

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

# Οπτική Μοντελοποίηση Ανθρώπινου Προσώπου με Εφαρμογές σε Αναγνώριση

Επαμεινώνδας Π. Αντωνάκος

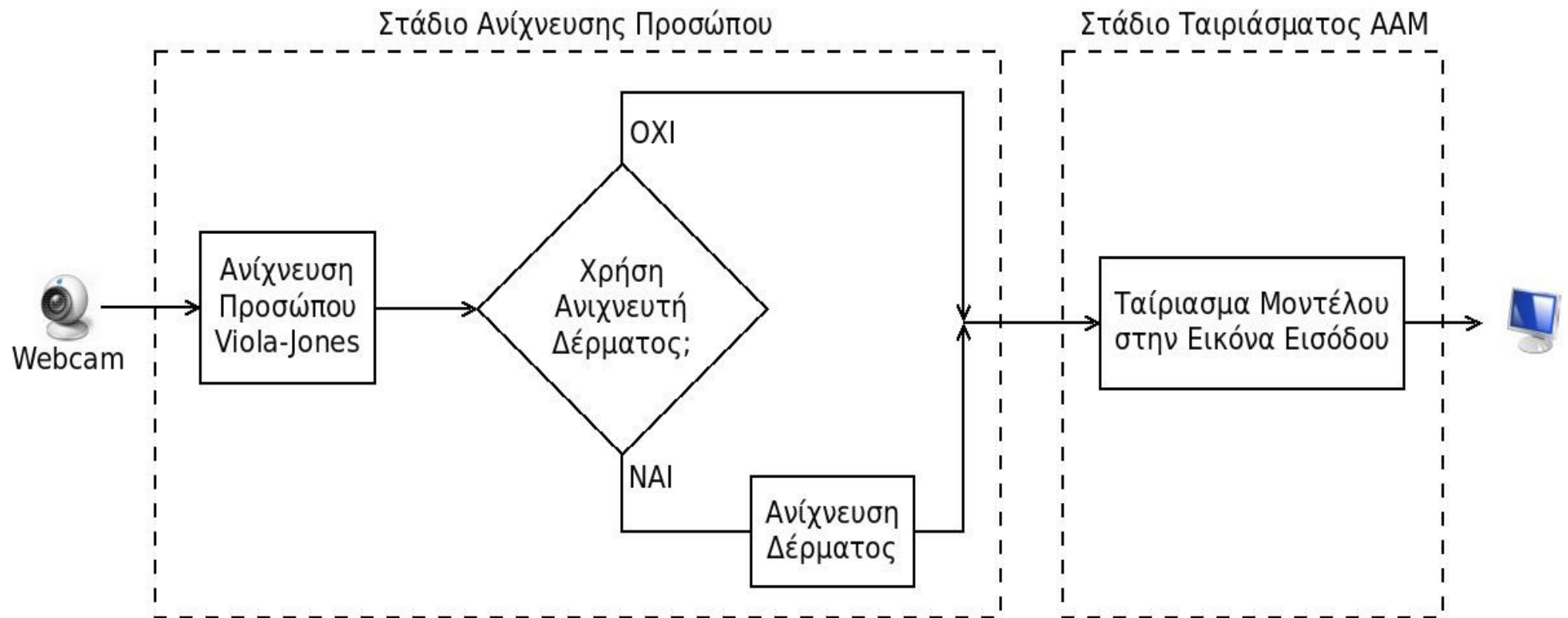
Επιβλέπων: καθ. Πέτρος Μαραγκός

# Ορισμός Προβλήματος

Η Οπτική Μοντελοποίηση είναι το πρόβλημα της δημιουργίας ενός παραμετρικού και προσαρμοζόμενου μοντέλου ανθρώπινου προσώπου και το ταίριασμα του μοντέλου σε νέα πρόσωπα

- Όραση Υπολογιστών και Αναγνώριση Προτύπων
- Αναγνώριση προσώπου, αναγνώριση συναισθημάτων, επικοινωνία ανθρώπου-ρομπότ, εξαγωγή μοντέλων συναισθηματικής προσοχής από δεδομένα ταινιών
- Ασφάλεια αυτοκινήτου, εικονική πραγματικότητα, αυτοματοποιημένη διάγνωση, ρομπότ οικιακής χρήσης

# Διάγραμμα Ροής Συστήματος

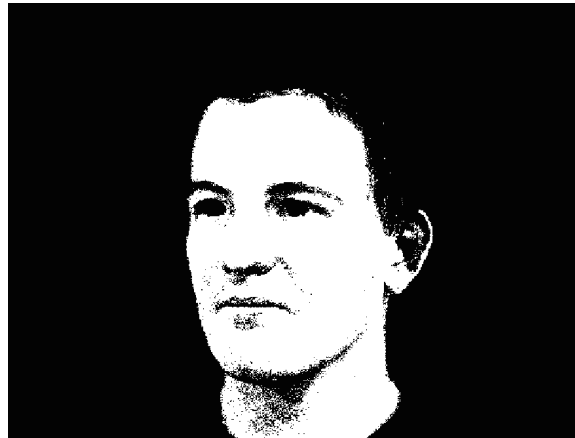


# Διάγραμμα Ροής Συστήματος

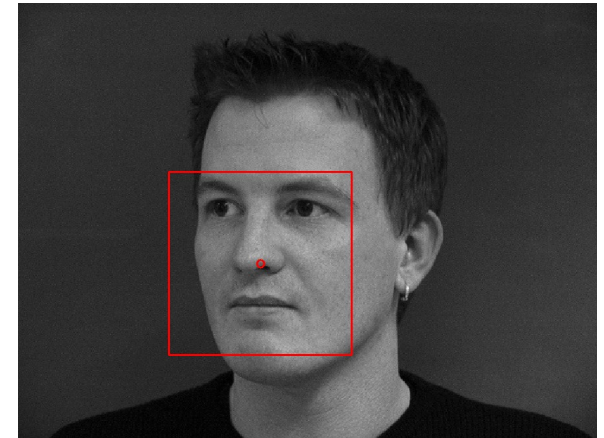
Εικόνα από κάμερα



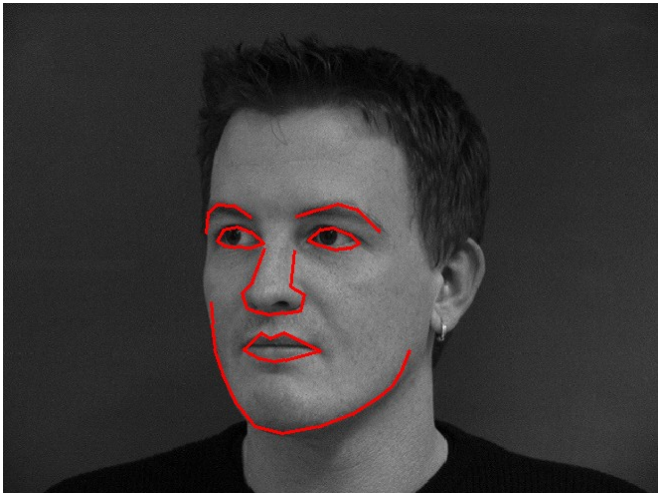
Ανίχνευση δέρματος



Ανίχνευση προσώπου



Τελική εικόνα



Ταίριασμα μοντέλου σχήματος  
και υφής προσώπου



# Δομή της Παρουσίασης

- (1) **Ανίχνευση προσώπου** με προαιρετική ανίχνευση δέρματος σε πραγματικό χρόνο
- (2) **Οπτική μοντελοποίηση προσώπου** με Active Appearance Models σε πραγματικό χρόνο
- (3) Εφαρμογή σε **αναγνώριση** εκφράσεων προσώπου και συναισθημάτων

# Ανίχνευση Προσώπου με προαιρετική Ανίχνευση Δέρματος

# Ανίχνευση Δέρματος με μίγμα Gaussian κατανομών

- Gaussian κατανομή ανθρώπινου δέρματος:

$$P(c|skin) = \frac{1}{2\pi |\Sigma_s|^{\frac{1}{2}}} e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1} (c-\mu_s)}$$

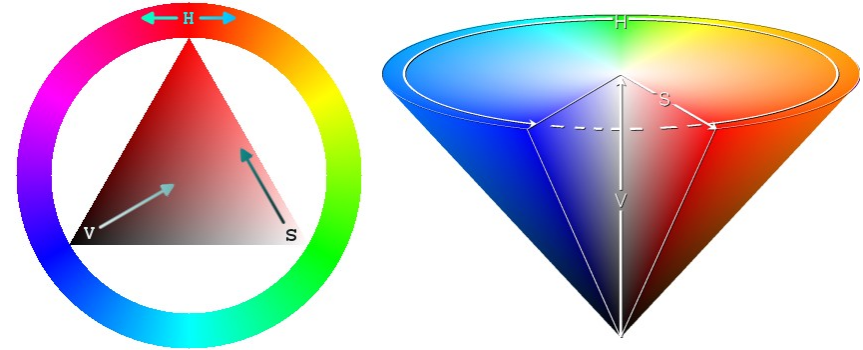
- Μοντέλο μίγματος Gaussian κατανομών με σ.π.π.:

$$P(c|skin) = \sum_{i=1}^k \pi_i P_i(c|skin)$$

- Εκπαίδευση του μοντέλου με k=8 κατανομές και χρήση του αλγορίθμου EM (Expectation-Maximization).
- Ταχύτητα εκτέλεσης: 0.037sec/frame = 27frames/sec

# Ανίχνευση Δέρματος με χρήση χρωματικού χώρου HSV

- Κυκλικός χώρος, απόχρωση (H), κορεσμός (S) και φωτεινότητα (V).



