(\chi_k)$:

$$F_k^b \equiv \mathcal{F}(B(x_k), B(\chi_k)) \quad (7.6)$$

Η μέση ομοιότητα μάσκας κατάστασης \bar{F}^b δίνεται από τη μέση τιμή του συνόλου τιμών της ακολουθίας $\langle F_k^b : k = 1, \dots, N \rangle$:

$$\bar{F}^b \equiv \frac{1}{N} \sum_{k=1}^N F_k^b \quad (7.7)$$

Το μέτρο αυτό εκφράζει την (κατά μέσο όρο) ομοιότητα του πλαισίου οριοθέτησης στην κατάσταση που υπολογίστηκε με αυτό στην πραγματική κατάσταση (ground truth).

- Το μέτρο *ομοιότητας χρωματικής μάσκας* F_k^c τη χρονική στιγμή k δίνεται από το F-score του μέρους της μάσκας της εκτιμηθείσας κατάστασης $B(x_k)$ που επικαλύπτει τη χρωματική μάσκα του αντικειμένου B_k^c στο πλαίσιο I_k ($B(x_k) \cap B_k^c$) ως προς το μέρος της μάσκας της πραγματικής κατάστασης $B(\chi_k)$ που επίσης επικαλύπτει τη χρωματική μάσκα του αντικειμένου ($B(\chi_k) \cap B_k^c$):

$$F_k^c \equiv \mathcal{F}(B(x_k) \cap B_k^c, B(\chi_k) \cap B_k^c) \quad (7.8)$$

Η μέση ομοιότητα χρωματικής μάσκας \bar{F}^c δίνεται από τη μέση τιμή του συνόλου τιμών της ακολουθίας $\langle F_k^c : k = 1, \dots, N \rangle$:

$$\bar{F}^c \equiv \frac{1}{N} \sum_{k=1}^N F_k^c \quad (7.9)$$

Το μέτρο αυτό εκφράζει την (κατά μέσο όρο) ομοιότητα του μέρους της χρωματικής μάσκας του αντικειμένου που οριοθετείται από την εκτιμηθείσα κατάσταση με το μέρος αυτής που οριοθετείται από την πραγματική κατάσταση (ground truth). Η ομοιότητα χρωματικής μάσκας παρέχει σε ορισμένες περιπτώσεις μια περισσότερο διαισθητική εποπτεία της ποιότητας ενός αποτελέσματος παρακολούθησης, καθώς περιορίζει τη σύγκριση εντός της χρωματικής μάσκας του αντικειμένου. Σημειώνουμε ότι το μέτρο F_k^c μπορεί να είναι μεγαλύτερο ή μικρότερο του F_k^b , ανάλογα με τη μορφή της χρωματικής μάσκας και με τη σχέση της με τα πλαίσια οριοθέτησης που ορίζουν οι υπό σύγκριση καταστάσεις.

7.2.2 Πειραματική διαδικασία

Θα επικεντρώσουμε το ενδιαφέρον μας στα ακόλουθα δύο σύνολα πειραματικών δοκιμών:

- **Πείραμα 1:** Αξιολόγηση των μεθόδων παρακολούθησης και μεταξύ τους σύγκριση.

Το πείραμα αυτό αποσκοπεί στην ποσοτική αξιολόγηση των αποτελεσμάτων που παράγουν οι μέθοδοι που αναπτύξαμε, με βάση τα μετρικά που ορίσαμε προηγουμένως με δεδομένα τα πραγματικά (ground truth) αποτελέσματα παρακολούθησης. Για κάθε μια από τις τρεις ακολουθίες βίντεο, υπολογίζουμε, για τις δύο μεθόδους μας, τα τρία μέτρα (σφάλμα θέσης e_k^p , ομοιότητα μάσκας κατάστασης F_k^b και ομοιότητα χρωματικής μάσκας F_k^c) ως συναρτήσεις του χρόνου/πλασίου του βίντεο. Με βάση τις χρονικές μέσες τιμές των τριών μέτρων, αξιολογούμε τις επιδόσεις της κύριας μεθόδου παρακολούθησής μας έχοντας ως αναφορά το Kalman Filter. Για όλες τις δοκιμές, η αρχικοποίηση των αλγορίθμων (ορισμός της αρχικής κατάστασης του συστήματος) έγινε χειροκίνητα, με τον τρόπο που περιγράψαμε στο Κεφάλαιο 6. Για τη μέθοδο παρακολούθησης με Particle Filter, χρησιμοποιήσαμε πλήθος σταθμισμένων δειγμάτων (particles) ίσο με $N_s = 30$.

- **Πείραμα 2:** Επίδραση του μεγέθους του συνόλου σταθμισμένων δειγμάτων.

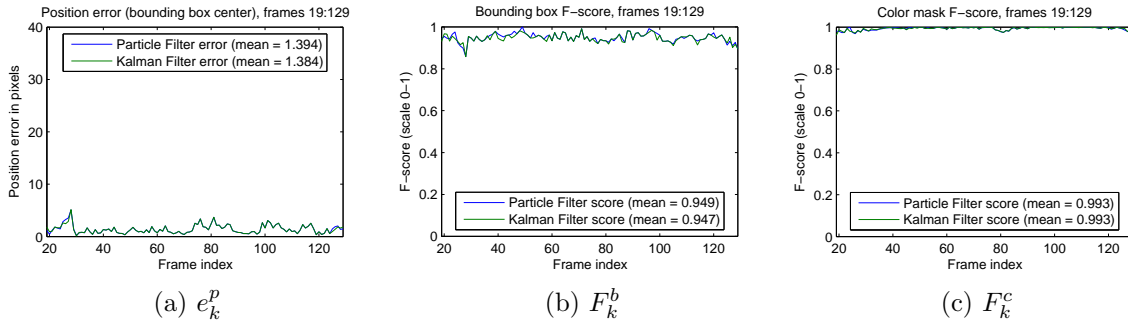
Το πείραμα αυτό αφορά μόνο την κύρια μέθοδο παρακολούθησης, αυτήν που υλοποιεί ένα Particle Filter. Θυμίζουμε ότι το σύνολο των σταθμισμένων δειγμάτων στο k -οστό χρονικό βήμα, $\{(x_k^i, w_k^i) : i = 1, \dots, N_s\}$, αποτελεί, σύμφωνα με τις υποθέσεις που πραγματοποιήσαμε στο Κεφάλαιο 5, μια διακριτή προσέγγιση της συνάρτησης πυκνότητας $p(x_k|y_{1:k})$. Για μεγαλύτερο μέγεθος N_s αυτού του συνόλου, έχουμε καλύτερη προσέγγιση της πυκνότητας $p(x_k|y_{1:k})$. Εφόσον το μοντέλο μετάβασης που θεωρήσαμε ανταποκρίνεται στοιχειωδώς στην πραγματική δυναμική του υπό παρακολούθηση συστήματος, για μεγαλύτερο N_s , αναμένεται να έχουμε ακριβέστερα αποτελέσματα παρακολούθησης. Για κάθε μια από τις ακολουθίες, πραγματοποιούμε προσομοιώσεις για $N_s \in \{10, 20, 30, 50, 100, 150\}$ και εξετάζουμε τη συμπεριφορά των χρονικών μέσων τιμών των τριών μετρικών αξιολόγησης που ορίσαμε ως συναρτήσεις του μεγέθους N_s . Για κάθε ακολουθία, χρησιμοποιήθηκε η ίδια, χειροκίνητα ορισμένη αρχικοποίηση του αλγορίθμου για όλες της δοκιμές.

Καθώς όλες οι ακολουθίες βίντεο που εξετάζουμε αφορούν παρόμοια σενάρια παρακολούθησης χεριού, χρησιμοποιήσαμε σε όλα τα πειράματά μας την τιμή $N_c = 4$ για το πλήθος των κλάσεων στις καταταμίσεις με βάση την κίνηση (ενότητα 6.2) που πραγματοποιούν οι αλγόριθμοί μας (χέρια, πρόσωπο, παρασκήνιο). Σε όλες τις προσομοιώσεις για τη μέθοδο που βασίζεται σε Particle Filter, χρησιμοποιήσαμε τιμή για το ελάχιστο ενεργό μέγεθος δείγματος N_T ίση με $\lfloor \frac{N_s}{2} \rfloor$. Θυμίζουμε ότι το N_T λειτουργεί σαν κατώφλι για το ενεργό μέγεθος δείγματος \hat{N}_{eff} και καθορίζει πότε θα πραγματοποιηθεί επαναδειγματοληψία προς αποφυγή του εκφυλισμού του συνόλου των σταθμισμένων δειγμάτων.

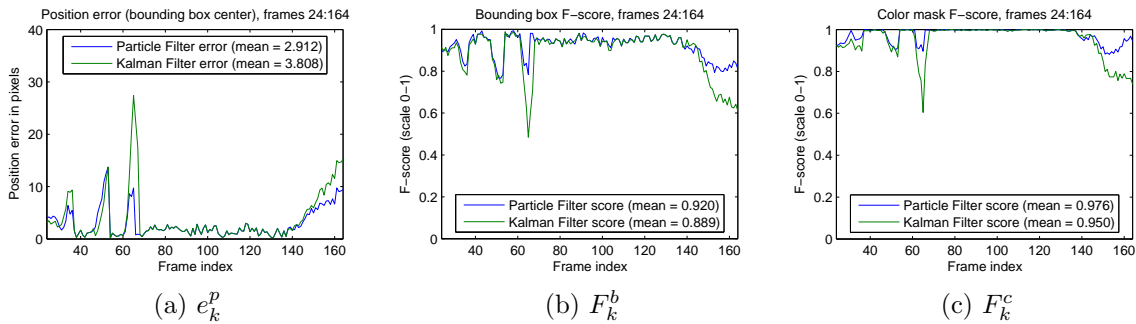
7.3 Πειραματικά αποτελέσματα

Πείραμα 1. Στα Σχήματα 7.2, 7.3 και 7.4 απεικονίζουμε τα αποτελέσματα του πρώτου συνόλου δοκιμών μας για τις τρεις ακολουθίες βίντεο. Σε κάθε περίπτωση και σε κοινά διαγράμματα για τους δύο αλγορίθμους μας, απεικονίζουμε, από αριστερά προς τα δεξιά, το σφάλμα θέσης e_k^p , το μέτρο ομοιότητας μάσκας κατάστασης F_k^b και την ομοιότητα χρωματικής μάσκας F_k^c ως συναρτήσεις του διακριτού χρόνου (πλασίου της ακολουθίας). Στον Πίνακα 7.1, παραθέτουμε

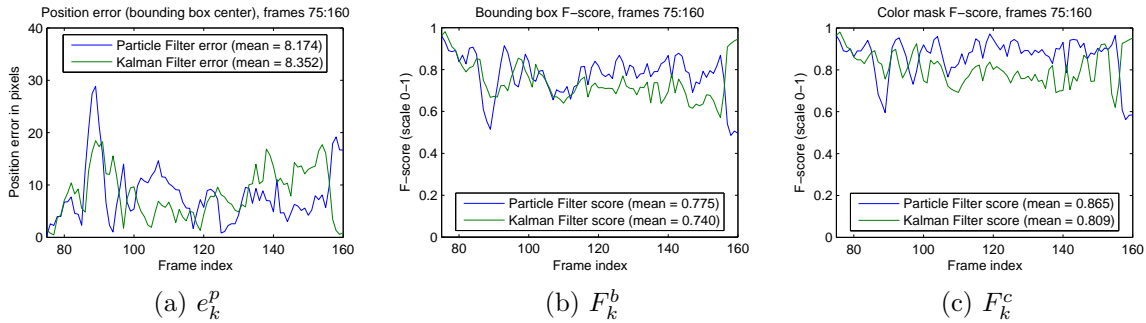
συγκεντρωτικά τις χρονικές μέσες τιμές των μέτρων αξιολόγησης για κάθε μία από τις δοκιμές μας.



Σχήμα 7.2: Ακολουθία GSL10_5 (πείραμα 1).



Σχήμα 7.3: Ακολουθία GSL10_6 (πείραμα 1).

