Παρακολούθηση αντικειμένων σε εικονοσειρές με μικτές κανονικές κατανομές

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης

του Τμήματος Πληροφορικής Εξεταστική Επιτροπή

από τον

Βασίλειο Καραβασίλη

ως μέρος των Υποχρεώσεων για τη λήψη του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ

ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ – ΕΦΑΡΜΟΓΕΣ

Ιανουάριος 2009

Περιεχομενα

Ευρετηριο Σχηματων

Ευρετήριο Πινάκων

Ευρετηριο Αλγοριθμων

Περιληψη

Βασίλειος Καραβασίλης του Χρήστου και της Ειρήνης. MSc, Τμήμα Πληροφορικής Πανεπιστημίου Ιωαννίνων, Φεβρουάριος, 2009. Παρακολούθηση αντικειμένων σε εικονοσειρές με μικτές κανονικές κατανομές. Επιβλέποντας: Χριστόφορος Νίκου

Η εργασία διαπραγματεύεται την παρακολούθηση της κίνησης (tracking) αντικειμένων σε μεγάλες σειρές εικόνων (video). Στο πρώτο μέρος, μελετώνται τα φίλτρα Kalman που αποτελούν ένα βασικό μοντέλο για την παρακολούθηση αντικειμένων και εφαρμόζονται σε περιπτώσεις που υπάρχει εκ των προτέρων γνώση για την μορφή της κίνησης και ο θόρυβος που επηρεάζει τόσο τη μέτρηση όσο και την κατάσταση του αντικειμένου είναι κανονικής κατανομής. Έπειτα αναλύεται ο αλγόριθμος διάδοσης της υπό συνθήκης πιθανότητας (Conditional Density Propagation – CONDENSATION) που αποτελεί μια γενικευμένη περίπτωση των φίλτρων Kalman για περιπτώσεις που ο θόρυβος δεν ακολουθεί κανονική κατανομή. Στο δεύτερο μέρος παρουσιάζονται εκτιμητές της κίνησης που βασίζονται στην προσέγγιση χαρακτηριστικών του αντικειμένου από ιστογράμματα. Οι εκτιμητές αυτοί προσδιορίζουνε τη θέση του αντικειμένου σε κάθε εικόνα με την ελαχιστοποίηση μια συνάρτησης απόστασης μεταξύ του προτύπου ιστογράμματος του αντικειμένου και των ιστογραμμάτων που προκύπτουν από τις πιθανές θέσεις του αντικειμένου στις διάφορες εικόνες. Σε αυτή την κατηγορία ανήκει ο αλγόριθμος μέσης μετατόπισης (Mean Shift) και ο αλγόριθμος DEMD (Differential EMD). Το πλεονέκτημα αυτών των αλγορίθμων είναι ότι δεν χρειάζεται καμία γνώση για την μορφή της κίνησης του αντικειμένου. Σε αυτό το πλαίσιο, προτείνεται η επέκταση του αλγορίθμου DEMD, με αναπαράσταση του αντικειμένου από μικτές κανονικές κατανομές. Η επέκταση αυτή οδηγεί δε εξίσου αποτελεσματική αλλά ταχύτερη εκτίμηση της κίνησης του αντικειμένου σε σχέση με τον κλασσικό αλγόριθμο.

Extended Abstract in English

Karavasilis, Vasileios. MSc, Computer Science Department, University of Ioannina, Greece. February, 2009. Object tracking in image sequences using Gaussian Mixture model. Thesis Supervisor: Christoforos Nikou.

One important field in computer vision is object tracking. Tracking is the problem of generating an inference about the motion of an object given a sequence of images. Solutions to this problem have a variety of applications, some of them being: surveillance, where we are looking for unusual movements, targeting and recognition from motion. In tracking problems we assume that we know the model of the object and in every image we can get a set of measurements (provided from the object of interest or any other object). Based on that, we want to find the object's true position. In this work, we review some methods that have been proposed and suggest a new one.

In the first category of methods we assume that the moving object has an internal state which is measured in some way. Combining those measurements efficiently, we can get the object's true position. The first method of that category is the Kalman filter. In order to apply Kalman filtering we must know the type of object's movement and assume that the noise which affects the object and the observation is Gaussian. Kalman filters successfully track objects even in the case of occlusions if the assumed type of movement is correctly modeled. An other family of method is Condensation and ICondensation algorithms. These are more general than Kalman filters, because they do not assume a specific type of densities and have the ability to predict an object's location again if it is occluded.

The above methods have the disadvantage that we must assume the type of object's movement. Other methods that are based on histogram representation of objects do not have this drawback. In this category, we choose some features of the object (color, texture) to create the histograms. Usually, the features are spatially masked with an isotopic kernel. Estimators in this category find the object's position by minimizing the cost function between the models histogram and the candidate histograms in the next image. The first method of this category is Kernel-Based Object Tracking using Mean Shift algorithm. In this method, the object is supposed to be inside an ellipse and the histogram is constructed from pixel values inside that ellipse. The initial histogram is known. In the next, image we initialize the center and we use Mean Shift to move the center to the true object's center (the method uses the Bhattacharyya coefficient). Another similar method is Differential Earth Mover's Distance (EMD). In this method the object is also represented by histograms in an ellipse, but the distance among them is the Earth Mover's Distance.

Finally, the new method that is proposed in this work is based on the Differential EMD algorithm. The object is bounded by an ellipse and the features are spatially masked with a kernel. The features of the object are modeled by a Gaussian Mixture Model instead of a simple histogram. This approach is as efficient as the original method but leads to significant improvement in terms of execution time.

Κεφαλαιο 1

Εισαγωγη

1.1 Εισαγωγή

Η ανίχνευση της κίνησης των αντικειμένων (tracking) είναι ένα σημαντικό κομμάτι της περιοχής της υπολογιστικής όρασης. Οι αυξανόμενες δυνατότητες των υπολογιστών και των ψηφιακών καμερών σε συνάρτηση με το μειωμένο οικονομικό κόστος και η απαίτηση να υπάρχει αυτοματοποιημένη ανάλυση σε εικονοσειρές έχουν οδηγήσει σε ανάπτυξη την περιοχή της υπολογιστικής όρασης. Υπάρχουν τρία βήματα για την ανάλυση των εικονοσειρών: ο καθορισμός του αντικειμένου που μας ενδιαφέρει, η ανίχνευση της θέσης του από εικόνα σε εικόνα και η εξαγωγή συμπερασμάτων από την κίνηση. Επομένως η ανίχνευση κινούμενων αντικειμένων είναι σχετική με εργασίες όπως:

- αναγνώριση με βάση την κίνηση, όπως για παράδειγμα η αναγνώριση ανθρώπων με βάση τον βηματισμό.
- αυτοματοποιημένη επίβλεψη, δηλαδή η παρακολούθηση μιας περιοχής για ύποπτες κινήσεις ή απίθανα γεγονότα.
- επικοινωνία ανθρώπου με υπολογιστή, όπως για παράδειγμα η ανίχνευση κινήσεων των χεριών για την καθοδήγηση του υπολογιστή.
- επίβλεψη της κυκλοφορίας, δηλαδή παρακολούθηση σε πραγματικό χρόνο της ροής οχημάτων και καταγραφή στατιστικών στοιχείων.
- καθοδήγηση οχήματος, δηλαδή εύρεση πορείας και αποφυγή εμποδίων με βάση τις παρατηρήσεις από κάμερα.

- κίνηση χαρακτήρων, δηλαδή καταγραφή της κίνηση από ένα μοντέλο και δημιουργία ενός φανταστικού χαρακτήρα που εκτελεί αυτή την κίνηση.

- πρόβλεψη της επόμενης θέσης ενός αντικειμένου με βάση την μέχρι τώρα πορεία του.

Στην πιο απλή περίπτωση η ανίχνευση της κίνησης ορίζεται ως η εκτίμηση της τροχιάς ενός αντικειμένου που κινείται σε μια περιοχή και αποτυπώνεται σε μια σειρά εικόνων. Με άλλα λόγια πρέπει να βρεθεί η θέση του αντικειμένου μέσα σε διαδοχικές εικόνες. Επιπροσθέτως μπορεί να χρειάζεται να βρεθεί και η κατεύθυνση του αντικειμένου. Η ανίχνευση του αντικειμένου μπορεί να είναι δύσκολη για διάφορους λόγους:

- απώλεια πληροφορίας κατά την προβολή από τις τρείς διαστάσεις του πραγματικού κόσμου στις δύο διαστάσεις της εικόνας.

- θόρυβος κατά την αποτύπωση της εικόνας.

- σύνθετη κίνηση του αντικειμένου.

- μη συμπαγή αντικείμενα.

- αλλαγή στον φωτισμό της σκηνής.

- απαιτήσεις πραγματικού χρόνου.

Η ανίχνευση μπορεί να γίνει βάζοντας περιορισμούς στο μοντέλο της κίνησης ή στην μορφή του αντικειμένου. Για παράδειγμα όλες οι μέθοδοι υποθέτουν ότι η κίνηση είναι ομαλή χωρίς απότομες αλλαγές. Ένας επιπλέον περιορισμός είναι η υπόθεση για σταθερή ταχύτητα ή σταθερή επιτάχυνση λόγο της εκ των προτέρων γνώσης μας για την κίνηση των αντικειμένων. Εκ ων προτέρων γνώση για το πλήθος των αντικειμένων ή την μορφή των αντικειμένων μπορεί επίσης να οδηγήσει σε απλούστευση του προβλήματος.

Διάφορες προσεγγίσεις για την ανίχνευση κίνησης έχουν προταθεί. Η διάκριση μεταξύ τους γίνεται με βάση τις απαντήσεις που δίνουν στα ακόλουθα ερωτήματα: Ποια αναπαράσταση είναι κατάλληλη για το αντικείμενο; Ποια χαρακτηριστικά του αντικειμένου θα λάβουμε υπόψη; Πώς θα μοντελοποιηθεί η κίνηση και το σχήμα του αντικειμένου; Οι απαντήσεις σε αυτά τα ερωτήματα εξαρτώνται από περιβάλλον που γίνεται η ανίχνευση και από τη χρήση που θέλουμε να κάνουμε. Πολλές μέθοδοι έχουν προταθεί και επιχειρούν να απαντήσουν στα παραπάνω ερωτήματα για ένα πλήθος σεναρίων. Ο σκοπός της εργασίας είναι να παρουσιάσει κάποιες από αυτές τις μεθόδους. Οι μέθοδοι που θα παρουσιαστούν εφαρμόζονται σε γενικές περιπτώσεις και δεν είναι σχεδιασμένοι για να αντιμετωπίζουν συγκεκριμένα αντικείμενα. Η ανίχνευση αρθρωτών αντικειμένων, όπως ανθρώπων, παρουσιάζεται στα [1] και [8].

Στο κεφάλαιο 2 παρουσιάζονται τα φίλτρα Kalman που είναι μια από τις πρώτες μεθόδους που αναπτύχτηκαν για παρακολούθηση αντικειμένων και βασίζεται στην γνώση της κίνησης του αντικειμένου. Στο κεφάλαιο 3 παρουσιάζεται η μέθοδος Condensation που είναι μια πιο γενική περίπτωση των φίλτρων Kalman. Στο κεφάλαιο 4 παρουσιάζεται η αναπαράσταση αντικειμένων με χρήση ιστογραμμάτων και η εφαρμογή του αλγορίθμου της μέσης μετατόπισης για την ανίχνευση αντικειμένων. Στο κεφάλαιο 5 παρουσιάζεται ο αλγόριθμος DEMD που επίσης βασίζεται στην χρήση ιστογραμμάτων. Τέλος στο κεφάλαιο 6 προτείνεται μια εναλλακτική αναπαράσταση των αντικειμένων με χρήση μικτών κανονικών κατανομών και η εφαρμογή τους σε μια παραλλαγή του αλγορίθμου DEMD.

Κεφάλαιο 2

Φίλτρα Kalman

2.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται η ιδέα των φίλτρων Kalman και η χρήση τους στο πεδίο της υπολογιστικής όρασης [7], [14], [6], [12], [15]. Στην παράγραφο 2.2 περιγράφεται το μοντέλο παρατήρησης. Στην παράγραφο 2.3 περιγράφεται η περίπτωση των γραμμικών φίλτρων Kalman. Στην παράγραφο 2.3 περιγράφονται τα εκτεταμένα φίλτρα Kalman που χρησιμοποιούνται για μη γραμμικές περιπτώσεις. Τέλος στην παράγραφο 2.4 υπάρχουν κάποια παραδείγματα της εφαρμογής των φίλτρων Kalman στην παρακολούθηση αντικειμένων σε εικονοσειρές.

2.2 Μοντέλο Παρατήρησης

Στη διαδικασία παραγωγής των δεδομένων υποθέτουμε ότι υπάρχει ένα γραμμικό σύστημα το οποίο διαθέτει μια εσωτερική κατάσταση (την οποία δεν γνωρίζουμε με ακρίβεια) και το οποίο παράγει κάποιες παρατηρήσεις (τις οποίες γνωρίζουμε). Πιο αναλυτικά το σύστημα είναι διακριτού χρόνου και η κατάσταση τη χρονική στιγμή n περιγράφεται από το x_n (το x_n μπορεί να είναι είτε διάνυσμα είτε βαθμωτό μέγεθος, όπως αναλύεται στις επόμενες παραγράφους). Η κατάσταση του συστήματος την επόμενη χρονική στιγμή $n+1$ περιγράφεται από την εξίσωση

$$x_{n+1} = F_n x_n + w_n \tag{2.1}$$

Στην εξίσωση (2.1) το x_n και x_{n+1} είναι η κατάσταση του συστήματος τη χρονική στιγμή n και $n+1$ αντίστοιχα, το F_n είναι (στη γενική περίπτωση) ο πίνακας μετάβασης από την κατάσταση x_n στην κατάσταση x_{n+1} και το w_n είναι ο θόρυβος που επιδρά στο σύστημα τη χρονική στιγμή n. Για το θόρυβο w_n υποθέτουμε ότι είναι λευκός γκαουσιανός με μηδενική μέση τιμή και πίνακα συμμεταβλητότητας που ορίζεται στην

$$E[w_i w_j^T] = \begin{cases} Q_i & , i = j \\ 0 & , i \neq j \end{cases} \tag{2.2}$$

Κάθε χρονική στιγμή παρατηρούμε το σύστημα και προκύπτουν διάφορες μετρήσεις. Οι μετρήσεις προκύπτουν από την κατάσταση του συστήματος, αλλά δεν αναπαριστούν ακριβώς την κατάσταση του συστήματος (δηλαδή στην γενική περίπτωση είναι ένα διάνυσμα διαφορετικής διάστασης). Στη γενική περίπτωση η παρατήρηση τη χρονική στιγμή n προκύπτει από την εξίσωση

$$z_n = H_n x_n + v_n \tag{2.3}$$

Στην εξίσωση (2.3) το z_n είναι η παρατήρηση, το x_n είναι η κατάσταση του συστήματος, το H_n είναι ο πίνακας μέτρησης και το v_n είναι ο θόρυβος που επιδρά στην παρατήρηση τη χρονική στιγμή n. Για το θόρυβο v_n υποθέτουμε ότι είναι λευκός γκαουσιανός με μηδενική μέση τιμή και πίνακα συμμεταβλητότητας που ορίζεται στην

$$E[v_i v_j^T] = \begin{cases} R_i & , i = j \\ 0 & , i \neq j \end{cases} \tag{2.4}$$

Κάποια σημεία που χρειάζονται προσοχή είναι ότι στην εξίσωση (2.1) η κατάσταση x_{n+1} εξαρτάται μόνο από την κατάσταση x_n και όχι από προηγούμενες καταστάσεις. Αυτό σημαίνει ότι για να προκύψει η νέα κατάσταση δεν χρειάζονται όλες οι προηγούμενες (και επομένως δεν χρειάζεται να αποθηκεύονται) αλλά μόνο η αμέσως προηγούμενη. Επιπλέον για να ξεκινήσει το σύστημα χρειάζεται μόνο την αρχική κατάσταση (δηλαδή την κατάσταση x_{-1}). Αυτή η κατάσταση μπορεί να θεωρηθεί γνωστή ή να προσεγγιστεί από τη μέση τιμή της κατανομής που θεωρούμε ότι ακολουθεί η αρχική κατάσταση.

Στην εξίσωση (2.3) πρέπει να επισημάνουμε ότι η παρατήρηση z_n εξαρτάται μόνο από την κατάσταση x_n, δηλαδή εξαρτάται από την τρέχουσα κατάσταση μόνο και όχι από προηγούμενες. Επίσης ο θόρυβος v_n είναι ανεξάρτητος από το θόρυβο w_n τη εξίσωσης (2.1).

Το πρόβλημα που καλείται να λύσει το φίλτρο Kalman είναι η εύρεση της κατάστασης του συστήματος μέσω της παρατήρησης των μετρήσεων. Γνωρίζουμε (δηλαδή πρέπει να καθορίσουμε):

- Την φύση του συστήματος που είναι το μέγεθος και τον τύπο του διανύσματος κατάστασης x_n αλλά δεν γνωρίζουμε τις ακριβές τιμές των διαφόρων συνιστωσών του διανύσματος.

- Πως προκύπτει η κατάσταση x_{n+1} από την x_n δηλαδή γνωρίζουμε τον πίνακα μετάβασης F_n (ο πίνακας μετάβαση στην γενική περίπτωση επιτρέπεται να αλλάζει στο χρόνο, αλλά συνήθως είναι σταθερός). Για παράδειγμα αν το αντικείμενο εκτελεί ευθύγραμμη ομαλή κίνηση τότε η κατάσταση περιγράφεται από την θέση του αντικειμένου και την ταχύτητά του (δηλαδή το ποσό που μεταβάλλεται η θέση του). Σε αυτή την περίπτωση η νέα θέση προκύπτει σύμφωνα με τον τύπο

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \\ \Delta x_{n+1} \\ \Delta y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \\ \Delta x_n \\ \Delta y_n \end{bmatrix} + w_n$$

  όπου x, y είναι η συντεταγμένες,$\Delta x, \Delta y$ η ταχύτητα σε κάθε άξονα και w_n είναι ο θόρυβος.

- Τον πίνακα συμμεταβλητότητας Q_n του θορύβου w_n, που είναι ο θόρυβος που επιδρά στην κατάσταση του συστήματος (ο πίνακας συμμεταβλητότητας επιτρέπεται να αλλάζει στο χρόνο, αλλά συνήθως είναι σταθερός).

- Την αρχική κατάσταση x_{-1} του συστήματος. Αν δεν ξέρουμε την ακριβή αρχική κατάσταση τότε γνωρίζουμε την κατανομή που αυτή ακολουθεί και έτσι την προσεγγίζουμε από τη μέση τιμή της κατανομής.

- Πως προκύπτει η παρατήρηση z_n από την κατάσταση x_n δηλαδή γνωρίζουμε τον πίνακα μετρήσεων H_n (ο πίνακας μετρήσεων επιτρέπεται να μεταβάλλεται στο χρόνο, αλλά συνήθως είναι σταθερός). Για παράδειγμα όταν παρατηρούμε ένα αντικείμενο που εκτελεί ευθύγραμμη ομαλή κίνηση μπορούμε να αντιληφθούμε μόνο την θέση του, δηλαδή η παρατήρηση θα είναι της μορφής

$$\begin{bmatrix} Z_{x_n} \\ Z_{y_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \\ \Delta x_n \\ \Delta y_n \end{bmatrix} + v_n$$

- Τον πίνακα συμμεταβλητότητας R_n του θορύβου v_n, που είναι ο θόρυβος που επιδρά στην παρατήρηση της κατάστασης του συστήματος (ο πίνακας συμμεταβλητότητας επιτρέπεται να αλλάζει στο χρόνο, αλλά συνήθως είναι σταθερός).

Το μόνο που δεν γνωρίζουμε (και θέλουμε να βρούμε) είναι η κατάσταση του συστήματος $x_n, \quad 0 \leq n \leq N$. Στην περίπτωση που για να βρούμε την κατάσταση x_n χρησιμοποιούμε παρατηρήσεις

- $z_i, \quad 0 \leq i \leq n$ το πρόβλημα λέγεται φιλτράρισμα (filtering).

- $z_i, \quad 0 \leq i \leq n-1$ το πρόβλημα λέγεται πρόβλεψη (prediction).

- $z_i, \quad 0 \le i \le N$ το πρόβλημα λέγεται εξομάλυνση (smoothing).

Το πρόβλημα που θέλουμε να λύσουμε είναι η εύρεση της κατάστασης x_n που ταιριάζει (βέλτιστα) στις παραμέτρους που έχουμε καθορίσει.

Πριν προχωρήσουμε πρέπει να ορίσουμε την έννοια της βέλτιστης κατάστασης. Η ανάλυση αναφέρεται σε βαθμωτά μεγέθη, αλλά η ίδια ανάλυση γίνεται και για διανύσματα. Έστω ότι έχουμε τις παρατηρήσεις $z_i, \quad 0 \le i \le n$ και έστω $\hat{x}_n$ είναι η εκ των υστέρων εκτίμηση της κατάστασης x_n. Στην γενική περίπτωση η εκτίμηση $\hat{x}_n$ είναι διαφορετική από την πραγματική κατάσταση x_n. Ορίζουμε τη συνάρτηση μέσου τετραγωνικού σφάλματος (mean square error):

$$J_n = E[(x_n - \hat{x}_n)^2] = E[(\tilde{x}_n)^2] \tag{2.5}$$

Όπου $\tilde{x}_n$ είναι το σφάλμα εκτίμησης. Η συνάρτηση (2.5) είναι μη αρνητική και μη φθίνουσα. Η εξάρτηση της συνάρτησης κόστους J_n από το n δίνει έμφαση στην μη σταθερή φύση της αναδρομικής διαδικασίας. Για να βρούμε την βέλτιστη εκτίμηση $\hat{x}_n$ θα χρειαστούμε τα παρακάτω δύο θεωρήματα [6].

Θεώρημα 2.1 (Υπό συνθήκη εκτιμητής μέσης τιμής)**.** *Αν οι στοχαστικές διαδικασίες $\{x_n\}$ και $\{z_n\}$ είναι από κοινού κανονικές, τότε ο βέλτιστος εκτιμητής $\hat{x}_n$ που ελαχιστοποιεί το σφάλμα J_n είναι ο υπό συνθήκη εκτιμητής μέσης τιμής*

$$\hat{x}_n = E[x_n | z_0, \ldots, z_n] \tag{2.6}$$

Θεώρημα 2.2 (Αρχή της ορθογωνικότητας)**.** *Έστω οι στοχαστικές διαδικασίες $\{x_n\}$ και $\{z_n\}$ έχουν μέση τιμή μηδέν, δηλαδή*

$$E[x_n] = E[z_n] = 0, \qquad \forall n \tag{2.7}$$

τότε αν ο βέλτιστος εκτιμητής $\hat{x}_n$ είναι γραμμική συνάρτηση των παρατηρήσεων και η συνάρτηση κόστους είναι η συνάρτηση μέσου τετραγωνικού σφάλματος, τότε η βέλτιστη εκτίμηση $\hat{x}_n$ δοσμένων των παρατηρήσεων $z_0, \ldots, z_n$ είναι η ορθογώνια προβολή του x_n στον χώρο των παρατηρήσεων.

