

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

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ¹ 2020-21

Καθηγητής Πέτρος Μαραγκός, E-mail: maragos@cs.ntua.gr
Εργαστήριο: <http://cvsp.cs.ntua.gr>, <http://robotics.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 με μια εκτίμηση του τρέχοντος μέσου όρου, ii) το βιογραφικό τους, και iii) τον αριθμό των οφειλομένων μαθημάτων αν βρίσκονται στο 4^ο ή ανώτερο έτος σπουδών. Η αποστολή της πληροφορίας αυτής να γίνεται στην Γραμματέα του Εργαστηρίου κα. Βίκυ Πλατίτσα (email: <vickyplatitsa@gmail.com>).

Προαπαιτούμενα μαθήματα (Ροής Σ) με επιτυχία και συνεπή παρακολούθηση και πολύ καλή απόδοση:

- Για θέματα N: ΨΕΣ και Αναγνώριση Προτύπων-ΑΠ (*).
- Για θέματα V: Οραση Υπολογιστών (ΟΥ), ή πιθανώς και ΑΠ (*).
- Για θέματα R που συνδυάζουν Ρομποτική με CV/SP/ML: ΟΥ και Ρομποτική Ι ή ΑΠ (*).
- Για θέματα AVM: ΨΕΣ, ΟΥ και ΑΠ (*).
- Για θέματα L: ΨΕΣ, ΟΥ και Αναγνώριση Προτύπων (*).
- Για θέματα S: ΨΕΣ και κάποιο μάθημα ΣΑΕ απο Ροη Σ, ή πιθανώς και Αναγνώριση Προτύπων (*).
- Για θέματα T: ΟΥ και ΑΠ (*).

* = μπορεί να παρακολουθείται ταυτόχρονα με διπλωματική.

- Γενικά, για να υπάρχει ποικιλία ερευνητικών θεμάτων, δίνονται περισσότερα θέματα από τις θέσεις που μπορούν να επιβλεφθούν εντός ενός έτους.
- Μετά την εκδήλωση ενδιαφέροντος οι ενδιαφερόμενοι φοιτητές θα συναντηθούν μαζί με τον υπεύθυνο καθ. Π. Μαραγκό και Επιστημονικούς Συνεργάτες καθώς και Υποψήφιους Διδάκτορες της ερευνητικής ομάδας και θα γίνει προσπάθεια να αντιστοιχισθούν θέματα και φοιτητές με όσο το δυνατόν καλύτερο ταίριασμα ενδιαφερόντων και ικανοτήτων.
- Η τελική αποδοχή αίτησης για εκπόνηση διπλωματικής θα εξαρτηθεί από την επίδοση στα σχετικά μαθήματα του εργαστηρίου, τον αριθμό των οφειλομένων μαθημάτων, και την χρονική διαθεσιμότητα του θέματος.

Δίκτυα Αισθητήρων & Big Data Processing: Εκμάθηση, Ανίχνευση, Αναγνώριση

N1. Καταγραφή & ανάλυση βιομετρικών δεικτών για υποστήριξη ασθενών σε θέματα ψυχικής υγείας

Η τεχνολογική πρόοδος που έχει σημειωθεί τα τελευταία χρόνια στις φορητές συσκευές (κινητά τηλέφωνα, smart-watch, κτλ.) έχει πλέον επιτρέψει την αξιόπιστη καταγραφή πληθώρας βιομετρικών δεικτών, οι οποίοι ήδη χρησιμοποιούνται στον αθλητισμό και την ιατρική. Με τη διπλωματική αυτή στοχεύουμε στην ανάπτυξη ενός ευφυούς συστήματος επεξεργασίας δεδομένων (big data) και ειδικών αλγορίθμων για τη μηχανική μάθηση μετρήσεων βιομετρικών δεικτών, τα οποία θα καταγραφούν από πραγματικούς χρήστες μέσω φορητών αισθητήρων, για τον εντοπισμό μεταβολών και κατάσεων και την έγκαιρη λήψη αποφάσεων βάσει στατιστικών εκτιμήσεων από την επεξεργασία τους.

