

Προτάσεις Διπλωματικών Εργασιών  
Σε Θέματα Ορασης Υπολογιστών και Επικοινωνίας Ανθρώπου-Ρομπότ

Computer Vision, Speech Communication & Signal Processing Group

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

Επιβλέπων: Καθ. Π. Μαραγκός ([maragos@cs.ntua.gr](mailto:maragos@cs.ntua.gr))

Συνεργάτες: Δρ. Β. Πιτσικάλης, Δρ. Γ. Ευαγγελόπουλος, Α. Ρούσσοι, Σ. Θεοδωράκης  
[gevag, vpitsik, troussos, sth][@cs.ntua.gr](mailto:>@cs.ntua.gr)

20-12-2009

1. “Μοντελοποίηση και αυτόματη αναγνώριση εκφράσεων του προσώπου”
2. “Αυτόματη Αναγνώριση Νοηματικής Γλώσσας: Μοντελοποίηση στοιχειωδών μονάδων κίνησης και θέσης ”
3. “Αυτόματη Αναγνώριση Συνεχούς Νοηματικής Γλώσσας: Παύση, Επένθεση και Μετάβαση”
4. “Εντοπισμός και παρακολούθηση χεριών σε ακολουθίες εικόνων νοηματικής γλώσσας”
5. “Αλληλεπίδραση ανθρώπου-ρομπότ μέσω αναγνώρισης κινήσεων και χειρονομιών”

## “Μοντελοποίηση και αυτόματη αναγνώριση εκφράσεων του προσώπου ”

Περιγραφή:

Η αναγνώριση ανθρώπινων προσώπων από στατικές εικόνες ή βίντεο στοχεύει στην λήψη αυτόματων αποφάσεων για την ύπαρξη ανθρώπων σε μια σκηνή, τη θέση τους, την ταυτοποίησή τους και σηματοδοτεί γεγονότα όπως για παράδειγμα ομιλία, διάλογοι, δράσεις, χειρονομίες, αφηγηματικά περιστατικά κ.λπ. Μια ιδιαίτερη υποκατηγορία του γενικότερου προβλήματος είναι η μοντελοποίηση και αναγνώριση των εκφράσεων του προσώπου που με εφαρμογές στις περιοχές της αναγνώρισης φωνής, μελέτης συμπεριφοράς, αναγνώρισης δράσεων, επικοινωνίας ανθρώπου-ρομπότ, γραφικά με υπολογιστές και συναισθηματική υπολογιστική (συναίσθηση, ανίχνευση και ερμηνεία των ανθρώπινων συναισθηματικών καταστάσεων).

Οι εκφράσεις του προσώπου αποτελούν την οπτική εκδήλωση της συναισθηματικής κατάστασης, της γνωσιακής δραστηριότητας, της πρόθεσης, της προσωπικότητας ή της ψυχολογικής κατάστασης. Για την αυτόματη αναγνώριση το μεγαλύτερο τμήμα της τρέχουσας βιβλιογραφίας έχει εμπνευστεί από το Σύστημα Κωδικοποίησης Δράσεων του Προσώπου (Facial Action Coding System, FACS) που εισάχθηκε στην συμπεριφοριστική επιστήμη από τους Ekman και Friesen. Βασίζεται σε ένα πρωτότυπο των βασικών ανθρώπινων εκφράσεων και επιτρέπει την μελέτη των εκφράσεων με βάση την ανατομική ανάλυση των κινήσεων του προσώπου.

