



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Ομάδα Όρασης Υπολογιστών, Επικοινωνίας Λόγου & Επεξεργασίας Σήματος

<http://cvsp.cs.ntua.gr>

Μη-γραμμική Διάχυση στην Όραση Υπολογιστών και Στατιστικά Μοντέλα Σχήματος με Εφαρμογές στην Ανάλυση Εικόνων Αρθρωτών Φωνητικού και Νοηματικού Λόγου

Παρουσίαση Διδακτορικής Διατριβής

Αναστάσιος Ρούσσος

Ηλεκτρολόγος Μηχανικός & Μηχανικός Η/Υ,

Υποψήφιος Διδάκτορας ΕΜΠ

6 Οκτωβρίου 2010

Περιεχόμενα

- **ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών**
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης
- Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας
- Σύνοψη Προόδου και Μελλοντική Έρευνα

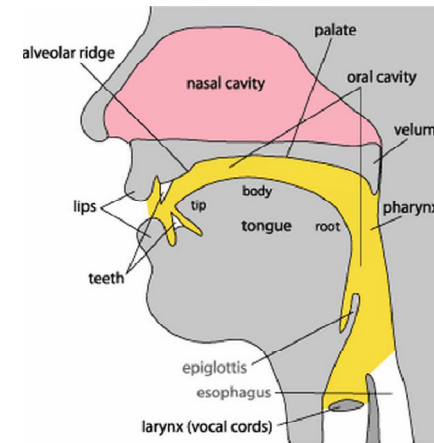
ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών

- Όραση Υπολογιστών (ΟΥ)
 - Εξαγωγή συμβολικής πληροφορίας από μία εικόνα / σύνολο εικόνων
- Μερικές Διαφορικές Εξισώσεις (ΜΔΕ) στην Όραση Υπολογιστών και στην Επεξεργασία Εικόνας
 - Σχετικά πρόσφατη τάση (τέλη '80)
 - Είναι πλέον δημοφιλής λόγω διαφόρων πλεονεκτημάτων
 - Ανάπτυξη των Χώρων Κλίμακας
 - Μη-γραμμική διάχυση σε προβλήματα της ΟΥ
 - Ενεργά Περιγράμματα για Κατάτμηση εικόνων
- Στατιστικά Μοντέλα Σχήματος
 - Εφαρμόζουν πρότερη πληροφορία σχήματος
 - Είναι αναγεννητικά και παραμορφώσιμα
 - Εντοπισμός & κατηγοριοποίηση αντικειμένων: ταίριασμα μοντέλου
 - Ενεργά Μοντέλα Σχήματος, Ενεργά Μοντέλα Εμφάνισης

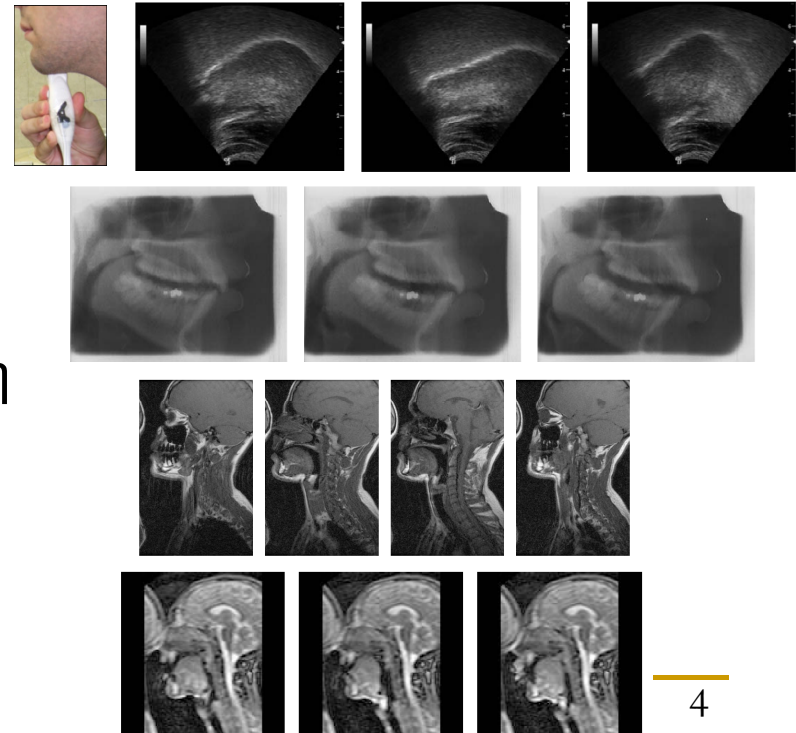
Έρευνα του Συστήματος Παραγωγής Φωνής

- Υποπροβλήματα
 - Αρθρωτή Σύνθεση Φωνής
 - Οπτικοακουσική Αντιστροφή Φωνής

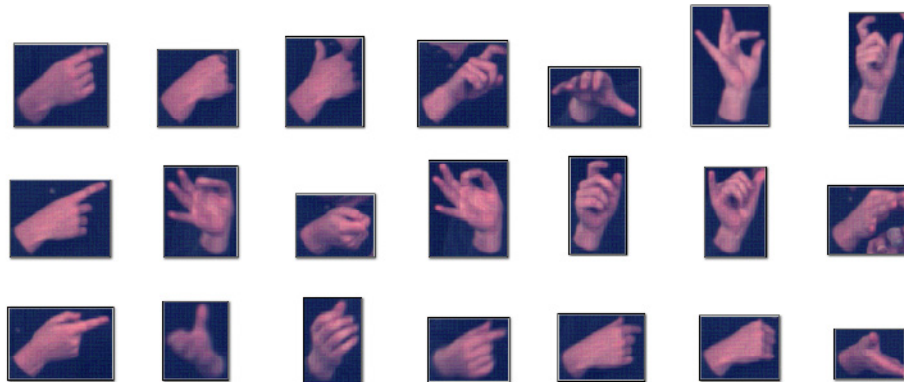
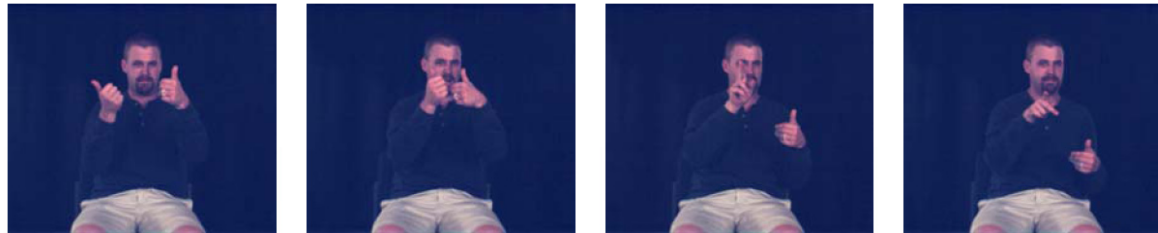
- Οπτικά δεδομένα αρθρωτών κατά την ομιλία
 - Τεχνικές απεικόνισης
 - Βελτίωση ποιότητας με ψηφιακή επεξεργασία
 - Ανάλυση των εικόνων για την εξαγωγή της γεωμετρικής πληροφορίας



Αρθρωτές Λόγου & Φωνητική Οδός



Αυτόματη Αναγνώριση Νοηματικής Γλώσσας

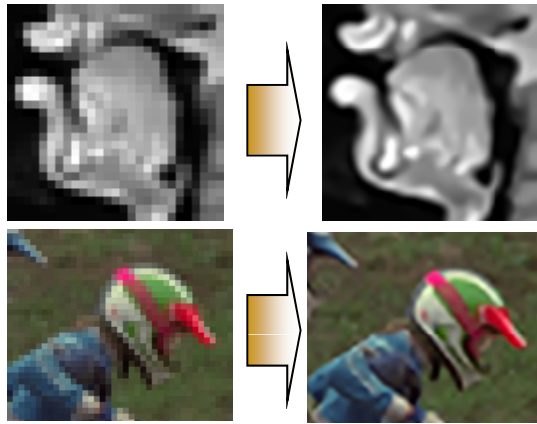


- Υποπροβλήματα
 - Εντοπισμός & παρακολούθηση των χεριών + κεφαλιού του νοηματιστή
 - Εξαγωγή χαρακτηριστικών που να περιγράφουν αξιόπιστα τις χειρομορφές
- Δυσκολίες
 - γρήγορη κίνηση των χεριών
 - αλληλεπικαλύψεις
 - υψηλή μεταβλητότητα στην πόζα και το σχήμα του χεριού

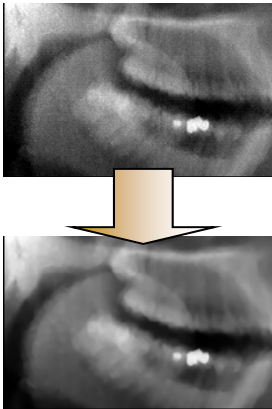
Ερευνητικές Συνεισφορές

Μέθοδοι μη-γραμμικής διάχυσης για την βελτίωση της ποιότητας εικόνων

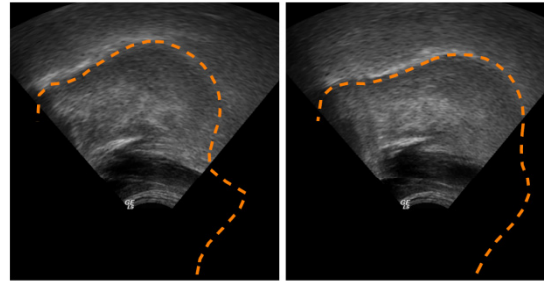
Μη-γραμμική διάχυση για την παρεμβολή διανυσματικών εικόνων



Γενικευμένο μεταβολικό πλαίσιο για ομαλοποιημένη μη-γραμμική διάχυση

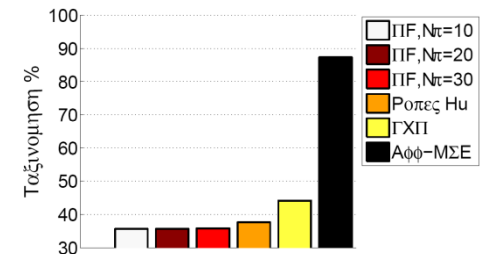
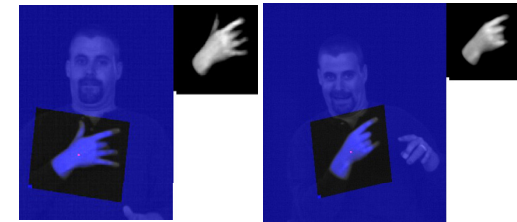


Παρακολούθηση της γλώσσας με Ενεργά Μοντέλα Εμφάνισης



Στατιστικά Μοντέλα Σχήματος

Αφρική Μοντελοποίηση Σχήματος-Εμφάνισης



Έρευνα του Συστήματος Παραγωγής Φωνής

Αναγνώριση Νοηματικής Γλώσσας

Περιεχόμενα

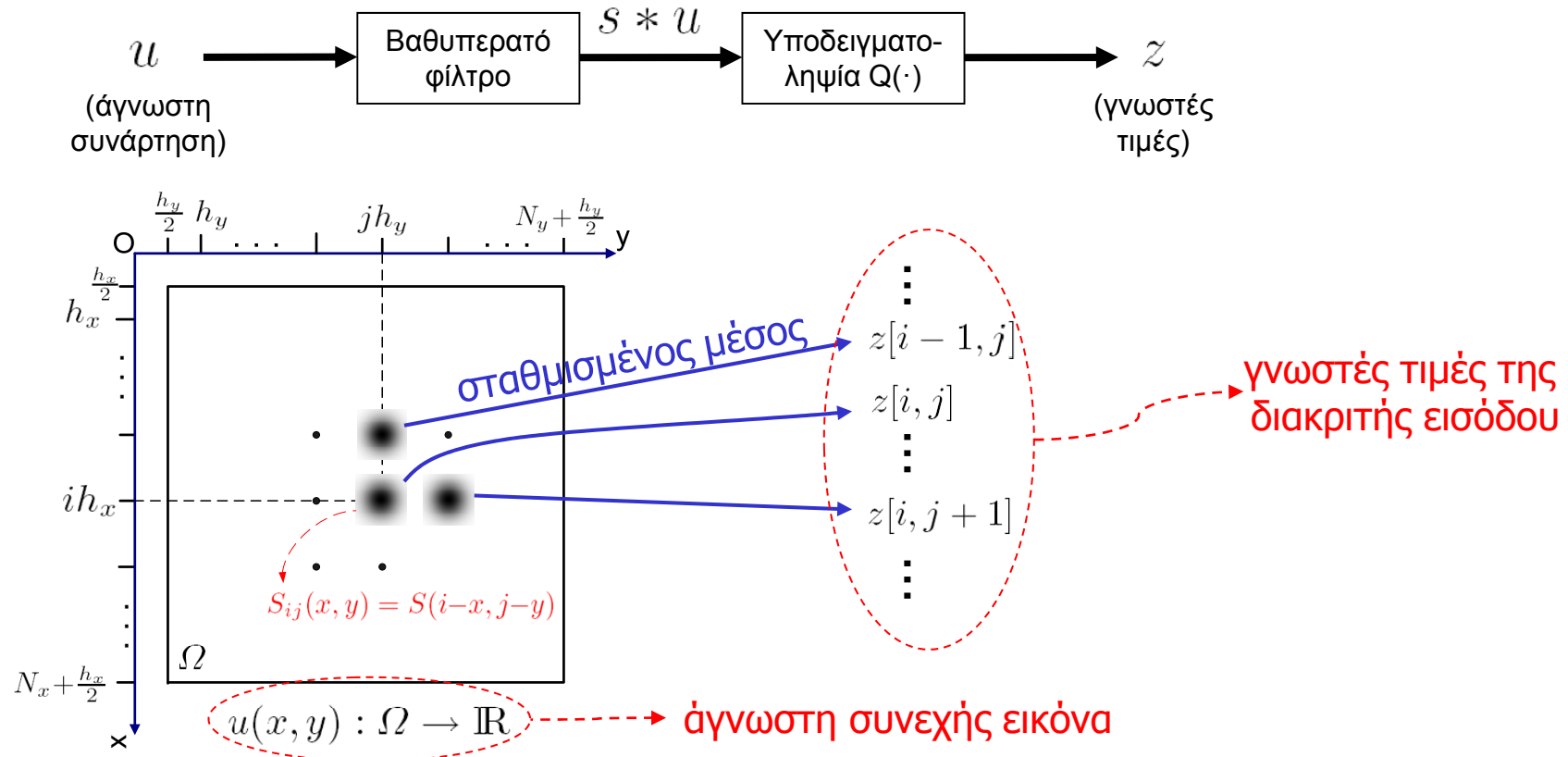
- ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- **Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης**
- Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας
- Σύνοψη Προόδου και Μελλοντική Έρευνα

Παρεμβολή Εικόνων

- Μπορεί να οριστεί ως η διαδικασία η οποία:
 - λαμβάνει σαν είσοδο μία **διακριτή εικόνα** και
 - ανακατασκευάζει μία **συνεχή εικόνα** ή μία διακριτή **υψηλότερης ευκρίνειας**
- **Θεμελιώδες πρόβλημα** της Επεξεργασίας Εικόνων με **ποικίλες εφαρμογές**:
 - επεξεργασία βιοϊατρικών, εναέριων και δορυφορικών εικόνων, αναγνώριση κειμένου, βίντεο υψηλής ευκρίνειας, υψηλής ποιότητας απεικόνιση και εκτύπωση,...
- Βήμα προ-επεξεργασίας σε διάφορα προβλήματα ΟΥ
 - κατάτμηση, εξαγωγή χαρακτηριστικών, αναγνώριση αντικειμένων, ανάλυση κίνησης
- Κατηγορίες Μεθόδων
 - Κλασικές **Γραμμικές** μέθοδοι
 - Προσαρμοστικές **Μη-γραμμικές** μέθοδοι

