

ΕΚΤΙΜΗΣΗ ΤΟΥ ΥΠΟΒΑΘΡΟΥ ΣΚΗΝΗΣ ΣΕ ΕΙΚΟΝΟΣΕΙΡΕΣ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Μηχανικών Η/Υ και Πληροφορικής
Εξεταστική Επιτροπή

από τον

Κωνσταντίνο Καλογερόπουλο

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Νοέμβριος 2013

ΑΦΙΕΡΩΣΗ

Θα ήθελα να αφιερώσω αυτή την εργασία στην οικογένειά μου για τη στήριξη που μου έχει δείξει όλα αυτά τα χρόνια.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά για τη βοήθεια, τη στήριξη και τις γνώσεις που μου παρείχαν κατά τη διάρκεια εκπόνησης της μεταπτυχιακής μου εργασίας εξειδίκευσης, τον αναπληρωτή καθηγητή κ. Χριστόφορο Νίκου και το διδακτορικό φοιτητή κ. Μιχάλη Βρίγκα.

Επίσης, το τμήμα των Μηχανικών Η/Υ και Πληροφορικής, τόσο για τις γνώσεις που απέκτησα σε προπτυχιακό επίπεδο, όσο και για το μεταπτυχιακό του πρόγραμμα, που είναι αξιόλογο και πραγματοποιείται χωρίς την απαίτηση διδάκτρων, γεγονός που διαδραματίζει σημαντικό ρόλο στη σημερινή οικονομική κατάσταση.

ΠΡΟΛΟΓΟΣ

Στην εποχή που ζούμε και ιδιαίτερα τώρα εν καιρώ κρίσης, η ποιότητα της παιδείας φθίνει όλο και περισσότερο με το πέρασμα των χρόνων. Η φθίνουσα αυτή πορεία της παιδείας δεν έχει αφήσει ανεπηρέαστες τις ευκαιρίες για ανώτερη μόρφωση, όπως οι πανεπιστημιακές σπουδές και οι μεταπτυχιακές σπουδές. Είναι σημαντικό λοιπόν, για την προώθηση και την ανάδειξη μελλοντικών επιστημόνων, να δίνονται ευκαιρίες για ανώτερη μόρφωση χωρίς την απαίτηση διδασκτρων. Η εφικτότητα, βέβαια, μιας τέτοιας κίνησης στην Ελλάδα, με την ολοένα και μεγαλύτερη μείωση των κονδυλίων στα πανεπιστήμια, είναι άξια συζήτησης.

Έστω ότι κάποιος, όντας στην καλύτερη των περιπτώσεων ευνοημένος από την οικονομική ενίσχυση ενός επιδοτούμενου προγράμματος έρευνας, καταφέρνει να εισαχθεί σε ένα μεταπτυχιακό πρόγραμμα σπουδών ή ακόμα καλύτερα σε ένα διδακτορικό πρόγραμμα σπουδών, προκειμένου να μπορέσει εν τέλει να επωμιστεί το μεταπτυχιακό δίπλωμα ειδίκευσης ή το διδακτορικό δίπλωμα ειδίκευσης. Το μείζον θέμα όμως που προκύπτει, είναι το κατά πόσον αυτό το επιπλέον εφόδιο με το οποίο θα οπλιστεί, θα του ανοίξει πόρτες για μελλοντική επαγγελματική αποκατάσταση στην Ελλάδα. Γιατί, για το εξωτερικό δεν αμφισβητώ την ύπαρξη προσφοράς θέσεων εργασίας, αλλά αναλόγως το άτομο, καθώς και για λόγους οικογενειακούς, πατριωτικούς, ιδεολογικούς, νοοτροπίας, τρόπους ζωής και άλλους, μπορεί να μην επιθυμεί να ξενιτευτεί.

Όσον αφορά την προσφορά θέσεων εργασίας στην Ελλάδα, αυτή είναι περισσότερο εμφανής στην πρωτεύουσα, στη συμπρωτεύουσα και σε άλλα μεγάλα αστικά κέντρα. Με το πέρασμα βέβαια του χρόνου, οι θέσεις εργασίας μειώνονται και ο ανταγωνισμός ανάμεσα στους υποψήφιους για την κάλυψη αυτών των θέσεων αυξάνεται. Η μεταβολή αυτή που παρατηρείται, οφείλεται στο γεγονός ότι όλο και

περισσότερες επιχειρήσεις και εταιρείες κλείνουν, με αποτέλεσμα τη σταδιακή έκλειψη της ζήτησης για παροχή υπηρεσιών από τον κλάδο μας. Όπως επίσης παρατηρείται, οι υπάρχουσες επιχειρήσεις και εταιρείες που βρίσκονται σε κατάσταση «ζόμπι», δε χρειάζονται πλέον υπηρεσίες, όπως το μάρκετινγκ, λόγω των μειωμένων κερδών που έχουν. Κατ' επέκταση, δουλειές οι οποίες μας έδιναν ευκαιρίες βιοπορισμού, όπως η διαφήμιση ή η προώθηση προϊόντων μέσω ιστοσελίδων, παρουσιάζουν πλέον δραματική μείωση.

Τέλος, ελπίζω αυτή η μεταπτυχιακή εργασία εξειδίκευσης και κατά συνέπεια το μεταπτυχιακό δίπλωμα ειδίκευσης που θα επωμιστώ μέσω αυτής, να μην καταλήξει επί ματαίω, παρά να σταθεί ως ο ακρογωνιαίος λίθος της ενιαίας μου κατάρτισης και της επαγγελματικής μου ανέλιξης.

Με εκτίμηση και σεβασμό,

Κωνσταντίνος Καλογερόπουλος.

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΑΦΙΕΡΩΣΗ	ii
ΕΥΧΑΡΙΣΤΙΕΣ	iii
ΠΡΟΛΟΓΟΣ	iv
ΠΕΡΙΕΧΟΜΕΝΑ	vi
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vii
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	viii
ΠΕΡΙΛΗΨΗ	ix
EXTENDED ABSTRACT IN ENGLISH	x
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1. Δομή της Διατριβής	3
ΚΕΦΑΛΑΙΟ 2. ΑΦΑΙΡΕΣΗ ΥΠΟΒΑΘΡΟΥ ΣΚΗΝΗΣ ΜΕ ΕΚΤΙΜΗΣΗ ΤΗΣ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ	4
2.1. Μοντέλο Μεικτών Κανονικών Κατανομών	4
2.2. Εκτίμηση της Πυκνότητας Πιθανότητας με Συνάρτηση Πυρήνα	7
ΚΕΦΑΛΑΙΟ 3. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΥΠΟΒΑΘΡΟΥ ΜΕ ΤΥΧΑΙΑ ΠΕΔΙΑ MARKOV	11
3.1. Ορισμοί	12
3.2. Κριτήριο MAP για MRF	13
3.3. Αλγόριθμος Ελάχιστης Διαμέρισης-Μέγιστης Ροής	15
3.3.1. Αλγόριθμος Αυξητικών Μονοπατιών για Εύρεση Μέγιστης Ροής	17
3.3.2. Επαναπαραμετροποίηση	20
3.4. Συμπερασματολογία MAP για τα MRF	22
3.5. Αλγόριθμος α -expansion	26
3.6. Αλγόριθμος α - β swap	30
3.7. Υπό Συνθήκη Τυχαία Πεδία (CRF) και Διακριτά Τυχαία Πεδία (DRF)	32
3.8. Εκτίμηση του Υποβάθρου με MRF και CRF	36
ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	39
ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ	55
ΑΝΑΦΟΡΕΣ	56
ΠΑΡΑΡΤΗΜΑ	58
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	74

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 4.1 Συγκεντρωτικά αποτελέσματα των μεθόδων σε όλες τις βάσεις	46
Πίνακας 4.2 Συγκριτικά αποτελέσματα Baseline της βάσης ChangeDetection	47
Πίνακας 4.3 Συγκριτικά αποτελέσματα Shadow της βάσης ChangeDetection	47
Πίνακας Π.1 Αποτελέσματα των μεθόδων στη βάση IPPR	58
Πίνακας Π.2 Αποτελέσματα των μεθόδων στη βάση Wallflower	59
Πίνακας Π.3 Συγκριτικά αποτελέσματα βίντεο TimeOfDay βάσης Wallflower	60
Πίνακας Π.4 Συγκριτικά αποτελέσματα βίντεο WavingTrees βάσης Wallflower	60
Πίνακας Π.5 Αποτελέσματα των μεθόδων στη βάση Perception	61
Πίνακας Π.6 Συγκριτικά αποτελέσματα ομοιότητας στη βάση Perception	62
Πίνακας Π.7 Συγκριτικά αποτελέσματα F_1 (%) στη βάση Perception	62
Πίνακας Π.8 Αποτελέσματα των μεθόδων στη βάση Weizmann	63
Πίνακας Π.9 Αποτελέσματα των μεθόδων στη βάση Baseline ChangeDetection	64
Πίνακας Π.10 Αποτελέσματα των μεθόδων στη βάση Camera Jitter	65
Πίνακας Π.11 Αποτελέσματα των μεθόδων στη βάση Dynamic Background	66
Πίνακας Π.12 Αποτελέσματα των μεθόδων στη βάση Intermittent Object Motion	67
Πίνακας Π.13 Αποτελέσματα των μεθόδων στη βάση Shadow ChangeDetection	68

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 3.1 Πρόβλημα Μέγιστης Ροής	16
Σχήμα 3.2 Κατασκευή Γραφήματος	19
Σχήμα 3.3 Επαναπαραμετροποίηση	21
Σχήμα 3.4 Γράφημα για multi-label MRF	22
Σχήμα 3.5 Περιορισμός για Submodular multi-label MRF	25
Σχήμα 3.6 Κυρτά και Μη-Κυρτά Δυναμικά	25
Σχήμα 3.7 Αλγόριθμος α -expansion	27
Σχήμα 3.8 Κατασκευή Γραφήματος Αλγόριθμου α -expansion	28
Σχήμα 3.9 Αποκοπή Αλγόριθμου α -expansion	29
Σχήμα 3.10 Αλγόριθμος α - β swap	31
Σχήμα 3.11 Κατασκευή Γραφήματος Αλγόριθμου α - β swap	31
Σχήμα 4.1 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά fountain της βάσης Perception	51
Σχήμα 4.2 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά campus της βάσης Perception	52
Σχήμα 4.3 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά Pets2006 της κατηγορίας Baseline της βάσης ChangeDetection	53
Σχήμα 4.4 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά busStation της κατηγορίας Shadow της βάσης ChangeDetection	54
Σχήμα Π.1 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά data1 της βάσης IPPR	69
Σχήμα Π.2 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά original_moonwalk της βάσης Weizmann	70
Σχήμα Π.3 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά sidewalk της κατηγορίας Camera Jitter της βάσης ChangeDetection	71
Σχήμα Π.4 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά fountain01 της κατηγορίας Dynamic Background της βάσης ChangeDetection	72
Σχήμα Π.5 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά abandonedBox της κατηγορίας Intermittent Object Motion της βάσης ChangeDetection	73

ΠΕΡΙΛΗΨΗ

Κωνσταντίνος Καλογερόπουλος του Σωτηρίου και της Ευφροσύνης.
MSc, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων,
Νοέμβριος, 2013.
ΕΚΤΙΜΗΣΗ ΤΟΥ ΥΠΟΒΑΘΡΟΥ ΣΚΗΝΗΣ ΣΕ ΕΙΚΟΝΟΣΕΙΡΕΣ.
Επιβλέπωντας: Χριστόφορος Νίκου.

Η απαλοιφή του υποβάθρου είναι μία ευρέως χρησιμοποιούμενη προσέγγιση και συχνά το πρώτο βήμα που απαιτείται σε πολλές εφαρμογές που αφορούν τον τομέα της υπολογιστικής όρασης, όπως η ανίχνευση κινούμενων αντικειμένων από στατικές κάμερες συστημάτων επιτήρησης βίντεο, ανάλυση βίντεο αθλημάτων, καταγραφή κίνησης και αυτοματισμός σε εργοστάσια. Πολλές μέθοδοι έχουν προταθεί για την αντιμετώπιση αυτού του προβλήματος και διαφέρουν στον τύπο του μοντέλου για το υπόβαθρο που χρησιμοποιείται και στη διαδικασία με την οποία αυτό το μοντέλο ανανεώνεται. Αυτή η εργασία πραγματεύεται τις εξής μεθόδους: τη μοντελοποίηση του υποβάθρου μέσω μεικτών κανονικών κατανομών, μέσω μίας συνάρτησης πυρήνα για εκτίμηση της πυκνότητας πιθανότητας, μέσω τυχαίων πεδίων *Markov* (*MRF-Markov Random Field*) και με υπό συνθήκη τυχαία πεδία (*CRF-Conditional Random Field*). Αυτές οι μέθοδοι έχουν εφαρμοσθεί σε μία επιλογή αντιπροσωπευτικών βίντεο δοκιμών και τα αποτελέσματα παρουσιάζονται και σχολιάζονται έτσι ώστε να μπορεί να ληφθεί μία απόφαση για το ποια μέθοδος για απαλοιφή του υποβάθρου είναι κατάλληλη.

EXTENDED ABSTRACT IN ENGLISH

Kalogeropoulos Konstantinos, Sotirios, Efrosyni.

MSc, Department of Computer Science and Engineering, University of Ioannina, Greece. November, 2013.

Background Subtraction From Video.

Thesis Supervisor: Christoforos Nikou.

Background subtraction is a widely used approach and often the first step of many computer vision applications to detect moving objects from static cameras for video surveillance, sport video analysis, motion capture and industrial automation.

Background subtraction involves creating a background model to represent the scene with no foreground objects present. Each image in the sequence is then compared with the background model. Pixels where the current image differs from the estimated background are classified as foreground. Various approaches have been proposed in order to overcome this problem and they differ in the type of background model used and the procedure used to update the model.

For example in the *mixture of Gaussians* model, the Gaussian distributions of the adaptive mixture model are then evaluated to determine which are most likely to result from a background process. Each pixel is classified based on whether the Gaussian distribution which represents it most effectively is considered part of the background model. The *Kernel Density Estimation (KDE)* method is a non-parametric method which estimates the true distribution of colors for a pixel over a given history and can describe repeating backgrounds. Each pixel is classified by calculating whether it has a high probability of being described by this distribution. The *MRF (Markov Random Field)* method represents the image with a graph and each pixel is connected with its neighborhoods and with one node that terms sink (background) and another that terms source (foreground). Then with a *min-cut/max-flow* algorithm the nodes of the graph are classified as foreground or background.

A background subtraction method must address many challenges. A robust system should not depend on careful placement of cameras. So, *camera jitter*, caused by wind moving the camera, is a challenge in outdoor scenes and may affect the accuracy of the foreground segmentation. The method must remain accurate over long periods of time, automatically adjusting to changing scene conditions. This is important with highway surveillance footage, which is long running and exhibits fast lighting changes caused by varying cloud coverage during time of the day or the presence of a light switch, as well as changing environmental conditions, such as rain, and varying

vehicle densities, which range from empty roads to congestion during peak traffic. Also, background motion, such as that caused by tree branches blowing in the wind, must be discarded, as it does not constitute a target foreground object. Additional, permanent variations of the objects in the scene, such as cars that park (or depart after a long period) that are called *foreground aperture*, things like movements of objects in the background that leave parts of it different from the background model or multiple objects moving in the scene both for long and short periods must be discarded.

This work discusses as a background model these methods: mixture of Gaussians, KDE: Kernel Density Estimation, MRF (Markov Random Field) and CRF (Conditional Random Field). These methods are applied to a selection of representative test footage and the results are presented and discussed so that an informed decision can be made when choosing a background subtraction method for some computer vision application.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Δομή της Διατριβής

Η απαλοιφή του υποβάθρου (*background subtraction*) [16] είναι μία ευρέως χρησιμοποιούμενη προσέγγιση και συχνά το πρώτο βήμα που απαιτείται σε πολλές εφαρμογές που αφορούν τον τομέα της υπολογιστικής όρασης, όπως η ανίχνευση κινούμενων αντικειμένων από στατικές κάμερες συστημάτων επιτήρησης βίντεο (*video surveillance*), ανάλυση βίντεο αθλημάτων, καταγραφή κίνησης (*motion capture*) και αυτοματισμός σε εργοστάσια.

Αυτή η προσέγγιση εμπεριέχει τη δημιουργία ενός μοντέλου για την αναπαράσταση του υποβάθρου, χωρίς την παρουσία των αντικειμένων που βρίσκονται στο προσκήνιο. Κάθε εικόνα συγκρίνεται με αυτό το μοντέλο. Τα εικονοστοιχεία στα οποία η τρέχουσα εικόνα διαφέρει από το εκτιμώμενο υπόβαθρο, κατηγοριοποιούνται σαν εικονοστοιχεία προσκηνίου. Πολλές μέθοδοι έχουν προταθεί για την αντιμετώπιση αυτού του προβλήματος και διαφέρουν στον τύπο του μοντέλου που χρησιμοποιείται και στη διαδικασία με την οποία αυτό το μοντέλο ανανεώνεται.

Παραδείγματος χάριν, στο μοντέλο μεικτών κανονικών κατανομών (*mixture of Gaussians*) που προτάθηκε από τους Stauffer και Grimson [19], η κάθε κανονική κατανομή του προσαρμοζόμενου μεικτού μοντέλου εκτιμάται έτσι ώστε να προσδιοριστεί ποια είναι πιο πιθανό να προέκυψε από το υπόβαθρο. Κάθε εικονοστοιχείο κατηγοριοποιείται με βάση εάν η κανονική κατανομή που το αντιπροσωπεύει, μπορεί αποδοτικότερα να θεωρηθεί ως μέρος του μοντέλου για το υπόβαθρο. Η μέθοδος με συνάρτηση πυρήνα για εκτίμηση της πυκνότητας

πιθανότητας (*KDE: Kernel Density Estimation*) που προτάθηκε από τους Mittal και Paragios [17], είναι μια μη παραμετρική μέθοδος η οποία εκτιμάει την πραγματική κατανομή των χρωμάτων ενός εικονοστοιχείου δοθέντος ενός ιστορικού και μπορεί να περιγράψει επαναλαμβανόμενα υπόβαθρα. Κάθε εικονοστοιχείο κατηγοριοποιείται υπολογίζοντας εάν έχει μεγάλη πιθανότητα περιγραφής από αυτή την κατανομή. Η μέθοδος με *MRF (Markov Random Field)* που αναφέρεται στο Κεφάλαιο 12 του βιβλίου: **“Computer vision: models, learning and inference”** του Prince [18] και ο αντίστοιχος κώδικας παρέχεται από το [8], αναπαριστά την εικόνα με ένα γράφημα στο οποίο τα εικονοστοιχεία συνδέονται με τα γειτονικά τους και με έναν κόμβο που ορίζει μια καταβόθρα (υπόβαθρο) και έναν που ορίζει μια πηγή (προσκήνιο). Ύστερα, με έναν αλγόριθμο ελάχιστης διαμέρισης-μέγιστης ροής (*min-cut/max-flow*) οι κόμβοι του γραφήματος κατηγοριοποιούνται ως προσκήνιο ή υπόβαθρο.

Μια μέθοδος πρέπει να είναι σε θέση να μπορεί να αντιμετωπίσει διάφορες προκλήσεις. Ένα εύρωστο (*robust*) σύστημα πρέπει να μην εξαρτάται από προσεκτική τοποθέτηση των καμερών. Επομένως, η μετατόπιση της κάμερας (*camera jitter*), που προκαλείται από τον άνεμο σε εξωτερικούς χώρους μπορεί να επηρεάσει την ακρίβεια της μεθόδου. Η μέθοδος πρέπει να παραμένει ακριβής σε μεγάλες περιόδους χρόνου και να προσαρμόζεται αυτόματα στις αλλαγές συνθηκών. Αυτό είναι σημαντικό σε βίντεο παρακολούθησης αυτοκινητόδρομων, τα οποία είναι μεγάλα σε διάρκεια και παρουσιάζονται γρήγορες αλλαγές στο φωτισμό που έχουν προκληθεί από εναλλαγές σύννεφων κατά τη διάρκεια της ημέρας ή την ύπαρξη ενός διακόπτη φωτός, καθώς και από αλλαγή των συνθηκών του περιβάλλοντος, όπως βροχή και ποικίλη πυκνότητα κυκλοφορίας, η οποία διακυμαίνεται από άδειους δρόμους έως συμφόρηση σε περιόδους αιχμής της κυκλοφορίας. Επίσης, κίνηση στο φόντο (*background motion*), όπως κλαδιά δέντρων που κουνιούνται από τον άνεμο, πρέπει να απορρίπτονται, καθώς δεν αποτελούν αντικείμενο του προσκηνίου που να μας ενδιαφέρει. Επιπλέον, μόνιμες μετατοπίσεις αντικειμένων, όπως όταν ένα όχημα παρκάρει (ή αναχωρεί μετά από πολλή ώρα), κινήσεις αντικειμένων που αφήνουν μέρη τους διαφορετικά από το μοντέλο του υποβάθρου ή ταυτόχρονα πολλά κινούμενα αντικείμενα για μικρές ή μεγάλες περιόδους πρέπει να απορρίπτονται.

1.1. Δομή της Διατριβής

Η συνέχεια της διατριβής περιέχει τα ακόλουθα μέρη:

- Στο Κεφάλαιο 2 αναλύεται η απαλοιφή του υποβάθρου μέσω μεικτών κανονικών κατανομών και μέσω μίας συνάρτησης πυρήνα για εκτίμηση της πυκνότητας πιθανότητας.
- Στο Κεφάλαιο 3 αναλύεται η μέθοδος απαλοιφής του υποβάθρου μέσω στοχαστικών (τυχαίων) πεδίων Markov (MRF) και υπό συνθήκη τυχαίων πεδίων (CRF) και παρουσιάζεται και η εκτίμηση του υποβάθρου με αυτές τις μεθόδους.
- Στο Κεφάλαιο 4 παρουσιάζονται και σχολιάζονται τα πειράματα που πραγματοποιήθηκαν με τη χρήση των παραπάνω μεθόδων που αναφέρθηκαν σε μία επιλογή αντιπροσωπευτικών βίντεο δοκιμών.
- Τέλος, στο Κεφάλαιο 5 υπάρχουν τα συμπεράσματα που προέκυψαν ύστερα από την παρουσίαση και το σχολιασμό των πειραμάτων των μεθόδων στο Κεφάλαιο 4.

ΚΕΦΑΛΑΙΟ 2. ΑΦΑΙΡΕΣΗ ΥΠΟΒΑΘΡΟΥ ΣΚΗΝΗΣ ΜΕ ΕΚΤΙΜΗΣΗ ΤΗΣ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

2.1 Μοντέλο Μεικτών Κανονικών Κατανομών

2.2 Εκτίμηση της Πυκνότητας Πιθανότητας με Συνάρτηση Πυρήνα

2.1. Μοντέλο Μεικτών Κανονικών Κατανομών

Σύμφωνα με τους Stauffer και Grimson [19], αντί να μοντελοποιήσουμε ρητά όλα τα εικονοστοιχεία με μία συγκεκριμένη κατανομή, απλά μοντελοποιούμε ένα συγκεκριμένο εικονοστοιχείο με μεικτές κανονικές κατανομές. Βασιζόμενοι στο βάρος και στη διακύμανση κάθε κατανομής, καθορίζουμε ποιες από αυτές μπορεί να αντιστοιχούν σε χρώματα του υποβάθρου. Τα εικονοστοιχεία που δεν ταιριάζουν σε κατανομές που αντιστοιχούν σε υπόβαθρο θεωρούνται ως εικονοστοιχεία προσκηνίου έως ότου υπάρξει κατανομή που να τα περιλαμβάνει.

Η μέθοδος αυτή προσαρμόζεται για να αντιμετωπίσει εύρωστα αλλαγές του φωτισμού, επαναλαμβανόμενες κινήσεις αντικειμένων της σκηνής, αργά κινούμενα αντικείμενα και αντικείμενα που εισέρχονται ή απομακρύνονται από τη σκηνή. Αργά κινούμενα αντικείμενα χρειάζονται περισσότερο χρόνο για να ενσωματωθούν στο υπόβαθρο, επειδή το χρώμα τους έχει μεγαλύτερη διακύμανση από το υπόβαθρο.

Αρχικά, κάθε εικονοστοιχείο χαρακτηρίζεται από την ένταση του στον χώρο χρωμάτων (RGB). Ύστερα, θεωρούμε την πιθανότητα παρακολούθησης της τρέχουσας τιμής ενός εικονοστοιχείου X_t που δίνεται από τον εξής τύπο:

$$P(X_t) = \sum_{i=1}^K w_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}), \quad (2.1)$$

όπου το K αντιστοιχεί στον αριθμό των κατανομών (οι Stauffer και W.E.L Grimson προτείνουν το K να τίθεται μεταξύ 3 και 5), το $w_{i,t}$ σε ένα βάρος για την i -οστή κανονική κατανομή τη χρονική στιγμή t , το $\mu_{i,t}$ στη μέση τιμή (*mean*) της i -οστής κανονικής κατανομής τη χρονική στιγμή t , το $\Sigma_{i,t}$ στον πίνακα συνδιακύμανσης της i -οστής κανονικής κατανομής τη χρονική στιγμή t και όπου η είναι η συνάρτηση πυκνότητας πιθανότητας (*pdf*) της κανονικής κατανομής:

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} (X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (2.2)$$

Για υπολογιστικούς λόγους, οι Stauffer και W.E.L Grimson θεώρησαν πως οι τιμές των εικονοστοιχείων στον χώρο χρωμάτων είναι ανεξάρτητες και έχουν τις ίδιες διακυμάνσεις. Έτσι, ο πίνακας συνδιακύμανσης είναι της μορφής:

$$\Sigma_{i,t} = \sigma_{i,t}^2 I \quad (2.3)$$

Για την εκτίμηση των βαρών, των μέσων τιμών και του πίνακα συνδιακύμανσης των κανονικών κατανομών χρησιμοποιώ τον αλγόριθμο *EM* (*expectation maximization*). Οι Stauffer και W.E.L Grimson χρησιμοποίησαν τον αλγόριθμο *K-means* για μελέτη σε πραγματικό χρόνο.