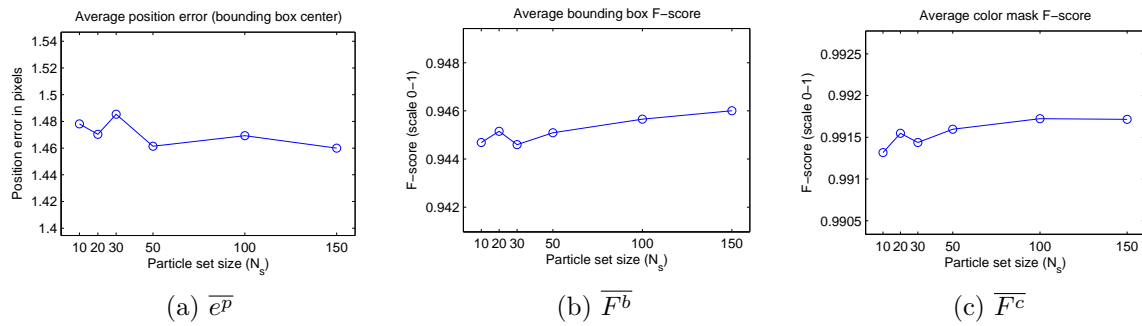


Σχήμα 7.4: Ακολουθία ASL (πείραμα 1).

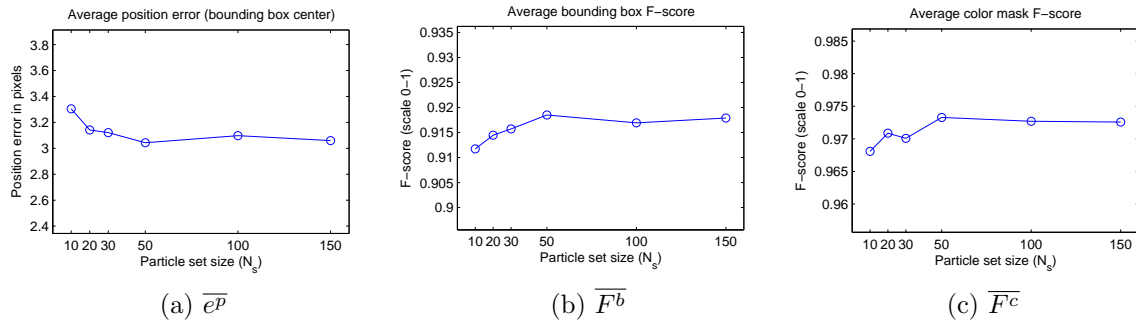
Πείραμα 2. Στα Σχήματα 7.5, 7.6 και 7.7 απεικονίζουμε τα αποτελέσματα του δεύτερου συνόλου δοκιμών μας για τις τρεις ακολουθίες βίντεο. Σε κάθε περίπτωση, απεικονίζουμε, από αριστερά προς τα δεξιά, το μέσο σφάλμα θέσης \bar{e}^p , τη μέση ομοιότητα μάσκας κατάστασης \bar{F}^b και τη μέση ομοιότητα χρωματικής μάσκας \bar{F}^c ως συναρτήσεις του πλήθους N_s των σταθμισμένων δειγμάτων (particles) που χρησιμοποιεί η μέθοδος παρακολούθησης. Για κάθε τιμή του μεγέθους N_s , πραγματοποιήσαμε 5 προσομοιώσεις και, για κάθε ένα από τα (μέσα) μέτρα αξιολόγησης, κρατήσαμε ως τελικό αποτέλεσμα το μέσο όρο των αποτελεσμάτων των προσομοιώσεων.

	Particle Filter			Kalman Filter		
	$\overline{e^p}$	$\overline{F^b}$	$\overline{F^c}$	$\overline{e^p}$	$\overline{F^b}$	$\overline{F^c}$
GSL10_5	1.394	0.949	0.993	1.384	0.947	0.993
GSL10_6	2.912	0.920	0.976	3.808	0.889	0.950
ASL	8.174	0.775	0.865	8.352	0.740	0.809

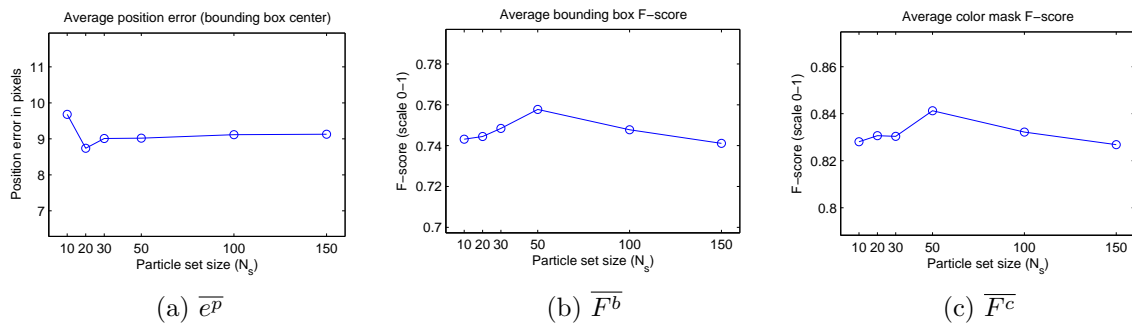
Πίνακας 7.1: Χρονικές μέσες τιμές των τριών μέτρων αξιολόγησης για κάθε ακολουθία και για τις δύο μεθόδους παρακολούθησης (πείραμα 1).



Σχήμα 7.5: Ακολουθία GSL10_5 (πείραμα 2).



Σχήμα 7.6: Ακολουθία GSL10_6 (πείραμα 2).



Σχήμα 7.7: Ακολουθία ASL (πείραμα 2).

7.4 Συμπεράσματα

Με βάση τα πειραματικά αποτελέσματα της προηγούμενης ενότητας, εξάγουμε ένα σύνολο συμπερασμάτων για κάθε μία από τις δοκιμές που πραγματοποιήσαμε.

Πείραμα 1. Τα αποτελέσματα του πρώτου πειράματος υποδεικνύουν ότι οι αλγόριθμοι που αναπτύξαμε μπορούν να εφαρμοστούν με επιτυχία για την οπτική παρακολούθηση ενός αντικειμένου σε πραγματικές ακολουθίες βίντεο. Με δεδομένη την αρκετά γενική φύση των μεθόδων μας ως προς τις υποθέσεις όπου στηρίζονται και την πρότερη γνώση που απαιτούν για το πρόβλημα όπου εφαρμόζονται, θεωρούμε ότι, και για τις τρεις ακολουθίες των πειραμάτων μας, το σφάλμα θέσης λαμβάνει μάλλον χαμηλές τιμές, ενώ τα μέτρα ομοιότητας μάσκας κατάστασης και χρωματικής μάσκας είναι εύλογα υψηλά. Συγκεκριμένα, εξετάζοντας τα διαγράμματα της χρονικής εξέλιξης του σφάλματος θέσης για τις τρεις δοκιμές μας, δεν παρατηρούμε το φαινόμενο της σταδιακής ολίσθησης (drifting) της θέσης του εκτιμηθέντος πλαισίου οριοθέτησης μακριά από την πραγματική προβολή του αντικειμένου στο επίπεδο της εικόνας. Αντίστοιχο συμπέρασμα μπορούμε να εξάγουμε από τα υπόλοιπα διαγράμματα, ιδιαίτερα για την περίπτωση του Particle Filter, και για τις διαστάσεις μεγέθους του πλαισίου οριοθέτησης. Το φαινόμενο της ολίσθησης εμφανίζεται συχνά σε αποτελέσματα αυτόματης οπτικής παρακολούθησης και οφείλεται στη συσσώρευση σφαλμάτων εκτίμησης κατά τη διάρκεια της διαδικασίας. Το γεγονός ότι δεν παρατηρείται στα αποτελέσματά μας, παρά τη σύνθετη, παραμορφώσιμη και με φαινόμενα αυτο-απόκρυψης (self-occlusion) κίνηση του υπό παρακολούθηση χεριού και στις τρεις δοκιμές, μας οδηγεί στο συμπέρασμα ότι η βασική μας υπόθεση, αυτή της κατά κύριο λόγο μεταφορικής κίνησης του αντικειμένου, είναι ρεαλιστική για το σενάριο της παρακολούθησης χεριού που εξετάσαμε.

Παρατηρώντας τις χρονικές μέσες τιμές των μέτρων αξιολόγησής μας στον Πίνακα 7.1, βλέπουμε ότι η μέθοδος που βασίζεται σε Particle Filter δίνει, κατά κανόνα, ακριβέστερες εκτιμήσεις. Το γεγονός αυτό ήταν αναμενόμενο, καθώς η κύρια μέθοδος μας λειτουργεί σε ένα σημαντικά γενικότερο θεωρητικό πλαίσιο, εκμεταλλευόμενη γενικότερα μοντέλα μετάβασης και παρατήρησης, και ο τρόπος με τον οποίο ορίσαμε το δυναμικό μοντέλο του Kalman Filter μας ήταν, ουσιαστικά, ως μια “ειδική περίπτωση” αυτού του Particle Filter. Επιπλέον, θυμίζουμε ότι το μοντέλο παρατήρησης του Kalman Filter δεν παρέχει κάποια αναπληροφόρηση για τις διαστάσεις μεγέθους του πλαισίου οριοθέτησης, εκτός αν είναι εφαρμόσιμη η τεχνική της παρακολούθησης χωρίς αβεβαιότητα που περιγράψαμε στην ενότητα 6.5. Η αδυναμία αυτή επιβεβαιώνεται πειραματικά από τα χρονικά διαγράμματα των μέτρων ομοιότητας της μάσκας κατάστασης F_k^b και της χρωματικής μάσκας F_k^c για την ακολουθία ASL (Σχήμα 7.4), όπου το Kalman Filter, από κάποιο σημείο και μετά, γίνεται συστηματικά λιγότερο ακριβές στις διαστάσεις μεγέθους του πλαισίου οριοθέτησης και παρουσιάζει τάσεις ολίσθησης ως προς αυτές.

Τέλος, ενδιαφέρουν παρουσιάζει η λειτουργία της παρακολούθησης χωρίς αβεβαιότητα που περιγράψαμε στην ενότητα 6.5. Η τεχνική αυτή εφαρμόζεται στο σύνολο, σχεδόν, της διάρκειας της ακολουθίας GSL10_5, για την οποία τα αποτελέσματα της παρακολούθησης είναι σχεδόν ιδανικά και για τους δύο αλγόριθμους (Σχήμα 7.2). Οποιοσδήποτε αποκλίσεις παρατηρούνται οφείλονται σε σφάλματα υπολογισμού των χρωματικών μασκών των πλαισίων της ακολουθίας ή/και σε ατέλειες της χειροκίνητης επισημείωσης που πραγματοποιήθηκε για την εξαγωγή των πραγματικών πλαισίων οριοθέτησης (ground truth). Στα χρονικά διαγράμματα του Σχήματος 7.3 για την ακολουθία GSL10_6, η δυναμική εφαρμογή της τεχνικής αυτής είναι εμφανής: στα διαστήματα όπου η τεχνική είναι εφαρμόσιμη, το σφάλμα θέσης λαμβάνει τιμές κοντά στο μηδέν και, αντίστοιχα, τα μέτρα ομοιότητας της μάσκας κατάστασης και της χρωματικής μάσκας λαμβάνουν τιμές κοντά στη μονάδα και για τις δύο μεθόδους. Όπως αναφέραμε και προηγουμένως, η τεχνική αυτή δεν

είναι εφαρμόσιμη σε καμία χρονική στιγμή της ακολουθίας ASL.

Πείραμα 2. Τα διαγράμματα 7.5, 7.6 και 7.7 του δεύτερου πειράματος, επιβεβαιώνουν, σε κάποιο βαθμό, τη γενική προσδοκία μας για μια ευθεία σχέση ανάμεσα στο μέγεθος N_s του συνόλου των σταθμισμένων δειγμάτων (particle set) και την ποιότητα των αποτελεσμάτων παρακολούθησης. Συγκεκριμένα, σε όλες τις περιπτώσεις, παρατηρείται μια τάση μείωσης του μέσου σφάλματος θέσης και ταυτόχρονα μια αυξητική τάση για τις μέσες ομοιότητες της μάρσας κατάστασης και της χρωματικής μάρσας. Επιπλέον, παρατηρούμε τα ακόλουθα:

- Δεν παρατηρείται πρακτικά καμία βελτίωση της ακρίβειας των αποτελεσμάτων για πλήθος δειγμάτων (particles) μεγαλύτερο από 50, πράγμα που σημαίνει ότι η διακριτή προσέγγιση της θεωρούμενης συνάρτησης πυκνότητας $p(x_k|y_{1:k})$ έχει πρακτικά συγκλίνει στην πραγματική πυκνότητα για $N_s = 50$. Το γεγονός ότι το φαινόμενο αυτό παρατηρείται για μικρό, σχετικά, μέγεθος συνόλου σταθμισμένων δειγμάτων έχει να κάνει με το ότι το μεγαλύτερο μέρος της μάζας της $p(x_k|y_{1:k})$ είναι συγκεντρωμένο σε μικρού όγκου χωρίο του χώρου καταστάσεων.
- Το γεγονός ότι έχουμε αποδεκτά αποτελέσματα παρακολούθησης ακόμα για τις μικρότερες τιμές του πλήθους N_s που δοκιμάστηκαν συνιστά ένδειξη του ότι το θεωρούμενης μορφής μοντέλο μετάβασης, στο οποίο βασίζεται η μέθοδός μας, πράγματι αποτελεί καλή επιλογή κατανομής σημαντικότητας για το Particle Filter.