Συνθήκη Αντιστρεψιμότητας για την Παρεμβολή

- Σκοπιά παρόμοια με [Malgouyres, Guichard, SIAM J. Num. Anal. '01]
- Εύρεση μιας λύσης για το **αντίστροφο πρόβλημα**:



$$(S * u)(i_1 h_x, i_2 h_y) = z[i_1, i_2], \quad \forall (i_1, i_2) \in \{1, \dots, N_x\} \times \{1, \dots, N_y\}$$

Προτεινόμενη Μέθοδος Παρεμβολής

[Roussos,Maragos SSVM 07], [Roussos,Maragos IJCV 09]

- Ροή Μη-γραμμικής Διάχυσης, η οποία:
 - εξελίσσεται στον υποχώρο $\mathcal{U}_{z,S}$ των συναρτήσεων που ικανοποιούν την *Συνθήκη Αντιστρεψιμότητας*
 - εφαρμόζει *προσαρμοστική εξομάλυνση*, οδηγώντας σε στοιχεία του $\mathcal{U}_{z,S}$ με «καλύτερη» οπτική ποιότητα

ΜΔΕ Προτεινόμενης Μεθόδου

$$\frac{\partial u_m(\mathbf{x}, t)}{\partial t} = P_{\mathcal{U}_{0,S}} \left\{ \operatorname{div} \left(T(J_\rho(\nabla \mathbf{u}_\sigma)) \nabla u_m \right) \right\}, \quad m=1, \dots, M$$

τεχνητός χρόνος τελεστής προβολής 2x2 τανυστής διάχυσης

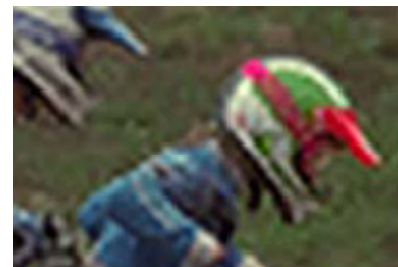
$\mathbf{u}(\mathbf{x}, 0)$ = μηδενικό γέμισμα υψηλών συχνοτήτων ($\in \mathcal{U}_{z,S}$)



είσοδος z



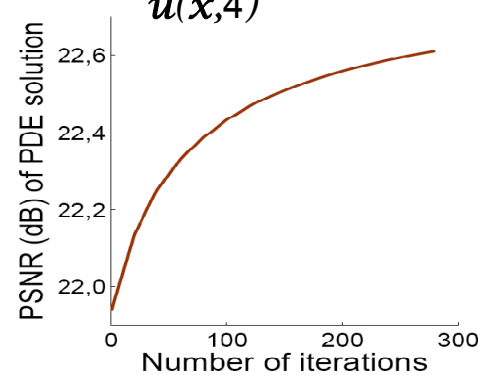
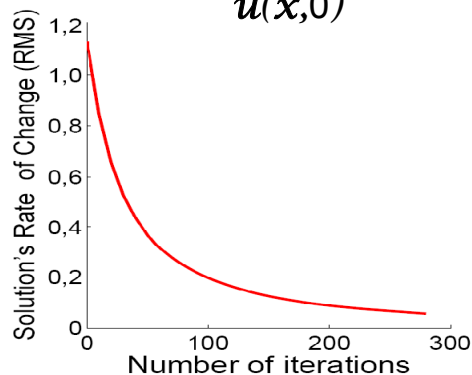
$u(x,0)$



$u(x,4)$



$u(x,56)$



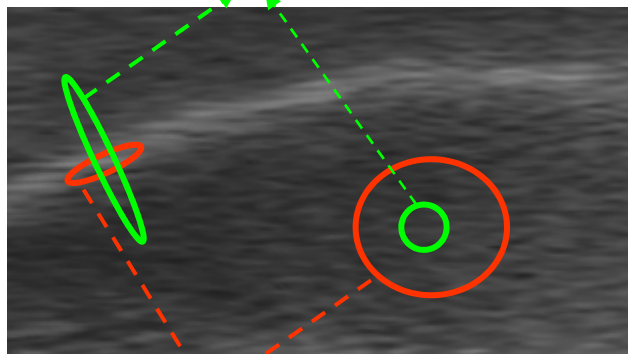
ΜΔΕ Προτεινόμενης Μεθόδου

$$\frac{\partial u_m(\mathbf{x}, t)}{\partial t} = P_{\mathcal{U}_{0,S}} \left\{ \operatorname{div} \left(\underbrace{T(J_\rho(\nabla \mathbf{u}_\sigma))}_{\text{2x2 τανυστής διάχυσης}} \nabla u_m \right) \right\}, \quad m=1, \dots, M$$

τεχνητός χρόνος τελεστής προβολής 2x2 τανυστής διάχυσης

$\mathbf{u}(\mathbf{x}, 0)$ = μηδενικό γέμισμα υψηλών συχνοτήτων ($\in \mathcal{U}_{z,S}$)

δομικός τανυστής $J_\rho(\nabla \mathbf{u}_\sigma)$



τανυστής διάχυσης $T(J_\rho(\nabla \mathbf{u}_\sigma))$

$$J_\rho(\nabla \mathbf{u}_\sigma) = G_\rho * \sum_{m=1}^M \nabla(G_\sigma * u_m) (\nabla(G_\sigma * u_m))^T$$

$$T(J_\rho(\nabla \mathbf{u}_\sigma)) = [1 + (\mathcal{N}/K)^2]^{-\frac{1}{2}} \mathbf{w}_- \mathbf{w}_-^T + [1 + (\mathcal{N}/K)^2]^{-1} \mathbf{w}_+ \mathbf{w}_+^T$$

$$\mathcal{N} = \sqrt{\lambda_+ + \lambda_-}$$

ΜΔΕ Προτεινόμενης Μεθόδου

$$\frac{\partial u_m(\mathbf{x}, t)}{\partial t} = P_{\mathcal{U}_{0,S}} \left\{ \operatorname{div} \left(T(J_\rho(\nabla \mathbf{u}_\sigma)) \nabla u_m \right) \right\}, \quad m=1, \dots, M$$

τεχνητός χρόνος
τελεστής προβολής
2x2 τανυστής διάχυσης

$\mathbf{u}(\mathbf{x}, 0)$ = μηδενικό γέμισμα υψηλών συχνοτήτων ($\in \mathcal{U}_{z,S}$)

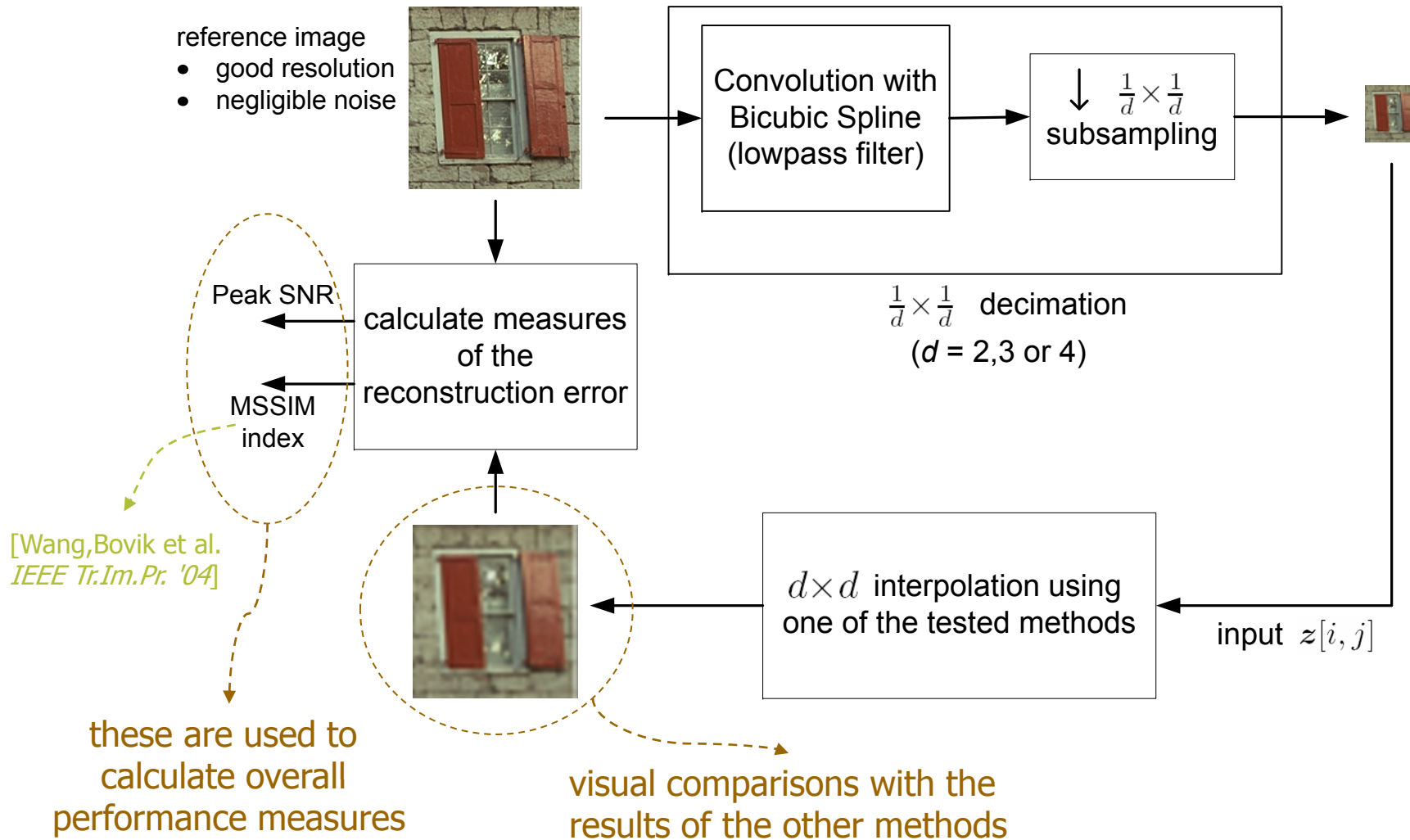
$$u \in \mathcal{U}_{z,S} \iff \sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{S} \left(\frac{2\pi}{\tilde{N}_x} (n_1 + k_1 \tilde{N}_x), \frac{2\pi}{\tilde{N}_y} (n_2 + k_2 \tilde{N}_y) \right) \cdot \hat{u}_{n_1+k_1 \tilde{N}_x, n_2+k_2 \tilde{N}_y} = \hat{z}_{n_1, n_2}$$

$$\left\{ \begin{aligned}
 P_{\mathcal{U}_{0,S}} \{v\} &= v(\mathbf{x}) - w(\mathbf{x}), \\
 \hat{w}_{m_1, m_2} &= \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x} + k_1 2\pi, \frac{2\pi m_2}{\tilde{N}_y} + k_2 2\pi \right) \cdot \hat{v}_{m_1+k_1 \tilde{N}_x, m_2+k_2 \tilde{N}_y} \right\} \cdot \hat{\phi} \left(\frac{2\pi m_1}{\tilde{N}_x}, \frac{2\pi m_2}{\tilde{N}_y} \right) \\
 \hat{\phi}(\omega_1, \omega_2) &= \left\{ \sum_{(k_1, k_2) \in \mathbb{Z}^2} \left| \hat{S}(\omega_1 + k_1 2\pi, \omega_2 + k_2 2\pi) \right|^2 \right\}^{-\frac{1}{2}} \cdot \overline{\hat{S}(\omega_1, \omega_2)}
 \end{aligned} \right.$$

Προηγούμενες Μέθοδοι Παρεμβολής με ΜΔΕ

- Παρεμβολή Βασισμένη στην Ολική Μεταβολή (OM)
[Malgouyres, Guichard, *SIAM J. Num. Anal.* 01]
- Μέθοδος Belahmidi-Guichard (BG)
[Belahmidi, Guichard, *ICIP* 04]
- Παρεμβολή Tschumperle-Deriche (TD)
[Tschumperle, Deriche, *IEEE-PAMI* 05]

Πειράματα Παρεμβολής: Πρωτόκολλο



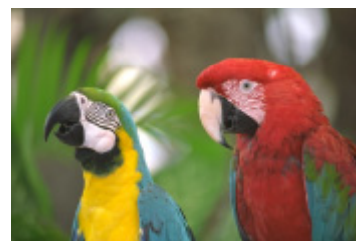
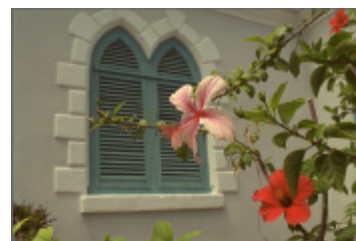
Πειράματα Παρεμβολής: Βάση Εικόνων

- Το πρωτόκολλο επαναλήφθηκε για εικόνες αναφοράς από την βάση εικόνων του *CIPR*:

www.cipr.rpi.edu/resource/stills/kodak.html

23 φυσικές εικόνες
μεγέθους 768x512
εικονοστοιχεία

- Χρησιμοποιήθηκαν τόσο οι ασπρόμαυρες όσο και οι έγχρωμες εκδοχές των εικόνων



8 από τις 23 εικόνες της βάσης

Παράδειγμα Αποτελεσμάτων Ασπρόμαυρης Παρεμβολής (4x4)



(α) Είσοδος (μεγεθυμένη με ZOH)

PSNR=25.58, MSSIM=0.758



(β) Παρεμβολή bicubic

PSNR=26.95, MSSIM=0.815



(γ) OM, πυρήνας sinc

PSNR=27.92, MSSIM=0.846



(δ) OM, πυρήνας μέσης τιμής

PSNR=27.27, MSSIM=0.831



(ε) Παρεμβολή BG

PSNR=26.89, MSSIM=0.818



(ζ) Προτεινόμενη μέθοδος

PSNR=28.54, MSSIM=0.868

Παράδειγμα Αποτελεσμάτων Έγχρωμης Παρεμβολής (4x4)



(α) Είσοδος (μεγεθυμένη με ZOH)