Όταν οι παραπάνω παράμετροι εκτιμηθούν, μια πρώτη ανίχνευση προσκηνίου μπορεί να γίνει και ύστερα οι παράμετροι ενημερώνονται. Αρχικά, οι Stauffer και W.E.L Grimson χρησιμοποίησαν ως κριτήριο την αναλογία $r_j = w_j/\sigma_j$ για να διατάξουν τις K κανονικές κατανομές. Αυτή η διάταξη υποθέτει πως ένα εικονοστοιχείο που ανήκει στο υπόβαθρο αντιστοιχεί σε κατανομή με υψηλό βάρος και μικρή διακύμανση,

εξαιτίας του γεγονότος ότι το υπόβαθρο είναι περισσότερο παρόν από τα κινούμενα αντικείμενα και για αυτό πρακτικά η τιμή του παραμένει σταθερή. Οι πρώτες b κανονικές κατανομές οι οποίες ξεπερνούν ένα συγκεκριμένο κατώφλι (*threshold*) T (μεγαλύτερο κατώφλι για επαναλαμβανόμενες κινήσεις του υποβάθρου όπως δέντρα φύλλων, κυματιζόμενες σημαίες και κύματα της θάλασσας), διατηρούνται ως κατανομές υποβάθρου:

$$B = \operatorname{argmin}_b \left(\sum_{i=1}^b w_{i,t} > T \right) \quad (2.4)$$

Οι άλλες κατανομές θεωρούνται ότι αναπαριστούν κατανομές προσκήνιου. Ύστερα, όταν έρθει το επόμενο πλαίσιο του βίντεο τη χρονική στιγμή $t+1$, οι κανονικές κατανομές που χαρακτηρίζουν καλύτερα κάθε εικονοστοιχείο, μπορούν να βρεθούν. Ένα εικονοστοιχείο αναπαρίσταται από μια κατανομή αν η απόσταση *Mahalanobis*:

$$\sqrt{(X_{t+1} - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_{t+1} - \mu_{i,t})} < k \sigma_{i,t}, \quad (2.5)$$

όπου k είναι ένα σταθερό κατώφλι ίσο με 2.5 και $\sigma_{i,t}$ η διακύμανση της i -οστής κανονικής κατανομής τη χρονική στιγμή t .

Μετά από αυτό, δύο περιπτώσεις μπορούν να συμβούν:

- Η κατανομή έχει βρεθεί ανάμεσα στις K κατανομές. Σε αυτή την περίπτωση, αν είναι κατανομή υποβάθρου, το εικονοστοιχείο κατηγοριοποιείται ως υπόβαθρο αλλιώς ως προσκήνιο.
- Καμία κατανομή δεν έχει βρεθεί ανάμεσα στις K κατανομές. Σε αυτή την περίπτωση, το εικονοστοιχείο κατηγοριοποιείται ως προσκήνιο.

Τώρα, πρέπει να ενημερωθούν κατάλληλα οι παράμετροι των κατανομών. Αν η κατανομή έχει βρεθεί τότε οι παράμετροι ενημερώνονται ως εξής:

$$w_{i,t+1} = (1-\alpha)w_{i,t} + \alpha, \text{ όπου } \alpha \text{ είναι μια σταθερά μάθησης.} \quad (2.6)$$

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1} \quad (2.7)$$

$$\sigma_{i,t+1}^2 = (1-\rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1})(X_{t+1} - \mu_{i,t+1})^T, \text{ όπου } \rho = \alpha\eta(X_{t+1}, \mu_i, \Sigma_i) \quad (2.8)$$

Αν η κατανομή δεν έχει βρεθεί, τότε μόνο τα βάρη αλλάζουν ως εξής:

$$w_{j,t+1} = (1-\alpha)w_{j,t} \quad (2.9)$$

Επίσης, μια καινούρια κατανομή δημιουργείται με μέση τιμή ίση με την τρέχουσα τιμή του εικονοστοιχείου, ένα χαμηλό βάρος και μια υψηλή διακύμανση. Αυτή η καινούρια κατανομή αντικαθιστά την κατανομή με το μικρότερο βάρος και τη μεγαλύτερη διακύμανση. Μετά από αυτή την ενημέρωση, όλα τα βάρη κανονικοποιούνται.

Ένα από τα σημαντικότερα πλεονεκτήματα αυτής της μεθόδου είναι ότι όταν επιτρέπεται κάτι να ενσωματωθεί στο υπόβαθρο, δεν καταστρέφεται το ήδη υπάρχον μοντέλο του υποβάθρου. Το αρχικό χρώμα του υποβάθρου παραμένει μέχρις ότου γίνει το K πιο πιθανό και ένα καινούριο χρώμα παρατηρηθεί. Για αυτό το λόγο, αν ένα αντικείμενο παραμένει στάσιμο για αρκετό χρονικό διάστημα, ώστε να ενσωματωθεί στο υπόβαθρο και μετά κινηθεί, η κατανομή που περιγράφει το προηγούμενο υπόβαθρο υπάρχει ακόμη με την ίδια μέση τιμή και διακύμανση, αλλά με μικρότερο βάρος και γρήγορα θα ενσωματωθεί εκ νέου στο υπόβαθρο.

2.2. Εκτίμηση της Πυκνότητας Πιθανότητας με Συνάρτηση Πυρήνα

Σε υπαίθριους χώρους, τυπικά υπάρχει ένα μεγάλο εύρος αλλαγών, οι οποίες μπορεί να συμβαίνουν πολύ γρήγορα. Οι υπαίθριοι χώροι συνήθως περιέχουν δυναμικές περιοχές, όπως κινούμενα δέντρα και θάμνους, κυματιστά νερά και κύματα ωκεανών. Τέτοιες ταχύτατες αλλαγές είναι μέρος του υποβάθρου. Η μοντελοποίηση τέτοιων

δυναμικών περιοχών απαιτεί μία πιο ευέλικτη αναπαράσταση της κατανομής πιθανότητας του υποβάθρου σε κάθε εικονοστοιχείο. Αυτό παρακινεί τη χρησιμοποίηση μη παραμετρικών εκτιμήσεων της πυκνότητας πιθανότητας για τη μοντελοποίηση του υποβάθρου.

Η εκτίμηση της πυκνότητας πιθανότητας ενός εικονοστοιχείου με μία συνάρτηση πυρήνα (*KDE*) που προτάθηκε από τους Mittal και Paragios [17], είναι μία μη παραμετρική μέθοδος η οποία εκτιμάει την πραγματική κατανομή των χρωμάτων για κάθε εικονοστοιχείο δοθέντος ενός ιστορικού και μπορεί να περιγράψει επαναλαμβανόμενα υπόβαθρα. Κάθε εικονοστοιχείο κατηγοριοποιείται υπολογίζοντας εάν έχει μεγάλη πιθανότητα περιγραφής από αυτή την κατανομή.

Μία ελκυστική μέθοδος που χρησιμοποιείται για τη μοντελοποίηση κατανομών πιθανότητας με πολλές μεταβλητές από δείγματα σημείων, είναι η εκτίμηση της πυκνότητας πιθανότητας με συνάρτηση πυρήνα (*kernel-based density estimation*, που επίσης ονομάζεται *Parzen-Rosenblatt* παράθυρα στην αναγνώριση προτύπων). Μία τέτοια επιλογή είναι ακόμα πιο κατάλληλη όταν τα δείγματα σημείων σχετίζονται με ευμετάβλητες αβεβαιότητες, καθώς το πλαίσιο παρέχει ένα δομημένο τρόπο αξιοποίησης τέτοιων αβεβαιοτήτων.

Εάν $\{X_1, X_2, \dots, X_t\} = \{I_t(x, y) : t - N \leq i \leq t\}$, είναι το ιστορικό των τιμών των εικονοστοιχείων των τελευταίων N εικόνων της ακολουθίας, τότε η πιθανότητα του εικονοστοιχείου X_t να έχει συγκεκριμένη τιμή χρώματος τη χρονική στιγμή t , δίνεται από τον εξής τύπο:

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N K_{\sigma}(X_t - X_i) \quad (2.10)$$

Μία ποικιλία συναρτήσεων πυρήνα με διαφορετικές ιδιότητες έχει χρησιμοποιηθεί στις μελέτες με μη παραμετρικές εκτιμήσεις της πυκνότητας πιθανότητας. Τυπικά, οι συναρτήσεις πυρήνα είναι συμμετρικές και μονότροπες συναρτήσεις που μειώνονται ραγδαία ως το μηδέν μακριά από το κέντρο τους, δηλαδή, θα πρέπει να έχουν πεπερασμένη τοπική υποστήριξη και σημεία πέρα από ένα συγκεκριμένο παράθυρο

να μην έχουν καθόλου συνεισφορά. Η Γκαουσιανή συνάρτηση τυπικά χρησιμοποιείται λόγω της συνέχειας, της διαφορισιμότητας και των τοπικών ιδιοτήτων της, παρόλο που παραβιάζει το κριτήριο της πεπερασμένης τοπικής υποστήριξης. Σε αυτό το σημείο να επισημάνω ότι επιλέγοντας τη Γκαουσιανή κατανομή ως συνάρτηση πυρήνα είναι διαφορετικό από το να προσαρμόσουμε την κατανομή σε ένα Γκαουσιανό μοντέλο (κανονική κατανομή). Εδώ, η Γκαουσιανή χρησιμοποιείται μόνο ως μία συνάρτηση για να δώσει βαρύτητα στα σημεία των δεδομένων. Σε αντίθεση με την παραμετρική προσαρμογή μεικτών κανονικών κατανομών, η εκτίμηση της πυκνότητας πιθανότητας με συνάρτηση πυρήνα είναι μία πιο γενική προσέγγιση, η οποία δεν υποθέτει κάποια συγκεκριμένη μορφή για τη συνάρτηση πυκνότητας πιθανότητας. Άρα, η συνάρτηση πυρήνα, K , με εύρος ζώνης σ , επιλέγεται να είναι μία Γκαουσιανή συνάρτηση και η εκτίμηση της πυκνότητας πιθανότητας δίνεται τελικά από τον εξής τύπο:

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(X_{ij} - X_{ij})^2}{\sigma_j^2}} \quad (2.11)$$

Ύστερα, κάθε εικονοστοιχείο κατηγοριοποιείται ως προσκίνητο αν η πιθανότητα, $P(X_t)$, είναι μικρότερη ενός κατωφλίου (*threshold*), T , αλλιώς κατηγοριοποιείται ως υπόβαθρο.

Ένα μείζον ζήτημα που είναι αναγκαίο να επισημανθεί όταν χρησιμοποιείται αυτή η μέθοδος, είναι η επιλογή του κατάλληλου εύρους του πυρήνα. Θεωρητικά, όσο ο αριθμός των δειγμάτων αυξάνεται και προσεγγίζει το άπειρο, η επιλογή του εύρους ζώνης είναι ασήμαντη και η εκτίμηση θα προσεγγίσει την πραγματική πυκνότητα πιθανότητας. Πρακτικά όμως, επειδή πεπερασμένος αριθμός δειγμάτων χρησιμοποιείται και ο υπολογισμός πρέπει να εκτελείται σε πραγματικό χρόνο, η επιλογή του κατάλληλου εύρους του πυρήνα είναι ουσιώδης. Ένα πολύ μικρό εύρος πυρήνα θα οδηγήσει σε μία χονδρική εκτίμηση της πυκνότητας πιθανότητας, ενώ ένα πολύ μεγάλο θα οδηγήσει σε μία υπερομαλοποιημένη εκτίμηση της πυκνότητας πιθανότητας. Το εύρος του πυρήνα πρέπει να εκτιμηθεί με ακρίβεια και καθορίζει το ποσό της επιρροής που τα γειτονικά δείγματα έχουν το ένα στο άλλο.

Ένα από τα μειονεκτήματα αυτής της μεθόδου είναι η ανάγκη για την αποθήκευση ενός μεγάλου αριθμού ιστορικού δειγμάτων για κάθε εικονοστοιχείο. Για αυτό το λόγο χρησιμοποιείται ένα σχετικά μικρό ιστορικό για τη διατήρηση του υπολογιστικού κόστους σε ένα εφικτό επίπεδο, κάνοντας τη μέθοδο να προσαρμόζεται ταχέως σε αλλαγές, αλλά να μην είναι αντιπροσωπευτική για μεγάλη περίοδο ιστορικού.

ΚΕΦΑΛΑΙΟ 3. ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΥΠΟΒΑΘΡΟΥ ΜΕ ΤΥΧΑΙΑ ΠΕΔΙΑ MARKOV

- 3.1 Ορισμοί
 - 3.2 Κριτήριο MAP για MRF
 - 3.3 Αλγόριθμος Ελάχιστης Διαμέρισης-Μέγιστης Ροής
 - 3.4 Συμπερασματολογία MAP για τα MRF
 - 3.5 Αλγόριθμος α -expansion
 - 3.6 Αλγόριθμος α - β swap
 - 3.7 Υπό Συνθήκη Τυχαία Πεδία (CRF) και Διακριτά Τυχαία Πεδία (DRF)
 - 3.8 Εκτίμηση του Υποβάθρου με MRF και CRF
-

Η απαλοιφή του υποβάθρου παρέχει έναν αποτελεσματικό μέσο για την κατάτμηση κινούμενων αντικειμένων, ενώπιο μιας στατικής κάμερας για πολλές εφαρμογές ανίχνευσης κινούμενων αντικειμένων (*tracking*) και συστημάτων επιτήρησης (*surveillance*). Οι ερευνητές παραδοσιακά, έχουν χρησιμοποιήσει ένα συνδυασμό μορφολογικών λειτουργιών για να αφαιρέσουν τον ενυπάρχων θόρυβο στο τελικό αποτέλεσμα της απαλοιφής. Τέτοιες τεχνικές μπορούν να απομονώσουν αποδοτικά αντικείμενα του προσκηνίου, αλλά τείνουν να χάνουν ακρίβεια γύρω από τα όρια ενός αντικειμένου, ειδικά για θορυβώδη δεδομένα. Σε αυτό το κεφάλαιο αναλύουμε τη χρησιμότητα ενός αλγόριθμου ελάχιστης διαμέρισης-μέγιστης ροής (*min-cut/max-flow*) πάνω σε ένα *MRF* (*Markov Random Field*) [18] και χρησιμοποιούμε *CRF* για την κατάτμηση του προσκηνίου, έχοντας ως αποτέλεσμα καθαρότερες ποιοτικά και ποσοτικά κατατμήσεις.

3.1. Ορισμοί

Ένα MRF ορίζεται ως εξής:

- Ένα σύνολο θέσεων $S = \{1 \dots N\}$. Αυτοί αντιστοιχούν στις N θέσεις των εικονοστοιχείων.
- Ένα σύνολο τυχαίων μεταβλητών $\{w_n\}_{n=1}^N$ που συνδέονται με καθένα από τους χώρους.
- Ένα σύνολο γειτόνων $\{N_n\}_{n=1}^N$ για καθένα από τους N χώρους.

Για να είναι ένας χώρος, χώρος MRF, το μοντέλο πρέπει να υπακούει στην ιδιότητα *Markov*:

$$\Pr(w_n | w_{S \setminus n}) = \Pr(w_n | w_{N_n}) \quad (3.1)$$

Με άλλα λόγια, το μοντέλο θα πρέπει να είναι υπό συνθήκη ανεξάρτητο όλων των άλλων δοθέντων μεταβλητών των γειτόνων του. Έτσι ακριβώς λειτουργεί η ανεξαρτησία υπό συνθήκη, σε ένα μη κατευθυνόμενο μοντέλο γραφήματος.

Επομένως, μπορούμε να θεωρήσουμε ένα MRF ως ένα μη κατευθυνόμενο μοντέλο γραφήματος, το οποίο περιγράφει την από κοινού πιθανότητα των μεταβλητών ως ένα γινόμενο *συναρτήσεων δυναμικού*, έτσι ώστε:

$$\Pr(w) = \frac{1}{Z} \prod_{j=1}^J \varphi_j[w_{C_j}], \quad (3.2)$$

όπου $\varphi_j[\cdot]$ είναι η j -οστή συνάρτηση δυναμικού και πάντα επιστρέφει μία μη αρνητική τιμή. Αυτή η τιμή εξαρτάται από την κατάσταση του υποσυνόλου των μεταβλητών $C_j \subset \{1, \dots, N\}$. Στο γενικό πλαίσιο των γραφημάτων, αυτό το υποσύνολο είναι

γνωστό ως *κλίκα* (*clique*). Ο όρος Z ονομάζεται συνάρτηση διαμέρισης (*partition function*) και είναι μία σταθερά κανονικοποίησης που εξασφαλίζει ότι το αποτέλεσμα είναι μία έγκυρη κατανομή πιθανότητας.

Μία επιλογή για την από κοινού πιθανότητα μεταβλητών είναι η κατανομή *Gibbs*:

$$\Pr(w) = \frac{1}{Z} \exp[-\sum_{j=1}^J \psi_j[w_{C_j}]], \quad (3.3)$$

όπου η $\psi_j[\cdot] = -\log[\phi_j[\cdot]]$ είναι γνωστή ως μία συνάρτηση κόστους και επιστρέφει είτε θετικές είτε αρνητικές τιμές.

3.2. Κριτήριο MAP για MRF

Στην απαλοιφή υποβάθρου σκηνής, στοχεύουμε στη σύνδεση μιας δυαδικής ετικέτας (*binary label*) $\{w_n\}_{n=1}^N$ σε καθένα από τα N εικονοστοιχεία της εικόνας, που θα υποδηλώνει αν το συγκεκριμένο εικονοστοιχείο ανήκει στο προσκήνιο ή στο υπόβαθρο, βασιζόμενοι στα παρατηρούμενα στον χώρο χρωμάτων (RGB) δεδομένα $\{x_n\}_{n=1}^N$ σε κάθε εικονοστοιχείο.

Αρχικά λοιπόν, θα υπολογίσουμε την εκ των υστέρων πιθανότητα (*posterior probability*) επί της άγνωστης δυαδικής ετικέτας w χρησιμοποιώντας τον κανόνα του *Bayes*:

$$\Pr(w_{1...N} | x_{1...N}) = \frac{\prod_{n=1}^N \Pr(x_n | w_n) \Pr(w_{1...N})}{\Pr(x_{1...N})}, \quad (3.4)$$

όπου έχουμε υποθέσει ότι η υπό συνθήκη πιθανότητα (*conditional probability*) $\Pr(w_{1...N} | x_{1...N})$ παραγοντοποιείται σε ένα γινόμενο όρων που συνδέονται με κάθε εικονοστοιχείο.

Για να επιτύχουμε κατάτμηση του προσκηνίου (*foreground segmentation*), εκτιμούμε τις μεταβλητές $\{x_n\}_{n=1}^N$ χρησιμοποιώντας το κριτήριο *MAP*, στοχεύοντας να βρούμε το σετ των δυαδικών ετικετών $\{w_n\}_{n=1}^N$ που μεγιστοποιούν τη μεταγενέστερη πιθανότητα $\Pr(w_{1..N} | x_{1..N})$ έτσι ώστε:

$$\begin{aligned}\hat{w}_{1..N} &= \operatorname{argmax}_{w_{1..N}} [\Pr(w_{1..N} | x_{1..N})] \\ &= \operatorname{argmax}_{w_{1..N}} [\prod_{n=1}^N \Pr(x_n | w_n) \Pr(w_{1..N})] \\ &= \operatorname{argmax}_{w_{1..N}} [\sum_{n=1}^N \log[\Pr(x_n | w_n)] + \log[\Pr(w_{1..N})]] ,\end{aligned}\quad (3.5)$$

όπου έχουμε εφαρμόσει τον κανόνα του *Bayes* και έχουμε λογαριθμήσει. Επειδή έχουμε ένα MRF με συνδέσεις μεταξύ των γειτονικών εικονοστοιχείων μπορούμε να εκφράσουμε την παραπάνω σχέση ως εξής:

$$\begin{aligned}\hat{w}_{1..N} &= \operatorname{argmax}_{w_{1..N}} [\sum_{n=1}^N \log[\Pr(x_n | w_n)] - \sum_{(m,n) \in C} \psi[w_m, w_n, \theta]] \\ &= \operatorname{argmax}_{w_{1..N}} [\sum_{n=1}^N -\log[\Pr(x_n | w_n)] + \sum_{(m,n) \in C} \psi[w_m, w_n, \theta]] \\ &= \operatorname{argmax}_{w_{1..N}} [\sum_{n=1}^N U_n(w_n) + \sum_{(m,n) \in C} P_{mn}(w_m, w_n)] ,\end{aligned}\quad (3.6)$$

όπου έχουμε υποθέσει ότι το κόστος της κλίκας $\psi[\cdot]$ είναι το ίδιο για κάθε (w_m, w_n) . Η παράμετρος θ ορίζει το κόστος $\psi[\cdot]$ για κάθε δυνατό συνδυασμό γειτονικών τιμών, έτσι ώστε όταν στην πρώτη μεταβλητή w_m της κλίκας ανατίθεται η ετικέτα j και στη δεύτερη μεταβλητή w_n ανατίθεται η ετικέτα k , πληρώνουμε κόστος ανάλογο του θ_{jk} .

Επίσης, το $U_n(w_n)$ υποδηλώνει το μοναδιαίο όρο (*unary term*) στο εικονοστοιχείο n . Αυτό είναι το κόστος για την παρατήρηση του εικονοστοιχείου n δοθείσας της κατάστασης w_n και είναι ο αρνητικός λογαριθμικός όρος πιθανότητας. Παρομοίως, το $P_{mn}(w_m, w_n)$ υποδηλώνει τον όρο μεταξύ των γειτονικών εικονοστοιχείων (*pairwise term*). Αυτό είναι το κόστος για την ανάθεση ετικετών w_m και w_n στις γειτονικές

θέσεις m και n και οφείλεται στα κόστη κλίκας $\psi[w_m, w_n, \theta]$. Να σημειώσουμε ότι έχουμε παραλείψει τον όρο $-\log[Z]$ από τον ορισμό του MRF, αφού είναι σταθερά σε σχέση με τις καταστάσεις $\{w_n\}_{n=1}^N$ και ως εκ τούτου, δεν επηρεάζει τη βέλτιστη λύση.

Η συνάρτηση κόστους (Εξίσωση 3.6) μπορεί να βελτιστοποιηθεί αξιοποιώντας ένα σύνολο τεχνικών, συλλογικά γνωστές ως διαμερίσεις γραφήματος (*graph-cuts*) [3]. Η διαμέριση του γραφήματος βασίζεται σε τεχνικές μέγιστης ροής που αρχικά προτάθηκαν από τους Ford και Fullkerson [5].

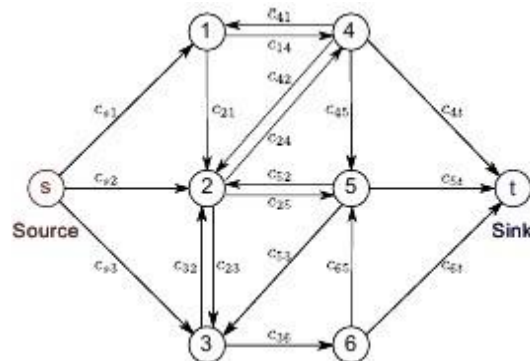
Θα θεωρήσουμε τρεις περιπτώσεις:

- Δυναδικά MRFs (Binary MRFs), (π.χ $w_i \in \{0,1\}$) όπου τα κόστη για διαφορετικούς συνδυασμούς των παρακείμενων ετικετών είναι “*submodular*” (θα εξηγήσουμε τι σημαίνει αυτό αργότερα στο κεφάλαιο). Ακριβής βέλτιστη λύση μπορεί να εντοπιστεί σε αυτή την περίπτωση.
- Πολλαπλών ετικετών MRFs (multi-label MRFs), (π.χ $w_i \in \{1,2,\dots,K\}$) όπου τα κόστη είναι “*submodular*”. Για μία ακόμη φορά, είναι δυνατή η εύρεση της ακριβούς βέλτιστης λύσης.
- Πολλαπλών ετικετών MRFs, όπου τα κόστη είναι πιο γενικά. Η εύρεση της ακριβούς βέλτιστης λύσης σε αυτή την περίπτωση είναι αδύνατη, όμως καλές προσεγγιστικές λύσεις μπορούν να εντοπιστούν σε μερικές περιπτώσεις.

3.3. Αλγόριθμος Ελάχιστης Διαμέρισης-Μέγιστης Ροής

Για να λυθούν τα ζητήματα που αναφέρθηκαν παραπάνω, θα τα μεταφράσουμε στη μορφή μέγιστης ροής (*maximum flow* ή *max-flow*) προβλημάτων. Αυτού του είδους τα προβλήματα έχουν μελετηθεί καλά και υπάρχουν ακριβείς αλγόριθμοι πολυωνυμικού χρόνου. Σε αυτή την ενότητα θα περιγράψουμε το πρόβλημα μέγιστης ροής και τη λύση του.

Θεωρούμε ένα γράφημα $G = (V, E)$ με κορυφές V και κατευθυνόμενες ακμές E , συνδέοντας τις όπως φαίνεται στο Σχήμα 3.1. Κάθε ακμή έχει μία μη-αρνητική χωρητικότητα, έτσι ώστε η ακμή μεταξύ των κορυφών m και n να έχει χωρητικότητα c_{mn} . Δύο εκ των κορυφών θεωρούνται ιδιαίτερες και ορίζονται ως πηγή (*source*) και καταβόθρα (*sink*).



Σχήμα 3.1 Πρόβλημα Μέγιστης Ροής. Η εικόνα έχει αναπαραχθεί από το [18].

