



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Ηλεκτρονικών Μηχανικών & Μηχανικών Υπολογιστών

Διπλωματική Εργασία

“ Μέθοδοι προβολής και ανακατασκευής κίνησης στον τρισδιάστατο
χώρο ”

Κοντίσης Μιχαήλ

©Χανιά, Σεπτέμβριος 2001

Περιεχόμενα

Εισαγωγικό σημείωμα	1
---------------------	---

Κεφάλαιο 1

Γεωμετρία εικόνας

1.1 Μοντέλα σχηματισμού εικόνας (προοπτική προβολή κλπ)	2
1.1.1 Εισαγωγή	2
1.1.2 Προοπτική προβολή	2
1.1.3 Επίδραση της απόστασης (ως χαρακτηριστικό της προοπτικής προβολής)	4
1.2 Βασικοί μετασχηματισμοί	6
1.2.1 Μετασχηματισμοί μετατόπισης – περιστροφής σημείων	6
1.2.2 Επέκταση σε αντικείμενα	8

Κεφάλαιο 2

Προβολή και ανακατασκευή 3Δ χώρου

2.1 Αντιστοίχιση 3Δ \rightarrow 2Δ (forward perspective mapping)	10
2.1.1 Εισαγωγή	10
2.1.2 Στατική κάμερα – στατικό περιβάλλον	10
2.1.3 Στατική κάμερα – κινούμενο αντικείμενο	12
2.1.4 Κινούμενη κάμερα – στατικό περιβάλλον	14
2.1.5 Κινούμενη κάμερα – κινούμενο αντικείμενο	15
2.1.6 Γενίκευση για αντικείμενα	16
2.1.7 Αντιστοίχιση σε pixel	16
2.2 Αντίστροφη προοπτική προβολή (inverse perspective mapping)	18
2.2.1 Αντιστροφή μετασχηματισμών (για σημεία)	18
2.2.2 Επέκταση για αντικείμενα	19
2.2.3 Παράδειγμα αντιστροφής	19
2.2.4 Αναλυτικότερη περιγραφή του αλγόριθμου αντιστροφής της προοπτικής προβολής (Inverse perspective mapping)	21
2.2.5 Αποτελέσματα και συζήτηση του αλγόριθμου αντιστροφής προοπτικής προβολής (Inverse perspective mapping)	24

2.3	Στερεοψία	25
2.4	Λήψη αντικειμένου από διαφορετικά σημεία	27
2.5	Λήψη αντικειμένου σε διαδοχικές χρονικές στιγμές	29
2.6	Χωρικοί περιορισμοί αντικειμένων	29

Κεφάλαιο 3

Εντοπισμός και παρακολούθηση αντικειμένου

3.1	Εισαγωγή	31
3.2	Μέθοδοι εντοπισμού κινούμενων αντικειμένων	31
3.3	Μέθοδοι παρακολούθησης χαρακτηριστικών (feature based) ..	32
3.3.1	Εισαγωγή	32
3.3.2	Εξαγωγή και ταυτοποίηση χαρακτηριστικών σημείων	32
3.3.3	Εξαγωγή και ταυτοποίηση χαρακτηριστικών γραμμών	33
3.3.4	Χρήση διαφορετικών μεθόδων με βάση τα χαρακτηριστικά (features)...	35
3.4	Μέθοδοι παρακολούθησης βασισμένες σε μοντέλα (model based)	37
3.4.1	Εισαγωγή	37
3.4.2	Δημιουργία 3D μοντέλου	37
3.4.3	Ταυτοποίηση σε δύο διαστάσεις (2D matching)	38
3.4.4	Ταυτοποίηση σε τρεις διαστάσεις (3D matching)	39

Κεφάλαιο 4

Ανάπτυξη και Υλοποίηση αλγορίθμων Δυναμικής Όρασης (Dynamic Vision) (Στατική κάμερα – κινούμενο αντικείμενο)

4.1	Εισαγωγή	40
4.2	Ταυτοποίηση χαρακτηριστικών σημείων (pixel feature matching)	41
4.2.1	Εισαγωγή	41
4.2.2	Περιγραφή αλγόριθμου	41
4.2.3	Αναλυτικότερη περιγραφή αλγόριθμου	45
4.2.4	Αποτελέσματα και συζήτηση	47
4.3	Ταυτοποίηση μοντέλου (model matching)	48
4.3.1	Εισαγωγή	48
4.3.2	Περιγραφή αλγόριθμου	48

4.3.3	Αναλυτικότερη περιγραφή του αλγόριθμου	52
4.4	Επεκτάσεις του αλγόριθμου ταυτοποίησης μοντέλου	54
4.4.1	Αλλαγή μοντέλου	54
4.4.2	Εναλλαγή επιπέδων κίνησης	55
4.4.3	Χρήση μοντέλου μη γραμμικών ακμών	57
4.4.4	Αναλυτικότερη περιγραφή του αλγόριθμου ταυτοποίησης μοντέλου και των επεκτάσεων αυτού	59
4.4.5	Αποτελέσματα και συζήτηση	62
	Μελλοντικές εφαρμογές και επεκτάσεις	65
	ΠΑΡΑΡΤΗΜΑ	
	Λύση εξισώσεων αντιστροφής της προοπτικής προβολής	66
	ΒΙΒΛΙΟΓΡΑΦΙΑ	69

Εισαγωγικό σημείωμα

Ένα ενδιαφέρον πεδίο της μηχανικής όρασης είναι η εξαγωγή συμπερασμάτων για τα αντικείμενα και τον τρόπο που αυτά κινούνται μέσα σε κάποιο συγκεκριμένο περιβάλλον. Η παρούσα εργασία είναι μια αρχική προσέγγιση στο θέμα. Στα πλαίσιά της, παρουσιάζονται μέθοδοι ανακατασκευής του τρισδιάστατου χώρου και παρακολούθησης αντικειμένων από δεδομένα της εικόνας. Ενδεικτικά αναφέρονται η στερεοοπτική όραση (stereopsis), η λήψη του αντικειμένου από διαφορετικά σημεία ή σε διαδοχικές χρονικές στιγμές και η εφαρμογή χωρικών περιορισμών για τα αντικείμενα, ενώ περιλαμβάνεται και η υλοποίηση ενός αλγορίθμου αντιστροφής της προοπτικής προβολής. Όσον αφορά τις μεθόδους παρακολούθησης, παρουσιάζονται και υλοποιούνται δύο διαφορετικές προσεγγίσεις. Η μια αφορά την ταυτοποίηση κάποιων χαρακτηριστικών της εικόνας (feature matching), ενώ η άλλη την ταυτοποίηση με βάση ένα 3D μοντέλο για το υπό παρακολούθηση αντικείμενο (model based). Τέλος προτείνονται βελτιώσεις και επεκτάσεις των αλγορίθμων με χρήση μοντέλων μη γραμμικών ακμών.

Αναλυτικότερα, στο πρώτο κεφάλαιο, παρουσιάζεται η γεωμετρία της εικόνας. Αναφέρονται ο τρόπος που αυτή σχηματίζεται με τη χρήση της προοπτικής προβολής και οι μετασχηματισμοί που είναι δυνατό να περιγράψουν την κίνηση αναλύοντάς την σε μετατόπιση και περιστροφή. Μετά από αυτό, στο κεφάλαιο 2, αναφέρονται οι τρόποι με τους οποίους συνδέεται ο τρισδιάστατος χώρος με την δισδιάστατη εικόνα του καθώς επίσης και οι μέθοδοι που είναι διαθέσιμες για την ανακατασκευή του πρώτου από τα δεδομένα που καταγράφονται στη δεύτερη. Επίσης παρουσιάζεται και η υλοποίηση ενός τέτοιου αλγόριθμου ανακατασκευής. Προχωρώντας στο τρίτο κεφάλαιο, γίνεται αναφορά στις μεθόδους παρακολούθησης αντικειμένων, ενώ στο τέταρτο γίνεται αναλύονται εκτενώς οι αντίστοιχοι αλγόριθμοι που υλοποιήθηκαν στα πλαίσια της διπλωματικής εργασίας. Ειδικότερα, για τον αλγόριθμο που βασίζεται σε 3D μοντέλο, αναφέρονται και δύο επεκτάσεις που υλοποιήθηκαν με την πρώτη από αυτές να αφορά τον τρόπο κίνησης του αντικειμένου και την δεύτερη το σχήμα και τις δυνατότητες γενίκευσης του ίδιου του 3D μοντέλου. Στο τέλος, γίνεται λόγος για τις πιθανές μελλοντικές επεκτάσεις και εφαρμογές.

Κεφάλαιο 1

Γεωμετρία εικόνας

1.1 Μοντέλο σχηματισμού εικόνας

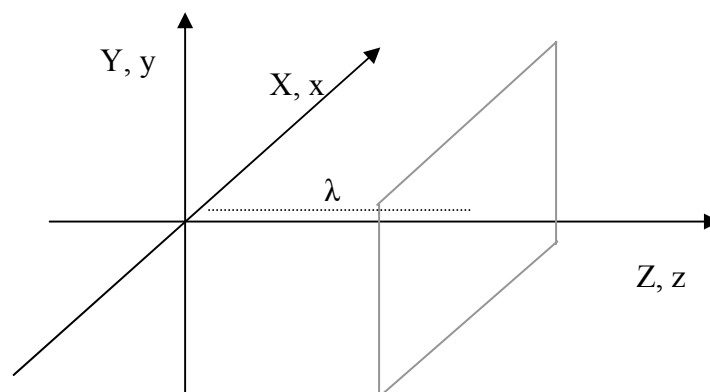
1.1.1 Εισαγωγή

Στην ενότητα που ακολουθεί εξετάζεται ο τρόπος με τον οποίο θεωρούμε ότι σχηματίζεται η εικόνα σε ένα σύστημα μηχανικής όρασης. Η συνήθης πρακτική είναι να αγνοούνται φαινόμενα όπως η διάθλαση, η παραμόρφωση του φακού και να υιοθετείται ένα απλό μαθηματικό μοντέλο με βάση την γεωμετρική οπτική. Υπάρχουν αρκετά τέτοια μοντέλα με κύρια αυτά της παράλληλης και της προοπτικής προβολής. Στις εφαρμογές μηχανικής όρασης η δεύτερη είναι προτιμότερη αφού, όπως θα φανεί και στις επόμενες παραγράφους, υπό την προοπτική προβολή, η εικόνα υπόκειται στην επίδραση της απόστασης, δίνοντας και κάποια «αντίληψη» του βάθους.

1.1.2 Προοπτική Προβολή

Ο προοπτικός μετασχηματισμός προβάλλει σημεία του 3Δ χώρου πάνω σε ένα επίπεδο, παρέχοντας έτσι μια καλή προσέγγιση του τρόπου με τον οποίο σχηματίζεται μια εικόνα παρατηρώντας μια 3Δ σκηνή.

Θεωρώντας την απλή περίπτωση όπου το σύστημα συντεταγμένων του 3Δ χώρου $w(X,Y,Z)$ συμπίπτει με αυτό της κάμερας $c(x,y,z)$,



Σχήμα 1.1 : Μοντέλο προοπτικής προβολής

προκύπτει από την προοπτική προβολή [1] :

$$x = \frac{\lambda X}{\lambda - Z} \quad (\text{εξ 1.1}), \quad y = \frac{\lambda Y}{\lambda - Z} \quad (\text{εξ 1.2})$$

όπου λ η εστιακή απόσταση του φακού,

ή σε μορφή κανονικών (ομογενών) συντεταγμένων :

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\lambda} & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow c_h = P w_h \quad (\text{εξ 1.3})$$

όπου P ο πίνακας του προοπτικού μετασχηματισμού
(perspective transformation matrix)

Για να προκύψουν οι καρτεσιανές συντεταγμένες από τις ομογενείς, θα πρέπει κάθε όρος του διανύσματος να διαιρεθεί με τον τελευταίο. Δηλαδή :

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\lambda} & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ \frac{\lambda - Z}{\lambda} \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda X / (\lambda - Z) \\ \lambda Y / (\lambda - Z) \\ \lambda Z / (\lambda - Z) \\ 1 \end{bmatrix}$$

Η πιο πάνω θεώρηση της προοπτικής προβολής περιγράφει καλύτερα αυτό που συμβαίνει στο φακό και στην φωτοευαίσθητη επιφάνεια του οπτικού αισθητήρα. Το είδωλο πράγματι σχηματίζεται στο επίπεδο της εικόνας αντεστραμμένο. Η εικόνα, όμως, που λαμβάνουμε ως έξοδο από την κάμερα, έχει υποστεί μια επεξεργασία από τα ηλεκτρονικά της συσκευής με αποτέλεσμα το είδωλο του αντικειμένου να

εμφανίζεται κανονικά (μη αντεστραμμένο). Έχοντας αυτά υπ' όψιν, οι εξισώσεις που προσεγγίζουν καλύτερα την εικόνα που τελικά έρχεται προς επεξεργασία, τμηματοποίηση και ερμηνεία γενικότερα, είναι οι :

$$x = \lambda X/Z \quad (\text{εξ 1.4}), \quad y = \lambda Y/Z \quad (\text{εξ 1.5})$$

Ο πίνακας P του προοπτικού μετασχηματισμού που αντιστοιχεί σ' αυτή την περίπτωση είναι :

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\lambda} & 0 \end{bmatrix}$$

Οι παραπάνω τύποι, δίνουν καλές προσεγγίσεις όταν $Z \rightarrow \infty$ ή πρακτικά όταν $Z \gg \lambda$.

1.1.3 Επίδραση της απόστασης στην εικόνα.

Θεωρώντας τις εξισώσεις της προοπτικής προβολής σαν συναρτήσεις, ως προς Z, των αντίστοιχων συντεταγμένων (x,y) του επιπέδου της εικόνας παρατηρούμε ότι αυτές είναι της μορφής $x = F(Z) = K/Z$ και $y = F(Z) = L/Z$ όπου K, L σταθερές.

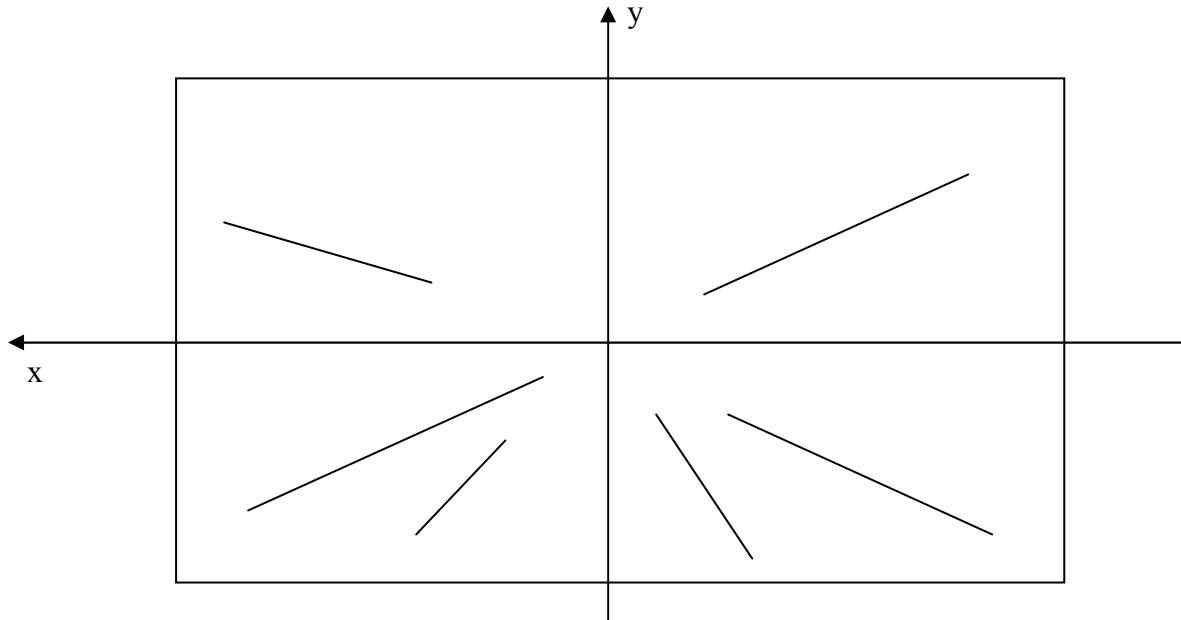
Είναι εμφανές ότι όσο το Z αυξάνει τα x και y τείνουν στο μηδέν. Αυτό σημαίνει πως ένα σημείο που απομακρύνεται από την κάμερα διατηρώντας ίδιες τις X και Y συντεταγμένες του στο χώρο, θα απεικονίζεται συνεχώς πλησιέστερα στο σημείο (0,0) του επιπέδου της εικόνας. Δηλαδή οι ευθείες που είναι κάθετες στο επίπεδο της εικόνας προβάλλονται σ' αυτό ως ευθείες που, προεκτεινόμενες, θα τέμνονταν στο (0,0), όπως φαίνεται στο Σχήμα 1.2. Οι ευθείες αυτές θα μπορούσαν να ανήκουν σε λωρίδες δρόμου, ακμές αντικειμένων ή στοιχείων του περιβάλλοντος όπως τοίχοι, δάπεδα και λοιπά. Το σημείο αυτό ονομάζεται focus of expansion (FOE) ή focus of contraction (FOC), ανάλογα με το αν θεωρούμε ότι οι γραμμές ξεκινούν ή καταλήγουν σ' αυτό αντίστοιχα. Η πρώτη περίπτωση εμφανίζεται όταν η κάμερα κινείται «εισερχόμενη» στο χώρο ενώ η δεύτερη όταν «εξέρχεται» από αυτόν. Θα πρέπει επίσης να τονιστεί ότι το FOE ή το FOC δεν ταυτίζονται πάντα με το (0,0). Αυτό συμβαίνει μόνο όταν η κάμερα δεν έχει περιστραφεί. Σε κάθε άλλη περίπτωση, η θέση του FOE (ή του FOC) εξαρτάται από την διεύθυνση της μετατόπισής της.

Επεκτείνοντας την ιδιότητα αυτή της προοπτικής προβολής για ομάδες σημείων που ανήκουν σε αντικείμενα, προκύπτει ότι όσο αυτά απομακρύνονται από την κάμερα, κινούμενα πάντα πάνω σε ευθείες κάθετες στο επίπεδο της κάμερας, εμφανίζονται να μετατοπίζονται προς το (0,0) συνεχώς σμικρυνόμενα.

Για να γίνει η σμίκρυνση κατανοητή, ας θεωρήσουμε το παρακάτω παράδειγμα :

Έστω ότι μια ράβδος παράλληλη προς τον X άξονα του επιπέδου της εικόνας, μετατοπίζεται σε μια διεύθυνση παράλληλη προς τον Z άξονα. Ας θεωρήσουμε τώρα ότι, αρχικά, τα σημεία που αποτελούν τα άκρα μιας από τις ακμές της ράβδου είναι τα

$(X1, Y, Z1)$ και $(X2, Y, Z2)$ ενώ, στην επόμενη χρονική στιγμή αυτά έχουν μετατοπιστεί για να βρεθούν, αντίστοιχα, στις θέσεις $(X3, Y, Z3)$ και $(X4, Y, Z4)$.



Σχήμα 1.2 : Απεικόνιση παράλληλων ευθειών του 3Δ χώρου που είναι κάθετες στο επίπεδο της εικόνας

Κάτω από την προοπτική προβολή τα σημεία αυτά αντιστοιχίζονται στα $(x1,y)$, $(x2,y)$, $(x3,y)$, $(x4,y)$ του επιπέδου της εικόνας,

όπου $x_i = \lambda X_i / Z_i$, $y = \lambda Y / Z_i$, $i = 1, 2, 3, 4$.

Η τεταγμένη των σημείων είναι κοινή αφού η ράβδος είναι παράλληλη στον x άξονα.

Ξεκινώντας από το γεγονός ότι το φυσικό μέγεθος της ράβδου παραμένει σταθερό,

$$|X1 - X2| \equiv |X3 - X4| \quad (\text{αφού το } Y \text{ είναι κοινό})$$

και χρησιμοποιώντας τις σχέσεις που προκύπτουν από την προοπτική προβολή

$$\frac{Z1}{\lambda} |x1 - x2| = \frac{Z2}{\lambda} |x3 - x4| \Rightarrow$$

$$\frac{|x1 - x2|}{|x3 - x4|} = \frac{Z2}{Z1}$$

προκύπτει ότι ο λόγος των μηκών των ακμών, όπως αυτές απεικονίζονται, είναι αντιστρόφως ανάλογος του λόγου των αντίστοιχων αποστάσεων από το επίπεδο της εικόνας. Έτσι η ακμή που βρίσκεται πλησιέστερα θα εμφανίζεται μεγαλύτερη.

1.2 Βασικοί Μετασχηματισμοί

1.2.1 Μετασχηματισμοί μετατόπισης – περιστροφής σημείων

Σε πολλές εφαρμογές στην επεξεργασία εικόνας εμφανίζεται η ανάγκη να αντιμετωπιστούν προβλήματα όπως η περιστροφή (rotation), η μετατόπιση (translation) και η κλιμάκωση (scaling). Παρακάτω παρουσιάζονται οι τρόποι με τους οποίους επιχειρείται η αντιμετώπισή τους [1].

Μετατόπιση (translation)

Έστω ότι ένα σημείο που ανήκει στο κινούμενο αντικείμενο μετακινείται από ένα σημείο με συντεταγμένες $w_1 (X_1 Y_1 Z_1 I)^T$ κατά X_0 , Y_0 και Z_0 ως προς τους αντίστοιχους άξονες. Η νέα θέση του $w_2 (X_2 Y_2 Z_2 I)^T$ συνδέεται με την αρχική με την σχέση :

$$w_2 = T w_1$$

$$\text{όπου } T = \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ ο πίνακας μετατόπισης}$$

Κλιμάκωση (scaling)

Η κλιμάκωση κατά S_x , S_y , S_z κατά μήκος των αντίστοιχων αξόνων X , Y , Z , δίνεται από τον παρακάτω πίνακα μετασχηματισμού :

$$S = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Περιστροφή (rotation)

Η απλούστερη μορφή αυτού του μετασχηματισμού είναι η περιστροφή ενός σημείου του χώρου γύρω από κάποιον από τους άξονες του συστήματος συντεταγμένων.

Η περιστροφή ενός σημείου γύρω από τον Z άξονα κατά γωνία θ , δίνεται από τον μετασχηματισμό :

$$R_{\theta} = \begin{bmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Αντίστοιχα γύρω από τον X άξονα κατά γωνία α

$$R_{\alpha} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

και γύρω από τον Y άξονα κατά γωνία β

$$R_{\beta} = \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Οι γωνίες θ , α , β μετρώνται θετικά κατά την φορά των δεικτών του ρολογιού (clockwise) καθώς κοιτάζουμε την αρχή του συστήματος συντεταγμένων από μια θέση πάνω στον +Z, +X, +Y άξονα αντίστοιχα.

Κάπως πολυπλοκότερη διαδικασία είναι η περιστροφή σημείου γύρω από κάποιο άλλο, αυθαίρετο σημείο του χώρου (X_0 , Y_0 , Z_0). Αυτό θα απαιτούσε τρεις μετασχηματισμούς :

ο πρώτος μετατοπίζει το αυθαίρετο σημείο στην αρχή των αξόνων

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

ο δεύτερος εφαρμόζει την περιστροφή

$$R = R_\alpha R_\theta R_\beta \quad (\text{ή με οποιαδήποτε άλλη σειρά εκτελούνται οι περιστροφές})$$

και ο τρίτος μετατοπίζει το σημείο πάλι πίσω στην αρχική του θέση

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Έτσι η περιστροφή ενός σημείου γύρω από ένα άλλο τυχαίο (X_0, Y_0, Z_0) επιτυγχάνεται με τον πίνακα :

$$R_p = T_2 R T_1 \quad (\text{εξ 1.6}).$$

1.2.2 Μετασχηματισμοί για συμπαγή (rigid) αντικείμενα

Οι μετασχηματισμοί που παρουσιάστηκαν στην προηγούμενη παράγραφο, αφορούν σημεία του χώρου. Η επέκτασή τους για αντικείμενα είναι εφικτή αν αυτά θεωρηθούν ως πεπερασμένα σύνολα σημείων. Για την αναπαράσταση των αντικειμένων είναι βολικό να θεωρήσουμε πίνακα του οποίου οι στήλες είναι διανύσματα, που περιέχουν τις συντεταγμένες των σημείων του τρισδιάστατου χώρου που αποτελούν το κάθε αντικείμενο. Ένας τέτοιος πίνακας, $Object_w$, που περιγράφει αντικείμενο αποτελούμενο από N σημεία, φαίνεται παρακάτω :

$$Object_w = \begin{bmatrix} X_0 & X_1 & \dots & X_{N-1} \\ Y_0 & Y_1 & \dots & Y_{N-1} \\ Z_0 & Z_1 & \dots & Z_{N-1} \\ 1 & 1 & \dots & 1 \end{bmatrix} \quad \text{στον 3Δ χώρο.}$$

Κατά αναλογία, ο πίνακας $Object_c$, που περιγράφει το αντικείμενο στο επίπεδο της εικόνας, έχει ως εξής :

$$Object_c = \begin{bmatrix} x_0 & x_1 & \cdots & x_{N-1} \\ y_0 & y_1 & \cdots & y_{N-1} \\ z_0 & z_1 & \cdots & z_{N-1} \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Εκμεταλλευόμενοι την συνθήκη της ακαμψίας (rigidity) του αντικειμένου, προκύπτει ότι οποιονδήποτε μετασχηματισμό προοπτικής, μετατόπισης ή περιστροφής και αν υποστεί το αντικείμενο, αυτός θα είναι κοινός για όλα τα σημεία που το απαρτίζουν. Μεταφέροντας την παραπάνω διαπίστωση στο συμβολισμό των πινάκων προκύπτει :

για την μετατόπιση

$$Object2_w = T \cdot Object1_w,$$

για την περιστροφή γύρω από τον άξονα X

$$Object2_w = R_a Object1_w,$$

για την περιστροφή γύρω από τον άξονα Y

$$Object2_w = R_\beta Object1_w,$$

για την περιστροφή γύρω από τον άξονα Z

$$Object2_w = R_\theta Object1_w$$

για την περιστροφή γύρω από σημείο p του 3Δ χώρου

$$Object2_w = R_p Object1_w$$

όπου $Object2_w$ ο πίνακας που περιέχει τις συντεταγμένες των σημείων του αντικειμένου στην νέα θέση, ενώ οι πίνακες T και R είναι αυτοί που ορίστηκαν στην ενότητα 1.2.1 .

Κατά αντιστοιχία το είδωλο ενός μετατοπισμένου και περιστρεμμένου αντικειμένου στην εικόνα θα δίνεται από τον πίνακα :

$$Object_c = P R T Object_w \quad (\text{εξ } 1.7)$$

Κεφάλαιο 2

Προβολή και ανακατασκευή 3Δ χώρου

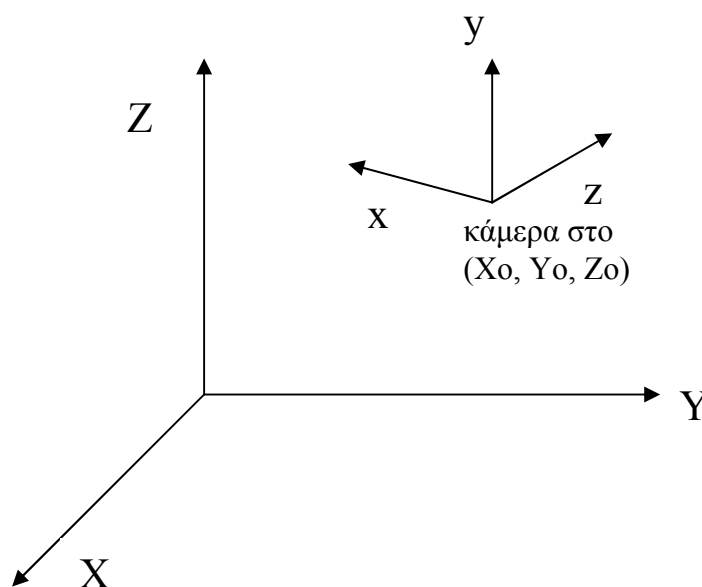
2.1 Αντιστοίχιση 3Δ → 2Δ (forward perspective mapping)

2.1.1 Εισαγωγή

Έχοντας υπόψη τους μετασχηματισμούς του προηγούμενου κεφαλαίου, ας προσπαθήσουμε να προσεγγίσουμε τον τρόπο με τον οποίο αυτοί χρησιμοποιούνται στις τέσσερις πιθανές διαμορφώσεις του συστήματος κάμερας – περιβάλλοντος (αντικειμένου).