PSNR=20.87, MSSIM=0.523



(β) Παρεμβολή bicubic

PSNR=21.85, MSSIM=0.579



(γ) Παρεμβολή TD

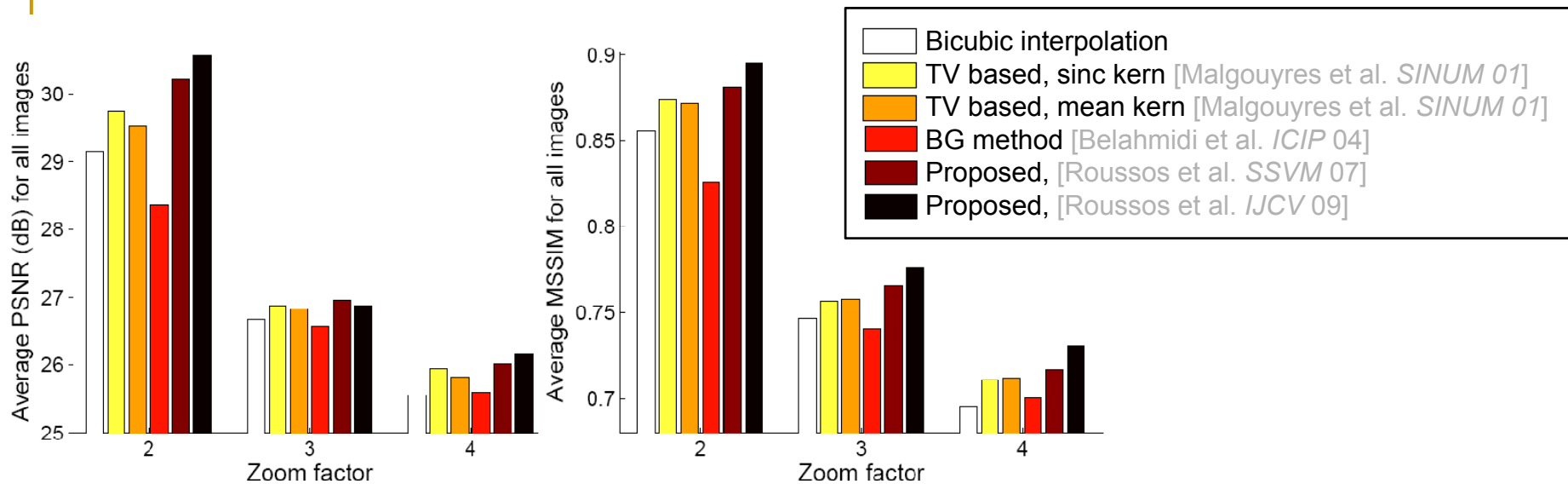
PSNR=19.89, MSSIM=0.458



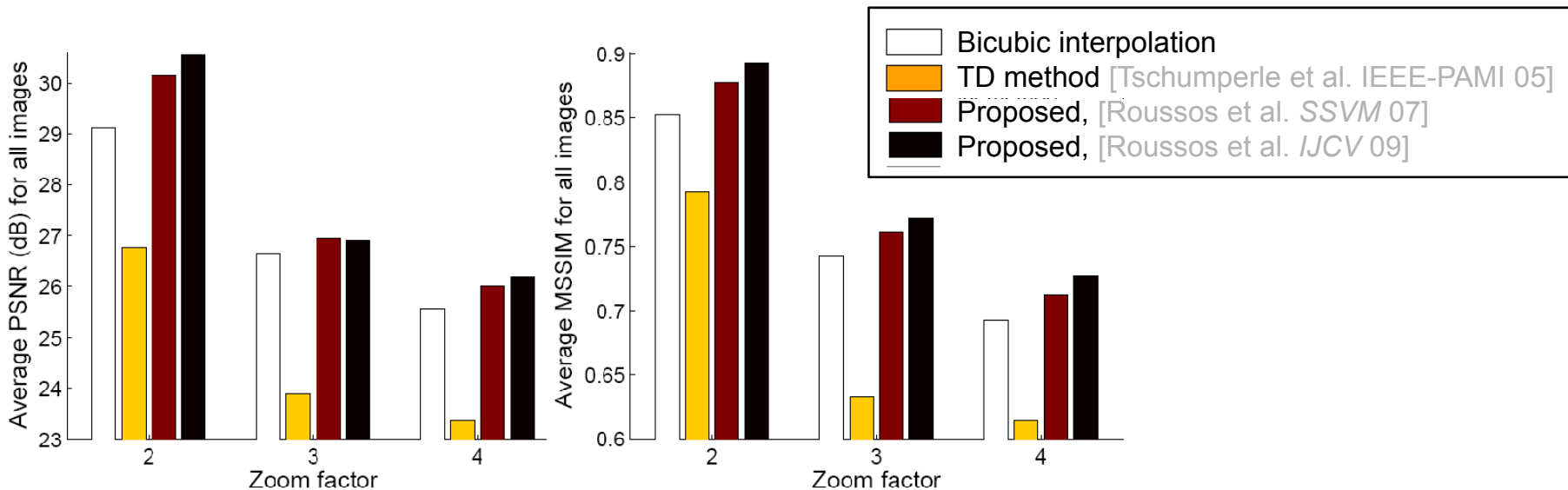
(δ) Προτεινόμενη μέθοδος

PSNR=22.63, MSSIM=0.652

Συγκεντρωτικές Μετρήσεις Απόδοσης



(α) Πειράματα με εικόνες γκριζας κλίμακας



(β) Πειράματα με έγχρωμες εικόνες













Πλήρες Σύνολο Αποτελεσμάτων στο Διαδίκτυο

cvsp.cs.ntua.gr/~tassos/PDEinterp/ssvm07res

- Συγκριτική παράθεση όλων των ~830 εικόνων-αποτελεσμάτων

[Go Back](#)

Table with links to all the experimental results

Image	Zoom factor d			Image	Zoom factor d		
	$d = 2$	$d = 3$	$d = 4$		$d = 2$	$d = 3$	$d = 4$
01. 	[color] [gray]	[color] [gray]	[color] [gray]	13. 	[color] [gray]	[color] [gray]	[color] [gray]
02. 	[color] [gray]	[color] [gray]	[color] [gray]	14. 	[color] [gray]	[color] [gray]	[color] [gray]
03. 	[color] [gray]	[color] [gray]	[color] [gray]	15. 	[color] [gray]	[color] [gray]	[color] [gray]
04. 	[color] [gray]	[color] [gray]	[color] [gray]	16. 	[color] [gray]	[color] [gray]	[color] [gray]
05. 	[color] [gray]	[color] [gray]	[color] [gray]	17. 	[color] [gray]	[color] [gray]	[color] [gray]
06. 	[color] [gray]	[color] [gray]	[color] [gray]	18. 	[color] [gray]	[color] [gray]	[color] [gray]

[Go Back to Results Table](#)

[<< Previous](#) [Next >>](#)

Img.Type = 'Color', Img.No = 14, ZoomFact = 3



[Our method, version of \[2\]](#)



[Our method, version of \[1\]](#)



[Bicubic](#)



[Tschumperle-Deriche method](#)



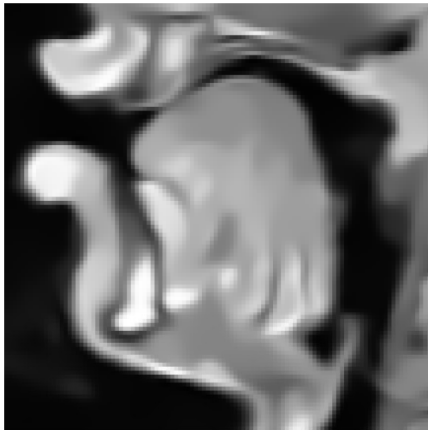
[Input, enlarged by simple ZOH](#)



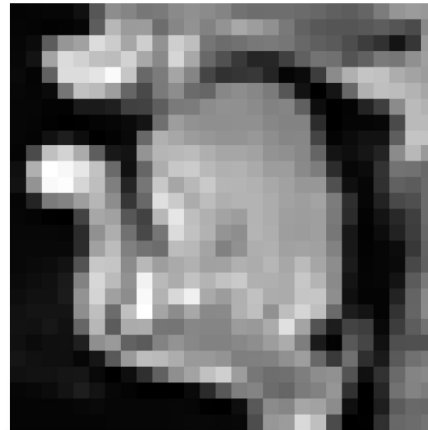
[Reference Image](#)



Παράδειγμα Αποτελεσμάτων Παρεμβολής Εικόνας του Φωνητικού Σωλήνα (4x4)



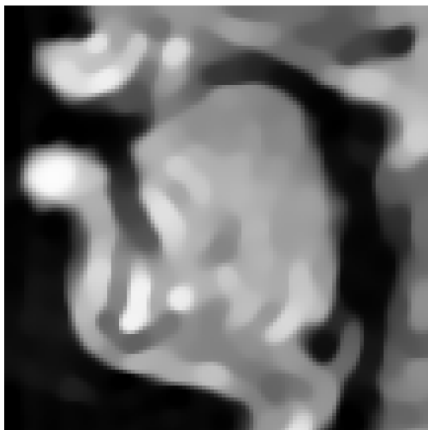
(α) Εικόνα αναφοράς
(108×108 εικονοστοιχεία)



(β) Είσοδος (μεγεθυμένη με ZOH)
PSNR=21.60, MSSIM=0.713



(γ) Παρεμβολή bicubic
PSNR=25.39, MSSIM=0.852



(δ) OM, πυρήνας sinc
PSNR=26.14, MSSIM=0.870



(ε) Παρεμβολή BG
PSNR=25.88, MSSIM=0.870



(ζ) Προτεινόμενη μέθοδος
PSNR=27.69, MSSIM=0.904

Περιεχόμενα

- ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης
- **Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές**
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας
- Σύνοψη Προόδου και Μελλοντική Έρευνα

Μεταβολικά Πλαίσια Διάχυσης: Κίνητρα (1/2)

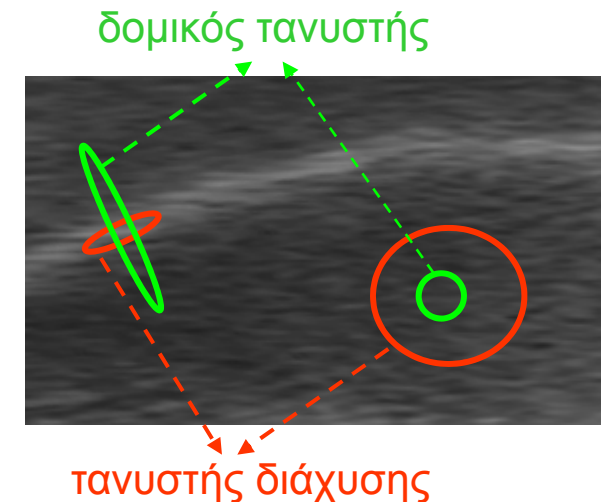
- Μέθοδοι Μη-γραμμικής Διάχυσης για την ΟΥ
 - **Κατηγορία Α: ΜΔΕ Σχεδιασμένες Κατευθείαν**
 - Μέθοδος Perona-Malik [ieeeT-PAMI'90]
 - Ομαλοποιημένη διάχυση των CLMC [Catté et al, siamJNA'92]
 - Διάχυση Βελτίωσης της Συνοχής [Weickert, IJCV'99]
 - Μέθοδος του [Tschumperlé & Deriche, ieeeT-PAMI'05]
 - ⋮
 - **Κατηγορία Β: Μεταβολικές Μέθοδοι**
 - Ολική Μεταβολή [Rudin, Osher & Fatemi, PhysicaD'92]
 - Διανυσματική Ολική Μεταβολή [Sapiro, CVIU'97]
 - Έγχρωμη Ολική Μεταβολή [Blomgren & Chan, ieeeT-IP'98]
 - Ροή Beltrami [Sochen, Kimmel & Maladi, ieeeT-IP'98]
 - ⋮
- Για **κάποιες** μεθόδους της Κατηγορίας Α: **γνωστή σύνδεση με την Κατηγορία Β**, π.χ. :

$$\begin{aligned} \square \text{ Μέθοδος Perona-Malik } \frac{\partial u(x, y, t)}{\partial t} &= \operatorname{div} (g(\|\nabla u\|^2) \nabla u) \\ \square \min_u \int_{\Omega} \varphi(\|\nabla u\|^2) dx &\rightsquigarrow \frac{\partial u}{\partial t} = \operatorname{div} (2\varphi'(\|\nabla u\|^2) \nabla u) \end{aligned} \quad \begin{array}{l} \swarrow \\ \searrow \end{array} \quad g(s^2) = 2\varphi'(s^2)$$

- Αλλά, για αρκετούς τύπους μεθόδων διάχυσης με ΜΔΕ **δεν υπήρχε καμία μεταβολική ερμηνεία**

Μεταβολικά Πλαίσια Διάχυσης: Κίνητρα (2/2)

- Πλεονεκτήματα της **μεταβολικής διατύπωσης** των μεθόδων διάχυσης
 - εννοιολογικά καθαρός φορμαλισμός
 - ευνοεί την μείωση των παραμέτρων
 - ευκολότερη εφαρμογή σε προβλήματα που μπορούν να διατυπωθούν σαν ελαχιστοποίηση με περιορισμούς, π.χ.:
 - αποκατάσταση, inpainting, παρεμβολή εικόνων
 - μπορεί να οδηγήσει σε αποδοτικές υλοποιήσεις βασισμένες σε τεχνικές βελτιστοποίησης
- Πλεονεκτήματα της χρήσης **τανυστών** για την διάχυση εικόνων
 - **Δομικός τανυστής**
αξιόπιστη μέτρηση της μεταβολής & γεωμετρίας της εικόνας στην γειτονιά του κάθε σημείου
 - **Τανυστής διάχυσης**
ευέλικτη προσαρμογή στις δομές της εικόνας



Γενίκευση του Συναρτησιακού Beltrami (1/2)

■ Κλασική Ροή Beltrami

[Sochen, Kimmel & Maladi, IEEE T-IP 98]

- **Ερμηνεία** μιας διανυσματικής εικόνας u με n κανάλια σαν την **2Δ επιφάνεια εμβυθισμένη** στον \mathbb{R}^{n+2} :

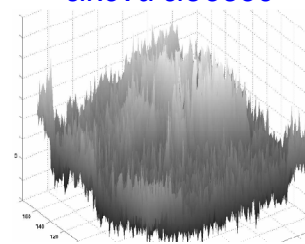
$$(x, y) \longrightarrow (x, y, u_1(x, y), u_2(x, y), \dots, u_n(x, y))$$

- Ροή προς την **ελαχιστοποίηση του εμβαδού της επιφάνειας**: διάχυση βασισμένη σε τανυστές
- Προσφέρει έναν κομψό τρόπο να:
 - συζευχθούν τα διαφορετικά κανάλια της εικόνας
 - επεκταθούν οι ιδιότητες της Ολικής Μεταβολής στις διανυσματικές εικόνες
- Όμως ο τανυστής διάχυσης **δεν είναι ομαλοποιημένος**
→ **περιορισμοί** στην ευρωστία απέναντι στον θόρυβο και στην βελτίωση των ακμών της εικόνας

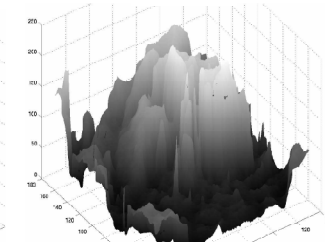
- Για να αποφύγουμε τους περιορισμούς αυτούς, γενικεύουμε το Συναρτησιακό Beltrami ...