Θεωρούμε ότι θέλουμε να μεταφέρουμε κάποια ποσότητα μέσω του δικτύου από την πηγή στην καταβόθρα. Ο στόχος του αλγόριθμου μέγιστης ροής είναι να υπολογίσει τη μέγιστη ποσότητα ροής που είναι εφικτό να μεταφερθεί διά μέσου του δικτύου, χωρίς να υπερβεί καμία από τις χωρητικότητες των ακμών.

Όταν η μέγιστη δυνατή ροή έχει μεταφερθεί, η λεγόμενη λύση της μέγιστης ροής, κάθε μονοπάτι από την πηγή στην καταβόθρα πρέπει να εμπεριέχει μια κορεσμένη ακμή. Εάν δεν έχει, τότε θα μπορούσαμε να ωθήσουμε λίγη ακόμα ροή προς αυτό το μονοπάτι και έτσι εξ' ορισμού αυτή δεν είναι η λύση της μέγιστης ροής.

Έπεται ότι μία εναλλακτική σκέψη για το πρόβλημα είναι να λάβουμε υπόψιν τις ακμές που έχουν κορεστεί. Ορίζουμε ως διαμέριση (*cut*) σε ένα γράφημα να είναι το ελάχιστο σύνολο κορυφών που διαχωρίζουν την πηγή από την καταβόθρα. Με άλλα λόγια, όταν αυτές οι ακμές αφαιρεθούν, δεν υπάρχει μονοπάτι από την πηγή στην καταβόθρα. Πιο συγκεκριμένα, μία διαμέριση διαιρεί τις κορυφές σε δύο ομάδες: τις κορυφές που μπορούν να προσπελαστούν μέσω κάποιου μονοπατιού από την πηγή, αλλά δεν μπορούν να προσπελάσουν την καταβόθρα, και τις κορυφές που δεν

μπορούν να προσπελαστούν από την πηγή, αλλά μπορούν να προσπελάσουν την καταβόθρα μέσω κάποιου μονοπατιού. Εν συντομία, θα αναφερόμαστε σε μία διαμέριση ως ένα ‘διαχωριστικό’ μεταξύ της πηγής και της καταβόθρας. Σε κάθε διαμέριση ανατίθεται ένα κόστος σύνδεσης, το οποίο είναι το άθροισμα των χωρητικότητων των αποκομμένων ακμών.

Καθώς οι κορεσμένες ακμές στη λύση της μέγιστης ροής διαχωρίζουν την πηγή από την καταβόθρα, συγκροτούν μία διαμέριση στο γράφημα. Στην πραγματικότητα, αυτή η ιδιαίτερη επιλογή της διαμέρισης έχει το ελάχιστο δυνατό κόστος και αναφέρεται ως ελάχιστης διαμέρισης (*minimum cut*) λύση. Ως εκ τούτου, το πρόβλημα της μέγιστης ροής και της ελάχιστης διαμέρισης μπορούν να θεωρηθούν εναλλακτικές λύσεις του ίδιου ζητήματος.

3.3.1. Αλγόριθμος Αυξητικών Μονοπατιών για Εύρεση Μέγιστης Ροής

Υπάρχουν πολλοί αλγόριθμοι για να υπολογίσουμε τη μέγιστη ροή, αλλά το να τους περιγράψουμε κατάλληλα είναι πέραν του σκοπού αυτού του κεφαλαίου. Ωστόσο, για πληρότητα, σας παρουσιάζουμε τον αλγόριθμο των αυξητικών μονοπατιών όπως παρουσιάζεται στο [18].

Θεωρήστε ότι διαλέγουμε ένα οποιοδήποτε μονοπάτι από την πηγή στην καταβόθρα και ωθούμε τη μέγιστη δυνατή ποσότητα ροής κατά μήκος του. Αυτή η ροή, είναι περιορισμένη από την ακμή σε αυτό το μονοπάτι που έχει τη μικρότερη χωρητικότητα και η οποία θα κορεστεί. Αφαιρούμε αυτό το ποσό της ροής από τις χωρητικότητες όλων των ακμών κατά μήκος του μονοπατιού, προξενώντας έτσι στην κορεσμένη ακμή να έχει νέα χωρητικότητα ίση με μηδέν. Επαναλαμβάνουμε αυτή τη διαδικασία, βρίσκοντας ένα δεύτερο μονοπάτι από την πηγή στην καταβόθρα, ωθώντας όση περισσότερη ροή είναι δυνατόν κατά μήκος του και ενημερώνοντας τις χωρητικότητες. Συνεχίζουμε αυτή τη διαδικασία μέχρι να μην υπάρχει κάποιο μονοπάτι από την πηγή στην καταβόθρα, χωρίς τουλάχιστον μία κορεσμένη ακμή. Η συνολική ροή που έχουμε μεταφέρει αποτελεί τη μέγιστη ροή και οι κορεσμένες ακμές συγκροτούν την ελάχιστη διαμέριση.

Στον πλήρη αλγόριθμο υπάρχουν κάποιες επιπλέον επιπλοκές: παραδείγματος χάριν, αν υπάρχει ήδη κάποια ροή κατά μήκος της ακμής $i-j$, ενδέχεται να υπάρχει ένα μονοπάτι από την πηγή στην καταβόθρα το οποίο εμπεριέχει την ακμή $j-i$. Σε αυτή την περίπτωση μειώνουμε τη ροή στην ακμή $i-j$ πριν προσθέσουμε τη ροή στην ακμή $j-i$.

Αν επιλέγουμε το μονοπάτι με τη μεγαλύτερη απομένουσα χωρητικότητα σε κάθε βήμα, ο αλγόριθμος είναι εγγυημένο ότι θα συγκλίνει και έχει πολυπλοκότητα $O(|E|^2|V|)$, όπου $|E|$ είναι ο αριθμός των ακμών και $|V|$ είναι ο αριθμός των κορυφών στο γράφημα. Από εδώ και στο εξής, θα υποθέσουμε ότι το πρόβλημα της μέγιστης ροής-ελάχιστης διαμέρισης (*max-flow/min-cut*) μπορεί να επιλυθεί και θα επικεντρωθούμε στο πως να μετατρέψουμε προβλήματα εκτίμησης μέσω του κριτηρίου MAP (μεγιστοποίηση της εκ των υστέρων πιθανότητας) με MRFs σε αυτή τη μορφή.

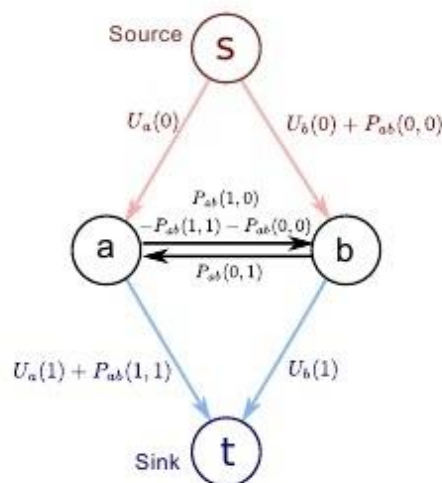
Προκειμένου να βρούμε μία MAP λύση στο πρόβλημα της τοποθέτησης ετικετών σε μία εικόνα, πρέπει να προσδιορίσουμε τις τιμές των ετικετών (Εξίσωση 3.6). Ας θεωρήσουμε τα κόστη μεταξύ των γειτονικών εικονοστοιχείων (*pairwise costs*) ως εξής:

$$\begin{aligned} P_{nm}(0,0) &= \theta_{00} & P_{nm}(1,0) &= \theta_{10} \\ P_{nm}(0,1) &= \theta_{01} & P_{nm}(1,1) &= \theta_{11} \end{aligned} \tag{3.7}$$

Η βασική ιδέα είναι να κατασκευάσουμε ένα κατευθυνόμενο γράφημα $G = (V,E)$ και να επισυνάψουμε βάρη στις ακμές του, έτσι ώστε η ελάχιστη διαμέριση σε αυτό το γράφημα να αντιστοιχεί στην MAP λύση. Πιο συγκεκριμένα, κατασκευάζουμε ένα γράφημα με μία κορυφή ανά εικονοστοιχείο και ένα ζεύγος κατευθυνόμενων ακμών μεταξύ γειτονικών κορυφών στο πλέγμα των εικονοστοιχείων. Επιπλέον, υπάρχει μία κατευθυνόμενη ακμή από την πηγή σε κάθε κορυφή και μία κατευθυνόμενη ακμή από κάθε κορυφή στην καταβόθρα.

Τώρα, θεωρούμε μία διαμέριση στο γράφημα. Σε οποιαδήποτε διαμέριση πρέπει είτε να αφαιρέσουμε την ακμή που συνδέει την πηγή με μία κορυφή εικονοστοιχείου, είτε την ακμή που συνδέει την κορυφή εικονοστοιχείου με την καταβόθρα ή και τις δύο. Εάν δεν το κάνουμε αυτό, τότε θα υπάρχει ακόμη ένα μονοπάτι από την πηγή στην καταβόθρα και δε θα είναι έγκυρη διαμέριση. Για την ελάχιστη διαμέριση, δεν πρέπει ποτέ να αφαιρέσουμε και τις δύο ακμές (υποθέτοντας τη γενική περίπτωση όπου οι δύο ακμές έχουν διαφορετικές χωρητικότητες) – αυτό είναι περιττό και αναπόφευκτα θα επιφέρει ένα μεγαλύτερο κόστος από το να αφαιρέσουμε τη μία ή την άλλη ακμή. Θα τοποθετήσουμε σε εικονοστοιχεία που η ακμή που τα συνδέει με την πηγή έχει αφαιρεθεί την ετικέτα $w_n=0$ και σε εικονοστοιχεία που η ακμή που τα συνδέει με την καταβόθρα έχει αφαιρεθεί την ετικέτα $w_n=1$. Επομένως, κάθε ταιριαστή ελάχιστη διαμέριση συσχετίζεται με την τοποθέτηση ετικετών στα εικονοστοιχεία.

Τώρα λοιπόν, ο στόχος μας είναι να αναθέσουμε χωρητικότητες στις ακμές, ώστε το κόστος κάθε διαμέρισης να ταιριάζει στο κόστος των συσχετιζόμενων ετικετών όπως ορίζεται στην Εξίσωση 3.6. Για απλότητα, το επεξηγούμε με μία μονοδιάστατη εικόνα με μόνο δύο εικονοστοιχεία (Σχήμα 3.2), αλλά τονίζουμε ότι όλες οι ιδέες είναι επίσης έγκυρες για δυσδιάστατες εικόνες και υψηλότερης διάστασης κατασκευές.



Σχήμα 3.2 Κατασκευή Γραφήματος. Η εικόνα έχει αναπαραχθεί από το [18].

Θεωρούμε τη λύση $(a=0, b=0)$. Πρέπει να αποκόψουμε τις ακμές $s - a$ και $s - b$ προσδίδοντας ένα συνολικό κόστος $U_a(0) + U_b(0) + P_{ab}(0,0)$. Για τη λύση $(a=1, b=0)$ πρέπει να αποκόψουμε τις ακμές $a - t$ και $s - b$ προσδίδοντας ένα συνολικό κόστος $U_a(1) + U_b(0) + P_{ab}(1,0)$. Παρομοίως, οι διαμερίσεις που αντιστοιχούν στις λύσεις $(a=0, b=1)$ και $(a=1, b=1)$ σε αυτό το γράφημα έχουν κόστη μεταξύ των γειτονικών εικονοστοιχείων $P_{ab}(0,1)$ και $P_{ab}(1,1)$ αντίστοιχα.

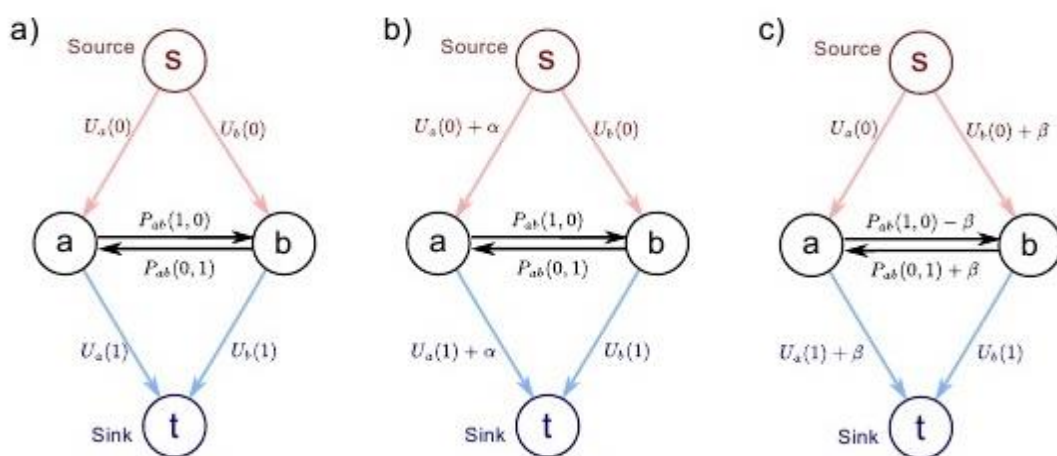
Παρατηρήστε ότι έχουμε προσθέσει το κόστος μεταξύ των γειτονικών εικονοστοιχείων $P_{ab}(0,0)$ στην ακμή $s - b$. Θα πρέπει να πληρώσουμε αυτό το κόστος κατάλληλα όπου $w_a=0$ και $w_b=0$. Δυστυχώς, θα το πληρώσουμε επίσης στην περίπτωση που $w_a=1$ και $w_b=0$. Για αυτό, αφαιρούμε το ίδιο κόστος από την ακμή $a - b$ η οποία πρέπει επίσης να αποκοπεί σε αυτή τη λύση. Με παρόμοια λογική προσθέτουμε το κόστος $P_{ab}(1,1)$ στην ακμή $a - t$ και το αφαιρούμε από την ακμή $a - b$. Με αυτό τον τρόπο συσχετίζουμε τα σωστά κόστη με κάθε τοποθέτηση ετικετών.

3.3.2. Επαναπαραμετροποίηση

Στην παραπάνω συζήτηση υποθέσαμε ότι όλα τα κόστη των ακμών είναι μη-αρνητικά και μπορούν να υπάρξουν έγκυρες χωρητικότητες. Εάν δεν είναι, τότε δεν είναι δυνατόν να υπολογίσουμε την MAP λύση. Δυστυχώς, είναι συχνή περίπτωση να υπάρξουν αρνητικά κόστη. Ακόμα και αν οι πρωτότυποι μοναδιαίοι (*unary*) και οι όροι μεταξύ των γειτονικών εικονοστοιχείων είναι θετικοί, η ακμή $a - b$ (Σχήμα 3.2) έχει κόστος $P_{ab}(1,0) - P_{ab}(1,1) - P_{ab}(0,0)$ που μπορεί να είναι αρνητικό. Η λύση σε αυτό το πρόβλημα είναι η επαναπαραμετροποίηση.

Ο στόχος της επαναπαραμετροποίησης είναι να μετασχηματίσει τα κόστη που σχετίζονται με τις ακμές στο γράφημα, με τέτοιο τρόπο ώστε η MAP λύση να μην αλλάξει. Ειδικότερα, θα προσαρμόσουμε τις χωρητικότητες των ακμών, ώστε κάθε δυνατή λύση να έχει ένα σταθερό κόστος προστιθέμενο σε αυτή. Αυτό δεν αλλάζει ποια λύση έχει η ελάχιστη διαμέριση και έτσι η τοποθέτηση ετικετών μέσω MAP θα είναι αμετάβλητη.

Θεωρούμε δύο επαναπαραμετροποιήσεις (Σχήμα 3.3). Πρώτα, θεωρούμε ότι προσθέτουμε ένα σταθερό κόστος a , στην ακμή που αντιστοιχεί στη σύνδεση από ένα εικονοστοιχείο στην πηγή και στην ακμή από το ίδιο εικονοστοιχείο στην καταβόθρα. Επειδή κάθε λύση αφαιρεί ακριβώς μία από αυτές τις ακμές, το συνολικό κόστος κάθε λύσης επαυξάνεται με a . Αυτό μπορούμε να το χρησιμοποιήσουμε ώστε να εξασφαλίσουμε ότι καμία από τις ακμές που συνδέουν τα εικονοστοιχεία με την πηγή και την καταβόθρα, δεν έχει αρνητικό κόστος: απλώς προσθέτουμε μία επαρκώς μεγάλη θετική τιμή a καθιστώντας όλα τα κόστη των ακμών μη-αρνητικά.



Σχήμα 3.3 Επαναπαραμετροποίηση. Η εικόνα έχει αναπαραχθεί από το [18].

Ένας πιο ευφυής τύπος επαναπαραμετροποίησης απεικονίζεται στο Σχήμα 3.3c. Αλλάζοντας τα κόστη με αυτόν τον τρόπο, αυξάνουμε το συνολικό κόστος κάθε δυνατής λύσης με β . Για παράδειγμα, στην ανάθεση ($w_a=0, w_b=0$) πρέπει να αφαιρέσουμε τις συνδέσεις $s-a$, $b-a$ και $b-t$ πληρώνοντας ένα συνολικό κόστος $U_a(0) + U_b(1) + P_{ab}(0,1) + \beta$.

Εφαρμόζοντας την επαναπαραμετροποίηση στο Σχήμα 3.3c στη γενική κατασκευή του Σχήματος 3.2, πρέπει να εξασφαλίσουμε ότι οι χωρητικότητες των ακμών μεταξύ των κόμβων των εικονοστοιχείων είναι μη-αρνητικές, έτσι ώστε:

$$\theta_{10} - \theta_{11} - \theta_{00} - \beta \geq 0 \quad (3.8)$$

$$\theta_{01} + \beta \geq 0 \quad (3.9)$$

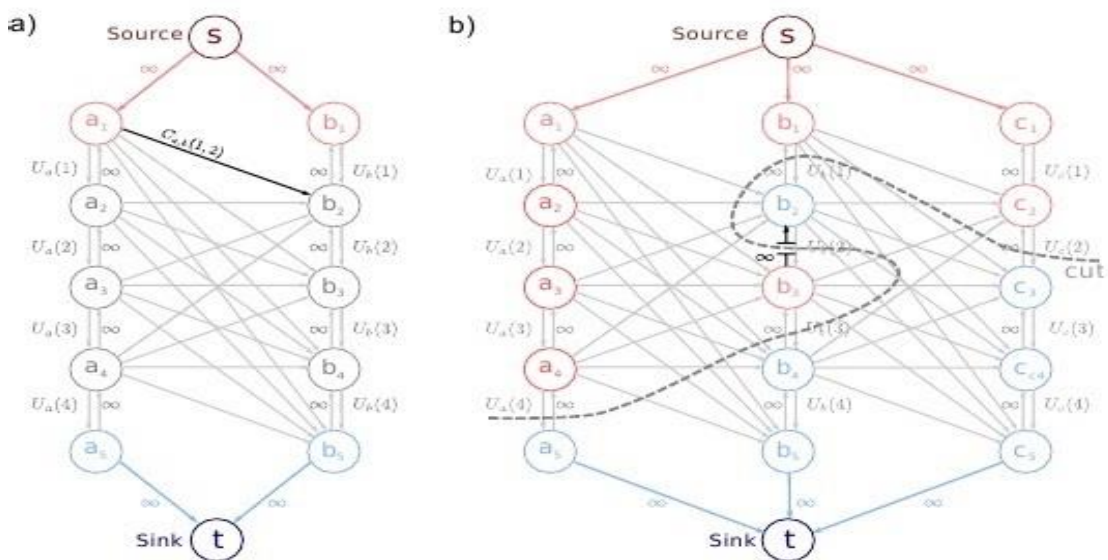
Προσθέτοντας αυτές τις εξισώσεις μαζί, μπορούμε να εξαλείψουμε το β για να καταλήξουμε στην ανισότητα:

$$\theta_{01} + \theta_{10} - \theta_{11} - \theta_{00} \geq 0 \quad (3.10)$$

Αν αυτή η συνθήκη ισχύει, το πρόβλημα είναι “*submodular*” και το γράφημα μπορεί να επαναπαραμετροποιηθεί έτσι ώστε να έχει μόνο μη-αρνητικά κόστη. Τότε μπορεί να λυθεί σε πολυωνυμικό χρόνο χρησιμοποιώντας τον αλγόριθμο ελάχιστης διαμέρισης-μέγιστης ροής. Εάν η συνθήκη δεν ισχύει, τότε αυτή η προσέγγιση δεν μπορεί να χρησιμοποιηθεί και γενικά το πρόβλημα είναι *NP-hard*.

3.4. Συμπερασματολογία MAP για τα MRF

Τώρα, θα εξετάσουμε την περίπτωση στην οποία χρησιμοποιούμε MRF δυναμικά (priors) με συνδέσεις μεταξύ των γειτονικών εικονοστοιχείων, όταν η κατάσταση w_n σε κάθε εικονοστοιχείο μπορεί να πάρει πολλαπλές ετικέτες $\{1, 2, \dots, K\}$. Για την επίλυση αυτού του προβλήματος, μεταβάλουμε την κατασκευή του γραφήματος (Σχήμα 3.4) με K ετικέτες και N εικονοστοιχεία, εισάγοντας $(K+1)N$ κορυφές στο γράφημα.



Σχήμα 3.4 Γράφημα για multi-label MRF. Η εικόνα έχει αναπαραχθεί από το [18].

Έχουμε δύο εικονοστοιχεία (a,b) και τέσσερις ετικέτες $(1,2,3,4)$. Υπάρχει μία αλυσίδα πέντε ακμών που συνδέεται με κάθε εικονοστοιχείο. Ανάμεσα στις τέσσερις κάθετες ακμές ανατίθενται τα μοναδιαία κόστη (*unary costs*) για κάθε μία από τις ετικέτες. Η ελάχιστη διαμέριση πρέπει να ‘σπάσει’ αυτή την αλυσίδα, ώστε να διαχωρίσει την πηγή από την καταβόθρα και η ετικέτα ανατίθεται σύμφωνα με το σημείο που ‘έσπασε’ η αλυσίδα. Κάθετες περιοριστικές ακμές άπειρης χωρητικότητας ρέουν ανάμεσα στις τέσσερις κορυφές, στην αντίθετη κατεύθυνση. Υπάρχουν επίσης διαγώνιες κορυφές ανάμεσα στην i -οστή κορυφή ενός εικονοστοιχείου a και την j -οστή κορυφή ενός εικονοστοιχείου b με κόστη $C_{ab}(i,j)$. Οι κάθετες περιοριστικές ακμές εμποδίζουν λύσεις σαν το παραπάνω παράδειγμα με τα τρία εικονοστοιχεία (Σχήμα 3.4b). Εδώ, η αλυσίδα των ακμών συνδέεται με το κεντρικό εικονοστοιχείο και ‘κόβεται’ σε περισσότερα του ενός σημεία. Έτσι η ετικετοποίηση δεν έχει σαφή ερμηνεία. Ωστόσο, για να συμβεί αυτό, μία περιοριστική σύνδεση πρέπει να ‘κοπεί’ και επομένως αυτή η λύση έχει άπειρο κόστος.

Στις διαγώνιες ακμές επιλέγουμε τα κόστη $C_{ab}(i,j)$ ως εξής:

$$C_{ab}(i,j) = P_{ab}(i,j-1) + P_{ab}(i-1,j) - P_{ab}(i,j) - P_{ab}(i-1,j-1), \quad (3.11)$$

όπου ορίζουμε κάθε περιττά κόστη μεταξύ των γειτονικών εικονοστοιχείων που συνδέονται με τις ανύπαρκτες ετικέτες 0 ή $K+1$ να είναι μηδέν, έτσι ώστε:

$$\begin{aligned} P_{ab}(i,0) &= 0 & P_{ab}(i,K+1) &= 0 & \forall i \in \{0 \dots K+1\} \\ P_{ab}(0,j) &= 0 & P_{ab}(K+1,j) &= 0 & \forall j \in \{0 \dots K+1\} \end{aligned} \quad (3.12)$$

Όταν η ετικέτα I ανατίθεται στο εικονοστοιχείο a και η ετικέτα J στο εικονοστοιχείο b , πρέπει να ‘κόψουμε’ όλες τις συνδέσεις από τις κορυφές $a_1 \dots a_I$ προς τις κορυφές $b_{J+1} \dots b_{K+1}$, ώστε να διαχωρίσουμε την πηγή από την καταβόθρα. Έτσι, το συνολικό κόστος λόγω των μεταξύ των εικονοστοιχείων ακμών για την ανάθεση της ετικέτας I στο εικονοστοιχείο a και της ετικέτας J στο εικονοστοιχείο b είναι:

$$\begin{aligned} \sum_{i=1}^I \sum_{j=J+1}^{K+1} C_{ab}(i, j) &= \sum_{i=1}^I \sum_{j=1}^{K+1} P_{ab}(i, j-1) + P_{ab}(i-1, j) - P_{ab}(i, j) - P_{ab}(i-1, j-1) = \\ &= P_{ab}(I, J) + P_{ab}(0, K+1) - P_{ab}(I, K+1) - P_{ab}(0, K+1) = \\ &= P_{ab}(I, J) \end{aligned} \quad (3.13)$$

Προσθέτοντας τους μοναδιαίους όρους, το συνολικό κόστος είναι $U_a(I) + U_b(J) + P_{ab}(I, J)$, όπως απαιτείται.

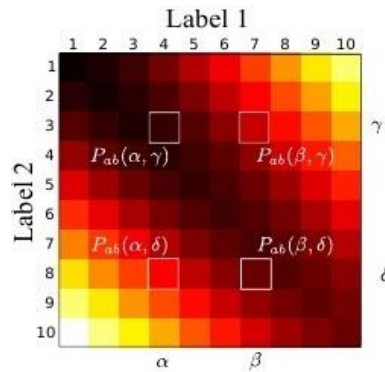
Άλλη μία φορά, έμμεσα κάναμε την υπόθεση ότι τα κόστη των ακμών είναι μη-αρνητικά. Αν οι όροι των κατακόρυφων ακμών έχουν αρνητικά κόστη, είναι δυνατή η επαναπαραμετροποίηση του γραφήματος προσθέτοντας μία σταθερά a σε όλους τους μοναδιαίους όρους. Επειδή το τελικό κόστος περιέχει ακριβώς ένα μοναδιαίο όρο για κάθε εικονοστοιχείο, κάθε δυνατή λύση αυξάνεται από τη σταθερά a και η MAP λύση μένει ανεπηρέαστη.