2.3 Γραμμικό Kalman Φίλτρο

Το φίλτρο Kalman είναι ένας αναδρομικός τύπος για την εύρεση της βέλτιστης λύσης σε γραμμικά προβλήματα φιλτραρίσματος. Η λύση προκύπτει από την προηγούμενη λύση και τα νέα δεδομένα που εισέρχονται. Ένα πλεονέκτημα είναι ότι δεν είναι αναγκαίο να αποθηκεύονται όλα τα δεδομένα, καθώς επίσης ότι για να κάνουμε την επεξεργασία δεν είναι αναγκαίο να έχουμε όλα τα δεδομένα, αλλά μόνο αυτά του παρελθόντος.

Υποθέτουμε ότι το μοντέλο είναι αυτό που περιγράφεται στην παράγραφο 2.2. Ο στόχος είναι να χρησιμοποιηθεί η πληροφορία από τις παρατηρήσεις z_n για να βελτιώσουμε την εκτίμηση της άγνωστης κατάστασης z_n. Έστω $\hat{x}_n^-$ η εκ των προτέρων εκτίμηση της κατάστασης του συστήματος (χωρίς την παρατήρηση z_n) την χρονική στιγμή n. Μπορούμε να εκφράσουμε την εκ των υστέρων εκτίμηση της κατάστασης $\hat{x}_n$ (με την χρήση της παρατήρησης z_n ως

$$\hat{x}_n = G_n^{(1)}\hat{x}_n^- + G_n z_n \tag{2.8}$$

όπου οι πίνακες $G_n^{(1)}$ και G_n πρέπει να βρεθούν. Το λάθος κατάστασης (η απόκλιση από την κανονική κατάσταση) δίνεται ως

$$\tilde{x}_n = x_n - \hat{x}_n \tag{2.9}$$

Εφαρμόζοντας την αρχή της ορθογωνικότητας μπορούμε να γράψουμε

$$E[\tilde{x}_n z_i^T] = 0, \qquad \forall i = 0, \dots, n-1 \tag{2.10}$$

Χρησιμοποιώντας τις εξισώσεις (2.3), (2.8), (2.9), (2.10) παίρνουμε

$$E[(x_n - G_n^{(1)}\hat{x}_n^- - G_n H_n x_n - G_n v_n) z_i^T] = 0, \qquad \forall i = 0, \dots, n-1 \tag{2.11}$$

Επειδή ο θόρυβος w_n και ο v_n είναι ανεξάρτητοι προκύπτει ότι

$$E[v_n z_n^T] = 0 \tag{2.12}$$

Χρησιμοποιώντας την εξίσωση (2.12) και προσθέτοντας το $G_n^{(1)}x_n - G_n^{(1)}x_n$ μπορούμε να γράψουμε την εξίσωση (2.11) ως

$$E[(I - G_n H_n - G_n^{(1)}) x_n z_i^T + G_n^{(1)}(x_n - \hat{x}_n^-) z_i^T] = 0, \qquad \forall i = 0, \dots, n-1 \tag{2.13}$$

Όπου I είναι ο μοναδιαίος πίνακας. Από την αρχή της ορθογωνικότητας ισχύσει ότι $E[(x_n - \hat{x}_n^-) z_i^T] = 0$ η εξίσωση (2.13) απλοποιείται σε

$$(I - G_n H_n - G_n^{(1)}) E[x_n z_i^T] = 0, \qquad \forall i = 0, \dots, n-1 \tag{2.14}$$

Για τυχαίες τιμές της κατάστασης x_n και z_i η εξίσωση (2.14) μπορεί να ικανοποιηθεί μόνο αν για τους συντελεστές $G_n^{(1)}$ και G_n ισχύει ότι

$$I - G_n H_n - G_n^{(1)} = 0 \quad \Leftrightarrow \quad G_n^{(1)} = I - G_n H_n \tag{2.15}$$

Αντικαθιστώντας την εξίσωση (2.15) στην (2.8) μπορούμε να εκφράσουμε την εκ των υστέρων εκτίμηση της κατάστασης ως

$$\hat{x}_n = \hat{x}_n^- + G_n(z_n - H_n \hat{x}_n^-) \tag{2.16}$$

όπου ο πίνακας G_n καλείται κέρδος Kalman.

Τώρα παραμένει το πρόβλημα της εύρεσης ενός τύπου για το G_n. Από την αρχή της ορθογωνικότητας έχουμε

$$E[(x_n - \hat{x}_n) z_i^T] \quad \Leftrightarrow \quad E[(x_n - \hat{x}_n)\hat{z}_i^T] \tag{2.17}$$

όπου $\hat{z}_n^T$ είναι η εκτίμηση της z_n από τις προηγούμενες μετρήσεις $z_0, \ldots, z_{n-1}$. Ορίζουμε το $\tilde{z}_n = z_n - \hat{z}_n$, το οποίο αναπαριστά ένα μέτρο της νέας πληροφορίας που περιέχεται στο z_n, που μπορεί να αναπαρασταθεί ως

$$\tilde{z}_n = z_n - H_n \hat{x}_n^- = H_n x_n + v_n - H_n \hat{x}_n^- = v_n + H_n \tilde{x}_n^- \tag{2.18}$$

Αφαιρώντας τα δύο μέλη της ισοδυναμίας (2.17) προκύπτει

$$E[(x_n - \hat{x}_n)\tilde{z}_n^T] = 0 \tag{2.19}$$

Χρησιμοποιώντας τις εξισώσεις (2.3) και (2.16) μπορούμε να εκφράσουμε το σφάλμα $x_n - \hat{x}_n$ ως

$$x_n - \hat{x}_n = \tilde{x}_n^- - G_n(H_n \tilde{x}_n^- + v_n) = (I - G_n H_n)\tilde{x}_n^- - G_n v_n \tag{2.20}$$

Αντικαθιστώντας τις εξισώσεις (2.18) και (2.20) στην (2.19) προκύπτει

$$E[\{(I - G_n H_n)\tilde{x}_n^- - G_n v_n\}(H_n \tilde{x}_n^- + v_n)] = 0 \tag{2.21}$$

Επειδή ο θόρυβος v_n είναι ανεξάρτητος της κατάστασης x_n και έτσι το σφάλμα $\tilde{x}_n^-$ η εκτίμηση της εξίσωσης (2.21) γίνεται

$$(I - G_n H_n)E[\tilde{x}_n^- \tilde{x}_n^{-T}]H_n^T - G_n E[v_n v_n^T] = 0 \tag{2.22}$$

Ορίζουμε την εκ των προτέρων πίνακα συμμεταβλητότητας

$$P_n^- = E[(x_n - \hat{x}_n^-)(x_n - \hat{x}_n^-)^T] = E[\tilde{x}_n^- \tilde{x}_n^{-T}] \tag{2.23}$$

Έτσι συμπεριλαμβάνοντας τις εξισώσεις (2.4) και (2.23) μπορούμε να γράψουμε την εξίσωση (2.22) ως

$$(I - G_n H_n)P_n^- H_n^T - G_n R_n = 0 \tag{2.24}$$

Και λύνοντας ως προς G_n προκύπτει

$$G_n = P_n^- H_n^T [H_n P_n^- H_n^T + R_n]^{-1} \tag{2.25}$$

Η εξίσωση (2.25) δίνει ένα τρόπο υπολογισμού του G_n, που δίνεται ως συνάρτηση του εκ των προτέρων πίνακα συμμεταβλητότητας P_n^-. Για να τελειώσουμε πρέπει να υπολογίσουμε την διάδοση σφάλματος του πίνακα συμμεταβλητότητας, που περιγράφει την επίδραση των σφαλμάτων εκτίμησης στον πίνακα συμμεταβλητότητας. Αυτή η διάδοση εμπεριέχει δύο βήματα:

1. Ο εκ των προτέρων πίνακας συμμεταβλητότητας P_n^- τη χρονική στιγμή n δίνεται από την εξίσωση (2.23). Δοσμένου του πίνακα P_n^- υπολογίζουμε των εκ των υστέρων πίνακα συμμεταβλητότητας P_n που δίνεται από τον τύπο

$$P_n = E[(x_n - \hat{x}_n)(x_n - \hat{x}_n)^T] = E[\tilde{x}_n \tilde{x}_n^T] \tag{2.26}$$

2. Δοσμένου του προηγούμενου εκ των υστέρων πίνακα συμμεταβλητότητας P_{n-1} υπολογίζουμε τον ενημερωμένο εκ των προτέρων πίνακα συμμεταβλητότητας P_n^-.

Για το πρώτο βήμα αντικαθιστούμε την εξίσωση (2.20) στην (2.26) και επειδή ο θόρυβος v_n είναι ανεξάρτητος της σκ των προτέρων σφάλματος εκτίμησης $\tilde{x}_n^-$ προκύπτει

$$\begin{aligned} P_n &= (I - G_n H_n) E[\tilde{x}_n^- \tilde{x}_n^{-T}](I - G_n H_n)^T + G_n E[v_n v_n^T] G_n^T \\ &= (I - G_n H_n) P_n^- (I - G_n H_n)^T + G_n R_n G_n^T \end{aligned} \tag{2.27}$$

Αναλύοντας την εξίσωση (2.27) και χρησιμοποιώντας την (2.25) προκύπτει

$$\begin{aligned} P_n &= (I - G_n H_n) P_n^- - (I - G_n H_n) P_n^- H_n^T G_n^T + G_n R_n G_n^T \\ &= (I - G_n H_n) P_n^- - G_n R_n G_n^T + G_n R_n G_n^T \\ &= (I - G_n H_n) P_n^- \end{aligned} \tag{2.28}$$

Για το δεύτερο βήμα της διάδοσης του σφάλματος πρώτα πρέπει να παρατηρήσουμε ότι η εκ των προτέρων εκτίμηση της κατάστασης εκφράζεται σε σχέση με την προηγούμενη εκτίμηση ως

$$\tilde{x}_n^- = F_n \hat{x}_{n-1} \tag{2.29}$$

Μπορούμε να χρησιμοποιήσουμε την εξίσωση (2.1) και (2.29) για να εκφράσουμε την εκ των προτέρων εκτίμηση του σφάλματος ως

$$\begin{aligned} \tilde{x}_n^- &= x_n - \hat{x}_n^- \\ &= (F_n x_{n-1} + w_{n-1}) - (F_n \hat{x}_{n-1}) \\ &= F_n (x_{n-1} - \hat{x}_{n-1}) + w_{n-1} \\ &= F_n \tilde{x}_{n-1} + w_{n-1} \end{aligned} \tag{2.30}$$

Αντικαθιστώντας την εξίσωση (2.30) στην (2.23) και επειδή ο θόρυβος w_n είναι ανεξάρτητος του $\hat{x}_{n-1}$ προκύπτει

$$\begin{aligned} P_n^- &= F_n E[\tilde{x}_{n-1} \tilde{x}_{n-1}^T] F_n^T + E[w_{n-1} w_{n-1}^T] \\ &= F_n P_{n-1} F_n^T + Q_{n-1} \end{aligned} \tag{2.31}$$

όπου ορίζουμε των εκ των προτέρων πίνακα συμμεταβλητότητας P_n^- από τον προηγούμενο εκ των υστέρων πίνακα συμμεταβλητότητας P_{n-1}.

Με τις εξισώσεις (2.29), (2.31), (2.25), (2.16) και (2.28) μπορούμε να παρουσιάζουμε τον αναδρομικό αλγόριθμο ενημέρωσης της εκτίμησης.

Στην αρχικοποίηση επιλέγουμε την αρχική κατάσταση ως $\hat{x}_0 = E[x_0]$ και τον αρχικό πίνακα συμμεταβλητότητας ως $P_0 = E[(x_0 - E[x_0])(x_0 - E[x_0)^T]$, γιατί δεν έχουμε άλλη πληροφορία για την κατανομή.

Τα φίλτρα Kalman χρησιμοποιούν γκαουσιανές κατανομές πιθανότητας στην διαδικασία διάδοσης, η διάχυση είναι γραμμική και η συνάρτηση πυκνότητας πιθανότητας εξελίσσεται στο χρόνο σαν γκαουσιανός παλμός που μετατοπίζεται, ανοίγει και δυναμώνει αλλά παραμένει γκαουσιανός.

Η τυχαία συνιστώσα w_n του μοντέλου οδηγεί σε διάχυση (δηλαδή αύξηση της αβεβαιότητας), ενώ η ντετερμινιστική συνιστώσα $F_n x_n$ προκαλεί την συνάρτηση να μετατοπιστεί. Το αποτέλεσμα της παρατήρησης z_n είναι να υπερθέσει μία αντίδραση στην διάχυση και τελικά η πυκνότητα πιθανότητας εμφανίζει μέγιστο στην γειτονιά της παρατήρησης.

Αλγόριθμος 1 Φίλτρα Kalman

1 Αρχικοποίηση:

$$\hat{x}_0 = E[x_0]$$

$$P_0 = E[(x_0 - E[x_0])(x_0 - E[x_0])^T]$$

2 Πρόβλεψη:

$$\hat{x}_n^- = F_n \hat{x}_{n-1}$$

$$P_n^- = F_n P_{n-1} F_n^T + Q_n$$

$$G_n = P_n^- H_n^T [H_n P_n^- H_n^T + R_n]^{-1}$$

3 Εκτίμηση:

$$\hat{x}_n = \hat{x}_n^- + G_n(z_n - H_n \hat{x}_n^-)$$

$$P_n = (I - G_n H_n) P_n^-$$

Επιστρέφουμε στο βήμα της εκτίμησης για την επόμενη χρονική στιγμή

2.4 Εκτεταμένο Kalman Φίλτρο

Όταν το σύστημα δεν είναι γραμμικό, το γραμμικό φίλτρο Kalman δεν αποτελεί τη βέλτιστη λύση. Μπορούμε να χρησιμοποιήσουμε το φίλτρο Kalman, αν προσεγγίσουμε το σύστημα με γραμμικό. Το αποτέλεσμα αναφέρεται ως εκτεταμένο Kalman φίλτρο (extended Kalman filter). Μια τέτοια επέκταση είναι εφικτή επειδή τα φίλτρα Kalman εκφράζονται σαν διαφορικές εξισώσεις διακριτού χρόνου.

Έστω ότι το μη γραμμικό σύστημα περιγράφεται από τις εξισώσεις

$$x_{n+1} = f(n, x_n) + w_n \tag{2.32}$$

$$z_n = h(n, x_n) + v_n \tag{2.33}$$

όπου τα w_n και v_n είναι ανεξάρτητοι γκαουσιανοί θόρυβοι με μηδενική μέση τιμή και πίνακες συμμεταβλητότητας R_n και Q_n αντίστοιχα. Εδώ ωστόσο οι συναρτήσεις $f(n, x_n)$ και $h(n, x_n)$ υποδηλώνουν ένα μη γραμμικό πίνακα μετασχηματισμού που εξαρτάται πιθανώς από το χρόνο. Η βασική ιδέα του εκτεταμένου φίλτρου Kalman είναι να κάνουμε γραμμικό το μοντέλο που περιγράφεται από τις εξισώσεις (2.32) και (2.33) κάθε χρονική στιγμή γύρω από την πιο πρόσφατη εκτίμηση της κατάστασης, που μπορεί να είναι είτε η $\hat{x}_n$ είτε η $\hat{x}_n^-$. Αφού γίνει η αρχικοποίηση το γραμμικό φίλτρο Kalman (παράγραφος 2.3) μπορεί να εφαρμοστεί.

Πιο αναλυτικά η προσέγγιση μπορεί να γίνει σε δύο στάδια.

1. Δημιουργούνται οι πίνακες

$$F_n = \left. \frac{\partial f(n, x)}{\partial x} \right|_{x = \hat{x}_n} \tag{2.34}$$

$$H_n = \left.\frac{\partial h(n,x)}{\partial x}\right|_{x=\hat{x}_n^-} \tag{2.35}$$

Το ij στοιχείο του πίνακα F_n είναι ίσο με την μερική παράγωγο της $i-οστής$ συνιστώσας της $f(n,x)$ ως προς την $j-οστή$ συνιστώσα του x. Όμοια το ij στοιχείο του πίνακα H_n είναι ίσο με την μερική παράγωγο της $i-οστής$ συνιστώσας της $h(n,x)$ ως προς την $j-οστή$ συνιστώσα του x. Οι παράγωγοι μπορούν να υπολογιστούν είτε στο σημείο $\hat{x}_n$ είτε στο $\hat{x}_n^-$, ανάλογα ποια είναι διαθέσιμη. Οι πίνακες F_n και H_n είναι με αυτό τον τρόπο γνωστοί τη χρονική στιγμή n.

2. Αφού έχουν υπολογιστεί οι πίνακες F_n και H_n, αναπτύσσουμε τις $f(n,x)$ και $h(n,x)$ σαν σειρά Taylor γύρω από το σημείο $\hat{x}_n$ ή $\hat{x}_n^-$ αντίστοιχα, οπότε προκύπτει

$$f(n,x_n) \approx f(n,\hat{x}_n) + F_n \cdot (x - \hat{x}_n) \tag{2.36}$$

$$h(n,x_n) \approx f(n,\hat{x}_n^-) + F_n \cdot (x - \hat{x}_n^-) \tag{2.37}$$

Έχοντας τα παραπάνω αποτελέσματα των βημάτων 1 και 2 μπορούμε να προσεγγίσουμε γραμμικά τις εξισώσεις (2.32) και (2.33) ως

$$x_{n+1} = F_n x_n + w_n + d_n \tag{2.38}$$

$$z_n = H_n x_n + v_n + \bar{z}_n \tag{2.39}$$

όπου

$$d_n = f(n,\hat{x}_n) - F_n \hat{x}_n \tag{2.40}$$

και

$$\bar{z}_n = h(n,\hat{x}_n^-) - H_n \hat{x}_n^- \tag{2.41}$$

Όλα τα στοιχεία στην (2.41) είναι γνωστά τη χρονική στιγμή n και το $\hat{z}_n^-$ μπορεί να εκφραστεί ως η παρατήρηση τη χρονική στιγμή n. Επίσης όλα τα στοιχεία στην εξίσωση (2.39) είναι γνωστά τη χρονική στιγμή n.

Με τις εξισώσεις (2.38), (2.39), (2.40), (2.41) μπορούμε να παρουσιάζουμε τον αναδρομικό αλγόριθμο ενημέρωσης της εκτίμησης για τη μη γραμμική περίπτωση.

Αλγόριθμος 2 Εκτεταμένα Φίλτρα Kalman

1 Αρχικοποίηση:

$$\hat{x}_0 = E[x_0]$$

$$P_0 = E[(x_0 - E[x_0])(x_0 - E[x_0])^T]$$

2 Πρόβλεψη:

$$\hat{x}_n^- = f(n, \hat{x}_{n-1})$$

$$F_{n-1} = \left.\frac{\partial f(n-1, x)}{\partial x}\right|_{x=\hat{x}_{n-1}}$$

$$H_n = \left.\frac{\partial h(n, x)}{\partial x}\right|_{x=\hat{x}_n^-}$$

$$P_n^- = F_{n-1} P_{n-1} F_{n-1}^T + Q_n$$

$$G_n = P_n^- H_n^T [H_n P_n^- H_n^T + R_n]^{-1}$$

3 Εκτίμηση:

$$\hat{x}_n = \hat{x}_n^- + G_n z_n - h(n, \hat{x}_n^-)$$

$$P_n = (I - G_n H_n) P_n^-$$

Επιστρέφουμε στο βήμα της εκτίμησης για την επόμενη χρονική στιγμή

Κεφαλαιο 3

Διαδοση υπο συνθηκης πυκνοτητας πιθανοτητας

3.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται ο αλγόριθμος Condensation και η επέκτασή του ICondensation. Οι αλγόριθμοι ανήκουν στην κατηγορία σμήνη σωματιδίων [2], [15]. Το όνομα Condensation [9] προέρχεται από την αγγλική φράση conditional density propagation. Στο 3.2 περιγράφεται ο αλγόριθμος Condensation και στο 3.3 περιγράφεται ο ICondensation [10].

3.2 Condensation

Τα φίλτρα Kalman παρέχουν μια γραμμική εκτίμηση αναδρομικά και εφαρμόζονται μόνο όταν υπάρχουν γκαουσιανές κατανομές. Επίσης αν σε κάποια στιγμή η κατάσταση αποκλίνει πολύ από την πραγματική κατάσταση (δηλαδή αν χαθεί το αντικείμενο που παρακολουθείται) τότε είναι πολύ δύσκολο να επανέρθει στην σωστή κατάσταση. Αυτά τα προβλήματα προσπαθεί να αντιμετωπίσει ο αλγόριθμος Condensation [9].

3.2.1 Διάδοση Πυκνότητας Πιθανότητας της Κατάστασης Διακριτού Χρόνου

Όπως στην περίπτωση των φίλτρων Kalman στην παράγραφο 2.2 υποθέτουμε ότι οι παρατηρήσεις γίνονται σε διακριτές χρονικές στιγμές n, υπάρχει η πραγματική κατάσταση x_n και η αντίστοιχη παρατήρηση z_n. Δεν υπάρχει καμία υπόθεση για τις κατανομές των x_n και z_n. Η πραγματική κατάσταση εξαρτάται μόνο από την αμέσως προηγούμενη και όχι από όλες τις προηγούμενες, δηλαδή

$$p(x_{n+1}|x_1,\ldots,x_n) = p(x_{n+1}|x_n) \tag{3.1}$$

Επίσης οι παρατηρήσεις z_n είναι ανεξάρτητες μεταξύ τους και εξαρτώνται μόνο από την τρέχουσα κατάσταση, δηλαδή

$$p(z_1,\ldots,z_n,x_{n+1}|x_1,\ldots,x_n) = p(x_{n+1}|x1,\ldots,x_n)\prod_{i=1}^{n} p(z_i|x_i) \tag{3.2}$$

Και ολοκληρώνοντας ως προς x_{n+1} προκύπτει

$$p(z_1,\ldots,z_{n+1}|x_1,\ldots,x_{n+1}) = \prod_{i=1}^{n+1} p(z_i|x_i) \tag{3.3}$$

Η διαδικασία παρατήρησης προσδιορίζεται μόνο από το $p(z_n|x_n)$ σε κάθε χρονική στιγμή n. Η πιθανότητα μιας κατάστασης x_n δίνεται από την εξίσωση

$$p(x_n|z_1,\ldots,z_n) = k_n p(z_n|x_n) p(x_n|z_1,\ldots,z_{n-1}) \tag{3.4}$$

Όπου

$$p(x_n|z_1,\ldots,z_{n-1}) = \int_{x_{n-1}} p(x_n|x_{n-1}) p(x_{n-1}|z_1,\ldots,z_{n-1}) \tag{3.5}$$

Το k_n στην εξίσωση (3.4) είναι μια σταθερά κανονικοποίησης, που είναι ανεξάρτητη από το x_n. Η εξίσωση (3.4) είναι ο κανόνας του Bayes για την εκ των υστέρων πιθανότητα. Η εκ των προτέρων πιθανότητα που δίνεται από το $p(x_n|z_1,\ldots,z_{n-1})$ της εξίσωσης (3.5) είναι μια πρόβλεψη της κατάστασης x_n που υπολογίζεται από την εκ των υστέρων $p(x_{n-1}|z_1,\ldots,z_{n-1})$ της προηγούμενης κατάστασης x_{n-1} και την πιθανότητα μετάβασης ενός βήματος από την κατάσταση x_{n-1} στην x_n. Στην (3.4) πολλαπλασιάζουμε με το $p(z_n|x_n)$ για να λάβουμε υπόψη την παρατήρηση z_n. Οι παραπάνω πιθανότητες στην γενική περίπτωση είναι μη γκαουσιανές και πρέπει να εφαρμοστεί ένα μη γραμμικό φίλτρο στο χρόνο. Για να είναι υπολογιστικά εφικτό αυτό καταφεύγουμε σε προσεγγίσεις.

