

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

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ¹ 2022-2023

Καθηγητής Πέτρος Μαραγκός, E-mail: maragos@cs.ntua.gr
Εργαστήριο: CVSP: <http://cvsp.cs.ntua.gr>, IRAL: <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 επιτυχή και συνεπή παρακολούθηση και πολύ καλή απόδοση:

- Για θέματα N: ΨΕΣ, και (*) Μηχανική Μάθηση-MM ή Αναγνώριση Προτύπων-ΑΠ .
- Για θέματα V: Οραση Υπολογιστών (ΟΥ), ή πιθανώς και MM ή ΑΠ (*).
- Για θέματα R που συνδυάζουν Ρομποτική με CV/SP/ML: ΟΥ και Ρομποτική I ή MM/ΑΠ (*).
- Για θέματα AVM: ΨΕΣ, ΟΥ και MM/ΑΠ (*)
- Για θέματα L: ΨΕΣ, ΟΥ και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα S: ΨΕΣ και κάποιο μάθημα ΣΑΕ απο Ροη Σ, ή πιθανώς και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα T: ΟΥ και MM/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. Τα ενεργά θέματα εντός του 2022-2023 είναι αυτά με το σήμα (*) στον τίτλο τους.
- 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. Ανάλυση χρονοσειρών βιομετρικών δεικτών ή ηχητικών σημάτων 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 πολλών και διαφορετικών βιομετρικών δεικτών, καθώς επίσης και ακουστικών σημάτων, τα οποία πλέον χρησιμοποιούνται και στην ιατρική. Με τη διπλωματική αυτή στοχεύουμε στην ανάπτυξη ενός συστήματος επεξεργασίας δεδομένων (big data) και ειδικών αλγορίθμων μηχανικής μάθησης μετρήσεων βιομετρικών δεικτών, τα οποία καταγράφονται από πραγματικούς χρήστες μέσω φορητών αισθητήρων, για τον εντοπισμό μεταβολών και τάσεων και την έγκαιρη λήψη αποφάσεων βάσει στατιστικών εκτιμήσεων από την επεξεργασία τους.

(Συνεργάτες: ΥΔ Χρήστος Γαρούφης, ΥΔ Νίκη Ευθυμίου, Δρ. Νάνσυ Ζλατίντση, ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

References:

- [1] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, “[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#),” *Sensors*, vol. 22, no. 19, Oct. 2022.
- [2] 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.

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

- Int'l Conf. on Artificial Intelligence Applications and Innovation (AIAI-2020), Halkidiki, Greece, 5-7 June, 2020.
- [3] 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.
- [4] P. P. Filntisis, A. Zlatintsi, N. Efthymiou, E. Kalisperakis, T. Karantinos, M. Lazaridi, N. Smyrnis and P. Maragos, Identifying differences in physical activity and autonomic function patterns between psychotic patients and controls over a long period of continuous monitoring using wearable sensors, [arXiv:2011.02285](#), 2020.
- [5] 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.
- [6] 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.
- [7] C. Garoufis, A. Zlatintsi, P. P. Filntisis, N. Efthymiou, E. Kalisperakis, V. Garyfalli, T. Karantinos, L. Mantonakis, N. Smyrnis and P. Maragos, [“An Unsupervised Learning Approach for Detecting Relapses from Spontaneous Speech in Patients with Psychosis”](#), in Proc. IEEE-EMBS Int'l Conf. ON Biomedical and Health Informatics (BHI'21) Jointly Organized with the 17th IEEE-EMBS Int'l Conf. on Wearable and Implantable Body Sensor Networks (BSN'21) (BHI-BSN-21), 2021.
- [8] J. Gideon, K. Matton, S. Anderau, M.G. McInnis, and E.M. Provost, [“When to Intervene: Detecting Abnormal Mood using Everyday Smart-phone Conversations”](#), arXiv preprint arXiv:1909.11248, 2019.
- [9] K.-Y. Huang, C.-H. Wu, and M.-H. Su, [“Attention-Based Convolutional Neural Network and Long Short-Term Memory for Short-Term Detection of Mood Disorders Based on Elicited Speech Responses”](#), Pattern Recognition, vol. 88, pp. 668–678, 2019.
- [10] Z. Pan, C. Gui, J. Zhang, J. Zhu, and D. Cui, [“Detecting Manic State of Bipolar Disorder Based on Support Vector Machine and Gaussian Mixture Model using Spontaneous Speech”](#), Psychiatry investigation, vol. 15, no. 7, pp. 695, 2018.