Οι διαγώνιες μεταξύ των εικονοστοιχείων ακμές είναι πιο προβληματικές. Είναι εφικτό να αφαιρέσουμε τις ακμές που αφήνουν τον κόμβο a_1 και τις ακμές που φθάνουν στον b_K , προσθέτοντας όρους στις ακμές μεταξύ των εικονοστοιχείων που συσχετίζονται με τους μοναδιαίους όρους. Αυτές οι ακμές μπορούν έπειτα να επαναπαραμετροποιηθούν, όπως περιγράφηκε παραπάνω, εάν είναι αναγκαίο. Δυστυχώς, δεν μπορούμε ούτε να αφαιρέσουμε ούτε να επαναπαραμετροποιήσουμε τις υπόλοιπες ακμές, έτσι απαιτούμε:

$$C_{ab}(i, j) \geq 0 \quad (3.14)$$

Από μαθηματική επαγωγή παίρνουμε το πιο γενικό αποτέλεσμα (Σχήμα 3.5),

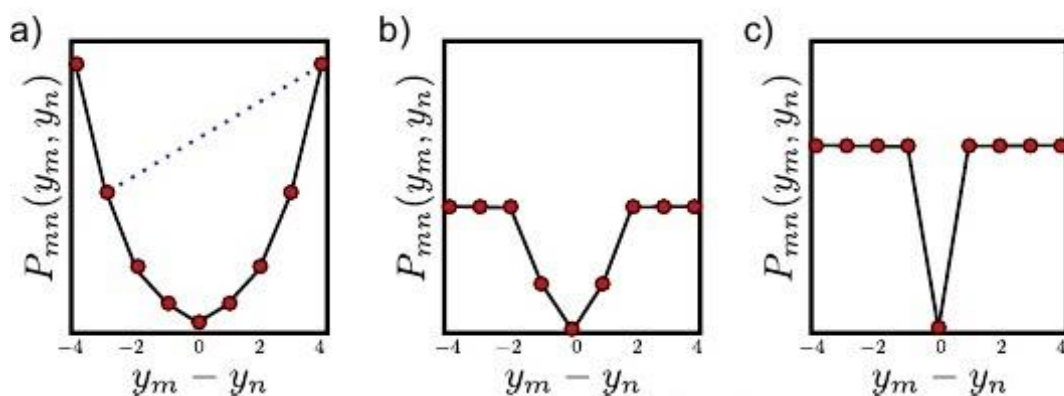
$$P_{ab}(\beta, \gamma) + P_{ab}(\alpha, \delta) - P_{ab}(\beta, \delta) - P_{ab}(\alpha, \gamma) \geq 0, \quad (3.15)$$



Σχήμα 3.5 Περιορισμός για Submodular multi-label MRF. Η εικόνα έχει αναπαραχθεί από το [18].

όπου τα $\alpha, \beta, \gamma, \delta$ είναι οποιεσδήποτε τέσσερις τιμές της κατάστασης y τέτοιες ώστε $\beta > \alpha$ και $\delta > \gamma$. Αυτή είναι μία γενίκευση της “submodular” συνθήκης με πολλαπλές ετικέτες (Εξίσωση 3.10). Αν η Εξίσωση 3.10 δεν ισχύει τότε το πρόβλημα είναι *NP-hard*.

Μία σημαντική κατηγορία κοστών μεταξύ των γειτονικών εικονοστοιχείων που είναι “submodular”, είναι αυτά που είναι κυρτά στην απόλυτη διαφορά $|w_i - w_j|$ ανάμεσα στις ετικέτες σε γειτονικά εικονοστοιχεία (Σχήμα 3.6a). Εδώ, η ομαλότητα ενθαρρύνεται όσο η ποινή γίνεται όλο και πιο αυστηρή, όταν τα άλματα μεταξύ των ετικετών μεγαλώνουν.

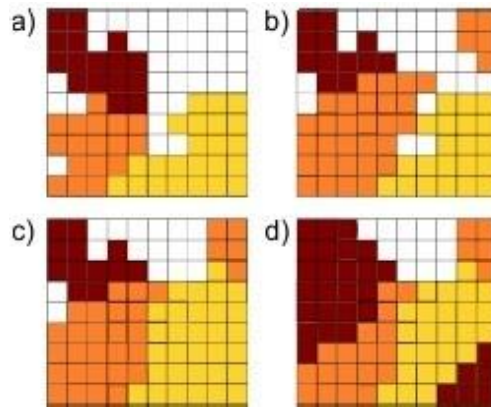


Σχήμα 3.6 Κυρτά και Μη-Κυρτά Δυναμικά. Η εικόνα έχει αναπαραχθεί από το [18].

Δυστυχώς, αυτή η κατηγορία των κυρτών κοστών μεταξύ των γειτονικών εικονοστοιχείων δεν είναι πάντοτε κατάλληλη. Σε πολλές περιπτώσεις, πρέπει να εργαστούμε με αλληλεπιδράσεις που είναι μη-κυρτές στην απόλυτη διαφορά των ετικετών, όπως η αποκομμένη τετραγωνική συνάρτηση ή το μοντέλο του *Potts* (Σχήμα 3.6b-c). Αυτές ευνοούν μικρές αλλαγές στις ετικέτες και βάζουν ποινή εξίσου ή σχεδόν εξίσου στις μεγάλες αλλαγές. Αυτό αντικατοπτρίζει το γεγονός ότι το ακριβές μέγεθος μίας απότομης αναπήδησης σε μία ετικέτα είναι σχετικά ασήμαντο. Δυστυχώς, αυτά τα κόστη μεταξύ των γειτονικών εικονοστοιχείων δεν ικανοποιούν τη “*submodular*” συνθήκη (Εξίσωση 3.10). Εδώ, η MAP λύση εν γένει δεν μπορεί να βρεθεί ακριβώς με τη μέθοδο που περιγράφηκε παραπάνω και για αυτό το πρόβλημα είναι *NP-hard*. Ευτυχώς, υπάρχουν καλές προσεγγιστικές μέθοδοι για τη βελτιστοποίηση τέτοιων προβλημάτων, όπως ο αλγόριθμος *α -expansion* [18] ή ο αλγόριθμος α - β αντιμετάθεσης (*swap*) [3].

3.5. Αλγόριθμος α -expansion

Ο αλγόριθμος α -expansion λειτουργεί διασπώντας τη λύση σε μία σειρά δυαδικών (*binary*) προβλημάτων, κάθε ένα από τα οποία μπορεί να επιλυθεί ακριβώς. Σε κάθε επανάληψη επιλέγουμε μία τιμή ετικέτας α και για κάθε εικονοστοιχείο σκεπτόμαστε είτε να διατηρήσουμε την τρέχουσα ετικέτα είτε να την αλλάξουμε στην τιμή α . Το όνομα α -expansion απορρέει από το γεγονός ότι ο χώρος που καταλαμβάνει η ετικέτα α στη λύση επεκτείνεται σε κάθε επανάληψη (Σχήμα 3.7). Η διαδικασία εκτελείται επαναληπτικά μέχρις ότου να μην υπάρχει επέκταση της ετικέτας α που να προκαλεί οποιαδήποτε αλλαγή. Κάθε επέκταση είναι εγγυημένο ότι θα μειώνει την τιμή της συνολικής αντικειμενικής συνάρτησης, παρόλο που το τελικό αποτέλεσμα δεν είναι εγγυημένο ότι θα είναι το ολικό ελάχιστο.



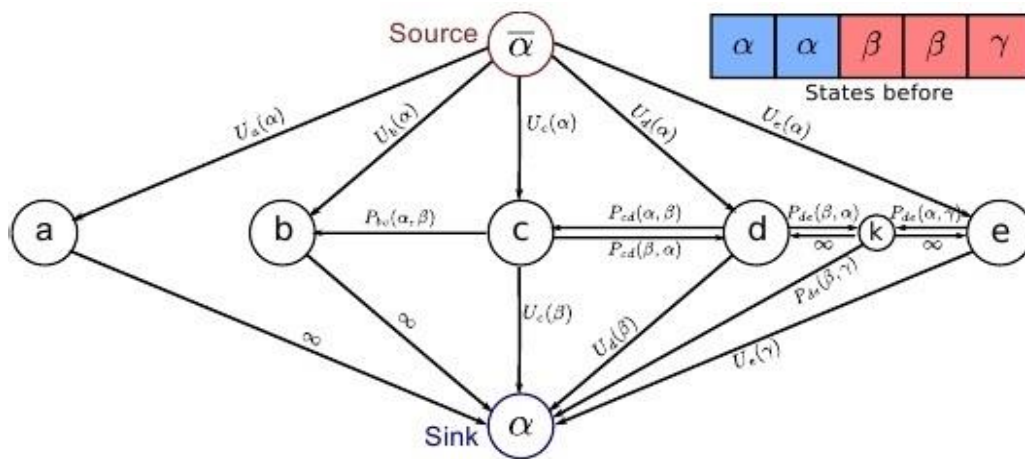
Σχήμα 3.7 Αλγόριθμος α -expansion. Η εικόνα έχει αναπαραχθεί από το [18].

Για να λειτουργήσει ο αλγόριθμος α -expansion, απαιτείται τα κόστη των ακμών να αποτελούν μία μετρική απόστασης. Με άλλα λόγια, απαιτούμε:

$$\begin{aligned}
 P(\alpha, \beta) &= 0 \Leftrightarrow \alpha = \beta \\
 P(\alpha, \beta) &= P(\beta, \alpha) \geq 0 \\
 P(\alpha, \beta) &\leq P(\alpha, \gamma) + P(\gamma, \beta)
 \end{aligned}
 \tag{3.11}$$

Αυτές οι υποθέσεις είναι λογικές για πολλές εφαρμογές και μας επιτρέπουν να μοντελοποιήσουμε μη-κυρτά δυναμικά.

Κατά την κατασκευή του γραφήματος (Σχήμα 3.8), υπάρχει ένας κόμβος που συνδέεται με κάθε εικονοστοιχείο. Κάθε ένας από αυτούς τους κόμβους συνδέεται με την πηγή (που συμβολίζει τη διατήρηση της αρχικής ετικέτας ή όχι) και την καταβόθρα (που συμβολίζει την ετικέτα α). Για να διαχωρίσουμε την πηγή από την καταβόθρα, πρέπει να αποκόψουμε μία εξ' αυτών των δύο ακμών σε κάθε εικονοστοιχείο. Η επιλογή της ακμής θα καθορίσει εάν θα κρατήσουμε την αρχική ετικέτα ή θα της θέσουμε τιμή α . Επομένως, συνδέουμε τα μοναδιαία για κάθε ακμή, θέτοντας τους τιμή α ή την αρχική τους ετικέτα με δύο συνδέσεις από κάθε εικονοστοιχείο. Αν το εικονοστοιχείο έχει ήδη την ετικέτα α , τότε θέτουμε το κόστος σε ∞ .



Σχήμα 3.8 Κατασκευή Γραφήματος Αλγόριθμου α -expansion. Η εικόνα έχει αναπαραχθεί από το [18].

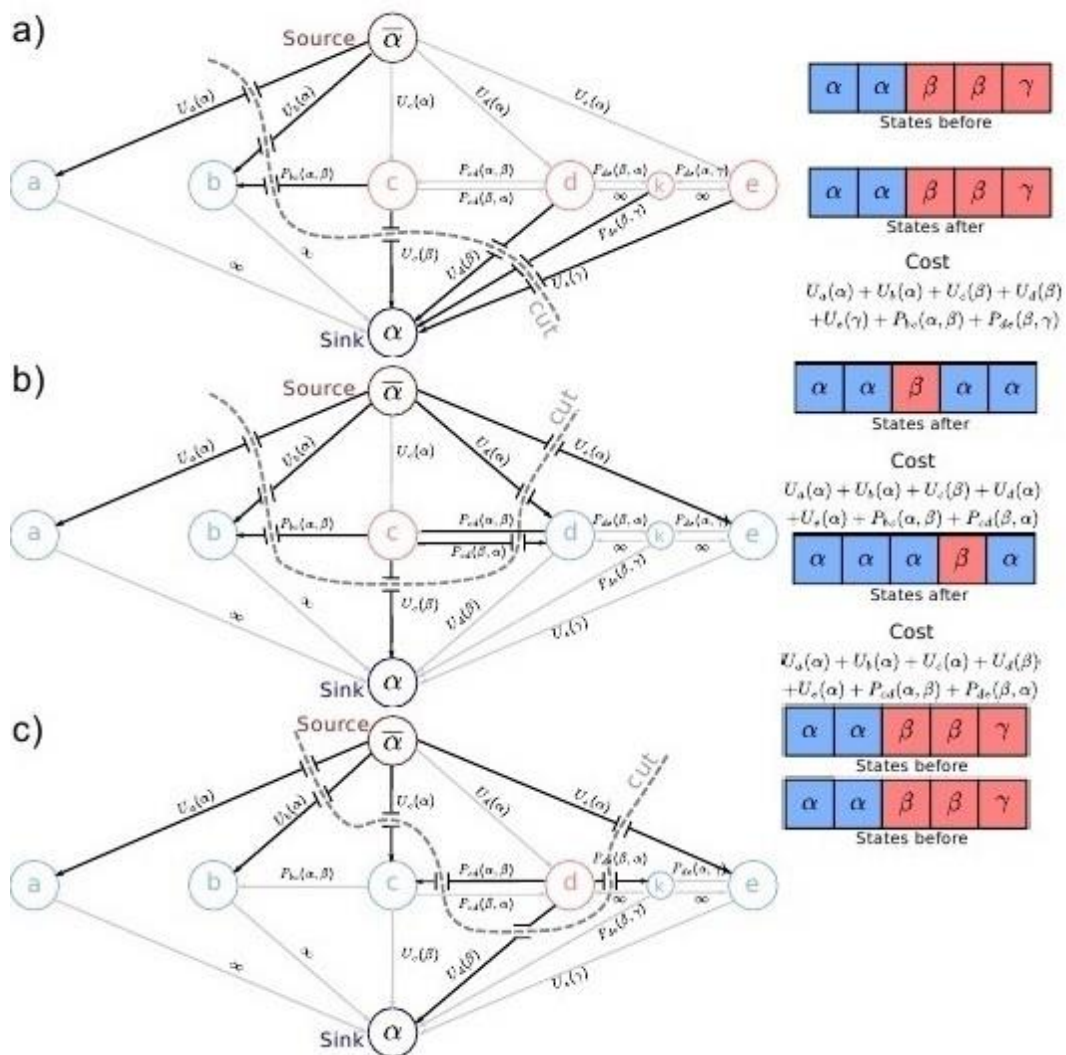
Η υπόλοιπη κατασκευή του γραφήματος είναι δυναμική: αλλάζει σε κάθε επανάληψη ανάλογα με την επιλογή του α και των τρεχουσών ετικετών. Υπάρχουν τέσσερις πιθανές σχέσεις ανάμεσα στα γειτονικά εικονοστοιχεία:

- Το εικονοστοιχείο i έχει την ετικέτα α και το εικονοστοιχείο j έχει την ετικέτα α . Εδώ, η τελική λύση είναι αναπόφευκτα $\alpha - \alpha$ και έτσι το κόστος μεταξύ των γειτονικών εικονοστοιχείων είναι μηδέν και δεν υπάρχει ανάγκη για πρόσθεση περαιτέρω ακμών που να συνδέουν τους κόμβους i και j στο γράφημα. Τα εικονοστοιχεία a και b στο Σχήμα 3.8 έχουν αυτή τη σχέση.
- Το πρώτο εικονοστοιχείο έχει την ετικέτα α αλλά το δεύτερο εικονοστοιχείο έχει μία διαφορετική ετικέτα β . Εδώ η τελική λύση μπορεί να είναι $\alpha - \alpha$ με μηδενικό κόστος ή $\alpha - \beta$ με κόστος $P_{ij}(\alpha, \beta)$. Εδώ προσθέτουμε μία μονή ακμή που συνδέει το εικονοστοιχείο j με το εικονοστοιχείο i με κόστος $P_{ij}(\alpha, \beta)$. Τα εικονοστοιχεία b και c στο Σχήμα 3.8 έχουν αυτή τη σχέση.
- Και τα δύο εικονοστοιχεία i και j παίρνουν την ίδια ετικέτα β . Εδώ η τελική λύση μπορεί να είναι $\alpha - \alpha$ με μηδενικό κόστος, $\beta - \beta$ με μηδενικό κόστος, $\alpha - \beta$ με κόστος $P_{ij}(\alpha, \beta)$ ή $\beta - \alpha$ με κόστος $P_{ij}(\beta, \alpha)$. Προσθέτουμε δύο ακμές

ανάμεσα στα εικονοστοιχεία που αντιπροσωπεύουν τα δύο μη-μηδενικά κόστη. Τα εικονοστοιχεία c και d στο Σχήμα 3.8 έχουν αυτή τη σχέση.

- Το εικονοστοιχείο i παίρνει την ετικέτα β και το εικονοστοιχείο j παίρνει μία δεύτερη ετικέτα γ . Εδώ η τελική λύση μπορεί να είναι $\alpha - \alpha$ με μηδενικό κόστος, $\beta - \gamma$ με κόστος $P_{ij}(\beta, \gamma)$, $\beta - \alpha$ με κόστος $P_{ij}(\beta, \alpha)$ ή $\alpha - \gamma$ με κόστος $P_{ij}(\alpha, \gamma)$. Προσθέτουμε ένα νέο βοηθητικό κόμβο k ανάμεσα στις κορυφές i και j και προσθέτουμε τρία μη αρνητικά κόστη στις ακμές $k - \alpha$, $i - k$ και $j - k$ αντίστοιχα. Τα εικονοστοιχεία d και e στο Σχήμα 3.8 έχουν αυτή τη σχέση.

Τρία παραδείγματα αποκοπής (διαμέρισης) παρουσιάζονται στο Σχήμα 3.9.



Σχήμα 3.9 Αποκοπή Αλγόριθμου α -expansion. Η εικόνα αναπαράχθηκε από το [18].

Να σημειώσουμε ότι αυτή η κατασκευή στηρίζεται κρίσιμα στην τριγωνική ανισότητα (Εξίσωση 3.11). Για παράδειγμα, θεωρούμε τα εικονοστοιχεία d και e στο Σχήμα 3.9a. Αν η τριγωνική ανισότητα δεν ισχύει έτσι ώστε:

$$P_{de}(\beta, \gamma) > P_{de}(\beta, \alpha) + P_{de}(\alpha, \gamma) ,$$

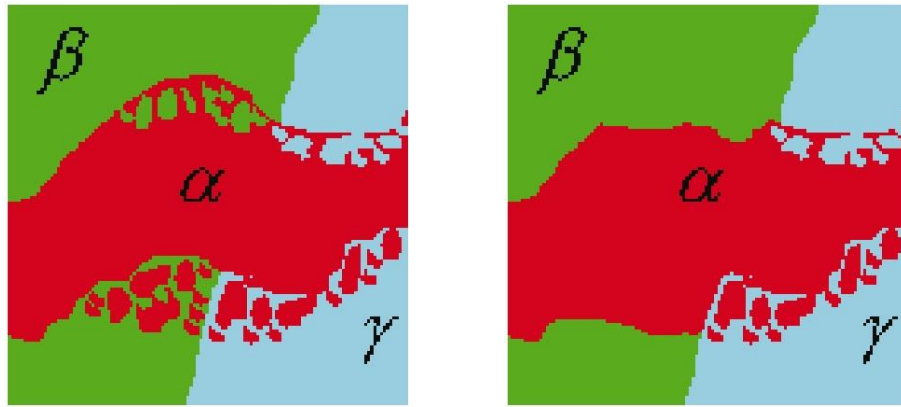
τότε τα λάθος κόστη θα ανατεθούν, αντί της σύνδεσης $k - \alpha$, οι δύο συνδέσεις $d - k$ και $e - k$ θα αποκοπούν. Στην πράξη, μερικές φορές είναι πιθανό να αγνοήσουμε αυτόν τον περιορισμό αφαιρώντας το κόστος $P_{ij}(\beta, \gamma)$ και εκτελώντας τον αλγόριθμο όπως συνήθως. Αφού γίνει η αποκοπή (διαμέριση), η αληθινή αντικειμενική συνάρτηση (το άθροισμα των μοναδιαίων και των κοστών μεταξύ των γειτονικών εικονοστοιχείων) μπορεί να υπολογιστεί για το καινούριο χάρτη ετικετών και η λύση είναι αποδεκτή αν το κόστος έχει μειωθεί.

Θα πρέπει να τονίσουμε ότι παρόλο που κάθε βήμα ενημερώνει βέλτιστα την αντικειμενική συνάρτηση, αυτός ο αλγόριθμος δεν είναι εγγυημένο ότι θα συγκλίνει στο ολικό ελάχιστο. Ωστόσο, μπορεί να αποδειχθεί ότι το αποτέλεσμα είναι εντός ενός παράγοντα του δύο του ελάχιστου και συχνά συμπεριφέρεται πολύ καλύτερα.

3.6. Αλγόριθμος α - β swap

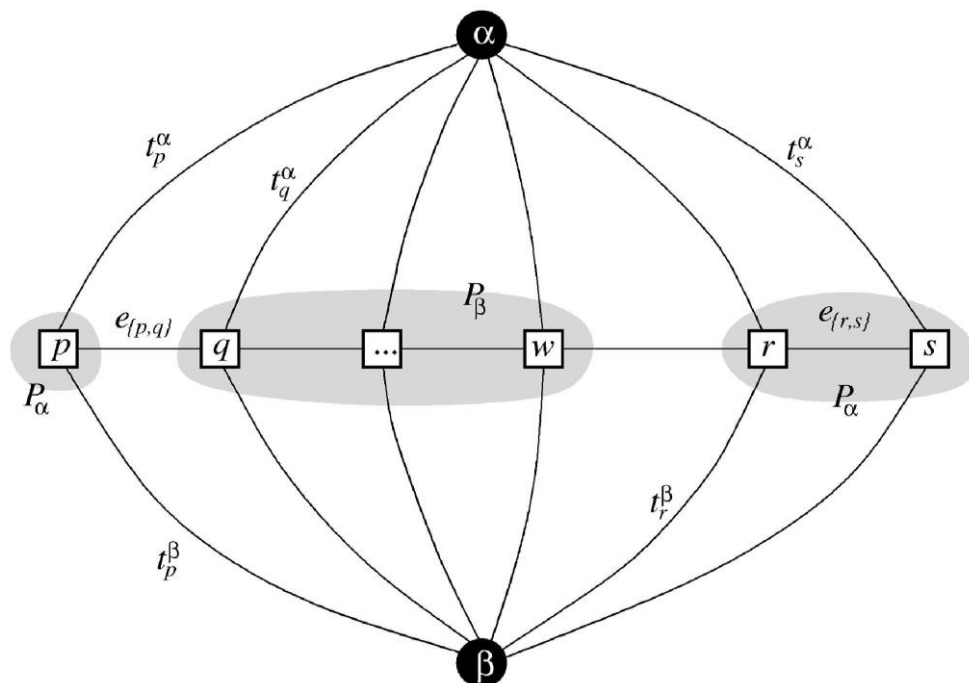
Θεωρούμε ότι σε κάθε εικονοστοιχείο $\mathbf{p} \in \mathbf{P}$ πρέπει να ανατεθεί μία ετικέτα l από κάποιο σύνολο $W = \{w_n\}_{n=1}^N$, όπου το N είναι ο αριθμός των ετικετών. Οποιαδήποτε ετικετοποίηση w μπορεί να αντιπροσωπευτεί μοναδικά από μία διαμέριση των εικονοστοιχείων της εικόνας $P = \{P_w | w \in W\}$, όπου $P_w = \{p \in P | l = w\}$ είναι ένα υποσύνολο εικονοστοιχείων με την ετικέτα w .

Δοθέντος ενός ζεύγους ετικετών α, β , μία κίνηση από την ετικετοποίηση w σε μία νέα ετικετοποίηση w' ονομάζεται α - β αντιμετάθεση [3] αν $P_{w'} = P'_w$ για κάθε ετικέτα που ισχύει $w \neq \alpha, \beta$. Ένα παράδειγμα α - β αντιμετάθεσης απεικονίζεται στο Σχήμα 3.10.



Σχήμα 3.10 Αλγόριθμος α-β swap. Η εικόνα έχει αναπαραχθεί από το [3].

Ο αλγόριθμος α-β swap βασίζεται στον υπολογισμό μιας ετικετοποίησης που να είναι αντίστοιχη σε μία ελάχιστη διαμέριση σε ένα γράφημα $G = (V, E)$. Η κατασκευή του γραφήματος είναι δυναμική και εξαρτάται από την τρέχουσα διαμέριση και τις ετικέτες α, β. Ένα παράδειγμα γραφήματος για μονοδιάστατη εικόνα παρουσιάζεται στο Σχήμα 3.11.



Σχήμα 3.11 Κατασκευή Γραφήματος Αλγόριθμου α-β swap. Η εικόνα έχει αναπαραχθεί από το [3].