εικόνα εισόδου



εμβυθισμένη επιφάνεια



στιγμιότυπο από την ροή

παράδειγμα για την απλούστερη περίπτωση $n=1$

Figure from [Tschumperle, Thesis'02]

Γενίκευση του Συναρτησιακού Beltrami (2/2)

- Χρησιμοποιούμε απεικονίσεις μεγαλύτερων διαστάσεων της μορφής:

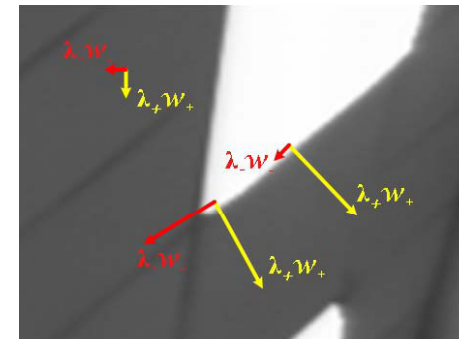
$$\mathbf{x} \rightarrow (\mathbf{x}, \mathcal{P}^u(\mathbf{x}))$$

- $\mathcal{P}^u(\mathbf{x})$: τεμάχιο της εικόνας (patch) [Tschumperle & Brun, ICIP'09], που περιέχει σταθμισμένες τιμές της εικόνας σε ένα πλέγμα που περιβάλλει το σημείο \mathbf{x}
- έτσι, κάθε σημείο συνεισφέρει στο εμβαδόν της επιφάνειας λαμβάνοντας υπόψιν την μεταβολή της εικόνας στην γειτονιά του
- Εάν το βήμα δειγματοληψίας $\rightarrow 0$, το εμβαδόν της εμβυθισμένης επιφάνειας τείνει στο:

$$A[\mathbf{u}] = \int_{\Omega} \sqrt{(\alpha^2 + \lambda_1)(\alpha^2 + \lambda_2)} d\mathbf{x}$$


- $\lambda_i = \lambda_i(J_K(\nabla \mathbf{u}))$ είναι οι ιδιοτιμές του δομικού τανυστή:

$$J_K(\nabla \mathbf{u}) = K * \sum_{i=1}^N \nabla u_i \otimes \nabla u_i$$



Γενικευμένο Συναρτησιακό Βασισμένο στον Δομικό Τανυστή

- $E[\mathbf{u}] = \int_{\Omega} \psi(\lambda_1(J_K(\nabla \mathbf{u})), \lambda_2(J_K(\nabla \mathbf{u}))) \, d\mathbf{x}$
 - $\psi(\lambda_1, \lambda_2)$: συνάρτηση κόστους (αύξουσα)
 - $J_K(\nabla \mathbf{u}) = K * \sum_{i=1}^N \nabla u_i \otimes \nabla u_i$: 2x2 δομικός τανυστής με:
 - ιδιοτιμές λ_1, λ_2 , ιδιοδιανύσματα $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ (εξαρτώνται από τον K)
- Δυσκολία στην θεωρητική ανάλυση:
Σε αντίθεση με τις περισσότερες μεταβολικές μεθόδους, οι **Euler-Lagrange** εξισώσεις δεν μπορούν να εφαρμοστούν εδώ
- Δείξαμε ότι η **ελαχιστοποίηση του συναρτησιακού** οδηγεί στο:

$$\left[\begin{array}{l} \partial u_i / \partial t = \operatorname{div}(D_K \nabla u_i), \quad i = 1, \dots, N, \\ D_K = K * \left(2 \frac{\partial \psi}{\partial \lambda_1} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + 2 \frac{\partial \psi}{\partial \lambda_2} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \end{array} \right.$$


νέος γενικός τύπος ανισοτροπικής διάχυσης

Τανυστική Ολική Μεταβολή

- 1^η ειδική περίπτωση του νέου γενικευμένου συναρτησιακού:

$$E[\mathbf{u}] = \int_{\Omega} \psi(\lambda_1(J_K(\nabla \mathbf{u})), \lambda_2(J_K(\nabla \mathbf{u}))) dx$$

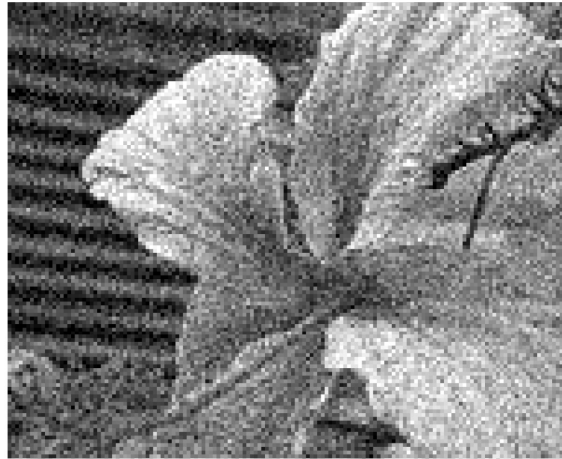
$$\text{με } \psi(\lambda_1, \lambda_2) = \sqrt{\lambda_1} + \sqrt{\lambda_2}$$

- Ροή ταχύτερης καθόδου:

$$\frac{\partial u_i}{\partial t} = \text{div} \left(\left[K * \left(\frac{1}{\sqrt{\lambda_1}} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + \frac{1}{\sqrt{\lambda_2}} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \right] \nabla u_i \right), \quad i = 1, \dots, N$$

- Κλασική ΟΜ: ειδική υπο-περίπτωση με: $N=1$ και $K = \delta(\mathbf{x})$
- Εξομαλύνει προσαρμοστικά την εικόνα
- Συνδυάζει τα πλεονεκτήματα της μεθόδου ΟΜ και των μεθόδων διάχυσης με τανυστές

Τανυστική Ολική Μεταβολή: Παράδειγμα (1)



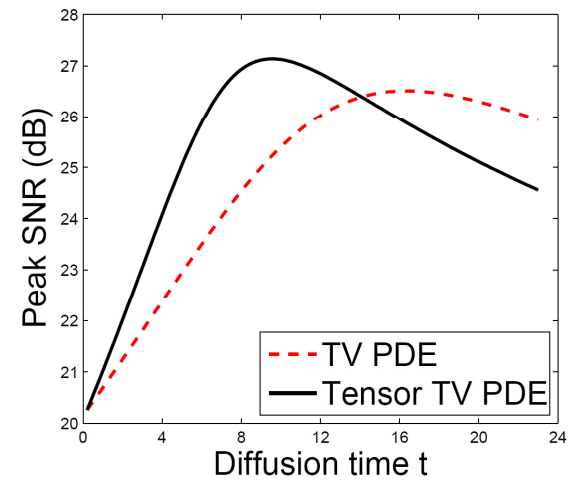
(α') Θορυβώδης Είσοδος (PSNR= 20.1 dB)



(β') ΜΔΕ Ολικής Μεταβολής (*TV PDE*) (PSNR= 26.5 dB)



(γ') ΜΔΕ Τανυστικής Ολικής Μεταβολής (*Tensor TV PDE*) (PSNR= 27.1 dB)



(δ') PSNRs κατά την εξέλιξη των ροών

Τανυστική Ολική Μεταβολή: Παράδειγμα (2)



Ακολουθία εισόδου



Ακολουθία εξόδου

Εφαρμογή της Τανυστικής ΟΜ σε μια ακολουθία εικόνων ακτίνων Χ της φωνητικής οδού ομιλητή.

Γενικευμένη Ροή Beltrami

- 2^η ειδική περίπτωση του νέου γενικευμένου συναρτησιακού :

$$E[\mathbf{u}] = \int_{\Omega} \psi(\lambda_1(J_K(\nabla \mathbf{u})), \lambda_2(J_K(\nabla \mathbf{u}))) dx$$

$$\text{με } \psi(\lambda_1, \lambda_2) = \sqrt{(\alpha^2 + \lambda_1)(\alpha^2 + \lambda_2)}$$

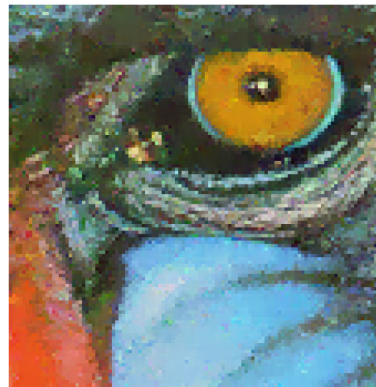
- Ροή ταχύτερης καθόδου:

$$\frac{\partial u_i}{\partial t} = \text{div} \left(\left[K * \left(\sqrt{\frac{\alpha^2 + \lambda_2}{\alpha^2 + \lambda_1}} \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_1 + \sqrt{\frac{\alpha^2 + \lambda_1}{\alpha^2 + \lambda_2}} \boldsymbol{\theta}_2 \otimes \boldsymbol{\theta}_2 \right) \right] \nabla u_i \right)$$

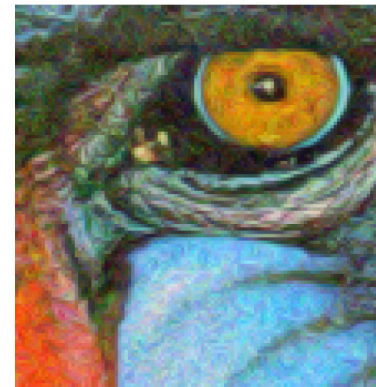
- Κλασική Ροή Beltrami [Sochen et. al, IEEE T-IP 98]: ειδική υπο-περίπτωση με $K = \delta(x)$ και ελαχιστοποίηση στον χώρο των εμβυθίσεων



(α) Θορυβώδης Είσοδος
(PSNR=19.9 dB)



(β) Ροή Beltrami
(PSNR=23.4 dB)



(γ) Γενικευμένη Ροή Beltrami
(PSNR=24.0 dB)

Άλλες Ενδιαφέρουσες Ειδικές Περιπτώσεις

- Άλλες **ειδικές περιπτώσεις** του νέου γενικευμένου συναρτησιακού:

$$E[\mathbf{u}] = \int_{\Omega} \psi(\lambda_1(J_K(\nabla \mathbf{u})), \lambda_2(J_K(\nabla \mathbf{u}))) dx, \text{ με:}$$

- $\psi(\lambda_1, \lambda_2) = \phi(\lambda_1 + \lambda_2)$: Ροή ταχύτερης καθόδου:

$$\partial u_i / \partial t = \operatorname{div} (2 [K * \varphi'(K * \|\nabla \mathbf{u}\|^2)] \nabla u_i)$$

→ νέα ομαλοποίηση του μοντέλου των Perona-Malik
εναλλακτική της κλασικής CLMC [Catté et al, *siamJNA*'92]

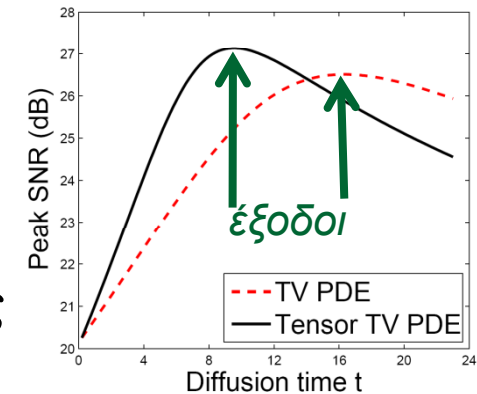
- $K = \delta(\mathbf{x})$ (χωρίς συνέλιξη ομαλοποίησης):

- Έχει μελετηθεί στα [Blomgren & Chan T-IP'98, Tschumperlé & Deriche, T-PAMI'05]
- Η διάχυση είναι ανισοτροπική μόνο αν τα κανάλια της εικόνας είναι $N \geq 2$
- Δεν γίνεται χρήση της πληροφορίας της γειτονιάς κάθε σημείου

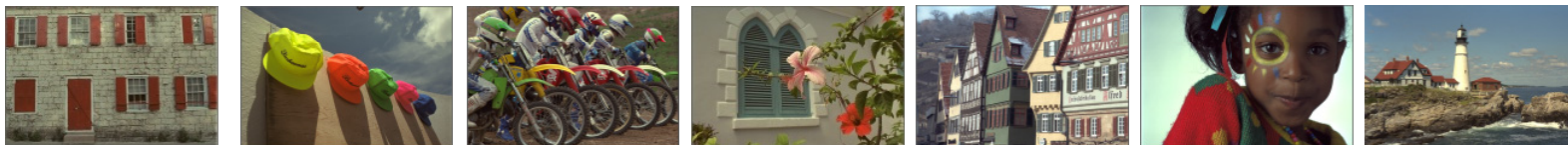
Πειράματα Αποθρομβοποίησης: Πλαίσιο

■ Πλαίσιο Πειραμάτων

- εικόνα αναφοράς με αμελητέο θόρυβο
- προσθήκη γκαουσιανού θορύβου
- χρήση σαν είσοδο στις μεθόδους διάχυσης
- υπολογισμός PSNR κατά τη διάρκεια της ροής κάθε ΜΔΕ, έξοδος: εικόνα με μέγιστο PSNR

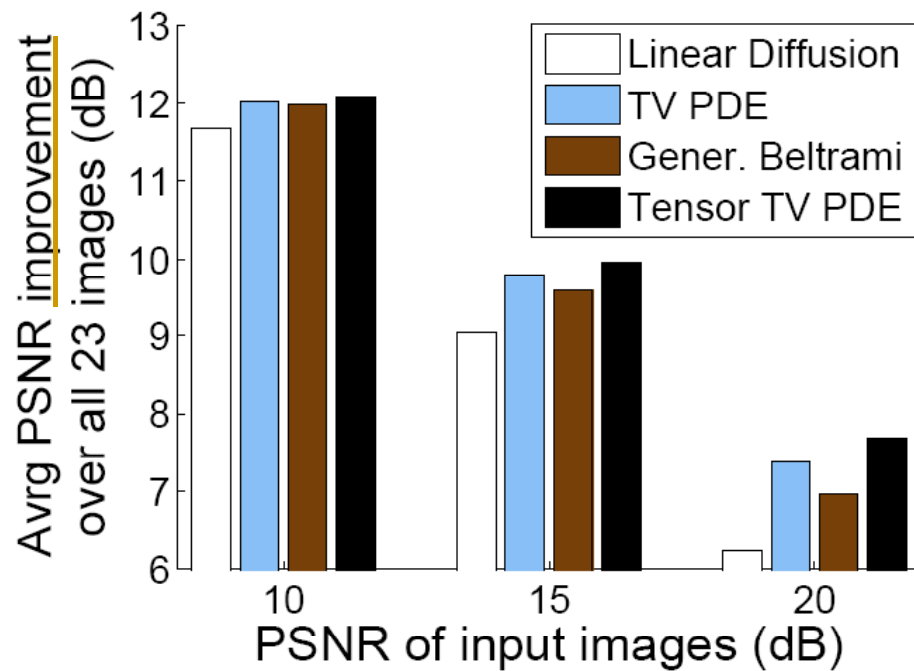


- Επανάληψη αυτού του πλαισίου για εικόνες αναφοράς από την βάση του [CIPR](http://www.cipr.rpi.edu/resource/stills/kodak.html):
www.cipr.rpi.edu/resource/stills/kodak.html

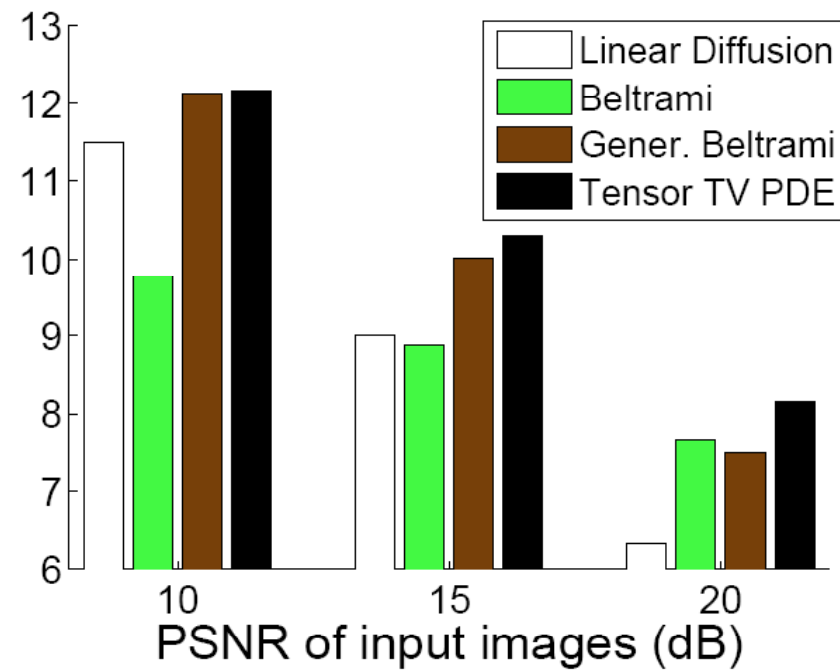


- Χρησιμοποιήθηκαν τόσο οι ασπρόμαυρες όσο και οι έγχρωμες εκδοχές των εικόνων

Πειράματα Αποθουροποίησης: Συγκεντρωτικές μετρήσεις PSNR



(a) Graylevel images

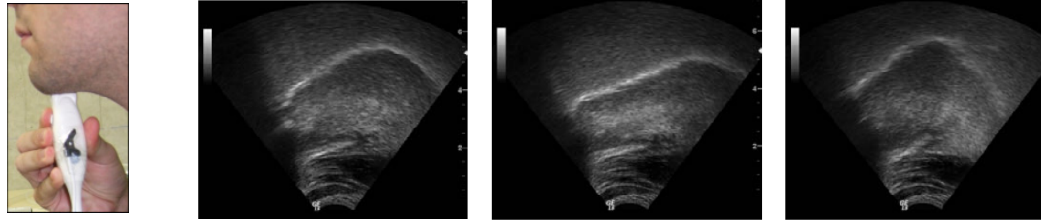


(b) Color images

Περιεχόμενα

- ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης
- Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας
- Σύνοψη Προόδου και Μελλοντική Έρευνα

