Ανίχνευση μεγέθους αντικειμένου με χρήση ζωντανού βίντεο

Κωνσταντίνος Σκουρογιάννης Ηρακλής Παλληκάρης

kostskouros@outlook.com iraklis.pallikaris@gmail.com

Γενικά

Αυτή η εργασία στοχεύει στη χρήση κάμερας ώστε να πετύχουμε να υπολογίζουμε το μέγεθος του αντικειμένου, το οποίο ένας άνθρωπος προσπαθεί να πιάσει με το χέρι του. Το αντικείμενο αυτό θα πρέπει να είναι ανάμεσα σε ένα ακόμα ή και περισσότερα αντικείμενα, ενώ το πρόγραμμα θα πρέπει να προβλέπει ποιό από αυτά προσπαθεί να πιάσει ο χρήστης, ενώ να προβλέπει και το μέγεθος του αντικειμένου αυτού, βάση του ανοίγματος του χεριού του. Μια τέτοια εφαρμογή έχει πολλές μεθόδους για να επιτευχθεί, αλλά εμείς επιλέξαμε την δυνατότητα livestreaming της κάμερας στην εφαρμογή ως πρώτη προτεραιότητα. Αυτό σημαίνει πως υλοποιήσαμε την μέθοδο αυτή με τους λιγότερους υπολογισμούς δυνατόν, κρατώντας όμως την ακρίβεια της εφαρμογής σε όσο υψηλότερο επίπεδο μπορούσαμε. Για αυτό εφαρμόζουμε Hand detection βασισμένο σε προ-εκπαιδευμένο μοντέλο Machine Learning (ML) και έπειτα απλούς αριθμητικούς υπολογισμούς σε αυτά τα δεδομένα και την ίδια της εικόνα της κάμερας ώστε να πάρουμε το αντικείμενο που θέλει να πιάσει ο χρήστης, το συντομότερο δυνατόν.

Εισαγωγή

Το πεδίο του computer vision έχει αναπτυχθεί πολύ τα τελευταία χρόνια καθώς και η υπολογιστική δύναμη που χρειάζεται, για την επεξεργασία τεράστιων δεδομένων, είναι ανάλογα αυξημένη. Οι εφαρμογές που αναγνωρίζουν συναισθήματα, γκριμάτσες, χειρονομίες κ.α. έχουν εκτοξευθεί τόσο σε πλήθος όσο και σε ποιότητα. Σήμερα οι περισσότερες εφαρμογές του είδους μπορούν να εκτελέσουν εργασίες με απλούς υπολογισμούς και προ εκπαιδευμένα μοντέλα σε συσκευές όπως ένα απλό κινητό, ειδικά εκείνες οι εφαρμογές που χρησιμοποιούν την κάμερα. Έτσι και εμείς θα προσπαθήσουμε να φτιάξουμε μια εφαρμογή η οποία δεν θα είναι απαιτητική ως προς τους πόρους που χρειάζεται, ενώ θα μπορεί να καταφέρνει με επιτυχία το έργο της εύστοχα.

Γνωστικό Υπόβαθρο

Η έρευνα γύρω από το συγκεκριμένο πρόβλημα δεν είναι τόσο μεγάλη όσο σε άλλα πεδία, καθώς η λύση αυτού του προβλήματος μπορεί να χρησιμοποιηθεί σε πολύ συγκεκριμένες εφαρμογές, κυρίως σε αυτές που προορίζονται για ασφαλή και ευχάριστη συνεργασία ανθρώπου-ρομπότ. Σε τέτοιες εφαρμογές, συχνά χρησιμοποιείται πιο εκλεπτυσμένος εξοπλισμός όπως κάμερες με σένσορες βάθους (RGB-D) ή ακόμα και κάμερες εγγραφής κίνησης (Motion Capture) που λόγω της φύσης τους μπορούν να καταγράψουν περισσότερες πληροφορίες, άρα και να χρησιμοποιηθούν σε εφαρμογές που κάνουν πιο πολύπλοκους υπολογισμούς.

