

ALMA MATER STUDIORUM
UNIVERSITÀ DEGLI STUDI DI BOLOGNA
CAMPUS DI CESENA

Scuola di Ingegneria e Architettura
Corso di Laurea Magistrale in Ingegneria Informatica

AUTOCALIBRAZIONE DI UNA TELECAMERA
UTILIZZANDO UNA SINGOLA VISTA

Elaborata nel corso di:
Elaborazione Delle Immagini LM

Tesi di Laurea di:
ANDREA ARNOFFI

Relatore:
Prof. ALESSANDRO
BEVILACQUA

Co-relatore:
ALESSANDRO GHERARDI

ANNO ACCADEMICO 2013–2014
SESSIONE III

PAROLE CHIAVE

Computer Vision

Visione artificiale

Geometria proiettiva

Autocalibrazione

Punti di fuga

Indice

Introduzione	vii
Lavori precedenti	viii
Contributo	viii
Struttura della tesi	ix
1 Calibrazione della telecamera	1
1.1 Matrice di calibrazione	1
1.1.1 Modello Pinhole	1
1.1.2 Mapping di proiezione centrale	4
1.1.3 Posizione e orientamento della telecamera	6
1.1.4 Modello generale di telecamera proiettiva	7
1.1.5 Distorsione radiale	8
1.2 Calibrazione dai punti di fuga	9
1.2.1 Elementi di geometria proiettiva	10
1.2.2 Relazione di ortogonalità tra punti di fuga	16
1.2.3 Calibrazione da tre punti di fuga ortogonali	18
2 Autocalibrazione dai punti di fuga	21
2.1 Estrazione dei segmenti da un'immagine	24
2.1.1 Rilevamento dei bordi	25
2.1.2 Estrazione dei segmenti	25
2.1.3 Raffinamento dei segmenti	28
2.1.4 Considerazioni sull'estrazione dei segmenti	29
2.2 Clusterizzazione dei segmenti	31
2.2.1 Determinazione dei punti di fuga potenziali iniziali	31
2.2.2 Consistenza tra i punti di fuga potenziali e i segmenti	32
2.2.3 Clusterizzazione iniziale e aggregazione iterativa	35

2.2.4	Considerazioni sulla clusterizzazione dei segmenti . . .	37
2.3	Stima dei punti di fuga da ogni cluster	39
2.3.1	Stima per minimizzazione dell'errore	40
2.3.2	Stima tramite mediana	44
2.3.3	Stima per massimizzazione della consistenza	46
2.4	Raffinamento dei punti di fuga stimati	47
2.5	Determinazione dei punti di fuga ortogonali	49
2.5.1	Calcolo della matrice di calibrazione potenziale . . .	50
2.5.2	Errore di ortogonalità	50
2.5.3	Errore del punto principale	52
2.5.4	Errore delle euristiche	52
3	Risultati sperimentali	57
3.1	Testbed	57
3.2	Obiettivi dei test	59
3.3	Determinazione della ground truth	59
3.4	Immagini dei test	60
3.5	Risultati ottenuti	68
3.5.1	Misurazione intrinseca della qualità	68
3.5.2	Misurazione statistica della qualità	79
3.6	Confronto con altre metodologie	89
3.6.1	Confronto col metodo proposto da Xu, Oh e Hoogs .	89
3.6.2	Confronto col metodo proposto da Nieto e Salgado .	90
4	Conclusioni	103

Introduzione

Nell'ambito della computer vision, la calibrazione di una telecamera è un passo fondamentale per poter impiegare il sensore della telecamera come uno strumento di analisi quantitativa. Grazie alla calibrazione infatti, da una scena rappresentata in un'immagine 2D è possibile risalire alla struttura geometrica degli oggetti inquadrati e quindi ricavare misure quantitative della scena tridimensionale, come la distanza di un oggetto dalla telecamera oppure le sue dimensioni, partendo da un'immagine bidimensionale.

Essendo questo un aspetto centrale, esiste una vasta letteratura dedicata al tema della calibrazione e sono state sviluppate diverse tecniche per raggiungere lo scopo. Molte di queste di queste richiedono l'impiego di oggetti o pattern di calibrazione, ossia degli oggetti fisici con caratteristiche ben definite, che consentano di determinare un mapping preciso tra l'oggetto reale inquadrato e il risultato rappresentato dalla sua proiezione nell'immagine. Altre tecniche invece si basano sull'impiego di più immagini, scattate da diverse angolazioni, per ricavare informazioni sulla struttura geometrica degli oggetti inquadrati e la posizione della telecamera, mediante tecniche di triangolazione.

Questi metodi, benché ormai consolidati e in grado di fornire risultati piuttosto precisi e affidabili, presentano anche diverse limitazioni che in certi contesti non possono essere tollerate. Ad esempio talora non sia possibile ricorrere all'impiego di oggetti di calibrazione, o nel caso in cui si voglia ridurre al minimo l'intervento dell'utente, oppure dove una calibrazione manuale sia impraticabile. Nel caso di una telecamera fissata, l'approccio di sfruttare più immagini prese da angolazioni diverse non è percorribile, mentre con ottiche a lunghezza focale variabile (telecamere zoom) non è possibile ricorrere ad una calibrazione manuale per ogni possibile impostazione. Per superarle è possibile ricorrere a tecniche di autocalibrazione che desumano le informazioni dall'analisi di una singola immagine.

Lavori precedenti

Poiché l'autocalibrazione è un argomento di primaria di importanza, negli anni è stato indagato da più parti e col tempo sono state avanzate diverse proposte per risolvere il problema.

Alcuni si affidano alla trasformata di Hough dei segmenti sulla sfera di Gauss [6], per poi affidarsi a classificare i segmenti in base alla loro orientazione. Questo approccio si è dimostrato molto sensibile al rumore e poco adatto a gestire gli outlier, portando così ad un'errata classificazione dei segmenti. Per limitare questi effetti, spesso viene impiegato un approccio iterativo di tipo EM [5][7][9] per raffinare la classificazione.

L'approccio Bayesiano invece si è rivelato molto più robusto verso gli outlier e le misure affette da rumore. Secondo questo tipo di approccio, le probabilità non vengono interpretate come frequenze, ma come livelli di fiducia. Mentre in alcuni lavori [2] i punti di fuga vengono desunti implicitamente, nella maggior parte dei casi vengono trattati in modo esplicito come il risultato dell'intersezione di segmenti individuati nell'immagine. Nel tal caso viene impiegata una funzione di misura della consistenza [5][7][9] per valutare la bontà dell'associazione di un segmento ad un punto di fuga.

Contributo

La tesi propone un metodo di autocalibrazione degli intrinseci della telecamera utilizzando una singola vista, sfruttando i punti di fuga riconosciuti nell'immagine, che permetta di sopperire ai limiti dei metodi di calibrazione convenzionali e che garantisca una precisione sufficiente a fornire prestazioni adatte ai compiti più comuni della computer vision. Vengono proposti anche nuovi approcci per migliorare lo stato dell'arte dell'autocalibrazione con i punti di fuga di una singola vista.

In particolare viene fornita una nuova metodologia per la determinazione dei punti di fuga, dai cluster di segmenti identificati. Inoltre vengono proposti degli approcci euristici che favoriscono la selezione della terna corretta di punti di fuga ortogonali.

Un ulteriore obiettivo che ci si propone, è di poter inserire tale tecnica di autocalibrazione all'interno di una libreria open-source. Per questo motivo è stata di particolare importanza la progettazione di un metodo modulare, per garantirne la massima flessibilità e possibilità di intervento.

Viene posta attenzione anche alle prestazioni del sistema progettato, permettendo di intervenire sui singoli moduli per trovare il giusto compromesso tra precisione e prestazioni. Questo aspetto è molto importante per consentire l'impiego della tecnica per attività in tempo reale.

Struttura della tesi

Il primo capitolo presenta in modo formale il problema legato all'autocalibrazione, descrivendo i concetti fondamentali e le problematiche da affrontare per riuscire a ricavare la matrice di calibrazione della telecamera.

Nel secondo capitolo viene proposto e sviluppato l'approccio scelto, presentando un caso di studio. Vengono anche descritte alcune delle alternative valutate, esaminandone le criticità.

Nel terzo capitolo vengono discussi i risultati ottenuti dall'approccio proposto, confrontandoli con le prestazioni ottenibili dalla calibrazione classica e da altri metodi di autocalibrazione a singola vista.

Il quarto capitolo è dedicato alle conclusioni e a gli sviluppi futuri.

Capitolo 1

Calibrazione della telecamera

In questo capitolo viene spiegato cosa si intende per calibrazione di una telecamera e qual è la matrice di calibrazione di una telecamera generica. Vengono esplicitati anche i concetti fondamentali di geometria prospettica, e come possa venire sfruttata la prospettiva per dedurre la matrice di calibrazione a partire da un'immagine.

1.1 Matrice di calibrazione

Una telecamera fornisce un mapping da un mondo tridimensionale ad un'immagine piana bidimensionale. Tale mapping può essere modellato attraverso una matrice di proiezione della telecamera.

Questa matrice sfrutta a sua volta una matrice di calibrazione, che descrive i parametri intrinseci della telecamera e permette di determinarne la posizione in relazione all'immagine.

Il fine della calibrazione è di individuare i valori corretti della matrice di calibrazione.

1.1.1 Modello Pinhole

Una telecamera pinhole è fondamentalmente una scatola, con un piccolo foro (detto pinhole) in uno dei suoi lati in grado di lasciar passare la luce. Se ipotizziamo che il foro sia molto piccolo, idealmente come un punto, otterremo che per ogni punto della scena può esserci solamente un raggio passante per il foro. Tale raggio andrà ad intersecarsi con il lato interno

della scatola, opposto a quello del foro, producendo un'immagine capovolta della scena.

Il problema del capovolgimento può essere corretto facilmente se, invece di considerare il piano immagine, viene considerata un'immagine virtuale come rappresentato in figura 1.1.

Questa immagine virtuale si troverà in un piano virtuale parallelo al piano immagine, ma dal lato opposto rispetto al pinhole.

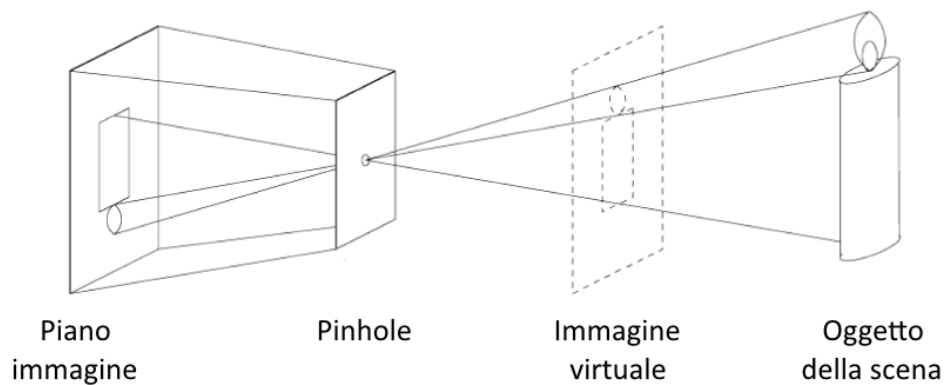


Figura 1.1: Rappresentazione schematica degli elementi fondamentali del modello Pinhole

Se poniamo il centro di proiezione all'origine di un sistema di coordinate euclideo, otteniamo il modello rappresentato in figura 1.2.

Il piano $Z = f$ verrà chiamato *piano focale* o *piano immagine*.

Il punto \mathbf{C} è il centro di proiezione ed è chiamato anche *centro ottico*, o anche *centro della telecamera*.

La retta passante per il centro della telecamera e perpendicolare al piano immagine, è chiamata *asse principale* o *raggio principale* della telecamera.

Il punto \mathbf{p} , intersezione tra l'asse principale e il piano immagine, è detto *punto principale*.

Il piano parallelo al piano immagine, passante per il centro della telecamera, è chiamato *piano principale* della telecamera.

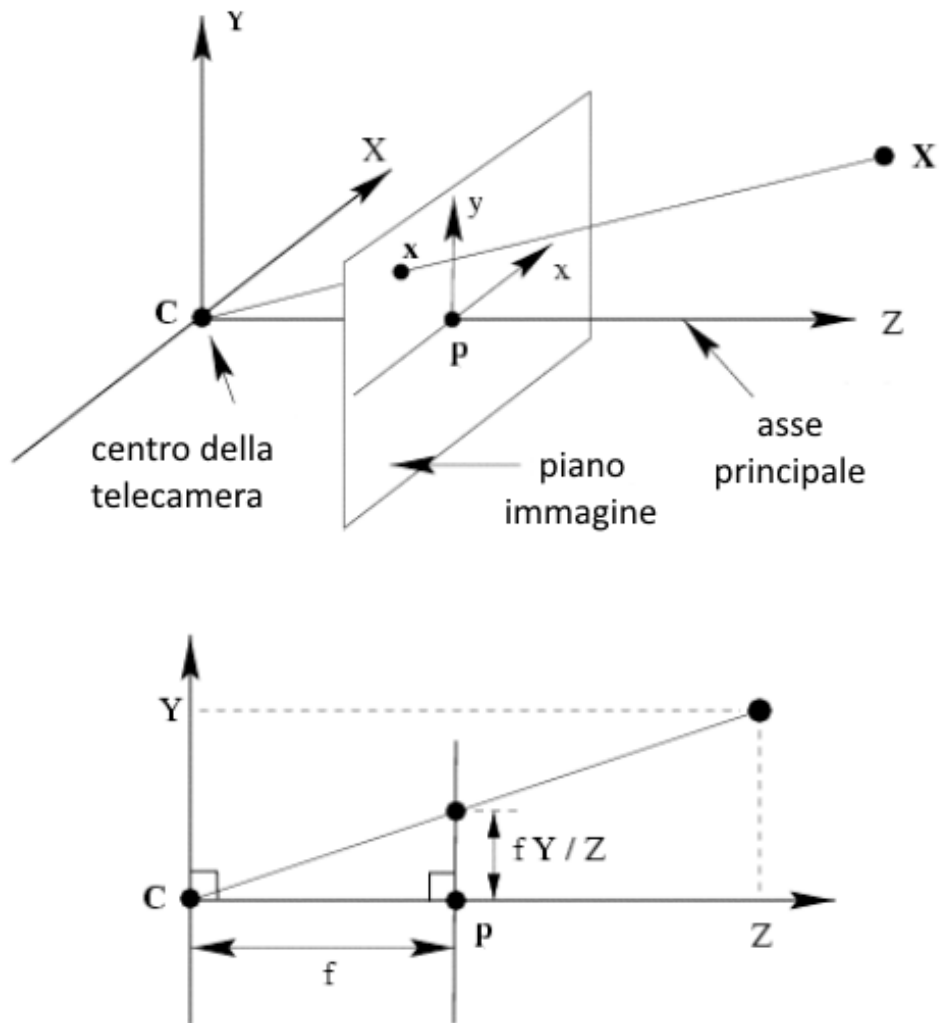


Figura 1.2: Geometria della telecamera pinhole: il centro della telecamera è posizionato all'origine del sistema di riferimento e il piano immagine è rappresentato di fronte al centro della camera, in corrispondenza del piano dell'immagine virtuale.

