



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΚΙΝΗΣΗΣ ΟΦΘΑΛΜΟΥ ΜΕ ΧΡΗΣΗ ΣΥΜΒΑΤΙΚΩΝ ΣΥΣΚΕΥΩΝ ΟΠΤΙΚΗΣ ΚΑΤΑΓΡΑΦΗΣ

Ευθύμιος Χριστόφορος

*Επιβλέπων καθηγητής
Χαράλαμπος Δημούλας*

Εισαγωγή

Θεώρηση του προβλήματος

Η ανίχνευση του ανθρώπινου οφθαλμού βρίσκει εφαρμογή σε διαφόρων ειδών τομείς, από την ιατρική έως και το μάρκετινγκ. Έχει διατυπωθεί ό,τι η παρατήρηση και η εκπαίδευση των κινήσεων των οφθαλμών βοηθούν στη διάγνωση και την αγωγή ασθενειών όπως η Διαταραχή Ελλειμματικής Προσοχής και Υπερκινητικότητας (ΔΕΠΥ) και Διαταραχές του Φάσματος του Αυτισμού (ΔΦΑ) μεταξύ άλλων [1][5][6]. Γενικότερα, γίνεται δυνατή η ανίχνευση της απόσπασης αλλά και της εστίασης προσοχής, γεγονός το οποίο καθιστά τέτοια τεχνολογία χρήσιμη σε διαδικασίες αξιολόγησης σχεδίων διεπαφής χρηστών σε εμπορικές εφαρμογές αλλά και αποδοτικότητας διαφημιστικών μηνυμάτων. Επιπλέον, αποτελεί ένα τρόπο διεπαφής και επικοινωνίας με ηλεκτρονικό υπολογιστή τόσο για άτομα με ειδικές ανάγκες, όσο και για τον καθημερινό χρήστη που επιθυμεί μία πιο ξεκούραστη εμπειρία.

Οι πιο προσβάσιμες και ακριβείς υλοποιήσεις συσκευών ανίχνευσης των κινήσεων του οφθαλμού και απεικόνισης στίγματος στην οθόνη συσκευής χρησιμοποιούν εξειδικευμένο υλικό σε συνδυασμό με αντίστοιχο λογισμικό. Κάποια παραδείγματα είναι συσκευή με κάμερα τοποθετημένη στο κεφάλι [1], κάμερα υψηλής ανάλυσης για καλύτερο εντοπισμό στοιχείων του προσώπου ή κάμερα υπέρυθρης ακτινοβολίας σε συνδυασμό με πηγή υπέρυθρης ακτινοβολίας για εύκολο εντοπισμό της ίριδας του ματιού [7]. Τέτοιες υλοποιήσεις είναι ιδιαίτερα ακριβές και δύσκολες στην απόκτηση για τον περισσότερο κόσμο. Από την άλλη, υπάρχουν πολλές φθηνές εναλλακτικές, οι οποίες όμως μειονεκτούν στην ακρίβεια, στην ελαστικότητα όσον αφορά τις συνθήκες φωτισμού και την ελευθερία κινήσεων του χρήστη, καθώς και στην ευκολία χρήσης και βαθμονόμησης. Τα παραπάνω καθιστούν αναγκαία την ανάπτυξη ενός αξιόπιστου και εύκολου στη χρήση συστήματος το οποίο θα μπορεί να λειτουργεί με μία απλή κάμερα, χωρίς πρόσθετες υλικές συσκευές.

Βιβλιογραφική επισκόπηση

Η διαδικασία του υπολογισμού των συντεταγμένων του σημείου στην οθόνη στο οποίο κοιτάει ο χρήστης από βίντεο απλής συσκευής καταγραφής μπορεί να χωριστεί σε τρία επιμέρους κομμάτια. Αρχικά, έχουμε τον εντοπισμό προσώπου ο οποίος βοηθά στον υπολογισμό της θέσης, της απόστασης και της γωνίας του κεφαλιού σε σχέση με την οθόνη. Στη συνέχεια απαιτείται ο εντοπισμός της κόρης ή ίριδας του ματιού ώστε να εξαχθεί η πληροφορία της γωνίας του βλέμματος και τέλος, από τις παραπάνω πληροφορίες, πρέπει να υπολογίσουμε το σημείο (x , y) στην οθόνη στο οποίο το βλέμμα του χρήστη συναντά την επιφάνειά της. Αξίζει να σημειωθεί ότι υπάρχουν μέθοδοι οι οποίες μπορούν με είσοδο την εικόνα να υπολογίσουν απευθείας τις προαναφερθείσες συντεταγμένες όπως θα δούμε παρακάτω. Στην παρούσα διπλωματική θα εστιάσουμε στα κομμάτια υλοποίησης του εντοπισμού οφθαλμού και του υπολογισμού των συντεταγμένων, έχοντας διαθέσιμες τις απαραίτητες πληροφορίες εντοπισμού προσώπου από κάποιο έτοιμο εργαλείο.