Κεφάλαιο 8

Πορίσματα της εργασίας

8.1 Συμβολή της εργασίας

Το βασικό προϊόν της παρούσας εργασίας είναι η πρόταση και υλοποίηση μιας μεθόδου οπτικής παρακολούθησης ενός αντικειμένου σε μία ακολουθία κινούμενων εικόνων. Η μέθοδός μας υλοποιεί ένα Particle Filter και εκμεταλλεύεται στοιχεία χρώματος και κίνησης που χαρακτηρίζουν τις εικόνες της ακολουθίας. Σε κάθε βήμα της ακολουθιακής εκτίμησης που πραγματοποιεί, η μέθοδός μας *μαθαίνει* ένα σύνολο παραμέτρων για το υποτιθέμενης μορφής μοντέλο μετάβασης του υπό παρακολούθηση συστήματος από ένα σύνολο ενδιάμεσων μεγεθών που προκύπτουν από ανάλυση του χρώματος και της κίνησης που χαρακτηρίζει/συνδέει το προηγούμενο και το τρέχον πλαίσιο.

Θυμίζουμε ότι η θεωρούμενη μορφή του μοντέλου μετάβασης, που ορίζει και την πυκνότητα σημαντικότητας του φίλτρου, αντιστοιχεί στην υπόθεση ότι η κίνηση του υπό παρακολούθηση αντικειμένου είναι *κατά κύριο λόγο* μεταφορική. Η μέθοδός μας δεν στηρίζεται σε κάποια περισσότερο περιοριστική υπόθεση για την κίνηση του αντικειμένου, το οποίο, για παράδειγμα, μπορεί να είναι παραμορφώσιμο και να εκτελεί σύνθετη κίνηση. Επιπλέον, καθώς δε βασιζόμαστε αποκλειστικά στο χρώμα, δεν πραγματοποιούμε κάποια υπόθεση για τη μορφή της χρωματικής μάσκας του αντικειμένου, η οποία προκύπτει από τη σχέση της κατανομής του χρώματός του με αυτήν του παρασκήνιου των εικόνων. Τέλος, η εφαρμογή της μεθόδου προϋποθέτει εκ των προτέρων γνώση ενός *χρωματικού μοντέλου* για το αντικείμενο (και, προαιρετικά, για το παρασκήνιο) και το *χειροκίνητο εντοπισμό* του αντικειμένου, μέσω ενός ορθογώνιου πλαισίου οριοθέτησης, στο πρώτο πλαίσιο της ακολουθίας.

Πιστεύουμε ότι τα περισσότερα αξιολογικά πλεονεκτήματα της προτεινόμενης μεθόδου, και, κατά συνέπεια, στοιχεία συμβολής της εργασίας αυτής, μπορούν να συνοψιστούν στα ακόλουθα σημεία:

- Σε θεωρητικό επίπεδο, όπως αναφέραμε προηγουμένως, η μόνη υπόθεση που πραγματοποιούμε είναι αυτή της *κατά κύριο λόγο* μεταφορικής κίνησης του υπό παρακολούθηση αντικειμένου. Το όχι ιδιαίτερα περιοριστικό σύνολο υποθέσεων στις οποίες στηρίζεται η προτεινόμενη μέθοδος, σε συνδυασμό με την πλούσια πληροφορία που παρέχουν από κοινού τα στοιχεία χρώματος και κίνησης που χρησιμοποιούμε, καθιστούν τη μέθοδό μας αρκετά *γενική* και εν δυνάμει εφαρμόσιμη σε ένα ευρύ φάσμα σεναρίων οπτικής παρακολούθησης.
- Στα πειράματα που πραγματοποιήσαμε στο Κεφάλαιο 7, η μέθοδός μας παρήγαγε ικανοποιητικής ακρίβειας, για τα θεωρούμενα μετρικά, αποτελέσματα παρακολούθησης. Καθώς στα

πειράματά μας περιλάβαμε πολλαπλής δυσκολίας σενάρια του προβλήματος, μπορούμε να συμπεράνουμε ότι η προτεινόμενη μέθοδος αποδεικνύεται *ικανή* στην πράξη για την επιτυχή παρακολούθηση ενός κινούμενου αντικειμένου σε ένα εύρος εφαρμογών.

- Η θεώρηση από το φίλτρο ενός *προσαρμοστικού* δυναμικού μοντέλου, μέσω της *μάθησης* των παραμέτρων του θεωρούμενης μορφής μοντέλου μετάβασης σε κάθε χρονικό βήμα, διευκολύνει σημαντικά την αναζήτηση στο χώρο των καταστάσεων. Η χρήση ενός “πληροφορημένου” δυναμικού μοντέλου περιορίζει την αναζήτηση σε ένα σύνολο περισσότερο εκ των προτέρων πιθανών καταστάσεων, καθιστώντας την πιο *αποδοτική* σε σχέση με αυτή που θα πραγματοποιούταν από ένα γενικότερο μοντέλο μετάβασης (για παράδειγμα, από ένα απλό μοντέλο σταθερής ταχύτητας). Άμεσες συνέπειες της παραπάνω ιδιότητας είναι ότι η προτεινόμενη μέθοδος λειτουργεί με επιτυχία χρησιμοποιώντας μόνο ένα μικρό αριθμό από δείγματα (particles) για την αναπαράσταση της υπό εκτίμηση συνάρτησης πυκνότητας και η μη απαίτηση ενός ιδιαίτερα εκλεπτυσμένου μοντέλου παρατήρησης για το στάδιο της διόρθωσης/ενημέρωσης βαρών στάθμισης των δειγμάτων (particle weights).
- Η προτεινόμενη μέθοδος είναι *ικανή* να ανιχνεύει, στο χρόνο εκτέλεσης, αν το πρόβλημα της παρακολούθησης λαμβάνει *ειδικές μορφές* που επιτρέπουν την *χωρίς αβεβαιότητα* επίλυσή του. Συγκεκριμένα, με τον τρόπο που περιγράψαμε στην ενότητα 6.5, ελέγχεται αν στο τρέχον χρονικό βήμα ο εντοπισμός του αντικειμένου μπορεί να γίνει μόνο με βάση το χρώμα και, στην περίπτωση όπου αυτό είναι εφικτό, πραγματοποιείται η αντίστοιχη εκτίμηση με μηδενική αβεβαιότητα. Ο έλεγχος για την ειδική αυτή περίπτωση έχει ενσωματωθεί σαν κομμάτι του δυναμικού μοντέλου του συστήματος που θεωρεί το φίλτρο και δεν οδηγεί σε παραβίαση του θεωρητικού μας πλαισίου.

8.2 Μελλοντικές κατευθύνσεις

Στην ενότητα αυτή, παραθέτουμε μια μη εξαντλητική λίστα από κατευθύνσεις μελλοντικής έρευνας για πιθανές επεκτάσεις και βελτιώσεις των περιορισμών της προτεινόμενης μεθόδου:

- Η μέθοδός μας στηρίζεται σε στοιχεία *χρώματος* και *κίνησης* για τον προσδιορισμό του μοντέλου μετάβασης και του μοντέλου παρατήρησης που υποθέτει. Χάρη στη γενικότητα του θεωρητικού πλαισίου που ακολουθεί, είναι εφικτή, κάτω από όχι ιδιαίτερα περιοριστικές υποθέσεις, η ενσωμάτωση με πιθανοτικό χαρακτήρα σε αυτήν και άλλων χαρακτηριστικών που μπορεί να περιέχουν πληροφορία χρήσιμη για το πρόβλημα της οπτικής παρακολούθησης. Μια τέτοια επέκταση θα επέτρεπε καλύτερη εκμετάλλευση από τον αλγόριθμο παρακολούθησης της γνώσης που μπορεί να έχουμε για το προς παρακολούθηση αντικείμενο, καθώς αυτός θα μπορεί να λειτουργεί βασιζόμενος σε ισχυρότερα, περισσότερο περιγραφικά μοντέλα του αντικειμένου από ό,τι η στατιστική κατανομή των χρωμάτων του (π.χ., μπορεί να έχουμε στη διάθεσή μας ένα bag of words μοντέλο από περιγραφείς SIFT τοπικών χαρακτηριστικών γνωρισμάτων [27] για το αντικείμενό μας).
- Η προτεινόμενη μέθοδος προϋποθέτει τη γνώση ενός χρωματικού μοντέλου για το προς παρακολούθηση αντικείμενο. Στα πειράματά που πραγματοποιήσαμε, η εκπαίδευση του μοντέλου αυτού έγινε πάνω σε δεδομένα που προέκυψαν από offline, χειροκίνητη επισήμειωση του αντικειμένου σε κάποια από τα πλαίσια της ακολουθίας βίντεο που θα αποτελούσε την είσοδο του αλγόριθμου μας. Η ανάγκη για εκ των προτέρων γνώση ενός ακριβούς χρωματικού μοντέλου για το αντικείμενο μπορεί να εξαλειφθεί αν χρησιμοποιούσαμε αρχικά ένα

αδρό, γενικό μοντέλο χρώματος για αυτό και στη συνέχεια εφαρμόζαμε, στο χρόνο εκτέλεσης του αλγορίθμου μας (π.χ., σε ένα ξεχωριστό νήμα εκτέλεσης), τεχνικές *δυναμικής ενημέρωσης/προσαρμογής* (online adaptation) του χρωματικού μοντέλου με βάση κάποια από τα πρόσφατα ακολουθιακά δεδομένα. Ένα αρχικό μοντέλο μπορεί να κατασκευαστεί από την επισημείωση του αντικείμενου στο πρώτο πλαίσιο της ακολουθίας. Προς διευκόλυνση της διαδικασίας, σε κάποιες ειδικές περιπτώσεις μπορούμε να αποκτήσουμε μια “ακατέργαστη” χρωματική μάσκα για το αντικείμενο χρησιμοποιώντας κάποια πρότερη, γενική γνώση για αυτό. Για παράδειγμα, σε ένα σενάριο παρακολούθησης χεριού, μπορεί να χρησιμοποιηθεί ο αλγόριθμος 1 για την εξαγωγή μιας αδρής μάσκας των περιοχών της εικόνας που απεικονίζουν ανθρώπινο δέρμα και το αρχικό μοντέλο μπορεί να εκπαιδευθεί από τα δεδομένα που υποδεικνύονται από το συνδυασμό της μάσκας αυτής και της επισημείωσης στο πρώτο πλαίσιο. Θυμίζουμε ότι την ίδια διαδικασία ακολουθήσαμε, σε ένα offline πλαίσιο, για την εκπαίδευση των χρωματικών μοντέλων στα πειράματα που πραγματοποιήσαμε σε ακολουθίες βίντεο νοηματικής γλώσσας. Μια τεχνική online προσαρμογής μειγμάτων Γκαουσιανών κατανομών (εκτίμηση παραμέτρων μέγιστης ύστερης πιθανότητας, MAP estimation) μπορεί να βρεθεί στο [79], ενώ μια σχετική εφαρμογή παρουσιάζεται στο [80].