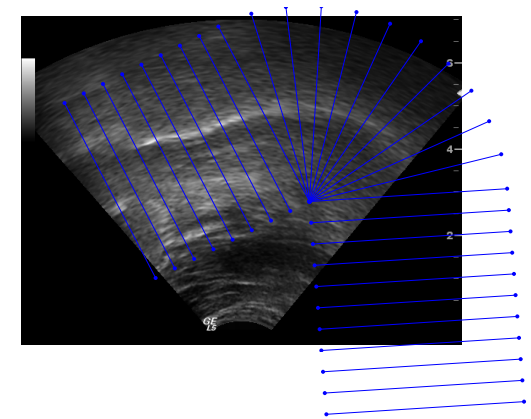
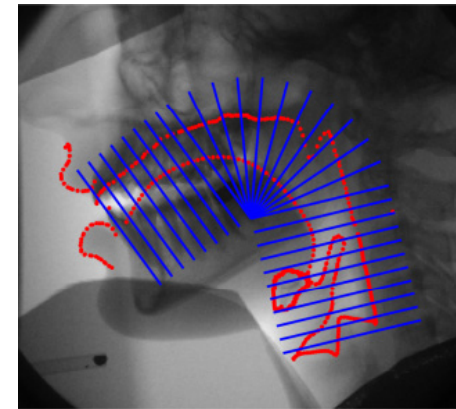
Παρακολούθηση Γλώσσας σε Εικόνες Υπερήχων



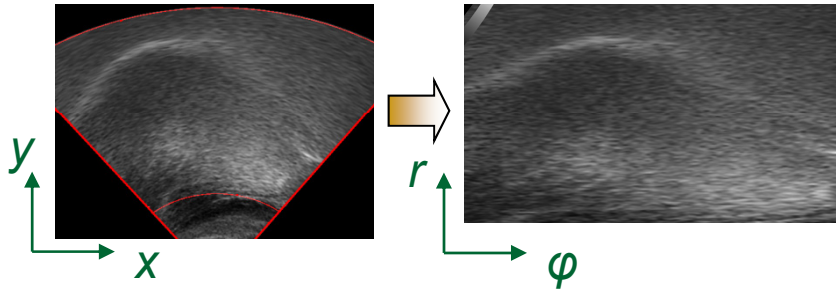
- Αυτόματη παρακολούθηση γλώσσας σε εικόνες υπερήχων
 - ιδιαίτερα χρήσιμη για μεγάλες βάσεις από βίντεο υπερήχων
 - δυσκολίες: υψηλές ποσότητες θορύβου speckle, μη-ορατά μέρη της γλώσσας, έντονη κίνηση & παραμόρφωση της γλώσσας
- Προτείνουμε μία νέα μέθοδο παρακολούθησης
 - βασισμένη σε παραλλαγή των Ενεργών Μοντέλων Εμφάνισης (Active Appearance Models - AAM)
 - ενσωματώνει πρότερη γνώση σχετικά με το σχήμα της γλώσσας

Παρακολούθηση Γλώσσας: Αξιοποίηση Δεδομένων

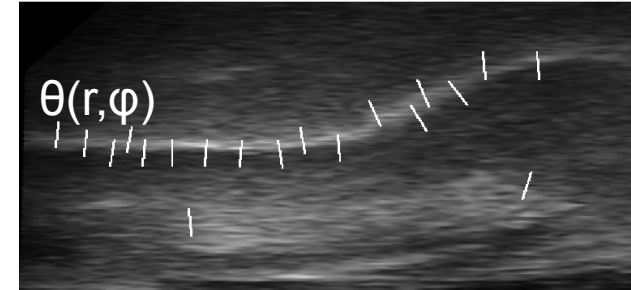
- Καταγεγραμμένα δεδομένα του **ίδιου ομιλητή**:
Υπέρηχοι, ΗΜ αισθητήρες, MRI, Βίντεο ακτίνων Χ
- Εκμετάλλευση των **ακτίνων Χ** για την **μοντελοποίηση του σχήματος** της γλώσσας
 - Χρήση ενός **Πλέγματος Φωνητικής Οδού (ΦΟ)** για την αναπαράσταση του σχήματος της γλώσσας [Maeda, BookChap'90]
- **Εκτίμηση της πόζας του πλέγματος ΦΟ** σε κάθε πλαίσιο υπερήχων, χρησιμοποιώντας δεδομένα των ΗΜ αισθητήρων και MRI του κεφαλιού



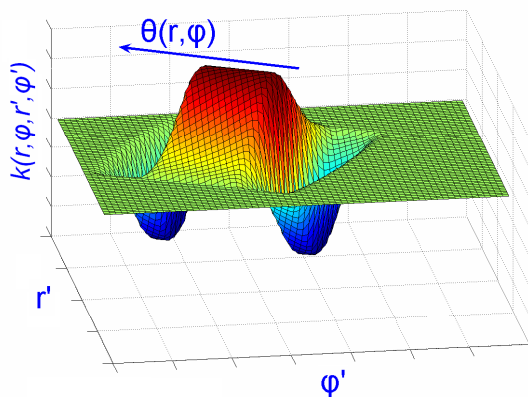
Φιλτράρισμα των πλαισίων υπερήχων



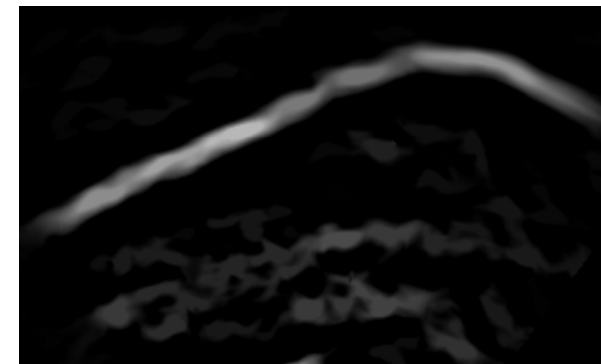
1. Μετατροπή $u(x,y) \rightarrow u(r,\varphi)$



2. Εύρωστη εκτίμηση του προσανατολισμού $\theta(r,\varphi)$ που είναι κάθετος στις ακμές

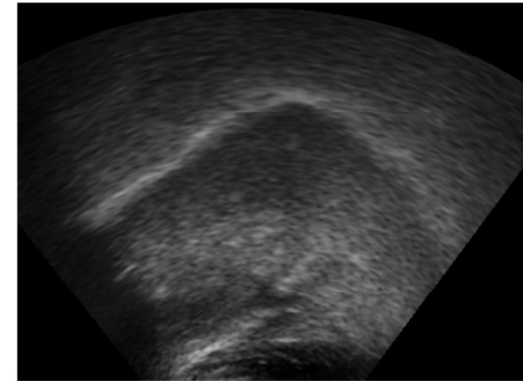
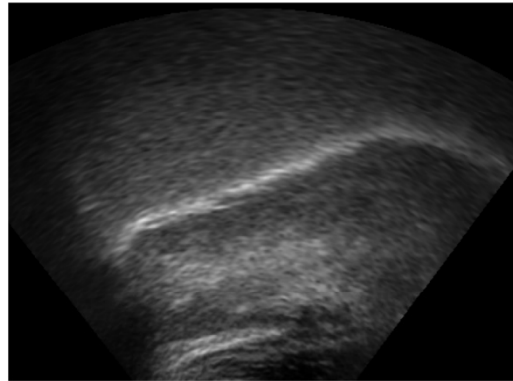
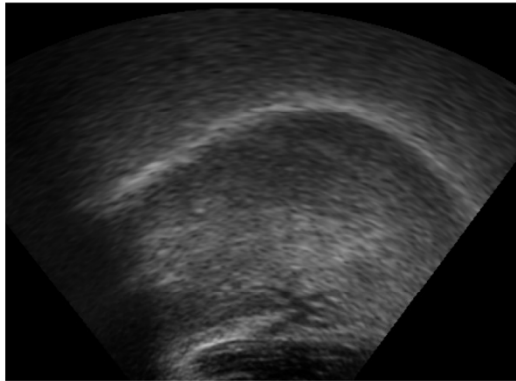


3. Συσχέτιση της $u(r,\varphi)$ με έναν χωρικά μεταβαλλόμενο πυρήνα $k(r,\varphi;r',\varphi')$, ευθυγραμμισμένο με την $\theta(r,\varphi)$

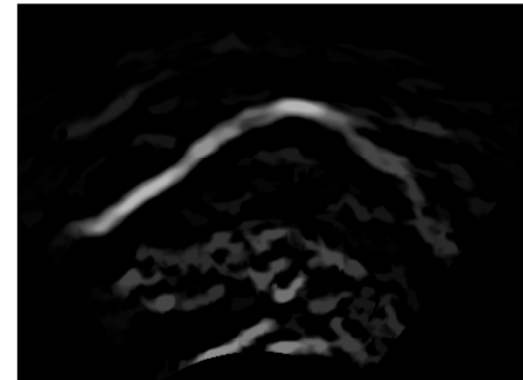
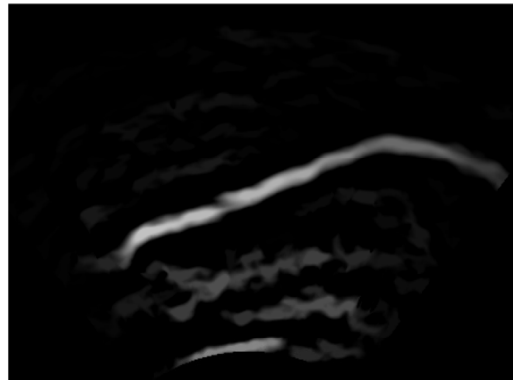
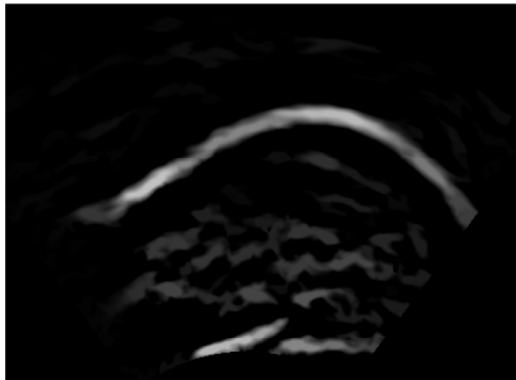


4. Απόρριψη των τιμών <0 , επιστροφή σε (x,y) συντεταγμένες & εφαρμογή *Area Opening*

Φιλτράρισμα των πλαισίων υπερήχων: Παράδειγμα



Ακολουθία εισόδου



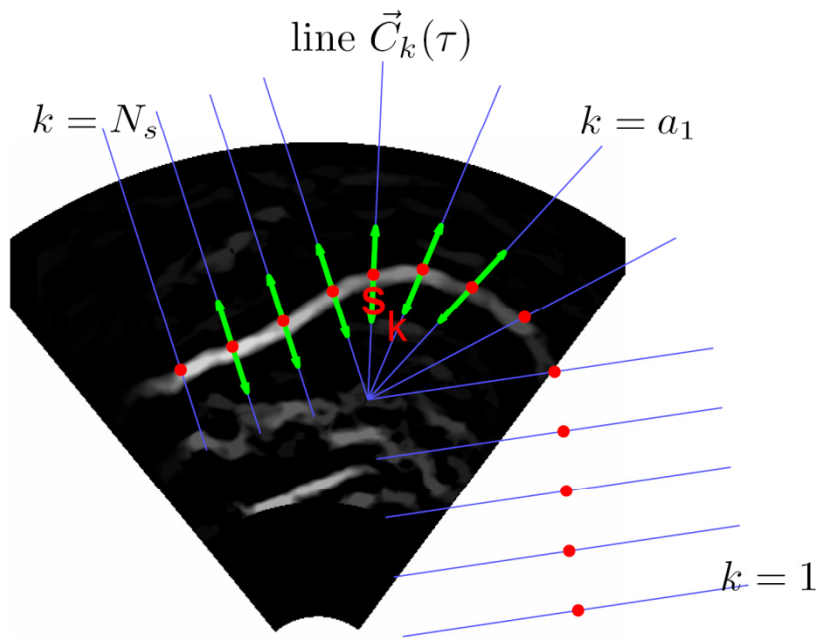
Ακολουθία εξόδου

Πλαίσιο #50

Πλαίσιο #100

Πλαίσιο #150

Αναπαράσταση της Εμφάνισης της Γλώσσας



- **Σχήμα** $\mathbf{s} = [s_1, \dots, s_{N_s}]^T$

- **Υφή** $\mathbf{g}(\mathbf{s}) = \left[\underbrace{[u_{a_1}(s_{a_1}+t)]_{t \in W}^T}_{1 \times N_W} \cdots \underbrace{[u_{a_{N_a}}(s_{a_{N_a}}+t)]_{t \in W}^T}_{1 \times N_W} \right]^T$

- μόνο οι **γραμμές πλέγματος ενεργού υφής** G_{act} χρησιμοποιούνται στην υφή

- $W = \{-d, -d+1, \dots, d\} \cdot \delta l$: παράθυρο δειγματοληψίας

- $u_k(\tau) = u(\vec{C}_k(\tau))$: περιορισμός της εικόνας στην κ-οστή γραμμή πλέγματος

- **Διαφορές** από τα κλασικά AAMs

- Διάφορες τροποποιήσεις για να εκμεταλλευτούμε τα χαρακτηριστικά της συγκεκριμένης εφαρμογής
- Μειωμένη πολυπλοκότητα της αναπαράστασης και μοντελοποίησης
- Ελαφρύτερο πρόβλημα βελτιστοποίησης για το ταίριασμα του μοντέλου

Μοντελοποίηση της Μεταβολής της Εμφάνισης

■ Μοντέλο Σχήματος

$$\mathbf{s} \approx \mathbf{s}_0 + \mathbf{Q}_s \mathbf{b}$$

- \mathbf{b} : κανονικοποιημένες παράμετροι σχήματος με $p(\mathbf{b}) = \mathcal{N}(\mathbf{b} | \mathbf{0}, \mathbf{I}_{N_b})$
- PCA για την εκμάθηση των \mathbf{s}_0 , \mathbf{Q}_s
 - διανύσματα εκπαίδευσης από επισημειωμένες καμπύλες γλώσσας σε 700 πλαίσια ακτίνων X

■ Μοντέλο Υφής

$$\mathbf{g} = \mathbf{g}_0 + \mathbf{Q}_g \boldsymbol{\lambda} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\lambda}$: παράμετροι υφής με $p(\boldsymbol{\lambda}) = \mathcal{N}(\boldsymbol{\lambda} | \mathbf{0}, \mathbf{I}_{N_\lambda})$
- $\boldsymbol{\varepsilon}$: σφάλμα ανακατασκευής υφής με:

$$p(\boldsymbol{\varepsilon}) = \mathcal{N}(\boldsymbol{\varepsilon} | \mathbf{0}, \boldsymbol{\Sigma}_\varepsilon), \quad \boldsymbol{\Sigma}_\varepsilon = \tilde{\mathbf{Q}}_g \text{diag}(\rho_1, \dots, \rho_{N_g}) \tilde{\mathbf{Q}}_g^T$$