2.1.2 Στατική κάμερα – στατικό αντικείμενο

Στην προηγούμενη ανάλυση, θεωρήσαμε ότι τα συστήματα συντεταγμένων της κάμερας και του 3Δ χώρου συμπίπτουν. Ας δούμε τώρα την γενικότερη περίπτωση, όπου η κάμερα είναι τοποθετημένη σε κάποιο σημείο του χώρου (X_0, Y_0, Z_0) έχοντας παράλληλα στραφεί προς κάποια κατεύθυνση, ώστε να καλύπτει καλύτερα τον συγκεκριμένο χώρο. Η κατεύθυνση μπορεί να αναλυθεί σε μια περιστροφή γύρω από τον Z άξονα κατά γωνία θ και σε μία γύρω από τον X κατά γωνία α . Η θετική φορά διαγραφής των γωνιών αυτών είναι αντίθετη με αυτή των δεικτών του ρολογιού (anti – clockwise). Αυτό απορρέει από την προηγούμενη σύμβαση, σύμφωνα με την οποία, θετικές θεωρούνται οι γωνίες όταν τα σημεία του χώρου περιστρέφονται κατά την φορά των δεικτών του ρολογιού (clockwise). Μια περιστροφή των σημείων του χώρου κατά την φορά των δεικτών του ρολογιού (clockwise), αντιστοιχεί σε περιστροφή της κάμερας κατά την αντίθετη έννοια.



Σχήμα 2.1 : Συστήματα συντεταγμένων χώρου και κάμερας

Η μετατόπιση επιτυγχάνεται με την χρήση του πίνακα μετασχηματισμού :

$$T = \begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Μετά την εφαρμογή του παραπάνω μετασχηματισμού, το σημείο (X_0, Y_0, Z_0) βρίσκεται στην αρχή των αξόνων του συστήματος συντεταγμένων της κάμερας. Η περιστροφή περιγράφεται από τον πίνακα :

$$R = R_\alpha R_\theta$$

όπου R_α, R_θ οι πίνακες που αναφέρονται στην προηγούμενη ενότητα. Δηλαδή :

$$R = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta \cos\alpha & \cos\alpha \cos\theta & \sin\alpha & 0 \\ \sin\theta \cos\alpha & -\cos\theta \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Τελικά ένα σημείο του 3Δ χώρου με ομογενείς συντεταγμένες $w(X, Y, Z, 1)^T$ αντιστοιχίζεται στο σημείο $c(x, y, z, 1)^T$ του επιπέδου της κάμερας σύμφωνα με την σχέση :

$$c = PRTw \quad (\text{εξ 2.1})$$

όπου P ο πίνακας του προοπτικού μετασχηματισμού.

Όπως είδαμε και στην ενότητα που αναφέρεται στην προοπτική προβολή, υπάρχουν τουλάχιστον δύο επιλογές ως προς τον πίνακα P .

Εκτελώντας τους πολλαπλασιασμούς των πινάκων για κάθε διαφορετικό P και μετατρέποντας σε καρτεσιανές συντεταγμένες, διαιρώντας το πρώτο και το δεύτερο στοιχείο του c με το τέταρτο, προκύπτει για τα σημεία (x, y) της εικόνας [1]:

$$\text{για } P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\lambda} & 1 \end{bmatrix}$$

$$x = \lambda \frac{(X - X_0) \cos \theta + (Y - Y_0) \sin \theta}{-(X - X_0) \sin \theta \cos \alpha + (Y - Y_0) \cos \theta \sin \alpha - (Z - Z_0) \cos \alpha + \lambda}$$

(εξ 2.2)

$$y = \lambda \frac{-(X - X_0) \sin \theta \cos \alpha + (Y - Y_0) \cos \theta \cos \alpha + (Z - Z_0) \sin \alpha}{-(X - X_0) \sin \theta \cos \alpha + (Y - Y_0) \cos \theta \sin \alpha - (Z - Z_0) \cos \alpha + \lambda}$$

(εξ 2.3)

ενώ για

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\lambda} & 0 \end{bmatrix}$$

$$x = \lambda \frac{(X - X_0) \cos \theta + (Y - Y_0) \sin \theta}{(X - X_0) \sin \theta \cos \alpha - (Y - Y_0) \cos \theta \sin \alpha + (Z - Z_0) \cos \alpha}$$

(εξ 2.4)

$$y = \lambda \frac{-(X - X_0) \sin \theta \cos \alpha + (Y - Y_0) \cos \theta \cos \alpha + (Z - Z_0) \sin \alpha}{(X - X_0) \sin \theta \cos \alpha - (Y - Y_0) \cos \theta \sin \alpha + (Z - Z_0) \cos \alpha}$$

(εξ 2.5)

2.1.3 Στατική κάμερα – κινούμενο αντικείμενο

Έστω ένα συμπαγές αντικείμενο το οποίο μετακινείται από μία θέση του 3Δ χώρου σε κάποια άλλη. Τότε κάθε ένα σημείο του κινούμενου αντικειμένου με αρχικές συντεταγμένες $w_1 (X_1 Y_1 Z_1 I)^T$ μετατοπίζεται σε κάποιο $w_2 (X_2 Y_2 Z_2 I)^T$. Αυτό που είναι ενδιαφέρον να βρεθεί, είναι η σχέση που συνδέει τις συντεταγμένες των ειδώλων των δύο αυτών σημείων πάνω στο επίπεδο της εικόνας. Αν θεωρήσουμε τα σημεία $c_1(x_1 y_1 z_1 I)^T$, $c_2(x_2 y_2 z_2 I)^T$ ως τις προβολές των w_1 , w_2 αντίστοιχα, τότε:

$$\text{αρχικά} \quad c_1 = P w_1 \quad (\text{εξ 2.6})$$

$$\text{τελικά} \quad c_2 = P w_2 = P R S T w_1 \quad (\text{εξ 2.7})$$

αφαιρώντας κατά μέλη τις (εξ 2.6), (εξ 2.7):

$$c_2 - c_1 = P (R S T - I) w_1$$

$$c_2 = c_1 + P (R S T - I) w_1 \quad (\text{εξ 2.8})$$

όπου P ο πίνακας του προοπτικού μετασχηματισμού,

$$T = \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} : \text{πίνακας μετατόπισης}$$

X_0, Y_0, Z_0 οι τιμές της γραμμικής μετατόπισης του αντικειμένου κατά τους αντίστοιχους άξονες,

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$S = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} : \text{πίνακας κλιμάκωσης,}$$

όπου μια συνήθης επιλογή είναι $S_x=S_y=I$ και $S_z=Z_1/Z_2$,

$$R = R_{\beta_0} R_{\alpha_0} R_{\theta_0}$$

όπως αυτοί ορίζονται στην ενότητα των βασικών μετασχηματισμών

Ιδιαίτερης προσοχής χρήζει η σειρά με την οποία εκτελούνται οι μετασχηματισμοί της μετατόπισης αλλά κυρίως της περιστροφής. Οι πίνακες $R_{\beta_0}, R_{\alpha_0}, R_{\theta_0}$, γενικώς, δεν αντιμετατίθενται, με αποτέλεσμα να έχει σημασία ο τρόπος με

τον οποίο θεωρούμε ότι αναλύεται η κίνηση (π.χ. αν το αντικείμενο περιστρέφεται πρώτα γύρω από τον Z και μετά γύρω από τον X και τον Y άξονα ή όχι).

2.1.4 Κινούμενη κάμερα σε σταθερό περιβάλλον

Σ' αυτή την περίπτωση ισχύουν τα ίδια με τα προαναφερθέντα μόνο που οι τιμές της γραμμικής μετατόπισης και της περιστροφής λαμβάνονται κατά την αντίθετη έννοια.

Έτσι, ο πίνακας μετατόπισης είναι :

$$T = \begin{bmatrix} 1 & 0 & 0 & -Xc \\ 0 & 1 & 0 & -Yc \\ 0 & 0 & 1 & -Zc \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Xc , Yc , Zc οι τιμές της γραμμικής μετατόπισης της κάμερας κατά τους αντίστοιχους άξονες

και οι πίνακες περιστροφής :

$$R_\theta = \begin{bmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad R_\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R_\beta = \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ενώ οι γωνίες θ , α και β θεωρούνται θετικές κατά την αντίθετη φορά των δεικτών του ρολογιού (anti-clockwise), αφού από σύμβαση έχουμε θεωρήσει θετικές τις γωνίες που διαγράφουν τα σημεία του χώρου περιστρεφόμενα κατά την φορά των δεικτών του ρολογιού (clockwise). Η περιστροφή της κάμερας κατά οποιονδήποτε τρόπο ισοδυναμεί με περιστροφή των σημείων του χώρου κατά την αντίστροφη φορά.

Η παραπάνω ανάλυση προσεγγίζει ικανοποιητικά την πραγματικότητα, όταν πρόκειται για σημεία του χώρου που βρίσκονται κοντά στην κάμερα. Στην περίπτωση των απομακρυσμένων σημείων θα ήταν δυνατή η απλοποίηση του όλου

μετασχηματισμού, με την απαλοιφή του πίνακα μετατόπισης. Η προσέγγιση αυτή εδράζεται στην παραδοχή, ότι για σημεία που απέχουν μεγάλη απόσταση από την κάμερα, τυπικά πάνω από ένα (1) χιλιόμετρο, η μετατόπιση της κάμερας δεν έχει ικανή επίδραση στον τρόπο που αυτά απεικονίζονται. Η μόνη επίδραση έχει να κάνει με την περιστροφή της κάμερας. Έτσι προκύπτει για τον πίνακα μετατόπισης :

$$T = I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ενώ οι πίνακες περιστροφής παραμένουν ως έχουν.

2.1.5 Κινούμενη κάμερα -- κινούμενο αντικείμενο

Στην περίπτωση αυτή οι τιμές της γραμμικής μετατόπισης και της περιστροφής λαμβάνονται σχετικά :

Έτσι :

$$T = \begin{bmatrix} 1 & 0 & 0 & X_o - X_c \\ 0 & 1 & 0 & Y_o - Y_c \\ 0 & 0 & 1 & Z_o - Z_c \\ 0 & 0 & 0 & 1 \end{bmatrix} : \text{πίνακας μετατόπισης}$$

όπου X_c, Y_c, Z_c οι τιμές της γραμμικής μετατόπισης της κάμερας κατά τους αντίστοιχους άξονες και X_o, Y_o, Z_o οι τιμές της γραμμικής μετατόπισης του αντικειμένου κατά τους αντίστοιχους άξονες.

ενώ οι πίνακες περιστροφής είναι οι $R_\theta, R_\alpha, R_\beta$ όπως ορίστηκαν προηγουμένως, μόνο που για τις γωνίες ισχύει : $\theta = \theta_o + \theta_c, \alpha = \alpha_o + \alpha_c, \beta = \beta_o + \beta_c$, αντίστοιχα.

Οι γωνίες $\theta, \alpha, \beta, \theta_o, \alpha_o$ και β_o θεωρούνται θετικές κατά την φορά των δεικτών του ρολογιού (clockwise), ενώ οι $\theta_c, \alpha_c, \beta_c$ κατά την αντίθετη φορά (anti - clockwise). Αν θεωρήσουμε κοινό σύστημα αναφοράς για τις φορές των γωνιών, τότε ισχύει :

$$\theta = \theta_o - \theta_c, \quad \alpha = \alpha_o - \alpha_c, \quad \beta = \beta_o - \beta_c.$$

2.1.6 Γενίκευση για αντικείμενα

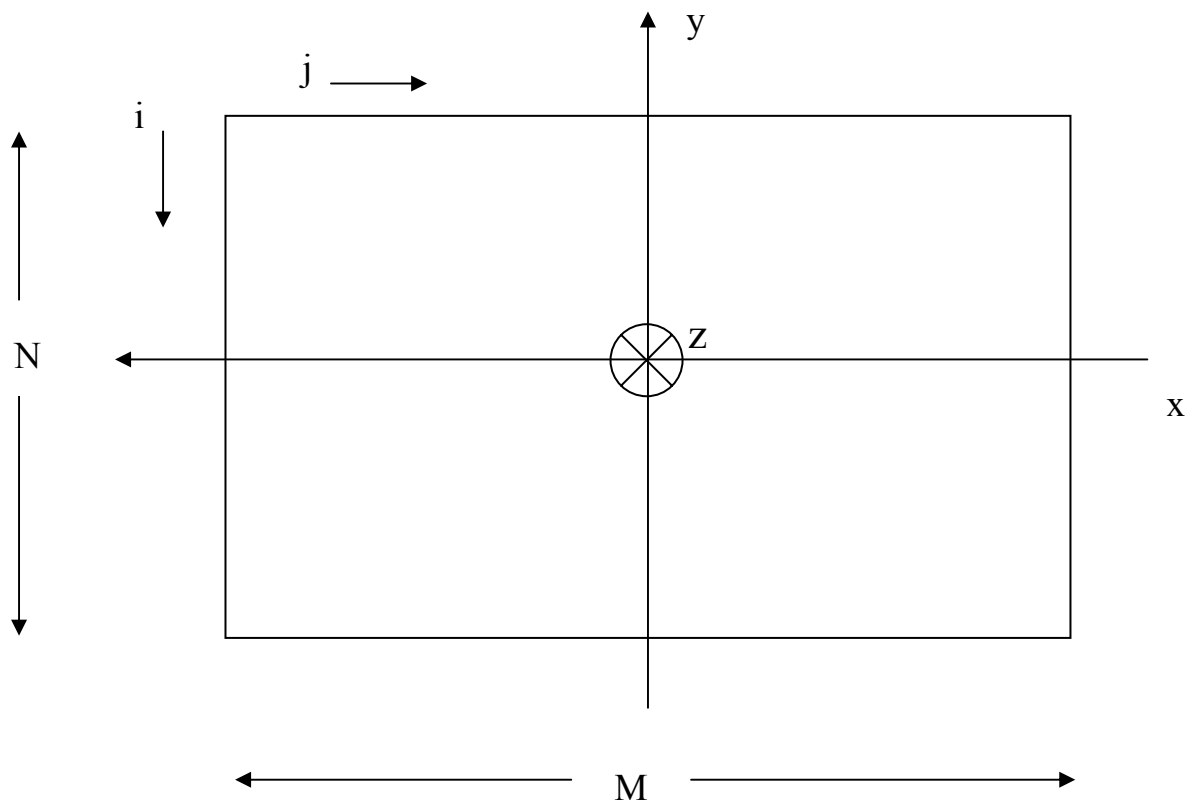
Όλοι οι προηγούμενοι μετασχηματισμοί έχουν εφαρμογή και στην περίπτωση των συμπαγών αντικειμένων αρκεί αυτοί να επιδρούν πάνω στους πίνακες που τα περιγράφουν. Δηλαδή σε κάθε περίπτωση :

$$Object_C = P R T Object_W \quad (\text{εξ 2.9})$$

όπου P ο πίνακας του προοπτικού μετασχηματισμού ενώ T και R οι πίνακες μετατόπισης και περιστροφής αντίστοιχα.

2.1.7 Αντιστοιχία pixel και συντεταγμένων εικόνας

Μέχρι στιγμής έχουμε δει πως συνδέονται τα σημεία του 3Δ χώρου με αυτά του 2Δ επιπέδου της εικόνας. Απομένει ένα ακόμη βήμα για να μπορεί να χρησιμοποιηθεί αυτή η αντιστοίχιση στην πράξη. Θα πρέπει να ληφθεί υπόψη ότι τα δεδομένα τα οποία θα κληθεί το οποιοδήποτε σύστημα να επεξεργαστεί έχουν ψηφιοποιηθεί. Η μορφή που θα έχουν είναι αυτή ενός πίνακα $N \times M$, όπως στο παρακάτω σχήμα.



Σχήμα 2.2 : Σχέση pixel και συντεταγμένων του επιπέδου της εικόνας

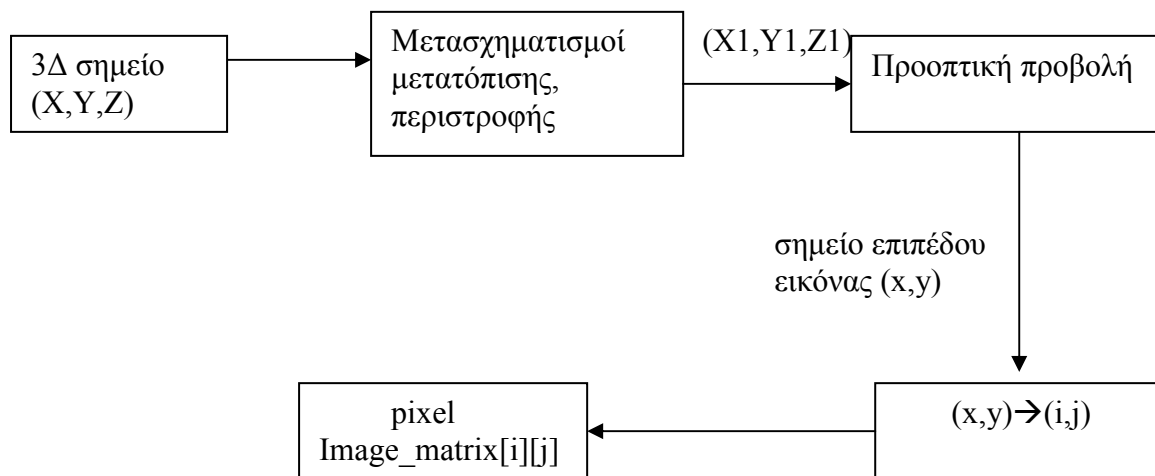
Οι συντεταγμένες (i, j) κάθε pixel αντιστοιχούν σε σημεία (x,y) του επιπέδου της εικόνας σύμφωνα με τις σχέσεις [2]:

$$x = j - \frac{M-1}{2} \quad (\text{εξ 2.10}), \quad y = -i + \frac{N-1}{2} \quad (\text{εξ 2.11})$$

$$j = x + \frac{M-1}{2} \quad (\text{εξ 2.12}), \quad i = -y + \frac{N-1}{2} \quad (\text{εξ 2.13})$$

Συνυπολογίζοντας ότι τα x και y έχουν φυσική διάσταση μήκους, ενώ τα i, j είναι αδιάστατα, θα πρέπει να πολλαπλασιαστούν τα τελευταία επί έναν συντελεστή μήκους ανά pixel S_i, S_j αντίστοιχα. Οι συντελεστές αυτοί εξαρτώνται από το μέγεθος της φωτοευαίσθητης επιφάνειας του οπτικού αισθητήρα.

Έχοντας ολοκληρώσει και αυτή την τελευταία αντιστοίχιση μπορούμε να δώσουμε την όλη διαδικασία της μετάβασης από τον 3Δ χώρο στον 2Δ πίνακα των pixel, σε ένα δομικό διάγραμμα (block diagram) της μορφής :



Σχήμα 2.3 : Δομικό διάγραμμα δημιουργίας της ψηφιακής εικόνας.

2.2 Αντίστροφη προοπτική προβολή (inverse perspective mapping)

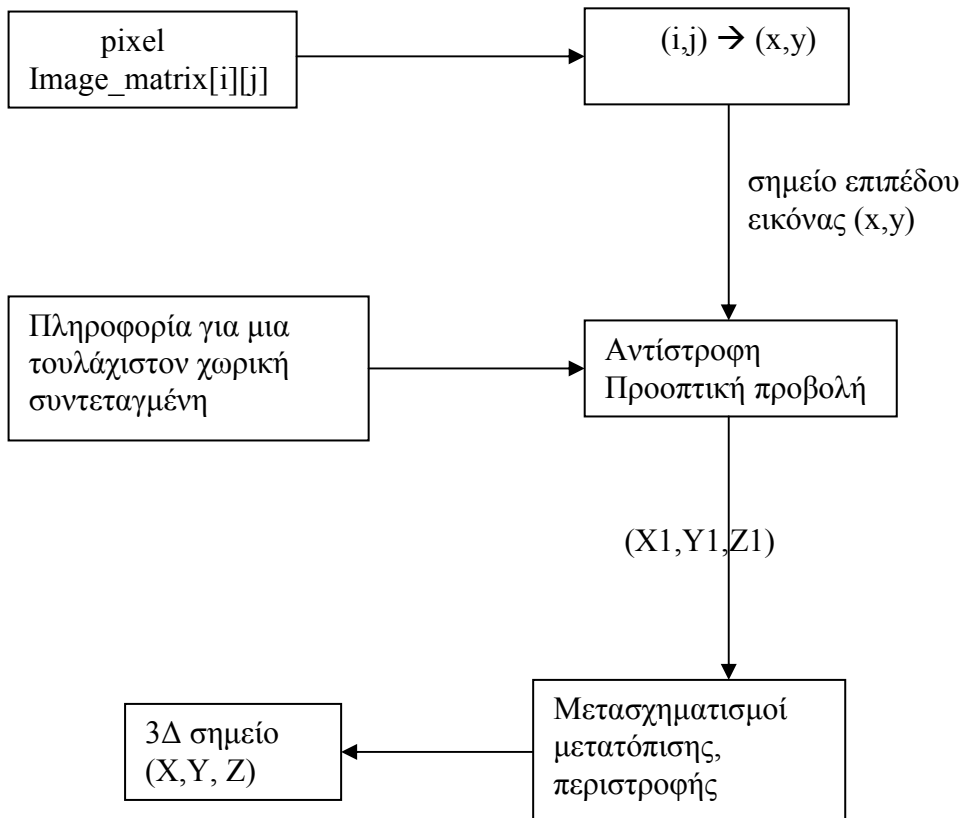
2.2.1 Αντιστροφή μετασχηματισμών (για σημεία)

Οι συντεταγμένες του επιπέδου της εικόνας (x,y) είναι δυνατόν να αντιστοιχηθούν με αυτές του χώρου (X, Y, Z) , αν αντιστραφούν οι πίνακες μετατόπισης, περιστροφής και κλιμάκωσης (με την προϋπόθεση ότι ο πίνακας $P R S T$ είναι αντιστρέψιμος).

$$c = P R S T w \quad (\text{εξ 2.14}) \Leftrightarrow w = (P R S T)^{-1} c \quad (\text{εξ 2.15})$$

Η διαδικασία φαίνεται, σε μορφή δομικού διαγράμματος, στο Σχήμα 2.5.

Όπως θα μπορούσε να υποψιαστεί κανείς, αυτή η αντιστοίχιση δεν είναι μονοσήμαντη. Πράγματι, σε κάθε σημείο $c (x,y)$ της εικόνας αντιστοιχούν τα σημεία της ευθείας του χώρου που διέρχεται από το (x,y) και το κέντρο της προβολής. Μια ένα προς ένα αντιστοίχιση θα απαιτούσε την γνώση μιας από τις χωρικές συντεταγμένες κάθε προβαλλόμενου σημείου (X, Y, Z) . Γι' αυτό, στις περισσότερες εφαρμογές, τα προς εντοπισμό αντικείμενα κινούνται πάνω σε συγκεκριμένο, γνωστό επίπεδο ή άλλο γεωμετρικό χώρο, ενώ είναι γνωστή και η δομή τους ώστε να προκύπτουν χωρικοί περιορισμοί για τα σημεία που τα αποτελούν.



Σχήμα 2.5 : Διάγραμμα της μετάβασης από τον πίνακα των pixel στον 3Δ χώρο

2.2.2 Επέκταση για αντικείμενα

Παρόλο ότι οι εξισώσεις αυτές αναφέρονται σε σημεία, μπορούν να γενικευτούν για την περίπτωση των αντικειμένων αν τα τελευταία θεωρηθούν ότι περιγράφονται ή προσεγγίζονται από ένα πλήθος σημείων. Στις εφαρμογές το πλήθος αυτό είναι πεπερασμένο και αποτελείται συχνά από κάποια χαρακτηριστικά σημεία του αντικειμένου. Αν υποθεθεί ότι το αντικείμενο αναπαρίσταται από N σημεία, τότε σε μορφή πίνακα θα ήταν :

$$Object_w = \begin{bmatrix} X_0 & X_1 & \dots & X_{N-1} \\ Y_0 & Y_1 & \dots & Y_{N-1} \\ Z_0 & Z_1 & \dots & Z_{N-1} \\ 1 & 1 & \dots & 1 \end{bmatrix} \text{ στον 3Δ χώρο}$$

και

$$Object_c = \begin{bmatrix} x_0 & x_1 & \dots & x_{N-1} \\ y_0 & y_1 & \dots & y_{N-1} \\ z_0 & z_1 & \dots & z_{N-1} \\ 1 & 1 & \dots & 1 \end{bmatrix} \text{ στο επίπεδο της εικόνας.}$$

Για τους δύο αυτούς πίνακες ισχύει :

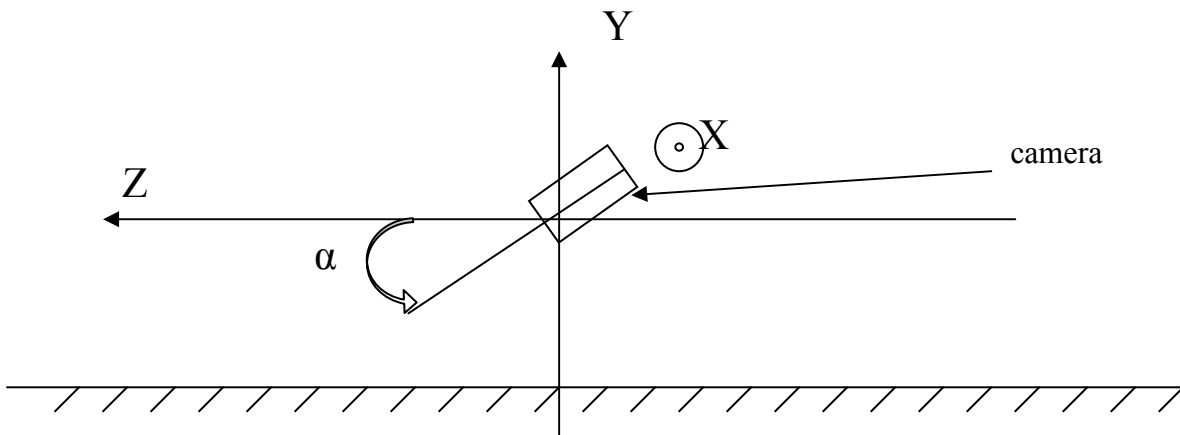
$$Object_c = P R S T Object_w \quad (\text{εξ 2.16}) \Leftrightarrow$$

$$Object_w = (P R S T)^{-1} Object_c \quad (\text{εξ 2.17}).$$

2.2.3 Παράδειγμα αντιστροφής

Ας δούμε πως θα μπορούσε να υλοποιηθεί μια αντιστοίχιση σημείων από το επίπεδο της εικόνας στον 3Δ χώρο, με την βοήθεια κάποιας επιπλέον πληροφορίας. Έστω, ότι θέλουμε να ανακατασκευάσουμε την εικόνα που λαμβάνουμε με τέτοιο τρόπο ώστε τα σημεία που ανήκουν σε ένα συγκεκριμένο επίπεδο του 3Δ χώρου να εμφανίζονται ισαπέχοντα. Δηλαδή, να αφαιρεθεί από την εικόνα η επίδραση της προοπτικής προβολής λαμβάνοντας υπόψη και την περιστροφή ή μετατόπιση που έχει υποστεί η κάμερα.

