

Ε.Μ.Π. - Σχολή Ηλεκτρολόγων Μηχ/κών & Μηχ/κών Υπολογιστών

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ¹ 2013-2014

Καθηγητής Πέτρος Μαραγκός, E-mail: maragos@cs.ntua.gr

Εργαστήριο: <http://cvsp.cs.ntua.gr>

Οι ενδιαφερόμενοι/ες παρακαλούνται να αποστείλουν ηλεκτρονικά σε ένα Zip με το όνομά τους τα εξής:
i) την αναλυτική τους βαθμολογία από ΣΗΜΜΥ-ΕΜΠ, ii) ένα σύντομο βιογραφικό τους, και iii) τον αριθμό των οφειλομένων μαθημάτων αν βρίσκονται στο 4^ο ή ανώτερο έτος σπουδών.

Προαπαιτούμενα μαθήματα με επιτυχία και συνεπή παρακολούθηση:

- Για θέματα Α: ΨΕΣ, ή πιθανώς και Αναγνώριση Προτύπων (* = μπορεί να παρακολουθείται ταυτόχρονα με διπλωματική).
- Για θέματα Σ: ΨΕΣ, ή πιθανώς και Αναγνώριση Προτύπων (*).
- Για θέματα V: Οραση Υπολογιστών, ή πιθανώς και Αναγνώριση Προτύπων (*).
- Για θέματα AV: ΨΕΣ, Οραση Υπολογιστών, ή πιθανώς και Αναγνώριση Προτύπων (*).

Γενικά για να υπάρχει ποικιλία ερευνητικών θεμάτων, δίνονται περισσότερα θέματα από τις θέσεις που μπορούν να επιβλεφθούν εντός ενός έτους. Θα γίνει προσπάθεια να αντιστοιχιστούν θέματα και φοιτητές με όσο το δυνατόν καλύτερο ταίριασμα ενδιαφερόντων και ικανοτήτων.

Οι διπλωματικές που επισημειώνονται με το σύμβολο ① θα πραγματοποιηθούν υπό τη συνεπίβλεψη του Δρ. Νάσου Κατσαμάνη, συνεργάτη ερευνητή του Εργαστηρίου Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων (<http://cvsp.cs.ntua.gr/~nassos>).

Επεξεργασία Φωνής-Γλώσσας και Πολυμεσικών Σημάτων:

A1. *Perceptually Motivated Audio Processing Front-end with application in Audio Saliency Computation, Acoustic Event Detection, or Audio/Speech Recognition* ①

Ανάπτυξη βελτιωμένου αλγορίθμου για ανίχνευση ακουστικά σημαντικών σημείων που ελκύουν την προσοχή σε πολυμεσικά βίντεο και ανίχνευση/κατηγοριοποίηση αντίστοιχων ακουστικών γεγονότων χρησιμοποιώντας σχετικά κίνητρα από ανθρώπινη αντίληψη.

References:

- C. Kayser et al., "[Mechanisms for allocating auditory attention: an auditory saliency map](#)", Current Biology, 2005.
- V. D. Delmotte, "[Computation auditory saliency](#)", PhD Thesis, Georgia Institute of Technology, Nov. 2012.
- S. Shamma, "[On the role of space and time in auditory processing](#)", Trends in Cognitive Science, 2001.

A2. *Automatic Audio (Speech or Music) or Text Summarization*

Αυτόματη συνόψιση ηχητικών ή γλωσσικών δεδομένων. (Μπορεί να χρησιμοποιηθούν και ιδέες από A1.)

References:

- A. Zlatintsi, P. Maragos A. Potamianos and G. Evangelopoulos, "[A Saliency-Based Approach to Audio Event Detection and Summarization](#)", Proc. EUSIPCO-2012, Bucharest, Romania, Aug. 2012.
- G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, "[Multimodal Saliency and Fusion for Movie Summarization based on Aural, Visual, and Textual Attention](#)", *IEEE Trans. Multimedia*, Nov. 2013.
- K. Spärck Jones "[Automatic summarising: The state of the art](#)", Information Processing & Management, 2007.

A3. *Deep Neural Networks for Speech Recognition* ①

Αναγνώριση ομιλίας με τη χρήση της τεχνολογίας των βαθιών νευρωνικών δικτύων (τεχνολογία που χρησιμοποιείται από τις Google, Microsoft στα συστήματα αναγνώρισης ομιλίας που αναπτύσσουν).