(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση, ΥΔ Παναγιώτης Φιλντίσης ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

References:

- [1] I. Maglogiannis, A. Zlatintsi, A. Menychtas, D. Papadimitos, P.P. Filntisis, N. Efthymiou, G. Retsinas, P. Tsanakas, and P. Maragos, [An intelligent cloud-based platform for effective monitoring of patients with psychotic disorders](#), in Proc. Int'l Conf. on Artificial Intelligence Applications and Innovation (AIAI-2020), Halkidiki, Greece, 5-7 June, 2020.
- [2] G. Retsinas, P.P. Filntisis, N. Efthymiou, E. Theodosis, A. Zlatintsi and P. Maragos, [Person Identification Using Deep Convolutional Neural Networks On Short-term Signals From Wearable Sensors](#), in Proc. 45th IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020), May 2020.
- [3] P. P. Filntisis, A. Zlatintsi, N. Efthymiou, E. Kalisperakis, T. Karantinos, M. Lazaridi, N. Smyrnis and P. Maragos,

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

Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors, Submitted to IEEE Int'l Conf. on Acoustics, Speech and Signal Processing, (ICASSP-2021).

[4] D. A. Adler et al., [Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks](#), JMIR mHealth and uHealth, Vol. 8, iss. 8, 2020.

[5] D. Bone, C.-C. Lee, T. Chaspari, J. Gibson, and S. Narayanan, [Signal Processing and Machine Learning for Mental Health Research and Clinical Applications](#). Perspectives, *IEEE Signal Processing Magazine*, Sep. 2017.

Επικοινωνία Ανθρώπου-Ρομπότ και Αυτόνομη Πλοήγηση: Πολυτροπικές Ευφυείς Διεπαφές, Ανάλυση Οπτικο-Ακουστικών Σημάτων, Ανίχνευση, Αναγνώριση, Πλοήγηση:

R1. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ σε πραγματικό χρόνο:

Στόχος είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών, η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων, όπως RGB-D ή/και ανθρώπινη πόζα.

Εφαρμογές σε assistive & social robotics.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης, ΥΔ Νίκη Ευθυμίου)

References:

[1] I. Rodomagoulakis, N. Kardaris, V. Pitsikalis, E. Mavroudi, A. Katsamanis, A. Tsiami and P. Maragos, "[Multimodal Human Action Recognition in Assistive Human-Robot Interaction](#)", Proc. ICASSP-2016, Shanghai, China, Mar. 2016.

[2] A. Zlatintsi, I. Rodomagoulakis, V. Pitsikalis, P. Koutras, N. Kardaris, X. Papageorgiou, C. Tzafestas and P. Maragos, "[Social Human-Robot Interaction for the Elderly: Two Real-life Use Cases](#)", Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI-2017), Vienna, Austria, March 2017.

[3] A. Tsiami, P. Koutras, N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, "[Multi3: Multi-sensory Perception System for Multi-modal Child Interaction with Multiple Robots](#)", Proc. IEEE Int'l Conference on Robotics and Automation (ICRA-2018), Brisbane, Australia, May 2018.

[4] J. Carreira, & A. Zisserman, "[Quo vadis, action recognition? a new model and the kinetics dataset](#)". Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017), Honolulu, Hawaii, July 2017

[5] M. Xu, M. Gao, Y. Chen, L. S. Davis, D. J. Crandall, "[Temporal recurrent networks for online action detection](#)", Proc. IEEE Int'l Conference on Computer Vision (ICCV-2019), Seoul, Korea, Oct.-Nov. 2019.

[6] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, Ahmed A. A. Osman, D. Tzionas, and M. J. Black, "[Expressive Body Capture: 3D Hands, Face, and Body from a Single Image](#)", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2019), Long Beach, CA, United States, June 2019.

[7] G. Chalvatzaki, P. Koutras, A. Tsiami, C. S. Tzafestas and P. Maragos, "[i-Walk Intelligent Assessment System: Activity, Mobility, Intention, Communication](#)", Proc. 16th European Computer Vision Conference Workshops (ECCVW) – 8th Int'l Workshop on Assistive Computer Vision and Robotics (ACVR-2020), Aug. 2020.

R2. Visual SLAM systems with geometric and/or semantic enhancements:

Τα συστήματα vSLAM[1,2] επιτυγχάνουν ταυτόχρονα την 3D ανακατασκευή του άγνωστου χώρου στον οποίο κινούνται, καθώς και την εκτίμηση της τροχιάς που διαγράφουν. Το vSLAM αποτελεί τη real-time εκδοχή του κλασικού προβλήματος της ΟΥ, Structure from Motion (SfM), ενώ έχει επίσης μελετηθεί παλιότερα υπό τη μορφή Time-of-Flight (ToF) μεθόδων (lasers ή sonars) και στο άλλο κλασικό πρόβλημα της Ρομποτικής, το SLAM. Πλέον, τα παραπάνω ενοποιούνται σε ένα σύστημα, με χρήση mono, stereo ή RGBD καμερών, καθώς και είναι δυνατό να συνδυαστούν με πληροφορία από LIDARs ή IMUs[6] οδηγώντας σε παραλλαγές συστημάτων vSLAM. Έμφαση δίνεται στην αξιοποίηση συνθετότερων γεωμετρικών δομών της εικόνας (ευθείες, επίπεδα, κωνικές)[3,4,5] ή και σημασιολογικής πληροφορίας[4], με σκοπό την αύξηση της ευρωστίας, της ακρίβειας, του εύρους των περιβάλλοντων εφαρμογής, ή τη μείωση της υπολογιστικής πολυπλοκότητας.

