

Όραση Υπολογιστών

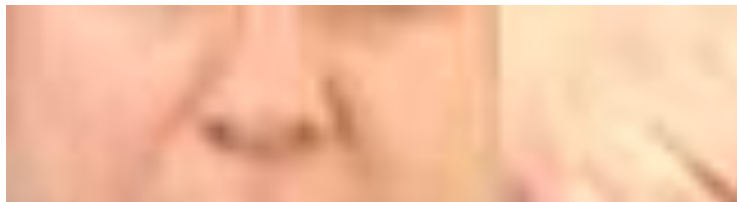
Βελεντζάς Γεώργιος
Ζωγόπουλος Γεώργιος

Θέμα : Παρακολούθηση προσώπου με Χρήση του Πεδίου Οπτικής Ροής (OpticalFlow) με τη Μέθοδο των Lucas-Kanade

Μέρος 1ο: Ανίχνευση Δέρματος Προσώπου

1.1 Εκπαίδευση Συστήματος

Αρχικά πρέπει να συστήσουμε μια κατανομή πιθανοτήτων με βάση την οποία θα κρίνεται αν ένα pixel της εικόνας ανήκει σε περιοχή δέρματος ή όχι. Μας δόθηκε έγχρωμη εικόνα δείγματος στο πρότυπο RGB. Καθώς το χρώμα στο πρότυπο αυτό εξαρτάται και από τα 3 κανάλια χρώματος, συμφέρει να μετατρέψουμε την εικόνα στο πρότυπο YCbCr. Με τον τρόπο αυτό μπορούμε να εξετάζουμε μόνον δύο κανάλια πληροφορίας για κάθε pixel για να αποφανθούμε για το σε ποια περιοχή ανήκει.



Σχήμα 1: Δείγμα δέρματος

Μετά, λοιπόν τη μετατροπή του δείγματος σε YCbCr, εξάγουμε τον μέσο όρο των τιμών σε κάθε κανάλι χρώματος, τη διακύμανση σε κάθε κανάλι χρώματος, αλλά και τη συνδιακύμανση μεταξύ των δύο καναλιών. Έτσι, μοντελοποιούμε την κατανομή του χρώματος δέρματος με την γκαουσιανή κατανομή:

$$P(\mathbf{c} = \text{skin}) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^2} e^{-\frac{1}{2}(\mathbf{c}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{c}-\boldsymbol{\mu})'} \quad (1.1)$$

Όπου

\mathbf{c} το 1x2 διάνυσμα τιμών C_b και C_r για κάθε pixel

Σ ο 2x2 πίνακας συνδιακύμανσης των τιμών C_b και C_r

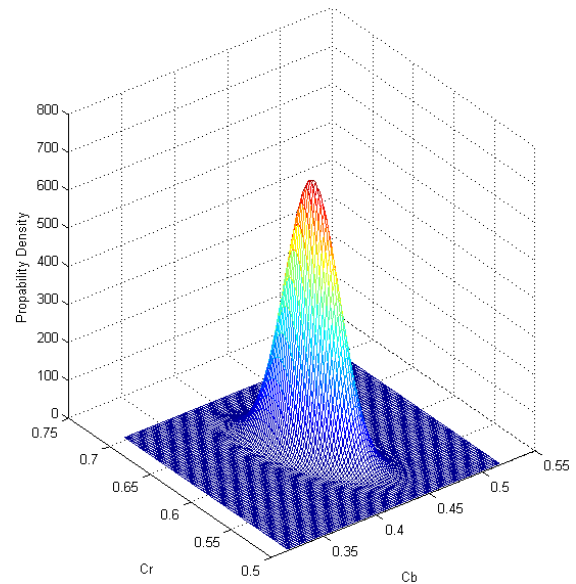
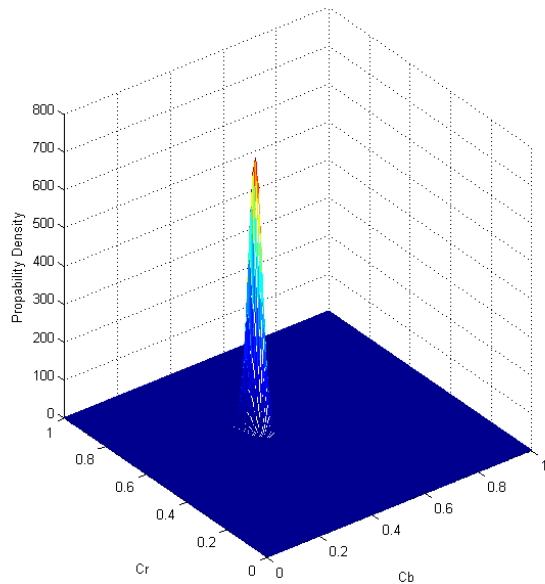
$\boldsymbol{\mu}$ το 1x2 διάνυσμα των μέσων τιμών C_b και C_r