Σημαντική είναι και η πρόοδος στα συστήματα που μπορούν να αναγνωρίζουν τον άνθρωπο και τα μέρη του σώματος του. Η ικανότητα αντίληψης του σχήματος και της κίνησης των χεριών μπορεί να είναι ζωτικής σημασίας για τη βελτίωση της εμπειρίας του χρήστη σε διάφορους τεχνολογικούς τομείς και πλατφόρμες. Για παράδειγμα, μπορεί να αποτελέσει τη βάση για την κατανόηση της νοηματικής γλώσσας και τον έλεγχο των χειρονομιών, και μπορεί επίσης να επιτρέψει την επικάλυψη ψηφιακού περιεχομένου

και πληροφοριών πάνω από τον φυσικό κόσμο στην επαυξημένη πραγματικότητα. Ενώ έρχεται φυσικά στους ανθρώπους, η ισχυρή αντίληψη των χεριών σε πραγματικό χρόνο είναι μια αναμφισβήτητα προκλητική εργασία όρασης υπολογιστή, καθώς τα χέρια συχνά φράζουν τον εαυτό τους ή το ένα το άλλο και δεν έχουν μοτίβα υψηλής αντίθεσης. (π.χ. αποφράξεις δακτύλων/παλάμης και κουνήματα χεριών)

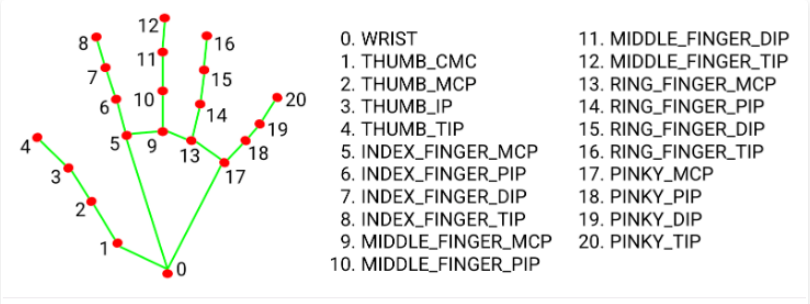
Το MediaPipe Hands είναι μια λύση παρακολούθησης χεριών και δακτύλων υψηλής πιστότητας. Χρησιμοποιεί μηχανική μάθηση (ML) για να συμπεράνει 21 τρισδιάστατα ορόσημα ενός χεριού από ένα μόνο πλαίσιο. Ενώ οι τρέχουσες προσεγγίσεις αιχμής βασίζονται κυρίως σε ισχυρά περιβάλλοντα επιτραπέζιων υπολογιστών για συμπέρασμα, η μέθοδός αυτή επιτυγχάνει απόδοση σε πραγματικό χρόνο σε ένα κινητό τηλέφωνο, ακόμη και κλιμάκωση σε πολλαπλά χέρια. Η παροχή αυτής της λειτουργικότητας αντίληψης χεριών στην ευρύτερη κοινότητα έρευνας και ανάπτυξης έχει ως αποτέλεσμα την εμφάνιση περιπτώσεων δημιουργικής χρήσης, την τόνωση νέων εφαρμογών και νέων ερευνητικών οδών. Για αυτό τον λόγο και τη χρησιμοποιούμε εμείς στον πειραματισμό μας για τον οποίο θα μιλήσουμε αναλυτικότερα παρακάτω.

