

Ανίχνευση Κινούμενων Αντικειμένων σε Βίντεο με Μικτά Μοντέλα Παλινδρόμησης

Ιωάννα Φλεντζούρη

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

— ♦ —

Ιωάννινα, Ιούλιος 2009



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF IOANNINA



**ΑΝΙΧΝΕΥΣΗ ΚΙΝΟΥΜΕΝΩΝ ΑΝΤΙΚΕΙΜΕΝΩΝ ΣΕ VIDEO ΜΕ
ΜΙΚΤΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ**

**Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ**

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

από την

Ιωάννα Φλεντζούρη

ως μέρος των Υποχρεώσεων

για τη λήψη

του

**ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ**

Ιούλιος 2009



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΙΝΣΤΙΤΟΥΤΟ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΕΚΔΟΣΕΩΝ ΔΙΔΑΚΤΙΚΩΝ ΒΙΒΛΙΩΝ

Η

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΥΠΟΧΡΕΩΤΙΚΟ ΜΑΘΗΜΑ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ
ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ
ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΟΛΟΓΙΑ ΤΗΣ ΕΠΙΣΤΗΜΗΣ

ΕΠΙΣΤΗΜΟΛΟΓΙΑ



ΑΦΙΕΡΩΣΗ

Στην οικογένειά μου



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα αρχικά να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Κωνσταντίνο Μπλέκα, για την ουσιαστική βοήθεια και την πολύτιμη καθοδήγησή του σε όλη τη διάρκεια της εργασίας αυτής. Επίσης, θα ήθελα να ευχαριστήσω ιδιαιτέρως τους γονείς μου για την υποστήριξη και τη μεγάλη υπομονή που έδειξαν όλο αυτό το διάστημα. Τέλος, θα ήθελα να ευχαριστήσω όλους τους φίλους μου για τη συμπαράστασή τους σε κάθε δύσκολη στιγμή.



ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΠΕΡΙΕΧΟΜΕΝΑ	iv
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vi
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	vii
ΕΠΕΞΗΓΗΣΕΙΣ ΣΥΜΒΟΛΙΣΜΩΝ	viii
ΠΕΡΙΛΗΨΗ	1
EXTENDED ABSTRACT IN ENGLISH	2
 ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	 4
1.1. Σκοπός της εργασίας	4
1.2. Διάρθρωση των κεφαλαίων	6
 ΚΕΦΑΛΑΙΟ 2. ΕΠΙΣΚΟΠΗΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΟΜΑΔΟΠΟΙΗΣΗΣ (CLUSTERING)	 7
2.1. Τι είναι η ομαδοποίηση;	7
2.2. Βασικές τεχνικές ομαδοποίησης στατικών δεδομένων	8
2.3. Βασικές τεχνικές ομαδοποίησης χρονοσειρών	10
2.4. Βασικές τεχνικές ανίχνευσης κίνησης σε βίντεο	13
 ΚΕΦΑΛΑΙΟ 3. ΟΜΑΔΟΠΟΙΗΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕ ΜΙΚΤΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	 17
3.1. Εισαγωγή	17
3.2. Προηγούμενες Μελέτες	17
3.3. Πολυωνυμικό μικτό μοντέλο παλινδρόμησης (polynomial regression mixtures)	19
3.3.1. Ορισμός Μοντέλου	19
3.3.2. Αλγόριθμος ΕΜ για Πολυωνυμικά Μικτά Μοντέλα Παλινδρόμησης	20
3.3.3. Ομαδοποίηση μονοδιάστατων χρονοσειρών	22
3.3.4. Ομαδοποίηση δισδιάστατων χρονοσειρών	25
3.4. Άλλα μοντέλα παλινδρόμησης	28
3.4.1. Κατά τμήματα πολυωνυμικά μικτά μοντέλα παλινδρόμησης	28
3.4.2. Μικτό μοντέλο παλινδρόμησης με συναρτήσεις πυρήνα	31
 ΚΕΦΑΛΑΙΟ 4. ΟΜΑΔΟΠΟΙΗΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕ ΒΑΣΗ ΤΟ ΜΟΝΤΕΛΟ ΜΕΤΑΤΟΠΙΣΗΣ ΣΤΗ ΔΙΑΣΤΑΣΗ ΤΟΥ ΧΩΡΟΥ	 32
4.1. Εισαγωγή	32



4.2. Προηγούμενες Μελέτες	32
4.3. Μοντέλο Μετατόπισης στη διάσταση του χώρου	33
4.3.1. Ορισμός Μοντέλου Μετατόπισης	34
4.4. Ομαδοποίηση με το μοντέλο της μετατόπισης	36
4.4.1. Από κοινού, περιθώρια και πιθανοφάνεια	37
4.4.2. Αλγόριθμος ΕΜ για το μοντέλο της μετατόπισης	38
4.4.3. Ομαδοποίηση μονοδιάστατων χρονοσειρών	40
ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	43
5.1. Εισαγωγή	43
5.2. Τεχνητά Δεδομένα	44
5.3. Πραγματικά Δεδομένα	58
ΚΕΦΑΛΑΙΟ 6. Επίλογος	67
6.1. Συμπεράσματα	67
6.2. Μελλοντικές Επεκτάσεις	68
ΑΝΑΦΟΡΕΣ	70
ΠΑΡΑΡΤΗΜΑ	73
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	80



ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 5.1 Ακρίβεια μεθόδου για 1 ^ο βίντεο	54
Πίνακας 5.2 Ακρίβεια μεθόδου για 2 ^ο βίντεο	55
Πίνακας 5.3 Ακρίβεια μεθόδου για 3 ^ο βίντεο	55
Πίνακας 5.4 Ακρίβεια μεθόδου για 4 ^ο βίντεο	55



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 2.1 Αναπαράσταση διανυσμάτων κίνησης	16
Σχήμα 3.1 Αλγόριθμος EM για PRM μοντέλο (μονοδιάστατες χρονοσειρές)	25
Σχήμα 3.2 Αλγόριθμος EM για PRM μοντέλο (δισδιάστατες χρονοσειρές)	28
Σχήμα 4.1 Φαινόμενο μετατόπισης	34
Σχήμα 4.2 Αλγόριθμος EM για μοντέλο μετατόπισης (δισδιάστατες χρονοσειρές)	42
Σχήμα 6.1 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 1	46
Σχήμα 6.2 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 1	47
Σχήμα 6.3 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 2	48
Σχήμα 6.4 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 2	49
Σχήμα 6.5 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 3	50
Σχήμα 6.6 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 3	51
Σχήμα 6.7 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 4	52
Σχήμα 6.8 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 4	53
Σχήμα 6.9 Σύγκριση ακρίβειας	57
Σχήμα 6.10 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 1	59
Σχήμα 6.11 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 1	60
Σχήμα 6.12 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 2	61
Σχήμα 6.13 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 2	62
Σχήμα 6.14 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 3	63
Σχήμα 6.15 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 3	64
Σχήμα 6.16 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 4	65
Σχήμα 6.17 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 4	66



ΕΠΕΞΗΓΗΣΕΙΣ ΣΥΜΒΟΛΙΣΜΩΝ

Στην ενότητα αυτή περιγράφονται εν συντομία οι συμβολισμοί οι οποίοι χρησιμοποιούνται στην παρούσα εργασία.

Διανύσματα και πίνακες

Ένα διάνυσμα αναπαρίσταται με έντονη γραφή ως $\mathbf{x}$ ή $\mathbf{y}$. Ένα διάνυσμα με όλα τα στοιχεία του μηδέν συμβολίζεται με $\mathbf{0}$, ενώ ένα διάνυσμα με όλα τα στοιχεία του άσσους συμβολίζεται ως $\mathbf{1}$. Η απόστροφος χρησιμοποιείται για την αναπαράσταση του αντιστρόφου, έτσι ώστε το $\mathbf{x}'$ να αναπαριστά ένα διάνυσμα γραμμή. Οι πίνακες αναπαρίστανται με κεφαλαία και έντονα γράμματα ως $\mathbf{X}$ ή $\mathbf{V}$. Ο μοναδιαίος πίνακας συμβολίζεται ως $\mathbf{I}$, ενώ ο πίνακας με όλα τα στοιχεία του άσσους συμβολίζεται ως $\mathbf{1}$.

Σύνολα καμπυλών (curves)

Στην παρούσα εργασία οι καμπύλες αναπαρίστανται ως διανύσματα σταθερού μήκους. Έτσι, $\mathbf{y}_i$ είναι μια καμπύλη η οποία αποτελείται από μια ακολουθία T παρατηρήσεων ή μετρήσεων. Η j -th μέτρηση της $\mathbf{y}_i$ συμβολίζεται ως y_{ij} . Ο χρόνος συμβολίζεται ως x_i , οπότε το x_{ij} μας δείχνει την χρονική στιγμή στην οποία παρατηρήθηκε το y_{ij} . Με $\mathbf{Y}$ συμβολίζεται το σύνολο $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ των N καμπυλών. Παρόμοια, με $\mathbf{X}$ συμβολίζεται το σύνολο $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$.

Κατανομές πιθανότητας

Μια ακαθόριστη πυκνότητα πιθανότητας αναπαρίσταται ως $p(\mathbf{y}_i|\theta)$. Η Gaussian κατανομή με μέσο μ και διακύμανση σ^2 συμβολίζεται ως $\mathcal{N}(\mu, \sigma^2)$. Οπότε μια τυχαία Gaussian μεταβλητή x συμβολίζεται ως $x \sim \mathcal{N}(\mu, \sigma^2)$ ή $p(x) = \mathcal{N}(x|\mu, \sigma^2)$. Γενικά, κάθε πυκνότητα εξαρτάται από ένα σύνολο παραμέτρων. Για



παράδειγμα, αν $p(\mathbf{y}_i) = \mathcal{N}(\mathbf{y}_i | \mathbf{X}_i \boldsymbol{\beta}, \Sigma)$, τότε το αριστερό μέρος εξαρτάται από το σύνολο παραμέτρων $\theta = \{\boldsymbol{\beta}, \Sigma\}$ και από τον πίνακα $\mathbf{X}_i$.

Μοντέλα Παλινδρόμησης

Το απλό μοντέλο παλινδρόμησης p βαθμού για την καμπύλη $\mathbf{y}_i$ γράφεται ως

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

όπου $\boldsymbol{\beta}$ είναι ένα $(p+1) \times 1$ διάνυσμα των συντελεστών παλινδρόμησης, $\boldsymbol{\epsilon}_i$ είναι ένα $n_i \times 1$ διάνυσμα θορύβου και ο πίνακας $\mathbf{X}_i$ είναι ο $n_i \times (p+1)$ πίνακας παλινδρόμησης Vandermonde:

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_{i1} & x_{i1}^2 & \dots & x_{i1}^p \\ 1 & x_{i2} & x_{i2}^2 & \dots & x_{i2}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i} & x_{in_i}^2 & \dots & x_{in_i}^p \end{bmatrix}$$



ΠΕΡΙΛΗΨΗ

Ιωάννα Φλεντζούρη του Δημητρίου και της Παρασκευής. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούλιος, 2009. Ανίχνευση κινούμενων αντικειμένων σε video με μικτά μοντέλα παλινδρόμησης. Επιβλέπωντας: Κωνσταντίνος Μπλέκας.

Στην παρούσα εργασία παρουσιάζεται μια νέα προσέγγιση ανίχνευσης αντικειμένων σε ακολουθίες εικόνων (βίντεο) με τη χρήση μεθόδων ομαδοποίησης χρονοσειρών. Θεωρώντας ότι το κάθε αντικείμενο απαρτίζεται από ένα σύνολο από pixels, μπορούμε, χρησιμοποιώντας την οπτική ροή μεταξύ όλων των ζευγών εικονοπλαισίων (frames), να συνθέσουμε τις τελικές τροχιές κίνησης για κάθε ένα από τα pixels. Οι τροχιές αυτές μπορούν να αναπαρασταθούν ως χρονοσειρές, δίνοντας έτσι τη δυνατότητα της περιγραφής των αντικειμένων ως ομάδες τροχιών κίνησης με κοινές ιδιότητες. Για την αντιμετώπιση του μεγάλου όγκου χρονοσειρών, καθώς μέσω της οπτικής ροής υπολογίζονται διανύσματα κίνησης για το σύνολο των pixels σε ένα frame, χρησιμοποιήσαμε ένα κατώφλι με βάση το οποίο αφαιρούνται οι χρονοσειρές εκείνες που ανήκουν στο παρασκήνιο, επιτρέποντας έτσι την έμμεση εξαγωγή του παρασκηνίου. Στη συνέχεια, η μεθοδολογία που προτείνουμε για την ομαδοποίηση των εναπομεινάντων χρονοσειρών βασίζεται σε μικτά πολυωνυμικά μοντέλα παλινδρόμησης, όπου για την εκτίμηση των παραμέτρων του μοντέλου χρησιμοποιείται ο αλγόριθμος Εκτίμησης Αναμενόμενης Τιμής – Μεγιστοποίησης (*Expectation – Maximization, EM*). Επίσης, προτείνεται και μια βελτίωση του απλού πολυωνυμικού μοντέλου, εισάγοντας σε αυτό και την έννοια της μετατόπισης των χρονοσειρών στο χώρο. Η πειραματική μελέτη έγινε σε δύο τύπους δεδομένων, σε τεχνητά και σε πραγματικά βίντεο. Τα αποτελέσματα είναι αρκετά ικανοποιητικά, καθώς δείχνουν την αποτελεσματικότητα της παράλληλης ανίχνευσης τόσο των κινούμενων αντικειμένων, όσο και των τροχιών τους.



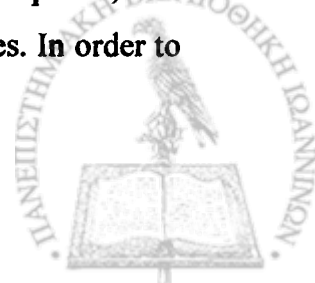
EXTENDED ABSTRACT IN ENGLISH

Flentzouri, Ioanna. MSc, Computer Science Department, University of Ioannina, Greece. July, 2009. Object tracking in image sequences using regression mixture models. Thesis Supervisor: Konstantinos Blekas.

Clustering is an important and challenging problem in many scientific areas. It is typically used as a tool for understanding and exploring large data sets. An enormous number of methodologies has been proposed to address the problem of clustering. These methodologies can be divided in two large categories, depending on the type of data sets in which they focus on. The first category operates on static data sets, that are data sets for which time does not affect their values, whereas the second category operates on dynamic data sets, that are data sets for which time is crucial, because they depend on it. An example of dynamic data sets is time-series.

A special category of time-series is the motion of an object throughout a sequence of frames. Object tracking has recently got a considerable interest to researchers, due to the rapid technological achievements that persistently force for the development of efficient video analysis systems. The problem of tracking can be stated as determining the appearance and location of a particular object in the sequence of frames. Many methodologies have been proposed to address the problem of object tracking. Among those, the most well known are: frame differencing, background subtraction and optical flow.

In this work, we propose a method that combines the two above stated problems, that is the problem of object tracking with the problem of clustering time-series. More specifically, each object in a sequence of frames can be seen as a set of pixels, that each of them follows a specific track throughout the sequence of frames. In order to



extract the track for every pixel we use the methodology of optical flow, with which we compute the motion vectors. So, the overall track of every pixel is just the composition of those motion vectors. This set of tracks consists the data set, that is the time-series set that we want to cluster.

To cope with the bulk of time-series that are extracted as we described above, we use a threshold so as to remove those time-series that are actually background pixels, leaving only those time-series that represent the objects' pixels.

The proposed clustering methodology is based on probabilistic mixture model, and specifically on polynomial regression mixture models, because it is one of the best methods to approach time-series data. In order to estimate the model parameters the Expectation – Maximization (EM) algorithm is used. As an extension of the simple polynomial regression model we propose a new methodology that incorporates the notion of space translation of time series, which can be a consequence of misalignment due to noise. Finally, the performance of the proposed methodologies was evaluated in two types of data sets, artificial (drawing sequences of frames) and real videos.



ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Σκοπός της εργασίας

1.2 Διάρθρωση κεφαλαίων

1.1. Σκοπός της εργασίας

Η αποτελεσματική ομαδοποίηση (*clustering*) δεδομένων, με την ευρεία έννοια, αποτελεί ένα από τα πιο σημαντικά προβλήματα, που έχει απασχολήσει και συνεχίζει να απασχολεί την ερευνητική κοινότητα. Πληθώρα μεθόδων έχει προταθεί ούτως ώστε να αντιμετωπιστεί το πρόβλημα αυτό. Αναφορικά με τον τύπο δεδομένων, οι μέθοδοι αυτές μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες: μέθοδοι που επεξεργάζονται στατικά δεδομένα, δηλαδή δεδομένα που δεν επηρεάζονται από τον παράγοντα χρόνο, και μέθοδοι που επεξεργάζονται δυναμικά δεδομένα, δηλαδή δεδομένα τα οποία αλλάζουν με την πάροδο του χρόνου, όπως είναι οι χρονοσειρές.

Μια ειδική κατηγορία χρονοσειρών αποτελούν και οι τροχιές κίνησης αντικειμένων σε ένα βίντεο. Η ανίχνευση της κίνησης των αντικειμένων (*tracking*) λαμβάνει αυξανόμενο ενδιαφέρον τα τελευταία χρόνια, καθώς οι ραγδαίοι ρυθμοί των τεχνολογικών επιτευγμάτων αυξάνουν συνεχώς τις προκλήσεις ανάπτυξης αποτελεσματικών συστημάτων ανάλυσης βίντεο. Το πρόβλημα της ανίχνευσης μπορεί να οριστεί απλά ως ο προσδιορισμός της εμφάνισης και της θέσης ενός συγκεκριμένου αντικειμένου σε μια ακολουθία εικόνων (*frames*). Διάφορες προσεγγίσεις έχουν προταθεί κατά καιρούς για την ανίχνευση της κίνησης αντικειμένων, με κυριότερες εκείνες που βασίζονται στην διαφορά μεταξύ των *frames*, στην αφαίρεση του παρασκηνίου και στην οπτική ροή.



Στην παρούσα εργασία γίνεται μια προσπάθεια αναγωγής του προβλήματος της ανίχνευσης κίνησης σε πρόβλημα ομαδοποίησης χρονοσειρών, θεωρώντας ότι το βίντεο προέρχεται από σταθερή κάμερα. Πιο συγκεκριμένα, το κάθε αντικείμενο σε μια ακολουθία από frames μπορεί να θεωρηθεί ως ένα σύνολο από pixels, που το καθένα από αυτά ακολουθεί μια συγκεκριμένη τροχιά στη ακολουθία των frames που συγκροτούν το βίντεο. Για την εξαγωγή των τροχιών κίνησης χρησιμοποιήσαμε τη μεθοδολογία της οπτικής ροής, με τη βοήθεια της οποίας υπολογίσαμε τα διανύσματα κίνησης για κάθε ζεύγος frames και στη συνέχεια από τα επιμέρους διανύσματα κίνησης συνθέσαμε τη συνολική τροχιά κίνησης για κάθε pixel. Το σύνολο των τροχιών κίνησης αποτελεί το σύνολο δεδομένων, δηλαδή χρονοσειρών, προς ομαδοποίηση.

Για την αντιμετώπιση του μεγάλου πλήθους χρονοσειρών που εξάγονται, όπως περιγράφηκε παραπάνω, που ουσιαστικά οφείλεται στο γεγονός ότι υπολογίζονται χρονοσειρές και για τα pixels τα οποία ανήκουν στο παρασκήνιο, χρησιμοποιήσαμε ένα κατώφλι με βάση το οποίο αφαιρούνται, ως ένα βαθμό, οι χρονοσειρές εκείνες που ανήκουν στο παρασκήνιο, αφήνοντας ως τελικές χρονοσειρές προς ομαδοποίηση εκείνες που ουσιαστικά ανήκουν στα pixels των αντικειμένων.

Όσον αφορά την ομαδοποίηση των τελικών χρονοσειρών, επιλέξαμε να χρησιμοποιήσουμε ένα μικτό μοντέλο κατανομών, καθώς τέτοια μοντέλα μπορούν να προσεγγίσουν αποδοτικότερα την διαφορετικότητα των υποκειμένων ομάδων. Πιο συγκεκριμένα, προσεγγίσαμε το σύνολο των χρονοσειρών με ένα μικτό πολωνυμικό μοντέλο παλινδρόμησης, εφόσον μπορεί να θεωρηθεί ως το πλέον κατάλληλο μοντέλο για την προσέγγιση χρονοσειρών. Για την εκτίμηση των παραμέτρων του μοντέλου χρησιμοποιήθηκε ο αλγόριθμος Εκτίμησης Αναμενόμενης Τιμής – Μεγιστοποίησης (*Expectation – Maximization, EM*). Στη συνέχεια, επεκτείναμε το απλό πολωνυμικό μοντέλο εισάγοντας σε αυτό και την έννοια της μετατόπισης των χρονοσειρών, η οποία μπορεί να προκληθεί από διάφορους εξωτερικούς παράγοντες, όπως για παράδειγμα ο θόρυβος. Τέλος, η απόδοση των προτεινόμενων μοντέλων εκτιμήθηκε εφαρμόζοντάς τα σε δύο τύπους δεδομένων, σε τεχνητά και σε πραγματικά βίντεο.



1.2. Διάρθρωση των κεφαλαίων

Η συνέχεια της εργασίας δομείται ως εξής. Στο Κεφάλαιο 2 γίνεται μια επισκόπηση του προβλήματος της ομαδοποίησης, όπου παρουσιάζονται οι σημαντικότερες προσεγγίσεις για τους δύο τύπους δεδομένων, δηλαδή στατικών και δυναμικών. Επίσης στο κεφάλαιο αυτό παρουσιάζεται και το πρόβλημα της ανίχνευσης της κίνησης αντικειμένων, καθώς επίσης και κάποιες από τις πιο γνωστές προσεγγίσεις αναφορικά με το πρόβλημα αυτό. Στο Κεφάλαιο 3 παρουσιάζεται το πρόβλημα της ομαδοποίησης χρονοσειρών με μικτά μοντέλα παλινδρόμησης κα κατ' επέκταση, με πολυωνμικά μοντέλα παλινδρόμησης. Ακόμη, παρουσιάζονται και δύο παραδείγματα της εφαρμογής της μεθοδολογίας αυτής σε χρονοσειρές μιας διάστασης και δύο διαστάσεων. Στο Κεφάλαιο 4 παρουσιάζεται η επέκταση του απλού πολυωνμικού μοντέλο, ενσωματώνοντας σε αυτό και την έννοια της μετατόπισης χρονοσειρών. Επίσης παρουσιάζεται και ένα παράδειγμα της εφαρμογής της μεθοδολογίας αυτής σε χρονοσειρές μιας διάστασης. Στο Κεφάλαιο 5 παρουσιάζονται τα αποτελέσματα από την πειραματική μελέτη, με βάση δύο τύπους δεδομένων, τεχνητά και πραγματικά βίντεο. Τέλος, στο Κεφάλαιο 6 εκτίθενται τα συμπεράσματα της εργασίας αυτής, καθώς επίσης και μελλοντικές επεκτάσεις που μπορούν να γίνουν.



ΚΕΦΑΛΑΙΟ 2. ΕΠΙΣΚΟΠΗΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΟΜΑΔΟΠΟΙΗΣΗΣ (CLUSTERING)

2.1 Τι είναι η ομαδοποίηση;

2.2 Βασικές τεχνικές ομαδοποίησης στατικών δεδομένων

2.3 Βασικές τεχνικές ομαδοποίησης χρονοσειρών

2.4 Βασικές τεχνικές ανίχνευσης κίνησης σε βίντεο

2.1. Τι είναι η ομαδοποίηση;

Με τον όρο ομαδοποίηση αναφερόμαστε στη διαδικασία εκείνη κατά την οποία προσπαθούμε να εντοπίσουμε εσωτερικές δομές σε ένα σύνολο δεδομένων, αγνώστου κατηγορίας, οργανώνοντας τα δεδομένα αυτά σε ομογενείς ομάδες, έτσι ώστε η ομοιότητα των δεδομένων που ανήκουν στην ίδια ομάδα να είναι μεγάλη, ενώ αντίθετα, η ομοιότητα δεδομένων που ανήκουν σε διαφορετικές ομάδες να είναι μικρή.

Η ομαδοποίηση αποτελεί ένα ερευνητικό πρόβλημα το οποίο χρησιμοποιείται σε πολλούς τομείς. Για παράδειγμα, στην βιολογία η ομαδοποίηση χρησιμοποιείται στην ανάλυση των ακολουθιών του γονιδιώματος ή στην κατηγοριοποίηση κάποιου αγνώστου είδους σε μια γνωστή ομάδα, με βάση κάποια χαρακτηριστικά. Στην ιατρική, η ομαδοποίηση χρησιμοποιείται στην επεξεργασία ιατρικών εικόνων, για παράδειγμα σε ένα PET scan για την διαφοροποίηση των ιστών από το αίμα. Στην έρευνα αγοράς, η ομαδοποίηση χρησιμοποιείται για την τμηματοποίηση της αγοράς σε ομάδες στόχους ή για την αντιστοίχιση ενός νέου προϊόντος σε μια συγκεκριμένη



ομάδα. Στη γεωλογία, για την ομαδοποίηση των επικέντρων σεισμικών δονήσεων για την εύρεση επικίνδυνων ζωνών.

Γενικά, μπορούμε να πούμε ότι η ομαδοποίηση χρησιμοποιείται για την περιγραφή πολύπλοκων δεδομένων με έναν τρόπο περισσότερο κατανοητό και χρήσιμο. Τα δεδομένα προς ομαδοποίηση μπορεί να είναι στατικά, εάν οι τιμές των χαρακτηριστικών του δεν αλλάζουν με την πάροδο του χρόνου, ή δυναμικά, εάν οι τιμές αλλάζουν. Η πλειονότητα των τεχνικών ομαδοποίησης που έχουν αναπτυχθεί αναφέρεται σε στατικά δεδομένα. Τα τελευταία όμως χρόνια παρατηρείται αύξηση του ενδιαφέροντος σχετικά με τις τεχνικές ομαδοποίησης δυναμικών δεδομένων ή χρονοσειρών, όπως συχνά αποκαλούνται.

2.2. Βασικές τεχνικές ομαδοποίησης στατικών δεδομένων

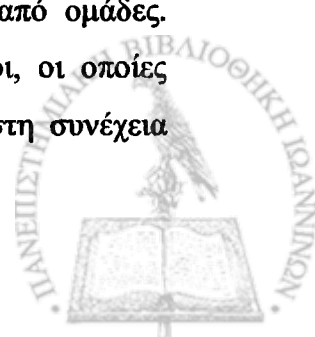
Υπάρχει μια πληθώρα μεθόδων και τεχνικών που έχουν αναπτυχθεί για την αντιμετώπιση του προβλήματος της ομαδοποίησης, όσον αφορά τα στατικά δεδομένα. Μια ενδιαφέρουσα κατηγοριοποίηση των διαθέσιμων τεχνικών έχουν κάνει οι Han και Kamber [14], χωρίζοντάς τες σε έξι μεγάλες κατηγορίες, οι οποίες παρουσιάζονται εν συντομία στη συνέχεια.

Μέθοδοι διαμέρισης (*partitioning methods*)

Δοθέντος ενός συνόλου n αντικειμένων, αγνώστου κατηγορίας, μια μέθοδος διαμέρισης κατασκευάζει k διαμερίσεις του συγκεκριμένου συνόλου, όπου κάθε διαμέριση αναπαριστά μια ομάδα που περιέχει τουλάχιστον ένα αντικείμενο. Μια διαμέριση είναι σαφής (*crisp*) αν κάθε αντικείμενο ανήκει σε ακριβώς μια ομάδα, ή ασαφής (*fuzzy*) αν ανήκει σε περισσότερες από μια ομάδα. Δύο πολύ γνωστές μέθοδοι σαφούς διαμέρισης είναι ο αλγόριθμος k -means και k -medoids [31]. Ενώ ασαφούς διαμέρισης είναι ο αλγόριθμος *fuzzy c-means* και *fuzzy c-medoids* [31].