- Εκπαίδευση του μοντέλου
 - Επισημειώσεις σε 400 πλαίσια υπερήχων. Αυτό το σύνολο εκπαίδευσης χωρίζεται σε 3 2 υποσύνολα T1 και T2
 - Από το υποσύνολο T1: εκμάθηση των \mathbf{g}_0 and \mathbf{Q}_g με PCA
 - Από το υποσύνολο T2: εκμάθηση των βέλτιστων παραμέτρων $\rho_1, \dots, \rho_{N_g}$

Παρακολούθηση μέσω του Ταιριάσματος του Μοντέλου

- Ταιρίασμα του Μοντέλου Εμφάνισης σε κάθε πλαίσιο υπερήχων
- Εκτίμηση με κριτήριο *maximum a posteriori* των παραμέτρων σχήματος \mathbf{b} και υφής $\boldsymbol{\lambda}$ μέσω της μεγιστοποίησης:

$$p(\mathbf{b}, \boldsymbol{\lambda} | u(x, y)) \propto p(u | \mathbf{b}, \boldsymbol{\lambda}) p(\mathbf{b}, \boldsymbol{\lambda}) = p(\boldsymbol{\varepsilon}) p(\mathbf{b}) p(\boldsymbol{\lambda})$$

$$\boldsymbol{\varepsilon} = \mathbf{g}(\mathbf{s}(\mathbf{b})) - \mathbf{g}_0 - \mathbf{Q}_g \boldsymbol{\lambda}$$

- Ισοδύναμα: ελαχιστοποίηση της ενέργειας:

$$E(\mathbf{b}, \boldsymbol{\lambda}) = -\ln p(\mathbf{b}, \boldsymbol{\lambda} | u) = C + \frac{1}{2} \{ \|\mathbf{b}\|^2 + \|\boldsymbol{\lambda}\|^2 + \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon} \}$$

- Κλίσεις της ενέργειας: $\nabla_{\mathbf{b}} E = \mathbf{b} + \mathbf{Q}_s^T (\partial \mathbf{g} / \partial \mathbf{s})^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}$

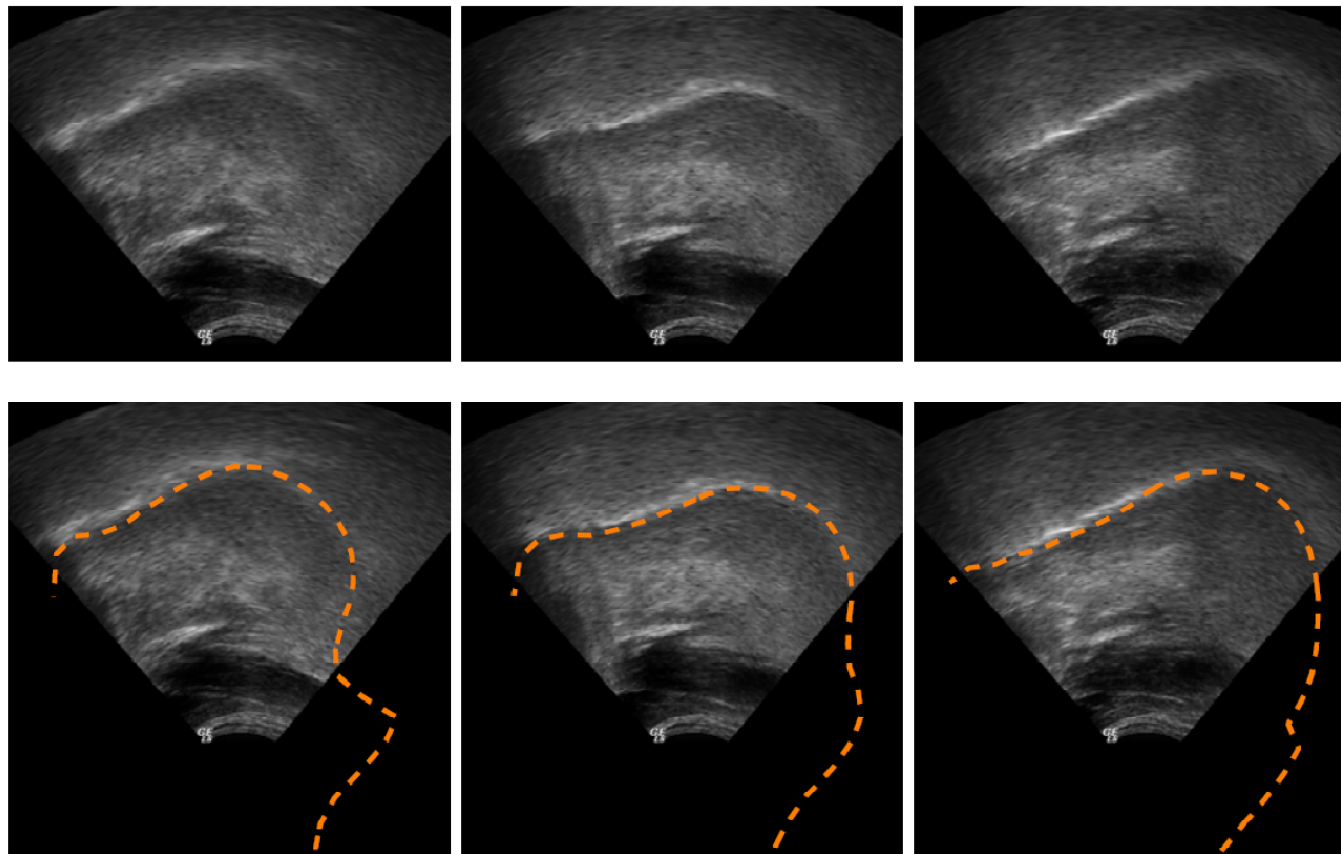
$$\nabla_{\boldsymbol{\lambda}} E = \boldsymbol{\lambda} - \mathbf{Q}_g^T \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}$$

$$\text{όπου: } \frac{\partial \mathbf{g}}{\partial s_k} = \begin{cases} [0 \cdots 0]^T, & \text{if } k \notin G_{act} \\ \left[\underbrace{0 \cdots 0}_{(k-1)N_W} \quad [u'_k(s_k+t)]_{t \in W}^T \quad \underbrace{0 \cdots 0}_{(N_s-k)N_W} \right]^T, & \text{if } k \in G_{act} \end{cases}$$

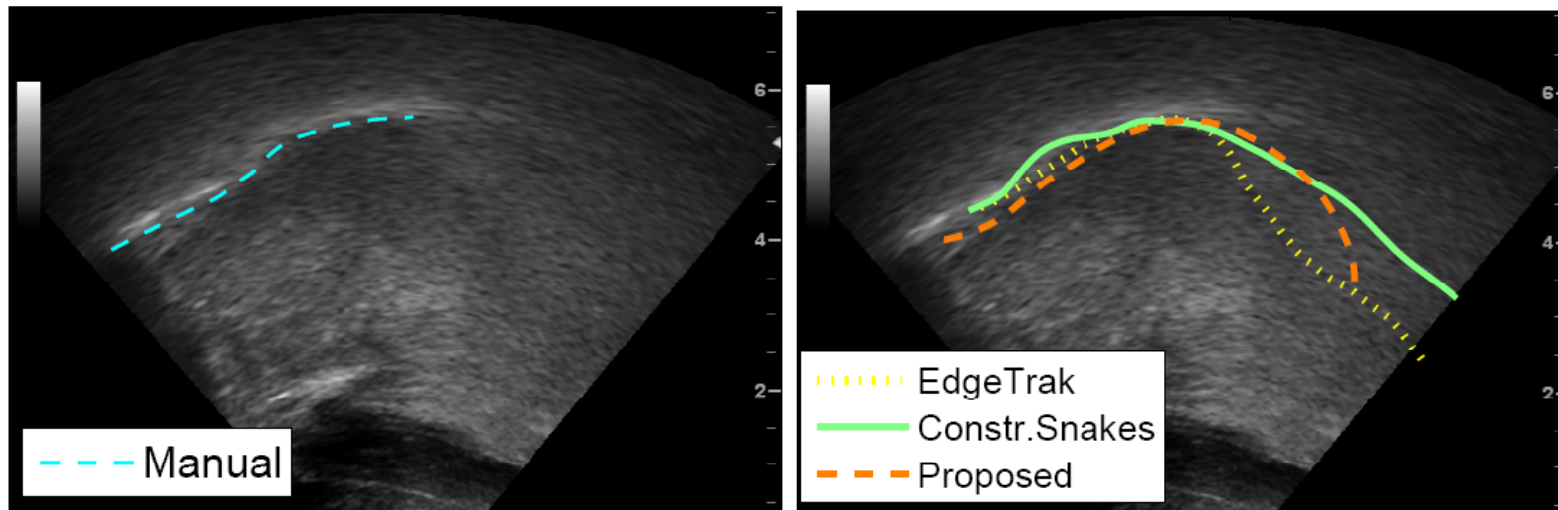
- Αλγόριθμος βελτιστοποίησης
 - Ροή ταχύτερης καθόδου
 - Αρχικοποίηση των παραμέτρων:
 - \mathbf{b}_0 : από το αποτέλεσμα του προηγούμενου πλαισίου
 - $\boldsymbol{\lambda}_0$: μεγιστοποίηση της ύστερης πιθανότητας $p(\boldsymbol{\lambda} | \mathbf{g}(\mathbf{s}(\mathbf{b}_0)))$

Παράδειγμα Εφαρμογής Προτεινόμενης Μεθόδου

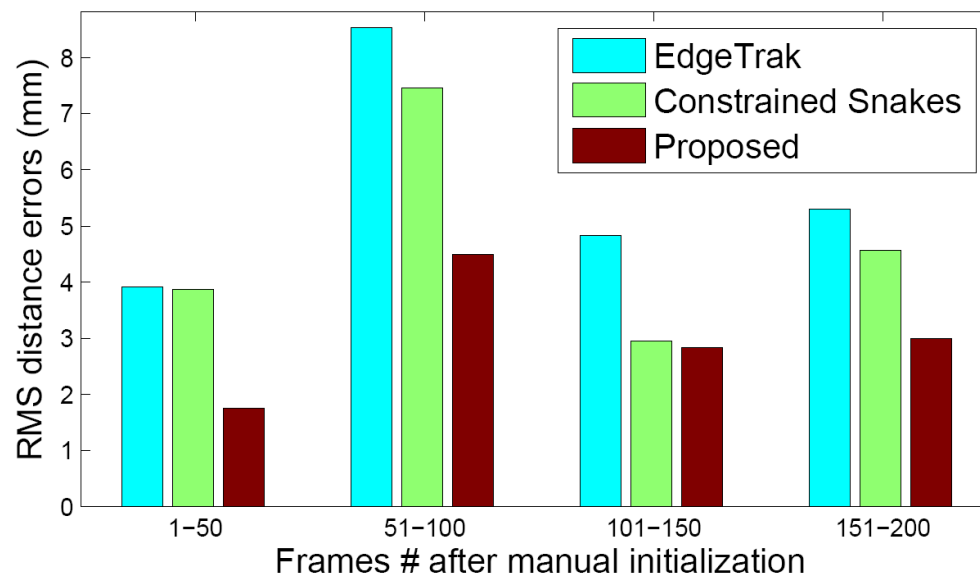
	Dimensionality of original vector	Number of model parameters	Variance explained (% of the total)
Shape	30	6	96%
Texture	1215	35	93%



Πειραματικές Συγκρίσεις



$$e_d = \sqrt{(d_{om}^2 + d_{mo}^2)/2}$$



Περιεχόμενα

- ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης
- Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- **Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας**
- Σύνοψη Προόδου και Μελλοντική Έρευνα

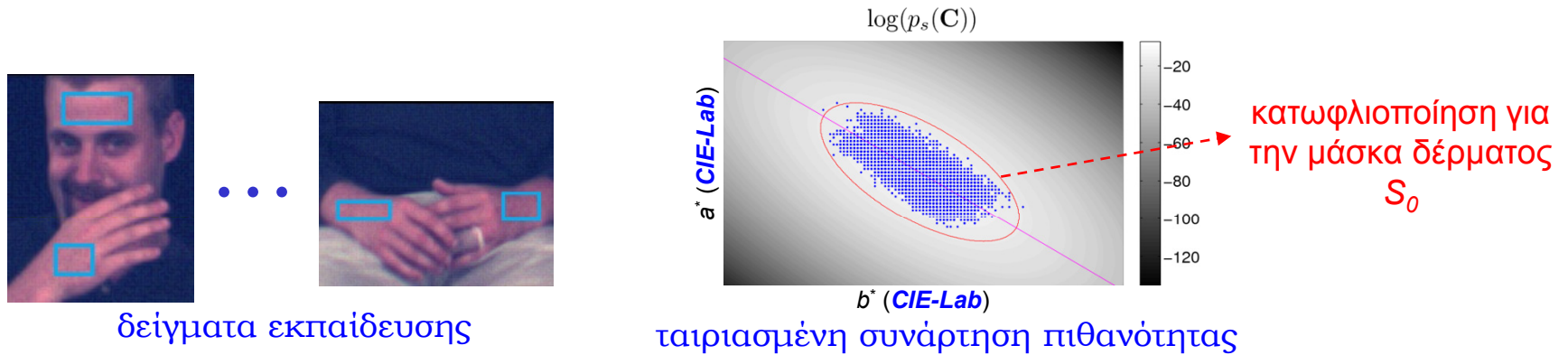
Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας



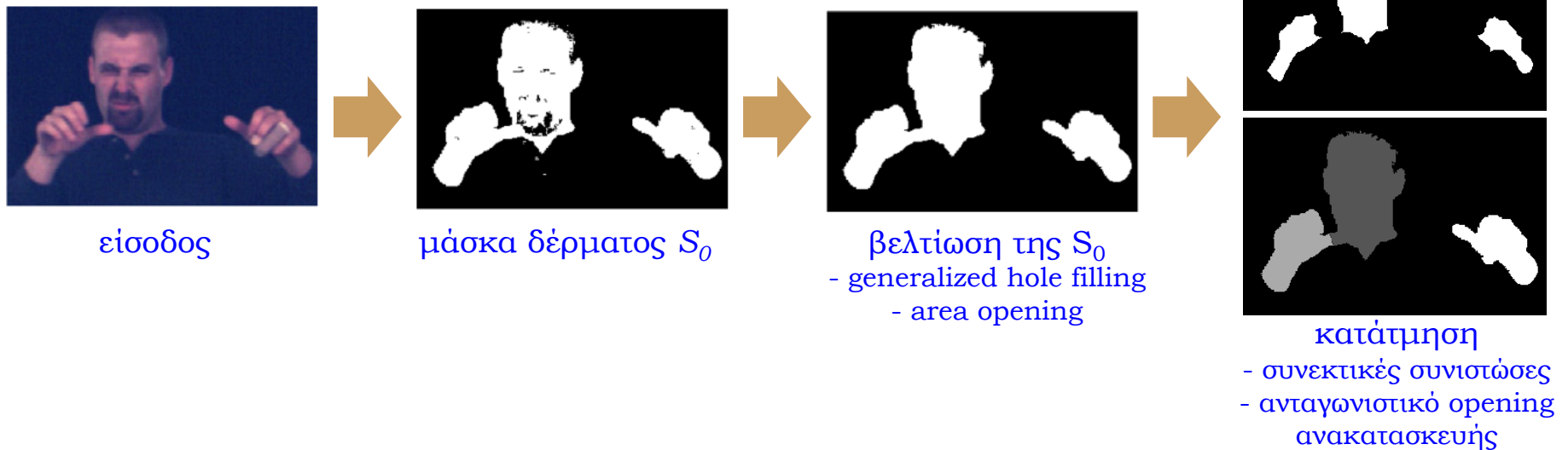
- Ανάλυση βίντεο συνεχούς νοηματισμού
- Στόχοι:
 - εντοπισμός & παρακολούθηση των χεριών + κεφαλιού του νοηματιστή
 - εξαγωγή χαρακτηριστικών που να περιγράφουν αξιόπιστα την πόζα & διαμόρφωση των χεριών του νοηματιστή
- Απώτερος στόχος:
 - αυτόματη αναγνώριση νοηματικής γλώσσας