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας)

References:

[1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "[ORB-SLAM: A Versatile and Accurate Monocular SLAM System](#)", *IEEE Transactions on Robotics*, 2015.

- [2] D. Schlegel, M. Colosi and G. Grisetti, “ProSLAM: Graph SLAM from a Programmer's Perspective”, *IEEE International Conference on Robotics and Automation*, 2018.
- [3] F. Nardi, B. D. Corte and G. Grisetti, “Unified Representation and Registration of Heterogeneous Sets of Geometric Primitives”, *IEEE Robotics and Automation Letters*, 2019.
- [4] L. Nicholson, M. Milford and N. Sünderhauf, “QuadricSLAM: Dual Quadrics From Object Detections as Landmarks in Object-Oriented SLAM”, *IEEE Robotics and Automation Letters*, 2019.
- [5] M. Hosseinzadeh, Y. Latif, T. Pham, N. Sünderhauf and I. Reid, “Structure Aware SLAM Using Quadrics and Planes”, *ACCV*, 2018.
- [6] A. Rosinol, M. Abate, Y. Chang and L. Carlone, “Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping”, *IEEE International Conference on Robotics and Automation*, 2020.

Οραση Υπολογιστών:

V1. Αναγνώριση συναισθηματικής κατάστασης μέσω της ανθρώπινης πόζας, προσώπου και περιβάλλοντος χώρου, με χρήση βαθέων νευρωνικών δικτύων.

(Συνεργάτες: ΥΔ Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] P. P. Filtntisis, N. Efthymiou, P. Koutras, G. Potamianos and P. Maragos, "[Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child-Robot Interaction](#)," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4011-4018, Oct. 2019. DOI: [10.1109/LRA.2019.2930434](https://doi.org/10.1109/LRA.2019.2930434) .
- [2] R. Kostic, Alvarez J., Adria Recasens, and Agata Lapedriza. "[Emotion Recognition in Context](#)." In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1667-1675, 2017.
- [3] Y. Luo, J., R. Adams, J. Li, M. Newman & J. Z. Wang. [ARBEE: Towards automated recognition of bodily expression of emotion in the wild](#). *International Journal of Computer Vision*, **128**, 1-25 (2020). <https://doi.org/10.1007/s11263-019-01215-y>
- [4] H. Gunes, & M. Pantic, “Automatic, dimensional and continuous emotion recognition”, *ACM International Journal of Synthetic Emotions*, 1(1), 68-99, 2010. <https://dl.acm.org/doi/10.4018/jse.2010101605> .
- [5] F. Noroozi, D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera and G. Anbarjafari, "[Survey on Emotional Body Gesture Recognition](#)," *IEEE Transactions on Affective Computing*, 2018. DOI: [10.1109/TAFFC.2018.2874986](https://doi.org/10.1109/TAFFC.2018.2874986) .
- [6] P. P. Filtntisis, N. Efthymiou, G. Potamianos, P. Maragos, “[Emotion Understanding in Videos Through Body, Context, and Visual-Semantic Embedding Loss](#)”, *Proc. ECCV Workshop*, 2020.

V2. Three-dimensional Shape: Analysis, Modeling, Matching

References:

- [1] A. Bronstein, M. Bronstein, and R. Kimmel. [Numerical geometry of non-rigid shapes](#). Springer, 2008.
- [2] M. Breuss, A. Bruckstein and P. Maragos (Eds.), [Innovations for Shape Analysis: Models and Algorithms](#), Springer, 2013.
- [3] J. Sun, M. Ovsjanikov, L. Guibas, “[A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion](#)”, *Eurographics Symposium on Geometry Processing* 2009.