Για κάθε εικόνα, το επαγόμενο γράφημα περιέχει ένα σύνολο κορυφών V με δύο ιδιαίτερους κόμβους α και β , που ονομάζονται τερματικοί κόμβοι και N εικονοστοιχεία, που ανήκουν στα σύνολα P_α και P_β . Με αυτόν τον τρόπο, το V περιέχει τους κόμβους α , β και $P_{\alpha\beta} = P_\alpha \cup P_\beta$. Κάθε εικονοστοιχείο p συνδέεται με τους τερματικούς κόμβους α και β μέσω ακμών t_ρ^α και t_ρ^β , αντίστοιχα. Θα αναφερόμαστε σε αυτές τις ακμές ως τερματικές συνδέσεις (t -links). Κάθε ζεύγος γειτονικών εικονοστοιχείων $\{p,q\} \subset P_{\alpha\beta}$ συνδέεται μέσω μίας ακμής που ονομάζεται γειτονική σύνδεση (n -links). Κάθε αποκοπή (διαμέριση) C πρέπει να περιέχει ακριβώς μία τερματική σύνδεση για κάθε εικονοστοιχείο $p \in P_{\alpha\beta}$. Αν δεν περιέχει καμία τερματική σύνδεση, τότε θα υπάρχει διαφορετικό μονοπάτι ανάμεσα στους τερματικούς κόμβους. Από την άλλη μεριά, αν και οι δύο τερματικές συνδέσεις υπάρχουν στην αποκοπή τότε αυτή δεν είναι αποδεκτή λύση.

Τα παραπάνω οδηγούν σε μία φυσική ετικετοποίηση που αντιστοιχεί σε μία αποκοπή. Αν το εικονοστοιχείο p ανήκει στο P_α , τότε στο p ανατίθεται η ετικέτα α όταν η αποκοπή διαχωρίζει το p από τον τερματικό κόμβο α . Παρομοίως, στο p ανατίθεται η ετικέτα β όταν η αποκοπή διαχωρίζει το p από τον τερματικό κόμβο β . Αν το p δεν ανήκει στο $P_{\alpha\beta}$, τότε διατηρούμε την αρχική ετικέτα w . Από αυτό συνεπάγεται ο ακόλουθος κανόνας:

$$w' = \begin{cases} \alpha & \text{αν } t_\rho^\alpha \in C \text{ και } t_\rho^\beta \notin C \text{ για } p \in P_\alpha \\ \beta & \text{αν } t_\rho^\beta \in C \text{ και } t_\rho^\alpha \notin C \text{ για } p \in P_\beta \\ w & \text{για } p \in P, p \notin P_{\alpha\beta} \end{cases} \quad (3.12)$$

3.7. Υπό Συνθήκη Τυχαία Πεδία (CRF) και Διακριτά Τυχαία Πεδία (DRF)

Τα υπό συνθήκη τυχαία πεδία (*conditional random fields* - CRFs [12]) είναι μία οικογένεια διακριτών μοντέλων (*discriminative model*) που μπορούν να χρησιμοποιηθούν για δομημένη πρόβλεψη. Τα CRFs λαμβάνουν υπόψιν το περιεχόμενο και προβλέπουν ακολουθίες ετικετών για ακολουθίες των δειγμάτων

εισόδου. Χρησιμοποιούνται για την κωδικοποίηση γνωστών σχέσεων μεταξύ των παρατηρήσεων και για την κατασκευή συνεπών ερμηνειών.

Υποθέτουμε ότι $x = \{x_1, x_2, \dots, x_n\}$ είναι μία τυχαία μεταβλητή πάνω από τις ακολουθίες των δεδομένων που είναι για ετικετοποίηση και $w = \{w_1, w_2, \dots, w_n\}$ είναι μία τυχαία μεταβλητή πάνω από τις αντίστοιχες ετικέτες. Ας θεωρήσουμε το πρόβλημα της δυαδικής ταξινόμησης (δηλαδή, $w_i \in \{-1, 1\}$). Τώρα, θεωρούμε ότι $G(V, E)$ είναι ένα γράφημα τέτοιο ώστε $w = (w_v)_{v \in V}$, έτσι ώστε το w να συνδέεται με τις κορυφές του G . Τότε, το (x, w) είναι ένα CRF σε περίπτωση που, όταν εξαρτάται από το x , οι τυχαίες μεταβλητές w_v υπακούουν στην ιδιότητα Markov (Εξίσωση 3.1).

Για αυτό, ένα CRF είναι ένα τυχαίο πεδίο (*random field*) που εξαρτάται καθολικά από τις παρατηρήσεις x . Η κοινή κατανομή πάνω στις ετικέτες w , δοθέντων των παρατηρήσεων x , μπορεί να γραφεί ως εξής:

$$\begin{aligned} P(w|x) &= \frac{1}{Z} \exp\left[\sum_{n \in V, k} \mu_k g_k(n, w_n, x) + \sum_{m \in E, k} \lambda_k f_k(m, w_m, x)\right] \\ &= \frac{1}{Z} \exp\left[\sum_{n \in V} U_n(w_n, x) + \sum_{n \in V} \sum_{m \in N_n} P_{nm}(w_n, w_m, x)\right], \end{aligned} \quad (3.13)$$

όπου Z είναι η συνάρτηση διαμέρισης και $U_n(w_n, x)$, $P_{nm}(w_n, w_m, x)$ είναι τα μοναδιαία δυναμικά και τα δυναμικά μεταξύ των γειτονικών εικονοστοιχείων αντίστοιχα. Επίσης, υποθέτουμε ότι τα χαρακτηριστικά g_k και f_k δίνονται. Το πρόβλημα είναι να εκτιμήσουμε τις παραμέτρους μ_k και λ_k από τα δεδομένα εκπαίδευσης $D = \{x_{(i)}, w_{(i)}\}_{i=1}^N$ χρησιμοποιώντας μία εμπειρική κατανομή $\tilde{P}(x, w)$. Το μοναδιαίο δυναμικό (*potential*) μπορεί να ενθαρρύνει συγκεκριμένες διαμορφώσεις του πεδίου των ετικετών, ενώ το δυναμικό μεταξύ των γειτονικών εικονοστοιχείων μπορεί να ενθαρρύνει συμφωνία μεταξύ των δεδομένων και του πεδίου των ετικετών.

Υπάρχουν δύο ουσιώδεις διαφορές ανάμεσα στο CRF μοντέλο (Εξίσωση 3.13) και στο MRF μοντέλο (Εξίσωση 3.6). Πρώτα, στο CRF το μοναδιαίο δυναμικό είναι μία συνάρτηση όλων των παρατηρήσεων $x_1, x_2, \dots, x_n$ και της ετικέτας w_i , ενώ στο MRF το

μοναδιαίο δυναμικό είναι μια συνάρτηση της παρατήρησης μόνο στη θέση x_i και της ετικέτας της. Δεύτερον, το δυναμικό μεταξύ των γειτονικών εικονοστοιχείων του CRF είναι μία συνάρτηση όλων των παρατηρήσεων x_n και των ετικετών τους, ενώ στο MRF το δυναμικό μεταξύ των γειτονικών εικονοστοιχείων είναι μια συνάρτηση των ετικετών μόνο. Επομένως, ένα CRF μπορεί να είναι κατάλληλο για την αντιμετώπιση καταστάσεων όπου η πιθανοφάνεια ενός MRF δεν είναι σε παραγοντική μορφή, τέτοια ώστε όλα τα x_n ($\forall n \in V$) να μπορούν να υπάρχουν ρητά και στα μοναδιαία δυναμικά και στα δυναμικά μεταξύ των γειτονικών εικονοστοιχείων. Επιπλέον, σε αντίθεση με ένα MRF, όπου τα x_n μπορούν να επηρεάσουν τα w_m έμμεσα μέσω του συστήματος των γειτόνων, σε ένα CRF, αυτό γίνεται άμεσα μέσω του συνδέσμου ανάμεσα στα x_n και στα w_m .

Το CRF μέχρι στιγμής έχει χρησιμοποιηθεί κυρίως σε ανάλυση (μονοδιάστατο σήμα) του λόγου. Μπορεί να επεκταθεί σε διακριτού μοντέλου τυχαίο πεδίο (*discriminative random field - DRF*) [11] για ανάλυση εικόνων όπως περιγράφεται παρακάτω.

Το DRF είναι ένα διακριτό μοντέλο που έχει χρησιμοποιηθεί για την άμεση ταξινόμηση προτύπων σε εικόνες. Το DRF είναι ένας ειδικός τύπος CRF με δύο επεκτάσεις. Πρώτα, ένα DRF ορίζεται πάνω σε δυσδιάστατα πλέγματα, τα οποία εν γένει περιέχουν γραφήματα με βρόγχους. Δεύτερον, τα μοναδιαία δυναμικά και τα δυναμικά μεταξύ των γειτονικών εικονοστοιχείων έχουν σχεδιαστεί χρησιμοποιώντας τοπικούς διακρίνοντες ταξινομητές.

Η δεσμευμένη κατανομή (*conditional distribution*) πάνω στις ετικέτες w δοθέντων των παρατηρήσεων x δίνεται από:

$$P(w|x) = \frac{1}{Z} \exp \left[\sum_{n \in V} U_n(w_n, x) + \sum_{n \in V} \sum_{m \in N_n} P_{nm}(w_n, w_m, x) \right], \quad (3.14)$$

όπου Z είναι η συνάρτηση διαμέρισης και $U_n(w_n, x)$, $P_{nm}(w_n, w_m, x)$ είναι τα μοναδιαία δυναμικά και τα δυναμικά μεταξύ των γειτονικών εικονοστοιχείων αντίστοιχα.

Ας θεωρήσουμε ότι η $f(\cdot)$ είναι μία συνάρτηση που χαρτογραφεί ένα αυθαίρετο τμήμα μίας εικόνας σε διάνυσμα χαρακτηριστικών. Τότε, το μοναδιαίο δυναμικό $U_n(w_n, x)$ μοντελοποιείται ως εξής:

$$U_n(w_n, x) = \log [\sigma(w_n \mu^T h_n(x))] , \quad (3.15)$$

όπου $h_n(x) = [I, \varphi_1(f_n(x)), \dots, \varphi_R(f_n(x))]^T$ είναι ένα μετασχηματισμένο διάνυσμα χαρακτηριστικών, $\varphi_i(\cdot)$ είναι μία μη γραμμική συνάρτηση χαρτογράφησης πυρήνα που χαρτογραφεί το πρωτότυπο διάνυσμα χαρακτηριστικών σε έναν υψηλής διάστασης χώρο, $\sigma(\cdot)$ είναι μία λογιστική συνάρτηση και μ είναι οι παράμετροι του μοντέλου.

Ορίζουμε μία καινούρια συνάρτηση $\psi(\cdot)$, που χαρτογραφεί ένα αυθαίρετο τμήμα μίας εικόνας σε διάνυσμα χαρακτηριστικών. Να σημειωθεί ότι χρησιμοποιώντας μία συνάρτηση $\psi(\cdot)$ διαφορετική από την $f(\cdot)$, είμαστε σε θέση να μοντελοποιήσουμε σχέσεις ανάμεσα σε ζεύγη θέσεων. Τώρα, το δυναμικό μεταξύ των γειτονικών εικονοστοιχείων μοντελοποιείται ως:

$$P_{nm}(w_n, w_m, x) = \beta [K w_n w_m + (1 - K)(2\sigma(w_n w_m^T g_{nm}(x)) - 1)] , \quad (3.16)$$

όπου $g_{nm}(x) = [I, \varphi_1(\psi_n(x), \psi_m(x)), \dots, \varphi_R(\psi_n(x), \psi_m(x))]^T$ είναι ένα καινούριο μετασχηματισμένο διάνυσμα χαρακτηριστικών, $\varphi_i(\cdot, \cdot)$ είναι μία μη γραμμική συνάρτηση χαρτογράφησης πυρήνα που χαρτογραφεί το πρωτότυπο διάνυσμα χαρακτηριστικών σε έναν υψηλής διάστασης χώρο και v είναι οι παράμετροι του μοντέλου. Ο πρώτος όρος στην Εξίσωση 3.16 είναι ένας όρος εξομάλυνσης, ενώ ο δεύτερος όρος είναι ένας όρος χαρτογράφησης της λογιστικής συνάρτησης μεταξύ των γειτονικών εικονοστοιχείων. Η παράμετρος K , για την οποία ισχύει $0 \leq K \leq 1$, προσθέτει στο μοντέλο ευελιξία και β είναι ένας συντελεστής αλληλεπίδρασης που ελέγχει το βαθμό της εξομάλυνσης.

Μία μέγιστης πιθανοφάνειας προσέγγιση μπορεί να υιοθετηθεί με σκοπό να μάθουμε τις παραμέτρους του DRF $\theta = \{\mu, \nu, \beta, K\}$, παρόλο που αυτό συνεπάγει την εκτίμηση της συνάρτησης διαμέρισης Z , η οποία εν γένει είναι ένα *NP-hard* πρόβλημα. Ωστόσο, είτε τεχνικές δειγματοληψίας είτε προσεγγίσεις (π.χ. ψευδοπιθανοφάνεια), μπορούν να εφαρμοσθούν για την εκτίμηση των παραμέτρων.

Η εκτίμηση της συνάρτησης διαμέρισης Z , είναι εν γένει, ένα *NP-hard* πρόβλημα. Οι βέλτιστες ετικέτες $\hat{w}_{1...N}$ μπορούν να βρεθούν μεγιστοποιώντας μία συνάρτηση κόστους της μορφής:

$$\hat{w}_{1...N} = \underset{w_{1...N}}{argmax} [\sum_{n \in V} U_n(w_n, x) + \sum_{n \in V} \sum_{m \in N_n} P_{nm}(w_n, w_m, x)], \quad (3.17)$$

η οποία μπορεί να βελτιστοποιηθεί χρησιμοποιώντας τεχνικές διαμερίσεις γραφήματος [3], όπως στην περίπτωση των MRFs.

Για δυαδικές ταξινομήσεις, μία εκτίμηση MAP μπορεί να υπολογιστεί χρησιμοποιώντας τον αλγόριθμο ελάχιστης διαμέρισης-μέγιστης ροής [18]. Μία εναλλακτική λύση είναι εύρεση της μέγιστης εκ των υστέρων περιθώριας πιθανότητας [1], όπου είτε τεχνικές δειγματοληψίας, είτε διάδοση πεποιθήσεων (*belief propagation*) [1] ή διάδοση πεποιθήσεων σε βρόγχο (*loopy belief propagation*) [10] μπορούν να εφαρμοσθούν για την εύρεση της μέγιστης εκ των υστέρων περιθώριας πιθανότητας.

3.8. Εκτίμηση του Υποβάθρου με MRF και CRF

Κάθε εικονοστοιχείο p_{ij} στην εικόνα παράγει μία αντίστοιχη κορυφή v_{ij} στο γράφημα. Δύο πρόσθετες κορυφές συγκροτούν την πηγή και την καταβόθρα, αντιπροσωπεύοντας το προσκήνιο και το υπόβαθρο αντίστοιχα. Η πηγή και η καταβόθρα συνδέονται με όλες τις κορυφές. Στο γράφημα που παράγεται για κάθε εικονοστοιχείο, λαμβάνουμε υπ' όψιν για το δυναμικό μεταξύ των γειτονικών

εικονοστοιχείων μία μικρή χωρική και χρονική (*spatial* και *temporal*) 3x3 γειτονιά της εικόνας. Μία τυπική κορυφή στο γράφημα συνδέεται με ακριβώς 26 άλλες κορυφές: με τις 8 χωρικά γειτονικές της κορυφές και με τις 18 χρονικά γειτονικές κορυφές (9 από το προηγούμενο πλαίσιο (*frame*) και 9 από το επόμενο).

Ως μοναδιαίο δυναμικό χρησιμοποιήσαμε την αρνητική λογαριθμική πιθανοφάνεια, από τις μεικτές κανονικές κατανομές στη μία περίπτωση και από την εκτίμηση της πυκνότητας πιθανότητας με συνάρτηση πυρήνα στην άλλη περίπτωση. Για δυναμικό μεταξύ των γειτονικών εικονοστοιχείων χρησιμοποιήσαμε τις εξής συναρτήσεις:

$$\Phi_{ij} = \exp(-||r_i - r_j||^2/\sigma) \quad (3.18)$$

και

$$\Psi_{ij} = \exp(-||c_i - c_j||^2/\sigma) \exp(-||r_i - r_j||^2/\sigma), \quad (3.19)$$

όπου c_i η τιμή του εικονοστοιχείου, c_j η τιμή της καθεμίας κορυφής από τις 26 που αναφέρθηκαν παραπάνω και συνδέονται με το εικονοστοιχείο, r_i η ετικέτα του εικονοστοιχείου, r_j η ετικέτα της καθεμίας κορυφής από τις 26 που συνδέονται με το εικονοστοιχείο και σ η παράμετρος της τυπικής απόκλισης που ελέγχει την ισχύ του όρου.

Άλλες παράμετροι που χρησιμοποιούνται σε αυτόν τον αλγόριθμο είναι οι παρακάτω:

- Μία παράμετρος που ελέγχει το βαθμό λεπτομέρειας (*clumpiness*) της διαμέρισης του γραφήματος και αντιστοιχεί στο επίπεδο πάνω από το οποίο τα εικονοστοιχεία συσχετίζονται πιο έντονα με το προσκήνιο παρά με το υπόβαθρο και παίρνει την ίδια τιμή για όλα τα εικονοστοιχεία.
- Μία παράμετρος που ελέγχει την ισχύ της ομάδας των γειτονικών εικονοστοιχείων. Εάν είναι μικρή, τότε τα γειτονικά εικονοστοιχεία είναι ασθενώς συνδεδεμένα μεταξύ τους και το τελικό αποτέλεσμα θα μοιάζει πολύ με αυτό που θα λαμβανόταν με απλή κατωφλίωση του αποτελέσματος από την απαλοιφή του υποβάθρου. Αντίστροφα, υψηλή τιμή αυτής της παραμέτρου έχει ως αποτέλεσμα την ισχυρή σύνδεση μεταξύ γειτονικών

εικονοστοιχείων και το τελικό αποτέλεσμα να περιέχει μεγαλύτερες περιοχές ομοιογένειας. Για αυτό το λόγο, θορυβώδη δεδομένα εισόδου τείνουν να απαιτούν μεγαλύτερες τιμές αυτής της παραμέτρου, προκειμένου να ομαλοποιηθούν μεγάλες περιοχές θορυβώδη εικονοστοιχείων.

Αφού κατασκευαστεί, πρότυπες μέθοδοι που βασίζονται στη ροή γραφημάτων θα βρουν τη βέλτιστη (ελάχιστου κόστους) διαμέριση διαχωρίζοντας την πηγή από την καταβόθρα. Ο Goldberg [4] έχει φτιάξει βελτιστοποιημένο κώδικα για αυτόν τον υπολογισμό που είναι διαθέσιμος στο Διαδίκτυο (<http://www.avglab.com/andrew/soft.html>). Κάθε κόμβος στο γράφημα θα βρίσκεται είτε στη μία είτε στην άλλη μεριά της βέλτιστης αποκοπής (διαμέρισης), παραμένοντας συνδεδεμένος μόνο στην πηγή ή στην καταβόθρα. Ο αλγόριθμος θα δώσει την ετικέτα προσκηνίου σε αυτούς τους κόμβους που είναι ακόμα συνδεδεμένοι στην πηγή και σε αυτούς που είναι στην καταβόθρα την ετικέτα υποβάθρου.

ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Προτείνω τη μέθοδο με MRF και CRF και εξετάζω διάφορες παραλλαγές για το μοναδιαίο δυναμικό και το δυναμικό μεταξύ γειτονικών εικονοστοιχείων. Στο πλαίσιο αυτό, συγκρίνουμε τις παρακάτω εκδοχές των μεθόδων:

- GMM: η μέθοδος των μεικτών κανονικών κατανομών όπως αναλύθηκε στο Κεφάλαιο 2.
- KDE: η μέθοδος της εκτίμησης της πυκνότητας πιθανότητας με Γκαουσιανή συνάρτηση πυρήνα που αναλύθηκε στο Κεφάλαιο 2.
- MRF(GMM) και MRF(KDE): οι μέθοδοι με MRF που αναλύθηκαν στο υποκεφάλαιο 3.8 και χρησιμοποιούν ως μοναδιαίο δυναμικό την αρνητική λογαριθμική πιθανοφάνεια από τις μεικτές κανονικές κατανομές και από την εκτίμηση της πυκνότητας πιθανότητας με Γκαουσιανή συνάρτηση πυρήνα αντίστοιχα και για δυναμικό μεταξύ των γειτονικών εικονοστοιχείων χρησιμοποιούν τη συνάρτηση της Εξίσωσης **3.18**.
- CRF(GMM) και CRF(KDE): οι μέθοδοι με CRF που αναλύθηκαν στο υποκεφάλαιο 3.8 και χρησιμοποιούν ως μοναδιαίο δυναμικό την αρνητική λογαριθμική πιθανοφάνεια από τις μεικτές κανονικές κατανομές και από την εκτίμηση της πυκνότητας πιθανότητας με Γκαουσιανή συνάρτηση πυρήνα αντίστοιχα και για δυναμικό μεταξύ των γειτονικών εικονοστοιχείων χρησιμοποιούν τη συνάρτηση της Εξίσωσης **3.19**.

Οι παραπάνω μέθοδοι έχουν εφαρμοσθεί σε μία επιλογή αντιπροσωπευτικών βίντεο δοκιμών, που είναι τα εξής:

- Βάση IPPR (<http://archer.ee.nctu.edu.tw/contest/>): Αποτελείται από μία σειρά τριών δύσκολων στην αναγνώριση του υποβάθρου βίντεο (data1-κλειστό γυμναστήριο, data2-εσωτερικός διάδρομος και data3-υπαίθριος χώρος στάθμευσης) με μια κάμερα επιτήρησης. Παρέχονται όλα τα πλαίσια ετικετοποιημένα με το 'χέρι' (*ground truth*).
- Βάση Wallflower (<http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm>) [20]: Αποτελείται από επτά ακολουθίες βίντεο που αντιπροσωπεύουν ένα διαφορετικό, πιθανώς προβληματικό σενάριο για διατήρηση του υποβάθρου. Πιο αναλυτικά, έχουμε εφαρμόσει τις μεθόδους μας στις εξής ακολουθίες: την Bootstrap σε μία πολυσύχναστη καφετέρια, την Camouflage στην οποία ένα άτομο περπατάει μπροστά από μία οθόνη η οποία έχει κυλιόμενες μπάρες παρεμβολής που περιέχουν παρόμοιο χρώμα με την ενδυματολογία του ατόμου, τη Foreground Aperture όπου ένα άτομο με μία μπλούζα ενός χρώματος ξυπνάει και αρχίζει να κινείται αργά, την TimeOfDay όπου ο φωτισμός σε ένα δωμάτιο αλλάζει σταδιακά από σκοτάδι σε έντονο φως και ύστερα ένα άτομο εισέρχεται στο δωμάτιο και κάθεται και τέλος τη WavingTrees με ένα δέντρο που ταλαντεύεται και ένα άτομο περπατάει μπροστά από το δέντρο. Παρέχεται το τελευταίο πλαίσιο ετικετοποιημένο με το 'χέρι' για κάθε βίντεο.
- Βάση Perception (<http://perception.i2r.a-star.edu.sg/>) [13]: Αποτελείται από εννέα βίντεο που περιέχουν σύνθετα υπόβαθρα. Αυτά τα βίντεο δοκιμών είναι δύσκολα βίντεο που περιλαμβάνουν πολυπληθείς μετακινήσεις ανθρώπων, κινούμενες σκιές κ.λπ. Τα περισσότερα από τα βίντεο δοκιμών περιέχουν μερικές χιλιάδες πλαίσια, με 20 πλαίσια ετικετοποιημένα με το 'χέρι' για κάθε βίντεο.

- Βάση Weizmann (<http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>) [2], [6]: Αποτελείται από έντεκα ακολουθίες βίντεο ανθρώπων που περπατούν σε διάφορα δύσκολα σενάρια μπροστά από διαφορετικά μη-ομοιόμορφα υπόβαθρα. Παρέχονται πλαίσια ετικετοποιημένα με το 'χέρι' για όλα τα πλαίσια των βίντεο.
- Βάση ChangeDetection (www.changedetection.net) [7]: Αποτελείται από πέντε κατηγορίες βίντεο δοκιμών από τέσσερα μέχρι έξι βίντεο σε κάθε κατηγορία. Έχουμε εφαρμόσει τις μεθόδους μας στις εξής κατηγορίες: την Baseline που είναι μία σκηνή με στατικό υπόβαθρο και στατική κάμερα, την Camera Jitter όπου έχει τοποθετηθεί μία κάμερα που μετατοπίζεται από τον άνεμο σε εξωτερικούς χώρους, την Dynamic Background που είναι μία σκηνή με δυναμικό υπόβαθρο και στατική κάμερα, την Intermittent Object Motion με μία στατική κάμερα που έχει στο οπτικό της πεδίο ένα αντικείμενο και αυτό το αντικείμενο κινείται αλλά η κίνηση του διακόπτεται κάποια χρονική στιγμή και τέλος την κατηγορία Shadow η οποία περιέχει ακολουθίες βίντεο με εμφανής σκιές. Παρέχονται πλαίσια ετικετοποιημένα με το 'χέρι' για όλα τα πλαίσια των βίντεο.

Οι παράμετροι που επηρεάζουν την ακρίβεια στην απόδοση των παραπάνω μεθόδων είναι οι εξής:

- Στη μέθοδο GMM: η παράμετρος του αριθμού των κανονικών κατανομών και η παράμετρος του αριθμού των συνιστωσών του υποβάθρου που παίρνουν την ίδια τιμή και κυμαίνεται συνήθως μεταξύ τρία και τέσσερα, εκτός από τις εξής περιπτώσεις: στο βίντεο Lobby της βάσης Perception παίρνει την τιμή δύο λόγω της γρήγορης προσαρμογής που απαιτείται στις απότομες αλλαγές στις συνθήκες φωτισμού και τέλος στο βίντεο traffic της κατηγορίας Camera Jitter και στα βίντεο fountain01 και fountain02 της κατηγορίας Dynamic Background της βάσης ChangeDetection παίρνουν την τιμή δύο για να αντιμετωπίσουμε τη μετατόπιση της κάμερας και να προσαρμοστούμε στο δυναμικό υπόβαθρο αντίστοιχα.