- Σχεδίαση τιμών Hue και Saturation σημείων δέρματος για επιλογή κατάλληλων ορίων.
- Για να είναι ένα σημείο ανθρώπινο δέρμα, πρέπει να έχει τιμές στα όρια:

$$0 \leq H \leq 0.11 \quad \text{και} \quad 0.2 \leq S \leq 0.7$$

- Ταχύτητα εκτέλεσης:  $0.127\text{sec/frame} = 7.9\text{frames/sec}$



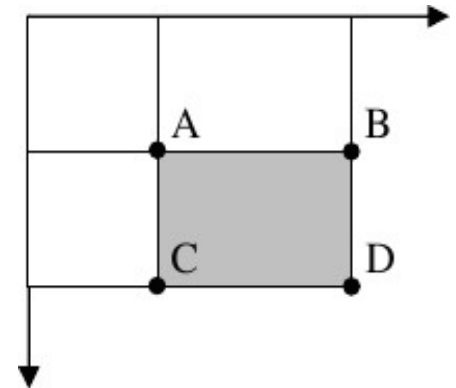
# Σύγκριση μεθόδων Ανίχνευσης Δέρματος

- Η *πιθανοτική μέθοδος* είναι καλύτερη στις περισσότερες συνθήκες φωτισμού και πιο γρήγορη.
- Η *πιθανοτική μέθοδος* έχει μια μεταβλητή κατωφλίου, ενώ η μέθοδος HSV όχι.
- Η μέθοδος HSV λειτουργεί καλύτερα υπό φυσικές συνθήκες έντονου φωτισμού.
- Η μέθοδος HSV υστερεί όταν η εικόνα εισόδου είναι από κάμερα χαμηλής ποιότητας (π.χ. Webcam) γιατί αλλοιώνονται τα χρώματα.

# Ανίχνευση Προσώπου

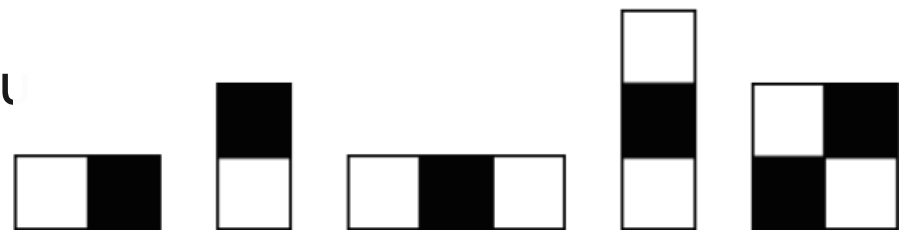
- Αλγόριθμος Viola και Jones, 2001.
- Χρήση **ολοκληρωτικής εικόνας** όπου το άθροισμα των *pixels* ενός ορθογωνίου της υπολογίζεται από τη σχέση:

$$S_{ABCD} = I_{int}(D) - [I_{int}(B) + I_{int}(C)] + I_{int}(A)$$



- *Haar Χαρακτηριστικά*

Απλά ορθογώνια σχήματα που η τιμή τους ορίζεται ως η διαφορά των αθροίσματος των *pixels* της λευκής από τη μαύρη περιοχή.



- Η ανίχνευση προσώπων σε όλες τις κλίμακες δεν γίνεται με την υπολογιστικά ακριβή κλιμάκωση της εικόνας, αλλά με την κλιμάκωση των χαρακτηριστικών, οπότε προκύπτουν ~160000 χαρακτηριστικά.

# Ανίχνευση Προσώπου

- *Αλγόριθμος AdaBoost*

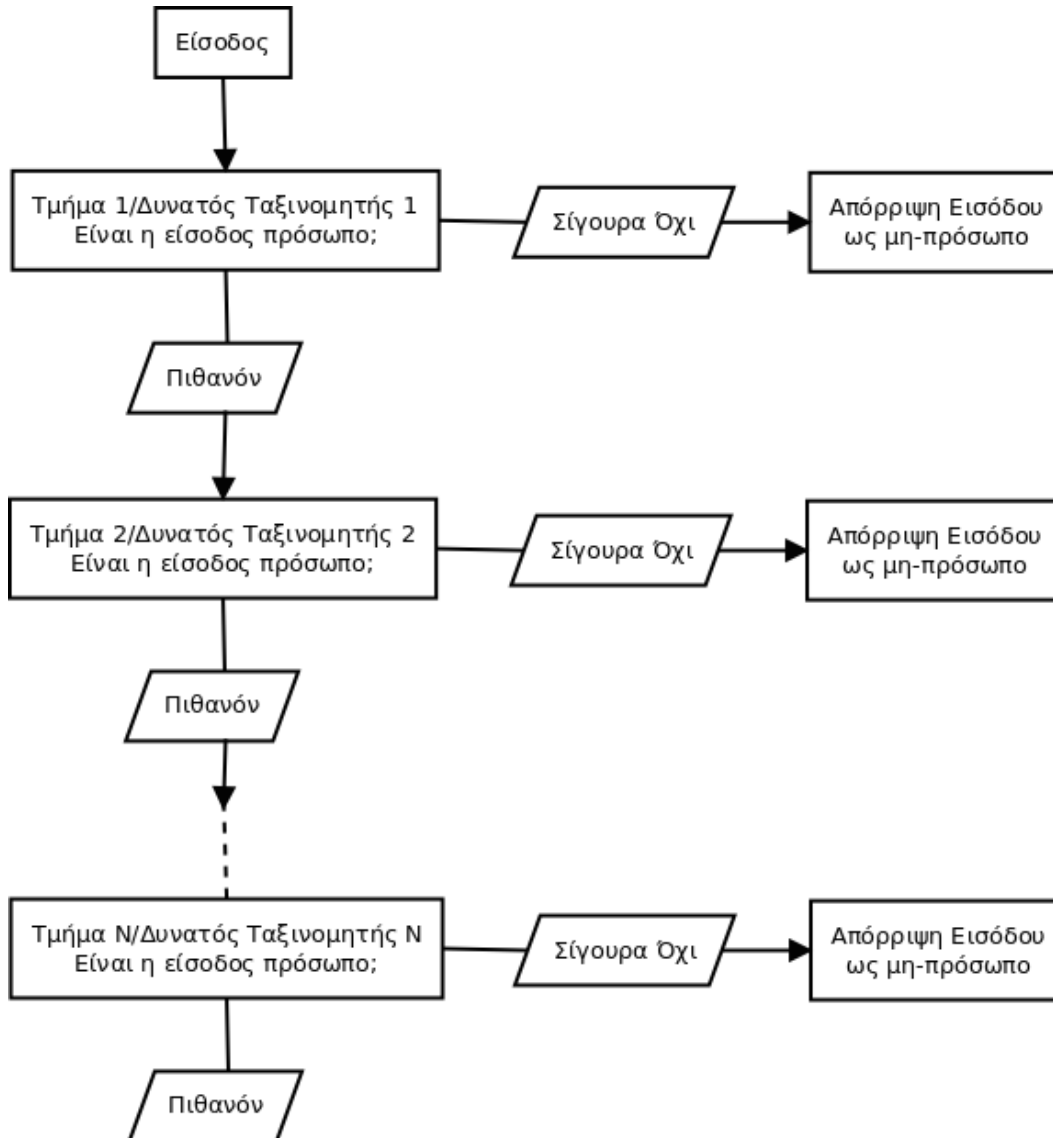
*Από ένα σύνολο αδύναμων ταξινομητών (σύνολο απλών δυαδικών τιμών) σχηματίζει ένα δυνατό ταξινομητή. Τα δείγματα εκπαίδευσης είναι θετικά και αρνητικά.*

- Η χρήση ενός δυνατού ταξινομητή είναι αναποτελεσματική, αφού η χρονική πολυπλοκότητα παραμένει η ίδια, ανεξάρτητα από τον αριθμό των προσώπων ή το μέγεθός τους.