Οι μετρήσεις που πάρθηκαν έδωσαν τα εξής αποτελέσματα:

$$\boldsymbol{\mu} = [0.4167 \ 0.6019]$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.1336 & -0.1426 \\ -0.1426 & 0.5256 \end{bmatrix} \cdot 10^{-3}$$

Οπτικά, η κατανομή της γκαουσσισιανής στον τρισδιάστατο χώρο φαίνεται στην παρακάτω εικόνα, συνολικά για το εύρος τιμών Cb/Cr, αλλά και περιορισμένη στην κεντρική περιοχή της.



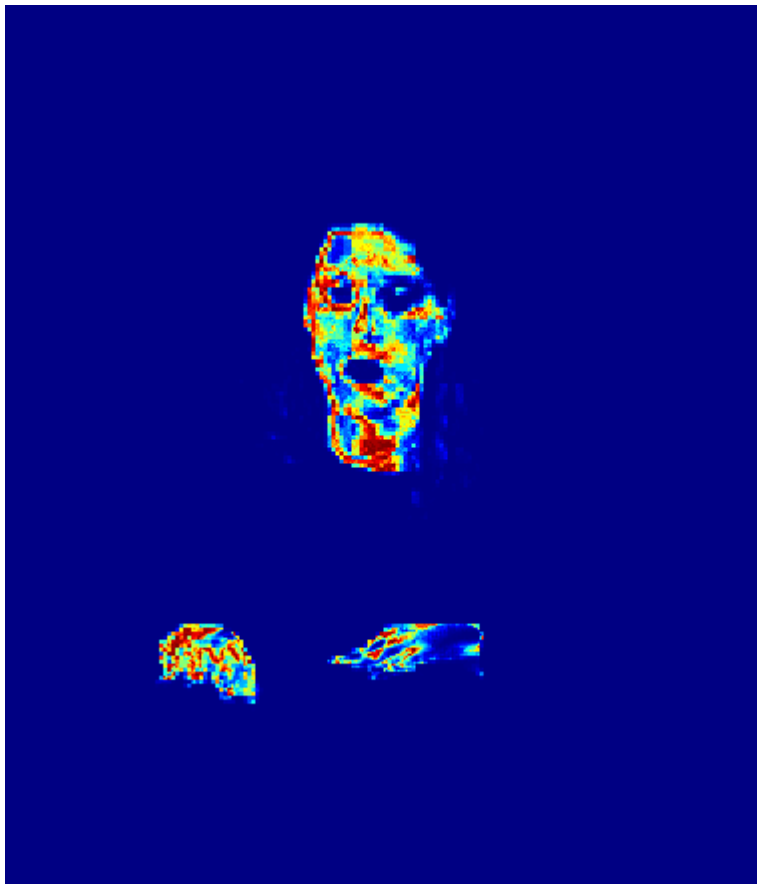