Ιεραρχικές Μέθοδοι (*hierarchical methods*)

Οι ιεραρχικές μέθοδοι ομαδοποιούν τα αντικείμενα σε ένα δέντρο από ομάδες. Υπάρχουν δύο τύποι ιεραρχικών μεθόδων: οι *agglomerative* μέθοδοι, οι οποίες ξεκινάνε τοποθετώντας κάθε αντικείμενο στη δικιά του ομάδα και στη συνέχεια



συγχωνεύουν τις μικρές αυτές ομάδες σε μεγαλύτερες μέχρι να καταλήξουν σε μια μόνο μεγάλη ομάδα ή μέχρι να φτάσουν στο επιθυμητό αριθμό ομάδων και οι divisive μέθοδοι, οι οποίες κάνουν ακριβώς το αντίθετο, δηλαδή ξεκινάνε από μια μεγάλη ομάδα και καταλήγουν σε μικρότερες. Οι πιο γνωστοί αλγόριθμοι της κατηγορίας αυτής είναι οι CURE και BIRCH [31].

Μέθοδοι βασισμένες στην πυκνότητα (density-based methods)

Οι μέθοδοι αυτές προσπαθούν να δημιουργήσουν ομάδες, βασιζόμενες στην πυκνότητα των σημείων των δεδομένων σε μια περιοχή, δηλαδή ομάδες αποτελούν οι πυκνές περιοχές του χώρου των δεδομένων. Χαρακτηριστική μέθοδος στην κατηγορία αυτή είναι ο αλγόριθμος DBSCAN [31].

Μέθοδοι βασισμένες σε πλέγμα (grid-based methods)

Αυτές οι μέθοδοι χωρίζουν το χώρο των δεδομένων σε ένα πεπερασμένο αριθμό κελιών, τα οποία σχηματίζουν ένα πλέγμα. Τα κελιά τα οποία περιέχουν αριθμό αντικειμένων μεγαλύτερο από ένα κατώφλι, θεωρούνται πυκνά και στη συνέχεια τα συγκεκριμένα κελιά ενώνονται ούτως ώστε να δημιουργήσουν ομάδες. Γνωστοί αλγόριθμοι σε αυτή την κατηγορία είναι οι STING και CLIQUE [31].

Μέθοδοι βασισμένες σε κάποιο μοντέλο (model-based methods)

Σε αυτή την κατηγορία ανήκουν οι μέθοδοι οι οποίες υποθέτουν ένα μοντέλο για κάθε μια από τις ομάδες και στη συνέχεια προσπαθούν να επιτύχουν την καλύτερη προσαρμογή των δεδομένων στα υποτιθέμενα μοντέλα. Υπάρχουν δύο μεγάλες προσεγγίσεις των μεθόδων αυτών: η στατιστική προσέγγιση και η προσέγγιση που βασίζεται σε νευρωνικά δίκτυα. Η πιο διαδεδομένη είναι η πρώτη προσέγγιση, επέκταση της οποίας αποτελεί η αμέσως επόμενη κατηγορία.

Μέθοδοι βασισμένες σε πεπερασμένα μικτά μοντέλα κατανομών

Τα πεπερασμένα μικτά μοντέλα αποτελούν μια ημι-παραμετρική μορφή της εκτίμησης πιθανότητας, συμφωνά με την οποία υποθέτουμε ότι το σύνολο των δεδομένων δεν περιγράφεται από μια μόνη κατανομή, αλλά κάθε υποπεριοχή των δεδομένων μπορεί να περιγραφεί με μια ατομική συνάρτηση πυκνότητας πιθανότητας (PDF), η οποία χαρακτηρίζεται από ένα διάνυσμα παραμέτρων θ_k .



Τότε, η συνολική πιθανότητα θα αποτελείται από έναν συνδυασμό όλων των επιμέρους PDFs.

Στη βασική μορφή, μπορούμε να περιγράψουμε το δεδομένο y_i με την ακόλουθη μικτή πυκνότητα:

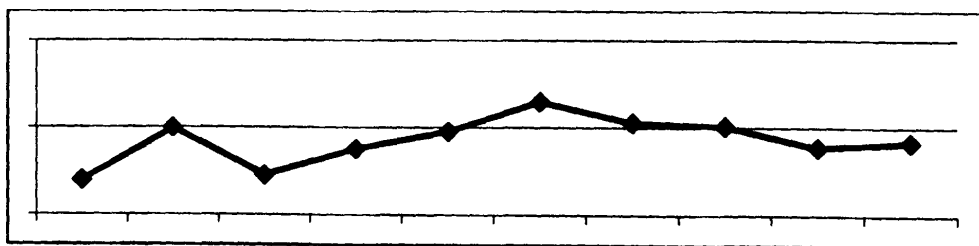
$$p(y_i|\Theta) = \sum_k^K a_k p_k(y_i|\theta_k)$$

όπου, a_k είναι το βάρος της k -οστής ομάδας και p_k είναι η πυκνότητα της k -οστής ομάδας με διάνυσμα παραμέτρων θ_k . Τα μικτά βάρη a_k έχουν άθροισμα ένα και είναι θετικά. Οι συναρτήσεις πυκνότητας $p_k(\cdot)$ περιγράφουν συγκεκριμένες υποπεριοχές (ομάδες) της πυκνότητας, ενώ η μικτή πιθανότητα $p(\cdot)$ εκφράζει τη συνολική πιθανότητα για το δεδομένο y_i .

Στη συνέχεια θα πρέπει να γίνει η προσαρμογή του συνόλου παραμέτρων Θ της κατανομής, ούτως ώστε η παραχθείσα πυκνότητα να “ταιριάζει” με την πραγματική κατανομή των δεδομένων. (Στο Παράρτημα Β περιγράφεται η γενική μεθοδολογία των μικτών μοντέλων κατανομής)

2.3. Βασικές τεχνικές ομαδοποίησης χρονοσειρών

Σε αντίθεση με τα στατικά δεδομένα, οι χρονοσειρές αποτελούν ακολουθίες παρατηρήσεων ή μετρήσεων στο χρόνο, για ένα ή περισσότερα χαρακτηριστικά ενός αντικειμένου (όπως φαίνεται και στο παρακάτω σχήμα).



Δοθέντος ενός συνόλου χρονοσειρών, αγνώστου κατηγορίας, απαιτείται πολλές φορές ο προσδιορισμός ομάδων όμοιων χρονοσειρών. Πολλοί αλγόριθμοι έχουν αναπτυχθεί για την αντιμετώπιση αυτού του προβλήματος, που αν τους δούμε από μια γενική σκοπιά, αποτελούν μια προσπάθεια προσαρμογής των υπαρχόντων αλγορίθμων (για στατικά δεδομένα), έτσι ώστε να διαχειριστούν αποτελεσματικά την ιδιαίτερη φύση των χρονοσειρών. Μια ενδιαφέρουσα κατηγοριοποίηση των μεθόδων ομαδοποίησης χρονοσειρών έχει κάνει ο Warren Liao [33], σύμφωνα με τον οποίο οι υπάρχουσες μέθοδοι μπορούν να χωριστούν σε τρεις μεγάλες κατηγορίες, οι οποίες και περιγράφονται στη συνέχεια.

Μέθοδοι για ανεπεξέργαστα δεδομένα

Σε αυτή την κατηγορία ανήκουν οι μέθοδοι που διαχειρίζονται όλο το σύνολο των χρονοσειρών, χωρίς να έχουν κάνει πρώτα κάποια επεξεργασία. Δύο χρονοσειρές προς σύγκριση επιλέγονται από το ίδιο χρονικό διάστημα, χωρίς να είναι απαραίτητα ιδίου μήκους.

Οι κυριότεροι αλγόριθμοι, που ανήκουν στην κατηγορία αυτή, αποτελούν προσαρμογή των αλγορίθμων που διαχειρίζονται στατικά δεδομένα. Οι πιο γνωστές παραλλαγές που υπάρχουν είναι οι εξής: fuzzy c- Means (Golay et al., Policker & Geva), k- Means and fuzzy c-Means (Liao), Agglomerative hierarchical clustering (Kakizawa et al., Kumar et al., Shumway, Van Wijk & Van Selow). Οι παραπάνω τεχνικές έχουν χρησιμοποιηθεί σε διάφορες εφαρμογές, όπως για παράδειγμα MRI εγκεφάλου, σεισμοί, προσομοίωση πεδίου μάχης, εποχιακά μοντέλα κ.τ.λ.

Μέθοδοι βασισμένες σε χαρακτηριστικά

Η ομαδοποίηση ανεπεξέργαστων χρονοσειρών, συνεπάγεται την αντιμετώπιση δεδομένων μεγάλης διάστασης, γεγονός το οποίο δυσχεραίνει την όλη διαδικασία. Για το λόγο αυτό, αναπτύχθηκαν πολλές μέθοδοι που προσπαθούν να ομαδοποιήσουν τις χρονοσειρές, με βάση κάποια από τα χαρακτηριστικά τους. Για παράδειγμα, οι Goutte et al. χρησιμοποίησαν, για την ομαδοποίηση χρονοσειρών fMRI, παραλλαγή του αλγορίθμου k-Means και hierarchical clustering, παίρνοντας όμως σαν είσοδο όχι όλο το σύνολο των χρονοσειρών αλλά μόνο εκείνες που πληρούσαν μια συγκεκριμένη συνθήκη, με βάση κάποια συνάρτηση



ετεροσυσχέτισης (cross-correlation function). Οι Fu et al. χρησιμοποίησαν την έννοια της χαρτογράφησης των χρονοσειρών, όπου με τη βοήθεια ενός συνεχώς κυλιόμενου παραθύρου προσπάθησαν να βρουν χρονικά μοτίβα διάσπαρτα στο σύνολο των χρονοσειρών και στη συνέχεια εφάρμοσαν πάνω σε αυτά τα μοτίβα μια παραλλαγή του k-Means για την ομαδοποίησή τους. Τέλος, οι Vlachos et al. χρησιμοποίησαν το μετασχηματισμό κυματομορφών Haar για να αποσυνθέσουν τις χρονοσειρές σε απλούστερες, ξεκινώντας πρώτα από χαμηλή ανάλυση και καταλήγοντας σε υψηλότερη, και στη συνέχεια εφάρμοσαν τον αλγόριθμο k-Means για την ομαδοποίησή τους.

Μέθοδοι βασισμένες σε κάποιο μοντέλο

Στην κατηγορία αυτή ανήκουν οι μέθοδοι που θεωρούν ότι κάθε χρονοσειρά έχει παραχθεί από ένα πιθανοτικό μοντέλο. Οι περισσότερες τεχνικές που έχουν αναπτυχθεί βασίζονται σε μοντέλα παλινδρόμησης (regression models), εκ των οποίων τα μοντέλα AR (Autoregressive models) και ARMA (Autoregressive Moving Average models) είναι τα πιο γνωστά. Τα μοντέλα αυτά ορίζουν μια μεταβλητή (εξαρτημένη) ως συνάρτηση κάποιων άλλων ανεξάρτητων μεταβλητών. Αυτό που αλλάζει κάθε φορά σε κάθε μέθοδο είναι η μετρική η οποία θα χρησιμοποιηθεί, ούτως ώστε να προσδιοριστεί η ομοιότητα μεταξύ των χρονοσειρών. Για παράδειγμα, ο Piccolo χρησιμοποιεί την Ευκλείδεια απόσταση μεταξύ των αυτοπαλινδρούμενων επεκτάσεων των χρονοσειρών. Ο Baragona χρησιμοποιεί την αυτοσυσχέτιση των καταλοίπων του μοντέλου ως μετρική. Ο Kalpakis et al., χρησιμοποιεί την Ευκλείδεια απόσταση μεταξύ της γραμμικής προγνωστικής κωδικοποίησης cepstra (Linear Predictive Coding cepstra) δύο χρονοσειρών, για να προσδιορίσει την ομοιότητά τους.

Μια άλλη τεχνική που έχει χρησιμοποιηθεί αρκετά (Ramoni et al., Li et al., Wang et al.) βασίζεται στην υπόθεση ότι οι χρονοσειρές έχουν ιδιότητες παρόμοιες με εκείνες των αλυσίδων Markov, μετασχηματίζοντας έτσι το πρόβλημα της ομαδοποίησης χρονοσειρών σε πρόβλημα ομαδοποίησης αλυσίδων Markov.

Μια επίσης γνωστή τεχνική είναι και η αναπαράσταση των ομάδων χρονοσειρών με μια Gaussian κατανομή εκτιμώντας στη συνέχεια τις παραμέτρους της κατανομής.



όπως για παράδειγμα ο Biernacki et al., ο οποίος χρησιμοποίησε τον αλγόριθμο E(xpectation) - M(aximization) και την πιθανοφάνεια ως κριτήριο, ενώ οι Tran & Wagner χρησιμοποίησαν μια παραλλαγή του αλγορίθμου fuzzy c-means.

2.4. Βασικές τεχνικές ανίχνευσης κίνησης σε βίντεο

Τα τελευταία χρόνια παρατηρείται ένα αυξημένο ενδιαφέρον σχετικά με την ανίχνευση (*tracking*) κινούμενων αντικειμένων σε ένα βίντεο. Με απλά λόγια, η ανίχνευση ενός κινούμενου αντικειμένου ισοδυναμεί με την εκτίμηση της τροχιάς τους μέσα από ένα σύνολο διαδοχικών εικόνων, οι οποίες συνθέτουν το τελικό βίντεο. Το πρόβλημα της ομαδοποίησης σχετίζεται άμεσα με το πρόβλημα της ανίχνευσης κίνησης, καθώς τα κινούμενα αντικείμενα μπορεί να θεωρηθούν ως περιοχές όμοιων χαρακτηριστικών και ως τέτοιες να αποτελούν ομάδα.

Πολλές τεχνικές έχουν αναπτυχθεί για την αποτελεσματική ανίχνευση της κίνησης αντικειμένων. Στη συνέχεια παρουσιάζονται συνοπτικά μερικές από τις πιο γνωστές τεχνικές.

Ανίχνευση με την αφαίρεση παρασκήνιου (background)

Η πιο διαδεδομένη τεχνική ανίχνευσης είναι η αφαίρεση του παρασκήνιου. Σύμφωνα με την τεχνική αυτή κάθε pixel χαρακτηρίζεται ως pixel παρασκήνιου ή ως pixel προσκήνιου, με βάση κάποιους κανόνες, και στη συνέχεια, με τη χρήση διαφόρων μεθόδων, γίνεται η ομαδοποίηση των pixels προσκήνιου, που ουσιαστικά είναι εκείνα από τα οποία απαρτίζονται τα αντικείμενα.

Ο Haritaoglu et al. [15] για να επιτύχει τον παραπάνω διαχωρισμό, χρησιμοποιεί ένα μοντέλο για το παρασκήνιο όταν είναι κενό, δηλαδή όταν δεν υπάρχουν κινούμενα αντικείμενα. Συμφωνά με το μοντέλο αυτό κάθε pixel του παρασκήνιου αναπαριστάται από τρεις τιμές: από την χαμηλότερη (M) και υψηλότερη τιμή (N) της έντασης και από την υψηλότερη τιμή διαφοράς έντασης (D) μεταξύ δύο διαδοχικών εικόνων. Στη συνέχεια, όταν στην εικόνα I θα υπάρχει και κάποιο αντικείμενο εκτός από το παρασκήνιο, για να χαρακτηριστεί ένα pixel x ως pixel



προσκηνίου θα πρέπει να ισχύει: $|M(x) - I(x)| > D(x)$ ή $|N(x) - I(x)| > D(x)$. Τέλος, αφού γίνει η εκκαθάριση των pixel προσκηνίου από τον θόρυβο, γίνεται η τμηματοποίηση τους και κάθε τμήμα συγκρίνεται με κάποια υπάρχοντα σχήματα για να προσδιοριστούν διάφορα μέρη του σώματος (όπως κεφάλι, σώμα, χέρια κ.τ.λ.).

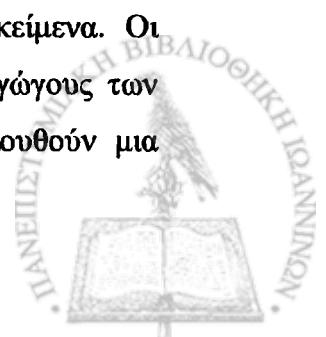
Παρόμοια διαδικασία ακολουθούν και οι Cucchiara et al. [2], οι οποίοι χρησιμοποιούν και εκείνοι ένα μοντέλο για το παρασκήνιο όταν αυτό είναι κενό και στη συνέχεια αφού χαρακτηριστεί κάθε pixel ως προσκηνίου ή παρασκηνίου, γίνεται οι συγχώνευση των pixel προσκηνίου σε ομάδες με βάση κάποια μετρική απόστασης. Άλλες τεχνικές οι οποίες βασίζονται στην αφαίρεση του παρασκηνίου παρουσιάζουν και οι McKenna et al. [22], Seki et al. [27], Jabri et al. [17].

Ανίχνευση με τη χρήση στατιστικών μοντέλων

Μια επίσης γνωστή τεχνική για την ανίχνευση κινούμενων αντικειμένων είναι και η χρησιμοποίηση στατιστικών μοντέλων για την αναπαράσταση της εικόνας. Η υπόθεση που γίνεται είναι η εξής: κάθε pixel είναι μια τυχαία μεταβλητή με μια συγκεκριμένη κατανομή πιθανότητας. Οι Wren et al. [34] θεωρούν ότι τα κινούμενα αντικείμενα αποτελούν ένα σύνολο από Gaussian κατανομές και στη συνέχεια, ως μέτρηση πιθανοφάνειας για κάθε pixel χρησιμοποιείται η απόσταση Mahalanobis από κάθε ομάδα. Παρόμοια τεχνική, χρησιμοποιώντας Gaussian κατανομές, ακολουθούν και οι Stauffer et al. [30] και οι McKenna et al. [21].

Ανίχνευση βασισμένη στη χωρο-χρονική (spatio-temporal) ανάλυση των frames

Μια άλλη κατηγορία τεχνικών ανίχνευσης, βασίζεται στη χωρο-χρονική ανάλυση των frames του βίντεο. Οι τεχνικές αυτές προσπαθούν να ανιχνεύσουν τα αντικείμενα λαμβάνοντας υπόψη όχι μόνο τη χρονική εξέλιξη (από frame σε frame) της έντασης και του χρώματος του κάθε pixel, αλλά επίσης και τις χωρικές του ιδιότητες, δηλαδή λαμβάνουν υπόψη και την εξέλιξη των γειτονικών pixels. Οι Ma & Zhang [35] υπολογίζοντας τις διακυμάνσεις των χρωμάτων μεταξύ διαδοχικών frames δημιούργησαν μια χωρο-χρονική εικόνα, με βάση την εντροπία και στη συνέχεια χρησιμοποιώντας μορφολογική ανάλυση εντόπιζαν τα αντικείμενα. Οι Souvenir et al. [29], χρησιμοποίησαν τις χωρικές και χρονικές παραγώγους των pixels και στη συνέχεια υποθέτοντας ότι οι παράγωγοι αυτοί ακολουθούν μια



Gaussian κατανομή, γινόταν η ανίχνευση των αντικειμένων χρησιμοποιώντας την απόσταση Mahalanobis.

Ανίχνευση βασισμένη στην οπτική ροή (optical flow)

Εκτός από τις προηγούμενες κατηγορίες, αρκετοί επιστήμονες έχουν ασχοληθεί με την ανίχνευση κινούμενων αντικειμένων χρησιμοποιώντας την οπτική ροή μεταξύ των frames του βίντεο (Jeongho et al. [18], Okada et al. [25], Kalafatic et al. [19]).

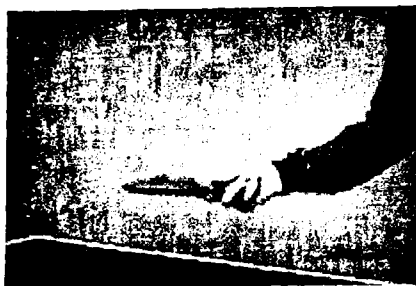
Κάθε pixel σε μια εικόνα (frame) χαρακτηρίζεται από μια τιμή που αντιστοιχεί στη φωτεινότητά του. Ένα αντικείμενο αποτελείται από ένα σύνολο pixels. Οπότε, για να βρούμε την κίνηση του αντικειμένου, μεταξύ δύο διαδοχικών frames, αρκεί να βρούμε την κίνηση του κάθε pixel που απαρτίζει το αντικείμενο αυτό.

Η κίνηση αυτή αναπαριστάται από ένα δισδιάστατο διάνυσμα, το οποίο ονομάζεται *διάνυσμα κίνησης*, όπου η πρώτη διάσταση αντιστοιχεί στην κίνηση κατά τη φορά του οριζοντίου άξονα (x) της εικόνας, ενώ η δεύτερη κατά τη φορά του κάθετου άξονα (y).

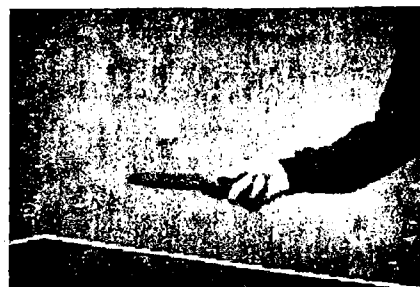
Το διάνυσμα αυτό αναπαριστά τη μετατόπιση του pixel σε ένα πολύ μικρό χρονικό διάστημα. Για να υπολογιστεί η ταχύτητα κίνησης, αρκεί να διαιρέσουμε το διάνυσμα κίνησης με το χρονικό διάστημα. (Στο Παράρτημα C περιγράφεται αναλυτικά η μεθοδολογία υπολογισμού της οπτικής ροής).

Ένα παράδειγμα αναπαράστασης των διανυσμάτων κίνησης ανάμεσα σε δύο διαδοχικά frames φαίνεται και στην παρακάτω εικόνα.

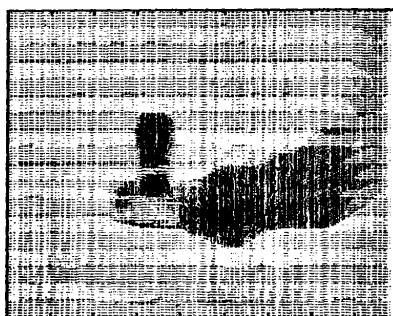




frame 1



frame 2



Διανύσματα κίνησης

Σχήμα 2.1 Αναπαράσταση διανυσμάτων κίνησης

Στην παρούσα εργασία χρησιμοποιείται η μεθοδολογία υπολογισμού των διανυσμάτων κίνησης των pixels μεταξύ όλων των frames ενός βίντεο, με βάση τα οποία συντίθενται στη συνέχεια οι χρονοσειρές κίνησης, δηλαδή οι τροχιές που ακολουθούν τα pixels στη διάρκεια του βίντεο. Οι χρονοσειρές αυτές αποτελούν είσοδο των δύο μεθοδολογιών ομαδοποίησης που περιγράφονται στα επόμενα δύο κεφάλαια.

ΚΕΦΑΛΑΙΟ 3. ΟΜΑΔΟΠΟΙΗΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕ ΜΙΚΤΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

3.1 Εισαγωγή

3.2 Προηγούμενες Μελέτες

3.3 Πολυωνυμικό μικτό μοντέλο παλινδρόμησης

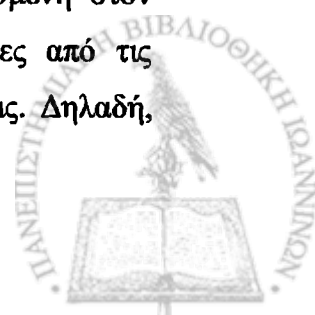
3.4 Άλλα μοντέλα παλινδρόμησης

3.1. Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζουμε ένα μικτό μοντέλο ομαδοποίησης χρονοσειρών, το οποίο βασίζεται στο πολυωνυμικό μοντέλο παλινδρόμησης. Πιο συγκεκριμένα, στην ενότητα 3.2 παρουσιάζουμε κάποιες προηγούμενες μελέτες στις οποίες χρησιμοποιούνται μικτά μοντέλα παλινδρόμησης για την αντιμετώπιση του προβλήματος της ομαδοποίησης. Στην ενότητα 3.3 ορίζουμε το πολυωνυμικό μικτό μοντέλο παλινδρόμησης και στη συνέχεια παρουσιάζουμε την εφαρμογή του στο πρόβλημα της ομαδοποίησης χρονοσειρών. Τέλος, στην ενότητα 3.4 παρουσιάζουμε άλλα δύο μοντέλα παλινδρόμησης, τα οποία μπορούν να χρησιμοποιηθούν στο συγκεκριμένο πρόβλημα ομαδοποίησης.

3.2. Προηγούμενες Μελέτες

Η ομαδοποίηση δεδομένων βασισμένη σε μοντέλα παλινδρόμησης έχει απασχολήσει από πολύ νωρίς τους επιστήμονες, ξεκινώντας από το απλό μοντέλο, στο οποίο υπάρχουν μόνο δύο ομάδες, μέχρι και τη γενική μεθοδολογία, βασισμένη στον αλγόριθμο ΕΜ, για την περίπτωση των K ομάδων. Οι περισσότερες από τις προηγούμενες μελέτες επικεντρώνονταν σε μονομεταβλητές περιπτώσεις. Δηλαδή,



το σύνολο δεδομένων αποτελούταν από μεμονωμένες, μονομεταβλητές παρατηρήσεις, οι οποίες θεωρούνταν ότι είχαν προέλθει από κάποια καμπύλη παλινδρόμησης.

Ένας από τους πρώτους που ασχολήθηκε με τα μοντέλα παλινδρόμησης ήταν ο Quandt(1972) [26], ο οποίος πρότεινε ένα μοντέλο πιθανοφάνειας με δύο συνιστώσες, το λεγόμενο μοντέλο εναλλαγής καταστάσεων (switching regression). Η μεθοδολογία του αυτή ήταν ικανή να βρει τις υποκείμενες ομάδες μεγιστοποιώντας την πιθανοφάνεια, με τη χρήση της μεθόδου των συζυγών παραγώγων.

Παρόμοια μελέτη με τον Quandt έκανε και ο Hosmer(1974) [16], ο οποίος επίσης πρότεινε ένα μικτό μοντέλο παλινδρόμησης με δύο συνιστώσες, αλλά χρησιμοποίησε τη μεγιστοποίηση της πιθανοφάνειας για να εκτιμήσει τις παραμέτρους του μοντέλου με τη χρήση μιας επαναληπτικής διαδικασίας. Τελικά, παρήγαγε έναν αλγόριθμο EM για ένα μοντέλο παλινδρόμησης με δύο ομάδες.

Αργότερα, οι DeSarbo και Cron (1988) [6] παρουσίασαν μια πιο γενική διαδικασία EM για μικτά μοντέλα γραμμικής παλινδρόμησης, με οποιοδήποτε αριθμό ομάδων. Όλες όμως αυτές οι μελέτες ασχολήθηκαν, όπως προαναφέρθηκε, με μονομεταβλητές περιπτώσεις. Μια προσπάθεια να επεκτείνει την παραπάνω διαδικασία και σε πολυμεταβλητά δεδομένα, έκανε ο McLachlan (1992) [23], ο οποίος παρουσίασε ένα μικτό μοντέλο παλινδρόμησης βασισμένο σε διανυσματικά δεδομένα.