- *Ταξινομητής Cascade*

*Ο αλγόριθμος απορρίπτει μη-πρόσωπα, αντί να βρίσκει πρόσωπα. Αποτελείται από σειριακά τμήματα δυνατών καταχωρητών. Όσα περισσότερα τμήματα καταφέρει να περάσει ένα παράθυρο, τόσο μεγαλύτερη η πιθανότητα να είναι τελικά πρόσωπο.*

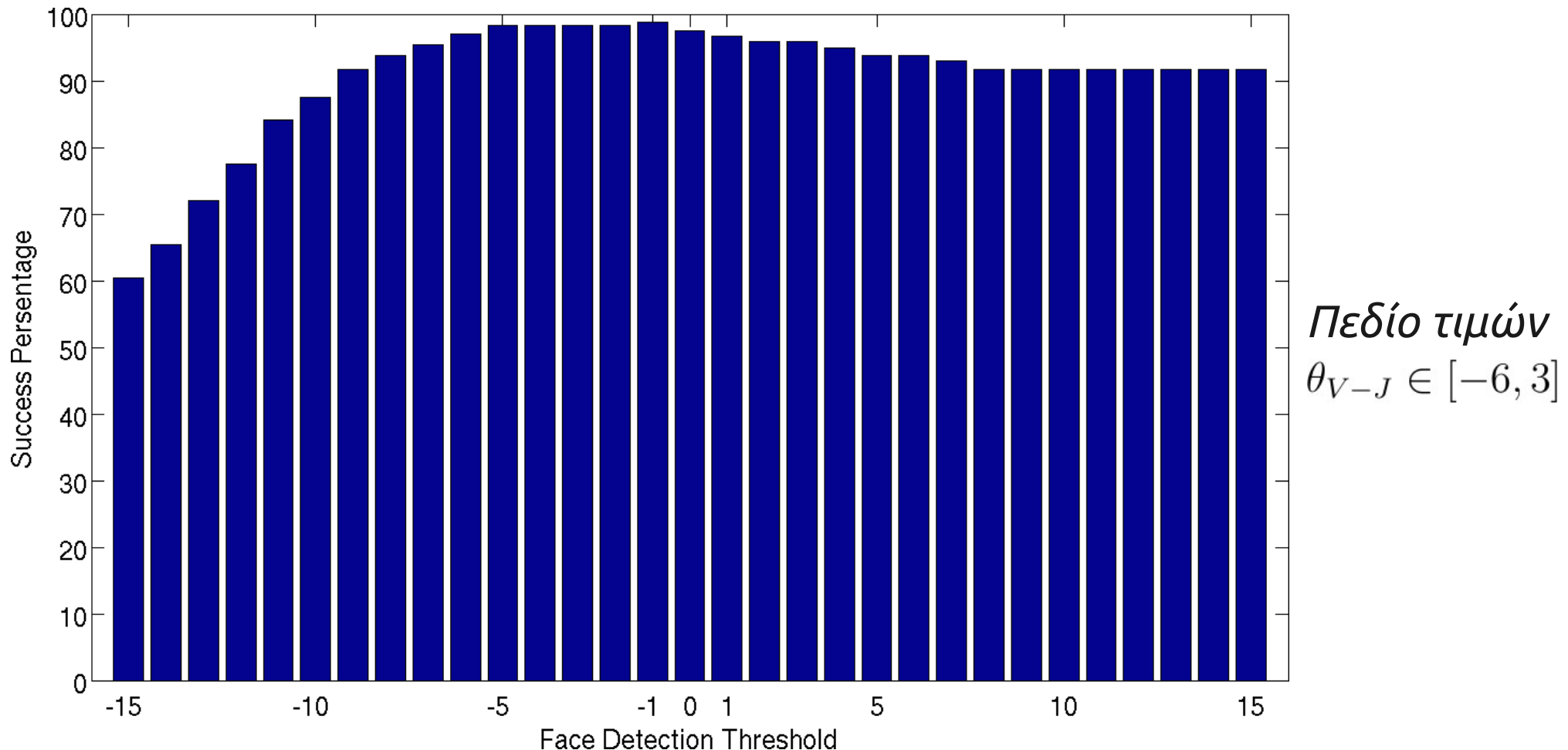
# Ανίχνευση Προσώπου



- Μείωση των false-positive εικόνων, γιατί αναμένεται να απορριφθούν σε κάποιο τμήμα
- Μεγαλύτερη υπολογιστική ισχύς σε περιοχές της εικόνας που υπάρχει μεγάλη πιθανότητα να έχουν ένα ή περισσότερα πρόσωπα

# Ανίχνευση Προσώπου

- Κατώφλι αλγορίθμου:  $\theta_{V-J}$
- Ποσοστά επιτυχίας στη βάση δεδομένων IMM (240 εικόνες).



# Ανίχνευση Προσώπου

Ταχύτητα εκτέλεσης:  $0.024 \text{ sec/frame} = 41.7 \text{ frames/sec}$  (1 πρόσωπο)  
 $0.110 \text{ sec/frame} = 9.1 \text{ frames/sec}$  (~30 πρόσωπα)



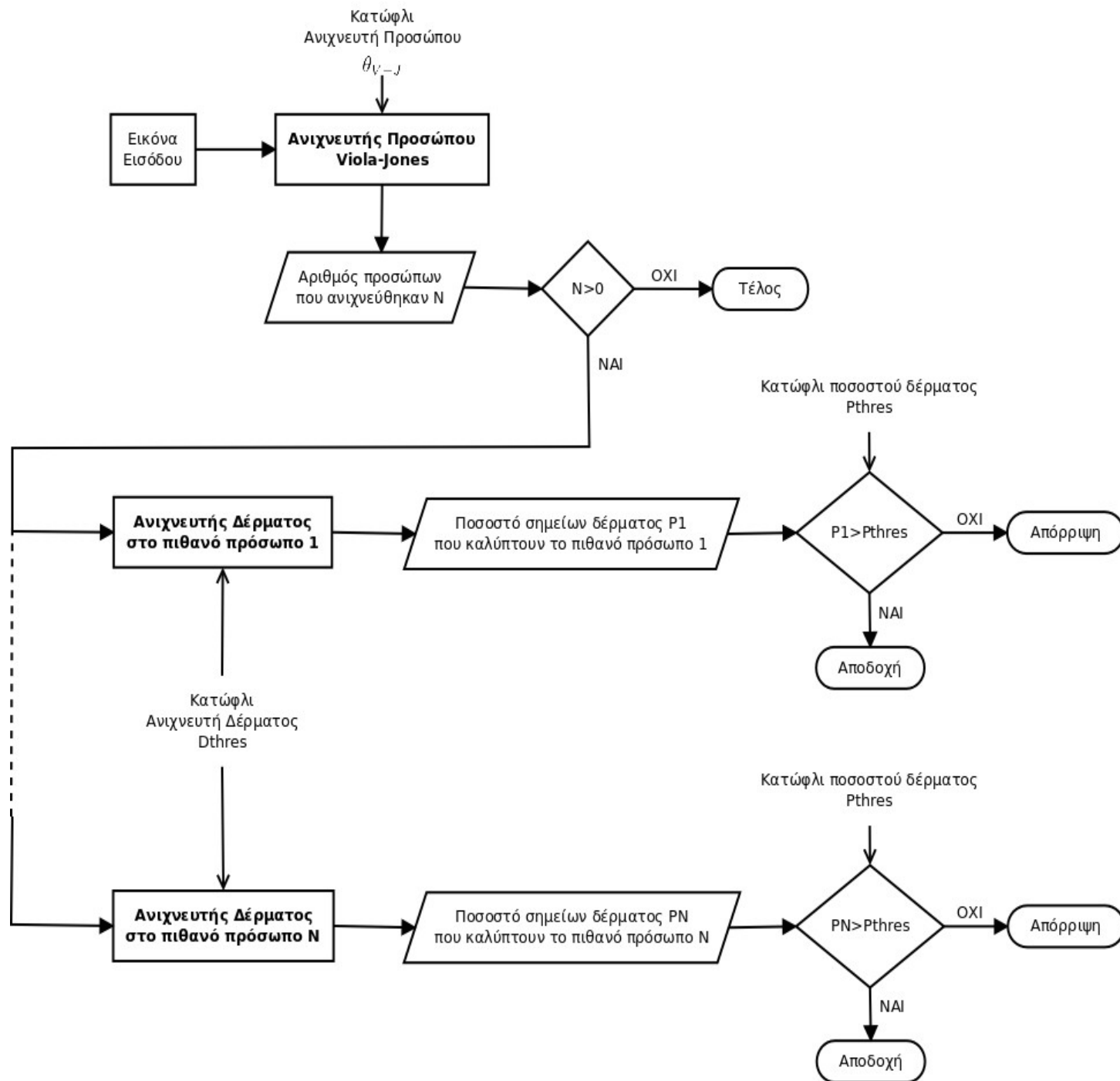
- Επιτυχία σε περίπτωση εμποδίων είτε στο στόμα είτε στο ένα μάτι.
- Αποτυχία σε περίπτωση μεγάλης περιστροφής του προσώπου.

# Συνδυασμός Ανίχνευσης Προσώπου και Ανίχνευσης Δέρματος

- Στόχος η αποφυγή false-positive ανιχνεύσεων.
- Εφαρμογή όταν το παρασκήνιο είναι πολύπλοκο.
- Εκτελείται πρώτα η ανίχνευση προσώπου και μετά οι περιοχές ελέγχονται για το ποσοστό δέρματος που περιέχουν.
- Τρεις **παράμετροι**:
  - (1) Κατώφλι ανιχνευτή προσώπου Viola-Jones
  - (2) Κατώφλι πιθανοτικού ανιχνευτή δέρματος
  - (3) Κατώφλι ποσοστού δέρματος
- Ταχύτητα εκτέλεσης:  $0.06 \text{ sec/frame} = 16.7 \text{ frames/sec}$

Μεγάλη τιμή στο κατώφλι ανιχνευτή δέρματος για να ανιχνευτούν τα true-positive πρόσωπα και τα false-positive θα απορριφθούν μέσω της παραμέτρου ποσοστού δέρματος.

