



Εθνικό Μετσόβιο Πολυτεχνείο
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Πληροφορικής

Διδακτορική Διατριβή

Ανάλυση Ακολουθιών Εικόνων για Χαρακτηρισμό, Προεπισκόπηση και Αναζήτηση με Βάση το Περιεχόμενο

Ιωάννης Αβρίθης

Αθήνα, Μάρτιος 2001

Ερευνητικό Αντικείμενο

- *Παραδοσιακή ανάκληση πληροφορίας: κατανόηση πληροφορίας κειμένου, εξαγωγή λέξεων-κλειδιών, κατηγοριοποίηση, δημιουργία περίληψης εγγράφων – με βάση το γραπτό κείμενο*
- *Εφαρμογές πολυμέσων: αποθήκευση, μετάδοση, αναζήτηση, επαναχρησιμοποίηση οπτικοακουστικής πληροφορίας (εικόνα, βίντεο, γραφικά, 3-Δ μοντέλα, κείμενο, ήχος, ομιλία) από υπολογιστικά συστήματα*
- *Ανάπτυξη νέων εργαλείων αυτόματου χαρακτηρισμού, δεικτοδότησης, εξαγωγής περίληψης, προεπισκόπησης, πλοήγησης, αναζήτησης με βάση το περιεχόμενο*

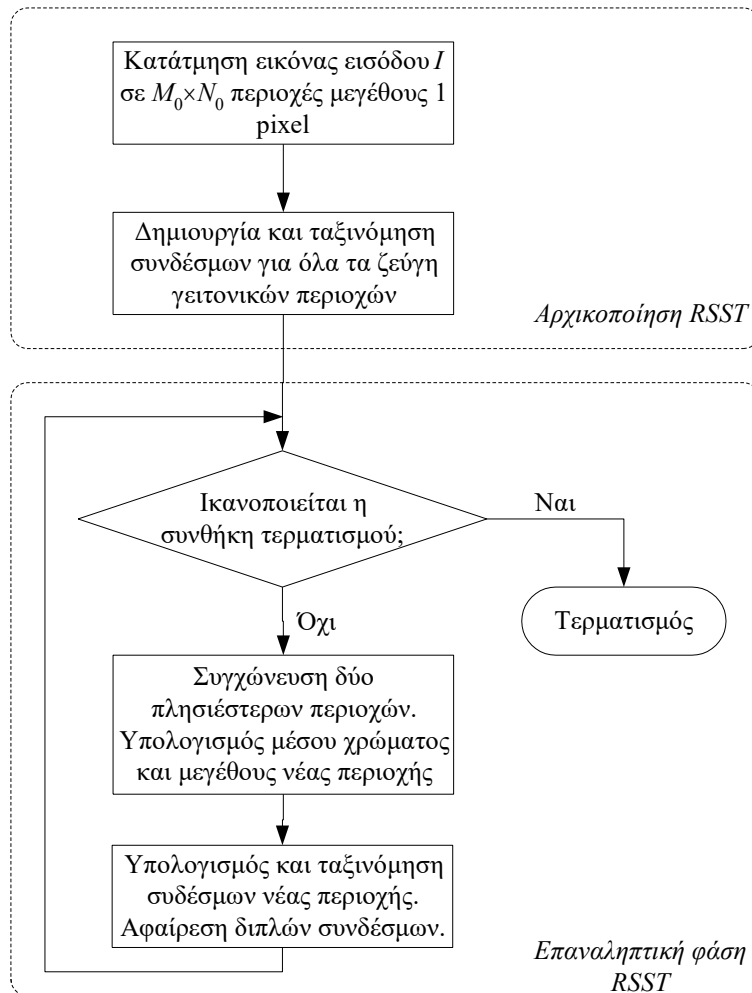
Δομή Διατριβής

- *Μέρος 1: Ανάλυση ακολουθιών σε αντικείμενα με βάση το χρώμα, την κίνηση & το πεδίο βάθους (σε στερεοσκοπικές ακολουθίες)*
- *Μέρος 2: Χαρακτηρισμός (υπολογισμός διανυσμάτων περιγραφής), εξαγωγή περίληψης / προεπισκόπηση (επιλογή χαρακτηριστικών πλάνων / καρτέ), αναζήτηση με βάση το περιεχόμενο*
- *Μέρος 3: Περιγραφή σχήματος αντικειμένων, κανονικοποίηση περιγραμμάτων, αναζήτηση με βάση το σχήμα*
- *Μέρος 4: Ανάλυση δελτίων ειδήσεων, εντοπισμός προσώπων, αναζήτηση με βάση τα πρόσωπα*

Κατάτμηση Ακολουθιών Εικόνων σε Αντικείμενα

- Ταχεία σύγκριση / υπολογισμός μέτρου ομοιότητας μεταξύ εικόνων, καρέ ή πλάνων ακολουθιών
- *Διανύσματα περιγραφής:*
 - ολικά (global) χαρακτηριστικά (π.χ. ιστογράμματα)
 - τοπικά χαρακτηριστικά (π.χ. κατάτμηση σε αντικείμενα)
- *Σημασιολογική κατάτμηση:* προσδιορισμός παρουσίας γνωστών αντικειμένων (εξειδικευμένες εφαρμογές, εκ των προτέρων γνώση)
- *Κατάτμηση με βάση το χρώμα, την κίνηση και το πεδίο βάθους* (στερεοσκοπικές ακολουθίες)
- *Συγχώνευση αντικειμένων*

Ο Αλγόριθμος RSST

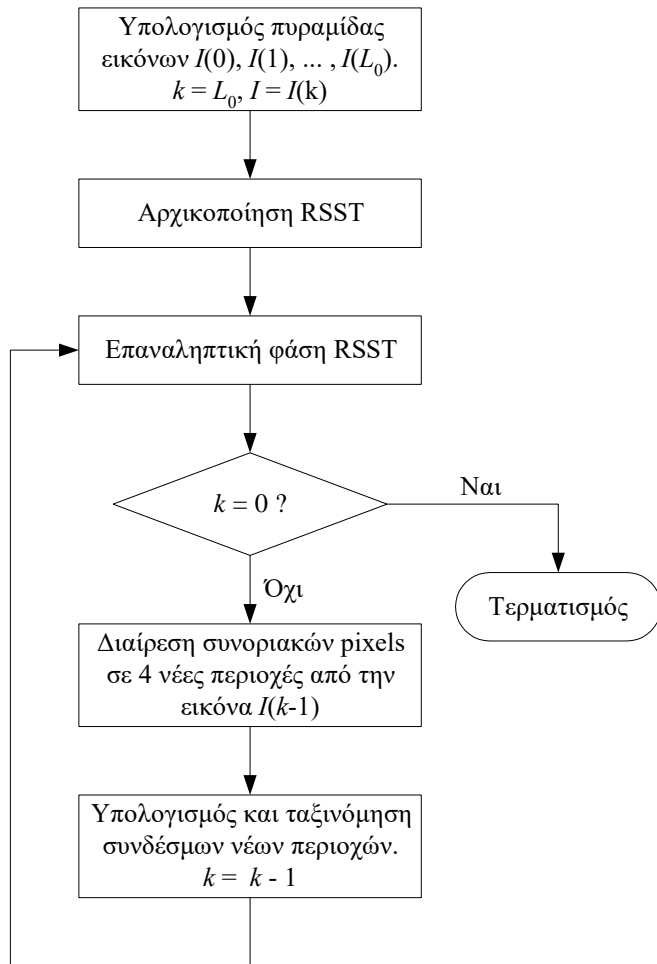


- Αρχικοποίηση με δημιουργία περιοχών μεγέθους 1 pixel
- 4-connected pixels
- Βάρος συνδέσμων:

$$\delta_c(X, Y) = \|\mathbf{c}(X) - \mathbf{c}(Y)\| \frac{a(X)a(Y)}{a(X) + a(Y)}$$

- Κατώφλι απόστασης
- Υπολογιστικό κόστος ταξινόμησης

Ο Αλγόριθμος M-RSST



- Ταχεία, πολυδιακριτική υλοποίηση
- Μεταβολή συνοριακών περιοχών μόνο σε κάθε επανάληψη
- Εξάλειψη πολύ μικρών περιοχών
- Εξαγωγή περιγραμμάτων με ακρίβεια
- Μερική αποκωδικοποίηση σε ακολουθίες MPEG

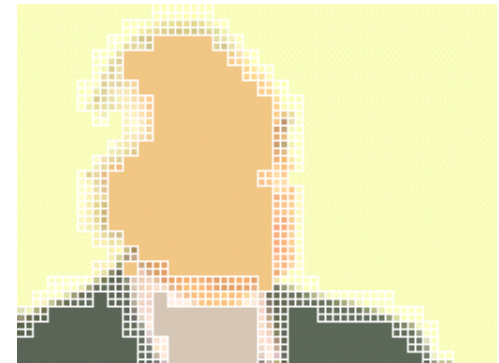
Κατάτμηση με Βάση το Χρώμα (1)



Αρχική Εικόνα



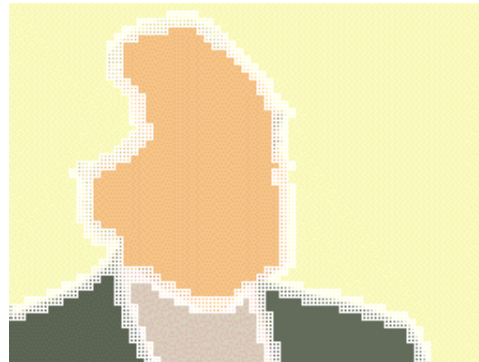
Επίπεδο 3 (merge)



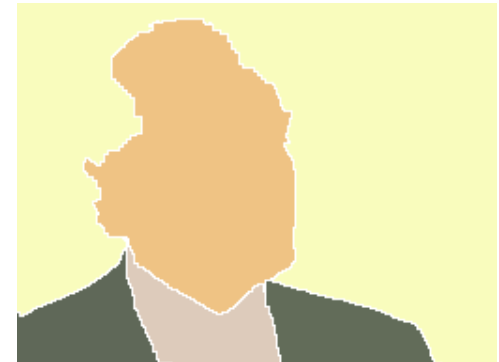
Επίπεδο 3 (split)



Επίπεδο 2 (merge)



Επίπεδο 2 (split)

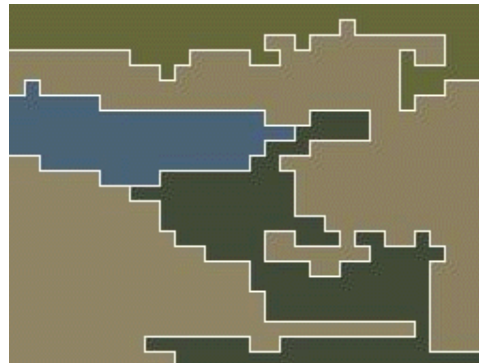


Επίπεδο 0 (merge)

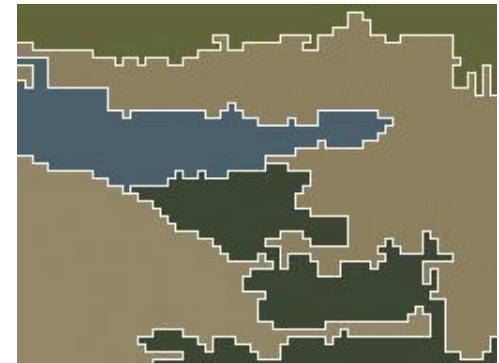
Κατάτμηση με Βάση το Χρώμα (2)