1.1.2 Mapping di proiezione centrale

Un punto nello spazio \mathbb{R}^3 , con coordinate $\mathbf{X} = (x, y, z)^\top$, viene mappato in un punto \mathbf{x} del piano immagine, in corrispondenza dell'intersezione con la retta che congiunge il punto \mathbf{X} al centro di proiezione \mathbf{C} .

In virtù di ciò, le coordinate del punto sul piano immagine varranno $\mathbf{x} = \left(\frac{f \cdot x}{z}, \frac{f \cdot y}{z}, f\right)^\top$. Ignorando l'ultima coordinata, otteniamo un mapping da \mathbb{R}^3 a \mathbb{R}^2 , ossia dalle coordinate della scena alle coordinate immagine. Questo mapping può essere espresso come:

$$(x, y, z)^\top \mapsto \left(\frac{f \cdot x}{z}, \frac{f \cdot y}{z}\right)^\top$$

In coordinate omogenee il mapping può essere scritto come:

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f \cdot x \\ f \cdot y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \text{diag}(f, f, 1) [\mathbf{I} | \mathbf{0}] \mathbf{X} \quad ,$$

dove $\text{diag}(f, f, 1)$ è una matrice diagonale, mentre $[\mathbf{I} | \mathbf{0}]$ è una matrice 3×4 , composta da una matrice identità 3×3 e un vettore colonna formato da zeri.

Ponendo $\mathbf{P} = \text{diag}(f, f, 1) [\mathbf{I} | \mathbf{0}]$, l'equazione può essere scritta come:

$$\mathbf{x} = \mathbf{P} \cdot \mathbf{X}$$

dove \mathbf{X} corrisponde alle coordinate di un punto della scena, rappresentato tramite un vettore di coordinate omogenee $(x, y, z, 1)^\top$, mentre \mathbf{x} è il punto immagine di coordinate omogenee $(f \cdot x, f \cdot y, z)^\top$.

La matrice omogenea \mathbf{P} è una matrice 3×4 ed è detta *matrice di proiezione* della telecamera.

Assumendo che l'origine delle coordinate nel piano immagine coincida col punto principale, si può così derivare il mapping di proiezione centrale:

$$(x, y, z)^\top \mapsto \left(\frac{f \cdot x}{z}, \frac{f \cdot y}{z}\right)^\top$$

Nel caso in cui invece l'origine non coincida col punto principale, di coordinate $(p_x, p_y)^\top$, il mapping sarà:

$$(x, y, z)^\top \mapsto \left(\frac{f \cdot x}{z} + p_x, \frac{f \cdot y}{z} + p_y\right)^\top$$

In coordinate omogenee il mapping può essere così espresso:

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f \cdot x + z p_x \\ f \cdot y + z p_y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{X}_{cam}$$

Perciò:

$$\mathbf{x} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{X}_{cam}$$

dove \mathbf{K} è una matrice 3×3 detta *matrice di calibrazione* della telecamera e verrà approfondita in seguito.

\mathbf{X}_{cam} invece è il vettore omogeneo che rappresenta le coordinate del punto della scena: viene enfatizzato il fatto di assumere che la telecamera si trovi posizionata all'origine del sistema di riferimento, con l'asse principale puntato nella stessa direzione dell'asse z .

La figura 1.3 esplica la relazione tra il sistema di coordinate della telecamera e il sistema di coordinate del piano immagine.

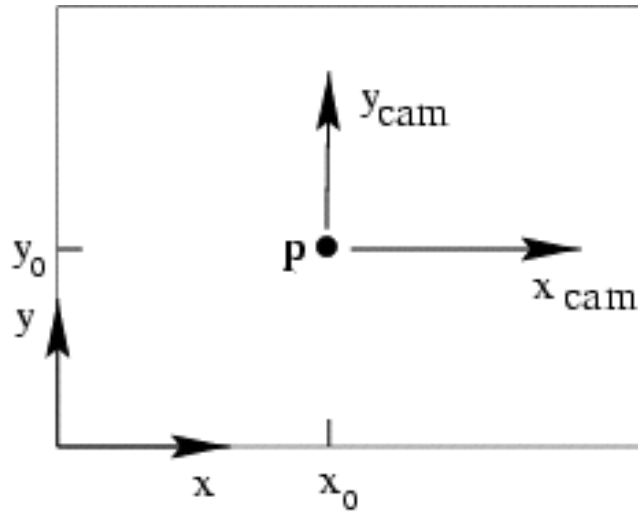


Figura 1.3: Relazione tra il sistema di coordinate dell'immagine (x, y) e quello della telecamera (x_{cam}, y_{cam})

1.1.3 Posizione e orientamento della telecamera

Come detto precedentemente, viene assunto che la telecamera si trovi all'origine del sistema di riferimento di coordinate euclidee, con l'asse principale puntato lungo la direzione dell'asse z . Tuttavia i punti \mathbb{R}^3 della scena possono essere espressi, in generale, con un differente sistema di riferimento mediante coordinate globali (indipendenti dalla posizione della telecamera). Questi due sistemi di riferimento, esemplificati in figura 1.4, sono riconducibili attraverso una rotazione e una traslazione.

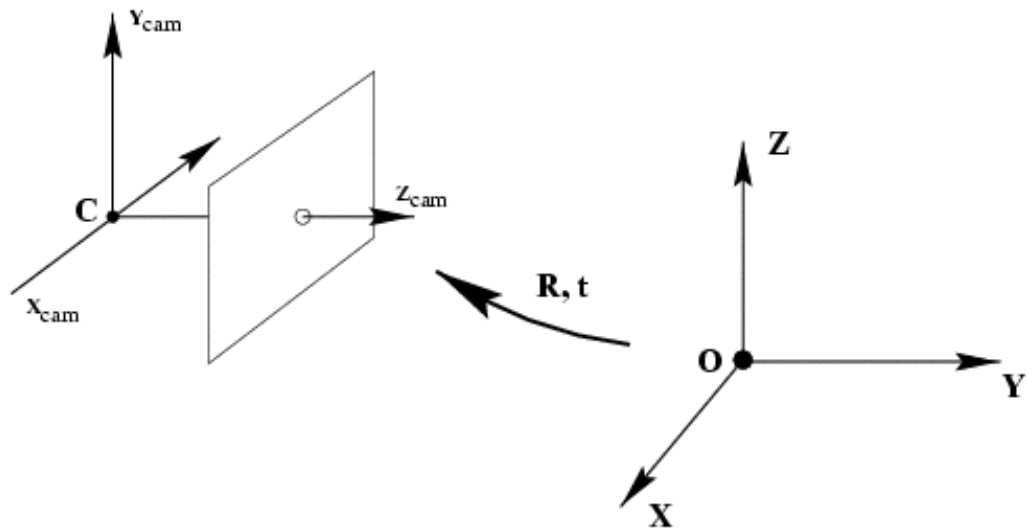


Figura 1.4: Trasformazione dal sistema di riferimento globale al sistema di riferimento della telecamera

Se $\tilde{\mathbf{X}}$ rappresenta le coordinate non omogenee globali di un punto e $\tilde{\mathbf{X}}_{cam}$ le rispettive coordinate di quel punto secondo il sistema di riferimento della telecamera, allora si può scrivere:

$$\tilde{\mathbf{X}}_{cam} = \mathbf{R}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$$

dove $\tilde{\mathbf{C}}$ rappresenta le coordinate della telecamera secondo il sistema di riferimento a coordinate globali ed $\tilde{\mathbf{R}}$ è una matrice 3×3 di rotazione che esprime l'orientamento del sistema di riferimento della telecamera.

In coordinate omogenee l'equazione diventa:

$$\tilde{\mathbf{X}}_{cam} = \begin{pmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{pmatrix} \mathbf{X}$$

dove \mathbf{X} viene espresso in coordinate globali.

In generale perciò, il mapping $\mathbf{x} \mapsto \mathbf{X}$ di una telecamera pinhole è definito da:

$$\mathbf{x} = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}] \cdot \mathbf{X}$$

con $\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}]$ che ha 9 gradi di libertà: 3 per \mathbf{K} , 3 per \mathbf{R} e 3 per $\tilde{\mathbf{C}}$.

Spesso conviene non esplicitare il centro della camera nella conversione da coordinate globali a coordinate della telecamera, perciò si preferisce rappresentarle come:

$$\tilde{\mathbf{X}}_{cam} = \mathbf{R}\tilde{\mathbf{X}} + \mathbf{t} \quad , \quad \text{con } \mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}}$$

In questo modo la matrice della telecamera diventa:

$$\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}] = \mathbf{K}[\mathbf{R} | \mathbf{t}]$$

1.1.4 Modello generale di telecamera proiettiva

Nel modello Pinhole, è stato assunto che le coordinate immagine adottassero la medesima scala per entrambi gli assi. Questo però non è sempre vero, come ad esempio nelle telecamere CCD.

Aggiungendo questa informazione nel modello, la matrice di calibrazione ha un grado di libertà in più e diventa:

$$\begin{aligned} \mathbf{K} &= \begin{pmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} f \cdot m_x & 0 & p_x \cdot m_x \\ 0 & f \cdot m_y & p_y \cdot m_y \\ 0 & 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Perciò in questo modo la matrice della telecamera risulta avere 10 gradi di libertà.

Per completare il modello, consideriamo il caso in cui gli assi del piano immagine non siano perpendicolari. Questo caso può essere considerato introducendo un ulteriore parametro s nella matrice di calibrazione, chiamato *deviazione*. Quando gli assi sono perpendicolari, $s = 0$.

Quindi, nel modello generale di telecamera proiettiva, la matrice di calibrazione vale:

$$\mathbf{K} = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

La matrice della telecamera viene portata in questo modo ad avere 11 gradi di libertà.

1.1.5 Distorsione radiale

Nel modello Pinhole si è assunto che la lente della telecamera fosse lineare, ossia che il punto della scena, il punto sull'immagine e il centro ottico fossero collineari. E che quindi le linee della scena venissero mantenute tali anche sul piano immagine.

In realtà, le lenti reali soffrono di un effetto di distorsione radiale, che diventa più evidente tanto più la lunghezza focale della lente diminuisce o è economica. L'effetto è rappresentato in figura 1.5.

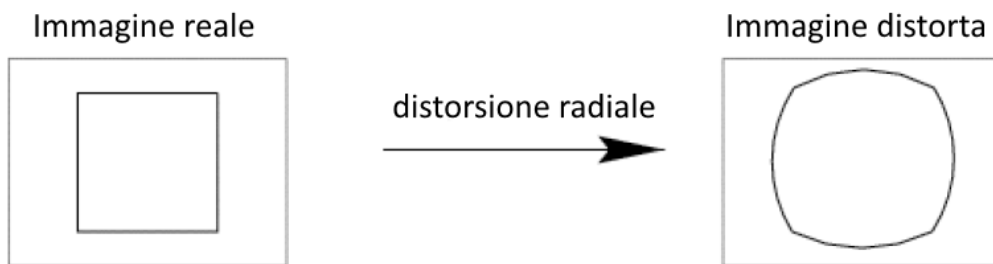


Figura 1.5: L'immagine reale, come sarebbe ottenuta attraverso una lente perfettamente lineare, viene distorta sotto l'effetto della distorsione radiale a cui sono soggette le lenti reali

Se poniamo che $(\tilde{x}, \tilde{y})^\top$ siano le coordinate immagine di un punto sotto una proiezione pinhole ideale (quindi senza distorsioni), allora abbiamo che:

$$(\tilde{x}, \tilde{y})^\top = \mathbf{K}[\mathbf{I}|\mathbf{0}] \cdot \mathbf{X}_{cam}$$

Perciò possiamo dire che il punto reale proiettato $(x_d, y_d)^\top$ sia correlato al punto ideale proiettato $(\tilde{x}, \tilde{y})^\top$ attraverso una relazione di distorsione di tipo radiale.

Questa distorsione può essere modellata in questo modo:

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = L(\tilde{r}) \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}$$

dove $L(\tilde{r})$ è il fattore di distorsione e \tilde{r} è la distanza radiale $\sqrt{\tilde{x}^2 + \tilde{y}^2}$.

Conoscendo $L(r)$, con $r = \sqrt{(x - x_c)^2 + (y - y_c)^2}$, e il centro di distorsione radiale $(x_c, y_c)^\top$, è possibile correggere la distorsione applicata al punto misurato $(x, y)^\top$ ed ottenere quindi le coordinate corrette del punto $(\hat{x}, \hat{y})^\top$.

$$\hat{x} = x_c + L(r)(x - x_c) = x_c + (x - x_c)(1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots)$$

$$\hat{y} = y_c + L(r)(y - y_c) = y_c + (y - y_c)(1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots)$$

I parametri $\{k_1, k_2, k_3, \dots, x_c, y_c\}$ vengono considerati parte della calibrazione della telecamera.

Per individuare questi parametri si può definire una funzione costo sulle linee proiettate sul piano immagine, che verrà iterativamente minimizzata in funzione dei parametri k_i e del centro della distorsione radiale.

1.2 Calibrazione dai punti di fuga

La calibrazione classica di una telecamera è un'operazione che sfrutta una conoscenza a priori della scena per computare la matrice \mathbf{P} . Avendo un certo numero di corrispondenze del tipo $\mathbf{x}_i \leftrightarrow \mathbf{X}_i$ (figura 1.6), dove \mathbf{X}_i è un punto della scena e \mathbf{x}_i è la sua immagine, è possibile determinare $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ tale che $\mathbf{x}_i = \mathbf{P} \cdot \mathbf{X}_i$.

Per questo motivo vengono impiegati degli oggetti di calibrazione (come

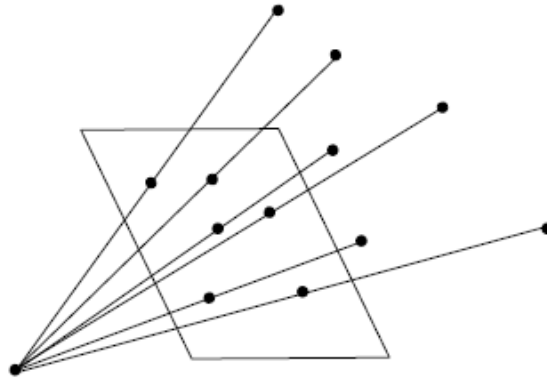


Figura 1.6: Corrispondenza tra punti della scena e rispettivi punti immagine. Conoscendo l'esatta posizione dei punti della scena e la loro immagine, è possibile risalire alla matrice di proiezione della telecamera

griglie colorate o altri supporti che possano essere misurati con precisione). Nel caso però che questi oggetti non siano utilizzabili, si può cercare di risalire alla posizione dei rispettivi punti nella scena sfruttando gli elementi prospettici presenti nell'immagine e facendo qualche assunzione.

