Αλγόριθμοι Ευθυγράμμισης Τρισδιάστατων Αντικειμένων

Νικόλαος Καραμπατζιάκης*

Τμήμα Πληροφορικής & Τηλεπικοινωνιών, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών
email: N.Karampatziakis@di.uoa.gr

Περίληψη Η εύρεση του στερεού μετασχηματισμού που ευθυγραμμίζει δύο αντικείμενα της ίδιας κλάσης είναι πολύ χρήσιμη στο πεδίο της τεχνητής όρασης γιατί επιτρέπει την επίλυση διάφορων άλλων προβλημάτων χρησιμοποιώντας τις ευθυγραμμισμένες αναπαραστάσεις των αντικειμένων. Σε αυτή την εργασία αναπτύχθηκαν, προσαρμόστηκαν, συνδυάστηκαν και αξιολογήθηκαν τρεις καθιερωμένοι αλγόριθμοι ευθυγράμμισης, τα Spin Images, ο ICP και οι εικόνες βάθους. Οι αλγόριθμοι αξιολογήθηκαν στην αναγνώριση τρισδιάστατων προσώπων με εκφράσεις και ενσωματώθηκαν σε ένα σύστημα αναγνώρισης προσώπου το οποίο συμμετέχει στο διαγωνισμό FRVT όπου η ακρίβεια της ευθυγράμμισης και η ανοχή στο θόρυβο είναι πολύ σημαντικές.

1 Εισαγωγή

Πολύ συχνά στο πεδίο της τεχνητής όρασης αναζητούμε ένα γεωμετρικό μετασχηματισμό που να ευθυγραμμίζει δύο τρισδιάστατες αναπαραστάσεις ενός αντικειμένου. Έχοντας ένα τέτοιο μετασχηματισμό, πολλά προβλήματα μπορούν να λυθούν ευκολότερα. Παραδείγματα τέτοιων προβλημάτων είναι:

Συνένωση δεδομένων από διαφορετικές μετρήσεις. Συνήθως, δεδομένα που λαμβάνονται με δειγματοληψία της ίδιας σκηνής από διαφορετικές οπτικές γωνίες θα πρέπει να ενοποιηθούν. Η ευθυγράμμιση είναι μια διαδικασία που μετασχηματίζει τα διαφορετικά σύνολα δεδομένων σε ένα σύστημα συντεταγμένων ώστε να είναι ευκολότερη η συνένωση τους σε ένα ενιαίο σύνολο.

Τρισδιάστατη αναγνώριση αντικειμένων. Ένα κεντρικό συστατικό σε πολλές προσεγγίσεις του προβλήματος της αναγνώρισης ενός αντικείμενου είναι αρχικά η εύρεση μιας καλής ευθυγράμμισης με κάποιο μοντέλο αντικειμένου.

Συγχώνευση δεδομένων από διαφορετικούς αισθητήρες. Μερικές φορές, συμπληρωματικά δεδομένα για ένα αντικείμενο είναι διαθέσιμα από διάφορα είδη αισθητήρων π.χ. φωτογραφικές μηχανές ορατού και υπέρυθρου φάσματος. Μέσω της ευθυγράμμισης, τα συμπληρωματικά δεδομένα μπορούν να συγχωνευθούν για την περαιτέρω επεξεργασία τους. Η οπτικοποίηση των ευθυγραμμισμένων δεδομένων παρέχει συνήθως περισσότερες πληροφορίες από την ανεξάρτητη απεικόνισή τους.

* Επιβλέπων Καθηγητής: Θεοχάρης Θεοχάρης.

Η ευθυγράμμιση έχει ως στόχο την δημιουργία αντιστοιχήσεων μεταξύ των χαρακτηριστικών δύο αντικειμένων και την εύρεση του κατάλληλου στερεού (rigid) μετασχηματισμού που θα φέρει αυτά τα χαρακτηριστικά κοντά μεταξύ τους. Ένας στερεός μετασχηματισμός δεν αλλάζει το σχήμα ενός αντικειμένου και μπορεί να εκφραστεί ως μία περιστροφή ακολουθούμενη από μία μεταφορά. Συνήθως τα δύο αντικείμενα καλούνται μοντέλο και σκηνή. Η σύμβαση είναι ότι αν ο μετασχηματισμός εφαρμοστεί πάνω στο μοντέλο, αυτό θα ευθυγραμμιστεί με τη σκηνή.

Σε αυτή την εργασία αναπτύχθηκαν, εφαρμόστηκαν, συνδυάστηκαν και αποτιμήθηκαν τρεις αλγόριθμοι, οι εικόνες περιστροφής (Spin Images), ο ICP και οι εικόνες βάθους, για τη λύση του προβλήματος της ευθυγράμμισης τρισδιάστατων προσώπων. Οι αλγόριθμοι αυτοί μπορούν μάλιστα να εκτελεστούν σε σειρά για να εκμεταλλευτούμε τα πλεονεκτήματα και των τριών. Στην επόμενη ενότητα περιγράφονται οι τρεις αλγόριθμοι και οι απαραίτητες προσαρμογές που έγιναν για το πρόβλημα της ευθυγράμμισης προσώπων. Στην ενότητα 3 παρουσιάζονται τα υπόλοιπα τμήματα του συστήματος αναγνώρισης και τα πειραματικά αποτελέσματα μας. Τέλος, στην ενότητα 4 συγκεντρώνονται τα συμπεράσματα της εργασίας.