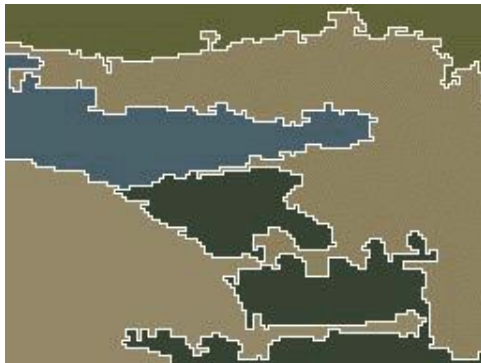
Αρχική εικόνα



Επίπεδο 3



Επίπεδο 2



Επίπεδο 1



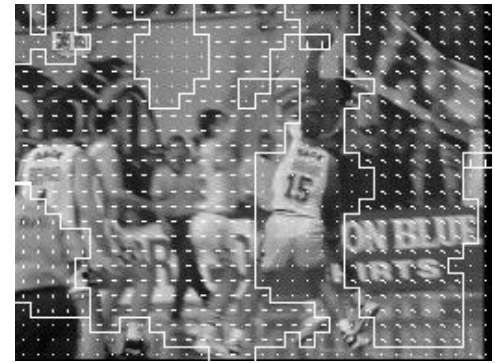
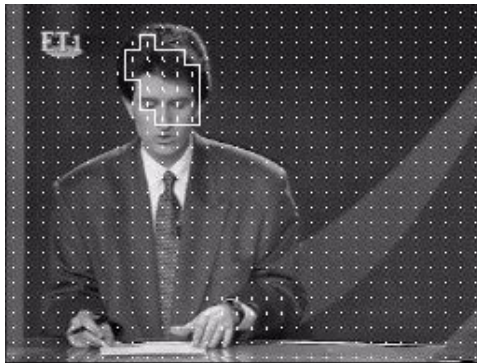
Επίπεδο 0



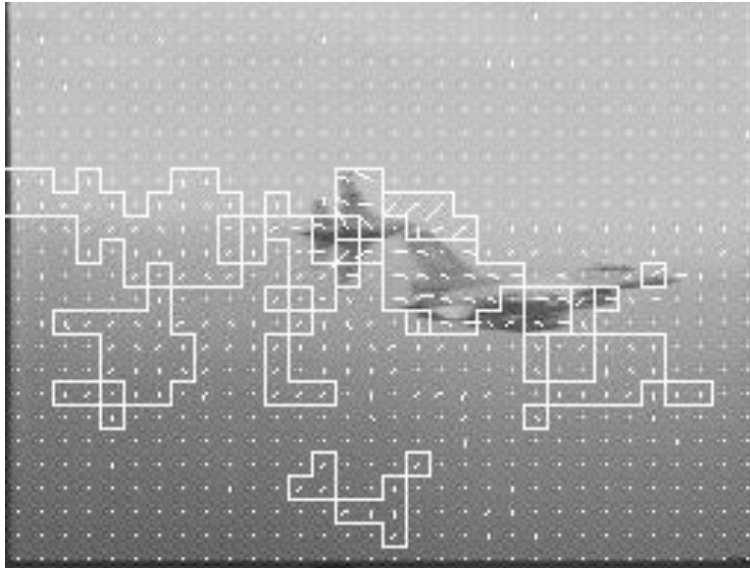
Τελική κατάτμηση

Κατάτμηση με Βάση την Κίνηση

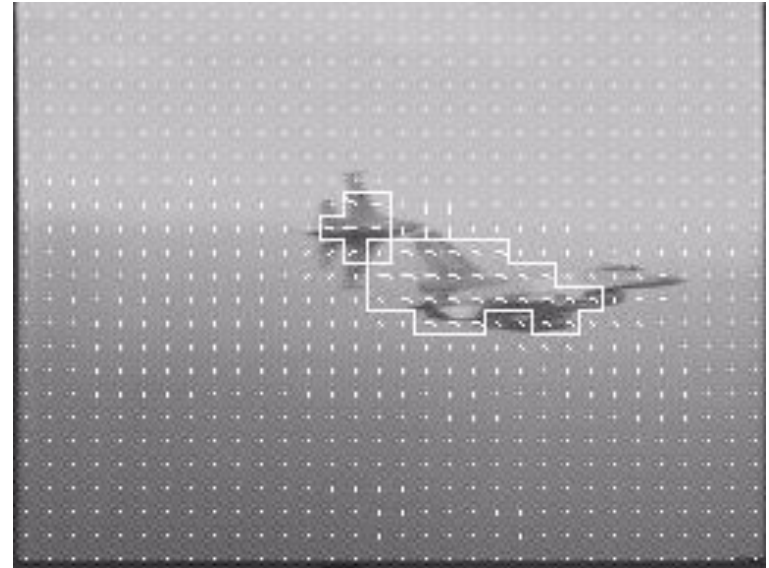
- Εκμετάλλευση διανυσμάτων κίνησης MPEG
- Επίπεδο ανάλυσης 3 (blocks 8×8)
- Εξομάλυνση με φίλτρο median
- Βάρος συνδέσμων: $\delta_m(X, Y) = \left\| \mathbf{v}(X) - \mathbf{v}(Y) \right\| \frac{a(X)a(Y)}{a(X) + a(Y)}$



Εξομάλυνση Διανυσμάτων Κίνησης



Χωρίς εξομάλυνση

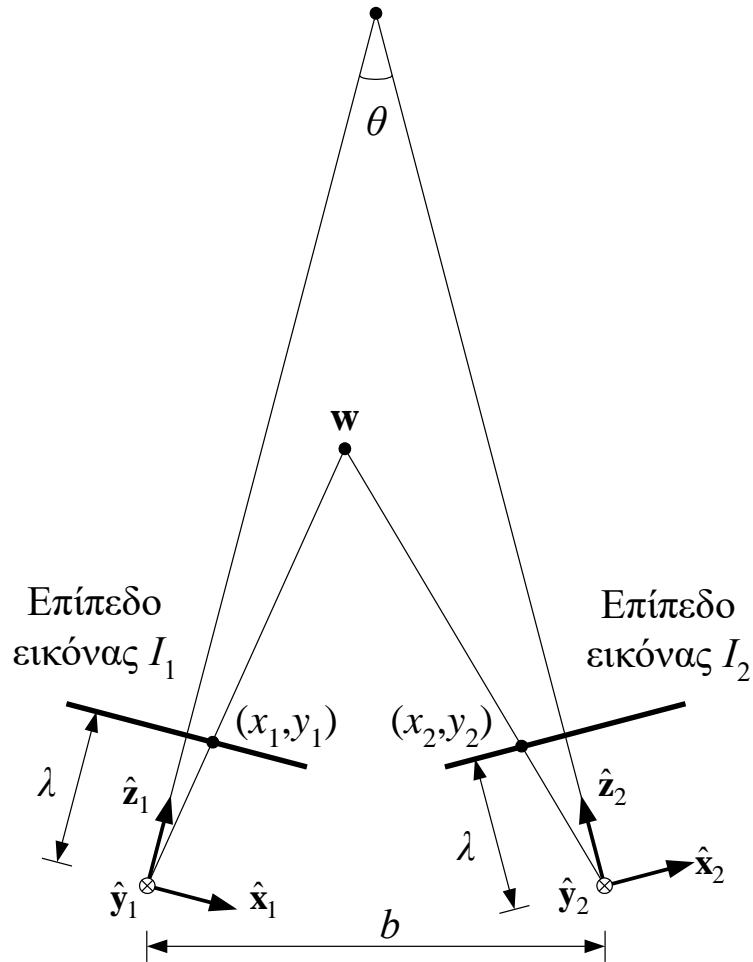


Με εξομάλυνση

Στερεοσκοπικές Ακολουθίες Βίντεο

- Ανάπτυξη συσκευών σύλληψης 3-Δ δεδομένων, στερεοσκοπικών συστημάτων κάμερας & συστημάτων 3-Δ απεικόνισης
- Εκμετάλλευση πληροφορίας βάθους: σημασιολογικά αντικείμενα συνήθως στο ίδιο επίπεδο βάθους
- Στερεοσκοπική ανάλυση: εκτίμηση πεδίου απόκλισης και βάθους, ανίχνευση & αντιστάθμιση επικάλυψης
- Κατάτμηση με βάση το πεδίο βάθους: ικανοποιητική προσέγγιση πραγματικών αντικειμένων, αναξιόπιστα περιγράμματα
- Συγχώνευση αντικειμένων χρώματος – βάθους

Στερεοσκοπική Ανάλυση



- Στερεοσκοπικό σύστημα κάμερας
- Προοπτική προβολή:

$$x_2 = \lambda \frac{(\lambda s + x_1 c)Z - \lambda b c'}{(\lambda c - x_1 s)Z + \lambda b s'}$$

$$y_2 = \lambda \frac{y_1 Z}{(\lambda c - x_1 s)Z + \lambda b s'}$$

- Πεδίο απόκλισης:

$$d_x = d_x(x_1, y_1) = x_2 - x_1$$

$$d_y = d_y(x_1, y_1) = y_2 - y_1$$

Εκτίμηση Πεδίου Βάθους και Απόκλισης



Αριστερό κανάλι



Δεξί κανάλι

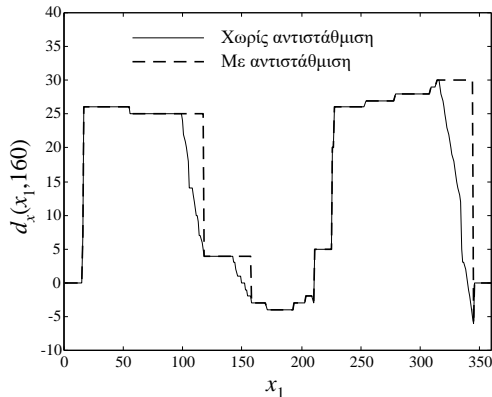


Πεδίο απόκλισης



Πεδίο βάθους

Ανίχνευση / Αντιστάθμιση Επικάλυψης



1-Δ αντιστάθμιση



Επικαλυπτόμενες περιοχές

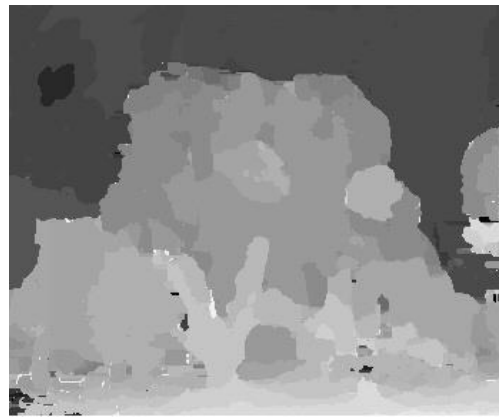
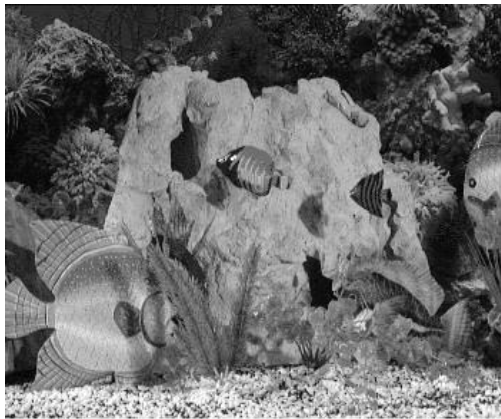