1.2 Εντοπισμός Προσώπου

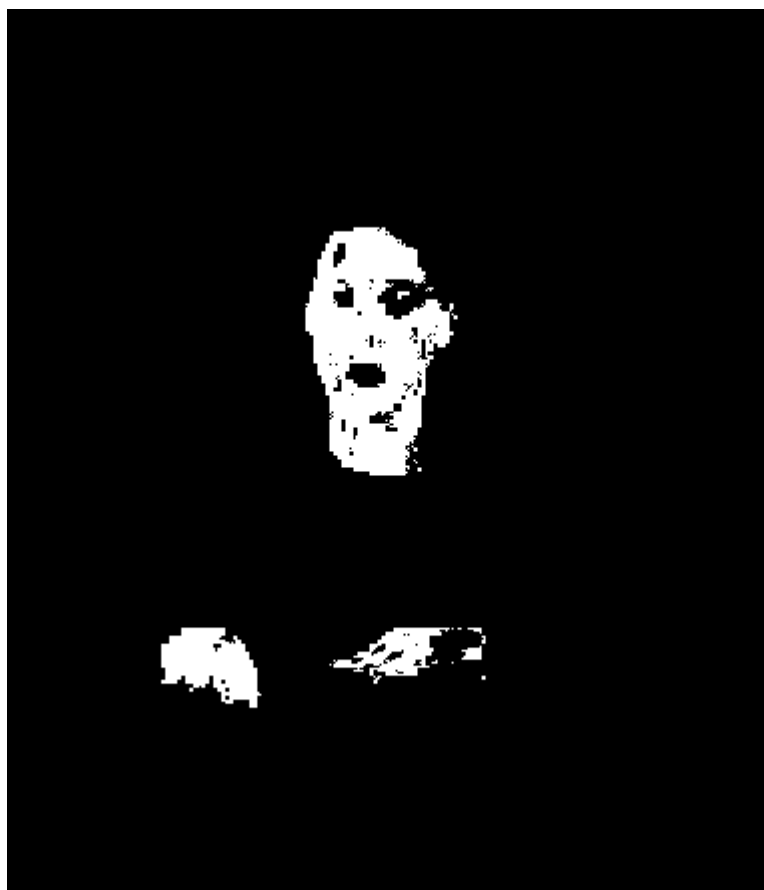
Με γνώση της πληροφορίας αυτής, κατασκευάζουμε τη συνάρτηση **fd(I, μ, Σ)**, η οποία δέχεται ως ορίσματα την εικόνα προς επεξεργασία, τη μέση τιμή και τη συνδιακύμανση του χρώματος του δέρματος και επιστρέφει τις συντεταγμένες και το μέγεθος ενός ορθογωνίου που περιβάλλει το πρόσωπο του ομιλητή.

Η συνάρτηση αυτή τελεί τα εξής βήματα:

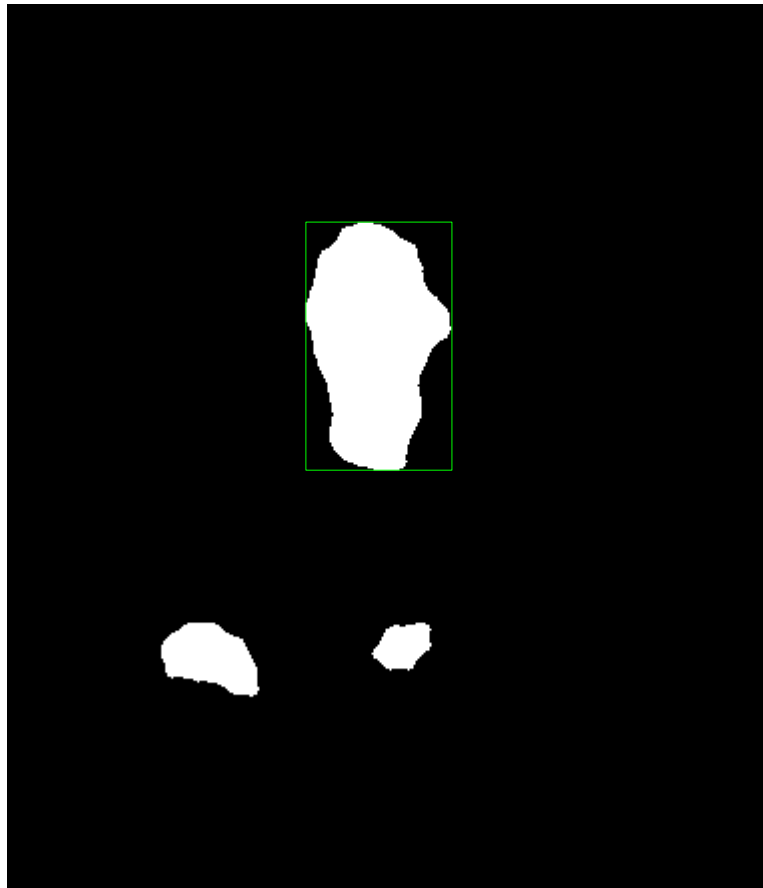
1. Μετατρέπει την εικόνα εισόδου σε YCbCr πρότυπο
2. Υπολογίζει την πιθανότητα κάθε pixel να ανήκει σε περιοχή δέρματος, βάσει των εισόδων μ και Σ και την αντίστοιχη γκαουσιανή κατανομή.
3. Κατωφλιώνει την πιθανότητα στο 0.15 και εξάγει δυαδική εικόνα με βάση αυτή τη σύγκριση.
4. Τελεί opening με δίσκο ακτίνας 3pixel, για την εξάλειψη θορύβου και μικρών προεξοχών.
5. Τελεί closing με δίσκο 15 pixel, για την ομαλοποίηση των περιγραμμάτων
6. Ονοματίζει τις διακριτές περιοχές με χρήση της `bwlabel()`
7. Εξάγει στοιχεία για τις ενιαίες περιοχές που προκύπτουν με χρήση της `regionprops()`.
8. Βρίσκει την περιοχή με το μεγαλύτερο εμβαδό και εντοπίζει την αντίστοιχη πληροφορία για τον περιβάλλον ορθογώνιο.

Η πληροφορία που επιστρέφει η **fd()** είναι της μορφής `[x, y, x_width, y_width]`, όπου `x,y` οι συντεταγμένες της άνω αριστερά γωνία του ορθογωνίου και `x_width, y_width` οι διαστάσεις του.





Σχήμα 2: Κατωφλιοποιημένες περιοχές δέρματος



Σχήμα 3: Περιοχές δέρματος μετά μορφολογικού φιλτραρίσματος και επιλογή προσώπου

Μέρος 2ο: Παρακολούθηση Προσώπου

A. Παρουσίαση της λύσης

Πρώτο μέλημα στο 2^ο μέρος της εργασίας ήταν η κατασκευή της συνάρτησης $[d_x, d_y] = lk(I1, I2, \rho, \epsilon, d_x0, d_y0)$ που υλοποιεί την εκτέλεση του αλγορίθμου Lucas-Kanade για ένα frame της εικόνας. Η συνάρτηση δέχεται κατά σειρά ορίσματα για το προηγούμενο frame του προσώπου, για το επόμενο, τη διακύμανση της γκαουσιανής, τη σταθερά ϵ , και την αρχικοποίηση της οπτικής ροής. Επιστρέφει τους συντελεστές της οπτικής ροής στον x και y άξονα.