Αρχικά ας θεωρήσουμε την διάταξη του Σχήματος 2.4.



Σχήμα 2.4 : Διάταξη για το παράδειγμα της αντιστροφής της προοπτικής προβολής.

Η κάμερα, δηλαδή, είναι τοποθετημένη σε κάποιο υπερυψωμένο σημείο και έχει περιστραφεί γύρω από τον άξονα των X κατά γωνία α . Αυτή η διαμόρφωση είναι συνηθισμένη σε εφαρμογές ελέγχου και διαχείρισης της κυκλοφορίας σε εθνικές οδούς.

Όπως αναφέρθηκε και προηγουμένως, υπάρχει η δυνατότητα εξαγωγής των 3Δ συντεταγμένων ενός σημείου από την θέση του ειδώλου του στην εικόνα αν είναι γνωστό κάτι για την θέση του στο 3Δ χώρο. Η πληροφορία αυτή αντλείται από το γεγονός ότι τα σημεία που μας ενδιαφέρουν ανήκουν σε κάποιον συγκεκριμένο γεωμετρικό τόπο. Αν αυτός ο τόπος είναι το επίπεδο $AX+BY+CZ=1$, τότε σύμφωνα με ό,τι έχει αναφερθεί ως τώρα για τους μετασχηματισμούς περιστροφής, μετατόπισης και προοπτικής προβολής, θα πρέπει να ισχύουν ταυτόχρονα οι εξής εξισώσεις :

$$x = \lambda \frac{X}{-Y \sin a + Z \cos a} \quad (\text{εξ 2.18})$$

$$y = \lambda \frac{Y \cos a + Z \sin a}{-Y \sin a + Z \cos a} \quad (\text{εξ 2.19})$$

$$AX + BY + CZ = 1 \quad (\text{εξ 2.20})$$

όπου (x,y) οι συντεταγμένες στο επίπεδο της εικόνας και (X,Y,Z) οι αντίστοιχες στον 3Δ χώρο.

Από την λύση αυτού του 3x3 συστήματος (βλ Παράρτημα) προκύπτουν οι (X,Y,Z) συντεταγμένες που αντιστοιχούν στο σημείο (x,y) της εικόνας, αν στο τελευταίο πράγματι απεικονίζεται σημείο του επιπέδου $AX+BY+CZ=1$ και όχι κάποιο άλλο σημείο που βρίσκεται ανάμεσα στο (X,Y,Z) και στο κέντρο της προβολής πάνω

στην ευθεία που αυτά ορίζουν. Έτσι, τα pixel που απεικονίζουν σημεία τα οποία δεν ανήκουν στον υποτιθέμενο γεωμετρικό χώρο, δεν θα αναδιαταχθούν σωστά.

Έχοντας υπολογίσει τις χωρικές συντεταγμένες που μας ενδιαφέρουν, σειρά έχει η διαδικασία της απεικόνισής τους. Τα pixel που αντιστοιχούν σε σημεία του επιπέδου $AX+BY+CZ=I$ και θα αποτελέσουν την ανακατασκευασμένη εικόνα σαρώνονται για να αποδοθεί στο καθένα η τιμή που θα του αντιστοιχούσε με δεδομένη την αρχική εικόνα. Αυτό συμβαίνει, διότι, πολλά σημεία της αρχικής εικόνας, δεν έχουν ένα προς ένα αντιστοιχία με αυτά της ανακατασκευασμένης. Το δομικό διάγραμμα (block diagram) στο Σχήμα 2.7 περιγράφει την όλη διαδικασία.

2.2.4 Αναλυτικότερη περιγραφή του αλγόριθμου αντιστροφής της προοπτικής προβολής (Inverse perspective mapping)

Αρχικά, καθορίζονται οι σταθερές που εξαρτώνται από τα χαρακτηριστικά της κάμερας. Οι σταθερές αυτές κρατούν τους συντελεστές μήκους ανά pixel πάνω στο επίπεδο της εικόνας.

$\text{pixel_per_meter_H} = \text{Οριζόντια ανάλυση (pixels)} / \text{Μήκος επιπέδου εικόνας (m)}$

$\text{pixel_per_meter_V} = \text{Κατακόρυφη ανάλυση (pixels)} / \text{Πλάτος επιπέδου εικόνας (m)}$

Είναι προτιμότερο οι τιμές αυτές να υπολογιστούν πειραματικά τοποθετώντας ένα αντικείμενο γνωστού μεγέθους σε γνωστή απόσταση από την κάμερα και παρατηρώντας τον τρόπο που αυτό αποτυπώνεται.

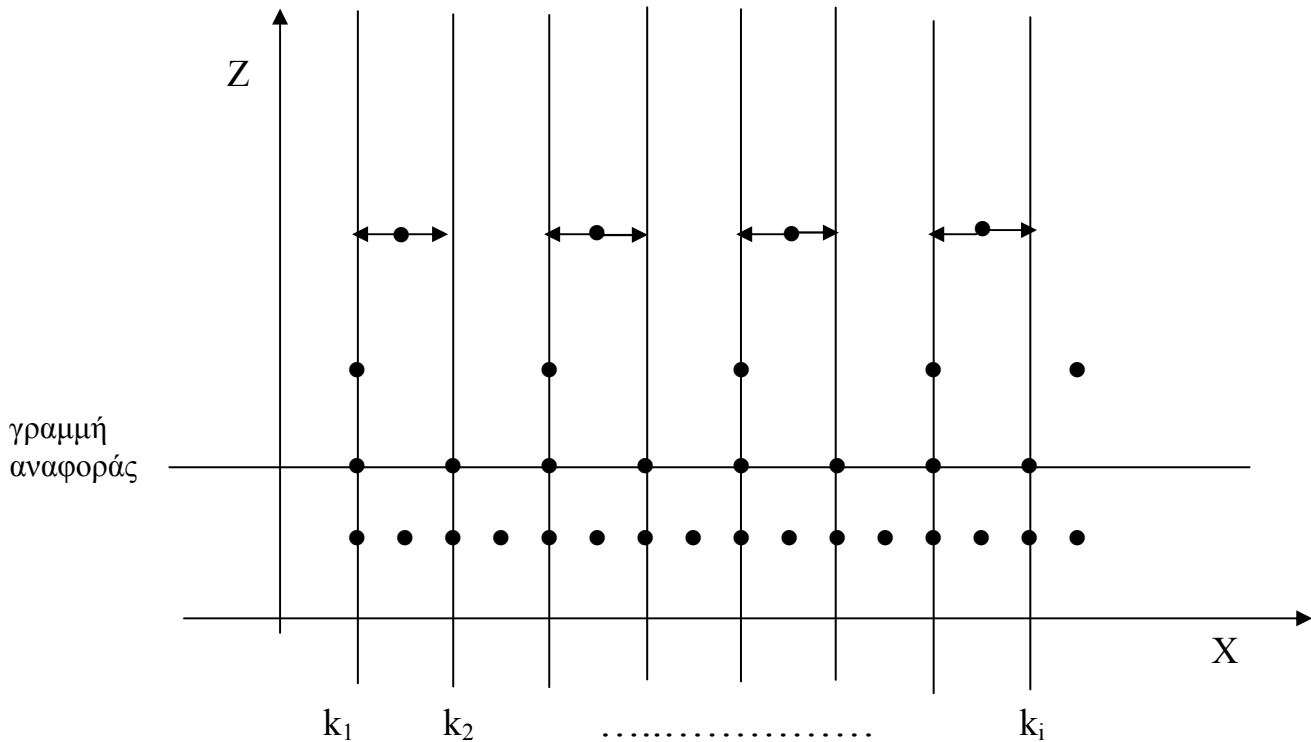
Η συνάρτηση που υλοποιεί την αντιστροφή της προοπτικής προβολής και αναδιατάσσει την εικόνα είναι η :

```
void inverse_main(unsigned char **Im,unsigned char **target,double A,double B,double C,double f,double R[3]),
```

όπου Im, target οι εικόνες εισόδου και εξόδου αντίστοιχα, τα A, B, C είναι οι παράμετροι του επιπέδου με εξίσωση $AX+BY+CZ=I$, ενώ το διάνυσμα R περιέχει την τιμή της γωνίας κατά την οποία έχει περιστραφεί η κάμερα γύρω από τον X (R[1]) άξονα του γενικού συστήματος συντεταγμένων. Οι άλλες δύο θέσεις του διανύσματος (R[0]), (R[2]) υπάρχουν με την προοπτική της λύσης ενός μελλοντικού προβλήματος που θα απαιτούσε τη στροφή της κάμερας και ως προς τους Z, και Y άξονες αντίστοιχα.

Πριν την περιγραφή του αλγόριθμου, θα ήταν καλό να καθορίσουμε την μορφή της αναδιαταγμένης εικόνας. Η ανακατασκευή, γενικά, ακολουθεί ένα μοντέλο που επιλέγεται έτσι ώστε να εξυπηρετεί τις ανάγκες κάθε εφαρμογής (π.χ. μοντέλο δρόμου). Επειδή κάτι τέτοιο δεν είναι διαθέσιμο στην δική μας περίπτωση, υιοθετείται ως περιορισμός, συνθήκη τέτοια ώστε σε κάθε γραμμή της εικόνα αυτής, κάθε δύο διαδοχικά pixel να αντιστοιχούν σε σημεία που απέχουν μια σταθερή απόσταση στον 3D χώρο. Η απόσταση αυτή είναι ίδια για όλες τις γραμμές. Αυτή η συνθήκη επιβάλλεται ουσιαστικά στην εικόνα επιλέγοντας για ανακατασκευή μόνο τα pixel που απεικονίζουν σημεία του 3D χώρου που βρίσκονται πάνω στις ευθείες $X=k_i$ (k_i σταθερές) του επιπέδου $AX+BY+CZ=I$. Για να καθοριστούν τα k_i επιλέγεται

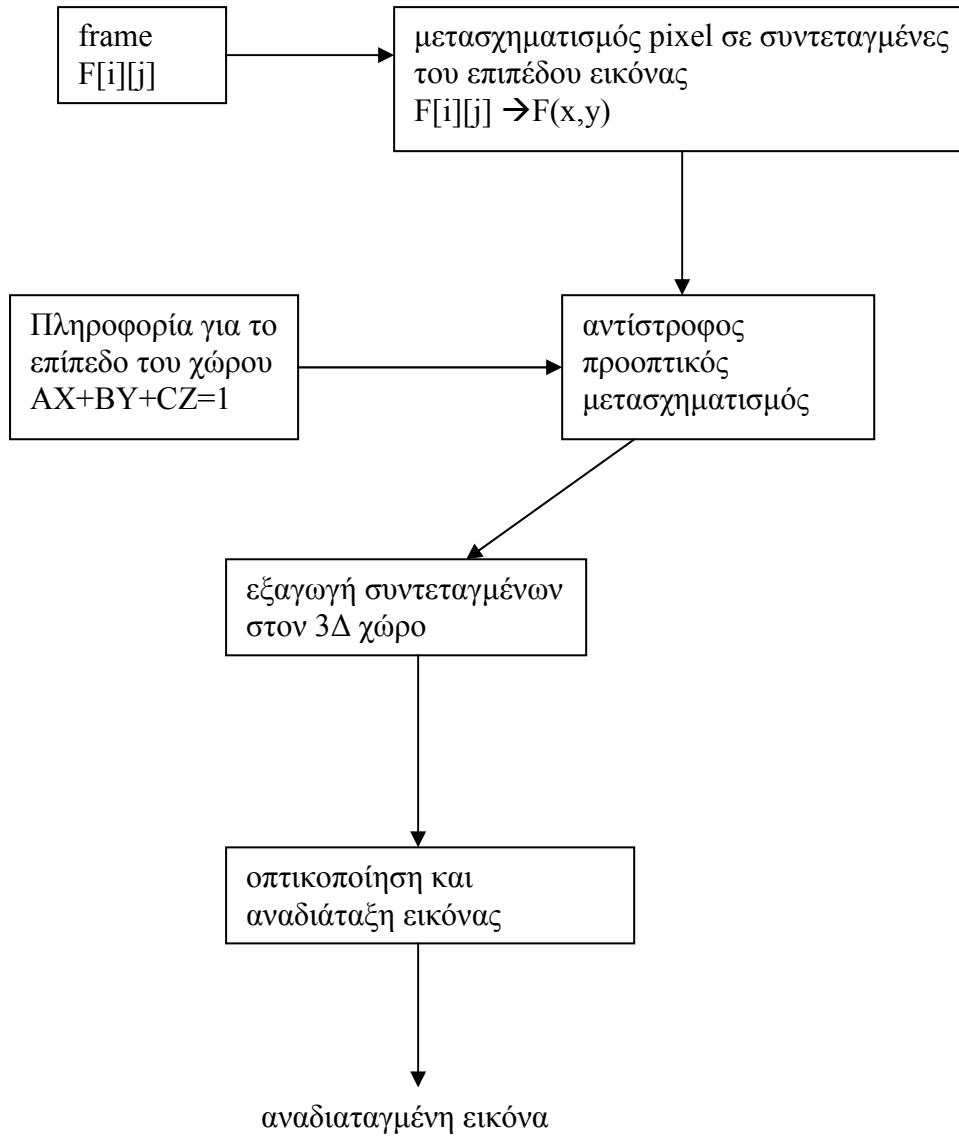
η γραμμή της εικόνας η οποία θα αποτελέσει την γραμμή αναφοράς και που σύμφωνα με τις συντεταγμένες X στον 3Δ χώρο που αντιστοιχούν στα pixel της θα αναδιαταχθεί η εικόνα. Η αναγκαιότητα της «γραμμής αναφοράς» γίνεται σαφής αν παρατηρηθεί το Σχήμα 2.6, όπου εμφανίζεται η διάταξη των σημείων του 3Δ χώρου που απεικονίζονται στην αρχική εικόνα. Είναι εμφανές, ότι η πυκνότητα των σημείων μειώνεται όσο αυξάνει το Z με αποτέλεσμα να μην είναι διαθέσιμη αρκετή πληροφορία έτσι ώστε να είναι εφικτή η ανακατασκευή όλων των σημείων του επιπέδου $AX+BY+CZ=1$. Έτσι, αρκούμαστε στην ανακατασκευή μόνο των $X=k_i$ ευθειών.



Σχήμα 2.6 : Διάταξη σημείων του 3Δ χώρου που απεικονίζονται στην αρχική εικόνα

Η γραμμή αναφοράς επιλέγεται έτσι ώστε να βρίσκεται στο μέσο της περιοχής της εικόνας (όχι του 3Δ χώρου). Η επιλογή πρέπει να γίνει προσεκτικά διότι η απόσταση στον 3Δ χώρο που αντιστοιχεί ανάμεσα το πρώτο και το τελευταίο pixel μιας γραμμής αυξάνεται για σημεία που βρίσκονται μακριά ενώ, ταυτόχρονα, η ποσότητα της πληροφορίας, που είναι διαθέσιμη μέσω της εικόνας, μειώνεται. Έτσι, αν επιλεγεί μια γραμμή που αντιστοιχεί σε σημεία πολύ κοντά στην κάμερα, ώστε να υπάρχει αρκετή πληροφορία για την ανακατασκευή, είναι πιθανό κάποια σημεία που βρίσκονται σε μεγαλύτερη απόσταση να μην αποτυπωθούν, αφού θα είναι έξω από τα στενά όρια που καθορίζει η γραμμή αναφοράς.

Ο αλγόριθμος, λοιπόν, σαρώνει τις γραμμές της αναδιαταγμένης εικόνας και προσπαθεί να βρει τα σημεία εκείνα της αρχικής, που έχουν συντεταγμένες X κοντά σε αυτές των σημείων του 3Δ χώρου που απεικονίζονται στην γραμμή αναφοράς. Κάθε σημείο της τελικής εικόνας παίρνει την τιμή του pixel της αρχικής που θα προκύψει από την προηγούμενη σάρωση.

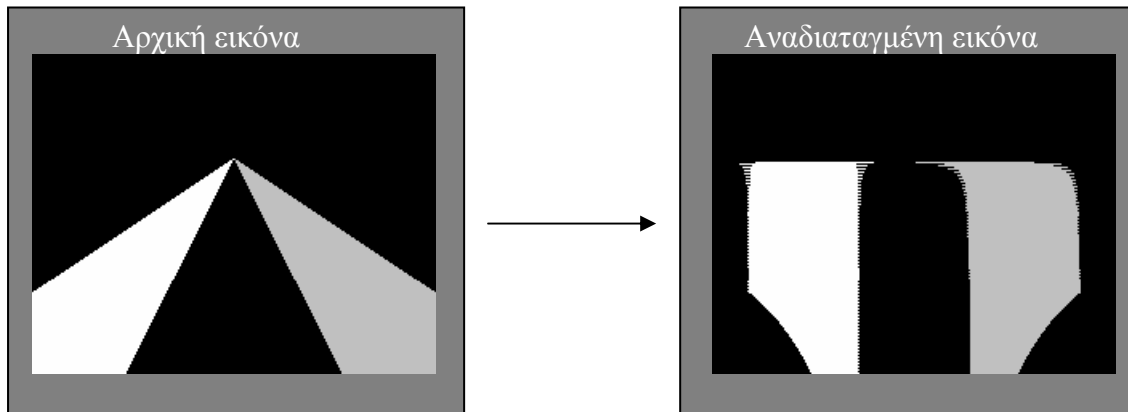


Σχήμα 2.7 : Διαδικασία αναδιάταξης εικόνας

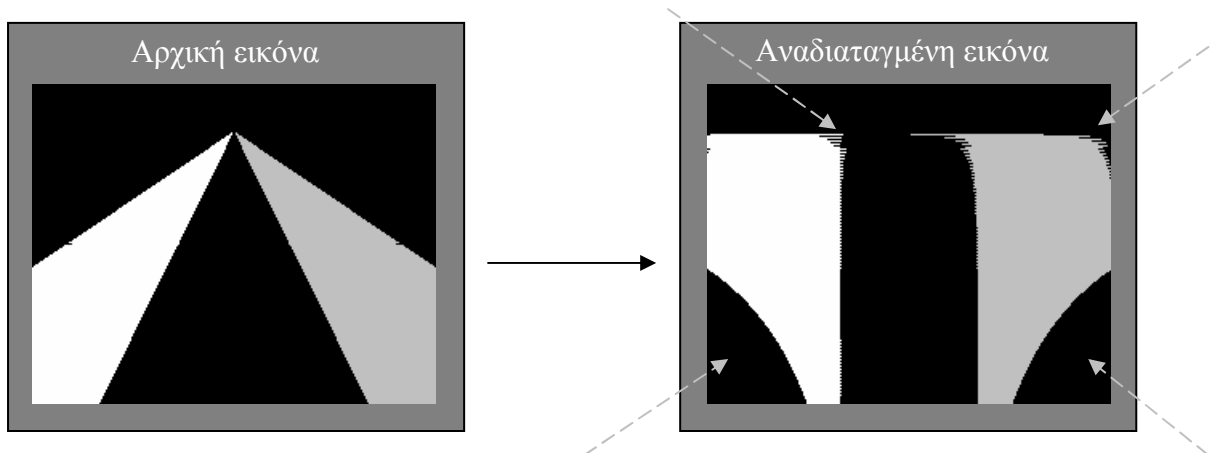
2.2.5 Αποτελέσματα και συζήτηση του αλγόριθμου αντιστροφής προοπτικής προβολής (Inverse perspective matching)

Για να δοκιμαστεί ο αλγόριθμος δημιουργήθηκαν εικόνες μοντέλα που απεικονίζουν λωρίδες πλάτους ενός μέτρου (1m) πάνω σε ένα επίπεδο το οποίο βρίσκεται ένα μέτρο (1m) κάτω από την κάμερα.

Για γωνία $\alpha = 5^\circ$ και επίπεδο $Y=-1$:



Για γωνία $\alpha = 10^\circ$ και επίπεδο $Y=-1$:



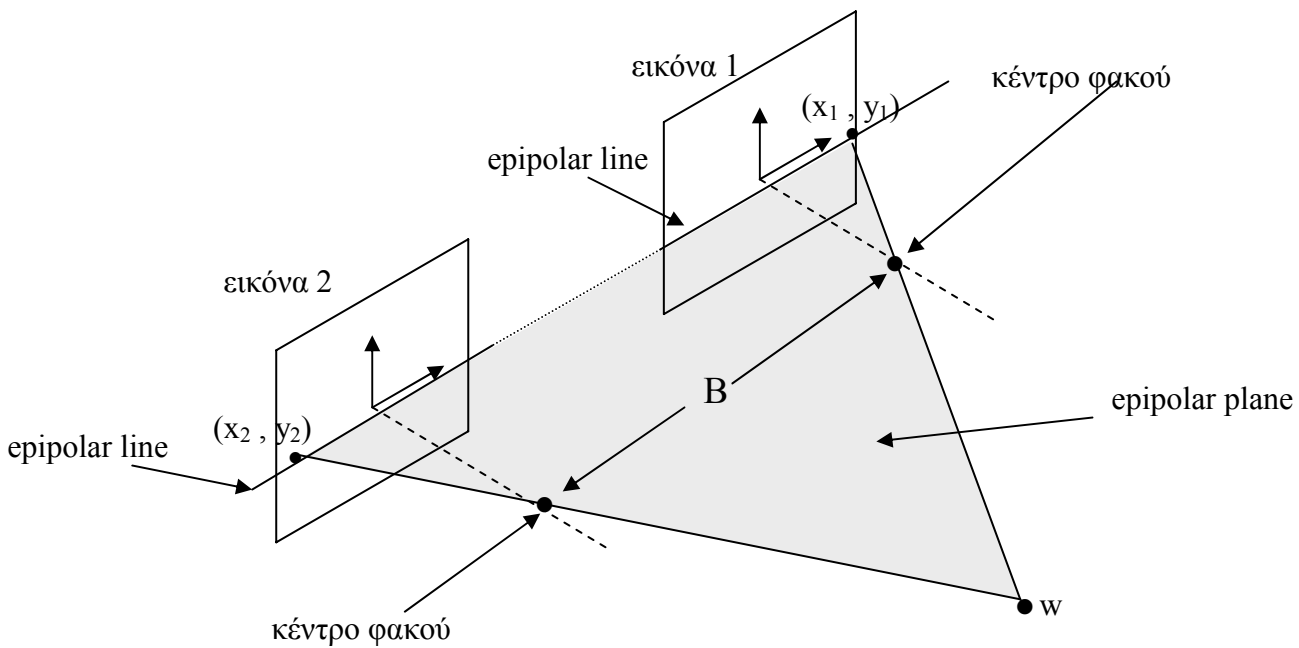
Εικόνα 2.1 : Αποτελέσματα αλγόριθμου αντιστροφής προοπτικής προβολής

Αξίζουν να αναφερθούν δύο παρατηρήσεις όσον αφορά τις αναδιαταγμένες εικόνες. Η πρώτη έχει να κάνει με τις περιοχές στα κάτω άκρα των εικόνων, οι οποίες εμφανίζονται «κενές» αφού η αρχική εικόνα δεν περιέχει πληροφορία για αυτές. Η άλλη αφορά στην ασάφεια που παρατηρείται στα πάνω άκρα των εικόνων τα οποία αντιστοιχούν σε σημεία που βρίσκονται μακριά, οπότε και η πληροφορία που είναι διαθέσιμη είναι ανεπαρκής.

2.3 Στερεοψία

Ένας άλλος τρόπος για να επιτευχθεί ένα προς ένα αντιστοίχιση μεταξύ των σημείων του επιπέδου της εικόνας και αυτών του 3Δ χώρου είναι η χρήση στερεοοπτικών τεχνικών (stereo ή binocular vision). Η λήψη, δηλαδή, του ίδιου χώρου από δύο ξεχωριστές, αλλά πανομοιότυπες, κάμερες οι οποίες έχουν τοποθετηθεί έτσι ώστε τα συστήματα συντεταγμένων τους να είναι τέλεια ευθυγραμμισμένα και τα κέντρα των φακών τους να απέχουν κατά B . Το επίπεδο που ορίζουν το τυχαίο σημείο w του χώρου που βρίσκεται υπό παρατήρηση και τα κέντρα των δύο φακών, ονομάζεται epipolar plane και η τομή του με το επίπεδο της εικόνας ορίζει την γραμμή epipolar line.

Το παρακάτω σχήμα περιγράφει την διάταξη.



Σχήμα 2.8 : Μοντέλο στερεοοπτικής όρασης

Σε αυτήν την περίπτωση, ένα σημείο w του χώρου θα απεικονίζεται στο (x_1, y_1) και (x_2, y_2) στις εικόνες ένα και δύο αντίστοιχα. Θεωρώντας ότι το σύστημα συντεταγμένων του 3Δ χώρου συμπίπτει με αυτό της πρώτης εικόνας, προκύπτει για την τετμημένη του σημείου w :

$$x_1 = \lambda \frac{X_1}{\lambda - Z_1} \Leftrightarrow X_1 = \frac{x_1}{\lambda} (\lambda - Z_1)$$

Κάνοντας το ίδιο και για την άλλη κάμερα :

$$x_2 = \lambda \frac{X_2}{\lambda - Z_2} \Leftrightarrow X_2 = \frac{x_2}{\lambda} (\lambda - Z_2)$$

όπου $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2)$ οι συντεταγμένες του σημείου w για κάθε περίπτωση.

Από την διάταξη των δύο μηχανών λήψης, προκύπτει

$$X_2 = X_1 + B \quad \text{και} \quad Z_2 = Z_1 = Z$$

Αντικαθιστώντας τα X_2, Z_1, Z_2 :

$$X_1 = \frac{x_1}{\lambda}(\lambda - Z)$$

και

$$X_1 + B = \frac{x_2}{\lambda}(\lambda - Z)$$

Αφαιρώντας κατά μέλη και λύνοντας ως προς Z :

$$Z = \lambda - \frac{\lambda B}{x_2 - x_1} \quad (\text{εξ. 2.21})$$

Από την παραπάνω σχέση, φαίνεται ότι, αν είναι γνωστή η εστιακή απόσταση λ του φακού, η απόσταση B (*baseline*) μεταξύ των δύο εικόνων και η διαφορά των αντίστοιχων συντεταγμένων στο επίπεδο της εικόνας x_1, x_2 , τότε ο υπολογισμός του Z για το σημείο w είναι εύκολος. Έχοντας υπολογίσει το Z , οι συντεταγμένες του σημείου w στον 3D χώρο X και Y , προκύπτουν από τους τύπους της προοπτικής προβολής. Με αυτόν τον τρόπο, κάθε σημείο του επιπέδου της εικόνας αντιστοιχίζεται με μοναδικό τρόπο στον 3D χώρο. Η ίδια διαδικασία εφαρμόζεται και στην περίπτωση όπου το ζητούμενο είναι η αντιστοίχιση των σημείων της εικόνας με αυτά ενός αντικείμενου. Σ' αυτή την περίπτωση, το Z από την προηγούμενη σχέση θα πρέπει να υπολογιστεί για όλα τα σημεία που απαρτίζουν το αντικείμενο.

Το δυσκολότερο σημείο στην όλη διαδικασία είναι η εύρεση των δύο αυτών σημείων στις διαφορετικές εικόνες που πραγματικά απεικονίζουν το ίδιο σημείο του 3D χώρου. Μια, συχνά χρησιμοποιούμενη τεχνική, είναι η επιλογή ενός σημείου μέσα σε μια μικρή περιοχή της μιας εικόνας και κατόπιν η εύρεση της περιοχής στην άλλη εικόνα με την οποία ταιριάζει καλύτερα. Η σύγκριση ομοιότητας των περιοχών γίνεται με τεχνικές συσχέτισης (*correlation*). Στην περίπτωση που η απεικονιζόμενη σκηνή περιέχει αντικείμενα με σαφή χαρακτηριστικά, όπως π.χ. γωνίες, μια τεχνική ταιριάσματος χαρακτηριστικών (*feature matching*) θα ήταν αποτελεσματικότερη.

