

ΠΡΟΣ

- 1) Όλα τα μέλη ΔΕΠ του Τμήματος Επιστήμης Υπολογιστών
- 2) Τους εκπροσώπους των Μεταπτυχιακών φοιτητών του Τμήματος Επιστήμης Υπολογιστών
- 3) Την Επταμελή Εξεταστική Επιτροπή
- 4) Όλα τα μέλη της Πανεπιστημιακής Κοινότητας

Πρόσκληση σε Δημόσια Παρουσίαση της Διδακτορικής Διατριβής της κας. Παυλίδη Δέσποινας

Την Δευτέρα, 26 Φεβρουαρίου 2018 και ώρα 12:00 στην αίθουσα Τηλεδιάσκεψης Κ206 του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής της υποψηφίου διδάκτορας του Τμήματος Επιστήμης Υπολογιστών κας. Παυλίδη Δέσποινας με θέμα:

**"DRACOSS: ένα ολοκληρωμένο πλαίσιο για εκτίμηση πλήθους και κατεύθυνσης άφιξης
πολλαπλών ηχητικών πηγών με συστοιχίες μικροφώνων"**

ΠΕΡΙΛΗΨΗ

Οι τεχνολογικές εξελίξεις έχουν διεισδύσει στην καθημερινότητά μας όπως ποτέ άλλοτε. Συσκευές υψηλής νοημοσύνης, εξοπλισμένες με αλγορίθμους τελευταίας τεχνολογίας, διευκολύνουν, ενισχύουν και ανυψώνουν τον τρόπο ζωής μας, κάνοντας το μέλλον να φαντάζει παρόν. Συστήματα αυτοματισμού για έξυπνα σπίτια, ακουστικά βαρηκοΐας επόμενης γενιάς, ρομπότ με αυτόνομα συστήματα πλοήγησης φέρνουν στο προσκήνιο της επιστημονικής κοινότητας προβλήματα επεξεργασίας σημάτων ήχου. Ένα από αυτά τα προβλήματα είναι η εκτίμηση του πλήθους και των κατευθύνσεων από τις οποίες προέρχεται ο ήχος, αυτό που συνήθως ονομάζουμε *εκτίμηση κατεύθυνσης άφιξης*.

Το πρόβλημα της εκτίμησης κατεύθυνσης άφιξης είναι ενεργό για πάνω από τριάντα χρόνια, συνεπώς πληθώρα αλγορίθμων έχει προταθεί στη σχετική βιβλιογραφία. Μερικοί από αυτούς τους αλγορίθμους προέρχονται από την περιοχή των τηλεπικοινωνιών. Σε αυτήν την κατηγορία ανήκουν οι τεχνικές διαμόρφωσης δέσμης, στις οποίες χρησιμοποιώντας κατάλληλα βάρη, ένα άθροισμα των σημάτων μίας

συστοιχίας μικροφώνων σχηματίζει μία δέσμη δέκτη που σαρώνει το χώρο και ανιχνεύει περιοχές ακουστικής δραστηριότητας. Οι προσεγγίσεις υποχώρων, όπως ο διάσημος αλγόριθμος MUSIC, διαμορφώνουν μία χωρική συνάρτηση η οποία μεγιστοποιείται στα σημεία που ανιχνεύεται δραστηριότητα χρησιμοποιώντας την αποσύνθεση της δειγματικής μήτρας ετεροδιακύμανσης της συστοιχίας. Άλλοι αλγόριθμοι έχουν αναδυθεί μέσα από τις προσπάθειες διαχωρισμού μειγμάτων ηχητικών σημάτων. Σε αυτήν την κατηγορία ανήκουν οι μέθοδοι που στηρίζονται στην ανάλυση ανεξάρτητων στοιχείων, στοχεύοντας στην εκτίμηση μίας μήτρας διαχωρισμού που εμπεριέχει πληροφορία για την κατεύθυνση άφιξης, καθώς και οι μέθοδοι ανάλυσης αραιών στοιχείων που εκμεταλλεύονται την αραιότητα της δραστηριότητας των πηγών σε κάποιον κατάλληλα επιλεγμένο χώρο. Μία προσέγγιση που έχει τραβήξει πρόσφατα το ενδιαφέρον είναι αυτή της εκτίμησης του διανύσματος έντασης του ηχητικού πεδίου, το οποίο έχει κατεύθυνση προς την καθαρή ροή ηχητικής ενέργειας, συνεπώς μπορεί να παρέχει την κατεύθυνση άφιξης της γεννήτριας ηχητικής πηγής.