Έπειτα, με βάση το 1^ο μέρος της εργασίας, για το πρώτο frame του βίντεο, ανιχνεύθηκε το μέγεθος και η θέση της περιοχής του προσώπου το ομιλητή. Χρησιμοποιώντας την περιοχή αυτή για την αρχικοποίηση του αλγορίθμου Lucas-Kanade και ακολουθώντας τις υποδείξεις της εκφώνησης, υλοποιήθηκε ο επαναληπτικός αλγόριθμος που εξάγει την οπτική ροή της παραθυρωμένης εικόνας, γύρω από το πρόσωπο του ομιλητή κατά τη διάρκεια όλου του βίντεο.

Για την καλύτερη εξαγωγή αποτελεσμάτων από το πεδίο οπτικής ροής, τα pixel που συμμετείχαν στον καθορισμό της συνολικής κατεύθυνσης του παραθύρου πέρασαν από φιλτράρισμα. Λήφθηκαν υπ' όψιν μόνον εκείνα των οποίων το μέτρο της οπτικής ροής ήταν μεγαλύτερο του 80% του μέγιστου μέτρου της οπτικής ροής για το εκάστοτε καρέ. Υπεύθυνη για την κατωφλιοποίηση αυτή ήταν η συνάρτηση $[displ_x, displ_y] = displ(d_x, d_y)$.

Κατά τη διάρκεια της εκτέλεσης, ο αριθμός των επαναλήψεων τέθηκε σταθερός.

Σε κάθε νέο καρέ η οπτική ροή αρχικοποιήτο στο μηδέν.

Όλα τα σχετικά αποτελέσματα αποθηκεύτηκαν σε πίνακες για μετέπειτα προβολή. Οι ποσότητες που αποθηκεύτηκαν είναι για κάθε καρέ είναι:

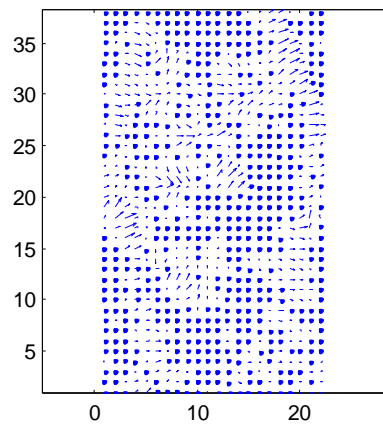
- Η οπτική ροή εντός του boundingbox.
- Η ενέργεια της οπτικής ροής εντός του boundingbox.
- Το συνολικό διάνυσμα μετατόπισης μετά από την κατωφλιοποίηση και ολοκλήρωση.
- Η θέση του boundingbox σύμφωνα με το διάνυσμα μετατόπισης

Παρακάτω εμφανίζονται ενδεικτικά αποτελέσματα της ανάλυσης για ένα τυχαίο καρέ.