Στόχος της διπλωματικής είναι η ανάπτυξη τεχνικών και αλγορίθμων στην κατεύθυνση της αυτόματης ανάλυσης και αναγνώρισης δράσεων του προσώπου (facial actions) με χρήση τεχνικών της Όρασης Υπολογιστών. Η προσπάθεια μπορεί να έχει ως αφετηρία τη βελτίωση της επίδοσης τεχνικών βασισμένων σε ειδικά διαμορφωμένα, αντικειμενοστραφή μοντέλα όπως είναι τα Μοντέλα Ενεργής Εμφάνισης (Active Appearance Models, AAMs). Έμφαση θα δοθεί στην ανάπτυξη μηχανισμών προσαρμογής και την εξαγωγή περιγραφικών ανεξάρτητων της ταυτότητας του προσώπου, με σκοπό την ανάδειξη χαρακτηριστικών της καθολικότητας των εκφράσεων του ανθρώπινου προσώπου. Πιθανές εφαρμογές περιλαμβάνουν την αναγνώριση νοηματικής γλώσσας, τη σύνθεση φωνής με συναισθηματική χροιά, και την εξαγωγή μοντέλων συναισθηματικής προσοχής και σημαντικότητας από δεδομένα ταινιών.

Λέξεις κλειδιά:

Περιγραφείς προσώπου, Σημεία ενδιαφέροντος, Παραμορφώσιμα μοντέλα, Αναγνώριση δράσεων, Εντοπισμός και αναγνώριση προσώπων, Ανάλυση συναισθημάτων.

Ενδεικτικές Δημοσιεύσεις:

- [1]Ekman, P. and Friesen, W., “Facial action coding system”. Palo Alto, CA, USA, 1978.
- [2]Donato, G., and Bartlett, M., and Hager, J., and Ekman, P., and Sejnowski, T., “Classifying Facial Actions”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(10):974-989, Oct. 1999.
- [3]Cootes, T. F. and Edwards, G. J. and Taylor, C. J., “Active Appearance Models”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.23, no.6, pp. 681-685, 2001.
- [4] Papandreou, G. and Maragos, P., “Adaptive and Constrained Algorithms for Inverse Compositional Active Appearance Model Fitting”, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2008), Anchorage, AL, U.S.A., June 2008.

## “Αυτόματη Αναγνώριση Νοηματικής Γλώσσας : Μοντελοποίηση στοιχειωδών μονάδων κίνησης και θέσης ”

Περιγραφή:

Οι νοηματικές γλώσσες αποτελούν ένα πλήρη κώδικα επικοινωνίας μεταξύ ατόμων με προβλήματα ακοής και την αποκλειστική γλώσσα επικοινωνίας της κοινότητας Κωφών. Οι νοηματικές γλώσσες χαρακτηρίζονται από κάποια ιδιαίτερα στοιχεία όπως παράλληλες ροές χαρακτηριστικών (π.χ. θέση-κίνηση των χεριών, είδος χειρομορφής κ.τ.λ) και την ύπαρξη ενός 3-διάστατου χώρου νοηματισμού.

Η μελέτη των νοηματικών γλωσσών υστερεί κατά πολύ των αντίστοιχων προφορικών γλωσσών. Στα πλαίσια αυτής παρουσιάζονται μείζονες σύγχρονες ερευνητικές προκλήσεις. Στον πυρήνα των προκλήσεων αυτών βρίσκονται κάποιες κυρίαρχες διαφορές κατά την νοηματική γλώσσα σε σχέση με τον προφορικό λόγο: κατά τον προφορικό λόγο συνεχόμενες καλώς ορισμένες μονάδες (φωνήματα) ήδη μελετημένες διεπιστημονικά συνδυάζονται γραμμικά προς σχηματισμό σύνθετων δομών. Σε αντίθεση στην νοηματική γλώσσα η αντίστοιχη μοντελοποίηση υστερεί κατά πολύ καθώς υπάρχουν διεπιστημονικά μη-κατασταλαγμένες ακόμη απόψεις για την μοντελοποίησή τους. Πρόσφατα έχουν δημοσιευθεί προτάσεις για την δεδομενο-κεντρική (data-driven) αυτόματη μοντελοποίηση σε επίπεδο υπο-μονάδων. Επιπρόσθετες προκλήσεις εμφανίζονται 1) λόγω του συνδυασμού πολλαπλών ροών πληροφορίας όπως κίνηση, χειρομορφή, κατεύθυνση και σημείο επαφής, 2) λόγω των δυνατοτήτων οι οποίες δημιουργούνται στον τρι-διάστατο χώρο νοηματισμού.