- Η μέθοδός μας στηρίζεται σε μια ιδιαίτερα απλή μορφή κατάτμησης με βάση την κίνηση, η οποία προκύπτει άμεσα από την ταξινόμηση (clustering) των διανυσμάτων κίνησης. Στην υλοποίησή μας, χρησιμοποιήσαμε τη μέθοδο ταξινόμησης k -means, η οποία προϋποθέτει εκ των προτέρων γνώση του πλήθους των κλάσεων (ομάδων) της κατάτμησης. Η ανάγκη αυτή μπορεί να εξαλειφθεί αν αντί του αλγορίθμου k -means υιοθετήσουμε κάποια μέθοδο φασματικής ταξινόμησης (spectral clustering) [65], εκτιμώντας το πλήθος των κλάσεων με βάση το κριτήριο του μέγιστου χάσματος ιδιοτιμών (maximum eigengap), ή τον αλγόριθμο ταξινόμησης affinity propagation [81], ο οποίος δεν απαιτεί τη λήψη κάποιας απόφασης για το πλήθος των κλάσεων.
- Η μέθοδος που προτείνουμε, όντας online και ακολουθιακή, μπορεί επί της αρχής να υλοποιηθεί με τρόπο που να επιτρέπει την εφαρμογή της σε σενάρια παρακολούθησης *πραγματικού χρόνου* (real-time tracking). Το υπολογιστικά απαιτητικότερο τμήμα της μεθόδου μας είναι ο υπολογισμός της (προς τα εμπρός και προς τα πίσω) οπτικής ροής ανάμεσα σε διαδοχικά πλαίσια της ακολουθίας εισόδου. Παρότι η παρούσα υλοποίησή μας στηρίζεται σε μια βελτιστοποιημένη βιβλιοθήκη σε C++ [78] για αυτούς τους υπολογισμούς, η εκτίμηση της οπτικής ροής που συνδέει δύο εικόνες απαιτεί χρόνο της τάξης μερικών δεκάδων δευτερολέπτων σε έναν τυπικό σύγχρονο υπολογιστή. Μια ιδιαίτερα υποσχόμενη λύση είναι η υλοποίηση του υπολογισμού της οπτικής ροής σε GPU (μαζικά παράλληλος επεξεργαστής γραφικών). Ενδεικτικές επιτυχημένες τέτοιες προσεγγίσεις παρουσιάζονται στις εργασίες [39, 82], ενώ ένας αριθμός σχετικών δημοσιεύσεων και διαθέσιμων υλοποιήσεων μπορεί να βρεθεί στο [83].
- Πλήρης αυτοματοποίηση της μεθόδου μέσω του αυτόματου εντοπισμού του προς παρακολούθηση αντικείμενου στο πρώτο πλαίσιο της ακολουθίας βίντεο. Αυτό μπορεί να γίνει εύκολα σε ειδικές περιπτώσεις (για παράδειγμα, μέσω της χρωματικής μάσκας του αντικείμενου, αν η κατανομή των χρωμάτων του δεν επικαλύπτει αυτήν του παρασκήνιου) ή μπορεί να απαιτεί πρότερη γνώση ενός ισχυρότερου, περισσότερο περιγραφικού μοντέλου για το αντικείμενο.

Το θεωρητικό πλαίσιο του Bayesian tracking/filtering και συγκεκριμένα τα Particle filters προσφέρουν μια ισχυρή και γενική βάση για την προσέγγιση του προβλήματος της οπτικής παρακολούθησης. Πιστεύουμε ότι όλες οι παραπάνω προτάσεις είναι εφικτές και ότι η ενσωμάτωσή τους στη μέθοδο που αναπτύξαμε ή η τροποποίηση της τελευταίας σύμφωνα με αυτές θα οδηγήσει σε μία αποδοτικότερη, γενικότερη και καλύτερων επιδόσεων μέθοδο οπτικής παρακολούθησης με ένα αξιολογικά ευρύ φάσμα πιθανών εφαρμογών.

Παράρτημα Α

Ταξινόμηση k -means

Η ταξινόμηση k -means αποτελεί την πιο δημοφιλή μέθοδο “μη εποπτευόμενης” οργάνωσης δεδομένων σε ομάδες (unsupervised clustering). Έστω $\{x_i : i = 1, \dots, n\} \subset \mathbb{R}^d$ ένα σύνολο από n το πλήθος σημεία σε ένα d -διάστατο Ευκλείδειο χώρο. Στόχος μας είναι η διαμέριση του συνόλου σε k ομάδες (clusters, “συμπλέγματα”), έτσι ώστε η Ευκλείδεια απόσταση ενός σημείου από άλλα σημεία που ανήκουν στην ίδια ομάδα να είναι μικρή σχετικά με την απόστασή του από σημεία εκτός αυτής (που ανήκουν σε άλλες ομάδες). Το μέγεθος k της διαμέρισης (πλήθος των ομάδων) θεωρείται εξαρχής γνωστό.

Για τον τυπικό ορισμό της βελτιστότητας της διαμέρισης που αναζητούμε, ορίζουμε τις παρακάτω ποσότητες:

- Έστω $\mu_j \in \mathbb{R}^d$, $j = 1, \dots, k$, το κεντροειδές της ομάδας j : το μ_j αναπαριστά το κέντρο της ομάδας j στον \mathbb{R}^d και λειτουργεί ως πρωτότυπο (αντιπρόσωπος) για αυτή.
- Έστω $r_{ij} \in \{0, 1\}$, $i = 1, \dots, n$, $j = 1, \dots, k$, ένα σύνολο δυαδικών μεταβλητών, ενδεικτικών του σε ποια ομάδα ανήκει κάθε σημείο: το σημείο x_i ανήκει στην ομάδα j ακριβώς όταν $r_{ij} = 1$ και $r_{im} = 0$ για κάθε $m \neq j$.

Με βάση τα παραπάνω, στόχος μας είναι η ελαχιστοποίηση της αντικειμενικής συνάρτησης:

$$J = \sum_{i=1}^n \sum_{j=1}^k r_{ij} \|x_i - \mu_j\|^2, \quad (\text{A.1})$$

υπό τους περιορισμούς για τις μεταβλητές ανάθεσης r_{ij} που αναφέρθηκαν. Η παραπάνω ποσότητα αναπαριστά το άθροισμα των τετραγώνων των αποστάσεων κάθε σημείου από το κεντροειδές της ομάδας στην οποία ανατίθεται (within-cluster sum of squares). Το ζητούμενο είναι, συνεπώς, ο προσδιορισμός τιμών για τα μ_j και r_{ij} , για $i = 1, \dots, n$ και $j = 1, \dots, k$, για τις οποίες η J γίνεται ελάχιστη. Στη γενική περίπτωση, το πρόβλημα της ελαχιστοποίησης της J είναι NP-hard.

Για την ελαχιστοποίηση της J , στην πράξη χρησιμοποιείται μια απλή επαναληπτική μέθοδος, η οποία περιλαμβάνει την ανάθεση αρχικών τιμών (με αυθαίρετο τρόπο) στα μ_j και κάθε επανάληψη της οποίας συνίσταται από δύο βήματα. Στο πρώτο βήμα, γίνεται ελαχιστοποίηση της J ως προς τα r_{ij} , με τα μ_j σταθερά. Στο δεύτερο βήμα, γίνεται ελαχιστοποίηση της J ως προς τα μ_j , με τα r_{ij} σταθερά. Το σχήμα αυτό είναι γνωστό ως ο αλγόριθμος του Lloyd και αποτελεί στιγμιότυπο του αλγορίθμου EM (Παράρτημα Β), με το πρώτο βήμα (ενημέρωση των αναθέσεων r_{ij}) να αντιστοιχεί στο Ε στάδιο και το δεύτερο (ενημέρωση των κεντροειδών μ_j) στο Μ στάδιο.

Για το πρώτο βήμα του αλγορίθμου (Ε στάδιο), από τη μορφή της [A.1](#) και τους περιορισμούς μας για τις δυαδικές μεταβλητές ανάθεσης r_{ij} , είναι εύκολο να δούμε ποιες τιμές των τελευταίων

ελαχιστοποιούν την J για σταθερά κεντροειδή μ_j . Για κάθε $i = 1, \dots, n$, θέτουμε $r_{ij} = 1$ για την ομάδα j της οποίας το κεντροειδές μ_j είναι πιο κοντά στο σημείο x_i :

$$r_{ij} = \begin{cases} 1 & \text{αν } j = \operatorname{argmin}_{m=1, \dots, k} \|x_i - \mu_m\|^2 \\ 0 & \text{διαφορετικά} \end{cases} \quad (\text{A.2})$$

Για το δεύτερο βήμα του αλγορίθμου (M στάδιο), αναζητούμε τις βέλτιστες τιμές για τα μ_j , με τα r_{ij} σταθερά. Θέτοντας την παράγωγο της A.1 (τετραγωνική συνάρτηση των μ_j) στο 0, λαμβάνουμε:

$$2 \sum_{i=1}^n r_{ij}(x_i - \mu_j) = 0 \Rightarrow \mu_j = \frac{\sum_{i=1}^n r_{ij}x_i}{\sum_{i=1}^n r_{ij}} \quad (\text{A.3})$$

Σημειώνουμε ότι το κεντροειδές της ομάδας j , $j = 1, \dots, k$, δεν είναι παρά η μέση τιμή των σημείων που έχουν ανατεθεί σε αυτή: στην παραπάνω έκφραση, ο αριθμητής ισούται με το άθροισμα των σημείων που ανήκουν στην ομάδα j και ο παρονομαστής είναι απλώς το πλήθος τους. Μια συνολική περιγραφή του επαναληπτικού σχήματος δίνεται στον αλγόριθμο 12.

Algorithm 12 k -means clustering (Lloyd's algorithm)

Input: $\{x_i : i = 1, \dots, n\}$, k

Output: $\{\mu_j : j = 1, \dots, k\}$, $\{r_{ij} : i = 1, \dots, n, j = 1, \dots, k\}$

1: Initialize μ_j , $j = 1, \dots, k$

2: **repeat**

3: /* **E step:** cluster assignments */

4: **for** $i = 1 : n$ **do**

5: Assign $r_{ij} = \begin{cases} 1 & \text{if } j = \operatorname{argmin}_{m=1, \dots, k} \|x_i - \mu_m\|^2 \\ 0 & \text{otherwise} \end{cases}$

6: **end for**

7: /* **M step:** centroid updates */

8: **for** $j = 1 : k$ **do**

9: Update $\mu_j = \frac{\sum_{i=1}^n r_{ij}x_i}{\sum_{i=1}^n r_{ij}}$

10: **end for**

11: **until** the assignment variables r_{ij} do not change

Καθώς κάθε βήμα, σε οποιαδήποτε επανάληψη, μπορεί μόνο να ελαττώσει την τιμή της αντικειμενικής συνάρτησης J , η σύγκλιση της μεθόδου είναι εγγυημένη. Παρόλα αυτά, η βελτιστότητα της επιστρεφόμενης λύσης δεν είναι εγγυημένη, καθώς ο αλγόριθμος μπορεί να συγκλίνει σε κάποιο τοπικό ελάχιστο της J , ανάλογα με την αρχικοποίηση των μ_j που πραγματοποιείται. Μια συνήθης πρακτική για τη μετρίαση του φαινομένου είναι η πολλαπλή εκτέλεση του αλγορίθμου, με διαφορετική αρχικοποίηση κάθε φορά, και διατήρηση της λύσης που φέρει τη μικρότερη τιμή για την J .

Παράρτημα Β

Ο αλγόριθμος EM για μείγματα Γκαουσιανών κατανομών

B.1 Μείγματα Γκαουσιανών κατανομών

Ένα μείγμα από Γκαουσιανές κατανομές (Gaussian Mixture Model, GMM) είναι μία συνεχής κατανομή πιθανότητας, η πυκνότητα της οποίας είναι μια γραμμική υπέρθεση (συγκεκριμένα, ένας κυρτός συνδυασμός) Γκαουσιανών πυκνοτήτων. Οι γραμμικοί συνδυασμοί Γκαουσιανών πυκνοτήτων ορίζουν μια ιδιαίτερα πλούσια κλάση συναρτήσεων πυκνοτήτων πιθανότητας. Ενσωματώνοντας επαρκή αριθμό Γκαουσιανών πυκνοτήτων με κατάλληλους συντελεστές μείξης και κατάλληλες παραμέτρους (μέσες τιμές και πίνακες συμμεταβλητότητας), σχεδόν οποιαδήποτε συνεχής πυκνότητα πιθανότητας μπορεί να προσεγγισθεί με αυθαίρετα μεγάλη ακρίβεια.