3.2.2 Σταθμευμένη Δειγματοληψία με Βάρη (Factored Sampling)

Αρχικά θα μελετήσουμε τη δειγματοληψία με βάρη (factored sampling). Σε αυτή την περίπτωση το ζητούμενο είναι να βρούμε κάποιο αντικείμενο που περιγράφεται από το διάνυσμα x με εκ των προτέρων πιθανότητα $p(x)$ και έχοντας παρατηρήσεις που περιγράφονται από το z και λαμβάνονται από μια μόνο εικόνα. Η εκ των υστέρων πιθανότητα $p(x|z)$ δίνει την εκτίμηση της κατάστασης x αν λάβουμε υπόψη την παρατήρηση. Από τον κανόνα του Bayes προκύπτει

$$p(x|z) = kp(z|x)p(x) \tag{3.6}$$

Όπου το k είναι όρος κανονικοποίησης ανεξάρτητος από το x.

Ο αλγόριθμος δειγματοληψίας με βάρη δημιουργεί μια τυχαία μεταβλητή x από την κατανομή $\tilde{p}(x)$ που προσεγγίζει την $p(x|z)$. Πρώτα δημιουργείται ένα σύνολο M δειγμάτων $s^1, \ldots, s^M$ από την εκ των προτέρων πιθανότητα $p(x)$. Στη συνέχεια ένας δείκτης $i \in \{1, \ldots, M\}$ επιλέγεται με πιθανότητα π_i, όπου

$$\pi^i = \frac{p(z|s^i)}{\sum_{j=1}^{M} p(z|s^j)} \tag{3.7}$$

Το $x' = x_i$ που επιλέγεται με αυτό τον τρόπο έχει κατανομή που προσεγγίζει την $p(x|z)$ καθώς το N αυξάνεται. Η εκ των υστέρων $E[g(x)|z]$ μπορεί να παραχθεί απευθείας από τα δείγματα $s^1, \ldots, s^M$ χρησιμοποιώντας το βάρη $p(z|x)$ και προκύπτει

$$E[g(x)|z] \approx \frac{\sum_{i=1}^{M} g(s^i)p(z|s^i)}{\sum_{i=1}^{M} p(z|s^i)} \tag{3.8}$$

Για παράδειγμα η μέση τιμή μπορεί να προσεγγιστεί αν χρησιμοποιήσουμε την συνάρτηση $g(x) = x$ και η διασπορά με την συνάρτηση $g(x) = xx^T$.

3.2.3 Ο Αλγόριθμος Condensation

Ο αλγόριθμος Condensation [9] βασίζεται στην δειγματοληψία με βάρη, αλλά την επεκτείνει ώστε να εφαρμόζεται αναδρομικά σε διαδοχικές εικόνες. Αυτό που συμβαίνει είναι ότι σε κάθε χρονική στιγμή n χρησιμοποιείται ο αλγόριθμος της δειγματοληψίας με βάρη για κάθε εικόνα και το αποτέλεσμα είναι το ζυγισμένο άθροισμα των $\{s_n^i, i = 1, \ldots, M\}$ με βάρη π_n^i που προσεγγίζει την κατανομή $p(x_n|z_1, \ldots, z_n)$. Για την δημιουργία των συνόλου $\{s_n^i, i = 1, \ldots, M\}$ πρέπει να ξέρουμε την εκ των προτέρων πιθανότητα $p(x_n|z_1, \ldots, z_{n-1})$. Αυτή η συνάρτηση πιθανότητας τις περισσότερες φορές δεν είναι σε κλειστή μαθηματική μορφή. Προσεγγίζουμε από το σύνολο $\{(s_{n-1}^i, \pi_{n-1}^i), i = 1, \ldots, M\}$ την πιθανότητα $p(x_{n-1}|z_1, \ldots, z_{n-1})$, που είναι το αποτέλεσμα του αλγορίθμου Condensation από την προηγούμενη χρονική στιγμή $n-1$. Με αυτά μπορεί να υπολογιστεί η πρόβλεψη $p(x_n|z_1, \ldots, z_{n-1})$ από την εξίσωση (3.5).

Ο σκοπός είναι να διατηρήσουμε σε διαδοχικές χρονικές στιγμές σύνολα $\{(s_n^i, \pi_n^i), i = 1, \ldots, M\}$ τα οποία προσεγγίζουν όσο καλύτερα γίνεται την $p(x_n|z_1, \ldots, z_n)$ και έχουν μικρό υπολογιστικό κόστος. Το πρώτο πράγμα που γίνεται είναι η δειγματοληψία (με αντικατάσταση) M φορές από το σύνολο $\{(s_{n-1}^i, \pi_{n-1}^i), i = 1, \ldots, M\}$, επιλέγοντας ένα συγκεκριμένο στοιχείο i με πιθανότητα p_{n-1}^i. Μερικά στοιχεία, ιδίως αυτά με μεγάλα βάρη, μπορούν να επιλεγούν αρκετές φορές, δίνοντας πολλαπλά αντίγραφα του ίδιου στοιχείου στο νέο σύνολο. Από την άλλη τα στοιχεία με μικρό βάρος μπορούν να μην επιλεγούν καθόλου.

Κάθε στοιχείο που επιλέχτηκε και μπήκε στο νέο σύνολο τώρα μπαίνει στο βήμα πρόβλεψης. Πρώτα το κάθε στοιχείο του συνόλου υφίσταται μία μικρή μεταβολή η οποία είναι ντετερμινιστική για όλα τα στοιχεία. Με αυτό τον τρόπο ίδια στοιχεία υφίστανται την ίδια μεταβολή. Στη συνέχεια του βήματος πρόβλεψης γίνεται μετατόπιση των στοιχείων με τυχαίο τρόπο, και με αυτό τον τρόπο ίδια στοιχεία καταλήγουν σε διαφορετικά. Σε αυτό το σημείο το σύνολο $\{s_n^i, i = 1, \ldots, M\}$ έχει δημιουργηθεί, αλλά δεν υπάρχουν ακόμη τα βάρη $\{\pi_n^i, i = 1, \ldots, M\}$. Αυτό το σύνολο είναι μια καλή προσέγγιση της εκ των προτέρων πιθανότητας $p(x_n|z_1, \ldots, z_{n-1})$ για τη χρονική στιγμή n. Τελικά με το βήμα της παρατήρησης από την δειγματοληψία με βάρη δημιουργούνται τα βάρη από την συνάρτηση πιθανότητας $p(z_n|x_n)$ και προκύπτει το σύνολο $\{(s_n^i, \pi_n^i), i = 1, \ldots, M\}$ για την χρονική στιγμή n. Ακολουθεί η περιγραφή του αλγορίθμου.

Αλγόριθμος 3 Condensation

Από το προηγούμενο σύνολο $\{(s_{n-1}^i, \pi_{n-1}^i, c_{n-1}^i), i = 1, \ldots, M\}$ του βήματος $n-1$ δημιουργείται το νέο σύνολο $\{(s_n^i, \pi_n^i, c_n^i), i = 1, \ldots, M\}$ του βήματος n.
Δημιουργούμε το $i - οστό$ δείγμα με τον παρακάτω τρόπο:

1 Επιλογή: επιλέγουμε ένα δείγμα s'^i_n δημιουργώντας έναν αριθμό $r \in [0, 1]$ και επιλέγοντας το s_{n-1}^j για το οποίο ισχύει ότι το j είναι το μικρότερο που ικανοποιεί την $c_{n-1}^j \geq r$.

2 Πρόβλεψη: δειγματοληπτούμε από την $p(x_n|x_{n-1} = s'^i_n)$ και προκύπτει το s_n^i.

3 Μέτρηση: παίρνουμε την επόμενη παρατήρηση και βρίσκουμε το βάρος του δείγματος $\pi_n^i = p(z_n|x_n = s_n^i)$. Στη συνέχεια κοινωνικοποιούμε έτσι ώστε $\sum_i \pi_n^i = 1$ και υπολογίσουμε το c_n^i όπου
$c_n^0 = 0$ ή $c_n^i = c_{n-1}^i + \pi_n^i, (i = 1, \ldots, M)$.

Αφού παραχθούν τα M δείγματα η εκτίμηση της κατάστασης της χρονική στιγμή n μπορεί να παραχωθεί από $E[f(x_n)] = \sum_{i=1}^{M} \pi_n^i f(s_n^i)$ (για τη μέση τιμή μπορεί να χρησιμοποιηθεί η συνάρτηση $f(x) = x$).

Στον αλγόριθμο χρησιμοποιούνται τα αθροιστικά βάρη c_{n-1}^i (τα οποία δημιουργούνται στο βήμα 3) για να γίνει αποτελεσματική δειγματοληψία στο βήμα 1. Επίσης στο βήμα της πρόβλεψης η πιθανότητα μετάβασης ενός βήματος μπορεί να δίνεται από κάποιο τύπο της μορφής $s_n^i = As'^i_n + Bw_n^i$, όπου w_n^i είναι διάνυσμα με γκαουσιανή κατανομή.

Ένα πλεονέκτημα του αλγορίθμου condensation είναι η απλότητά του σε σχέση με τα φίλτρα Kalman, παρά την γενικότητά του. Για την χρήση του αλγορίθμου πρέπει να:

- καθοριστεί η αρχική κατάσταση $\{(s_0^i, \pi_0^i), i = 1, \ldots, M\}$.

- να επιλεγεί κατάλληλα η πιθανότητα μετάβασης ενός βήματος $p(x_n = s_n^i|x_{n-1}^j)$. Στην πραγματικότητα δεν είναι απαραίτητο να υπολογίσουμε την τιμή της $p(x_n = s_n^i|x_{n-1}^j)$, αλλά αρκεί να δειγματοληπτούμε από αυτή. Για παράδειγμα αν η σχέση που συνδέει τη νέα θέση του αντικειμένου με την παλιά είναι $x_{n+1} - \bar{x} = A(x_n - \hat{x}) + Bw_n$, όπου $\bar{x}$ είναι η μέση τιμή της κατάστασης, οι A, B πίνακες και ο w_n θόρυβος, τότε η δειγματοληψία μπορεί να γίνει ως

$$x_{n+1} = \bar{x} + A(x_n - \hat{x}) + Bw_n$$

- να επιλεγεί κατάλληλα η συνάρτηση πυκνότητας πιθανότητας $p(z_n|x_n) = x_n^i$. Για παράδειγμα αν υποθέσουμε ότι η μέτρηση είναι κανονικά κατανεμημένη γύρω από την θέση x_n, τότε μπορεί να χρησιμοποιηθεί ο τύπος

$$p(z_n|x_n) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(z_n - x_n)^2}{2\sigma^2}$$

όπου σ είναι η διασπορά της κατανομής στην περίπτωση της μιας διάστασης.

3.3 ICondensation

Στον αλγόριθμο Condensation [9] οι θέσεις των δειγμάτων s_n^i είναι καθορισμένες από τα δείγματα του προηγούμενου βήματος $\{(s_{n-1}^i, \pi_{n-1}^i)\}$ και τη μετάβαση ενός βήματος $p(x_n = s_n^i|x_{n-1}^j)$. Τα τμήματα της εικόνας που θα δειγματοληπτηθούν στο βήμα της μέτρησης καθορίζονται πριν γίνει οποιαδήποτε μέτρηση. Αυτό είναι αποδεκτό όταν τα δείγματα s_n^i προσεγγίζουν καλά την συνάρτηση πυκνότητας πιθανότητας. Όμως στην γενική περίπτωση η συνάρτηση αλλάζει με τον χρόνο και η τυχαία κίνηση του αντικειμένου δίνει μη μηδενική πιθανότητα σε αρκετές περιοχές τριγύρω του. Αυτό στην καλύτερη περίπτωση θα δώσει δείγματα με αρκετή πιθανότητα που θα βρίσκονται κοντά στο αντικείμενο και με αυτό τον τρόπο το θα μπορούσαμε να το παρακολουθήσουμε. Αλλά το πεπερασμένο πλήθος δειγμάτων θα δώσει δείγματα μόνο στις πιο πιθανές περιοχές. Αυτό θα δώσει πολλά δείγματα σε μία περιοχή (τα οποία θα έχουν μεγάλη εκ των προτέρων πιθανότητα) και λίγα ή και καθόλου σε περιοχές με μικρή εκ των προτέρων πιθανότητα. Για να κάνουμε τον αλγόριθμο πιο ευσταθή σε ξαφνικές κινήσεις του αντικειμένου μπορούμε να παίρνουμε περισσότερα δείγματα σε μεγαλύτερη περιοχή γύρω από τα αρχικά δείγματα. Για να το κάνουμε αυτό πρέπει να αυξήσουμε το πλήθος των δειγμάτων, κάνοντας τον αλγόριθμο πιο αργό.

3.3.1 Δειγματοληψία με βάση την Σημαντικότητα

Η τεχνική της δειγματοληψίας με βάση την σημαντικότητα βελτιώνει την τεχνική της δειγματοληψίας με βάρη (παράγραφος 3.2.2). Εφαρμόζεται όταν υπάρχει γνώση για την μορφή της συνάρτησης ενδιαφέροντος $g(x)$, δηλαδή ξέρουμε ποιες περιοχές του χώρου είναι πιο πιθανές να περιέχουν το αντικείμενο. Η ιδέα είναι να παράγουμε δείγματα s^i σε αυτές τις περιοχές δειγματοληπτόντας από την $g(x)$ και όχι από την εκ των προτέρων συνάρτηση πυκνότητας πιθανότητας $p(x)$. Το επιθυμητό αποτέλεσμα είναι να αποφύγουμε όσο το δυνατόν να δημιουργήσουμε δείγματα που θα έχουν μικρά βάρη και δεν θα επιλεγούν. Ένας όρος f/g πρέπει να προστεθεί στα βάρη των δειγμάτων και προκύπτει ο τύπος για τα βάρη

$$\pi^i = \frac{f(s^i)}{g(s^i)} p(z|x = s^i) \tag{3.9}$$

για να εξισορροπήσουμε τα ακανόνιστα δείγματα. Αυτός ο όρος διόρθωσης διασφαλίζει ότι για μεγάλο αριθμό δειγμάτων M η δειγματοληψία με βάση την σημαντικότητα δεν έχει καμία επίδραση στην εκτίμηση $\tilde{p}(x|z)$. Οποιαδήποτε συνάρτηση ενδιαφέροντος $g(x)$ μπορεί να επιλεγεί και αν το M είναι αρκετά μεγάλο τότε η $\tilde{p}(x|z)$ είναι μια καλή εκτίμηση της $p(x|z)$. Ο σκοπός της δειγματοληψίας με βάση την σημαντικότητα είναι να μειώσει την διακύμανση την εκτιμήσεων για καθορισμένο M και να αυξήσει την ευστοχία της $\tilde{p}(x|z)$ για μικρό M. Αφού τα δείγματα παράγονται από την $g(x)$ παίζει το ρόλο της συνάρτησης πυκνότητας πιθανότητας, αλλά δεν είναι απαραίτητο να έχει τη μορφή κάποιας συγκεκριμένης κατανομής.

3.3.2 Ο αλγόριθμος ICondensation

Η δειγματοληψία με βάση την σημαντικότητα μπορεί να εφαρμοστεί στον αλγόριθμο Condensation (παράγραφο 3.2.3) και με αυτό τον τρόπο παράγεται ο αλγόριθμος ICondensation (Importance Condensation) [10]. Η συνάρτηση ενδιαφέροντος τη χρονική στιγμή n δίνεται από την $g_n(x_n)$. Στον αλγόριθμο ICondensation οι θέσεις των δειγμάτων παράγονται από την

$$f_n(s_n^i) = \tilde{p}(x_n = s_n^i|z_1, \ldots, z_{n-1}) = \sum_{j=1}^{M} \pi_{n-1}^j p(x_n = s_n^i|x_{n-1} = s_{n-1}) \tag{3.10}$$

Πρέπει να σημειωθεί ότι ενώ το σύνολο $\{(s_n^i, \pi_n^i)\}$ παρέχει μια διακριτή αναπαράσταση μιας κατανομής η πρόβλεψη $\tilde{p}(x_n|z_1, \ldots, z_{n-1})$ είναι μία μεικτή συνεχής κατανομή από πυρήνες $p(x_n|x_{n-1})$ που αναπαριστούν την εξέλιξη του μοντέλου. Αντί να δειγματοληπτούμε από την $\tilde{p}(x_n|z_1, \ldots, z_{n-1})$, τα δείγματα s_n^i μπορούν να παραχθούν από κάποια $g_n(x_n)$ και τα βάρη καθορίζονται ως

$$\pi_n^i = \frac{f_n(s_n^i)}{g_n(s_n^i)} p(z_n|x_n = s_n^i) \tag{3.11}$$

Το αποτέλεσμα του όρου διόρθωσης είναι να διατηρηθεί η συνεχόμενη κίνηση που αναπαριστάται από το μοντέλο. Παρόλο που τα δείγματα είναι τοποθετημένα με βάση την g_n, η συνάρτηση που εκτιμάται από το σύνολο $\{(s_n^i, \pi_n^i)\}$ εξακολουθεί να προσεγγίζει την $p(x_n|z_n)$. Η δειγματοληψία με βάση την σημαντικότητα προσπαθεί να βελτιώσει την αποτελεσματικότητα της αναπαράστασης των δειγμάτων, χωρίς να αλλάζει το μοντέλο πιθανοτήτων. Πρέπει να σημειωθεί ότι η (3.10) επιβάλει ένα περιορισμό στην μορφή του μοντέλου. Στον αλγόριθμο Condensation είναι αρκετό να δειγματοληπτούμε από την $p(x_n|x_{n-1})$, αλλά στον ICondensation πρέπει να βρεθεί και η ακριβής τιμή της. Αν χρησιμοποιηθούν γκαουσιανές κατανομές, τότε ο υπολογισμός μπορεί να γίνει άμεσα. Το άθροισμα στην (3.10) πρέπει να υπολογιστεί στην (3.11) για κάθε $i = 1, \ldots, M$ και αυτό κάνει την πολυπλοκότητα του αλγορίθμου να είναι $O(M^2)$ (από $O(M)^*$).

Στην πράξη η συνάρτηση ενδιαφέροντος θα εξαχθεί από μια ατελή διαδικασία, και μπορεί να παραλείψει κάποια μέγιστα της $p(z_n|x_n)$. Είναι λοιπόν συνετό να δημιουργήσουμε κάποια δείγματα χρησιμοποιώντας την δειγματοληψία με βάρη και κάποια χρησιμοποιώντας την συνάρτηση ενδιαφέροντος g_n. Όσο η $\tilde{p}(x_n|z_{n-1})$ και η g_n δεν αποτυγχάνουν ταυτόχρονα στο να βρουν την θέση του αντικειμένου, η παρακολούθηση του θα είναι επιτυχής.

Είναι επίσης καλό να αυξήσουμε το μοντέλο ώστε να περιέχει με κάποια πιθανότητα q επαναρχικοποίηση, δηλαδή επανατοποθέτηση του αντικειμένου σύμφωνα με μια πιθανότητα $p(x_n)$ που είναι ανεξάρτητη από τις παρατηρήσεις z_i. Αυτό επιτρέπει να βρεθεί ένα αντικείμενο που εισέρχεται στην σκηνή ή να ξαναβρεθεί ένα αντικείμενο που έχει χαθεί λόγο κακών μετρήσεων, πχ αν είχε κρυφτεί πίσω από κάποιο άλλο αντικείμενο. Το τροποποιημένο μοντέλο είναι της μορφής

$$\tilde{p}'(x_n|z_{n-1}) = (1-q)\tilde{p}(x_n|z_{n-1}) + qp(x_n) \tag{3.12}$$

Όπου το $p(x_n)$ είναι η εκ των προτέρων πιθανότητα της αρχικοποίησης. Με αυτό το μοντέλο έχουμε δύο καταστάσεις. Μπορούμε να το επεκτείνουμε με το να κάνουμε δειγματοληψία με βάρη με πιθανότητα $1-q-r$, δειγματοληψία με βάση την σημαντικότητα με πιθανότητα r και αρχικοποίηση με πιθανότητα q. Στην απουσία οποιουδήποτε γνώσης για την αρχικοποίηση μπορούμε να χρησιμοποιήσουμε $p(x_n) = g_n(x_n)$.

Για την χρήση του αλγορίθμου πρέπει να:

- καθοριστεί η αρχική κατάσταση $\{(s_0^i, \pi_0^i), i = 1, \ldots, M\}$.
- να επιλεγεί κατάλληλα η πιθανότητα μετάβασης ενός βήματος $p(x_n|x_{n-1} = s_{n-1}^i)$.
- να επιλεγεί κατάλληλα η συνάρτηση πυκνότητας πιθανότητας $p(z_n|x_n = x_n^i)$.
- να επιλεγεί κατάλληλη συνάρτηση ενδιαφέροντος $g_n(x_n)$.
- να καθοριστούν οι τιμές των q και r.

Αλγόριθμος 4 ICondensation

Από τα προηγούμενο σύνολο δειγμάτων $\{(s^i_{n-1}, \pi^i_{n-1}), i = 1, \ldots, M\}$ της χρονικής στιγμής $n-1$ παράγεται το νέο σύνολο $\{(s^i_n, \pi^i_n), i = 1, \ldots, M\}$ της χρονικής στιγμής n. Η συνάρτηση ενδιαφέροντος $g_n(x_n)$ για τη χρονική στιγμή n θεωρούμε ότι είναι γνωστή. Δημιουργούμε το $i - o\sigma\tau\acute{o}$ δείγμα με τον παρακάτω τρόπο:

1 Επιλογή: επιλέγουμε την μέθοδο δειγματοληψίας παράγοντας έναν ομοιόμορφο τυχαίο αριθμό $a \in [0, 1)$.

2 Δειγματοληψία: δειγματοληπτούμε από την $\tilde{p}'(x_n|z_{n-1})$ ως εξής:

1. Αν $a < q$ χρησιμοποιούμε την εκ των προτέρων κατανομή αρχικοποίησης. Διαλέγουμε το s^i_n δειγματοληπτόντας από την $g_n(x_n)$ και θέτουμε τον διορθωτικό όρο σημαντικότητας $\lambda^i_n = 1$.
2. Αν $q \leq a \leq q + r$ χρησιμοποιούμε δειγματοληψία με βάση την σημαντικότητα. Διαλέγουμε το s^i_n δειγματοληπτόντας από την $g_n(x_n)$ και θέτουμε τον διορθωτικό όρο σημαντικότητας $\lambda^i_n = f_n(s^i_n)/g_n(s^i_n)$, όπου

$$f_n(s^i_n) = \sum_{j=1}^{M} \pi^j_{n-1} p(x_n = s^i_n | x_{n-1} = s^j_{n-1})$$

3. Αν $a \geq q+r$ χρησιμοποιούμε τη δειγματοληψία με βάρη του αλγορίθμου Condensation. Διαλέγουμε ένα αρχικό δείγμα s^i_{n-1} με πιθανότητα π^i_{n-1}, διαλέγουμε το s^i_n δειγματοληπτόντας από την $p(x_n|x_{n-1} = s^i_{n-1})$ και θέτοντας τον διορθωτικό όρο σημαντικότητας $\lambda^i_n = 1$.