Αντ. πεδίο απόκλισης

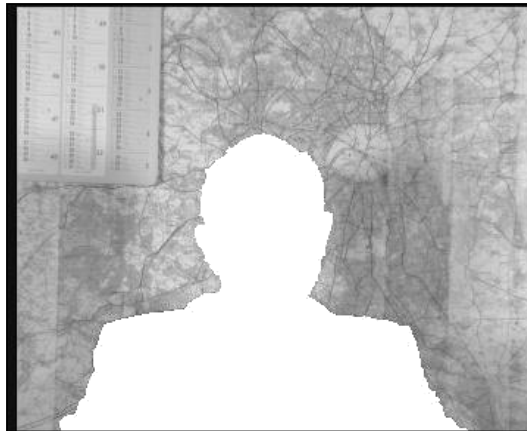
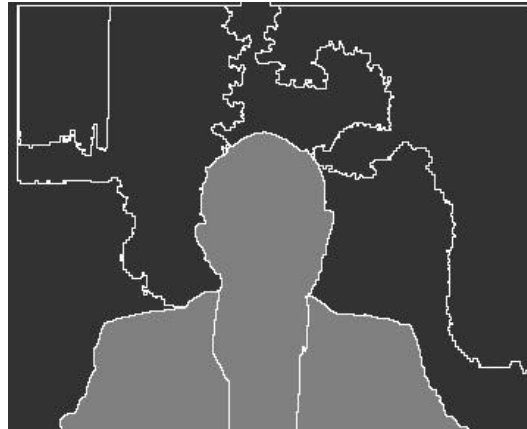


Αντ. πεδίο βάθους

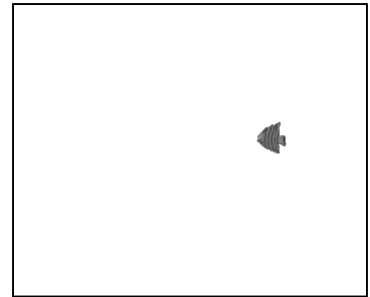
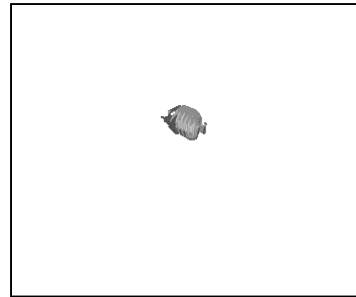
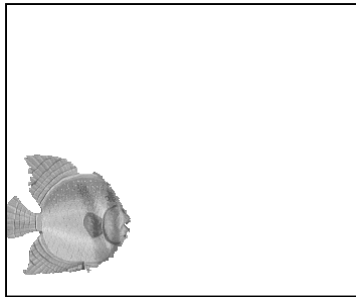
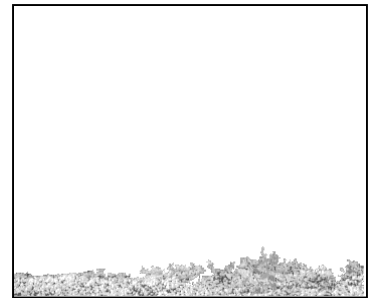
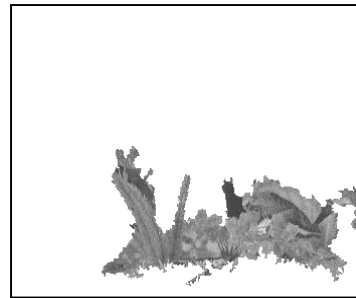
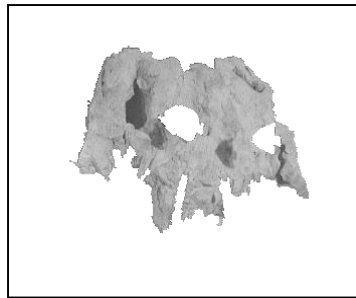
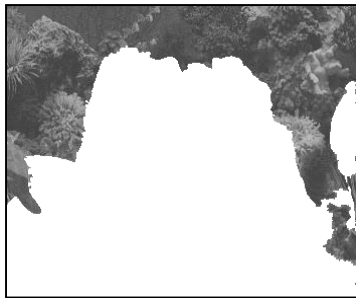
Κατάτμηση με Βάση το Πεδίο Βάθους



Συγχώνευση Αντικειμένων



Σημαιολογική Κατάτμηση

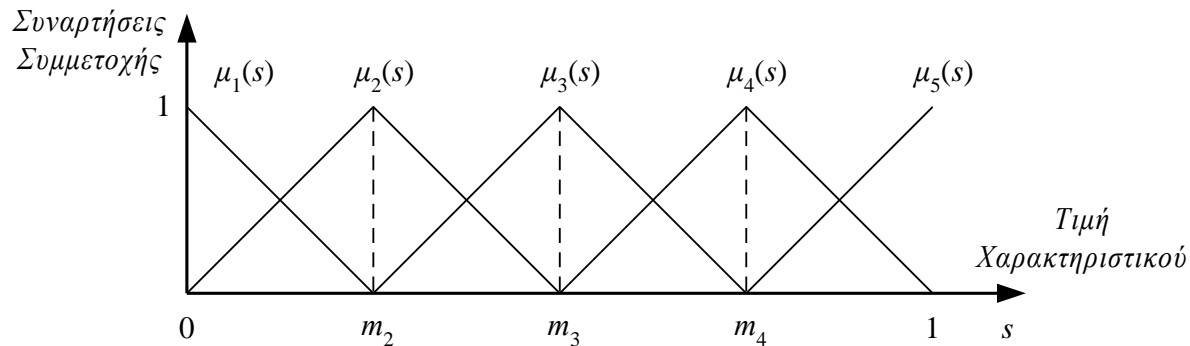


Αναπαράσταση Ακολουθιών Εικόνων

- Υπολογισμός χαρακτηριστικών (πλήθος, θέση, μέγεθος, χρώμα, κίνηση, βάθος) αντικειμένων
- Ταξινόμηση αντικειμένων ως προς τα χαρακτηριστικά τους σε προκαθορισμένες κλάσεις: πολυδιάστατα ιστογράμματα
- Ανάθεση βαθμού συμμετοχής αντικειμένων σε κλάσεις: ασαφής ταξινόμηση
- Κατασκευή πολυδιάστατων διανυσμάτων περιγραφής
- Επιλογή χαρακτηριστικών πλάνων / καρτέ

Μονοδιάστατη Ταξινόμηση

- Μονοδιάστατο χαρακτηριστικό $s \in [0,1]$
- Διαμέριση πεδίου τιμών μέσω Q συναρτήσεων συμμετοχής $\mu_n(s)$, $n = 1,2,\dots,Q$



- Υπολογισμός ασαφούς ιστογράμματος από τα δείγματα s_i , $i = 1,2,\dots,K$:

$$H(n) = \frac{1}{K} \sum_{i=1}^K \mu_n(s_i), \quad n = 1,2,\dots,Q$$

Πολυδιάστατη Ταξινόμηση

- Υπολογισμός διανύσματος χαρακτηριστικών \mathbf{s}_i διαστάσεων $L \times 1$ για κάθε τμήμα S_i , $i = 1, 2, \dots, K$, μιας εικόνας: $\mathbf{s}_i = [\mathbf{c}^T(S_i) \mathbf{I}^T(S_i) a(S_i)]^T = [s_{i,1} \ s_{i,2} \ \dots \ s_{i,L}]^T$
- Διαμέριση πεδίου τιμών κάθε στοιχείου μέσω Q συναρτήσεων συμμετοχής $\mu_{n_j}(s_{i,j})$, $n_j = 1, 2, \dots, Q$
- Βαθμός συμμετοχής διανύσματος \mathbf{s}_i στην πολυδιάστατη κλάση $\mathbf{n} = [n_1 \ n_2 \ \dots \ n_L]^T$:

$$\mu_{\mathbf{n}}(\mathbf{s}_i) = \prod_{j=1}^L \mu_{n_j}(s_{i,j}), \quad i = 1, 2, \dots, K$$

Διανύσματα Περιγραφής

- Κατασκευή πολυδιάστατου ασαφούς ιστογράμματος από τα δείγματα $\mathbf{s}_i, i = 1, 2, \dots, K$:

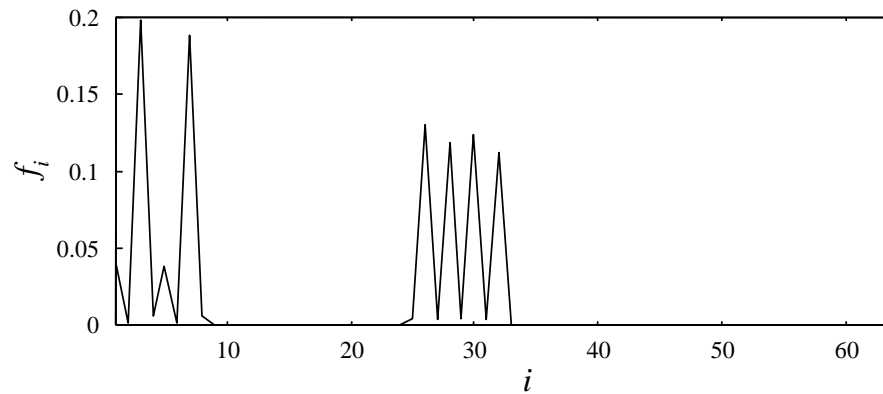
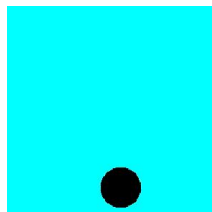
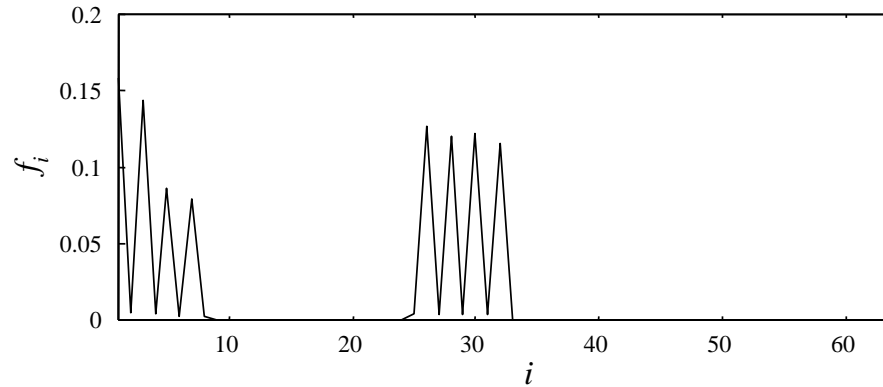
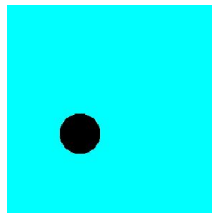
$$H(\mathbf{n}) = \frac{1}{K} \sum_{i=1}^K \mu_{\mathbf{n}}(\mathbf{s}_i) = \frac{1}{K} \sum_{i=1}^K \prod_{j=1}^L \mu_{n_j}(s_{i,j})$$

- Ορισμός συνάρτησης δεικτών:

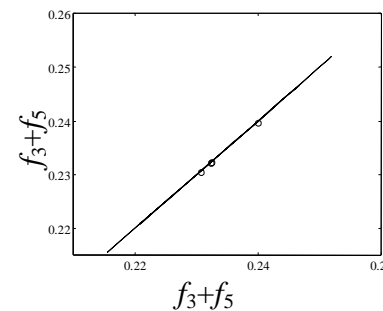
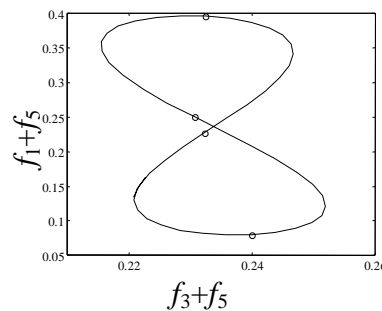
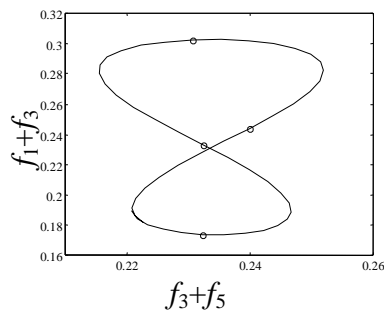
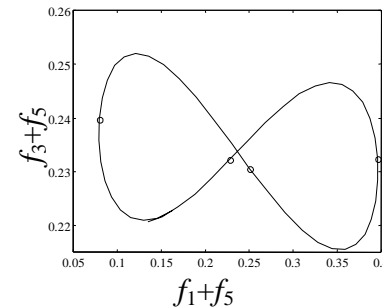
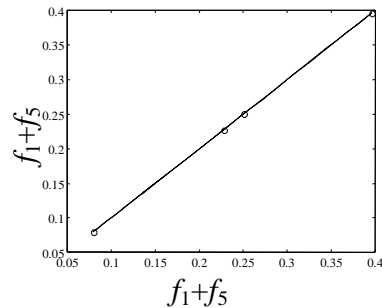
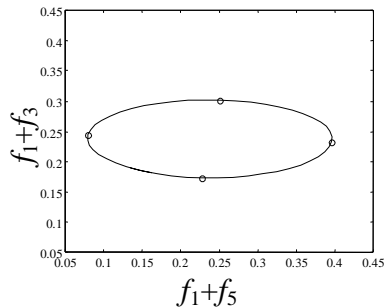
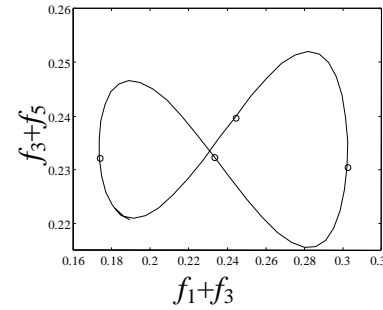
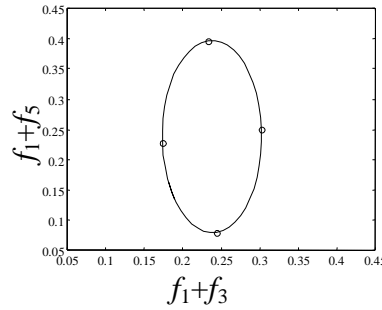
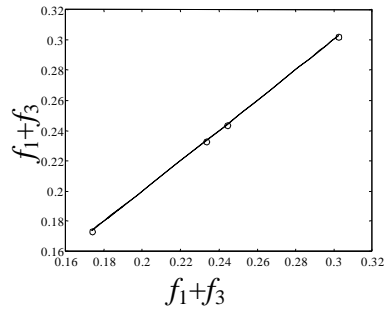
$$z(\mathbf{n}) = 1 + \sum_{j=1}^L n_j Q^{L-j}$$

- Υπολογισμός διανύσματος περιγραφής, μήκους $C = Q^L$, $\mathbf{f} = [f_1 f_2 \dots f_C]^T$, έτσι ώστε $f_{z(\mathbf{n})} = H(\mathbf{n})$ για κάθε κλάση \mathbf{n}

Παράδειγμα Συνθετικής Ακολουθίας (1)



Παράδειγμα Συνθετικής Ακολουθίας (2)

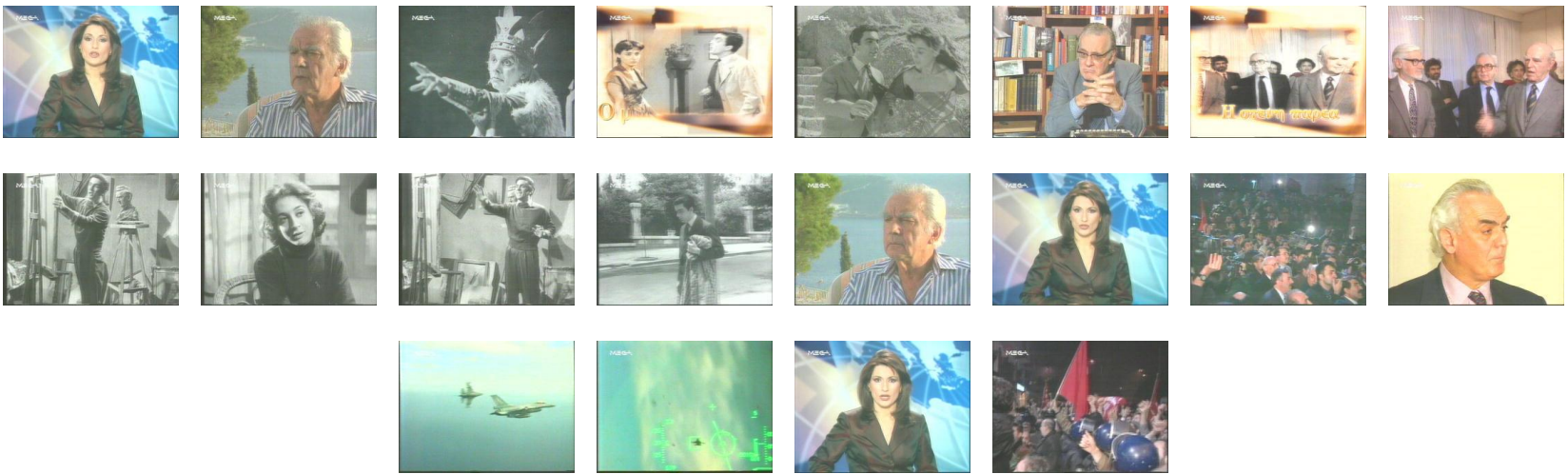


Επιλογή Χαρακτηριστικών Πλάνων

- *Ανίχνευση αλλαγής πλάνων μέσω σφάλματος εκτίμησης κίνησης*
- *Χαρακτηρισμός πλάνων με υπολογισμό διανυσμάτων περιγραφής*
- *Ομαδοποίηση (clustering) πλάνων παρόμοιου οπτικού περιεχομένου στο χώρο των διανυσμάτων περιγραφής*
- *Υλοποίηση με τον αλγόριθμο Lloyd-Max ή K-means*
- *Επιλογή ενός αντιπροσωπευτικού πλάνου από κάθε ομάδα*

Παράδειγμα: Ακολουθία Δελτίου Ειδήσεων

- Ακολουθία δελτίων ειδήσεων διάρκειας 2.5 λεπτών (3750 καρέ) που αποτελείται από $N_S = 20$ πλάνα
- Χρήση $Q = 2$ τριγωνικών συναρτήσεων συμμετοχής με 50% επικάλυψη
- Επιλογή $K_S = 4$ αντιπροσωπευτικών πλάνων



Ομαδοποίηση Πλάνων

Ομάδα 1



Ομάδα 2



Ομάδα 3



Ομάδα 4



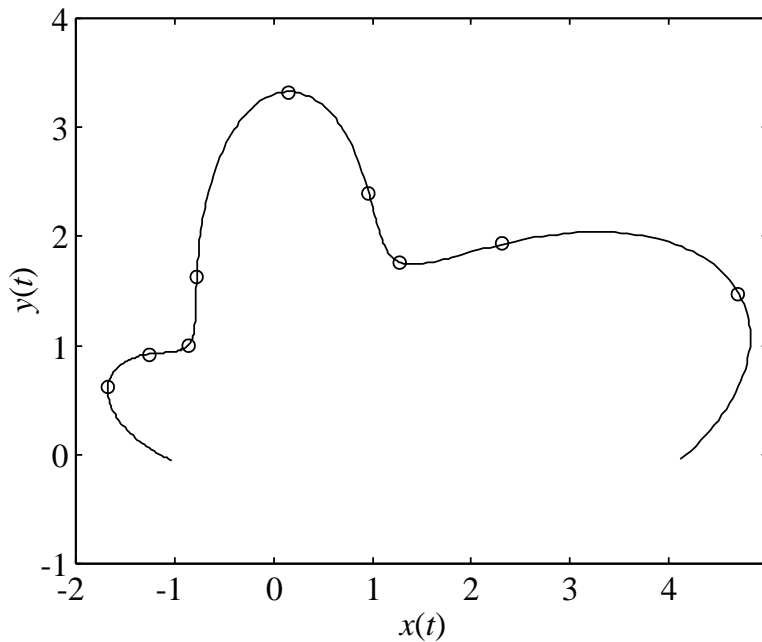
Χαρακτηριστικά πλάνα:



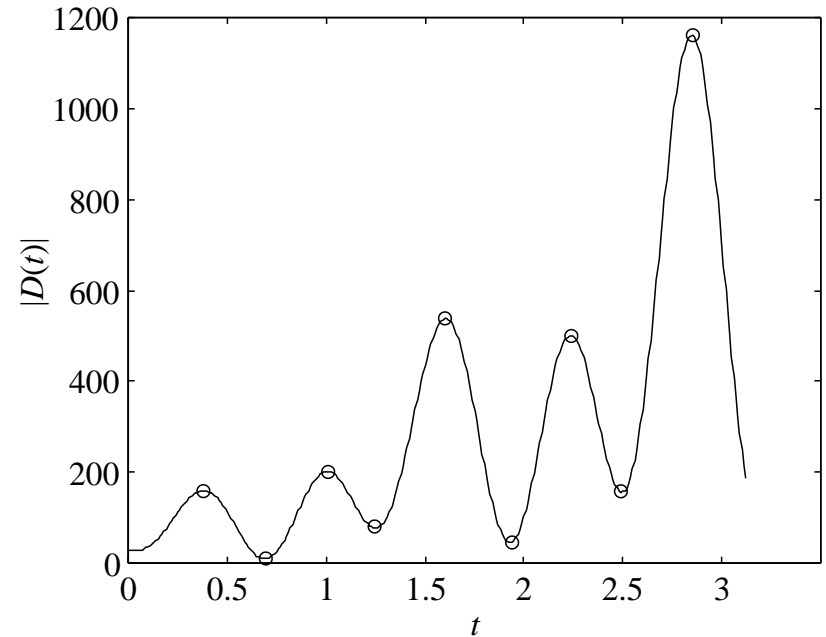
Επιλογή Χαρακτηριστικών Καρέ με Βάση τη Χρονική Μεταβολή

- Με τη χρονική μεταβολή του διανύσματος περιγραφής σχηματίζεται μία τροχιά σε ένα πολυδιάστατο χώρο
- Επιλογή χαρακτηριστικών καρέ (key frames) σε χρονικές στιγμές που αντιστοιχούν σε ακραία σημεία της τροχιάς
- Χρήση μέτρου καμπυλότητας για τον εντοπισμό ακραίων σημείων: μέτρο 2ης παραγώγου, $|D(t)|$
- Χρονικό φιλτράρισμα για την απομάκρυνση του θορύβου