V3. Graph-theoretic Methods for Clustering and Segmentation

References:

- [1] C. G. Bampis, P. Maragos and A. C. Bovik, “[Graph-Driven Diffusion and Random Walk Schemes for Image Segmentation](#),” *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 35-50, Jan. 2017.
- [2] C. Sakaridis, K. Drakopoulos and P. Maragos, “[Theoretical Analysis of Active Contours on Graphs](#),” *SIAM J. Imaging Sciences*, vol.10, no. 3, pp. 1475-1510, 2017.

V4. Sign-Language Recognition using Deep Learning: Αναγνώριση χειρονομιών και συνεχών βίντεο νοηματικής γλώσσας με την χρήση βαθειών νευρωνικών δικτύων.

References:

- [1] S. Theodorakis, V. Pitsikalis and P. Maragos, “[Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition](#)”, *Image and Vision Computing*, vol.32, no.8, pp.533-549, Aug. 2014.

- [2] 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. 1627-1663, June 2013.
- [3] O. Koller, N.C. Camgoz, H. Ney, and R. Bowden, “[Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos](#),” *IEEE Trans. PAMI*, 2020. DOI: [10.1109/TPAMI.2019.2911077](#)
- [4] M. Parelli, K. Papadimitriou, G. Potamianos, G. Pavlakos and P. Maragos, “[Exploiting 3D Hand Pose Estimation in Deep Learning-based Sign Language Recognition from RGB Videos](#)”, *Proc. ECCV Workshop on Sign Language Recognition, Translation and Production (SLRTP 2020)*, Aug. 2020.
- [5] A. Kratimenos, G. Pavlakos and P. Maragos, “Independent Sign Language Recognition With 3d Body, Hands, And Face Reconstruction”, submitted to ICASSP-2021. Also [arXiv:2012.05698](#).
- [6] C. C. de Amorim, D. Macêdo and C. Zanchettin, “[Spatial-Temporal Graph Convolutional Networks for Sign Language Recognition](#)”, <https://arxiv.org/abs/1901.11164>, 2019.
- [7] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, <http://www.deeplearningbook.org>
- [8] [IEEE-EURASIP Summer School on Signal Processing Meets Deep Learning](#), Capri, Italy, Sep. 2017.

V5. Shape from Shading. 3D Shape reconstruction from polarized images of unknown lighting and albedo:

References:

- [1] W. A.P. Smith, R. Ramamoorthi, S. Tozza, “Height-from-Polarisation with Unknown Lighting or Albedo”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [2] S. Tozza, W. A.P. Smith, D. Zhu, R. Ramamoorthi, E. R. Hancock, “[Linear Differential Constraints for Photo-polarimetric Height Estimation](#)”, *Proc. ICCV* 2017.
- [3] S. Tozza, M. Falcone, “[Analysis and approximation of some Shape-from-Shading models for non-Lambertian surfaces](#)”, *Journal of Mathematical Imaging and Vision*, 55(2): 153-178, 2016.
- [4] R. Or – El. G. Rosman, A. Wetzler, R. Kimmel, A. Bruckstein, “[RGBD-Fusion: Real-Time High Precision Depth Recovery](#)”, *Proc. CVPR* 2015.

V6. 3D Computer Vision for Faces & Hands and Other Deformable Objects

(Συνεργάτες: Δρ. Αναστάσιος Ρούσσοσ, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ΙΤΕ, και ΥΔ Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] J. Booth, A. Roussos, ..., S. Zafeiriou, “[Large Scale 3D Morphable Face Models](#)”, *IJCV* 2017.
- [2] J. Booth, A. Roussos et al, “3D reconstruction of in-the-Wild Faces in Images and Videos”, *PAMI* 2018.
- [3] M.R. Koujan, ..., A. Roussos, “[Real-time Facial Expression Recognition “In The Wild” by Disentangling 3D Expression from Identity](#)”, *Proc. Int’l Conf. Automatic Face and Gesture Recognition (FG 2020)*.
- [4] S. Ploumpis et al, “Towards a complete 3D morphable model of the human head”, <https://arxiv.org/abs/1911.08008> (CVPR and PAMI reprint).

Επεξεργασία Ακουστικών σημάτων (π.χ. μουσικής, φωνής) ή Οπτικών σημάτων σε Πολυμεσικά περιβάλλοντα:

AVM1. Υπολογιστική επεξεργασία μουσικών σημάτων και γενικότερα μουσικής πληροφορίας με εφαρμογές όπως:

Περίληψη μουσικής (highlight extraction) [1 - 3].

Ανάπτυξη εργαλείων σύνθεσης μουσικής βάσει επιθυμητών χαρακτηριστικών (χαμηλού και υψηλού επιπέδου) [4 - 6].