3 Μέτρηση: λαμβάνουμε την νέα μέτρηση και καθορίζουμε το βάρος του νέου δείγματος και τον όρο διόρθωσης ως

$$\pi^i_n = \lambda^i_n p(z_n | x_n = s^i_n)$$

Και μετά κανονικοποιούμε έτσι ώστε $\sum_i \pi^i_n = 1$ και το αποθηκεύουμε στο $\{(s^i_n, \pi^i_n)\}$.

Κεφάλαιο 4

Ο αλγόριθμος Μέσης Μετατόπισης

4.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται ο αλγόριθμος της μέσης μετατόπισης [4], [5]. Ο αλγόριθμος της μέσης μετατόπισης είναι μια ευσταθής μέθοδος για την εύρεση τοπικών μεγίστων μιας συνάρτησης. Αυτό για συνεχείς συναρτήσεις είναι προφανές, αλλά για διακριτά σύνολα δεδομένων δεν είναι εύκολο. Ο προσδιορισμός ευσταθής είναι από στατιστικής πλευράς και αναφέρεται στο ότι ο αλγόριθμος αγνοεί δεδομένα που αποκλίνουν πολύ από το μέσο όρο (outliers). Επίσης λαμβάνει υπόψη δεδομένα που βρίσκονται μέσα σε μια συγκεκριμένη περιοχή και αγνοεί τα υπόλοιπα. Αυτό που κάνει είναι ότι προσπαθεί επαναληπτικά να βρει το τοπικό μέγιστο της συνάρτησης. Στο 4.2 περιγράφεται η αναπαράσταση των αντικειμένων με μορφή ιστογραμμάτων. Στο 4.3 ορίζεται η απόσταση ιστογραμμάτων βασισμένη στον συντελεστή Bhattacharyya. Στο 4.4 παρουσιάζεται ο αλγόριθμος μέσης μετατόπισης για παρακολούθηση αντικειμένων. Στο 4.5 παρουσιάζεται η επέκταση του αλγορίθμου ώστε να συμπεριλεμβάνει και πληροφορίες που εμφανίζονται στο φόντο του αντικειμένου. Στο 4.6 παρουσιάζονται κάποια παραδείγματα εφαρμογής του αλγορίθμου.

4.2 Αναπαράσταση Αντικειμένων

Για τον χαρακτηρισμό του αντικειμένου πρέπει πρώτα να επιλεγεί ο χώρος των χαρακτηριστικών. Η αναφορά στον αρχικό στόχο (target model) αναπαριστάται από μια συνάρτηση πυκνότητας πιθανότητας q στον χώρο των χαρακτηριστικών. Για παράδειγμα η συνάρτηση q μπορεί να παίρνει ως όρισμα το χρώμα του αντικειμένου (δηλαδή ο χώρος γνωρισμάτων είναι το χρώμα που είναι μονοδιάστατος μέχρι και τρισδιάστατος). Χωρίς βλάβη της γενικότητας θεωρούμε ότι το κέντρο του αντικειμένου βρίσκεται στην χωρική θέση $\mathbf{0}$ (τα έντονα σύμβολα υποδηλώνουν διανύσματα). Στην επόμενη εικόνα ένας υποψήφιος στόχος (target candidate) ορίζεται στην θέση $\mathbf{y}$ και αναπαριστάται από την συνάρτηση πυκνότητας πιθανότητας $p(\mathbf{y})$. Η τιμή και των δύο συναρτήσεων μπορούν να υπολογιστούν από τα δεδομένα. Για να μπορεί ο υπολογισμός να γίνει σε πραγματικό χρόνο χρησιμοποιούμε ιστογράμματα m κάδων. Έτσι έχουμε

$$\begin{array}{rll} \textit{target model}: & \hat{\mathbf{q}} = \{\hat{q}_u\}_{u=1\ldots m} & \sum_{u=1}^{m} \hat{q}_u = 1 \\ \textit{target candidate}: & \hat{\mathbf{p}}(\mathbf{y}) = \{\hat{p}_u(\mathbf{y})\}_{u=1\ldots m} & \sum_{u=1}^{m} \hat{p}_u = 1 \end{array}$$

Θα αναπαριστούμε την συνάρτηση ομοιότητας ανάμεσα στα $\hat{\mathbf{p}}$ και $\hat{\mathbf{q}}$ με το

$$\hat{\rho}(\mathbf{y}) \equiv \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] \tag{4.1}$$

Η συνάρτηση $\hat{\rho}(\mathbf{y})$παίζει τον ρόλο της πιθανοφάνειας και το τοπικό μέγιστο της μέσα στην εικόνα αναπαριστά το αντικείμενο στην δεύτερη εικόνα που είναι πιο όμοιο με το $\hat{\mathbf{q}}$ της πρώτης εικόνας. Αν χρησιμοποιηθούν μόνο φασματικά χαρακτηριστικά, τότε η συνάρτηση ομοιότητας θα έχει μεγάλες μεταβολές για γειτονικές περιοχές για γειτονικές περιοχές της εικόνας και η χωρική πληροφορία χάνεται. Για να βρούμε το μέγιστο τέτοιων συναρτήσεων μέθοδοι ελαχιστοποίησης που χρησιμοποιούν παραγώγους είναι δύσκολο να εφαρμοστούν και μπορεί να γίνει μόνο εξαντλητική αναζήτηση. Κανονικοποιούμε την συνάρτηση ομοιότητας χρησιμοποιώντας μία ισοτοπική συνάρτηση πυρήνα γύρω από το αντικείμενο στις χωρικές πεδίο ορισμού. Όταν τα βάρη του πυρήνα που εμπεριέχουν συνεχή χωρική πληροφορία χρησιμοποιούνται για να ορίσουμε το πεδίο ορισμού των χαρακτηριστικών, η $\hat{\rho}(\mathbf{y})$ γίνεται μια ομαλή συνάρτηση στο $\mathbf{y}$.

4.2.1 Μοντέλο αναφοράς

Το αντικείμενο που μας ενδιαφέρει (στο εξής θα αναφέρεται ως στόχος) αναπαριστάται από μια ελλειπτική περιοχή της εικόνας. Για να μειώσουμε την επίδραση των διαφορετικών μεγεθών των αξόνων της έλλειψης, κανονικοποιούμε την έλλειψη σε μοναδιαία σφαίρα. Αυτό γίνεται διαιρώντας με το μέγεθος του κάθε άξονα (h_x και h_y) τις χωρικές συντεταγμένες κάθε σημείου του αντικειμένου.

Έστω ότι $\{\mathbf{x}_i^*\}_{i=1\ldots n}$ είναι οι κανονικοποιημένες συντεταγμένες των εικονοστοιχείων (pixels) που περιέχονται στον στόχο. Ένας ισοτροπικός πυρήνας με κυρτό και γνησίως φθίνουσα κατανομή πυρήνα $k(x)$ [1], η οποία βάζει μικρότερα βάρη στα εικονοστοιχεία που

[1]Η κατανομή του πυρήνα K ορίζεται ως μια συνάρτηση $k : [0, \infty) \to \Re$ τέτοια ώστε $K(x) = k(\|x\|^2)$.

είναι πιο μακριά από το κέντρο. Με τη χρήση βαρών αυξάνεται η ευστάθεια, καθώς τα περιφερειακά εικονοστοιχεία είναι λιγότερο αξιόπιστα, και συχνά επηρεάζονται από επικαλύψεις με άλλα αντικείμενα ή παρεμβολές από το φόντο.

Η συνάρτηση $b : \Re^2 \rightarrow \{1 \dots m\}$ συσχετίζει κάθε εικονοστοιχείο που βρίσκεται στη θέση x_i^* με το δείκτη $b(\mathbf{x}_i^*)$ του κάδου του ιστογράμματος του χώρου των χαρακτηριστικών στο οποίο ανήκει. Η πιθανότητα του κάδου των χαρακτηριστικών $u = 1 \dots m$ του στόχου υπολογίζεται ως

$$\hat{q}_u = C \sum_{i=1}^{n} k(\|\mathbf{x}_i^*\|^2)\delta[b(\mathbf{x}_i^*) - u] \tag{4.2}$$

όπου δ είναι η συνάρτηση δέλτα του Kronecker. Ο όρος κανονικοποίησης C υπολογίζεται θέτοντας τον περιορισμό $\sum_{u=1}^{m} \hat{q}_u = 1$, και προκύπτει

$$C = \frac{1}{\sum_{i=1}^{n} k(\|\mathbf{x}_i^*\|^2)} \tag{4.3}$$

επειδή το άθροισμα της δέλτα συνάρτησης για $u = 1 \dots m$ είναι ίσο με 1.

4.2.2 Υποψήφιοι στόχοι

Έστω ότι $\{\mathbf{x}_i\}_{1 \dots n_h}$ είναι οι κανονικοποιημένες συντεταγμένες των εικονοστοιχείων στον υποψήφιο στόχο, που το κέντρο του βρίσκεται στη θέση $\mathbf{y}$ της τρέχουσα εικόνας. Η κανονικοποίηση των συντεταγμένων κληρονομείται από την εικόνα που έχει τον αρχικό στόχο. Χρησιμοποιώντας την ίδια κατανομή πυρήνα $k(x)$, αλλά με εύρος h, η πιθανότητα του κάδου των χαρακτηριστικών $u = 1 \dots m$ του υποψήφιου στόχου υπολογίζεται ως

$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right)\delta[b(\mathbf{x}_i) - u] \tag{4.4}$$

όπου

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right)} \tag{4.5}$$

είναι η σταθερά κανονικοποίηση. Πρέπει να σημειωθεί ότι το C_h δεν εξαρτάται από το $\mathbf{y}$, αφού τα $\mathbf{x}_i$ είναι οργανωμένα σε ένα κανονικοποιημένο πλέγμα και το $\mathbf{y}$ είναι ένα σημείο αυτού του πλέγματος. Οπότε το C_h μπορεί να υπολογιστεί εκ των προτέρων για κάποιο πυρήνα και διαφορετικές τιμές του h. Το εύρος h ορίζει την κλιμάκωση του υποψήφιου στόχου.

4.2.3 Ομαλότητα της συνάρτησης ομοιότητας

Η ομαλότητα της συνάρτησης (4.1) κληρονομεί τις ιδιότητες της κατανομής του πυρήνα $k(x)$ όταν ο αρχικός στόχος και ο υποψήφιος στόχος αναπαριστώνται από τις (4.2) και (4.4). Μια παραγωγίσιμη κατανομή πυρήνα οδηγεί σε παραγωγίσιμη συνάρτηση ομοιότητας και αποτελεσματικές μέθοδοι που χρησιμοποιούν την παράγωγο μπορούν να χρησιμοποιηθούν για να βρεθεί το τοπικό μέγιστο. Η παρουσία συνεχούς πυρήνα εισάγει μια διαδικασία παρεμβολής ανάμεσα στις θέσεις του πλέγματος του αντικειμένου. Η επιλεγμένη αναπαράσταση του αντικειμένου ως έλλειψη δεν περιορίζει τον τρόπο που η ομοιότητα μετριέται και πολλές συναρτήσεις μπορούν να χρησιμοποιηθούν για την ρ.

4.3 Απόσταση Ιστογραμμάτων

Η συνάρτηση ομοιότητας ορίζει την απόσταση ανάμεσα στο αρχικό στόχο και τον υποψήφιο στόχο. Για να διευθετηθούν συγκρίσεις ανάμεσα σε διάφορους στόχους, αυτή η απόσταση πρέπει να έχει μια μετρική μορφή. Ορίζουμε την απόσταση ανάμεσα σε δύο διακριτές κατανομές ως

$$d(\mathbf{y}) = \sqrt{1 - \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}]} \tag{4.6}$$

όπου επιλέγουμε

$$\hat{\rho}(\mathbf{y}) \equiv \rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u} \tag{4.7}$$

ως εκτιμητή τον συντελεστή Bhattacharyya [11]ανάμεσα στις κατανομές $\mathbf{p}$ και $\mathbf{q}$.

Ο συντελεστής Bhattacharyya είναι μέτρο διακριτού τύπου που έχει μια γεωμετρική ερμηνεία. Είναι το συνημίτονο της γωνίας ανάμεσα στα $m-$ *διάστατα* μοναδιαία διανύσματα $(\sqrt{\hat{p}_1}, \ldots, \sqrt{\hat{p}_m})^T$ και $(\sqrt{\hat{q}_1}, \ldots, \sqrt{\hat{q}_m})^T$. Το γεγονός ότι τα $\mathbf{p}$ και $\mathbf{q}$ είναι κατανομές είναι αυτό που λαμβάνεται υπ' όψη με το να τις αναπαριστούμε στην μοναδιαία σφαίρα. Ταυτόχρονα μπορούμε να ερμηνεύσουμε την (4.7) σαν την (κανονικοποιημένη) συσχέτιση ανάμεσα στα διανύσματα $(\sqrt{\hat{p}_1}, \ldots, \sqrt{\hat{p}_m})^T$ και $(\sqrt{\hat{q}_1}, \ldots, \sqrt{\hat{q}_m})^T$.

Το μέτρο που ορίζεται στην (4.6) έχει τις παρακάτω επιθυμητές ιδιότητες:

1. Έχει τις ιδιότητες συνάρτηση μέτρου (θετική και ίση με 0 μόνο για την απόσταση του διανύσματος από τον εαυτό του, τριγωνομετρική ανισότητα).

2. Έχει μια καθαρή γεωμετρική αναπαράσταση.

3. Χρησιμοποιεί διακριτές συχνότητες, οπότε είναι ανεξάρτητο από την κλιμάκωση του στόχου.

4. Είναι έγκυρη για αυθαίρετες κατανομές ιστογραμμάτων.

4.4 Παρακολούθηση του Αντικειμένου με τον Αλγόριθμο της Μέσης Μετατόπισης

Για να βρούμε την θέση του στόχου στην τρέχουσα εικόνα, η απόσταση (4.6) πρέπει να ελαχιστοποιηθεί ως συνάρτηση του $\mathbf{y}$. Η διαδικασία εντοπισμού αρχίζει από την θέση του στόχου στην προηγούμενη εικόνα και ψάχνει στην γειτονιά του. Αφού η συνάρτηση απόστασης είναι ομαλή χρησιμοποιούμε την παράγωγο που παρέχεται από το διάνυσμα της μέσης μετατόπισης.

Η πληροφορία που χρησιμοποιείται για τα ιστογράμματα είναι το χρώμα, αλλά το ίδιο μπορεί να εφαρμοστεί για υφή ή οποιοδήποτε συνδυασμό αυτών. Υποθέτουμε ότι οι επόμενες πληροφορίες είναι διαθέσιμες

1. Η αναγνώριση και ο εντοπισμός του αντικειμένου που μας ενδιαφέρει στην αρχική εικόνα.

2. Περιοδική ανάλυση του κάθε αντικειμένου για να αναγνωρίσουμε πιθανές αλλαγές στην αναπαράσταση του στόχου, που οφείλονται στην αλλαγή του χρώματος του αντικειμένου.

4.4.1 Ελαχιστοποίηση της απόστασης

Η ελαχιστοποίηση της απόστασης (4.6) είναι ισοδύναμη με τη μεγιστοποίηση του συντελεστή Bhattacharyya $\hat{\rho}(\mathbf{y})$. Η έρευνα για τη νέα τοποθεσία του στόχου στην τρέχουσα εικόνα αρχίζει από την θέση $\hat{\mathbf{y}}_0$, την οποία είχε ο στόχος στην προηγούμενη εικόνα. Για αυτό το λόγο οι πιθανότητες $\{\hat{p}_u(\hat{\mathbf{y}}_0)\}_{u=1\ldots m}$ του στόχου στην θέση $\hat{\mathbf{y}}_0$ στην τρέχουσα εικόνα πρέπει να υπολογιστούν πρώτα. Αναπτύσσοντας σε σειρά Taylor γύρω από τις τιμές $\hat{\rho}_u(\hat{\mathbf{y}}_0)$, η γραμμική προσέγγιση του συντελεστή Bhattacharyya (4.7) μετά από πράξεις είναι

$$\rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m}\hat{p}_u(\mathbf{y})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}} \tag{4.8}$$

Η προσέγγιση είναι ικανοποιητική όταν ο υποψήφιος στόχος $\{\hat{p}_u(\hat{\mathbf{y}})\}_{u=1\ldots m}$ δεν αλλάζει δραματικά από τον αρχικό $\{\hat{p}_u(\hat{\mathbf{y}}_0)\}_{u=1\ldots m'}$, η οποία είναι σωστή υπόθεση στις περισσότερες περιπτώσεις. Ο περιορισμός $\hat{p}_u(\hat{\mathbf{y}}_0) > 0$ (ή ένα μεγαλύτερο όριο) για όλα τα $u = 1 \ldots m$ μπορούν πάντα να διασφαλιστούν με το να μην χρησιμοποιηθούν οι τιμές που τον παραβιάζουν. Από την (4.4) προκύπτει

$$\rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}] \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{C_h}{2}\sum_{u=1}^{n_h} w_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right) \tag{4.9}$$

όπου

$$w_i = \sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}}\delta[b(\mathbf{x}_i) - u] \tag{4.10}$$

οπότε για να ελαχιστοποιήσουμε την απόσταση (4.6) πρέπει να μεγιστοποιηθεί ο δεύτερος όρος της εξίσωσης (4.9), αφού ο πρώτος όρος είναι ανεξάρτητος από το $\mathbf{y}$. Πρέπει να παρατηρήσουμε ότι ο δεύτερος όρος αναπαριστά την εκτίμηση της πυκνότητα που υπολογίζεται με την κατανομή πυρήνα $k(x)$ στην θέση $\mathbf{y}$ της τρέχουσας εικόνας με τα δεδομένα να έχουν βάρη (4.10). Η επικρατούσα τιμή αυτής της συνάρτησης πυκνότητας πιθανότητας στη γειτονιά είναι το επιθυμητό μέγιστο το οποίο μπορεί να βρεθεί με την μέθοδο της μέσης μετατόπισης [5]. Σε αυτή τη διαδικασία ο πυρήνας κινείται αναδρομικά από την τρέχουσα θέση $\hat{\mathbf{y}}_0$ στην νέα θέση $\hat{\mathbf{y}}_1$ σύμφωνα με την σχέση

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g\left(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h}\right\|^2\right)} \tag{4.11}$$

όπου $g(x) = -k'(x)$, υποθέτοντας ότι η παράγωγος της $k(x)$ υπάρχει για όλα τα $x \in [0, \infty)$, εκτός από ένα πεπερασμένο πλήθος σημείων. Ολόκληρος ο αλγόριθμος για την εύρεση του στόχου περιγράφεται παρακάτω.

Αλγόριθμος 5 Μεγιστοποίηση του συντελεστή Bhattacharyya $\rho[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}]$

Δίνονται ο αρχικός στόχος $\{\hat{q}_u\}_{u=1\ldots m}$ και η θέση του $\mathbf{y}_0$ της προηγούμενης εικόνας.

1 Αρχικοποίηση της θέσης του στόχου στην τρέχουσα εικόνα στο $\mathbf{y}_0$, υπολογισμός του $\{\hat{p}_u(\hat{\mathbf{y}}_0)\}_{u=1\ldots m'}$ και του

$$\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{\mathbf{q}}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u}$$

2 Παράγονται τα βάρη $\{w_i\}_{i=1\ldots n_h}$ σύμφωνα με το (4.10).

3 Εύρεση την επόμενης θέσης του υποψήφιου στόχου σύμφωνα με την (4.11).

4 Υπολογισμός των $\{\hat{p}_u(\hat{\mathbf{y}}_1)\}_{u=1\ldots m'}$ και του

$$\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_1)\hat{q}_u}$$

5 Όσο $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}] < \rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{\mathbf{q}}]$
ανάθεση $\hat{\mathbf{y}}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$
υπολογισμός $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}]$

6 Αν $\|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\| < \epsilon$ τερματίζουμε.
Αλλιώς θέτουμε $\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1$ και πάμε στο βήμα 2.

4.4.2 Υλοποίηση του αλγορίθμου

Το κριτήριο τερματισμού με το όριο ϵ που χρησιμοποιείται στο βήμα 6 παράγεται από τον περιορισμό ότι τα διανύσματα $\hat{\mathbf{y}}_0$ και $\hat{\mathbf{y}}_1$ πρέπει να είναι στο ίδιο εικονοστοιχείο στις αρχικές (όχι τις κανονικοποιημένες) χωρικές συντεταγμένες. Ένα μικρότερο όριο θα απαιτούσε ακρίβεια μικρότερη από το εικονοστοιχείο. Επίσης για να εξασφαλίσουμε την εκτέλεση του αλγορίθμου σε πραγματικό χρόνο θέτουμε ένα άνω όριο στο πλήθος των επαναλήψεων του αλγορίθμου.

Στην υλοποίηση του ο αλγόριθμος μπορεί να απλουστευθεί επιπλέον. Ο σκοπός του βήματος 5 είναι να αποφευχθούν πιθανά αριθμητικά προβλήματα στην μεγιστοποίηση του αλγορίθμου της μέσης μετατόπισης. Τα προβλήματα μπορούν να εμφανιστούν λόγο της γραμμικής προσέγγισης του συντελεστή Bhattacharyya, δηλαδή η τιμή της συνάρτησης στη νέα θέση να μειώνεται αντί να αυξάνεται. Αυτό όμως συμβαίνει σπάνια (στο 0.1% των περιπτώσεων) και για αυτό το βήμα 5 μπορεί να παραληφθεί. Αν το βήμα 5 δεν συμπεριφερθεί τότε δεν είναι απαραίτητος ο υπολογισμός του συντελεστή Bhattacharyya στα βήματα 1 και 4.

Στην πραγματική υλοποίηση του αλγορίθμου κάνουμε επαναλήψεις με το να υπολογίζουμε τα βάρη του βήματος 2 , να βρίσκουμε τη νέα θέση στο βήμα 3 και να ελέγχουμε το μέγεθος του πυρήνα στο βήμα 6. Ο συντελεστής Bhattacharyya υπολογίζεται μόνο μετά το τέλος του αλγορίθμου ανάμεσα στοω αρχικό στόχο και τον επιλεγμένο υποψήφιο στόχο.

Πυρήνες με κατανομή Epanechnikov [4]

$$k(x) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-x) & αν \quad x \leq 1 \\ 0 & αλλιώς \end{cases} \tag{4.12}$$

είναι καλή στην χρήση. Σε αυτή την περίπτωση η παράγωγος της κατανομής του πυρήνα $g(x)$ είναι σταθερή και η σχέση (4.11) απλοποιείται στην

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i}{\sum_{i=1}^{n_h} w_i} \tag{4.13}$$

δηλαδή είναι ένα απλό σταθμισμένο άθροισμα.

Ορίζουμε ως λειτουργική λεκάνη έλξης την περιοχή στην τρέχουσα εικόνα στην οποία η νέα θέση του αντικειμένου μπορεί να βρεθεί με χρήση του αλγορίθμου της μέσης μετατόπισης. Λόγω της χρήσης πυρήνα, η λεκάνη είναι τουλάχιστον ίση σε μέγεθος με τον αρχικό στόχο. Δηλαδή στην τρέχουσα εικόνα αν το κέντρο του στόχου παραμένει μέσα στην περιοχή της εικόνας που καλύπτεται από τον αρχικό στόχο στην προηγούμενη εικόνα, τότε το τοπικό μέγιστο του συντελεστή Bhattacharyya είναι αξιόπιστος δείκτης για την νέα θέση του αντικειμένου. Υποθέτουμε ότι η αναπαράσταση του στόχου παρέχει αρκετή διάκριση, έτσι ώστε ο συντελεστής Bhattacharyya να δίνει ένα μοναδικό μέγιστο στην γειτονική περιοχή.