Στην παραπάνω ανάλυση χρησιμοποιήθηκε για την προοπτική προβολή ο τύπος $x = \lambda \frac{X}{\lambda - Z}$. Αντίστοιχα, αν είχε θεωρηθεί ο τύπος $x = \lambda \frac{X}{Z}$, τότε θα προέκυπτε:

$$Z = \frac{\lambda B}{x_2 - x_1} \quad (\text{εξ. 2.22}).$$

Η προηγούμενη διαδικασία μπορεί να δοθεί και σε μορφή πινάκων. Σε αυτή την περίπτωση, το σημείο του 3Δ χώρου $w[X, Y, Z, 1]^T$ απεικονίζεται στο σημείο $c_1[x_1, y_1, z_1, 1]^T$ στην πρώτη και στο $c_2[x_2, y_2, z_2, 1]^T$ στην δεύτερη εικόνα με τον ακόλουθο τρόπο :

$$c_1 = Pw \quad (\text{εξ. 2.23})$$

$$c_2 = PTw \quad (\text{εξ. 2.24})$$

όπου P ο πίνακας του προοπτικού μετασχηματισμού και T ο πίνακας του μετασχηματισμού μετατόπισης, ο οποίος και δίνεται παρακάτω :

$$T = \begin{bmatrix} 1 & 0 & 0 & B \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

όπου B η απόσταση μεταξύ των κέντρων της προβολής των δύο εικόνων (baseline).

Έχει θεωρηθεί ότι το σύστημα συντεταγμένων του 3Δ χώρου ταυτίζεται με αυτό της πρώτης κάμερας. Ως εκ τούτου, το σύστημα συντεταγμένων της δεύτερης κάμερας είναι μετατοπισμένο κατά $-B$ στην διεύθυνση του άξονα X . Επιλέγοντας την πρώτη γραμμή από καθεμιά από τις σχέσεις (εξ. 2.23), (εξ. 2.24) που συνδέουν τα c_1, c_2 με το w , προκύπτουν οι ίδιες εξισώσεις για τα x_1, x_2, Z που αναφέρθηκαν και προηγουμένως.

2.4 Λήψη αντικειμένου από διαφορετικά σημεία

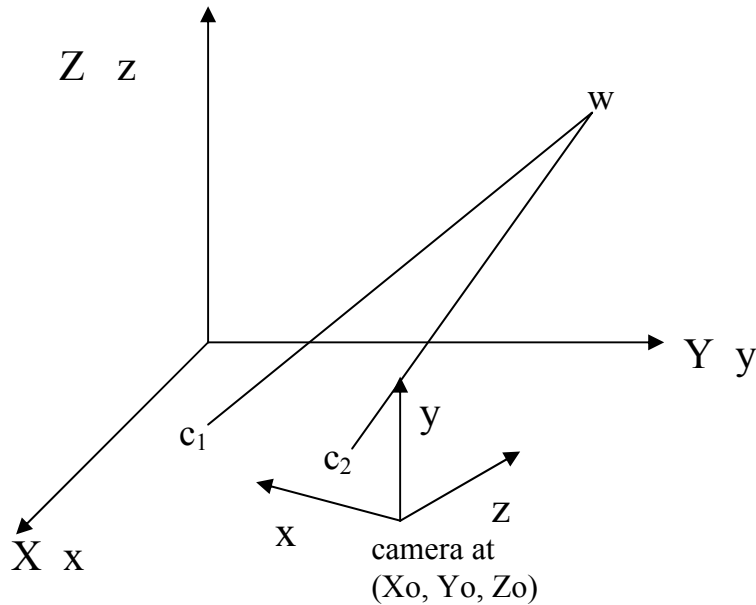
Παρόμοια με την μέθοδο της στερεοπτικής όρασης, είναι η μέθοδος της λήψης πολλαπλών εικόνων του ίδιου αντικειμένου από διαφορετικά σημεία του χώρου. Αρχικά λαμβάνεται μια εικόνα από την αρχική θέση της κάμερας, στην οποία τα συστήματα συντεταγμένων της κάμερας και του χώρου συμπίπτουν. Έτσι, το σημείο w του 3Δ χώρου αντιστοιχίζεται στο επίπεδο της εικόνας στο σημείο c_1 σύμφωνα με την σχέση :

$$c_1 = P w \quad (\text{εξ. 2.25}) \quad \text{όπου } P \text{ ο πίνακας της προοπτικής προβολής}$$

Μετατοπίζοντας και περιστρέφοντας την κάμερα λαμβάνεται ακόμα μια εικόνα, στην οποία το ίδιο σημείο w του 3Δ χώρου αντιστοιχίζεται στο επίπεδο της εικόνας στο σημείο c_2 σύμφωνα με την σχέση :

$$c_2 = P R T w \quad (\text{εξ. 2.26})$$

όπου R και T οι πίνακες των μετασχηματισμών περιστροφής και μετατόπισης αντίστοιχα.



Σχήμα 2.9 : Μοντέλο λήψης από διαφορετικές θέσεις.

Όπως θα μπορούσε να υποψιαστεί κανείς, προκύπτει και εδώ το πρόβλημα της αντιστοίχισης που είχε παρουσιαστεί και στην περίπτωση της στερεοοπτικής όρασης. Με την προϋπόθεση ότι αυτό έχει λυθεί, από τις παραπάνω σχέσεις (εξ. 2.25), (εξ. 2.26) επιλέγονται οι δύο πρώτες γραμμές της πρώτης και μια από τις αντίστοιχες της δεύτερης, για να προκύψει ένα 3×3 σύστημα της μορφής :

$$\begin{aligned} x_1 &= \lambda X / Z \\ y_1 &= \lambda Y / Z \\ x_2 &= f(X, Y, Z) \end{aligned}$$

η λύση του οποίου θα δώσει τις χωρικές συντεταγμένες του σημείου w .

Η μέθοδος επεκτείνεται για αντικείμενα αντικαθιστώντας τα c_1, c_2, w με τα $Object_{c1}, Object_{c2}, Object_w$, οπότε προκύπτουν N το πλήθος 3×3 συστήματα, η λύση των οποίων θα δώσει τις συντεταγμένες των N σημείων που περιγράφουν το αντικείμενο.

2.5 Λήψη αντικειμένου σε διαδοχικές χρονικές στιγμές

Στην προσπάθεια να ξεπεραστεί η εγγενής αδυναμία ακριβούς ανακατασκευής του 3D χώρου από μία 2D εικόνα, χρησιμοποιείται ακόμα μια τεχνική. Αντί, όπως προηγουμένως, να λαμβάνονται όψεις του αντικειμένου από διαφορετικά χωρικά σημεία, λαμβάνονται χρονικά διαδοχικές εικόνες του (frames). Έστω λοιπόν, το σημείο w του 3D χώρου το οποίο την χρονική στιγμή t , απεικονίζεται στο σημείο c_t του επιπέδου της εικόνας, ενώ την στιγμή $t+1$ στο σημείο c_{t+1} . Καλό θα ήταν να σημειωθεί, ότι η συνήθης πρακτική, όσον αφορά τις διαδοχικές λήψεις, είναι να γίνονται σε τακτά χρονικά διαστήματα. Έτσι, ο χρόνος δεν έχει απόλυτη έννοια και ο συμβολισμός $t+1$ απλά δηλώνει ότι το σημείο c_{t+1} ανήκει στην αμέσως επόμενη εικόνα μετά την t . Θεωρώντας ότι το σύστημα συντεταγμένων του 3D χώρου ταυτίζεται με αυτό της εικόνας, προκύπτει για το σημείο c_t :

$$c_t = P w \quad (\text{εξ. 2.27})$$

Αν υποθεθεί ότι στον χρόνο που μεσολαβεί μέχρι την λήψη της επόμενης εικόνας της ακολουθίας, το σημείο w έχει βρεθεί σε μια νέα θέση έχοντας υποστεί μετατόπιση και περιστροφή που δίνονται από τους αντίστοιχους πίνακες T και R , τότε :

$$c_{t+1} = P R T w \quad (\text{εξ. 2.28})$$

Βασικό πρόβλημα εδώ, όπως και στην περίπτωση της στερεοοπτικής όρασης, είναι η αντιστοίχιση του σημείου c_t με το c_{t+1} . Η εύρεση δηλαδή του σημείου c_{t+1} στο οποίο απεικονίζεται το ίδιο σημείο w .

Σε αναλογία με την περίπτωση της παραγράφου 2.4, επιλέγοντας τις δύο πρώτες γραμμές της πρώτης (εξ. 2.27) και μια από τις αντίστοιχες της δεύτερης (εξ. 2.28) σχέσης, προκύπτει ένα 3×3 σύστημα η λύση του οποίου, θα δώσει τις χωρικές συντεταγμένες του σημείου w .

Η μέθοδος επεκτείνεται για αντικείμενα αντικαθιστώντας τα c_t , c_{t+1} , w με τα $Object_t$, $Object_{t+1}$, $Object_w$, οπότε προκύπτουν N το πλήθος 3×3 συστήματα, η λύση των οποίων θα δώσει τις συντεταγμένες των N σημείων που περιγράφουν το αντικείμενο.

2.6 Χωρικοί περιορισμοί αντικειμένων

Μια διαφορετική προσέγγιση στο ίδιο πρόβλημα της ανακατασκευής είναι η χρήση των περιορισμών που προκύπτουν για τα σημεία ενός αντικειμένου από το γεγονός ότι ανήκουν σ' αυτό. Προφανώς, για να έχει εφαρμογή κάτι τέτοιο, θα πρέπει να υπάρχει αρκετή πληροφορία για το υπό εξέταση αντικείμενο, η οποία να μπορεί να δώσει ένα ικανοποιητικό 3D μοντέλο. Με δεδομένα, λοιπόν, κάποια χαρακτηριστικά του αντικειμένου είναι δυνατό να καθοριστούν οι συντεταγμένες των σημείων που το αποτελούν, σε σχέση με ένα σημείο που θα διαδραματίζει το ρόλο του σημείου αναφοράς. Στην γενική περίπτωση, δηλαδή, οι συντεταγμένες των σημείων, θα είναι συναρτήσεις του σημείου αναφοράς. Το σημείο αυτό είναι δυνατό να ταυτίζεται με το κέντρο βάρους του αντικειμένου ή με όποιο άλλο σημείο είναι σε θέση να εκμεταλλευτεί την όποια συμμετρία ή άλλο χαρακτηριστικό του.

Με τα προαναφερθέντα υπ' όψιν, αν το σημείο αναφοράς του αντικειμένου στον 3Δ χώρο είναι το $w_{ref} [x_{ref}, y_{ref}, z_{ref}]^T$, ο πίνακας που το περιγράφει έχει ως εξής :

$$Object_w = \begin{bmatrix} f_0(x_{ref}) & f_1(x_{ref}) & \cdots & f_{N-1}(x_{ref}) \\ g_0(y_{ref}) & g_1(y_{ref}) & \cdots & g_{N-1}(y_{ref}) \\ h_0(z_{ref}) & h_1(z_{ref}) & \cdots & h_{N-1}(z_{ref}) \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Οι συναρτήσεις f_i , g_i , h_i θα μπορούσαν να είναι συναρτήσεις και των τριών συντεταγμένων του σημείου αναφοράς (π.χ. $f_i(x_{ref}, y_{ref}, z_{ref})$ κ.ο.κ.). Κάτι τέτοιο, όμως, αυξάνει την πολυπλοκότητα και θα πρέπει να εφαρμόζεται μόνο όταν οι συναρτήσεις μιας μεταβλητής αδυνατούν να περιγράψουν με ικανοποιητική ακρίβεια το μοντέλο του αντικείμενου.

Η μορφή αυτή του πίνακα που περιγράφει το αντικείμενο, δίνει την δυνατότητα εύρεσης των συντεταγμένων των σημείων του, αν είναι γνωστές οι αντίστοιχες του σημείου αναφοράς ή είναι δυνατή κάποιοι είδους εκτίμησή τους. Μια τέτοια μέθοδος που βασίζεται στην εκτίμηση των συντεταγμένων του σημείου αναφοράς για τον προσδιορισμό της θέσης του αντικειμένου, παρουσιάζεται σε επόμενο κεφάλαιο.

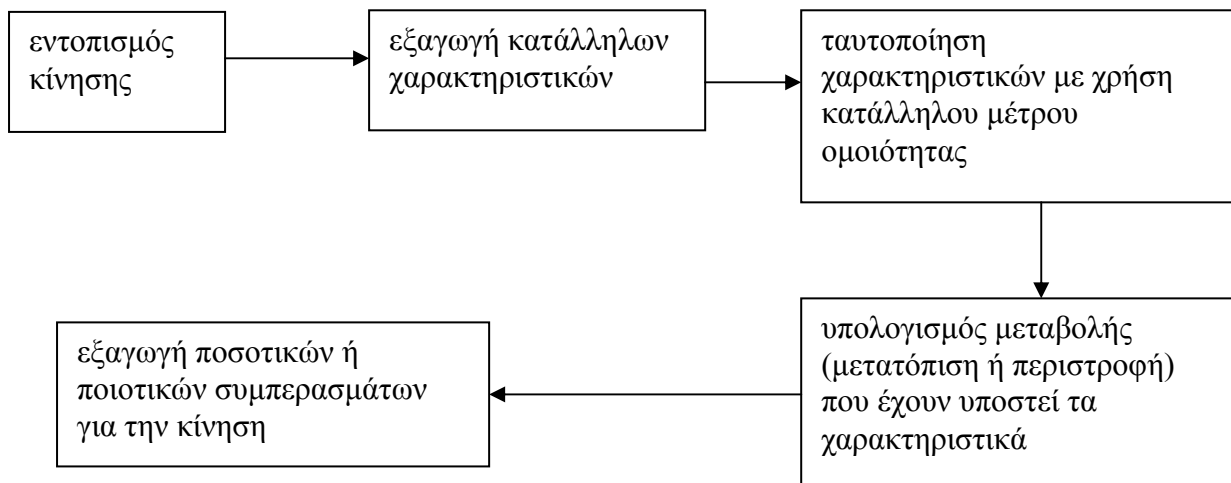
Κεφάλαιο 3

Εντοπισμός και παρακολούθηση αντικειμένου

3.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικότερες μέθοδοι εντοπισμού και παρακολούθησης αντικειμένων, οι οποίες και διακρίνονται σε δύο κατηγορίες. Η πρώτη, αφορά μεθόδους που βασίζονται στην παρακολούθηση κάποιων χαρακτηριστικών του αντικειμένου (feature based) κατά τη διάρκεια μιας ακολουθίας εικόνων. Στην άλλη κατηγορία, ανήκουν οι τεχνικές εκείνες που στηρίζονται στην ταυτοποίηση των δεδομένων της εικόνας με κάποιο μοντέλο που περιγράφει το αντικείμενο (model based).

Στο παρακάτω διάγραμμα φαίνεται η διαδικασία που ακολουθείται γενικότερα για τον εντοπισμό και την παρακολούθηση αντικειμένων :



Σχήμα 3.1 : Δομικό διάγραμμα των γενικότερων σταδίων στην διαδικασία του εντοπισμού και της παρακολούθησης της κίνησης

3.2 Μέθοδοι εντοπισμού κινούμενων αντικειμένων

Το στάδιο του εντοπισμού της περιοχής ενδιαφέροντος προηγείται της διαδικασίας παρακολούθησης του αντικειμένου. Είναι λογικό να θέλουμε να απομονώσουμε την ή τις περιοχές της εικόνας στις οποίες εμφανίζεται η κίνηση. Η συνήθης πρακτική, κυρίως όταν η κάμερα είναι ακίνητη, είναι να πάρουμε την διαφορά ανάμεσα σε δυο διαδοχικά frame.

Η διαφορά αυτή θα είναι μηδέν για σημεία που δεν έχουν κινηθεί, ενώ είναι διάφορη του μηδενός για pixel που αντιστοιχούν σε κινούμενα σημεία. Για να γίνει

σαφέστερος αυτός ο διαχωρισμός εφαρμόζεται ένα κατώφλι τ στις τιμές της διαφοράς ώστε να προκύψει η εικόνα :

$$DP_{jk}(x, y) = \begin{cases} 1 & \text{αν } |F(x, y, 1) - F(x, y, 2)| > \tau \\ 0 & \text{αλλιώς} \end{cases} \quad (\text{εξ. 3.1})$$

Σε μια τέτοια εικόνα, τα σημεία με τιμή ένα (1) είναι δυνατό να θεωρηθούν ότι είναι το αποτέλεσμα κίνησης ή αλλαγής στο φωτισμό, σε αντίθεση με τα σημεία με τιμή μηδέν (0) που θεωρούνται στατικά.

Όπως θα μπορούσε να φανταστεί κανείς, η επιλογή του κατωφλίου τ διαδραματίζει σημαντικό ρόλο. Ένα συγκεκριμένο κατώφλι δεν μπορεί να εφαρμοστεί για εικόνες που περιέχουν κίνηση ή μεταβολή φωτεινότητας με αργό ρυθμό. Έτσι, θα πρέπει, όπως αναφέρεται και στο τέλος της εργασίας στην παράγραφο που αφορά τις επεκτάσεις των μεθόδων, να επιλέγεται κάθε φορά ένα διαφορετικό κατώφλι, είτε από τον σχεδιαστή του συστήματος, είτε από κάποιον προσαρμοστικό (adaptive) αλγόριθμο.

3.3 Μέθοδοι παρακολούθησης χαρακτηριστικών (feature based)

3.3.1 Εισαγωγή

Ο στόχος αυτών των μεθόδων είναι η εξαγωγή των στοιχείων κίνησης ενός αντικείμενου, παρακολουθώντας την συμπεριφορά στον χώρο και στον χρόνο, κάποιων χαρακτηριστικών του σημείων. Το πρώτο βήμα, λοιπόν, είναι ο καθορισμός αυτών των χαρακτηριστικών. Η επιλογή είναι δυνατό να βασίζεται σε στατιστικά στοιχεία π.χ. επιλογή των pixel γύρω από τα οποία εμφανίζεται η μέγιστη διακύμανση φωτεινότητας, ή σε μορφολογικά χαρακτηριστικά του αντικείμενου π.χ. γωνίες, καμπύλες, πρότυπα (pattern), γραμμές.

3.3.2 Εξαγωγή και ταυτοποίηση χαρακτηριστικών σημείων [2]

Ανάλογα με την προηγούμενη επιλογή, προσδιορίζεται ο κατάλληλος αλγόριθμος εξαγωγής και ταυτοποίησης (matching) των χαρακτηριστικών. Έτσι, για την περίπτωση που τα χαρακτηριστικά οριστούν να είναι σημεία της εικόνας με συγκεκριμένα στατιστικές ιδιότητες, εφαρμόζεται στην εικόνα ο αντίστοιχος τελεστής δημιουργώντας ένα «χάρτη» από τιμές. Για παράδειγμα, κάθε θέση του «χάρτη» θα μπορούσε να περιέχει την τιμή διασποράς της φωτεινότητας στο αντίστοιχο pixel της εικόνας. Με βάση αυτές τις τιμές, επιλέγονται για παρακολούθηση τα σημεία που ικανοποιούν τις εκάστοτε καθορισμένες απαιτήσεις και συνήθως εμφανίζονται ως τοπικά ελάχιστα ή μέγιστα του «χάρτη».

Για την ταυτοποίηση (matching) των χαρακτηριστικών σημείων από μία εικόνα (frame) της ακολουθίας (sequence) στην επόμενη, χρησιμοποιούνται μέτρα ομοιότητας των περιοχών που περιβάλλουν τα υπό εξέταση σημεία. Ένα τέτοιο μέτρο

είναι και η συσχέτιση των τιμών φωτεινότητας των pixel κάθε περιοχής. Δύο όμοιες περιοχές αναμένεται να έχουν υψηλό δείκτη συσχέτισης σε σχέση με τις υπόλοιπες υποψήφιες προς ταυτοποίηση. Στο τέλος αυτής της διαδικασίας, έχουν δημιουργηθεί ζεύγη χαρακτηριστικών σημείων. Κάθε μέλος ενός ζεύγους, είναι η εικόνα του ίδιου σημείου του 3D αντικειμένου, σε διαφορετική χρονική στιγμή. Από την μετατόπιση (disparity) μεταξύ τους και την χρονική διαφορά μέσα στην οποία προέκυψε, μπορούν να εξαχθούν συμπεράσματα για τα χαρακτηριστικά της κίνησης. Υπενθυμίζεται, ότι η χρονική αυτή διαφορά ισοδυναμεί με την αντίστροφο της συχνότητας λήψης εικόνων της κάμερας.

Στο σημείο αυτό, θα πρέπει να αναφερθούν και τα ενδεχόμενα προβλήματα που παρουσιάζονται τόσο κατά την διαδικασία της ταυτοποίησης, όσο και κατά την εκτίμηση των ιδιοτήτων της κίνησης. Η διαδικασία της εξαγωγής χαρακτηριστικών δεν εγγυάται, στην γενική περίπτωση, ότι από κάθε εικόνα θα εξάγεται πάντα ένας σταθερός αριθμός σημείων. Το γεγονός αυτό, εμποδίζει την ένα προς ένα αντιστοίχιση όλων αυτών των σημείων, αφού για κάποια σημεία της μίας εικόνας δεν θα υπάρχουν αντίστοιχα στην άλλη. Ακόμα όμως και στην περίπτωση όπου, με κάποιον τρόπο, επιβληθεί ένας συγκεκριμένος αριθμός σημείων ανά εικόνα, είναι αρκετά πιθανό να αποκλείονται σημεία που όντως έχουν αντίστοιχο στην άλλη εικόνα, με αποτέλεσμα την εσφαλμένη ταυτοποίηση. Επακόλουθο όλων αυτών είναι να εισάγεται ένας βαθμός ανακρίβειας στην εκτίμηση των χαρακτηριστικών της κίνησης.

Πέρα, όμως, από αυτά, η εκτίμησης της κίνησης υπόκειται και σε άλλης μορφής σφάλματα. Λόγω του φαινομένου της κλιμάκωσης (scaling) που εισάγει η αλλαγή της απόστασης από την κάμερα, ένα σημείο του αντικειμένου στην εικόνα που έχει ληφθεί, όταν αυτό απείχε περισσότερο από την μηχανή λήψης, αντιστοιχεί σε μία ολόκληρη περιοχή της εικόνας που παρουσιάζει το αντικείμενο σε μικρότερη απόσταση. Συνέπεια αυτού, είναι η αδυναμία εξαγωγής αξιόπιστων συμπερασμάτων για την κίνηση, αφού τόσο η διαδικασία εξαγωγής χαρακτηριστικών όσο και η αντίστοιχη της ταυτοποίησης θα επηρεαστούν δίνοντας αποτελέσματα αμφίβολης ποιότητας. Για να αποφευχθεί η κατάσταση αυτή, φαίνεται λογικός ο περιορισμός της κίνησης του αντικειμένου σε κάποιο γνωστό επίπεδο παράλληλο, ή έστω σχεδόν παράλληλο, με αυτό της εικόνας. Με αυτόν τον τρόπο, το φαινόμενο της κλιμάκωσης είτε δεν υφίσταται είτε επηρεάζει ελάχιστα την όλη διαδικασία.

3.3.3 Εξαγωγή και ταυτοποίηση χαρακτηριστικών γραμμών

Αντί των σημείων, μπορούν να χρησιμοποιηθούν οι γραμμές που υπάρχουν σε μια εικόνα ως χαρακτηριστικά προς παρακολούθηση, ειδικά όταν τα αντικείμενα που εμφανίζονται στην ακολουθία των εικόνων χαρακτηρίζονται μορφολογικά από την ύπαρξη γραμμών. Η εξαγωγή των γραμμών από την εικόνα, γίνεται συνήθως εφαρμόζοντας σ' αυτήν κάποιον από τους τελεστές Prewitt ή Sobel. Εναλλακτικά, θα μπορούσε να βρει χρήση το φίλτρο Canny ή ο μετασχηματισμός Hough δίνοντας καλύτερα αποτελέσματα. Όπως είναι εύκολο να φανταστεί κανείς, η παρουσία του θορύβου ενδεχομένως να επηρεάσει την πιο πάνω διαδικασία, με αποτέλεσμα την εξαγωγή αποσπασματικών γραμμών.

Με δεδομένες τις γραμμές της εικόνας, ο αλγόριθμος εισέρχεται στο στάδιο της ταυτοποίησης. Στην παρακάτω προσέγγιση έχει υποτεθεί ότι τα αντικείμενα που ενδιαφέρεται το σύστημα να αναγνωρίσει, αποτελούνται από επίπεδες επιφάνειες διαχωριζόμενες από ευθύγραμμα τμήματα (3D line segments). Έτσι, ο χώρος

αναπαριστάται από επίπεδα τα οποία χαρακτηρίζονται από τις παρακάτω παραμέτρους [3]:

- $\sigma \rightarrow$ (slant) η γωνία που σχηματίζει η κάθετος στο επίπεδο του αντικειμένου με την κάθετο στο επίπεδο της εικόνας (άξονας z)
- $\tau \rightarrow$ (tilt) η γωνία που σχηματίζεται από την προβολή στο επίπεδο της εικόνας της καθέτου στο επίπεδο του αντικειμένου σε σχέση με το σύστημα αναφοράς της εικόνας
- $z \rightarrow$ απόσταση του επιπέδου του αντικειμένου από την αρχή των αξόνων

Οι τρεις αυτές παράμετροι, καθορίζουν με μοναδικό τρόπο κάθε επίπεδο στον 3Δ χώρο. Για κάθε σημείο (X, Y, Z) του 3Δ χώρου που ανήκει στο επίπεδο (σ, τ, z) και προβάλλεται στο (x_0, y_0) σημείο της εικόνας, ισχύει :

$$3D \rightarrow \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x_0 \cos \tau - y_0 \sin \tau \cos \sigma \\ x_0 \sin \tau + y_0 \cos \tau \cos \sigma \\ z - x_0 \sin \sigma \end{bmatrix} \leftarrow 2D (x_0, y_0) \quad (\text{εξ. 3.1})$$

Θεωρώντας τώρα ότι ένα 3Δ ευθύγραμμο τμήμα δίνεται από τα άκρα του (X_0, Y_0, Z_0) και (X_1, Y_1, Z_1) τα οποία ανήκουν στο επίπεδο (σ, τ, z) , προκύπτει ένα σύστημα 6 εξισώσεων με 7 αγνώστους. Η δέσμη επιπέδων που αντιστοιχεί στις λύσεις του προηγούμενου συστήματος, περιγράφεται από τις παρακάτω εξισώσεις :

$$\sigma = \arctan \left[\frac{Z_0 - Z_1}{(Y_0 - Y_1) \cos \tau + (X_0 - X_1) \sin \tau} \right] \quad (\text{εξ. 3.2})$$

$$z = Z_0 + \left[\frac{Z_0 - Z_1}{(Y_0 - Y_1) \cos \tau + (X_0 - X_1) \sin \tau} \right] (Y_0 \cos \tau - X_0 \sin \tau) \quad (\text{εξ. 3.3})$$

Με δεδομένα δύο 3Δ ευθύγραμμο τμήματα, είναι δυνατό να βρεθεί το επίπεδο στο οποίο ανήκουν υπολογίζοντας το σημείο τομής των δύο καμπυλών που περιγράφονται από τις προηγούμενες εξισώσεις. Αυτό γίνεται με την βοήθεια μιας διαδικασίας «ψηφοφορίας» (voting process) μέσα σε έναν 2Δ χώρο $(\sigma - \tau)$, όπου κάθε σημείο (σ, τ) περιέχει πληροφορία για την παράμετρο z και την ετικέτα (label) του «ψηφίζοντος» ευθύγραμμου τμήματος. Τα επίπεδα ανιχνεύονται ως τοπικά μέγιστα αυτού του χώρου.

Με αυτόν τον τρόπο, φαίνεται πως είναι δυνατό να εξαχθεί πληροφορία για την θέση ενός αντικειμένου, χωρίς η διαδικασία της ταυτοποίησης να γίνεται μεταξύ των γραμμών δυο διαδοχικών εικόνων, αλλά μεταξύ των δεδομένων γραμμών και των επιπέδων του 3Δ χώρου σε κάθε μια ξεχωριστά.