2 Αλγόριθμοι Ευθυγράμμισης

2.1 Η Βασική Λύση

Ας υποθέσουμε ότι μας δίνονται οι συντεταγμένες ενός πλήθους σημείων όπως μετρήθηκαν σε δύο διαφορετικά Καρτεσιανά συστήματα συντεταγμένων. Η εύρεση του στερεού μετασχηματισμού μεταξύ των δύο συστημάτων είναι το απλούστερο και βασικότερο πρόβλημα ευθυγράμμισης με το οποίο θα ασχοληθούμε.

Για τον προσδιορισμό ενός στερεού μετασχηματισμού χρειάζονται τρία τουλάχιστον ζεύγη σημείων. Γενικά έχουμε παραπάνω ζεύγη σημείων, πιθανώς με σφάλματα στις μετρήσεις, οπότε αναζητούμε τον μετασχηματισμό που θα ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων κάθε ζεύγους σημείων [1].

Ας θεωρήσουμε n σημεία και ας ονομάσουμε τα δύο συστήματα συντεταγμένων αριστερό και δεξιό. Οι συντεταγμένες των σημείων στα δύο συστήματα θα συμβολίζονται ως $\{p_{l,i}\}$ και $\{p_{r,i}\}$, $i = 1 \ldots n$. Αναζητούμε ένα μετασχηματισμό της μορφής $p_r = R(p_l) + p_0$ από το αριστερό στο δεξιό σύστημα συντεταγμένων όπου $R(p_l)$ είναι το περιστραμμένο σύνολο σημείων p_l και p_0 είναι η μεταφορά. Θα ελαχιστοποιήσουμε το άθροισμα των τετραγώνων των σφαλμάτων $\sum ||e_i||^2 = \sum ||p_{r,i} - R(p_{l,i}) - p_0||^2$.

Αποδεικνύεται εύκολα [1] ότι η βέλτιστη μεταφορά είναι απλώς η διαφορά της μέσης τιμής των σημείων στο δεξιό σύστημα από την περιστραμμένη μέση τιμή των σημείων στο αριστερό σύστημα $p_0 = \bar{p}_r - R(\bar{p}_l)$ όπου $\bar{p}_l = \frac{1}{n}\sum p_{l,i}, \bar{p}_r = \frac{1}{n}\sum p_{r,i}$.

Η περιστροφή αναπαρίσταται με μοναδιαία quaternions τα οποία έχουν πολλά πλεονεκτήματα. Αποδεικνύεται [1] ότι η βέλτιστη περιστροφή δίνεται από το μοναδιαίο quaternion $q = [q_0, q_x, q_y, q_z]^T$ που μεγιστοποιεί το $q^T N q$ όπου

$$N = \begin{bmatrix} S_{11}+S_{22}+S_{33} & S_{23}-S_{32} & S_{31}-S_{13} & S_{12}-S_{21} \\ S_{23}-S_{32} & S_{11}-S_{22}-S_{33} & S_{12}+S_{21} & S_{31}+S_{13} \\ S_{31}-S_{13} & S_{12}+S_{21} & -S_{11}+S_{22}-S_{33} & S_{23}+S_{32} \\ S_{12}-S_{21} & S_{31}+S_{13} & S_{23}+S_{32} & -S_{11}-S_{22}+S_{33} \end{bmatrix}$$

με τον πίνακα $S = \sum (p_{l,i} - \bar{p}_l)(p_{r,i} - \bar{p}_r)^T$. Είναι γνωστό ότι η τετραγωνική μορφή $q^T N q$ με $||q|| = 1$ μεγιστοποιείται όταν το q είναι το ιδιοδιάνυσμα που αντιστοιχεί στη μέγιστη ιδιοτιμή του N. Ο πίνακας περιστροφής R είναι τότε

$$R = \begin{bmatrix} q_0^2 + q_x^2 - q_y^2 - q_z^2 & 2(q_x q_y - q_0 q_z) & 2(q_x q_z + q_0 q_y) \\ 2(q_x q_y + q_0 q_z) & q_0^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q_0 q_x) \\ 2(q_x q_z - q_0 q_y) & 2(q_y q_z + q_0 q_x) & q_0^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix}$$

και με τη βοήθεια της $p_0 = \bar{p}_r - R(\bar{p}_l)$ υπολογίζεται και η μεταφορά. Μία επέκταση αυτής της λύσης προκύπτει αν στην αντικειμενική συνάρτηση προστεθούν βάρη [1], δηλαδή αν ελαχιστοποιούμε το $\sum w_i ||e_i||^2$. Τότε χρησιμοποιούνται οι σταθμισμένες μέσες τιμές $\bar{p}_l = \frac{\sum w_i p_{l,i}}{\sum w_i}$, $\bar{p}_r = \frac{\sum w_i p_{r,i}}{\sum w_i}$ και η βέλτιστη περιστροφή υπολογίζεται χρησιμοποιώντας τον πίνακα $S = \sum w_i (p_{l,i} - \bar{p}_l)(p_{r,i} - \bar{p}_r)^T$.

2.2 Εικόνες περιστροφής

Όταν δεν γνωρίζουμε για τα σημεία του μοντέλου τα σημεία της σκηνής στα οποία αντιστοιχούν, αυτή η αντιστοίχηση πρέπει να υπολογιστεί. Μία γενική μέθοδος εύρεσης τέτοιων αντιστοιχήσεων είναι μέσω εικόνων περιστροφής [2].