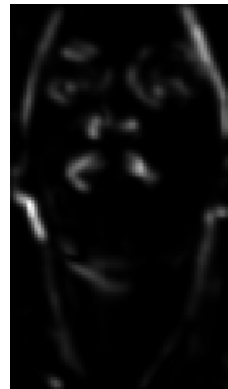
Image



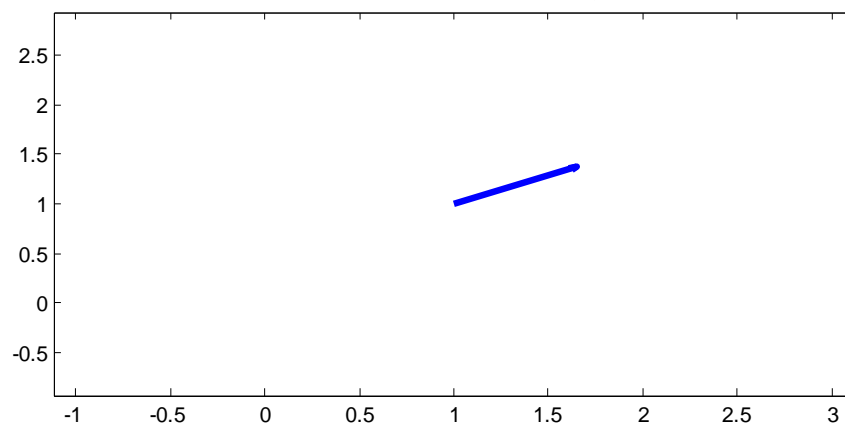
Optical Flow



Optical Flow Energy



Estimated Move of the Face. Pixels Down : -0.41126, Pixels Left : -0.72577



B. Αξιολόγηση Αποτελεσμάτων

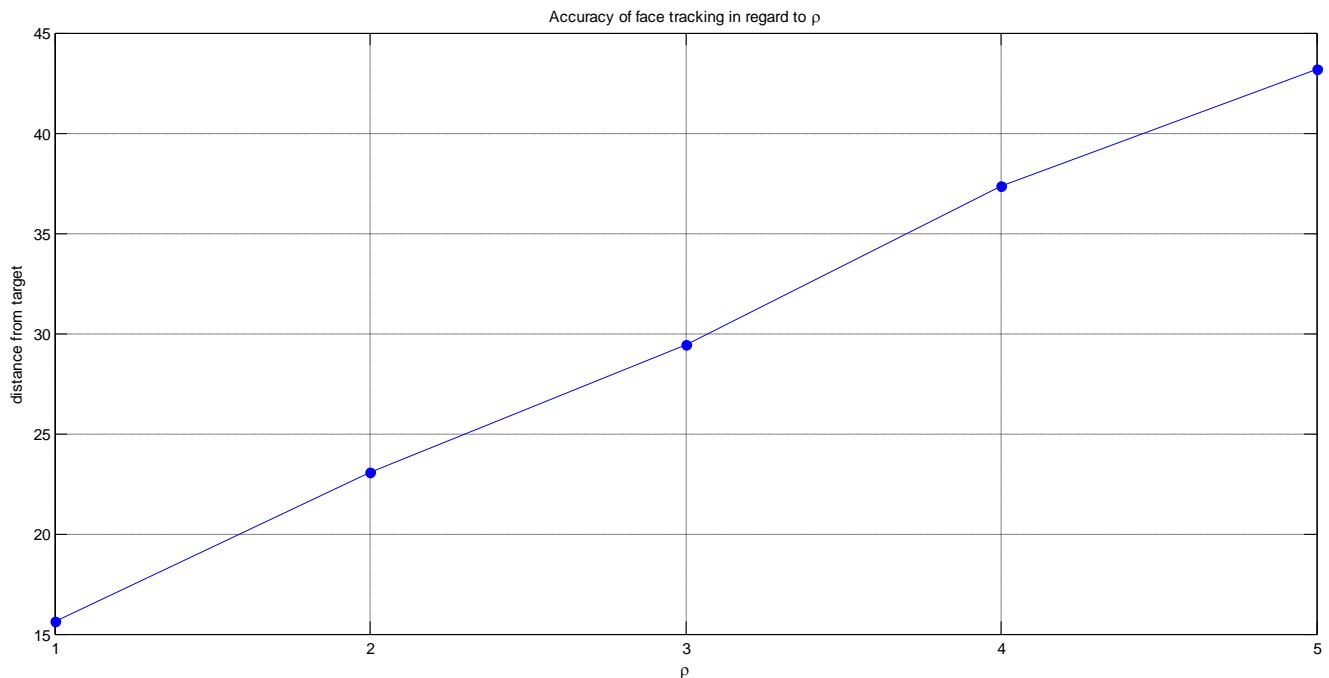
Προκειμένου να ποσοτικοποιηθεί η επίδραση των παραμέτρων ρ (διακύμανση γκαουσιανής), ε και του κατωφλίου της ενέργειας οπτικής ροής θ_ε , χρησιμοποιήθηκε το εξής κριτήριο:

d : Απόσταση του κέντρου του boundingbox στο τελικό καρέ από το πραγματικό κέντρο του προσώπου του ομιλητή.