Οι μέθοδοι βασισμένες στην εμφάνιση του ματιού (appearance - based) χρησιμοποιούν τις φωτομετρικές ιδιότητες της περιοχής του ματιού σε συνδυασμό με μοντέλα μηχανικής μάθησης για να εντοπίσουν τα σημεία ενδιαφέροντος του ματιού. Αυτό σημαίνει ότι τα μοντέλα αυτά, με είσοδο τις τιμές φωτεινότητας της εικόνας στην περιοχή του ματιού, μπορούν να δώσουν σαν έξοδο την κατεύθυνση του βλέμματος ή, με κατάλληλη εκπαίδευση, και τις συντεταγμένες του σημείου στην

οθόνη απευθείας. Για να επιτευχθούν τα παραπάνω, τα μοντέλα χρειάζονται εκτενή εκπαίδευση με μεγάλα και ποικίλα σετ δεδομένων έτσι ώστε να μπορούν να είναι αποδοτικά σε διάφορα είδη οφθαλμών, συνθηκών φωτισμού και ανάλυσης εικόνας. Οι πιο συνηθισμένοι αλγόριθμοι που χρησιμοποιούνται για το σκοπό αυτό είναι οι KNN (K-Nearest-Neighbor), παλινδρόμηση RF (Random Forest), παλινδρόμηση GP (Gaussian Processers), SVM (Support Vector Machines), ANN (Artificial Neural Networks) και CNN (Convolutional Neural Networks) [4]. Παρά την απλότητα χρήσης τους και την ευελιξία τους όσον αφορά τις συνθήκες φωτισμού και ποιότητας κάμερας, τέτοιες μέθοδοι περιορίζονται από παράγοντες κλίμακας και περιστροφής του κεφαλιού και λόγω αυτού μπορούν να παρουσιάσουν μειωμένη ακρίβεια[2].

Οι μέθοδοι βασισμένες στο στα στοιχεία του ματιού (feature - based) προσπαθούν να εντοπίσουν συγκεκριμένα τοπικά στοιχεία ενδιαφέροντος του ματιού – όπως το κέντρο του, γωνίες ή περιγράμματα- ή και να ταιριάζει γεωμετρικά σχήματα – όπως κύκλοι ή ελλείψεις - στα στοιχεία αυτά. Στη δεύτερη περίπτωση, τα σχήματα τα οποία θα αντιστοιχηθούν στα διάφορα στοιχεία του ματιού μπορεί να είναι είτε απλά, όπως μία απλή έλλειψη στην κόρη ή την ίριδα του ματιού, είτε πιο σύνθετα που να περιλαμβάνουν τα βλέφαρα και τις γωνίες του ματιού [2]. Οι μέθοδοι αυτές εφαρμόζονται σε μικρό κομμάτι της εικόνας το οποίο περιλαμβάνει αποκλειστικά τον οφθαλμό έτσι ώστε να αποφεύγονται αστοχίες στην αναγνώριση των σχημάτων και λανθασμένος εντοπισμός στοιχείων του ματιού εκεί που δεν υπάρχουν [2]. Είναι προφανές ότι για να μπορέσουμε να εξάγουμε πληροφορίες από ένα μικρό κομμάτι της εικόνας που περιέχει τον οφθαλμό χρειαζόμαστε επαρκώς φωτισμένη εικόνα υψηλής ανάλυσης, γεγονός το οποίο περιορίζει αρκετά τις προδιαγραφές του συστήματος.

Για τον υπολογισμό της γωνίας όρασης και του στίγματος στην επιφάνεια της οθόνης μπορούν να χρησιμοποιηθούν μέθοδοι είτε γεωμετρικές, είτε παλινδρόμησης. Οι γεωμετρικές μέθοδοι μπορούν με τη σειρά τους να διαχωριστούν σε δυσδιάστατες και τρισδιάστατες. Οι δυσδιάστατες μέθοδοι χρησιμοποιούν μία συνάρτηση αντιστοίχισης της θέσης/γωνίας των ματιών και του κεφαλιού που εκλήφθηκαν από την εικόνα, στο σημείο στο οποία κοιτάει ο χρήστης στην οθόνη. Το θετικό των μεθόδων αυτών είναι ότι έχουν αρκετά μικρή πολυπλοκότητα, η συνάρτηση αντιστοίχισης είναι σχετικά απλή και προσφέρει αξιόπιστα αποτελέσματα [1]. Οι τρισδιάστατες γεωμετρικές μέθοδοι έχουν μεγαλύτερη πολυπλοκότητα, αφού υπολογίζουν τα διανύσματα από τα μάτια προς το σημείο όρασης στην οθόνη με χρήση της γωνίας του κεφαλιού, της γωνίας όψης των ματιών και της θέσης και της απόστασης από την οθόνη. Πέρα από την αυξημένη πολυπλοκότητά τους, οι γεωμετρικές μέθοδοι απαιτούν εικόνα υψηλής ανάλυσης για να πετύχουν επαρκή ακρίβεια στους υπολογισμούς τους [1]. Τέλος, έχουμε τις μεθόδους παλινδρόμησης οι οποίες μπορούν να λαμβάνουν σαν είσοδο μέρη της εικόνας και να παράγουν απευθείας τις συντεταγμένες του στίγματος που μας ενδιαφέρει, ή να λαμβάνουν παραμέτρους σχετικές με τη θέση/γωνία των ματιών και του κεφαλιού από αυτές να παράγει το εν λόγω αποτέλεσμα [3]. Οι μέθοδοι παλινδρόμησης όπως είδαμε είναι αρκετά ευέλικτες αλλά απαιτούν πολλή εκπαίδευση για να παρουσιάσουν ακριβή αποτελέσματα [2].

Αναφορές

- [1] [Video-oculography eye tracking towards clinical applications: A review - ScienceDirect](#)
- [2] [Unobtrusive and pervasive video-based eye-gaze tracking - ScienceDirect](#)
- [3] [On visual gaze tracking based on a single low cost camera - ScienceDirect](#)

- [4] [Appearance-Based Gaze Tracking: A Brief Review | SpringerLink](#)
- [5] [Predicting ASD diagnosis in children with synthetic and image-based eye gaze data - ScienceDirect](#)
- [6] [Detecting boredom from eye gaze and EEG - ScienceDirect](#)
- [7] [Deep Neural Networks for Low-Cost Eye Tracking - ScienceDirect](#)