



ΠΑΡΟΥΣΙΑΣΗ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

ΗΜΕΡΟΜΗΝΙΑ: Παρασκευή, 12 Δεκεμβρίου 2025

ΩΡΑ: 10:00 – 12:00

ΑΙΘΟΥΣΑ: Αίθουσα Σεμιναρίων ΤΜΗΥΠ

ΟΜΙΛΗΤΗΣ: Ιάσων-Ιωάννης Πανάγος

Θ έ μ α

«Lightweight Methods and Models for Practical Visual Speech Recognition from Video Sequences»

Επταμελής Εξεταστική Επιτροπή:

1. **Χριστόφορος Νίκου**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
2. **Γεώργιος Σφήκας**, Επίκουρος Καθηγητής, Τμήμα Μηχανικών Τοπογραφίας & Γεωπληροφορικής, Πανεπιστήμιο Δυτικής Αττικής
3. **Λυσίμαχος-Πάυλος Κόντης**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
4. **Αριστείδης Λύκας**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
5. **Κωνσταντίνος Μπλέκας**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
6. **Αναστάσιος Κεσίδης**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Τοπογραφίας & Γεωπληροφορικής, Πανεπιστήμιο Δυτικής Αττικής
7. **Ιωάννης Κακογεωργίου**, Ερευνητής Γ', Ινστιτούτο Πληροφορικής & Τηλεπικοινωνιών, Εθνικό Κέντρο Ερευνας Φυσικών Επιστημών «Δημόκριτος».



Περίληψη:

Η αναγνώριση ομιλίας είναι ένα πρόβλημα υπολογιστικής όρασης που έχει ως σκοπό την αναγνώριση λεκτικών μονάδων (γράμματα, λέξεις ή φράσεις) που παρουσιάζονται σε ψηφιακά μέσα με οπτικο-ακουστικό περιεχόμενο. Οι εφαρμογές της αναγνώρισης ομιλίας είναι πολυδιάστατες και συναντώνται σε πολλούς τομείς της καθημερινότητας προσφέροντας πλήθος από πρακτικά οφέλη. Παραδείγματα αποτελούν εφαρμογές στον κλάδο της ιατρικής για την υποβοήθηση ατόμων με προβλήματα επικοινωνίας, στην ψηφιακή αρχειοθέτηση περιεχομένου για την διατήρηση της πολιτιστικής κληρονομιάς, λύσεις ασφαλείας σε δημοσίους χώρους αλλά και σε προσωπικό επίπεδο, κ.α.

Οι μέθοδοι που προτείνονται στην βιβλιογραφία αποτελούνται από πολύπλοκα μοντέλα και αλγορίθμους με μεγάλο υπολογιστικό κόστος, γεγονός που περιορίζει σημαντικά ή καθιστά αδύνατη την εφαρμογή τους σε πραγματικές συνθήκες και σενάρια. Στόχος αυτής της διατριβής είναι η σχεδίαση και ανάπτυξη αποδοτικών μοντέλων με χαμηλές απαιτήσεις σε υλικό και υπολογιστικούς πόρους για την αναγνώριση ομιλίας από ακολουθίες βίντεο χωρίς ήχο προσφέροντας μεγαλύτερη ευελιξία και δυνατότητες εφαρμογής.

Για τον σκοπό αυτό, αρχικά εξετάζονται μέθοδοι μείωσης της πολυπλοκότητας των υφιστάμενων μοντέλων, ακολουθώντας δύο προσεγγίσεις. Χρησιμοποιούνται δομές που μειώνουν το σύνολο των παραμέτρων των μοντέλων, μειώνοντας τις απαιτήσεις σε πόρους συστήματος, και αξιοποιούνται υπολογιστικές μονάδες που μειώνουν το μέγεθος και την πολυπλοκότητα των δικτύων. Παράλληλα προτείνεται μια μέθοδος κατασκευής μοντέλων με πολύ μικρό κόστος, που είναι ιδανικά για εφαρμογές σε περιβάλλοντα και συσκευές με περιορισμούς στο υλικό.

Στη συνέχεια, σχεδιάζονται και αναπτύσσονται ενιαίες αρχιτεκτονικές με υψηλή απόδοση και μικρό υπολογιστικό κόστος που εκμεταλλεύονται την πρόοδο της έρευνας στον τομέα της υπολογιστικής όρασης. Τα προτεινόμενα μοντέλα πετυχαίνουν ποιότητα αναγνώρισης παραπλήσια με μεθόδους της βιβλιογραφίας ενώ ταυτόχρονα είναι πολλές φορές πιο αποδοτικά σε παραμέτρους και πλήθος υπολογισμών. Επιπρόσθετα, ξεπερνούν σημαντικά τις επιδόσεις μοντέλων με παρόμοιο μέγεθος και πολυπλοκότητα που έχουν δημοσιευθεί.

Τέλος, αξιοποιούνται μέθοδοι βελτίωσης της ποιότητας αναγνώρισης χωρίς να αυξάνεται το υπολογιστικό κόστος των αρχιτεκτονικών. Στα αναπτυχθέντα ενιαία μοντέλα προστίθενται μηχανισμοί προσοχής με σκοπό την αποδοτικότερη εξαγωγή χαρακτηριστικών και την βελτίωση της ακρίβειας των εκτιμήσεων. Επιπλέον, κατά τη διάρκεια της εκπαίδευσης των μοντέλων, εισάγονται τεχνικές κανονικοποίησης ώστε να βελτιωθεί η ικανότητα μάθησής τους που και αυτή θα οδηγήσει σε υψηλότερα ποσοστά ακρίβειας.