Πειραματισμός

Στο πείραμα που κάναμε είχαμε ένα dataset από 8 αντικείμενα στα οποία είχαμε δώσει εξαρχής στον κώδικα τόσο το όνομα τους όσο και το πλάτος τους σύμφωνα με μέτρηση που έγινε με μέτρο

Καταρχάς το πρώτο θέμα που θα πρέπει να σχολιαστεί είναι το πως τοποθετήσαμε την κάμερα και για ποιόν λόγο έγινε αυτή η τοποθέτηση. Η κάμερα είναι σε κατακόρυφη θέση από τα αντικείμενα και το ύψος που έχουν μεταξύ τους είναι 80 cm. Ο λόγος που έχουμε κατακόρυφα την κάμερα από τα αντικείμενα είναι διότι με αυτόν τον τρόπο αποφεύγουμε το πρόβλημα της αλλαγής στην απόσταση μεταξύ του αντίχειρα και του δείκτη όταν το χέρι έρχεται πιο κοντά ή μακριά σε αυτήν. Όσον αφορά το ύψος που έχουμε επιλέξει, τοποθετήσαμε την κάμερα από τα αντικείμενα σε αυτή την απόσταση είναι διότι έτσι φαίνεται όλος ο χώρος του γραφείου που χρησιμοποιήσαμε, στη διεπαφή του χρήστη. Επίσης στην απόσταση του ύψους των 80 cm βασίζεται η συνάρτηση που μετατρέπει τα pixels σε cm σύμφωνα με της προδιάγραφες του πειράματος μας.

Για το hand detection (αναγνώριση χεριού) χρησιμοποιούμε την βιβλιοθήκη της mediapipe για python όπου μέσω αυτής καταλαβαίνουμε τόσο που είναι το χέρι, όσο μέσω τον landmarks που μας δίνει μπορούμε να καταλάβουμε σε ποιο σημείο (x,y,z) βρίσκονται στην οθόνη τα 21 σημεία του χεριού που ορίζει η βιβλιοθήκη που χρησιμοποιούμε.



Εικόνα 1: 21 σημεία του σκελετού που αναγνωρίζει και μας επιστρέφει το Media Pipe

Στο δικό μας πείραμα κάναμε κάποιες αλλαγές στις παραμέτρους του hand detection μοντέλου της Media Pipe όπου αυτές ήταν οι εξής:

1) Επιτρέπουμε αναγνώριση ενός μόνο χεριού.
2) Ανεβάσαμε το minimum detection confidence στο 0.7 από 0.5 που είναι το default ώστε να έχουμε πιο ακριβή αποτελέσματα, καθώς ο χώρος πειραματισμού μας είναι με καθαρό άσπρο background.

Καθώς πλέον έχουμε τα σημεία (x,y) για τα σημεία ενδιαφέροντος του χεριού πρέπει να υπολογίσουμε την απόσταση μεταξύ τους. Για αυτόν τον υπολογισμό παίρνουμε την απόσταση Manhattan ώστε να μας ορίσει την απόσταση σε pixels τον δυο σημείων (id=4 και id=8). Έπειτα μετατρέπουμε αυτήν την απόσταση σε cm σύμφωνα με τις προδιαγραφές του πειράματος μας (ύψος κάμερας = 80 cm) για την μετατροπή αυτή κάνουμε μια απλή μαθηματική πράξη που έχει βγει μέσω της παρατήρησης για το συγκεκριμένο πείραμα.

Μέσω αυτής της απόστασης των δακτύλων και το πλάτος των αντικειμένων του dataset που αναφέραμε δημιουργείται ένα score που είναι απολυτή τιμή τις αφαιρέσεις των αποστάσεων. Αυτό έχει ως αποτέλεσμα να έχουμε ένα score για κάθε αντικείμενο. Αυτό με το πιο μικρό score είναι και το αντικείμενο που θεωρούμε ότι θα πιάσει ο χρήστης.

