



*Università di Pisa*

TESI:

SVILUPPO DI UN ALGORITMO DI  
POSE ESTIMATION ROBUSTO  
PER APPLICAZIONI AERONAUTICHE

Relatori:

*Innocenti Mario*

*Andrea Caiti*

*Pollini Lorenzo*

*Laureando:*

*Raul Pra Levis*

# 1. INTRODUZIONE

Nell'ambito della *computer vision*, sistemi di visione artificiale, una delle questioni fondamentali consiste nella possibilità di stimare istantaneamente la posizione e l'orientamento di un oggetto tridimensionale in movimento osservato attraverso un dispositivo video. Il funzionamento di questo sistema si basa sull'utilizzo di una telecamera che viene posta in modo da individuare e puntare un oggetto, così detto target, ovvero l'oggetto da inseguire.

Nel contempo si estraggono dal filmato i singoli frame che vengono analizzati in tempo reale, per ricavare la posizione relativa dell'obbiettivo; il dato così ottenuto consente di configurare l'oggetto osservato non soltanto attraverso le sue coordinate spaziali, ma anche, attraverso il suo orientamento, di prevederne i futuri spostamenti nel frame successivo, in modo tale da poter adeguare la dinamica del osservatore a quella dell'obbiettivo.

La precisione della stima dipende sostanzialmente da due fattori:

- dal maggior grado di conoscenza dell'oggetto da rilevare;
- dalla quantità e qualità dei sensori a disposizione;

Sul primo fattore si rileva che la conoscenza dell'oggetto raramente potrà essere totalmente esaustiva; si dovrà pertanto ricorrere a modelli matematici più o meno complessi, che sintetizzino al meglio la conformazione del target; tali approssimazioni sono ovviamente dipendenti dalla fattispecie di applicazione.

Riguardo invece al secondo fattore, per sensori sono da intendersi tutte le fonti di informazione utili ad ottenere informazioni spaziali sul target: una o più telecamere, rilevatori di distanza laser, ad infrarossi o ad ultrasuoni.

Questo perché una delle nozioni fondamentali per stimare l'orientamento dell'obbiettivo è individuare la posizione nello spazio dei punti rilevati: maggiori sono le informazioni ottenibili, maggiore sarà la possibilità di ricostruire tali punti con accuratezza.

Ovviamente non possiamo esimerci dal considerare che nelle situazioni contingenti si debba fare i conti con la scarsità dei sensori disponibili rispetto alla situazione ottimale.

Nel caso più propriamente detto di visione artificiale, il numero di telecamere è fondamentale per la semplificazione del problema. Le tecniche di computer vision si basano sull'utilizzo di un numero esiguo di telecamere: la grande differenza sta nell'utilizzarne una o più di una.

Questo perché, similmente alla vista umana, due telecamere, poste in una configurazione nota, che inquadrino lo stesso oggetto, permettono di stabilire agevolmente la profondità di quest'ultimo.

Nel caso di una sola telecamera, che sarà la fattispecie trattata in questa tesi, ciò costituirà un problema importante.

Oltre a ciò, si consideri che nell'oggetto di questa analisi, si farà riferimento, quali informazioni conosciute dell'oggetto, soltanto ad alcuni punti prescelti.

Nonostante le difficoltà di ricerca e sviluppo dei sistemi di visione artificiale, questi acquisiscono una crescente importanza nell'ambito della ricerca informatica e delle sue infinite applicazioni.

## **1.1 L'AMBITO DELLE APPLICAZIONI PRATICHE**

Tra le possibilità di applicazione di questo tipo di analisi si citano a

titolo di esempio tutte le situazioni pratiche in cui si voglia configurare il movimento autonomo.

Infatti, le finalità di tale studio sono molteplici: alcuni esempi sono l'assistenza (se non automatizzazione) delle procedure di rifornimento in volo, le capacità di inseguimento di un obiettivo in movimento e il volo autonomo in formazione.

Altri campi di applicazione sono lo studio della dinamica morfologica delle espressioni facciali, utili soprattutto per il ramo dell'automazione specializzato all'interazione uomo-macchina.

E ancora si consideri l'opportunità di utilizzare questo sistema di analisi nella medicina, nella diagnosi medica e nella microchirurgia applicata.

In ultimo, si consideri la necessità di monitoraggio di movimenti, che incrementerebbe l'utilizzo di questa applicazione dall'ambito industriale a quello dei sistemi di sicurezza.

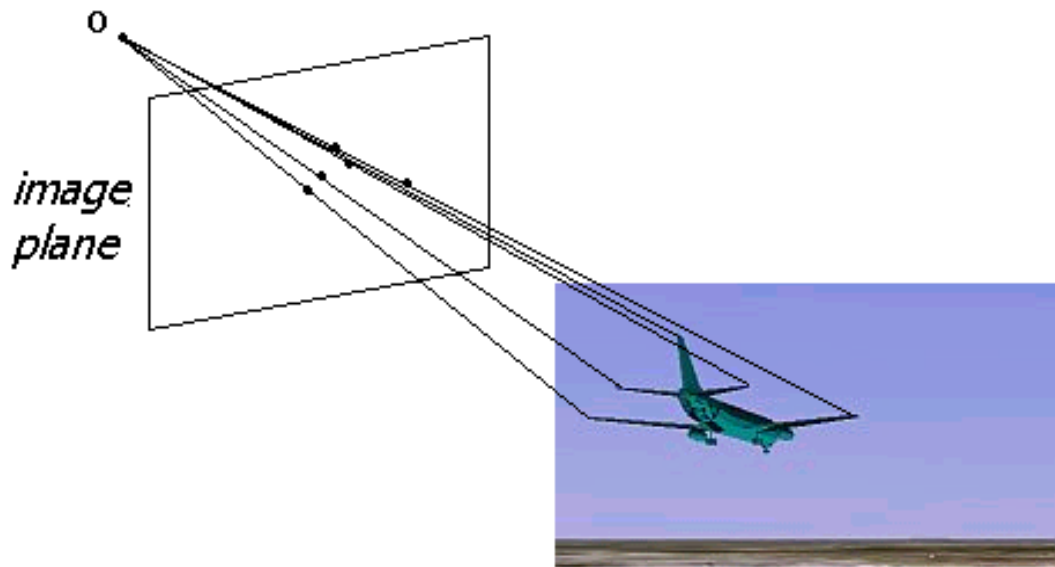
## **1. LE PROBLEMATICHE DI ANALISI**

Il primo problema, detto *pose estimation*, è di carattere prevalentemente geometrico: consiste nel ricavare la rototraslazione che porta la configurazione geometrica dei punti rilevati all'istante osservato a coincidere con la posizione assunta in una configurazione di riferimento.