Ανάπτυξη συστήματος εξατομικευμένων μουσικών προτάσεων (music recommendation) [7-9] βασισμένο σε μουσική ομοιότητα (music similarity) [10, 11].

Αυτόματη αναγνώριση και εντοπισμός μουσικών χαρακτηριστικών υψηλού επιπέδου, σχετικών με το μουσικό η σημασιολογικό περιεχόμενο (auto-tagging) [12 - 15].

Ευθυγράμμιση μεταξύ του μουσικού σήματος και του περιεχομένου του (audio/lyrics - audio/transcript alignment) [16, 17].

Εξαγωγή της κυρίαρχης μελωδίας από πολυφωνικά μουσικά σήματα (melody extraction) [18, 19].

(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση και ΥΔ Χρήστος Γαρούφης, ΕΜΠ)

References:

Music Summarization / Highlight extraction

- [1] A. Zlatintsi, E. Iosif, P. Maragos and A. Potamianos, [Audio Salient Event Detection and Summarization using Audio and Text Modalities](#), Proc. European Conf. EUSIPCO-2015, Nice, France, 2015.
- [2] 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.
- [3] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos and P. Maragos, ["COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization"](#), *EURASIP Journal on Image and Video Processing* (2017) 2017:54.

Automatic Music Synthesis

- [4] C. Garoufis, A. Zlatintsi, and P. Maragos, [An LSTM-Based Dynamic Chord Progression Generation System for Interactive Music Performance](#), in Proc. 45th Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP-20), 2020.
- [5] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, [Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment](#). In Proc. AAAI Conf. on Artificial Intelligence, New Orleans, LA, USA, 2018.
- [6] G. Hadjeres, F. Pachet, and F. Nielsen. [Deepbach: a steerable model for bach chorales generation](#). In Proc. Int'l Conf. on Machine Learning, Sydney, Australia, 2017

Music Recommendation / Music Similarity

- [7] Lee et al., [Can we Listen to it Together? Factors Influencing Reception of Recommendations and Post-Recommendation Behavior](#), ISMIR-19
- [8] Zangerle et al., [Hit Song Prediction: Leveraging Low And High-Level Audio Features](#), ISMIR-19
- [9] Behrooz et al., [Augmenting Music Listening Experience on Voice Assistants](#), ISMIR-19
- [10] B. McFee, L. Barrington, and G. R. G. Lanckriet, ["Learning Content Similarity for Music Recommendation"](#), *IEEE Transactions on Audio, Speech & Language Processing* 20, no.8 (2012): 2207-2218
- [11] Prétet, L., G. Richard, and G. Peeters. ["Learning to Rank Music Tracks using Triplet Loss."](#) In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

Auto-tagging

- [12] A. Kratimenos, K. Avramidis, C. Garoufis, A. Zlatintsi, and P. Maragos, [Augmentation Methods On Monophonic Audio For Instrument Classification In Polyphonic Music](#), to appear in Proc. 28th European Signal Proc. Conf., (EUSIPCO-20).
- [13] Choi, Keunwoo, György Fazekas, Mark Sandler, and Kyunghyun Cho. ["Convolutional Recurrent Neural Networks for Music Classification."](#) In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, Louisiana, USA, 2017.
- [14] Kim, T., Lee, J., & Nam, J. ["Sample-level cnn architectures for music auto-tagging using raw waveforms."](#) In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AL, Canada, 2018.
- [15] Won, Minz, Sanghyuk Chun, Oriol Nieto, and Xavier Serra. ["Data-driven Harmonic Filters for Audio Representation Learning."](#) In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

Audio-lyrics / audio-transcript alignment

- [16] Stoller, Daniel, Simon Durand, and Sebastian Ewert. ["End-to-end lyrics alignment for polyphonic music using an audio-to-character recognition model."](#) *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019.
- [17] Mesaros, Annamaria, and Tuomas Virtanen. ["Automatic alignment of music audio and lyrics."](#) *Proceedings of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*. 2008.

Melody Extraction - f0 Estimation

- [18] Bittner, Rachel M., Brian McFee, Justin Salamon, Peter Li, and Juan Pablo Bello. ["Deep Saliency Representations for F0 Estimation in Polyphonic Music."](#) In Proc. *ISMIR 2017*, Suzhou, China, 2017.

[19] Kim, Jong Wook, Justin Salamon, Peter Li, and Juan Pablo Bello. "[CREPE: A convolutional representation for pitch estimation](#)." In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AL, Canada, 2018.