Ο αλγόριθμος της μέσης μετατόπισης βρίσκει μια ρίζα της παραγώγου σαν συνάρτηση της θέσης, που μπορεί ωστόσο να αντιστοιχεί σε σαγματικό σημείο της επιφάνειας ομοιότητας. Τα σαγματικά σημεία είναι μη σταθερές λύσεις και επειδή ο θόρυβος στην εικόνα δρα σαν ανεξάρτητος παράγοντας μεταβολής των δεδομένων, δεν θα επηρεάσει σημαντικά τα αποτελέσματα σε μια σειρά εικόνων.

4.4.3 Προσαρμοσμένη κλιμάκωση

Ο αλγόριθμος που περιγράφεται στην παράγραφο 4.4.1, για ένα γνωστό αρχικό στόχο, βρίσκει μια θέση του στόχου στην τρέχουσα εικόνα η οποία ελαχιστοποιεί την απόσταση

(4.6) στην γειτονιά της προηγούμενης θέσης. Ωστόσο η κλιμάκωση του στόχου αλλάζει στον χρόνο, οπότε το εύρος h στην (4.4) πρέπει να προσαρμοστεί κατάλληλα. Αυτό είναι εφικτό λόγω του ότι η (4.6) δεν επιρεάζεται από το h.

Συμβολίζουμε με h_{prev} το εύρος στην προηγούμενη εικόνα. Βρίσκουμε το νέο εύρος h_{opt} στην τρέχουσα εικόνα με το να τρέξουμε τον αλγόριθμο τρείς φορές με $h = h_{prev}$, $h = h_{prev} + \Delta h$ και $h = h_{prev} - \Delta h$. Μια ενδεικτική τιμή τιμή είναι $\Delta h = 0.1 h_{prev}$. Το καλύτερο αποτέλεσμα h_{opt} που δίνει το μεγαλύτερο αποτέλεσμα για το συντελεστή Bhattacharyya διατηρείται. Για να αποφύγουμε μεγάλες μεταβολές στην κλιμάκωση, το εύρος που συσχετίζεται με την τρέχουσα εικόνα παίρνεται από φιλτράρισμα, με τον τύπο

$$h_{new} = \gamma h_{opt} + (1 - \gamma) h_{prev} \tag{4.14}$$

όπου μια ενδεικτική τιμή για το γ είναι το 0.1.

4.5 Σταθμισμένο Ιστόγραμμα Φόντου

Η πληροφορία που παρέχεται από το φόντο (background) είναι σημαντική για δύο τουλάχιστον λόγους. Πρώτον, αν κάποια από τα χαρακτηριστικά του στόχου υπάρχει και στο φόντο, η συνάφεια τους οδηγεί σε χειρότερη εύρεση του στόχου. Δεύτερον, σε πολλές περιπτώσεις είναι δύσκολο να περιγραφεί επακριβώς ο στόχος και το μοντέλο μπορεί να περιέχει χαρακτηριστικά του φόντου επίσης. Ταυτόχρονα, η ακατάλληλη χρήση της πληροφορίας του φόντου μπορεί να επηρεάσει την επιλογή της κλιμάκωσης του αντικειμένου, κάνοντας αδύνατη την μέτρηση της ομοιότητας ανάμεσα στις διάφορες τιμές κλιμάκωσης, και επομένως τον εντοπισμό του στόχου. Ο τρόπος που παρουσιάζεται παρακάτω είναι να παίρνεται μια απλή αναπαράσταση των χαρακτηριστικών του φόντου και να χρησιμοποιούνται για επιλογή μόνο τα αξιοσημείωτα μέρη της αναπαράστασης του αρχικού στόχου και των υποψήφιων στόχων.

Έστω ότι το $\{\hat{o_u}\}_{u=1\ldots m}$ (με $\sum_{u=1}^{m} \hat{o}_u = 1$) είναι μια διακριτή αναπαράσταση (ιστόγραμμα) του φόντου στο χώρο των χαρακτηριστικών και $\hat{o}^*$ η μικρότερη μη μηδενική τιμή. Αυτή η αναπαράσταση υπολογίζεται σε μια περιοχή τριγύρω από το στόχο. Το μέγεθος της περιοχής εξαρτάται από την εφαρμογή και μπορούμε να χρησιμοποιούμε μια περιοχή τρείς φορές μεγαλύτερη από την περιοχή του στόχου. Τα βάρη

$$\left\{ \upsilon_u = min\left(\frac{\hat{o}^*}{\hat{o}_u}, 1 \right) \right\}_{u=1\ldots m} \tag{4.15}$$

ορίζουν ένα μετασχηματισμό για τη αναπαράσταση του αρχικού στόχου και των υποψήφιων στόχων. Ο μετασχηματισμός μειώνει την σπουδαιότητα από αυτά τα χαρακτηριστικά που έχουν μικρό υ_u, δηλαδή αυτά που είναι χαρακτηριστικά στο φόντο. Ο νέος αρχικός στόχος ορίζεται ως

$$\hat{q}_u = C \upsilon_u \sum_{i=1}^{n} k(\|\mathbf{x}_i^*\|^2) \delta[b(\mathbf{x}_i^*) - u] \tag{4.16}$$

με τον όρο κανονικοποίηση C να ορίζεται ως

$$C = \frac{1}{\sum_{i=1}^{n} k(\|\mathbf{x}_i^*\|^2) \sum_{u=1}^{m} \upsilon_u \delta[b(\mathbf{x}_i^*) - u]} \tag{4.17}$$

Όμοια ο νέος υποψήφιος στόχος αναπαριστάται ως

$$\hat{p}_u(\mathbf{y}) = C_h \upsilon_u \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right) \delta[b(\mathbf{x}_i) - u] \tag{4.18}$$

όπου

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right) \sum_{u=1}^{m} \upsilon_u \delta[b(\mathbf{x}_i) - u]} \tag{4.19}$$

4.6 Πειραματικά Αποτελέσματα

Παρακάτω παρατίθενται διάφορα αποτελέσματα από την εφαρμογή του βασικού αλγορίθμου της μέσης μετατόπισης σε διάφορες εικονοσειρές. Τα πειράματα έχουν γίνει σε υπολογιστή με core 2 Duo επεξεργαστή στα 1.6 GHz με 2GB μνήμη RAM και χρησιμοποιήθηκε το πρόγραμμα Matlab.

Οι εικονοσειρά του σχήματος 4.1 αποτελείται από 61 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 5.367 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 4.2 αποτελείται από 22 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 23.228 δευτερόλεπτα.

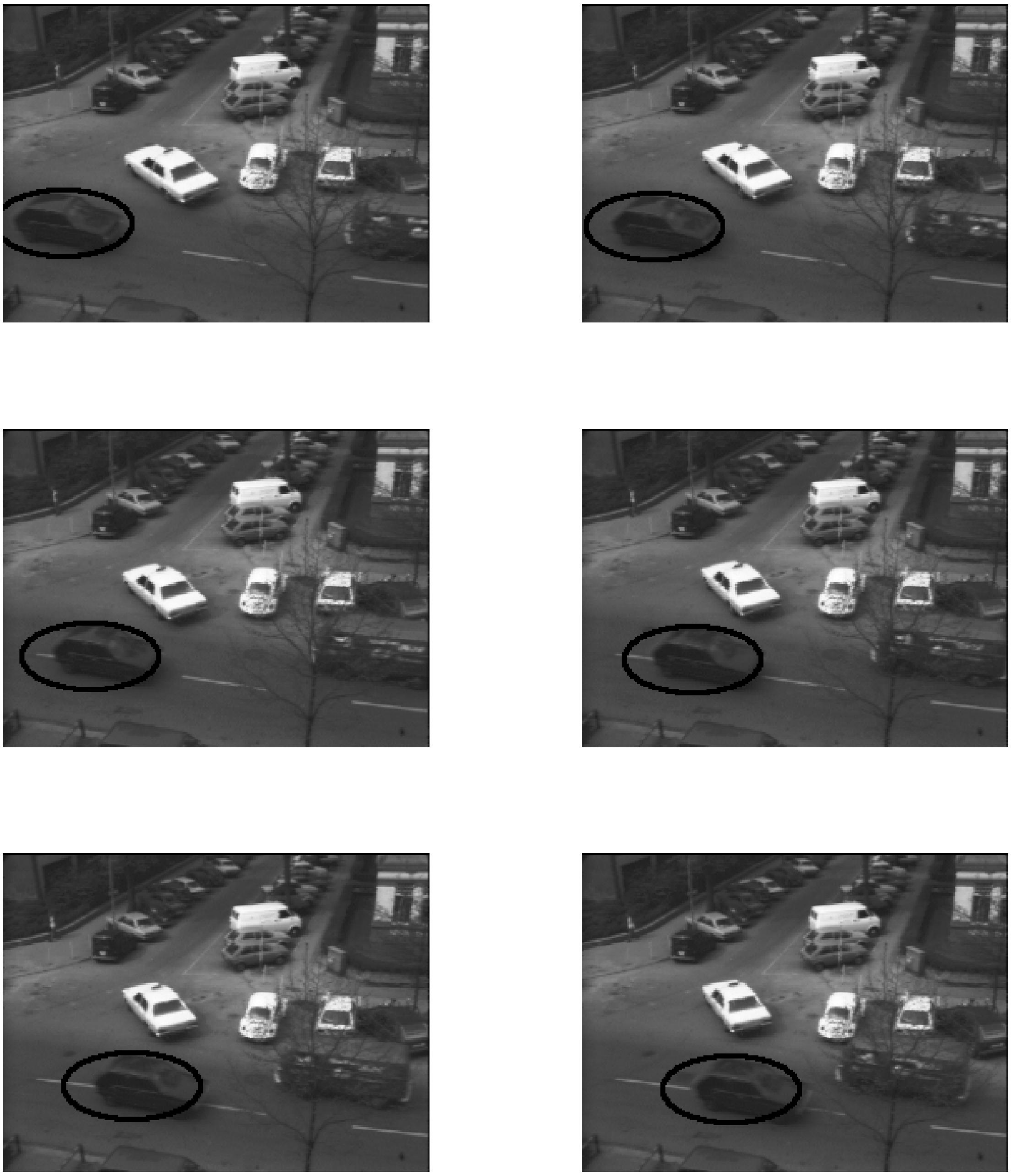
Οι εικονοσειρά του σχήματος 4.3 αποτελείται από 16 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 89.588 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 4.4 αποτελείται από 81 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 60.325 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 4.5 αποτελείται από 101 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 54.470 δευτερόλεπτα.

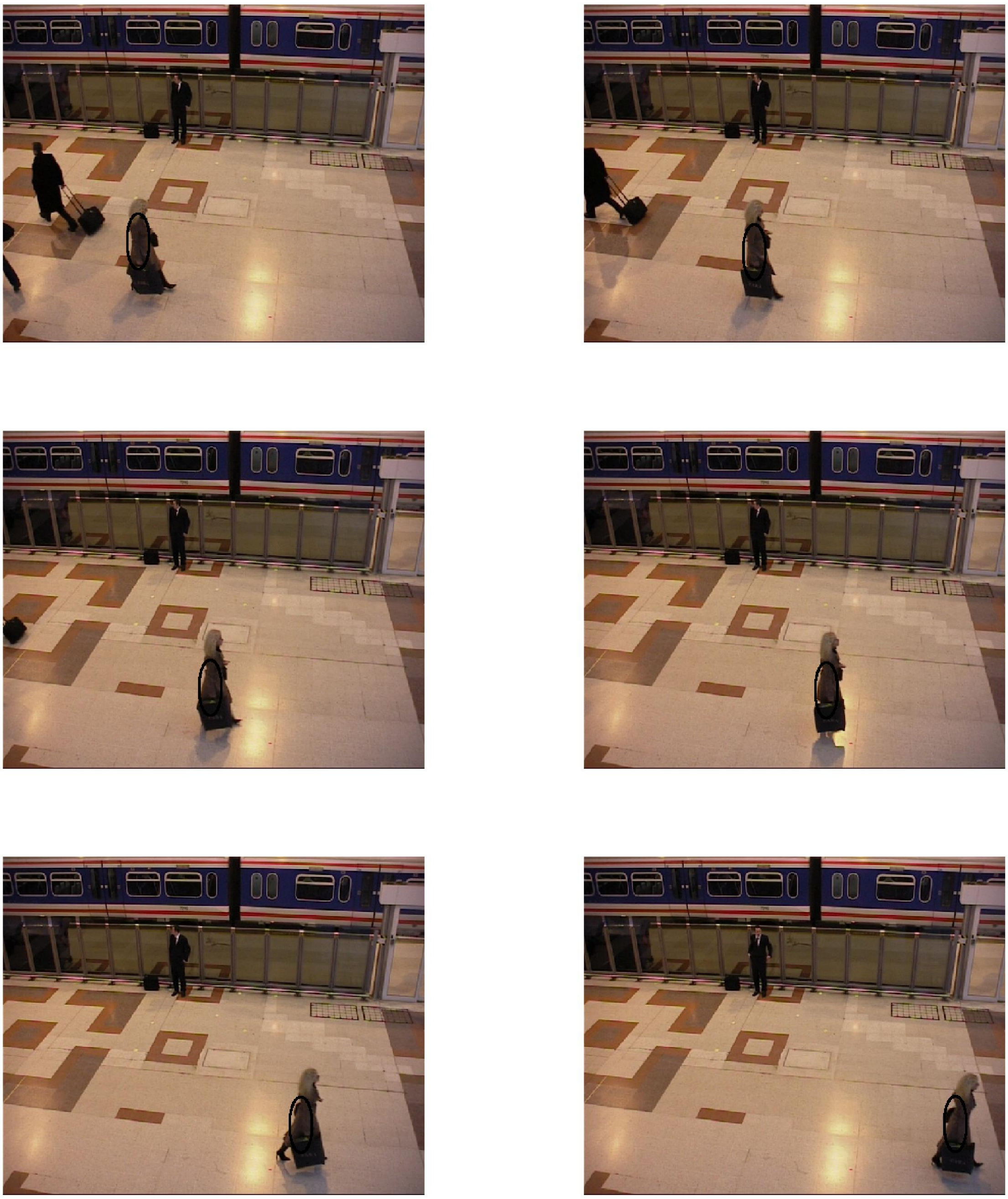
Οι εικονοσειρά του σχήματος 4.6 αποτελείται από 71 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 112.400 δευτερόλεπτα.

Σχήμα 4.1: Κήπος.



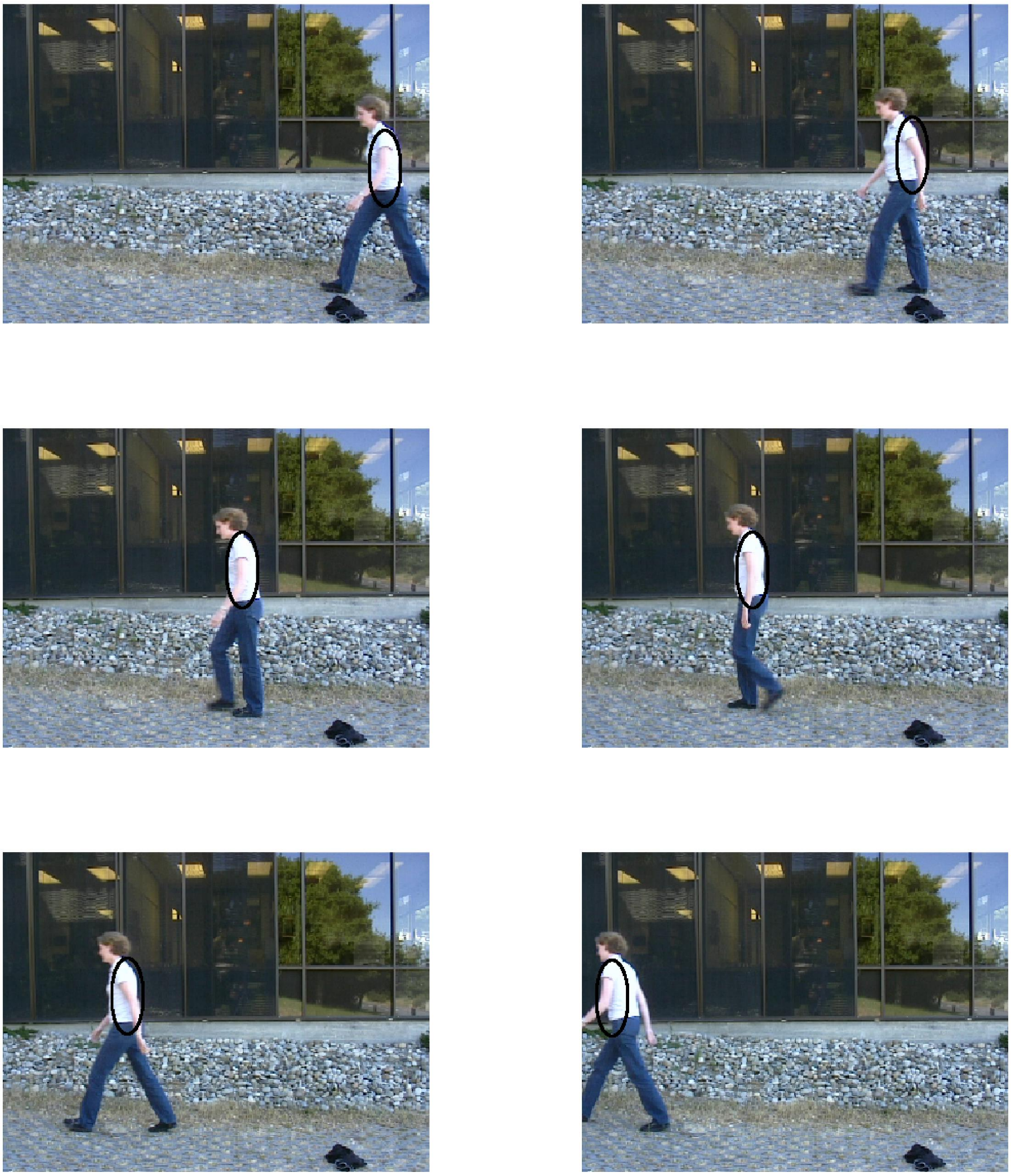
Σχήμα 4.2: Κινούμενο αυτοκίνητο 1.

Σχήμα 4.3: Κινούμενο αυτοκίνητο 2.



Σχήμα 4.4: Υπόγειος σιδηρόδρομος 1.

Σχήμα 4.5: Υπόγειος σιδηρόδρομος 2.



Σχήμα 4.6: Άνθρωπος που περπατά.

Κεφαλαιο 5

Αλγοριθμος Διαφορικης αποστασης EMD

5.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται ο αλγόριθμος εύρεσης θέσης με διαφορική Earth Mover's Distance (DEMD) [16]. Η EMD είναι ένα μέτρο σύγκρισης που είναι αρκετά ευσταθές σε αλλαγές του φωτισμού. Ωστόσο ο υπολογισμός της είναι υπολογιστικά ακριβός. Για αυτό ο αλγόριθμος DEMD υπολογίζει την παράγωγο της EMD σε σχέση με τη θέση, οπότε δεν είναι απαραίτητος ο υπολογισμός της EMD σε κάθε σημείο. Στο 5.2 περιγράφεται το μέτρο σύγκρισης EMD. Στο 5.3 παρουσιάζεται ο αλγόριθμος DEMD. Στο 5.4 παρουσιάζονται επεκτάσεις του DEMD ώστε να περιέχει και πληροφορία του φόντου. Στο 5.5 παρουσιάζονται κάποια παραδείγματα εφαρμογής του αλγορίθμου.

5.2 Η Απόσταση Κατανομών EMD

Η Earth Mover's Distance (EMD) παίρνει το όνομά της από την ιδέα ότι αν υπάρχουν δύο κατανομές, η μία μπορεί να αναπαρασταθεί ως λόφοι χώματος κατάλληλα τοποθετημένοι στον χώρο και η άλλη ως τρύπες διασκορπισμένες στον χώρο. Η EMD μετρά το ελάχιστο ποσό έργου που χρειάζεται να παραχθεί για να γεμίσουν όλες οι τρύπες με το χώμα από τους λόφους, όπου μια μονάδα έργου αντιστοιχεί στην μεταφορά μιας μονάδας χώματος σε απόσταση μιας μονάδας.

Η EMD χρησιμοποιείται για να συγκρίνει τις κατανομές του χρώματος ανάμεσα στον αρχικό στόχο και τον υποψήφιο στόχο [13], [16]. Οι κατανομές περιγράφονται σε μορφή υπογραφών. Οι υπογραφές είναι ένα σύνολο ομάδων χαρακτηριστικών και ορίζονται ως

$$\mathbf{s} = \{s_u\}_{u=1\ldots m}, \quad s_u = (a_u, w_u) \tag{5.1}$$

όπου το m είναι το πλήθος των ομάδων της υπογραφής, a_u είναι το μέσο της ομάδας u και το w_u το βάρος της ίδιας ομάδας.

Αναπαριστώντας τον αρχικό στόχο με την αρχική υπογραφή και τον υποψήφιο στόχο με την υποψήφια υπογραφή, μπορούμε να ορίσουμε την απόσταση ανάμεσα στην $u - οστή$ ομάδα της αρχικής υπογραφής και την $\upsilon - οστή$ ομάδα της υποψήφιας ως $d_{u\upsilon}$ και την ροή (δηλαδή το ποσό του μεταφερόμενου χώματος) ανάμεσά τους ως $f_{u\upsilon}(y)$. Ο σκοπός είναι να βρεθεί η θέση y που αντιστοιχεί στην μικρότερη τιμή EMD

$$arg\min_y(\min_{f_{u\upsilon}} Z(f_{u\upsilon}(y))) \tag{5.2}$$

όπου $EMD(f_{u\upsilon}(y)) = \min_{f_{u\upsilon}} Z(f_{u\upsilon}(y))$. Στην εξίσωση (5.2) πρώτα χρειάζεται μια ελαχιστοποίηση ώστε να βρεθεί η EMD για κάθε τοποθεσία και μετά μια επιπλέον για να βρούμε την καλύτερη θέση. Η EMD ορίζεται ως

$$Z(f_{u\upsilon}(y)) = \sum_{u=1}^{m^M} \sum_{\upsilon=1}^{m^C} d_{u\upsilon} f_{u\upsilon}(y)$$

υπό τους περιορισμούς

$$\begin{gathered}
\textstyle\sum_{u=1}^{m^M} f_{u\upsilon}(y) = w_\upsilon^C(y), \qquad 1 \le \upsilon \le m^C \\
\textstyle\sum_{\upsilon=1}^{m^C} f_{u\upsilon}(y) = w_u^M, \qquad 1 \le u \le m^M \\
\textstyle\sum_{u=1}^{m^M} \sum_{\upsilon=1}^{m^C} f_{u\upsilon}(y) = 1 \\
f_{u\upsilon}(y) \ge 0, \quad 1 \le u \le m^M, \quad 1 \le \upsilon \le m^C
\end{gathered}$$

Στις παραπάνω εξισώσεις το M δηλώνει το αρχικό αντικείμενο και το C το υποψήφιο αντικείμενο. Το w_u^M είναι το βάρος της $u - οστής$ ομάδας στην αρχική υπογραφή και το w_υ^C είναι το βάρος της $\upsilon - οστής$ ομάδας στην υποψήφια. Το m^M είναι το πλήθος ομάδων στην αρχική υπογραφή και το m^C το πλήθος ομάδων στην υποψήφια.