Σε αυτήν τη διατριβή προτείνουμε ένα ολοκληρωμένο πλαίσιο για την επίλυση του προβλήματος εκτίμησης πλήθους και κατεύθυνσης άφιξης πολλαπλών, ταυτόχρονα ενεργών πηγών με τη χρήση συστοιχιών μικροφώνων. Το πλαίσιο, το οποίο ονομάζουμε εφεξής DRACOSS, αναπτύσσεται σε διδιάστατους και τριδιάστατους χώρους, χρησιμοποιώντας μία ομοιόμορφη, κυκλική συστοιχία και μία σφαιρική συστοιχία μικροφώνων αντίστοιχα. Το DRACOSS αποτελεί ουσιαστικά μία διαδικασία τεσσάρων ευκρινών βημάτων: (α) εκμετάλλευση της αραιότητας των ηχητικών σημάτων, (β) τοπική εκτίμηση κατεύθυνσης άφιξης μίας πηγής, (γ) σχηματισμός ιστογράμματος, και (δ) επεξεργασία του ιστογράμματος. Ανιχνεύουμε την αραιότητα των ηχητικών σημάτων στο πεδίο των χρονο-συχνοτήτων χρησιμοποιώντας μία χαλαρή υπόθεση αραιότητας που στηρίζεται στην εκτίμηση ενός συντελεστή μέσης συσχέτισης μεταξύ σημάτων ζευγών μικροφώνων. Σε επόμενο βήμα συλλέγουμε τοπικές εκτιμήσεις κατεύθυνσης άφιξης από όλες τις περιοχές μοναδιαίας δραστηριότητας, τις οποίες και χρησιμοποιούμε για να σχηματίσουμε ιστογράμματα. Σε διδιάστατους χώρους, ως τοπικό εκτιμητή κατεύθυνσης χρησιμοποιούμε έναν αλγόριθμο ειδικά σχεδιασμένο για κυκλικές συστοιχίες και σχηματίζουμε μονοδιάστατα ιστογράμματα, ενώ για τους τριδιάστατους χώρους χρησιμοποιούμε εκτιμήσεις του διανύσματος ηχητικής έντασης και σχηματίζουμε διδιάστατα ιστογράμματα. Και στις δύο περιπτώσεις με περαιτέρω επεξεργασία των ιστογραμμάτων παρέχουμε εκτιμήσεις του πλήθους και των κατευθύνσεων άφιξης όλων των ενεργών ηχητικών πηγών. Το DRACOSS παρουσιάζει εύρωστη απόδοση τόσο σε προσομοιωμένα, όσο και σε πραγματικά σενάρια, για διάφορες συνθήκες θορύβου και ανακλάσεων και για διάφορα πλήθη εμπλεκόμενων πηγών. Επίσης, το προτεινόμενο πλαίσιο υπερέρχει πολλών, γενικώς αναγνωρισμένων μεθόδων της βιβλιογραφίας. Επιπροσθέτως προτείνουμε την ανάπτυξη δύο κλασικών μεθόδων εκτίμησης άφιξης, της μεθόδου σχηματισμού δέσμης και του αλγορίθμου MUSIC, υπό το προτεινόμενο πλαίσιο DRACOSS, βελτιώνοντας έτσι σημαντικά την απόδοσή τους. Αποσκοπώντας στη συνεχή βελτίωση της προσέγγισής μας, ακολουθώντας, δε,

τις τελευταίες τεχνολογικές τάσεις, παρουσιάζουμε πρόσφατα και πολλά υποσχόμενα αποτελέσματα αναφορικά με την εκτίμηση πλήθους ενεργών πηγών, χρησιμοποιώντας βαθιά νευρωνικά δίκτυα.

Επόπτης Διδακτορικής Διατριβής: Αναπληρωτής Καθηγητής Αθανάσιος Μουχτάρης

“DRACOSS: a framework for direction of arrival estimation and counting of multiple sound sources with microphone arrays”

Abstract

Technological advances have infiltrated our everyday life more than ever before. High intelligence devices and gadgets, equipped with cutting-edge technology algorithms, facilitate, empower, and elevate our lifestyle, bringing the future to the present. Smart-home automation, next generation hearing aids, robots with autonomous navigation systems have brought to the foreground of the research community audio signal processing problems. One such problem is the estimation of the number of sources and the directions from which sound originates, what we most frequently call *direction of arrival (DOA)* estimation.

The problem of DOA estimation is active for more than thirty years, consequently a plethora of algorithms have been proposed in the literature. Some of them can be considered classic and frequently come from the telecommunications research area. Beamforming techniques belong in this category, where an appropriately weighted sum of the signals of a microphone array is used to form a receiving beam, which scans the space and detects areas of activity. Subspace approaches, such as the famous MUSIC algorithm, formulate a spatial function that gets maximized when activity is detected, relying on the decomposition of the array sample covariance matrix. Other algorithms stemmed from research activity on blindly separating mixtures of audio signals, i.e., the blind source separation (BSS) problem. Independent component analysis methods, where the goal is to estimate a demixing matrix, which reveals DOA information, and sparse component analysis methods, which exploit the sparsity of activity of the sources in some appropriately chosen domain, both fall into the BSS category. A recently emerging category is that of estimating the intensity vector, which points towards the net flow of sound energy, hence, revealing the corresponding DOA of the generating sound source.

In this thesis we propose the DRACOSS framework, i.e., an integrated framework for tackling the problem of DOA estimation and counting of multiple, simultaneously active, sound sources utilizing microphone arrays. DRACOSS is developed in two-dimensional (2D) and three-dimensional (3D) spaces, using a uniform circular array and a spherical microphone array respectively. DRACOSS constitutes a procedure of four distinct steps: (a) sparsity of sound signals, (b) local single-source DOA estimation, (c) histogram formation, and (d) post-processing of the histogram. We detect the sparsity of involved sound signals in the time-frequency domain by utilizing a relaxed sparsity assumption, which relies on the estimation of a mean correlation coefficient between pairs of microphones. We proceed with the collection of local DOA estimates in detected single-activity areas, which will then be used to form histograms. For the 2D case we employ a local DOA estimator, designed specifically for circular arrays and form one-dimensional histograms. For the 3D case we use an intensity vector estimator and then form two-dimensional histograms. In both cases, by post-processing the histograms we provide counting and DOA estimation results for all active sound sources. DRACOSS performs robustly under a wide collection of simulated and real scenarios in terms of noise and reverberation conditions, in terms of the number of simultaneously active sources and in comparison with state-of-the-art methods. We also propose the formulation of two classic DOA methods, i.e., beamforming and MUSIC, through the DRACOSS framework, which manages to significantly improve their performance. Aiming at constantly improving our approach and following the vivid technological stream, we show recent, very promising results on counting by utilizing deep neural networks.

Supervisor: Associate Professor Athanasios Mouchtaris