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

(*) R1. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ, με εφαρμογές σε assistive robotics:

Μια από τις προκλήσεις για την ανάπτυξη ευφύων ρομπότ που αλληλεπιδρούν με φυσικό τρόπο με το περιβάλλον τους, είναι η ικανότητά τους να αναγνωρίζουν τις δράσεις και τις χειρονομίες που εκτελούν οι άνθρωποι ανά πάσα στιγμή. Στόχος της διπλωματικής αυτής είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών σε πραγματικό χρόνο, η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων, όπως RGB-D ή/και ανθρώπινη πόζα. Οι μέθοδοι αυτές μπορούν να χρησιμοποιηθούν σε μεγάλες βάσεις δεδομένων ή/και σε ρομποτικές διατάξεις υποβοήθησης ηλικιωμένων και ατόμων με κινητικά προβλήματα.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης)

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.
- [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] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, [“NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding”](#), in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2019

- [8] 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.
- [9] V. Vasileiou, N. Kardaris, P. Maragos, [Exploring Temporal Context and Human Movement Dynamics for Online Action Detection in Videos](#), in *Proc. 29th European Signal Processing Conference (EUSIPCO 2021)*, Dublin, Ireland, 2021.
- [10] G. Moustiris, N. Kardaris, A. Tsiami, G. Chalvatzaki, P. Koutras, A. Dometios, P. Oikonomou, C. Tzafestas, P. Maragos, E. Efthymiou, X. Papageorgiou, S.-E. Fotinea, Y. Koumpouros, A. Vacalopoulou, E. Papageorgiou, A. Karavasili, F. Koureta, D. Dimou, A. Nikolakakis, K. Karaiskos and P. Mavridis, “[The i-Walk Lightweight Assistive Rollator: First Evaluation Study](#)”, in *Frontiers in Robotics and AI*, vol. 8, pp. 272, Sep. 2021.

(* R2. Αναγνώριση ανθρώπινων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ σε πραγματικό χρόνο, με εφαρμογές σε social robotics

Στόχος είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών, η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων (πχ. RGB-D ή/και ανθρώπινη πόζα) . Επιπρόσθετα, η παραπάνω πληροφορία μπορεί να αξιοποιηθεί για την εξαγωγή χαρακτηριστικών high-level που συμβάλλουν στην πιο φυσική επικοινωνία ανθρώπου-ρομπότ, πχ εκτίμηση engagement κατά τη διάρκεια της αλληλεπίδρασης. Εφαρμογές σε social robotics & child-robot interaction.

(Συνεργάτες: ΥΔ Νίκη Ευθυμίου, ΥΔ Νίκος Κάρδαρης)

References:

- [1] 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.
- [2] N. Efthymiou, P.P Filntisis, P. Koutras, A. Tsiami, J. Hadfield, G. Potamianos, P. Maragos, “[ChildBot: Multi-Robot Perception and Interaction with Children](#)”, *Robotics and Autonomous Systems*, vol. 150, April 2022, 103975.
- [3] D. Anagnostopoulou, N. Efthymiou, C. Papailiou, P. Maragos, “[Engagement Estimation During Child Robot Interaction Using Deep Convolutional Networks Focusing on ASD Children](#)”, *Proc. IEEE Int’l Conference on Robotics and Automation (ICRA-2021)*, Xi’an, China, June 2021.
- [4] N. Efthymiou, P. Koutras, P.P. Filntisis, G. Potamianos and P. Maragos, “[Multi-view Fusion for Action Recognition in Child-Robot Interaction](#)”, in *Proc. of IEEE Int’l Conf. on Image Processing (ICIP-18)*, Athens, Greece, Oct. 2018.
- [5] E. Marinoiu, M. Zanzir, V. Olaru, and C. Sminchisescu. “[3D human sensing, action and emotion recognition in robot assisted therapy of children with autism](#)”, *Proc. CVPR 2018*.
- [6] O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, and R. W. Picard, “[CultureNet: A deep learning approach for engagement intensity estimation from face images of children with autism](#),” in *Proc. IROS. IEEE*, 2018.