References:

¹ Αρκετά από τα ανωτέρω θέματα έχουν προοπτική για Διδακτορικό με οικονομική υποστήριξη από ερευνητικά προγράμματα.

- G. Hinton et al., "[Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups](#)", IEEE Signal Processing Magazine, Nov. 2012.
- <http://deeplearning.net>

A4. Microphone-Array Speech and Audio Processing: Theory, Applications

References:

- M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer 2001.
- J. Benesty, J. Chen and Y. Huang, *Microphone Array Signal Processing*, Springer 2008.
- P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer 2010.

Συστήματα και Αυτόματα (με εφαρμογές σε λόγο ή όραση):

S1. Μη-γραμμικά δυναμικά συστήματα που χρησιμοποιούν max-plus άλγεβρα και finite-state automata με εφαρμογή σε language processing and recognition.

References (επίσης και αυτά του S2):

- Hori, Takaaki, and Atsushi Nakamura. *Speech Recognition Algorithms Using Weighted Finite-State Transducers*. Morgan & Claypool, 2013.

S2. Μη-γραμμικά δυναμικά συστήματα (που χρησιμοποιούν max-plus άλγεβρα και finite-state automata, και πιθανώς Bayesian inference) με εφαρμογή σε προβλήματα ελέγχου οπτικής προσοχής και αντιστοίχισης μεταξύ ανίχνευσης (με μεθόδους όρασης υπολογιστών) οπτικά σημαντικών γεγονότων σε βίντεο και σημασιολογίας (semantics).

References (για S2 και S1):

- R. Cuninghame-Green, *Minimax Algebra*, Springer-Verlag, New York, 1979.
- http://cvsp.cs.ntua.gr/publications/jpubl+bchap/Maragos_LatImProcUnifMorfFuzAlgSyst_JMIV2005.pdf
- http://cvsp.cs.ntua.gr/publications/jpubl+bchap/MaragosTzafestas_MinmaxControlApplicDEDS_bchap1999.pdf
- D. Jurafsky and J. H. Martin, *Speech and Language Processing*, Prentice-Hall, 2000.
- M. Jordan, *Graphical Models*, Lecture Notes, Stanford University.
- G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, [Multimodal Saliency and Fusion for Movie Summarization based on Aural, Visual, and Textual Attention](#), *IEEE Trans. Multimedia*, Nov. 2013.

Όραση Υπολογιστών:

V1. Visual Saliency Computation and Eye-Gaze Tracking

Ανάπτυξη βελτιωμένου αλγορίθμου για ανίχνευση οπτικά σημαντικών σημείων που ελκύουν την προσοχή σε εικόνα/βίντεο και πειραματισμός με σύστημα για παρακολούθηση βλέμματος.

References:

- L. Itti and C. Koch. "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, 40:1489–1506, 2000.
- http://cvsp.cs.ntua.gr/publications/confr/GkioulekasEvangMaragos_SpatialBayesianSurpriseImageSaliency_ICIP10.pdf
- G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, [Multimodal Saliency and Fusion for Movie Summarization based on Aural, Visual, and Textual Attention](#), *IEEE Trans. Multimedia*, Nov. 2013.

V2. Face Detection/Recognition in Videos and Semantic Analysis

Of particular interest for the semantic analysis of videos (images + text) is the task of person-related analysis, which involves face detection and recognition by using the visual information stream and possibly textual information. The challenge is to recover this information automatically and 'in the wild' while dealing with global and local variation.

References:

- G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. "Labeled faces in the wild: A database for studying face recognition in unconstrained environments". Proc. ECCV-2008 Workshop on Faces in Real-Life Images.
- M. Everingham, J. Sivic, and A. Zisserman "Hello! My name is... Buffy" - Automatic naming of characters in TV video". Proc. BMVC, 2006.
- T. Cour, B. Sapp, A. Nagle, B. Taskar, "Talking Pictures: Temporal Grouping and Dialog-Supervised Person Recognition". Proc. CVPR, 2010.