Οι εικόνες περιστροφής είναι ένας τρόπος περιγραφής στις δύο διαστάσεις του σχήματος ενός τρισδιάστατου αντικειμένου M θεωρώντας ως βάση ένα προσανατολισμένο σημείο $O = (p, n)$ όπου p ένα σημείο στην επιφάνεια του M και n το κανονικό διάνυσμα στο σημείο p. Το n το υπολογίζουμε ως το μέσο όρο των κανονικών διανυσμάτων των τριγώνων στα οποία συμμετέχει το p [3]. Κάθε σημείο x απεικονίζεται στις δύο διαστάσεις με δύο από τις κυλινδρικές του συντεταγμένες:

$$S_O(x) = (r, z) = \left(\sqrt{||x - p||^2 - (n \cdot (x - p))^2}, n \cdot (x - p) \right) \tag{1}$$

Μία εικόνα περιστροφής I^{OA} είναι το αποτέλεσμα της εφαρμογής της (1) στα σημεία ενός πολυγωνικού αντικειμένου A. Είναι προφανές ότι η απεικόνιση S_O είναι αναλλοίωτη σε μεταφορές και περιστροφές του αντικειμένου και επομένως μία εικόνα περιστροφής δεν εξαρτάται από τη θέση του αντικειμένου στον χώρο.

Κάθε εικόνα περιστροφής αναπαρίσταται με ένα διδιάστατο πίνακα ο οποίος είναι ένα ιστόγραμμα της κατανομής των τιμών r και z. Κάθε κάδος του ιστογράμματος έχει μέγεθος ίσο με την ανάλυση του μοντέλου (η οποία εδώ μετράται ως ο διάμεσος (median) των μηκών των ακμών του). Οι τιμές των κάδων υπολογίζονται σύμφωνα με τη διαδικασία του πίνακα 1. Η αξία κάθε σημείου διαμοιράζεται στους τέσσερις γειτονικούς κάδους για να υπάρχει μεγαλύτερη ανοχή στις ακριβείς θέσεις των σημείων.

Οι εικόνες περιστροφής που προέρχονται από δύο σημεία τα οποία θα ήταν πολύ κοντά αν δύο αντικείμενα ήταν ευθυγραμμισμένα, δεν θα είναι ακριβώς ίδιες λόγω θορύβου, διαφορετικής ψηφιοποίησης ή ελαφρώς διαφορετικού σχήματος. Αναμένεται όμως περίπου το ίδιο ποσοστό σημείων να απεικονίζεται σε αντίστοιχους κάδους. Ένας τρόπος σύγκρισης δύο γραμμικά συσχετισμένων εικόνων P και Q

Πίνακας 1. Διαδικασία κατασκευής μιας εικόνας περιστροφής. Το BinSize είναι το μέγεθος του κάδου. Το Support είναι το γινόμενο του αριθμού των pixel σε μια γραμμή ή στήλη μιας εικόνας περιστροφής επί το μέγεθος του κάδου. Τα σημεία για τα οποία $r >$ Support και $2|z| >$ Support αγνοούνται. Στην υλοποίηση το Support είναι ίσο με την μεγαλύτερη διάσταση του bounding box του μοντέλου με το οποίο ευθυγραμμίζονται όλα τα αντικείμενα.

```
Είσοδος: Ένα προσανατολισμένο σημείο O και ένα αντικείμενο M
Έξοδος: Η εικόνα περιστροφής I^{OM}
CreateSpinImage(O,M)
I = 0
foreach point x in M
    (r, z) = S_O(x)
    (a, b) = ((0.5Support − z)/BinSize, r/BinSize)
    (i, j) = (⌊a⌋, ⌊b⌋)
    (u, v) = (a − i, b − j)
    I_{i,j} += (1 − u)(1 − v); I_{i+1,j} += u(1 − v); I_{i,j+1} += (1 − u)v; I_{i+1,j+1} += uv
```

με N pixel είναι ο κανονικοποιημένος συντελεστής γραμμικής συσχέτισης

$$R(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{\left(N\sum p_i^2 - \left(\sum p_i\right)^2\right)\left(N\sum q_i^2 - \left(\sum q_i\right)^2\right)}} \tag{2}$$

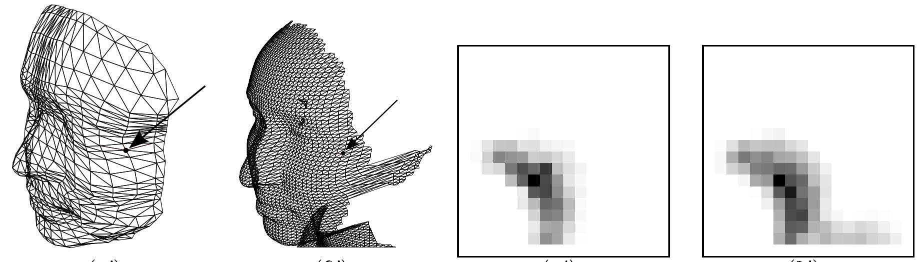
όπου τα p_i, q_i συμβολίζουν τις εντάσεις των i-οστών pixel των εικόνων P και Q. Στην περίπτωση που υπάρχουν ελλειπή δεδομένα, οι εικόνες περιστροφής δεν θα μοιάζουν σε όλη τους την έκταση (σχήμα 1). Αυτό αντιμετωπίζεται αν τα δεδομένα που χρησιμοποιούνται για τον υπολογισμό της (2) λαμβάνονται μόνο από τις περιοχές όπου οι δύο εικόνες επικαλύπτονται. Έτσι η μετρική (2) τροποποιείται σε

$$C(P,Q) = (\operatorname{atanh}(R(P,Q)))^2 - \lambda\left(\frac{1}{K-3}\right) \tag{3}$$

όπου K είναι ο αριθμός των κάδων στους οποίους και η P και η Q έχουν μη μηδενικές τιμές. Το λ τίθεται ίσο με τη μέση επικάλυψη που έχουν οι εικόνες περιστροφής στο μοντέλο. Το $\frac{1}{K-3}$ είναι η διασπορά του $\operatorname{atanh}(R(P,Q))$ [2].