(* R3. 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 [7] οδηγώντας σε παραλλαγές συστημάτων vSLAM. Έμφαση δίνεται στην αξιοποίηση συνθετότερων γεωμετρικών δομών της εικόνας (ευθείες, επίπεδα, κωνικές) [3,4,6], σημείων φυγής για αποσύζευξη μεταφορικής και περιστροφικής συνιστώσας της πόζας [8] ή και σημασιολογικής πληροφορίας [4,5], με σκοπό την αύξηση της ευρωστίας, της ακρίβειας, του εύρους των περιβαλλόντων εφαρμογής, ή τη μείωση της υπολογιστικής πολυπλοκότητας. Ευρέως χρησιμοποιούμενα σύνολα δεδομένων σε εξωτερικούς χώρους είναι το [9] και σε εσωτερικούς τα [10,11].

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμυγκας)

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] I. Asmanis, P. Mermigkas, G. Chalvatzaki, J. Peters and P. Maragos, “[A Semantic Enhancement of Unified Geometric Representations for Improving Indoor Visual SLAM](#)”, *19th International Conference on Ubiquitous Robots (UR)*, 2022.
- [6] M. Hosseinzadeh, Y. Latif, T. Pham, N. Sünderhauf and I. Reid, “[Structure Aware SLAM Using Quadrics and Planes](#)”, *ACCV*, 2018.
- [7] 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.
- [8] J. Liu and Z. Meng, “[Visual SLAM With Drift-Free Rotation Estimation in Manhattan World](#),” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6512-6519, Oct. 2020.
- [9] KIT, Visual Odometry/SLAM Evaluation, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php
- [10] TUM, RGB-D SLAM Dataset and Benchmark, [Online]. Available: <https://vision.in.tum.de/data/datasets/rgbd-dataset>
- [11] ICL-NUIM, Dataset for benchmarking VO and SLAM algorithms, [Online]. Available: <https://www.doc.ic.ac.uk/~ahanda/VaFRIC/iclnuim.html>

Οραση Υπολογιστών:

(*) **V1. Υλοποίηση συστήματος αναγνώρισης συμπεριφοράς (συναισθηματικής κατάστασης και χειρονομιών) με χρήση βαθέων νευρωνικών δικτύων στο Robot Operating System, για αλληλεπίδραση ανθρώπου-ρομπότ.**

(Συνεργάτες: ΥΔ Παναγιώτης Φιλντίσης, ΕΜΠ, και Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] P. P. Filntisis, 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] <https://www.ros.org/>
- [2] 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> .
- [3] P. P. Filntisis, N. Efthymiou, G. Potamianos, P. Maragos, “[Emotion Understanding in Videos Through Body, Context, and Visual-Semantic Embedding Loss](#)”, *Proc. ECCV Workshop*, 2020.
- [4] I. Pikoulis, P. P. Filntisis, and P. Maragos. “[Leveraging Semantic Scene Characteristics and Multi-Stream Convolutional Architectures in a Contextual Approach for Video-Based Visual Emotion Recognition in the Wild](#).” *Proc. FG Conference*, 2021
- [5] P. Antoniadis, I. Pikoulis, P. P. Filntisis, and P. Maragos. [An audiovisual and contextual approach for categorical and continuous emotion recognition in-the-wild](#). *Proc. ICCV Workshop*, 2021.
- [7] L. Wang et al. "Temporal segment networks: Towards good practices for deep action recognition." *Proc ECCV* 2016.