3.3.4 Χρήση διαφορετικών μεθόδων με βάση τα χαρακτηριστικά (features)

Εναλλακτικές προσεγγίσεις στο ίδιο πρόβλημα, θα μπορούσαν να προκύψουν από την επιλογή διαφορετικών χαρακτηριστικών για ταυτοποίηση και παρακολούθηση. Όπως θα μπορούσε να αντιληφθεί κανείς, κάτι τέτοιο θα εξαρτιόταν άμεσα από την μορφή του αντικειμένου. Έτσι, το σύνολο των χαρακτηριστικών (feature set) θα ήταν δυνατό να περιλαμβάνει ιδιότητες όπως την θέση, το εμβαδόν, το χρώμα ή την περίμετρο μιας περιοχής, ενώ η διαδικασία της ταυτοποίησης θα προσπαθούσε να βρει περιοχές σε διαδοχικές εικόνες με παρόμοιες τιμές για όλες ή για κάποιες από αυτές τις ιδιότητες.

Μια διαφορετική μέθοδος θα προέκυπτε βασιζόμενη στην παρατήρηση ότι ένα αντικείμενο αναπαριστάται, όχι μόνο από τα εξαγόμενα χαρακτηριστικά αυτά καθαυτά, αλλά και από τις σχέσεις που έχουν μεταξύ τους. Οι σχέσεις αυτές μπορούν να είναι χωρικές ή κάποιου άλλου τύπου. Σε αυτή την περίπτωση, το αντικείμενο αναπαριστάται από ένα γράφο, οι κόμβοι του οποίου αντιστοιχούν στα χαρακτηριστικά (features) και οι ακμές που συνδέουν τους κόμβους αντιπροσωπεύουν τις μεταξύ τους σχέσεις. Με αυτό τον τρόπο, το πρόβλημα της αναγνώρισης του αντικειμένου καταλήγει να είναι πρόβλημα ταυτοποίησης γράφου (graph matching problem). Το πρότυπο για αυτή την αναγνώριση – ταυτοποίηση είναι δυνατό, είτε να προέρχεται από κάποια βάση δεδομένων, είτε να θεωρηθεί ως τέτοιο το αντικείμενο που ανιχνεύεται στην εικόνα (frame), που προηγείται χρονικά της εξεταζόμενης.

Με δεδομένους δύο γράφους G_1, G_2 που περιέχουν κόμβους N_{ij} , όπου i και j δηλώνουν τον αριθμό του γράφου και του κόμβου αντίστοιχα, οι σχέσεις μεταξύ ενός κόμβου j και ενός άλλου k είναι R_{ijk} . Το επόμενο βήμα είναι να καθοριστεί ένα μέτρο ομοιότητας για τους γράφους που να παίρνει υπόψη του τις ομοιότητες μεταξύ των κόμβων αλλά και μεταξύ των σχέσεων. Η επιλογή του μέτρου αυτού συνδέεται άμεσα με τις ιδιαιτερότητες της εκάστοτε εφαρμογής.

Αυτή είναι η γενικότερη ιδέα και ο συμβολισμός που χρησιμοποιείται για να περιγράψει τα βασικά δομικά στοιχεία ενός γράφου. Παρακάτω θα υιοθετηθεί ένας βολικότερος συμβολισμός για να παρουσιαστούν κάποια συγκεκριμένα μέτρα «ομοιότητας». Συμβολίζοντας, λοιπόν, τους δύο γράφους με (V_1, E_1) και (V_2, E_2) , τα σύνολα που περιέχουν τους κόμβους θ_i, θ_j , είναι τα V_1, V_2 ενώ E_1, E_2 τα σύνολα που αποτελούνται από τις ακμές $\varepsilon_1, \varepsilon_2$ του γράφου. Στόχος ενός τέτοιου μέτρου «ομοιότητας» ή μιας διαδικασίας ταυτοποίησης αν προτιμάτε, θα μπορούσε να είναι η αναζήτηση μιας ένα προς ένα (1:1) αντιστοιχίας (ισομορφισμού) f ανάμεσα στα V_1, V_2 τέτοια ώστε :

$$\text{για } \theta_1, \theta_2 \in V_1, V_2 \quad f(\theta_1) = \theta_2$$

$$\text{και } \forall \varepsilon_i \in E_1 \text{ που συνδέει οποιοδήποτε ζεύγος κόμβων } \theta_1, \theta'_1 \in V_1$$

$$\exists \varepsilon_j \in E_2 \text{ που συνδέει τα } f(\theta_1) \text{ και } f(\theta'_1).$$

Η διαδικασία αυτή αναφέρεται ως ισομορφισμός γράφου (graph isomorphism) και χρησιμοποιείται μόνο στις περιπτώσεις όπου το αντικείμενο είναι ορατό στην ολότητά του.

Στην περίπτωση ενός μερικώς ορατού αντικειμένου βρίσκει εφαρμογή ο ισομορφισμός υπο – γράφου (subgraph isomorphism). Σε αυτήν την περίπτωση αναζητείται ισομορφισμός ανάμεσα σε ένα γράφο (V_1, E_1) και ένα υπο – γράφο ενός άλλου (V_2, E_2) .

Ένα πρόβλημα των προσεγγίσεων αυτών είναι ότι ο ισομορφισμός γράφου είναι NP πρόβλημα. Έτσι, ακόμα και για έναν λογικού μεγέθους γράφο, ο χρόνος που απαιτείται για την ταυτοποίηση είναι δυνατό να αποδειχθεί απαγορευτικός. Αυτό που μπορεί να βοηθήσει την κατάσταση είναι η επιπλέον πληροφορία που είναι διαθέσιμη μέσω των ιδιοτήτων που έχουν οι κόμβοι (υπενθυμίζεται ότι οι κόμβοι αναπαριστούν χαρακτηριστικά (features) του αντικειμένου). Η πληροφορία αυτή χρησιμοποιείται από διάφορους ευρετικούς (heuristic) αλγορίθμους, που έχουν προταθεί για την λύση του προβλήματος της ταυτοποίησης του γράφου (graph matching problem). Ένας τέτοιος αλγόριθμος θα πρέπει να συνυπολογίζει :

- Την μεταβλητότητα των ιδιοτήτων και των σχέσεων
- Την απουσία ιδιοτήτων ή σχέσεων
- Το γεγονός ότι ένα «μοντέλο – πρότυπο» είναι μια αφαιρετική οντότητα (abstraction) από κλάσεις αντικειμένων
- Το γεγονός ότι οι διάφορες «στάσεις» (instances) μπορεί να περιέχουν επιπλέον πληροφορία.

Με όλα τα παραπάνω υπόψη, ένας τρόπος για να οριστεί ένα μέτρο ομοιότητας είναι αυτός που παρουσιάζεται και στο [2]. Σε αυτή την θεώρηση οι ακμές του γράφου αντιμετωπίζονται ως ελατήρια που ενώνουν τις μάζες που βρίσκονται τοποθετημένες στους κόμβους. Τότε, η ποιότητα μιας ταυτοποίησης είναι μια συνάρτηση του πόσο καλά τοποθετούνται τα χαρακτηριστικά (features) πάνω στο μοντέλο και της ποσότητας της «ενέργειας» που απαιτείται για να «τεντώσουν» τα ελατήρια έτσι ώστε να συμπίσουν τα δεδομένα με το μοντέλο – πρότυπο.

Δηλαδή :

$$\begin{aligned}
 C = & \sum_{d \in R_1} \text{κόστος χαρακτηριστικού}(d, F(d)) \\
 & + \sum_{(d,e) \in R_2} \text{κόστος ελατηρίου}(F(d), F(e)) \quad (\text{εξ. 3.4}) \\
 & + \sum_{c \in R_3} \text{κόστος απουσίας χαρακτηριστικού}(c)
 \end{aligned}$$

όπου R_1 το σύνολο των δεδομένων d που βρήκαν αντίστοιχο στο πρότυπο $F(d)$, R_2 το σύνολο $R_1 \times \{\text{δεδομένα που βρίσκονται στην εικόνα}\}$ και $R_3 = \{\text{απόντα από το πρότυπο}\} \cup \{\text{απόντα από την εικόνα}\}$.

Είναι προφανές ότι η περιγραφή αυτή είναι πολύ γενική. Έτσι, τα κόστη που εμφανίζονται στα αθροίσματα παίρνουν την μορφή που θα καθορίσει η κάθε εφαρμογή.

3.4 Μέθοδοι παρακολούθησης βασισμένες σε μοντέλα (model based)

3.4.1 Εισαγωγή

Οι τεχνικές αυτής της παραγράφου έχουν στόχο, όπως και προηγουμένως, να εξάγουν από μια ακολουθία εικόνων, πληροφορία θέσης και ταχύτητας για το αντικείμενο που φαίνεται σ' αυτές. Η διαδικασία αυτή, τώρα, θα βασιστεί σε κάποια εκ των προτέρων γνώση για την μορφή του αντικειμένου, με την βοήθεια της οποίας, δημιουργείται ένα 3Δ μοντέλο. Ο τύπος του μοντέλου ποικίλει ανάλογα με την περίπτωση και μπορεί να περιλαμβάνει σημεία, γραμμές ή οποιοδήποτε άλλο χαρακτηριστικό γνώρισμα του αντικειμένου. Τελικώς το ζητούμενο, δηλαδή η πληροφορία θέσης του αντικειμένου, προκύπτει από την ταυτοποίηση των δεδομένων που λαμβάνονται από την εικόνα με τα αντίστοιχα χαρακτηριστικά του μοντέλου.

3.4.2 Δημιουργία 3Δ μοντέλου

Το πρώτο βήμα για κάθε μέθοδο παρακολούθησης που βασίζεται σε μοντέλο είναι η δημιουργία του τελευταίου. Όπως προαναφέρθηκε, ανάλογα με την εφαρμογή, η πληροφορία που περιέχεται σ' αυτό είναι διαφορετική. Συνήθως μια τεχνική που αποδεικνύεται βολική, είναι να θεωρηθεί 3Δ μοντέλο σημείων που ενώνονται με ευθύγραμμα τμήματα. Ως τέτοιο, είναι δυνατό να δοθεί σε μορφή πίνακα όπως αναφέρθηκε στο 2^ο κεφάλαιο και φαίνεται παρακάτω :

$$Model_w = \begin{bmatrix} X_0 & X_1 & \dots\dots & X_{N-1} \\ Y_0 & Y_1 & \dots\dots & Y_{N-1} \\ Z_0 & Z_1 & \dots\dots & Z_{N-1} \\ 1 & 1 & \dots\dots & 1 \end{bmatrix}.$$

Σε αυτή την μορφή είναι δυνατό να υποστεί οποιονδήποτε μετασχηματισμό μετατόπισης και περιστροφής, ενώ η προβολή του στο επίπεδο της εικόνας $Model_c$ προκύπτει εύκολα πολλαπλασιάζοντας από αριστερά με τον πίνακα του προοπτικού μετασχηματισμού.

$$Model_c = P Model_w \quad (\text{εξ. 3.5})$$

Όπως θα φανεί και παρακάτω, η διαδικασία της προβολής είναι κομβικής σημασίας για την ταυτοποίηση της εικόνας και του μοντέλου.

3.4.3 Ταυτοποίηση σε δύο διαστάσεις (2D matching)

Έχοντας εξάγει τα δεδομένα από την εικόνα, υπολογίζεται μια αντιστοιχία μοντέλου και δεδομένων, με βάση την οποία βρίσκεται η βέλτιστη τοποθέτηση (best fit position) του πρώτου ως προς τα δεύτερα. Προηγουμένως, το μοντέλο έχει ήδη προβληθεί στο επίπεδο της εικόνας. Η προβολή αυτή δεν εξαντλείται στην εφαρμογή του πίνακα του προοπτικού μετασχηματισμού P στον πίνακα που περιγράφει το μοντέλο στον 3D χώρο, αλλά συμπληρώνεται από κάποια τεχνική απόκρυψης των ακμών του μοντέλου που δεν είναι ορατές από την συγκεκριμένη θέση της κάμερας. Η όλη διαδικασία για να επιτευχθεί ταυτοποίηση σε 2D, διακρίνεται στα παρακάτω επιμέρους βήματα [4] :

1. Καθορισμός χώρου αναζήτησης πιθανών αντιστοιχιών μοντέλου και δεδομένων

Ένας τέτοιος χώρος αναζήτησης θα μπορούσε να είναι γράφος, οι κόμβοι του οποίου αντιστοιχούν σε μια συγκεκριμένη αντιστοιχία δεδομένων και μοντέλου και αποτελούν τις καταστάσεις (state). Οι κόμβοι συνδέονται με γραμμές που αντιπροσωπεύουν μεταβάσεις. Κάθε μετάβαση προσθέτει ή αφαιρεί k γραμμές από την αντιστοιχία μοντέλου \leftrightarrow δεδομένων.

Η επιλογή της αρχικής κατάστασης μπορεί να γίνει :

- A) με εύρεση χαρακτηριστικών (key features)
- B) με γενικευμένο μετασχηματισμό Hough και επιλεγμένα key features

2. Καθορισμός μεθόδου επιλογής της βέλτιστης χωρικής τοποθέτησης (fit) με δεδομένη μια αντιστοιχία μοντέλου και δεδομένων

Για την βελτιστοποίηση της τοποθέτησης (fit), υπολογίζονται σε κάθε κατάσταση (state) τα s, R, T (scale, Rotation, Translation) που ελαχιστοποιούν το σφάλμα τοποθέτησης του μοντέλου στις μέχρι εκείνη τη στιγμή δεδομένες γραμμές. Μια μορφή σφάλματος είναι και η παρακάτω :

$$E_{fit} = \sum_i (l_i (n_i \cdot (R_{(p_i)} + T) - s \rho_i)^2) \quad (\text{εξ. 3.6})$$

Κάθε όρος του αθροίσματος αντιστοιχεί στο τετράγωνο της απόστασης από την δεδομένη γραμμή ως την γραμμή του μοντέλου. Το ρ_i είναι η απόσταση από την αρχή των αξόνων ως την γραμμή δεδομένων L_i στην κατεύθυνση του μοναδιαίου διανύσματος n_i . Το n_i είναι το κάθετο διάνυσμα στο επίπεδο του χώρου που βρίσκεται η γραμμή. Η ποσότητα l_i είναι το μήκος της γραμμής.

3. Επιλογή αντικειμενικού μέτρου για την αξιολόγηση της χωρικής τοποθέτησης, δεδομένης μιας αντιστοιχίας μοντέλου και δεδομένων

Η αντικειμενική συνάρτηση καθορίζει το σφάλμα ανάμεσα στο βέλτιστο τοποθετημένο μοντέλο και στις αντίστοιχες δεδομένες γραμμές. Υπάρχουν δύο τύποι σφάλματος που συνυπολογίζονται :

- A) οι δεδομένες γραμμές (data lines) είναι μετατοπισμένες σε σχέση με αυτές του μοντέλου
 B) μέρος του μοντέλου λείπει από τα data

$$E_{match} = \sum (E_{fit}(L) + \alpha E_{omission}(L)) = E_{fit} + \alpha \sum E_{omission}(L) \quad (\text{εξ. 3.7})$$

όπου $E_{omission} = e^{\beta P}$ με το P να εκφράζει το ποσοστό της γραμμής μοντέλου που λείπει από τα δεδομένα (data).

Οι συντελεστές α , β είναι σταθεροί με τυπικές τιμές 0.5 και 4 αντίστοιχα ($\alpha=0.5$, $\beta=4$).

3.4.4 Ταυτοποίηση σε τρεις διαστάσεις (3D matching)

Εναλλακτική της μεθόδου που θέλει την διαδικασία της ταυτοποίησης να γίνεται στο επίπεδο της εικόνας είναι η τεχνική της ταυτοποίησης στον 3D χώρο. Αντί, τώρα, να προβάλλεται το μοντέλο στο επίπεδο της εικόνας, γίνεται προσπάθεια να ανακατασκευαστεί το αντικείμενο στον χώρο από τα δεδομένα που έχουν εξαχθεί από την εικόνα. Η διαδικασία αυτή, απαιτεί την χρήση κάποιας από τις μεθόδους που αναπτύχθηκαν στο κεφάλαιο 2 και αφορούν στην ανακατασκευή του 3D χώρου από τα δεδομένα της εικόνας. Αυτό κρίνεται απαραίτητο, αφού, σε αντίθεση με την ταυτοποίηση σε δύο διαστάσεις όπου χρειάζεται απλά μια προβολή του μοντέλου, η ποσότητα πληροφορίας που είναι διαθέσιμη από μια μόνο εικόνα δεν είναι επαρκής.

Έχοντας, λοιπόν, διαθέσιμα την ανακατασκευή του αντικειμένου και το μοντέλο, ακολουθεί η ταυτοποίησή τους με χρήση κάποιων μέτρων απόκλισης που οφείλεται σε μετατόπιση ή σε περιστροφή. Το απλούστερο από όλα αυτά είναι η ευκλείδεια απόσταση μεταξύ των αντίστοιχων σημείων ή ακμών του μοντέλου και των δεδομένων, υπό την προϋπόθεση ότι μια τέτοια αντιστοίχιση είναι εφικτή. Εναλλακτικά, θα μπορούσε να χρησιμοποιηθεί μια πολυπλοκότερη συνάρτηση σφάλματος τοποθέτησης, στα πρότυπα όσων αναφέρθηκαν στην παράγραφο 3.3.3. Συμπληρωματικά η συνάρτηση θα επαναυπολογίζεται για διαφορετικές εκτιμήσεις των διανυσμάτων περιστροφής και μετατόπισης του μοντέλου έτσι ώστε να επιλέγεται ή βέλτιστη τοποθέτηση του πάνω στα δεδομένα.

Γενικά, η επίλυση παρόμοιων προβλημάτων στον 3D χώρο παρουσιάζει αυξημένη δυσκολία σε σχέση πάντα με την λύση τους στον 2D χώρο. Η διαδικασία της ανακατασκευής, πέρα από χρονοβόρα, δεν είναι απαλλαγμένη από πιθανά σφάλματα δίνοντας έτσι ισχυρή ένδειξη ότι το εγγενές έλλειμμα πληροφορίας που την διακρίνει δεν είναι εύκολο να ξεπεραστεί. Σε αντιστοιχία, η διαδικασία ταυτοποίησης παρουσιάζει και αυτή μεγαλύτερη πολυπλοκότητα αφού η όποια συνάρτηση σφάλματος τοποθέτησης χρησιμοποιηθεί, λογικά, θα είναι συνθετότερη με προφανείς επιπτώσεις στην ταχύτητα υπολογισμού. Οι παραπάνω είναι αρκετοί λόγοι για τους οποίους προτιμάται συνήθως η ταυτοποίηση σε δύο και όχι σε τρεις διαστάσεις. Παρόλα αυτά, η ταυτοποίηση σε τρεις διαστάσεις βρίσκει εφαρμογή σε περιπτώσεις όπου είναι απαραίτητη μια λεπτομερέστερη περιγραφή του χώρου. Η αρτιότητα στην ανακατασκευή του 3D χώρου θα μπορούσε να είναι ένα βήμα μπροστά προς την γενικότερη κατανόηση του θέματος της όρασης, μηχανικής ή μη.

Κεφάλαιο 4

Ανάπτυξη και Υλοποίηση αλγορίθμων Δυναμικής Όρασης (Dynamic Vision)

(Στατική κάμερα – κινούμενο αντικείμενο)

4.1 Εισαγωγή

Κάθε σύστημα δυναμικής όρασης, όπως αυτό που επιχειρεί να εξετάσει αυτή η εργασία, αποτελείται από τρία βασικά στάδια ανάλυσης [2] :

- Περιφερειακή ανάλυση (Peripheral)
- Ανάλυση εστίασης της προσοχής (Attentive)
- Γνωστικό επίπεδο (Cognitive)

Στο στάδιο της περιφερειακής ανάλυσης εξάγεται, προσεγγιστικά, πληροφορία η οποία θα αποδειχτεί χρήσιμη στα επόμενα στάδια. Αυτή η πληροφορία είναι ενδεικτική της δραστηριότητας στην εικόνα και μπορεί να χρησιμοποιηθεί για να εντοπιστούν οι περιοχές της εικόνας που χρειάζονται προσεκτική ανάλυση.

Στο επόμενο στάδιο της εστίασης της προσοχής (Attentive), η ανάλυση επικεντρώνεται στα «ενεργά» μέρη της εικόνας και εξάγει συμπεράσματα τα οποία θα μπορούσαν να φανούν χρήσιμα στην αναγνώριση κάποιων αντικειμένων, στην ανάλυση των χαρακτηριστικών της κίνησής τους ή σε άλλες σχετικές ιδιότητες.

Το τελικό στάδιο της ανάλυσης στο γνωστικό επίπεδο (Cognitive), συνδυάζει την γνώση που έχει το σύστημα για τα αντικείμενα, την κίνησή τους ή ενδεχομένως για οποιαδήποτε άλλη έννοια που μπορεί να εξαρτάται από την φύση της κάθε εφαρμογής, έτσι ώστε να αναλύσει την 3Δ σκηνή ως προς τα αντικείμενα που είναι παρόντα και τα γεγονότα που συμβαίνουν.

Η είσοδος σε ένα τέτοιο σύστημα είναι μια ακολουθία από εικόνες (frame sequence) αναπαριστώμενες από μια συνάρτηση της μορφής $F(x,y,t)$ όπου x,y οι χωρικές συντεταγμένες στο επίπεδο της εικόνας την χρονική στιγμή t . Επειδή κατά κανόνα, η δειγματοληψία των διαδοχικών εικόνων γίνεται με σταθερό ρυθμό, θα ήταν καλύτερα να θεωρηθεί ότι το t αντιστοιχεί όχι σε απόλυτο χρόνο, αλλά στην t -οστή εικόνα μιας ακολουθίας.

Με τα παραπάνω υπόψη, υλοποιήθηκαν δύο διαφορετικές προσεγγίσεις στο θέμα. Η πρώτη αφορά στην εξαγωγή κάποιων χαρακτηριστικών pixel από την εικόνα και την παρακολούθησή τους, ενώ η δεύτερη στην ανίχνευση χαρακτηριστικών ακμών και ταίριασμά τους με αυτές ενός μοντέλου.

Ας δούμε αναλυτικά με τη σειρά αυτές τις προσεγγίσεις.

4.2 Ταυτοποίηση χαρακτηριστικών σημείων (pixel feature matching)

4.2.1 Εισαγωγή

Όπως προαναφέρθηκε, η βασική ιδέα πίσω από την ταυτοποίηση χαρακτηριστικών σημείων είναι η δυνατότητα εξαγωγής συμπερασμάτων για τις ιδιότητες της κίνησης που περιλαμβάνεται στην εικόνα, μέσω της ανίχνευσης ενός συνόλου χαρακτηριστικών σημείων και την παρακολούθηση αυτών κατά τη διάρκεια της ακολουθίας των εικόνων.

4.2.2 Περιγραφή αλγορίθμου

Στις παραγράφους που ακολουθούν, επιχειρείται η περιγραφή ενός αλγορίθμου ταιριάσματος (matching) χαρακτηριστικών σημείων μέσα από τα στάδια ανάλυσης που αυτός ακολουθεί.

i) Στάδιο περιφερειακής ανάλυσης (*peripheral analysis*)

Η ανίχνευση της περιοχής της εικόνας όπου υπάρχει δραστηριότητα (στάδιο της περιφερειακής (αρχικής) ανάλυσης) γίνεται υπολογίζοντας την διαφορά ανάμεσα σε δύο διαδοχικά frame. Έπειτα εφαρμόζεται ένα κατώφλι (threshold) στην εικόνα που προέκυψε από την αφαίρεση, έτσι ώστε να ξεχωρίσει το κινούμενο πια αντικείμενο. Αφού γίνει και αυτό, η περιοχή που μας ενδιαφέρει περικλείεται από ένα παραλληλόγραμμο (Minimum Bounding Rectangle).

ii) Στάδιο εστίασης της προσοχής (*attentive analysis*)

Μέσα σε αυτό το MBR εφαρμόζεται ένας τελεστής «ενδιαφέροντος» [2] ο οποίος αναζητά pixel που θα μπορούσαν να χρησιμοποιηθούν ως σημεία αναφοράς για ταιρίασμα (matching) ανάμεσα σε δύο frames. Ως τέτοια θα μπορούσαν να θεωρηθούν τα σημεία της εικόνας με υψηλή διασπορά αφού, προφανώς, τα σημεία των ομοιόμορφων περιοχών δεν είναι κατάλληλα. Ο τελεστής εξετάζει κάθε pixel τοποθετώντας το στο κέντρο ενός παραθύρου S και υπολογίζοντας την διασπορά προς κάθε κατεύθυνση.

$$I_1 = \sum_{(x,y) \in S} [f(x,y) - f(x,y+1)]^2 \quad (\text{εξ. 4.1})$$

$$I_2 = \sum_{(x,y) \in S} [f(x,y) - f(x+1,y)]^2 \quad (\text{εξ. 4.2})$$

$$I_3 = \sum_{(x,y) \in S} [f(x,y) - f(x+1,y+1)]^2 \quad (\text{εξ. 4.3})$$

$$I_4 = \sum_{(x,y) \in S} [f(x,y) - f(x+1,y-1)]^2 \quad (\text{εξ. 4.4})$$

Το παράθυρο S μετακινείται πάνω στην εικόνα όπως ακριβώς μια «μάσκα», ενώ συνήθη μεγέθη είναι 5×5 έως 11×11 ενώ για την παρούσα εφαρμογή επιλέχθηκε να είναι 7×7 .

Η τιμή που δίνει τελικά ο τελεστής στο κεντρικό pixel (x_c, y_c) είναι το ελάχιστο των διασπορών με στόχο να αποφευχθούν σημεία που βρίσκονται ακριβώς πάνω σε ακμές. Τα σημεία αυτά σε μία τουλάχιστον κατεύθυνση, την κατεύθυνση της ακμής, θα έχουν πολύ χαμηλή, αν όχι μηδενική, διασπορά. Ο λόγος για τον οποίο απορρίπτονται τα σημεία των ακμών, είναι ότι ένα τέτοιο σημείο σε ένα frame θα μπορούσε να αντιστοιχεί με όλα τα σημεία της ακμής στο επόμενο, καθιστώντας δυσχερή τον υπολογισμό της μετατόπισης.

Έτσι προκύπτει :

$$I(x_c^s, y_c^s) = \min(I_1, I_2, I_3, I_4) \quad (\text{εξ. 4.5})$$

Η διαδικασία αυτή, είναι δυνατό να αναδειξεί ως «ενδιαφέροντα» περισσότερα από ένα σημεία για κάποιο χαρακτηριστικό (feature). Προς αποφυγή αυτού, επιλέγονται τα pixel αυτά στα οποία ο τελεστής εμφανίζει τοπικά μέγιστα. Επιπλέον, ένα τέτοιο σημείο θεωρείται καλό και τελικώς επιλέγεται, αν αυτό το τοπικό μέγιστο βρίσκεται πάνω από μια συγκεκριμένη, προκαθορισμένη τιμή κατωφλίου.

Έχοντας εξάγει τα χαρακτηριστικά σημεία S_i^1, S_j^2 από δύο διαδοχικά frames, ακολουθεί η διαδικασία του ταιριάσματος (matching) συγκρίνοντας ως προς την ομοιότητά τους τις δύο περιοχές γύρω από αυτά τα pixel. Το μέτρο της ομοιότητας δίνεται από τον παρακάτω συντελεστή συσχέτισης (correlation coefficient):

$$r(d_x, d_y) = \frac{\sum_{(x,y) \in S} [f_1(x, y) - \bar{f}_1][f_2(x + d_x, y + d_y) - \bar{f}_2]}{\left\{ \sum_{(x,y) \in S} [f_1(x, y) - \bar{f}_1]^2 \sum_{(x,y) \in S} [f_2(x + d_x, y + d_y) - \bar{f}_2]^2 \right\}^{1/2}} \quad (\text{εξ. 4.6})$$

όπου d_x, d_y οι αποστάσεις των δύο pixel κατά τους αντίστοιχους άξονες και \bar{f}_1, \bar{f}_2 οι μέσες εντάσεις των pixel των συγκρινόμενων περιοχών.