1.2.1 Elementi di geometria proiettiva

La prospettiva è un tipo di proiezione che traspone una scena tridimensionale in un'immagine bidimensionale, cercando di riprodurre artificialmente il senso di profondità come viene percepita dall'occhio umano.

Per farlo, di primaria importanza è il punto di vista, che verrà assunto dal centro della telecamera. Inoltre questo tipo di proiezione si basa sul fatto che gli oggetti più vicini vengono percepiti più grandi rispetto a gli oggetti più distanti.

Dal punto di vista geometrico, questo significa che tutte le linee parallele convergono, o meglio fuggono, verso i medesimi punti, chiamati per l'appunto punti di fuga.

Se invece di considerare le linee parallele si considerano i piani paralleli si ottiene l'estensione del concetto dei punti di fuga, dato dalle linee di fuga.

Punti di fuga

I punti ideali nella scena si trovano all'infinito (infatti sono detti anche punti all'infinito) e hanno coordinate omogenee nella forma $(x, y, z, 0)^T$. Queste coordinate hanno solo due parametri liberi, poiché la scala di un vettore omogeneo è arbitraria.

L'immagine di un punto ideale, sotto un mapping proiettivo, è chiamata *punto di fuga*. Poiché due linee della scena parallele in uno spazio proiettivo si incontrano solo in un punto ideale (all'infinito), le immagini di queste linee convergeranno nel punto di fuga (figura 1.7). Ciò può essere esteso considerando anche più di due linee della scena parallele.

Geometricamente il punto di fuga si ottiene intersecando il piano immagine con un raggio passante per il centro della telecamera, parallelo alla linea della scena.

Chiaramente, a orientazioni diverse delle linee della scena corrisponderanno diversi punti di fuga. Inoltre, questi punti di fuga possono trovarsi anche all'infinito: nel caso in cui le linee della scena siano parallele al piano immagine, allora il relativo punto di fuga si troverà all'infinito.

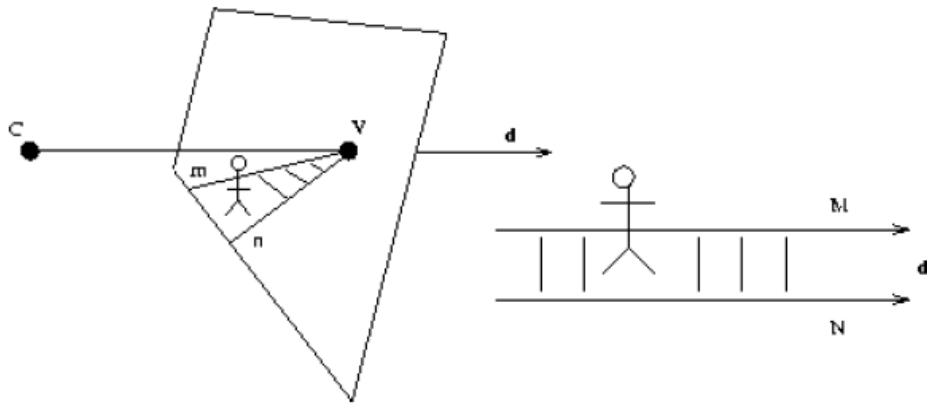


Figura 1.7: Formazione di un punto di fuga: due linee della scena parallele, quando proiettate convergono in un punto di fuga sul piano immagine

Un'altra considerazione che può essere fatta è che se consideriamo dei punti equidistanti su una linea della scena, la distanza delle loro immagini diminuirà tanto più i punti saranno distanti dal centro della telecamera, come si può vedere in figura 1.8. Portando l'esempio al limite, per $\mathbf{X} \rightarrow \infty$ il punto della scena verrà mappato nel punto $\mathbf{x} = \mathbf{v}$ lungo la linea immagine verticale della figura.

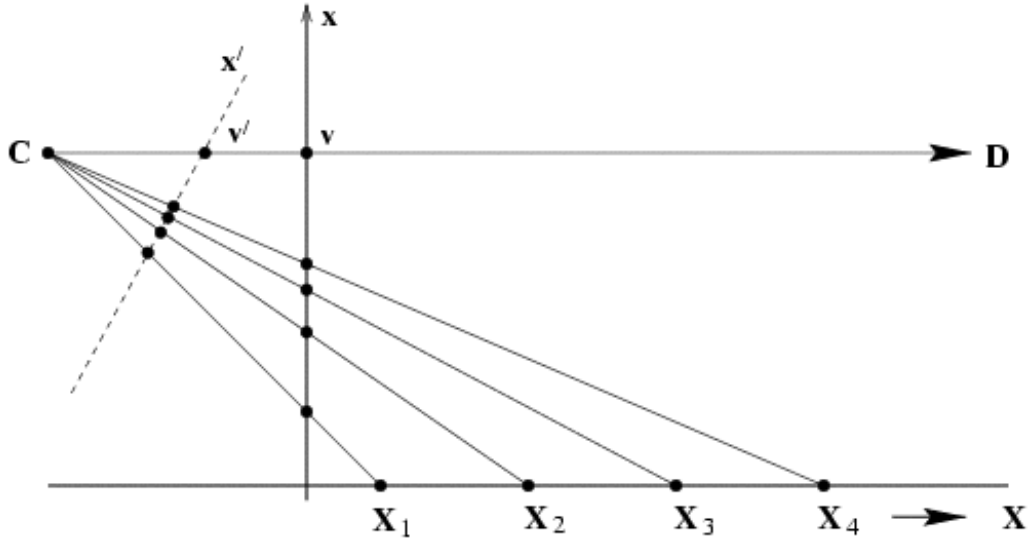


Figura 1.8: I punti \mathbf{X}_i sono equidistanti nella scena, ma la distanza delle loro immagini decresce sempre più. Al limite, con $\mathbf{X} \rightarrow \infty$, il punto della scena verrà mappato nel punto $\mathbf{x} = \mathbf{v}$

Si può dimostrare che, se $\mathbf{P} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ è la matrice di proiezione della telecamera, il punto di fuga individuato dalle linee della scena con direzione \mathbf{d} è il punto di intersezione \mathbf{v} , tra il piano immagine e un raggio avente direzione \mathbf{d} passante per il centro della telecamera (figura 1.9).

Questo punto di fuga è dato da:

$$\mathbf{v} = \mathbf{K} \cdot \mathbf{d}$$

Per dimostrarlo, consideriamo una linea \mathbf{L} della scena, che collega un punto \mathbf{A} e un punto all'infinito $\mathbf{D} = (\mathbf{d}^\top, 0)^\top$.

Se \mathbf{X} è un punto arbitrario della linea \mathbf{L} , allora \mathbf{X} può essere riscritto come:

$$\mathbf{X}(\lambda) = \mathbf{A} + \lambda \cdot \mathbf{D}$$

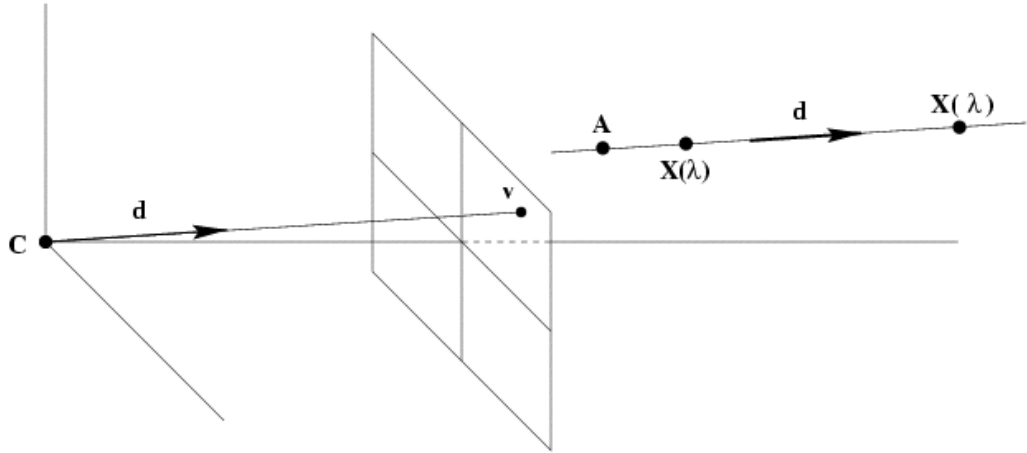


Figura 1.9: Il punto di fuga \mathbf{v} della linea con direzione \mathbf{d} è il punto di intersezione tra il piano immagine e il raggio parallelo a \mathbf{d} passante per \mathbf{C}

Mentre λ varia da 0 a ∞ , $\mathbf{X}(\lambda)$ varierà da \mathbf{A} a \mathbf{D} .

Ora consideriamo l'immagine del punto $\mathbf{X}(\lambda)$ pari a $\mathbf{x}(\lambda) = \mathbf{P} \cdot \mathbf{X}(\lambda)$. Possiamo scrivere che:

$$\mathbf{x}(\lambda) = \mathbf{P} \cdot \mathbf{X}(\lambda) = \mathbf{P} \cdot \mathbf{A} + \lambda \cdot \mathbf{P} \cdot \mathbf{D} = \mathbf{a} + \lambda \cdot \mathbf{K} \cdot \mathbf{d}$$

Se adesso poniamo $\lambda \rightarrow \infty$, otteniamo:

$$\mathbf{v} = \lim_{\lambda \rightarrow \infty} \mathbf{x}(\lambda) = \mathbf{K} \cdot \mathbf{d}$$

Di conseguenza il raggio che passa per il centro della telecamera e per il punto \mathbf{v} avrà direzione:

$$\mathbf{d} = \mathbf{K}^{-1} \cdot \mathbf{v}$$

Linee di fuga

L'insieme di tutti i punti ideali costituisce un piano, denominato *piano ideale*, di coordinate omogenee $(x, y, z, 0)^\top$. Una qualsiasi linea di questo piano è detta *linea ideale* e la sua immagine sottoposta a mapping proiettivo viene chiamata *linea di fuga* (figura 1.10).

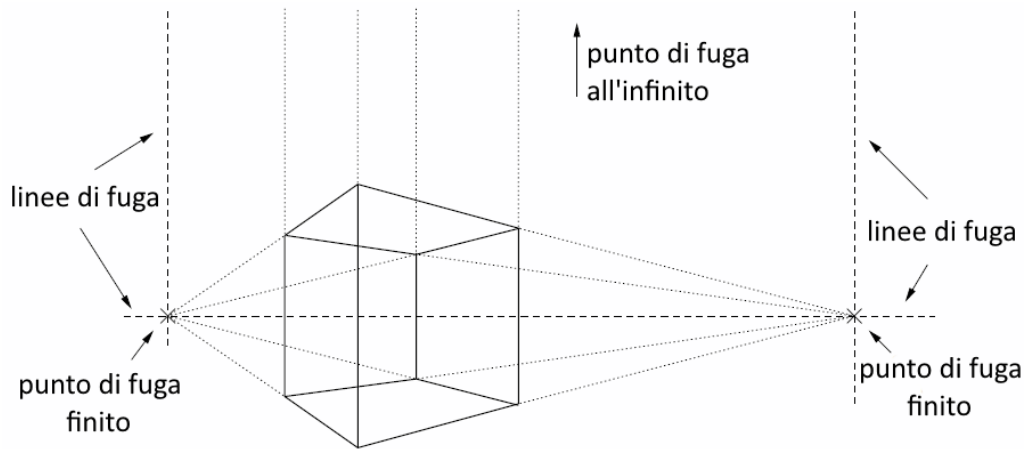


Figura 1.10: I tre punti di fuga e le linee di fuga di un cubo

Dato un insieme di linee parallele su un piano, il loro punto di fuga si troverà sulla linea di fuga del piano. Infatti, analogamente a quello che succede coi punti, se consideriamo delle linee parallele equidistanti su un piano della scena (figura 1.11), la distanza delle loro immagini diminuirà con l'aumentare della distanza dal centro della telecamera. Al limite, con i punti sul piano distanti all'infinito dalla telecamera, la loro immagine si troverà sulla linea di fuga.

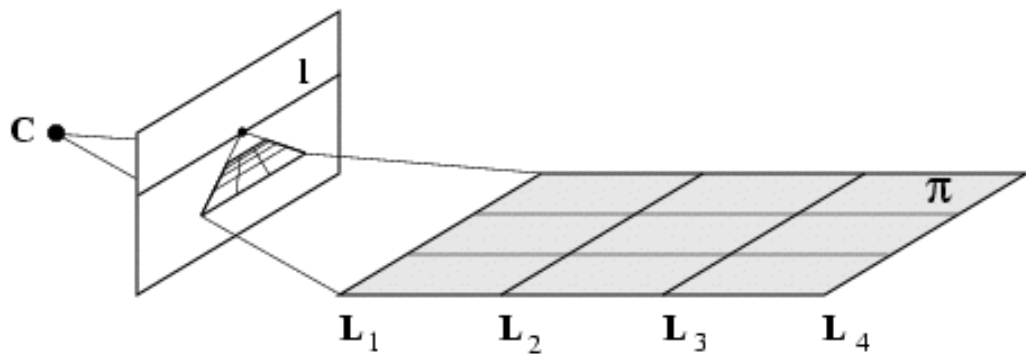


Figura 1.11: Le linee L_i sono equidistanti nella scena, ma la distanza delle loro immagini decresce sempre più. Al limite, con $L \rightarrow \infty$, la linea della scena verrà mappata nella linea di fuga l

Geometricamente una linea di fuga si costruisce intersecando il piano immagine con un piano parallelo alla scena passante per il centro ottico, come rappresentato in figura 1.12.