Η πυκνότητα πιθανότητας ενός μείγματος k Γκαουσιανών κατανομών έχει τη μορφή:

$$p(x) = \sum_{j=1}^k \pi_j \mathcal{N}(x; \mu_j, \Sigma_j) \quad (\text{B.1})$$

Κάθε Γκαουσιανή πυκνότητα $\mathcal{N}(x; \mu_j, \Sigma_j)$, $j = 1, \dots, k$, είναι μια *συνιστώσα* του μείγματος και χαρακτηρίζεται από τη δική της μέση τιμή μ_j και πίνακα συμμεταβλητότητας Σ_j . Οι παράμετροι π_j , $j = 1, \dots, k$, στην B.1 ονομάζονται *συντελεστές μείξης* και ικανοποιούν:

$$\sum_{j=1}^k \pi_j = 1 \quad (\text{B.2})$$

$$0 \leq \pi_j \leq 1, \text{ για κάθε } j = 1, \dots, k \quad (\text{B.3})$$

Η B.1 μπορεί να γραφεί σαν περιθώρια κατανομή της $p(x, z)$, όπου $z = (z_1, \dots, z_k) \in \{0, 1\}^k$ είναι ένα διάνυσμα δυαδικών μεταβλητών από τις οποίες μόνο μία συνιστώσα z_j είναι ίση με 1 και όλες οι υπόλοιπες είναι ίσες με 0. Καθώς υπάρχουν k διακριτές καταστάσεις για την κρυφή μεταβλητή z , που αντιστοιχούν στο ποια συνιστώσα z_j αυτής είναι μη μηδενική, η $p(x)$ εκφράζεται σαν περιθώρια της $p(x, z)$ ως ακολούθως:

$$p(x) = \sum_{j=1}^k p(x, z_j = 1) = \sum_{j=1}^k p(x|z_j = 1)P(z_j = 1) \quad (\text{B.4})$$

Η κρυφή τυχαία μεταβλητή z προσδιορίζει την Γκαουσιανή συνιστώσα (από τις k) που παρήγαγε την παρατήρηση x . Αντιπαραβάλλοντας τις B.1 και B.4, για $j = 1, \dots, k$, έχουμε:

$$\pi_j = P(z_j = 1)$$

$$p(x|z_j = 1) = \mathcal{N}(x; \mu_j, \Sigma_j)$$

Από τις παραπάνω προκύπτει μια απλή ερμηνεία των συντελεστών μείξης: οι τελευταίοι ορίζουν την περιθώρια κατανομή $P(z)$ της $p(x, z)$ και, πιο συγκεκριμένα, ο συντελεστής π_j ερμηνεύεται ως η *πρότερη* πιθανότητα μια παρατήρηση x να προέρχεται από τη συνιστώσα j του μείγματος.

Μια άλλη χρήσιμη ποσότητα για τη διατύπωση του αλγορίθμου EM για μείγματα Γκαουσιανών κατανομών είναι η πιθανότητα $P(z|x)$. Συμβολίζοντας με γ_j την πιθανότητα $P(z_j = 1|x)$ και χρησιμοποιώντας το θεώρημα του Bayes, έχουμε:

$$\gamma_j = P(z_j = 1|x) = \frac{P(z_j = 1)p(x|z_j = 1)}{p(x)} = \frac{\pi_j \mathcal{N}(x; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x; \mu_m, \Sigma_m)}$$

Η ποσότητα γ_j είναι η *ύστερη* πιθανότητα η παρατήρηση x να προέρχεται από τη συνιστώσα j του μείγματος (εφόσον έχει παρατηρηθεί η τιμή της x) και ονομάζεται *υπευθυνότητα* της συνιστώσας j για την ερμηνεία της παρατήρησης x .

B.2 Εκτίμηση παραμέτρων μέγιστης πιθανοφάνειας

Έστω $X = \{x_i : i = 1, \dots, n\} \subset \mathbb{R}^d$ ένα σύνολο παρατηρήσεων που υποθέτουμε ότι αποτελείται από ανεξάρτητα δείγματα (i.i.d.) από ένα μείγμα Γκαουσιανών κατανομών με k συνιστώσες. Αν $\pi = \{\pi_j : j = 1, \dots, k\}$, $\mu = \{\mu_j : j = 1, \dots, k\}$ και $\Sigma = \{\Sigma_j : j = 1, \dots, k\}$ είναι οι παράμετροι του μοντέλου (συντελεστές μείξης, μέσες τιμές και πίνακες συμμεταβλητότητας των Γκαουσιανών συνιστωσών, αντίστοιχα), η πιθανοφάνεια (likelihood) του δείγματος, με βάση την υπόθεση ανεξαρτησίας μας, δίνεται από:

$$p(X; \pi, \mu, \Sigma) = \prod_{i=1}^n \left\{ \sum_{j=1}^k \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j) \right\} \quad (\text{B.5})$$

Το ζητούμενο είναι η εκτίμηση των παραμέτρων π , μ και Σ που μεγιστοποιούν την πιθανοφάνεια B.5 του δείγματος X (εκτιμήτριες μέγιστης πιθανοφάνειας, Maximum Likelihood estimates), ή, ισοδύναμα, το λογάριθμο αυτής (log-likelihood):

$$\ln p(X; \pi, \mu, \Sigma) = \sum_{i=1}^n \ln \left\{ \sum_{j=1}^k \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j) \right\} \quad (\text{B.6})$$

Προσαρμόζοντας τους συμβολισμούς της προηγούμενης ενότητας, έστω $z_i = (z_{i1}, \dots, z_{ik}) \in \{0, 1\}^k$ η κρυφή τυχαία μεταβλητή που προσδιορίζει την Γκαουσιανή συνιστώσα του μείγματος που παρήγαγε την παρατήρηση x_i : για κάθε $i = 1, \dots, n$, ακριβώς μία συνιστώσα z_{ij} της z_i είναι ίση με 1 και όλες οι υπόλοιπες είναι ίσες με 0. Η υπευθυνότητα γ_{ij} της j -οστής συνιστώσας του μείγματος για την παρατήρηση x_i τότε γράφεται:

$$\gamma_{ij} = P(z_{ij} = 1|x_i) = \frac{P(z_{ij} = 1)p(x_i|z_{ij} = 1)}{p(x_i)} = \frac{\pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x_i; \mu_m, \Sigma_m)} \quad (\text{B.7})$$

Ορίζουμε, επιπλέον, το μέγεθος:

$$N_j = \sum_{i=1}^n \gamma_{ij},$$

το οποίο ερμηνεύεται σαν ο ενεργός αριθμός σημείων που “ανατίθενται” στη συνιστώσα j .

Οι τιμές των μ_j , $j = 1, \dots, k$, που μεγιστοποιούν την B.6 απαιτείται να μηδενίζουν την παράγωγο της τελευταίας, από όπου, χρησιμοποιώντας τους παραπάνω ορισμούς και υποθέτοντας ότι οι πίνακες συμμεταβλητότητας όλων των συνιστωσών είναι αντιστρέψιμοι, λαμβάνουμε, για $j = 1, \dots, k$:

$$\sum_{i=1}^n \frac{\pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x_i; \mu_m, \Sigma_m)} \Sigma_j^{-1} (x_i - \mu_j) = 0 \Rightarrow \mu_j = \frac{1}{N_j} \sum_{i=1}^n \gamma_{ij} x_i \quad (\text{B.8})$$

Παρατηρούμε ότι η μέση τιμή μ_j της j -οστής συνιστώσας του μείγματος είναι ένας σταθμισμένος μέσος όλων των δειγμάτων του X , όπου ο συντελεστής στάθμισης του σημείου x_i δεν είναι παρά η ύστερη πιθανότητα γ_{ij} η συνιστώσα j να είναι υπεύθυνη για την παραγωγή της παρατήρησης x_i . Απαιτώντας την ικανοποίηση της ίδιας συνθήκης βελτιστότητας για τους πίνακες συμμεταβλητότητας Σ_j , για $j = 1, \dots, k$, λαμβάνουμε:

$$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^n \gamma_{ij} (x_i - \mu_j)(x_i - \mu_j)^T \quad (\text{B.9})$$

Το αποτέλεσμα θυμίζει και πάλι την εκτιμήτρια μέγιστης πιθανοφάνειας για τον πίνακα συμμεταβλητότητας μιας απλής Γκαουσιανής κατανομής, με τη διαφορά ότι και πάλι η συνεισφορά κάθε σημείου σταθμίζεται από την αντίστοιχη ύστερη πιθανότητα γ_{ij} και ότι ο παρονομαστής αφορά το ενεργό πλήθος σημείων που αντιστοιχούν στην εκάστοτε συνιστώσα. Για την εύρεση των βέλτιστων συντελεστών μείξης του μείγματος, πρέπει να λάβουμε υπόψη τον περιορισμό κυρτότητας B.2 που διατυπώσαμε. Ο μηδενισμός της παραγώγου της Λαγκρανζιανής:

$$L = \ln p(X; \pi, \mu, \Sigma) + \lambda \left(\sum_{j=1}^k \pi_j - 1 \right)$$

δίνει, για $j = 1, \dots, k$:

$$\sum_{i=1}^n \frac{\mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x_i; \mu_m, \Sigma_m)} + \lambda = 0 \Leftrightarrow \sum_{i=1}^n \frac{\pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x_i; \mu_m, \Sigma_m)} + \pi_j \lambda = 0$$

Αθροίζοντας τις παραπάνω για $j = 1, \dots, k$ και κάνοντας χρήση της B.2, παίρνουμε ότι $\lambda = -N$ και από την παραπάνω, για $j = 1, \dots, k$, έχουμε:

$$\sum_{i=1}^n \gamma_{ij} - \pi_j N = 0 \Rightarrow \pi_j = \frac{N_j}{N} \quad (\text{B.10})$$

Σημειώνουμε ότι οι συνθήκες βελτιστότητας B.8, B.9 και B.10 δεν οδηγούν σε κάποια κλειστή λύση για τις παραμέτρους του μείγματος, καθώς οι υπευθυνότητες γ_{ij} εξαρτώνται με περίπλοκο τρόπο από αυτές (εξίσωση B.7). Παρόλα αυτά, τα παραπάνω αποτελέσματα υποδεικνύουν την

εφαρμογή μιας απλής επαναληπτικής μεθόδου για την εύρεση των εκτιμητριών μέγιστης πιθανοφάνειας για ένα μείγμα Γκαουσιανών κατανομών, η οποία προκύπτει ότι αποτελεί ένα στιγμιότυπο του γενικού αλγορίθμου EM (Expectation Maximization) [84]. Η μέθοδος περιλαμβάνει την (εν γένει, αυθαίρετη) αρχικοποίηση των παραμέτρων του μείγματος (μ , Σ , π) και κάθε επανάληψη της συνίσταται από δύο βήματα: στο E (Expectation) βήμα, χρησιμοποιούνται οι τρέχουσες τιμές των παραμέτρων του μείγματος για τον υπολογισμό των υπευθυνοτήτων (B.7) και στο M (Maximization) βήμα, οι τιμές αυτές χρησιμοποιούνται για την επανεκτίμηση των παραμέτρων του μείγματος σύμφωνα με τις B.8, B.9 και B.10. Συγκεκριμένα, η ενημέρωση των πινάκων συμμεταβλητότητας B.9 γίνεται με βάση τις νέες μέσες τιμές από την B.8. Μια συνολική περιγραφή του επαναληπτικού αυτού σχήματος, το οποίο αποτελεί τον αλγόριθμο EM για μείγματα Γκαουσιανών κατανομών, δίνεται στον αλγόριθμο 13.

Algorithm 13 EM algorithm for Gaussian mixtures

Input: $\{x_i : i = 1, \dots, n\}$, k

Output: $\{(\mu_j, \Sigma_j) : j = 1, \dots, k\}$, $\{\pi_j : j = 1, \dots, k\}$

- 1: Initialize μ_j , Σ_j and π_j for $j = 1, \dots, k$ and evaluate the initial value of the log-likelihood according to B.6
 - 2: **repeat**
 - 3: /* **E step:** calculate responsibilities for current parameter values */
 - 4: **for** $i = 1 : n$ **do**
 - 5:
$$\gamma_{ij} = \frac{\pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{m=1}^k \pi_m \mathcal{N}(x_i; \mu_m, \Sigma_m)}$$
 - 6: **end for**
 - 7: /* Calculate effective number of points assigned to each cluster */
 - 8: **for** $j = 1 : k$ **do**
 - 9:
$$N_j = \sum_{i=1}^n \gamma_{ij}$$
 - 10: **end for**
 - 11: /* **M step:** re-estimate the parameters for the current responsibilities */
 - 12: **for** $j = 1 : k$ **do**
 - 13:
$$\mu_j = \frac{1}{N_j} \sum_{i=1}^n \gamma_{ij} x_i$$
 - 14:
$$\Sigma_j = \frac{1}{N_j} \sum_{i=1}^n \gamma_{ij} (x_i - \mu_j)(x_i - \mu_j)^T$$
 - 15:
$$\pi_j = \frac{N_j}{N}$$
 - 16: **end for**
 - 17: Evaluate the value of the log-likelihood for the updated parameters according to B.6
 - 18: **until** convergence of either the parameters or the log-likelihood is achieved
-

Μπορεί εύκολα ναδειχθεί ότι κάθε ενημέρωση των παραμέτρων του μείγματος που πραγματοποιείται από ένα M βήμα (το οποίο έπεται ενός E βήματος) πάντοτε προκαλεί μονότονη αύξηση της λογαριθμικής πιθανοφάνειας. Κατά συνέπεια, η σύγκλιση του αλγορίθμου σε κάποιο τοπικό μέγιστο της πιθανοφάνειας είναι εγγυημένη. Στην πράξη, ο αλγόριθμος θεωρείται ότι έχει συγκλίνει αν κάποιο μέτρο μεταβολής της λογαριθμικής πιθανοφάνειας ή των υπό εκτίμηση παραμέτρων του μείγματος ανάμεσα σε δύο διαδοχικές επαναλήψεις είναι μικρότερο από κάποιο κατώφλι.