Η πληρέστερη μελέτη των μοντέλων παλινδρόμησης, όσον αφορά πολυμεταβλητά δεδομένα, έχει γίνει από τους Gaffney & Smyth (1999) [11], οι οποίοι εστιάζουν στο σύνολο των καμπυλών και όχι απλά στις μεμονωμένες παρατηρήσεις. Επεκτείνοντας τη μελέτη των DeSarbo και Cron (1988) [6], παρουσίασαν ένα πιο γενικό πλαίσιο, στο οποίο ενσωμάτωσαν την έννοια της καμπύλης και της ιδιότητας μέλους μιας καμπύλης σε μια ομάδα.



3.3. Πολυωνυμικό μικτό μοντέλο παλινδρόμησης (polynomial regression mixtures)

Στην ενότητα αυτή περιγράφεται μια μεθοδολογία ομαδοποίησης χρονοσειρών βασισμένη στο πολυωνυμικό μικτό μοντέλο παλινδρόμησης (PRM). Το μοντέλο PRM χρησιμοποιεί για τον ορισμό των PDFs το πολυωνυμικό μοντέλο παλινδρόμησης προσθέτοντας και έναν όρο θορύβου (Gaussian error). Η χρήση αυτών των μοντέλων παλινδρόμησης στο γενικό πλαίσιο ομαδοποίησης καμπυλών, οδηγεί σε αποτελεσματικούς αλγορίθμους EM, με βάση τους οποίους μπορεί να γίνει η ομαδοποίηση.

Στην ενότητα 3.3.1 παρουσιάζεται ο ορισμός του μοντέλου PRM, ενώ στην ενότητα 3.3.2 παρουσιάζεται ο αλγόριθμος EM που βασίζεται στο μοντέλο αυτό.

3.3.1. Ορισμός Μοντέλου

Έστω ότι έχουμε ένα σύνολο Y από n χρονοσειρές $\{y_1, \dots, y_i, \dots, y_n\}$. Κάθε χρονοσειρά έχει μήκος n_i , δηλαδή n_i μετρήσεις στα (χρονικά) σημεία του x_i . Υποθέτουμε ότι υπάρχει μια πολυωνυμική παλινδρομική σχέση βαθμού p , ανάμεσα στο y_i και x_i , συν έναν όρο σφάλματος (Gaussian error) για την απόδοση των εξωγενών επιδράσεων.

Η παλινδρόμηση του y_i στο x_i μπορεί έτσι να οριστεί ως εξής:

$$y_i = X_i \beta + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I) \quad \text{Εξ. 3.1}$$

όπου ο $n_i \times p$ πίνακας παλινδρόμησης X_i είναι ο πίνακας Vandermonde για κάθε x_i , ενώ το β είναι το p -διάνυσμα των συντελεστών παλινδρόμησης. Ο πίνακας Vandermonde βαθμού p εκτιμώμενος στο x_i είναι ο ακόλουθος:

$$X_i = \begin{bmatrix} 1 & x_{i1} & x_{i1}^2 & \dots & x_{i1}^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{in_i} & x_{in_i}^2 & \dots & x_{in_i}^p \end{bmatrix}$$



Η παλινδρομική αυτή εξίσωση, μαζί και με το μοντέλο θορύβου, καθορίζουν την υπό συνθήκη συνάρτηση πυκνότητας πιθανότητας (conditional PDF) του y_i δοθέντος του x_i ως $\mathcal{N}(y_i | \mathbf{X}_i \boldsymbol{\beta}, \sigma^2 \mathbf{I})$, η οποία χαρακτηρίζεται από το διάνυσμα παραμέτρων $\theta = \{\boldsymbol{\beta}, \sigma^2\}$.

Μπορούμε να ενσωματώσουμε την παραπάνω PDF σε ένα μικτό μοντέλο (βλέπε Παράρτημα Β), απλά προσθέτοντας και την εξάρτηση από την ομάδα k . Στη σημειογραφία, η εξάρτηση αυτή δηλώνεται με τη μορφή δεικτών στις παραμέτρους ως $\{\boldsymbol{\beta}_k, \sigma_k^2\}$. Η ενσωμάτωση αυτών των εξαρτημένων PDFs στην υπό συνθήκη μικτή πυκνότητα της Εξίσωσης (Β.1) (Παράρτημα Β) μας δίνει τον ορισμό του μικτού μοντέλου παλινδρόμησης (PRM) ως εξής:

$$\begin{aligned} p(y_i | \mathbf{x}_i, \boldsymbol{\theta}) &= \sum_k^K a_k p_k(y_i | \mathbf{x}_i, \theta_k) \\ &= \sum_k^K a_k \mathcal{N}(y_i | \mathbf{X}_i \boldsymbol{\beta}_k, \sigma_k^2 \mathbf{I}) \end{aligned} \quad \text{Εξ. 3.2}$$

Η συνάρτηση λογαριθμικής πιθανοφάνειας θα είναι τότε το άθροισμα των υπό συνθήκη πυκνοτήτων όλων των n καμπυλών. Δηλαδή:

$$\log p(Y | \mathbf{X}, \boldsymbol{\theta}) = \sum_i \log \sum_k^K a_k p_k(y_i | \mathbf{x}_i, \theta_k) \quad \text{Εξ. 3.3}$$

Με βάση το μοντέλο αυτό μπορούμε να περιγράψουμε τον αλγόριθμο EM για την ομαδοποίηση καμπυλών χρησιμοποιώντας PRMs.

3.3.2. Αλγόριθμος EM για Πολυωνομικά Μικτά Μοντέλα Παλινδρόμησης

Στην ενότητα αυτή περιγράφεται ο αλγόριθμος EM για τα PRMs. (Η βασική θεωρία του αλγορίθμου EM περιγράφεται στο Παράρτημα Α)



Έστω ότι το z_i δηλώνει τη συμμετοχή σε μια ομάδα (cluster membership) για την καμπύλη i και η από κοινού πυκνότητα των y_i και z_i γράφεται ως:

$$\begin{aligned} p(y_i, z_i | x_i) &= a_{z_i} p_{z_i}(y_i | x_i) \\ &= a_{z_i} \mathcal{N}(y_i | X_i \beta_{z_i}, \sigma_{z_i}^2 \mathbf{I}) \end{aligned} \quad \text{Εξ. 3.4}$$

Η συμμετοχή σε μια ομάδα $\{z_i\}$ θεωρείται κρυμμένη πληροφορία. Τότε, η πυκνότητα των κρυμμένων δεδομένων γίνεται η εκ των υστέρων (posterior) πιθανότητα $p(z_i | y_i, x_i)$. Η συνολική συνάρτηση λογαριθμικής πιθανοφάνειας $\mathcal{L}_c$ μπορεί να υπολογιστεί παίρνοντας το άθροισμα, για όλες τις καμπύλες, των λογαρίθμων των από κοινού πυκνοτήτων της Εξίσωσης (3.4):

$$\mathcal{L}_c = \sum_i \log a_{z_i} \mathcal{N}(y_i | X_i \beta_{z_i}, \sigma_{z_i}^2 \mathbf{I}) \quad \text{Εξ. 3.5}$$

Βήμα Υπολογισμού Αναμενόμενης Τιμής (E-step)

Στο E-step, υπολογίζουμε την εκ των υστέρων πιθανότητα της κρυμμένης πληροφορίας $p(z_i | y_i, x_i)$ η οποία δίνει την πιθανότητα η καμπύλη i να έχει παραχθεί από την ομάδα z_i . Τότε η πιθανότητα αυτή ορίζεται ως:

$$\begin{aligned} w_{ik} &= p(z_i = k | y_i, x_i) \propto a_{z_i} p_{z_i}(y_i | x_i) \\ &= \frac{a_k \mathcal{N}(y_i | X_i \beta_k, \sigma_k^2 \mathbf{I})}{\sum_k^K a_k \mathcal{N}(y_i | X_i \beta_k, \sigma_k^2 \mathbf{I})} \end{aligned} \quad \text{Εξ. 3.6}$$

όπου οι πιθανότητες μέλους για την κάθε ομάδα πληρούν τις εξής προϋποθέσεις: $a_k \geq 0$ και $\sum_{k=1}^K a_k = 1$.

Η εκ των υστέρων αναμενόμενη τιμή της $\mathcal{L}_c$ στην Εξίσωση (3.5) λαμβάνεται με βάση και την παραπάνω εκ των υστέρων πιθανότητα και υπολογίζεται η συνάρτηση Q ως εξής:



Βήμα Μεγιστοποίησης (M-step)

Στο M-step μεγιστοποιούμε τη συνάρτηση Q ως προς τις παραμέτρους $\{\beta_k, \sigma_k^2, \alpha_k\}$.

Οι λύσεις οι οποίες προκύπτουν έχουν ως εξής:

$$\hat{\beta}_k = \left[\sum_i w_{ik} X_i' X_i \right]^{-1} \sum_i w_{ik} X_i' y_i \quad \text{Εξ. 3.8}$$

$$\hat{\sigma}_k^2 = \frac{1}{\sum_i w_{ik}} \sum_i w_{ik} \|y_i - X_i \beta_k\|^2 \quad \text{Εξ. 3.9}$$

$$\hat{\alpha}_k = \frac{1}{n} \sum_i w_{ik} \quad \text{Εξ. 3.10}$$

Η υπολογιστική πολυπλοκότητα του αλγορίθμου EM είναι γραμμική στον αριθμό των καμπυλών. Η αρχικοποίηση γίνεται επιλέγοντας τυχαία τιμές για τις πιθανότητες μέλους στις ομάδες και στη συνέχεια γίνονται οι επαναλήψεις του M-step. Η σύγκλιση του αλγορίθμου επιτυγχάνεται όταν η βελτίωση της λογαριθμικής πιθανοφάνειας, ανάμεσα σε δύο διαδοχικά βήματα είναι μικρότερη από ένα κατώφλι (π.χ. 1×10^{-6}).

Στη συνέχεια παρουσιάζεται η εφαρμογή του πολυωνυμικού μοντέλου παλινδρόμησης για την ομαδοποίηση μονοδιάστατων και δισδιάστατων χρονοσειρών.

3.3.3. Ομαδοποίηση μονοδιάστατων χρονοσειρών

Έστω ότι έχουμε ένα σύνολο Y από N μονοδιάστατες χρονοσειρές $\{y_1, \dots, y_i, \dots, y_N\}$. Θεωρούμε ότι κάθε χρονοσειρά έχει μήκος T , δηλαδή T μετρήσεις στον άξονα του χρόνου t .

Τότε το βασικό μοντέλο παλινδρόμησης έχει ως εξής:

$$y_t = X\beta_k + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma_k^2 \mathbf{I})$$

Εξ. 3.11



$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_k + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbf{I}) \quad \text{Εξ. 3.11}$$

θεωρώντας ότι ο πίνακας παλινδρόμησης $\mathbf{X}$ είναι ο $N \times p$ πίνακας Vandermonde και είναι σταθερός για κάθε καμπύλη, ενώ το $\boldsymbol{\beta}$ είναι το p -διάνυσμα των συντελεστών παλινδρόμησης. Ο πίνακας Vandermonde βαθμού p εκτιμώμενος στον άξονα του χρόνου t είναι ο ακόλουθος:

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & t_T & t_T^2 & \dots & t_T^p \end{bmatrix}$$

Με βάση το μοντέλο αυτό οι παραπάνω εξισώσεις του αλγορίθμου EM προσαρμόζονται ως εξής:

Βήμα Υπολογισμού Αναμενόμενης Τιμής (E-step)

Η εκ των υστέρων πιθανότητα της κρυμμένης πληροφορίας $p(z_i | \mathbf{y}_i, t)$ ορίζεται ως:

$$\begin{aligned} w_{ik} &= p(z_i = k | \mathbf{y}_i, t) \propto a_{z_i} p_{z_i}(\mathbf{y}_i | t) \\ &= \frac{a_k \mathcal{N}(\mathbf{y}_i | \mathbf{X}\boldsymbol{\beta}_k, \sigma_k^2 \mathbf{I})}{\sum_k a_k \mathcal{N}(\mathbf{y}_i | \mathbf{X}\boldsymbol{\beta}_k, \sigma_k^2 \mathbf{I})} \end{aligned} \quad \text{Εξ. 3.12}$$

Και η συνάρτηση Q ως εξής:

$$Q = E[\mathcal{L}_c | \mathbf{y}_i, t] = \sum_i \sum_k w_{ik} \log a_k \mathcal{N}(\mathbf{y}_i | \mathbf{X}\boldsymbol{\beta}_k, \sigma_k^2 \mathbf{I}) \quad \text{Εξ. 3.13}$$

Βήμα Μεγιστοποίησης (M-step)

Στο M-step μεγιστοποιούμε τη συνάρτηση Q ως προς τις παραμέτρους $\{\boldsymbol{\beta}_k, \sigma_k^2, a_k\}$.

Οι λύσεις οι οποίες προκύπτουν έχουν ως εξής:

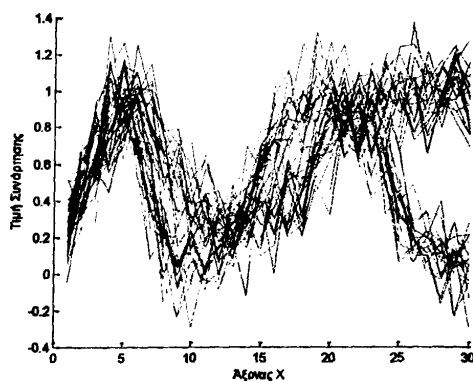
$$\hat{\boldsymbol{\beta}}_k = \left[\sum_i w_{ik} \mathbf{X}' \mathbf{X} \right]^{-1} \sum_i w_{ik} \mathbf{X}' \mathbf{y}_i \quad \text{Εξ. 3.14}$$



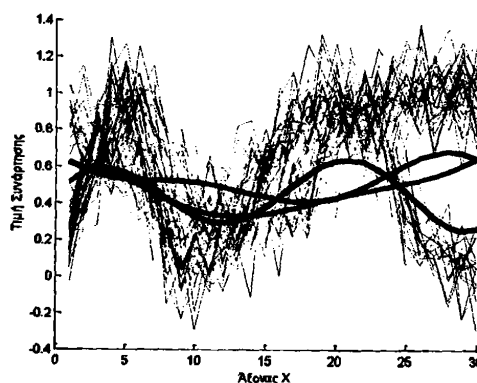
$$\hat{a}_k = \frac{1}{N} \sum_i w_{ik} \quad \text{Εξ. 3.16}$$

Στο Σχήμα 3.1 παρουσιάζεται ένα παράδειγμα της εκτέλεσης του αλγορίθμου EM για ένα πολυωνυμικό μικτό μοντέλο παλινδρόμησης, για μονοδιάστατες χρονοσειρές. Το σύνολο των δεδομένων αποτελείται από τρεις ομάδες των 20 χρονοσειρών η κάθε μία, παραγόμενες από τρία διαφορετικά πολυώνυμα με την προσθήκη θορύβου. Το μήκος των χρονοσειρών είναι 30.

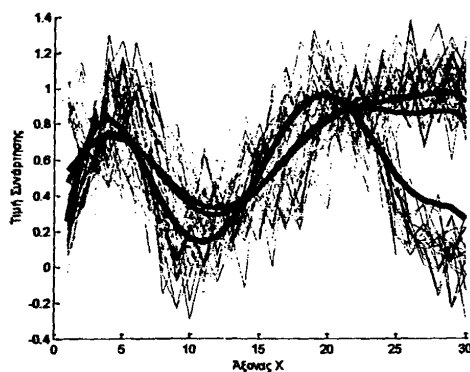
Το Σχήμα 3.1(α) δείχνει όλες τις καμπύλες που παίρνει ως είσοδο ο EM. Οι καμπύλες παρουσιάζονται με διαφορετικά χρώματα για να φανούν οι ομάδες, αλλά αυτή η πληροφορία δεν δίνεται στον αλγόριθμο. Το Σχήμα 3.1(β) δείχνει την αρχικοποίηση του αλγορίθμου για τις τρεις ομάδες. Το Σχήμα 3.1(γ) δείχνει τα κέντρα των ομάδων μετά από μια επανάληψη του αλγορίθμου και το Σχήμα 3.1(δ) την τελική ομαδοποίηση, η οποία προκύπτει μετά από τέσσερις επαναλήψεις.



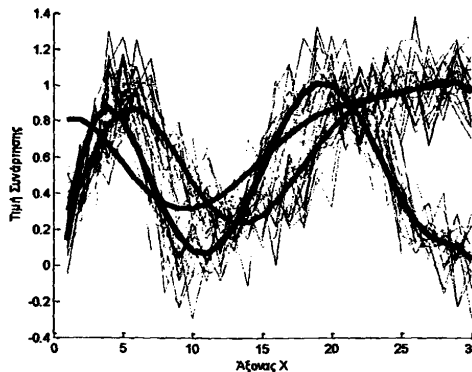
(α) Πραγματικά Δεδομένα



(β) Αρχικοποίηση του EM



(γ) Κέντρα ομάδων μετά από μια επανάληψη



(δ) Τελικές ομάδες (πέντε επαναλήψεις)

Σχήμα 3.1 Αλγόριθμος EM για PRM μοντέλο (μονοδιάστατες χρονοσειρές)

3.3.4. Ομαδοποίηση δισδιάστατων χρονοσειρών

Η διαφορά των δισδιάστατων από τις μονοδιάστατες χρονοσειρές έγκειται στην αναπαράστασή τους στο χώρο. Δηλαδή, μια μονοδιάστατη χρονοσειρά απεικονίζει μια μόνο τιμή για κάθε παρατήρηση του χρόνου, ενώ αντίθετα μια δισδιάστατη χρονοσειρά απεικονίζει, για κάθε παρατήρηση του χρόνου, δύο τιμές, μια για τη διάσταση x και μια για τη διάσταση y . Οπότε ουσιαστικά, μπορούμε να πούμε ότι μια δισδιάστατη χρονοσειρά y_i αποτελεί σύνθεση δύο μονοδιάστατων y_{ix} και y_{iy} . Το μοντέλο που χρησιμοποιείται είναι το ίδιο με εκείνο που χρησιμοποιήθηκε και για την ομαδοποίηση μονοδιάστατων καμπυλών, με την μόνη διαφορά ότι οι παραπάνω τύποι εφαρμόζονται και στις δύο διαστάσεις των χρονοσειρών και κατά x και κατά y .

Οπότε τα δύο βήματα του αλγορίθμου EM έχουν ως εξής:

Βήμα Υπολογισμού Αναμενόμενης Τιμής (E-step)

Η εκ των υστέρων πιθανότητα της κρυμμένης πληροφορίας $p(z_i|y_i, t)$ ορίζεται ως:

$$\begin{aligned} w_{ik} &= p(z_i = k|y_i, t) \propto a_{z_i} p_{z_i}(y_i|t) \\ &= a_k \mathcal{N}(y_{ix} | \mathbf{X}\beta_{kx}, \sigma_{kx}^2 \mathbf{I}) \mathcal{N}(y_{iy} | \mathbf{X}\beta_{ky}, \sigma_{ky}^2 \mathbf{I}) \end{aligned}$$

Εξ. 3.17



Και η συνάρτηση Q ως εξής:

$$Q = E[\mathcal{L}_c | \mathbf{y}_i, t] \\ = \sum_i \sum_k w_{ik} \log a_k \mathcal{N}(\mathbf{y}_{ix} | \mathbf{X}\boldsymbol{\beta}_{kx}, \sigma_{kx}^2 \mathbf{I}) \mathcal{N}(\mathbf{y}_{iy} | \mathbf{X}\boldsymbol{\beta}_{ky}, \sigma_{ky}^2 \mathbf{I}) \quad \text{Εξ. 3.18}$$

Βήμα Μεγιστοποίησης (M-step)

Στο M-step μεγιστοποιούμε τη συνάρτηση Q ως προς τις παραμέτρους $\{\boldsymbol{\beta}_{kx}, \sigma_{kx}^2, \boldsymbol{\beta}_{ky}, \sigma_{ky}^2, a_k\}$. Οι λύσεις οι οποίες προκύπτουν έχουν ως εξής:

$$\hat{\boldsymbol{\beta}}_{kx} = \left[\sum_i w_{ik} \mathbf{X}'\mathbf{X} \right]^{-1} \sum_i w_{ik} \mathbf{X}'\mathbf{y}_{ix} \quad \text{Εξ. 3.19}$$

$$\hat{\boldsymbol{\beta}}_{ky} = \left[\sum_i w_{ik} \mathbf{X}'\mathbf{X} \right]^{-1} \sum_i w_{ik} \mathbf{X}'\mathbf{y}_{iy} \quad \text{Εξ. 3.20}$$

$$\hat{\sigma}_{kx}^2 = \frac{1}{\sum_i w_{ik}} \sum_i w_{ik} \|\mathbf{y}_{ix} - \mathbf{X}\boldsymbol{\beta}_{kx}\|^2 \quad \text{Εξ. 3.21}$$

$$\hat{\sigma}_{ky}^2 = \frac{1}{\sum_i w_{ik}} \sum_i w_{ik} \|\mathbf{y}_{iy} - \mathbf{X}\boldsymbol{\beta}_{ky}\|^2 \quad \text{Εξ. 3.22}$$

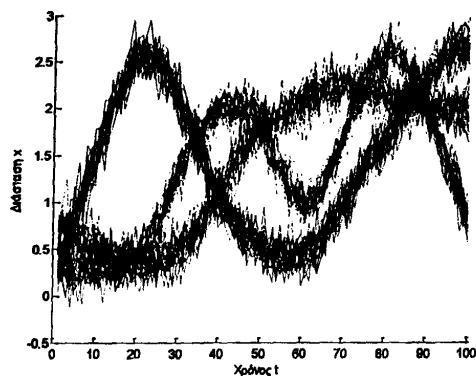
$$\hat{a}_k = \frac{1}{N} \sum_i w_{ik} \quad \text{Εξ. 3.23}$$

Στο Σχήμα 3.2 παρουσιάζεται ένα παράδειγμα της εκτέλεσης του αλγορίθμου EM για ένα πολυωνμικό μικτό μοντέλο παλινδρόμησης, για δισδιάστατες χρονοσειρές. Το σύνολο των δεδομένων αποτελείται από τρεις ομάδες, των 30 χρονοσειρών η κάθε μία, παραγόμενες από τρία διαφορετικά πολυώνυμα με την προσθήκη θορύβου. Το μήκος των χρονοσειρών είναι 100.

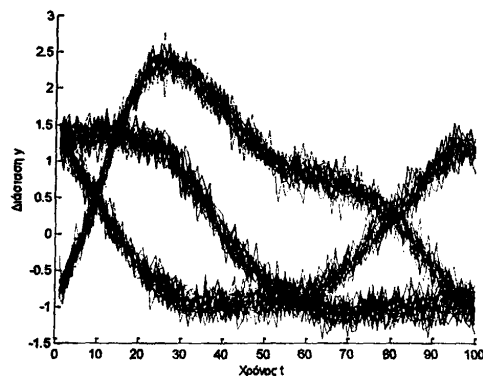
Το Σχήμα 3.2(α) δείχνει όλες τις χρονοσειρές στον άξονα x , ενώ το Σχήμα 3.2(β) δείχνει όλες τις χρονοσειρές στον άξονα y . Το Σχήμα 3.2(γ) δείχνει συνολικά τις καμπύλες και στις δύο διαστάσεις τις οποίες παίρνει ως είσοδο ο EM. Οι καμπύλες



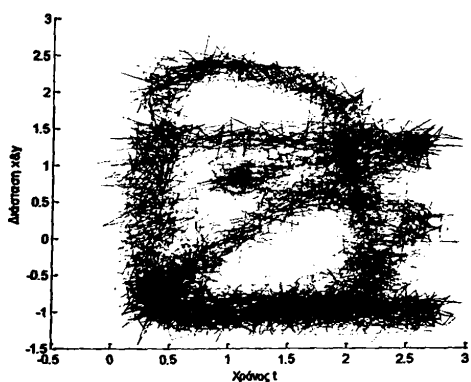
παρουσιάζονται με διαφορετικά χρώματα για να φανούν οι ομάδες, αλλά αυτή η πληροφορία δεν δίνεται στον αλγόριθμο. Το Σχήμα 3.2(δ) δείχνει την αρχικοποίηση του αλγορίθμου για τις τρεις ομάδες. Το Σχήμα 3.2(ε) δείχνει τα κέντρα των ομάδων μετά από μια επανάληψη του αλγορίθμου και το Σχήμα 3.2(στ) την τελική ομαδοποίηση, η οποία προκύπτει μετά από τρεις επαναλήψεις.



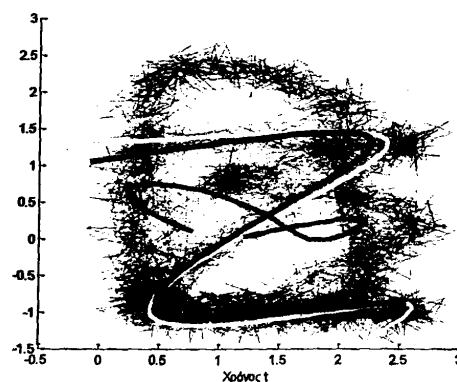
(α) Πραγματικά Δεδομένα x



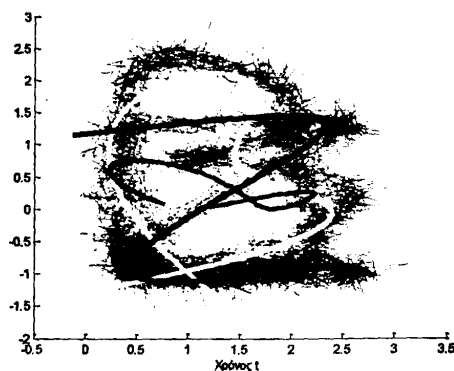
(β) Πραγματικά Δεδομένα y



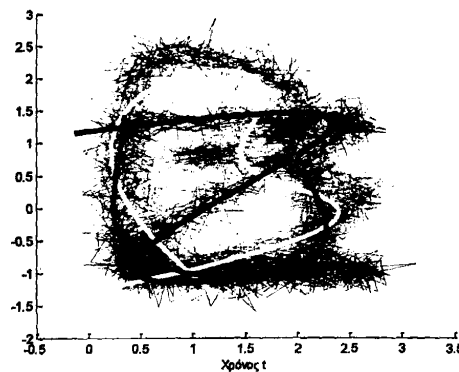
(γ) Πραγματικά Δεδομένα κατά x και y



(δ) Αρχικοποίηση του EM



(ε) Κέντρα ομάδων μετά από μια επανάληψη



(στ) Τελικές ομάδες (τρεις επαναλήψεις)

Σχήμα 3.2 Αλγόριθμος EM για PRM μοντέλο (δισδιάστατες χρονοσειρές)

3.4. Άλλα μοντέλα παλινδρόμησης

Στην ενότητα αυτή περιγράφονται δυο άλλα μικτά μοντέλα παλινδρόμησης τα οποία χρησιμοποιούνται για την ομαδοποίηση χρονοσειρών. Το κατά τμήματα πολυωνυμικό μικτό μοντέλο παλινδρόμησης (spline regression mixtures, SRM), το οποίο είναι ένα ημι-παραμετρικό μοντέλο ομαδοποίησης και το μικτό μοντέλο παλινδρόμησης το οποίο χρησιμοποιεί συναρτήσεις πυρήνα (kernel regression mixtures, KRM), το οποίο είναι ένα μη-παραμετρικό μοντέλο.

3.4.1. Κατά τμήματα πολυωνυμικά μικτά μοντέλα παλινδρόμησης

Για διάφορους λόγους, είναι προτιμότερο πολλές φορές να προσεγγίζεται μια χρονοσειρά με περισσότερα του ενός πολώνυμα, δηλαδή ανά τμήματα να περιγράφεται η χρονοσειρά με διαφορετικά πολώνυμα. Η τεχνική αυτή της κατά τμήματα πολυωνυμικής προσέγγισης καλείται spline [3].