- Στη μέθοδο KDE: η παράμετρος για κατωφλίωση της πιθανότητας ενός εικονοστοιχείου και κατηγοριοποίησής του ως εικονοστοιχείου προσκηνίου αν είναι μικρότερη από την παράμετρο ή ως εικονοστοιχείο υποβάθρου αν είναι μεγαλύτερη από την παράμετρο και παίρνει συνήθως την τιμή $4 \cdot 10^{-3}$, εκτός από τις εξής περιπτώσεις: στο βίντεο data2 της βάσης IPPR που παίρνει την τιμή $2 \cdot 10^{-2}$, στη βάση Weizmann όπου η τιμή της κυμαίνεται από $4 \cdot 10^{-3}$ μέχρι $11 \cdot 10^{-3}$ και στο βίντεο busStation της κατηγορίας Shadow της βάσης ChangeDetection που παίρνει την τιμή $1 \cdot 10^{-2}$.
- Στις μεθόδους MRF(GMM) και CRF(GMM): η παράμετρος που ελέγχει το βαθμό λεπτομέρειας (*clumpiness*) της διαμέρισης του γραφήματος και αντιστοιχεί στο επίπεδο πάνω από το οποίο τα εικονοστοιχεία συσχετίζονται πιο έντονα με το προσκήνιο παρά με το υπόβαθρο είναι ίδια για όλα τα βίντεο δοκιμών και παίρνει την τιμή 0.55. Επίσης, σε ορισμένες περιπτώσεις αντί να χρησιμοποιήσουμε στο μοναδιαίο δυναμικό της αρνητικής λογαριθμικής πιθανοφάνειας από τις μεικτές κανονικές κατανομές τον χώρο χρωμάτων (RGB) χρησιμοποιήσαμε την κλίμακα του γκρι (grayscale) καθώς είχαμε καλύτερα αποτελέσματα. Αυτές οι περιπτώσεις είναι οι παρακάτω: στο βίντεο data1 της βάσης IPPR, στα βίντεο Escalator, Fountain, Hall, Lobby και WaterSurface της βάσης Perception, στα βίντεο original_briefcase και original_dog της βάσης Weizmann και τέλος στο βίντεο traffic της κατηγορίας Camera Jitter, στα βίντεο sofa και winterDriveway της κατηγορίας Intermitten Object Motion της βάσης ChangeDetection.
- Στις μεθόδους MRF(KDE) και CRF(KDE): η παράμετρος που ελέγχει το βαθμό λεπτομέρειας (*clumpiness*) της διαμέρισης του γραφήματος και αντιστοιχεί στο επίπεδο πάνω από το οποίο τα εικονοστοιχεία συσχετίζονται πιο έντονα με το προσκήνιο παρά με το υπόβαθρο και παίρνει τις τιμές 0.5, 0.6 και 0.7.

Να επισημάνω ότι στα αποτελέσματα έχει γίνει επεξεργασία (*post-processing*) ώστε να βελτιωθεί η ακρίβεια. Παρατηρώντας την ετικέτα των 8 χωρικά γειτονικών εικονοστοιχείων, αλλάζουμε την ετικέτα του κάθε εικονοστοιχείου. Με αυτόν τον τρόπο είτε εξαλείφουμε κομμάτια του υποβάθρου (*outliers*), είτε γεμίζουμε κενά που υπάρχουν στο προσκήνιο.

Για να μετρήσουμε την ακρίβεια χρησιμοποιήσαμε τους εξής δείκτες:

$$Recall = \frac{tp}{tp + fn}, \quad (4.1)$$

επίσης γνωστός ως *ποσοστό ανίχνευσης* (*detection rate*), που υπολογίζει το ποσοστό των εικονοστοιχείων προσκηνίου τα οποία έχουν ορθώς κατηγοριοποιηθεί ως εικονοστοιχεία προσκηνίου (*true positives-tp*) σε σύγκριση με το συνολικό αριθμό των εικονοστοιχείων προσκηνίου για την επαλήθευση των αποτελεσμάτων (όπου *false negatives-fn* ο αριθμός των εικονοστοιχείων προσκηνίου που έχουν εσφαλμένα κατηγοριοποιηθεί ως εικονοστοιχεία υποβάθρου). Υψηλή τιμή του δείκτη *Recall* υποδηλώνει ότι η μέθοδος ταξινομήσε ορθώς τα περισσότερα από τα εικονοστοιχεία προσκηνίου για την επαλήθευση ως εικονοστοιχεία προσκηνίου.

Ο δείκτης *Recall* δεν είναι αρκετός από μόνος του για να συγκρίνουμε διαφορετικές μεθόδους και γενικά χρησιμοποιείται σε συνδυασμό με το δείκτη *Precision* (Εξίσωση 4.2), επίσης γνωστό ως *θετική πρόβλεψη* (*positive prediction*), που υπολογίζει το ποσοστό των εικονοστοιχείων προσκηνίου τα οποία έχουν ορθώς κατηγοριοποιηθεί ως εικονοστοιχεία προσκηνίου σε σύγκριση με το συνολικό αριθμό των εικονοστοιχείων που ανιχνεύθηκαν από τη μέθοδο (όπου *false positives-fp* ο αριθμός των εικονοστοιχείων υποβάθρου που έχουν εσφαλμένα κατηγοριοποιηθεί ως εικονοστοιχεία προσκηνίου). Υψηλή τιμή του δείκτη *Precision* υποδηλώνει ότι από

τα εικονοστοιχεία που κατηγοριοποιήθηκαν ως εικονοστοιχεία προσκηνίου από τη μέθοδο, η πλειονότητα εξ' αυτών έχει κατηγοριοποιηθεί ορθά.

$$Precision = \frac{tp}{tp + fp}, \quad (4.2)$$

Γενικά, χρησιμοποιώντας τις προαναφερόμενες ποσότητες, μία μέθοδος θεωρείται καλή αν έχει υψηλές τιμές στην ποσότητα *Recall*, χωρίς να θυσιάζεται η ποσότητα *Precision*. Επιπλέον, λαμβάνουμε υπ' όψιν την ποσότητα F_1 (Εξίσωση 4.3), επίσης γνωστή ως *συντελεστής κέρδους* (*Figure of Merit*) ή *F-measure*, που είναι ο σταθμισμένος αρμονικός μέσος (*weighted harmonic mean*) των δεικτών *Precision* και *Recall* και υπολογίζει την ακρίβεια της κατάτμησης.

$$F_1 = \frac{2Recall \cdot Precision}{Recall + Precision}, \quad (4.3)$$

Τέλος, χρησιμοποιούμε την *ομοιότητα* (*Similarity*), που ορίζεται ως εξής:

$$Similarity = \frac{tp}{tp + fn + fp}. \quad (4.4)$$

Οι μέθοδοι μας συγκρίνονται με τις μεθόδους από τις εξής δημοσιεύσεις:

- Της δημοσίευσης [9] που βασίζεται στα βίντεο της βάσης ChangeDetection [7]. Η μοντελοποίηση του υποβάθρου γίνεται με ένα πλέγμα τοπικών αυτό-ομοιότητας περιγραφών (*local self-similarity* - *LSS*). Ο περιγραφέας κάθε εικονοστοιχείου με τη μεγαλύτερη συχνότητα επιλέγεται για τη μοντελοποίηση μια περιοχής εικονοστοιχείων υποβάθρου. Για το προσκηνίο, για κάθε εικονοστοιχείο υπολογίζεται η Ευκλείδεια απόσταση μεταξύ του

περιγραφέα του και του περιγραφέα που αντιστοιχεί στο υπόβαθρο και αν είναι μεγαλύτερη ενός κατωφλίου θεωρείται ως εικονοστοιχείο προσκηνίου. Τέλος, χρησιμοποιείται μία τεχνική για την καταστολή των σκιών των αντικειμένων του προσκηνίου (*shadow suppression*) και έχει γίνει επεξεργασία (*post-processing*) στα αποτελέσματα της μεθόδου.

- Της δημοσίευσης [14] που βασίζεται στα βίντεο της βάσης Perception [13] και στηριζόμενη σε τοπικά δυαδικά πρότυπα (*local binary pattern – LBP*) προτείνει αμετάβλητης κλίμακας τοπικά τριαδικά πρότυπα (*scale invariant local ternary pattern – SILTP*) για τη μοντελοποίηση του υποβάθρου. Επειδή αυτά τα τοπικά πρότυπα δεν είναι κοινές αριθμητικές τιμές, δεν μπορούν να μοντελοποιηθούν απευθείας σε πατροπαράδοτες συναρτήσεις πυκνότητας και προτείνουν επίσης ένα πρότυπο εκτίμησης της πυκνότητας πιθανότητας με συνάρτηση πυρήνα (*Pattern Kernel Density Estimation – PKDE*).
- Της δημοσίευσης [15] που βασίζεται στα βίντεο της βάσης Perception [13] και Wallflower [20]. Η απαλοιφή του υποβάθρου βασίζεται στην αυτόματη παραγωγή του μοντέλου του υποβάθρου με μία αυτό-οργανωτική μέθοδο (*SOBS*) χωρίς προγενέστερη γνώση για τα εμπλεκόμενα πρότυπα, σύμφωνα με προκαθορισμένα χαρακτηριστικά όπως τα επίπεδα του γκρι, η κίνηση και το σχήμα. Η δημοσίευση συγκρίνεται με τη μέθοδο *Pfinder* (αντίστοιχη με τη *GMM*), τη *VSAM* (ενσωμάτωση της ανάλυσης των εικονοστοιχείων και των περιοχών για την ανίχνευση κίνησης μέσω μίας πεπερασμένης μηχανής καταστάσεων) όπου χρησιμοποιούνται 19 υπολογιστές παράλληλα και 3 επιμέρους μέθοδοι ανάλογα το υπόβαθρο (στατικό ή δυναμικό κ.λπ.) και τον αλγόριθμο *CB* (αντίστοιχου με το *SOBS*). Όλες οι παραπάνω μέθοδοι, εφαρμόζουν την τεχνική για την καταστολή των σκιών των αντικειμένων του προσκηνίου και κάνουν επεξεργασία στα αποτελέσματα τους.

Ακολουθεί ένας συγκεντρωτικός πίνακας με τα αποτελέσματα των βίντεο δοκιμών σε κάθε βάση. Στον πίνακα αυτό, δίνεται η μέση τιμή και η τυπική απόκλιση.

Πίνακας 4.1 Συγκεντρωτικά αποτελέσματα των μεθόδων σε όλες τις βάσεις.

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
IPPR	0.82 ± 0.05	0.63 ± 0.12	0.84 ± 0.06	0.84 ± 0.06	0.69 ± 0.07	0.69 ± 0.07
Wallflower	0.78 ± 0.19	0.55 ± 0.27	0.79 ± 0.18	0.79 ± 0.18	0.61 ± 0.25	0.61 ± 0.25
Perception	0.63 ± 0.18	0.52 ± 0.16	0.64 ± 0.18	0.64 ± 0.18	0.63 ± 0.16	0.63 ± 0.16
Weizmann	0.82 ± 0.03	0.73 ± 0.06	0.82 ± 0.03	0.82 ± 0.03	0.78 ± 0.05	0.78 ± 0.05
Baseline	0.70 ± 0.14	0.56 ± 0.15	0.71 ± 0.08	0.71 ± 0.08	0.64 ± 0.19	0.64 ± 0.19
Camera Jitter	0.70 ± 0.19	0.72 ± 0.12	0.52 ± 0.17	0.52 ± 0.17	0.76 ± 0.12	0.76 ± 0.12
Dynamic Background	0.84 ± 0.09	0.88 ± 0.13	0.84 ± 0.06	0.84 ± 0.06	0.95 ± 0.04	0.95 ± 0.04
Intermittent Object Motion	0.38 ± 0.20	0.33 ± 0.19	0.40 ± 0.19	0.40 ± 0.19	0.34 ± 0.16	0.34 ± 0.16
Shadow	0.84 ± 0.15	0.66 ± 0.20	0.84 ± 0.16	0.84 ± 0.16	0.67 ± 0.23	0.67 ± 0.23
Precision						
IPPR	0.64 ± 0.05	0.38 ± 0.16	0.64 ± 0.06	0.64 ± 0.06	0.41 ± 0.21	0.41 ± 0.21
Wallflower	0.73 ± 0.17	0.71 ± 0.18	0.73 ± 0.22	0.73 ± 0.22	0.73 ± 0.19	0.73 ± 0.19
Perception	0.65 ± 0.19	0.41 ± 0.25	0.66 ± 0.19	0.66 ± 0.19	0.44 ± 0.27	0.44 ± 0.27
Weizmann	0.76 ± 0.10	0.73 ± 0.06	0.78 ± 0.09	0.78 ± 0.09	0.74 ± 0.07	0.74 ± 0.07
Baseline	0.81 ± 0.13	0.59 ± 0.19	0.80 ± 0.12	0.80 ± 0.12	0.68 ± 0.16	0.68 ± 0.16
Camera Jitter	0.34 ± 0.04	0.15 ± 0.04	0.51 ± 0.18	0.51 ± 0.18	0.16 ± 0.06	0.16 ± 0.06
Dynamic Background	0.16 ± 0.10	0.07 ± 0.05	0.29 ± 0.09	0.20 ± 0.09	0.07 ± 0.04	0.07 ± 0.04
Intermittent Object Motion	0.70 ± 0.19	0.44 ± 0.15	0.68 ± 0.14	0.68 ± 0.14	0.66 ± 0.28	0.66 ± 0.28
Shadow	0.48 ± 0.12	0.37 ± 0.14	0.51 ± 0.12	0.51 ± 0.12	0.51 ± 0.15	0.51 ± 0.15
F₁						
IPPR	0.73 ± 0.03	0.43 ± 0.13	0.74 ± 0.03	0.74 ± 0.03	0.43 ± 0.24	0.43 ± 0.24
Wallflower	0.74 ± 0.14	0.57 ± 0.07	0.75 ± 0.17	0.75 ± 0.17	0.62 ± 0.08	0.62 ± 0.08
Perception	0.62 ± 0.15	0.41 ± 0.16	0.64 ± 0.16	0.64 ± 0.16	0.47 ± 0.19	0.47 ± 0.19
Weizmann	0.78 ± 0.05	0.73 ± 0.03	0.80 ± 0.05	0.80 ± 0.05	0.76 ± 0.02	0.76 ± 0.02
Baseline	0.74 ± 0.07	0.56 ± 0.13	0.75 ± 0.08	0.75 ± 0.08	0.64 ± 0.14	0.64 ± 0.14
Camera Jitter	0.45 ± 0.07	0.24 ± 0.05	0.49 ± 0.09	0.49 ± 0.09	0.26 ± 0.06	0.26 ± 0.06
Dynamic Background	0.26 ± 0.14	0.13 ± 0.08	0.43 ± 0.10	0.43 ± 0.10	0.13 ± 0.08	0.13 ± 0.08
Intermittent Object Motion	0.47 ± 0.17	0.32 ± 0.10	0.49 ± 0.17	0.49 ± 0.17	0.41 ± 0.15	0.41 ± 0.15
Shadow	0.60 ± 0.11	0.45 ± 0.08	0.62 ± 0.09	0.62 ± 0.09	0.54 ± 0.08	0.54 ± 0.08
Similarity						
IPPR	0.56 ± 0.03	0.28 ± 0.08	0.58 ± 0.03	0.58 ± 0.03	0.30 ± 0.14	0.30 ± 0.14
Wallflower	0.61 ± 0.17	0.39 ± 0.07	0.62 ± 0.21	0.62 ± 0.21	0.45 ± 0.08	0.45 ± 0.08
Perception	0.46 ± 0.16	0.26 ± 0.13	0.48 ± 0.17	0.48 ± 0.17	0.31 ± 0.15	0.31 ± 0.15
Weizmann	0.64 ± 0.08	0.54 ± 0.05	0.65 ± 0.08	0.65 ± 0.08	0.59 ± 0.04	0.59 ± 0.04
Baseline	0.56 ± 0.11	0.39 ± 0.13	0.57 ± 0.11	0.57 ± 0.11	0.47 ± 0.16	0.47 ± 0.16
Camera Jitter	0.26 ± 0.04	0.11 ± 0.03	0.28 ± 0.05	0.28 ± 0.05	0.11 ± 0.03	0.11 ± 0.03
Dynamic Background	0.12 ± 0.07	0.06 ± 0.03	0.23 ± 0.07	0.23 ± 0.07	0.06 ± 0.04	0.06 ± 0.04
Intermittent Object Motion	0.31 ± 0.14	0.16 ± 0.03	0.33 ± 0.15	0.33 ± 0.15	0.25 ± 0.14	0.25 ± 0.14
Shadow	0.40 ± 0.13	0.21 ± 0.03	0.42 ± 0.12	0.42 ± 0.12	0.30 ± 0.07	0.30 ± 0.07

Πίνακας 4.2 Συγκριτικά αποτελέσματα Baseline της βάσης ChangeDetection [7].

Ο πίνακας έχει αναπαραχθεί από το [9].

	Recall	Precision	F₁
SOBS	0.9193	0.9313	0.9251
KDE	0.8969	0.9223	0.9092
ViBe	0.8204	0.9288	0.8700
Mahalanobis Distance	0.8872	0.9071	0.8954
Euclidean Distance	0.8385	0.9114	0.8720
GMM KaewTraKulPong	0.5863	0.9532	0.7119
LSS	0.9732	0.7564	0.8494
GMM Zivkovic	0.8085	0.8993	0.8382
GMM Stauffer & Grimson	0.8180	0.8461	0.8245

Πίνακας 4.3 Συγκριτικά αποτελέσματα Shadow της βάσης ChangeDetection [7].

Ο πίνακας έχει αναπαραχθεί από το [9].

	Recall	Precision	F₁
ViBe	0.7833	0.8342	0.8032
KDE	0.8536	0.7660	0.8028
GMM KaewTraKulPong	0.6323	0.8577	0.7176
SOBS	0.8350	0.7219	0.7716
GMM Stauffer & Grimson	0.7956	0.7156	0.7367
GMM Zivkovic	0.7770	0.7232	0.7319
LSS	0.9584	0.4673	0.5951
Euclidean Distance	0.8001	0.6112	0.6785
Mahalanobis Distance	0.7845	0.5685	0.6348

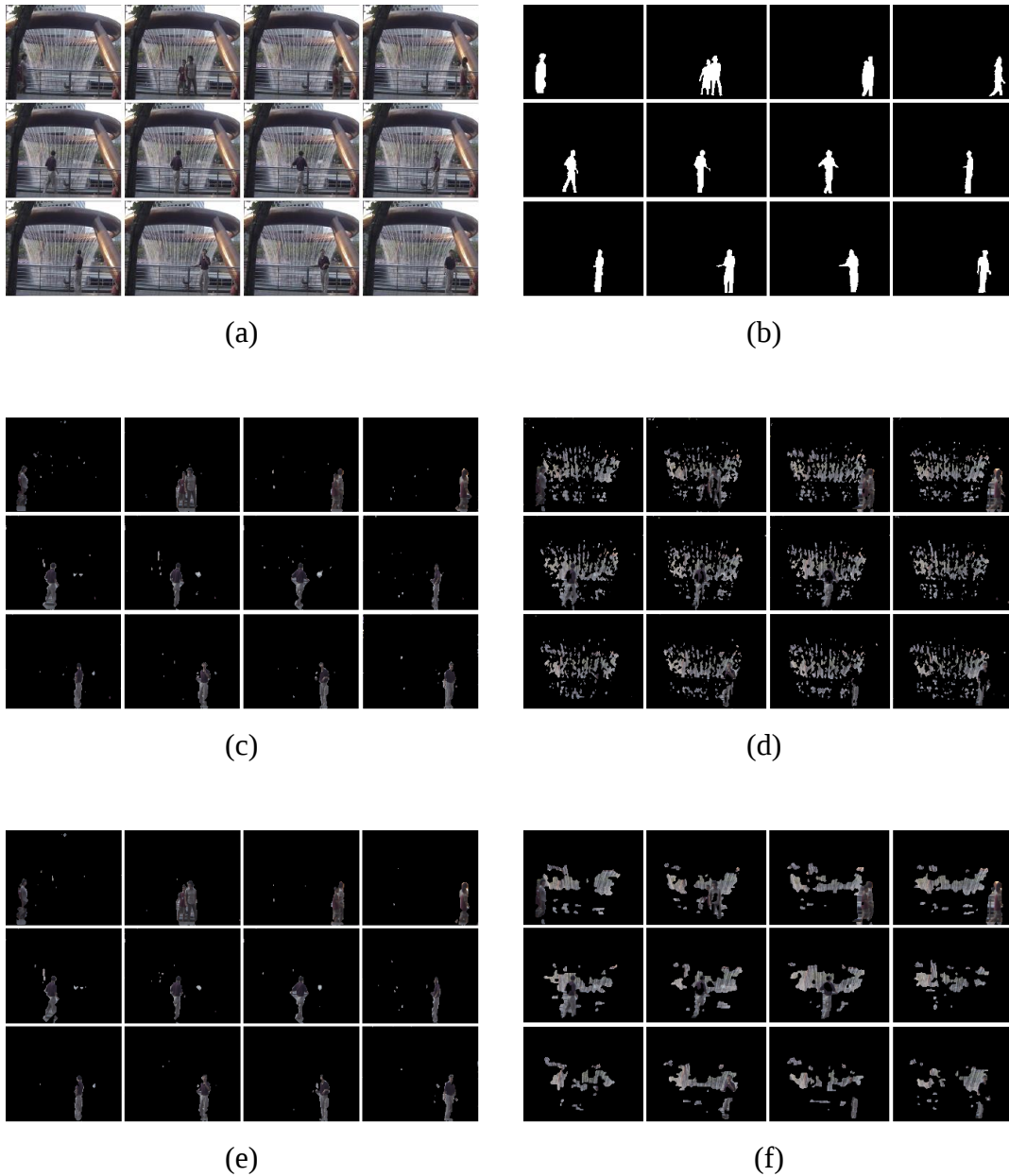
Από τα αποτελέσματα προκύπτουν τα εξής:

- Οι μέθοδοι MRF(GMM), CRF(GMM) δεν έχουν διαφορά στα αποτελέσματα λόγω του ότι χρησιμοποιούν διαφορετικό δυναμικό μεταξύ των γειτονικών εικονοστοιχείων. Το ίδιο ισχύει και με τις μεθόδους MRF(KDE), CRF(KDE).
- Στη βάση IPPR: όπως φαίνεται στον πίνακα 4.1, οι μέθοδοι MRF(GMM) και CRF(GMM) υπερಿಸχύουν των άλλων μεθόδων σε όλους τους δείκτες και τις ποσότητες στο μέσο όρο.
- Στη βάση Wallflower [20]: εδώ τα αποτελέσματα έχουν περισσότερο ενδιαφέρον όπως φαίνεται στον πίνακα 4.1. Οι μέθοδοι MRF(GMM) και CRF(GMM) υπερτερούν στο μέσο όρο με αμελητέα διαφορά, εκτός του δείκτη *Precision* που ισοψηφούν με τις μεθόδους GMM, MRF(KDE) και CRF(KDE). Πιο αναλυτικά (πίνακας Π.2): οι μέθοδοι MRF(KDE) και CRF(KDE) υπερτερούν στο δείκτη *Recall*, στα βίντεο Bootstrap και Waving Trees με μικρή διαφορά λόγω του ότι λαμβάνουν υπ' όψιν το ιστορικό των πλαισίων και έτσι αντιμετωπίζουν καλύτερα δύσκολες καταστάσεις δυναμικού υποβάθρου (κλαδιά ή φύλλα δέντρου που κουνιούνται από τον αέρα). Στο δείκτη *Precision*, στα βίντεο Bootstrap, Camouflage και Foreground Aperture υπερτερούν με σχετικά μεγάλη διαφορά. Τέλος, στην ποσότητα F_1 και στην ομοιότητα κερδίζουν με αρκετή διαφορά μόνο στο βίντεο Bootstrap. Συγκριτικά με τα αποτελέσματα από το [15] (πίνακες Π.3 και Π.4), όσον αφορά το δείκτη *Recall*, στο βίντεο TimeOfDay είμαστε στην προτελευταία θέση, ενώ στο βίντεο WavingTrees καταλαμβάνουμε τη δεύτερη θέση. Στο δείκτη *Precision*, στο βίντεο TimeOfDay βρισκόμαστε στη δεύτερη θέση, ενώ στο βίντεο WavingTrees είμαστε στην τελευταία θέση. Τέλος, στην ποσότητα F_1 και στην ομοιότητα, στο βίντεο TimeOfDay βρισκόμαστε στη δεύτερη και τρίτη θέση αντίστοιχα, ενώ στο βίντεο WavingTrees είμαστε στην προτελευταία θέση.