Όσο καλύτερα ακολουθεί την κίνηση του προσώπου μέσω της οπτικής ροής το boundingbox, τόσο μικρότερο αναμένεται να είναι και το σφάλμα παρακολούθησης.

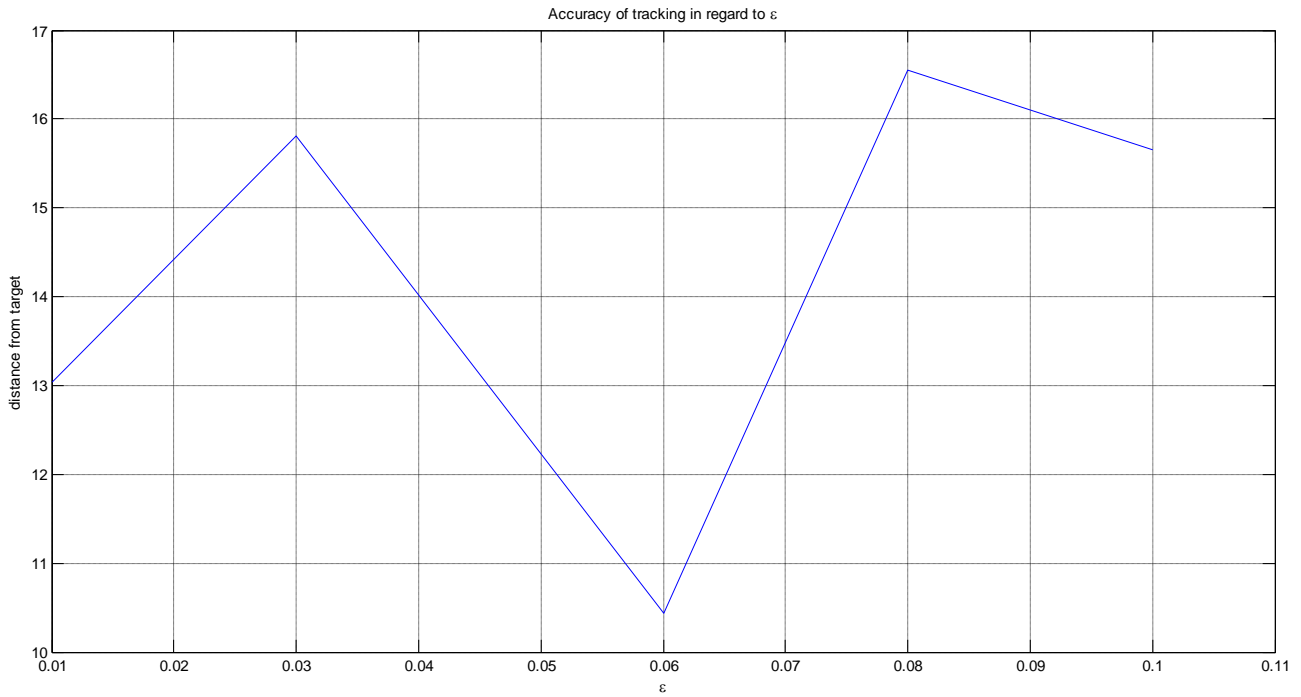
Οι αρχικές τιμές του διανύσματος παραμέτρων $(\rho, \varepsilon, \theta_\varepsilon)$ ήταν $(1.5, 0.1, 0.8)$.