Avendo come strumento di visione una unica videocamera, le informazioni a disposizione provenienti dall'acquisizione di una sola immagine sono punti bidimensionali, nel nostro caso è necessario dover ricostruire a partire dai suddetti punti le corrispondenti posizioni nello spazio.

Il secondo problema più grande, detto di *matching*, è dato dal riconoscimento automatico della associazione tra punti immagine

rilevati e quelli della configurazione di riferimento; la complessità di tale problema aumenta molto qualora i punti tridimensionali nello spazio assumano configurazioni singolari (è il caso per esempio di due o più punti le cui proiezioni sul piano immagine risultino molto vicine), o in presenza di rumore.



Un esempio di massima

### **1.3 STATO DELL'ARTE**

La base di partenza di questa tesi è il sistema sviluppato presso il dipartimento di Automatica e Sistemi Elettrici della facoltà di Ingegneria di Pisa. La tecnica per la stima di posizione utilizzata è l'algoritmo di Iterazione Ortogonale sviluppato da *Lu, Hager* e *Mjolness* [1](in seguito per comodità verrà denominato LHM), mentre per la fase di assegnamento delle feature si utilizza l'algoritmo comunemente detto dell'Ungherese, ovvero l'ottimizzazione delle distanze euclidee tra i punti immagine estratti dal frame precedente e i punti rilevati sul frame attuale. Per ovviare ad alcuni problemi di corrispondenza, è stato sviluppato un ulteriore meccanismo denominato *Feasible Matching*, che consiste nell'analizzare le distanze fornite all'ungherese alla ricerca di configurazioni sub-ottime, per le quali alcune corrispondenze sono caratterizzate da una distanza comunque piccola dal punto di riferimento. La ricerca di tali configurazioni avviene per mezzo di un insieme di regole euristiche.

### **1.4 OBIETTIVO DELLA TESI**

Scopo di questa tesi è di sviluppare un algoritmo di matching robusto, e di studiare la possibile integrazione della fase di matching con quella di pose estimation.

## 2 IL PUNTO DI PARTENZA: L'ALGORITMO DI ITERAZIONE ORTOGONALE

I punti nel sistema di riferimento dell'oggetto ( $p_i$ ) e i punti tridimensionali nel sistema di riferimento della camera ( $q_i$ ) sono legati da una traslazione rigida:

$$q_i = Rp_i + t, \quad \text{con } R \in SO(3), t \in R^3 \quad (1)$$

Il punto immagine  $v_i = [u_i, v_i, 1]$  è costituito dalle proiezioni prospettiche del punto  $p_i$  sul piano immagine normalizzato. Il centro delle proiezioni della piano della camera è posto nell'origine (Fig.1).

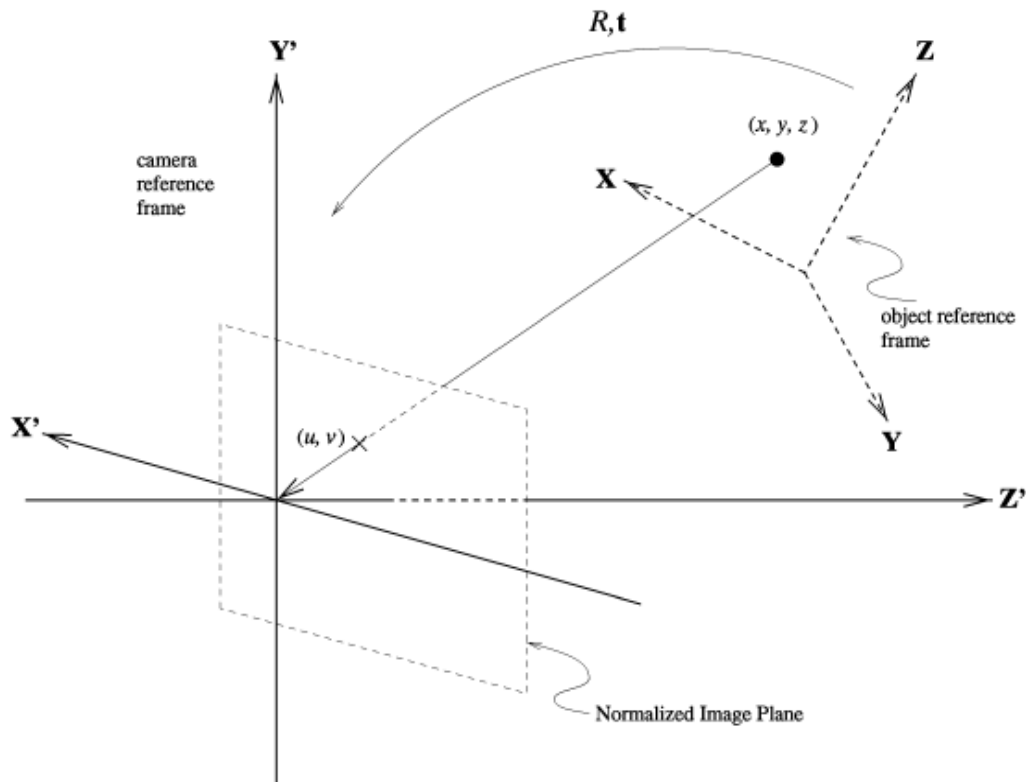


Fig.1

Sia associato al punto immagine  $v_i$  l'operatore di proiezione  $V_i$  che riproietta il punto tridimensionale a cui è applicato sulla linea di vista definita dal punto immagine  $v_i$ . Vale la relazione:

$$(Rp_i + t) = V_i(Rp_i + t) \quad (2)$$

Che specifica il fatto che i punti  $v_i$ ,  $q_i$  e il centro delle proiezioni sono colineari.