Στόχος της εργασίας αυτής είναι η αντιμετώπιση προβλημάτων αυτόματης δεδομενο-κεντρικής (data-driven) μοντελοποίησης σε επίπεδο υπο-μονάδων τα διαφορετικά είδη κίνησης-θέσης των χεριών ενός νοηματιστή κατά την διάρκεια εκτέλεσης ενός νοήματος (π.χ. ευθεία τροχιά με διαφορετικές κατευθύνσεις, κυκλική τροχιά κ.τ.λ) και η αναγνώριση τους, με απώτερο σκοπό την αναγνώριση νοηματικής γλώσσας.

Λέξεις κλειδιά:

Hidden Markov Models (HMMs), Regression Models, DTW, Segmental K-means

Ενδεικτική Βιβλιογραφία:

- [1] *Automatic sign language analysis: A survey and the future beyond lexical meaning*, [S. Ong](#) and [S. Ranganath](#) (2005), in: *IEEE Trans. Pattern Anal. Mach. Intellig.*, 27:6(873-891)
- [2] *Handshapes and Movements: Multiple-Channel American Sign Language Recognition*, [C. Vogler](#) and [D. Metaxas](#), in: *Gesture Workshop*, pages 247-258, 2003
- [3] *A novel approach to automatically extracting basic units from Chinese sign language*, [G. Fang](#), [X. Gao](#), [W. Gao](#) and [Y. Chen](#), 2004
- [4] *An Approach Based on Phonemes to Large Vocabulary Chinese Sign Language Recognition*, [C. Wang](#), [W. Gao](#) and [S. Shan](#), in: *FGR '02*
- [5] *Towards an Automatic Sign Language Recognition System Using Subunits*, [B. Bauer](#) and [K. -F. Kraiss](#), in: *Int'l Gesture Workshop*, pages 64-75, 2001
- [6] *Trajectory Clustering with Mixtures of Regression Models*, [S. Gaffney](#) and [P. Smyth](#), in: *KDD*, pages 63-72, 1999
- [7] *Joint Probabilistic Curve Clustering and Alignment* [S. Gaffney](#) and [P. Smyth](#), in: *NIPS*, 2004

## “Αυτόματη Αναγνώριση Συνεχούς Νοηματικής Γλώσσας : Παύση, Επένθεση και Μετάβαση”

Περιγραφή:

Οι νοηματικές γλώσσες αποτελούν ένα πλήρη κώδικα επικοινωνίας μεταξύ ατόμων με προβλήματα ακοής και την αποκλειστική γλώσσα επικοινωνίας της κοινότητας Κωφών. Οι νοηματικές γλώσσες χαρακτηρίζονται από κάποια ιδιαίτερα στοιχεία όπως οι παράλληλες ροές χαρακτηριστικών και η ύπαρξη ενός 3-διάστατου χώρου νοηματισμού.

Η μελέτη των νοηματικών γλωσσών υστερεί κατά πολύ των αντίστοιχων προφορικών γλωσσών. Στα πλαίσια αυτής παρουσιάζονται μείζονες σύγχρονες ερευνητικές προκλήσεις οι οποίες απέχουν αρκετά από το να έχουν ακόμη αντιμετωπιστεί. Μία από αυτές αποτελεί η μοντελοποίηση συνεχούς νοηματικής γλώσσας ως επόμενο στάδιο μετά την μοντελοποίηση απομονωμένων νοημάτων. Σε αυτά τα πλαίσια τα κυριότερα προβλήματα είναι η απουσία καλώς ορισμένων φωνημάτων όπως στον προφορικό λόγο και επιπλέον φαινόμενα μεταβάσεων, συν-άρθρωσης και επένθεσης. Τα φαινόμενα αυτά περιπλέκονται καθώς στις νοηματικές γλώσσες συνυπάρχουν παράλληλα πολλαπλές ροές πληροφορίας (η κίνηση, η χειρομορφή, η κατεύθυνση και το σημείο επαφής) οι οποίες αρθρώνονται παράλληλα κατά την παραγωγή των νοημάτων και γίνονται απευθείας αντιληπτές οπτικά.