Η σύγκλιση του αλγορίθμου EM απαιτεί, εν γένει, πολύ περισσότερες επαναλήψεις σε σύγκριση με την εφαρμογή του αλγορίθμου 12 για την ταξινόμηση k -means στο ίδιο σύνολο δε-

δομένων και κάθε επανάληψη είναι υπολογιστικά απαιτητικότερη [85]. Για την επιτάχυνση της σύγκλισης του EM, μια συνήθης πρακτική είναι η αρχικοποίηση των παραμέτρων του μείγματος με βάση το αποτέλεσμα της ταξινόμησης k -means: κάθε ομάδα της διαμέρισης αντιστοιχίζεται σε μια Γκαουσιανή συνιστώσα του μείγματος και η μέση τιμή της συνιστώσας τίθεται ίση με το κεντροειδές της ομάδας, ο πίνακας συμμεταβλητότητας της συνιστώσας ίσος με τη συμμεταβλητότητα δείγματος των σημείων που ανήκουν στην ομάδα και ο συντελεστής μείξης της συνιστώσας στο κλάσμα των σημείων που ανατέθηκαν στην ομάδα.

Σημειώνουμε ότι η ταξινόμηση k -means μπορεί να διατυπωθεί σαν το όριο ενός ειδικού στιγμιότυπου του αλγορίθμου EM για μείγματα Γκαουσιανών. Συγκεκριμένα, αν όλες οι συνιστώσες ενός μείγματος έχουν κοινό πίνακα συμμεταβλητότητας, ίσο με ϵI (όπου I ο $d \times d$ μοναδιαίος πίνακας), με το ϵ να είναι μια σταθερά και όχι παράμετρος προς εκτίμηση, είναι εύκολο ναδειχθεί [85] ότι καθώς $\epsilon \rightarrow 0$, $\gamma_{ij} \rightarrow r_{ij}$, με τα r_{ij} όπως ορίζονται από την Α.2. Με άλλα λόγια, για αυτή την ειδική μορφή μείγματος και στο όριο $\epsilon \rightarrow 0$, οι χαλαρές αναθέσεις (soft assignments) σε συνιστώσες που ορίζουν οι υπευθυνότητες γ_{ij} μετατρέπονται στις απόλυτες αναθέσεις (hard assignments) σε ομάδες της ταξινόμησης k -means.

Παράρτημα C

Ενδεικτικά ενδιάμεσα αποτελέσματα και μεγέθη

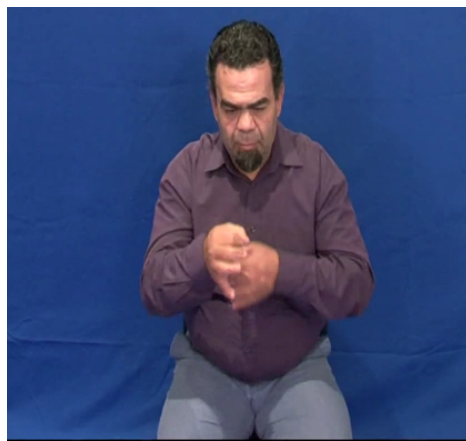
Στο παράρτημα αυτό, παραθέτουμε ένα σύνολο απεικονίσεων βοηθητικών μεγεθών και ενδιάμεσων αποτελεσμάτων, προκειμένου να παρέχουμε μια καλύτερη διαίσθηση για τη λειτουργία των αλγορίθμων που αναπτύξαμε. Οι απεικονίσεις μας αποτελούν παραδείγματα για τη μορφή που έχουν στην πράξη κάποια από τα μεγέθη που ορίζουμε στο Κεφάλαιο 6 και βασίζονται στην ακολουθία βίντεο `GSL10_6`, μια από τις ακολουθίες που χρησιμοποιήσαμε για την πειραματική αξιολόγηση των αλγορίθμων μας στο Κεφάλαιο 7. Μια σύντομη περιγραφή της ακολουθίας αυτής μπορεί να βρεθεί στην ενότητα 7.1.

C.1 Μεγέθη χρώματος και κίνησης

Στην ενότητα αυτή απεικονίζουμε κάποια από τα μεγέθη που εισάγαμε στην ενότητα 6.2 για την ακολουθία της δοκιμής μας. Επικεντρώνουμε το ενδιαφέρον μας στα πλαίσια 48 και 49 της ακολουθίας: τα ενδιάμεσα αποτελέσματα που θα παρουσιάσουμε θα χρησιμοποιηθούν για την εκτίμηση της κατάστασης τη χρονική στιγμή $k = 49$. Ο συμβολισμός που χρησιμοποιούμε και ο τρόπος υπολογισμού κάθε οντότητας έχουν περιγραφεί με λεπτομέρεια στην ενότητα 6.2.

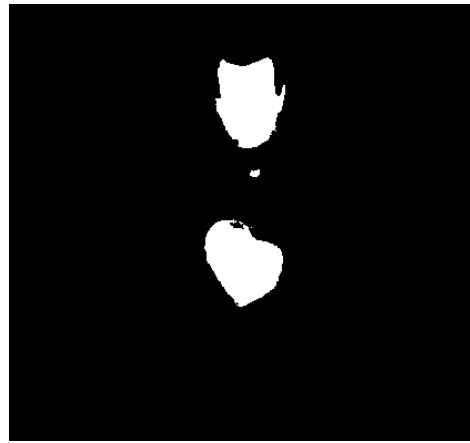
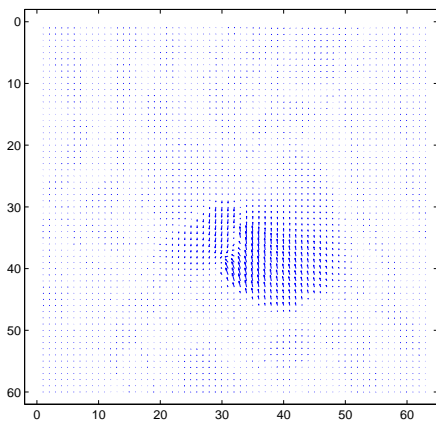
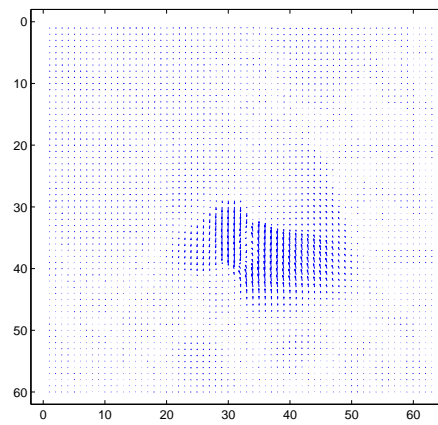
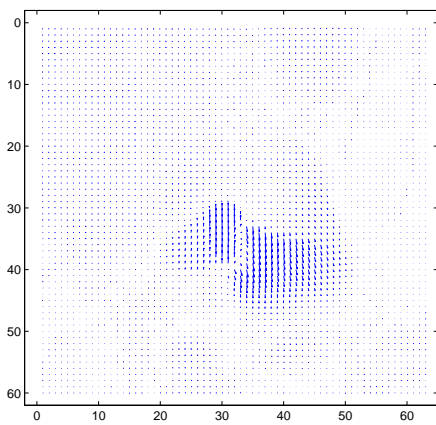
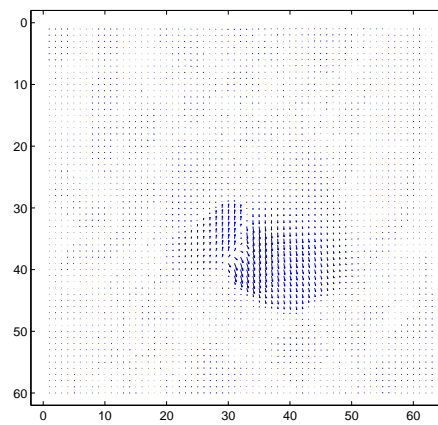


(a) I_{48}

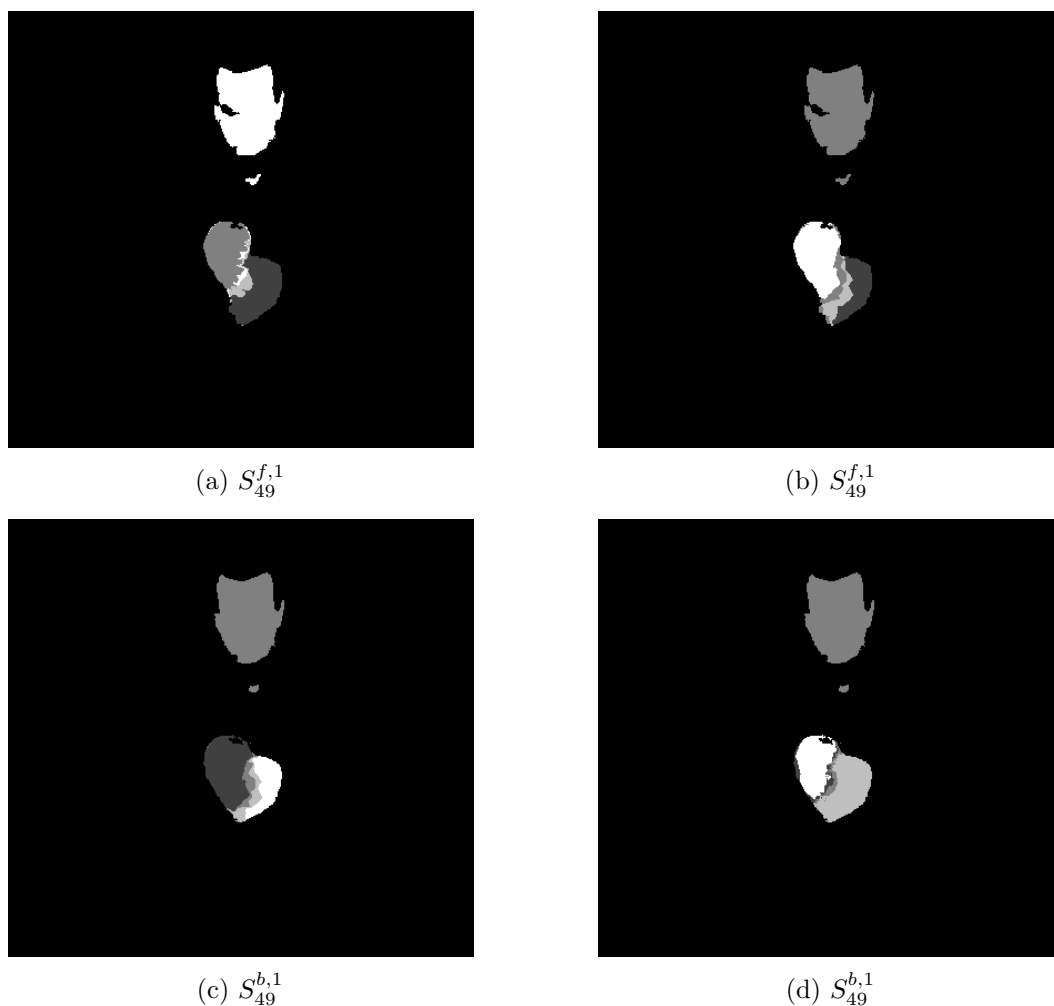


(b) I_{49}

Σχήμα C.1: Διαδοχικά πλαίσια της ακολουθίας τις στιγμές $k = 48$ και $k = 49$.