Οι splines χρονοσειρές περιγράφονται παραμετρικά από μια εξίσωση της μορφής $r = r(t)$. Το πεδίο ορισμού της παραμέτρου t είναι το διάστημα $[a, b]$, το οποίο θεωρείται χωρισμένο σε k τμήματα, όχι κατ' ανάγκη ίσα μεταξύ τους. Δηλαδή, έχουμε $k + 1$ τιμές $t_0, t_1, \dots, t_k$ της παραμέτρου t με $t_0 = a, t_k = b$ και $t_{i-1} < t_i$.



$i = 1, \dots, k$. Έτσι, μια πολυωνυμική spline τάξεως $p + 1$, είναι κάθε χρονοσειρά $r = r(t)$ η οποία έχει τις παρακάτω δύο ιδιότητες:

- Σε κάθε διάστημα $[t_{i-1}, t_i]$, $i = 1, \dots, k$, η συνάρτηση $r(t)$ είναι ένα πολυώνυμο βαθμού p ως προς t .
- Η συνολική χρονοσειρά, που προκύπτει από την ένωση των πολυωνυμικών τμημάτων, είναι συνεχής και έχει συνεχείς παραγώγους μέχρι τάξεως $p - 1$ σε όλα τα ενδιάμεσα σημεία t_i , $i = 1, \dots, k - 1$



Όπως φαίνεται και από το παραπάνω σχήμα, οι χρονοσειρές οι οποίες προκύπτουν από την τεχνική αυτή έχουν πιο ομαλή μορφή από τις απλές πολυωνυμικές χρονοσειρές. Το σύνολο των σημείων τα οποία ορίζουν μια spline καμπύλη ονομάζονται *κόμβοι*.

Οι πιο γνωστές spline χρονοσειρές είναι οι B-splines [4]. Οι χρονοσειρές αυτές είναι αρκετά αποτελεσματικές λόγω του μπλοκ διαγώνιου παραγόμενου πίνακα. Έστω ότι $t = \{t_1 < \dots < t_N\}$ είναι μια αυξανόμενη ακολουθία κόμβων και $[t_m, t_{m+1})$ είναι το ημι-ανοιχτό διάστημα από t_m μέχρι t_{m+1} . Τότε η καμπύλη B-spline B_{mp} τάξεως p είναι ένα τμηματικό πολυώνυμο το οποίο έχει πεπερασμένη τιμή στο διάστημα $[t_m, t_{m+p})$ και είναι μηδέν οπουδήποτε αλλού. Τα πολυωνυμικά τμήματα της B_{mp} είναι βαθμού $p - 1$. Γενικά η B_{mp} ορίζεται ως εξής:

$$\sum_m^L B_{mp}(x) = 1$$

όπου $L = N - p$ μας δίνει τον αριθμό των χρονοσειρών B-splines που ορίζονται στην ακολουθία t των κόμβων.

Τότε η χρονοσειρά spline $s(x)$ ορίζεται ως ο γραμμικός συνδυασμός της B_{mp} :



$$s(x) = \sum_m^L B_{mp}(x) c_m \quad \text{Εξ. 3.24}$$

όπου c_m είναι ο συντελεστής της χρονοσειράς.

Για την αναπαράσταση μια χρονοσειράς y_i με μια spline, εξισώνουμε το j σημείο y_{ij} στην τιμή της συνάρτησης spline στο j σημείο του χρόνου x_{ij} . Δηλαδή, θέτουμε $y_{ij} = s(x_{ij})$ για κάθε $1 \leq j \leq n_i$.

Η παραπάνω εξίσωση για την χρονοσειρά y_i μπορεί να γραφεί με μορφή πινάκων. Σαν $B_{mp}(x_i)$ ορίζεται το n_i -διάνυσμα των μεμονωμένων χρονικών σημείων του x_i εκτιμώμενες από τη συνάρτηση B_{mp} . Τότε ο spline πίνακας B_i είναι ο $n_i \times L$ πίνακας:

$$B_i = [B_{1p}(x_i) \quad B_{1p}(x_i) \quad \dots \quad B_{1p}(x_i)] \quad \text{Εξ. 3.25}$$

Με τη βοήθεια του παραπάνω πίνακα μπορούμε να αναπαραστήσουμε την χρονοσειρά y_i χρησιμοποιώντας μια spline ως εξής:

$$y_i = B_i c \quad \text{Εξ. 3.26}$$

όπου c είναι το διάνυσμα των συντελεστών της χρονοσειράς spline.

Το μοντέλο των spline χρονοσειρών παρέχει μεγάλη ευελιξία στην ομαδοποίηση χρονοσειρών, ειδικά όταν οι χρονοσειρές δεν ακολουθούν μια καθαρά πολυωνυμική μορφή, με αποτέλεσμα να μη μπορούν να αναπαρασταθούν από ένα και μόνο πολυώνυμο. Παρόλα αυτά όμως, το βασικό μειονέκτημα του μοντέλου των spline είναι το μεγάλο μέγεθος του διανύσματος των συντελεστών των χρονοσειρών spline, το οποίο αυξάνει την υπολογιστική πολυπλοκότητα της μεθόδου.



3.4.2. Μικτό μοντέλο παλινδρόμησης με συναρτήσεις πυρήνα

Η μη-παραμετρική εκτίμηση συναρτήσεων έχει μελετηθεί με διάφορους τρόπους, όπως kernel smoothing (Wand & Jones, 1995) [32], local polynomial modelling (Fan & Gijbels, 1996) [7] και density estimation (Silverman, 1986) [28]. Όσον αφορά την ομαδοποίηση χρονοσειρών, η μοντελοποίηση των συναρτήσεων πυκνότητας με μη-παραμετρικά μοντέλα παλινδρόμησης, μας δίνει τη δυνατότητα να ομαδοποιήσουμε χρονοσειρές για τις οποίες η γενική σχέση μεταξύ της χρονοσειράς y και της μεταβλητής x είναι αβέβαιη, ή όταν δε θέλουμε να κάνουμε κάποια υπόθεση για τη σχέση αυτή.

Η βασική ιδέα πίσω από την παλινδρόμηση με συναρτήσεις πυρήνα είναι ότι μπορούμε να προσεγγίσουμε οποιαδήποτε αυθαίρετη συνάρτηση με ένα σύνολο απλών, τοπικά σταθμισμένων (locally-weighted) συναρτήσεων, όπως για παράδειγμα συναρτήσεις γραμμικής παλινδρόμησης. Για να προσεγγίσουμε την άγνωστη συνάρτηση σε ένα σημείο x_0 , εφαρμόζουμε μια τοπικά σταθμισμένη, γραμμική παλινδρόμηση (τάξης p) γύρω από το σημείο x_0 και θεωρούμε την εκτίμηση $\hat{y}$ ως το ύψος της προσαρμογής (fit) αυτής. Τα βάρη παράγονται από μια συμμετρική συνάρτηση πυρήνα (π.χ. Gaussian density) με κέντρο περίπου στο σημείο x_0 .

Το βασικό μειονέκτημα των KRMs είναι η μεγάλη υπολογιστική πολυπλοκότητα σε αντίθεση με το απλό πολυωνυμικό μοντέλο, το οποίο παρέχει παρόμοια ευελιξία.



ΚΕΦΑΛΑΙΟ 4. ΟΜΑΔΟΠΟΙΗΣΗ ΧΡΟΝΟΣΕΙΡΩΝ ΜΕ ΒΑΣΗ ΤΟ ΜΟΝΤΕΛΟ ΜΕΤΑΤΟΠΙΣΗΣ ΣΤΗ ΔΙΑΣΤΑΣΗ ΤΟΥ ΧΩΡΟΥ

4.1 Εισαγωγή

4.2 Προηγούμενες Μελέτες

4.3 Μοντέλο Μετατόπισης στη διάσταση του χώρου

4.4 Ομαδοποίηση με το μοντέλο μετατόπισης

4.1. Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζουμε μια επέκταση του πολυωνυμικού μικτού μοντέλου ομαδοποίησης χρονοσειρών, το οποίο περιγράψαμε στο Κεφάλαιο 3. Πιο συγκεκριμένα, ορίζουμε την έννοια της μετατόπισης - ευθυγράμμισης χρονοσειρών και ενσωματώνουμε το νέο αυτό παράγοντα στο πολυωνυμικό μοντέλο μας. Στην ενότητα 4.2 παρουσιάζουμε κάποιες προηγούμενες μελέτες που έχουν γίνει αναφορικά με το πρόβλημα της ευθυγράμμισης. Στην ενότητα 4.3 ορίζουμε το μοντέλο μετατόπισης στη διάσταση του χώρου. Τέλος, στην ενότητα 4.4 παρουσιάζουμε την εφαρμογή του μοντέλου αυτού στο πρόβλημα της ομαδοποίησης χρονοσειρών.

4.2. Προηγούμενες Μελέτες

Για το πρόβλημα της ευθυγράμμισης και στη συνέχεια ομαδοποίησης χρονοσειρών έχουν γίνει ελάχιστες μελέτες. Έχουν γίνει όμως παρόμοιες μελέτες σε άλλους ερευνητικούς τομείς. Αξιόλογη προσπάθεια, όσον αφορά την ευθυγράμμιση και την ομαδοποίηση, έχει γίνει στο χώρο της μοντελοποίησης εικόνων χρησιμοποιώντας



μικτές κατανομές. Παρότι στο χώρο αυτό δεν υπάρχει η έννοια της χρονοσειράς, το γενικό πλαίσιο είναι παρεμφερές. Για παράδειγμα, το μοντέλο των μετασχηματισμένων μικτών Gaussian κατανομών (transformed mixtures of Gaussians, TMG) χρησιμοποιεί μια πιθανοτική προσέγγιση και έναν αλγόριθμο EM για την εύρεση μικτών κατανομών εικόνων, υποκείμενες σε διαφορετικούς γραμμικούς μετασχηματισμούς (Frey & Jojic 1999, 2002, 2003) [8], [9], [10].

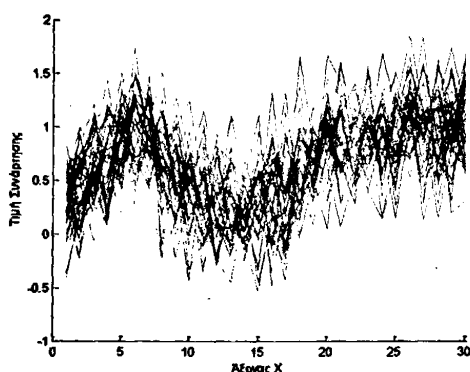
Μια από τις πιο ολοκληρωμένες και πρόσφατες μελέτες για την ευθυγράμμιση και την ομαδοποίηση χρονοσειρών, και όχι απλών συνόλων σημείων, έχει γίνει από τον Gaffney (2004) [12]. Στη μελέτη αυτή η μετατόπιση θεωρείται ως μια επιπλέον μεταβλητή η οποία εκτιμάται, όπως και οι υπόλοιπες μεταβλητές, με τη βοήθεια του αλγορίθμου EM. Στη συγκεκριμένη μελέτη βασιζόμαστε ούτως ώστε να ορίσουμε το μοντέλο της μετατόπισης, το οποίο περιγράφουμε στη συνέχεια.

4.3. Μοντέλο Μετατόπισης στη διάσταση του χώρου

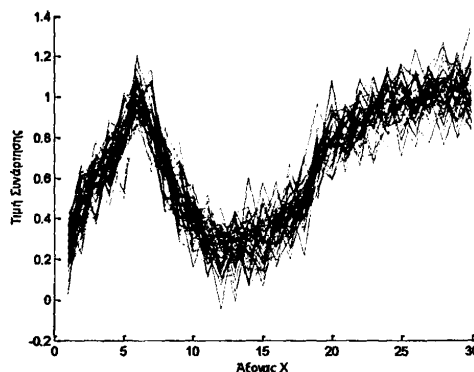
Είναι σύνηθες φαινόμενο ένα σύνολο χρονοσειρών να προκύψει από μετρήσεις που έχουν επηρεαστεί από εξωτερικούς παράγοντες, όπως για παράδειγμα από θόρυβο. Σε αυτές τις περιπτώσεις οι χρονοσειρές δε θα κατέχουν την πραγματική τους θέση αλλά θα είναι μετατοπισμένες είτε στο χώρο είτε στο χρόνο. Στην παρούσα εργασία μελετάμε μόνο την πρώτη περίπτωση, δηλαδή τη μετατόπιση στο χώρο.

Στο παρακάτω σχήμα φαίνεται ένα παράδειγμα το οποίο δείχνει την ευθυγράμμιση ενός συνόλου χρονοσειρών, οι οποίες έχουν επηρεαστεί από θόρυβο.





(α) Μη ευθυγραμμισμένες χρονοσειρές



(β) Ευθυγραμμισμένες χρονοσειρές

Σχήμα 4.1 Φαινόμενο μετατόπισης

Όπως φαίνεται και από το παραπάνω σχήμα το υποκείμενο σχήμα είναι πιο ξεκάθαρο όταν οι χρονοσειρές είναι ευθυγραμμισμένες, δηλαδή όταν μετατοπίζονται, με βάση κάποιο παράγοντα μετατόπισης, ούτως ώστε να προσεγγίζουν πληρέστερα την πραγματική τους θέση.

4.3.1. Ορισμός Μοντέλου Μετατόπισης

Για να ορίσουμε το γενικό πλαίσιο του μοντέλου μετατόπισης θα χρησιμοποιήσουμε το πολυωνυμικό μοντέλο παλινδρόμησης που περιγράψαμε στην ενότητα 3.3. Προσθέτουμε στο μοντέλο μας μια μεταβλητή d_i , η οποία δηλώνει την μετατόπιση της χρονοσειράς στο χώρο. Οπότε το νέο μοντέλο είναι:

$$y_i = X_i\beta + d_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I) \quad \text{Εξ. 4.1}$$

Με βάση το μοντέλο αυτό, η πιθανότητα της χρονοσειράς $p(y_i|\theta, d_i)$ εξαρτάται από τις καθολικές παραμέτρους $\theta = \{\beta, \sigma^2\}$, καθώς επίσης και από την άγνωστη μεταβλητή της μετατόπισης d_i . Για να προσεγγίσει το μοντέλο τα πραγματικά δεδομένα θα πρέπει να βρεθούν και οι παράμετροι θ αλλά και κάθε ένα από τα d_i .



Συνάρτηση πιθανότητας της d_i

Για να μπορέσουμε να εκτιμήσουμε τις μετατοπίσεις d_i θα πρέπει πρώτα να ορίσουμε ένα μοντέλο για τη συνάρτηση πιθανότητας. Υποθέτουμε ότι η d_i είναι μια τυχαία μεταβλητή η οποία έχει μια συγκεκριμένη πυκνότητα πιθανότητας. Θεωρούμε ότι η τυχαία αυτή μεταβλητή ακολουθεί την Gaussian κατανομή με μέσο μηδέν και κάποια διακύμανση u^2 . Δηλαδή:

$$p(d_i) = \mathcal{N}(d_i|0, u^2) \quad \text{Εξ. 4.2}$$

Η κατανομή αυτή μας επιτρέπει να κάνουμε δύο υποθέσεις:

- Η πιο πιθανή μετατόπιση είναι η μηδενική μετατόπιση
- Η αρνητική ή η θετική μετατόπιση είναι ισοδύναμες

Η διακύμανση u^2 αποτελεί μια ακόμη παράμετρο η οποία χρειάζεται να εκτιμηθεί.

Από κοινού (joint) και περιθώρια (marginal) πιθανότητα

Χρησιμοποιώντας το μοντέλο το οποίο ορίστηκε στην Εξίσωση 3.1 μπορούμε να ορίσουμε την υπό συνθήκη πυκνότητα πιθανότητας για το y_i , ως εξής:

$$p(y_i|d_i) = \mathcal{N}(y_i|\mathbf{X}_i\boldsymbol{\beta} + d_i, \sigma^2\mathbf{I}) \quad \text{Εξ. 4.3}$$

Υποθέτοντας όμως ότι η d_i είναι μια τυχαία μεταβλητή, όπως κάναμε στην παραπάνω παράγραφο, μπορούμε να ορίσουμε την από κοινού πυκνότητα πιθανότητας των y_i και d_i , ως εξής:

$$\begin{aligned} p(y_i, d_i) &= p(y_i|d_i)p(d_i) \\ &= \mathcal{N}(y_i|\mathbf{X}_i\boldsymbol{\beta} + d_i, \sigma^2\mathbf{I})\mathcal{N}(d_i|0, u^2) \end{aligned} \quad \text{Εξ. 4.4}$$

Με βάση την παραπάνω από κοινού πυκνότητα πιθανότητας μπορούμε να ορίσουμε και την περιθώρια πυκνότητα για το y_i , ολοκληρώνοντας ως προς d_i :

$$p(y_i) = \int p(y_i, d_i) dd_i$$



$$\begin{aligned}
&= \int p(\mathbf{y}_i | d_i) p(d_i) dd_i \\
&= \int \mathcal{N}(\mathbf{y}_i | \mathbf{X}_i \boldsymbol{\beta} + d_i, \sigma^2 \mathbf{I}) \mathcal{N}(d_i | 0, u^2) dd_i \\
&= \mathcal{N}(\mathbf{y}_i | \mathbf{X}_i \boldsymbol{\beta}, u^2 \mathbf{1} + \sigma^2 \mathbf{I})
\end{aligned}$$
Εξ. 4.5

όπου το $\mathbf{1}$ είναι ένας πίνακας $n_i \times n_i$ με άσσους. Η περιθώρια πυκνότητα εξαρτάται από τις παραμέτρους $\theta = \{\boldsymbol{\beta}, \sigma^2, u^2\}$ και από τον μη τυχαίο πίνακα $\mathbf{X}_i$.

Συνάρτηση Πιθανοφάνειας

Η παραπάνω περιθώρια πυκνότητα πιθανότητας μας βοηθάει να ορίσουμε την πιθανοφάνεια για το σύνολο $Y = \{\mathbf{y}_i\}_1^n$ των n καμπυλών. Η πιθανοφάνεια θα είναι το άθροισμα, για όλες τις καμπύλες, τις περιθώριες πιθανότητας του $\mathbf{y}_i$. Δηλαδή:

$$\log p(Y) = \sum_i \log \mathcal{N}(\mathbf{y}_i | \mathbf{X}_i \boldsymbol{\beta}, u^2 \mathbf{1} + \sigma^2 \mathbf{I})$$
Εξ. 4.6

Η πιθανοφάνεια είναι η εξίσωση που θα πρέπει να μεγιστοποιηθεί ούτως ώστε να γίνει η εκτίμηση των παραμέτρων $\theta = \{\boldsymbol{\beta}, \sigma^2, u^2\}$.

4.4. Ομαδοποίηση με το μοντέλο της μετατόπισης

Στην ενότητα αυτή θα χρησιμοποιήσουμε το μοντέλο της μετατόπισης, το οποίο περιγράφηκε παραπάνω, ούτως ώστε να δημιουργήσουμε το γενικό πλαίσιο για την ομαδοποίηση χρονοσειρών με βάση το μοντέλο της μετατόπισης.

Η προσθήκη, στο μοντέλο της Εξίσωσης 4.1, της εξάρτησης από τις ομάδες, σημαίνει ότι το μοντέλο αυτό θα πρέπει να επαναληφθεί για τις K διαφορετικές υποκείμενες ομάδες του συνόλου δεδομένων. Οπότε το γενικό μοντέλο θα είναι το εξής:

$$\mathbf{y}_l = \mathbf{X}_l \boldsymbol{\beta}_k + d_l + \epsilon_l, \quad \epsilon_l \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbf{I})$$
Εξ. 4.7



με $d_i \sim \mathcal{N}(0, u_k^2)$ και διάνυσμα παραμέτρων $\theta = \{\beta_k, \sigma_k^2, u_k^2\}$.

Η υπό συνθήκη και εξαρτημένη από την ομάδα πυκνότητα πιθανότητας για κάθε y_i είναι η εξής:

$$p_k(y_i|d_i) = \mathcal{N}(X_i\beta_k + d_i, \sigma_k^2\mathbf{I}) \quad \text{Εξ. 4.8}$$

Και επειδή η πιθανότητα μέλους της κάθε ομάδας είναι άγνωστη, οδηγούμαστε σε μια μικτή υπό συνθήκη κατανομή:

$$p(y_i|d_i) = \sum_k \alpha_k p_k(y_i|d_i) \quad \text{Εξ. 4.9}$$

όπου οι πιθανότητες μέλους για την κάθε ομάδα πληρούν τις εξής προϋποθέσεις: $\alpha_k \geq 0$ και $\sum_{k=1}^K \alpha_k = 1$.

Το παραπάνω μικτό μοντέλο υποθέτει ότι η παράμετρος μετατόπισης d_i είναι γνωστή. Στην επόμενη ενότητα ορίζονται οι πυκνότητες για την περίπτωση που η παράμετρος αυτή δεν είναι γνωστή.

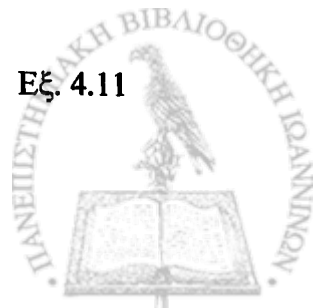
4.4.1. Από κοινού, περιθώρια και πιθανοφάνεια

Η από κοινού πυκνότητα των y_i και d_i είναι όπως ορίστηκε στην Εξίσωση 4.4, με τη διαφορά όμως ότι λόγω της εξάρτησης από τις ομάδες, θα έχουμε K ξεχωριστές πυκνότητες, μια για κάθε ομάδα. Δηλαδή:

$$\begin{aligned} p_k(y_i, d_i) &= p_k(y_i|d_i)p_k(d_i) \\ &= \mathcal{N}(y_i|X_i\beta_k + d_i, \sigma_k^2\mathbf{I})\mathcal{N}(d_i|0, u_k^2) \end{aligned} \quad \text{Εξ. 4.10}$$

Στη συνέχεια, ολοκληρώνοντας ως προς d_i μπορούμε να ορίσουμε και την περιθώρια πυκνότητα για το y_i :

$$\begin{aligned} p_k(y_i) &= \int p_k(y_i, d_i) dd_i \\ &= \mathcal{N}(y_i|X_i\beta_k, \mathbf{U}_k), \quad \mathbf{U}_k = u^2\mathbf{1} + \sigma^2\mathbf{I} \end{aligned} \quad \text{Εξ. 4.11}$$



Με βάση την παραπάνω περιθώρια πυκνότητα πιθανότητας μπορούμε να ορίσουμε τη συνάρτηση πιθανοφάνειας για το σύνολο $Y = \{y_i\}_1^n$ των n καμπυλών. Η συνάρτηση πιθανοφάνειας θα είναι το άθροισμα, για όλες τις καμπύλες, τις περιθώριας πιθανότητας του y_i . Δηλαδή:

$$\log p(Y) = \sum_i \log \sum_k \alpha_k \mathcal{N}(y_i | X_i \beta_k, U_k) \quad \text{Εξ. 4.12}$$

4.4.2. Αλγόριθμος EM για το μοντέλο της μετατόπισης

Στην ενότητα αυτή περιγράφεται ο αλγόριθμος EM για την ομαδοποίηση χρονοσειρών με βάση το μοντέλο της μετατόπισης, το οποίο ορίστηκε στην προηγούμενη ενότητα. Τα βήματα που ακολουθούνται είναι ανάλογα με εκείνα του αλγορίθμου EM για το απλό πολυωνυμικό μικτό μοντέλο παλινδρόμησης.

Αρχικά θεωρούμε ότι το z_i δηλώνει τη συμμετοχή σε μια ομάδα (cluster membership) για τη χρονοσειρά y_i . Τόσο η συμμετοχή σε μια ομάδα $\{z_i\}$, όσο και η παράμετρος d_i θεωρούνται κρυμμένη πληροφορία. Τότε, η πυκνότητα των κρυμμένων δεδομένων γίνεται η εκ των υστέρων (posterior) από κοινού πιθανότητα $p(z_i, d_i | y_i)$.

Στη συνέχεια η συνολική λογαριθμική συνάρτηση πιθανοφάνειας μπορεί να γραφεί ως το άθροισμα, για όλες τις καμπύλες, του λογαρίθμου του γινομένου των a_{z_i} και από κοινού πυκνοτήτων όπως ορίστηκαν στην Εξίσωση 4.10. Δηλαδή:

$$\mathcal{L}_c = \sum_i \log a_{z_i} p_{z_i}(y_i | d_i) p_{z_i}(d_i) \quad \text{Εξ. 4.13}$$

Βήμα Υπολογισμού Αναμενόμενης Τιμής (E-step)

Στο E-step, υπολογίζουμε την εκ των υστέρων, από κοινού πιθανότητα $p(z_i, d_i | y_i)$ και στη συνέχεια τη χρησιμοποιούμε για να υπολογίσουμε την αναμενόμενη τιμή για



την λογαριθμική συνάρτηση πιθανοφάνειας της Εξίσωσης 4.13. Η εκ των υστέρων πιθανότητα μπορεί να γραφεί ως το γινόμενο $p(z_i, d_i | y_i) = p(d_i | z_i, y_i)p(z_i | y_i)$.

Ο πρώτος όρος $p(d_i | z_i, y_i)$ μπορεί να γραφεί ως:

$$p(d_i | z_i = k, y_i) \propto p_k(y_i | d_i) p_k(d_i) \quad \text{Εξ. 4.14}$$

Οπότε έχουμε K διαφορετικές εκ των υστέρων πυκνότητες, μια για κάθε τιμή του z_i . Η κάθε μια από αυτές τις πυκνότητες μπορεί να θεωρηθεί ότι ακολουθεί μια μονομεταβλητή κανονική εκ των υστέρων πυκνότητα με μέσο:

$$\hat{d}_{ik} = V_{d_{ik}}(y_i - X_i \beta_k) U^{-1} \mathbf{1} \quad \text{Εξ. 4.15}$$

και διακύμανση:

$$V_{d_{ik}} = (\mathbf{1}' U_k^{-1} \mathbf{1} + 1/u_k^2)^{-1} \quad \text{Εξ. 4.16}$$

Όσον αφορά το δεύτερο όρο, δηλαδή την εκ των υστέρων πιθανότητα $p(z_i | y_i)$, αυτή είναι απλά η πιθανότητα μέλους w_{ik} , δηλαδή η πιθανότητα η καμπύλη y_i να προήλθε από την ομάδα $z_i = k$. Όπως έχουμε δείξει και στο Κεφ.3, η πιθανότητα αυτή είναι η ακόλουθη:

$$\begin{aligned} w_{ik} &= p(z_i = k | y_i) \propto p(y_i | z_i = k) p(z_i = k) \\ &= \alpha_k p_k(y_i) \\ &= \frac{\alpha_k \mathcal{N}(y_i | X_i \beta_k, U_k)}{\sum_k \alpha_k \mathcal{N}(y_i | X_i \beta_k, U_k)} \end{aligned} \quad \text{Εξ. 4.17}$$

Υπολογισμός της συνάρτησης Q

Για τον υπολογισμό της συνάρτησης Q θα πρέπει να υπολογίσουμε την εκ των υστέρων αναμενόμενη τιμή της Εξίσωσης (4.13) με βάση και την παραπάνω εκ των υστέρων πιθανότητα στην Εξίσωση (4.17). Οπότε συνολικά θα έχουμε:



$$Q = E[\mathcal{L}_c | \mathbf{y}_i] = \sum_i \sum_k w_{ik} \log a_k \mathcal{N}(\mathbf{y}_i | \mathbf{X}_i \boldsymbol{\beta}_k, \mathbf{U}_k) \quad \text{Εξ. 4.18}$$

Βήμα Μεγιστοποίησης (M-step)

Στο M-step μεγιστοποιούμε τη συνάρτηση Q ως προς τις παραμέτρους $\{\boldsymbol{\beta}_k, \sigma_k^2, u_k^2, \alpha_k\}$. Οι λύσεις που προκύπτουν έχουν ως εξής:

$$\hat{\boldsymbol{\beta}}_k = \left[\sum_i w_{ik} \mathbf{X}_i' \mathbf{X}_i \right]^{-1} \sum_i w_{ik} \mathbf{X}_i' (\mathbf{y}_i - \hat{d}_{ik}) \quad \text{Εξ. 4.19}$$

$$\hat{\sigma}_k^2 = \frac{1}{\sum_i w_{ik} n_i} \sum_i w_{ik} (\|\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k - \hat{d}_{ik}\|^2 + n_i V_{d_{ik}}) \quad \text{Εξ. 4.20}$$

$$\hat{u}_k^2 = \frac{1}{\sum_i w_{ik}} \sum_i w_{ik} [\hat{d}_{ik}^2 + V_{d_{ik}}] \quad \text{Εξ. 4.21}$$

$$\hat{\alpha}_k = \frac{1}{n} \sum_i w_{ik} \quad \text{Εξ. 4.22}$$

4.4.3. Ομαδοποίηση μονοδιάστατων χρονοσειρών

Για να παρουσιάσουμε ένα παράδειγμα της λειτουργίας του μοντέλου μετατόπισης, θα χρησιμοποιήσουμε το παράδειγμα της ενότητας 3.3.3.