Εντοπισμός Ακραίων Σημείων

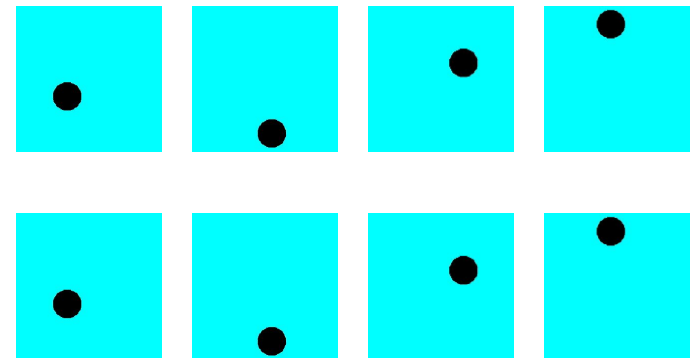
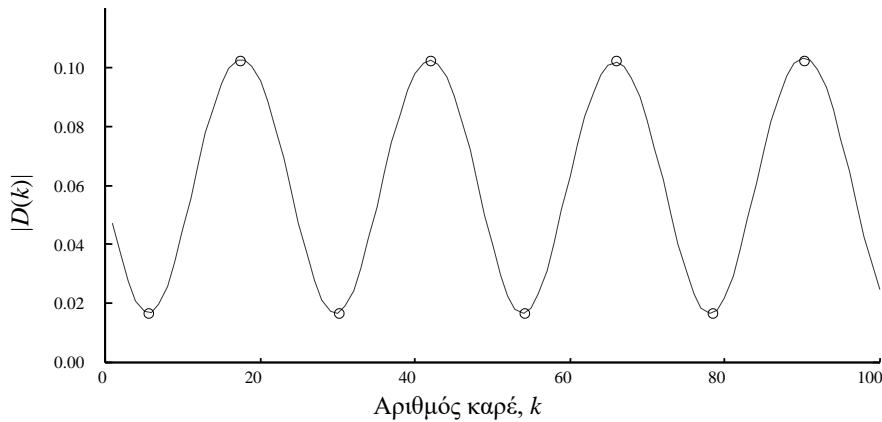
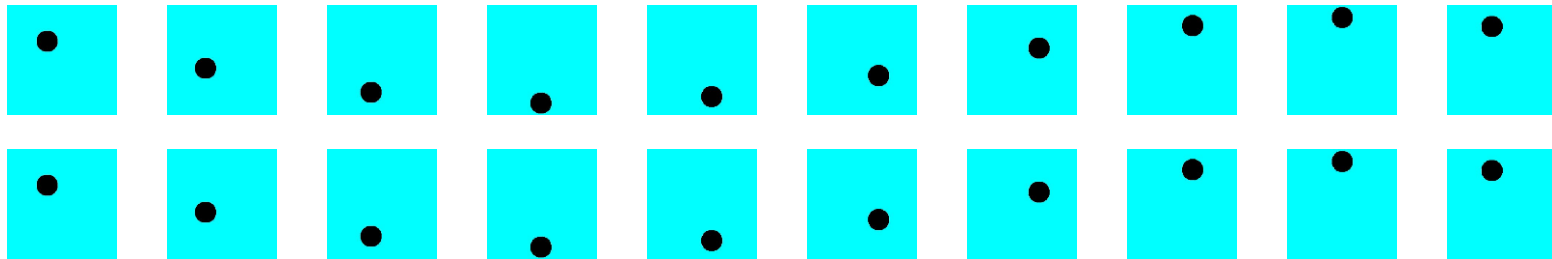


2-Δ συνεχής καμπύλη



Μέτρο καμπυλότητας

Παράδειγμα Συνθετικής Ακολουθίας (3)



Μέτρο καμπυλότητας

Χαρακτηριστικά καρέ

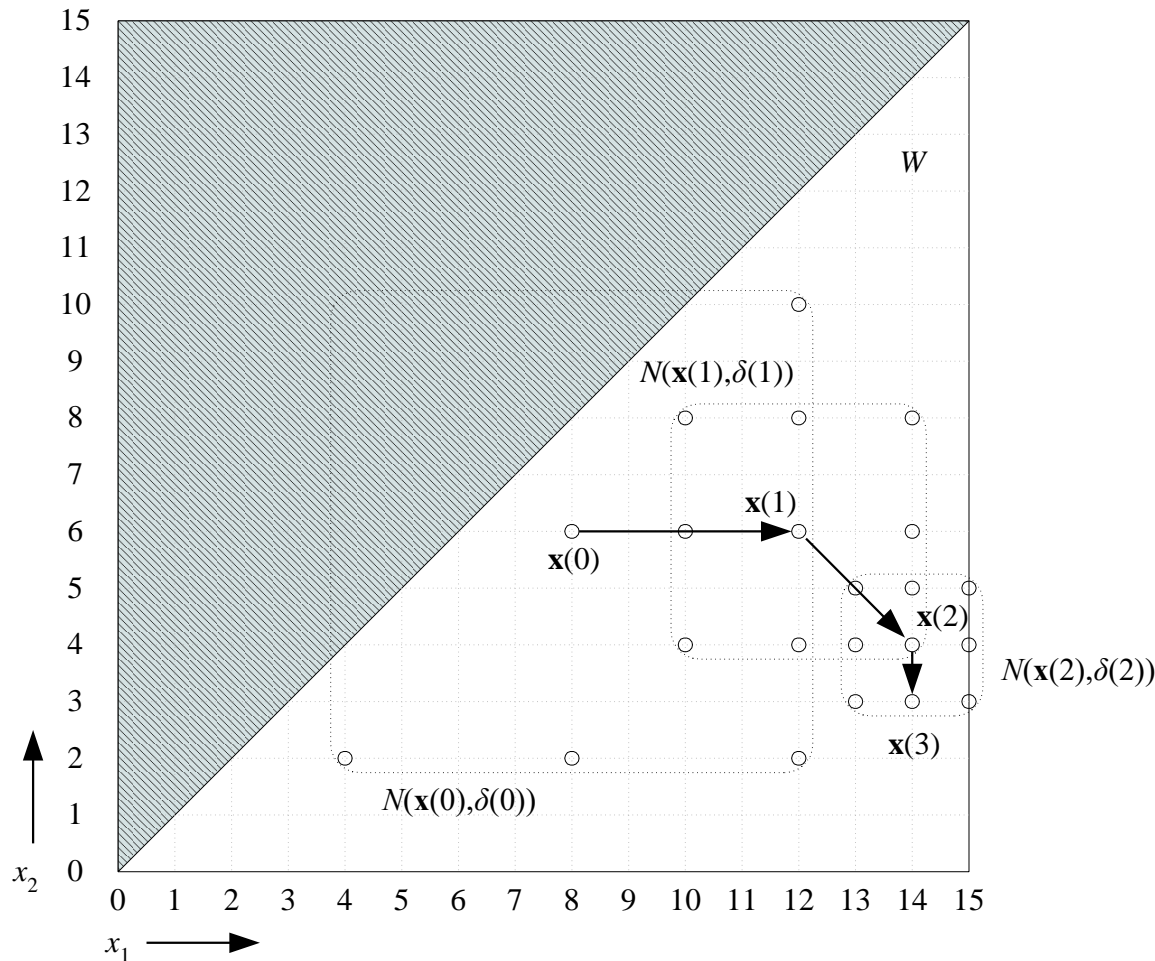
Βέλτιστη Επιλογή Χαρακτηριστικών Καρέ

- *Μέθοδος χρονικής μεταβολής: ανίχνευση πολλαπλών επαναλήψεων, αλλά διατήρηση περιττής πληροφορίας*
- *Μέθοδος συσχέτισης: βέλτιστη επιλογή καρέ που χαρακτηρίζουν το οπτικό περιεχόμενο και είναι ανόμοια μεταξύ τους, λαμβάνοντας υπόψη τη χρονική συσχέτιση των διανυσμάτων περιγραφής*
- *Συνδυαστικό πρόβλημα με απαγορευτικό υπολογιστικό κόστος: υλοποίηση με λογαριθμική αναζήτηση και γενετικό αλγόριθμο*

Μέθοδος Συσχέτισης

- Πλάνο μήκους N καρέ, επιλογή K χαρακτηριστικών
- Διάνυσμα περιγραφής k -οστού καρέ: $\mathbf{f}(k)$, $k \in V$, όπου $V = \{0, 1, \dots, N-1\}$
- Συσχέτιση διανυσμάτων περιγραφής $\mathbf{f}(k)$, $\mathbf{f}(l)$:
 $\rho_{k,l} = C_{k,l} / (\sigma_k \sigma_l)$ όπου $C_{k,l} = (\mathbf{f}(k) - \mathbf{m})^T (\mathbf{f}(l) - \mathbf{m})$, $k, l \in V$
- Διάνυσμα δεικτών: $\mathbf{x} = (x_1, \dots, x_K) \in W \subset V^K$, όπου $W = \{(x_1, \dots, x_K) \in V^K : x_1 < \dots < x_K\}$
- Μέτρο συσχέτισης:
$$R(\mathbf{x}) = R(x_1, \dots, x_K) = \frac{2}{K(K-1)} \sum_{i=1}^{K-1} \sum_{j=i+1}^K (\rho_{x_i, x_j})^2$$
- Ελαχιστοποίηση: $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_K) = \arg \min_{\mathbf{x} \in W} R(\mathbf{x})$

Λογαριθμική Αναζήτηση



Αναζήτηση με Γενετικό Αλγόριθμο

- *Λογαριθμική αναζήτηση*: υπο-βέλτιστη λύση, συχνός εγκλωβισμός σε τοπικά ελάχιστα
- *Γενετικός αλγόριθμος*: καθοδηγούμενη τυχαία αναζήτηση, παράλληλη αναζήτηση σε διαφορετικές περιοχές, με ένταση ανάλογη της πιθανότητας εύρεσης λύσης, μικρό υπολογιστικό κόστος
- Αναπαράσταση δυνατών λύσεων με *χρωμοσώματα*, κωδικοποίηση γονιδίων με διανύσματα δεικτών, κατασκευή αρχικού πληθυσμού
- Επαναληπτική εφαρμογή εκτίμησης ικανότητας (μέτρο συσχέτισης), επιλογής γονέων, ανταλλαγής γονιδίων, μετάλλαξης, αντικατάστασης πληθυσμού

Παράδειγμα Επιλογής Χαρακτηριστικών Καρέ

- Πλάνο ακολουθίας, μήκους $N = 223$ καρέ
Επιλογή $K = 4$ χαρακτηριστικών καρέ



Χαρακτηριστικά Καρέ



Log

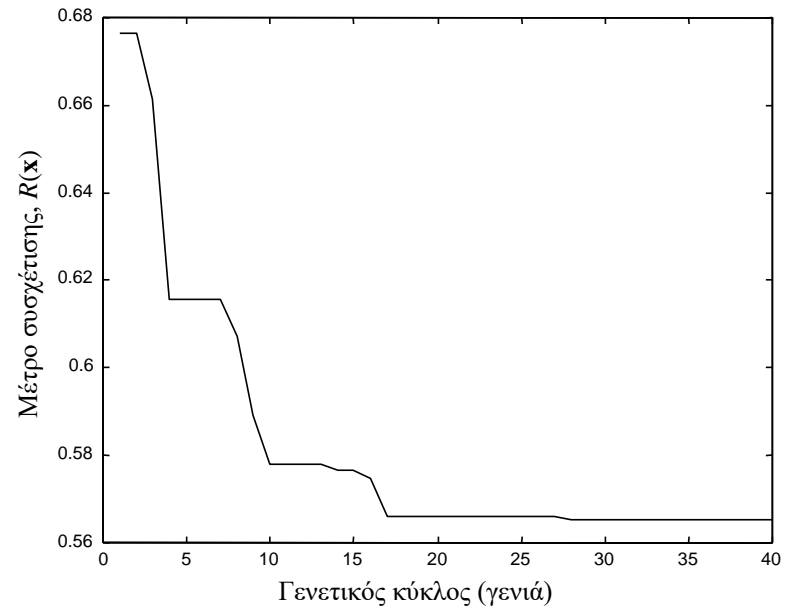
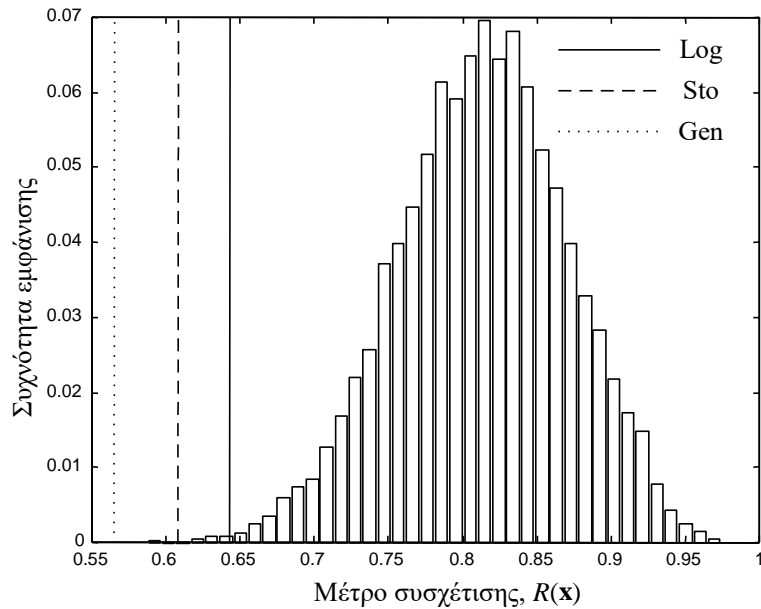