Για την ευθυγράμμιση του μοντέλου με τη σκηνή, αρχικά επιλέγεται ένα τυχαίο σημείο από τη σκηνή και κατασκευάζεται η αντίστοιχη εικόνα περιστροφής. Αυτή συγκρίνεται με την εικόνα περιστροφής κάθε σημείου του μοντέλου με χρήση της (3) και από τις αντίστοιχες τιμές κατασκευάζεται ένα ιστόγραμμα. Εκεί εντοπίζονται οι outliers (ασυνήθιστα μεγάλες τιμές) και δημιουργούνται αντιστοιχήσεις μεταξύ του σημείου της σκηνής και των σημείων που αντιπροσωπεύουν αυτοί. Η διαδικασία επαναλαμβάνεται για το 10% τυχαία επιλεγμένων σημείων της σκηνής.

Για να ελεγχθεί η συνέπεια των αντιστοιχήσεων αφαιρούνται καταρχάς οι αντιστοιχήσεις με τιμή της μετρικής C μικρότερη από το μισό της καλύτερης αντιστοίχησης για το ίδιο σημείο της σκηνής. Έπειτα χρησιμοποιείται η έννοια της γεωμετρικής συνέπειας η οποία δηλώνει τη συμβατότητα δύο αντιστοιχήσεων. Έστω οι



(α′) (β′) (γ′) (δ′)

Σχήμα 1. Στα πρόσωπα των σχημάτων (α′) και (β′) έχουν επιλεγεί δύο σημεία που μια καλή ευθυγράμμιση πρέπει να φέρει πολύ κοντά. Οι αντίστοιχες εικόνες περιστροφής φαίνονται στα σχήματα (γ′) και (δ′). Όπως είναι φανερό οι δύο εικόνες δεν είναι εντελώς όμοιες

αντιστοιχήσεις $C_1 = [s_1, m_1]$, $C_2 = [s_2, m_2]$ όπου τα s_1, s_2 και m_1, m_2 είναι σημεία της σκηνής και του μοντέλου αντίστοιχα, τότε όσο πιο μικρή είναι η ποσότητα

$$d_{gc}(C_1, C_2) = \frac{||S_{m_2}(m_1) - S_{s_2}(s_1)||}{0.5\,(||S_{m_2}(m_1)|| + ||S_{s_2}(s_1)||)}$$

όπου το $S_x(y)$ ορίζεται στην (1), τόσο πιο συμβατές είναι η C_1 και C_2. Έτσι η συνέπεια δύο αντιστοιχήσεων ορίζεται από τη σχέση $D_{gc} = \max(d_{gc}(C_1, C_2), d_{gc}(C_2, C_1))$.

Οι γεωμετρικά συνεπείς αντιστοιχήσεις στο σύνολο όλων των αντιστοιχήσεων υπολογίζονται ως εξής. Αρχικά υπολογίζεται ο πίνακας $D = D_{gc}(C_i, C_j)$ και απορρίπτονται οι αντιστοιχήσεις που δεν είναι αρκετά όμοιες με τουλάχιστον 25% από τις υπόλοιπες αντιστοιχήσεις. Έπειτα εφαρμόζουμε μια ομαδοποίηση (clustering) στις υπόλοιπες αντιστοιχήσεις με βάση τον πίνακα D [2]. Σε κάθε ομάδα τριών τουλάχιστον αντιστοιχήσεων εφαρμόζουμε τη βασική λύση (ενότητα 2.1) και παίρνουμε ένα μικρό σύνολο εύλογων στερεών μετασχηματισμών. Ο επόμενος αλγόριθμος αποφασίζει ποιος από αυτούς τους μετασχηματισμούς είναι ο καταλληλότερος.

2.3 Ο Αλγόριθμος ICP (Iterative Closest Point)

Όταν τα δύο αντικείμενα που θα ευθυγραμμιστούν είναι αρκετά κοντά τότε είναι πολύ πιθανό το σημείο της σκηνής το οποίο πραγματικά αντιστοιχεί σε ένα σημείο m του μοντέλου M να είναι το κοντινότερο σημείο της σκηνής S στο m. Αυτή είναι η βασική ιδέα πίσω από τον αλγόριθμο ICP. Ας θεωρήσουμε το σύνολο $T(A) = \{(m_i, c_i) : m_i \in A \cdot M,\ c_i = \text{argmin}_{s \in S} ||s - m_i||\}$ όπου A η τρέχουσα εκτίμηση του ζητούμενου μετασχηματισμού. Το $T(A)$ υπολογίζεται γρήγορα οργανώνοντας το S σε kd-δέντρο και ο υπολογισμός θα συμβολίζεται με τον τελεστή $\mathcal{C}$: $T(A) = \mathcal{C}(A, M, S)$. Δεδομένου του $T(A)$ ο μετασχηματισμός που ευθυγραμμίζει τα σημεία του M με τα σημεία του S δίνεται από τη βασική λύση της ενότητας 2.1 και θα τον συμβολίζουμε με τον τελεστή: $(B, d) = \mathcal{Q}(T(A))$ όπου B η νέα εκτίμηση του στερεού μετασχηματισμού και $d = \frac{1}{N_m} \sum_{i=1}^{N_m} ||Bm_i - t_i||^2$ το μέσο τετραγωνικό σφάλμα μεταξύ των ευθυγραμμισμένων αντικειμένων. Ο αλγόριθμος ICP [4]