## Διάγραμμα Ροής





# Οπτική Μοντελοποίηση Προσώπου

# Οπτική Μοντελοποίηση Προσώπου

Χρήση των αντικειμενοστραφών, παραμετροποιήσιμων και παραμορφώσιμων Active Appearance Models για την **εκπαίδευση** ενός μοντέλου σχήματος και ενός μοντέλου υφής προσώπου.

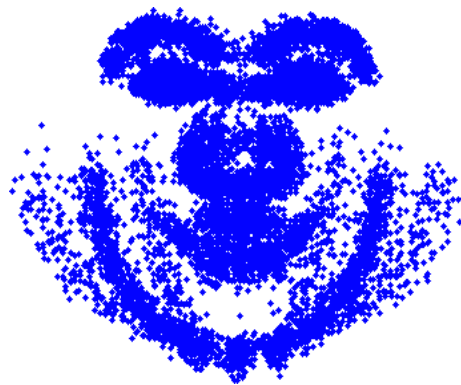
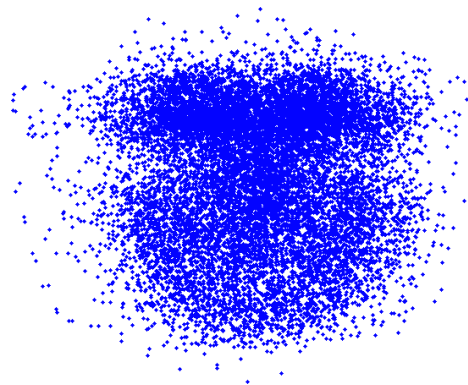
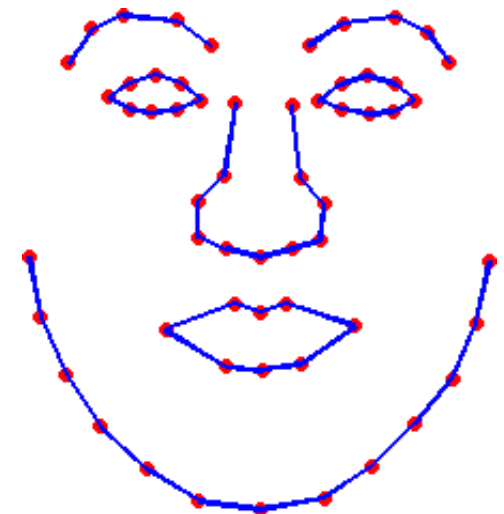
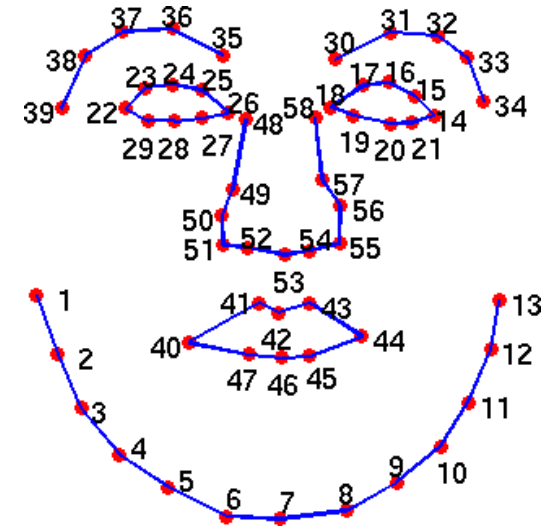
**Ταίριασμα** των μοντέλων σχήματος και υφής προσώπου Active Appearance Models σε νέες εικόνες.

# Βάσεις Δεδομένων Εικόνων

- *IMM Database, Technical University of Denmark*  
240 εικόνες, 40 άτομα με 6 εικόνες ανά άτομο,  
20% γυναίκες, 480x640, ειδικευμένη σε  
ανίχνευση και αναγνώριση προσώπου
- *Extended Cohn-Kanade Dataset (CK+), CMU*  
11734 εικόνες, 593 ακολουθίες εικόνων από 123  
άτομα, 69% γυναίκες, 640x490, ειδικευμένη σε  
αναγνώριση εκφράσεων προσώπου και  
συναισθημάτων

# Εκπαίδευση Μοντέλου Σχήματος Προσώπου AAM

- Το σχήμα του προσώπου ορίζεται σαν ένα διάνυσμα των  $L$  **landmark σημείων**:  
$$s = [x_1, y_1, x_2, y_2, \dots, x_{L-1}, y_{L-1}, x_L, y_L]^T$$
- Τα landmark σημεία είναι ορισμένα με το χέρι στις βάσεις δεδομένων.
- **Ευθυγράμμιση σχημάτων** προσώπου με Procrustes Analysis και υπολογισμός μέσου σχήματος προσώπου.



# Εκπαίδευση Μοντέλου Σχήματος Προσώπου AAM

- **Principal Component Analysis (PCA)** για μείωση των διαστάσεων του χώρου των διανυσμάτων σχήματος προσώπου της βάσης δεδομένων.
- **Αριθμός αξόνων μεταβλητότητας** διατηρώντας το 95% της συνολικής μεταβλητότητας είναι  $n=21$ .
- Σύνθεση **στιγμιότυπων** σχήματος προσώπου:

$$S_p = S_0 + \sum_{i=1}^n p_i S_i$$

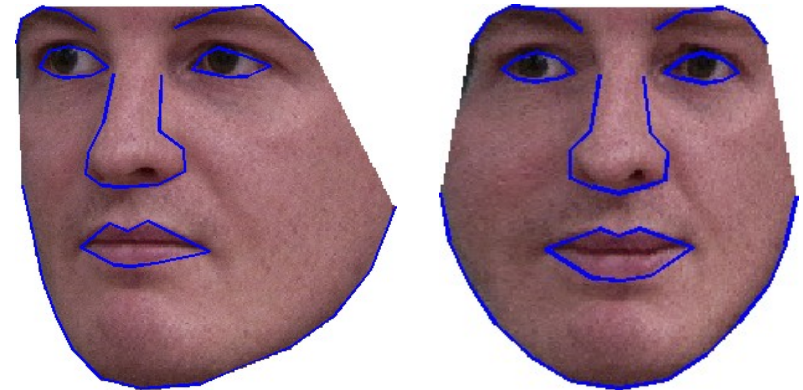
Ιδιοδιανύσματα	$S_i$
Πλήθος ιδιοδιανυσμάτων (modes of variation)	$n$
Πλήθος landmark σημείων	$L$
Παράμετροι σχήματος	$P_{n \times 1}$
Στιγμιότυπο σχήματος προσώπου	$S_p$

- **Μετασχηματισμός Ομοιότητας** τεσσάρων παραμέτρων (κλίμακα, περιστροφή και μετατόπιση  $x,y$ ) για την πόζα.

# Εκπαίδευση Μοντέλου Υφής Προσώπου AAM

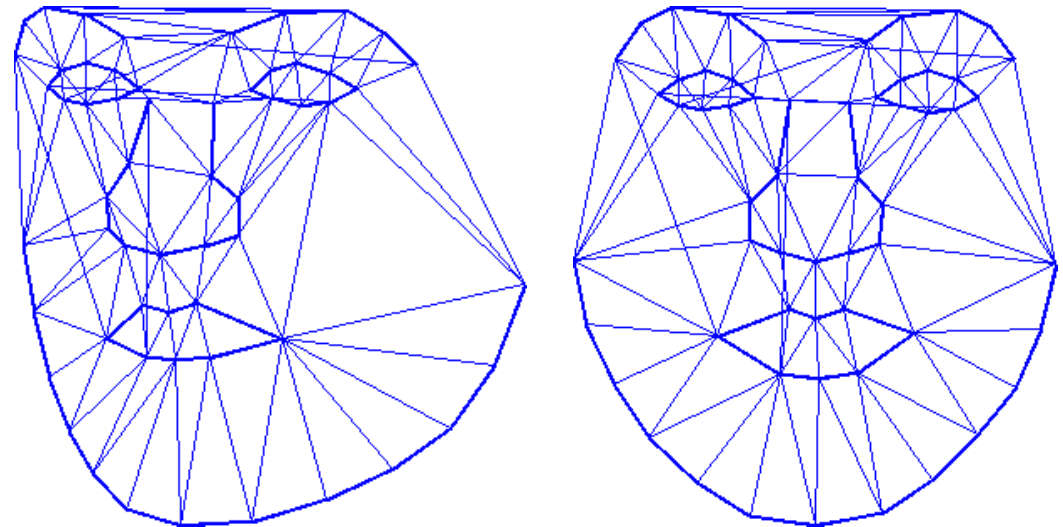
- Επιλέγεται το μέσο σχήμα προσώπου σαν **σχήμα αναφοράς**.

- **Χαρτογράφηση** όλων των διανυσμάτων υφής προσώπου στο σχήμα αναφοράς για να υπολογιστεί η μέση υφή προσώπου.