AVM2. Audio-Visual Multi-tasking Network: Action Recognition, Saliency Estimation and Video Summarization: Οπτικοακουστικό δίκτυο που πραγματοποιεί ταυτόχρονα αναγνώριση δράσεων, εκτίμηση του saliency και περίληψη

(Συνεργάτες: Δρ. Αντιγόνη Τσιάμη, ΕΜΠ)

References:

- [1] P. Koutras and P. Maragos, "[SUSiNet: See, Understand and Summarize It](#)", *Proc. CVPR Workshop* 2019.
- [2] A. Tsiami, P. Koutras and P. Maragos, "[STAViS: Spatio-Temporal AudioVisual Saliency Network](#)", *Proc. CVPR 2020*.
- [3] 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.
- [4] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos and P. Maragos, "[COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization](#)", *EURASIP Journal on Image and Video Processing* (2017) 2017:54.

AVM3. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα AM-FM και αλγορίθμους πολυζωνικής ενεργειακής αποδιαμόρφωσης, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "[Energy Separation in Signal Modulations with Application to Speech Analysis](#)", *IEEE Trans. Signal Processing*, vol.41, no.10, pp.3024-3051, Oct. 1993.
- [2] A. Potamianos and P. Maragos, "[Speech Processing Applications Using an AM-FM Modulation Model](#)", *Speech Communication*, vol.28, no.3, pp.195-209, July 1999.
- [3] D. Dimitriadis, P. Maragos, and A. Potamianos, "[Robust AM-FM Features for Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.12, no.9, pp.621-624, Sep. 2005.
- [4] J. H.L. Hansen and S. Patil, [Speech Under Stress: Analysis, Modeling and Recognition](#), Springer LNAI 4343, 2007.
- [5] D. Dimitriadis, P. Maragos and A. Potamianos, "[On the Effects of Filterbank Design and Energy Computation on Robust Speech Recognition](#)", *IEEE Trans. Audio, Speech and Language Processing*, vol.19, pp.1504-1516, Aug.2011.
- [6] A. Zlatintsi and P. Maragos, "[Comparison of Different Representations Based on Nonlinear Features for Music Genre Classification](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [7] T. Chaspari, D. Dimitriadis and P. Maragos, "[Emotion Classification of Speech Using Modulation Features](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [8] I. Rodomagoulakis and P. Maragos, "[On the Improvement of Modulation Features Using Multi-Microphone Energy Tracking for Robust Distant Speech Recognition](#)", *Proc. EUSIPCO-2017*, Kos, Greece, Aug. 2017.
- [9] H. B. Sailor, M. R. Kamble and H. A. Patil, "[Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection](#)", in *Proc. Interspeech 2018*.

AVM4. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα από φράκταλς και χάος, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos and A. Potamianos, "[Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition](#)", *Journal of Acoustical Society of America*, vol.105 (3), pp.1925--1932, March 1999.
- [2] V. Pitsikalis and P. Maragos, "[Filtered Dynamics and Fractal Dimensions for Noisy Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.13, no.11, pp.711-714, Nov. 2006..

- [3] V. Pitsikalis and P. Maragos, “[Analysis and Classification of Speech Signals by Generalized Fractal Dimension Features](#)”, *Speech Communication*, vol.51, no.12, pp.1206-1223, Dec. 2009.
- [4] A. Zlatintsi and P. Maragos, “[Multiscale Fractal Analysis of Musical Instrument Signals with Application to Recognition](#)”, *IEEE Trans. Audio, Speech and Language Processing*, vol.21, no.4, pp.737-748, Apr. 2013.
- [5] Karmele López-de-Ipina et al., “[Feature selection for spontaneous speech analysis to aid in Alzheimer’s disease diagnosis: A fractal dimension approach](#)”, *Computer Speech and Language* 30, pp.43–60, 2015.
- [6] E. Tzinis, G. Paraskevopoulos, C. Baziotis, A. Potamianos, “[Integrating Recurrence Dynamics for Speech Emotion Recognition](#)”, in *Proc. Interspeech 2018*.
- [7] K. Avramidis, A. Zlatintsi, C. Garoufis, and P. Maragos, “Multiscale Fractal Analysis of stimulated EEG Signals with Application to Emotion Classification”, submitted to ICASSP–2021.
- [8] W. Gilpin, “Deep Reconstruction of Strange Attractors from Time Series”, NIPS 2020, [arXiv:2002.05909 \(2020\)](#).