Il funzionale di LHM minimizza l'errore di colinearità:

$$\min_{R,t} \sum_i \|(I - V_i)(Rp_i + t)\|^2 = \min_{R,t} \sum_i \|Rp_i + t - V_i q_i\|^2 \quad (3)$$

Definendo  $q_i'$  e  $p_i'$  l'insieme dei punti  $q_i$  e  $p_i$  baricentrati, la soluzione della (3) fornita dagli autori (in forma sintetica) risulta:

$$M = \sum_i^N q_i' p_i'^T$$

$$[U, \Sigma, V] = SVD(M) \quad (4)$$

$$R^* = VU^T$$

e per la traslazione (ricavata in forma chiusa):

$$t(R) = \frac{1}{n} \left( I - \frac{1}{n} \sum_j V_j \right)^{-1} \sum_j (V_j - I) Rp_j \quad (5)$$

LHM sfrutta iterativamente la stima dei  $q_i$  ricavata al passo precedente come nuovo set di punti per la risoluzione della (3) al passo attuale.

Se i punti  $v_i$  risultano perturbati dal rumore, il nuovo funzionale di



LHM diventa [1]:

$$\sum_i (Rp_i + t)^T (I - V_i) \Lambda^{-1} (I - V_i) (Rp_i + t) \quad (6)$$

dove  $\Lambda$  è la matrice di covarianza associata a  $V_i$  dovuta al rumore sul punto immagine  $v_i$ . Questa nuova formulazione inibisce l'uso della tecnica utilizzata da LHM basata sulla decomposizione a valori singolari per la risoluzione.

### 3 SVILUPPI E MODIFICHE ALL'ALGORITMO

#### 3.1 IL MATCHING DA 3D A 3D

Come già detto in precedenza, un algoritmo di pose estimation iterativo si basa sulla stima successiva della rototraslazione tra due insiemi di punti in tre dimensioni: quelli del modello e quelli tridimensionali ricostruiti.

Si assuma che i punti del modello  $\hat{p}_i$  e quelli dell'oggetto  $\hat{x}_i$  siano corrotti da una componente addittiva di rumore:

$$\begin{aligned}x_i &= \hat{x}_i + \Delta x_i \\p_i &= \hat{p}_i + \Delta p_i\end{aligned}\tag{7}$$

Dove  $\Delta x_i$  e  $\Delta p_i$  sono variabili statistiche di covarianza nota:

$$\begin{aligned}Var[x_i] &= E[\Delta x_i \Delta x_i^T] \\Var[p_i] &= E[\Delta p_i \Delta p_i^T]\end{aligned}$$

La relazione tra i punti senza rumore prima e dopo la rotazione è data da:

$$\hat{x}_i = R\hat{p}_i + t, \quad i = 1, \dots, Np\tag{8}$$

L'espressione per la pose estimation più generale risulta quella formulata con la minimizzazione della distanza di Mahalanobis:

$$\begin{aligned}
 J(R, t) = & \sum_i^{Np} (\hat{p}_i - p_i)^T Var[p_i]^{-1} (\hat{p}_i - p_i) \\
 & + \sum_i^{Mp} (\hat{x}_i - x_i)^T Var[x_i]^{-1} (\hat{x}_i - x_i)
 \end{aligned} \tag{9}$$

$$R \in SO(3)$$

Si ponga ora  $\mathbf{M}_P = \mathbf{N}_P$  per ricondurci al caso di nostro interesse.

Per la risoluzione della (9), il problema viene scisso nelle due variabili da minimizzare (a tal proposito si veda riferimento [2]): da qui, saranno minimizzate indipendentemente la rotazione  $R$  e il termine di traslazione  $t$ .

Un approccio iterativo ricercherà la soluzione ottima in entrambi gli spazi.

### 3.2 TERMINE DI ROTAZIONE

Soffermandosi sulla rotazione, il modello descritto nella (8) si riduce dunque a:

$$\begin{aligned}\hat{x}_i &= R\hat{p}_i, \quad i = 1, \dots, Np \\ R &\in SO(3)\end{aligned}\tag{10}$$

Otha e Kanatani [3] hanno dimostrato che la minimizzazione della (9) soggetta alla (10) si riduce alla minimizzazione di:

$$J_R(R) = \sum_i^{Np} (x_i - Rp_i)^T W_i^R (x_i - Rp_i) \quad \hat{R} = \arg \min_{R \in SO(3)} J_R\tag{11}$$

dove:

$$W_i^R = \left( RVar[p_i] R^T + Var[X_i] \right)^{-1}$$

Gli stessi autori forniscono uno schema di ottimizzazione della soluzione basato sui quaternioni usando la tecnica detto *rinormalizzazione* di Kanatani [4], la cui implementazione è provata essere stabile e robusta.

Di seguito viene illustrata sinteticamente tale tecnica, soprattutto per chiarimenti implementativi; per gli approfondimenti teorici si rimanda alla bibliografia.

La non linearità nei vincoli (10) impedisce l'applicazione diretta della tecnica di rinormalizzazione: occorre linearizzarli.

Si consideri una rotazione di angolo  $\Omega$  intorno all'asse  $\mathbf{l}$  (vettore unitario). Si definisca uno scalare  $q_0$  e un vettore tridimensionale  $\mathbf{q}_l$  come:

$$q_0 = \cos \frac{\Omega}{2}, \quad \mathbf{q}_l = l \sin \frac{\Omega}{2}$$

Si noti che per definizione:

$$q_0^2 + \|\mathbf{q}_l\|^2 = 1$$

Al contrario, uno scalare  $q_0$  e un vettore tridimensionale  $\mathbf{q}_l$  determinano unicamente la rotazione  $\mathbf{R}$  attorno all'asse  $\mathbf{l}$  di un angolo  $\Omega$  nella forma:

$$\mathbf{R} = \left( q_0^2 - \|\mathbf{q}_l\|^2 \right) \mathbf{I} + 2 \left( \mathbf{q}_l \mathbf{q}_l^T + q_0 \mathbf{q}_l \times \mathbf{I} \right),$$

dove il prodotto  $a \times \mathbf{A}$  di un vettore  $a$  e una matrice  $\mathbf{A}$  è la matrice le cui colonne sono prodotti vettoriali di  $a$  e le corrispondenti colonne di  $\mathbf{A}$ .