Το βασικό μοντέλο παλινδρόμησης έχει ως εξής:

$$\mathbf{y}_i = \mathbf{X} \boldsymbol{\beta}_k + d_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbf{I}) \quad \text{Εξ. 4.23}$$

Με βάση το μοντέλο αυτό οι παραπάνω εξισώσεις του αλγορίθμου EM προσαρμόζονται ως εξής:



Βήμα Υπολογισμού Αναμενόμενης Τιμής (E-step)

Η συνάρτηση Q ως εξής:

$$Q = E[\mathcal{L}_c | \mathbf{y}_i] = \sum_i \sum_k w_{ik} \log a_k \mathcal{N}(\mathbf{y}_i | \mathbf{X} \boldsymbol{\beta}_k, \mathbf{U}_k) \quad \text{Εξ. 4.24}$$

Βήμα Μεγιστοποίησης (M-step)

Στο M-step μεγιστοποιούμε τη συνάρτηση Q ως προς τις παραμέτρους $\{\boldsymbol{\beta}_k, \sigma_k^2, u_k^2, \alpha_k\}$. Οι λύσεις οι οποίες προκύπτουν έχουν ως εξής:

$$\hat{\boldsymbol{\beta}}_k = \left[\sum_i w_{ik} \mathbf{X}' \mathbf{X} \right]^{-1} \sum_i w_{ik} \mathbf{X}' (\mathbf{y}_i - \hat{d}_{ik}) \quad \text{Εξ. 4.25}$$

$$\hat{\sigma}_k^2 = \frac{1}{\sum_i w_{ik} T} \sum_i w_{ik} (\|\mathbf{y}_i - \mathbf{X} \hat{\boldsymbol{\beta}}_k - \hat{d}_{ik}\|^2 + TV_{d_{ik}}) \quad \text{Εξ. 4.26}$$

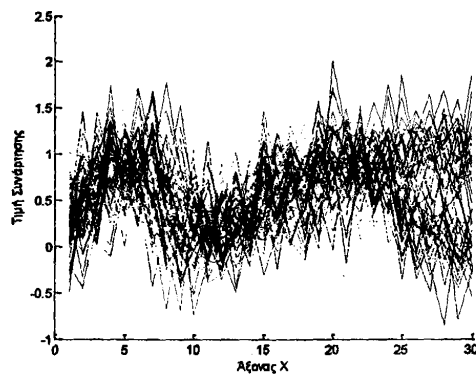
$$\hat{u}_k^2 = \frac{1}{\sum_i w_{ik}} \sum_i w_{ik} [\hat{d}_{ik}^2 + V_{d_{ik}}] \quad \text{Εξ. 4.27}$$

$$\hat{\alpha}_k = \frac{1}{N} \sum_i w_{ik} \quad \text{Εξ. 4.28}$$

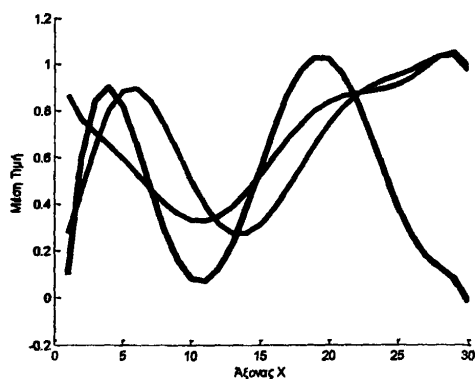
Στο Σχήμα 4.2 παρουσιάζεται ένα παράδειγμα της εκτέλεσης του αλγορίθμου EM για ένα πολυωνυμικό μικτό μοντέλο παλινδρόμησης, για μονοδιάστατες χρονοσειρές. Το σύνολο των δεδομένων αποτελείται από τρεις ομάδες των 30 χρονοσειρών η κάθε μία, παραγόμενες από τρία διαφορετικά πολυώνυμα με την προσθήκη θορύβου. Το μήκος των χρονοσειρών είναι 30.

Το Σχήμα 4.2(α) δείχνει όλες τις χρονοσειρές που παίρνει ως είσοδο ο EM. Οι καμπύλες παρουσιάζονται με διαφορετικά χρώματα για να φανούν οι ομάδες, αλλά αυτή η πληροφορία δεν δίνεται στον αλγόριθμο. Το Σχήμα 4.2(β) δείχνει τις ομάδες οι οποίες προκύπτουν από την εφαρμογή του μοντέλου της μετατόπισης και το Σχήμα 4.2(γ) δείχνει τις ομάδες που προκύπτουν από το απλό μοντέλο ομαδοποίησης, δηλαδή χωρίς μετατόπιση.

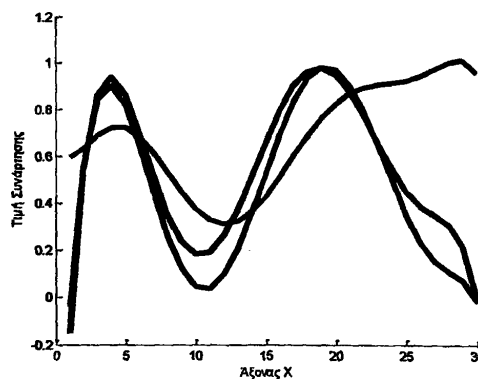




(α) Πραγματικά Δεδομένα



(β) Ομαδοποίηση με το μοντέλο της μετατόπισης



(γ) Ομαδοποίηση χωρίς μετατόπιση

Σχήμα 4.2 Αλγόριθμος EM για μοντέλο μετατόπισης (μονοδιάστατες καμπύλες)

Όπως φαίνεται από το παραπάνω σχήμα, οι ομάδες που προκύπτουν από το μοντέλο της μετατόπισης ανταποκρίνονται καλύτερα στις πραγματικές ομάδες του συνόλου χρονοσειρών. Επίσης μεγάλη διαφορά παρατηρείται και στην τιμή της πιθανοφάνειας μεταξύ των δύο μοντέλων. Η τελική τιμή πιθανοφάνειας στο τελευταίο βήμα του EM, για το μοντέλο της μετατόπισης, είναι 2113.03 ενώ για το απλό μοντέλο είναι 0.59.



ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Εισαγωγή

5.2 Τεχνητά Δεδομένα

5.3 Πραγματικά Δεδομένα

5.1. Εισαγωγή

Η υλοποίηση και των δύο μοντέλων ομαδοποίησης χρονοσειρών, δηλαδή με και χωρίς μετατόπιση, έγινε σε MATLAB. Η αρχικοποίηση και των δύο μοντέλων έγινε με παρόμοιο τρόπο. Αρχικά, επιλέχθηκαν τυχαία K χρονοσειρές από το σύνολο δεδομένων και στη συνέχεια, με τη βοήθεια της μεθόδου των ελαχίστων τετραγώνων, υπολογίσθηκαν οι $K \times (p + 1)$ συντελεστές παλινδρόμησης. Τα βάρη για κάθε ομάδα K θεωρήθηκαν ίσα, δηλαδή $1/K$, ενώ ως διακύμανση θεωρήθηκε η μέση τιμή των τυπικών αποκλίσεων των χρονοσειρών. Για την περαιτέρω προσαρμογή των παραμέτρων αυτών, εκτελέστηκε ένα βήμα του αλγορίθμου EM και υπολογίστηκε η πιθανοφάνεια. Τέλος, ως παράμετροι αρχικοποίησης χρησιμοποιήθηκαν εκείνες που είχαν τη μεγαλύτερη τιμή πιθανοφάνειας μετά από 100 επαναλήψεις των παραπάνω βημάτων αρχικοποίησης.

Για την επιλογή του βαθμού των πολωνύμων, εκτελέστηκαν τα πειράματα για βαθμούς από δύο (2) έως και επτά (7). Η αλλαγή του βαθμού πολωνύμου δεν έδειξε να επηρεάζει τη συμπεριφορά του αλγορίθμου, καθώς οι κινήσεις σε όλα τα βίντεο είναι σχεδόν ευθύγραμμες (με λίγες εξαιρέσεις). Η παράθεση των αποτελεσμάτων στα σχήματα που θα ακολουθήσουν αφορά μια ενδιάμεση κατάσταση, δηλαδή πολώνυμα βαθμού $p = 5$.



Προκειμένου να εκτιμηθεί η αποτελεσματικότητα των μοντέλων και να γίνει η μεταξύ τους σύγκριση, εφαρμόστηκαν σε δύο τύπους δεδομένων. Ο πρώτος τύπος αποτελείται από ένα σύνολο τεχνητών δεδομένων, δηλαδή βίντεο τα οποία έχουν δημιουργηθεί από τεχνητά frames. Ενώ ο δεύτερος τύπος αποτελείται από πραγματικά βίντεο, τα οποία προέρχονται είτε από εσωτερικό είτε από εξωτερικό χώρο. Στην Ενότητα 5.2. περιγράφονται τα τεχνητά δεδομένα και ακολούθως παρατίθενται τα αποτελέσματα της εφαρμογής του μοντέλου ομαδοποίησης με μετατόπιση όπως επίσης και τα αποτελέσματα από τη σύγκριση με το απλό μοντέλο ομαδοποίησης. Ενώ στη Ενότητα 5.3. περιγράφονται τα πραγματικά δεδομένα και στη συνέχεια παρουσιάζονται τα αποτελέσματα της εφαρμογής του μοντέλου ομαδοποίησης με μετατόπιση.

5.2. Τεχνητά Δεδομένα

Η μέθοδος εφαρμόστηκε σε έξι τεχνητά βίντεο, τα οποία δημιουργήθηκαν από συγκεκριμένες ακολουθίες εικόνων. Για το κάθε βίντεο χρησιμοποιήθηκε το πρόγραμμα ζωγραφικής MsPaint για τη δημιουργία διαδοχικών εικόνων (frames), συγκεκριμένων διαστάσεων, με κατάληξη .jpg, οι οποίες δίνουν την αίσθηση την ενιαίας κίνησης.

Αρχικά, ως κατώφλι για την απομάκρυνση των περιττών χρονοσειρών (background), χρησιμοποιήθηκε ο μέσος όρος της διακύμανσης των χρονοσειρών. Επίσης, ο αριθμός των ομάδων θεωρείται ότι είναι γνωστός εξ αρχής.

Για τη μέτρηση της απόδοσης του μοντέλου ομαδοποίησης, η μετρική η οποία χρησιμοποιήθηκε είναι το ποσοστό της ορθής ταξινόμησης (*accuracy*). Όπως έχουμε ήδη αναφέρει και στις προηγούμενες ενότητες, το σύνολο των χρονοσειρών που αποτελούν είσοδο του αλγορίθμου, επηρεάζεται από το κατώφλι (*threshold*), με βάση το οποίο γίνεται η επιλογή τους. Για να δούμε πόσο μεγάλη είναι η επιρροή αυτή, για διαφορετικές τιμές κατωφλίου, εφαρμόσαμε το μοντέλο μας δέκα φορές για κάθε τεχνητό βίντεο και υπολογίσαμε το μέσο όρο του ποσοστού ορθής ταξινόμησης για τις επαναλήψεις αυτές, καθώς και το μέσο όρο της πιθανοφάνειας.

Για τον υπολογισμό της ακρίβειας χρησιμοποιήθηκε ο παρακάτω πίνακας:



		Κατηγορίες σύμφωνα με τη μέθοδο			
Πραγματικές κατηγορίες		1	...	...	K
	1	A_{11}			
	...				
	...		A_{ij}		
	K				

όπου K είναι ο αριθμός των ομάδων. Στη συνέχεια αθροίζονται τα μέγιστα από κάθε γραμμή και υπολογίζεται η ακρίβεια ως εξής:

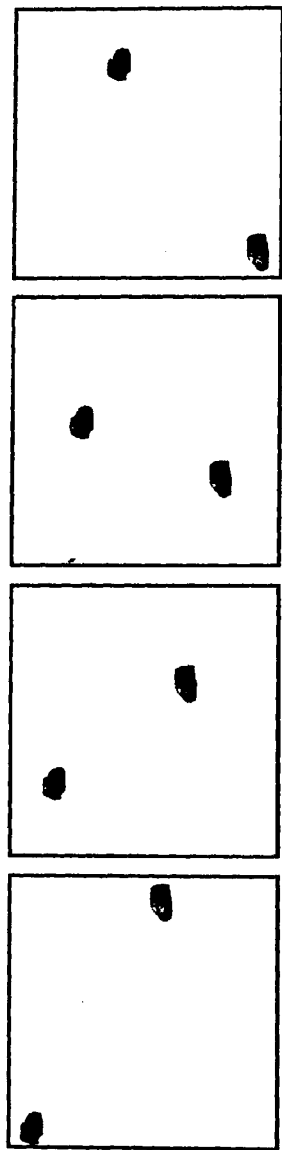
$$accuracy = \frac{\sum_i^K \max_{j=1..K} A_{ij}}{N} \quad \text{Εξ. 5.1}$$

όπου N είναι ο πραγματικός αριθμός των pixels των αντικειμένων.

Στα Σχήματα που ακολουθούν (Σχήμα 5.1 έως 5.8) παρατίθενται τα αποτελέσματα της εφαρμογής του μοντέλου ομαδοποίησης για τα τέσσερα από τα τεχνητά βίντεο. Το 1^ο βίντεο απεικονίζει δύο αυτοκίνητα τα οποία κινούνται ευθύγραμμα με αντίθετη φορά και αποτελείται από 15 frames, με διαστάσεις 300×300 pixels. Το 2^ο βίντεο απεικονίζει δύο μπάλες που κινούνται αντίθετα η μία από την άλλη και αποτελείται από 17 frames, με διαστάσεις 100×100 pixels. Το 3^ο βίντεο απεικονίζει ένα σαλιγκάρι και δύο πεταλούδες που κινούνται από αριστερά προς τα δεξιά αλλά όχι ευθύγραμμα και αποτελείται από 15 frames με διαστάσεις 300×300 pixels. Τέλος, το 4^ο βίντεο απεικονίζει τέσσερα αεροπλάνα τα οποία ξεκινώντας από το κέντρο κινούνται ευθύγραμμα προς τις τέσσερις γωνίες και αποτελείται από 18 frames με διαστάσεις 300×300 pixels. Στα δύο πρώτα βίντεο οι ομάδες των αντικειμένων είναι δύο, ενώ στα επόμενα δύο είναι τρεις και τέσσερις ομάδες αντίστοιχα. Για κάθε βίντεο παρουσιάζονται ενδεικτικά τέσσερα frames και στη συνέχεια τα αποτελέσματα ομαδοποίησης συνολικά για τις δύο διαστάσεις X και Y και τα επιμέρους αποτελέσματα για κάθε μια διάσταση ξεχωριστά.

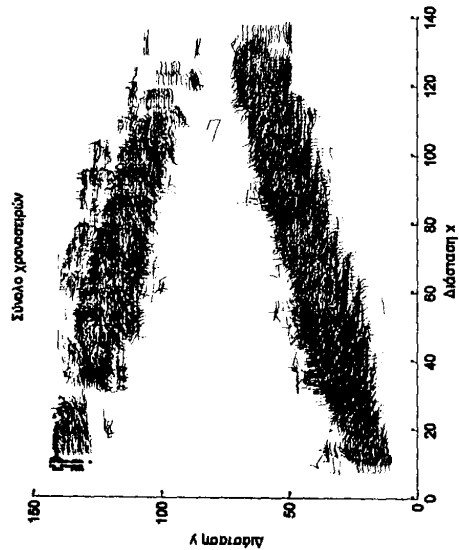


1ο βίντεο (Δύο αυτοκίνητα τα οποία προχωράνε αντίθετα)

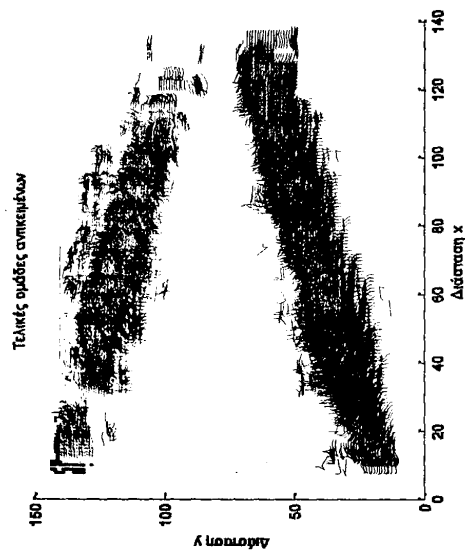


frame 1 frame 5 frame 10 frame 15

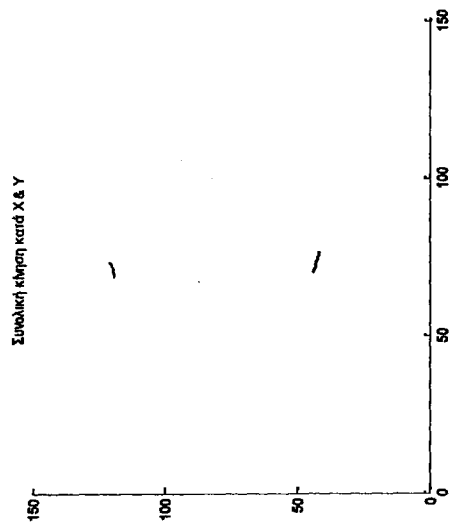
Σχήμα 5.1 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 1



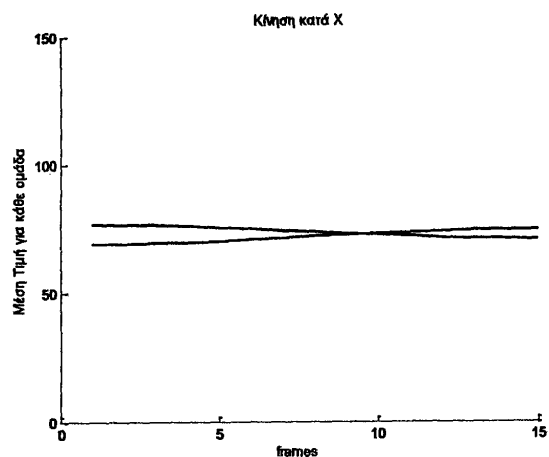
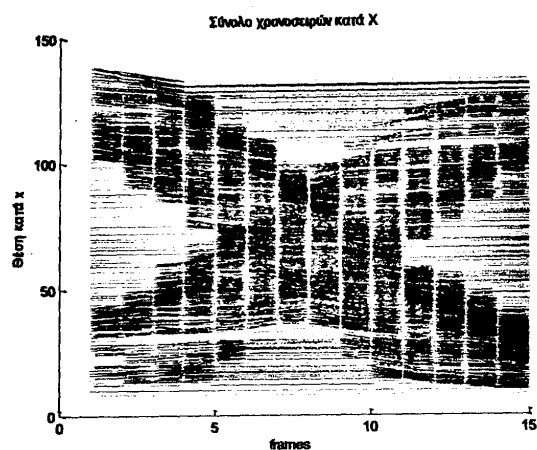
(α) Σύνολο χρονοσειρών



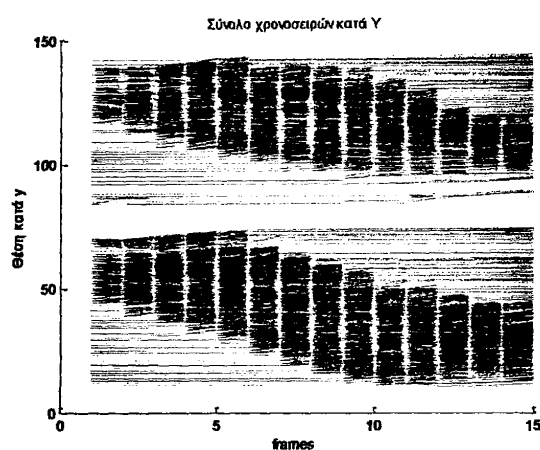
(β) Τελικές ομάδες αντικειμένων



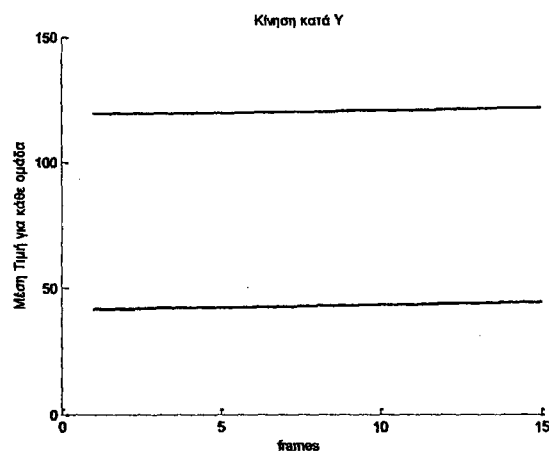
(γ) Συνολική κίνηση κατά X και Y



(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X

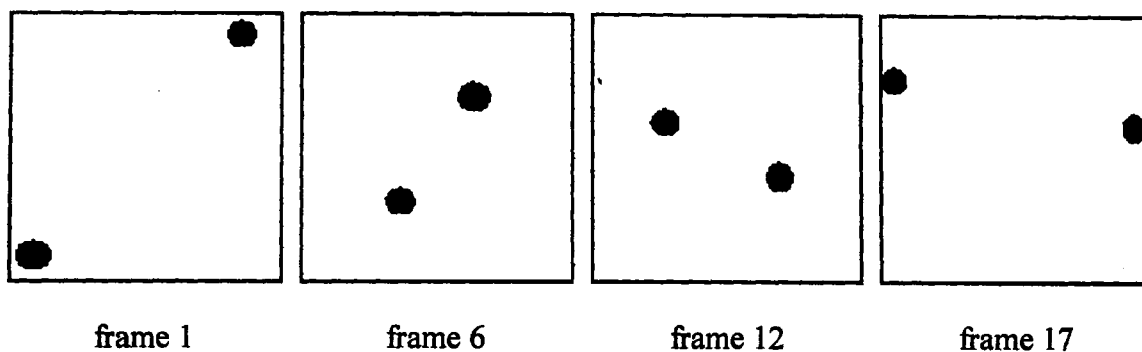


(στ) Σύνολο χρονοσειρών κατά Y

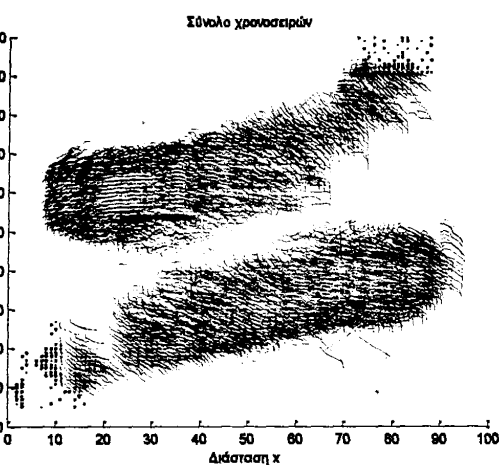
(ζ) Κίνηση κατά Y

Σχήμα 5.2 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 1

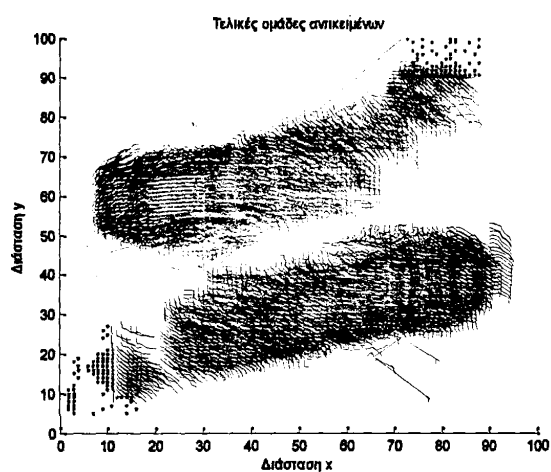
βίντεο (Δύο μπάλες οι οποίες προχωράνε αντίθετα)



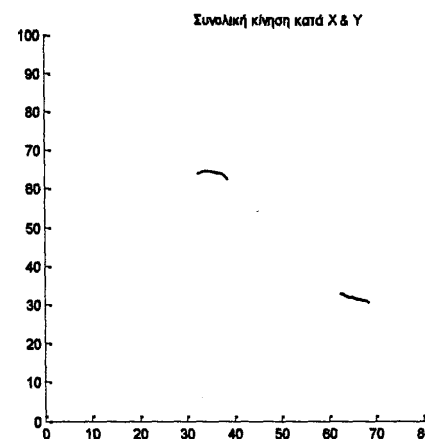
Σχήμα 5.3 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 2



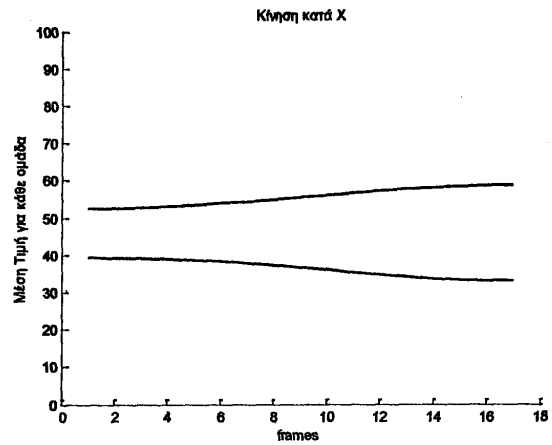
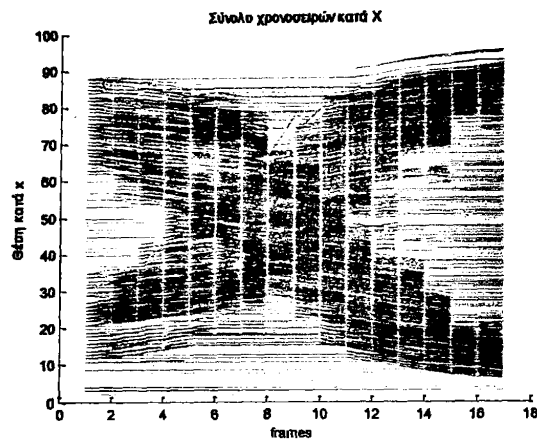
(α) Σύνολο χρονοσειρών