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”, ICASSP-2021.
- [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 and Other Deformable Objects

(*) V6.1 3D face modelling/reconstruction and applications

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

The human face is one of the most commonly-considered objects in Computer Vision and Graphics. Modelling and reconstructing the detailed 3D shape and dynamics of the human face has numerous applications, such as augmented reality, performance capture, computer games, visual effects, human-computer interaction, computer-aided craniofacial surgery, rehabilitation and research in psychology, to name a few. During the last years we have proposed state-of-the-art methods for 3D face modelling, 3D face reconstruction from in-the-wild images and videos, facial expression recognition and photo-realistic synthesis of facial videos, see e.g. [1-5]. There can be various interesting projects related to the aforementioned state-of-the-art methods. These methods have yielded promising results, but there are many ways that they can be improved, extended, studied further or applied to different problems.

References:

- [1] M.C. Doukas, M.R. Koujan, V. Sharmanska, A. Roussos, and S. Zafeiriou, “[Head2Head++: Deep Facial Attributes Re-Targeting](#)”, *IEEE Transactions on Biometrics, Behavior, and Identity Science (IEEE T-BIOM)*, Volume 3, Issue 1, pp 31-43, January 2021.
- [2] 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)*.
- [3] M.R. Koujan, A. Roussos and S. Zafeiriou. “[DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation](#)”, *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [4] J. Booth, A. Roussos et al, “[3D reconstruction of in-the-Wild Faces in Images and Videos](#)”, *PAMI* 2018.

- [5] J. Booth, A. Roussos, ..., S. Zafeiriou, “[Large Scale 3D Morphable Face Models](#)”, IJCV 2017.
- [6] S. Ploumpis et al, “Towards a complete 3D morphable model of the human head”, <https://arxiv.org/abs/1911.08008> (CVPR and PAMI reprint).

(* V6.2 Implicit neural representations for 3D shapes

(Συνεργάτης: Δρ. Αναστάσιος Ρούσσος, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ITE)

Implicit Neural Representations are a novel way to parameterize signals and 3D objects in particular. Conventional 3D shape representations are usually discrete - they are usually parameterized as grids of voxels, point clouds, or meshes. In contrast, Implicit Neural Representations parameterize a signal as a continuous function that maps the domain of the signal (i.e. the coordinates of any point in the 3D space) to whatever is at that coordinate (e.g. value of a signed distance function - SDF). This is a recent and exciting research topic with many open research questions. For example, how can we develop novel frameworks and formulations that will make the implicit neural representations easily and efficiently editable? Several project ideas can be derived based on a combination of mathematical frameworks from differential geometry and signal processing with the latest developments of deep learning.

References:

- [1] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., & Wetzstein, G. (2020). “[Implicit neural representations with periodic activation functions](#)”. Advances in Neural Information Processing Systems, 33. (*NeurIPS 2020 Oral*)
- [2] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R., 2020, August. “[NeRF: Representing scenes as neural radiance fields for view synthesis](#)”. In European conference on computer vision (ECCV) (pp. 405-421). Springer, Cham. (*ECCV 2020 Oral - Best Paper Honorable Mention*)
- [3] Sitzmann, V., Chan, E.R., Tucker, R., Snavely, N. and Wetzstein, G., 2020. “[MetaSDF: Meta-learning signed distance functions](#)”. NeurIPS 2020.
- [4] [List with recent Implicit Neural Representations works.](#)

(* V7. Semantic Image Segmentation with Deep Convolutional Neural Networks

The adoption of convolutional neural networks [1] for semantic image segmentation, i.e., pixel-level classification, has led to a tremendous improvement in performance on challenging, large-scale datasets, such as Cityscapes [2], MS COCO [3], and PASCAL VOC [4]. This has made it possible to use such models in cutting-edge applications including autonomous cars and biomedical image computing. Most of the related works [5, 6] focus primarily on architectural modifications of the employed networks in order to better combine global context aggregation and local detail preservation, and use a simple loss that is computed on individual pixels. The design of more sophisticated losses that take into account the structure which is present in semantic labelings has received significantly less attention [7]. This thesis aims to explore such priors for semantic segmentation and to exploit them in the supervision of state-of-the-art networks, in order to achieve results that better capture the regularity of real segmentations.