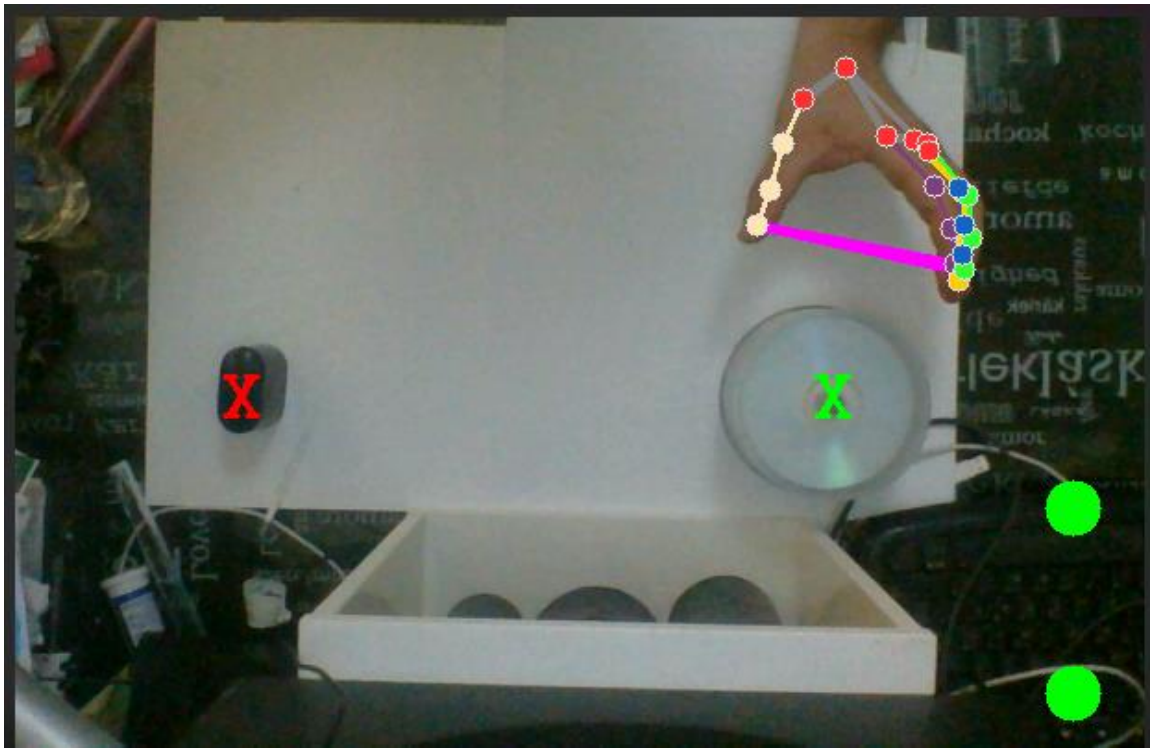
Πέρα από αυτό δημιουργούμε και ένα δεύτερο binary score όπου ελέγχει σε ποιο μέρος της οθόνης είναι ο καρπός καθώς με αυτό μπορούμε να αναγνωρίσουμε σε ποιο μισό της οθόνης (δεξιά ή αριστερά) είναι ο καρπός ώστε να κρίνουμε ποιο αντικείμενο προσπαθεί να πιάσει.

Τέλος κάνουμε ένα επιπλέον prediction με τον συνδυασμό των δυο score μας με την εξής λογική: το ένα score της απόσταση είναι ένα νούμερο που το χρησιμοποιούμε αυτούσιο, ενώ το δεύτερο score που είναι binary χρειάζεται μια μετατροπή σε αριθμό. Επιλέξαμε η μετατροπή αυτή να αποτελεί ένα αριθμητικό bonus (+2) σαν επιβράβευση καθώς το prediction από την θέση του καρπού είναι κάτι που δύσκολα αμφισβητείτε.