5.3 Ο Αλγόριθμος DEMD

5.3.1 Περίληψη του αλγορίθμου DEMD

Ο σκοπός είναι να βρεθεί ένας τύπος για την παράγωγο της απόστασης EMD σε σχέση με τη θέση. Επειδή ο τύπος για τον υπολογισμό της απόστασης EMD είναι ένα πρόβλημα γραμμικού προγραμματισμού, δεν μπορεί να υπολογιστεί η παράγωγος άμεσα. Για να ξεπεραστεί αυτό το πρόβλημα πρέπει να διασπαστεί ο υπολογισμός σε δύο στάδια [16].

Αναπαριστούμε την παράγωγο της EMD ως προς της θέση με $\nabla_y Z(y)$. Μπορούμε να την εκφράσουμε με τον κανόνα αλυσίδας ως την παράγωγο της Z προς τα βάρη $\frac{\partial Z(y)}{\partial w_\upsilon^C(y)}$ και την παράγωγο των βαρών προς την θέση $\nabla_y w_\upsilon^C(y)$, δηλαδή

$$\nabla_y Z(y) = \sum_{\upsilon=1}^{m^C} \frac{\partial Z(y)}{\partial w_\upsilon^C(y)} \nabla_y w_\upsilon^C(y) \tag{5.3}$$

όπου το w_υ^C είναι το βάρος της $\upsilon - o\sigma\tau\acute{\eta}\varsigma$ ομάδας και το m^C το πλήθος ομάδων στην υποψήφια υπογραφή.

5.3.2 Η μέθοδος simplex σε μορφή πινάκων

Μετασχηματίζουμε την μορφή της εξίσωσης (5.2) με χρήση πινάκων [3]. Υπάρχουν $m^M \times m^C$ μεταβλητές $f_{u\upsilon}(y)$ και $m^M \times m^C$ σταθερών $d_{u\upsilon}$. Θα χρησιμοποιούμε διανύσματα στηλών $\mathbf{f}(y)$ και $\mathbf{d}$ διάστασης $m^M \times m^C$ για να αναπαρασταθεί η ροή και η απόσταση. Παίρνοντας και τις τρεις εξισώσεις των περιορισμών της εξίσωσης (5.2) μαζί, μπορούμε να δημιουργήσουμε ένα δισδιάστατο πίνακα H, $m^M + m^C + 1$ γραμμών και $m^M + m^C$ στηλών του οποίου οι τιμές είναι 0 ή 1. Τέλος, συμβολίζουμε το διάνυσμα $[(\mathbf{w}^C(y))^T, (\mathbf{w}^M)^T, 1]^T$ σαν $\mathbf{b}(y)$ και $\mathbf{Z} = \mathbf{d}^T\mathbf{f}(y)$, οπότε προκύπτει η μορφή της εξίσωσης (5.2) ως

$$arg \min_y (\min_{\mathbf{f}} \mathbf{Z}) \tag{5.4}$$

υπό τους περιορισμούς

$$H\mathbf{f}(y) = \mathbf{b}(y)$$
$$\mathbf{f}(y) \geq \mathbf{0}$$

Για να γίνουν κάποιες πράξεις με τους πίνακες οι πίνακες αναδιατάσσονται. Αφού υπάρχουν $m^M \times m^C$ μεταβλητές και $m^M + m^C + 1$ περιορισμοί στο πρόβλημα, υπάρχουν $m^M + m^C + 1$ (δηλαδή μεταβλητές με μη μηδενική τιμή) και $m^M \times m^C - (m^M + m^C + 1)$ μη βασικών μεταβλητών. Ομαδοποιούμε όλες τις βασικές μεταβλητές και όλες τις μη βασικές μαζί και σπάμε το διάνυσμα $\mathbf{f}$ σε $[\mathbf{f}_B^T, \mathbf{f}_{NB}^T]^T$ όπου το N δηλώνει τις βασικές μεταβλητές και το NB τις μη βασικές μεταβλητές. Όμοια σπάμε το διάνυσμα $\mathbf{d}$ σε $[\mathbf{d}_B^T, \mathbf{d}_{NB}^T]^T$ και τον πίνακα H σε $[\mathbf{H}_B, \mathbf{H}_{NB}]$. Έτσι ο αρχικός πίνακας για την μέθοδο simplex μπορεί να γραφεί στη μορφή που παρουσιάζεται στον πίνακα 5.1. Στον πίνακα 5.1 το RHS αναφέρεται στο δεξί μέρος της εξίσωσης, η δεύτερη γραμμή αντιστοιχεί στην αντικειμενική συνάρτηση της εξίσωσης (5.4) και η τρίτη γραμμή στους περιορισμούς της εξίσωσης (5.4).

Κάνοντας πράξεις με τους πίνακες ο βέλτιστος πίνακας παίρνει του πίνακα 5.2.

5.3.3 Ανάλυση Ευαισθησίας της μεθόδου simplex

Με βάση τον βέλτιστο πίνακα 5.2 αναλύουμε την ευαισθησία της Z στην αλλαγή των βαρών των ομάδων των υπογραφών. Η ανάλυση μπορεί να εφαρμοστεί μόνο στα $\mathbf{w^C}(y)$, δηλαδή στα βάρη που εξαρτώνται από την θέση του υποψήφιου στόχου.

Πίνακας 5.1: Αρχικός πίνακας

Z	$\mathbf{f}_B$	$\mathbf{f}_{NB}$	**RHS**
1	$-\mathbf{d}_B^T$	$-\mathbf{d}_{NB}^T$	0
$\mathbf{0}$	H_B	H_{NB}	$\mathbf{b}$

Πίνακας 5.2: Βέλτιστος πίνακας

Z	$\mathbf{f}_B$	$\mathbf{f}_{NB}$	**RHS**
1	0	$-\mathbf{d}_{NB}^T + \mathbf{d}_B^T H_B^{-1} H_{NB}$	$\mathbf{d}_B^T H_B^{-1}\mathbf{b}$
$\mathbf{0}$	I	$H_B^{-1} H_{NB}$	$H_B^{-1}\mathbf{b}$

Για την δεύτερη γραμμή του πίνακα 5.2 έχουμε $Z = \mathbf{d}_B^T H_B^{-1}\mathbf{b}$. Υποθέτουμε ότι το $\mathbf{b}$ αλλάζει σε $\mathbf{b}'$, όπου το $\mathbf{b}'$ ορίζεται ως $b_i' = b_i + \Delta b_i, (1 \leq i \leq m^C)$, δηλαδή το βάρος i αλλάζει και τα υπόλοιπα βάρη $j \neq i$ δεν αλλάζουν. Η βέλτιστη λύση παίρνει τη μορφή

$$Z' = \mathbf{d}_B^T H_B^{-1}\mathbf{b}' = \mathbf{d}_B^T H_B^{-1}\mathbf{b} + \mathbf{d}_B^T H_B^{-1}[0 \ldots 0 \Delta b_i 0 \ldots 0]^T = \mathbf{d}_B^T H_B^{-1}\mathbf{b} + k_i \Delta b_i \tag{5.5}$$

όπου $k_i = \sum_{l=1}^{m^M+m^C+1} (\mathbf{d}_B)_l (H_B^{-1})_{li}$.

Οπότε προκύπτει

$$\frac{\partial Z}{\partial b_i} = \lim_{\Delta b_i \to 0} \frac{\Delta Z}{\Delta b_i} = \frac{k_i \Delta b_i}{\Delta b_i} = k_i \tag{5.6}$$

Επειδή το άθροισμα των βαρών των ομάδων πρέπει να είναι 1, η αλλαγή σε ένα βάρος πρέπει να επηρεάσει τα υπόλοιπα βάρη λόγο του περιορισμού κανονικοποίηση. Υπό αυτό τον περιορισμό προκύπτει ο τύπος

$$\frac{\partial Z}{\partial b_i} = k_i - \sum_{j \neq i} k_j \frac{b_j}{\sum_{l \neq i,j} b_l} \tag{5.7}$$

Η ιδέα είναι ότι προβάλουμε το k_i της εξίσωσης (5.6) από το χώρο διάστασης m^C στον χώρο διάστασης $m^C - 1$ λόγο του περιορισμού $\sum_{i=1}^{m^C} b_i = 1$. Η εξίσωση (5.7) παρέχει έναν αναλυτικό τύπο για τον υπολογισμό της παραγώγου της Z ως προς τις αλλαγές του χρώματος.

5.3.4 Αναπαράσταση του αντικειμένου με χρήση χαρακτηριστικών χρώματος

Χρησιμοποιούνται υπογραφές χρώματος αντί για ιστογράμματα, εξ αιτίας της περιεκτικότητάς τους. Χωρίς βλάβη της γενικότητας, θεωρούμε ότι το αντικείμενο στην αρχική εικόνα έχει κέντρο στο χωρικό σημείο 0 και υπάρχει μια συνάρτηση πυρήνα. Ορίζουμε την υπογραφή του στόχου στην αρχική εικόνα σύμφωνα με την εξίσωση (5.1) ως

$$\mathbf{s}^M = \{s_u^M\}_{u=1\ldots m^M}, \quad s_u^M = (a_u^M, w_u^M) \tag{5.8}$$

όπου

$$w_u^M = \beta \sum_{n=1}^{N} k\left(\left\|\frac{x_n}{h}\right\|^2\right)\delta[c(x_n) - u] \tag{5.9}$$

Στις εξισώσεις (5.8) και (5.9) το m^M είναι το πλήθος των ομάδων, το x_n δηλώνει τις δισδιάστατες συντεταγμένες της εικόνας και το πλήθος των εικονοστοιχείων είναι N. Η c είναι μια συνάρτηση που συσχετίζει την θέση x του εικονοστοιχείου στην ομάδα που είναι πιο κοντινή στο χρώμα του. Η $k(x)$ είναι μια ισοτροπική συνάρτηση κατανομής πυρήνα που αναθέτει μικρότερα βάρη στις θέσεις που είναι πιο μακριά από το κέντρο. Τα εικονοστοιχεία των αθροισμάτων περιέχονται σε ένα παράθυρο που είναι τριγύρω από το κέντρο του αντικειμένου., όπου το h αναπαριστά της ακτίνα της περιοχής. Η συνάρτηση δ είναι η Kronecker συνάρτηση δέλτα και το β είναι ένας όρος κανονικοποίηση.

Όμοια ορίζουμε την υπογραφή του υποψηφίου στόχου στην νέα εικόνα, το κέντρο του οποίου υποθέτουμε ότι βρίσκεται στο σημείο y, ως

$$\mathbf{s}^C(y) = \{s_\upsilon^C(y)\}_{\upsilon=1\ldots m^C}, \quad s_\upsilon^C(y) = (a_\upsilon^C(y), w_\upsilon^C(y)) \tag{5.10}$$

όπου

$$w_\upsilon^C(y) = \gamma \sum_{n=1}^{N} k\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\delta[c(x_n) - \upsilon] \tag{5.11}$$

όπου m^C είναι το πλήθος ομάδων και γ ένας όρος κανονικοποίηση.

5.3.5 Εκτίμηση της παραγώγου της συνάρτηση πυκνότητας

Παίρνοντας τις παραγώγους των βαρών της εξίσωσης (5.11) ως προς y έχουμε

$$\nabla_y w_\upsilon^C(y) = \frac{2\gamma}{h^2} \sum_{n=1}^{N} (x_n - y) g\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\delta[c(x_n) - \upsilon] \tag{5.12}$$

Στην παραπάνω εξίσωση (5.12) ορίζουμε $g(x) = -k'(x)$ όπου k είναι η συνάρτηση πυρήνα με πεδίο ορισμού $k : [0, \infty) \to \Re$ και $k(\|x\|^2) = K(x)$.

5.3.6 Κλειστή μορφή της παραγώγου της EMD ως προς τη θέση

Από την εξίσωση (5.3) είναι $\nabla_y Z(y) = \sum_{\upsilon=1}^{m^C} \frac{\partial Z(y)}{\partial w_\upsilon^C(y)} \nabla_y w_\upsilon^C(y)$. Αντικαθιστώντας το $\frac{\partial Z(y)}{\partial w_\upsilon^C(y)}$ από την εξίσωση (5.7) και το $\nabla_y w_\upsilon^C(y)$ από την (5.12) προκύπτει η

$$\nabla_y Z(y) = \frac{2\gamma}{h^2} \sum_{n=1}^{N} (x_n - y) g\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\pi_n \tag{5.13}$$

όπου το π_n ορίζεται ως

$$\pi_n = \sum_{\upsilon=1}^{m^C} \left(k_\upsilon - \sum_{j \neq \upsilon} k_j \frac{b_j}{\sum_{l \neq i,j} b_j}\right)\delta[c(x_n) - \upsilon] \tag{5.14}$$

όπου $k_i = \sum_{l=1}^{m^M + m^C + 1} (\mathbf{d}_B)_l (H_B^{-1})_{li}$. Οπότε η ελαχιστοποίηση μπορεί να γίνει με την εξίσωση (5.13) με τον παρακάτω αλγόριθμο [16].

Αλγόριθμος 6 Γρήγορος διαφορικός EMD (DEMD)

Είσοδος: Το κέντρο του αντικειμένου στην προηγούμενη εικόνα: $y_0 = y^{i-1}$

Έξοδος: Το κέντρο του αντικειμένου στην τρέχουσα εικόνα y^i

1 Αρχικοποίηση της θέσης του αντικειμένου στην τρέχουσα εικόνα στο y_0. Υπολογισμός της $EMD(y)$ με την εξίσωση (5.2).

2 Υπολογισμός των βαρών $\{\pi_i\}_{n=1...N}$ για όλα τα εικονοστοιχεία του παραθύρου σύμφωνα με την εξίσωση (5.14).

3 Υπολογισμός της παραγώγου $\nabla_y Z(y_0)$ σύμφωνα με την (5.13).

4 Κίνηση του αντικειμένου ένα από τα 8 γειτονικά εικονοστοιχεία, ανάλογα με το διάνυσμα της παραγώγου. Έστω y_1 το νέο εικονοστοιχείο. Υπολογισμός της $EMD(y_1)$ με χρίση της (5.2).

5 Αν $EMD(y_1) > EMD(y_0)$, θέτουμε $y_0^i \leftarrow y_0$ και σταματάμε. Αλλιώς θέτουμε $y_0 \leftarrow y_1$ και πάμε στο βήμα 2.

5.4 Επεκτάσεις

Ο αλγόριθμος DEMD κάνει καλή παρακολούθηση του αντικειμένου στις περισσότερες περιπτώσεις. Ωστόσο υπάρχουν περιπτώσεις που αλλάζει η κλιμάκωση του αντικειμένου, επικάλυψη με άλλα αντικείμενα ή το φόντο. Για να καθορίσουμε την κλιμάκωση και την θέση, ορίζουμε τα χαρακτηριστικά του τοπικού φόντου και του αντικειμένου και τις ομοιότητες των δύο. Ο σκοπός είναι να βρούμε την θέση y του αντικειμένου και της κλιμάκωσης h που αντιστοιχούν στο μικρότερο άθροισμα για το EDM του αντικειμένου και το EMD του φόντου, δηλαδή

$$arg \min_{y,h} (\min_{f_{u\upsilon}} Z(f_{u\upsilon}(y,h)) + \min_{f_{u\upsilon}^{Bg}} Z(f_{u\upsilon}^{Bg}(y,h))) \tag{5.15}$$

όπου το Bg δηλώνει το τοπικό φόντο. Ο τύπος για την EMD είναι παρόμοια με την εξίσωση (5.2). Ο γραμμικός συνδυασμός των δύο EMD είναι ένας καλός τρόπος για εξισορροπήσουμε την επίδραση του φόντου στο αντικείμενο.

Για να έχουμε εκτέλεση σε πραγματικό χρόνο, η αρχική θέση για την τρέχουσα εικόνα χρησιμοποιούμε τον γρήγορο αλγόριθμο DEMD της παραγράφου 5.3.6. Αυτό είναι ένα καλό αρχικό σημείο για τα επόμενα βήματα που η κλιμάκωση και η θέση πρέπει να καθοριστούν σύμφωνα με την εξίσωση (5.15).

Αλγόριθμος 7 DEMD με χρήση και του φόντου

Είσοδος: Το κέντρο του αντικειμένου: $y_0 = y^i$ που επιστρέφεται από τον αλγόριθμο της παραγράφου 5.3.6. Η κλιμάκωση από την προηγούμενη εικόνα $h_0 = h^{i-1}$.

Έξοδος: Το κέντρο του αντικειμένου y^i και η κλιμάκωση h^i για την τρέχουσα εικόνα.

1 Αρχικοποίηση της θέσης του αντικειμένου στην y_0, και του h_1 σε $h_0, h_0 + 10\%, h_0 - 10\%$ και καθορισμός ποια κλιμάκωση είναι καλύτερη σύμφωνα με την εξίσωση (5.15).

2 Αν η κλιμάκωση h_1 είναι ίση με την h_0, θέτουμε $h^i \leftarrow h_0$, $y^i \leftarrow y_0$ και σταματάμε. Αλλιώς θέτουμε $h_0 \leftarrow h_1$ και τρέχουμε τον αλγόριθμο της παραγράφου 5.3.6 για να βρούμε την θέση y_1.

3 Αν το y_1 είναι ίσο με y_0, θέτουμε $h^i \leftarrow h_0$, $y^i \leftarrow y_0$ και σταματάμε. Αλλιώς θέτουμε $y_0 \leftarrow y_1$ και πάμε στο βήμα 1.

5.5 Πειραματικά Αποτελέσματα

Παρακάτω παρατίθενται διάφορα αποτελέσματα από την εφαρμογή του βασικού αλγορίθμου DEMD σε διάφορες εικονοσειρές. Διάφορες επιλογές για τις διάφορες παραμέτρους του αλγορίθμου είναι παρόμοιες με αυτές της παραγράφου (4.4.2). Τα πειράματα έχουν γίνει σε υπολογιστή με core 2 Duo επεξεργαστή στα 1.6 GHz με 2GB μνήμη RAM και χρησιμοποιήθηκε το πρόγραμμα Matlab.

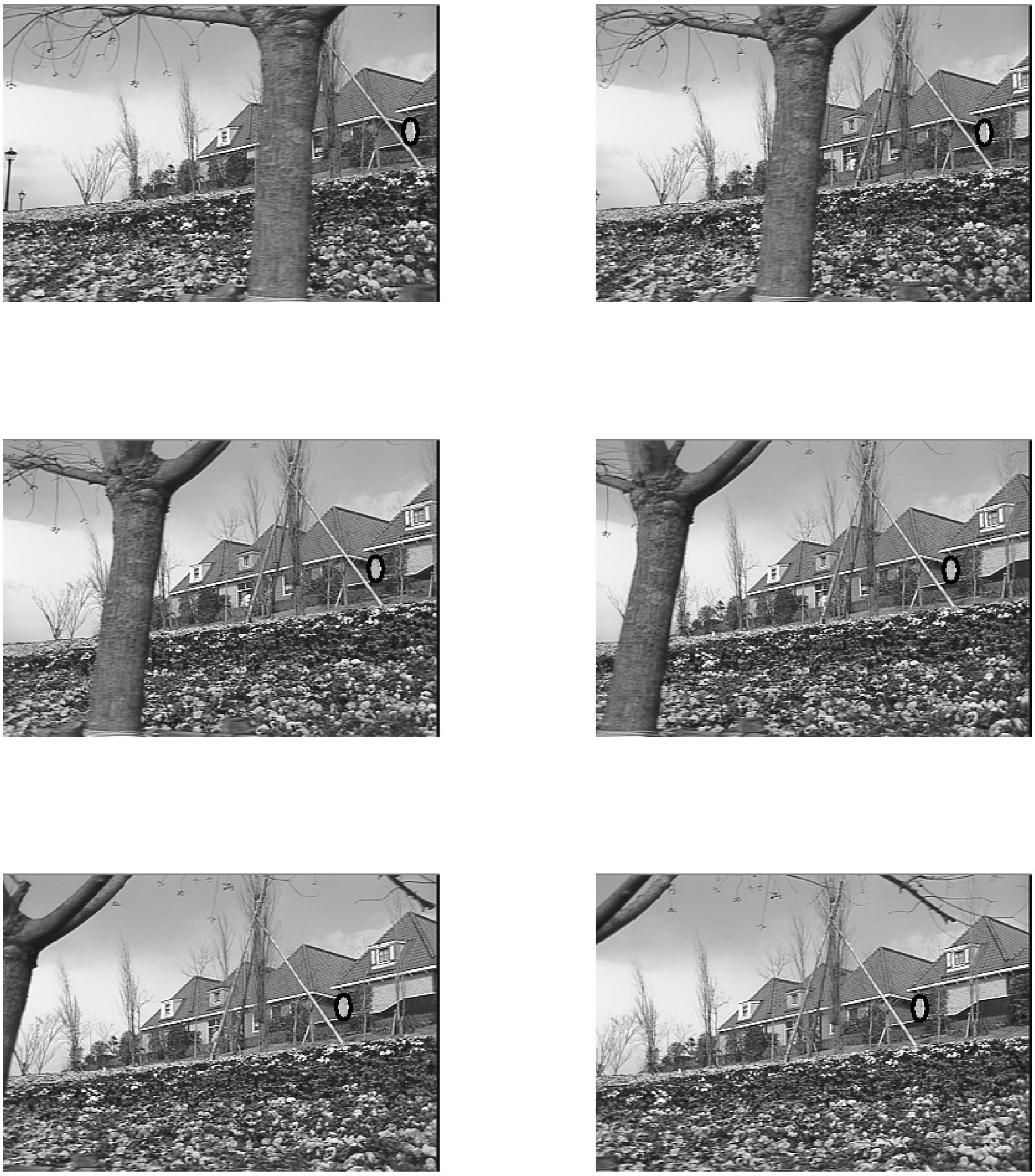
Οι εικονοσειρά του σχήματος 5.1 αποτελείται από 61 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 3.966 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 5.2 αποτελείται από 22 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 13.902 δευτερόλεπτα.