(Συνεργάτης: Δρ. Χρήστος Σακαρίδης, Μεταδιδακτορικός Ερευνητής, Εργαστήριο Όρασης Υπολογιστών, ETH Zurich)

References:

- [1] Jonathan Long, Evan Shelhamer, and Trevor Darrell. [Fully convolutional networks for semantic segmentation](#). In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. [The Cityscapes dataset for semantic urban scene understanding](#). In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C. Lawrence Zitnick. [Microsoft COCO: Common objects in context](#). In The European Conference on Computer Vision (ECCV), 2014.
- [4] Mark Everingham, S.M. Ali Eslami, Luc Van Gool, Christopher K.I. Williams, John Winn, and Andrew Zisserman. [The PASCAL Visual Object Classes challenge: a retrospective](#). International Journal of Computer Vision, 111:98-136, 2015.
- [5] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. [Encoder-decoder with atrous separable convolution for semantic image segmentation](#). In The European Conference on Computer Vision (ECCV), 2018.

[6] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. [Deep high-resolution representation learning for visual recognition](#). IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[7] Tsung-Wei Ke, Jyh-Jing Hwang, Ziwei Liu, and Stella X. Yu. [Adaptive affinity fields for semantic segmentation](#). In The European Conference on Computer Vision (ECCV), 2018.

(* V8. Style Adaptation from Real to Synthetic Images with Generative Adversarial Networks

Generative adversarial networks have been widely used to generate new, synthetic instances of data that can be used as real data in many problems and for many applications. SoftGrip is a European project that aims to facilitate delicate harvesting and boost production of mushrooms by developing a smart soft robotic gripper [1]. To accomplish that, the system must be able to estimate mushroom positions and poses. However, in order to develop such a system there is a great need for annotated mushroom data. These annotations would be tremendously time consuming and difficult to acquire. Therefore, the need for synthetic data is created. This thesis aims to leverage state of the art style transfer techniques [1,2,3] to produce high quality synthetic annotated data of mushrooms that can be used to train detection and pose estimation models. An existing simple pipeline to generate synthetic mushroom data will be extended and combined with real mushroom images to improve the resulting data quality.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, ΥΔ Δάφνη Αναγνωστοπούλου, ΕΜΠ)

References:

[1] Karras, Tero & Laine, Samuli & Aila, Timo. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. 4396-4405. 10.1109/CVPR.2019.00453.

[2] Patashnik, Or & Wu, Zongze & Shechtman, Eli & Cohen-Or, Daniel & Lischinski, Dani. (2021). StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery.

[3] P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[4] <https://www.softgrip-project.eu/>

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:

Ανάπτυξη συστήματος εξατομικευμένων μουσικών προτάσεων (music recommendation) [1-3, 6] βασισμένο σε μουσική ομοιότητα (music similarity) [4, 5].

Αυτόματη αναγνώριση και εντοπισμός μουσικών χαρακτηριστικών υψηλού επιπέδου, σχετικών με το μουσικό (για παράδειγμα αναγνώριση οργάνων) ή σημασιολογικό περιεχόμενο (όπως το συναίσθημα) (auto-tagging) [7 - 13].

Διαχωρισμός των φωνητικών, ή κάποιου συγκεκριμένου οργάνου, από ένα μουσικό κομμάτι (music source separation) [14 - 17]

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

References:

Music Recommendation / Music Similarity

[1] Lee et al., [Can we Listen to it Together? Factors Influencing Reception of Recommendations and Post-Recommendation Behavior](#), ISMIR-19

[2] Zangerle et al., [Hit Song Prediction: Leveraging Low And High-Level Audio Features](#), ISMIR-19

[3] Huang, Q., A. Jansen, L. Zhang, D.P. Ellis, R.A. Saurous, and J. Anderson, "[Large-scale weakly-supervised content embeddings for music recommendation and tagging](#)". In *Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

- [4] McFee, B., L. Barrington, and G. R. G. Lanckriet, "[Learning Content Similarity for Music Recommendation](#)", *IEEE Trans. on Audio, Speech & Language Processing* 20, no.8 (2012): 2207-2218
- [5] Pr  tet, L., G. Richard, and G. Peeters. "[Learning to Rank Music Tracks using Triplet Loss.](#)" *In Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.
- [6] Schedl, M., Z. Hamed, C. Ching-Wei, D. Yashar, and M. Elahi, "[Current Challenges and Visions in Music Recommender Systems Research](#)", *Int'l Journal of Multimedia Information Retrieval*, 7(2):95-116, 2018.