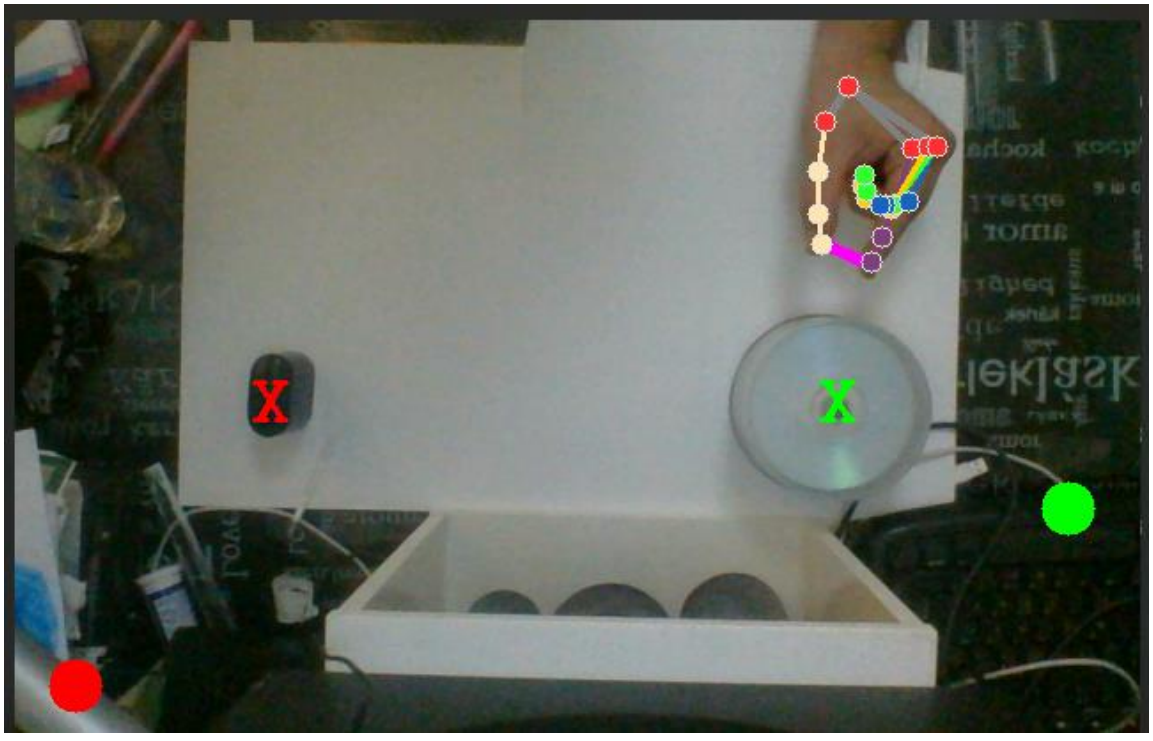
Αποτελέσματα

Καταρχάς πρέπει να γίνει μια διευκρίνηση, για το τι περιμένουμε να εμφανίζουν τα αποτελέσματα. Στην κάθε πλευρά υπάρχουν δυο τέλειες: η πάνω είναι αυτή που βασίζεται στο precision από την θέση του καρπού ενώ η τελεία κάτω βασίζεται από το precision του score της απόστασης των δακτύλων και του μήκους του εκάστοτε αντικειμένου.