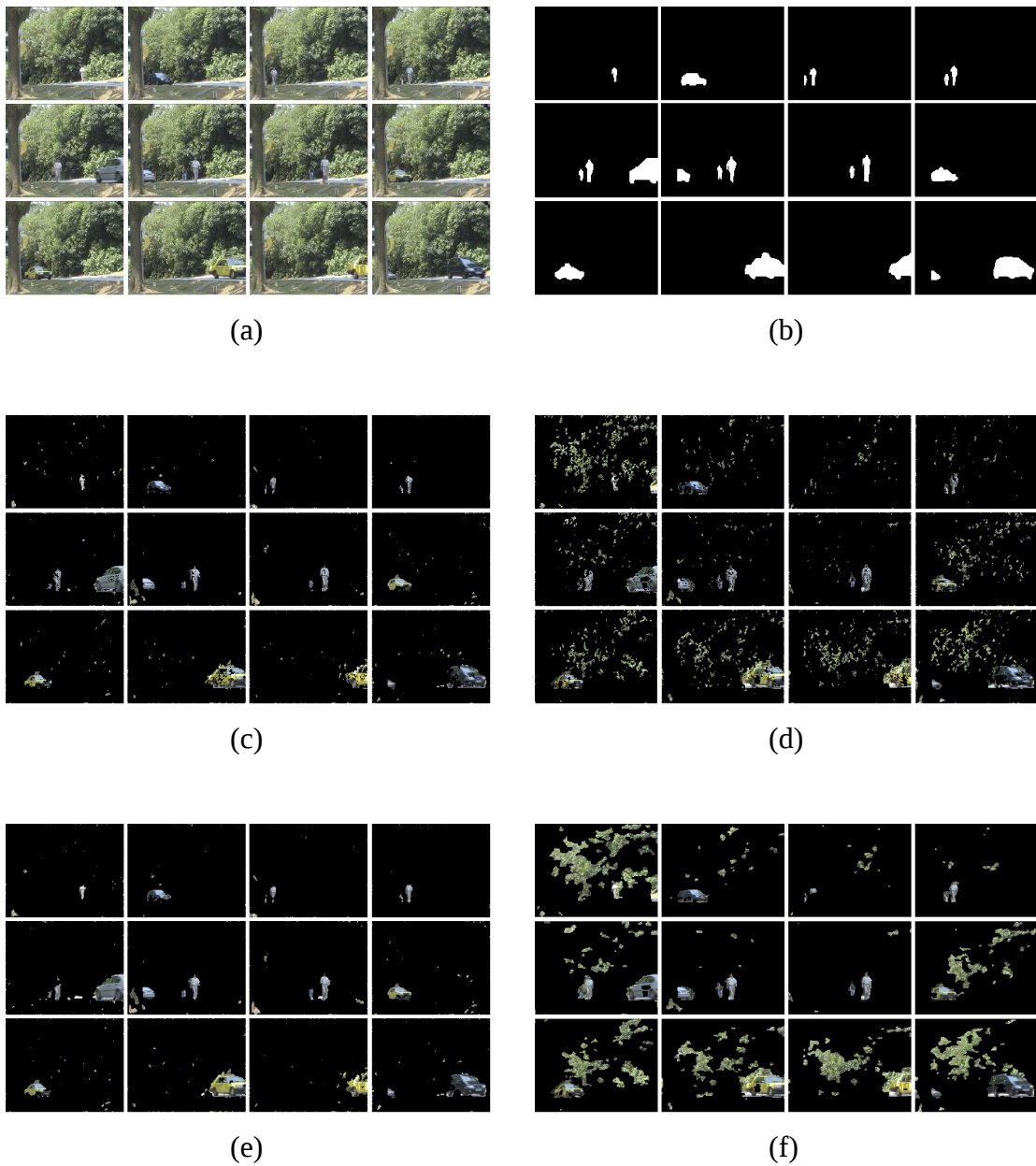
- Στη βάση Perception [13]: εδώ, επιμέρους δεν έχουμε σταθερά αποτελέσματα όπως φαίνεται στον πίνακα 4.1 και πιο αναλυτικά στον πίνακα Π.5. Σε όλους τους δείκτες και τις ποσότητες, υπερισχύουν οι μέθοδοι MRF(GMM) και CRF(GMM) στο μέσο όρο, με επιμέρους νίκες όμως των μεθόδων GMM, MRF(KDE) και CRF(KDE). Οι επιμέρους νίκες των μεθόδων MRF(KDE) και CRF(KDE) οφείλονται στο γεγονός ότι λαμβάνουν υπ' όψιν το ιστορικό των πλαισίων και έτσι αντιμετωπίζουν καλύτερα δύσκολες καταστάσεις που προκύπτουν από δυναμικό υπόβαθρο (κουρτίνα που μετακινείται από τον αέρα, κυλιόμενες σκάλες). Συγκριτικά με τα αποτελέσματα F_1 από το [14] (πίνακας Π.7), έχουμε μέτρια αποτελέσματα όσον αφορά το μέσο όρο (πέμπτη θέση), ενώ μεμονωμένα, στο βίντεο Curtain και στο βίντεο Shoppingmall βρισκόμαστε στην πέμπτη θέση και στο βίντεο Campus και στο βίντεο Fountain βρισκόμαστε στην τρίτη θέση. Συγκριτικά με τα αποτελέσματα ομοιότητας από το [15] (πίνακας Π.6), είμαστε στην τρίτη θέση στο μέσο όρο, ενώ ειδικότερα, στο βίντεο Shoppingmall βρισκόμαστε στη δεύτερη θέση και στο βίντεο Fountain στην πρώτη θέση.
- Στη βάση Weizmann [2], [6]: εδώ όπως φαίνεται στον πίνακα 4.1, εκτός του δείκτη *Recall* όπου έχουμε ισοψηφία μεταξύ των μεθόδων GMM, MRF(GMM) και CRF(GMM) στο μέσο όρο, στα υπόλοιπα υπερτερούν αμελητέα οι μέθοδοι MRF(GMM) και CRF(GMM). Αν και ειδικότερα (πίνακας Π.8), ανάλογα το βίντεο, ενίοτε, υπερισχύει μία από όλες τις μεθόδους ή έχουμε ισοψηφίες.
- Στη βάση ChangeDetection [7]: τα αποτελέσματα ποικίλουν ανάλογα την κατηγορία των βίντεο δοκιμών όπως φαίνεται στον πίνακα 4.1. Πιο αναλυτικά (πίνακες Π.9-Π.13): στην κατηγορία Baseline, στο μέσο όρο εκτός του δείκτη *Precision* όπου υπερτερεί η μέθοδος GMM, υπερτερούν οι μέθοδοι MRF(GMM) και CRF(GMM). Ειδικότερα, έχουμε ενίοτε νίκες των μεθόδων GMM, MRF(KDE) και CRF(KDE) ή ισοψηφίες. Στην κατηγορία Camera Jitter, στο δείκτη *Recall* κερδίζουν οι μέθοδοι MRF(KDE) και CRF(KDE) στο μέσο όρο, με νίκη όμως της μεθόδου GMM στο βίντεο badminton, στο δείκτη *Precision*, υπερισχύουν οι μέθοδοι MRF(GMM) και CRF(GMM) στο μέσο

όρο, στην ποσότητα F_1 υπερτερούν οι μέθοδοι MRF(GMM) και CRF(GMM) στο μέσο όρο και τέλος στην ομοιότητα υπερτερούν οι μέθοδοι MRF(GMM) και CRF(GMM) στο μέσο όρο. Οι επιμέρους νίκες των μεθόδων MRF(KDE) και CRF(KDE) οφείλονται στο γεγονός ότι λαμβάνουν υπ' όψιν το ιστορικό των πλαισίων και έτσι αντιμετωπίζουν καλύτερα δύσκολες καταστάσεις που προκύπτουν από μετατόπιση της κάμερας που προκαλείται από τον άνεμο σε εξωτερικούς χώρους. Στην κατηγορία Dynamic Background εκτός του δείκτη *Recall* όπου υπερτερούν οι μέθοδοι MRF(KDE) και CRF(KDE), έχουμε υπεροχή των μεθόδων MRF(GMM) και CRF(GMM) στο μέσο όρο. Η υπεροχή των μεθόδων MRF(KDE) και CRF(KDE) οφείλονται στο γεγονός ότι λαμβάνουν υπ' όψιν το ιστορικό των πλαισίων και έτσι αντιμετωπίζουν καλύτερα δύσκολες καταστάσεις που προκύπτουν από δυναμικό υπόβαθρο. Στην κατηγορία Intermittent Object Motion, σε όλους τους δείκτες εκτός του δείκτη *Precision* όπου υπερτερεί η μέθοδος GMM, υπερισχύουν οι μέθοδοι MRF(GMM) και CRF(GMM) στο μέσο όρο και σε ορισμένα βίντεο κερδίζουν οι μέθοδοι MRF(KDE) και CRF(KDE). Στην κατηγορία Shadow, σε όλους τους δείκτες και τις ποσότητες στο μέσο όρο υπερισχύουν οι μέθοδοι MRF(GMM) και CRF(GMM). Ειδικότερα, έχουμε υπεροχή σε κάποια βίντεο είτε της μεθόδου GMM είτε των μεθόδων MRF(KDE) και CRF(KDE). Συγκριτικά με τα αποτελέσματα από το [9], όπως φαίνεται στον πίνακα 4.2, στην κατηγορία Baseline, στο δείκτη *Recall* είμαστε με αρκετή διαφορά παραπάνω από την τελευταία θέση με τη μέθοδο GMM-KaweTraKulPong, ενώ στο δείκτη *Precision* και στην ποσότητα F_1 βρισκόμαστε με αρκετή διαφορά παραπάνω από την τελευταία θέση με τη μέθοδο που προτείνεται από τους συγγραφείς τις δημοσίευσης. Στην κατηγορία Shadow όπως φαίνεται στον πίνακα 4.3, στο δείκτη *Recall* καταλαμβάνουμε την τρίτη θέση, ενώ στο δείκτη *Precision* και στην ποσότητα F_1 βρισκόμαστε με αρκετή διαφορά παραπάνω από την τελευταία θέση με τη μέθοδο που προτείνεται από τους συγγραφείς τις δημοσίευσης.

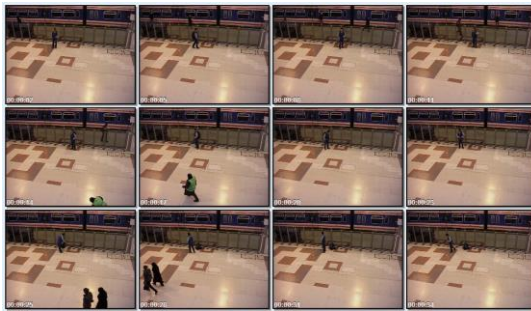
Ακολουθούν ενδεικτικές εικόνες από τα βίντεο δοκιμών με τα αποτελέσματα κάθε μεθόδου και από το βίντεο ή τα πλαίσια επαλήθευσης.



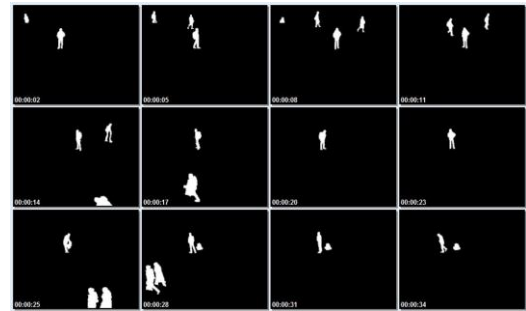
Σχήμα 4.1 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά fountain της βάσης Perception. Η εικονοσειρά έχει 410 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



Σχήμα 4.2 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά campus της βάσης Perception. Η εικονοσειρά έχει 1049 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



(a)



(b)



(c)



(d)



(e)

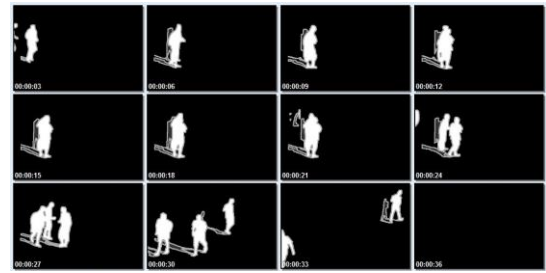


(f)

Σχήμα 4.3 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά Pets2006 της κατηγορίας Baseline της βάσης ChangeDetection. Η εικονοσειρά έχει 901 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



(a)



(b)



(c)



(d)



(e)



(f)

Σχήμα 4.4 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά busStation της κατηγορίας Shadow της βάσης ChangeDetection. Η εικονοσειρά έχει 951 εικονοπλαίσις. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).

ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Γενικότερα, έχω να σχολιάσω ότι ενώ δεν καταφέραμε να είμαστε ανταγωνιστικοί με τις ανταγωνιστικές μεθόδους που υπάρχουν, βγήκαν τα εξής συμπεράσματα:

Οι μέθοδοι MRF(GMM) και CRF(GMM) υπερίσχυσαν γενικά αλλά προέκυψαν και κάποια ενδιαφέροντα στοιχεία για τις μεθόδους MRF(KDE) και CRF(KDE). Στα βίντεο Bootstrap και WavingTrees της βάσης Wallflower, υπερτερούν στο δείκτη *Recall* με μικρή διαφορά λόγω του ότι λαμβάνουν υπ' όψιν το ιστορικό των πλαισίων και έτσι αντιμετωπίζουν καλύτερα δύσκολες καταστάσεις δυναμικού υποβάθρου (κλαδιά ή φύλλα δέντρου που κουνιούνται από τον αέρα). Ισοψηφούν στο δείκτη *Precision* στο μέσο όρο και ειδικότερα στα βίντεο Bootstrap, Camouflage και Foreground Aperture υπερτερούν με σχετικά μεγάλη διαφορά. Τέλος, στην ποσότητα F_1 και στην ομοιότητα κερδίζουν με αρκετή διαφορά μόνο στο βίντεο Bootstrap. Στη βάση Perception, όπου καλούμασταν να αντιμετωπίσουμε δύσκολες καταστάσεις που προκύπταν από δυναμικό υπόβαθρο (κουρτίνα που μετακινείται από τον αέρα, κυλιόμενες σκάλες) είχαμε σε ορισμένες περιπτώσεις υπεροχή τους. Στη βάση ChangeDetection, στην κατηγορία Camera Jitter και Dynamic Background υπερίσχυσαν στο δείκτη *Recall* και αντιμετώπισαν πιο αποτελεσματικά μετατόπιση της κάμερας από τον άνεμο σε εξωτερικούς και από δυναμικό υπόβαθρο αντίστοιχα.

Στο μέλλον, για τη βελτίωση των αποτελεσμάτων, μπορεί να χρησιμοποιηθεί η τεχνική για την καταστολή των σκιών των αντικειμένων του προσκηνίου.

Από αυτή την εργασία, ελπίζω να επωφεληθούν μελλοντικά άτομα που θέλουν να ασχοληθούν με τον τομέα και να μπορέσουν να προχωρήσουν σε μεθόδους πιο αποδοτικές από τις ήδη υπάρχουσες.

ΑΝΑΦΟΡΕΣ

- [1] C. M. Bishop. “*Pattern Recognition and Machine Learning*”. Springer, 2006.
- [2] M. Blank, L. Gorelick, E. Shechtman and M. Irani and R. Basri. “Actions as space-time shapes”. In *10th IEEE International Conference on Computer Vision (ICCV’05)*, Beijing, China, pages 1395-1402, 2005.
- [3] Y. Boykov, O. Veksler, and R. Zabih. “Fast approximate energy minimization via graph cuts”. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222-1239, November 2001.
- [4] B. V. Cherkassky and A. V. Goldberg. “On implementing the push-relabel method for the maximum flow problem”. In *Algorithmica*, 19(4):390-410, 1997.
- [5] L. Ford and D. Fullkerson. “*Flows in Networks*”. Princeton University Press, 1962.
- [6] L. Gorelick, M. Blank, E. Shechtman, M. Irani and R. Basri, “Actions as space-time shapes”. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2247-2253, December, 2007.
- [7] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad and P. Ishwar. “Changedetection.net: A new change detection benchmark dataset”. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, June 2012.
- [8] N. R. Howe and A. Deschamps. “*Better foreground segmentation through graph cuts*”. Technical Report, Smith College, July 2004.
(<http://www.cs.smith.edu/~nhowe/research/code/>)
- [9] J. P. Jodoin, G. A. Bilodeau and N. Saunier. “Background subtraction based on Local Shape”. In *IEEE Computer Vision and Pattern Recognition*, 2012.
- [10] N. Komodakis and G. Tziritas. “Image completion using efficient belief propagation via priority scheduling and dynamic pruning”. In *IEEE Transactions on Image Processing*, 16(11):2649-2661, November 2007.
- [11] S. Kumar and M. Hebert. “Discriminative random fields”. In *International Journal of Computer Vision*, 68(2):179-201, 2006.

- [12] J. D. Lafferty, A. McCallum and F. C. N. Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In *Proc. 18th International Conference on Machine Learning*, pages 282-289, 2001.
- [13] L. Li, W. Huang, I. Gu and Q. Tian. "Foreground object detection from videos containing complex background". In *Proceedings of the 11th ACM International Conference on Multimedia*, pages 2-10, New York, USA, 2003.
- [14] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen and S. Z. Li. "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes". In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, pages 1301-1306, 2010.
- [15] L. Maddalena and A. Petrosino. "A self-organizing approach to background subtraction for visual surveillance applications". In *IEEE Transactions on Image Processing*, volume 17(7):1168-1177, July 2008.
- [16] Z. Mayo and J. R. Tapamo. "Background subtraction survey for highway surveillance". In *Proc. of the Annual Symposium of the Pattern*, 2009.
- [17] A. Mittal and N. Paragios. "Motion-based background subtraction using kernel density estimation". In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, volume 2, pages 302-309, January 2004.
- [18] S. J. D. Prince. "*Computer vision: models, learning and inference*". Cambridge University Press, 2012.
- [19] C. Stauffer and W. E. L. Grimson. "Adaptive background mixture models for real-time tracking". In *12th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, volume 2, pages 246-252, June 1999.
- [20] K. Toyama, J. Krumm, B. Brumitt and B. Meyers. "Wallflower: Principles and practice of background maintenance". In *7th International Conference on Computer Vision*, Corfu, Greece, vol. 1, pp. 255-261, September 1999.

ΠΑΡΑΡΤΗΜΑ

Παρακάτω, δίνονται αναλυτικά τα αποτελέσματα για κάθε βάση και κάθε βίντεο, καθώς και συγκριτικά αποτελέσματα. Στους πίνακες που υπολογίσαμε, δίνεται η μέση τιμή και η τυπική απόκλιση.

Πίνακας Π.1 Αποτελέσματα των μεθόδων στη βάση IPPR.
(<http://archer.ee.nctu.edu.tw/contest/>)

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
data1	0.77	0.67	0.77	0.77	0.68	0.68
data2	0.83	0.73	0.87	0.87	0.76	0.76
data3	0.86	0.50	0.87	0.87	0.63	0.63
Total	0.82 ± 0.05	0.63 ± 0.12	0.84 ± 0.06	0.84 ± 0.06	0.69 ± 0.07	0.69 ± 0.07
Precision						
data1	0.66	0.43	0.68	0.68	0.54	0.54
data2	0.68	0.21	0.67	0.67	0.16	0.16
data3	0.58	0.51	0.58	0.58	0.52	0.52
Total	0.64 ± 0.05	0.38 ± 0.16	0.64 ± 0.06	0.64 ± 0.06	0.41 ± 0.21	0.41 ± 0.21
F₁						
data1	0.71	0.52	0.72	0.72	0.60	0.60
data2	0.75	0.33	0.76	0.76	0.26	0.26
data3	0.69	0.51	0.70	0.70	0.57	0.57
Total	0.73 ± 0.03	0.43 ± 0.13	0.74 ± 0.03	0.74 ± 0.03	0.43 ± 0.24	0.43 ± 0.24
Similarity						
data1	0.57	0.35	0.59	0.59	0.37	0.37
data2	0.59	0.20	0.60	0.60	0.14	0.14
data3	0.53	0.30	0.54	0.54	0.38	0.38
Total	0.56 ± 0.03	0.28 ± 0.08	0.58 ± 0.03	0.58 ± 0.03	0.30 ± 0.14	0.30 ± 0.14

Πίνακας Π.2 Αποτελέσματα των μεθόδων στη βάση Wallflower [20].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
Bootstrap	0.53	0.53	0.56	0.56	0.67	0.67
Camouflage	0.93	0.37	0.95	0.95	0.45	0.45
Foreground Aperture	0.81	0.34	0.82	0.82	0.35	0.35
TimeOfDay	0.64	0.51	0.64	0.64	0.60	0.60
WavingTrees	0.99	1	0.97	0.97	1	1
Total	0.78 ± 0.19	0.55 ± 0.27	0.79 ± 0.18	0.79 ± 0.18	0.61 ± 0.25	0.61 ± 0.25
Precision						
Bootstrap	0.52	0.60	0.43	0.43	0.70	0.70
Camouflage	0.78	0.95	0.77	0.77	0.97	0.97
Foreground Aperture	0.60	0.75	0.60	0.60	0.75	0.75
TimeOfDay	0.95	0.79	0.95	0.95	0.77	0.77
WavingTrees	0.79	0.47	0.90	0.90	0.45	0.45
Total	0.73 ± 0.17	0.71 ± 0.18	0.73 ± 0.22	0.73 ± 0.22	0.73 ± 0.19	0.73 ± 0.19
F₁						
Bootstrap	0.53	0.57	0.49	0.49	0.69	0.69
Camouflage	0.85	0.53	0.85	0.85	0.62	0.62
Foreground Aperture	0.69	0.47	0.69	0.69	0.48	0.48
TimeOfDay	0.77	0.62	0.77	0.77	0.68	0.68
WavingTrees	0.88	0.64	0.93	0.93	0.62	0.62
Total	0.74 ± 0.14	0.57 ± 0.07	0.75 ± 0.17	0.75 ± 0.17	0.62 ± 0.08	0.62 ± 0.08
Similarity						
Bootstrap	0.36	0.39	0.32	0.32	0.52	0.52
Camouflage	0.73	0.36	0.74	0.74	0.44	0.44
Foreground Aperture	0.53	0.30	0.53	0.53	0.31	0.31
TimeOfDay	0.62	0.45	0.62	0.62	0.51	0.51
WavingTrees	0.79	0.47	0.87	0.87	0.45	0.45
Total	0.61 ± 0.17	0.39 ± 0.07	0.62 ± 0.21	0.62 ± 0.21	0.45 ± 0.08	0.45 ± 0.08

Πίνακας Π.3 Συγκριτικά αποτελέσματα βίντεο TimeOfDay βάσης Wallflower [20].

Ο πίνακας έχει αναπαραχθεί από το [15].

	Recall	Precision	F₁	Similarity
SOBS	0.73	0.97	0.83	0.71
Pfinder	0.69	0.86	0.77	0.62
VSAM	0.61	0.95	0.75	0.60
CB	0.72	0.95	0.82	0.69

Πίνακας Π.4 Συγκριτικά αποτελέσματα βίντεο WavingTrees βάσης Wallflower [20].

Ο πίνακας έχει αναπαραχθεί από το [15].

	Recall	Precision	F₁	Similarity
SOBS	0.97	0.98	0.98	0.96
Pfinder	0.67	0.97	0.79	0.66
VSAM	0.99	0.94	0.96	0.93
CB	0.97	0.97	0.97	0.94

Πίνακας Π.5 Αποτελέσματα των μεθόδων στη βάση Perception [13].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
Bootstrap	0.53	0.53	0.56	0.56	0.67	0.67
Campus	0.71	0.72	0.71	0.71	0.91	0.91
Curtain	0.67	0.40	0.76	0.76	0.53	0.53
Escalator	0.48	0.61	0.45	0.45	0.72	0.72
Fountain	0.94	0.61	0.92	0.92	0.62	0.62
Hall	0.43	0.46	0.43	0.43	0.49	0.49
Shoppingmall	0.79	0.61	0.79	0.79	0.71	0.71
Watersurface	0.45	0.21	0.49	0.49	0.39	0.39
Total	0.63 ± 0.18	0.52 ± 0.16	0.64 ± 0.18	0.64 ± 0.18	0.63 ± 0.16	0.63 ± 0.16
Precision						
Bootstrap	0.52	0.60	0.43	0.43	0.70	0.70
Campus	0.58	0.26	0.58	0.58	0.25	0.25
Curtain	0.93	0.81	0.93	0.93	0.85	0.85
Escalator	0.32	0.16	0.42	0.42	0.16	0.16
Fountain	0.71	0.16	0.74	0.74	0.19	0.19
Hall	0.58	0.47	0.58	0.58	0.53	0.53
Shoppingmall	0.71	0.61	0.70	0.70	0.62	0.62
Watersurface	0.85	0.17	0.86	0.86	0.21	0.21
Total	0.65 ± 0.19	0.41 ± 0.25	0.66 ± 0.19	0.66 ± 0.19	0.44 ± 0.27	0.44 ± 0.27
F₁						
Bootstrap	0.53	0.57	0.49	0.49	0.69	0.69
Campus	0.64	0.38	0.64	0.64	0.39	0.39
Curtain	0.78	0.54	0.84	0.84	0.65	0.65
Escalator	0.38	0.25	0.44	0.44	0.26	0.26
Fountain	0.81	0.25	0.82	0.82	0.29	0.29
Hall	0.49	0.47	0.49	0.49	0.51	0.51
Shoppingmall	0.74	0.61	0.74	0.74	0.66	0.66
Watersurface	0.59	0.19	0.62	0.62	0.27	0.27
Total	0.62 ± 0.15	0.41 ± 0.16	0.64 ± 0.16	0.64 ± 0.16	0.47 ± 0.19	0.47 ± 0.19
Similarity						
Bootstrap	0.36	0.39	0.32	0.32	0.52	0.52
Campus	0.46	0.23	0.48	0.48	0.24	0.24
Curtain	0.64	0.35	0.71	0.71	0.41	0.41
Escalator	0.22	0.14	0.26	0.26	0.15	0.15
Fountain	0.68	0.15	0.70	0.70	0.18	0.18
Hall	0.34	0.31	0.35	0.35	0.33	0.33
Shoppingmall	0.59	0.43	0.59	0.59	0.48	0.48
Watersurface	0.41	0.10	0.44	0.44	0.13	0.13
Total	0.46 ± 0.16	0.26 ± 0.13	0.48 ± 0.17	0.48 ± 0.17	0.31 ± 0.15	0.31 ± 0.15

Πίνακας Π.6 Συγκριτικά αποτελέσματα ομοιότητας στη βάση Perception [13].
Ο πίνακας έχει αναπαραχθεί από το [15].

	Bootstrap	Campus	Curtain	Escalator	Fountain	Hall	Shoppingmall	Watersurface	Total
SOBS	0.6019	0.6960	0.8178	0.5770	0.6554	0.5943	0.6677	0.8247	0.6794
BNN	0.4779	0.5256	0.7368	0.4928	0.4636	0.3923	0.5696	0.7540	0.5516
MoG	0.3838	0.0757	0.7580	0.1388	0.6854	0.3335	0.5363	0.7948	0.4633
Li et al.	0.3079	0.1596	0.1841	0.1294	0.0999	0.1135	0.5209	0.0667	0.1978

Πίνακας Π.7 Συγκριτικά αποτελέσματα F_1 (%) στη βάση Perception [13].
Ο πίνακας έχει αναπαραχθεί από το [14].