Auto-tagging

- [7] Kratimenos, A., 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).
- [8] Choi, K., G. Fazekas, M. Sandler, and K. Cho. "[Convolutional Recurrent Neural Networks for Music Classification.](#)" *In Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, Louisiana, USA, 2017.
- [9] Kim, T., Lee, J., & Nam, J. "[Sample-level cnn architectures for music auto-tagging using raw waveforms.](#)" *In 2018 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AL, Canada, 2018.
- [10] W., Minz, S. Chun, O. Nieto, and X. Serra. "[Data-driven Harmonic Filters for Audio Representation Learning.](#)" *In 2020 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.
- [11] R. Panda, R. Malheiro, R.P. Paiva, "[Novel audio features for music emotion recognition](#)", *IEEE Trans. on Affective Computing*, 11(4):614-26, Mar. 2018.
- [12] S. Chowdhury, A. Vall, V. Haunschmid, G. Widmer, "[Towards explainable music emotion recognition: The route via mid-level features](#)", arXiv preprint arXiv:1907.03572. 2019 Jul 8.
- [13] Castellon, R., Donahue, C. and Liang, P., "[Codified audio language modeling learns useful representations for music information retrieval.](#)" To be presented in ISMIR 2021.

Music Source Separation

- [14] Stoller, D., S. Ewert, and S. Dixon. "[Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation.](#)" *In Int'l Soc. for Music Inform. Retrieval Conf. (ISMIR)*, Paris, France, 2018.
- [15] Samuel, D., A. Ganeshan, and J. Naradowsky, "[Meta-learning Extractors for Music Source Separation,](#)" in Proc. ICASSP 2020, Barcelona, Spain, 2020.
- [16] Choi, W., M. Kim, J. Chung, D. Lee, and S. Jung. "[Investigating u-nets with various intermediate blocks for spectrogram-based singing voice separation.](#)" *In Int'l Soc. for Music Inform. Retrieval Conf. (ISMIR)*, Montreal, Quebec, Canada, 2020 (held online).
- [17] Nachmani, E., Y. Adi, and L. Wolf. "[Voice separation with an unknown number of multiple speakers.](#)" *in Proc. ICML 2020*, held online.

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

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 on EEG Signals for Music-Induced Emotion Recognition](#)", *Proc. EUSIPCO, 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 Krivanek 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.

(* AVM6. Expressive Audio-Visual Speech Synthesis (Talking heads, Avatars)

(Συνεργάτες: ΥΔ Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] P. P. Filntisis, A. Katsamanis, P. Tsiakoulis, & P. Maragos. [Video-realistic expressive audio-visual speech synthesis for the Greek language](#). *Speech Communication*, 95, 137-152.
- [2] J. Thies, M. Elgharib, A. Tewari, C. Theobalt, & M. Nießner. “[Neural voice puppetry: Audio-driven facial reenactment](#)”. *Proc. ECCV 2020*.
- [3] A. H. Abdelaziz, A. P. Kumar, C. Seivwright, G. Fanelli, J. Binder, Y. Stylianou, Y., & S. Kajarekar, S. [Audiovisual Speech Synthesis using Tacotron2](#). arXiv preprint arXiv:2008.00620.
- [4] R. Anderson, B. Stenger, V. Wan, & R. Cipolla. [Expressive visual text-to-speech using active appearance models](#). In *Proc CVPR 2013*.

Datasets:

- [5] Kaisiyuan, Wang, et al. “[Mead: A large-scale audio-visual dataset for emotional talking-face generation](#).” *Proc ECCV, 2020*.
- [6] P. P. Filntisis, A. Katsamanis and P. Maragos, “[Photorealistic adaptation and interpolation of facial expressions using HMMS and AAMS for audio-visual speech synthesis](#),” *Proc ICIP 2017*.

Συστήματα, Δίκτυα, Γράφοι, Optimization & Learning:

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.