Στόχος της εργασίας αυτής είναι η αντιμετώπιση προβλημάτων μοντελοποίησης φαινομένων κατά τις μεταβάσεις μεταξύ νοημάτων στην συνεχή νοηματική γλώσσα. Ένα αντίστοιχο παράδειγμα στην αναγνώριση φωνής είναι το λεγόμενο πρόβλημα *speech vs silence*. Η μοντελοποίηση αυτών των μεταβάσεων σε συνεργασία με την μοντελοποίηση των φωνημάτων θα έχει ως απότερω σκοπό την αναγνώριση συνεχούς νοηματικού λόγου.

Λέξεις κλειδιά:

Hidden Markov Models (HMMs), Segmentation, Segmental K-means

Ενδεικτική Βιβλιογραφία:

- [1] *Automatic sign language analysis: A survey and the future beyond lexical meaning*, [S. Ong](#) and [S. Ranganath](#) (2005), in: *Trans. Pattern Anal. Mach. Intellig.*, 27:6(873-891)
- [2] *Handshapes and Movements: Multiple-Channel American Sign Language Recognition*, [C. Vogler](#) and [D. Metaxas](#), in: *Gesture Workshop*, pages 247-258, 2003
- [3] *Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition*, [W. Gao](#), [G. Fang](#), [D Zhao](#) and [Y. Chen](#)

## “Εντοπισμός και παρακολούθηση χεριών σε ακολουθίες εικόνων νοηματικής γλώσσας”

Περιγραφή:

Οι νοηματικές γλώσσες αποτελούν ένα πλήρη κώδικα επικοινωνίας μεταξύ ατόμων με προβλήματα ακοής και την αποκλειστική γλώσσα επικοινωνίας της κοινότητας Κωφών. Οι νοηματικές γλώσσες χαρακτηρίζονται από κάποια ιδιαίτερα στοιχεία όπως οι παράλληλες ροές χαρακτηριστικών και η ύπαρξη ενός 3Δ χώρου νοηματισμού.

Η μελέτη των νοηματικών γλωσσών υστερεί κατά πολύ των αντίστοιχων προφορικών γλωσσών. Στα πλαίσια αυτής παρουσιάζονται μείζονες σύγχρονες ερευνητικές προκλήσεις. Ο εντοπισμός και η παρακολούθηση των χεριών του νοηματιστή αποτελούν συνήθως τα αρχικά στάδια ενός συστήματος αυτόματης αναγνώρισης νοηματικής γλώσσας και είναι ιδιαίτερα κρίσιμα για την αποτελεσματικότητα του όλου συστήματος. Ο κύριος στόχος των σταδίων αυτών είναι η εκτίμηση των τμημάτων κάθε 2Δ εικόνας της ακολουθίας που αντιστοιχούν σε κάθε χέρι. Μερικά από τα υποστάδια που θα μπορούσαν να βοηθήσουν στην επίτευξη του στόχου αυτού είναι για παράδειγμα τα ακόλουθα: 1) μοντελοποίηση του χρώματος του δέρματος του νοηματιστή, 2) χρήση προγενέστερης πληροφορίας σχετικά με το σχήμα των χεριών και 3) εκτίμηση της οπτικής ροής μεταξύ διαδοχικών εικόνων.