- **Delaunay τριγωνισμός**

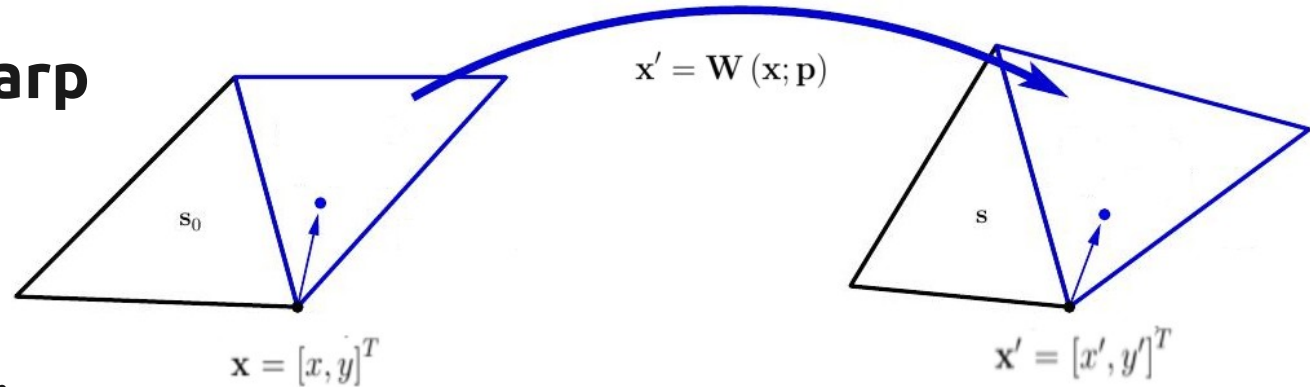
Κανένα από τα αρχικά σημεία να μην ανήκει εντός του περιγεγραμμένου κύκλου οποιουδήποτε τριγώνου του τριγωνισμού.



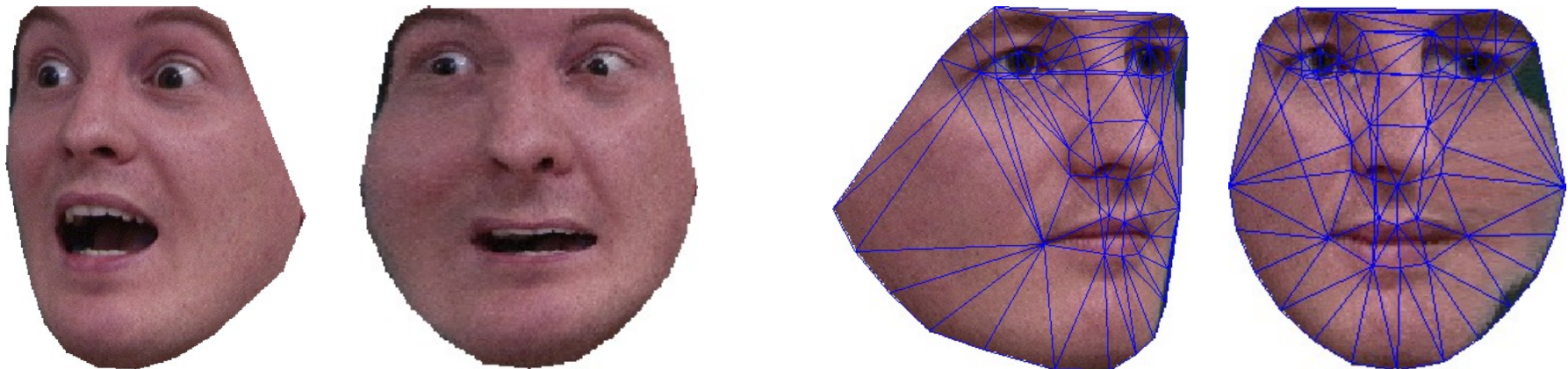
# Εκπαίδευση Μοντέλου Υφής Προσώπου AAM

- **Piecewise Affine Warp**

Συνάρτηση  
ομοπαράλληλης  
στρέβλωσης με  
αντιστοιχία σημείων.



- **Δειγματοληψία** εικόνας (αντιγραφή των τιμών των σημείων που ορίζονται από συνάρτηση warp) με **διγραμμική παρεμβολή** στα κενά.





# Εκπαίδευση Μοντέλου Υφής Προσώπου AAM

- **Principal Component Analysis (PCA)** για μείωση των διαστάσεων του χώρου των διανυσμάτων υφής προσώπου της βάσης δεδομένων.
- **Αριθμός αξόνων μεταβλητότητας** διατηρώντας το 96% της συνολικής μεταβλητότητας είναι  $m=93$ .
- Σύνθεση **στιγμιότυπων** υφής προσώπου:

$$A_{\lambda} = A_0 + \sum_{i=1}^m \lambda_i A_i$$

Ιδιοδιανύσματα	$A_i$
Πλήθος ιδιοδιανυσμάτων (modes of variation)	$m$
Πλήθος pixels μέσου σχήματος	$N$
Παράμετροι υφής	$\lambda_{m \times 1}$
Στιγμιότυπο υφής προσώπου	$A_{\lambda}$

- **Ομοπαράλληλικός Μετασχηματισμός Υφής** δυο παραμέτρων (κέρδος και αντιστάθμιση) για τις συνθήκες φωτισμού της υφής.



# Ταίριασμα Μοντέλου AAM

- Διάνυσμα παραμέτρων του μοντέλου AAM του σχήματος και της υφής προσώπου:

$$\mathbf{q} = \begin{bmatrix} \tilde{\mathbf{p}} \\ \tilde{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{t}_{4 \times 1} \\ \mathbf{p}_{n \times 1} \\ \mathbf{u}_{2 \times 1} \\ \boldsymbol{\lambda}_{m \times 1} \end{bmatrix}_{(n+m+6) \times 1}$$

Να βρεθεί το διάνυσμα παραμέτρων  $\mathbf{q}$  που ελαχιστοποιεί τη διαφορά μεταξύ της χαρτογραφημένης εικόνας υφής και της υφής του μοντέλου AAM.

Αλγόριθμοι Matthews & Baker 2004:

- Project-Out Inverse Compositional (POIC)*, ταχύς, μη ακριβής
- Simultaneous Inverse Compositional (SIC)*, αργός, ακριβής

# Ταίριασμα Μοντέλου AAM

**Αλγοριθμικές βελτιώσεις Παπανδρέου & Μαραγκός 2008:**

- Προσαρμογή πίνακα ταιριάσματος
- Ανανέωση της μέσης υφής προσώπου
- Ταίριασμα με περιορισμούς πρότερης (prior) πληροφορίας

**Επαναληπτικοί αλγόριθμοι με ανανέωση παραμέτρων.**

- Παράμετροι *υφής* προσώπου:  $\lambda \leftarrow \lambda + d\lambda$
- Παράμετροι *σχήματος* προσώπου:
  - (1) Ευθεία-προσθετική (FA) μέθοδος:  $\mathbf{p} \leftarrow \mathbf{p} + d\mathbf{p}$
  - (2) Αντίστροφη-συνθετική (IC) μέθοδος:  $\mathbf{W}(\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{p}) \circ \mathbf{W}^{-1}(d\mathbf{p})$

# Ταίριασμα Μοντέλου AAM: Προσαρμογή Πίνακα Ταιριάσματος

- Το κριτήριο ταιριάσματος του μοντέλου είναι η **εικόνα λάθους**  $E(\mathbf{x}; \mathbf{q}, d\mathbf{q}) = I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})) - A_{\lambda+d\lambda}(\mathbf{W}(\mathbf{x}; d\tilde{\mathbf{p}}))$



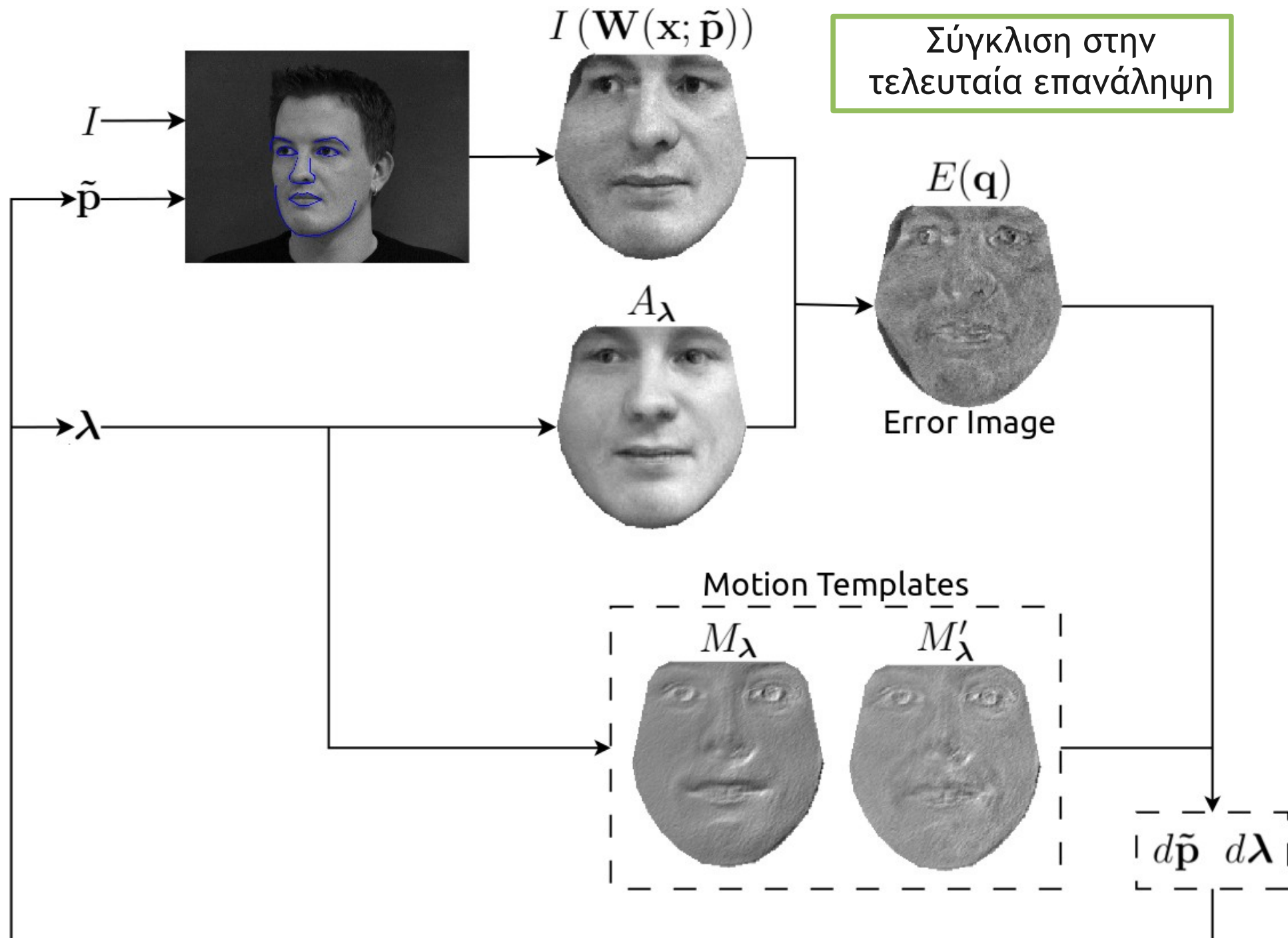
Εικόνα εισόδου  
στην περιοχή του  
μοντέλου σχήματος