Sto



Gen

Μέτρο Συσχέτισης



Πειραματικά Αποτελέσματα

Αλγόριθμος	\bar{R}	\bar{T} (sec)
Ομοιόμορφη επιλογή	0.85	–
Τυχαία αναζήτηση	0.52	55.38
Λογαριθμική αναζήτηση	0.63	1.92
Στοχαστική αναζήτηση	0.59	12.43
Γενετικός αλγόριθμος	0.44	0.54

Αναζήτηση με Βάση το Περιεχόμενο

- Αναπαράσταση περιεχομένου με διανύσματα περιγραφής χαρακτηριστικών πλάνων / καρτέ
- Μειωμένες απαιτήσεις αποθήκευσης / μετάδοσης
- Ιεραρχική προεπισκόπηση σε επίπεδο ακολουθίας, πλάνων ή καρτέ
- Η αναζήτηση από ακολουθίες βίντεο ανάγεται σε αναζήτηση ακίνητων εικόνων
- Αναζήτηση βάσει παραδείγματος ή σχεδιαγράμματος
- Χρήση παραμετρικής απόστασης: *ανάδραση*

$$d_{\mathbf{w}}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^N w_j (x_j - y_j)^2 = \sum_{j=1}^N w_j e_j^2$$

Παράδειγμα Αναζήτησης

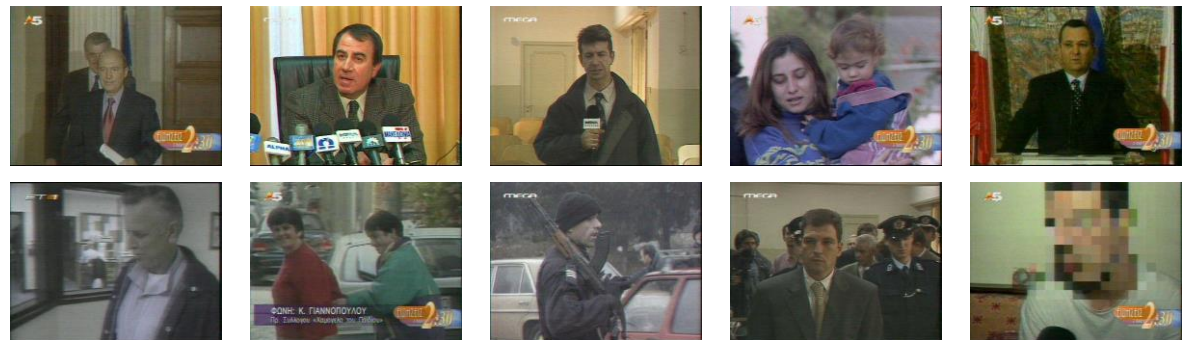
Εικόνα εισόδου



Αρχική ανάκληση



Ανάκληση με ανάδραση



Περιγραφή Σχήματος Αντικειμένων

- Ισχυρό εργαλείο στην αναζήτηση βάσει σχεδιαγράμματος
- Υπάρχουσα πληροφορία σχήματος στα σύγχρονα πρότυπα κωδικοποίησης με βάση τα αντικείμενα
- Αναπαράσταση περιγραμμάτων: chain coding, polygonal approximation, medial axis, Fourier descriptors, curvature scale spaces (MPEG-7), interest points, Legendre descriptors, Zernike moments
- Περιγραφή αναλλοίωτη στους *affine* μετασχηματισμούς:
 - Ενσωμάτωση στη διαδικασία αναγνώρισης / ταιριάσματος
 - Ταίριασμα καμπυλών με βελτιστοποίηση *affine* παραμέτρων
- Τεχνικές κανονικοποίησης

Κανονικοποίηση Περιγραμμάτων

- *Μοντελοποίηση με B-Splines*: αφαίρεση θορύβου, ομοιόμορφη δειγματοληψία ως προς μήκος τόξου
- *Κανονικοποίηση σε «πρότυπη» θέση, αναλλοίωτη σε affine μετασχηματισμούς & μετασχηματισμούς παραμέτρου, χωρίς καμία απώλεια πληροφορίας*
- *Μοναδικότητα / πληρότητα*
- *Αμελητέο υπολογιστικό κόστος*
- *Ταίριασμα χωρίς ταυτόχρονη γνώση δύο καμπυλών*
- *Εφαρμογή ως στάδιο προεπεξεργασίας με οποιαδήποτε τεχνική αναπαράστασης, ταξινόμησης, αναγνώρισης, ή ανάκλησης*

Ορθοκανονικοποίηση (1)

- 2-Δ κλειστή καμπύλη $\mathbf{s} = [\mathbf{s}_0 \dots \mathbf{s}_{N-1}]$ ($2 \times N$), αποτελείται από τα σημεία $\mathbf{s}_i = [x_i \ y_i]^T$, $i = 0, \dots, N-1$
- Συντεταγμένες $\mathbf{x} = [x_0 \dots x_{N-1}]$, $\mathbf{y} = [y_0 \dots y_{N-1}]$ ($1 \times N$)
- Ροπή βαθμού (p, q) :

$$m_{p,q}(\mathbf{s}) = \frac{1}{N} \sum_{i=0}^{N-1} x_i^p y_i^q$$

- Κανονικοποίηση κέντρου βάρους:
 $\mathbf{x}_1 = \mathbf{x} - \mu_x$, $\mathbf{y}_1 = \mathbf{y} - \mu_y$, όπου $\mu_x = m_{1,0}(\mathbf{s})$, $\mu_y = m_{0,1}(\mathbf{s})$
- 1η κανονικοποίηση κλίμακας: $\mathbf{x}_2 = \sigma_x \mathbf{x}_1$, $\mathbf{y}_2 = \sigma_y \mathbf{y}_1$,
όπου $\sigma_x = 1/\sqrt{m_{2,0}(\mathbf{s}_1)}$ $\sigma_y = 1/\sqrt{m_{0,2}(\mathbf{s}_1)}$

Ορθοκανονικοποίηση (2)

- Αριστερόστροφη περιστροφή κατά $\theta_0 = \pi/4$:

$$\mathbf{s}_3 = \mathbf{R}_{\pi/4} \mathbf{s}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{x}_2 - \mathbf{y}_2 \\ \mathbf{x}_2 + \mathbf{y}_2 \end{bmatrix}$$

- 2η κανονικοποίηση κλίμακας: $\mathbf{x}_4 = \tau_x \mathbf{x}_3$, $\mathbf{y}_4 = \tau_y \mathbf{y}_3$,
όπου $\tau_x = 1/\sqrt{m_{2,0}(\mathbf{s}_3)}$ $\tau_y = 1/\sqrt{m_{0,2}(\mathbf{s}_3)}$
- Ισοδύναμα,

$$n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s})(\mathbf{s} - \boldsymbol{\mu}(\mathbf{s})) = \frac{1}{\sqrt{2}} \begin{bmatrix} \tau_x & 0 \\ 0 & \tau_y \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \left(\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix} \right)$$

- Συνθήκες ορθοκανονικοποίησης:
 - $m_{1,0}(n_a(\mathbf{s})) = m_{0,1}(n_a(\mathbf{s})) = m_{1,1}(n_a(\mathbf{s})) = 0$
 - $m_{2,0}(n_a(\mathbf{s})) = m_{0,2}(n_a(\mathbf{s})) = 1$

Κανονικοποίηση Μετατόπισης, Κλίμακας και Κλίσης (Skew)

- Έστω δύο καμπύλες \mathbf{s}, \mathbf{s}' που σχετίζονται μέσω ενός affine μετασχηματισμού:

$$\mathbf{s}' = \mathbf{A}\mathbf{s} + \mathbf{t} = \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

- Υπάρχει 2×2 ορθογώνιος πίνακας \mathbf{Q} τέτοιος ώστε $n_a(\mathbf{s}') = \mathbf{Q} n_a(\mathbf{s})$ (με $\det \mathbf{Q} = \pm 1$)
- Ο affine μετασχηματισμός ανάγεται σε *ορθογώνιο*, που περιέχει μόνο περιστροφή / κατοπτρισμό
- Κανονικοποίηση χωρίς γνώση των παραμέτρων του μετασχηματισμού, χωρίς ταίριασμα των \mathbf{s}, \mathbf{s}'
- Καμία απώλεια πληροφορίας

Κανονικοποίηση Σημείου Αναφοράς

- Μιγαδική αναπαράσταση $\mathbf{z} = \mathbf{x} + j\mathbf{y} = [z_0 \dots z_{N-1}]^T$ ($N \times 1$), όπου $z_i = x_i + jy_i$, $i = 0, 1, \dots, N-1$
- Μετασχηματισμός *Fourier* της καμπύλης: $u_k = \sum_{i=0}^{N-1} z_i \omega^{-ki}$
- Πρωτεύον όρισμα: $a_k = \text{Arg } u_k$, $k = 0, 1, \dots, N-1$
- Πρότυπη κυκλική ολίσθηση:

$$p(\mathbf{z}) = \left\lfloor \frac{N}{4\pi} (a_1 - a_{N-1}) \right\rfloor \bmod N / 2$$

- Κανονικοποίηση: $n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z})$
- Έστω $\mathbf{z}' = S_m(\mathbf{z})$: $z'_i = z_{(i+m) \bmod N}$, $i = 0, \dots, N-1$. Τότε:
 - $p(n_p(\mathbf{z}')) = p(n_p(\mathbf{z})) = 0$
 - $n_p(\mathbf{z}') = n_p(\mathbf{z})$ ή $n_p(\mathbf{z}') = S_{N/2}(n_p(\mathbf{z}))$

Κανονικοποίηση Περιστροφής & Κατοπτρισμού

- Κανονικοποίηση περιστροφής: $\mathbf{z}_1 = \mathbf{z} e^{-jr(\mathbf{z})}$ όπου

$$r(\mathbf{z}) = \left(\frac{1}{2} (a_1 + a_{N-1}) \right) \bmod \pi$$

- Κανονικοποίηση κατοπτρισμού:

- $v(\mathbf{z}_1) = v_x(\mathbf{z}_1) + jv_y(\mathbf{z}_1) = \text{sgn}m_{1,2}(\mathbf{z}_1) + j\text{sgn}m_{2,1}(\mathbf{z}_1)$