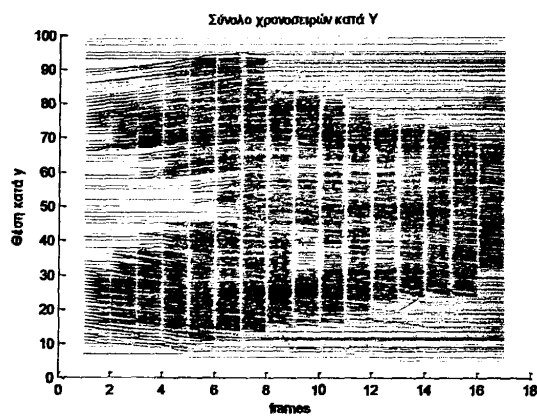
(β) Τελικές ομάδες αντικειμένων



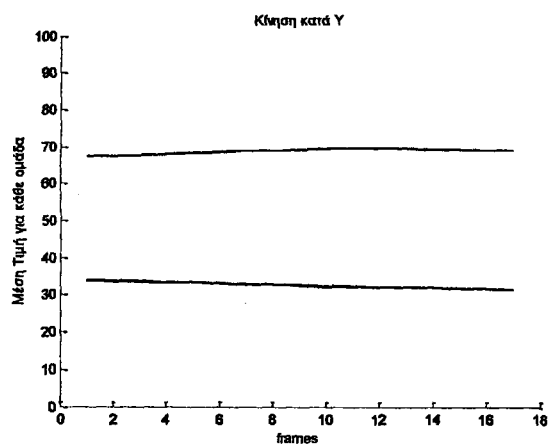
(γ) Συνολική κίνηση κατά X & Y



(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X

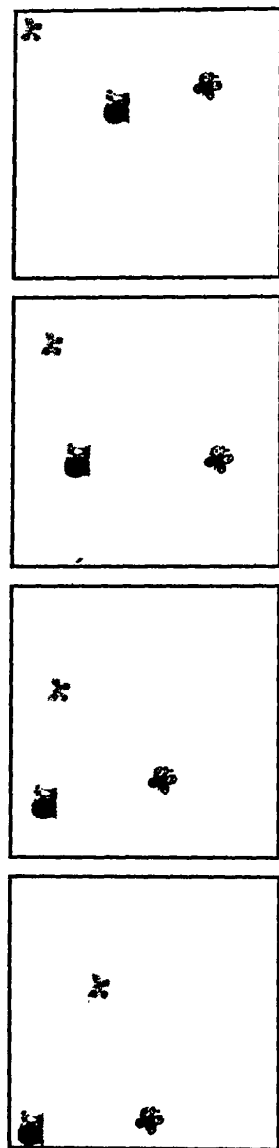


(στ) Σύνολο χρονοσειρών κατά Y

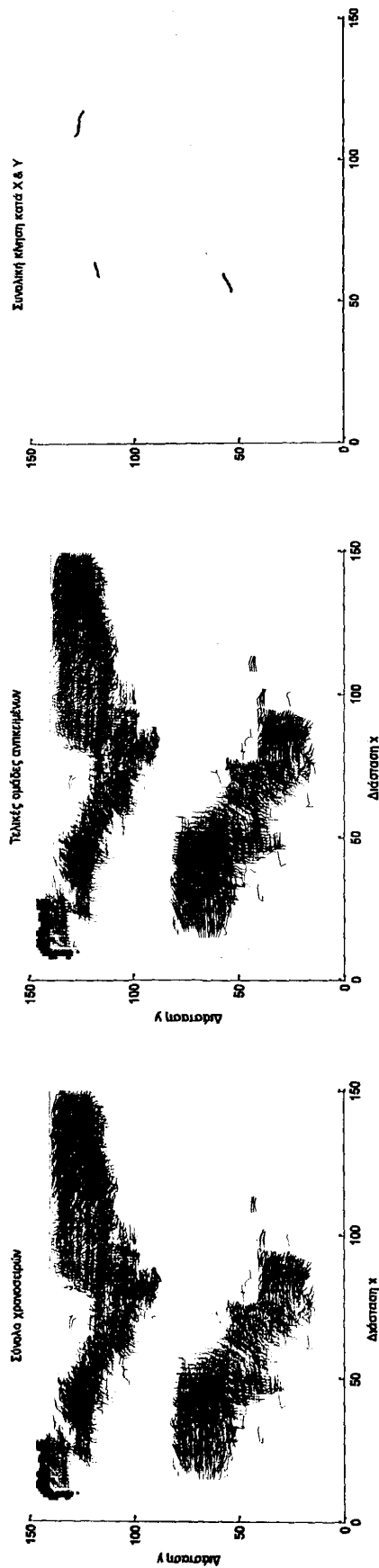
(ζ) Κίνηση κατά Y

Σχήμα 5.4 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 2

3ο βίντεο (Κίνηση δύο πεταλούδων και ενός σαλιγκαριού)



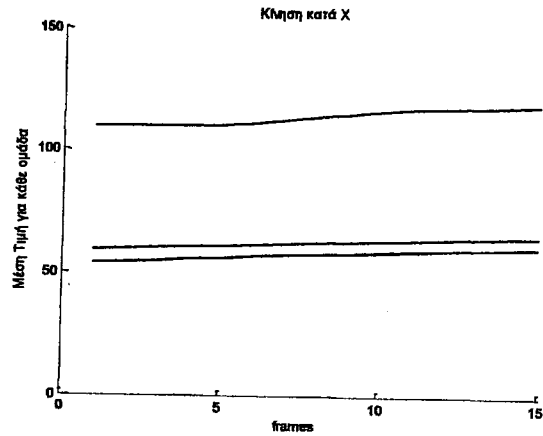
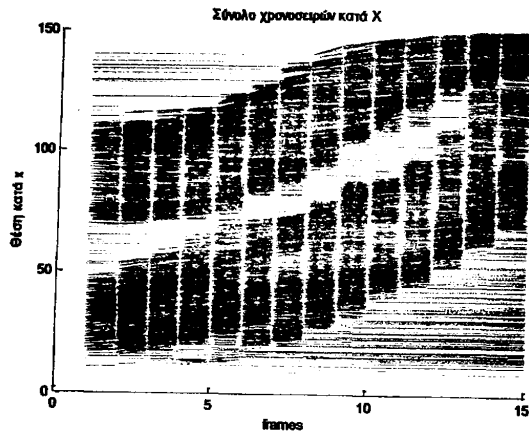
frame 1 frame 5 frame 10 frame 15
Σχήμα 5.5 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 3



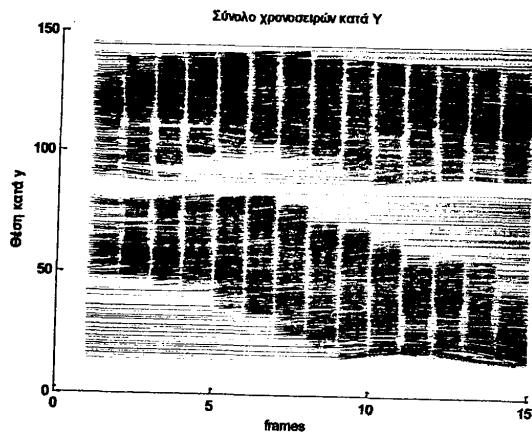
(α) Σύνολο χρονοσειρών

(β) Τελικές ομάδες αντικειμένων

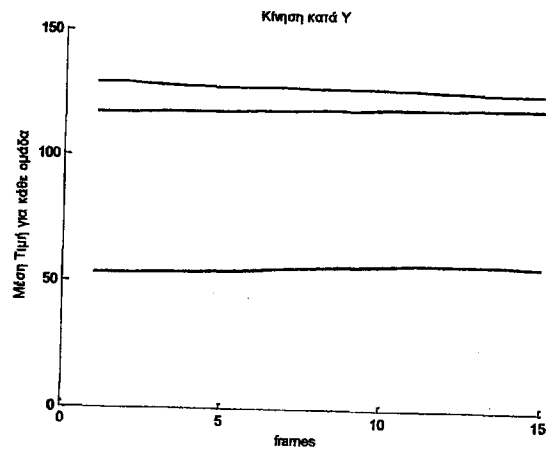
(γ) Συνολική Κίνηση κατά X και Y



(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X

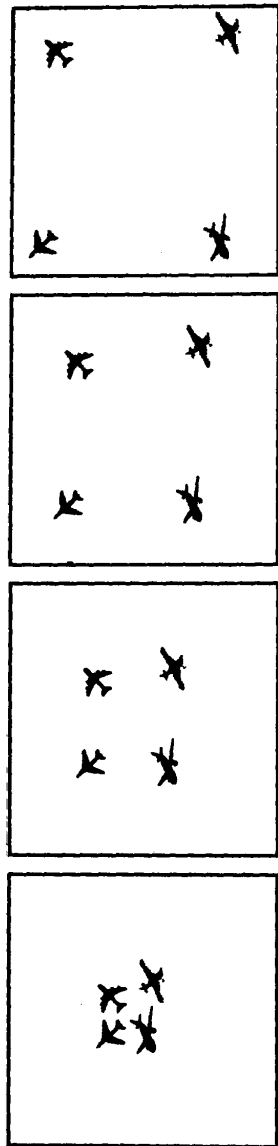


(στ) Σύνολο χρονοσειρών κατά Y

(ζ) Κίνηση κατά Y

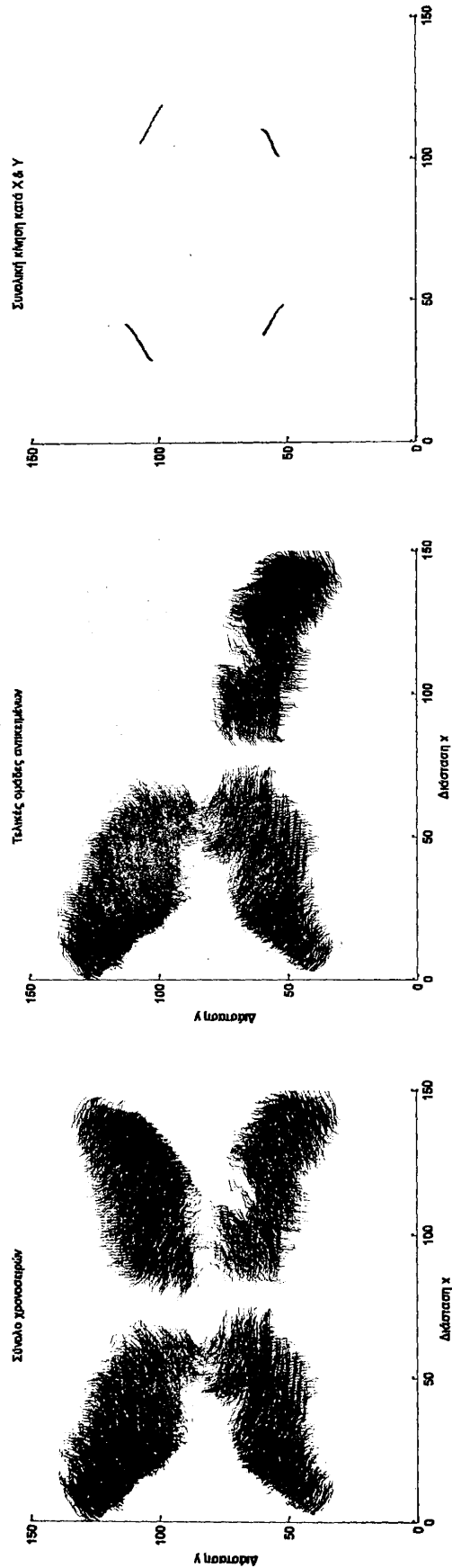
Σχήμα 5.6 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 3

4ο βίντεο (Κίνηση τεσσάρων αεροπλάνων)



frame 1 frame 6 frame 12 frame 18

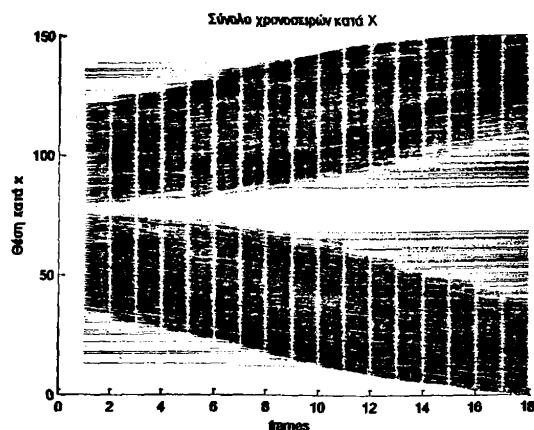
Σχήμα 5.7 Τέσσερα ενδεικτικά frames από το τεχνητό βίντεο 4



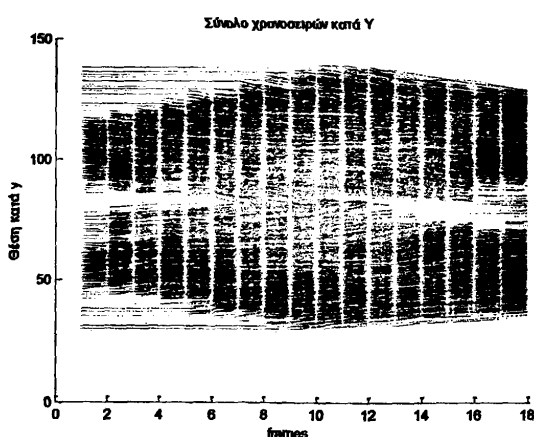
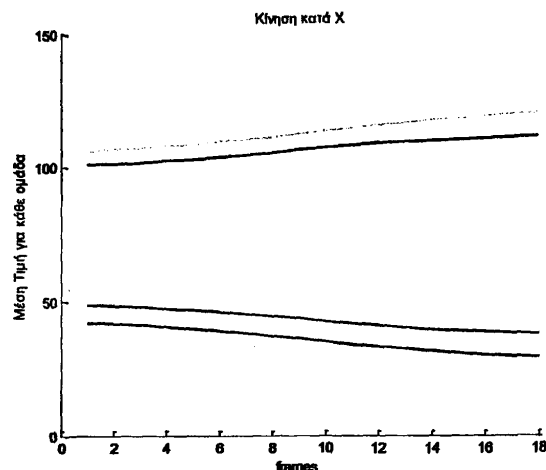
(α) Σύνολο χρονοσειρών

(β) Τελικές ομάδες αντικειμένων

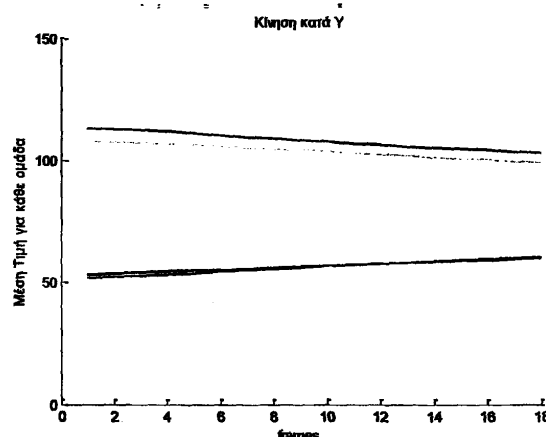
(γ) Συνολική Κίνηση κατά X και Y



(δ) Σύνολο χρονοσειρών κατά X



(στ) Σύνολο χρονοσειρών κατά Y



(ζ) Κίνηση κατά Y

Σχήμα 5.8 Ομαδοποίηση αντικειμένων του τεχνητού βίντεο 4

Όπως φαίνεται από τα παραπάνω σχήματα, το μοντέλο ομαδοποίησης με μετατόπιση βρίσκει σε κάθε ένα από τα βίντεο τις υποκείμενες ομάδες. Η συνολική κίνηση κατά X και Y δεν αντιστοιχεί στην πραγματική κίνηση των αντικειμένων, αλλά χαρακτηρίζει τη μέση κίνηση των χρονοσειρών που απαρτίζουν το κάθε αντικείμενο. Η κίνηση των αντικειμένων αναπαριστάται καλύτερα με τα διαγράμματα που παρουσιάζουν ξεχωριστά την κίνηση για κάθε διάσταση, δηλαδή κατά X και κατά Y.

Υπολογισμός Ποσοστού Ορθής Ταξινόμησης

Για τον υπολογισμό του ποσοστού ορθής ταξινόμησης χρησιμοποιήθηκαν ο πίνακας ορθής ταξινόμησης και η Εξίσωση 5.1, όπως αυτά περιγράφηκαν στην αρχή της ενότητας 5.2.

Ως αρχική τιμή για το κατώφλι (T) θεωρούμε το μέσο όρο της διακύμανσης των χρονοσειρών και σταδιακά το αυξάνουμε μέχρι και $3.5 \cdot T$ (λαμβάνοντας επίσης ως τιμή και το $0.5 \cdot T$). Για τις επτά διαφορετικές τιμές κατωφλίου εφαρμόστηκε το μοντέλο της ομαδοποίησης με μετατόπιση δέκα φορές σε κάθε ένα από τα τεχνητά βίντεο, τα οποία παρουσιάστηκαν παραπάνω. Από τις δέκα αυτές επαναλήψεις υπολογίστηκε ο μέσος όρος του ποσοστού ορθής ταξινόμησης, καθώς επίσης και ο μέσος όρος της πιθανοφάνειας. Επίσης, σημειώνεται και το ποσοστό των πραγματικών pixels που απαρτίζουν τα αντικείμενα σε κάθε βίντεο σε σχέση με το σύνολο των χρονοσειρών που αποτελούν είσοδο του αλγορίθμου.

Τα αποτελέσματα φαίνονται στους πίνακες που ακολουθούν:

1ο βίντεο (Δύο αυτοκίνητα τα οποία προχωράνε αντίθετα)

Πίνακας 5.1 Ακρίβεια μεθόδου για 1^ο βίντεο

Threshold	Accuracy (%)	Loglikelihood
0.7815 ($0.5 \cdot T$)	100	-1.9043e+006
1.5630 (T)	100	-1.5242e+006
2.3446 ($1.5 \cdot T$)	100	-1.2248e+006
3.1261 ($2 \cdot T$)	100	-1.0017e+006
3.9076 ($2.5 \cdot T$)	100	-0.8336e+006
4.6891 ($3 \cdot T$)	100	-0.7031e+006
5.4706 ($3.5 \cdot T$)	100	-0.5992e+006



2ο βίντεο (Δύο μπάλες οι οποίες προχωράνε αντίθετα)

Πίνακας 5.2 Ακρίβεια μεθόδου για 2ο βίντεο

Threshold	Accuracy (%)	Loglikelihood
1.3884 (0.5*T)	100	-10.230e+005
2.7767 (T)	100	-8.6364e+005
4.1653 (1.5*T)	100	-7.2460e+005
5.5537 (2*T)	100	-6.2796e+005
6.9421 (2.5*T)	100	-5.2419e+005
8.3306 (3*T)	100	-4.2675e+005
9.7190 (3.5*T)	100	-3.1018e+005

3ο βίντεο (Κίνηση δύο πεταλούδων και ενός σαλιγκαριού)

Πίνακας 5.3 Ακρίβεια μεθόδου για 3ο βίντεο

Threshold	Accuracy (%)	Loglikelihood
1.1424 (0.5*T)	98.20	-1.8220e+006
2.2848 (T)	99.37	-1.3640e+006
3.4272 (1.5*T)	97.69	-1.0741e+006
4.5696 (2*T)	92.19	-0.8724e+006
5.7120 (2.5*T)	89.05	-0.6982e+006
6.8544 (3*T)	84.38	-0.5730e+006
7.9968 (3.5*T)	88.16	-0.4865e+006

4ο βίντεο (Κίνηση τεσσάρων αεροπλάνων)

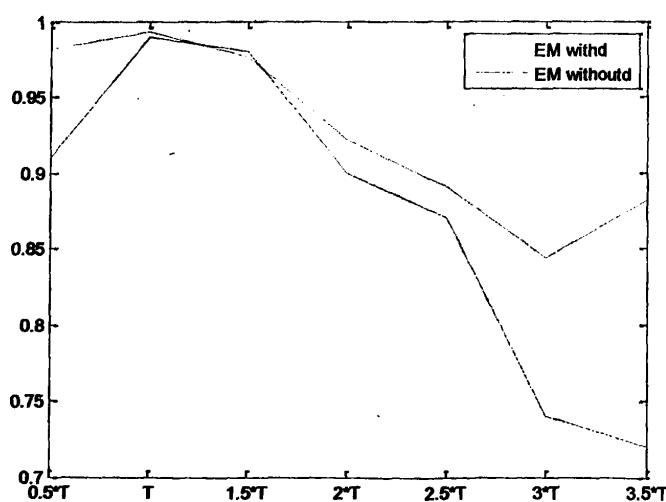
Πίνακας 5.4 Ακρίβεια μεθόδου για 4ο βίντεο

Threshold	Accuracy (%)	Loglikelihood
3.4624 (0.5*T)	85.06	-2.4074e+006
6.9247 (T)	82.18	-1.9033e+006
10.3871 (1.5*T)	84.63	-1.5400e+006
13.8495 (2*T)	86.98	-1.2328e+006
17.3118 (2.5*T)	75.90	-0.9870e+006
20.7742 (3*T)	64.18	-0.7823e+006
24.2366 (3.5*T)	60.79	-0.6063e+006



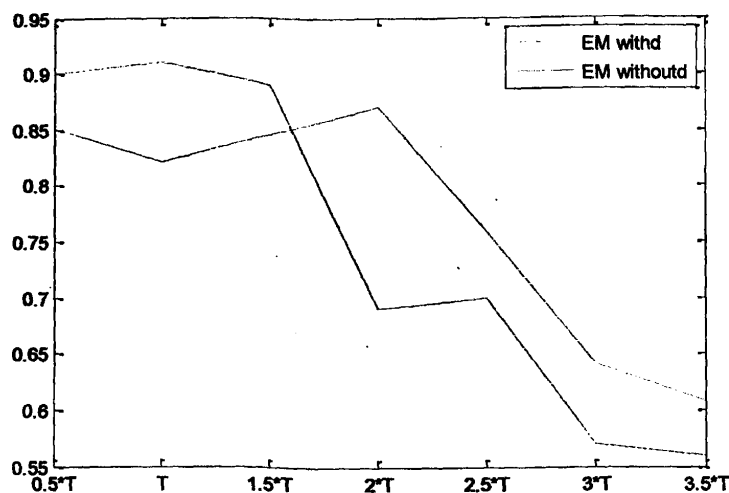
Όπως φαίνεται και από τα αποτελέσματα στους παραπάνω πίνακες, όταν έχουμε λίγα αντικείμενα (1^ο και 2^ο βίντεο) το κατώφλι φαίνεται να μην επηρεάζει την ακρίβεια του μοντέλου καθώς πάντα βρίσκει τις υποκείμενες ομάδες. Η ακρίβεια αρχίζει και μειώνεται όταν αυξάνεται ο αριθμός των ομάδων, όπως επίσης και το κατώφλι, με βάση το οποίο αφαιρούνται χρονοσειρές από το σύνολο δεδομένων εισόδου.

Στη συνέχεια, για να δούμε την ακρίβεια και του απλού μοντέλου χωρίς μετατόπιση, κάναμε το ίδιο πείραμα, με τα ίδια βίντεο και υπολογίσαμε το μέσο όρο για δέκα επαναλήψεις, για κάθε μια από τις επτά διαφορετικές τιμές του κατωφλίου. Στα παρακάτω σχήματα παρουσιάζονται τα αποτελέσματα (για το 3^ο και 4^ο βίντεο) από το σύγκριση του μοντέλου ομαδοποίησης με μετατόπιση με εκείνο χωρίς μετατόπιση. Για το 1^ο και 2^ο βίντεο η ακρίβεια είναι η ίδια και για τα δύο μοντέλα, καθώς ο αριθμός των αντικειμένων είναι μικρός οπότε και τα δύο μοντέλα έχουν παρόμοια συμπεριφορά. Για το λόγο αυτό δεν παρουσιάζουμε κάποιο σχήμα.



(α) 3^ο βίντεο





(β) 4° βίντεο

Σχήμα 5.9 Σύγκριση ακρίβειας

Όπως φαίνεται στο Σχήμα 5.9(α), το οποίο αναφέρεται στο 3^ο βίντεο, το ποσοστό ορθής ταξινόμησης είναι καλύτερο στο μοντέλο ομαδοποίησης με μετατόπιση. Με εξαίρεση στις τιμές κατωφλίου T και $1.5 \cdot T$ όπου τα δύο μοντέλα παρουσιάζουν παρόμοια συμπεριφορά. Στο Σχήμα 5.9(β), το οποίο αναφέρεται στο 4^ο βίντεο, αρχικά, δηλαδή για τιμές κατωφλίου από $0.5 \cdot T$ έως $1.5 \cdot T$, το μοντέλο ομαδοποίησης χωρίς μετατόπιση φαίνεται να δίνει καλύτερα αποτελέσματα, αλλά όσο αυξάνεται η τιμή του κατωφλίου το μοντέλο με τη μετατόπιση παρουσιάζει καλύτερη συμπεριφορά. Και στα δύο σχήματα φαίνεται ότι το μοντέλο με τη μετατόπιση παρουσιάζει καλύτερη συμπεριφορά καθώς αυξάνεται το κατώφλι, γεγονός το οποίο δείχνει το πώς η μετατόπιση μπορεί να οδηγήσει στην ευθυγράμμιση των χρονοσειρών που συγκροτούν το αντικείμενο και στη συνέχεια στη σωστή ομαδοποίησή τους, κάτι το οποίο αδυνατεί να κάνει το απλό μοντέλο, γι' αυτό καθώς μειώνεται ο αριθμός των χρονοσειρών (αυξάνεται δηλαδή το κατώφλι) μειώνεται και το ποσοστό ορθής ταξινόμησης.

5.3. Πραγματικά Δεδομένα

Η μέθοδος μας εφαρμόστηκε σε δέκα πραγματικά βίντεο, επτά από εσωτερικό χώρο και τρία από εξωτερικό. Τα τρία από αυτά είναι από το χώρο του πανεπιστημίου ενώ τα υπόλοιπα υπάρχουν διαθέσιμα σε ιστοσελίδες του διαδικτύου [36] και [37]. Ως κατώφλι, για την απομάκρυνση των περιττών χρονοσειρών (background), χρησιμοποιήθηκε ο μέσος όρος της διακύμανσης των χρονοσειρών. Ο αριθμός των ομάδων εκτιμήθηκε με βάση την παρατήρηση.

Στη συνέχεια παρατίθενται τα αποτελέσματα της εφαρμογής του μοντέλου ομαδοποίησης για τα τέσσερα από τα πραγματικά βίντεο. Το 1^ο βίντεο απεικονίζει δύο άτομα μέσα σε ένα δωμάτιο τα οποία κινούνται αντίθετα και αποτελείται από 23 frames, με διαστάσεις 720×576 pixels. Το 2^ο βίντεο απεικονίζει δύο άτομα μέσα στο μετρό τα οποία κινούνται σχεδόν παράλληλα και αποτελείται από 27 frames, με διαστάσεις 720×576 pixels. Το 3^ο βίντεο απεικονίζει τρία άτομα σε εξωτερικό χώρο τα οποία κινούνται σχεδόν προς την ίδια κατεύθυνση και αποτελείται από 25 frames με διαστάσεις 384×288 pixels. Τέλος, το 4^ο βίντεο απεικονίζει δύο άτομα που περπατούν, ένα άτομο με ποδήλατο και ένα αυτοκίνητο και αποτελείται από 28 frames με διαστάσεις 384×288 pixels.

Λόγω της μεγάλης πολυπλοκότητας που συνεπάγεται ο υπολογισμός των διανυσμάτων κίνησης για κάθε pixel των frames του βίντεο, μειώθηκαν οι αρχικές διαστάσεις των frames, εφαρμόζοντας το μοντέλο ομαδοποίησης σε frames ανάλυσης 30%, όσον αφορά τα δύο πρώτα βίντεο και 50% για το τρίτο και τέταρτο βίντεο. Το γεγονός αυτό αποτελεί και το βασικό μειονέκτημα της μεθόδου, καθώς η υψηλή ανάλυση οδηγεί στον υπολογισμό μεγάλου όγκου διανυσμάτων κίνησης και ως εκ τούτου αυξάνει την πολυπλοκότητα της μεθόδου.