Στιγμιότυπο  
μοντέλου υφής

- Πρόβλημα **ελαχίστων τετραγώνων**:

Εύρεση διανύσματος ανανέωσης παραμέτρων  $d\mathbf{q} = \begin{bmatrix} d\tilde{\mathbf{p}} \\ d\tilde{\lambda} \end{bmatrix}$   
που ελαχιστοποιεί τη νόρμα  $\|E(\mathbf{q}, d\mathbf{q})\|_2$



# Ταίριασμα Μοντέλου AAM: Προσαρμογή Πίνακα Ταιριάσματος

- Τα **πρότυπα κίνησης (motion templates)** περιγράφουν την αλλαγή στην υφή από την ανανέωση μιας παραμέτρου σχήματος.
- **Γραμμικός συνδυασμός** με βάρη τις παραμέτρους υφής. Αποτελούνται από  $n+4$  εικόνες (μία για κάθε παράμετρο σχήματος)

$$M_{\lambda} = M_0 + \sum_{i=1}^r \lambda_i M_i \quad \text{Μέγεθος: } N \times (n + 4)$$

- Τα **project-out motion templates** είναι προβολές στο συμπληρωματικό του υποχώρου που καλύπτεται από τα ιδιοδιανύσματα υφής:

$$M'_{\lambda} = M'_0 + \sum_{i=1}^r \lambda_i M'_i \quad M'_i = P M_i$$

- Υπολογισμός **εκτός του βρόχου επανάληψης (pre-computed)** ως:

$$M_i(\mathbf{x}) = \left. \frac{\partial A_i(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \right|_{\tilde{\mathbf{p}}=0}$$

# Ταίριασμα Μοντέλου AAM: Προσαρμογή Πίνακα Ταιριάσματος

- Η παράμετρος  $r$  είναι η **τάξη του αλγορίθμου**: καθορίζει το πλήθος των  $m$  ιδιοδιανυσμάτων υφής που χρησιμοποιούνται.



- Παράμετρος **καθολικής κλίμακας** του μοντέλου AAM:  $S_{global}$

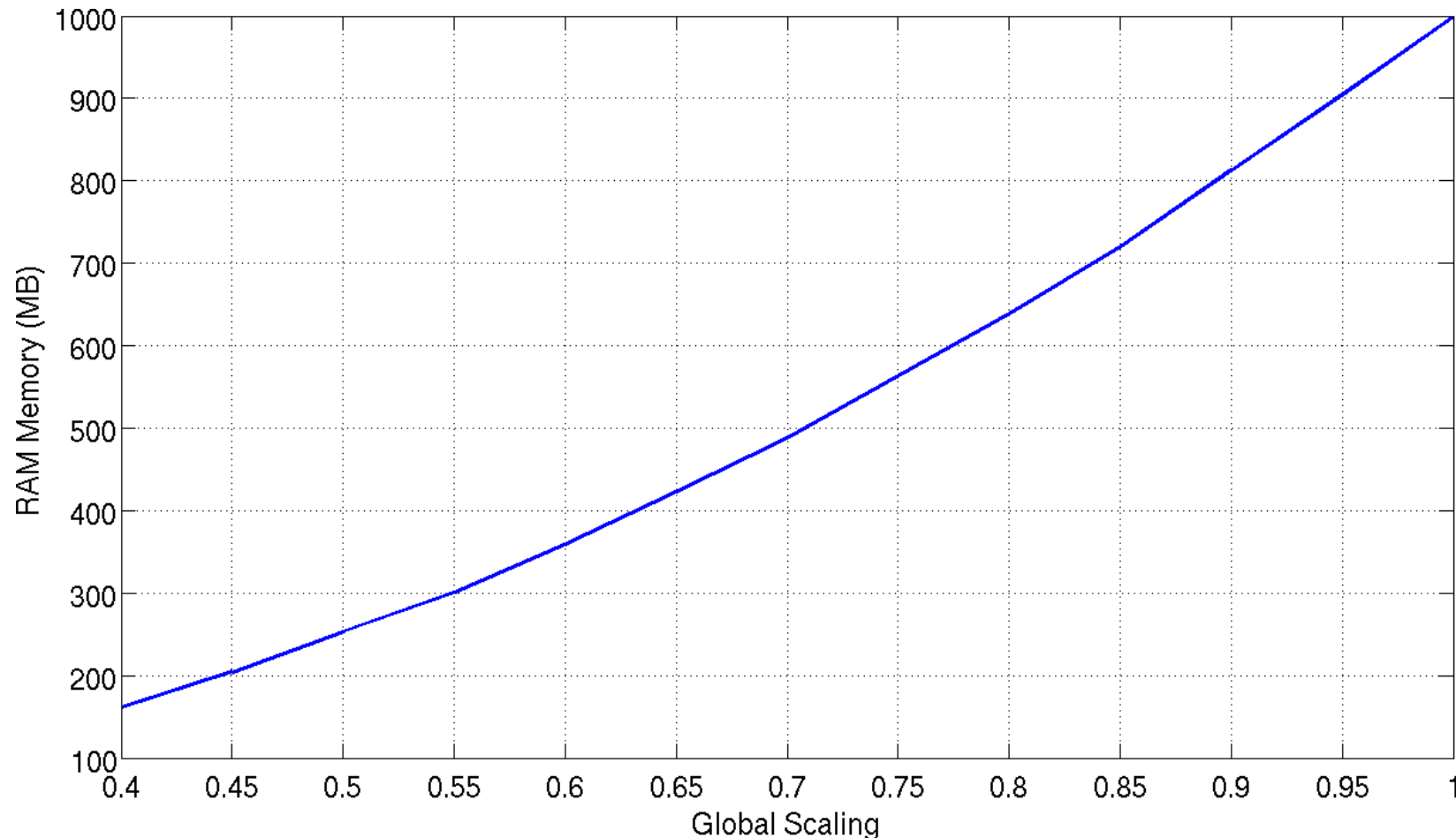
Καθορίζει την ανάλυση του μοντέλου, δηλαδή τον *αριθμό των pixels*  $N$  εσωτερικά του μέσου σχήματος προσώπου.

Δεν γίνεται κλιμάκωση στο αρχικό μοντέλο AAM. Εκπαιδεύεται το μοντέλο με *κλιμάκωση στις εικόνες της βάσης δεδομένων*.

Σημαντική παράμετρος για την *ταχύτητα εκτέλεσης* του αλγορίθμου ταιριάσματος και για την *απαιτούμενη μνήμη RAM*.