Ο εντοπισμός και παρακολούθηση των χεριών με έναν αξιόπιστο αλλά και υπολογιστικά αποδοτικό τρόπο αποτελούν ένα ανοικτό ερευνητικό πρόβλημα, λόγω διαφόρων δυσκολιών που σχετίζονται με 1) την γρήγορη κίνηση των χεριών κατά τον νοηματισμό, 2) τη γεωμετρική πολυπλοκότητα και μεταβολή του σχήματος των χεριών, 3) τη μεταβλητότητα του χρώματος του δέρματος από νοηματιστή σε νοηματιστή αλλά και από εικόνα σε εικόνα του ίδιου νοηματιστή λόγω σκίασης ή μεταβολών στη φωτεινότητα, 4) τις συχνές αμοιβαίες επικαλύψεις των χεριών (το δεξί χέρι, κατά την κίνησή του, μπορεί να “κρύψει” το αριστερό χέρι και το αντίστροφο) και 5) του μη ομοιόμορφου ή μη στατικού φόντου της ακολουθίας εικόνων.

Στόχος της εργασίας αυτής είναι η αντιμετώπιση προβλημάτων αυτόματου εντοπισμού και παρακολούθησης των χεριών.

Λέξεις κλειδιά: εντοπισμός (*localization*), παρακολούθηση (*tracking*), κατάτμηση (*segmentation*), χρώμα του δέρματος, προγενέστερη πληροφορία σχήματος (*prior shape information*), οπτική ροή (*optical flow*), επικάλυψη (*occlusion*), αφαίρεση φόντου (*background subtraction*).

Ενδεικτική Βιβλιογραφία:

[1] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss, “Recent developments in visual sign language recognition”, Springer Journal on Universal Access in the Information Society, 6(4):323–362, February 2008.

[2] X. Zabulis, H. Baltzakis, A.A. Argyros, "Vision-based Hand Gesture Recognition for Human-Computer Interaction", in "The Universal Access Handbook", Lawrence Erlbaum Associates, Inc. (LEA), Series on "Human Factors and Ergonomics", ISBN: 978-0-8058-6280-5, pp 34.1 – 34.30, Jun 2009.

[3] Y. Wu, T.S. Huang, “Vision-based gesture recognition: A review”, 3<sup>rd</sup> Gesture Workshop, 1999.

## “Μοντελοποίηση και αναγνώριση χειρομορφών στη νοηματική γλώσσα”

Περιγραφή:

Οι νοηματικές γλώσσες αποτελούν ένα πλήρη κώδικα επικοινωνίας μεταξύ ατόμων με προβλήματα ακοής και την αποκλειστική γλώσσα επικοινωνίας της κοινότητας Κωφών. Οι νοηματικές γλώσσες χαρακτηρίζονται από κάποια ιδιαίτερα στοιχεία όπως οι παράλληλες ροές χαρακτηριστικών και η ύπαρξη ενός 3Δ χώρου νοηματισμού.

Η μελέτη των νοηματικών γλωσσών υστερεί κατά πολύ των αντίστοιχων προφορικών γλωσσών. Στα πλαίσια αυτής παρουσιάζονται μείζονες σύγχρονες ερευνητικές προκλήσεις. Στον πυρήνα των προκλήσεων αυτών βρίσκονται κάποιες κυρίαρχες διαφορές κατά την νοηματική γλώσσα σε σχέση με τον προφορικό λόγο: κατά τις νοηματικές γλώσσες συνυπάρχουν παράλληλα πολλαπλές ροές πληροφορίας οι οποίες αρθρώνονται κατά την παραγωγή των νοημάτων και γίνονται απευθείας αντιληπτές οπτικά. Τέτοιες ροές πληροφορίας είναι η κίνηση, η χειρομορφή, η κατεύθυνση και το σημείο επαφής. Η μοντελοποίηση και αναγνώριση του σχήματος της χειρομορφής παρουσιάζει πληθώρα προκλήσεων καθώς υπάρχει εξαιρετικός πλούτος διαμορφώσεων χειρομορφών. Επιπλέον ενδιαφέροντα προβλήματα εμφανίζονται λόγω 1) των διακυμάνσεων της θέσης και του προσανατολισμού (πόζας) των χειρομορφών στον 3Δ χώρο, 2) των συχνών αμοιβαίων επικαλύψεων των χεριών και του κεφαλιού (π.χ. το δεξί χέρι, κατά την κίνησή του, μπορεί να “κρύψει” το αριστερό χέρι ή το κεφάλι) και 3) του μη ομοιόμορφου ή μη στατικού φόντου της ακολουθίας εικόνων.