Αρχικά εξετάζουμε την επίδραση του ρ . Δοκιμάσαμε τις τιμές $\{1, 2, 3, 4, 5\}$. Το αποτέλεσμα φαίνεται στο παρακάτω διάγραμμα:



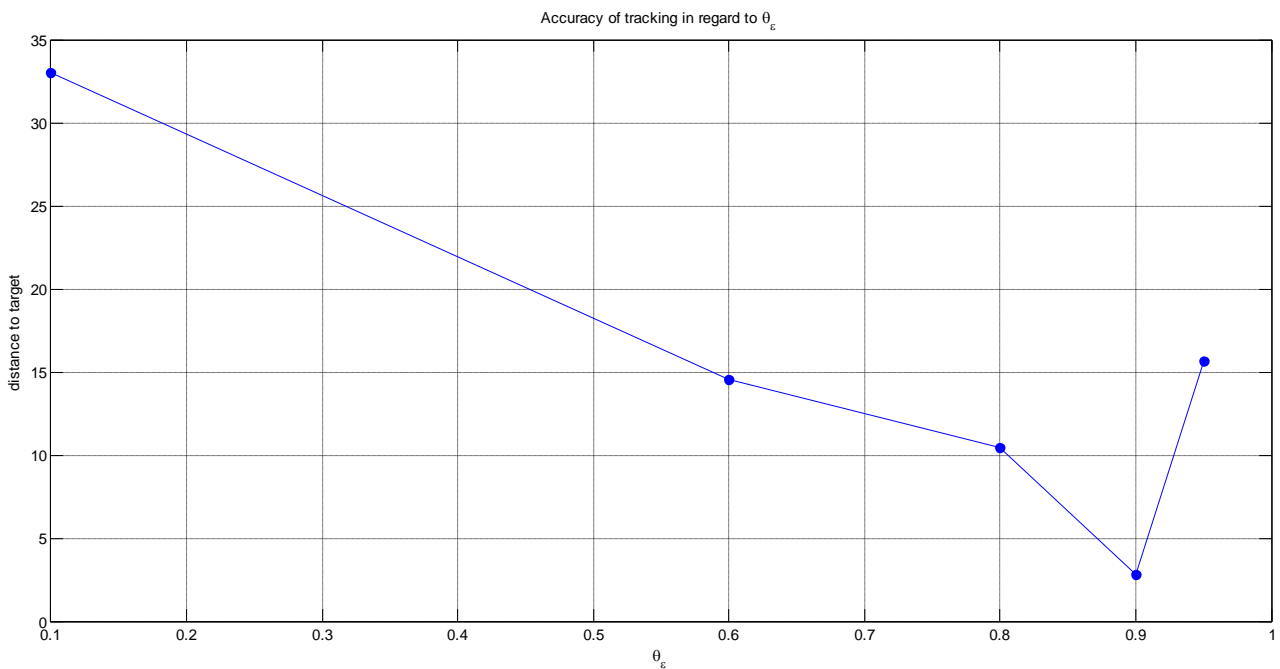
Κατά τα φαινόμενα, όσο μικρότερη είναι η τιμή του ρ , τόσο καλύτερη είναι η πρόβλεψη. Μάλιστα η σχέση φαίνεται να είναι γραμμική, τουλάχιστον στο διάστημα $[1, 5]$.

Έπειτα, εξετάζουμε την επίδραση του ε . Ως δείγμα, δοκιμάζουμε τις τιμές $\{0.01, 0.03, 0.06, 0.08, 0.01\}$. Για το ρ παίρνουμε την νέα βέλτιστη τιμή 1.



Παρατηρούμε πως μεταβολές του ε , τουλάχιστον στο διάστημα $[0.01 \ 0.1]$ δεν επηρεάζουν το αποτέλεσμα σημαντικά.

Τέλος, εξετάζουμε την επίδραση του κατωφλίου θ_ε . Θα λάβουμε μετρήσεις για $\theta_\varepsilon = \{0.1, 0.6, 0.8, 0.9, 0.95\}$ Για τις επόμενες μετρήσεις θέτουμε $\rho=1$ και $\varepsilon=0.06$.



Το πείραμα δείχνει πως μεγαλύτερα κατώφλια οδηγούν σε καλύτερα αποτελέσματα, μέχρι ενός ορίου όμως. Η καλύτερη επίδοση ήταν στο 90%.