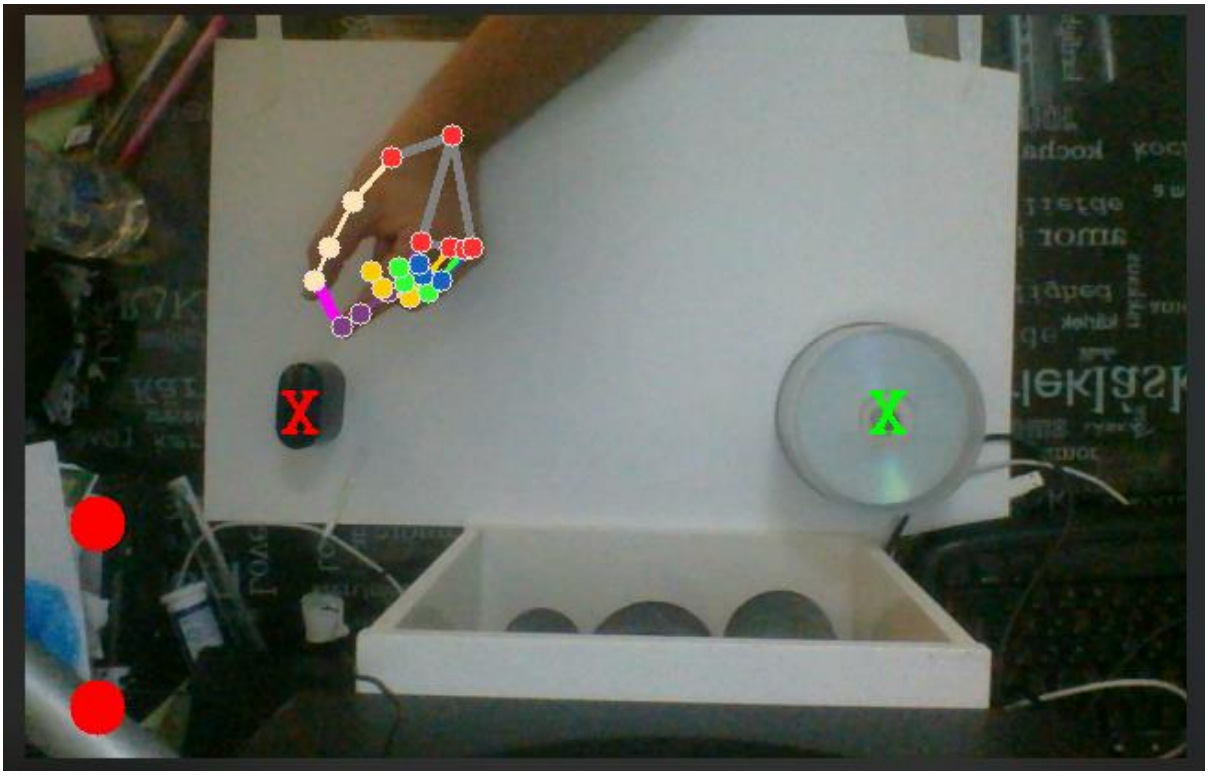
Στις εικόνες ακολουθούν cases από κάποια αποτελέσματα από prediction της εφαρμογής:



Εικόνα 2: το prediction της απόστασης και του καρπού δείχνουν το δεξιά αντικείμενο.



Εικόνα 3: Το prediction του καρπού δείχνει δεξιά ενώ της απόστασης δείχνει το αριστερά αντικείμενο.



Εικόνα 4: το prediction της απόστασης και του καρπού δείχνουν το αριστερό αντικείμενο.

Επίσης δείχνουμε δυο εικόνες για το αποτελέσματα της ένωσης των δύο score δηλαδή πως το binary score αλλάζει το αποτέλεσμα του distance score του αντικειμένου που κάνει prediction ο καρπός.

1) Σε αυτήν την εικόνα το αποτέλεσμα του καρπού συμφωνούσε με το αρχικό αποτέλεσμα των αποστάσεων και βλέπουμε ότι σε αυτήν την περίπτωση μικραίνει ακόμα περισσότερο την απόσταση του δεξιά αντικειμένου κάνοντας πιο σίγουρο το αποτέλεσμα.

```
karpos detections is right
old distance score for right object is 9.421052631578949
new distance score for right object is 7.421052631578949
distance score for left object is 19.42105263157895
```

Εικόνα 5: Εκτύπωση Terminal Scores Παράδειγμα 1ο

Από την άλλη μεριά σε αυτήν την εικόνα τα δυο score έχουν διαφορετικό precision όπως βλέπουμε μειώνει την απόσταση του αριστερά αντικειμένου κατά 2 cm αλλά και πάλι δεν είναι μικρότερο από την απόσταση score του δεξιά οπότε και πάλι θα διάλεγε το δεξιά εν τέλη.

```
karpos detections is left
old distance score for left object is 16.894736842105264
new distance score for left object is 14.894736842105264
distance score for right object is 6.894736842105264
```

Εικόνα 6: Εκτύπωση Terminal Scores Παράδειγμα 2ο

Συμπεράσματα

Τα αποτελέσματα μας καταφέρνουν να ολοκληρώσουν την δουλειά τους στο να πετυχαίνουν το αντικείμενο που έχει σκοπό ο χρήστης να πιάσει και κατά επέκταση να προβλέψουν το μέγεθος του αντικειμένου εφόσον υπάρχουν τα μεγέθη καταγεγραμμένα στη βάση δεδομένων μας.