(a) B_{48}^c (b) B_{49}^c Σχήμα C.2: Χρωματικές μάσκες των I_{48} και I_{49} .(a) $f_{49}^{f,1}$ (b) $f_{49}^{f,1}$ (c) $f_{49}^{b,1}$ (d) $f_{49}^{b,1}$

Σχήμα C.3: Πεδία οπτικής ροής που συνδέουν τις I_{48} και I_{49} (δύο εκδοχές ανά κατεύθυνση). Για τις ανάγκες της απεικόνισης, έχει πραγματοποιηθεί υποδειγματοληψία στο 15% του αρχικού μεγέθους.



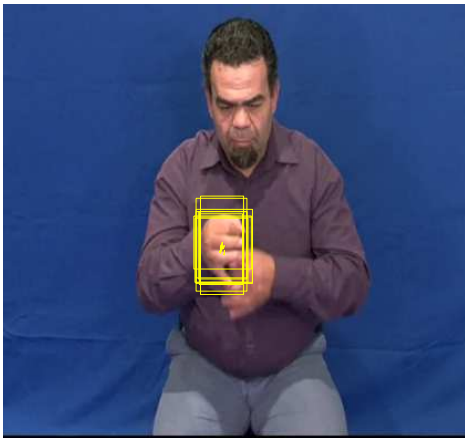
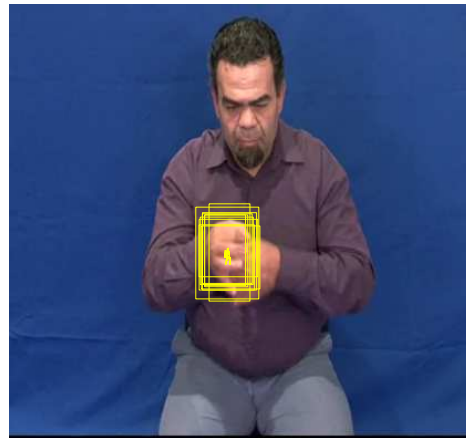
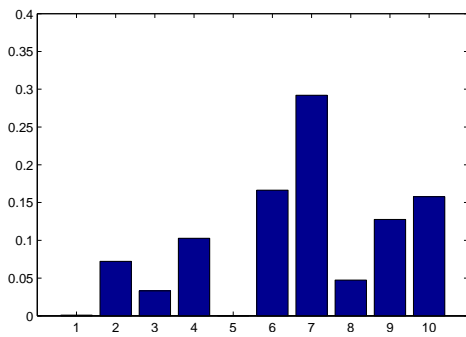
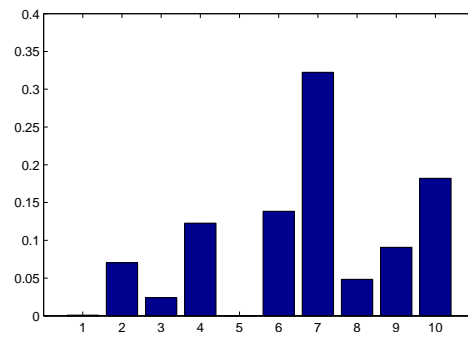
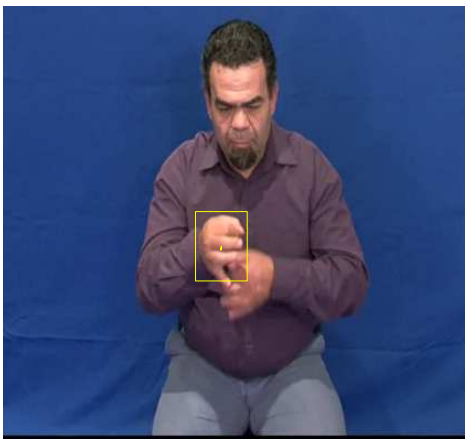
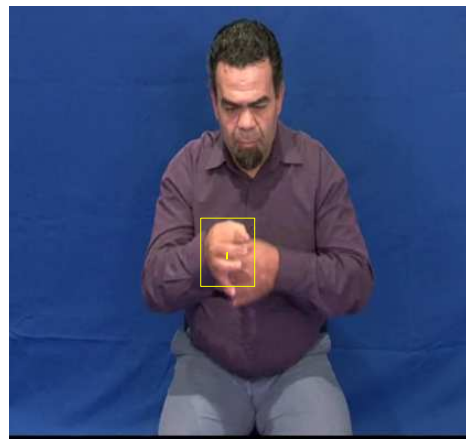
Σχήμα C.4: Κατατιμήσεις με βάση την κίνηση σε $N_c = 4$ κλάσεις για κάθε εκδοχή και κατεύθυνση πεδίου οπτικής ροής. Θυμίζουμε από την ενότητα 6.2 ότι ταξινομούνται μόνο τα διανύσματα κίνησης των οποίων η αφετηρία ανήκει στη χρωματική μάσκα B_{48}^c για την πάνω σειρά (προς τα εμπρός κίνηση) και τη χρωματική μάσκα B_{49}^c για την κάτω σειρά (προς τα πίσω κίνηση).

C.2 Κάποια ενδιάμεσα αποτελέσματα

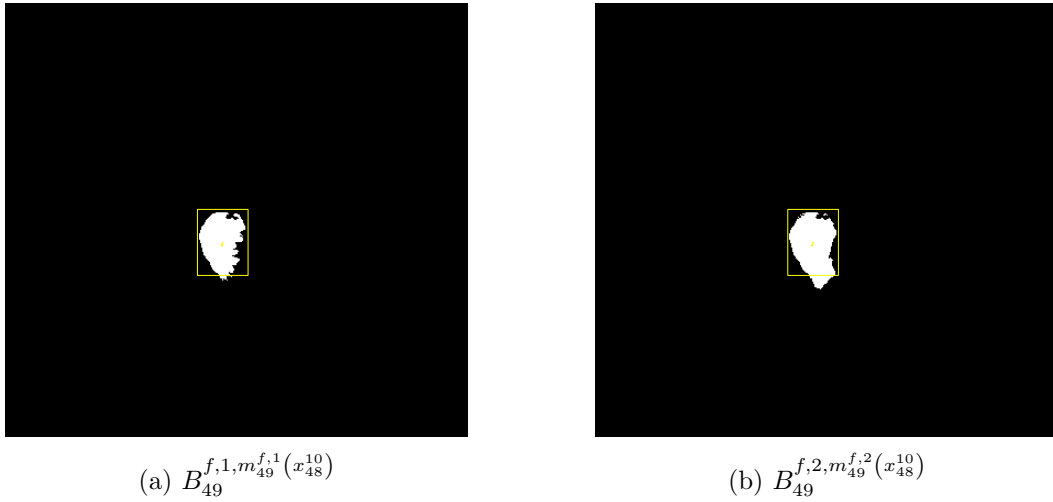
Σε αυτή την ενότητα απεικονίζουμε κάποια ενδιάμεσα αποτελέσματα από μια εκτέλεση του Particle Filter αλγορίθμου παρακολούθησης στην ακολουθία της δοκιμής μας, για οποία χρησιμοποιήσαμε πλήθος σταθμισμένων δειγμάτων (number of particles) ίσο με $N_s = 10$ και ελάχιστο ενεργό μέγεθος δείγματος ίσο με $N_T = 5$. Όπως και προηγουμένως, επικεντρώνουμε το ενδιαφέρον μας στα πλαίσια 48 και 49.

Στο Σχήμα C.5, απεικονίζουμε γραφικά τα particle sets τις χρονικές στιγμές $k = 48, 49$, τις τιμές των βαρών στάθμισης και τις αντίστοιχες εκτιμήσεις για τις καταστάσεις x_{48} και x_{49} . Θυμίζουμε ότι το σύνολο σταθμισμένων δειγμάτων τη χρονική στιγμή $k = 49$ προκύπτει από τη χρονική διάδοση και την ενημέρωση των βαρών στάθμισης των δειγμάτων τη χρονική στιγμή $k = 48$, με τον τρόπο που περιγράψαμε στην ενότητα 6.3.

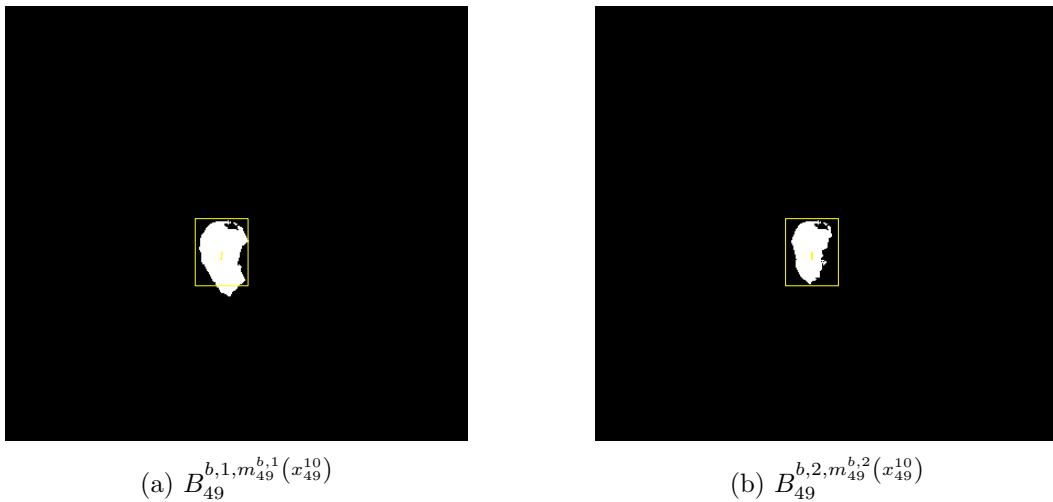
Στα Σχήματα C.6 και C.7 επικεντρώνουμε το ενδιαφέρον μας στην εξέλιξη του particle με δείκτη 10 για τις δύο αυτές χρονικές στιγμές (δείγματα x_{48}^{10} και x_{49}^{10}). Στο Σχήμα C.6, απεικονίζουμε

(a) $\{(x_{48}^i, w_{48}^i) : i = 1, \dots, 10\}$ (b) $\{(x_{49}^i, w_{49}^i) : i = 1, \dots, 10\}$ (c) Τιμές των βαρών στάθμισης w_{48}^i (d) Τιμές των βαρών στάθμισης w_{49}^i (e) Εκτίμηση της κατάστασης x_{48} (f) Εκτίμηση της κατάστασης x_{49}

Σχήμα C.5: Πρώτη σειρά: γραφική απεικόνιση του συνόλου σταθμισμένων δειγμάτων (particle set) τις χρονικές στιγμές $k = 48$ και $k = 49$. Για κάθε δείγμα/κατάσταση, σχεδιάζεται το πλαίσιο οριοθέτησης (bounding box) και το διάνυσμα ταχύτητας (με αφετηρία το κέντρο του πλαισίου οριοθέτησης) που του αντιστοιχούν. Τα πάχη των γραμμών της σχεδίασης είναι ενδεικτικά (ανάλογα) των βαρών στάθμισης. Δεύτερη σειρά: τιμές των βαρών στάθμισης w_k^i των δειγμάτων, για $i = 1, \dots, 10$ και για τις δύο χρονικές στιγμές. Τρίτη σειρά: εκτίμηση (προσδοκώμενη τιμή) της κατάστασης για τις δύο χρονικές στιγμές από τα αντίστοιχα particle sets (MMSE εκτιμήσεις, εξίσωση 6.9).



Σχήμα C.6: Δυαδικές εικόνες που αντιστοιχούν στις κύριες συνιστώσες κίνησης στην κατάσταση/δείγμα (particle) x_{48}^{10} με βάση τις κατατιμήσεις της προς τα εμπρός κίνησης $S_{49}^{f,\alpha}$ του Σχήματος C.4, για $\alpha = 1, 2$. Θυμίζουμε ότι είναι $m_{49}^{f,\alpha}(x_{48}^{10}) \equiv PMC(S_{49}^{f,\alpha}, B(x_{48}^{10}))$ (εξίσωση 6.12), ενώ χρησιμοποιούμε την αναπαράσταση 6.4 για τις $S_{49}^{f,\alpha}$ (για $\alpha = 1, 2$).