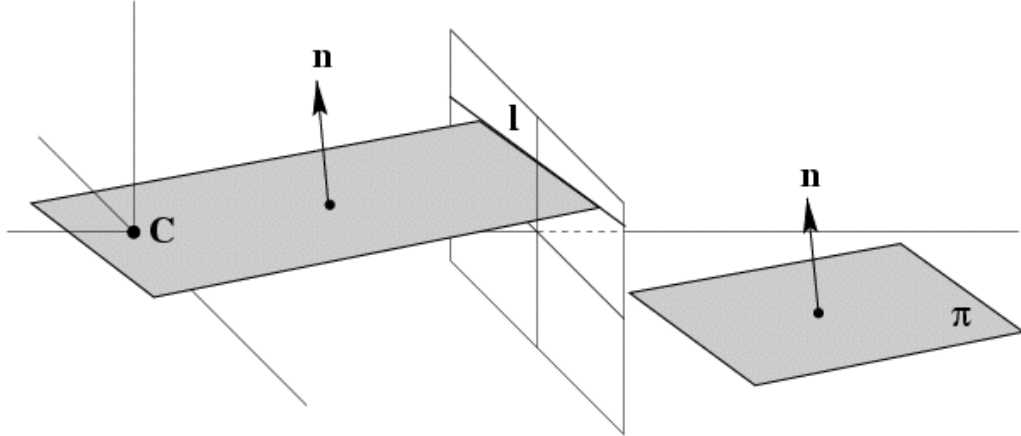


Figura 1.12: Formazione della linea di fuga: la linea di fuga l del piano π è ricavata intersecando il piano immagine con il piano parallelo a π passante per il centro della telecamera C

L'immagine di una linea l definisce un piano π passante per il centro della telecamera e avente una direzione normale $\mathbf{n} = \mathbf{K}^T \mathbf{l}$.

Lo si può dimostrare prendendo un punto \mathbf{x} sulla linea l e ponendo che \mathbf{n} sia la normale del piano π . La linea l verrà proiettata all'indietro nelle direzioni $\mathbf{d} = \mathbf{K}^{-1} \mathbf{x}$, che sono tutte ortogonali a \mathbf{n} . Perciò si ha che:

$$\mathbf{d}^T \mathbf{n} = 0 \Rightarrow \mathbf{x}^T \mathbf{K}^{-T} \mathbf{n} = 0$$

Poiché ogni punto \mathbf{x} di l soddisfa $\mathbf{x}^T \mathbf{l} = 0$, allora:

$$\begin{cases} \mathbf{x}^T \mathbf{l} = 0 \\ \mathbf{x}^T \mathbf{K}^{-T} \mathbf{n} = 0 \end{cases} \Rightarrow \mathbf{l} = \mathbf{K}^{-T} \mathbf{n} \Rightarrow \mathbf{n} = \mathbf{K}^T \mathbf{l}$$

Se \mathbf{n} è la normale di un piano π passante per il centro della telecamera C e l è la linea di intersezione tra il piano π e il piano immagine, allora l è la

linea di fuga dell'insieme dei piani perpendicolari a \mathbf{n} e dati da $\mathbf{l} = \mathbf{K}^{-\top} \cdot \mathbf{n}$. Infatti, poiché l'immagine di una linea \mathbf{l} definisce un piano π passante per il centro della telecamera e avente una direzione normale $\mathbf{n} = \mathbf{K}^{\top} \mathbf{l}$, allora possiamo scrivere:

$$\mathbf{n} = \mathbf{K}^{\top} \mathbf{l} \Rightarrow \mathbf{l} = \mathbf{K}^{-\top} \mathbf{n}$$

Se \mathbf{n} è la normale dell'insieme dei piani allora questi sono paralleli a π . Di conseguenza, \mathbf{l} è la linea di fuga dell'insieme dei piani normali a \mathbf{n} .

1.2.2 Relazione di ortogonalità tra punti di fuga

Un punto immagine proietta all'indietro (verso la scena) un raggio passante per il centro della camera. Se $\mathbf{D} = (x_d, y_d, z_d, 0)^{\top}$ è un punto all'infinito con $\mathbf{d} = (x_d, y_d, z_d)^{\top}$, allora possiamo scrivere che:

$$\mathbf{x} = \mathbf{P} \cdot \mathbf{D} = \mathbf{K} \cdot [\mathbf{I} | \mathbf{0}] \cdot \mathbf{D} = \mathbf{K} \cdot \mathbf{d}$$

Di conseguenza il raggio che passa per il punto immagine \mathbf{x} ha direzione:

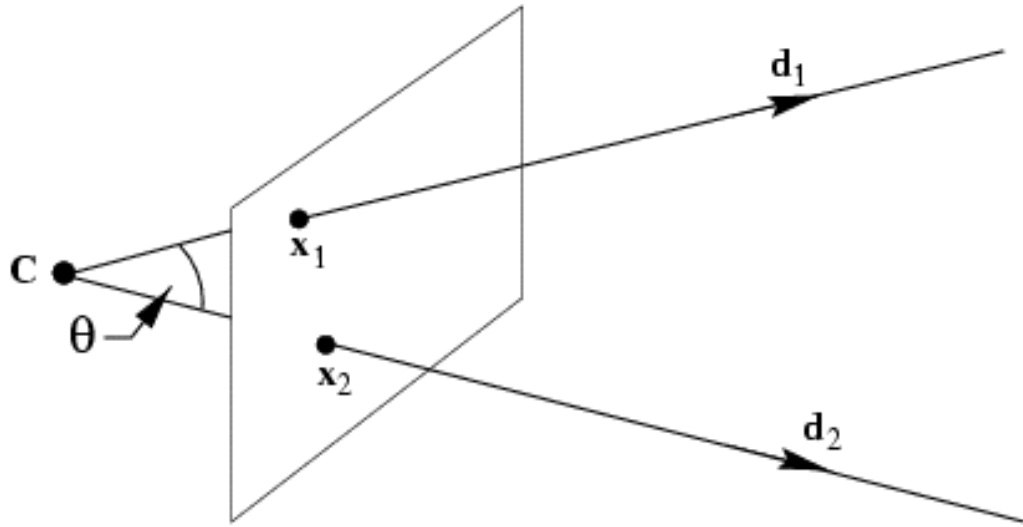
$$\mathbf{d} = \mathbf{K}^{-1} \cdot \mathbf{x}$$

A questo punto si può considerare l'angolo sotteso dal centro della telecamera e due raggi con direzione d_1 e d_2 , corrispondenti ai punti x_1 e x_2 , come in figura 1.13. Utilizzando la formula del coseno per l'angolo tra due vettori, con θ pari all'angolo da calcolare si ottiene:

$$\begin{aligned} \cos(\theta) &= \frac{\mathbf{d}_1^{\top} \mathbf{d}_2}{\sqrt{(\mathbf{d}_1^{\top} \mathbf{d}_1)(\mathbf{d}_2^{\top} \mathbf{d}_2)}} = \frac{(\mathbf{K}^{-1} \mathbf{x}_1)(\mathbf{K}^{-1} \mathbf{x}_2)}{\sqrt{[(\mathbf{K}^{-1} \mathbf{x}_1)^{\top} (\mathbf{K}^{-1} \mathbf{x}_1)] [(\mathbf{K}^{-1} \mathbf{x}_2)^{\top} (\mathbf{K}^{-1} \mathbf{x}_2)]}} \\ &= \frac{\mathbf{x}_1^{\top} (\mathbf{K}^{\top} \mathbf{K}^{-1}) \mathbf{x}_2}{\sqrt{[\mathbf{x}_1^{\top} (\mathbf{K}^{\top} \mathbf{K}^{-1}) \mathbf{x}_1] [\mathbf{x}_2^{\top} (\mathbf{K}^{\top} \mathbf{K}^{-1}) \mathbf{x}_2]}} \end{aligned}$$

Se la matrice \mathbf{K} è nota, allora lo è anche \mathbf{K}^{-1} . Ponendo $\omega = \mathbf{K}^{\top} \mathbf{K}^{-1}$, l'equazione diventa:

$$\cos(\theta) = \frac{\mathbf{x}_1^{\top} \omega \mathbf{x}_2}{\sqrt{(\mathbf{x}_1^{\top} \omega \mathbf{x}_1)(\mathbf{x}_2^{\top} \omega \mathbf{x}_2)}}$$

Figura 1.13: L'angolo θ tra i due raggi

Poiché il punto di fuga di una linea della scena proietta all'indietro un raggio parallelo a questa linea, si può misurare l'angolo tra due linee della scena partendo dai loro punti di fuga.

$$\cos(\theta) = \frac{\mathbf{v}_1^\top \omega \mathbf{v}_2}{\sqrt{(\mathbf{v}_1^\top \omega \mathbf{v}_1)(\mathbf{v}_2^\top \omega \mathbf{v}_2)}}$$

Da questa relazione si evince che se i punti di fuga sono ortogonali tra loro:

$$0 = \frac{\mathbf{v}_1^\top \omega \mathbf{v}_2}{\sqrt{(\mathbf{v}_1^\top \omega \mathbf{v}_1)(\mathbf{v}_2^\top \omega \mathbf{v}_2)}} \Rightarrow \mathbf{v}_1^\top \omega \mathbf{v}_2 = 0$$

Analogamente, con le linee di fuga è possibile misurare l'angolo tra due piani della scena. Con $\omega^* = \omega^{-1} = \mathbf{K} \cdot \mathbf{K}^\top$ si ha:

$$\cos(\theta) = \frac{\mathbf{l}_1^\top \omega^* \mathbf{l}_2}{\sqrt{(\mathbf{l}_1^\top \omega^* \mathbf{l}_1)(\mathbf{l}_2^\top \omega^* \mathbf{l}_2)}}$$

E di conseguenza, con linee di fuga ortogonali si ottiene:

$$0 = \frac{\mathbf{l}_1^\top \omega^* \mathbf{l}_2}{\sqrt{(\mathbf{l}_1^\top \omega^* \mathbf{l}_1)(\mathbf{l}_2^\top \omega^* \mathbf{l}_2)}} \Rightarrow \mathbf{l}_1^\top \omega^* \mathbf{l}_2 = 0$$

1.2.3 Calibrazione da tre punti di fuga ortogonali

Come è stato dimostrato, i punti di fuga \mathbf{v}_1 e \mathbf{v}_2 di due linee della scena, quando queste sono perpendicolari, soddisfano:

$$\mathbf{v}_1^\top \omega \mathbf{v}_2 = 0$$

dove $\omega = \mathbf{K}^\top \mathbf{K}^{-1}$ e \mathbf{K} è la matrice 3×3 dei parametri interni della telecamera.

Analogamente, le linee di fuga \mathbf{l}_1 e \mathbf{l}_2 di due piani della scena, quando questi sono perpendicolari, soddisfano:

$$\mathbf{l}_1^\top \omega^* \mathbf{l}_2 = 0$$

Se la telecamera ha deviazione zero ($s = 0$) e i pixel sono quadrati ($m_x = m_y$) allora ω può essere ricavata direttamente da una tripletta di punti di fuga con direzioni ortogonali.

Sappiamo che:

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

dove s è il coefficiente di deviazione tra asse x e asse y, mentre x_0 e y_0 sono le coordinate del punto principale. Invece $\alpha_x = f \cdot m_x$, $\alpha_y = f \cdot m_y$ rappresentano la lunghezza focale in termini di pixel, dove f è la lunghezza focale in termini di distanza, mentre m_x e m_y sono i fattori di scala dei pixel.

Generalmente le fotocamere moderne hanno pixel quadrati ($m = m_x = m_y$) e $s = 0$, quindi il modello può essere semplificato in:

$$\mathbf{K} = \begin{bmatrix} \alpha & 0 & x_0 \\ 0 & \alpha & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \alpha = f$$

A questo punto occorre ricavare il punto principale e la lunghezza focale.

Punto principale

Il punto principale può essere definito come l'ortocentro del triangolo formato dalla tripletta dei punti di fuga considerati, come indicato in Fig. 1.14.

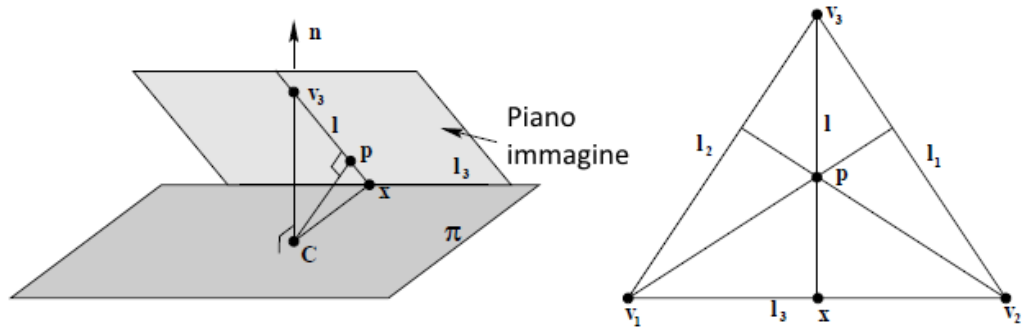


Figura 1.14: Costruzione geometrica del punto principale: il punto principale coincide con l'ortocentro di un triangolo formato dai tre punti di fuga

La linea di fuga l_3 viene proiettata all'indietro in un piano π con normale \mathbf{n} e il punto di fuga \mathbf{v}_3 viene proiettato all'indietro in una linea ortogonale al piano π . La normale \mathbf{n} del piano π attraverso il centro della camera \mathbf{C} e l'asse principale definiscono un piano che intercetta l'immagine nella linea $l = \mathbf{v}_3 \cdot \mathbf{x}$. La linea l_3 è l'intersezione tra il piano π e il piano immagine, mentre il punto \mathbf{v}_3 è l'intersezione tra la normale e il piano immagine. Il punto principale deve trovarsi ovviamente lungo l , inoltre l e l_3 sono perpendicolari al piano immagine.

La linea l corrisponde all'altezza di un triangolo, dove il punto di fuga \mathbf{v}_3 è uno dei vertici. Ripetendo l'operazione per gli altri due punti di fuga si ottengono altre due altezze che formano l'ortocentro del triangolo formato dai tre punti di fuga, e che determinano in modo univoco il punto principale.

Lunghezza focale

La lunghezza focale invece, riprendendo l'esempio precedente, può essere determinata come rappresentato in Fig. 1.15.

Considerando la stessa situazione, la lunghezza focale α è la distanza tra il centro della camera \mathbf{C} e il piano immagine.

Per la legge dei triangoli simili si ha che:

$$\alpha^2 = d(\mathbf{p}, \mathbf{v}_3)d(\mathbf{p}, \mathbf{x}) \Rightarrow \alpha = \sqrt{d(\mathbf{p}, \mathbf{v}_3)d(\mathbf{p}, \mathbf{x})}$$

con $d(\mathbf{u}, \mathbf{v})$ che corrisponde alla distanza tra i punti \mathbf{u} e \mathbf{v} .