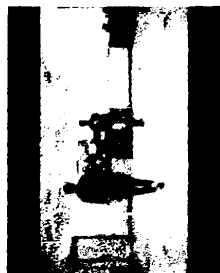
Στα Σχήματα που ακολουθούν (Σχήμα 5.10 έως 5.17) παρουσιάζονται ενδεικτικά τέσσερα frames και στη συνέχεια τα αποτελέσματα ομαδοποίησης συνολικά για τις δύο διαστάσεις X και Y και τα επιμέρους αποτελέσματα για κάθε μια διάσταση ξεχωριστά.



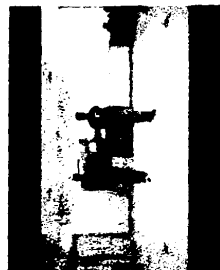
1ο βίντεο (Δύο άνθρωποι κινούνται αντίθετα)



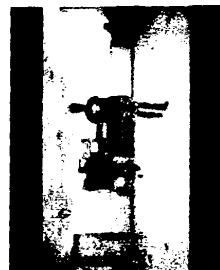
frame 1



frame 7

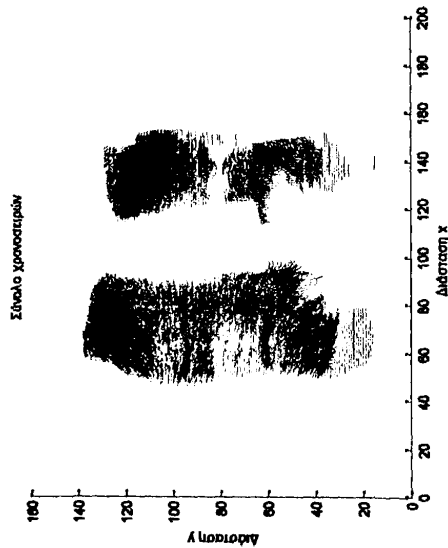


frame 16

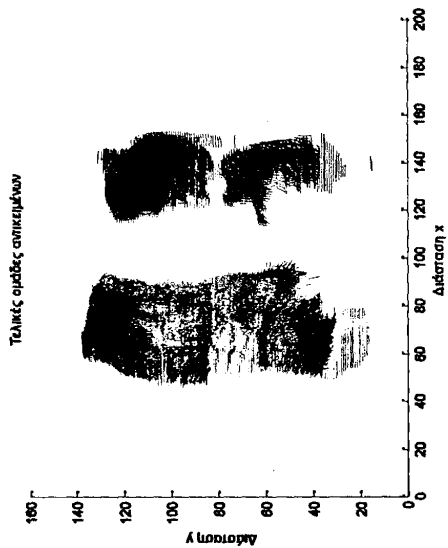


frame 23

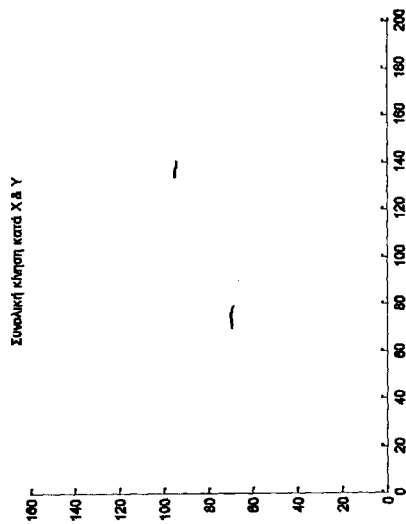
Σχήμα 5.10 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 1



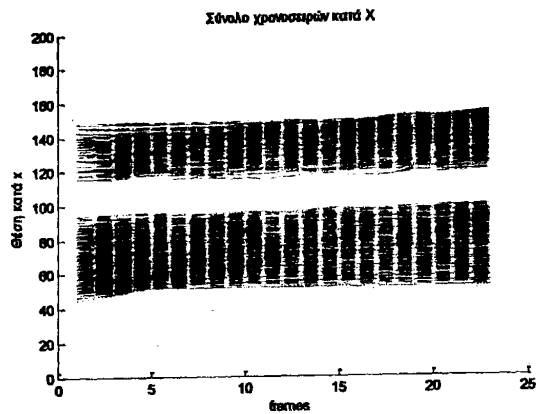
(α) Σύνολο χρονοσειρών



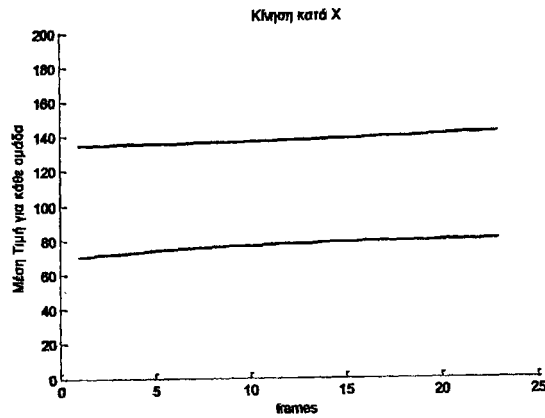
(β) Τελικές ομάδες αντικειμένων



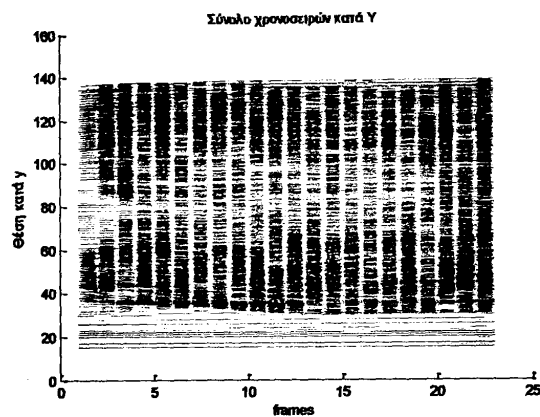
(γ) Συνολική κίνηση κατά X και Y



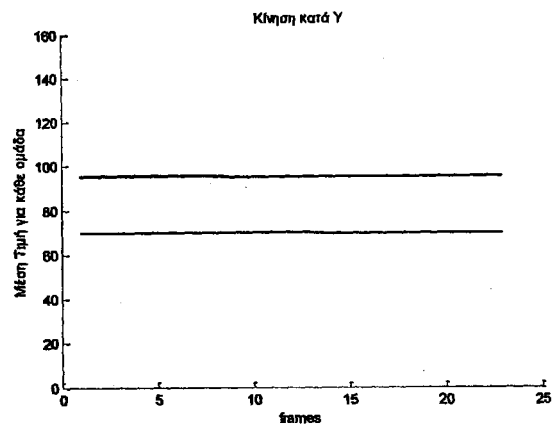
(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X



(στ) Σύνολο χρονοσειρών κατά Y



(ζ) Κίνηση κατά Y

Σχήμα 5.11 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 1



2ο βίντεο (Δύο άνθρωποι μέσα στο metro)



frame 1



frame 9

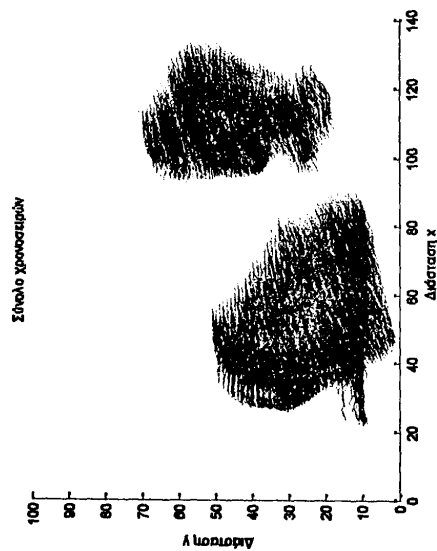


frame 18

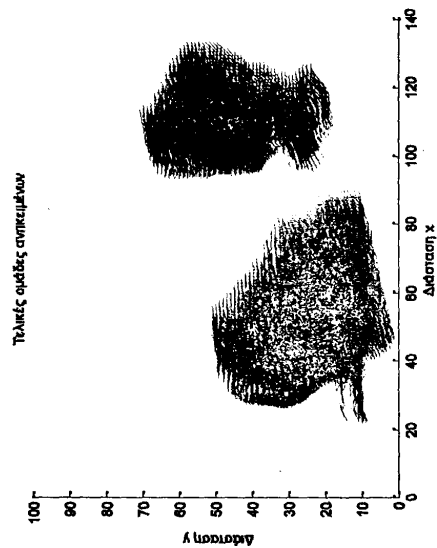


frame 27

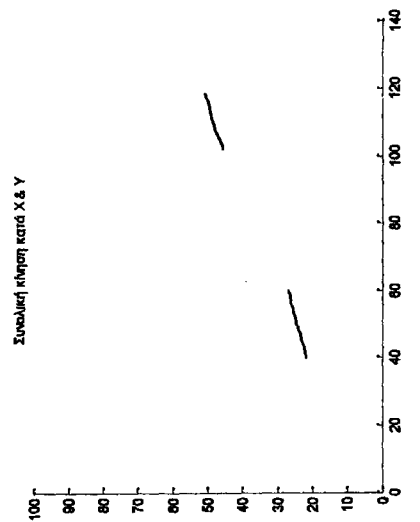
Σχήμα 5.12 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 2



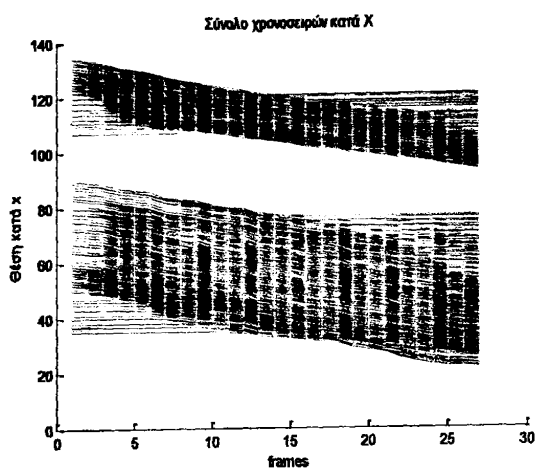
(α) Σύνολο χρονοσειρών



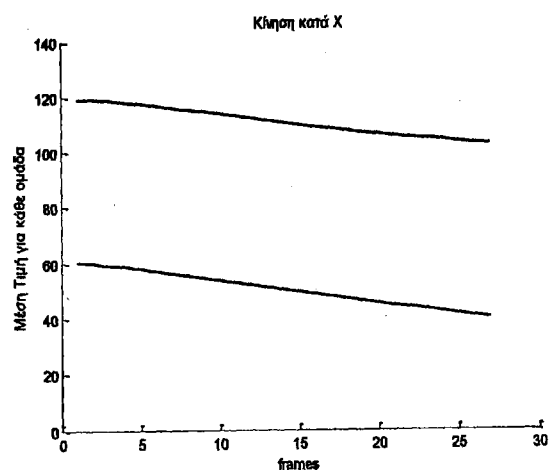
(β) Τελικές ομάδες αντικειμένων



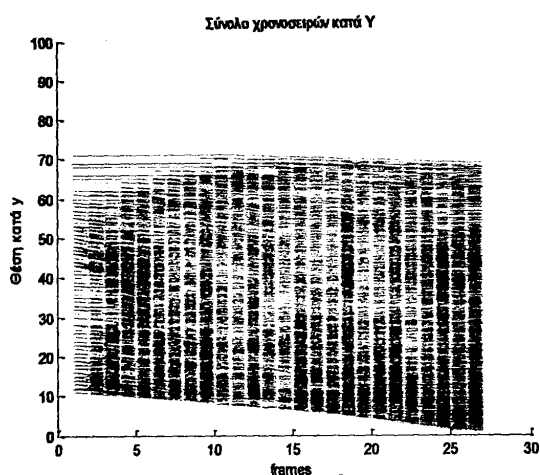
(γ) Συνολική κίνηση κατά X και Y



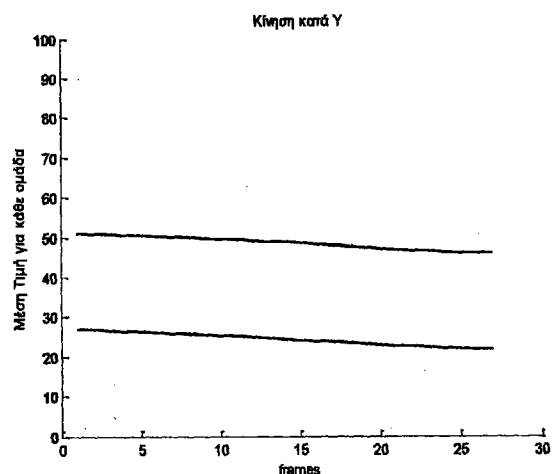
(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X



(στ) Σύνολο χρονοσειρών κατά Y



(ζ) Κίνηση κατά Y

Σχήμα 5.13 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 2



3ο βίντεο (Τρεις άνθρωποι κινούνται σε εξωτερικό χώρο)



frame 1

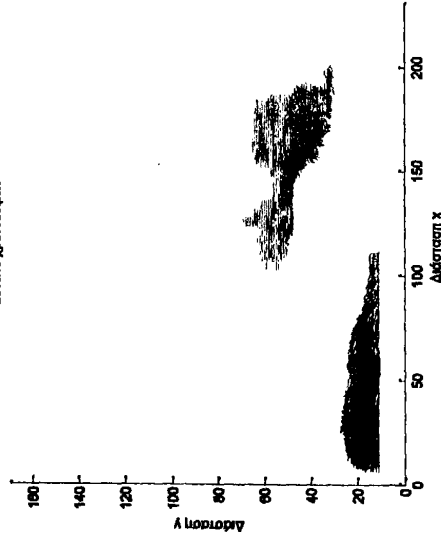
frame 10

frame 18

frame 25

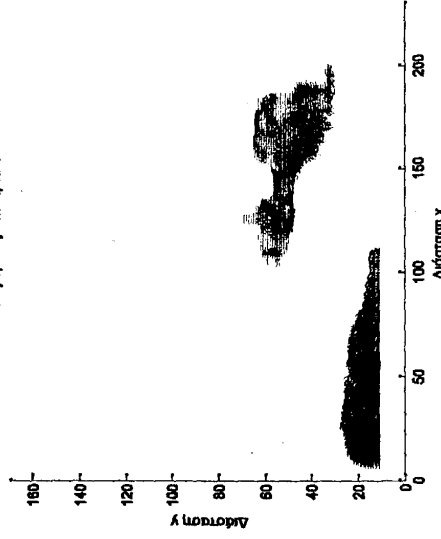
Σχήμα 5.14 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 3

Σύνολο χρονοσειρών



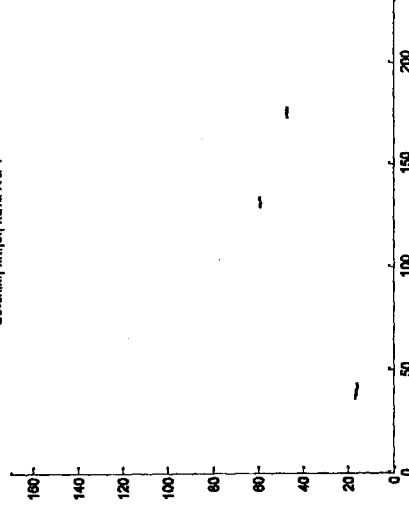
(α) Σύνολο χρονοσειρών

Τελικές ομάδες αντικειμένων

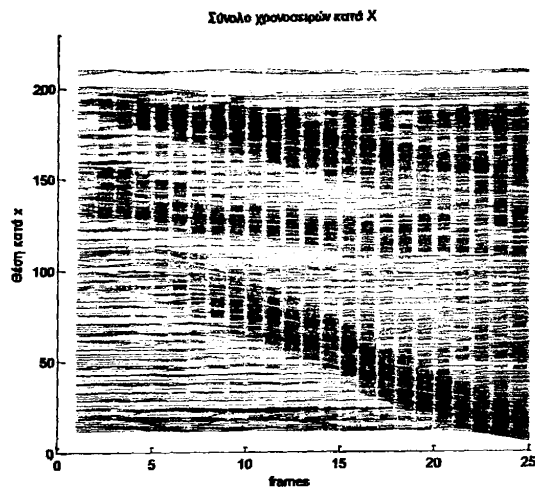


(β) Τελικές ομάδες αντικειμένων

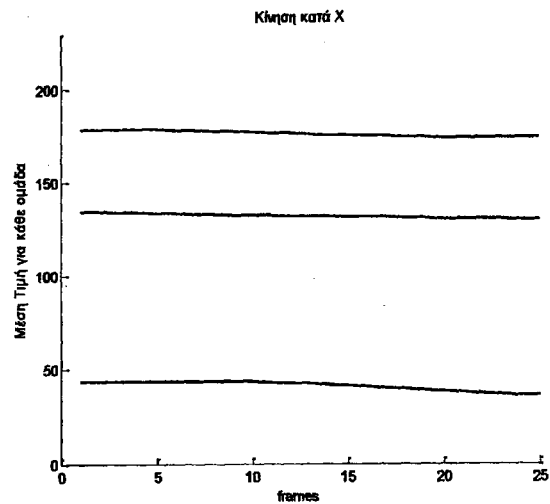
Συνολική κίνηση κατά X & Y



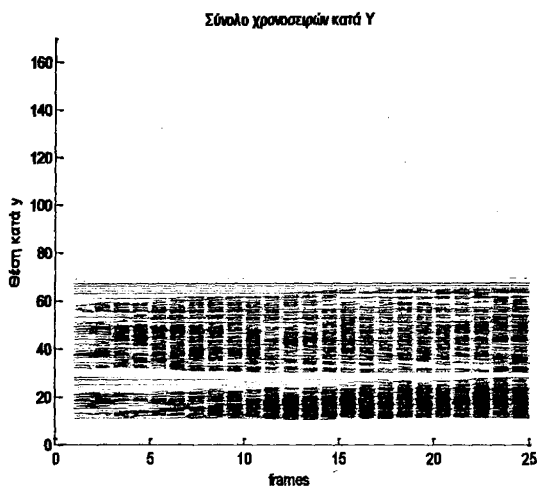
(γ) Συνολική κίνηση κατά X και Y



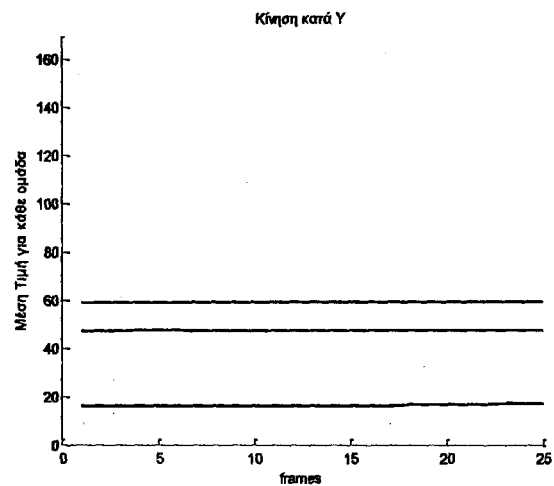
(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X



(στ) Σύνολο χρονοσειρών κατά Y

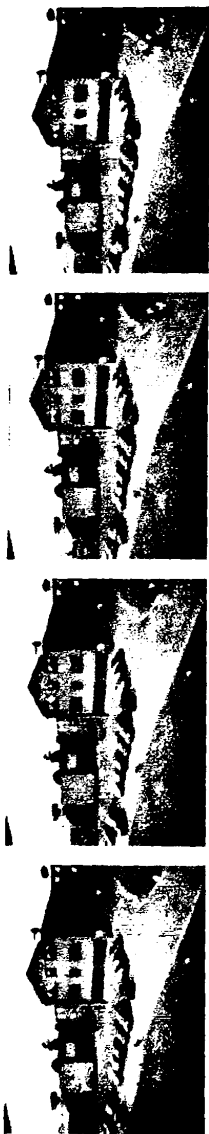


(ζ) Κίνηση κατά Y

Σχήμα 5.15 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 3



4ο βίντεο (Δύο άνθρωποι, ένα ποδήλατο και ένα αυτοκίνητο)



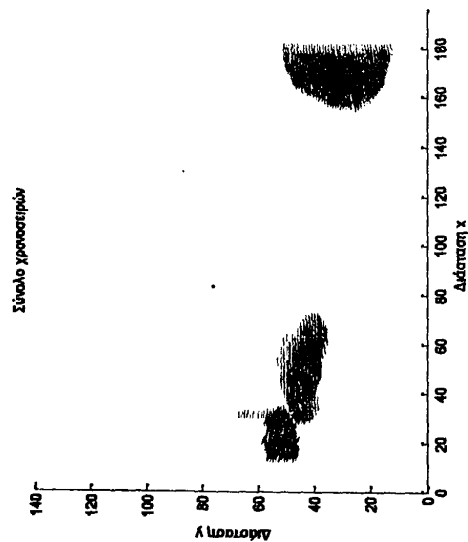
frame 1

frame 10

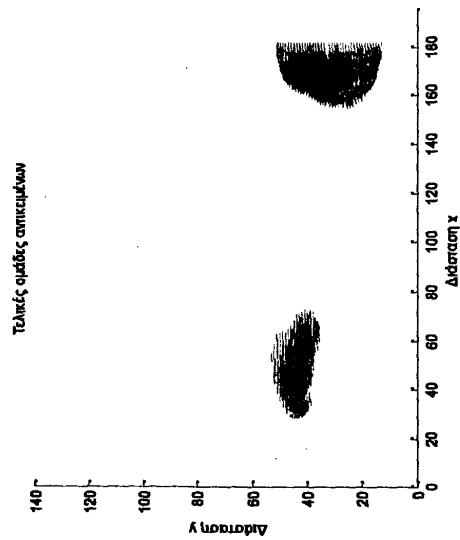
frame 19

frame 28

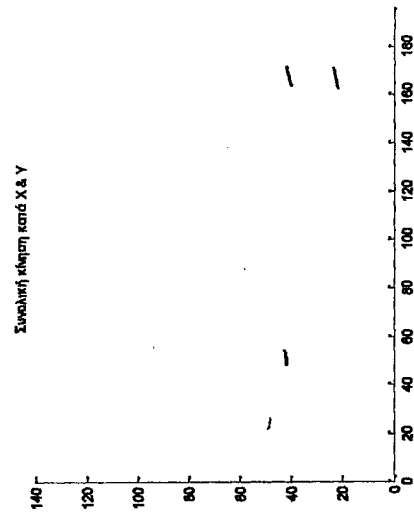
Σχήμα 5.16 Τέσσερα ενδεικτικά frames από το πραγματικό βίντεο 4



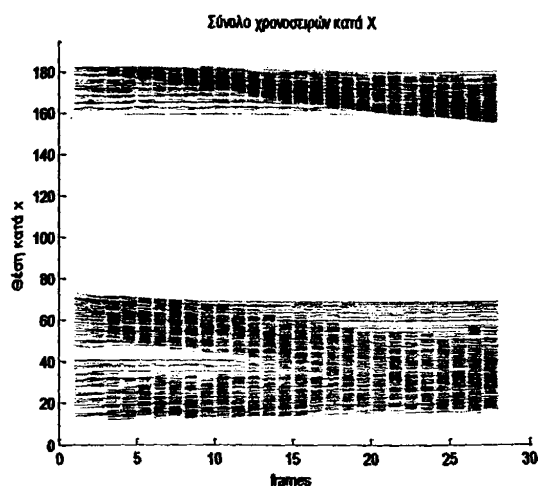
(α) Σύνολο χρονοσειρών



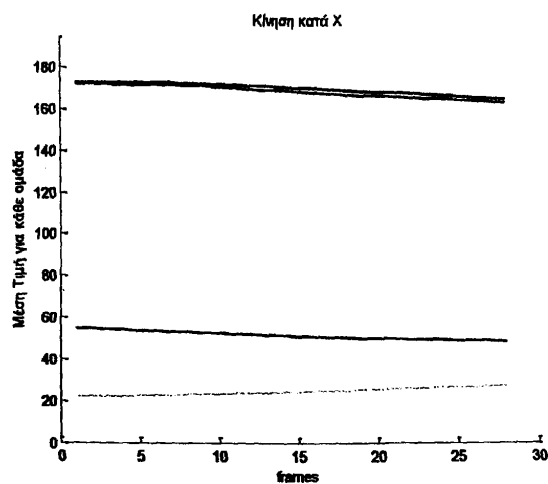
(β) Τελικές ομάδες αντικειμένων



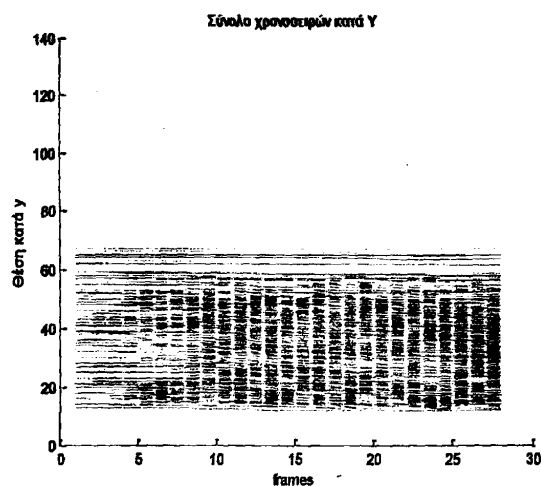
(γ) Συνολική κίνηση κατά X και Y



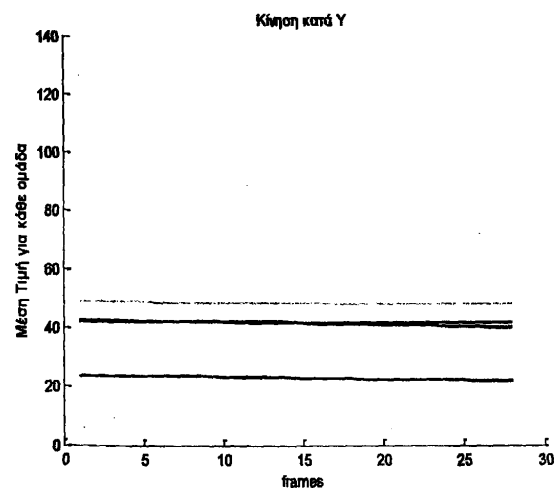
(δ) Σύνολο χρονοσειρών κατά X



(ε) Κίνηση κατά X



(στ) Σύνολο χρονοσειρών κατά Y



(ζ) Κίνηση κατά Y

Σχήμα 5.17 Ομαδοποίηση αντικειμένων του πραγματικού βίντεο 4

Όπως φαίνεται από τα παραπάνω σχήματα, το μοντέλο ομαδοποίησης βρίσκει κάθε φορά τις υποκείμενες ομάδες σε κάθε βίντεο. Η κίνηση των αντικειμένων είναι πιο ευδιάκριτη στα διαγράμματα που παρουσιάζουν τις τροχιές των αντικειμένων και στους δύο άξονες, δηλαδή και κατά X και κατά Y.

ΚΕΦΑΛΑΙΟ 6. ΕΠΙΛΟΓΟΣ

6.1 Συμπεράσματα

6.2 Μελλοντικές Επεκτάσεις

6.1. Συμπεράσματα

Στην παρούσα εργασία ασχοληθήκαμε με την εξαγωγή και ομαδοποίηση των τροχιών κίνησης κινούμενων αντικειμένων σε βίντεο προερχόμενα από σταθερή κάμερα, με σκοπό την αποτελεσματική ανίχνευση της συνολικής κίνησης των αντικειμένων. Αρχικά, χρησιμοποιήθηκε η μεθοδολογία της οπτικής ροής, με τη βοήθεια της οποίας έγινε η εξαγωγή των διανυσμάτων κίνησης για κάθε pixel μεταξύ όλων των διαδοχικών frames των βίντεο. Ακολούθως, τα διανύσματα κίνησης χρησιμοποιήθηκαν ούτως ώστε να συντεθεί η συνολική τροχιά κίνησης (χρονοσειρά) του κάθε pixel. Για την ομαδοποίηση των χρονοσειρών αυτών χρησιμοποιήθηκε το μικτό πολωνυμικό μοντέλο παλινδρόμησης. Για την αφαίρεση του παρασκηνίου χρησιμοποιήθηκε ένα κατώφλι T με βάση το οποίο έγινε η επιλογή των τελικών χρονοσειρών προς ομαδοποίηση. Τέλος, έγινε επέκταση του μικτού πολωνυμικού μοντέλου παλινδρόμησης, ενσωματώνοντας σε αυτό και την έννοια της μετατόπισης των χρονοσειρών.