Μια εναλλακτική μέθοδος στον υπολογισμό του συντελεστή συσχέτισης, η οποία και τελικά ακολουθήθηκε, είναι η χρήση του μέτρου του gradient, αντί της τιμής φωτεινότητας, σε κάθε pixel και η αντιστοίχιση των αποτελεσμάτων σε τρεις τιμές, -1, 0 και 1, με την χρήση δυο κατωφλίων. Με αυτόν τον τρόπο, ο συντελεστής συσχέτισης απλοποιείται στο άθροισμα των γινομένων των μέτρων του gradient των αντίστοιχων pixel.

$$r(d_x, d_y) = \sum_{(x,y) \in S} g_1(x, y) g_2(x + d_x, y + d_y) \quad (\text{εξ. 4.7})$$

όπου g_1, g_2 οι πίνακες που περιέχουν τις τιμές του gradient για τις εικόνες f_1, f_2 αντίστοιχα.

Από τους πίνακες $\mathcal{G}_1, \mathcal{G}_2$ δεν είναι απαραίτητο να υπολογιστούν, παρά μόνο η περιοχή εκείνη των στοιχείων που περιέχει τα χαρακτηριστικά pixel (feature points). Για κάθε λοιπόν πιθανό ζεύγος (match) χαρακτηριστικών σημείων (S_i^1, S_j^2) , υπολογίζεται ο συντελεστής συσχέτισης ανάμεσα στις δύο περιοχές $r_{i,j}(d_x, d_y)$. Ακολούθως, για κάθε χαρακτηριστικό σημείο του ενός frame επιλέγεται ως αντίστοιχό του το σημείο του άλλου frame για το οποίο μεγιστοποιείται ο συντελεστής συσχέτισης. Δηλαδή επιλέγεται το

$$\max_j [r_{i,j}(d_x, d_y)].$$

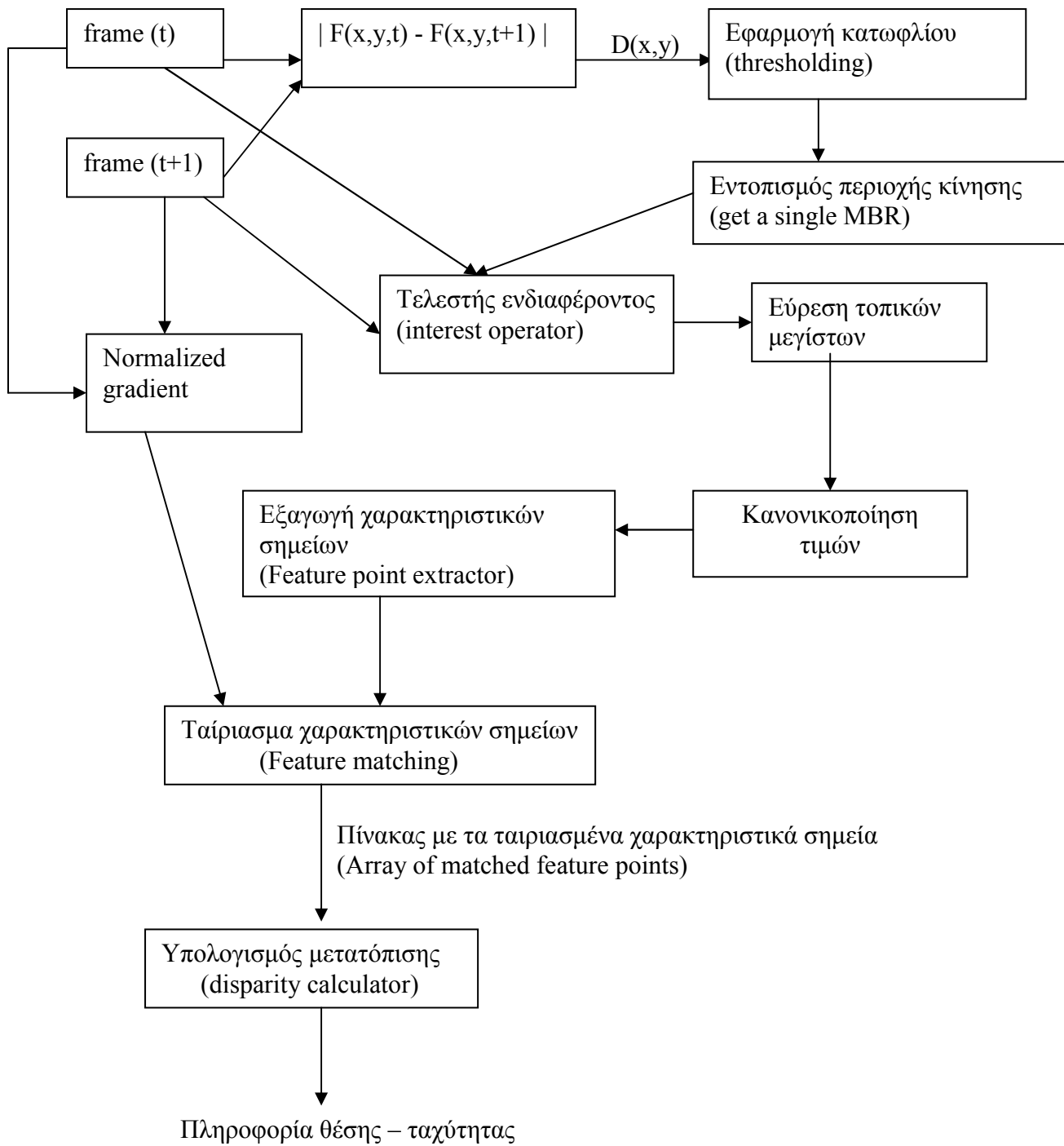
Αυτή η τακτική, όμως, δεν εγγυάται μία ένα προς ένα αντιστοίχιση. Για να ξεπεραστεί αυτό το πρόβλημα εισάγεται το κριτήριο της εγκυρότητας για κάθε αντιστοίχιση (match). Ένα ταίριασμα στοιχείων (match) θεωρείται έγκυρο αν κανένα από τα δύο του μέλη δεν συμμετέχει σε κάποιο άλλο ζεύγος με μεγαλύτερο συντελεστή συσχέτισης.

Στο τέλος επιλέγονται όλα τα έγκυρα ζεύγη σημείων για να αποτελέσουν στοιχεία ενός πίνακα αντιστοίχισης (matching array). Από το πίνακα αυτόν εξάγονται συμπεράσματα για την μετατόπιση του αντικειμένου για το χρονικό διάστημα ανάμεσα στη λήψη των δύο εικόνων. Η μετατόπιση (disparity) υπολογίζεται για κάθε ζεύγος του πίνακα και σταθμίζεται με τον συντελεστή συσχέτισης που αντιστοιχεί στο ζεύγος. Τελικά, η εκτιμώμενη μετατόπιση κατά την οριζόντια και την κατακόρυφη έννοια προκύπτει ως το κανονικοποιημένο άθροισμα των σταθμισμένων μετατοπίσεων που υπολογίστηκαν προηγουμένως, εκπεφρασμένη σε pixel.

iii) Γνωστικό επίπεδο ανάλυσης

Για να είναι ικανό το σύστημα να εισέλθει σ' αυτό το επίπεδο ανάλυσης και να εξάγει συμπεράσματα με κάποια σημασία, θα πρέπει να είναι διαθέσιμες πληροφορίες για το υπό παρακολούθηση αντικείμενο, το περιβάλλον κτλ. Με άλλα λόγια απαιτείται μια αυστηρώς προδιαγεγραμμένη εφαρμογή, η οποία δεν είναι το αντικείμενο αυτής της εργασίας.

Με την ολοκλήρωση της περιγραφής του αλγόριθμου, δίνεται στην επόμενη σελίδα στο Σχήμα 4.1, το δομικό διάγραμμά του (block diagram).



Σχήμα 4.1 : Δομικό διάγραμμα του αλγόριθμου ταιριάσματος χαρακτηριστικών σημείων (feature point matching)

4.2.3 Αναλυτικότερη περιγραφή του αλγόριθμου

Στην παράγραφο αυτή θα παρουσιαστεί ο αλγόριθμος ταιριάσματος (ταυτοποίησης) χαρακτηριστικών σημείων, σε επίπεδο υλοποίησης. Το πρόγραμμα `feature.exe`, δέχεται μια ακολουθία εικόνων `frame_00x.bmp`, όπου `x` αύξων αριθμός και υπολογίζει την μετατόπιση του αντικειμένου σε `pixel`, ανάμεσα σε δύο διαδοχικές εικόνες.

Όπως και στην περίπτωση της αντιστροφής της προοπτική προβολής, θα πρέπει αρχικά να καθοριστούν οι σταθερές που εξαρτώνται από την κάμερα :

`pixel_per_meter_H` = Οριζόντια ανάλυση (pixels) / Μήκος επιπέδου εικόνας (m)

`pixel_per_meter_V` = Κατακόρυφη ανάλυση (pixels) / Πλάτος επιπέδου εικόνας (m)

Το σημείο αυτό είναι κοινό και για τους δύο αλγόριθμους δυναμικής όρασης που υλοποιήθηκαν, όπως επίσης και κάποιες βασικές συναρτήσεις. Οι τελευταίες παρουσιάζονται ακολούθως :

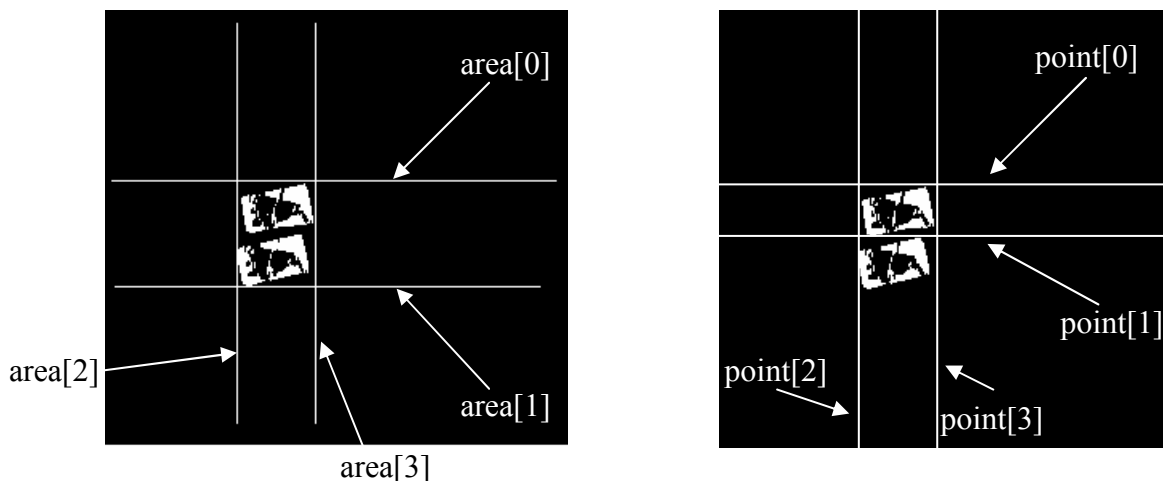
Βασικές συναρτήσεις

`void thres(int threshold,unsigned char **pic,unsigned char **target)` : εφαρμόζει ένα κατώφλι (`threshold`) σε μια εικόνα (`pic`) και αποδίδει την τελική εικόνα (`target`).

`void thres2(int threshold,unsigned char **pic,unsigned char **target)` : όμοια με την προηγούμενη ως προς τα ορίσματα και την λειτουργία, με την διαφορά ότι το κατώφλι εφαρμόζεται στην ανεστραμμένη εικόνα (`pic`). Χρήσιμη στην περίπτωση που το υπόβαθρο (`background`) είναι λευκό.

`void difference(unsigned char **Im1,unsigned char **Im2,unsigned char **target)` : αποδίδει την απόλυτη τιμή της διαφοράς δύο εικόνων (`Im1`, `Im2`) σε μια τρίτη (`target`).

`void get_a_single_MBR(unsigned char **Im1,int* area,int *point)` : αποδίδει το ελάχιστο περιβάλλον παραλληλόγραμμο (`point`) για ένα αντικείμενο μέσα σε μια περιοχή (`area`) της εικόνας (`Im1`).



Εικόνα 4.1 : Λειτουργία της συνάρτησης `get_a_single_MBR`

Συναρτήσεις που αφορούν τον αλγόριθμο (feature matching)

Πέρα από τις βασικές συναρτήσεις, υπάρχουν και αυτές που αφορούν αποκλειστικά τον αλγόριθμο ταύτισης χαρακτηριστικών σημείων (feature point matching). Οι συναρτήσεις αυτές παρουσιάζονται παρακάτω.

`void interest(int window,unsigned char **Im,float **target,int *area) :`

Εφαρμόζει τον τελεστή ενδιαφέροντος (interest operator) χρησιμοποιώντας παράθυρο μεγέθους (window)², μέσα σε μια περιοχή (area) της αρχικής εικόνας (Im1) και αποθηκεύει τα αποτελέσματα στον πίνακα (target).

`void normalize(float **farray,unsigned char **target) :`

Κανονικοποιεί τα στοιχεία του πίνακα (farray) στο διάστημα [0,255] και τοποθετεί τις αντίστοιχες τιμές στον πίνακα (target). Η διαδικασία είναι χρήσιμη για την εφαρμογή κατωφλίου στα δεδομένα.

`void local_max_nxn(float **farray,float **target,int *area,int n) :`

Αναζητεί τοπικά μέγιστα μέσα σε παράθυρο nxn ενός υποπίνακα (area) του farray και τα τοποθετεί στις αντίστοιχες θέσεις του πίνακα target.

`int get_feature_points(unsigned char **Im1,int **feature,unsigned char featureThreshold) :`

Μέσα σε μια εικόνα (Im1) βρίσκει τα σημεία για τα οποία ο τελεστής ενδιαφέροντος βρίσκεται πάνω από ένα κατώφλι (featureThreshold), κρατάει τις συντεταγμένες τους (i,j) σε πίνακα (feature) και επιστρέφει το πλήθος τους.

`void save_feature_points(int **feature,char *filename)`

Αποθηκεύει σε ένα αρχείο με όνομα (filename) ένα πίνακα με συντεταγμένες χαρακτηριστικών σημείων (feature). Η χρήση της είναι σχεδόν αποκλειστικά για τη διαδικασία διόρθωσης προγραμματιστικών σφαλμάτων (debugging).

`void normalized_gradient(unsigned char **Im,char **target,int T1,int T2,int*area)`

Σε μια περιοχή (MBR) της εικόνας (Im) υπολογίζει το gradient για κάθε pixel το κανονικοποιεί σε μείον ένα (-1) για τιμές μικρότερες του (T1), σε μηδέν (0) για τιμές ανάμεσα σε (T1) και (T2), σε ένα (1) για τιμές μεγαλύτερες του (T2) και το αποθηκεύει σε πίνακα (target).

`int **match_feature_norm(char **Im1, char **Im2,int **featureA,int **featureB,
int *num_of_featuresA,int num_of_featuresB,float*r_in)`

Η συνάρτηση αυτή παίρνει σαν είσοδο δύο εικόνες (Im1), (Im2), τους αντίστοιχους πίνακες χαρακτηριστικών σημείων για κάθε εικόνα (featureA), (featureB) και το πλήθος τους (num_of_featuresA), (num_of_featuresB) και επιστρέφει ένα πίνακα με τα ταιριασμένα σημεία μαζί με τις αντίστοιχες τιμές του συντελεστή συσχέτισης (r_in). Ο πίνακας που επιστρέφει η συνάρτηση έχει μέγεθος Nx2 όπου N ο αριθμός των έγκυρων ταιριασμάτων, ενώ τα δύο στοιχεία κάθε γραμμής είναι δείκτες στους αντίστοιχους πίνακες χαρακτηριστικών σημείων των εικόνων (featureA), (featureB).

`void print_features(unsigned char **frame1,unsigned char** frame2,
int **featuresA,int**featuresB,int num_of_features,int **match_array)`

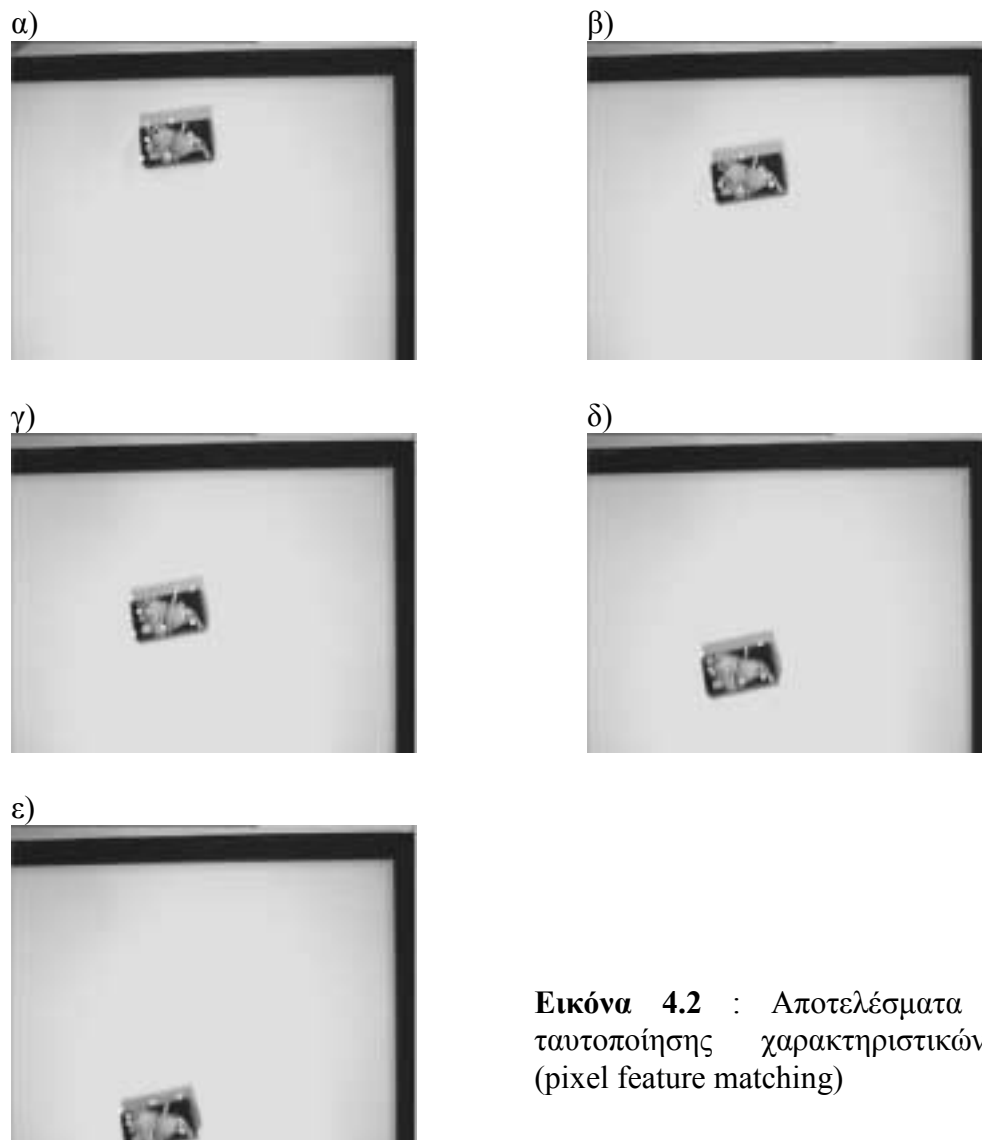
Αποτυπώνει τα χαρακτηριστικά σημεία (featuresA), (featuresB) που έχουν ταιριαστεί (match_array), πάνω σε δύο εικόνες (frame1), (frame2).

```
void calculate_disparity(int **featuresA,int **featuresB,int num_of_features,
                        int **match_array, float *di, float *dj, float *r)
```

Εισάγονται τα χαρακτηριστικά σημεία (featuresA), (featuresB), το πλήθος των πρώτων (num_of_features), ο πίνακας που περιέχει τις μεταξύ τους αντιστοιχίες (match_array) καθώς και ο πίνακας που περιέχει το συντελεστή συσχέτισης της κάθε μιας, ενώ υπολογίζονται και επιστρέφονται οι μετατοπίσεις στους άξονες i (d_i) και j (d_j).

4.2.4 Αποτελέσματα και συζήτηση

Η διαδικασία εφαρμόστηκε σε μια σειρά από frames με τα παρακάτω αποτελέσματα :



Εικόνα 4.2 : Αποτελέσματα αλγορίθμου ταυτοποίησης χαρακτηριστικών σημείων (pixel feature matching)

Στην ακολουθία των εικόνων φαίνονται σημειωμένα με κουκκίδες τα χαρακτηριστικά σημεία κάθε εικόνας. Ο αριθμός τους, όπως φαίνεται και με απλή επισκόπηση, δεν παραμένει σταθερός. Αυτό όμως δεν αποτελεί ουσιαστικό πρόβλημα υπό την προϋπόθεση ότι μερικά χαρακτηριστικά σημεία επιβιώνουν ανάμεσα σε δυο διαδοχικές εικόνες. Ένα άλλο σημείο που θα πρέπει να τονιστεί είναι ότι, για να εξαχθούν χρήσιμα συμπεράσματα για την κίνηση του αντικειμένου, θα πρέπει το επίπεδο κίνησης να σχηματίζει όσο το δυνατό μικρότερη γωνία με αυτό της εικόνας ώστε κατά την κίνηση, να μην αλλάζει η απόσταση του αντικειμένου από την κάμερα σημαντικά. Η μεταβολή της απόστασης εισάγει το φαινόμενο της κλιμάκωσης στο πρόβλημα με αποτέλεσμα οι περιοχές γύρω από τα χαρακτηριστικά σημεία να μην είναι συγκρίσιμες, αφού ένα pixel της εικόνας με το αντικείμενο στην πιο απομακρυσμένη θέση θα αντιστοιχούσε σε μια ομάδα pixel στην εικόνα με το αντικείμενο σε μια πλησιέστερη.

Ο περιορισμός αυτός θα μπορούσε να αρθεί αν σε κάθε χαρακτηριστικό σημείο (feature point) που εξάγεται, υπάρχει κάποια πληροφορία για το βάθος της σκηνής.

4.3 Ταυτοποίηση μοντέλου (model matching)

4.3.1 Εισαγωγή

Το επιθυμητό αποτέλεσμα για τον δεύτερο αλγόριθμο είναι η εκτίμηση των χαρακτηριστικών θέσης του αντικειμένου με την χρήση ενός τρισδιάστατου μοντέλου που το περιγράφει επαρκώς. Σε γενικές γραμμές, αναζητείται εκείνη η τοποθέτηση του μοντέλου στον 3D χώρο, για την οποία η προοπτική προβολή του στο επίπεδο της εικόνας, ταιριάζει καλύτερα στα δεδομένα που έχει συλλέξει η κάμερα.

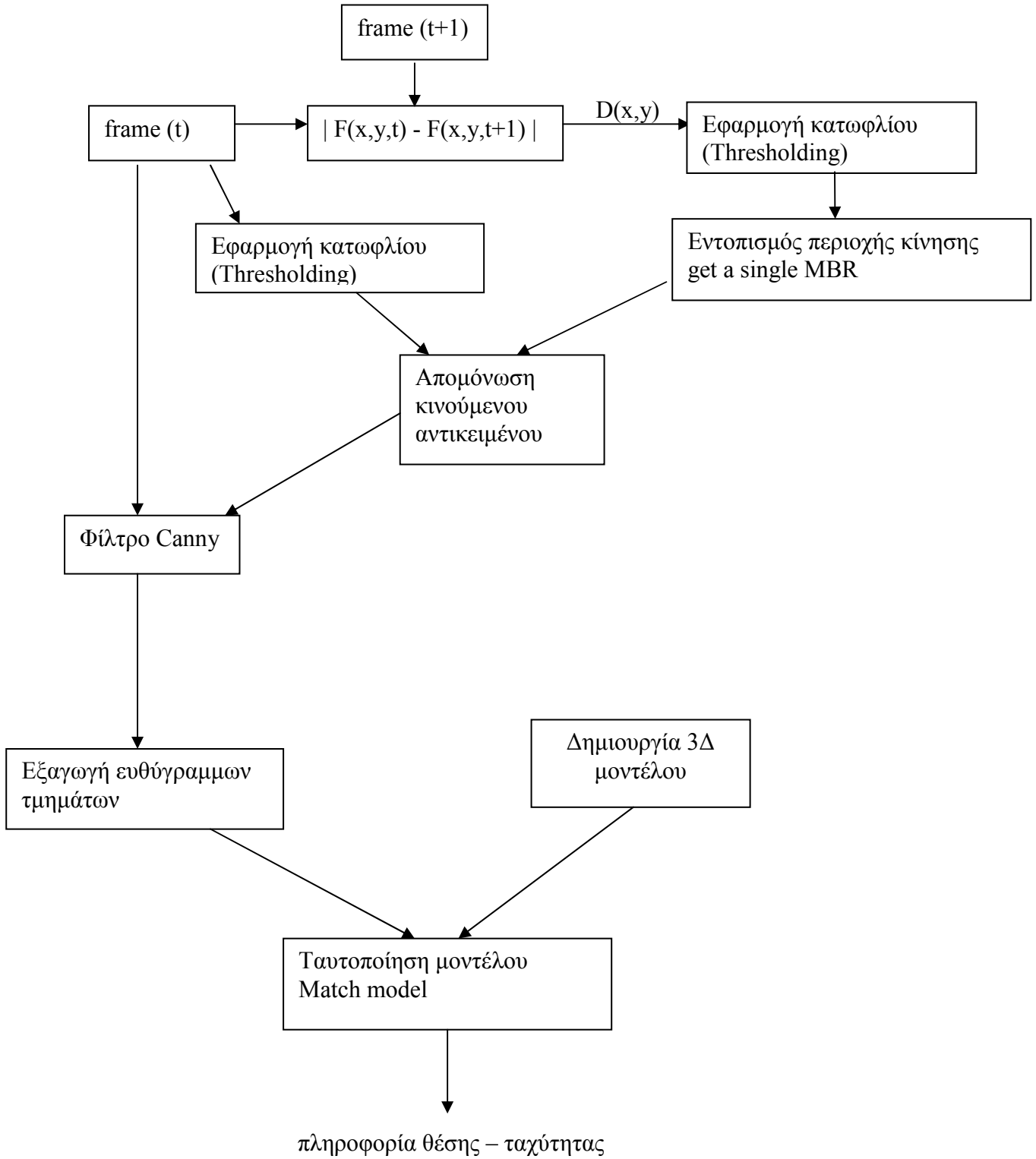
4.3.2 Περιγραφή αλγορίθμου

Στις παραγράφους που ακολουθούν, επιχειρείται η περιγραφή ενός αλγορίθμου ταιριάσματος (matching) χαρακτηριστικών σημείων πάλι μέσα από τα στάδια ανάλυσης που αυτός ακολουθεί.

Αναλυτικότερα :

i) Στάδιο περιφερειακής ανάλυσης (peripheral analysis)

Το στάδιο αυτό δε διαφέρει από το αντίστοιχο του προηγούμενου αλγορίθμου. Έτσι, αρχικά υπολογίζεται η διαφορά ανάμεσα σε δύο διαδοχικά frame. Έπειτα, εφαρμόζεται ένα κατώφλι (threshold) στην εικόνα που προέκυψε από την αφαίρεση, έτσι ώστε να ξεχωρίσει το κινούμενο αντικείμενο. Αφού γίνει και αυτό, η περιοχή που μας ενδιαφέρει περικλείεται από ένα παραλληλόγραμμο (Minimum Bounding Rectangle). Μέσα σε αυτήν την περιοχή, όμως, υπάρχουν δύο απεικονίσεις του αντικειμένου. Για να ξεχωρίσει αυτή που μας ενδιαφέρει, εφαρμόζουμε ένα κατώφλι στην αρχική εικόνα (frame) που περιέχει το αντικείμενο και αναζητούμε το



Σχήμα 4.2 : Δομικό διάγραμμα (block diagram) του αλγόριθμου ταιριάσματος μοντέλου (model matching).