Πίνακας 2. Ο αλγόριθμος ICP

Είσοδος: Ένα αντικείμενο μοντέλο M και ένα αντικείμενο σκηνή S
Έξοδος: Ένας στερεός μετασχηματισμός A
ICP(S,M)
Κατασκεύασε το kd-δέντρο για το σύνολο των σημείων του S
$A_0 = I_3$; $k = 0$; // Το $A_k \cdot M$ αντιπροσωπεύει την τρέχουσα θέση του μοντέλου
do
$T(A_k) = \mathcal{C}(A_k, M, S)$ // Υπολογισμός των κοντινότερων σημείων
$(A_{k+1}, d_{k+1}) = \mathcal{Q}(T(A_k))$ // Υπολογισμός του μετασχηματισμού
$k = k + 1$
while $d_{k-1} - d_k > \epsilon$ // Έλεγχος σύγκλισης
return A_k

φαίνεται στον πίνακα 2. Ο ICP αποδεικνύεται ότι συγκλίνει μονότονα σε ένα τοπικό ελάχιστο του σφάλματος d και στην πράξη είναι ιδιαίτερα αποτελεσματικός.

Αν τα δύο αντικείμενα επικαλύπτονται μερικώς τότε θα υπάρχουν σημεία για τα οποία δεν θα έπρεπε να υπήρχε καμία αντιστοίχηση. Ο ICP ωστόσο υπολογίζει τα κοντινότερα σημεία στο S για όλα τα σημεία του M και έτσι θα προκύπτουν λανθασμένες αντιστοιχήσεις. Μια απλή λύση γι΄ αυτό το πρόβλημα είναι στον υπολογισμό του βέλτιστου μετασχηματισμού να αγνοούνται όλες οι (συνήθως λανθασμένες) αντιστοιχήσεις των οποίων ένα τουλάχιστον σημείο ανήκει στο σύνορο κάποιας επιφάνειας [5]. Αυτή η τεχνική είναι αποτελεσματική, αλλά η ενσωμάτωση της στον αλγόριθμο αναιρεί την εγγύηση σύγκλισης προς ένα τοπικό ελάχιστο. Ωστόσο, απλές αλλαγές στο κριτήριο σύγκλισης αρκούν για να διορθώσουν το πρόβλημα.

Ο ICP έχει άριστη συμπεριφορά όταν τα δύο αντικείμενα είναι περίπου ευθυγραμμισμένα. Αν τα δύο αντικείμενα δεν είναι κοντά, είναι δυνατόν να παγιδευτεί σε κάποιο άσχημο τοπικό ελάχιστο. Για να αρχικοποιηθεί ο ICP με μια καλή ευθυγράμμιση, μπορούν να δοκιμαστούν για λίγες επαναλήψεις όλοι οι εύλογοι μετασχηματισμοί που προέκυψαν μέσω των εικόνων περιστροφής και τελικά να επιλεγεί αυτός που στο τέλος οδήγησε στο μικρότερο μέσο τετραγωνικό σφάλμα.

2.4 Παραλλαγές του ICP

Η απλότητα του ICP έχει ως αποτέλεσμα ο βασικός αλγόριθμος να έχει επεκταθεί με διάφορους τρόπους [6,7,8,9]. Μερικές παραλλαγές οι οποίες ήταν εφικτό να δοκιμαστούν στο δικό μας πειραματικό πλαίσιο είναι:

Βάρη στις Αντιστοιχήσεις Επειδή τα μοντέλα που χρησιμοποιήθηκαν στα πειράματα μας είχαν περιοχές στις οποίες έπρεπε η ευθυγράμμιση να είναι πολύ καλή (π.χ. στη μύτη), χρησιμοποιήσαμε τη λύση με βάρη στις αντιστοιχήσεις.

Δειγματοληψία Σημείων Τα πρόσωπα που είχαμε στη διάθεση μας ήταν αρκετά θορυβώδη και είχαν συχνά περιοχές όπου έλλειπαν δεδομένα τα οποία υπήρχαν στο μοντέλο. Τελικά αποφασίστηκε να χρησιμοποιούνται περιοχές στις οποίες τα δεδομένα ήταν πάντα αξιόπιστα (βλ. ενότητα 3).

Απόρριψη Ασθενών Αντιστοιχήσεων Για να βελτιωθεί η ευρωστία της υλοποίησης σε περιπτώσεις όπου το μοντέλο διαφέρει αρκετά από τα δεδομένα, απορρίπτεται το 20% των αντιστοιχήσεων που συνεισφέρει το μεγαλύτερο σφάλμα [7].

2.5 Εικόνες Βάθους

Αν η ευθυγράμμιση που προέκυψε από τον ICP δεν είναι αρκετά ικανοποιητική, μπορούμε να ευθυγραμμίσουμε τις επιφάνειες που ορίζουν τα σημεία των αντικειμένων [10,11]. Η βασική ιδέα είναι να χρησιμοποιηθούν οι τιμές που αποθηκεύονται στη μνήμη βάθους (z-buffer) κατά τον σχεδιασμό ενός αντικειμένου, τις οποίες θα καλούμε εικόνα βάθους, ως μια διακριτή περιγραφή της επιφάνειας του και να ελαχιστοποιηθεί η διαφορά μεταξύ των καμπυλοτήτων των δύο επιφανειών. Για αυτό το σκοπό χρησιμοποιούνται οι πρώτες παράγωγοι της εικόνας βάθους. Στα πειράματα μας η πλειονότητα των χαρακτηριστικών των προσώπων μπορεί να αποτυπωθεί σε μία εικόνα βάθους αντί για 6 που χρησιμοποιούνται στο [11].