V3. Perceptually-motivated Αναγνώριση Σχήματος με Εφαρμογές σε Gesture Recognition για επικοινωνία Ανθρώπου-Ρομπότ ή σε Αναγνώριση Νοηματικής Γλώσσας

References:

- U. Agris, J. Zieren, U. Canzler, B. Bauer, and K. F. Kraiss, "Recent developments in visual sign language recognition," *Universal Access in the Information Society*, vol. 6, pp. 323–362, 2008.
- A. Roussos, S. Theodorakis, V. Pitsikalis and P. Maragos, [Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos](#), *Journal of Machine Learning Research*, vol. 14, pp.1513-1539, 2013.
- de La Gorce, M., Fleet, D.J. and Paragios, N., "Model-based 3D hand pose estimation from monocular video", *IEEE Trans. Pattern Analysis and Machine Intelligence*,33(9):1793-1805, 2011. <http://www.cs.toronto.edu/~fleet/research/Papers/handTrackerPamiVideo.avi>

V4. Graph-based Models for Image Segmentation and Object Detection

References:

- Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts", *IEEE Trans. PAMI*, 23(11):1222–1239, 2001.
- Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, Sep. 2004.
- V.-T. Ta, A. Elmoataz, and O. Lezoray, "Nonlocal PDEs-Based Morphology on Weighted Graphs for Image and Data Processing", *IEEE Trans. Image Processing*, 20(6):1504–1516, June 2011.
- K. Drakopoulos and P. Maragos, [Active Contours on Graphs: Multiscale Morphology and Graphcuts](#), *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 7, pp.780-794, Nov. 2012.

V5. Clustering and Segmentation on Graphs with application in Social or other Networks

References:

- J. Kleinberg, [Networks, Crowds, and Markets: Reasoning About a Highly Connected World](#),

V6. Three-dimensional Shape: Analysis, Modeling, Matching

References:

- A. Bronstein, M. Bronstein, and R. Kimmel. [Numerical geometry of non-rigid shapes](#). Springer, 2008.
- K. Siddiqi and S. Pizer (eds.) [Medial Representations – Mathematics, Algorithms and Applications](#), Springer, 2008.
- M. Breuss, A. Bruckstein and P. Maragos (Eds.), [Innovations for Shape Analysis: Models and Algorithms](#), Springer, 2013.

V7. Web-based Photo-tour Application

The development of image-based modeling and rendering techniques has been an important recent trend in computer vision and computer graphics, using multiple viewpoint photographic shots for the automated 3-D reconstruction and photorealistic presentation of scenes. In the photo-tour application the goal of the thesis is to organize a large volume of photographs that we will take from an archeological site in a way that will convey to the user accessing the application on the web a convincing sense of the spatial relationship and interaction of buildings, wall-paintings, and other findings, thus providing a satisfying remote sightseeing experience.

References:

- N. Snavely, S. Seitz, and R. Szeliski, "Photo Tourism: Exploring photo collections in 3D," in *Proc. SIGGRAPH*, 2006.
- R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer 2010. <http://szeliski.org/Book/>
- O. Faugeras and Q.T. Luong, *Geometry of Multiple Images*, MIT Press, 2001.
- R. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge Univ. Press, 2004 <http://www.robots.ox.ac.uk/~vgg/hzbook.html>

V8. Κατάτμηση/μοντελοποίηση της φωνητικής οδού σε οπτικοακουστικά βίντεο μαγνητικής τομογραφίας με 3D Active Appearance Models ①

References:

- Mitchell et al., "3D Active Appearance Models: Segmentation of Cardiac MR and Ultrasound Images", *IEEE Trans. Medical Imaging*, Sep. 2002.
- A. Roussos, A. Katsamanis, P. Maragos, "[Tongue tracking in ultrasound images with active appearance models](#)", in Proc. IEEE ICIP 2009.

V9. Ανάλυση Οπτικής Κίνησης και Κατάτμηση Βίντεο με Point Clouds και Παρακολούθηση Τροχιών

References:

- T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories", ECCV 2010.
- M. Raptis, I. Kokkinos and S. Soatto, "Discovering Discriminative Action Parts from Mid-Level Video Representations", CVPR 2012.

Ανάλυση Οπτικο-Ακουστικών Σημάτων, Ανίχνευση, Αναγνώριση, Σύνθεση:

AV1. Συνδυασμός Πληροφορίας Πολύ-ασθητηριακών και Πολυμεσικών Σημάτων: Οπτικο-Ακουστική Συσχέτιση, Ανίχνευση Γεγονότων, Εντοπισμός Ομιλητών/Προσώπων.