αντικείμενο μέσα στην περιοχή που προέκυψε προηγουμένως. Τελικά το αντικείμενο περικλείεται από ένα παραλληλόγραμμο (Minimum Bounding Rectangle).

ii) *Στάδιο εστίασης της προσοχής (attentive analysis)*

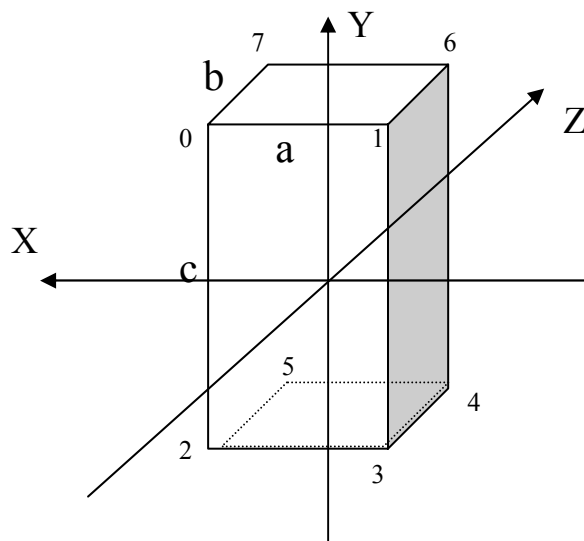
Σε αυτή τη φάση, εξάγονται χαρακτηριστικά από την εικόνα. Με δεδομένο ότι τα χαρακτηριστικά που αναζητούμε είναι ακμές, δοκιμάστηκαν αρκετά φίλτρα ανίχνευσης ακμών. Οι τελεστές Prewitt και Sobel δεν έδωσαν ικανοποιητικά αποτελέσματα, ενώ ο Laplacian επέδειξε την γνωστή ευαισθησία του στον θόρυβο. Η κατάσταση δεν βελτιώθηκε δραματικά ούτε με τον Laplacian of Gaussian (LoG). Παραβλέποντας, λοιπόν, το μειονέκτημα της πολυπλοκότητας, επιλέχτηκε και υλοποιήθηκε το φίλτρο Canny με κάποια μικρή παραλλαγή. Αυτή παρουσιάζεται σε δύο σημεία :

α) πριν από την σύνδεση των ακμών υπάρχει ένα φίλτρο μεγέθους το οποίο αποκλείει τις πολύ μικρές περιοχές που πιθανόν οφείλονται σε θόρυβο ή στην υφή του αντικειμένου. (Αυτό δεν σημαίνει ότι αντιμετωπίζονται επιτυχώς επιφάνειες αντικειμένων με έντονη υφή).

β) η έξοδος του φίλτρου δεν είναι μια δυαδική (binary) εικόνα αλλά κάθε pixel έχει τιμή grayscale ανάλογα με την διεύθυνση της ακμής στην οποία ανήκει. Αυτό το χαρακτηριστικό θα φανεί χρήσιμο αργότερα στην διαδικασία της σήμανσης (labeling) των ακμών.

Τα αποτελέσματα του Canny φίλτρου, εισάγονται στην ρουτίνα σήμανσης (labeling) όπου τους δίνεται αύξων αριθμός, εξάγονται ιδιότητες όπως προσανατολισμός, μέγεθος, θέση και αποθηκεύονται σε λίστα.

Επόμενο βήμα είναι ο ορισμός του 3Δ μοντέλου. Αυτό επιλέχτηκε να είναι ένα ορθογώνιο παραλληλεπίπεδο όπως το παρακάτω :



Σχήμα 4.3 : Μοντέλο 3Δ όπου φαίνονται οι ακμές και οι κορυφές του

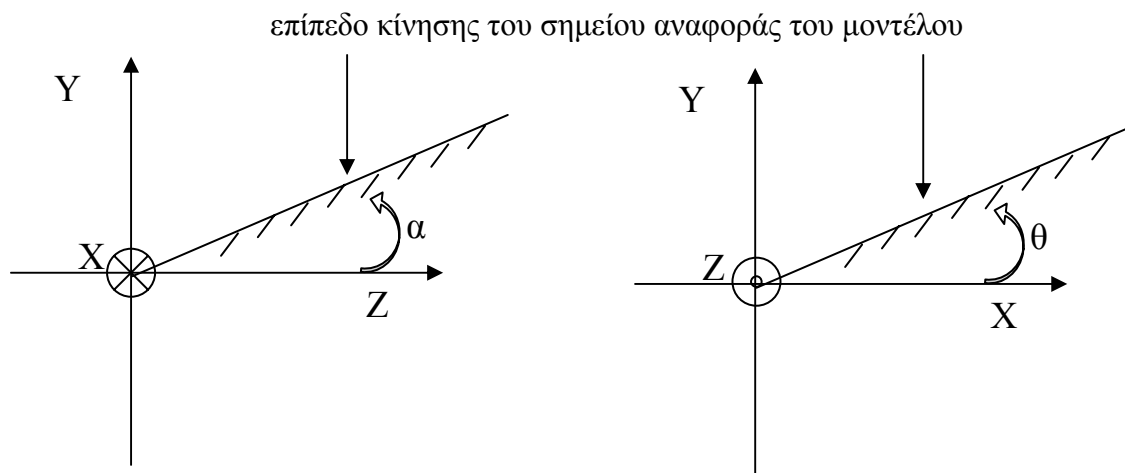
Τα a , b και c είναι προφανώς οι διαστάσεις του αντικειμένου οι οποίες μπορούν να είναι σταθερές ή να δίνονται ως παράμετροι. Το μοντέλο αποθηκεύεται στην μνήμη ως πίνακας που περιέχει τις 3Δ χωρικές συντεταγμένες των οκτώ κορυφών του, ενώ δημιουργείται έτσι ώστε το κέντρο βάρους και σημείο αναφοράς του να είναι το $(0,0,0)$. Έτσι οι συντεταγμένες των κορυφών είναι:

$$\begin{array}{ll}
 p_0(a/2, c/2, -b/2) & p_7(a/2, c/2, b/2) \\
 p_1(-a/2, c/2, -b/2) & p_6(-a/2, c/2, b/2) \\
 p_2(a/2, -c/2, -b/2) & p_5(a/2, -c/2, b/2) \\
 p_3(-a/2, -c/2, -b/2) & p_4(-a/2, -c/2, b/2)
 \end{array}$$

Το μοντέλο που έχει δημιουργηθεί και η λίστα με τις ακμές της εικόνας συνδυάζονται παρακάτω για να επιτευχθεί η βέλτιστη τοποθέτηση του πρώτου ως προς τις δεύτερες. Η διαδικασία, η οποία φαίνεται και στο Σχήμα 4.5, έχει ως εξής :

Αρχικά, εκτιμάται η θέση του κέντρου βάρους του αντικείμενου από τα δεδομένα της εικόνας. Η διαδικασία λαμβάνει υπόψη την θέση του ειδώλου στο επίπεδο της εικόνας και το επίπεδο πάνω στο οποίο κινείται το αντικείμενο. Μια τέτοια εκτίμηση είναι δυνατό να βρίσκεται αρκετά κοντά στην πραγματικότητα, αν οι διαστάσεις του αντικείμενου είναι τέτοιες ώστε να μπορούμε να θεωρήσουμε ότι το κέντρο βάρους δεν απέχει πολύ από το κέντρο της επιφάνειας που σχηματίζεται στην εικόνα. Οι συντεταγμένες του κέντρου βάρους ορίζονται ως αρχικές τιμές για το διάνυσμα $T = [X, Y, Z]$ που καθορίζει την θέση του αντικείμενου σε μια χρονική στιγμή.

Έπειτα, καθορίζεται το επίπεδο πάνω στο οποίο βρίσκεται το αντικείμενο, με την χρήση δύο εκ των γωνιών περιστροφής θ, α, β γύρω από τους Z, X, Y άξονες αντίστοιχα (βλ Σχήμα 4.4).



Σχήμα 4.4 : Γωνίες που καθορίζουν το επίπεδο κίνησης του σημείου αναφοράς του μοντέλου

Οι γωνίες που θα επιλεγούν ως σταθερές σε συνδυασμό με την «ελεύθερη» γωνία περιστροφής πάνω στο επίπεδο κίνησης την οποία και αναζητούμε, αποτελούν το διάνυσμα γωνιακής μετατόπισης $R=[\theta, \alpha, \beta]$. Εδώ θα πρέπει να διευκρινιστεί ότι τα διανύσματα T και R είναι σε πλήρη αντιστοιχία με αυτά που αναφέρθηκαν στο κεφάλαιο 2, μόνο που για λόγους οικονομίας ο αλγόριθμος αποθηκεύει μόνο τις τιμές των μεταβλητών.

Για κάθε τιμή της «ελεύθερης» γωνίας από $-\pi/2$ έως $\pi/2$ προβάλλεται το μοντέλο στο επίπεδο της εικόνας. Οι ακμές που προκύπτουν, συγκρίνονται αρχικά με τις δεδομένες ακμές ως προς τον προσανατολισμό τους. Η τιμή της «ελεύθερης» γωνίας εκείνης για την οποία το άθροισμα των μεγεθών των γραμμών δεδομένων που έχουν τον ίδιο προσανατολισμό ($\pm 5^\circ$) με αυτές του προβαλλόμενου μοντέλου είναι μεγαλύτερο, θεωρείται επικρατέστερη.

Από την διαδικασία αυτή προκύπτουν δύο τέτοιες τιμές οι οποίες διαφέρουν κατά $\pi/2$. Για να ξεχωρίσει η ορθότερη λύση, εισάγεται το κριτήριο κάλυψης του μοντέλου. Αρχικά, σε κάθε ακμή του μοντέλου αντιστοιχίζονται μόνο οι δεδομένες ακμές που απέχουν από αυτές όχι περισσότερο από τρία (3) pixel. Σε κάθε τέτοια αντιστοίχιση αφαιρείται από το μήκος της ακμής του μοντέλου αυτό της δεδομένης ακμής και η απόλυτη τιμή της διαφοράς αθροίζεται σε ένα μετρητή. Η τιμή του μετρητή αποτελεί ουσιαστικά ένδειξη του πόσο καλά καλύπτεται το μοντέλο από τα δεδομένα, με τις μικρότερες τιμές να αντιστοιχούν σε πληρέστερη κάλυψη. Αυτό υπό την προϋπόθεση ότι στα δεδομένα, το 50% τουλάχιστον υπάρχει και δεν έχει αποσπασθεί μακριά από τις πραγματικές γραμμές του αντικειμένου (outliers).

Έτσι, επιλέγεται η τιμή της «ελεύθερης» γωνίας για την οποία το άθροισμα των μεγεθών των γραμμών δεδομένων που έχουν τον ίδιο προσανατολισμό ($\pm 5^\circ$) με αυτές του προβαλλόμενου μοντέλου είναι μεγαλύτερο και ο μετρητής του κριτηρίου κάλυψης ελάχιστος.

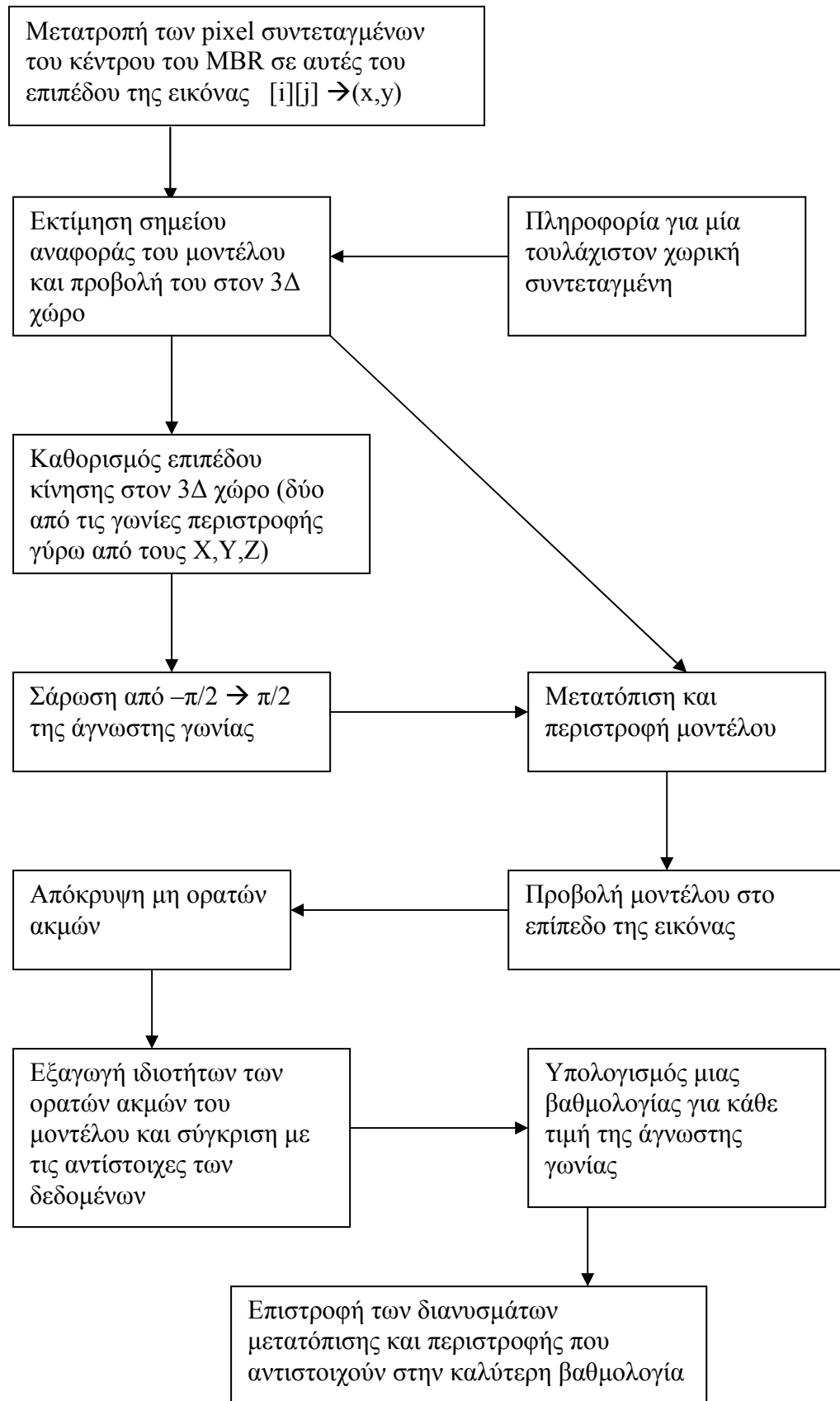
Στο τέλος αυτού του σταδίου είναι διαθέσιμα τα διανύσματα T και R τα οποία και περιέχουν τις τιμές γραμμικής και γωνιακής μετατόπισης του αρχικού μοντέλου γύρω από τους X, Y, Z άξονες.

Γνωστικό επίπεδο ανάλυσης

Όπως και στην περίπτωση του προηγούμενου αλγορίθμου έτσι και εδώ, αυτή η φάση της ανάλυσης εξαρτάται άμεσα από την εξειδίκευση της εφαρμογής.

4.3.3 Αναλυτικότερη περιγραφή του αλγόριθμου

Η αναλυτική περιγραφή του αλγορίθμου θα γίνει συνολικά, μετά την παρουσίαση και των επεκτάσεων που ακολουθούν.

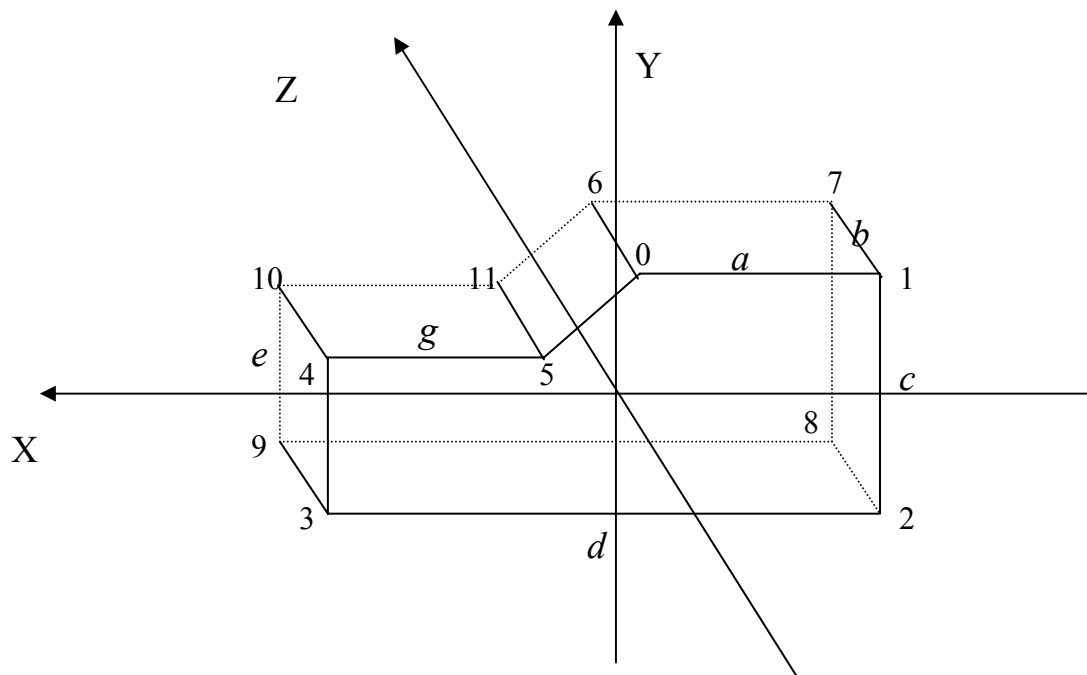


Σχήμα 4.5 : Λεπτομερέστερο δομικό διάγραμμα της διαδικασίας ταιριάσματος δεδομένων γραμμών και γραμμών μοντέλου

4.4 Επεκτάσεις του αλγόριθμου ταυτοποίησης μοντέλου

4.4.1 Αλλαγή μοντέλου

Ο αλγόριθμος εκτίμησης των χαρακτηριστικών κίνησης με βάση κάποιο μοντέλο, μπορεί να επεκταθεί για να προσαρμοστεί σε συνθετότερες συνθήκες. Μια τέτοια περίπτωση θα ήταν η χρήση ενός διαφορετικού μοντέλου που περιγράφει καλύτερα κάποιο αντικείμενο. Για παράδειγμα ας δούμε το 3Δ μοντέλο ενός «αυτοκινήτου» :



Σχήμα 4.6 : Μοντέλο 3Δ υποθετικού αυτοκινήτου

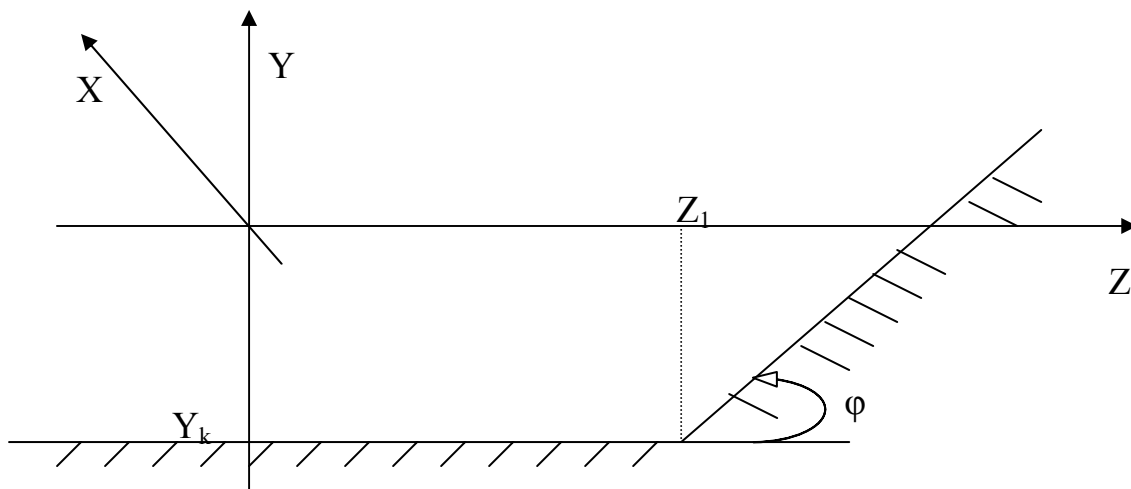
όπου τα a, b, c, d, e, g (το f απουσιάζει επίτηδες για αποφυγή σύγχυσης με το focal length) είναι τα μεγέθη που καθορίζουν το μοντέλο. Το τελευταίο δημιουργείται έτσι ώστε το σημείο αναφοράς του, που δεν είναι πια το κέντρο βάρους του, να ταυτίζεται με την αρχή των αξόνων. Έτσι τα σημεία 0...11 που περιγράφουν το μοντέλο έχουν συντεταγμένες :

$$\begin{array}{ll}
 p_0(a - d/2, c/2, -b/2) & p_6(a - d/2, c/2, b/2) \\
 p_1(-d/2, c/2, -b/2) & p_7(-d/2, c/2, b/2) \\
 p_2(-d/2, -c/2, -b/2) & p_8(-d/2, -c/2, b/2) \\
 p_3(d/2, -c/2, -b/2) & p_9(d/2, -c/2, b/2) \\
 p_4(d/2, e - c/2, -b/2) & p_{10}(d/2, e - c/2, b/2) \\
 p_5(d/2 - g, e - c/2, -b/2) & p_{11}(d/2 - g, e - c/2, b/2)
 \end{array}$$

Τα υπόλοιπα μέρη του αλγορίθμου παραμένουν τα ίδια, εφόσον έχουν ενημερωθεί οι ρουτίνες για την αλλαγή στον αριθμό των σημείων και των γραμμών του μοντέλου.

4.4.2 Εναλλαγή επιπέδων κίνησης

Μια ενδιαφέρουσα προσθήκη στον αλγόριθμο είναι η δυνατότητα του μοντέλου να μην κινείται πάντα πάνω σε ένα μόνο γεωμετρικό χώρο, αλλά να αλλάζει επίπεδο κίνησης από ένα σημείο και έπειτα. Για παράδειγμα, έστω η παρακάτω διαμόρφωση, όπου το σημείο αναφοράς του μοντέλου κινείται αρχικά στο οριζόντιο και εν συνεχεία στο κεκλιμένο επίπεδο, ενώ η κάμερα είναι τοποθετημένη έτσι ώστε το σύστημα συντεταγμένων της να συμπίπτει με αυτό του 3Δ χώρου.



Σχήμα 4.7 : Διάταξη για την περιγραφή της αλλαγής επιπέδου κίνησης.

Στο παραπάνω σχήμα το $Y=Y_k$ είναι το ένα επίπεδο κίνησης του σημείου αναφοράς του μοντέλου, ενώ το $Z=Z_l$ είναι το επίπεδο εκατέρωθεν του οποίου αλλάζει το επίπεδο κίνησης. Η γωνία φ είναι αυτή που σχηματίζεται ανάμεσα στα δύο επίπεδα. Έχοντας δεδομένες αυτές τις τρεις παραμέτρους, ας δούμε πως προκύπτουν οι 3Δ συντεταγμένες των σημείων που ανήκουν στα δύο επίπεδα. Ο υπολογισμός αυτός είναι απαραίτητος, αφού δίνει μια καλή εκτίμηση της θέσης του αντικειμένου.

Όσον αφορά το οριζόντιο επίπεδο, ισχύουν αυτά που αναφέρονται στην ενότητα 2.2.3 και στο Παράρτημα για $B = 1/Y_k$, $\alpha = A = C = \theta$. Στο Σχήμα 4.7

φαίνεται η ευθεία που αποτελεί την τομή του κεκλιμένου επιπέδου με το YZ επίπεδο. Η τομή αυτή θα δώσει μια σχέση ανάμεσα στα Y και Z η οποία, λόγω της περιστροφής του κεκλιμένου μόνο γύρω από τον X άξονα, θα ισχύει για κάθε X . Είναι προφανές ότι η ευθεία αυτή έχει κλίση ϕ και διέρχεται από το σημείο (Y_k, Z_l) του YZ επιπέδου. Οπότε :

$$\left. \begin{array}{l} Y = Z \tan \phi + b \\ Y_k = Z_l \tan \phi + b \end{array} \right\} \Rightarrow Y = Z \tan \phi + Y_k - Z_l \tan \phi$$

Αντικαθιστώντας το Y από με το ίσο του από την σχέση της προοπτικής προβολής :

$$\frac{yZ}{\lambda} = Z \tan \phi + Y_k - Z_l \tan \phi \Rightarrow yZ = \lambda Z \tan \phi + \lambda Y_k - \lambda Z_l \tan \phi \Rightarrow Z(y - \lambda \tan \phi) = \lambda(Y_k - Z_l \tan \phi)$$

και τελικά

$$Z = \lambda \frac{Y_k - Z_l \tan \phi}{y - \lambda \tan \phi} \quad (\text{εξ. 4.8})$$

Έχοντας υπολογίσει το Z , οι άλλες δύο συντεταγμένες προκύπτουν από τους τύπους της προοπτικής προβολής :

$$x = \lambda \frac{X}{Z} \Rightarrow X = \frac{xZ}{\lambda} \quad (\text{εξ. 4.9})$$

$$y = \lambda \frac{Y}{Z} \Rightarrow Y = \frac{yZ}{\lambda} \quad (\text{εξ. 4.10})$$

Πέρα από τους επιπλέον υπολογισμούς θα χρειαστεί και μια μικρή ρύθμιση του αλγορίθμου. Η αλλαγή στον αλγόριθμο έγκειται στο σημείο της εκτίμησης του σημείου αναφοράς του μοντέλου. Αρχικά, γίνεται μια εκτίμηση του σημείου αναφοράς το μοντέλου με βάση τις εξισώσεις του κεκλιμένου επιπέδου. Αν από τους υπολογισμούς προκύψει τιμή για το Z μικρότερη από Z_l , τότε απορρίπτεται και εκτιμάται νέα θέση χρησιμοποιώντας αυτή τη φορά τις εξισώσεις για το οριζόντιο επίπεδο. Από το στάδιο αυτό και πέρα, ο αλγόριθμος παραμένει όπως και προηγουμένως. Δηλαδή, θεωρείται πάλι ένα πλήθος τιμών για την «ελεύθερη» γωνία μέχρι να βρεθεί η τιμή εκείνη για την οποία τα δεδομένα ταιριάζουν καλύτερα με το μοντέλο.

4.4.3 Χρήση μοντέλου μη γραμμικών ακμών

Μέχρι τώρα το μοντέλο που χρησιμοποιήθηκε είχε για ακμές, τμήματα καμπύλων πρώτου βαθμού (ευθύγραμμα τμήματα). Στην προσπάθεια να περιγραφούν ακόμη πολυπλοκότερα αντικείμενα, ένα πιθανό βήμα είναι η αντικατάσταση των ευθύγραμμων τμημάτων από τμήματα καμπύλων ανώτερου βαθμού και συγκεκριμένα τρίτου. Κάθε τμήμα της κυβικής spline (cubic spline), είναι μια παραμετρική καμπύλη :

$$\begin{aligned} X(t) &= a_x t^3 + b_x t^2 + c_x t + d_x \\ Y(t) &= a_y t^3 + b_y t^2 + c_y t + d_y \\ Z(t) &= a_z t^3 + b_z t^2 + c_z t + d_z \end{aligned} \quad \text{για } 0 \leq t \leq 1$$

όπου X, Y, Z οι συντεταγμένες στον 3Δ χώρο των σημείων που ανήκουν στην καμπύλη και t η παράμετρος.