Quindi, la rotazione è unicamente rappresentata da una coppia  $[q_0, \mathbf{q}_l]$ , chiamata un quaternione.

Supponendo un punto  $\bar{r}_\alpha$  subisca una rotazione  $\mathbf{R}$  di un angolo  $\Omega$  attorno all'asse  $\mathbf{l}$ , finendo in una nuova posizione  $\bar{r}'_\alpha$ .

Dalla figura 1, si può notare che lo spostamento  $\bar{r}'_\alpha - \bar{r}_\alpha$  e il punto medio  $(\bar{r}_\alpha + \bar{r}'_\alpha)/2$  sono legate dalla relazione:

$$\bar{r}'_\alpha - \bar{r}_\alpha = 2 \tan \frac{\Omega}{2} l \times \frac{\bar{r}_\alpha + \bar{r}'_\alpha}{2} \quad (12)$$

Risolvendo quest'ultima per  $\bar{r}'_\alpha$  al variare di  $\bar{r}_\alpha$ , possiamo ottenere una relazione equivalente a (9) espressa nei termini dell'angolo  $\Omega$  e dell'asse  $\mathbf{l}$  della rotazione  $\mathbf{R}$ .

Da ciò segue che l'equazione (12) è equivalente alla relazione (9). Moltiplicando entrambi i lati dell'equazione (12) per  $\cos(\Omega/2)$ , otteniamo dopo alcuni passaggi:

$$q_0 (\bar{r}'_\alpha - \bar{r}_\alpha) + (\bar{r}'_\alpha + \bar{r}_\alpha) \times q_l = 0.$$

Si sottolinea che  $q_0$  corrisponde a  $p_i$  e  $q_l$  a  $x_i$ .

Si applichi questa sostituzione e si definiscano ora la seguente matrice  $3 \times 4$   $Y_i$  e il vettore unitario di quattro dimensioni  $q$  come:

$$Y_i = (x_i - p_i (x_i + p_i) \times I), \quad q_l = \begin{pmatrix} q_0 \\ q_l \end{pmatrix} \quad (13)$$

dove  $\times$  rappresenta prodotto vettoriale (o esterno). Si consideri  $\bar{Y}_i$ , il

valore di  $Y_i$  ottenuto sostituendo  $p_i$  e  $x_i$  rispettivamente con  $\hat{p}_i$  e  $\hat{x}_i$ .  
 Una rotazione può essere espressa dal quaternion  $q$ , di modo che i vincoli descritti dall'equazione (10) possono essere riscritti come:

$$\bar{Y}_i q = 0. \quad (14)$$

Risultando lineari.

Ora il problema è diventato la minimizzazione della (9) soggetta alla (14). Si dimostra che tale problema si può ridurre alla minimizzazione rispetto a  $q$  di:

$$J = (q, Mq) \rightarrow \min \quad (15)$$

Dove  $M$  è definita come:

$$M = \sum_{i=1}^{N_p} Y_i^T W_i Y_i \quad (16)$$

e  $W_i$  una matrice 3x3 definita come:

$$W_i = \left( q_0^2 (V[p_i] + V[x_i]) - 2q_0 S \left[ q_l \times (V[p_i] - V[x_i]) \right] + q_l \times (V[p_i] + V[x_i]) \times q_l \right)^{-1} \quad (17)$$

$S[ \cdot ]$  designa l'operazione simmetrizzazione:  $S[A] = (A + A^T)/2$ .

L'entità statistica  $E[\Delta M]$ , necessaria per applicare la tecnica di rinormalizzazione, può essere espressa come:

$$E[\Delta M] = \varepsilon^2 N$$

(18)

Definendo:

$$N = \begin{pmatrix} n_0 & n^T \\ n & N' \end{pmatrix} \quad (19)$$

Dove:

$$n_0 = \sum_{i=1}^{N_P} (W_i; V[p_i] + V[x_i]),$$

$$n = -2 \sum_{i=1}^{N_P} t_3 \left[ A \left[ W_i (V[p_i] - V[x_i]) \right] \right],$$

$$N' = \sum_{i=1}^{N_P} \left[ W_i \times (V[p_i] + V[x_i]) \right]$$

$(A;B)$  rappresenta il prodotto interno tra due matrici  $A$  e  $B$ .

L'operatore  $A[ \cdot ]$  denota asimmetrizzazione:  $A[B] = (B - B^T)/2$ .

La procedura di rinormalizzazione che si ottiene è la seguente:

1. Dall'insieme di dati  $\{ p_i \}$  e  $\{ x_i \}$  si computi  $Y_i$ ,  $i = 1, \dots, N_P$  a partire dalla prima delle (13).
2. Si setti  $c = 0$  e  $W_i = I$ ,  $i = 1, \dots, N_P$ .
3. Si computi la matrice  $M$  come dall'equazione (16).
4. Si computi la matrice  $N$  come dall'equazione (19).



5. Si calcoli il più piccolo autovalore  $\lambda$  della matrice:

$$M^* = M - cN$$

e il corrispondente autovettore  $\mathbf{q} = (q_0, q_1, q_2, q_3)^\top$ .

6. se  $\|\lambda\| < \delta$ , con  $\delta$  prossimo a 0 l'algoritmo ha concluso e ritorna  $\mathbf{q}$ . Altrimenti si aggiorni  $W_i$  come espresso dalla (17) con  $\mathbf{q}$  suddiviso come specificato dalla seconda delle (13) e  $c$  come segue:

$$c \leftarrow c + \frac{\lambda}{(\mathbf{q}, N\mathbf{q})}$$

7. si ritorni al passo 3.