Η μετρική σφάλματος που πρέπει να ελαχιστοποιηθεί δίνεται από την εξίσωση

$$\epsilon_d = \frac{1}{A}\iint_V \left|\frac{\partial d_1(x,y)}{\partial x} - \frac{\partial d_2(x,y)}{\partial x}\right| + \left|\frac{\partial d_1(x,y)}{\partial y} - \frac{\partial d_2(x,y)}{\partial y}\right| \mathrm{d}V \qquad (4)$$

όπου A είναι το εμβαδόν της επικάλυψης των επιφανειών την οποία συμβολίζουμε με V ενώ με $d_1(x,y)$ και $d_2(x,y)$ συμβολίζεται το βάθος της κάθε επιφάνειας στο σημείο (x,y). Επειδή η συνάρτηση βάθους αναπαρίσταται με εικόνες βάθους, χρησιμοποιήθηκε η διακριτή εκδοχή αυτής της μετρικής.

Το πρόβλημα με αυτή τη μετρική είναι ότι στα άκρα του αντικειμένου η παράγωγος παίρνει πολύ μεγάλες τιμές με αποτέλεσμα αυτές οι περιοχές να συνεισφέρουν πολύ στο σφάλμα και να επισκιάζουν τα σφάλματα που προέρχονται από την υπόλοιπη επιφάνεια. Αυτό διορθώνεται αν χρησιμοποιηθεί ένα άνω όριο ώστε να μην επιτρέπονται τιμές παραγώγων μεγαλύτερες από αυτό. Για την ελαχιστοποίηση της μετρικής της εξίσωσης (4) χρησιμοποιήθηκε η εμπλουτισμένη προσομοιωμένη ανόπτηση (enhanced simulated annealing) [12], η οποία είναι μια βελτιωμένη εκδοχή της κλασικής προσομοιωμένης ανόπτησης.

3 Μελέτη Περίπτωσης: Αναγνώριση Προσώπων

Η διαδικασία της ευθυγράμμισης αποτελεί το αρχικό βήμα σε ένα σύστημα αναγνώρισης αντικειμένων και η ακριβής αποτίμηση της ποιότητάς της είναι πολύ δύσκολη γιατί εξαρτάται από το πρόβλημα στο οποίο εφαρμόζεται. Στην περίπτωση των πειραμάτων αναγνώρισης προσώπου χρησιμοποιείται συχνά η ορολογία αρχείο (gallery) και έλεγχος (probe). Συνήθως κάποιες από τις 3Δ εικόνες ανήκουν στο αρχείο και οι υπόλοιπες στον έλεγχο. Το σύστημα αναγνώρισης έχει αντιστοιχήσει σε κάθε εικόνα που ανήκει στο αρχείο έναν άνθρωπο που θα πρέπει να αναγνωρίζει. Όταν παρουσιαστεί ένα πρόσωπο προς αναγνώριση, μια εικόνα ελέγχου, το σύστημα θα πρέπει να βρει με ποια από τις εικόνες αρχείου μοιάζει περισσότερο.

3.1 Το Σύστημα Αναγνώρισης UR3D

Το UR3D[1] χρησιμοποιεί ήδη τους παραπάνω αλγορίθμους και δίνει τη δυνατότητα στο χρήστη να χρησιμοποιήσει οποιοδήποτε συνδυασμό τους. Η προσέγγιση που ακολουθεί το UR3D στην αναγνώριση προσώπου είναι η εξής

1. Ευθυγράμμιση του υπό εξέταση προσώπου με ένα γενικό μοντέλο προσώπου.
2. Ελαστική παραμόρφωση του γενικού μοντέλου με χρήση ανθρωπομετρικών περιορισμών ώστε να προσαρμοστεί στη γεωμετρία του υπό εξέταση προσώπου.
3. Μετατροπή του 3Δ προσώπου σε 2Δ (γεωμετρική εικόνα, geometry image) μέσω γνωστής (u, v) παραμετροποίησης του μοντέλου.
4. Χρήση διάφορων δοκιμασμένων τεχνικών για 2Δ εικόνες όπως τα κυματίδια (wavelets) για τη λύση του απλούστερου αυτού προβλήματος.