Σχήμα C.7: Δυαδικές εικόνες που αντιστοιχούν στις κύριες συνιστώσες κίνησης στην κατάσταση/δείγμα (particle) x_{49}^{10} με βάση τις κατατιμήσεις της προς τα πίσω κίνησης $S_{49}^{b,\alpha}$ του Σχήματος C.4, για $\alpha = 1, 2$. Θυμίζουμε ότι είναι $m_{49}^{b,\alpha}(x_{49}^{10}) \equiv PMC(S_{49}^{b,\alpha}, B(x_{49}^{10}))$ (εξίσωση 6.28), ενώ χρησιμοποιούμε την αναπαράσταση 6.4 για τις $S_{49}^{b,\alpha}$ (για $\alpha = 1, 2$).

την κατάσταση x_{48}^{10} , υπερτεθειμένη στη δυαδική εικόνα που υποδεικνύει την κύρια συνιστώσα κίνησης στη μάσκα κατάστασης $B(x_{48}^{10})$ για κάθε μια από τις κατατιμήσεις με βάση την προς τα εμπρός κίνηση που απεικονίσαμε στο Σχήμα C.4. Θυμίζουμε ότι το κεντροειδές (διάλυση κίνησης) που αντιστοιχεί στις δύο εκδοχές κύριας συνιστώσας κίνησης που απεικονίζονται παίζει σημαντικό ρόλο στον ορισμό του δυναμικού μοντέλου $p(x_{49}^{10}|x_{48}^{10})$ που ακολουθεί το υπό εξέταση δείγμα (ενότητα 6.3.1).

Η κατάσταση x_{49}^{10} προκύπτει από τη χρονική διάδοση (propagation) της x_{48}^{10} , μέσω δειγματοληψίας από την κατανομή $p(x_{49}^{10}|x_{48}^{10})$. Στο Σχήμα C.7, απεικονίζουμε την κατάσταση x_{49}^{10} , υπερτεθειμένη στη δυαδική εικόνα που υποδεικνύει την κύρια συνιστώσα κίνησης στη μάσκα κατάστασης $B(x_{49}^{10})$ για κάθε μια από τις κατατμήσεις με βάση την προς τα πίσω κίνηση που απεικονίσαμε στο Σχήμα C.4. Το κεντροειδές (διάλυσμα κίνησης) που αντιστοιχεί στις δύο εκδοχές κύριας συνιστώσας κίνησης που απεικονίζονται χρησιμοποιείται στον ορισμό του μοντέλου παρατήρησης $p(y_{49}|x_{49}^{10})$ για το υπό εξέταση δείγμα (ενότητα 6.3.2, πιθανοφάνεια κίνησης).

Βιβλιογραφία

- [1] A. Yilmaz, O. Javed, and M. Shah, “Object Tracking: A Survey,” *ACM Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.
- [2] P. Maragos, *Image Analysis and Computer Vision*. National Technical University of Athens, 2005.
- [3] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2002.
- [4] R. Szeliski, *Computer Vision: Algorithms and Applications*. New York, NY, USA: Springer-Verlag New York, Inc., 1st ed., 2010.
- [5] G. Stockman and L. G. Shapiro, *Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1st ed., 2001.
- [6] B. D. R. Stenger, *Model-Based Hand Tracking Using A Hierarchical Bayesian Filter*. PhD thesis, University of Cambridge, 2004.
- [7] N. Paragios, Y. Chen, and O. D. Faugeras, *Handbook of Mathematical Models in Computer Vision*. Springer Publishing Company, Incorporated, 2006.
- [8] J. Shi and C. Tomasi, “Good Features to Track,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600, IEEE, 1994.
- [9] M. Kolsch and M. Turk, “Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration,” in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW’04. Conference on*, pp. 158–158, IEEE, 2004.
- [10] J. Hoey, “Tracking using Flocks of Features, with Application to Assisted Handwashing,” in *British Machine Vision Conference (BMVC)*, 2006.
- [11] D. Serby, E. Meier, and L. Van Gool, “Probabilistic Object Tracking Using Multiple Features,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, pp. 184–187, IEEE, 2004.
- [12] D. Comaniciu, V. Ramesh, and P. Meer, “Real-Time Tracking of Non-Rigid Objects using Mean Shift,” in *Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 142–149, IEEE, 2000.
- [13] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-Based Object Tracking,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 564–577, 2003.
- [14] A. A. Argyros and M. I. Lourakis, “Tracking Skin-colored Objects in Real-time,” *Cutting Edge Robotics*, pp. 77–90, 2005.

- [15] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, no. 1, pp. 99–110, 2003.
- [16] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: A multi-view approach," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pp. 73–80, IEEE, 1996.
- [17] M. Isard and A. Blake, "CONDENSATION—Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision (IJCV)*, vol. 29, no. 1, pp. 5–28, 1998.
- [18] D. Terzopoulos and R. Szeliski, "Active vision," ch. Tracking with Kalman Snakes, pp. 3–20, Cambridge, MA, USA: MIT Press, 1993.
- [19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
- [20] A. Baumberg and D. Hogg, "An Efficient Method for Contour Tracking using Active Shape Models," in *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pp. 194–199, IEEE, 1994.
- [21] T. F. Cootes, G. J. Edwards, C. J. Taylor, *et al.*, "Active Appearance Models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [22] T. Mathes and J. H. Piater, "Robust Non-Rigid Object Tracking Using Point Distribution Models," in *British Machine Vision Conference (BMVC)*, British Machine Vision Association, 2005.
- [23] M. J. Black and A. D. Jepson, "Eigentracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation," *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63–84, 1998.
- [24] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 696–710, 1997.
- [25] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1409–1422, 2012.
- [26] J. Canny, "A Computational Approach to Edge Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.
- [27] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [28] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.
- [29] H. Zhou, Y. Yuan, and C. Shi, "Object Tracking Using SIFT Features and Mean Shift," *Computer vision and image understanding*, vol. 113, no. 3, pp. 345–352, 2009.

- [30] M. B. Kaaniche and F. Bremond, "Tracking HOG Descriptors for Gesture Recognition," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pp. 140–145, IEEE, 2009.
- [31] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [32] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511, IEEE, 2001.
- [33] S. Avidan, "Ensemble Tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 2, pp. 261–271, 2007.
- [34] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," in *IJCAI*, vol. 81, pp. 674–679, 1981.
- [35] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.
- [36] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *Information Theory, IEEE Transactions on*, vol. 21, no. 1, pp. 32–40, 1975.
- [37] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [38] T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories," in *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, Springer, Sept. 2010.
- [39] N. Sundaram, T. Brox, and K. Keutzer, "Dense Point Trajectories by GPU-Accelerated Large Displacement Optical Flow," in *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, Springer, Sept. 2010.
- [40] T. Brox and J. Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011.
- [41] L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance," *JOSA A*, vol. 3, no. 1, pp. 29–33, 1986.
- [42] D. A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.
- [43] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Colour by correlation: A simple, unifying approach to colour constancy," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, pp. 835–842, IEEE, 1999.
- [44] G. D. Finlayson and G. Y. Tian, "Perfect color constancy vs. color normalization for object recognition," in *Industrial Lasers and Inspection (EUROPTO Series)*, pp. 195–203, International Society for Optics and Photonics, 1999.

- [45] G. D. Finlayson, B. Schiele, and J. L. Crowley, "Comprehensive Colour Image Normalization," in *Computer Vision—ECCV'98*, pp. 475–490, Springer, 1998.
- [46] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.
- [47] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, vol. 3, pp. 85–92, 2003.
- [48] S. L. Phung, A. Bouzerdoum Sr, and D. Chai Sr, "Skin segmentation using color pixel classification: analysis and comparison," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 1, pp. 148–154, 2005.
- [49] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-interscience, second ed., 2001.
- [50] R. J. Steele and A. E. Raftery, "Performance of Bayesian model selection criteria for Gaussian mixture models," *Dept. Stat., Univ. Washington, Washington, DC, Tech. Rep*, vol. 559, 2009.
- [51] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its application in image and video databases," in *Proc. SPIE: Storage and Retrieval for Image and Video Databases VII*, vol. 3656, pp. 458–466, 1999.
- [52] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," in *ACM Transactions on Graphics (TOG)*, vol. 23, pp. 309–314, ACM, 2004.
- [53] F. Gasparini and R. Schettini, "Skin segmentation using multiple thresholding," in *Electronic Imaging 2006*, pp. 60610F–60610F, International Society for Optics and Photonics, 2006.
- [54] S. Tsekeridou and I. Pitas, "Facial feature extraction in frontal views using biometric analogies," in *Proceedings of the IX European Signal Processing Conference*, vol. 1, pp. 315–318, 1998.
- [55] B. Jähne and H. Haussecker, *Computer Vision and Applications: A Guide for Students and Practitioners*. Academic Press, 1999.
- [56] B. K. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial intelligence*, vol. 17, no. 1, pp. 185–203, 1981.
- [57] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer vision and image understanding*, vol. 63, no. 1, pp. 75–104, 1996.
- [58] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," in *ECCV 2004*, pp. 25–36, Springer, 2004.
- [59] C. Zach, T. Pock, and H. Bischof, "A Duality Based Approach for Realtime TV- L^1 Optical Flow," in *Pattern Recognition*, pp. 214–223, Springer, 2007.

- [60] D. J. Fleet and A. D. Jepson, "Computation of Component Image Velocity from Local Phase Information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.
- [61] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of Optical Flow Techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [62] D. Sun, S. Roth, and M. J. Black, "Secrets of Optical Flow Estimation and Their Principles," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2432–2439, IEEE, 2010.
- [63] C.-S. Fuh and P. Maragos, "Motion displacement estimation using an affine model for image matching," *Optical Engineering*, vol. 30, no. 7, pp. 881–887, 1991.
- [64] M. Nixon and A. S. Aguado, *Feature Extraction and Image Processing*. Newnes, 2002.
- [65] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, Aug. 2000.
- [66] "Middlebury Optical Flow Dataset." <http://vision.middlebury.edu/flow/data/>.
- [67] K. P. Murphy, *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.
- [68] M. S. Arulampalam, S. Maskell, and N. Gordon, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.
- [69] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [70] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," 1995.
- [71] P. S. Maybeck, *Stochastic models, estimation, and control*, vol. 141 of *Mathematics in Science and Engineering*. Academic Press, Inc, 1979.
- [72] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [73] J. S. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, vol. 93, pp. 1032–1044, 1998.
- [74] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *Trans. Sig. Proc.*, vol. 50, pp. 736–746, Mar. 2002.
- [75] A. Doucet, N. De Freitas, and N. Gordon, eds., *Sequential Monte Carlo methods in practice*. 2001.
- [76] J. D. Hol, T. B. Schön, and F. Gustafsson, "On Resampling Algorithms for Particle Filters," in *Nonlinear Statistical Signal Processing Workshop*, 2006.

- [77] M. Bolic, P. M. Djuric, and S. Hong, "Resampling Algorithms for Particle Filters: A Computational Complexity Perspective," *EURASIP J. Adv. Sig. Proc.*, vol. 2004, no. 15, pp. 2267–2277, 2004.
- [78] P. Sand and S. Teller, "Particle Video: Long-Range Motion Estimation Using Point Trajectories," *International Journal of Computer Vision*, vol. 80, pp. 72–91, Oct. 2008.
- [79] D. Reynolds, "Gaussian Mixture Models," *Encyclopedia of Biometrics*, pp. 659–663, 2009.
- [80] S. J. McKenna, Y. Raja, and S. Gong, "Tracking colour objects using adaptive mixture models," *Image and vision computing*, vol. 17, no. 3, pp. 225–231, 1999.
- [81] D. Dueck, "Affinity Propagation: Clustering by Passing Messages," 2009.
- [82] C. Zach, T. Pock, and H. Bischof, "A Duality Based Approach for Realtime TV-L1 Optical Flow," in *Proceedings of the 29th DAGM Conference on Pattern Recognition*, (Berlin, Heidelberg), pp. 214–223, Springer-Verlag, 2007.
- [83] "GPU4Vision project." <http://www.gpu4vision.org/>.
- [84] S. Borman, "The Expectation Maximization Algorithm - A short tutorial," 2009.
- [85] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [86] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: fifteen years later," 2011.
- [87] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings F, Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.
- [88] A. Doucet, N. Defreitas, and N. Gordon, *An Introduction to Sequential Monte Carlo Methods*. New York: Springer-Verlag,, 2001.
- [89] O. Cappe, S. J. Godsill, and E. Moulines, "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, pp. 899–924, July 2007.
- [90] M. Bolic, S. Hong, and P. M. Djuric, "Performance and Complexity Analysis of Adaptive Particle Filtering for Tracking Applications," in *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, vol. 1, pp. 853–857, IEEE, 2002.