Η εφαρμογή αυτή με πολλούς τρόπους μπορεί να αναπτυχθεί ώστε να χρησιμοποιεί πιο εκλεπτυσμένα συστήματα είτε στον τρόπο με τον οποίο γίνεται η καταγραφή των δεδομένων, είτε στον τρόπο πρόβλεψης του μεγέθους.

Όσον αφορά τον τρόπο καταγραφής δεδομένων, έχει νόημα να εφαρμοστεί ένα σύστημα "calibration" το οποίο θα έχει τη λειτουργία του να δέχεται την απόσταση της κάμερας από τα αντικείμενα και να χρησιμοποιεί αυτή τη μοναδική τιμή για να υπολογίζει τις αποστάσεις των αντικειμένων και των δαχτύλων του χεριού. Αυτό θα είχε απώτερο σκοπό να μπορεί η εφαρμογή να προβλέψει και αντικείμενα τα οποία δεν έχει «ξαναδεί» ώστε να μπορεί να έχει και γενικότερη χρήση.

Όσον αφορά τον τρόπο υπολογισμού της πρόβλεψης του αντικειμένου, υπάρχει πολύ χώρος για αναβάθμιση. Αρχικά θα ήταν χρήσιμο να υπάρχει ένα μοντέλο ML το οποίο θα έχει εκπαιδευτεί σε δεδομένα που αφορούν την κίνηση του καρπού και των γύρο σημείων του, για να μπορούμε να χρησιμοποιήσουμε πιο αποδοτικά τις άλλες πληροφορίες που μας δίνονται πέρα από την απόσταση δείκτη και αντίχειρα. Σίγουρα και η καταγραφή του χώρου του οποίου «πιάνει» η κίνηση του χεριού κατά το «άρπαγμα» του κάθε αντικειμένου μπορεί να μας δώσει σημαντικές πληροφορίες για την πρόβλεψη που θέλουμε να γίνει. Η εκπαίδευση αυτών των δεδομένων σε μαθηματικό μοντέλο θα έδινε ακόμα περισσότερη ευστοχία κατά τη πρόβλεψη.

Εν κατακλείδι, το πρόβλημα της πρόβλεψης του μεγέθους ενός αντικειμένου χρησιμοποιώντας μια κάμερα που κάνει μόνο απεικόνιση δύο διαστάσεων έχει τα αρνητικά και τα θετικά του. Τόσο στη τεχνική εντοπισμού του ανθρώπινου χεριού όσο και στη τεχνική πρόβλεψης οι σημερινές τεχνολογίες δίνουν μια πληθώρα εργαλείων και οργάνων για να πετύχουμε αυτό τον σκοπό. Ο σωστά υπολογισμένος και με ορθό τρόπο οργανωμένος συνδυασμός όλης αυτής της γνώσης είναι εκείνος που θα μπορεί να παράξει το αποτέλεσμα που κυνηγάμε, μια εφαρμογή που με εκλεπτυσμένο τρόπο να πετυχαίνει τον σκοπό της, ενώ μπορεί να υλοποιείται από την πιο αδύναμη και απλή συσκευή.

References

Code Sources
Media Pipe: https://google.github.io/mediapipe/solutions/hands.html

Sources
Dagioglou, M., Soulounias, N., Giannakopoulos, T. (2022). Object Size Prediction from Hand Movement Using a Single RGB Sensor. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science(), vol 13336. Springer, Cham. https://doi.org/10.1007/978-3-031-05643-7_24

Ansuini, Caterina & Cavallo, Andrea & Koul, Atesh & Jacono, Marco & Yang, Yuan & Becchio, Cristina. (2015). Predicting Object Size from Hand Kinematics: A Temporal Perspective. PloS one. 10. e0120432. 10.1371/journal.pone.0120432.