- $n_r(\mathbf{z}) = \mathbf{z}_2 = v_x(\mathbf{z}_1) \mathbf{x}_1 + jv_y(\mathbf{z}_1) \mathbf{y}_1$

- Έστω $\mathbf{z}' = (s_x \mathbf{x} + js_y \mathbf{y}) e^{j\theta}$. Τότε:

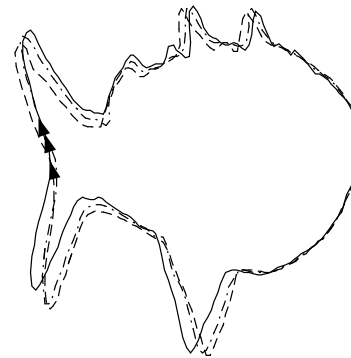
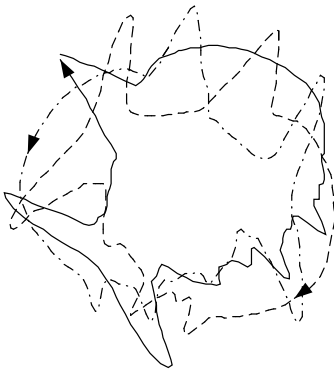
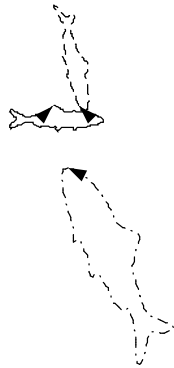
- $r(n_r(\mathbf{z}')) = r(n_r(\mathbf{z})) = 0$

- $v_x(n_r(\mathbf{z}')) = v_y(n_r(\mathbf{z}')) = v_x(n_r(\mathbf{z})) = v_y(n_r(\mathbf{z})) = 1$

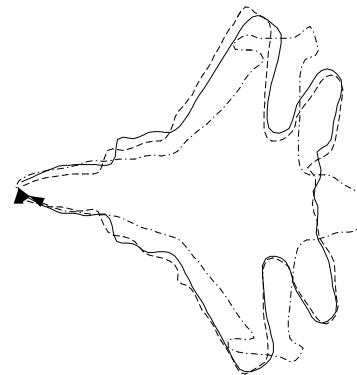
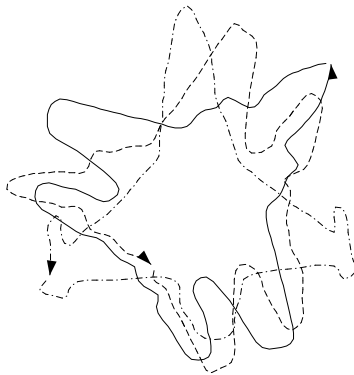
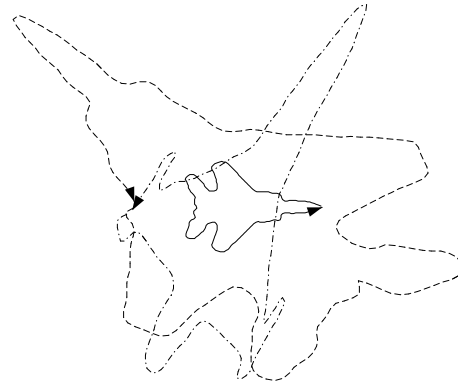
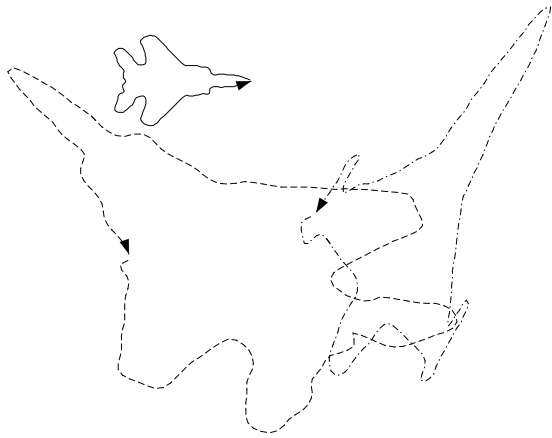
- $n_r(\mathbf{z}') = n_r(\mathbf{z})$

- Κανονικοποίηση φοράς διαγραφής (δεξιόστροφη)

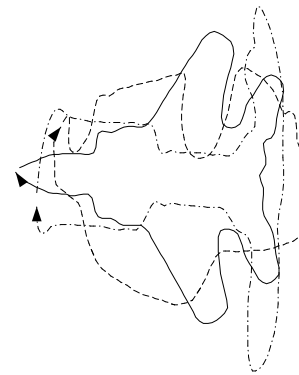
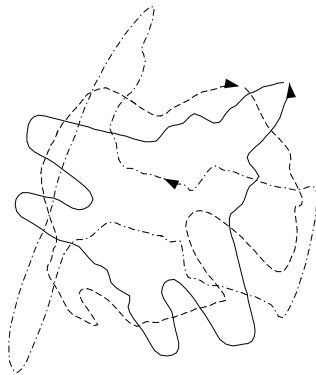
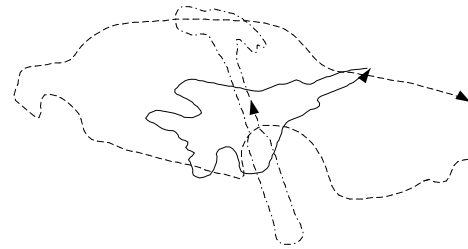
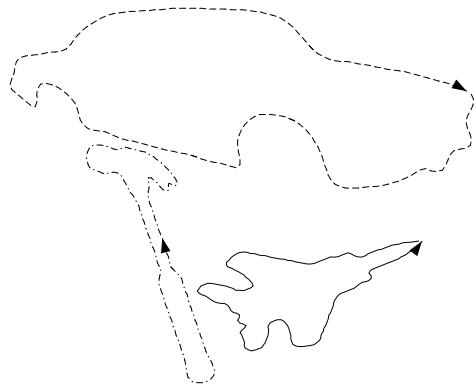
Παράδειγμα: Ίδιο Αντικείμενο



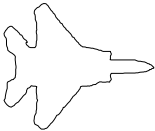

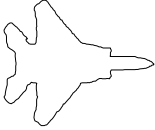
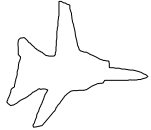
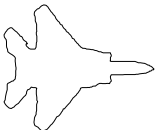

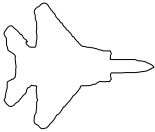

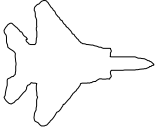

Παράδειγμα: Παρόμοια Αντικείμενα



Παράδειγμα: Ανόμοια Αντικείμενα



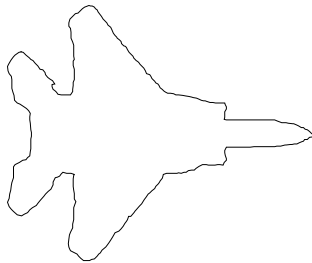
Απόσταση Κανονικοποιημένων Περιγραμμάτων

α/α	Περιγραμμα 1	Περιγραμμα 2	ΑΚΣ (αρχ)	ΑΚΣ (καν)	NFD	MFD
#1			0.69	0.01	0.02	0.01
#2			0.34	0.19	0.12	0.11
#3			0.81	0.75	0.41	0.57
#4			0.93	0.89	0.62	0.65
#5			0.84	0.76	0.25	0.32

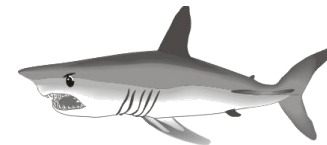
Αναζήτηση με Βάση το Σχήμα



Εικόνα εισόδου



Εξαγόμενο
περίγραμμα



Αποτελέσματα ανάκλησης

Εντοπισμός Προσώπων

- Απαραίτητος στην ανάλυση ακολουθιών δελτίων ειδήσεων
- Συνδεδεμένος με την αναγνώριση προσώπων – υψηλή ακρίβεια & υπολογιστικό κόστος, Α/Μ εικόνες
- Σύγχρονες εφαρμογές ανάλυσης – ταχεία υλοποίηση, ικανοποιητική ακρίβεια, έγχρωμες εικόνες
- Ανίχνευση χρώματος δέρματος: μοντελοποίηση κατανομής χρώματος δέρματος στο χώρο YCrCb
- Γενίκευση μοντέλου: Gaussian mixture models, εκτίμηση ιστογράμματος, μοντέλα Markov
- Συνδυασμός με χαρακτηριστικά σχήματος & υφής

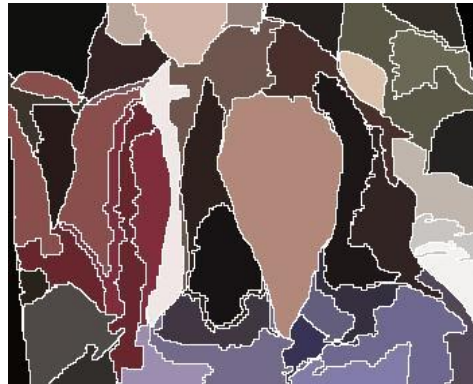
Προτεινόμενη Τεχνική

- Μοντελοποίηση χρώματος δέρματος με 2-Δ κανονική (Gaussian) κατανομή
- Επανεκτίμηση της μέσης τιμής της κατανομής σε ακολουθίες βίντεο
- Αναπροσαρμογή σε μεταβολές φωτισμού & νέα πρόσωπα, χωρίς απόκλιση από το γενικό μοντέλο
- Κατάτμηση με βάση το χρώμα, εκτίμηση πιθανοφάνειας δέρματος, συγχώνευση τμημάτων δέρματος: $\delta_f(X, Y) = [\max(1-p_s(X), 1-p_s(Y))]^2$
- Ανάλυση σχήματος: συμπαγές / επιμήκυνση
- Στάδιο επαλήθευσης: ταίριασμα με πρότυπο

Κατάτμηση – Συγχώνευση Τμημάτων Προσώπου



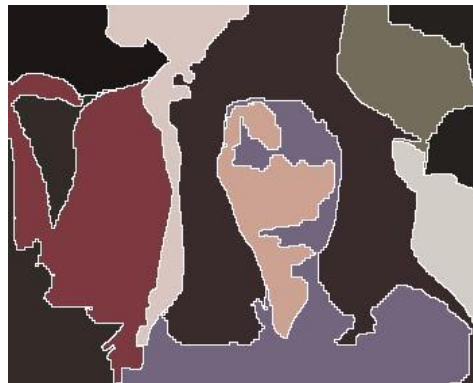
Εικόνα εισόδου



Με συγχώνευση



Υψηλό κατώφλι

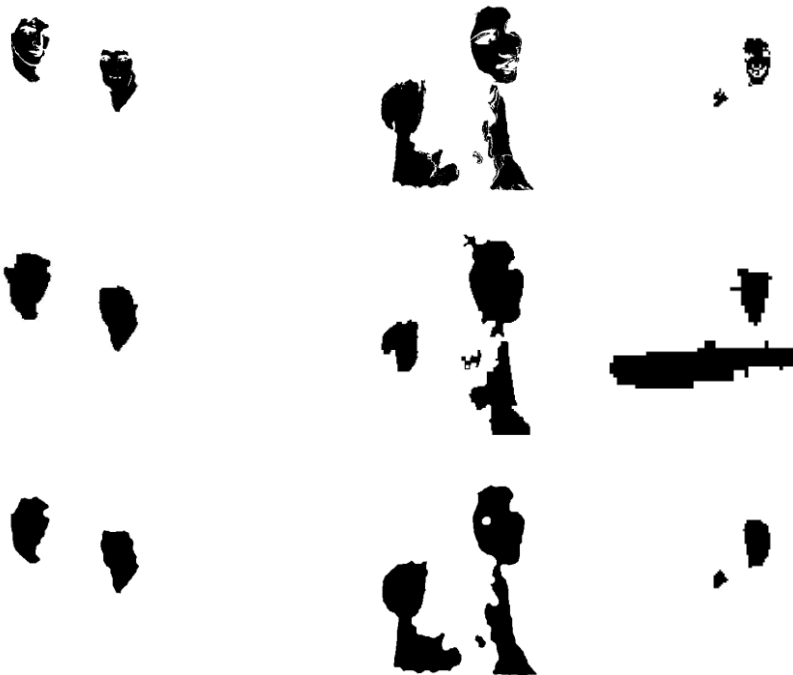


Χαμηλό κατώφλι

Πιθανοφάνεια Δέρματος / Προσώπου



Αποτελέσματα Ανίχνευσης Δέρματος στη Βάση Compaq CRL



Σύνολο (320 εικόνες)