	MoG	ACMMM03	LBP-B	LBP-P	PKDE _{ltp}	PKDE _{silt}	PKDE _{w=2 mb-silt}	PKDE _{w=3 mb-silt}	PKDE _{w=1+2+3 mb-silt}
Bootstrap	54.07	60.46	52.81	52.00	73.86	75.35	69.45	64.87	72.90
Campus	55.37	75.40	62.85	60.57	42.98	42.54	51.88	61.53	67.83
Curtain	50.53	56.08	66.08	71.42	74.19	91.16	89.37	87.97	92.40
Escalator	36.64	32.95	59.08	53.93	67.71	63.90	64.37	60.18	68.66
Fountain	77.85	56.49	70.52	75.33	81.05	83.45	81.17	77.60	85.04
Hall	57.86	50.18	47.73	50.29	62.13	68.14	65.87	63.60	68.02
Shoppingmall	66.95	67.84	54.67	62.92	73.91	79.62	77.75	74.49	79.65
Watersurface	63.52	63.66	76.80	82.21	41.46	74.30	81.08	83.51	83.15
Total	60.14	59.21	58.69	63.46	67.59	75.35	75.24	73.59	78.69

Πίνακας Π.8 Αποτελέσματα των μεθόδων στη βάση Weizmann [2], [6].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
Original_view_00	0.78	0.72	0.80	0.80	0.71	0.71
Original_bag	0.83	0.69	0.85	0.85	0.82	0.82
Original_briefcase	0.81	0.88	0.81	0.81	0.88	0.88
Original_dog	0.81	0.71	0.78	0.78	0.82	0.82
Original_kneesup	0.80	0.73	0.80	0.80	0.79	0.79
Original_limp	0.85	0.73	0.87	0.87	0.78	0.78
Original_moonwalk	0.85	0.65	0.85	0.85	0.79	0.79
Original_nofeet	0.82	0.72	0.82	0.82	0.83	0.83
Original_normwalk	0.81	0.74	0.82	0.82	0.73	0.73
Original_pole	0.77	0.72	0.78	0.78	0.71	0.71
Original_skirt	0.84	0.77	0.83	0.83	0.76	0.76
Total	0.82 ± 0.03	0.73 ± 0.06	0.82 ± 0.03	0.82 ± 0.03	0.78 ± 0.05	0.78 ± 0.05
Precision						
Original_view_00	0.93	0.85	0.94	0.94	0.86	0.86
Original_bag	0.82	0.75	0.81	0.81	0.73	0.73
Original_briefcase	0.57	0.61	0.60	0.60	0.58	0.58
Original_dog	0.76	0.80	0.79	0.79	0.75	0.75
Original_kneesup	0.71	0.72	0.78	0.78	0.73	0.73
Original_limp	0.82	0.79	0.83	0.83	0.76	0.76
Original_moonwalk	0.82	0.73	0.84	0.84	0.71	0.71
Original_nofeet	0.66	0.71	0.66	0.66	0.70	0.70
Original_normwalk	0.75	0.69	0.73	0.73	0.79	0.79
Original_pole	0.74	0.69	0.74	0.74	0.80	0.80
Original_skirt	0.73	0.73	0.81	0.81	0.74	0.74
Total	0.76 ± 0.10	0.73 ± 0.06	0.78 ± 0.09	0.78 ± 0.09	0.74 ± 0.07	0.74 ± 0.07
F₁						
Original_view_00	0.85	0.78	0.86	0.86	0.78	0.78
Original_bag	0.81	0.72	0.83	0.83	0.77	0.77
Original_briefcase	0.67	0.72	0.69	0.69	0.70	0.70
Original_dog	0.78	0.75	0.79	0.79	0.78	0.78
Original_kneesup	0.75	0.73	0.79	0.79	0.76	0.76
Original_limp	0.84	0.76	0.85	0.85	0.77	0.77
Original_moonwalk	0.84	0.69	0.85	0.85	0.75	0.75
Original_nofeet	0.73	0.72	0.73	0.73	0.76	0.76
Original_normwalk	0.78	0.71	0.78	0.78	0.76	0.76
Original_pole	0.76	0.71	0.76	0.76	0.75	0.75
Original_skirt	0.78	0.75	0.82	0.82	0.75	0.75
Total	0.78 ± 0.05	0.73 ± 0.03	0.80 ± 0.05	0.80 ± 0.05	0.76 ± 0.02	0.76 ± 0.02
Similarity						
Original_view_00	0.74	0.63	0.76	0.76	0.64	0.64
Original_bag	0.70	0.52	0.71	0.71	0.60	0.60
Original_briefcase	0.48	0.53	0.49	0.49	0.50	0.50
Original_dog	0.63	0.54	0.63	0.63	0.60	0.60
Original_kneesup	0.60	0.52	0.66	0.66	0.59	0.59
Original_limp	0.71	0.59	0.72	0.72	0.61	0.61
Original_moonwalk	0.71	0.49	0.72	0.72	0.57	0.57
Original_nofeet	0.56	0.49	0.56	0.56	0.58	0.58
Original_normwalk	0.62	0.50	0.62	0.62	0.60	0.60
Original_pole	0.60	0.50	0.60	0.60	0.59	0.59
Original_skirt	0.63	0.58	0.68	0.68	0.57	0.57
Total	0.64 ± 0.08	0.54 ± 0.05	0.65 ± 0.08	0.65 ± 0.08	0.59 ± 0.04	0.59 ± 0.04

Πίνακας Π.9 Αποτελέσματα των μεθόδων στη βάση Baseline ChangeDetection [7].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
highway	0.71	0.64	0.74	0.74	0.75	0.75
office	0.58	0.36	0.63	0.63	0.37	0.37
pedestrians	0.89	0.71	0.80	0.80	0.79	0.79
Pets2006	0.62	0.53	0.65	0.65	0.65	0.65
Total	0.70 ± 0.14	0.56 ± 0.15	0.71 ± 0.08	0.71 ± 0.08	0.64 ± 0.19	0.64 ± 0.19
Precision						
highway	0.97	0.86	0.95	0.95	0.89	0.89
office	0.84	0.55	0.78	0.78	0.62	0.62
pedestrians	0.67	0.52	0.79	0.79	0.52	0.52
Pets2006	0.76	0.43	0.67	0.67	0.67	0.67
Total	0.81 ± 0.13	0.59 ± 0.19	0.80 ± 0.12	0.80 ± 0.12	0.68 ± 0.16	0.68 ± 0.16
F₁						
highway	0.82	0.73	0.83	0.83	0.81	0.81
office	0.69	0.44	0.70	0.70	0.46	0.46
pedestrians	0.77	0.60	0.80	0.80	0.63	0.63
Pets2006	0.68	0.48	0.66	0.66	0.66	0.66
Total	0.74 ± 0.07	0.56 ± 0.13	0.75 ± 0.08	0.75 ± 0.08	0.64 ± 0.14	0.64 ± 0.14
Similarity						
highway	0.70	0.58	0.71	0.71	0.68	0.68
office	0.46	0.29	0.47	0.47	0.30	0.30
pedestrians	0.60	0.37	0.61	0.61	0.40	0.40
Pets2006	0.48	0.31	0.48	0.48	0.48	0.48
Total	0.56 ± 0.11	0.39 ± 0.13	0.57 ± 0.11	0.57 ± 0.11	0.47 ± 0.16	0.47 ± 0.16

Πίνακας Π.10 Αποτελέσματα των μεθόδων στη βάση Camera Jitter ChangeDetection [7].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
badminton	0.87	0.55	0.39	0.39	0.59	0.59
boulevard	0.78	0.76	0.76	0.76	0.81	0.81
sidewalk	0.43	0.76	0.42	0.42	0.80	0.80
traffic	0.72	0.82	0.52	0.52	0.84	0.84
Total	0.70 ± 0.19	0.72 ± 0.12	0.52 ± 0.17	0.52 ± 0.17	0.76 ± 0.12	0.76 ± 0.12
Precision						
badminton	0.31	0.18	0.72	0.72	0.22	0.22
boulevard	0.35	0.09	0.41	0.41	0.10	0.10
sidewalk	0.30	0.15	0.32	0.32	0.15	0.15
traffic	0.39	0.16	0.59	0.59	0.17	0.17
Total	0.34 ± 0.04	0.15 ± 0.04	0.51 ± 0.18	0.51 ± 0.18	0.16 ± 0.06	0.16 ± 0.06
F₁						
badminton	0.46	0.27	0.51	0.51	0.32	0.32
boulevard	0.48	0.16	0.53	0.53	0.18	0.18
sidewalk	0.35	0.25	0.36	0.36	0.25	0.25
traffic	0.51	0.27	0.55	0.55	0.28	0.28
Total	0.45 ± 0.07	0.24 ± 0.05	0.49 ± 0.09	0.49 ± 0.09	0.26 ± 0.06	0.26 ± 0.06
Similarity						
badminton	0.29	0.11	0.29	0.29	0.11	0.11
boulevard	0.25	0.07	0.26	0.26	0.08	0.08
sidewalk	0.21	0.09	0.22	0.22	0.10	0.10
traffic	0.29	0.15	0.33	0.33	0.15	0.15
Total	0.26 ± 0.04	0.11 ± 0.03	0.28 ± 0.05	0.28 ± 0.05	0.11 ± 0.03	0.11 ± 0.03

Πίνακας Π.11 Αποτελέσματα των μεθόδων στη βάση Dynamic Background ChangeDetection [7].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
canoe	0.74	0.97	0.77	0.77	0.97	0.97
fountain01	0.89	0.91	0.89	0.89	0.95	0.95
fountain02	0.93	0.94	0.90	0.90	0.97	0.97
overpass	0.78	0.69	0.81	0.81	0.89	0.89
Total	0.84 ± 0.09	0.88 ± 0.13	0.84 ± 0.06	0.84 ± 0.06	0.95 ± 0.04	0.95 ± 0.04
Precision						
canoe	0.26	0.09	0.26	0.26	0.09	0.09
fountain01	0.03	0.01	0.32	0.32	0.01	0.01
fountain02	0.17	0.06	0.39	0.39	0.06	0.06
overpass	0.18	0.12	0.18	0.18	0.11	0.11
Total	0.16 ± 0.10	0.07 ± 0.05	0.29 ± 0.09	0.20 ± 0.09	0.07 ± 0.04	0.07 ± 0.04
F₁						
canoe	0.39	0.17	0.39	0.39	0.17	0.17
fountain01	0.06	0.02	0.47	0.47	0.02	0.02
fountain02	0.29	0.11	0.54	0.54	0.11	0.11
overpass	0.29	0.20	0.30	0.30	0.20	0.20
Total	0.26 ± 0.14	0.13 ± 0.08	0.43 ± 0.10	0.43 ± 0.10	0.13 ± 0.08	0.13 ± 0.08
Similarity						
canoe	0.20	0.09	0.20	0.20	0.09	0.09
fountain01	0.03	0.01	0.27	0.27	0.01	0.01
fountain02	0.12	0.05	0.30	0.30	0.06	0.06
overpass	0.13	0.07	0.14	0.14	0.09	0.09
Total	0.12 ± 0.07	0.06 ± 0.03	0.23 ± 0.07	0.23 ± 0.07	0.06 ± 0.04	0.06 ± 0.04

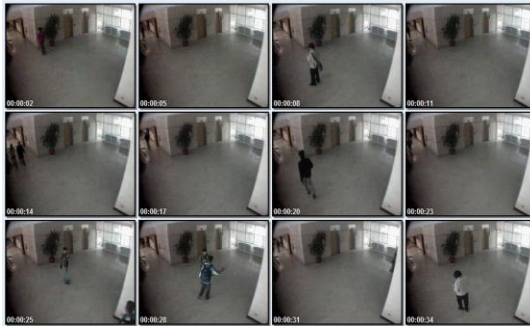
Πίνακας Π.12 Αποτελέσματα των μεθόδων στη βάση Intermittent Object Motion ChangeDetection [7].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
abandonedBox	0.27	0.26	0.27	0.27	0.29	0.29
parking	0.57	0.59	0.61	0.61	0.54	0.54
sofa	0.25	0.18	0.26	0.26	0.21	0.21
streetLight	0.29	0.26	0.37	0.37	0.27	0.27
tramstop	0.21	0.14	0.23	0.23	0.18	0.18
winterDriveway	0.68	0.56	0.66	0.66	0.55	0.55
Total	0.38 ± 0.20	0.33 ± 0.19	0.40 ± 0.19	0.40 ± 0.19	0.34 ± 0.16	0.34 ± 0.16
Precision						
abandonedBox	0.43	0.37	0.62	0.62	0.42	0.42
parking	0.97	0.47	0.97	0.97	0.98	0.98
sofa	0.87	0.53	0.61	0.61	0.87	0.87
streetLight	0.65	0.43	0.64	0.64	0.53	0.53
tramstop	0.62	0.64	0.61	0.61	0.87	0.87
winterDriveway	0.63	0.19	0.65	0.65	0.31	0.31
Total	0.70 ± 0.19	0.44 ± 0.15	0.68 ± 0.14	0.68 ± 0.14	0.66 ± 0.28	0.66 ± 0.28
F₁						
abandonedBox	0.33	0.31	0.38	0.38	0.34	0.34
parking	0.72	0.52	0.75	0.75	0.70	0.70
sofa	0.39	0.27	0.37	0.37	0.34	0.34
streetLight	0.40	0.32	0.47	0.47	0.36	0.36
tramstop	0.31	0.23	0.33	0.33	0.30	0.30
winterDriveway	0.65	0.28	0.66	0.66	0.40	0.40
Total	0.47 ± 0.17	0.32 ± 0.10	0.49 ± 0.17	0.49 ± 0.17	0.41 ± 0.15	0.41 ± 0.15
Similarity						
abandonedBox	0.21	0.20	0.23	0.23	0.22	0.22
parking	0.56	0.14	0.60	0.60	0.53	0.53
sofa	0.23	0.16	0.22	0.22	0.20	0.20
streetLight	0.24	0.17	0.29	0.29	0.20	0.20
tramstop	0.20	0.14	0.22	0.22	0.17	0.17
winterDriveway	0.40	0.12	0.40	0.40	0.18	0.18
Total	0.31 ± 0.14	0.16 ± 0.03	0.33 ± 0.15	0.33 ± 0.15	0.25 ± 0.14	0.25 ± 0.14

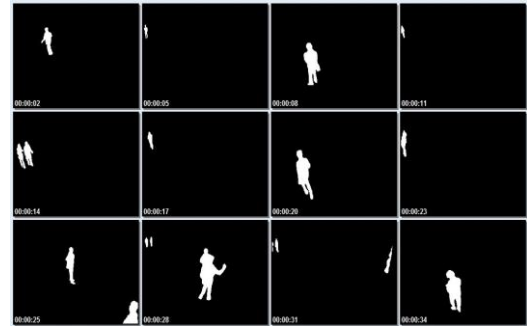
Πίνακας Π.13 Αποτελέσματα των μεθόδων στη βάση Shadow ChangeDetection [7].

	GMM	KDE	MRF(GMM)	CRF(GMM)	MRF(KDE)	CRF(KDE)
Recall						
backdoor	0.92	0.84	0.91	0.91	0.90	0.90
bungalows	0.96	0.82	0.96	0.96	0.83	0.83
busStation	0.63	0.43	0.61	0.61	0.42	0.42
peopleInShade	0.84	0.54	0.86	0.86	0.52	0.52
Total	0.84 ± 0.15	0.66 ± 0.20	0.84 ± 0.16	0.84 ± 0.16	0.67 ± 0.23	0.67 ± 0.23
Precision						
backdoor	0.31	0.30	0.34	0.34	0.31	0.31
bungalows	0.53	0.28	0.54	0.54	0.49	0.49
busStation	0.48	0.33	0.58	0.58	0.56	0.56
peopleInShade	0.58	0.57	0.58	0.58	0.67	0.67
Total	0.48 ± 0.12	0.37 ± 0.14	0.51 ± 0.12	0.51 ± 0.12	0.51 ± 0.15	0.51 ± 0.15
F₁						
backdoor	0.46	0.44	0.50	0.50	0.46	0.46
bungalows	0.68	0.42	0.69	0.69	0.62	0.62
busStation	0.55	0.37	0.60	0.60	0.48	0.48
peopleInShade	0.69	0.56	0.69	0.69	0.59	0.59
Total	0.60 ± 0.11	0.45 ± 0.08	0.62 ± 0.09	0.62 ± 0.09	0.54 ± 0.08	0.54 ± 0.08
Similarity						
backdoor	0.25	0.19	0.27	0.27	0.24	0.24
bungalows	0.51	0.19	0.52	0.52	0.40	0.40
busStation	0.33	0.20	0.37	0.37	0.25	0.25
peopleInShade	0.49	0.26	0.50	0.50	0.32	0.32
Total	0.40 ± 0.13	0.21 ± 0.03	0.42 ± 0.12	0.42 ± 0.12	0.30 ± 0.07	0.30 ± 0.07

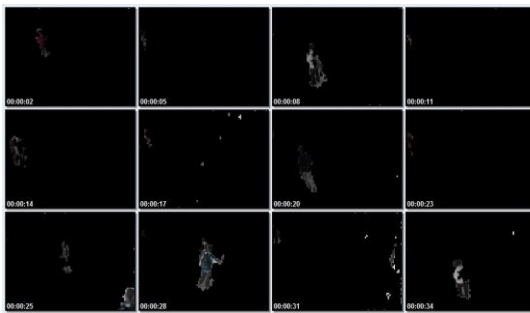
Ακολουθούν εικόνες με τα αποτελέσματα κάθε μεθόδου.



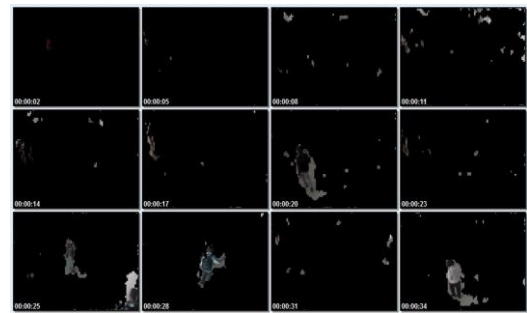
(a)



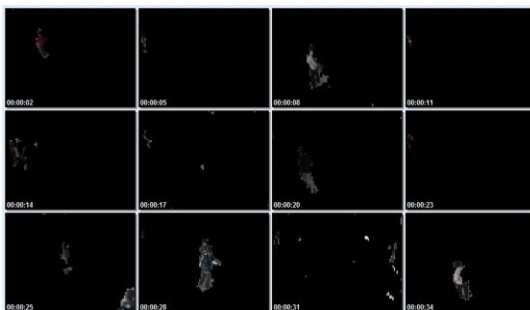
(b)



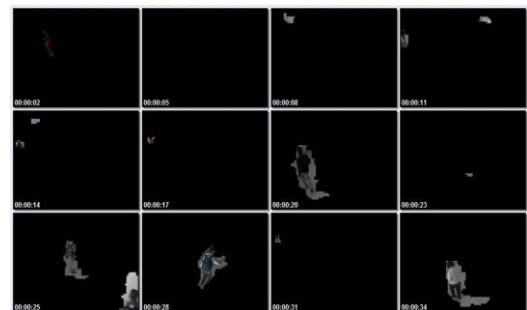
(c)



(d)



(e)

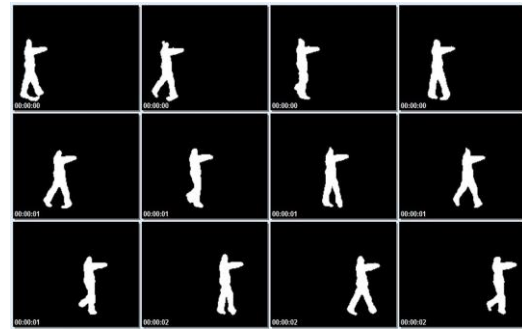


(f)

Σχήμα Π.1 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά data1 της βάσης IPPR. Η εικονοσειρά έχει 300 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



(a)



(b)



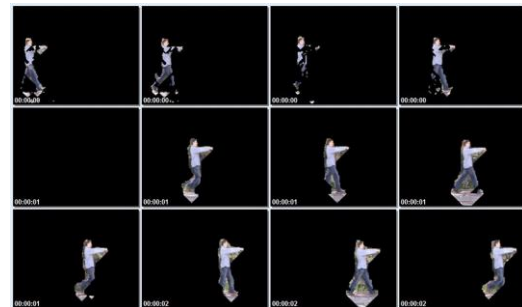
(c)



(d)



(e)



(f)

Σχήμα Π.2 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά original_moonwalk της βάσης Weizmann. Η εικονοσειρά έχει 66 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



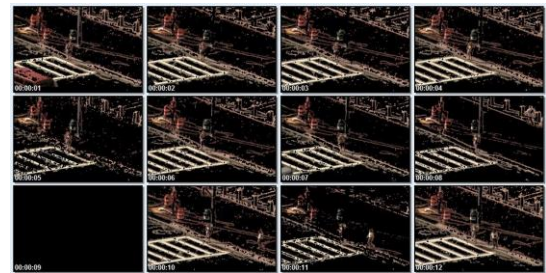
(a)



(b)



(c)



(d)



(e)

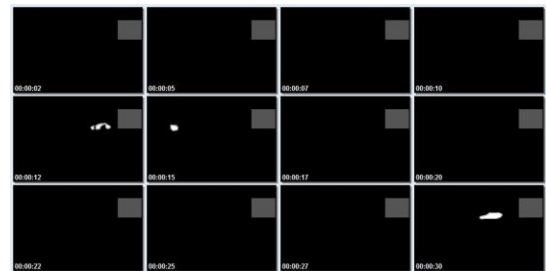


(f)

Σχήμα Π.3 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά sidewalk της κατηγορίας Camera Jitter της βάσης ChangeDetection. Η εικονοσειρά έχει 401 εικονοπλαίσια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



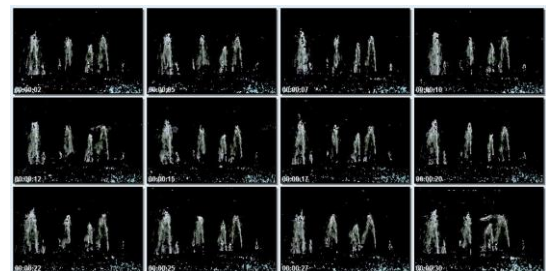
(a)



(b)



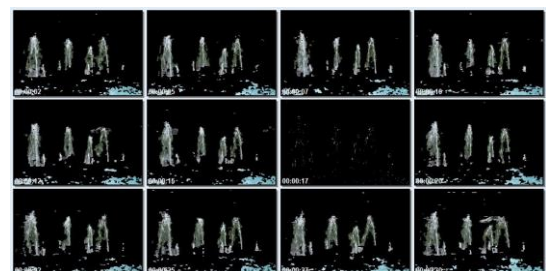
(c)



(d)



(e)

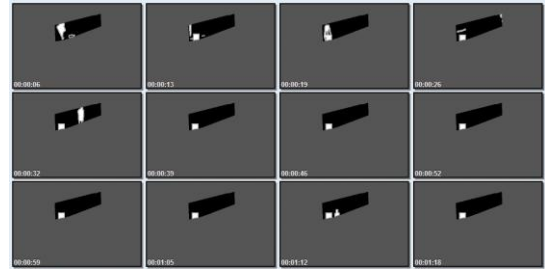


(f)

Σχήμα Π.4 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά fountain01 της κατηγορίας Dynamic Background της βάσης ChangeDetection. Η εικονοσειρά έχει 785 εικονοπλαίσις. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).



(a)



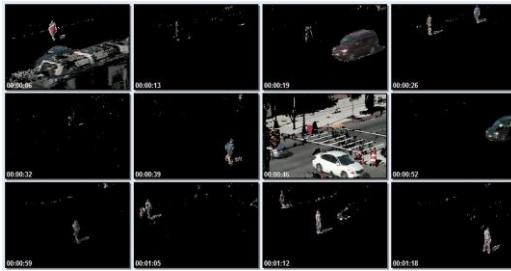
(b)



(c)



(d)



(e)



(f)

Σχήμα Π.5 Αντιπροσωπευτικά αποτελέσματα στην εικονοσειρά abandonedBox της κατηγορίας Intermittent Object Motion της βάσης ChangeDetection. Η εικονοσειρά έχει 2051 εικονοπλαίδια. (a) Εικονοσειρά, (b) Ground Truth, (c) GMM, (d) KDE, (e) CRF(GMM) και (f) CRF(KDE).

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

ΠΡΟΣΩΠΙΚΑ ΣΤΟΙΧΕΙΑ

Όνοματεπώνυμο: Κωνσταντίνος Καλογερόπουλος
Ημερομηνία Γέννησης: 05/08/1988
Διεύθυνση Κατοικίας: Σφακτηρίας 7, 24100, Καλαμάτα
Τηλέφωνο: 6978707533
Ηλεκτρονικό Ταχυδρομείο: kouloumout@hotmail.gr

ΕΚΠΑΙΔΕΥΣΗ

- 2006-2011
- Πτυχίο Πληροφορικής
- Τμήμα Πληροφορικής, Σχολή Θετικών Επιστημών, Πανεπιστήμιο Ιωαννίνων

ΞΕΝΕΣ ΓΛΩΣΣΕΣ

- Αγγλικά (*First Certificate in English*)

ΕΡΕΥΝΗΤΙΚΑ ΕΝΔΙΑΦΕΡΟΝΤΑ

- Επεξεργασία Εικόνas και Πολυμέσων
- Ευρύτερο ενδιαφέρον στους τομείς της τεχνολογίας και των εφαρμογών

ΑΝΑΦΟΡΕΣ

ΠΑΡΑΡΤΗΜΑ

ΒΙΟΓΡΑΦΙΚΟ