Θέμα AV1-A. Εντοπισμός Ομιλητών με χρήση Συστοιχίας Μικροφώνων και Kinect Καμερών

References:

- Special Session on Computer Vision Applications of Kinect, EUSIPCO-2012.

Θέμα AV1-B. Ανάλυση Κινήσεων Σώματος με αισθητήρες Motion-Capture και Εφαρμογές στην μοντελοποίηση γλώσσας σώματος σε στάση, βάδισμα, χορό, αθλητισμό ①

References:

- A. Metallinou, A. Katsamanis, S. Narayanan, "[Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information](#)", Image and Vision Computing, Feb. 2013.
- <http://viterbi.usc.edu/news/news/2010/quantifying-human-behavior.htm>

Θέμα AV1-Γ. Ανίχνευση Γεγονότων ή Προσώπων σε πολυμεσικά βίντεο με συσχέτισμό οπτικής, ακουστικής και γλωσσικής πληροφορίας ①

References:

- I. Laptev, M. Marszalek, C. Schmid and B. Rozenfeld, "Learning realistic human actions from movies", Proc. CVPR 2008.
- T. Cour, B. Sapp, A. Nagle, B. Taskar, "Talking Pictures: Temporal Grouping and Dialog-Supervised Person Recognition". Proc. CVPR, 2010.

AV2. Audio-Visual Fusion using Canonical-Correlation Analysis or other Machine Learning or Cognitive Methods

References:

- J. Hershey and J. Monevallon, "Audio-Vision: Using Audio-Visual Synchrony to Locate Sounds".
- E. Kidron, Y. Y. Schechner and M. Elad, "Pixels that Sound", Proc. CVPR 2005.
- Monaci G., Escoda, O. D. and Vandergheynst P, "Analysis of Multimodal Sequences Using Geometric Video Representations", *Signal Processing*, vol. 86, May 2006.
- Sargin M. E., Yemez Y., Erzin E., and Tekalp A. M., "Audiovisual Synchronization and Fusion Using Canonical Correlation Analysis", *IEEE Trans. On Multimedia*, Vol.9, No. 7, Nov. 2007.
- M.Slaney and M. Covell, "FaceSync: A Linear Operator for Measuring Synchronization of Video Facial Images and Audio Tracks", *Proc. of Neural Information Processing Society 13*, MIT Press, 2001.
- M. T. Wallace, G. E. Roberson, W. D. Hairston, B. E. Stein, J. W. Vaughan and J. A. Schirillo, "Unifying Multisensory Signals Across Time and Space", *J. of Exp. Brain Res.*, Vol. 158, 2004.

AV3. Affective Audiovisual Speech Synthesis using Active Appearance and Hidden Markov Models ①

Δημιουργία φωτο-ρεαλιστικού τεχνητού ομιλητή με συναισθήματα.

References:

- R. Anderson et al., "Expressive visual text-to-speech using active appearance models", Proc. CVPR 2013.
- <http://www.cam.ac.uk/research/news/face-of-the-future-rears-its-head>

AV4. Κατάτμηση/μοντελοποίηση της φωνητικής οδού σε οπτικοακουστικά βίντεο μαγνητικής τομογραφίας με αξιοποίηση και της ακουστικής πληροφορίας ①

References:

- A. Roussos, A. Katsamanis, P. Maragos, "[Tongue tracking in ultrasound images with active appearance models](#)", in Proc. IEEE ICIP 2009.
- A. Katsamanis, E. Bresch, V. Ramanarayanan, and S. Narayanan. "[Validating rt-MRI based articulatory representations via articulatory recognition](#)", in Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- S. Narayanan et al., "[A multimodal real-time mri articulatory corpus for speech research](#)", in Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- <http://sail.usc.edu/span/videos/thesearethemovts.mov>

AV5. Επεξεργασία EEG σημάτων και συσχέτισμός με συναισθηματική κατάσταση κατά την παρακολούθηση οπτικο-ακουστικών βίντεο ①

References:

Τα σήματα θα καταγράφονται από το EEG device: Emotiv EPOC Model 1.0 EEG Neuroheadset (high resolution multi-channel portable EEG, 14 biopotential sensors).