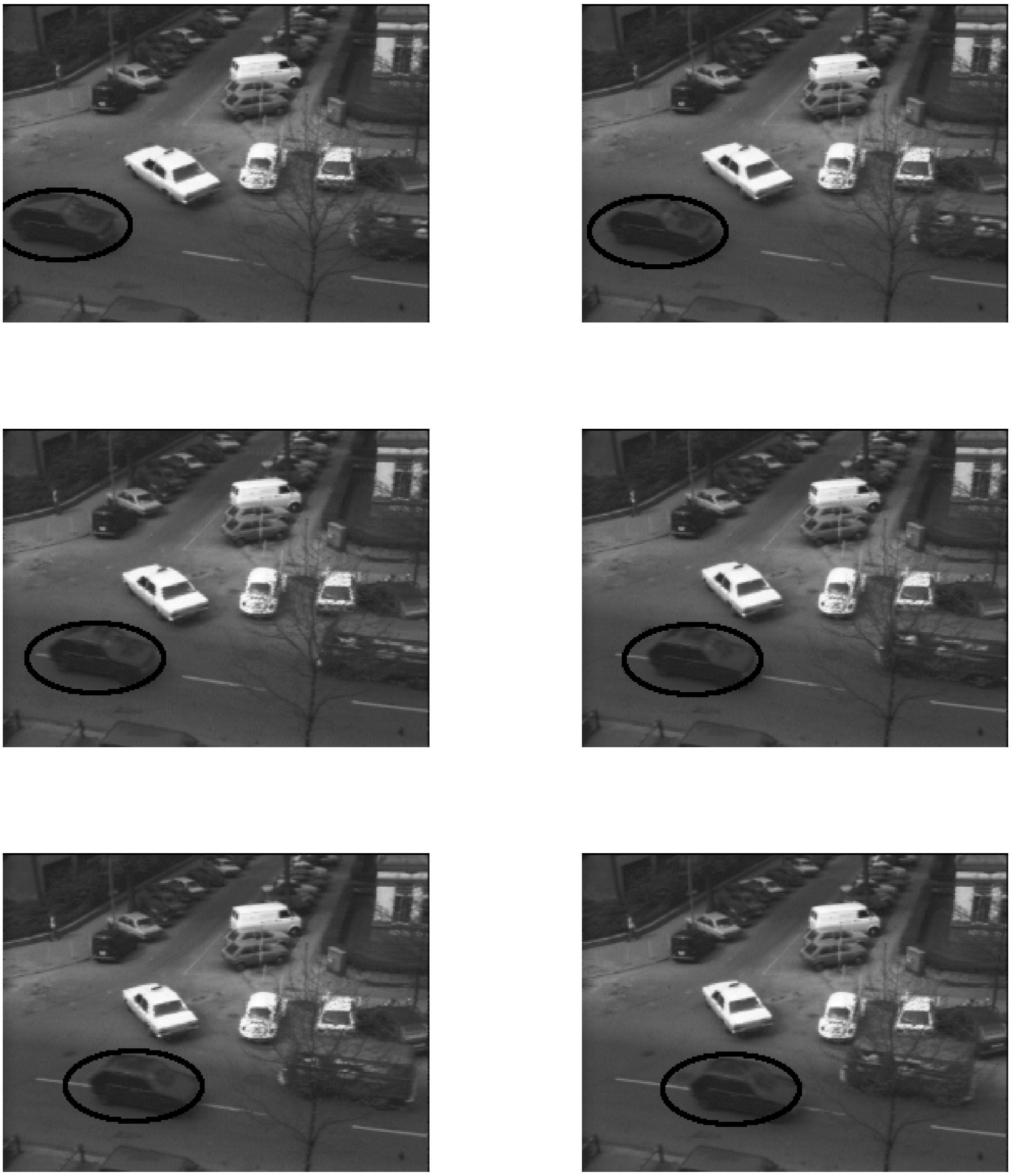
Οι εικονοσειρά του σχήματος 5.3 αποτελείται από 16 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 146.513 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 5.4 αποτελείται από 81 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 29.033 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 5.5 αποτελείται από 101 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 28.089 δευτερόλεπτα.

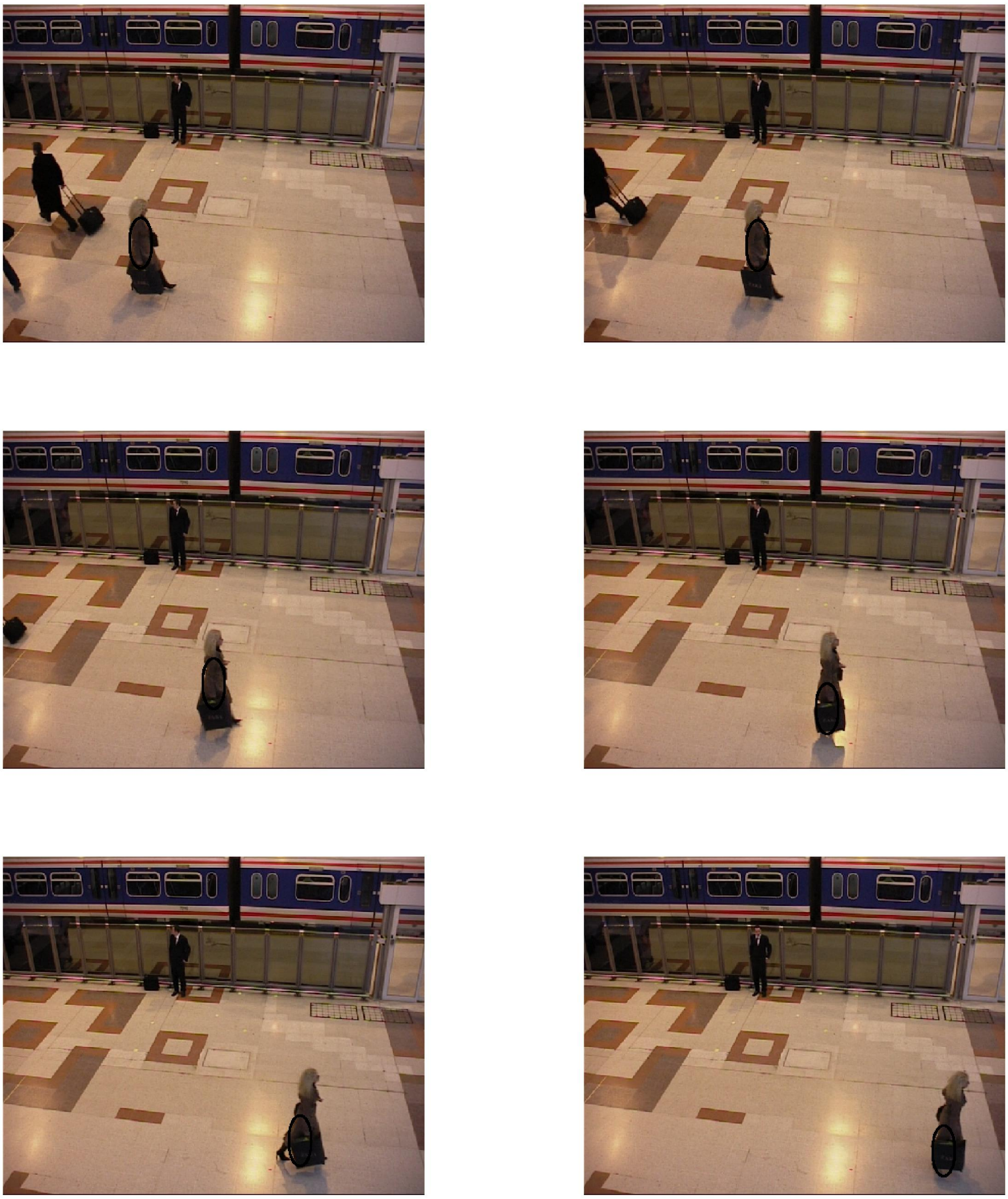


Σχήμα 5.1: Κήπος.



Σχήμα 5.2: Κινούμενο αυτοκίνητο 1.

Σχήμα 5.3: Κινούμενο αυτοκίνητο 2.



Σχήμα 5.4: Υπόγειος σιδηρόδρομος 1.

Σχήμα 5.5: Υπόγειος σιδηρόδρομος 2.

Κεφαλαιο 6

Διαφορικη EMD με μικτες κατανομες

6.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται ο συνδυασμός του αλγόριθμου εύρεσης θέσης με διαφορική Earth Mover's Distance (DEMD) [16] (που παρουσιάστηκε στο κεφάλαιο 5) με χρήση μικτών κανονικών κατανομών (Gaussian Mixture Model - GMM). Οι μικτές κανονικές κατανομές χρησιμοποιούνται για να αναπαραστήσουν τα χαρακτηριστικά του αντικειμένου που μας ενδιαφέρει. Στο 6.2 περιγράφεται το μέτρο σύγκρισης EMD για μικτές κανονικές κατανομές. Στο 6.3 παρουσιάζεται ο αλγόριθμος DEMD με μικτές κανονικές κατανομές. Στο 6.4 παρουσιάζονται κάποια παραδείγματα εφαρμογής του αλγορίθμου.

6.2 EMD και Μικτές Κανονικές Κατανομές

Στην παράγραφο 5.2 παρουσιάζεται η γενική ιδέα της Earth Mover's Distance (EMD) [2]. Στην παρούσα παράγραφο παρουσιάζεται η ειδική περίπτωση της EMD για μικτές κανονικές κατανομές.

6.2.1 Απόσταση ανάμεσα σε κανονικές κατανομές

Υπάρχουν πολλοί τρόποι για να ορίσουμε την απόσταση ανάμεσα σε δύο κατανομές. Στην παρακάτω παράγραφο περιγράφουμε την απόσταση ανάμεσα σε δύο κανονικές κατανομές η οποία βασίζεται στη απόσταση Kullback-Leibler.

Έστω δύο κανονικές κατανομές

$$f_1(x) = N(x|\mu_1, \sigma_1), \qquad f_2(x) = N(x|\mu_2, \sigma_2) \tag{6.1}$$

όπου μ είναι η μέση τιμή της κατανομή και σ η διασπορά της κατανομής. Απόσταση Kullback-Leibler ανάμεσα στις δύο κανονικές κατανομές της εξίσωσης (6.1) ορίζεται ως

$$D_{KL}(f_1|f_2) = \frac{1}{2}\left[\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{\sigma_1^2}{\sigma_2^2} + \frac{(\mu_2-\mu_1)^2}{\sigma_2^2} - 1\right] \tag{6.2}$$

Η απόσταση (6.2) δεν είναι συμμετρική, οπότε για να ξεπεράσουμε αυτό το πρόβλημα ορίζουμε την συμμετρική Kullback-Leibler ως

$$d(f_1, f_2) = D_{KL}(f_1|f_2) + D_{KL}(f_2|f_1) \tag{6.3}$$

και κάνοντας μερικές πράξεις έρχεται στη μορφή

$$d(f_1, f_2) = \frac{1}{2}\left[\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} + (\mu_1-\mu_2)^2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) - 2\right] \tag{6.4}$$

αυτή η απόσταση θα χρησιμοποιηθεί ως η απόσταση ανάμεσα σε κανονικές κατανομές.

6.2.2 EMD ανάμεσα σε μικτές κανονικές κατανομές

Στην παρακάτω παράγραφο ορίζεται η απόσταση Earth Mover's Distance (EMD) ανάμεσα σε μικτές κανονικές κατανομές. Η ανάλυση είναι παρόμοια με αυτή της παραγράφου 5.2. Μία μικτή κανονική κατανομή ορίζεται ως

$$p(x) = \sum_{i=1}^{m} \pi_i N(x|\mu_i, \sigma_i) \tag{6.5}$$

όπου m είναι το πλήθος των κατανομών, $N(x|\mu_i, \sigma_i)$ η $i-οστή$ κανονική κατανομή και π_i η πιθανότητα της $i-οστή$ κατανομής. Ορίζουμε την υπογραφή βασισμένη σε μικτές κανονικές κατανομές (GMM υπογραφή) ως

$$\mathbf{s} = \{s_i\}_{i=1\ldots m}, \quad s_i = (\mu_i, \sigma_i, \pi_i) \tag{6.6}$$

Αναπαριστούμε τον αρχικό στόχο της πρώτης εικόνας με την αρχική GMM υπογραφή $\mathbf{s}^M = \{s_u^M\}_{u=1\ldots m^M}$ και τον υποψήφιο στόχο της επόμενης εικόνας με την υποψήφια GMM υπογραφή $\mathbf{s}^C = \{s_\upsilon^C\}_{\upsilon=1\ldots m^C}$. Ως χαρακτηριστικά του στόχου θεωρείται ότι είναι η φωτεινότητα, αλλά μπορεί να είναι οτιδήποτε, όπως χρώμα ή υφή. Μπορούμε να ορίσουμε την απόσταση ανάμεσα στην $u-οστή$ ομάδα της αρχικής GMM υπογραφής και την $\upsilon-οστή$ ομάδα της υποψήφιας ως $d_{u\upsilon}$ (σύμφωνα με την παράγραφο 6.2.1) και την ροή ανάμεσά τους ως $f_{u\upsilon}(y)$. Ορίζουμε την απόσταση Earth Mover's Distance ανάμεσα σε δύο μικτές κανονικές κατανομές $\mathbf{s}^M = \{s_u^M\}_{u=1\ldots m^M}$ και $\mathbf{s}^C = \{s_\upsilon^C\}_{\upsilon=1\ldots m^C}$ ως

$$EMD(f_{u\upsilon}(y)) = \min_{f_{u\upsilon}} Z(f_{u\upsilon}(y)) \tag{6.7}$$

όπου

$$Z(f_{u\upsilon}(y)) = \sum_{u=1}^{m^M}\sum_{\upsilon=1}^{m^C} d_{u\upsilon} f_{u\upsilon}(y)$$

υπό τους περιορισμούς

$$\begin{array}{c} \sum_{u=1}^{m^M} f_{u\upsilon}(y) = \pi_\upsilon^C(y), \qquad 1 \leq \upsilon \leq m^C \\ \sum_{\upsilon=1}^{m^C} f_{u\upsilon}(y) = \pi_u^M, \qquad 1 \leq u \leq m^M \\ \sum_{u=1}^{m^M}\sum_{\upsilon=1}^{m^C} f_{u\upsilon}(y) = 1 \\ f_{u\upsilon}(y) \geq 0, \quad 1 \leq u \leq m^M, \quad 1 \leq \upsilon \leq m^C \end{array}$$

όπου m^M είναι το πλήθος κατανομών της αρχικής GMM υπογραφής και m^C το πλήθος κατανομών της υποψήφιας.

Ο σκοπός είναι να βρεθεί η θέση y που αντιστοιχεί στην μικρότερη τιμή EMD, δηλαδή

$$arg\min_y(\min_{f_{u\upsilon}} Z(f_{u\upsilon}(y))) \tag{6.8}$$

Δηλαδή ψάχνουμε την θέση που δίνει την υποψήφια υπογραφή, η οποία μοιάζει πιο πολύ με την αρχική.

6.3 Διαφορική DEMD με Μικτές Κανονικές Κατανομές

Η ανάλυση που ακολουθεί στην παρούσα παράγραφο είναι ίδια με αυτή της παραγράφου 5.3, με τη διαφορά ότι αναφέρεται σε GMM υπογραφές.

6.3.1 Παράγωγος EMD σε GMM ως προς τη θέση

Όμοια με την παράγραφο 5.3.1 διασπούμε τον υπολογισμό σε δύο στάδια [16], δηλαδή στον υπολογισμό της παραγώγου της Z προς τα βάρη $\frac{\partial Z(y)}{\partial \pi_\upsilon^C(y)}$ και στην παράγωγο των βαρών προς την θέση $\nabla_y \pi_\upsilon^C(y)$, δηλαδή

$$\nabla_y Z(y) = \sum_{\upsilon=1}^{m^C} \frac{\partial Z(y)}{\partial \pi_\upsilon^C(y)} \nabla_y \pi_\upsilon^C(y) \tag{6.9}$$

όπου το π_υ^C είναι το βάρος της $\upsilon - \textit{οστής}$ ομάδας και το m^C το πλήθος ομάδων στην υποψήφια υπογραφή.

6.3.2 Παράγωγος EMD ως προ βάρη

Ακολουθείται όμοιος συλλογισμός με τις παραγράφους 5.3.2 και 5.3.3. Μετασχηματίζουμε την μορφή της εξίσωσης (6.8) με χρήση πινάκων. Υπάρχουν $m^M \times m^C$ μεταβλητές $f_{u\upsilon}(y)$ και $m^M \times m^C$ σταθερών $d_{u\upsilon}$. Θα χρησιμοποιούμε διανύσματα στηλών $\mathbf{f}(y)$ και $\mathbf{d}$ διάστασης $m^M \times m^C$ για να αναπαρασταθεί η ροή και η απόσταση. Παίρνοντας και τις τρεις εξισώσεις των περιορισμών της εξίσωσης (6.8) μαζί, μπορούμε να δημιουργήσουμε ένα δισδιάστατο πίνακα H, $m^M + m^C + 1$ γραμμών και $m^M + m^C$ στηλών του οποίου οι τιμές είναι 0 ή 1. Τέλος, συμβολίζουμε το διάνυσμα $[(\boldsymbol{\pi}^C(y))^T, (\boldsymbol{\pi}^M)^T, 1]^T$ σαν $\mathbf{b}(y)$ και $\mathbf{Z} = \mathbf{d}^T\mathbf{f}(y)$, οπότε προκύπτει η μορφή της εξίσωσης (6.8) ως

$$arg\min_y(\min_{\mathbf{f}} \mathbf{Z}) \tag{6.10}$$

υπό τους περιορισμούς

$$H\mathbf{f}(y) = \mathbf{b}(y)$$
$$\mathbf{f}(y) \geq \mathbf{0}$$

Ακλουθώντας όμοιο συλλογισμό με την παράγραφο 5.3.3 καταλήγουμε στον τύπο για τον υπολογισμό της παραγώγου της Z προς τα βάρη $\frac{\partial Z(y)}{\partial b_i(y)}$, δηλαδή

$$\frac{\partial Z}{\partial b_i} = k_i - \sum_{j\neq i} k_j \frac{b_j}{\sum_{l\neq i,j} b_l} \tag{6.11}$$

όπου $k_i = \sum_{l=1}^{m^M+m^C+1}(\mathbf{d}_B)_l(H_B^{-1})_{li}$.

6.3.3 Αναπαράσταση του αντικειμένου με χρήση GMM υπογραφής φωτεινότητας

Χρησιμοποιούνται GMM υπογραφές για την αναπαράσταση των αντικειμένων. Θεωρούμε ότι το αντικείμενο στην αρχική εικόνα έχει κέντρο στο χωρικό σημείο 0 και υπάρχει μια συνάρτηση πυρήνα. Ορίζουμε την υπογραφή του στόχου στην αρχική εικόνα σύμφωνα με την εξίσωση (6.6)ως

$$\mathbf{s}^M = \{s_u^M\}_{u=1\ldots m^M}, \quad s_u^M = (\mu_u^M, \sigma_u^M, \pi_u^M) \tag{6.12}$$

όπου

$$\pi_u^M = \frac{1}{N}\sum_{n=1}^{N} k\left(\left\|\frac{x_n}{h}\right\|^2\right) \frac{\pi_u^M N(I(\frac{x_n-y}{h})|\mu_u^M, \sigma_u^M)}{\sum_{i=1}^{m^M} \pi_i^M N(I(\frac{x_n-y}{h})|\mu_i^M, \sigma_i^M)} \tag{6.13}$$

Στις εξισώσεις (6.12) και (6.13) το m^M είναι το πλήθος των ομάδων, το x_n δηλώνει τις δισ-διάστατες συντεταγμένες της εικόνας και το πλήθος των εικονοστοιχείων είναι N. Η $k(x)$ είναι μια ισοτροπική συνάρτηση κατανομής πυρήνα που αναθέτει μικρότερα βάρη στις θέσεις που είναι πιο μακριά από το κέντρο. Τα εικονοστοιχεία των αθροισμάτων περιέχονται σε ένα παράθυρο που είναι τριγύρω από το κέντρο του αντικειμένου, όπου το h αναπαριστά της ακτίνα της περιοχής. Η συνάρτηση $I(\frac{x_n-y}{h})$ αναπαριστά τη φωτεινότητα του αντικειμένου στο σημείο $\frac{x_n-y}{h}$ (που είναι οι κανονικοποιημένες συντεταγμένες του σημείου x_n). Πρέπει να παρατηρήσουμε ότι στην εξίσωση (6.13) το π_u^M υπάρχει και στο αριστερό και στο δεξί μέρος της εξίσωσης. Ουσιαστικά πρόκειται για την εξίσωση του αλγορίθμου EM που προσαρμόζει τα βάρη.

Όμοια ορίζουμε την υπογραφή του υποψηφίου στόχου στην νέα εικόνα, το κέντρο του οποίου υποθέτουμε ότι βρίσκεται στο σημείο y, ως

$$\mathbf{s}^C(y) = \{s_\upsilon^C(y)\}_{\upsilon=1\ldots m^C}, \quad s_\upsilon^C = (\mu_\upsilon^C(y), \sigma_\upsilon^C(y), \pi_\upsilon^C(y)) \tag{6.14}$$

όπου

$$\pi^C_\upsilon(y) = \frac{1}{N}\sum_{n=1}^{N} k\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\frac{\pi^M_\upsilon N(I(\frac{x_n-y}{h})|\mu^C_\upsilon, \sigma^C_\upsilon)}{\sum_{i=1}^{m^M}\pi^M_i N(I(\frac{x_n-y}{h})|\mu^C_i, \sigma^C_i)} \tag{6.15}$$

όπου m^C είναι το πλήθος ομάδων. Πρέπει να σημειωθεί ότι για μα μην υπάρχουν προβλήματα υποθέτουμε ότι $m^M = m^C$. Επίσης θεωρούμε ότι $\mu^M_u = \mu^C_\upsilon(y)$ και $\sigma^M_u = \sigma^C_\upsilon(y)$, δηλαδή μόνο τα βάρη αλλάζουν σε σχέση με τη θέση.

6.3.4 Παράγωγος βαρών ως προς τη θέση

Παίρνοντας τις παραγώγους των βαρών της εξίσωσης (6.15) ως προς y έχουμε

$$\begin{aligned}\nabla_y \pi^C_\upsilon(y) =& \frac{1}{N}\sum_{n=1}^{N} g\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\frac{x_n - y}{h}A_{n,\upsilon} + \\ & \frac{1}{N}\sum_{n=1}^{N} k\left(\left\|\frac{x_n - y}{h}\right\|^2\right)B_{n,\upsilon}\nabla_y I\Big(\frac{x_n - y}{h}\Big)\end{aligned} \tag{6.16}$$

Στην παραπάνω εξίσωση (5.12) ορίζουμε $g(x) = -k'(x)$ όπου k είναι η συνάρτηση πυρήνα με πεδίο ορισμού $k : [0, \infty) \to \Re$ και $k(\|x\|^2) = K(x)$. Το$\nabla_y \frac{I(x_n-y)}{h}$ είναι η παράγωγος της φωτεινότητας ως προς τη θέση. Τα $A_{n,\upsilon}$ και $B_{n,\upsilon}$ ορίζονται ως

$$A_{n,\upsilon} = \frac{2}{h}\frac{\pi^M_\upsilon N(I(\frac{x_n-y}{h})|\mu^C_\upsilon, \sigma^C_\upsilon)}{\sum_{i=1}^{m^M}\pi^M_i N(I(\frac{x_n-y}{h})|\mu^C_i, \sigma^C_i)} \tag{6.17}$$

και

$$B_{n,\upsilon} = \frac{\sum_{i=1}^{m^M}[\pi^M_\upsilon N(I(\frac{x_n-y}{h})|\mu^C_\upsilon, \sigma^C_\upsilon)\pi^M_i N(I(\frac{x_n-y}{h})|\mu^C_i, \sigma^C_i)\Delta I_{n,v,i}]}{[\sum_{i=1}^{m^M}\pi^M_i N(I(\frac{x_n-y}{h})|\mu^C_i, \sigma^C_i)]^2} \tag{6.18}$$

όπου

$$\Delta I_{n,v,i} = \frac{I\big(\frac{x_n-y}{h}\big) - \mu^C_\upsilon}{\sigma^{C\,2}_\upsilon} - \frac{I\big(\frac{x_n-y}{h}\big) - \mu^C_i}{\sigma^{C\,2}_i} \tag{6.19}$$

6.3.5 Τελική μορφή Παραγώγου

Από την εξίσωση (6.9) και κάνοντας χρήση των εξισώσεων (6.11) και (6.16) προκύπτει

$$\begin{aligned}\nabla_y Z(y) =& \frac{1}{N}\sum_{n=1}^{N}\left[g\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\left(\frac{x_n - y}{h}\right)R_n + \right. \\ & \left. k\left(\left\|\frac{x_n - y}{h}\right\|^2\right)\nabla_y I\left(\frac{x_n - y}{h}\right)P_n\right]\end{aligned} \tag{6.20}$$

όπου

$$R_n = \sum_{v=1}^{m^M}\Gamma_v A_{n,v} \tag{6.21}$$

$$P_n = \sum_{v=1}^{m^M} \Gamma_v B_{n,v} \tag{6.22}$$

και

$$\Gamma_v = k_v - \sum_{j \neq n} k_j \frac{b_j}{\sum_{l \neq v,j} b_l} \tag{6.23}$$

$$k_i = \sum_{l=1}^{m^M+m^C+1} (\mathbf{d}_B)_l (H_B^{-1})_{li} \tag{6.24}$$

6.3.6 Αλγόριθμος DEMD με GMM

Παρακάτω παρουσιάζεται ο αλγόριθμος DEMD με χρήση μικτών κανονικών κατανομών

Αλγόριθμος 8 DEMD με GMM

Είσοδος: Το κέντρο του αντικειμένου στην προηγούμενη εικόνα: $y_0 = y^{i-1}$

Έξοδος: Το κέντρο του αντικειμένου στην τρέχουσα εικόνα y_0^i

1 Αρχικοποίηση της θέσης του αντικειμένου στην τρέχουσα εικόνα στο y_0. Υπολογισμός της $EMD(y)$ με την εξίσωση (6.7).