AVM5. Speech recognition in Greek using Deep Learning

(Συνεργάτες: Δρ. Αντιγόνη Τσιάμη, ΕΜΠ)

References:

- [1] Kaldi framework, (<https://github.com/kaldi-asr/kaldi>)
- [2] Linhao Dong et al., “Speech-Transformer: a no-recurrence sequence-to-sequence model for speech recognition,” in *Proc. of ICASSP*, 2018, pp. 5884–5888.
- [3] Tara N Sainath et al., “Two-pass end-to-end speech recognition,” in *Proc. of Interspeech*, 2019, pp. 2773–2777.
- [4] Shigeki Karita et al., “Improving Transformer-based end-to-end speech recognition with connectionist temporal classification and language model integration,” in *Proc. of Interspeech*, 2019, pp. 1408–1412.
- [5] Samuel Kriman et al., “Quartznet: Deep automatic speech recognition with 1d time-channel separable convolutions,” in *Proc. of ICASSP*, 2020, pp. 6124–6128.
- [6] Wei Han et al., “ContextNet: Improving convolutional neural networks for automatic speech recognition with global context,” in *Proc. of Interspeech*, 2020.
- [7] Anmol Gulati et al., “Conformer: Convolution-augmented Transformer for speech recognition,” in *Proc. of Interspeech*, 2020.

Συστήματα, Δίκτυα, Αυτόματα, Γράφοι:

S1. Μη-γραμμικά δυναμικά συστήματα που χρησιμοποιούν max-plus άλγεβρα και finite-state automata με εφαρμογές σε ένα πρόβλημα από τις περιοχές ανίχνευσης, βελτιστοποίησης, δικτύων, ελέγχου, θεωρία γράφων, κ.ά.

References:

- [1] R. Cuninghame-Green, [Minimax Algebra](#), Springer-Verlag, New York, 1979.
- [2] P. Butkovič. [Max-linear Systems: Theory and Algorithms](#). Springer, 2010.
- [3] P. Maragos, “[Dynamical Systems on Weighted Lattices: General Theory](#)”, *Mathematics of Control, Signals and Systems*, 29:21, 2017.
- [4] I. Kordonis, P. Maragos, G. P. Papavassilopoulos, “[Stochastic Stability in Max-Product and Max-Plus Systems with Markovian Jumps](#),” *Automatica* 92 (2018) 123–132.
- [5] A. Tsiamis and P. Maragos, “[Sparsity in Max-plus Algebra](#)”, *Discrete Events Dynamic Systems*, 29:163–189, May 2019. <https://doi.org/10.1007/s10626-019-00281-1>

S2. Signal Processing on Graphs

References:

- [1] A. Sandryhaila and J. Moura, “Discrete Signal Processing on Graphs”, *IEEE Trans. Signal Processing*, Apr. 2013.
- [2] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The Emerging Field of Signal Processing on Graphs”, *IEEE Signal Processing Magazine*, May 2013.
- [3] G. Giannakis, Y. Shen and G. Karanikolas, “Topology Identification and Learning Over Graphs: Accounting for Nonlinearities and Dynamics”, *Proc. IEEE*, vol.106, May 2018.

Machine Learning:

L1. Machine Learning, Deep Neural Nets, and Tropical Geometry

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] V. Charisopoulos and P. Maragos, "[Morphological Perceptrons: Geometry and Training Algorithms](#)," *Proc. ISMM 2017*. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.
- [2] V. Charisopoulos and P. Maragos, "[A tropical approach to neural networks with piecewise linear activations](#)", arXiv:1805.08749, 2018.
- [3] E. Theodosis and P. Maragos, "[Analysis of the Viterbi Algorithm Using Tropical Algebra and Geometry](#)", Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC-18), Greece, June 2018.
- [4] D. Maclagan and B. Sturmfels, *Introduction to Tropical Geometry*, Amer. Math. Soc., Providence, RI, 2015.
- [5] P. Maragos and E. Theodosis, "[Multivariate Tropical Regression and Piecewise-Linear Surface Fitting](#)", *Proc. 45th IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020)*, May 2020.
- [6] G. Smyrnis and P. Maragos, "[Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation](#)", *Proc. 37th Int'l Conf. on Machine Learning (ICML-2020)*, July 2020.
- [7] P. Maragos and E. Theodosis, "[Tropical Geometry and Piecewise-Linear Approximation of Curves and Surfaces on Weighted Lattices](#)", in *Shape Analysis: Euclidean, Discrete and Algebraic Geometric Methods*, edited by M. Breuss, A. Bruckstein, C. Kiselman and P. Maragos, Springer, to appear.
- [8] N. Dimitriadis and P. Maragos, "[Advances in the training, pruning and enforcement of shape constraints of Morphological Neural Networks using Tropical Algebra](#)", arXiv, 2020.
- [9] N. Tsilivis, A. Tsiamis and P. Maragos, "[Sparse Approximate Solutions to Max-Plus Equations with Application to Multivariate Convex Regression](#)", arXiv, 2020.
- [10] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, <http://www.deeplearningbook.org>
- [11] [IEEE-EURASIP Summer School on Signal Processing Meets Deep Learning](#), Capri, Italy, Sep. 2017.