Μέτρηση	P (%)	R (%)
TAO	66.42	85.69
TAP	77.82	85.36
SAT	87.61	93.57
NST	85.77	88.46

	P (%)	R (%)
TAO	79.8	98.1
TAP	90.2	98.2
SAT	100	100
NST	100	100

	P (%)	R (%)
TAO	71.5	63.8
TAP	80.0	64.4
SAT	99.2	55.7
NST	75.0	66.7

	P (%)	R (%)
TAO	10.5	73.2
TAP	13.4	75.2
SAT	18.4	87.5
NST	50.0	50.0

Αποτελέσματα Ανίχνευσης Προσώπων

Μέτρηση	Εικόνες	Ακολουθίες
Συνολικές εικόνες / καρτέ	420	701
Εικόνες με πρόσωπα	305	643
Συνολικά πρόσωπα	409	736
Εντοπισμένα πρόσωπα	379	681
Σωστά εντοπισμένα πρόσωπα	368	667
Ακρίβεια (%)	97.10	97.94
Ανάκληση (%)	89.98	90.63
MET (sec)	1.9	1.5

Συντακτική Ανάλυση Δελτίων Ειδήσεων

- Σημαντική πηγή οπτικοακουστικής πληροφορίας
 - τυπική οργάνωση σε θεματικές ενότητες με επαναλαμβανόμενη εμφάνιση παρουσιαστών, ρεπορτάζ και εξωτερικών λήψεων: *συντακτική ανάλυση*
 - χειρονακτικές μέθοδοι χαρακτηρισμού & δεικτοδότησης: υψηλό κόστος
- Υφιστάμενα συστήματα ανάλυσης & πλοήγησης ειδήσεων, συστήματα News on Demand (NoD):
 - κυρίως *πληροφορία κειμένου* (υπότιτλοι, εγγραφές teletext)
 - απλή επεξεργασία οπτικοακουστικής πληροφορίας (ανίχνευση περιόδων σιωπής / μαύρης οθόνης)
 - ενσωμάτωση πληροφορίας: δυναμικός προγραμματισμός, *finite state machines, hidden Markov Models*

Προτεινόμενη Τεχνική

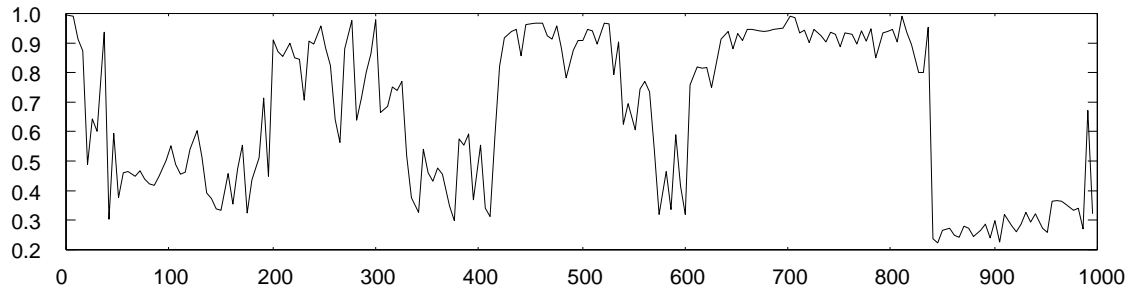
- Ανίχνευση αλλαγής πλάνου
- Εντοπισμός επικρατέστερων προσώπων
- Εκτίμηση κίνησης φόντου
- Ταξινόμηση πλάνων
 - *Παρουσιαστής*: κοντινά πλάνα προσώπου, στατικό φόντο
 - *Ρεπορτάζ / συνέντευξη*: κοντινά πλάνα προσώπου, κιν. φόντο
 - *Στατική εικόνα* (οικονομικά/ αθλητικά ρεπορτάζ, δελτία καιρού): χωρίς πρόσωπα, στατικό φόντο
 - *Εξωτερικό πλάνο*: μικρά ή καθόλου πρόσωπα, κιν. φόντο
- Ομαδοποίηση πλάνων παρουσιαστή
- Χρονική κατάτμηση: στοιχειώδεις ενότητες ειδήσεων

Παράδειγμα Ανάλυσης Δελτίου Ειδήσεων

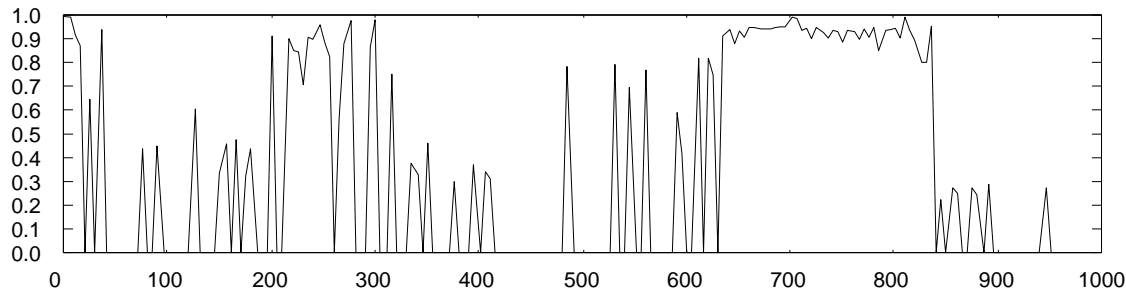
- Ακολουθία διάρκειας 100 sec (1000 καρέ) με 15 πλάνα, 1 πλάνο παρουσιαστή, άσχημες συνθήκες φωτισμού



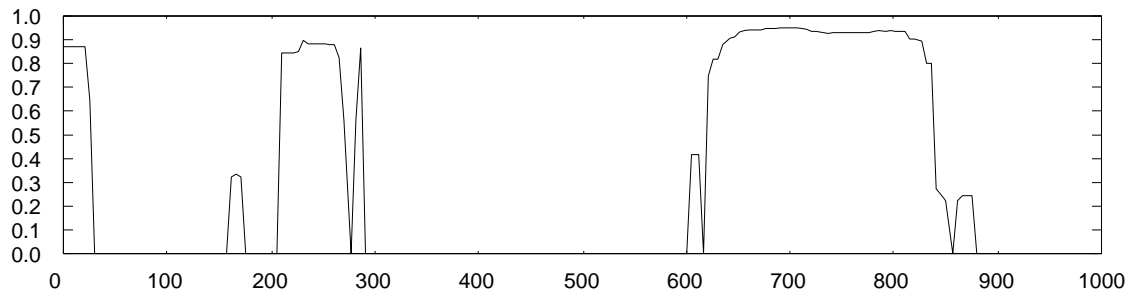
Εντοπισμός Παρουσιαστή / Συνέντευξης



Μέγιστη
πιθανοφάνεια
προσώπου



Μέγιστη
πιθανοφάνεια
επικρατέστερου
προσώπου



Χρονικό
φιλτράρισμα
(median)

Αξιολόγηση

Ακολουθία	Παρουσιαστής		Ρεπορτάζ/ Συνέντευξη		Στατική Εικόνα		Εξωτερικό Πλάνο	
	P	R	P	R	P	R	P	R
A5 (α)	0.92	0.94	0.65	0.82	–	–	0.75	0.85
A5 (β)	0.95	1.00	0.83	0.94	0.50	1.00	0.73	0.87
ANT1	0.93	0.94	0.77	0.88	0.75	0.66	0.85	0.86
ET-1	1.00	1.00	0.71	0.88	0.66	1.00	0.81	0.93
MEGA (α)	0.93	0.93	0.76	0.86	0.75	0.75	0.67	0.86
MEGA (β)	0.96	1.00	0.84	0.91	–	–	0.74	0.81
Σύνολο	0.95	0.97	0.76	0.88	0.67	0.85	0.76	0.86

Ανάκληση με Βάση τα Πρόσωπα (Χρώμα & Κλίμακα Προσώπων)



$s_c = 0.9985$



$s_c = 0.9981$



$s_c = 0.9976$



$s_c = 0.9901$



$s_c = 0.9845$



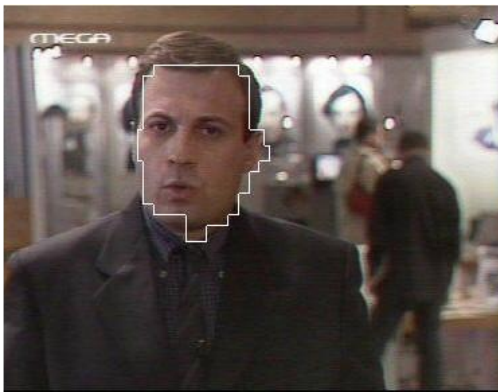
$s_c = 0.9803$



$s_c = 0.9774$



$s_c = 0.9618$



$s_r = 0.9994$



$s_r = 0.9973$



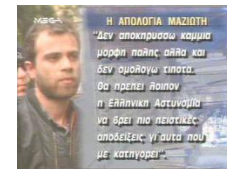
$s_r = 0.9909$



$s_r = 0.9814$



$s_r = 0.9811$



$s_r = 0.9807$



$s_r = 0.9693$



$s_r = 0.9649$

Ανάκληση με Βάση τα Πρόσωπα (Πλήθος Προσώπων)



$s_r = 0.9744$



$s_r = 0.9659$



$s_r = 0.9545$



$s_r = 0.9528$



$s_r = 0.9502$



$s_r = 0.9478$



$s_r = 0.9436$



$s_r = 0.9312$

Πλήθος προσώπων	Κανένα	1	2	3+
Συνολικές εικόνες	62	275	112	51
Ανακαλούμενες εικόνες	66	282	106	46
Σωστά ανακαλούμενες	59	273	101	42
Ακρίβεια (%)	89.34	96.81	93.57	91.30
Ανάκληση (%)	95.16	99.27	87.61	82.35

Συμπεράσματα

- Ικανοποιητική προσέγγιση σημασιολογικών αντικειμένων
- Συμπαγής αναπαράσταση χαρακτηριστικών των αντικειμένων, ταχεία & αξιόπιστη σύγκριση
- Μη γραμμική αναπαράσταση ακολουθιών ως προς το χρόνο – ιεραρχική προεπισκόπηση & πλοήγηση
- Αναζήτηση από βίντεο με βάση το περιεχόμενο, ισοδύναμη με την αναζήτηση ακίνητων εικόνων
- Κανονικοποίηση περιγραμμάτων χωρίς απώλεια πληροφορίας – συνδυασμός με οποιαδήποτε τεχνική
- Ανάλυση δελτίων ειδήσεων με πολύ απλούς κανόνες