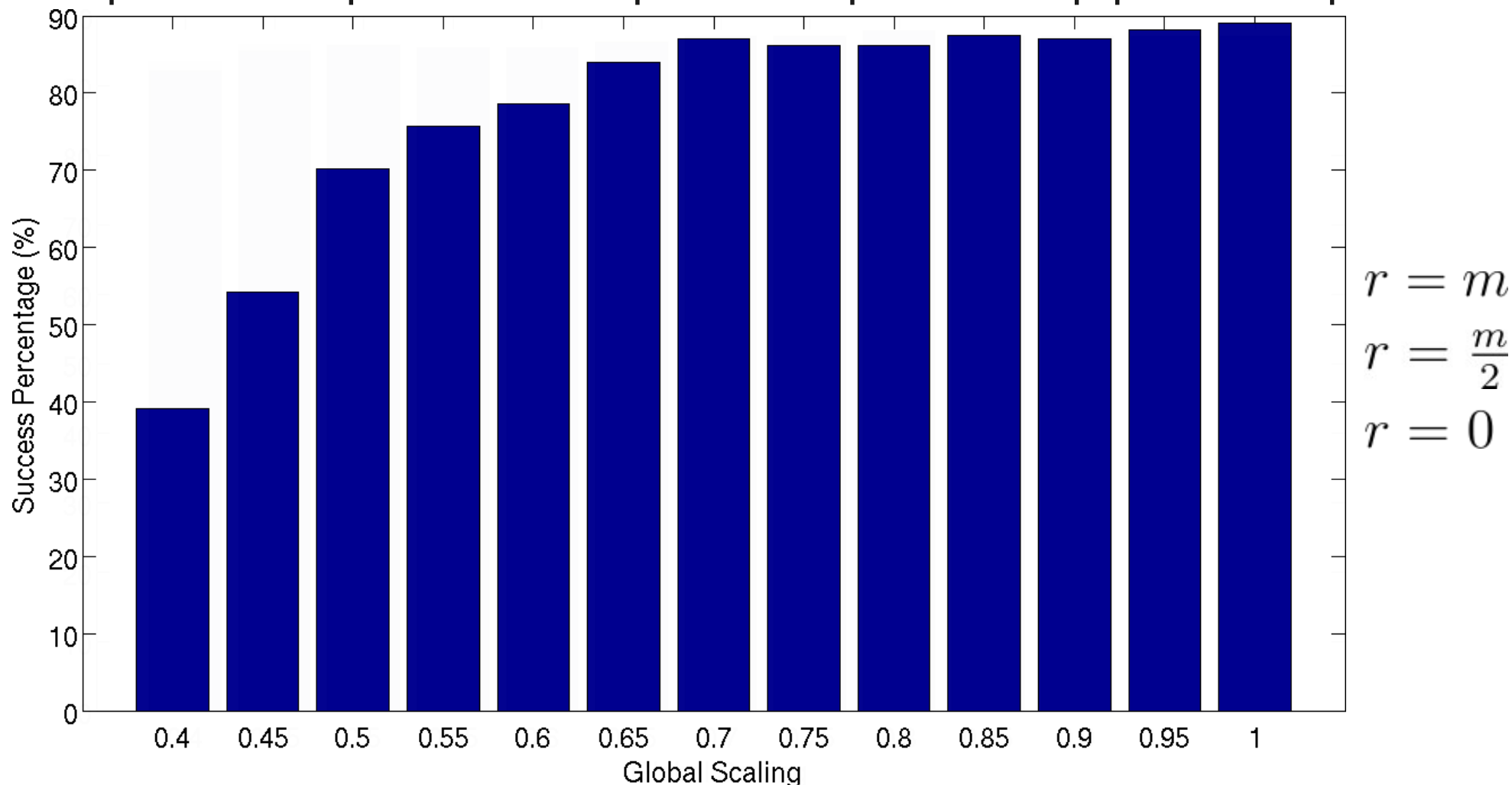
# Ταίριασμα Μοντέλου AAM: Προσαρμογή Πίνακα Ταιριάσματος

**Μνήμη RAM** που απαιτείται για τα μοντέλα AAM και τους πίνακες των motion templates ως προς την παράμετρο καθολικής κλίμακας



# Ταίριασμα Μοντέλου AAM: Ποσοστά Επιτυχίας

Ποσοστά επιτυχίας για διάφορες τιμές της καθολικής κλίμακας στη βάση δεδομένων IMM με εκπαίδευση σε 30 άτομα και δοκιμή σε 10 άτομα.

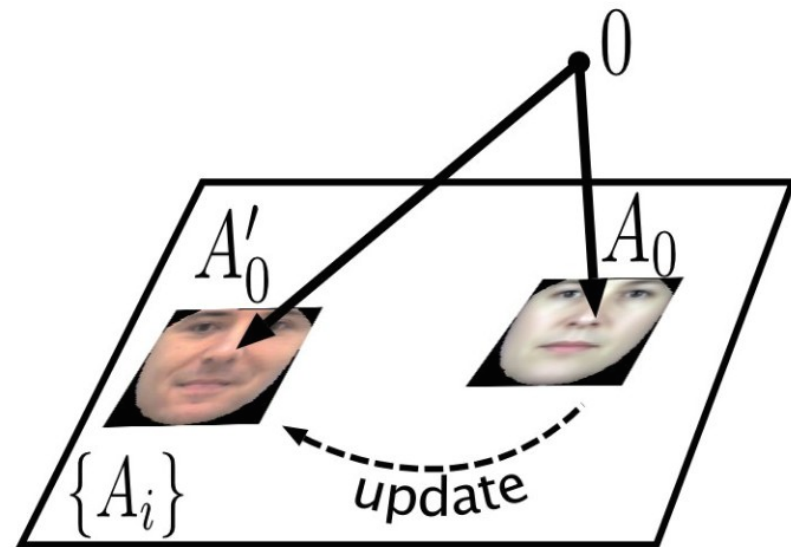




# Ταίριασμα Μοντέλου AAM: Ανανέωση Μέσης Υφής Προσώπου

- Μέθοδος βελτίωσης των ποσοστών για **χαμηλή τάξη**.
- Ανανεώνει τη μέση υφή του μοντέλου AAM με τα ιδιοδιανύσματα που δεν χρησιμοποιούνται:

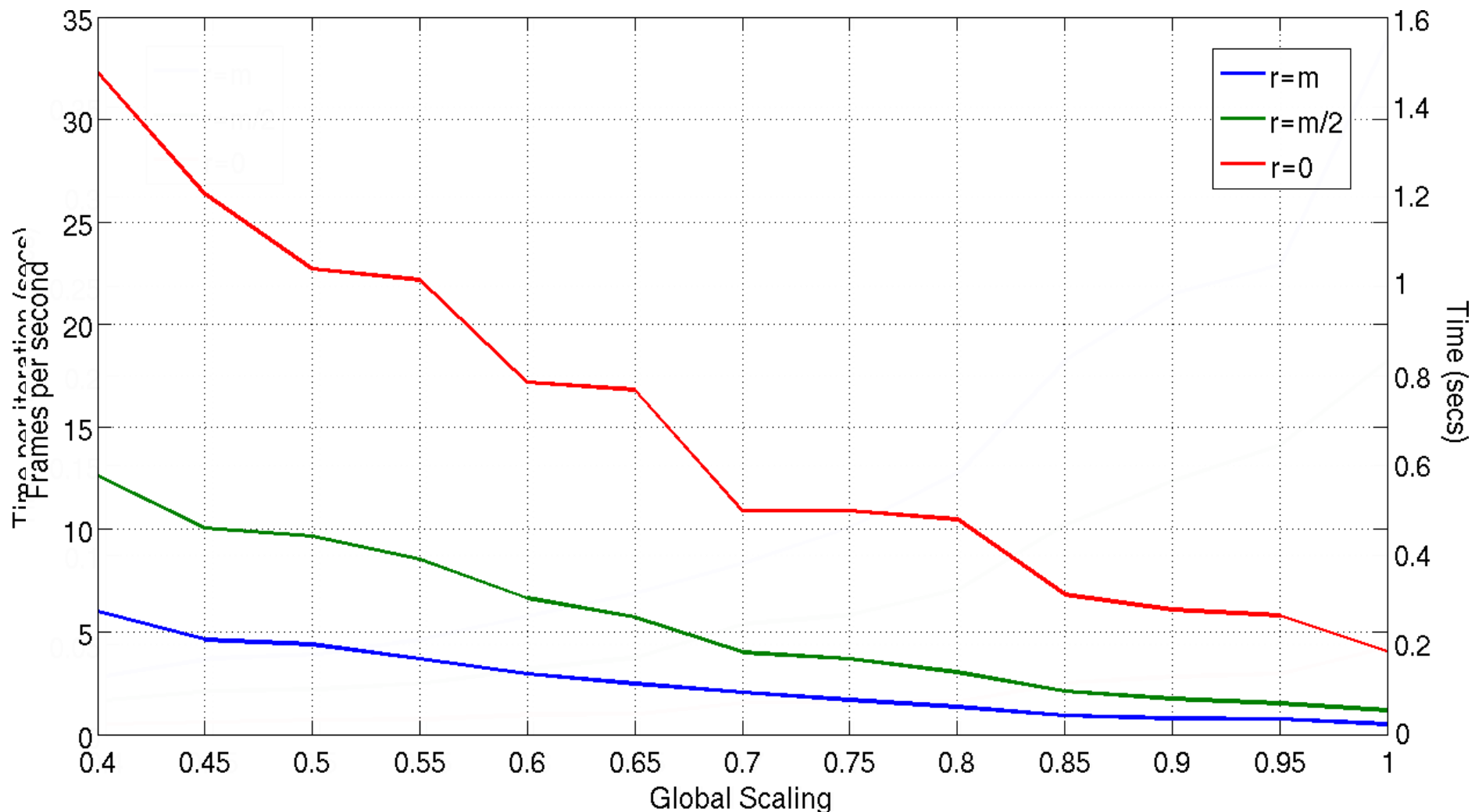
$$A'_0 = A_0 + \sum_{i=r+1}^m \lambda_i A_i$$