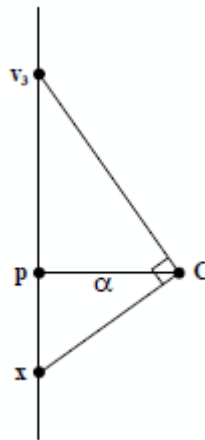


Figura 1.15: Costruzione geometrica della lunghezza focale: considerando il piano definito dal centro della telecamera, dal punto principale e da uno dei punti di fuga, i raggi da C al punto di fuga v_3 sono perpendicolari a gli altri. Quindi la lunghezza focale α è la distanza tra il centro della telecamera e il piano immagine.

A questo punto tutti i parametri interni della fotocamera sono determinati e la calibrazione può dirsi conclusa.

Capitolo 2

Autocalibrazione dai punti di fuga

Per risolvere il problema dell'autocalibrazione sono state avanzate negli anni diverse proposte. Nella quasi totalità dei casi che considerano l'autocalibrazione a singola vista, si ricorre all'impiego dei punti di fuga o delle linee di fuga ricavate dall'immagine.

Poiché tutte le tecniche sfruttano l'ortogonalità di questi punti, è necessario che nella scena siano dominanti linee parallele ortogonali tra loro. La garanzia di ortogonalità dei tre punti di fuga non può essere assicurata dall'analisi di una sola immagine della scena, ma spesso può essere assunta con una certa sicurezza quando si tratta di immagini di ambienti realizzati dall'uomo, che possono essere sia ambienti esterni (come edifici) sia di interni (come corridoi o stanze arredate).

L'approccio proposto prende spunto dai lavori che si sono dimostrati più efficaci, per proporre un'alternativa che ne permetta di superare le limitazioni. In particolare si è ricorso ad una clusterizzazione dei segmenti con la tecnica J-Linkage[7], impiegando la nuova misura di consistenza proposta[9] e proponendo un nuovo metodo per ricavare i punti di fuga da ogni cluster. Quindi vengono proposte alcune euristiche per stimare determinare la terna di punti di fuga ortogonali. Per calcolare la matrice di calibrazione viene seguito il metodo[3] descritto nel capitolo precedente.

La tecnica proposta, sintetizzata in figura 2.1, si compone essenzialmente di quattro fasi:

Estrazione dei segmenti

L'immagine viene analizzata e vengono estrapolati i segmenti fondamentali contenuti che la descrivono.

Clusterizzazione dei segmenti

I segmenti che concorrono allo stesso punto di fuga, vengono aggregati al medesimo cluster.

Stima di un punto di fuga da ogni cluster

Da ogni cluster vengono considerati i rispettivi segmenti e da questi viene ricavato il relativo punto di fuga.

Determinazione dei punti di fuga ortogonali

Dall'insieme dei punti di fuga ottenuti, viene ricavata una terna di punti di fuga ortogonali da cui poter determinare la matrice di calibrazione.

Ognuna di queste fasi è composta a sua volta di fasi intermedie che verranno approfondite nei paragrafi successivi. Per avere un raffronto immediato del contributo di ogni fase, verrà presentato un caso di studio con un'immagine (figura 2.2) che sarà analizzata passo per passo.

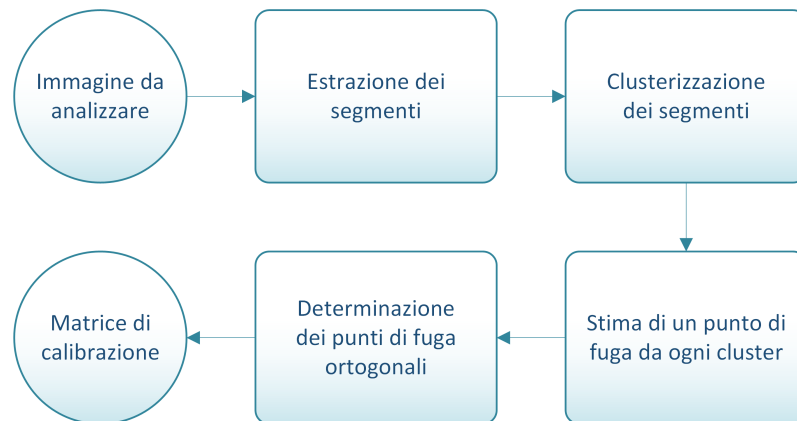


Figura 2.1: Le fasi dell'algorithm di autocalibrazione: da un'immagine da analizzare vengono ricavati i punti di fuga che poi verranno impiegati per determinare la matrice di calibrazione



Figura 2.2: L'immagine utilizzata per il caso di studio: le direzioni ortogonali sono semplici da individuare per la visione naturale, mentre per quella artificiale vi sono diverse difficoltà da affrontare

Si è cercato di progettare una metodologia modulare, per dotare l'approccio di maggior flessibilità e permettere eventualmente di sostituire all'occorrenza un modulo della metodologia con uno più efficace a seconda del contesto d'utilizzo o della precisione che si è disposti a sacrificare a fronte di una maggiore velocità d'esecuzione.

Quest'ultimo aspetto è di primaria importanza per l'impiego in contesti in tempo reale e perciò verranno effettuate considerazioni in merito a gli aspetti più critici che determinano la complessità computazionale dell'approccio progettato.

2.1 Estrazione dei segmenti da un'immagine

Il primo passo da compiere è l'individuazione ed estrazione dei segmenti, dall'immagine convertita in scala di grigi.

Per poter effettuare la segmentazione, è necessario iniziare con il rilevamento dei bordi dell'immagine. Quindi, una volta ottenuti i bordi fondamentali dell'immagine è possibile procedere con l'estrazione dei segmenti (figura 2.3).

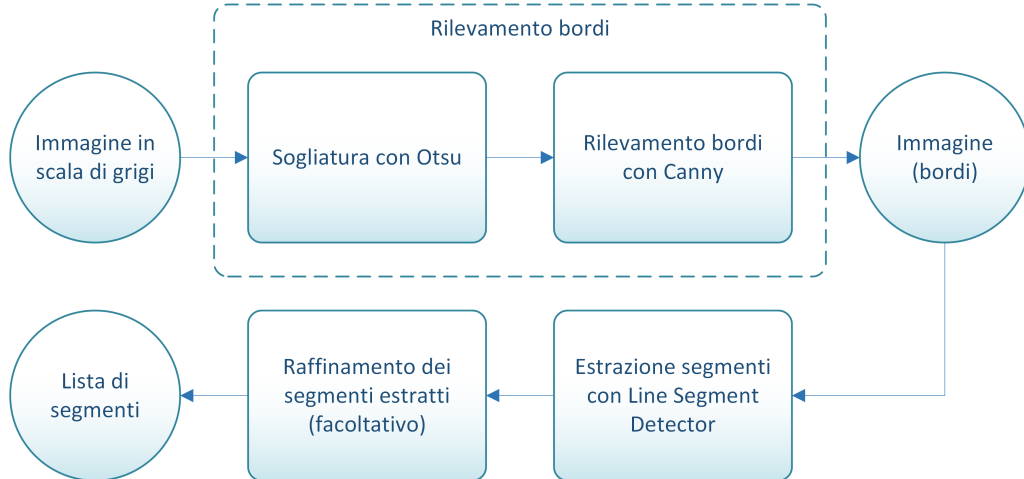


Figura 2.3: Estrazione dei segmenti: partendo da un'immagine in scala di grigi, viene ricavata un'immagine formata solo dai bordi dell'immagine precedente; infine i bordi vengono convertiti in segmenti ottenendo una lista che contiene tutti i segmenti dell'immagine

2.1.1 Rilevamento dei bordi

Per individuare i bordi dell'immagine, si è scelto di utilizzare un riconoscitore *Canny* in merito alla sua comprovata efficacia, testimoniata da un notevole numero di pubblicazioni che l'hanno reso uno strumento da considerare ormai consolidato. Poiché tale riconoscitore effettua una sogliatura per isteresi, richiede dei valori di sogliatura per meglio identificare ciò che è un bordo da ciò che non lo è. Quindi occorre determinare tali valori per permettere di distinguere gli oggetti di interesse da tutto ciò che invece non è interessante. Definendo la zona di certezza del bordo del riconoscitore Canny con il valore di soglia trovato dal sogliatore, il secondo valore di soglia (quello che definisce la zona di incertezza) viene assunto pari alla metà del valore trovato.

Per individuare il valore di soglia si possono impiegare diverse tecniche di sogliatura. Nel nostro caso è stata impiegata la tecnica Otsu, che permette di ottenere velocemente un valore di tale sogliatura. È bene ricordare però che Otsu funziona bene nel caso di immagini con distribuzione bimodale, ossia per immagini in cui gli oggetti di interesse spiccano nei confronti di quelli di non interesse.

Applicando il riconoscitore di bordi all'immagine del caso di studio, il risultato che si ottiene è quello in figura 2.4.

2.1.2 Estrazione dei segmenti

Ottenuta l'immagine coi bordi si può procedere con l'estrazione dei segmenti. In pratica si tratta di determinare il tipo di segmento che meglio riesce ad approssimare una linea che può essere anche curva. Per fare ciò le linee vengono trasformate nella somma di tanti segmenti più corti.

L'insieme di tutti i segmenti rilevati fornirà la lista di segmenti che rappresentano l'immagine processata e fornirà il punto di partenza per le fasi successive.

Per ricavare questi segmenti ci si è affidati ad un riconoscitore esterno[8], che si è dimostrato molto efficace. Il risultato è in figura 2.5.

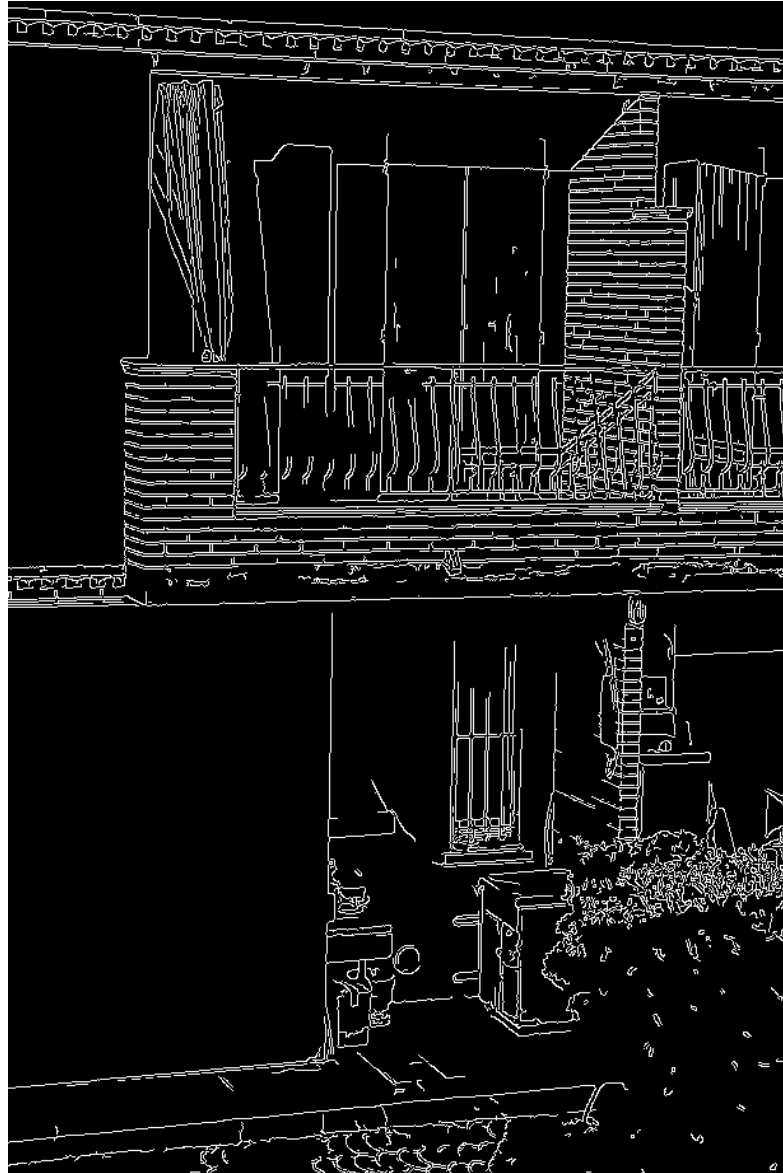


Figura 2.4: Rilevamento dei bordi: l'immagine è stata sottoposta al rilevamento dei bordi con Canny. Utilizzando la tecnica di sogliatura Otsu, i parametri usati per Canny risultano: $th_1 = 68$, $th_2 = 137$, $\sigma = 3$



Figura 2.5: Estrazione dei segmenti: i 500 segmenti estratti sono sovrapposti all'immagine originale, per meglio apprezzare la precisione con cui sono stati individuati. I segmenti individuati sono stati colorati per risaltare maggiormente

2.1.3 Raffinamento dei segmenti

Come passo ulteriore è possibile effettuare un raffinamento dei segmenti contenuti nel cluster, in quanto il riconoscitore non è perfetto e vi potrebbero essere segmenti sovrapposti o segmenti spezzati che dovrebbero essere uno solo.

Ciò aiuta la determinazione del punto di fuga, come si può vedere in figura 2.6, in quanto permette di eliminare potenziali punti di fuga che andrebbero a perturbare la stima del punto di fuga finale, peggiorandola. Infatti in questa fase quello che si fa è controllare dove confluiscono i segmenti del cluster, presi a coppie, e battezzare come punto di fuga quello che meglio approssima i punti individuati.

Un altro importante motivo è che tale raffinamento può permettere una diminuzione del numero di segmenti, per semplificare la risoluzione delle fasi successive.

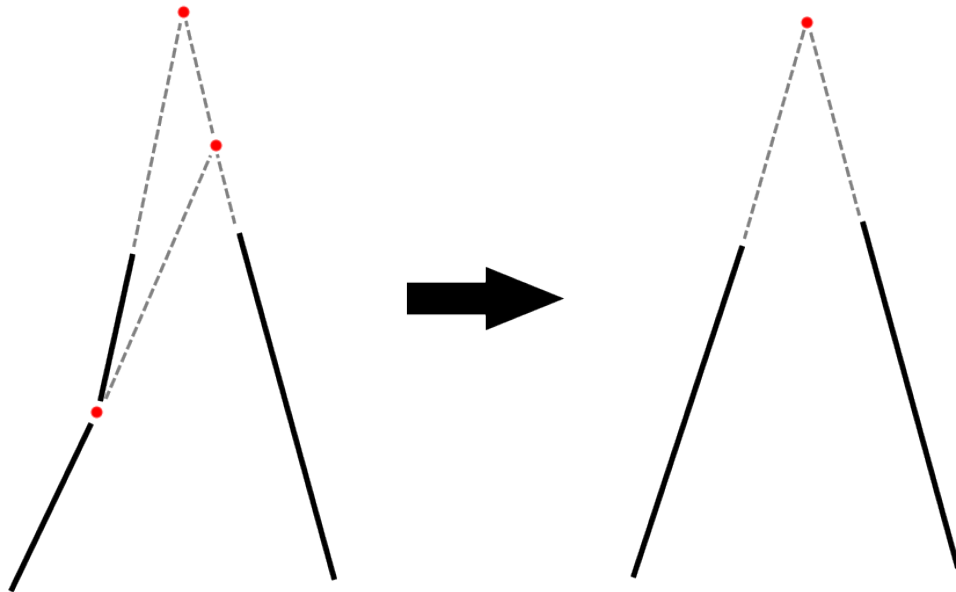


Figura 2.6: Vantaggi del raffinamento dei segmenti: diminuisce il numero di segmenti da gestire e si riduce il numero di punti di fuga potenziali

Euristiche considerate

Per fare ciò occorre determinare un qualche tipo di euristica che stabilisca quando due segmenti possano considerarsi sovrapponibili o collegabili, e in che modo.