### 3.3 TERMINE DI TRASLAZIONE

Riguardo al termine di traslazione, si ponga ora:

$$J_t(t) = \sum_i^{Np} (x_i - p_i - t)^T w_i^t (x_i - p_i - t) \quad \hat{t} = \arg \min_{t \in R^3} J_t \quad (20)$$

per cui la minimizzazione l'equazione (9) si riduce a:

$$\hat{t} = \left( \sum_i^{Np} w_i^t \right)^{-1} \sum_i^{Np} w_i^t (x_i - p_i) \quad (21)$$

dove:

$$w_i^t = \left( Var[p_i] + Var[x_i] \right)^{-1}$$

La cui soluzione risulta:

$$J(R, t) = \sum_i^{Np} (Rp_i + t - v_i q_i)^T \Lambda^{-1} (Rp_i + t - v_i q_i) \quad (22)$$

L'iter della tecnica risolutiva nel suo complesso risulta:

1. stimare R' come prima descritto usando  $\{ x_i' = x_i - t' \}$  e  $\{ p_i \}$ ;
2. ruotare i punti attraverso R':  $\{ p_i' = R' p_i \}$ ;

3. stimare  $t'$  usando  $\{ x_i \}$  e  $\{ p_i' \}$ ;

4. verificare la convergenza, altrimenti tornare al punto 1).

il quale rappresenta uno strumento molto potente per il calcolo della rototraslazione tra due insiemi di punti tridimensionali.

### 3.4 RICOSTRUZIONE DEI PUNTI IN 3D A PARTIRE DAI PUNTI IN 2D

Riguardo alla ricostruzione dei punti dell'oggetto a partire dalle loro proiezioni prospettiche, la scelta operata è la stessa usata dall'algoritmo LHM, mediante la riproiezione sulla linea di vista descritta dal punto  $v_i$  con l'operatore  $V_i$ ;

Si vuole quindi minimizzare l'errore di colinearità.

In virtù di ciò l'equazione (9) diventa (per il momento si trascura l'errore sui punti del modello  $p_i$  e quindi viene omessa la prima sommatoria dall'equazione):

$$J(R, t) = \sum_i^{Np} \left( Rp_i + t - v_i q_i \right)^T \Lambda^{-1} \left( Rp_i + t - v_i q_i \right) \quad (23)$$

Come già detto in precedenza,  $\Lambda_i$  è la matrice di covarianza associata alla matrice di proiezione  $V_i$ , causata dal rumore sul punto immagine  $v_i$ . Sia  $v_i$  il punto immagine senza rumore e  $v_i'$  quello corrotto, per cui:  $v_i' = v_i + \Delta v_i$ . Il rumore si ipotizzi Gaussiano con  $E(\Delta v_i) = 0$  (è da notare che  $v_i = [u_i, v_i, 1]$ ,  $\Delta v_i = [\Delta u_i, \Delta v_i, 0]$ ) e matrice di covarianza:

$$C_{2D} = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Sia  $V_i'$  l'operatore proiezione legato a  $v_i'$ .

$$V_i' = V_i + \Delta V_i$$

Data la struttura della matrice di proiezione, è possibile scrivere:  
e ponendo  $x_i$  la stima 3D del  $i$ -esimo punto:

$$V_i' x_i = V_i x_i + \Delta V_i x_i \quad (24)$$

Dove quindi le caratteristiche statistiche dell'entità  $\Lambda_i$   
sono legate a  $\Delta V_i$ , e quest'ultimo a quelle note sui  $v_i'$ .

Una tecnica per stimare tali informazioni può essere la seguente.  
Innanzitutto si definisca l'operazione di proiezione attuata da  $V_i$  come  
una funzione  $V_f(x_i, v_i)$ , che proietta il punto  $x_i$  lungo la linea di vista  
del punto  $v_i$ .

Si linearizzi tale funzione intorno al punto immagine rumoroso [5]:

$$V_i' x_i = V_f(x_i, v_i + \Delta v_i) = V_f(x_i, v_i) + \left[ \frac{\partial V_f}{\partial v} \right]_{x_i, v_i} \Delta v_i \quad (25)$$

Si assuma che la distanza tra la stima del punto immagine (la proiezione  
di  $q_i$  all'iterazione corrente) e il punto immagine corretto sia piuttosto  
piccola, cioè che valga la condizione:

$$V_i \mathbf{x}_i \cong V_f \left( \mathbf{x}_i^{est}, \mathbf{v}_i \right)$$

Allora si può considerare valida l'assunzione:

$$\Delta V_i \mathbf{x}_i = \left[ \frac{\partial V_f}{\partial \mathbf{v}} \right]_{\mathbf{x}_i^{est}, \mathbf{v}_i} \Delta \mathbf{v}_i$$

Per semplicità si scriverà  $\Delta V_i \mathbf{x}_i = M_i \Delta \mathbf{v}_i$ , dove  $M$  rappresenta lo Jacobiano di  $V$  valutato in  $(\mathbf{x}_i, \mathbf{v}_i)$ . La formula per la covarianza di  $\Lambda_i$ :

$$\begin{aligned} \Lambda_i &= E \left( \Delta V_i \mathbf{x}_i \Delta V_i \mathbf{x}_i^T \right) = E \left( M_i \Delta \mathbf{v}_i \Delta \mathbf{v}_i^T M_i^T \right) \\ &= M_i E \left( \Delta \mathbf{v}_i \Delta \mathbf{v}_i^T \right) M_i^T = M_i C_{2D} M_i^T . \end{aligned} \quad (26)$$

Ricavando così una relazione tra i parametri statistici dell'errore sui punti di ingresso e il conseguente errore sull'operatore di proiezione.

Se il matching non è corretto, la (23) mi fornisce la matrice di rototraslazione che minimizza l'errore di colinearità tra i punti del modello e quelli ricostruiti mediante proiezione.

Usando questa nuova tecnica per risolvere la minimizzazione dell'errore

di colinearità, è stato osservato che il calcolo della traslazione risulta più preciso e rapido se viene ricavato secondo una formulazione simile a quella derivata da LHM:

$$t(R) = \left( \sum_i^{N_p} V(V[x_i])^{-1} (I - V_i) \right)^{-1} \sum_i^{N_p} V(V[x_i])^{-1} (V_i - I) Rp_i \quad (27)$$

Inoltre considerando la versione completa dell'equazione (9), può essere contemplato anche l'errore di conoscenza della configurazione dei punti del modello.

## 4 VALUTAZIONI E TEST DELL'ALGORITMO

Per testare la robustezza e l'efficacia dell'algoritmo modificato, sono state effettuate diverse simulazioni di Monte-Carlo.