3.2 Πειραματικά Αποτελέσματα

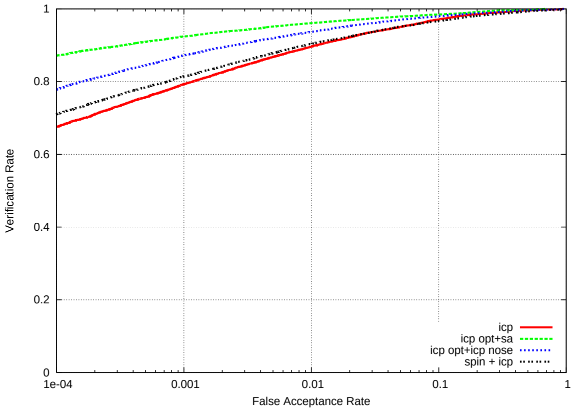
Τα αποτελέσματα που παρουσιάζονται εδώ προέρχονται από τρεις καθιερωμένες μετρήσεις του διαγωνισμού Face Recognition Vendor Test, FRVT 2006. Ο FRVT αποτελείται από μια σειρά δύσκολων πειραμάτων και το σύστημα UR3D συμμετέχει σε ένα πείραμα με τρισδιάστατη μόνο διαθέσιμη πληροφορία. Οι τρεις καθιερωμένες μετρήσεις βασίζονται στο FRGC2, ένα προκαθορισμένο σύνολο 4007 προσώπων με εκφράσεις που διατίθεται για πειραματισμό στους συμμετέχοντες του FRVT. Στις μετρήσεις αυτές ζητείται η μεγιστοποίηση του ποσοστού των true positives (verification rate) για τιμές ποσοστού false positives (false acceptance rate) ίσες με 1% και 0.1%. Οι μετρήσεις χρησιμοποιούν τρία υποσύνολα του FRGC2, αυξανόμενης δυσκολίας, που καλούνται πρώτη, δεύτερη και τρίτη μάσκα.

Για όλες τις μετρήσεις χρησιμοποιήθηκε ένα μοντέλο προσώπου στο οποίο η τρισδιάστατη πληροφορία είναι περισσότερο αξιόπιστη. Συγκεκριμένα, έχουν αφαιρεθεί η περιοχή του στόματος και η περιοχή του μετώπου οι οποίες ήταν ιδιαίτερα αναξιόπιστες. Από τις υπόλοιπες περιοχές, στη μύτη δόθηκε το μεγαλύτερο βάρος, λιγότερο στα μάτια, και ακόμα λιγότερο στα μάγουλα.

Στο σχήμα 2 παρουσιάζονται για οικονομία χώρου μόνο οι καμπύλες ROC για την τρίτη και δυσκολότερη μάσκα για διάφορες στρατηγικές ευθυγράμμισης. Τα αποτελέσματα για τις άλλες δύο μάσκες είναι παρόμοια. Όλες οι καμπύλες βρίσκονται λίγο ψηλότερα στη δεύτερη μάσκα και ακόμα πιο ψηλά στην πρώτη. Αρχικά δοκιμάστηκε μόνο ο ICP. Στη συνέχεια δοκιμάστηκε να αρχικοποιηθεί ο ICP με τα αποτελέσματα των εικόνων περιστροφής (spin + icp στο υπόμνημα). Η βελτίωση της επίδοσης είναι μικρή επειδή στο FRGC2 οι θέσεις των προσώπων περιορίζονται από τις δυνατές κινήσεις του αυχένα. Έτσι συνήθως η εκτέλεση του ICP συγκλίνει σε καλά τοπικά ελάχιστα για τα δεδομένα του FRGC2.

Δοκιμάστηκε επίσης η χρήση δύο εκτελέσεων του ICP. Με την πρώτη οδηγούμαστε σε μία καλή ευθυγράμμιση και με τη δεύτερη ευθυγραμμίζουμε την μύτη θέτοντας αμελητέα βάρη στα μάτια και στα μάγουλα. Η πρώτη ευθυγράμμιση βοηθά στην αποφυγή άσχημων τοπικών ελαχίστων. Αρχικά γινόταν χρησιμοποιώντας

[1] http://www.cbl.uh.edu/UR8D



Σχήμα 2. Οι καμπύλες ROC για τέσσερις στρατηγικές ευθυγράμμισης στην τρίτη μάσκα

Πίνακας 3. Τα αποτελέσματα των υπόλοιπων πειραμάτων με εικόνες βάθους. Ο trimmed icp είναι ο αλγόριθμος ICP με απόρριψη του 20% των αντιστοιχήσεων. Στον πίνακα φαίνεται το verification rate για false acceptance rate ίσο με 0.001 και 0.01

Μέθοδος ευθυγράμμισης	Πρώτη Μάσκα		Δεύτερη Μάσκα		Τρίτη Μάσκα	
	0.001	0.01	0.001	0.01	0.001	0.01
icp + sa	0.9379	0.9713	0.9310	0.9662	0.9233	0.9604
trimmed icp + sa	0.9354	0.9690	0.9288	0.9643	0.9208	0.9590
icp opt + icp nose + sa	0.9347	0.9686	0.9293	0.9649	0.9222	0.9602

την παραλλαγή που απορρίπτει το 20% των χειρότερων αντιστοιχήσεων (βλ. ενότητα 2.4) η οποία, είχε άλλοτε καλύτερα και άλλοτε χειρότερα αποτελέσματα από τον απλό ICP. Έτσι αποφασίστηκε να εκτελείται πρώτα ο απλός ICP και αν το μέσο τετραγωνικό σφάλμα είναι πάνω από ένα όριο να επανεκτελείται με την απόρριψη του 20% των αντιστοιχήσεων. Αυτή η τακτική δηλώνεται ως icp opt. Τελικά η ευθυγράμμιση της μύτης (icp opt + icp nose) είχε αρκετά καλύτερα αποτελέσματα.