Sostanzialmente vengono verificate le seguenti condizioni:

1. Se gli estremi di un segmento sono interni ad un altro segmento, allora il primo segmento viene rimosso.



2. Se due segmenti hanno una pendenza simile e un solo estremo di un segmento è contenuto all'interno del secondo segmento, allora entrambi i segmenti vengono rimossi e sostituiti da un nuovo segmento che gli approssimi entrambi.



3. Se due segmenti hanno una pendenza simile, con entrambi gli estremi esterni e si trovano sufficientemente vicini, allora entrambi i segmenti vengono rimossi e sostituiti da un nuovo segmento che gli approssimi entrambi.



2.1.4 Considerazioni sull'estrazione dei segmenti

L'estrazione dei segmenti è il primo passaggio dell'algoritmo e, come tale, influenzerà tutti i passaggi successivi. È bene perciò analizzare con attenzione i risultati ottenuti.

Come è stato anticipato nel capitolo precedente, con obiettivi di scarsa qualità o con lunghezze focali piuttosto corte è possibile notare il verificarsi di distorsioni ottiche, come ad esempio le distorsioni radiali. Osservando l'immagine in figura 2.7 è facilmente intuibile come queste possano compromettere il corretto riconoscimento dei segmenti.

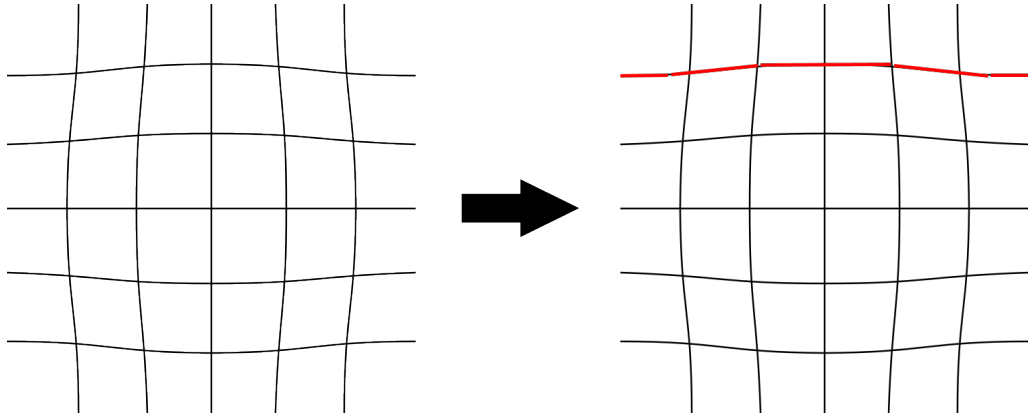


Figura 2.7: Effetto della distorsione radiale nel riconoscimento dei segmenti: le linee dritte vengono curvate e perciò il riconoscitore di segmenti estrarrà dall'immagine diversi segmenti inclinati invece di un unico segmento dritto

L'immagine mostra una griglia di linee verticali e orizzontali che, soggette alla distorsione, vengono deformate in linee curve. Questo porta un problema in fase di riconoscimento dei segmenti poiché queste linee, invece di essere riconosciute come un unico segmento, vengono identificate come più segmenti di diversa inclinazione. Di conseguenza alcuni segmenti che apparterebbero alla medesima linea, potrebbero venire associati a punti di fuga diversi.

Per limitare questo problema è possibile impiegare qualche tecnica di correzione delle distorsioni, oppure un meccanismo che, in base a qualche euristica, unisca i segmenti che probabilmente facevano parte della stessa linea, in modo da poter lavorare con una lista di segmenti più fedele a quella veritiera.

Questo problema non verrà affrontato e ci si baserà sull'assunzione di lavorare con immagini a limitata distorsione.

Un altro problema è dato dal caso di immagini troppo scure o particolarmente rumorose. In questo caso i bordi individuati dal riconoscitore saranno meno definiti e di conseguenza i segmenti estratti saranno meno affini alla realtà portando a prospettive diverse.

Infine, poiché il numero dei segmenti sarà determinante per la complessità delle fasi successive, in contesti di tempo reale può essere necessario scartare un certo numero di segmenti per semplificare il problema da affrontare. L'ideale sarebbe rimuovere piccoli segmenti o alcuni segmenti di un fascio numeroso, che non aggiungono particolare contributo, ma purtroppo il riconoscitore di segmenti utilizzato non permette di discriminare tra segmenti più importanti da quelli più incerti. Rimuovere segmenti casualmente aggiungerebbe un fattore randomico all'algoritmo, mentre troncando la lista si correrebbe il rischio di tagliare fuori intere aree di segmenti magari fondamentali, anche se l'operazione sarebbe meno onerosa.

Per limitare il rischio del verificarsi del secondo caso, si è deciso di rimuovere casualmente i segmenti eccedenti.

2.2 Clusterizzazione dei segmenti

Per clusterizzazione dei segmenti si intende una loro classificazione in base ad un qualche criterio, che assicuri che ad ogni classe corrispondano segmenti facenti riferimento al medesimo punto di fuga. Per operare questa clusterizzazione si è fatto affidamento alla tecnica J-Linkage[7] con la nuova misura di consistenza suggerita[9].

Per prima cosa è necessario identificare un insieme di punti di fuga potenziali sui quali procedere alla clusterizzazione dei segmenti. Dopodiché si procederà alla clusterizzazione vera e propria che produrrà un certo numero di cluster di segmenti.

Lo schema è rappresentato in figura 2.8.

2.2.1 Determinazione dei punti di fuga potenziali iniziali

La classificazione ha luogo nei punti di fuga e perciò prima di procedere è necessario determinare un primo insieme di punti di fuga ipotetici, che verranno impiegati nella prima fase di classificazione.

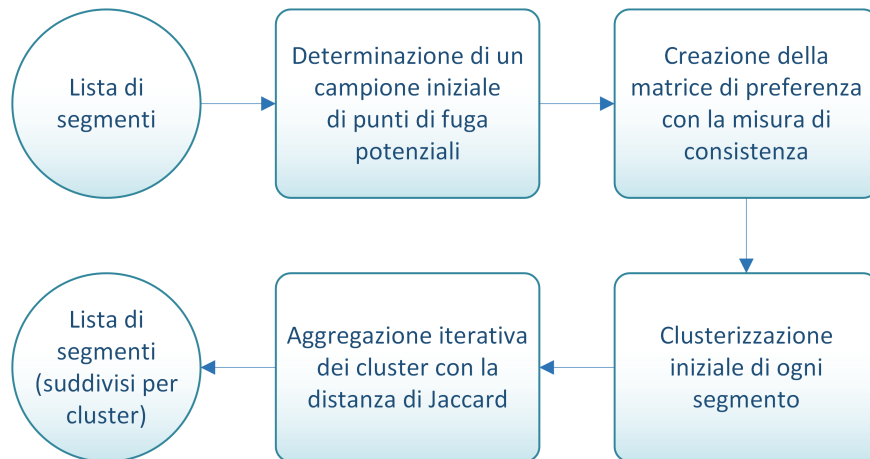


Figura 2.8: Clusterizzazione dei segmenti: la lista dei segmenti viene clusterizzata mediante un approccio iterativo e viene restituita la lista di segmenti suddivisa per cluster

L'idea è quella di partire da questi punti di fuga e poi scartarli (o meglio aggregarli) man mano, fino a rimanere solo con quelli maggiormente probabili. Poiché le linee convergono all'infinito nei punti di fuga, i veri punti di fuga si trovano in corrispondenza ad alcune intersezioni tra le rette corrispondenti ai segmenti rilevati.

Per questo motivo si può pensare di determinare i punti di fuga iniziali considerando tutte le intersezioni possibili tra le diverse coppie di segmenti.

2.2.2 Consistenza tra i punti di fuga potenziali e i segmenti

Dato un insieme di punti di fuga potenziali, viene stabilita una relazione di consistenza tra un segmento e il punto di fuga preso in considerazione. Questo porta alla creazione di una matrice booleana, chiamata *matrice di preferenza*, che esprime se un certo segmento è consistente con un certo punto di fuga o meno. Sostanzialmente viene confrontato ogni segmento per ogni punto di fuga potenziale e si tiene nota se questa combinazione è consistente o meno. Per farlo se ne misura il valore di consistenza e se tale valore è superiore ad una certa soglia allora la combinazione si considera consistente.

Scelta del modello di misura della consistenza

A questo punto è necessario definire un modo per stimare questa consistenza. Il senso è fornire una misura che quantifichi quanto un punto di fuga può essere compatibile con un segmento e questo può essere fatto valutando diversi aspetti, come la loro posizione, la dimensione e l'inclinazione del segmento e altri.

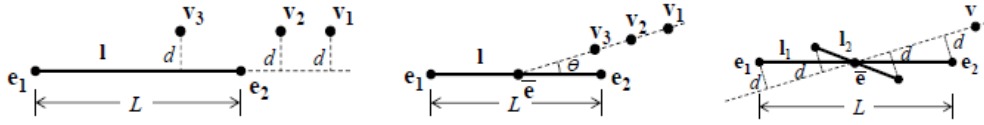


Figura 2.9: Esempi di misure di consistenza

In figura 2.9 sono rappresentate diverse tentativi possibili per valutare la consistenza.

Nel primo caso viene valutata la distanza ortogonale tra il punto di fuga e la retta corrispondente al segmento. Come si può notare, il problema di quest'approccio è che vengono valutati ugualmente diversi punti di fuga posti alla medesima distanza. Nell'esempio, il punto v_3 dovrebbe restituire una consistenza pari a zero, poiché all'infinito una linea si avvicina asintoticamente ad un punto di fuga, mentre la proiezione del punto v_3 sulla retta si trova proprio all'interno del segmento. Invece secondo questo modello, viene restituito lo stesso valore di consistenza di v_1 . E per lo stesso motivo anche v_2 dovrebbe restituire una consistenza minore di v_1 , mentre anche in questo caso gli viene attribuita la medesima consistenza.

Nel secondo caso invece viene misurato l'angolo tra la retta del segmento e quella formata dal punto di fuga e il punto medio del segmento. Anche in questo caso vale lo stesso ragionamento precedente: tutti i punti di fuga vengono valutati con la medesima consistenza, mentre v_1 dovrebbe essere più consistente di v_2 e a sua volta di v_3 .

Il terzo caso misura la distanza degli estremi del segmento, proiettata sulla retta data dal punto di fuga e il punto medio del segmento. Questo modello presenta il problema di non discriminare tra segmenti più lunghi e segmenti più corti: questi ultimi dovrebbero venire sfavoriti in quanto un segmento più corto è più soggetto a rumore e perciò dovrebbe restituire una consistenza meno affidabile di un segmento lungo.

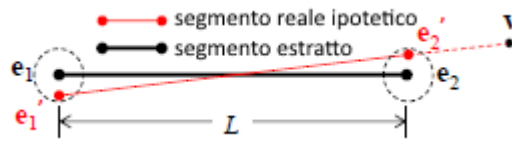


Figura 2.10: Modello di misura di consistenza usato

Per questi motivi si è preferito sfruttare un modello differente, rappresentato in figura 2.10, che adotta un approccio probabilistico.

Dato un segmento e un punto di fuga, viene valutato il segmento ipotetico che massimizzerebbe il valore di consistenza e quindi la distanza tra gli estremi del nuovo segmento e quelli originali, valutata come errore. In questo modo si valuta la possibilità che durante il riconoscimento dei segmenti, i loro estremi possano essere stati soggetti ad un errore di stima.

Se tra tutte le possibili orientazioni del segmento consideriamo solo l'insieme dei segmenti collineari al punto di fuga, la misura di consistenza può essere calcolata come l'integrale di tutte le linee di questo insieme. Ciò significa calcolare l'integrale del prodotto tra le due probabilità relative a gli estremi ipotizzati del nuovo segmento e le due ellissi di misura dell'errore (in due dimensioni).

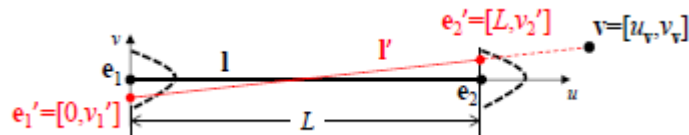


Figura 2.11: Modello di misura di consistenza approssimato

È possibile semplificare il modello in quello rappresentato in figura 2.11, considerando in questo modo solo una dimensione per la misurazione dell'errore. Ci si basa sul fatto che la relazione di collinearità dipende principalmente dalla distanza dagli estremi lungo la direzione ortogonale a quella del segmento. Inoltre la deviazione standard dell'errore è solitamente nell'ordine del sub-pixel e quindi significativamente minore della lunghezza del segmento, lungo la direzione tangente.

Per semplificare i calcoli il segmento viene allineato con gli assi dell'immagine e le coordinate del punto di fuga vengono modificate di conseguenza. Definendo così il punto di fuga come $\mathbf{v} = [u_{\mathbf{v}}, v_{\mathbf{v}}]$, se $u_{\mathbf{v}} > L$ e l' è collineare a \mathbf{v} , allora si ottiene $v'_2 = \frac{u_{\mathbf{v}} - L}{u_{\mathbf{v}}} v'_1 + \frac{v_{\mathbf{v}} L}{u_{\mathbf{v}}}$. Parametrizzando con $t \in \mathbb{R}$ si ottiene:

$$v'_1(t) = \frac{u_{\mathbf{v}} t}{\sqrt{u_{\mathbf{v}}^2 + (u_{\mathbf{v}} - L)^2}}$$

$$v'_2(t) = \frac{(u_{\mathbf{v}} - L)t}{\sqrt{u_{\mathbf{v}}^2 + (u_{\mathbf{v}} - L)^2}} + \frac{v_{\mathbf{v}} L}{u_{\mathbf{v}}}$$

Quindi la misura della consistenza vale:

$$\int_{-\infty}^{+\infty} f(v'_1(t); 0, \sigma^2) f(v'_2(t); 0, \sigma^2) dt = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{v_{\mathbf{v}}^2 L^2}{2\sigma^2(u_{\mathbf{v}}^2 + (u_{\mathbf{v}} - L)^2)}}$$

dove $f(\cdot; \mu, \sigma^2)$ è una funzione gaussiana di densità di probabilità.