Τα αποτελέσματα τα οποία προέκυψαν από την πειραματική μελέτη έδειξαν ότι το μικτό πολωνυμικό μοντέλο ομαδοποίησης με μετατόπιση είναι αρκετά αποτελεσματικό. Βέβαια, ένα σημαντικό μειονέκτημα του μοντέλου είναι ότι βασίζεται στην παρατήρηση για την εκτίμηση της καλύτερης τιμής τριών σημαντικών παραμέτρων, του αριθμού K των ομάδων, του βαθμού p των πολωνύμων και του κατωφλίου T . Επίσης, ένα εξίσου σημαντικό μειονέκτημα του μοντέλου είναι η μεγάλη πολυπλοκότητα που συνεπάγεται η εφαρμογή του σε



υψηλής ανάλυσης βίντεο, καθώς η υψηλή ανάλυση σημαίνει πολύ μεγάλο πλήθος pixels, γεγονός το οποίο αφενός επιβαρύνει τη διαδικασία εξαγωγής των διανυσμάτων κίνησης και αφετέρου οδηγεί στην παραγωγή μεγάλου όγκου διανυσμάτων κίνησης. Ο μεγάλος όγκος των διανυσμάτων κίνησης οδηγεί φυσικά σε μεγάλο όγκο δεδομένων, δηλαδή χρονοσειρών, κάτι το οποίο επιβραδύνει τη διαδικασία ομαδοποίησης. Δύο πιθανές λύσεις για την μείωση της πολυπλοκότητας της μεθόδου, λόγω της υψηλής ανάλυσης του βίντεο, είναι, πρώτον, η αύξηση του παραθύρου που χρησιμοποιείται για την εξαγωγή των διανυσμάτων κίνησης (βλέπε Παράρτημα C για τον υπολογισμό της οπτικής ροής) και δεύτερον, αντί να υπολογίσουμε τις χρονοσειρές για κάθε pixel, μπορούμε να χρησιμοποιήσουμε ένα παράθυρο $m \times m$ και να πάρουμε τη μέση τιμή των pixels του παραθύρου αυτού.

Πέραν των μειονεκτημάτων, τα οποία αναφέρθηκαν παραπάνω, το μοντέλο της ομαδοποίησης χρονοσειρών έχει και αρκετά δυνατά σημεία. Το σημαντικότερο ίσως πλεονέκτημά του αποτελεί ο ταυτόχρονος εντοπισμός των pixels, τα οποία απαρτίζουν τα αντικείμενα, και της κίνησης που διεξάγεται από αυτά, καθώς οι χρονοσειρές που συντίθενται από τα διανύσματα κίνησης απεικονίζουν και το αντικείμενο αλλά και την τροχιά κίνησής του. Επίσης, η χρησιμοποίηση κατωφλίου (ανεξάρτητα αν βασίζεται στην παρατήρηση ή όχι) για την αφαίρεση των χρονοσειρών που ανήκουν στο παρασκήνιο μας επιτρέπει να έχουμε μια πιο ξεκάθαρη εικόνα για τα αντικείμενα που εμφανίζονται στο βίντεο, με αποτέλεσμα η ομαδοποίησή τους να είναι πιο αποτελεσματική.

6.2. Μελλοντικές Επεκτάσεις

Υπάρχουν πολλά περιθώρια περαιτέρω βελτίωσης της μεθόδου, σύμφωνα με τις παρατηρήσεις που έγιναν στην παραπάνω ενότητα. Αρχικά, θα μπορούσε να χρησιμοποιηθεί κάποια τεχνική για την αποτελεσματική εκτίμηση του αριθμού των ομάδων K . Επίσης, θα μπορούσε να ενσωματωθεί στο μικτό μοντέλο ομαδοποίησης και μια επιπλέον συνιστώσα, η οποία θα αναφερόταν στο παρασκήνιο (αυξάνοντας έτσι τον αριθμό των ομάδων από K σε $K+1$). Με αυτό τον τρόπο μπορεί να αποφευχθεί η χρήση κατωφλίου για την αφαίρεση του παρασκηνίου. Όσον αφορά την εκτίμηση του βαθμού του πολυωνύμου, η χρήση αραιών μοντέλων

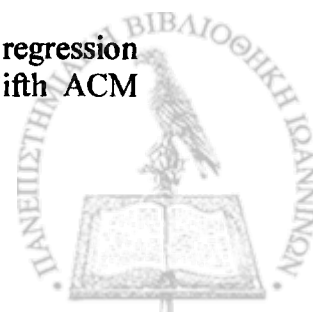


παλινδρόμησης μπορεί να μειώσει αποτελεσματικά την εξάρτηση από το βαθμό p . Τέλος, το μοντέλο μας αναπτύχθηκε και εφαρμόστηκε για βίντεο στα οποία η κάμερα είναι σταθερή. Μια πιθανή μελλοντική επέκταση θα μπορούσε να γίνει ούτως ώστε το μοντέλο να εφαρμοστεί και σε βίντεο προερχόμενα από κινούμενη κάμερα.



ΑΝΑΦΟΡΕΣ

- [1] Bezdek J. C., "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York and London, 1987.
- [2] Cucchiara R., Grana C., Piccardi M., and Prati A., "Detecting moving objects, ghosts and shadows in video streams," IEEE Trans. Pattern Anal. Machine Intell., vol. 25(10), pp 1337–1342, 2003.
- [3] de Boor C., "A practical guide to splines", New York, NY: Springer-Verlag, 1978.
- [4] de Boor C., "B(asic)-spline basics". In C. de Boor (Ed.), Extension of bspline curve algorithms to surfaces (Vols. Course #5, ACM SIGGRAPH 86, pp18-22). Dallas, TX: ACM, 1986.
- [5] Dempster A. P., Laird N. M., Rubin D. B., "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society B, Vol (39), pp 1-38, 1977.
- [6] DeSarbo W. S. & Cron W. L., "A maximum likelihood methodology for clusterwise linear regression", Journal of Classification, Vol5(1), pp 249-282, 1988.
- [7] Fan J., Gijbels I., "Local Polynomial modeling and its applications", Chapman and Hall, London, 1996.
- [8] Frey, B. J., & Jojic, N. "Estimating mixture models of images and inferring spatial transformations using the EM algorithm", In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp 416–422, 1999.
- [9] Frey, B. J., & Jojic, N. "Fast, large-scale transformation-invariant clustering", In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14. Cambridge, MA: MIT Press, 2002.
- [10] Frey, B. J., & Jojic, N. "Transformation-invariant clustering using the EM Algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 25 (1), pp 1–17, 2003.
- [11] Gaffney S. & Smyth P., "Trajectory clustering with mixtures of regression models", In S. Chaudhuri & D. Madigan (Eds.), Proceedings of the Fifth ACM



SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 63-72, August 1999.

[12] Gaffney S., "Probabilistic Curve – Aligned clustering and prediction with regression mixture models", University of California, Irvine, 2004.

[13] Gonzalez R. C. and Woods R. E., "Digital Image Processing". Prentice Hall, 2002.

[14] Han J., Kamber M., "Data Mining: Concepts and Techniques", Morgan Kaufmann, San Francisco, pp 346-389, 2001.

[15] Haritaoglu I, Harwood D, Davis L., "W-4: Real-time surveillance of people and their activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8) 809-30, 2000.

[16] Hosmer D., "Maximum likelihood estimates of the parameters of a mixture of two regression lines", Communication in Statistics, Vol 3(10), pp 995-1006, 1974.

[17] Jabri S., Duric Z., Wechsler H., Rosenfeld A., "Detection and location of people in video images using adaptive fusion of color and edge information," in Proceedings of Int'l Conference on Pattern Recognition, pp 627-630, 2000.

[18] Jeongho S., Sangjin K., Sangkyu K., Seong-Won L., Joonki P., Besma A., Mongi A., "Optical flow-based real-time object tracking using non-prior training active feature mode", ELSEVIER Real-Time Imaging, Vol (11), pp 204-218, USA, June 2005.

[19] Kalafatic Z., Ribaric S., Stanisavljevic V., "Real-time object tracking based on optical flow and active rays", Electrotechnical Conference, MELECON 2000. 10th Mediterranean, Vol(2), pp 542-545, 2000.

[20] Kaufman L., Rousseeuw P. J., "Finding Groups in Data: An Introduction to Cluster Analysis", Wiley, New York, 1990.

[21] McKenna S. J. and Gong S., "Tracking colour objects using adaptive mixture models," Image Vision Computing, vol. 17, pp 225-231, 1999.

[22] McKenna S.J., Jabri S., Duric Z., Rosenfeld A., Wechsler H., "Tracking groups of people", Computer Vision and Image Understanding, Vol. 1(80), pp 42-56, 2000.

[23] McLachlan G. J. & Basford K. E., "Mixtures models: Inference and applications to clustering", New York: Marcel Dekker, 1988.

[24] MacQueen, J. B., "Some methods for classification and analysis of multivariate observations" in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, pp 281-297, 1967.



- [25] Okada R., Shirai Y., Miura J., "Object Tracking based on optical flow and depth", Proc. Of IEEE/SICE/RSJ Int. Conf. on MFI, pp 565-571, 1996
- [26] Quandt R. E., "A new approach to estimating switching regressions", Journal of the American Statistical Association, 67, pp 306-310, 1972.
- [27] Seki M., Fujiwara H., Sumi K., "A robust background subtraction method for changing background", in Proceedings of IEEE Workshop on Applications of Computer Vision, pp 207-213, 2000.
- [28] Silverman B. W., "Density estimation for statistics and data analysis", NY: Chapman and Hall, New York, Vol 57(4), pp 673-689, 1986.
- [29] Souvenir R., Wright J., Pless R., "Spatio-temporal detection and isolation: Results on the PETS2005 datasets (2008)", In Proceedings of the IEEE Workshop on Performance Evaluation in Tracking and Surveillance, 2005.
- [30] Stauffer C., Eric W., Grimson L., (2000) "Learning patterns of activity using real-time tracking". IEEE Trans. Pattern Anal. Machine Intell., Vol. 22(8), pp 747-757, August 2000.
- [31] Tan P. N., Steinbach M., Kumar V., "Introduction to Data Mining", Addison-Wesley, February 2005.
- [32] Wand M. P., Jones M. C., "Kernel Smoothing", Chapman and Hall, London, 1995.
- [33] Warren L. T., "Clustering of time series data – a survey", Pattern Recognition Society, Elsevier Ltd, Vol (38), pp 1857-1874, 2005.
- [34] Wren C. R., Azarbayejani A., Darrell T., Pentland A. P., "Pfinder: Real-time tracking of the human body", IEEE Trans. Pattern Anal. Machine Intell., vol. 19(7), pp 780-785, July 1997.
- [35] Yu-Fei Ma, Hong-Jiang Zhang, "Detecting motion object by spatio-temporal entropy", IEEE International Conference on Multimedia and Expo (ICME'01), pp 68, 2001.
- [36] [HTTP://WWW.PETS2006.NET/](http://www.pets2006.net/)
- [37] [HTTP://DOMINO.RESEARCH.IBM.COM/COMM/RESEARCH_PROJECTS.NSF/PAGES/S3.PERFORMANCEEVALUATION.HTML](http://domino.research.ibm.com/comm/research_projects.nsf/pages/s3.performanceevaluation.html)
- [38] [HTTP://EN.WIKIPEDIA.ORG/WIKI/OPTIC_FLOW](http://en.wikipedia.org/wiki/Optic_flow)



ΠΑΡΑΡΤΗΜΑ

Α ΑΛΓΟΡΙΘΜΟΣ EM

Ο αλγόριθμος EM [5] αποτελεί μια επαναληπτική ML (Maximum Likelihood) διαδικασία, η οποία προσφέρει ένα γενικό και αποτελεσματικό πλαίσιο για την εκτίμηση των παραμέτρων ενός πιθανοτικού μοντέλου, σε περιπτώσεις ελλειπών ή κρυμμένων δεδομένων.

Κάθε επανάληψη του αλγόριθμου EM αποτελείται από δύο βήματα: Το βήμα υπολογισμού της αναμενόμενης τιμής (*E-step*) και το βήμα μεγιστοποίησης (*M-step*). Στο *E-step* γίνεται η εκτίμηση των κρυμμένων δεδομένων, δοθέντων των πραγματικών δεδομένων και των τρεχουσών εκτιμήσεων των παραμέτρων του μοντέλου. Στο *M-step* γίνεται η μεγιστοποίηση της συνάρτησης πιθανοφάνειας, κάνοντας την υπόθεση ότι τα κρυμμένα δεδομένα είναι γνωστά. Η εκτίμηση των κρυμμένων δεδομένων, από το *E-step*, χρησιμοποιείται αντί των πραγματικών κρυμμένων δεδομένων.

Στη συνέχεια περιγράφονται αναλυτικότερα τα δύο βήματα του αλγορίθμου EM.

Έστω ότι έχουμε τα δεδομένα Y , τα κρυμμένα δεδομένα Z και το διάνυσμα παραμέτρων θ . Τότε η πιθανοφάνεια του θ , δοθέντος του Y , στη γενική της μορφή, είναι η εξής:

$$\mathcal{L}(\theta|Y) = \int_Z p(Y|Z, \theta)p(Z|\theta) dZ \quad \text{Εξ. Α.1}$$

Ο υπολογισμός του παραπάνω ολοκληρώματος είναι αρκετά δύσκολος. Ο αλγόριθμος EM διευκολύνει τον υπολογισμό αυτό, ορίζοντας μια άλλη συνάρτηση



πιθανοφάνειας, γνωστή ως *complete-data likelihood*, η οποία περιέχει και τα κρυμμένα δεδομένα. Η νέα αυτή συνάρτηση είναι η ακόλουθη:

$$\mathcal{L}_c(\theta|Y, Z) = p(Y, Z|\theta) \quad \text{Εξ. A.2}$$

E-step

Οι πραγματικές τιμές του Z , στην παραπάνω συνάρτηση πιθανοφάνειας, είναι άγνωστες. Γι' αυτό ο αλγόριθμος EM συμπληρώνει τιμές για τα άγνωστα Z , υπολογίζοντας την αναμενόμενη τιμή της $\mathcal{L}_c$ με τη βοήθεια της εκ των υστέρων κατανομής $p(Z|Y, \theta')$, για ένα σταθερό διάνυσμα παραμέτρων θ' . Η εκ των υστέρων αυτή κατανομή είναι γνωστή και ως κατανομή των κρυμμένων δεδομένων (*hidden-data distribution*), εφόσον μας δίνει την κατανομή των κρυμμένων δεδομένων, δοθέντων των πραγματικών παρατηρήσεων Y και του τρέχοντος διανύσματος παραμέτρων θ' . Τότε η αναμενόμενη τιμή έχει ως εξής:

$$E[\mathcal{L}_c|Y] = Q(\theta, \theta') = \int p(Y, Z|\theta)p(Z|Y, \theta') dZ \quad \text{Εξ. A.3}$$

Η παραπάνω συνάρτηση ονομάζεται και Q -συνάρτηση.

M-step

Σε αυτό το βήμα γίνεται η μεγιστοποίηση της Q -συνάρτησης, για να υπολογιστούν οι εκτιμήσεις $\hat{\theta}$ των παραμέτρων θ . Δηλαδή:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} Q(\theta, \theta') \quad \text{Εξ. A.4}$$

Στην επόμενη επανάληψη του EM το θ' αντικαθίσταται από το νέο $\hat{\theta}$ στην κατανομή των κρυμμένων δεδομένων $p(Z|Y, \theta')$.

Επειδή σε κάθε επανάληψη ο αλγόριθμος EM αυξάνει την πιθανοφάνεια, η σύγκλιση επιτυγχάνεται όταν η αύξηση αυτή, ανάμεσα από δύο διαδοχικές επαναλήψεις, είναι μικρότερη από ένα κατώφλι.



B ΜΙΚΤΑ ΜΟΝΤΕΛΑ ΚΑΤΑΝΟΜΗΣ

Τα πεπερασμένα Μικτά Μοντέλα μπορούν να θεωρηθούν ως μια ημι-παραμετρική μορφή της εκτίμησης πιθανότητας, συμφωνά με την οποία υποθέτουμε ότι το σύνολο των δεδομένων δεν περιγράφεται από μια μόνη κατανομή, αλλά κάθε υποπεριοχή των δεδομένων μπορεί να περιγραφεί με μια ατομική συνάρτηση πυκνότητας πιθανότητας (PDF), η οποία χαρακτηρίζεται από ένα διάνυσμα παραμέτρων θ . Η συνολική πιθανότητα θα αποτελείται από έναν συνδυασμό όλων των PDFs. Ένα από τα βασικά πλεονεκτήματα του μικτού μοντέλου, είναι η ικανότητά του να περιγράφει πολύπλοκες συναρτήσεις πυκνότητας χρησιμοποιώντας απλούστερες μορφές.

Στη βασική μορφή, μπορούμε να περιγράψουμε το δεδομένο y_i με την ακόλουθη μικτή πυκνότητα:

$$p(y_i|\Theta) = \sum_k^K a_k p_k(y_i|\theta_k) \quad \text{Εξ. B.1}$$

όπου, a_k είναι το βάρος της k -οστής ομάδας και p_k είναι η πυκνότητα της k -οστής ομάδας με διάνυσμα παραμέτρων θ_k . Τα μικτά βάρη a_k έχουν άθροισμα ένα και είναι θετικά. Οι συναρτήσεις πυκνότητας $p_k(\cdot)$ περιγράφουν συγκεκριμένες υποπεριοχές (ομάδες) της πυκνότητας, ενώ η μικτή πιθανότητα $p(\cdot)$ συνοψίζει όλες τις υποπεριοχές με βάση τα μικτά βάρη a_k .

Εκτίμηση πυκνότητας πιθανότητας

Η εκτίμηση της πυκνότητας είναι μια βασική πιθανοτική τεχνική η οποία μπορεί να χρησιμοποιηθεί για την περιγραφή ενός συνόλου δεδομένων. Η τεχνική αυτή, αρχικά υποθέτει μια παραμετρική μορφή για τη συνάρτηση πυκνότητας πιθανότητας (PDF), όπως για παράδειγμα μια κανονική κατανομή, και στη συνέχεια προσαρμόζει (fit) το μοντέλο αυτό στα πραγματικά δεδομένα.



Η διαδικασία της προσαρμογής περιλαμβάνει την προσαρμογή ενός συνόλου παραμέτρων Θ της κατανομής (για παράδειγμα, ο μέσος και η διακύμανση), ούτως ώστε η παραχθείσα πυκνότητα να “ταιριάζει” με την πραγματική κατανομή των δεδομένων.

Για τη διαδικασία της προσαρμογής συνήθως χρησιμοποιείται μια από τις παρακάτω μεθόδους: μέθοδος των ροπών (*method of moments*), Εκτίμηση Μεγίστης Πιθανοφάνειας (*Maximum Likelihood Estimation, ML*), Εκτίμηση Μεγίστης εκ των υστέρων Πιθανότητας (*Maximum a posteriori Estimation, MAP*) ή Μπεϋζιανές Τεχνικές (*Bayesian techniques*).

Εκτίμηση Μέγιστης Πιθανοφάνειας (ML) και ο αλγόριθμος EM

Η πιθανοφάνεια ενός συνόλου δεδομένων $Y = \{y_1, \dots, y_n\}$ είναι μια συνάρτηση των παραμέτρων Θ , δοθέντων των δεδομένων Y , η οποία είναι ανάλογη της πιθανότητας $p(Y|\Theta)$ των δεδομένων, δηλαδή της εκ των υστέρων πιθανότητας των δεδομένων Y , δοθέντων των παραμέτρων.

Η λογαριθμική πιθανοφάνεια (*log-likelihood*) είναι ο λογάριθμος της πιθανοφάνειας και ορίζεται ως:

$$\begin{aligned}\mathcal{L}(\Theta|Y) &= \log p(Y|\Theta) \\ &= \sum_i \log p(y_i|\Theta) \\ &= \sum_i \log \sum_k a_k p_k(y_i|\theta_k)\end{aligned}\quad \text{Εξ. B.2}$$

υποθέτοντας ότι τα y_i είναι ανεξάρτητα και ομοιόμορφα κατανεμημένα. Η εκτίμηση ML του διανύσματος παραμέτρων Θ έγκειται στην εύρεση εκείνων των τιμών του Θ που μεγιστοποιούν την Εξίσωση (B.2).

Γενικά, οι ML εκτιμήσεις για το Θ , δεν παράγουν λύσεις κλειστής μορφής, εφόσον οι εκτιμήσεις εξαρτώνται μη-γραμμικά ή μια από την άλλη. Η πιο γνωστή



επαναληπτική ML διαδικασία μεγιστοποίησης της αναμενόμενης τιμής της πιθανοφάνειας, είναι ο αλγόριθμος *Expectation-Maximization (EM)* [5]. Ο αλγόριθμος EM ψάχνει επαναληπτικά για ένα σύνολο παραμέτρων $\hat{\Theta}$ το οποίο μεγιστοποιεί την πιθανότητα των παρατηρούμενων δεδομένων (βλέπε Παράρτημα Α).

Σ ΟΠΤΙΚΗ ΡΟΗ (OPTICAL FLOW)

Οι μέθοδοι της οπτικής ροής προσπαθούν να αναπαραστήσουν την κίνηση μεταξύ δύο εικόνων (frames) οι οποίες απέχουν χρόνο δt [38]. Αυτές οι μέθοδοι ονομάζονται διαφορικές επειδή βασίζονται σε προσεγγίσεις με σειρές Taylor, δηλαδή χρησιμοποιούν μερικές παραγώγους ως προς τις χωρικές και χρονικές συντεταγμένες.

Ένα pixel το οποίο βρίσκεται στη θέση (x, y, t) , δηλαδή τη χρονική στιγμή t η θέση του προσδιορίζεται από τις συντεταγμένες (x, y) , και έχει ένταση φωτεινότητας $I(x, y, t)$, θα έχει μετακινηθεί κατά δx , δy και δt , μεταξύ δύο διαδοχικών εικόνων. Οπότε η παρακάτω εξίσωση περιορισμών μπορεί να οριστεί:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad \text{Εξ. C.1}$$

Υποθέτοντας ότι η κίνηση είναι πολύ μικρή μπορούμε να την εκφράσουμε με μια σειρά Taylor:

$$\begin{aligned} I(x + \delta x, y + \delta y, t + \delta t) \\ = I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t + O.M.T \end{aligned} \quad \text{Εξ. C.2}$$

όπου $O.M.T$ σημαίνει όροι μεγαλύτερης τάξης, οι οποίοι είναι αρκετά μικροί ώστε να αγνοηθούν.

Θεωρώντας ότι η φωτεινότητα ενός αντικειμένου δεν αλλάζει, οπότε ούτε η φωτεινότητα των pixels αλλάζει, προκύπτει ότι:



$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) \quad \text{Εξ. C.3}$$

Από τις Εξισώσεις C.2 και C.3 προκύπτει ότι:

$$\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t = 0 \quad \text{Εξ. C.4}$$

Και αν διαιρέσουμε με δt έχουμε:

$$\frac{\partial I}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial I}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} \frac{\delta t}{\delta t} = 0 \quad \text{Εξ. C.5}$$

Η Εξίσωση C.5 μπορεί να γραφεί ως:

$$\frac{\partial I}{\partial x} V_x + \frac{\partial I}{\partial y} V_y + \frac{\partial I}{\partial t} = 0 \quad \text{Εξ. C.6}$$

όπου V_x και V_y είναι οι x και y συνιστώσες της ταχύτητας ή, όπως αλλιώς ονομάζεται, η *Οπτική Ροή* της $I(x, y, t)$ και $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, $\frac{\partial I}{\partial t}$ είναι οι παράγωγοι της εικόνας στο σημείο (x, y, t) . Εάν αντικαταστήσουμε τις παραγώγους με I_x , I_y και I_t αντίστοιχα, τότε η Εξίσωση C.6 μπορεί να γραφεί ως εξής:

$$I_x V_x + I_y V_y = -I_t \quad \text{Εξ. C.7}$$

Η Εξίσωση C.7 είναι μια εξίσωση με δύο αγνώστους και ως εκ τούτου δε μπορεί να λυθεί. Για να βρεθεί η οπτική ροή χρειάζονται και κάποιες άλλες εξισώσεις, οι οποίες εξάγονται από κάποιους επιπλέον περιορισμούς. Όλες οι διαθέσιμες μέθοδοι υπολογισμού της οπτικής ροής εισάγουν επιπλέον συνθήκες για να εκτιμήσουν την πραγματική ροή. Στη συνέχεια περιγράφεται μια από τις πιο γνωστές μεθόδους υπολογισμού οπτικής ροής, η Lucas-Kanade (που είναι και η μέθοδος που χρησιμοποιείται στην παρούσα εργασία).

Lucas-Canade Optical Flow

Η μέθοδος αυτή βασίζεται στην παραπάνω Εξίσωση Οπτικής Ροής. Ο επιπρόσθετος περιορισμός που χρησιμοποιείται είναι η υπόθεση ότι η ροή (V_x, V_y) είναι σταθερή



σε ένα μικρό παράθυρο μεγέθους $m \times m$ με $m > 1$, το οποίο έχει κέντρο το pixel x, y και αριθμώντας τα pixels μέσα στο παράθυρο ως $1 \dots n$, $n = m^2$, ένα σύνολο εξισώσεων μπορεί να οριστεί:

$$\begin{aligned} I_{x1}V_x + I_{y1}V_y &= -I_{t1} \\ I_{x2}V_x + I_{y2}V_y &= -I_{t2} \\ &\vdots \\ I_{xn}V_x + I_{yn}V_y &= -I_{tn} \end{aligned} \quad \text{Εξ. C.8}$$

Με τις παραπάνω εξισώσεις ορίζονται περισσότερες από δύο εξισώσεις για τους δύο αγνώστους, οπότε το σύστημα είναι υπερεκτιμημένο. Άρα:

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \vdots \\ -I_{tn} \end{bmatrix} \quad \text{Εξ. C.9}$$

ή

$$A\vec{v} = -b \quad \text{Εξ. C.10}$$

Για να λυθεί το παραπάνω σύστημα μπορεί να χρησιμοποιηθεί η μέθοδος των ελαχίστων τετραγώνων. Οπότε έχουμε:

$$\begin{aligned} A^T A \vec{v} &= A^T (-b) \Rightarrow \\ \vec{v} &= (A^T A)^{-1} A^T (-b) \Rightarrow \\ \begin{bmatrix} V_x \\ V_y \end{bmatrix} &= \begin{bmatrix} \sum I_{xi}^2 & \sum I_{xi}I_{yi} \\ \sum I_{xi}I_{yi} & \sum I_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{xi}I_{ti} \\ -\sum I_{yi}I_{ti} \end{bmatrix} \end{aligned} \quad \text{Εξ. C.11}$$

για $i = 1 \dots n$.



ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Ιωάννα Φλεντζούρη γεννήθηκε στο Αργυρόκαστρο το 1984. Το 2002 έγινε δεκτή στο τμήμα Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας, στη Θεσσαλονίκη, ολοκληρώνοντας τη φοίτησή της το 2006. Την επόμενη χρονιά, έγινε δεκτή στο Μεταπτυχιακό Πρόγραμμα του Τμήματος Πληροφορικής, του Πανεπιστημίου Ιωαννίνων. Τα ερευνητικά της ενδιαφέροντα περιλαμβάνουν την ομαδοποίηση χρονοσειρών και την ανάχνευση κινούμενων αντικειμένων σε βίντεο.