2 Υπολογισμός των R_n και P_n για όλα τα εικονοστοιχεία του παραθύρου σύμφωνα με τις εξισώσεις (6.21) και (6.22).

3 Υπολογισμός της παραγώγου $\nabla_y Z(y_0)$ σύμφωνα με την (6.20).

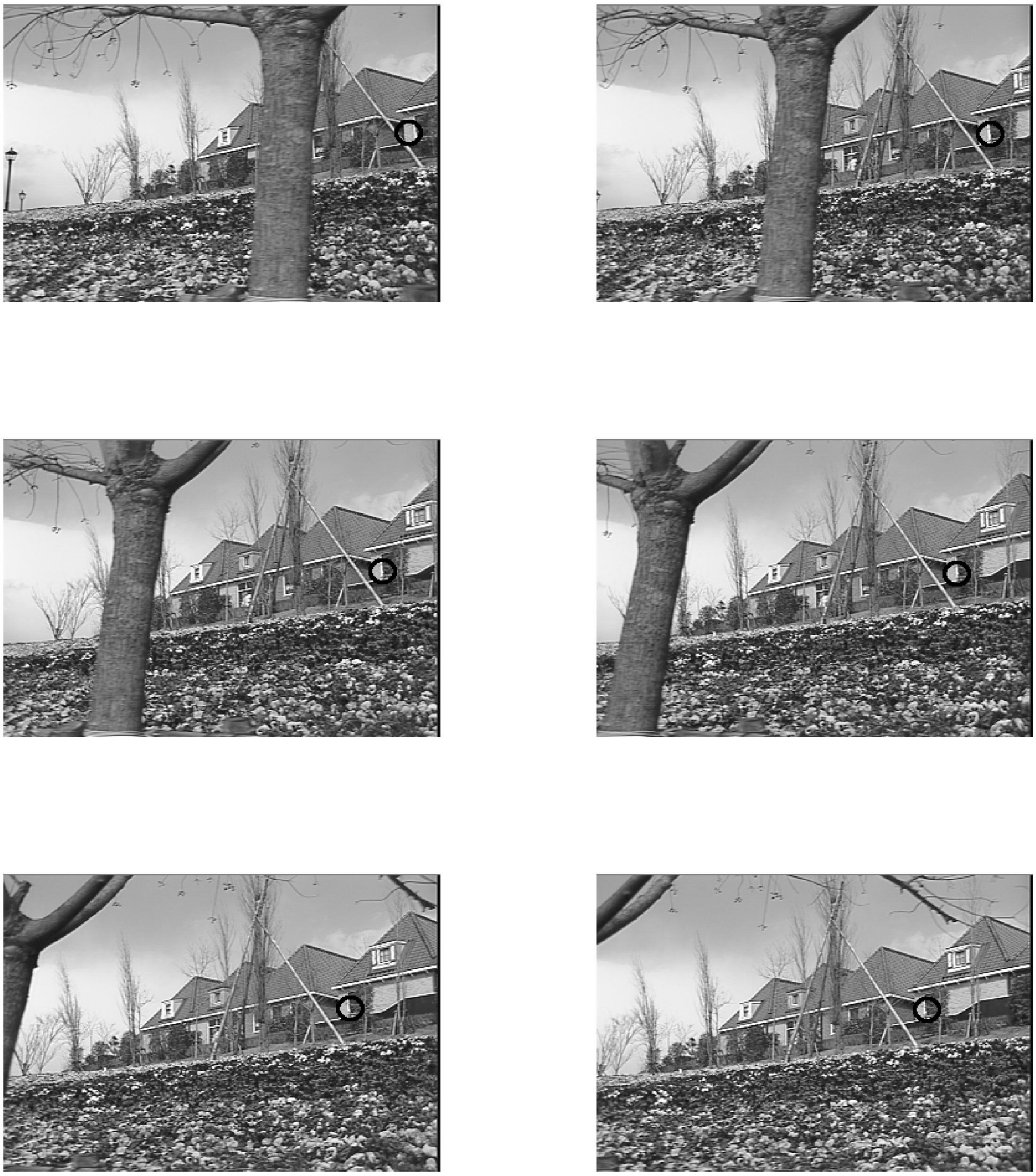
4 Κίνηση του αντικειμένου ένα από τα 8 γειτονικά εικονοστοιχεία, ανάλογα με το διάνυσμα της παραγώγου. Έστω y_1 το νέο εικονοστοιχείο. Υπολογισμός της $EMD(y_1)$ με χρίση της (6.7).

5 Αν $EMD(y_1) > EMD(y_0)$, θέτουμε $y_0^i \leftarrow y_0$ και σταματάμε. Αλλιώς θέτουμε $y_0 \leftarrow y_1$ και πάμε στο βήμα 2.

6.4 Πειραματικά Αποτελέσματα

Παρακάτω παρατίθενται διάφορα αποτελέσματα από την εφαρμογή του βασικού αλγορίθμου DEMD χρήση μικτών κανονικών κατανομών σε διάφορες εικονοσειρές. Διάφορες επιλογές για τις διάφορες παραμέτρους του αλγορίθμου είναι παρόμοιες με αυτές της παραγράφου (4.4.2). Τα πειράματα έχουν γίνει σε υπολογιστή με core 2 Duo επεξεργαστή στα 1.6 GHz με 2GB μνήμη RAM και χρησιμοποιήθηκε το πρόγραμμα Matlab.

Οι εικονοσειρά του σχήματος 6.1 αποτελείται από 61 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 3.776δευτερόλεπτα.



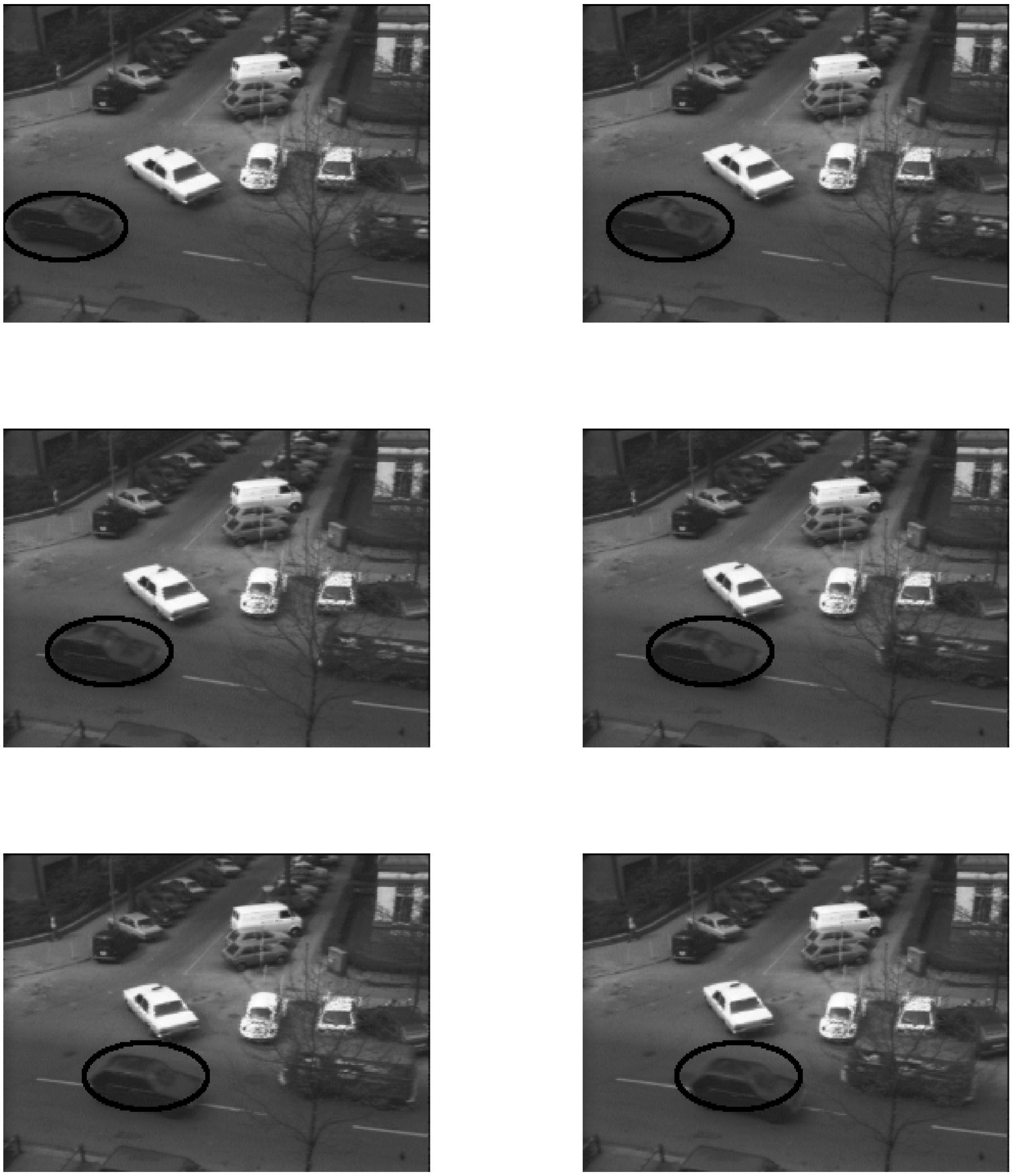
Σχήμα 6.1: Κήπος.

Οι εικονοσειρά του σχήματος 6.2 αποτελείται από 22 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 10.278 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 6.3 αποτελείται από 16 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 131.580 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 6.4 αποτελείται από 81 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 23.139 δευτερόλεπτα.

Οι εικονοσειρά του σχήματος 6.5 αποτελείται από 101 εικόνες και για την επεξεργασία

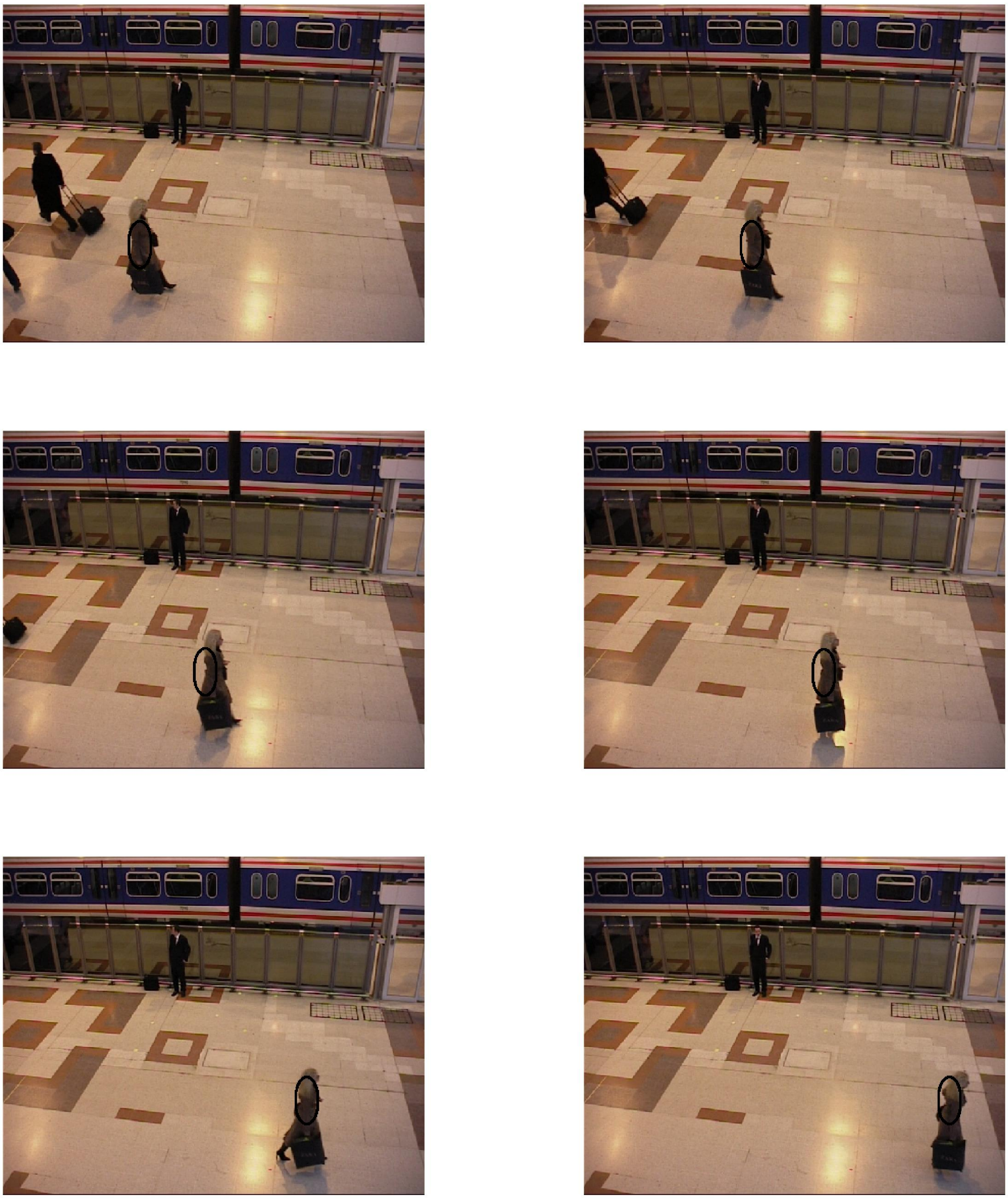


Σχήμα 6.2: Κινούμενο αυτοκίνητο 1.

τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 25.666 δευτερόλεπτα.

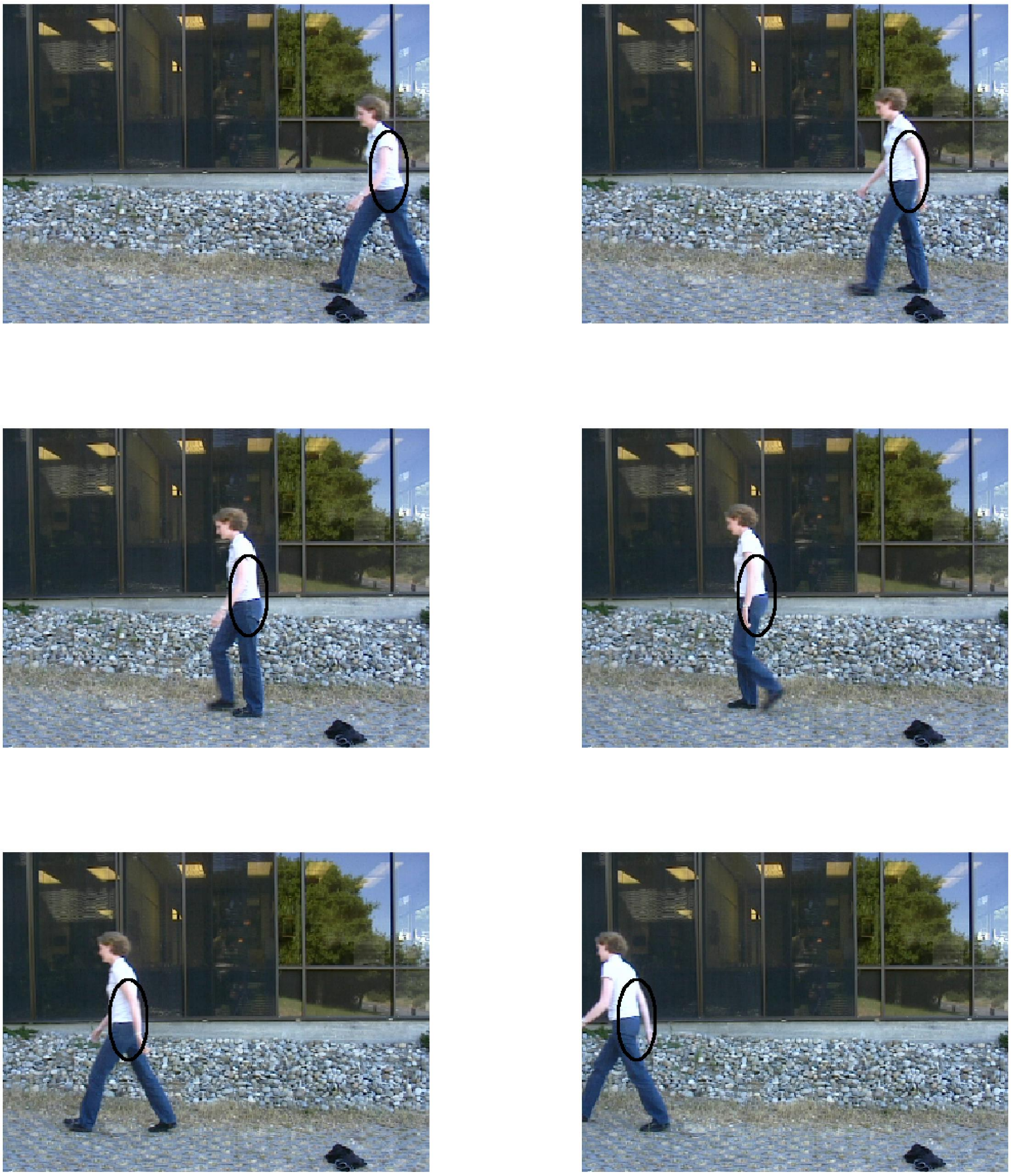
Οι εικονοσειρά του σχήματος 6.6 αποτελείται από 71 εικόνες και για την επεξεργασία τους ο αλγόριθμος της μέσης μετατόπισης χρειάστηκε 48.549 δευτερόλεπτα.

Σχήμα 6.3: Κινούμενο αυτοκίνητο 2.



Σχήμα 6.4: Υπόγειος σιδηρόδρομος 1.

Σχήμα 6.5: Υπόγειος σιδηρόδρομος 2.



Σχήμα 6.6: Άνθρωπος που περπατά.

Κεφαλαιο 7

Επιλογος

7.1 Επίλογος

Στην εργασία ασχοληθήκαμε με την εκτίμηση της κίνησης των αντικειμένων. Στο κεφάλαιο 2 παρουσιάστηκαν τα φίλτρα Kalman που είναι μια από τις πρώτες μεθόδους που αναπτύχτηκαν για ανίχνευση αντικειμένων και βασίζεται στην γνώση της κίνησης του αντικειμένου. Στο κεφάλαιο 3 παρουσιάστηκε η μέθοδος Condensation που είναι μια πιο γενική περίπτωση των φίλτρων Kalman. Στο κεφάλαιο 4 παρουσιάστηκε η αναπαράσταση αντικειμένων με χρήσης ιστογραμμάτων και η εφαρμογή του αλγορίθμου της μέσης μετατόπισης για την ανίχνευση αντικειμένων. Στο κεφάλαιο 5 παρουσιάστηκε ο αλγόριθμος DEMD που επίσης βασίζεται στην χρήση ιστογραμμάτων. Τέλος στο κεφάλαιο 6 προτάθηκε μια εναλλακτική αναπαράσταση των αντικειμένων με χρήση μικτών κανονικών κατανομών και η εφαρμογή τους σε μια παραλλαγή του αλγορίθμου DEMD.

Οι αλγόριθμοι των κεφαλαίων 4, 5 και 6 εφαρμόστηκαν σε ενδεικτικά σύνολα εικόνων και τα συμπεράσματα είναι ότι η συμπεριφορά τους στην γενική περίπτωση είναι αρκετά καλή. Σε ότι αφορά τον χρόνο εκτέλεσης παρατηρήσαμε ότι ο αλγόριθμος της μέσης μετατόπισης του κεφαλαίου 4 υστερεί σε σχέση με τον αλγόριθμο DEMD του κεφαλαίου 5 και ο οποίος με την σειρά του είναι πιο αργός από την επέκτασή του με μικτές κανονικές κατανομές του κεφαλαίου 6.

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3):428–440, 1999.

[2] C. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

[3] E. K. P. Chong and S. H. Zak. *An introduction to Optimazation.* Wiley Interscience, 2008.

[4] D. Comaniciu and P. Meer. Mean Shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[5] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[6] E. Cuevas, D. Zaldivar, and R. Rojas. Kalman filter for vision tracking. Technical report, Freier Universitat Berlin, Institut fur Informatik, 2005.

[7] D. Forsgth and J. Ponce. *Computer Vision: a modern approach.* Prentice Hall, 2003.

[8] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

[9] M. Isard and A. Blake. CONDENSATION - conditional density propagation for visual tracking. 29(1):5–28, 1998.

[10] M. Isard and A. Blake. ICONDENSATION - Unifying low-level and high-level tracking in a stochastic framework. In *Proceedings of the 5th European Conference on Computer Vision (ECCV)*, pages 893–908, 1998.

[11] T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15:52–60, 1967.

[12] S. Kay. *Fundamentals of statistical signal processing. Estimation Theory.* Prentice Hall, 1993.

[13] Y. Rubner, C. Tomasi, and L. J. Guibas. The Earth Mover's Distance as a metric for image retrieval. Technical report, Computer Science Department, Stanford University, 1998.

[14] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision.* Brooks, Cole, 1999.

[15] A. Yilmaz, O. Javed, and M. Shah. Object tracking: a survey. *ACM Computing Surveys*, 38(4):article No. 13, 2006.

[16] Q. Zhao, S. Brennan, and H. Tao. Differential EMD Tracking. Technical report, Department of Computer Engineering, University of California at Santa Cruz, 2007.

Βιογραφικο

Ο Βασίλειος Καραβασίλης γεννήθηκε στη Λάρισα το 1985. Αποφοίτησε το 2003 από το 4^o ενιαίο λύκειο Λάρισας. Οι βασικές σπουδές πραγματοποιήθηκαν στο Τμήμα Πληροφορικής του πανεπιστημίου Ιωαννίνων από όπου και αποφοίτησε το 2007. Συνέχισε για Μεταπτυχιακές σπουδές στο ίδιο Ίδρυμα και ασχολήθηκε με το πεδίο της τεχνητής όρασης. Για το διάστημα των μεταπτυχιακών σπουδών διατέλεσε και υπότροφος του Ιδρύματος Κρατικών Υποτροφιών.