Si può notare come questa formulazione incapsuli esplicitamente l'errore di misura σ e tenga conto della lunghezza L del segmento. Per decidere se un segmento è consistente ad un punto di fuga si valuta se il valore supera quello di una soglia $\eta = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}}$, che corrisponde misura della consistenza ad una deviazione standard dalla media.

2.2.3 Clusterizzazione iniziale e aggregazione iterativa

Una volta ottenuta la matrice di preferenza si può procedere con la clusterizzazione vera e propria. Quello che si fa è considerare tutti i segmenti e verificare se esistono tra di loro delle compatibilità, espresse come quantità di consistenze verso i medesimi punti di fuga. Questo significa riconoscere un certo punto di fuga come potenziale verso più segmenti.

La compatibilità viene espressa in termini di distanza di Jaccard, ossia una distanza che misura il grado di sovrapposizione di due insiemi. Per ogni segmento si controlla l'insieme dei punti di fuga consistenti e quindi si applica la distanza di Jaccard espressa come:

$$jd(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$



Figura 2.12: Clusterizzazione dei segmenti: i segmenti individuati sono stati colorati in modo diverso per evidenziare l'appartenenza a cluster diversi

Più i due insiemi si sovrappongono e più questa distanza è piccola, fino ad annullarsi quando i due insiemi coincidono. Quando invece i due insiemi sono completamente disgiunti questa distanza vale 1.

Per prima cosa perciò si parte costruendo un numero di cluster pari al numero di segmenti, ottenendo così tanti cluster da un elemento ciascuno. Per ognuno di questi cluster viene calcolata la distanza di jaccard nei confronti degli altri cluster, quindi questi cluster vengono poi aggregati man mano a due a due, partendo da quelli caratterizzati da una distanza di Jaccard più piccola, fino ad ottenere tutti cluster tra di loro disgiunti.

Fatto ciò si avrà un insieme di cluster, dove ognuno aggrega i segmenti che dovrebbero fare riferimento al medesimo punto di fuga e perciò ad ogni cluster dovrebbe corrispondere un punto di fuga diverso. Il risultato è rappresentato in figura 2.12. Chiaramente i cluster con un maggior numero di segmenti dovrebbero fornire una stima più affidabile, mentre in quelli più piccoli l'incertezza sarà maggiore. Per questo motivo si può decidere di classificare come outlier i cluster con troppi pochi segmenti e di rimuoverli.

2.2.4 Considerazioni sulla clusterizzazione dei segmenti

Per poter procedere alla creazione della matrice di preferenza è necessario partire con un insieme iniziale di punti di fuga, che viene determinato prendendo delle coppie di segmenti e valutandone l'intersezione delle due rette corrispondenti.

L'approccio più preciso per determinare i punti di fuga iniziali consiste nel considerare tutte le intersezioni possibili tra le diverse coppie di segmenti, ma purtroppo produce un numero di punti di fuga pari a:

$$numOfInitialVps = \binom{numOfSegments}{2}$$

In figura 2.13 si nota chiaramente come questo numero cresca molto rapidamente all'aumentare del numero di segmenti rilevati nell'immagine. Questo rende tale approccio inadeguato per un dispositivo dalle prestazioni limitate, come può essere un dispositivo mobile. Perciò si preferisce scegliere un numero arbitrario di punti di fuga iniziali, considerando coppie casuale di

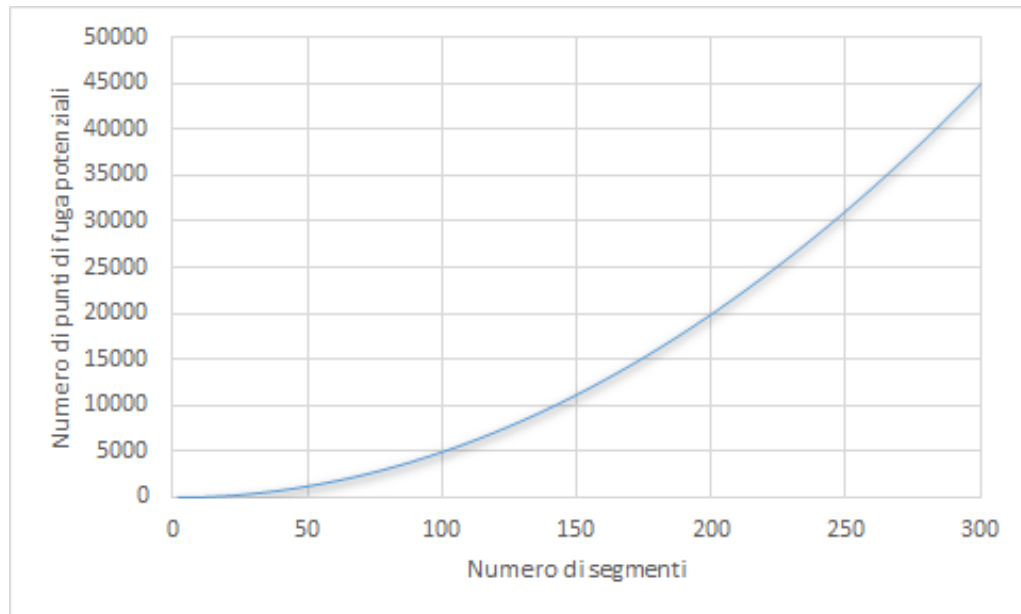


Figura 2.13: Relazione tra il numero di segmenti e punti di fuga potenziali: all'aumentare del numero di segmenti, il numero dei punti di fuga potenziali aumenta notevolmente

segmenti. Questo può portare a ottenere una moltitudine di punti di fuga coincidenti, che verranno poi aggregati in fase di clusterizzazione.

Il numero di punti di fuga iniziali deve essere scelto coerentemente, come compromesso tra precisione e prestazioni: un numero più elevato fornirà probabilità maggiori di ottenere un punto di fuga iniziale vicino a uno di quelli reali, ma questo significa anche moltiplicare il carico di lavoro da sostenere in fase di clusterizzazione.

Il problema di questa soluzione è l'introduzione di una variabilità casuale dell'efficacia dell'algoritmo. Questo perché la determinazione dei punti di fuga si basa sulla clusterizzazione, che a sua volta si basa sull'insieme dei punti di fuga iniziali. I passi successivi possono essere soltanto delle raffinazioni dei punti trovati, ma se la clusterizzazione è incompleta risulta complicato in un secondo momento risolvere il problema.

La selezione casuale con distribuzione uniforme, fornisce sì valori teoricamente sparsi uniformemente nello spazio delle possibilità, ma non assicura

che tra questi valori vi siano anche quelli più corretti che corrisponderebbero ai veri punti di fuga. Per questo motivo, utilizzando questo approccio, tra due esecuzioni si possono ottenere risultati anche molto diversi tra loro.

Un altro aspetto degno di nota è il processo di clusterizzazione. Questo porta inizialmente a costruire cluster che contengono un singolo segmento e poi ad accorparne tra loro due di volta in volta, in funzione della distanza di Jaccard calcolata in base alla matrice di preferenza.

Poiché ad ogni accorpamento è necessario ricalcolare la distanza di Jaccard tra il cluster accorpato e ogni altro cluster, la conseguenza è che nel caso peggiore (ossia quando tutti i cluster vengono aggregati in un unico cluster) sia necessario, con n pari al numero di segmenti, calcolare le distanze tra cluster un numero di volte pari a:

$$\text{cl}(n) = \underbrace{\binom{n}{2}}_{\text{inizializzazione}} + \underbrace{\binom{n-1}{2}}_{\text{aggregazione iterativa}}$$

Come si può vedere, si tratta di un valore che dipende fortemente dal numero di segmenti considerati. Per limitare questo problema si può decidere eliminare i segmenti meno significativi dalla lista di segmenti restituita dal riconoscitore di segmenti, nella fase precedente.

2.3 Stima dei punti di fuga da ogni cluster

Per ogni cluster deve essere determinato un unico punto di fuga e ciò può essere fatto analizzando i segmenti che contiene. L'idea è che ognuno di essi contribuisca a definire un unico punto di fuga di sintesi.

Considerando ogni coppia di segmenti del cluster, viene calcolata l'intersezione delle rette corrispondenti. Si ottengono così i punti di fuga potenziali del cluster, che serviranno per dedurre quello finale. Per farlo ci sono diversi modi: il più semplice è una media delle intersezioni trovate, ma vi sono sistemi più sofisticati che sono stati indagati e verranno presentati.

Lo schema è molto semplice ed è in figura 2.14, mentre il risultato di questa fase lo si può osservare in figura 2.15.

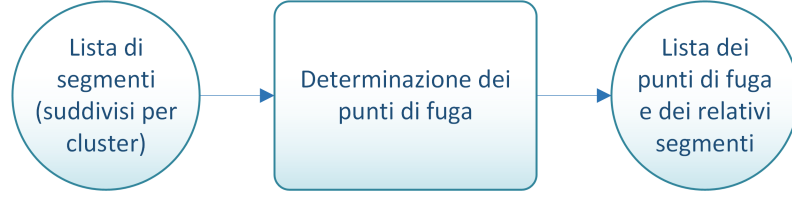


Figura 2.14: Determinazione dei punti di fuga

2.3.1 Stima per minimizzazione dell'errore

Questa tecnica si basa sull'approccio suggerito da Xu, Oh e Hoogs[9].

Dati due segmenti del cluster \mathbf{l}_j e \mathbf{l}_k e la loro intersezione $\mathbf{v}_{(j,k)}$, il punto d'intersezione trovato è tanto più affidabile quanto i due segmenti sono ortogonali tra loro. Ciò si può capire bene dalla figura 2.16.

Questo è dovuto al fatto che i segmenti riconosciuti possono essere affetti da errore e perciò è logico pensare che in realtà la loro posizione potesse essere leggermente diversa. Questo ovviamente si può ripercuotere sulle intersezioni trovate, in maniera diversa a seconda dell'inclinazione dei segmenti: in una coppia di segmenti con inclinazione simile è sufficiente un minimo errore per determinare una variazione consistente del punto di intersezione, mentre più i segmenti sono ortogonali tra loro e più questa variazione diventa contenuta a fronte del medesimo errore.

Ciò porta a pensare ad un modello di media pesata, che per la stima del punto di fuga impieghi dei pesi in base all'affidabilità di ogni singola intersezione.

Per farlo viene calcolata la matrice jacobiana \mathbf{G} della funzione $g(\cdot)$, usata per ottenere l'intersezione relativa ai segmenti \mathbf{l}_j e \mathbf{l}_k , partendo dai loro estremi.

$$\begin{aligned} \mathbf{v}_{(j,k)} &= \begin{bmatrix} u_{(j,k)} \\ v_{(j,k)} \end{bmatrix}^T \triangleq g(u_{j1}, v_{j1}, u_{j2}, v_{j2}, u_{k1}, v_{k1}, u_{k2}, v_{k2}) \\ &= \begin{bmatrix} \frac{(u_{k1}-u_{k2})(u_{j1}v_{j2}-u_{j2}v_{j1})-(u_{j1}-u_{j2})(u_{k1}v_{k2}-u_{k2}v_{k1})}{(u_{j1}-u_{j2})(v_{k1}-v_{k2})-(u_{k1}-u_{k2})(v_{j1}-v_{j2})} \\ \frac{(v_{k1}-v_{k2})(u_{j1}v_{j2}-u_{j2}v_{j1})-(v_{j1}-v_{j2})(u_{k1}v_{k2}-u_{k2}v_{k1})}{(u_{j1}-u_{j2})(v_{k1}-v_{k2})-(u_{k1}-u_{k2})(v_{j1}-v_{j2})} \end{bmatrix}^T \end{aligned}$$

In questo modo è possibile definire la matrice di covarianza del punto $\mathbf{v}_{(j,k)}$:

$$\Sigma^{\mathbf{v}_{(j,k)}} = \mathbf{G}_{(j,k)} \Sigma_{(j,k)} \mathbf{G}_{(j,k)}^T = \mathbf{G}_{(j,k)} \sigma^2 \mathbf{I}_8 \mathbf{G}_{(j,k)}^T = \sigma^2 \mathbf{G}_{(j,k)} \mathbf{G}_{(j,k)}^T$$

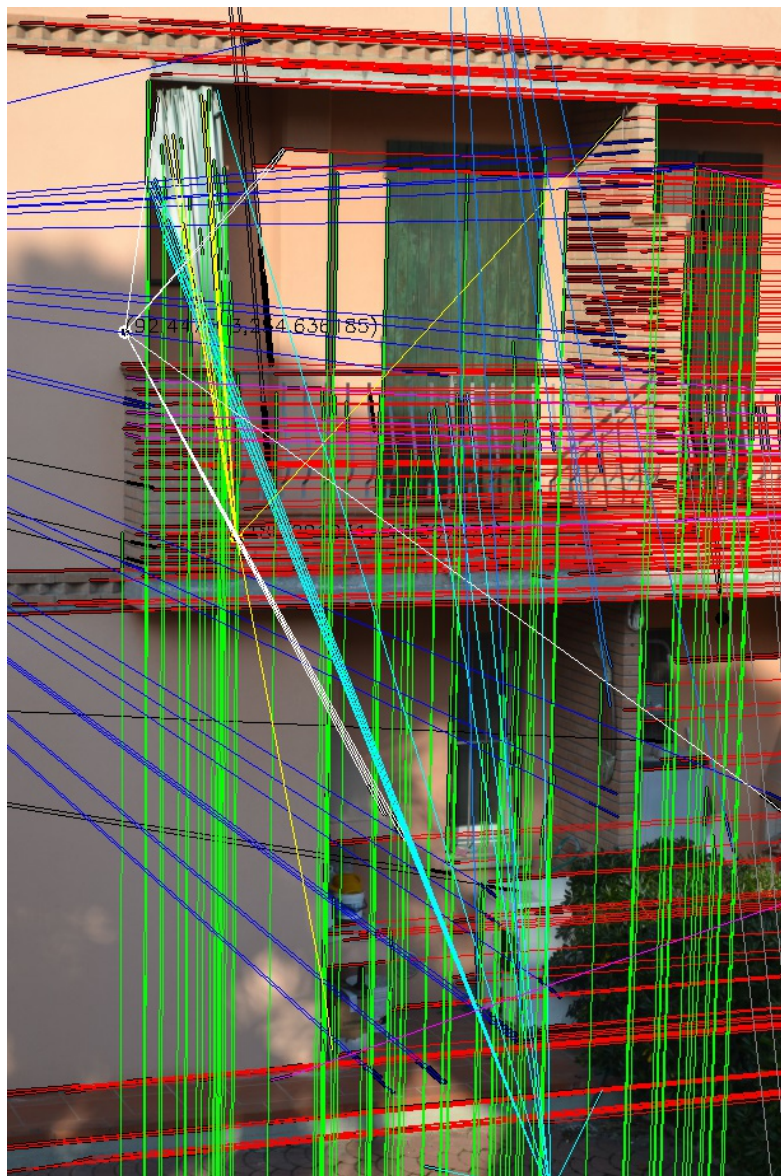


Figura 2.15: Individuazione dei punti di fuga: vengono individuati per ogni cluster il rispettivo punto di fuga. Da ogni segmento viene tracciata la linea di fuga che lo congiunge al rispettivo punto di fuga