Per ogni simulazione le coordinate dei punti nel sistema di riferimento dell'oggetto corrispondono agli otto vertici della cella  $[-5,5] \times [-5,5] \times [-5,5]$ . Sono generate casualmente una matrice di rotazione e un vettore di traslazione con componenti  $x$  e  $y$  nell'intervallo  $[5,15]$  e componente  $z$  in  $[10,50]$ ; i punti dell'oggetto sono poi trasformati attraverso tale rototraslazione, e di tali punti vengono poi calcolate le proiezioni prospettiche sul piano immagine normalizzato.

Riguardo all'inizializzazione dell'algoritmo, viene sfruttato lo stesso metodo usato da LHM, che consta nel compiere una iterazione sui punti immagine supponendo i punti ricostruiti in una approssimazione detta "weak perspective".

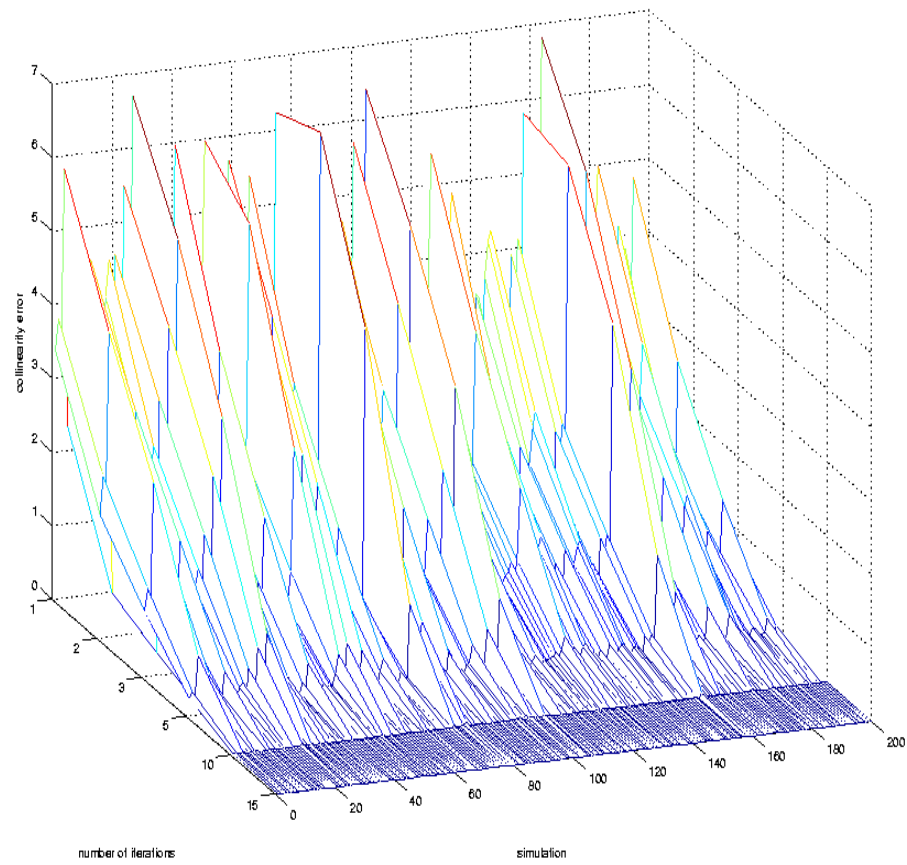
Ovviamente quello modificato ha ingresso anche la matrice di covarianza dell'errore sui punti (o perlomeno un ordine di grandezza). L'algoritmo di riferimento per il confronto è ovviamente LHM, inizializzato nello stesso modo.

### 4.1 Test T1

Il primo aspetto da testare è l'efficienza e la precisione della convergenza in presenza di un matching corretto (grafico G1). Si riportano di seguito i risultati ottenuti dall'effuttazione di 200 simulazioni, plottando l'errore di colinearità finale in funzione del numero delle iterazioni. Ciò per valutare non solo la precisione, ma



anche la velocità di convergenza.

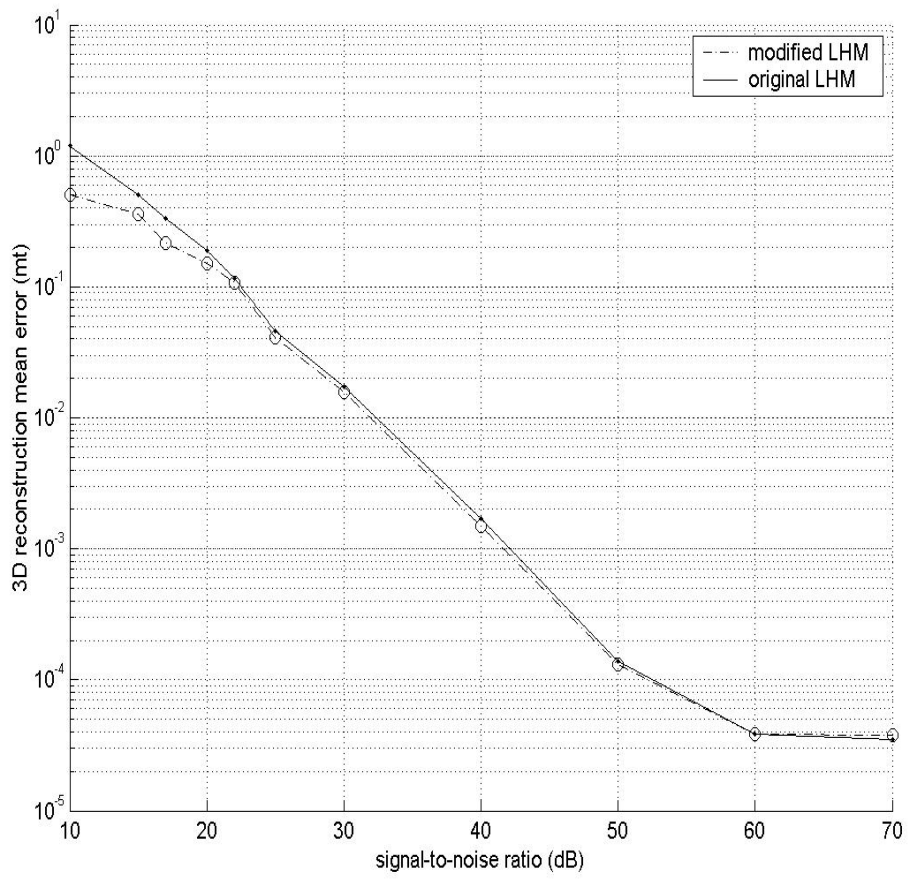


G1.