Αρχική Παρακολούθηση Χεριών-Κεφαλιού (1/2)

- Πιθανοτική μοντελοποίηση του χρώματος δέρματος

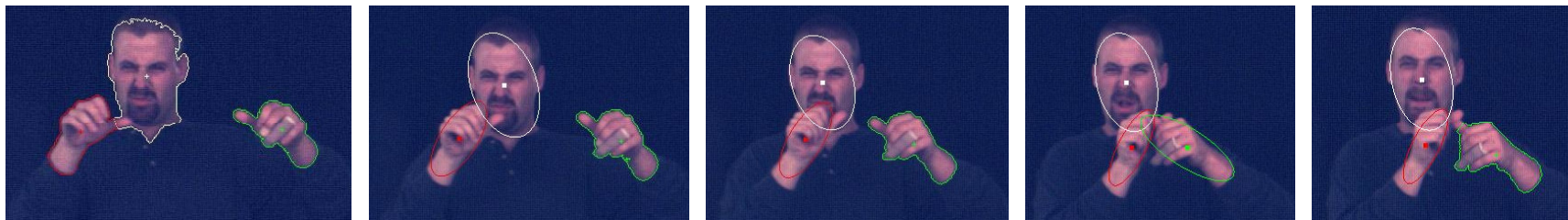


- Μορφολογική επεξεργασία της μάσκας δέρματος



Αρχική Παρακολούθηση Χεριών-Κεφαλιού (2/2)

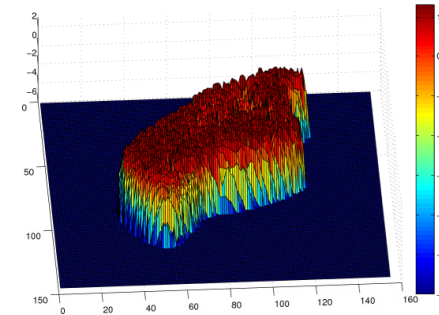
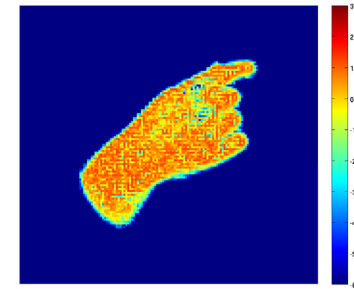
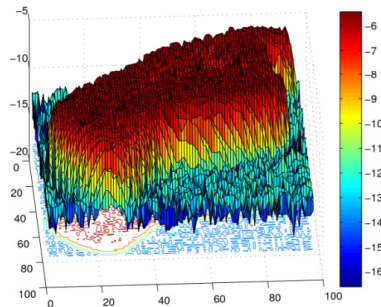
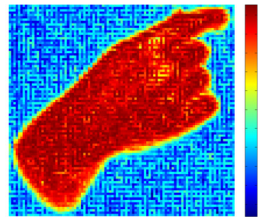
- Κύρια στοιχεία της αρχικής παρακολούθησης:
 - χρονικά αμφίδρομη πρόβλεψη (forward-backward),
 - template matching,
 - ταίριασμα ελλείψεων,
 - πιθανοτικοί περιορισμοί



- Έξοδος: σύνολο από μάσκες περιοχών δέρματος + απονομή ετικετών {H,R,L} στις μάσκες



Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ): Αναπαράσταση



$I(x)$

αρχική κομμένη
εικόνα του
χεριού

$g(I(x))$

προβολή του (a^*, b^*)
κάθε pixel στον
πρωτεύοντα άξονα της
γκαουσιανής $p_s(a^*, b^*)$

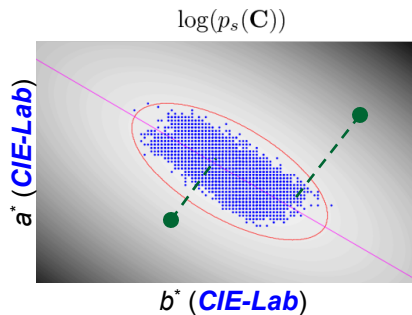
M

μάσκα
δέρματος

Εικόνα Σχήματος-Εμφάνισης:

$$f(x) = \begin{cases} g(I(x)), & \text{εάν } x \in M \\ -c_b, & \text{διαφορετικά} \end{cases}$$

σταθερά ισορροπίας ανάμεσα στο
σχήμα και την εμφάνιση



Μοντελοποίηση Σχήματος-Εμφάνισης

Εικόνα Σχήματος-Εμφάνισης

του χεριού

εικόνα βάσης

ιδιοεικόνες

$$f(W_{\mathbf{p}}(\mathbf{x})) \approx A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x})$$

2Δ αφινικός
μετασχηματισμός

$$W_{\mathbf{p}}(x, y) = \begin{pmatrix} 1 + p_1 & p_3 & p_5 \\ p_2 & 1 + p_4 & p_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ): Εκπαίδευση των $A_i(\mathbf{x})$ (1/2)

- Σύνολο Εκπαίδευσης
 - εξαγωγή εικόνων χεριού **χωρίς επικαλύψεις** από βίντεο εκπαίδευσης
 - **τυχαία επιλογή** 500 τέτοιων εικόνων

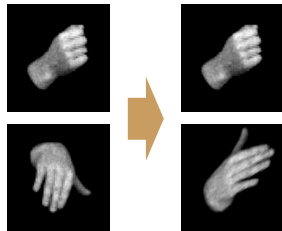


- Αφινική Ευθυγράμμιση του Συνόλου Εκπαίδευσης
.....

Αφινική Ευθυγράμμιση του Συνόλου Εκπαίδευσης



- Επίπεδο 1: 1-1 ευθυγράμμιση
 - Χρήση του Inverse-Compositional Αλγόριθμου [Gross,Matthews,Baker, IVC'05]



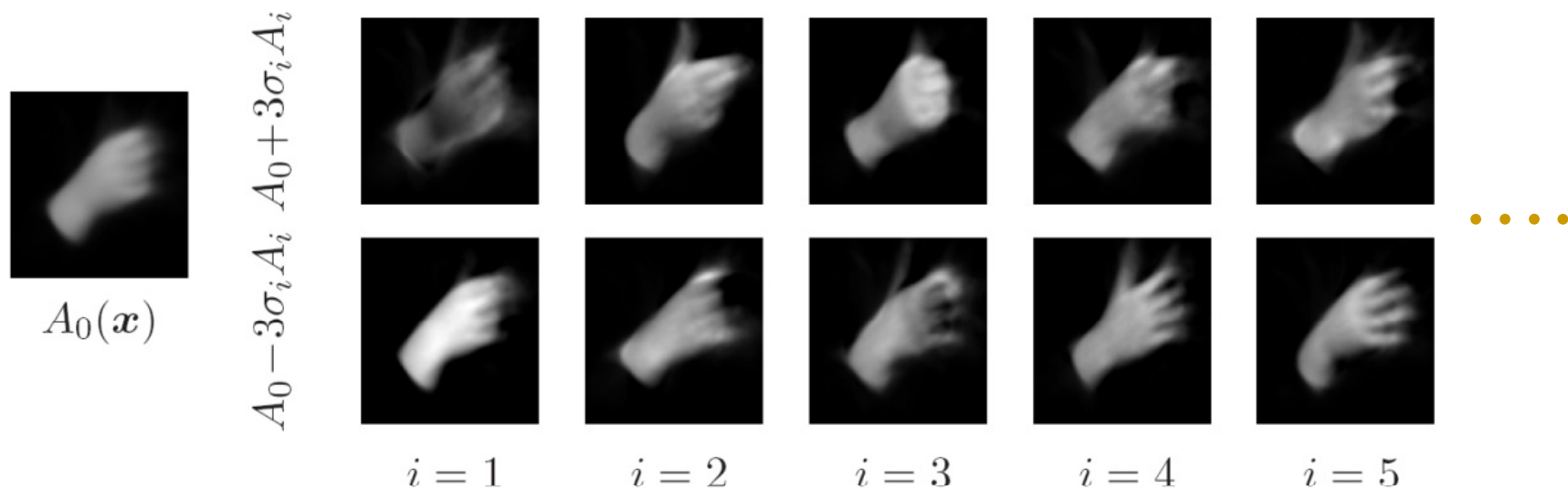
- Επίπεδο 2: Ευθυγράμμιση του συνόλου εκπαίδευσης
 - Γενίκευση της Προκρούστειας Ανάλυσης [Cootes,Taylor, TecRep'04]

- Επίπεδο 3: Επαναληπτική ανάδραση από τον χρήστη

	f_{ref}							
ΕΙΚΟΝΕΣ ΕΙΣΟΔΟΥ								...
ΕΠΙΠΕΔΟ 1	Ευθυγράμμιση συνόλου (Επίπεδο 2)							
Ανάδραση από τον χρήστη	✓	×	✓	✓	×	×	✓	
ΕΠΙΠΕΔΟ 2	Ευθυγράμμιση συνόλου (Επίπεδο 2)							
Ανάδραση από τον χρήστη	✓	✓	✓	✓	✓	×	✓	ΑΠΟΡΡΙΨΗ
ΕΠΙΠΕΔΟ 3	Ευθυγράμμιση συνόλου (Επίπεδο 2)							
Ανάδραση από τον χρήστη	✓	✓	✓	✓	✓		✓	

Μοντέλο Σχήματος-Εμφάνισης (ΜΣΕ): Εκπαίδευση των $A_i(\mathbf{x})$ (2/2)

- **Ανάλυση Πρωτεύουσών Συνιστωσών (PCA)** του αφινικά ευθυγραμμισμένου συνόλου εκπαίδευσης
- Κρατάμε 35 ιδιοεικόνες $A_i(\mathbf{x})$ που εξηγούν το 78% της μεταβλητότητας
- Η αφινική ευθυγράμμιση προσφέρει σημαντική μείωση της διακύμανσης



Μοντέλο Σχήματος-Εμφάνισης: Ταίριασμα

- Έξοδοι: **εύρωστη** παρακολούθηση & εξαγωγή χαρακτηριστικών χεριού
- Εύρεση παραμέτρων λ, p που ελαχιστοποιούν την **ομαλοποιημένη ενέργεια**:

$$E(\lambda, p) = E_{rec}(\lambda, p) + w_S E_S(\lambda, p) + w_D E_D(\lambda, p)$$

$$E_{rec}(\lambda, p) = \frac{1}{N_M} \sum_{\mathbf{x}} \left\{ A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x}) - f(W_p(\mathbf{x})) \right\}^2$$

μέσο τετραγωνικό σφάλμα ανακατασκευής

$$E_S(\lambda, p) = \frac{1}{N_c} \|\lambda - \lambda_0\|_{\Sigma_\lambda}^2 + \frac{1}{N_p} \|p - p_0\|_{\Sigma_p}^2$$

στατική πρότερη πληροφορία

$$E_D(\lambda, p) = \frac{1}{N_c} \|\lambda - \lambda^e\|_{\Sigma_{\epsilon_\lambda}}^2 + \frac{1}{N_p} \|p - p^e\|_{\Sigma_{\epsilon_p}}^2$$

δυναμική πρότερη πληροφορία

- Δυναμικά μοντέλα για **Γραμμική Πρόβλεψη**:

$$\lambda^e[n] = \sum_{\nu \in W(K)} A_{K,\nu} \lambda[n - \nu] \quad , \quad \tilde{p}^e[n] = \sum_{\nu \in W(K)} B_{K,\nu} \tilde{p}[n - \nu]$$

- $\Sigma_\lambda, \Sigma_p, \Sigma_{\epsilon_\lambda}, \Sigma_{\epsilon_p}$: μήτρες συμμεταβλητότητας
- λ_0, p_0 : μέσες τιμές

προσδιορισμός μέσω εκπαίδευσης με πλαίσια χωρίς επικαλύψεις

- Χρήση του **Inverse-Compositional Αλγόριθμου** [Baker et al, TecRep'04]



Frame #10116 (start)



Frame #10121



Frame #10126



Frame #10131



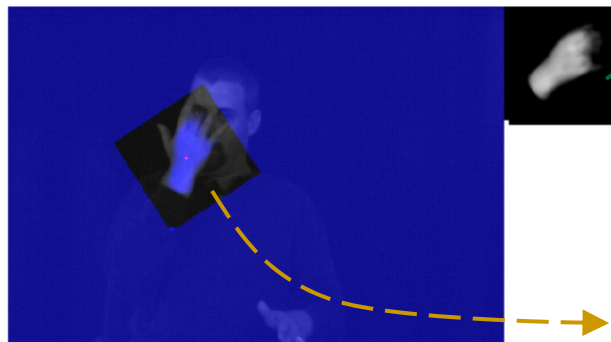
Frame #10136



Frame #10141



Frame #10146



Frame #10151

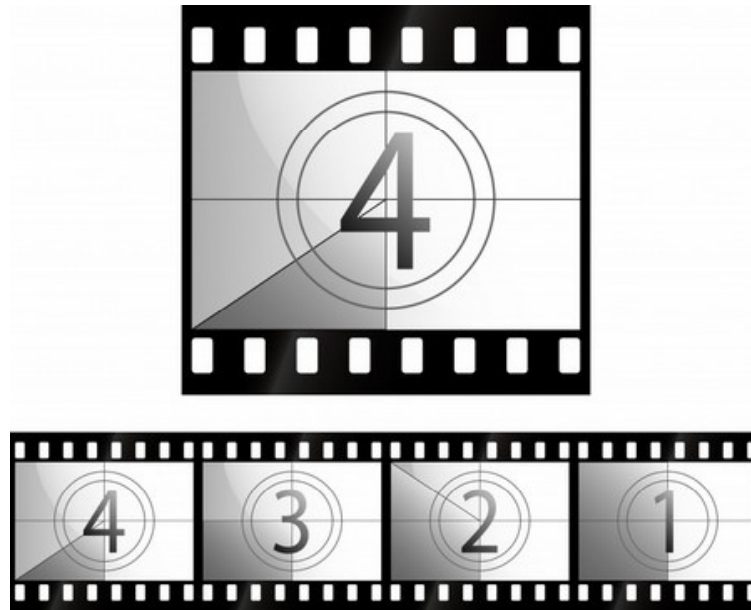
ανακατασκευή στο πεδίο του μοντέλου

$$A_0(\mathbf{x}) + \sum \lambda_i A_i(\mathbf{x})$$

ανακατασκευή στο πεδίο της εικόνας

$$A_0(W_p^{-1}(\mathbf{x})) + \sum \lambda_i A_i(W_p^{-1}(\mathbf{x}))$$

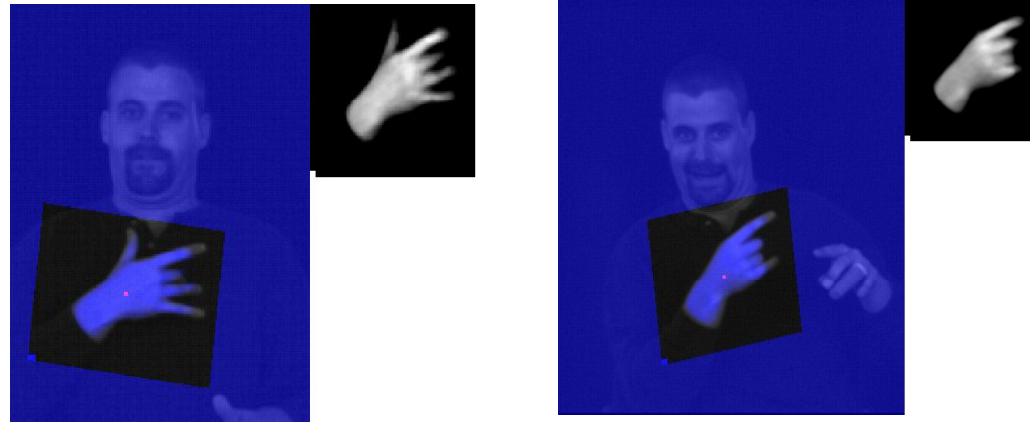
Παράδειγμα Ταιριάσματος Μοντέλου Σχήματος-Εμφάνισης