L2. Deep Neural Network Compression. Ελαχιστοποίηση χωρικών/χρονικών απαιτήσεων σύγχρονων νευρωνικών δικτύων. Ενδεικτικές κατευθύνσεις: pruning, quantization, weight sharing.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] G. Retsinas, A. Elafrou, G. Goumas and P. Maragos, "[RecNets: Depthwise-Recurrent Convolutional Neural Networks](#)", BMVC, 2019.
- [2] G. Smyrnis, G. Retsinas, P. Maragos, "[Maxpolynomial Division with Application to Neural Network Simplification](#)", ICASSP, 2020.
- [3] G. Retsinas, A. Elafrou, G. Goumas and P. Maragos. "[Weight Pruning via Adaptive Sparsity Loss](#)", arXiv preprint arXiv:2006.02768, 2020.
- [4] Y. Cheng, D. Wang, P. Zhou and T. Zhang, "[A survey of model compression and acceleration for deep neural networks](#)", arXiv preprint arXiv:1710.09282, 2017.
- [5] M. Rastegari, V. Ordonez, J. Redmon and A. Farhadi, "[Xnor-net: Imagenet classification using binary convolutional neural networks](#)", ECCV, 2016.

Επεξεργασία κειμένων και φυσικής γλώσσας:

T1. Εντοπισμός λέξεων-κλειδιών και αναγνώριση κειμένου σε εικόνες χειρόγραφων εγγράφων με τη χρήση βαθιών νευρωνικών δικτύων και αναπαράστασή τους σε πολύ χαμηλή διάσταση για εφαρμογές σε μεγάλο όγκο δεδομένων. *Εξρεύνηση της σημασιολογικής πληροφορίας του κειμένου για βελτιστοποίηση του αποτελεσμάτος (NLP).*

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] V. Tassopoulou, G. Retsinas and P. Maragos, "[Enhancing Handwritten Text Recognition with N-gram sequence decomposition and Multitask Learning](#)", ICPR, 2021.
- [2] G. Retsinas, G. Sfikas and P. Maragos. "[WSRNet: Joint Spotting and Recognition of Handwritten Words](#)", arXiv preprint arXiv:2008.07109, 2020.

[3] G. Retsinas, G. Louloudis, N. Stamatopoulos, G. Sfikas and B. Gatos, [“An Alternative Deep Feature Approach to Line Level Keyword Spotting”](#), CVPR, 2019.

[4] G. Retsinas et al. [“Nonlinear Manifold Embedding on Keyword Spotting using t-SNE”](#), ICDAR 2017.

[5] T. Mikolov et al. [“Efficient estimation of word representations in vector space”](#), ICLR 2013.

T2. Εντοπισμός και αναγνώριση επιγραφών κειμένου (scene text detection and recognition). Μελέτη τεχνικών βαθιάς μάθησης για την αναγνώριση κειμένου σε ρεαλιστικές σκηνές (π.χ. επιγραφές/πινακίδες) με έμφαση στην εφαρμογή τους σε πραγματικό χρόνο.

(Συνεργάτες: Δρ. Γιώργος Ρετινάς, ΕΜΠ)

References:

[1] J. Ye, Z. Chen, J. Liu and B. Du, [“TextFuseNet: Scene Text Detection with Richer Fused Features”](#), IJCAI, 2020.

[2] F. Zhan and S. Lu, [“Esir: End-to-end scene text recognition via iterative image rectification”](#), CVPR, 2019.

T3. Επεξεργασία κειμένων και φυσικής γλώσσας με αλγεβρικές και γεωμετρικές μεθόδους (vector spaces, graphs).

References:

[1] D. Widdows, “Geometry and Meaning”, Vol. 773. Stanford: CSLI publications, 2004.

[2] P. Gärdenfors, “Conceptual spaces: The Geometry of Thought”, MIT Press, 2004.

[3] T. Mikolov et al. [“Efficient estimation of word representations in vector space”](#), ICLR 2013.