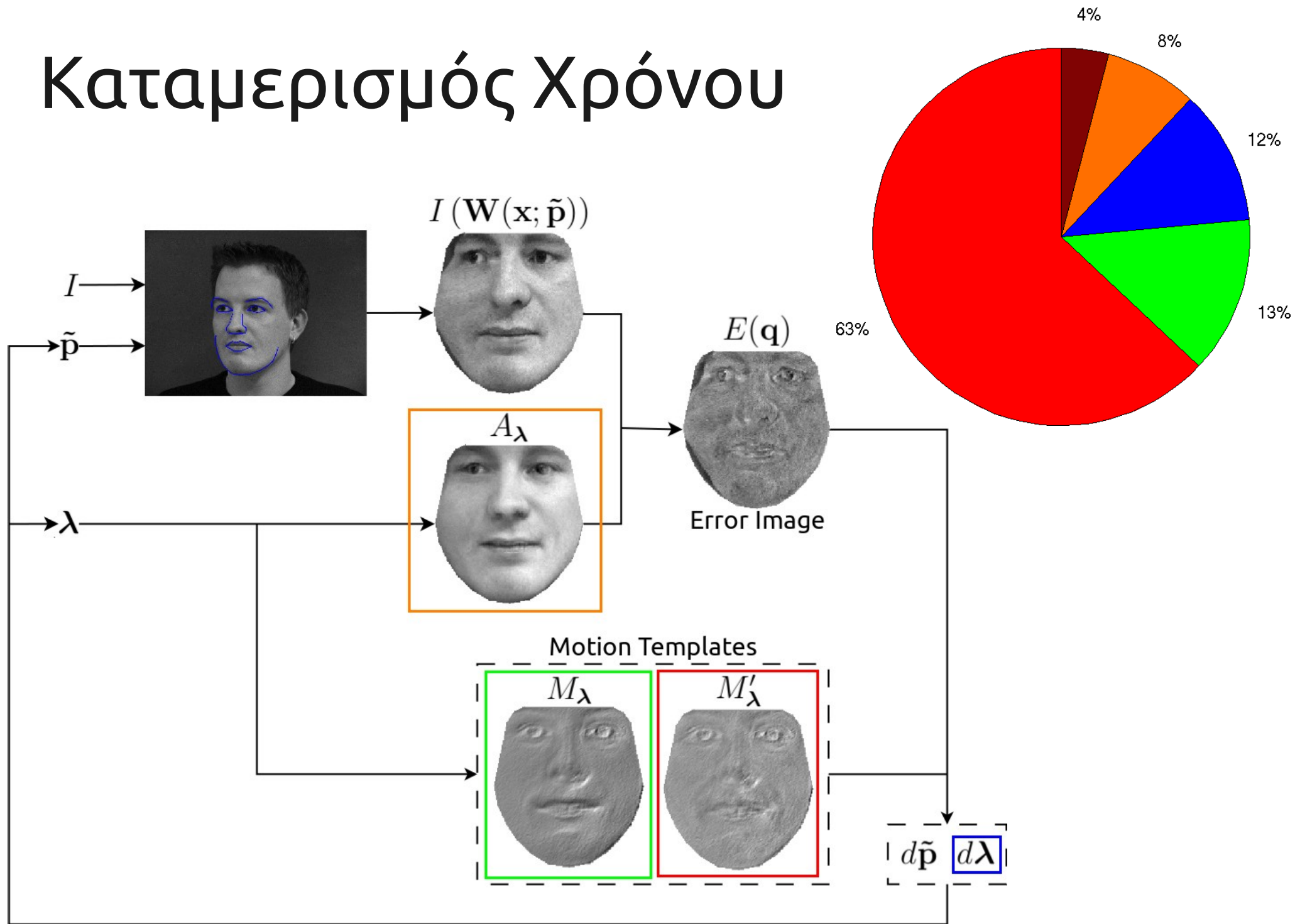
- **Περιοδική ανανέωση** μέσου σχήματος προσώπου για ταίριασμα του μοντέλου σε βίντεο (κάθε ~20 frames).
- Ίδια ποσοστά επιτυχίας με *αλγόριθμο πλήρους τάξης*.

# Ταίριασμα Μοντέλου AAM: Χρόνος Εκτέλεσης

Απαιτούνται ~0.16sec/frame επιπλέον για τη φόρτωση και την πολυπλοκότητα αλγορίθμου τάξης:  $O(r^2 n^2 + (n+r)N + n^3)$  απεικόνιση του frame με ανάλυση 640x480.

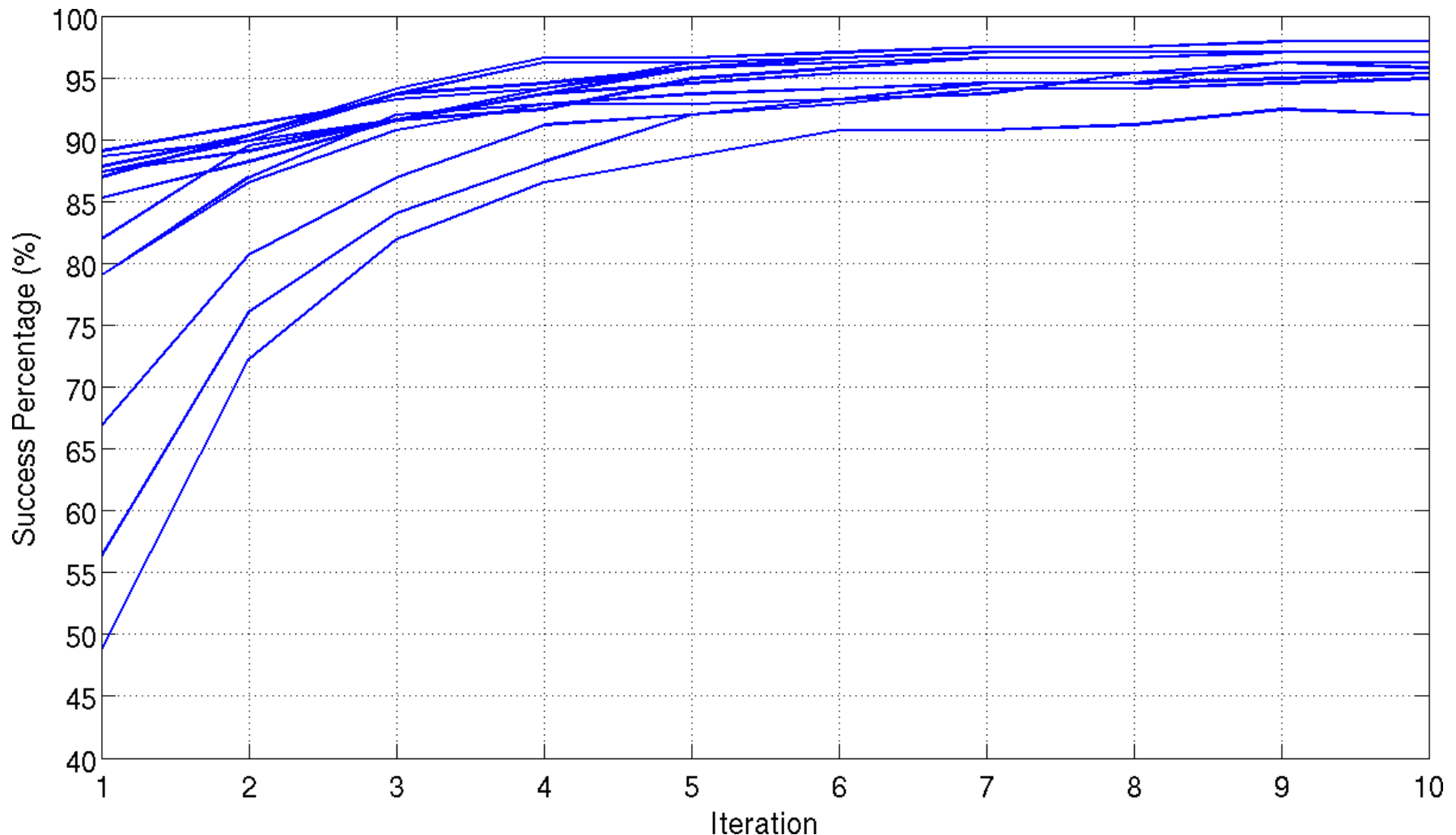


# Καταμερισμός Χρόνου



# Ταίριασμα Μοντέλου AAM: Παράμετροι Επιλογής

Ο αριθμός επαναλήψεων προκύπτει από δοκιμές:



# Ταίριασμα Μοντέλου AAM: Παράμετροι Επιλογής

- **Παράμετροι μοντέλου:**

- (1) Τάξη αλγορίθμου
- (2) Καθολική κλίμακα μοντέλου

- **Παράμετροι αρχικοποίησης:**

- (3) Κατώφλι ανίχνευσης προσώπου
- (4) Κλίμακα αρχικοποίησης: καθορίζει τις αρχικές τιμές των παραμέτρων του μετασχηματισμού ομοιότητας

- **Παράμετροι συνθήκης τερματισμού:**

- (5) Μέγιστος αριθμός επαναλήψεων
- (6) Μέγιστη μετακίνηση των σημείων του μοντέλου σχήματος

OR