## ***4.2 Test T2***

Si valuta ora la convergenza in presenza di rumore sui punti. Per fare ciò, dopo aver creato tutti i dati necessari come suddetto, alle proiezioni prospettiche viene sommato del rumore, con rapporto segnale/rumore pari a  $SNR_i = [10\ 15\ 17\ 20\ 22\ 25\ 30\ 40\ 50\ 60\ 70]$ ; per ognuno dei valori sono state effettuate più di 500 simulazioni.

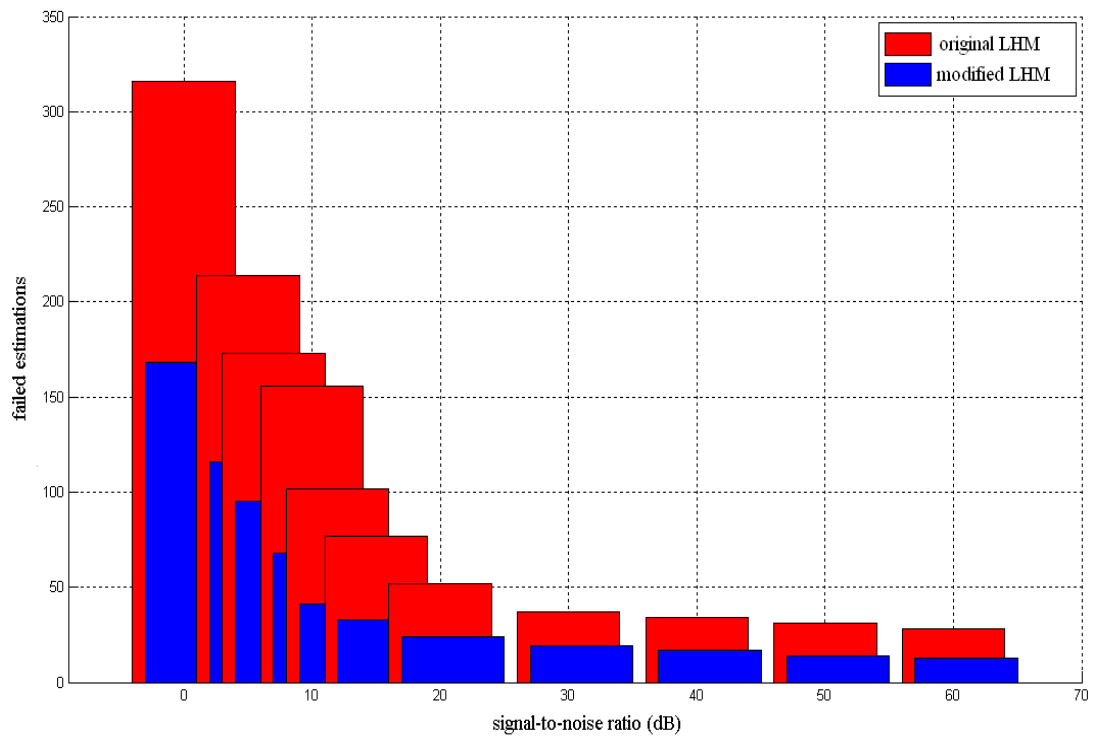
Per valutare la bontà della soluzione ottenuta, non si considera più l'errore di colinearità in quanto, trattandosi di due algoritmi convergenti, una configurazione che minimizza tale errore viene sempre raggiunta, bensì si computa l'errore tra la posizione dei punti tridimensionali corretti e quella effettivamente stimata dagli algoritmi. Il grafico (G2) è plottato in scala logaritmica.



G2

Si evince dal grafico che l'errore di ricostruzione dell'algoritmo modificato risulta minore nei casi in cui la componente di rumore è più consistente, mentre per rumore sui punti che tenda a 0 (SNR dell'ordine di 50), entrambi gli errori di ricostruzione tendono allo stesso valore. La precisione di entrambi gli algoritmi è comunque ampiamente soddisfacente.

La performance dell'algoritmo modificato risulta ancor più positiva se, sempre sugli stessi dati, si vada a considerare il numero di stime di posizione fallite (cioè le simulazioni in cui risultato converge, ma ad una stima completamente inesatta) di ciascun algoritmo (G3): il numero di convergenze positive della tecnica modificata risulta sensibilmente minore, anche in presenza di forte rumore.



G3

I principali vantaggi dell'algorithmo modificato sono:

- Soluzione attendibile nel caso di matching rumoroso;
- Miglior robustezza in presenza di rumore;
- L'algorithmo di stima accetta in ingresso ulteriori informazioni sul rumore attraverso le matrici di covarianza;
- Possibilità di gestire rumore sui punti del modello;

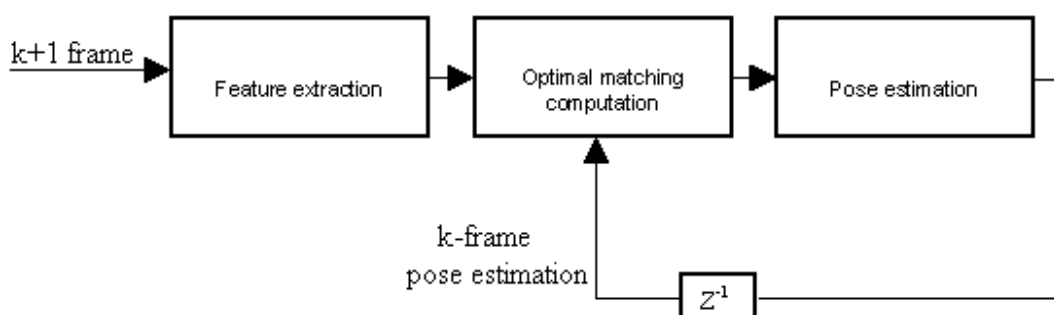
Di fronte a questi vantaggi è da sottolineare che non si riscontra comunque un apprezzabile decadimento delle prestazioni rispetto all'algorithmo originario.