Οι παραπάνω εξισώσεις μπορούν να γραφούν σε μορφή πινάκων :

$$\mathbf{p}(t) = \mathbf{a}t^3 + \mathbf{b}t^2 + \mathbf{c}t + \mathbf{d} \quad (\text{εξ. 4.11})$$

όπου

$$\begin{aligned} \mathbf{p}(t) &= (x(t), y(t), z(t))^T \\ \mathbf{a} &= [a_x, a_y, a_z]^T \\ \mathbf{b} &= [b_x, b_y, b_z]^T \\ \mathbf{c} &= [c_x, c_y, c_z]^T \\ \mathbf{d} &= [d_x, d_y, d_z]^T \end{aligned}$$

Η αναπαράσταση αυτή είναι δυνατό να καταλήξει στην παρακάτω μορφή :

$$\mathbf{p}(t) = [\mathbf{a} \quad \mathbf{b} \quad \mathbf{c} \quad \mathbf{d}] \begin{bmatrix} t^3 \\ t^2 \\ t \\ 1 \end{bmatrix} \quad (\text{εξ. 4.12}).$$

Με την εισαγωγή των κυβικών spline στο μοντέλο, οι έδρες του μετατρέπονται σε μη επίπεδες επιφάνειες. Στην καλύτερη περίπτωση μπορεί να διατηρηθεί επίπεδο το πολύ ένα ζεύγος παράλληλων εδρών. Για να περιγραφούν οι νέου τύπου έδρες, χρησιμοποιούνται κυβικές spline επιφάνειες (cubic spline surfaces).

Η κυβική spline επιφάνεια θα προκύψει υπό μορφή τανυστικού γινομένου από τον συνδυασμό δύο κυβικών spline καμπύλων, μιας για κάθε παραμετρική συντεταγμένη. Έχοντας υπόψη την προηγούμενη αναπαράσταση της κυβικής spline καμπύλης προκύπτει για την επιφάνεια :

$$\mathbf{p}(u, v) = \begin{bmatrix} u^3 & u^2 & u & 1 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix} \begin{bmatrix} v^3 \\ v^2 \\ v \\ 1 \end{bmatrix}$$

όπου κάθε \mathbf{a}_i είναι ένα διάνυσμα 1×3 , κάθε \mathbf{b}_i είναι ένα διάνυσμα 3×1 , ενώ τα u και v είναι οι παράμετροι.

Η παραμετρική επιφάνεια μπορεί να γραφεί τελικά ως :

$$\mathbf{p}(t, s) = U^T M V \quad (\text{εξ. 4.13})$$

όπου

$$U = \begin{bmatrix} u^3 \\ u^2 \\ u \\ 1 \end{bmatrix} \quad \text{και} \quad V = \begin{bmatrix} v^3 \\ v^2 \\ v \\ 1 \end{bmatrix}$$

και

$$M = \begin{bmatrix} \mathbf{a}_1 \mathbf{b}_1 & \mathbf{a}_1 \mathbf{b}_2 & \mathbf{a}_1 \mathbf{b}_3 & \mathbf{a}_1 \mathbf{b}_4 \\ \mathbf{a}_2 \mathbf{b}_1 & \mathbf{a}_2 \mathbf{b}_2 & \mathbf{a}_2 \mathbf{b}_3 & \mathbf{a}_2 \mathbf{b}_4 \\ \mathbf{a}_3 \mathbf{b}_1 & \mathbf{a}_3 \mathbf{b}_2 & \mathbf{a}_3 \mathbf{b}_3 & \mathbf{a}_3 \mathbf{b}_4 \\ \mathbf{a}_4 \mathbf{b}_1 & \mathbf{a}_4 \mathbf{b}_2 & \mathbf{a}_4 \mathbf{b}_3 & \mathbf{a}_4 \mathbf{b}_4 \end{bmatrix} .$$

Έχοντας τα παραπάνω υπόψη, δημιουργούνται κυβικές spline επιφάνειες από τις κυβικές spline καμπύλες που τις ορίζουν. Για να μην εμφανίζει το μοντέλο «κενά» θα πρέπει να προσεχτεί η συνέχεια των καμπύλων επιφανειών στα σημεία εκείνα που βρίσκονται στα «σύνορα» μεταξύ των εδρών.

Άλλο ένα σημείο του αλγορίθμου που επηρεάζεται από την εισαγωγή των κυβικών spline είναι η διαδικασία ταιριάσματος της 2Δ προβολής του μοντέλου με τα δεδομένα που συλλέγει η κάμερα. Πέραν του ότι για την προβολή θα πρέπει να χρησιμοποιηθεί κάποιος συνθετότερος αλγόριθμος παρακολούθησης ακτίνας (ray-tracing) που να λαμβάνει υπόψη του και μοντέλα φωτισμού της σκηνής, θα ήταν ενδεχομένως επιβεβλημένο να εξεταστούν και άλλα μέτρα επιτυχίας της προσαρμογής των pixel στις καμπύλες του μοντέλου (curve fitting). Τέτοια μέτρα θα μπορούσαν να είναι το μέσο τετραγωνικό σφάλμα (Mean Squared Error) :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n d_i^2 \quad \text{όπου } d_i \text{ η απόσταση του σημείου από την καμπύλη,}$$

ή το μέγιστο κανονικοποιημένο σφάλμα (Normalized maximum error) :

$$\varepsilon = \frac{\max_i |d_i|}{S} \quad \text{όπου } S \text{ το μήκος της καμπύλης.}$$

4.4.4 Αναλυτικότερη περιγραφή του αλγορίθμου ταυτοποίησης μοντέλου και των επεκτάσεων αυτού

Οι παρακάτω συναρτήσεις αποτελούν τα δομικά στοιχεία του αλγορίθμου ταιριάσματος μοντέλου (model matching).

Ανίχνευση και σήμανση (labeling) ευθύγραμμων τμημάτων (line segments).

Ξεκινώντας από τη συνάρτηση που επιτελεί την λειτουργία της εξαγωγής ευθύγραμμων τμημάτων από την εικόνα :

```
void canny(unsigned char **Im1,unsigned char **target,int *area)
```

η οποία εφαρμόζει το φίλτρο canny σε μια περιοχή (area) της εικόνας εισόδου (Im1) και τοποθετεί το αποτέλεσμα στην εικόνα εξόδου (target). Η εικόνα εξόδου έχει σε κάθε pixel τιμή grayscale ανάλογα με την διεύθυνση της ακμής στην οποία ανήκει :

Τιμή (grayscale)	Διεύθυνση ακμής
60	
111	\
192	-
253	/

προχωράμε στην συνάρτηση εξαγωγής χαρακτηριστικών :

```
void get_line_segments(unsigned char **Im1,struct line_segment **line_list,
int *area,int min_size)
```

Η συνάρτηση δέχεται ως είσοδο την εικόνα (Im) που προέκυψε από την εφαρμογή του προηγούμενου canny φίλτρου και αποδίδει μια λίστα (line_list) με τα

χαρακτηριστικά κάθε ευθύγραμμου τμήματος που βρίσκεται στην περιοχή (area) της εικόνας και έχει μέγεθος μεγαλύτερο από (min_size) pixel. Κάθε κόμβος της λίστας περιέχει πληροφορία για την ετικέτα του τμήματος, τα σημεία που το ορίζουν (end points) σε (i,j) και (x,y) συντεταγμένες, το μέγεθός του σε pixel, την κατεύθυνσή του σε ακτίνια, την τιμή που του έδωσε το canny φίλτρο και το μήκος του.

Ορισμός μοντέλου

Πέρα από τις σταθερές που εξαρτώνται από την κάμερα και αναφέρθηκαν και στην περιγραφή των προηγούμενων αλγορίθμων, θα πρέπει να καθοριστούν οι σταθερές που αφορούν στο πλήθος των σημείων και των γραμμών του μοντέλου :

num_of_points – αριθμός σημείων
num_of_lines – αριθμός ακμών

Για να οριστεί πλήρως το μοντέλο θα πρέπει να οριστούν οι (X, Y, Z) συντεταγμένες των σημείων που το αποτελούν, καθώς και ο τρόπος με τον οποίο συνδέονται μεταξύ τους. Η διαδικασία αυτή γίνεται με την συνάρτηση :

```
void create_3dmodel_car_at_Zo(double model[num_of_points][3],
                             int connect_m[][2], double a, double b, double c, double d,
                             double e, double g)
```

Η συνάρτηση αυτή έχει ως είσοδο τις παραμέτρους του μοντέλου (a,b,c,d,e,g) (το f έχει παραληφθεί επίτηδες για να μην υπάρξει σύγχυση με το focal length), και επιστρέφει ένα πίνακα (model) με τις συντεταγμένες των σημείων στον 3D χώρο και έναν άλλο πίνακα (connect) κάθε γραμμή του οποίου περιέχει τα δύο σημεία που ορίζουν μια ακμή του μοντέλου. Η συγκεκριμένη επιστρέφει το μοντέλο του «αυτοκινήτου». Εύκολα κάποιο άλλο μοντέλο θα ήταν δυνατό να δημιουργηθεί, αλλάζοντας τις τιμές των στοιχείων των πινάκων (model) και (connect) μέσα στο σώμα της συνάρτησης.

Πριν παρουσιαστεί η διαδικασία του ταιριάσματος (matching), θα ήταν σωστό να εξεταστούν κάποιες βασικές λειτουργίες που χρησιμοποιούνται από αυτή. Μία τέτοια λειτουργία είναι η περιστροφή και μετατόπιση του μοντέλου.

```
void translate_n_rotate_3D_model(double model3d_A[num_of_points][3],
                                 double model3d_B[][3], double *translate, double *rotate)
```

Η συνάρτηση δέχεται ένα μοντέλο (model3d_A), τα διανύσματα μετατόπισης (translate) και περιστροφής (rotate) και επιστρέφει τη νέα θέση των σημείων του μοντέλου (model3d_B). Το σώμα της συνάρτησης αποτελείται από τέσσερις for βρόχους (loops). Εναλλάσσοντας την σειρά εκτέλεσης των βρόχων, αλλάζει και η σειρά με την οποία το μοντέλο περιστρέφεται γύρω από το σημείο αναφοράς του και τους άξονες X, Y, Z .

Άλλη μια συνάρτηση που χρησιμοποιείται στην διαδικασία ταιριάσματος είναι η :

```
void project_3Dmodel_to_image_plane(double model3d[num_of_points][3],
                                    double model2d[num_of_points][2], double f),
```

η οποία, με δεδομένη την εστιακή απόσταση (f), προβάλλει τα σημεία του 3D μοντέλου (model3d) στο επίπεδο της εικόνας παράγοντας ένα 2D μοντέλο (model2d).

Την αντίστροφη λειτουργία έχει η :

```
void xy_to_XYZ(double x,double y,double f,double *X,double *Y,double Z),
```

η οποία αντιστοιχίζει ένα σημείο του επιπέδου της εικόνας (x,y), στον 3Δ χώρο (X,Y,Z), γνωρίζοντας την Z συντεταγμένη του. Ίδιου τύπου διαδικασίες, που αφορούν μετατροπή και αντιστοίχιση σημείων, διεκπεραιώνει η παρακάτω συνάρτηση :

```
void xy_to_ij(double x,double y,int *i,int *j,double pixel_per_meter_h,
             double pixel_per_meter_v)
```

μετατρέπει τις συνεχείς (x,y) συντεταγμένες του επιπέδου της εικόνας στις διακριτές [i, j] του πίνακα των pixel, συνυπολογίζοντας και τους συντελεστές (pixel_per_meter_h, pixel_per_meter_v) που εξαρτώνται από την κάμερα.

και η αντίστροφή της :

```
void ij_to_xy(int i,int j,double *x,double *y,double pixel_per_meter_h,
             double pixel_per_meter_v).
```

Χρήσιμη, επίσης, είναι η συνάρτηση που ελέγχει για το ποια σημεία του 3Δ μοντέλου είναι ορατά από την θέση που παρατηρούμε το μοντέλο.

```
void check_for_visible_points(double point_model3d[num_of_points][3],int *visible)
```

Έχοντας για είσοδο τα σημεία που αποτελούν το τρισδιάστατο μοντέλο, επιστρέφει πίνακα (visible) από σημεία. Αν το ν-οστό στοιχείο του πίνακα (visible) έχει τιμή ένα (1) τότε το ν-οστό σημείο του μοντέλου είναι ορατό, αν έχει τιμή μηδέν (0) τότε δεν είναι. Ο αλγόριθμος απόκρυψης των μη ορατών σημείων, ταξινομεί όλα τα σημεία ως προς την Z συντεταγμένη και αποκρύπτει εκείνο ή εκείνα που απέχουν περισσότερο από την κάμερα (μεγαλύτερο Z). Στις περισσότερες περιπτώσεις αυτό είναι αρκετό, ενώ ανακρίβειες παρατηρούνται μόνο για μικρές γωνίες περιστροφής του μοντέλου ($5^{\circ} \sim 10^{\circ}$). Επιλέχθηκε η ταξινόμηση μόνο κατά Z για λόγους απλότητας και ταχύτητας.

Έχοντας δει τις βασικές συναρτήσεις που χρησιμοποιούνται στη διαδικασία ταιριάσματος, ακολουθεί παρακάτω η κύρια συνάρτηση :

```
void match_model(double point_model3d[num_of_points][3],int connect[][2],
                struct line_segment *line_list,int area[4],int *match_table,double *R_m,
                double *T_m,double f)
```

Είσοδοι αυτής της συνάρτησης είναι το 3Δ μοντέλο (point_model3d), ο πίνακας σύνδεσης των σημείων (connect), η περιοχή της εικόνας που έχει εντοπιστεί κίνηση (area), η λίστα των ευθύγραμμων τμημάτων (line_list), η εστιακή απόσταση (f) και τέλος τα διανύσματα R_m, T_m των οποίων ο διττός ρόλος θα αναλυθεί αργότερα παρακάτω. Έξοδοι της είναι τα διανύσματα R_m, T_m τα οποία, μετά την εκτέλεση της ρουτίνας, περιέχουν τα στοιχεία των διανυσμάτων περιστροφής και μετατόπισης αντίστοιχα.

Όπως αναφέρθηκε και παραπάνω, τα διανύσματα R_m, T_m, έχουν διπλό ρόλο. Ο ένας, προφανώς, είναι να λειτουργούν ως μεταβλητές εξόδου. Η άλλη χρήση

τους είναι στην κλήση της ρουτίνας, όπου μέσω των στοιχείων $T_m[1]$ και $T_m[2]$ εισάγονται οι τιμές για τα μεγέθη Y_k και Z_1 αντίστοιχα, ενώ μέσω του $R[1]$ η γωνία ϕ (βλ Σχήμα 4.5). Το $R[0]$ εισάγονται με τιμή μηδέν (0) ενώ το $R[2]$ είναι η «ελεύθερη» γωνία της οποίας την τιμή αναζητά ο αλγόριθμος.

Αυτά όσον αφορά την εναλλαγή επιπέδων κίνησης. Στην περίπτωση όπου υπάρχει μόνο ένα επίπεδο κίνησης, δύο από τις γωνίες $R[0]$, $R[1]$, $R[2]$ καθορίζονται από το επίπεδο της κίνησης, ενώ αυτή που απομένει είναι η «ελεύθερη» γωνία περιστροφής του μοντέλου η οποία αποτελεί και την μεταβλητή του κυρίως for βρόχου της ρουτίνας `math_model`.

Για να οπτικοποιηθούν τα αποτελέσματα του ταιριάσματος (matching), χρησιμοποιούνται οι παρακάτω συναρτήσεις :

```
void print_line(int pointA[2],int pointB[2],unsigned char **Im1,int gray)
```

Εκτυπώνει σε μια εικόνα (`Im1`) το ευθύγραμμο τμήμα με άκρα τα (`pointA`) και (`pointB`) δίνοντας στα pixel που το αποτελούν τιμή γκριζου (`gray`).

```
void print_2dmodel(double point_model2d[num_of_points][2],
int connect[][2],unsigned char **Im1,int *visible)
```

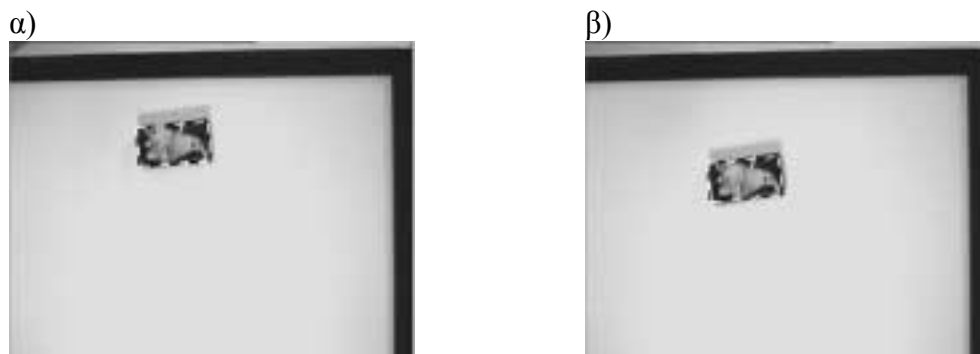
Στη συνάρτηση εισάγεται το δισδιάστατο μοντέλο (`point_model2d`), ο πίνακας σύνδεσης των σημείων (`connect`), το διάνυσμα ορατών σημείων (`visible`) και εξάγεται η εικόνα (`Im1`) που περιέχει το είδωλο του μοντέλου, όπως αυτό θα αποτυπωνόταν στην εικόνα αν υπήρχε πραγματικά στον 3Δ χώρο.

Με αυτές τις συναρτήσεις ολοκληρώνεται η περιγραφή των ρουτινών που χρησιμοποιήθηκαν στην υλοποίηση του αλγορίθμου ταιριάσματος μοντέλου (`model matching algorithm`).

4.4.5 Αποτελέσματα και συζήτηση

Αλγόριθμος ταιριάσματος μοντέλου (model matching)

Για την ευθεία αντιπαραβολή των δύο αλγορίθμων δυναμικής όρασης, η ίδια ακολουθία εικόνων χρησιμοποιήθηκε για την δοκιμή και του 2^{ου} αλγορίθμου με τα παρακάτω αποτελέσματα.



Εικόνα 4.3 : Αποτελέσματα αλγορίθμου ταυτοποίησης μοντέλου

γ)



δ)



Εικόνα 4.3 (συνέχεια): Αποτελέσματα αλγόριθμου ταυτοποίησης μοντέλου

Όπως φαίνεται και στις εικόνες, η τοποθέτηση του μοντέλου (διακεκομμένες γραμμές) είναι σχετικά καλή με κάποιες μικρές αποκλίσεις στον προσανατολισμό. Αυτές είναι πιθανό να οφείλονται σε ακμές που έχουν βρεθεί εκτός θέσης ή ανήκουν στο εσωτερικό του αντικειμένου (π.χ. ακμές από έντονη υφή). Ακόμα όμως και οι ακμές που είναι πραγματικά παράλληλες μεταξύ τους, στην εικόνα εμφανίζονται να έχουν μια μικρή διαφορά στις τιμές της κλίσης τους της τάξης των $2^\circ \sim 3^\circ$. Το γεγονός αυτό μειώνει την δυνατότητα του αλγορίθμου να βρει με απόλυτη ακρίβεια την γωνία περιστροφής του αντικειμένου. Παρόλα αυτά, το σφάλμα φαίνεται να διατηρείται κάτω από τις 5° που αποτελούν και το βήμα διακριτοποίησης της «ελεύθερης» γωνίας.

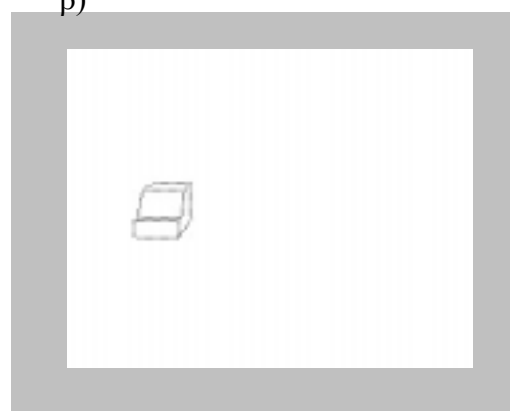
Αποτελέσματα χρήσης διαφορετικού μοντέλου και αλλαγής επιπέδου κίνησης

Οι παρακάτω εικόνες παρουσιάζουν τις θέσεις στις οποίες θα αναζητηθούν οι γραμμές του αντικειμένου, αν υποθεθεί ότι πρόκειται για ένα αυτοκίνητο το οποίο πλησιάζει την κάμερα κινούμενο πρώτα σε κεκλιμένο επίπεδο και καταλήγοντας σε οριζόντιο επίπεδο, όπως φαίνεται στο Σχήμα 4.5. Η γωνία μεταξύ των δύο επιπέδων είναι 10° .

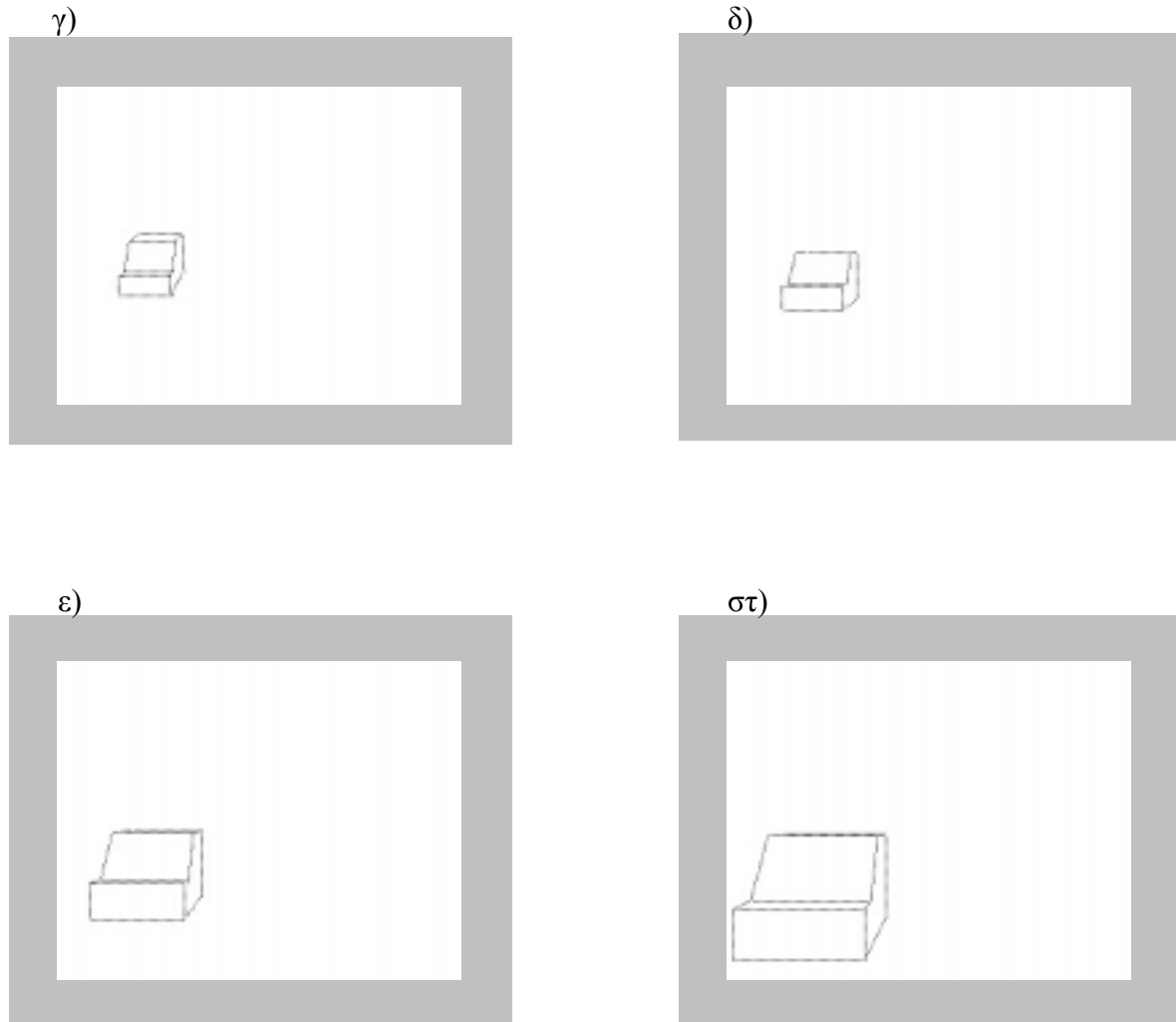
α)



β)



Εικόνα 4.4 : Θέσεις αναζήτησης δεδομένων γραμμών για την περίπτωση του μοντέλου του αυτοκινήτου που αλλάζει επίπεδο κίνησης.



Εικόνα 4.4 (συνέχεια): Θέσεις αναζήτησης δεδομένων γραμμών για την περίπτωση του μοντέλου του αυτοκινήτου που αλλάζει επίπεδο κίνησης.

Στην παραπάνω ακολουθία εικόνων φαίνεται τόσο η αλλαγή επιπέδου κίνησης από το τέταρτο frame και μετά, όσο και η επίδραση της απόστασης στο μέγεθος της προβολής του μοντέλου.

Μελλοντικές εφαρμογές και επεκτάσεις

Στις επεκτάσεις και βελτιώσεις τις οποίες μπορούν να δεχθούν οι αλγόριθμοι που υλοποιήθηκαν, περιλαμβάνονται :

- α) Ένας καλύτερος αλγόριθμος τμηματοποίησης, ο οποίος είτε θα προσαρμόζεται αυτόματα στις εκάστοτε συνθήκες, είτε θα είναι εξ αρχής σχεδιασμένος για τις απαιτήσεις της συγκεκριμένης εφαρμογής.
- β) Βελτιωμένοι αλγόριθμοι προβολής και συγκεκριμένα απόκρυψης των μη ορατών σημείων του κάθε μοντέλου, με ενδεχόμενη χρήση μεθόδων γραφικής (π.χ. Z-buffer).
- γ) Υλοποίηση μοντέλων μη γραμμικών ακμών (π.χ. κυβικές B-splines)
- δ) Δημιουργία βάσης δεδομένων από διάφορα μοντέλα για πληρέστερη περιγραφή και κατανόηση του χώρου.
- ε) Γενίκευση με χρήση όλων των παραπάνω σε συνδυασμό με μοντέλα φωτισμού της 3Δ σκηνής.

Κάθε γενίκευση είναι αυτονόητα ευπρόσδεκτη από ακαδημαϊκή άποψη, αλλά ενδεχομένως εισάγει εξαιρετικά μεγαλύτερη πολυπλοκότητα από αυτή που είναι διατεθειμένος να αποδεχθεί ο σχεδιαστής ενός συστήματος μηχανικής όρασης.

Το πλαίσιο των εφαρμογών αυτών των μεθόδων περικλείει :

- α) Την παρατήρηση και τον έλεγχο της κυκλοφορίας σε οδικά δίκτυα.
- β) Την παροχή πληροφορίας σε αλγόριθμους πλοήγησης αυτόνομων οχημάτων.
- γ) Την επιτήρηση εσωτερικών ή εξωτερικών χώρων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Rafael C. Gonzalez, Richard E. Woods. *Digital Image Processing*. Addison – Wesley, 1993
- [2] Ramesh Jain, Rangachar Kasturi, Brian G. Schunck. *Machine Vision*. McGraw Hill, 1995
- [3] Foresti, Murino, Regazzoni. *Vehicle Recognition and Tracking from Road Image Sequences*. IEEE Transactions on Vehicular Technology, Vol. 48, No.1, January 1999
- [4] C. Fennema, A. Hanson, E. Riseman, Beveridge, R. Kumar. *Model – Directed Mobile robot Navigation*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 20, NO.6 November/December 1990.