(* S2. Sparsity and Max-plus Algebra, Statistical Analysis of Tropical Estimators

(Συνεργάτες: Νίκος Τσιλιβής, Courant Institute of New York)

- [1] 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>
- [2] N. Tsilivis, A. Tsiamis and P. Maragos, “[Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression](#)”, *Proc. ICASSP-2021*, June 2021.
- [3] N. Tsilivis, A. Tsiamis and P. Maragos, “Sparse Approximate Solutions to Max-Plus Equations”, In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology: DGMM-2021*, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39

S3. 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 Geometry

(*) L1.1 Geometry, Pruning, Training, Applications of DNNs with PWL Activations and/or Morphological Networks for Classification or Regression Problems

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

- Τα τροπικά μαθηματικά μπορούν να δώσουν κλειστές λύσεις για piecewise-linear approximations [4]. Μετέπειτα εργασίες [9] επεκτείνουν το framework ώστε να γίνεται sparse approximation των πολυδιάστατων επιφανειών. Όμως, η επιλογή των στοιχείων για το sparse approximation γίνεται από μια δοσμένη “βάση” πιθανών στοιχείων. Στο προκειμένο θέμα εξετάζουμε τρόπους να μάθουμε από τα δεδομένα τη βάση των πιθανών στοιχείων για το sparse approximation.
- Ενώ τα τροπικά μαθηματικά μπορούν να χρησιμοποιηθούν για piecewise-linear approximations [4], τα approximations είναι πάντα convex. Κάτι τέτοιο περιορίζει την ευρεία εφαρμογή των ανεπτυγμένων μεθόδων, και παράλληλα δεν επιτρέπει το approximation με αυθαίρετη ευκρίνεια. Σε αυτό το θέμα εξετάζουμε τρόπους για να αρθεί αυτός ο περιορισμός (π.χ. difference of convex functions) ή τρόπους να χαρακτηριστεί το approximation error έναντι κάποιων βέλτιστων μεθόδων (π.χ. με upper και lower envelopes).
- Το pruning μεμονωμένων βαρών ενός νευρωνικού δικτύου, ενώ μειώνει το μέγεθος του δικτύου, δεν οδηγεί σε speedups κατά τη διάρκεια του inference [11]. Παράλληλα, pruning μέθοδοι που αντιμετωπίζουν το πρόβλημα globally συνήθως αγνοούν την εσωτερική δομή του δικτύου, και μέθοδοι που επιτρέπουν learnable sparsity για τα δίκτυα [12] οδηγούν σε καλύτερα αποτελέσματα. Έχοντας ως έναυσμα τα παραπάνω, στο προκειμένο θέμα διπλωματικής εξετάζουμε το pruning νευρωνικών δικτύων χρησιμοποιώντας learnable structured sparsity.

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] 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.
- [5] G. Smyrnis, G. Retsinas, P. Maragos, “[Maxpolynomial Division with Application to Neural Network Simplification](#)”, *Proc. IEEE ICASSP*, 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 Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)”, *Proc. ICASSP-2021*, June 2021.
- [9] N. Tsilivis, A. Tsiamis and P. Maragos, “[Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression](#)”, *Proc. ICASSP-2021*, June 2021.
- [10] N. Tsilivis, A. Tsiamis and P. Maragos, “Sparse Approximate Solutions to Max-Plus Equations”, In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology: DGMM-2021*, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39
- [11] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, “[Learning Structured Sparsity in Deep Neural Networks](#)”, *Advances of Neural Information Processing Systems*, 2016.
- [12] A. Kusupati, V. Ramanujan, R. Somani, M. Wortsman, P. Jain, S. Kakade, A. Farhadi, “[Soft Threshold Weight Reparameterization for Learnable Sparsity](#)”, *International Conference on Machine Learning*, 2020.
- [13] P. Maragos, V. Charisopoulos, E. Theodosis, “Tropical Geometry and Machine Learning”, *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>

(*) L1.2 Tropical Properties of Deep Neural Networks

- Μελέτη τροπικών ιδιοτήτων νευρωνικών δικτύων, με έμφαση σε βαθιές αρχιτεκτονικές, CNNs και ResNets.
- Χρήση τροπικών ιδιοτήτων για μελέτη συμπεριφοράς δικτύων σε προβλήματα μηχανικής μάθησης.
- Πρακτική εφαρμογή στα ανωτέρω προβλήματα (ενδεικτικό πρόβλημα: συμπίεση νευρωνικών δικτύων).