Στόχος της εργασίας αυτής είναι η αντιμετώπιση προβλημάτων μοντελοποίησης διαμορφώσεων χειρομορφών με απώτερο σκοπό την αυτόματη ταξινόμησή τους.

Λέξεις κλειδιά: χειρομορφή (*handshape*), διαμόρφωση χειρομορφής (*handshape configuration*), εξαγωγή χαρακτηριστικών (*feature extraction*), πόζα χεριού (*hand pose*), επικάλυψη (*occlusion*), μη ομοιόμορφο/ μη στατικό φόντο (*non uniform/non static background*), ταξινόμηση (*classification*), αναγνώριση (*recognition*).

Ενδεικτικές Δημοσιεύσεις:

[1] U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss, “Recent developments in visual sign language recognition”, Springer Journal on Universal Access in the Information Society, 6(4):323–362, February 2008.

[2] X. Zabulis, H. Baltzakis, A.A. Argyros, "Vision-based Hand Gesture Recognition for Human-Computer Interaction", in "The Universal Access Handbook", Lawrence Erlbaum Associates, Inc. (LEA), Series on "Human Factors and Ergonomics", ISBN: 978-0-8058-6280-5, pp 34.1 – 34.30, Jun 2009.

[3] Y. Wu, T.S. Huang, “Vision-based gesture recognition: A review”, 3<sup>rd</sup> Gesture Workshop, 1999.

## **“Αλληλεπίδραση ανθρώπου-ρομπότ μέσω αναγνώρισης κινήσεων και χειρονομιών”**

### Περιγραφή:

Στα πλαίσια της αλληλεπίδρασης Ανθρώπου-Ρομπότ δημιουργούνται πλήθος ερευνητικών προκλήσεων όσον αφορά στην διαπροσωπεία επικοινωνίας. Μεταξύ των πιο φυσικών ως προς τον άνθρωπο τροπικοτήτων περιλαμβάνονται η φωνή, τα νεύματα, οι κινήσεις και οι χειρονομίες. Οι κινήσεις και οι χειρονομίες βρίσκονται στο επίκεντρο ερευνητικού ενδιαφέροντος καθώς είναι παρούσες σε πολλές περιπτώσεις επικοινωνίας, είτε συνοδεύοντας τον προφορικό λόγο οι οποίες είναι ένθετες ή παράλληλες με αυτόν (πχ. προσωδιακού χαρακτήρα, έμφασης, επεξήγησης, περιγραφής, χωροθετικού χαρακτήρα, παραθέτοντας επιπλέον πληροφορία, μηχανικής επανάληψης), είτε στην ακραία περίπτωση των Νοηματικών Γλωσσών στην οποία οι χειρονομίες αποκτούν κυρίαρχο ρόλο σε ένα ευρύτερο γλωσσικό οικοδόμημα δίνοντας στην γλώσσα πληρότητα και μια εξαιρετική εκφραστική δυναμική.

Στόχος είναι με έμπνευση από την νοηματική γλώσσα να οριστεί ένας περιορισμένος κώδικας επικοινωνίας με απώτερο σκοπό την επικοινωνία Ανθρώπου-Ρομπότ, χρησιμοποιώντας στοιχεία όπως την κίνηση των χεριών, τις σχηματιζόμενες χειρομορφές και την κατεύθυνση. Μια γρήγορη υλοποίηση η οποία δύναται να λειτουργεί σε πραγματικό χρόνο μπορεί να χρησιμοποιηθεί περαιτέρω για μια πιο λειτουργική και φυσική προς τον άνθρωπο διαπροσωπεία με το μηχανικό-ρομποτικό σύστημα.

Λέξεις κλειδιά: Human-Robot interaction-interfaces, gesture – action – (handshape) recognition