Τέλος, δοκιμάστηκαν αρκετές στρατηγικές όπου ο μετασχηματισμός που υπολόγιζε ο ICP δινόταν ως αρχική ευθυγράμμιση στον αλγόριθμο των εικόνων βάθους. Τα αποτελέσματα αυτών των στρατηγικών ήταν πολύ καλύτερα από τις προηγούμενες στρατηγικές. Η καλύτερη στρατηγική μεταξύ αυτών φαίνεται στο σχήμα και χρησιμοποιεί την προαιρετική επανεκτέλεση του ICP με απόρριψη του 20% των αντιστοιχήσεων (icp opt + sa[2] στο υπόμνημα). Τα αποτελέσματα των υπόλοιπων στρατηγικών φαίνονται στον πίνακα 3. Αν και η μέθοδος των εικόνων βάθους έχει πολύ καλύτερα αποτελέσματα από τους υπόλοιπους αλγορίθμους, είναι ιδιαίτερα ευαίσθητη στις αρχικές συνθήκες. Γι΄ αυτό πρέπει πάντα να προηγείται η εκτέλεση κάποιου άλλου αλγορίθμου, όπως ο ICP, που θα φέρει κοντά τα δύο αντικείμενα.

[2] Το sa προέρχεται από τα αρχικά του Simulated Annealing, τη μέθοδο βελτιστοποίησης που χρησιμοποιείται στον αλγόριθμο των εικόνων βάθους

Παρατηρώντας το σχήμα 2 και τον πίνακα 3 συμπεραίνουμε ότι οι μέθοδοι ευθυγράμμισης είναι πολύ σημαντικές στην αποδοτικότητα ενός συστήματος αναγνώρισης προσώπου. Όπως είναι λογικό, υπάρχει μια μικρή πτώση της απόδοσης για πιο δύσκολα προβλήματα. Όμως η σχετική απόδοση του συστήματος για διαφορετικές στρατηγικές ευθυγράμμισης, η οποία δείχνει έμμεσα και την ποιότητα κάθε στρατηγικής, παραμένει αναλλοίωτη για τις τρεις μάσκες.

4 Συμπεράσματα και Μελλοντικές Κατευθύνσεις

Στην παρούσα εργασία αναπτύχθηκαν, συνδυάστηκαν και αποτιμήθηκαν διάφοροι αλγόριθμοι ευθυγράμμισης τρισδιάστατων προσώπων. Σε ένα τυπικό σύστημα, οι αλγόριθμοι αυτοί μπορούν να χρησιμοποιηθούν σε σειρά (πρώτα εικόνες περιστροφής, ύστερα ICP και τέλος εικόνες βάθους) για μεγιστοποίηση της απόδοσης. Τα αποτελέσματα σε ένα πραγματικό πείραμα έδειξαν ότι η στρατηγική ευθυγράμμισης επηρεάζει πολύ την τελική απόδοση. Ο καλύτερος συνδυασμός επιλογών που βρέθηκε για τη συγκεκριμένη περίπτωση ήταν η αρχική εκτέλεση του αλγορίθμου ICP και η προαιρετική επανεκτέλεση του αν το μέσο τετραγωνικό σφάλμα είναι αρκετά μεγάλο. Στη συνέχεια, η ευθυγράμμιση που υπολογίστηκε βελτιώνεται χρησιμοποιώντας εικόνες βάθους.

Μερικά από τα θέματα που προέκυψαν κατά τη διάρκεια της εργασίας και τα οποία χρήζουν της ανάλογης μελέτης είναι τα εξής. Πρώτον, η χρήση άλλων τεχνικών ευθυγράμμισης όπως ο μετασχηματισμός Hough και ο γεωμετρικός κατακερματισμός. Δεύτερον, η χρήση αναλλοίωτων χαρακτηριστικών, όπως οι καμπυλότητες, ώστε οι αλγόριθμοι να εντοπίζουν ευκολότερα σωστές αντιστοιχήσεις. Τέλος, η τροποποίηση της βασικής λύσης ώστε να λαμβάνει υπόψη και τα σφάλματα στα κανονικά διανύσματα. Τα άμεσα σχέδια μας περιλαμβάνουν τη διερεύνηση των παραπάνω κατευθύνσεων για να αποτιμηθεί η ενσωμάτωσή τους στο UR3D.

Αναφορές

1. Horn, B.: Closed form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America **4** (1987) 629–642
2. Johnson, A.: Spin-Images: A Representation for 3-D Surface Matching. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (1997)
3. Gouraud, H.: Continuous shading of curved surfaces. IEEE Transactions on Computers **20** (1971) 623–629
4. Besl, P., McKay, N.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 239–256
5. Turk, G., Levoy, M.: Zippered polygon meshes from range images. In: Proceedings of the SIGGRAPH. (1994)
6. Rusinkiewicz, S., Levoy, M.: Efficient variants of the icp algorithm. In: Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling. (2001)
7. Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P.: The trimmed iterative closest point algorithm. In: Proceedings of the International Conference on Pattern Recognition. Volume 3. (2002) 545–548

8. Sharp, G., Lee, S., Wehe, D.: Invariant features and the registration of rigid bodies. In: IEEE International Conference on Robotics and Automation. (1999) 932–937
9. Zinsser, T., Schmidt, J., Niemann, H.: A refined icp algorithm for robust 3-d correspondence estimation. In: Proceedings of the International Conference on Image Processing. (2003)
10. Vajramushti, N., Kakadiaris, I.A., Theoharis, T., Papaioannou, G.: Efficient 3d object retrieval using depth images. In: MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, New York, NY, USA, ACM Press (2004) 189–196
11. Papaioannou, G., Karabassi, E., Theoharis, T.: Reconstruction of threedimensional objects through matching of their parts. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 114–124
12. Siarry, P., Berthiau, G., Durbin, F., Haussy, J.: Enhanced simulated annealing for globally minimizing functions of many-continuous variables. ACM Transactions on Mathematical Software **23** (1997) 209–228