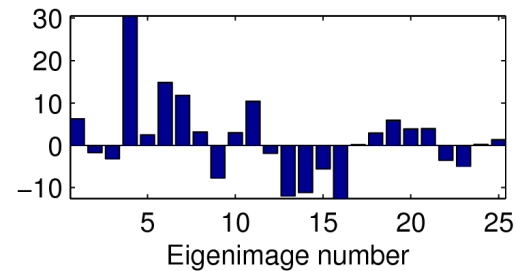
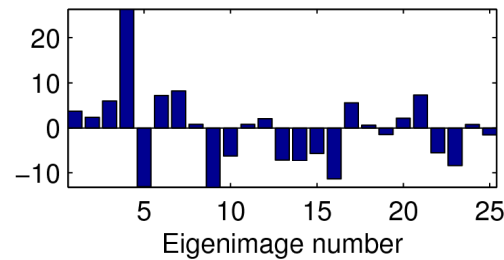
(Video)

Εξαγωγή Χαρακτηριστικών

πλαίσια εισόδου
+
ταιριάσματα του
μοντέλου



βάρη των
ιδιοεικόνων
 λ



χαρακτηριστικά
χειρομορφής

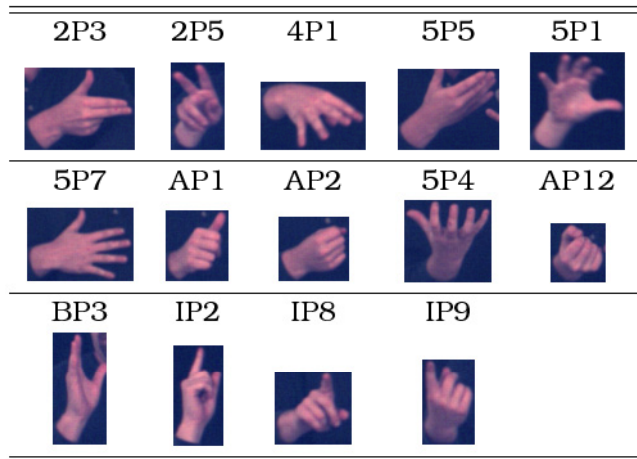
παράμετροι του
αφινικού μετ/μου

p

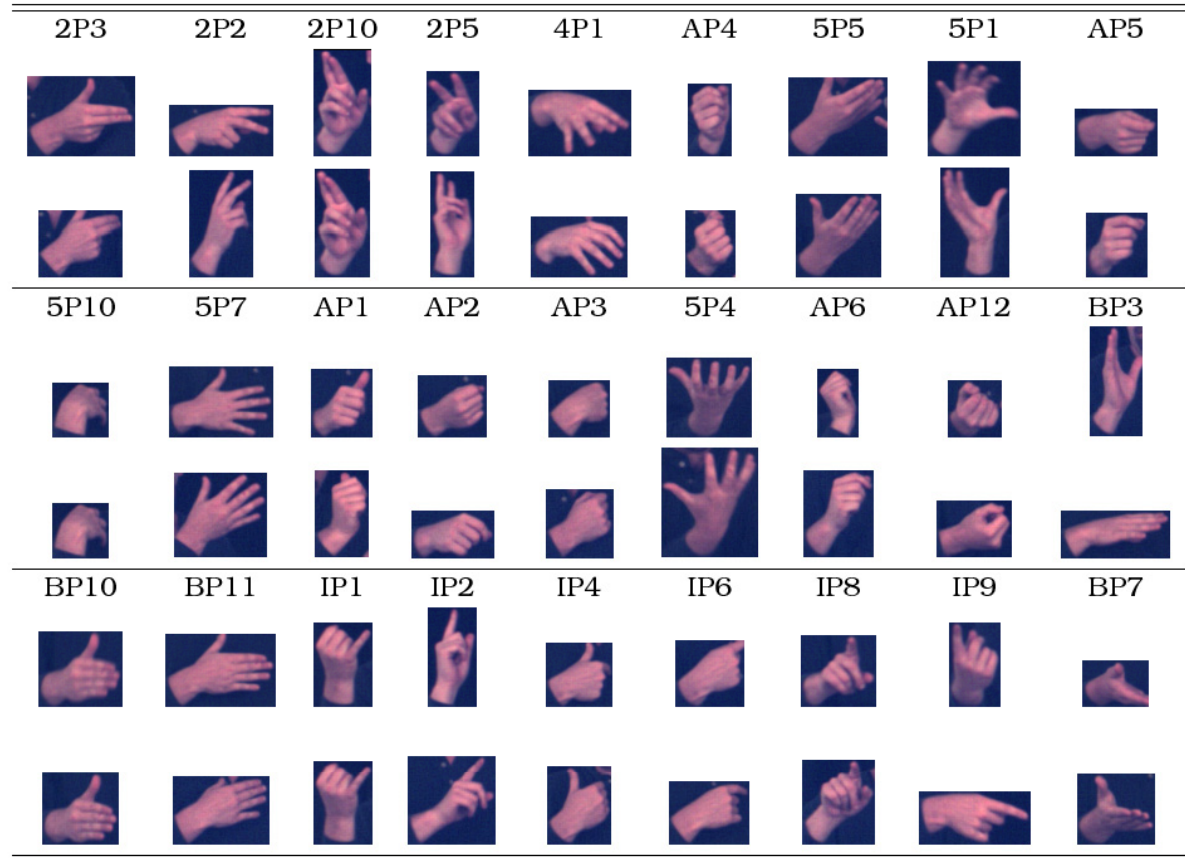
$$\begin{pmatrix} -0.0061 & -0.0944 & -78.1642 \\ 0.1033 & 0.0552 & -128.2917 \end{pmatrix}$$

$$\begin{pmatrix} -0.0185 & -0.0278 & -95.0785 \\ 0.1260 & -0.0499 & -139.1400 \end{pmatrix}$$

Πειράματα Ταξινόμησης Χειρομορφών



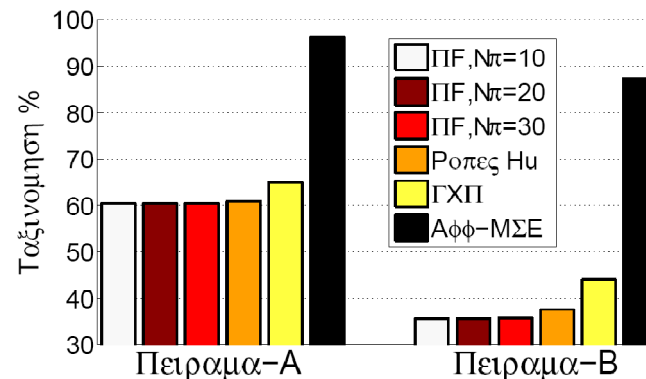
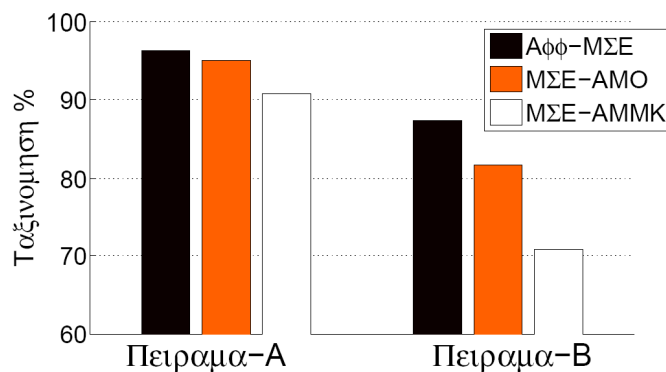
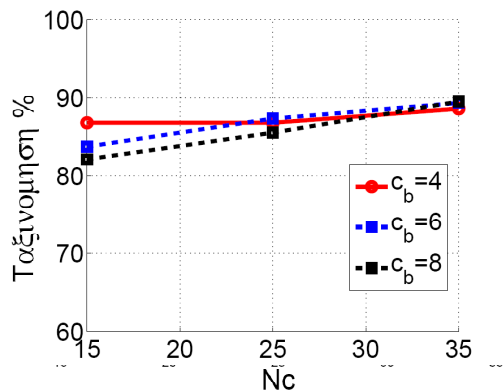
**Τάξεις
Πειράματος-A**



**Τάξεις
Πειράματος-B**

Αποτελέσματα Ταξινόμησης Χειρομορφών

- Ίδια διαδικασία εκπαίδευσης & ταξινόμησης παντού
- Μεταβολή **μόνο** του διανύσματος χαρακτηριστικών



Προτεινόμενη μέθοδος για το Πείραμα-B: Μεταβολή των κυρίων παραμέτρων:

- # των PCA συνιστωσών
- C_b : Σταθερά παρασκηνίου για τις εικόνες ΣΕ

Σύγκριση της προτεινόμενης μεθόδου με απλοποιημένες εκδοχές της:

- **Αφφ-ΜΣΕ**: Αφινική Μοντελοποίηση Σχήματος-Εμφάνισης (προτεινόμενη)
- **ΜΣΕ-ΑΜΟ**: Μοντελοποίηση ΣΕ με Απευθείας Μετασχηματισμούς Ομοιότητας
- **ΜΣΕ-ΑΜΜΚ**: Μοντελοποίηση ΣΕ με Απευθείας Μετασχηματισμούς Μετατόπισης & Κλιμάκωσης

Σύγκριση της προτεινόμενης μεθόδου με βασικές μεθόδους:

- **ΠΦ**: Περιγραφητές Fourier με 10,20,30 συντελεστές
- **Ροπές Hu**: αναλοιώτες Ροπές Hu της περιοχής του χεριού
- **ΓΧΠ**: Γεωμετρικά χαρακτηριστικά περιοχής του χεριού (εμβαδό, εκκεντρότητα, βαθμός συμπύκνωσης, μήκη ελάσσονος και μείζονος άξονα του χεριού)
- **Αφφ-ΜΣΕ**: προτεινόμενη

Περιεχόμενα

- ΜΔΕ & Μοντέλα Σχήματος στην Όραση Υπολογιστών
- Εφαρμογές στην Οπτική Ανάλυση Αρθρωτών
 - Έρευνα του Συστήματος Παραγωγής Φωνής
 - Αυτόματη Αναγνώριση Νοηματικής Γλώσσας
- Παρεμβολή Εικόνων μέσω Μη-γραμμικής Διάχυσης
- Μεταβολικά Πλαίσια για Μοντέλα Διάχυσης με Τανυστές
- Παρακολούθηση της Γλώσσας με Ενεργά Μοντέλα Εμφάνισης
- Μοντελοποίηση Χειρομορφών Νοηματικής Γλώσσας
- **Σύνοψη Προόδου και Μελλοντική Έρευνα**

Σύνοψη Προόδου

- Ανάπτυξη πρωτότυπων μεθόδων μη-γραμμικής διάχυσης για την βελτίωση της ποιότητας εικόνων
 - Μέθοδος ανισοτροπικής διάχυσης - προβολής για την βελτίωση της ευκρίνειας διανυσματικών εικόνων
 - Θεωρητικό πλαίσιο το οποίο βασίζεται στον δομικό τανυστή της εικόνας και γενικεύει διάφορες μεταβολικές μεθόδους μη-γραμμικής διάχυσης.
- Σχεδίαση στατιστικών μοντέλων σχήματος για την παρακολούθηση και ταξινόμηση αντικειμένων
 - Στατιστικό μοντέλο για την ανίχνευση της καμπύλης της γλώσσας ομιλητή
 - Αφινικά αναλλοίωτη μοντελοποίηση των χειρομορφών νοηματικής γλώσσας. Ομαλοποιημένη παρακολούθηση του χεριού και εξαγωγή χαρακτηριστικών χειρομορφής

Δημοσιεύσεις Διατριβής

1. A. Roussos and P. Maragos. Reversible interpolation of vectorial images by an anisotropic diffusion-projection PDE. *International Journal of Computer Vision*, 84(2), August 2009.
2. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Dynamic affine-invariant shape-appearance model for hand tracking and feature extraction in continuous sign language. Under preparation to be submitted to the *International Journal of Computer Vision*.
3. A. Roussos and P. Maragos. Vector-valued image interpolation by an anisotropic diffusion-projection PDE. In *Scale Space and Variational Methods in Computer Vision, First International Conference, SSVM-2007 Proceedings*, volume 4485 of *Lecture Notes in Computer Science*, pages 104–115. Springer-Verlag, 2007.
4. M. Aron, A. Roussos, M.-O. Berger, E. Kerrien, and P. Maragos. Multimodality Acquisition of Articulatory Data and Processing. In *Proceedings of the European Signal Processing Conference (EUSIPCO), Lausanne, 2008*.
5. A. Katsamanis, A. Roussos, P. Maragos, M. Aron, and M.-O. Berger. Inversion from audiovisual speech to articulatory information by exploiting multimodal data. In *International Seminar on Speech Production*, December 2008.
6. A. Roussos, A. Katsamanis, and P. Maragos. Tongue tracking in ultrasound images with active appearance models. In *Proceedings of the International Conference on Image Processing*, November 2009.
7. A. Roussos and P. Maragos. Tensor-based image diffusions derived from generalizations of the total variation and beltrami functionals. In *Proceedings of the International Conference on Image Processing*, September 2010.
8. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Affine-invariant modeling of shape-appearance images applied on sign language handshape classification. In *Proc. Int'l Conf. on Image Processing*, September 2010.
9. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. In *Proc. of Workshop on Sign, Gesture and Activity, 11th ECCV*, September 2010.

Κατευθύνσεις Μελλοντικής Έρευνας

- Ανάπτυξη **αποδοτικών αλγορίθμων** υλοποίησης της μεθόδου μη-γραμμικής διάχυσης για την παρεμβολή εικόνων
- Γενικό Μεταβολικό Πλαίσιο για μοντέλα διάχυσης με τανυστές και συνελίξεις ομαλοποίησης:
 - προοπτική **επέκτασης της χρήσης τους** σε διάφορα προβλήματα, όπως αποκατάσταση, deblurring, παρεμβολή και inpainting εικόνων
- **Θεωρητικά ζητήματα** σχετικά με μεθοδολογίες μη-γραμμικής διάχυσης: Ύπαρξη, μοναδικότητα, ...
- Μέθοδος παρακολούθησης της γλώσσας σε εικόνες υπερήχων:
 - **ενσωμάτωση πρότερης δυναμικής πληροφορίας**
- Μοντελοποίηση των Χειρομορφών νοηματικής γλώσσας
 - περαιτέρω **ποσοτική αξιολόγηση** & συγκρίσεις
 - χρήση της εξαγωγής χαρακτηριστικών χειρομορφών σε ένα σύστημα **αναγνώρισης νοηματικής γλώσσας**
 - **ανεξαρτησία** από τον νοηματιστή & συνθήκες καταγραφής βίντεο
 - εφαρμογή ίδιας μοντελοποίησης και στο **δευτερεύον χέρι και το κεφάλι** & συνδυασμός των τριών μοντέλων

Ευχαριστώ για την προσοχή σας!

Ερωτήσεις;



Ιστοσελίδα Ομάδας Όρασης Υπολογιστών,
Επικοινωνίας Λόγου & Επεξεργασίας Σήματος: cvsp.cs.ntua.gr

Προσωπική ιστοσελίδα: www.troussos.gr