## 5 SVILUPPI FUTURI

Riguardo alle future possibilità di sviluppo di questo tipo di sistema analitico, detto computer vision, è opportuno segnalare, tra le altre, in particolar modo la tecnica del Feasible Matching, poiché essa rappresenta uno strumento efficace per la soluzione di alcune casistiche contingenti, in particolar modo delle situazioni di matching dubbie, qualora l'oggetto assumesse quel tipo di configurazioni che si sono definite "ambigue" nella introduzione.

Le prestazioni di tale tecnica, infatti, diminuiscono con l'aumento di rumore sui punti immagine; pertanto al crescere delle possibili corrispondenze sub-ottime è necessario valutarne tutte le possibili combinazioni.

Non di meno, anche se ciò è innegabile, si tenga tuttavia presente, che tra le varie ipotesi di funzionamento dell'algorithmo LHM esiste anche quella di un Matching corretto, che si può tentare di estrapolare tra le altre facenti riferimento invece ai punti di rumore.



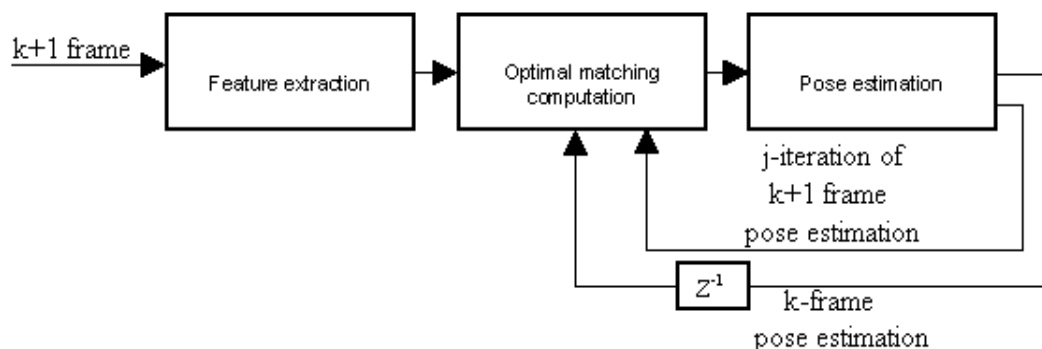
Il Feasible Matching non è comunque sia scevro da lacune, la principale delle quali è costituita dalla separazione della fase di matching da quella della pose estimation.

Per capire quanto sia importante integrarle, si tenga presente che l'unica relazione che lega tra loro i punti immagine corretti è la loro comune caratteristica di essere proiezioni di uno spostamento rigido di una configurazione di punti nota.

Ciò sottolinea come la rototraslazione stimata ad ogni iterazione dall'algoritmo possa essere una informazione utile per la soluzione del matching.

Questo lascia facilmente intuire quali potrebbero essere i futuri ambiti di ricerca, a partire su tale assunto, per sviluppare e incrementare l'efficacia del Feasible Matching, e le relative applicazioni pratiche.

Grazie alla robustezza dell' algoritmo di pose estimation sviluppato, che può trattare punti immagine rumorosi in ingresso, fornendo in uscita una stima di rototraslazione rigida, è possibile utilizzare una configurazione del tipo:





La stima di rototraslazione di ogni singola iterazione, quindi, potrebbe essere usata come feedback per l'algoritmo di matching, unitamente alla possibilità, che appartiene all'algoritmo modificato, di gestire le informazioni di carattere statistico del rumore.

In tal modo la tecnica, con gli annessi incrementi empirici, conseguirebbe un grado di perfezionamento e raffinatezza tali da poter disporre di superiori caratteristiche di precisione ed efficienza del sistema di analisi del Matching.

## 6 CONCLUSIONI

Durante questa tesi sono state affrontate le tematiche legate alla visione artificiale. Sono state studiate e vagliate un insieme di proposte alla ricerca di un sistema per aumentare l'efficienza e l'efficacia del sistema basato su LHM.

Dal punto di vista geometrico, gran parte delle proposte si basavano su l'uso di approssimazioni prospettiche (ortoprospectiche e paraprospectiche): ciò presentava degli svantaggi legati alla non perfetta rappresentazione dei rispettivi punti nello spazio.

Si è giunti alla conclusione che la tecnica di ricostruire i punti 3D sfruttando la colinearità espressa dall'operatore di proiezione lungo la linea di vista risulta sicuramente una delle proprietà più utili da sfruttare; questa infatti è stata presa come uno dei capisaldi di tutto il lavoro successivo.

Durante il periodo di studio si è consolidata la certezza che l'unico modo di avere una informazione certa per risolvere il matching era il loro legame con i punti tridimensionali.

Da qui il secondo punto prefissato: la possibilità di integrare tra fase di matching con la fase di pose estimation. Perché le due fasi potessero interloquire, si è scelto di cercare un algoritmo di stima di posizione che avesse un approccio statistico al problema. Da qui la soluzione proposta, che sfruttando lo *scheletro* geometrico di HLM, ricerca la

soluzione in maniera robusta. Ciò è confermato dai risultati delle simulazioni, e si evince dal basso errore di ricostruzione dei punti, e dal numero di stime totalmente scorrette.

A ciò si aggiunge comunque una velocità di convergenza paragonabile a quella già ottima di LHM.

## BIBLIOGRAFIA

- 1) C. Lu, G.D. Hager e E. Mjølness. *Fast and Globally Convergent Pose Estimation from Video Image*. IEEE transaction on pattern analysis and machine intelligence, 2000.
- 2) R.S. J. Estépar, A. Brun, C.F. Westin. *Robust Generalized Total Least Squares Iterative Closest Point Registration*. Harvard Medical School, Boston, MA, USA Department of Biomedical Engineering, Linköping University, Sweden, 2004.
- 3) N. Ohta e K. Kanatani. *Optimal Estimation of Three-dimensional Rotation and Reliability Evaluation*. IEEE Transactions on Information and Systems, E81-D(11):1247–1252, 1998.
- 4) K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, 1996.
- 5) W. Hoff, T. Vincent. *Analysis of Head Pose Accuracy in Augmented Reality*, IEEE transaction on visualization and computer graphics, 2000.