(Συνεργάτες: Γιώργος Σμυρνής, University of Texas at Austin, Δρ. Ιωάννης Κορδώνης, ΕΜΠ)

References:

- [1] V. Charisopoulos and P. Maragos, "[A tropical approach to neural networks with piecewise linear activations](#)", arXiv:1805.08749, 2018.
- [2] N. Dimitriadis and P. Maragos, "[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)", *Proc. IEEE ICASSP 2021*, June 2021.
- [3] P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "[Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes](#)", *Proc. ICLR 2022*, April 2022.
- [4] G. Smyrnis, G. Retsinas, P. Maragos, "[Maxpolynomial Division with Application to Neural Network Simplification](#)", *Proc. IEEE ICASSP 2020*, May 2020.
- [5] 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.
- [6] L. Zhang, G. Naitzat, and L.-H. Lim. "[Tropical geometry of deep neural networks](#)", *Proc. 35th Int'l Conf. on Machine Learning (ICML-2018)*, July 2018.
- [7] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>

(*) L2. Deep Neural Network Compression.

Ελαχιστοποίηση χωρικών/χρονικών απαιτήσεων σύγχρονων νευρωνικών δικτύων. Ενδεικτικές κατευθύνσεις: pruning, quantization, weight sharing. Αξιολόγηση σε δημοφιλείς βάσεις ταξινόμησης εικόνων:

<https://www.cs.toronto.edu/~kriz/cifar.html>,

<https://paperswithcode.com/sota/image-classification-on-cifar-100>.

(Συνεργάτες: Δρ. Γιώργος Ρετσίνας, 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.

(*) L3. Self-Supervision and contrastive learning.

Στο πλαίσιο αυτής της εργασίας θα γίνει μελέτη τεχνικών αυτο-επιβλεπομένης μάθησης (self-supervised learning), για την εκμάθηση αναπαραστάσεων δεδομένων εισόδου προς χρήση σε πολλαπλά tasks. Ενδεικτική κατεύθυνση είναι η ανάπτυξη τεχνικών στον κλάδο της μάθησης με αντιδιαστολή (contrastive learning), η οποία είναι ευρέως χρησιμοποιούμενη για την εκπαίδευση state of the art βαθιών νευρωνικών δικτύων.

(Συνεργάτες: Γιώργος Σμυρνής, University of Texas at Austin)

References:

- [1] T. Chen, S. Kornblith, M. Norouzi and G. Hinton, "[A simple framework for contrastive learning of visual representations](#)," *Proc. ICML*, 2020

- [2] J. D. Lee, Q. Lei, N. Saunshi and J. Zhuo, “[Predicting what you already know helps: Provable self-supervised learning](#)” *Proc. NeurIPS*, 2021.
- [3] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark and G. Krueger, “[Learning transferable visual models from natural language supervision](#)” *Proc. ICML*, 2021
- [4] T. Wang and P. Isola, “[Understanding contrastive representation learning through alignment and uniformity on the hypersphere](#)” *Proc. ICML*, 2020

Επεξεργασία κειμένων και φυσικής γλώσσας:

(*) T1. 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. *Εξερεύνηση της σημασιολογικής πληροφορίας του κειμένου για βελτιστοποίηση του αποτελέσματος (NLP).*

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] G. Retsinas, G. Sfikas, C. Nikou and P. Maragos, “[Deformation-invariant Networks for Handwritten Text Recognition](#)”, *ICIP*, 2021.
- [2] G. Retsinas, G. Sfikas and P. Maragos, “[From Seq2Seq recognition to handwritten word embeddings](#)”, *BMVC*, 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] J. Ye, Z. Chen, J. Liu and B. Du, “[TextFuseNet: Scene Text Detection with Richer Fused Features](#)”, *IJCAI*, 2020.
- [5] F. Zhan and S. Lu, “[Esir: End-to-end scene text recognition via iterative image rectification](#)”, *CVPR*, 2019.

T2. 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.