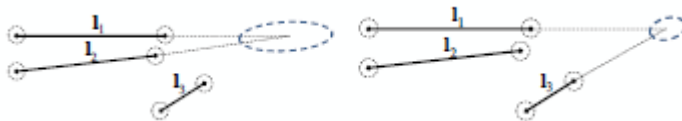


Figura 2.16: Errore dell'intersezione dei segmenti

La covarianza del punto di fuga però è meno significativa nello spazio immagine, a causa della distorsione prospettica della fotocamera. Infatti una grande covarianza di un punto di fuga all'infinito non implica necessariamente una stima inaccurata. Ciò che conta invece è la direzione di fuga, che può essere rappresentata come un punto nella sfera di Gauss. Il centro di questa sfera coincide col centro della fotocamera e i suoi due assi di rotazione (tilt e pan) sono paralleli a quelli del piano immagine (u e v). Rappresentando $\mathbf{v}_{(j,k)}$ sulla superficie della sfera di Gauss come $\tilde{\mathbf{v}} = [\theta, \phi]$, tramite coppia di rotazione pan e tilt, si ha:

$$\tilde{\mathbf{v}}_{(j,k)} = [\theta_{(j,k)}, \phi_{(j,k)}] \triangleq h(\mathbf{v}_{(j,k)}) = \left[\tan^{-1} \left(\frac{u_{(j,k)}}{f} \right), -\tan^{-1} \left(\frac{v_{(j,k)}}{\sqrt{u_{(j,k)}^2 + f^2}} \right) \right]$$

dove f è la lunghezza focale, che non essendo calibrata deve essere stimata. La maggior parte delle fotocamere hanno un angolo di campo che va dai $15\hat{\text{A}}^\circ$ ai $60\hat{\text{A}}^\circ$, perciò si può stimare che $f \in [0.28W, 3.80W]$, dove W è la dimensione del lato più lungo dell'immagine. In base a ciò, viene stimata una lunghezza focale pari a $f = 2W$.

Questo tipo di approssimazione è inaccettabile con approcci che si basano su una calibrazione accurata, ma in questo caso la soluzione finale verrà bilanciata dalla competizione delle altre soluzioni minime che vengono stimate con la medesima lunghezza focale.

Definendo $\rho_{(j,k)} = \sqrt{u_{(j,k)}^2 + f^2}$ e la matrice jacobiana di $h(\cdot)$ come:

$$\mathbf{H}_{(j,k)} = \begin{bmatrix} \frac{f}{\rho_{(j,k)}^2} & 0 \\ \frac{uv}{(\rho_{(j,k)}^2 + v_{(j,k)}^2)\rho_{(j,k)}} & \frac{-\rho_{(j,k)}}{\rho_{(j,k)}^2 + v_{(j,k)}^2} \end{bmatrix}$$

Il giacobiano dai vertici di $\mathbf{l}_j, \mathbf{l}_k$ ai rispettivi angoli di pan e tilt vale:

$$\mathbf{J}_{(j,k)} = \mathbf{H}_{(j,k)} \mathbf{G}_{(j,k)}$$

perciò la covarianza di $\tilde{\mathbf{v}}_{(j,k)}$ vale:

$$\Sigma^{\tilde{\mathbf{v}}_{(j,k)}} = \mathbf{J}_{(j,k)} \Sigma_{(j,k)}, \mathbf{J}_{(j,k)}^T = \sigma^2 \mathbf{J}_{(j,k)} \mathbf{J}_{(j,k)}^T$$

In questo modo abbiamo un metodo per discriminare, come in figura 2.16, le coppie di segmenti più influenti rispetto a quelle dove l'incertezza è maggiore.

Il passo successivo è definire ogni peso che andrà a contribuire al calcolo del punto di fuga finale. Dato un vettore di pesi $\mathbf{w} = [w_{(1,2)}, w_{(1,3)}, \dots, w_{(n-1,n)}]^T$, tale che $\sum_i \mathbf{w}_i = 1$ e $\forall \mathbf{w}_i \geq 0$, il punto di fuga finale è dato da:

$$\tilde{\mathbf{v}} = \sum_{j=1}^{n-1} \sum_{k=j+1}^n w_{(j,k)} \tilde{\mathbf{v}}_{(j,k)}$$

Lo scopo è trovare il \mathbf{w} che minimizza la covarianza di $\tilde{\mathbf{v}}$ e questo significa minimizzare la traccia di $\Sigma^{\tilde{\mathbf{v}}}$:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \text{tr}(\Sigma^{\tilde{\mathbf{v}}}) = \arg \min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w}$$

in cui \mathbf{A} è una matrice $n(n-1)/2 \times n(n-1)/2$ dove gli elementi della sua diagonale $\mathbf{A}(a, a) = \text{tr}(\Sigma^{\tilde{\mathbf{v}}})$ e a è l'indice di \mathbf{w} del peso corrispondente alla coppia dei segmenti $\mathbf{l}_j, \mathbf{l}_k$. Gli elementi esterni alla diagonale invece modellano la correlazione tra due diverse coppie di segmenti.

Poiché gli elementi esterni alla diagonale sono molto piccoli, si può semplificare il calcolo assumendo che le coppie di segmenti siano indipendenti e non abbiano alcuna correlazione. Approssimando \mathbf{A} ad una matrice diagonale si ottiene:

$$w_{(j,k)}^* = w_a^* = \frac{(\mathbf{A}(a, a))^{-1}}{\sum_{a=1}^{(n-1)n/2} (\mathbf{A}(a, a))^{-1}}$$

Attraverso \mathbf{w}^* è possibile calcolare il punto di fuga finale nella sfera di Gauss, come $\tilde{\mathbf{v}}^* = [\theta^*, \phi^*] = \sum_{j=1}^{n-1} \sum_{k=j+1}^n w_{(j,k)}^* \tilde{\mathbf{v}}_{(j,k)}$

Per individuare il punto di fuga finale in coordinate immagine basta utilizzare la formula inversa:

$$\mathbf{v}^* = [u^*, v^*] = \left[f \tan(\theta^*), -f \tan(\phi^*) \sqrt{\tan^2(\theta^*) + 1} \right]$$

2.3.2 Stima tramite mediana

La stima tramite minimizzazione dell'errore è molto potente in teoria, ma all'atto pratico si scontra con un problema: il riconoscimento dei segmenti è soggetto a rumore che può portare alla determinazione di segmenti con un'inclinazione o una posizione errata. Ciò viene mitigato dal merging dei segmenti, ma in certe situazioni si manifesta comunque chiaramente.

Quello che succede è ben rappresentato dalla figura 2.17.

Le intersezioni calcolate sono rappresentate dai punti rossi e il punto di fuga finale dovrebbe essere il punto verde. Tuttavia, se la misura è poco resistente a gli *outlier*, si verifica ciò che si vede: il punto rosso in basso condiziona tutto il calcolo e il punto di fuga individuato risulta essere quello blu. Eliminando quel punto rosso dai calcoli, il punto di fuga ottenuto diventa quello viola.

Da ciò si nota che se il tipo di stima che viene effettuato non tiene conto degli *outlier*, si ha una misurazione fortemente suscettibile a perturbazioni indotte. La stima tramite minimizzazione dell'errore cerca di rimediare mediante un sistema di pesi che può privilegiare o sfavorire certi punti di fuga dal calcolo complessivo, ma non contiene nella sua definizione la nozione di intersezioni errate. Questo significa che ogni intersezione contribuisce al calcolo e perciò è stato adottato un metodo di stima del punto di fuga alternativo, resistente a gli *outlier*.

Il metodo è molto semplice e si basa sulla resistenza del 50% dello strumento statistico *mediana* a gli *outlier*. Questo significa che il risultato tramite la *mediana* non viene influenzato dagli *outlier* a meno che essi non siano presenti all'interno del campione in quantità maggiore del 50%. L'idea è che generalmente per ogni cluster si ottengano molti punti di fuga potenziali concentrati attorno al vero punto di fuga, oltre ad alcuni punti sparsi (gli *outlier*).

Dato un insieme di punti di fuga potenziali, ottenuti dalle intersezioni dei segmenti di un cluster, quello si fa è prendere separatamente l'insieme dei valori di ciascuna coordinata, ordinarli in modo crescente. Poiché la maggiorparte dei valori è vicina al punto corretto, si otterranno due vettori dove gli *outlier* andranno a posizionarsi o all'inizio o alla fine, per cui nelle posizioni centrali invece la possibilità di trovare un valore prossimo a quello

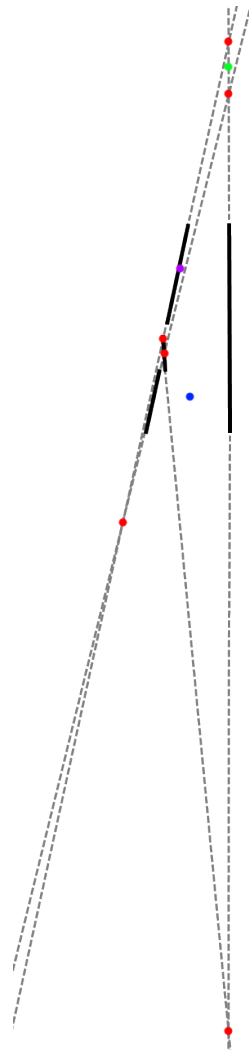


Figura 2.17: Perturbazione causata dai segmenti mal riconosciuti: la linea a sinistra, riconosciuta come tre segmenti porta il numero di punti di fuga potenziali (in rosso) da uno a sei. In verde vi è il punto di fuga corretto. La media dei punti di fuga potenziali restituisce il punto blu, che è molto distante dal risultato corretto. Eliminando il punto rosso più in basso si ottiene il punto viola, ancora molto distante. Questo perchè la media è un'operazione poco resistente a gli outlier

corretto sarà molto più alta. Perciò si procede scegliendo da entrambi i vettori i valori in posizione centrale, andando così ad ottenere il valore mediano di entrambe le coordinate e ottenendo così il valore stimato del punto di fuga del cluster.

Il difetto di quest'approccio è che, per ogni coordinata, il valore scelto appartiene ad una singola osservazione, mentre sarebbe logico pensare che tutte le osservazioni circostanti dovrebbero in qualche modo competere per determinare il valore più corretto, tralasciando le osservazioni più distanti. Questo porta a pensare di integrare il primo metodo di stima col secondo, utilizzando la stima con la mediana per ottenere un punto di riferimento e quindi sfruttare poi la stima con minimizzazione dell'errore per calcolare i contributi di tutte e sole le intersezioni che si trovano entro una certa distanza di tolleranza con il punto di riferimento prima determinato. In questo modo è possibile sfruttare le potenzialità di entrambi i metodi colmandone le reciproche lacune.

2.3.3 Stima per massimizzazione della consistenza

Quest'approccio si basa sulla stessa misura di consistenza impiegata per valutare la matrice di preferenza del punto precedente. L'idea è che se la misura di consistenza fornisce una stima corretta per valutare se un punto di fuga potenziale è un buon candidato, e se ogni cluster contiene i segmenti che appartengono al medesimo punto di fuga, allora il punto di fuga potenziale migliore è quello che massimizza il numero di segmenti consistenti con quel punto di fuga.

Per ogni segmento del cluster considerato, si valutano le intersezioni con le rispettive rette di ogni altro segmento del cluster: in questo modo si ottiene un certo numero di punti di fuga potenziali. Ognuno di questi punti di fuga viene sottoposto al test di consistenza con ogni segmento del cluster e quindi infine si sceglie il punto di fuga che massimizza il numero di segmenti consistenti.

Poiché i segmenti sono comunque soggetti a rumore, è lecito pensare che il vero punto di fuga possa comunque non trovarsi nel punto determinato dall'intersezione di due segmenti del cluster, ma in un intorno dello stesso. Per questo motivo si può pensare di raffinare ulteriormente il punto di fuga individuato, mediante una tecnica di ricerca locale che converga verso un

punto di massimo del valore di consistenza.

In questo caso si valuta non il numero di segmenti consistenti, ma la somma totale del valore di consistenza dato dai segmenti, poiché in due posizioni vicine potrebbe esserci un punto di fuga migliore dell'altro anche se rendono consistenti lo stesso numero di segmenti.

Dato un punto di fuga potenziale se ne valuta il valore di consistenza totale. Quindi si fa lo stesso per altri 8 punti di fuga potenziali posti in un intorno (nei punti cardinali) del punto iniziale. Se il punto di fuga che restituisce la maggiore consistenza risulta quello iniziale, o se la differenza con il valore di consistenza maggiore è sufficientemente contenuta, si accetta come punto di fuga finale quello centrale. Altrimenti si pone come punto di fuga iniziale quello di valore di consistenza massimo trovato.

In questo modo ci si sposta da un punto di fuga potenziale all'altro, andando a convergere verso quello che massimizza il valore di consistenza e di conseguenza andando a determinare il punto di fuga migliore per il cluster in esame.

2.4 Raffinamento dei punti di fuga stimati

Le stime individuate possono essere ulteriormente raffinate, nel tentativo di scremare un maggior numero di falsi positivi nell'individuazione dei punti di fuga. Quando si procede nella clusterizzazione, ogni segmento viene associato al cluster col punto di fuga che si rivela più consistente. Così quando si procede alla stima del punto di fuga del cluster, vengono coinvolti solo i segmenti di quel cluster, escludendone altri che potrebbero comunque dare il loro contributo.

La determinazione di nuovi punti di fuga in questa fase però potrebbe rimettere in discussione la clusterizzazione effettuata, portando a rendere consistenti verso il nuovo punto di fuga dei segmenti che prima non lo erano. È per questo che si può pensare di reiterare il processo sostituendo all'insieme iniziale di punti di fuga, ottenuti da un certo numero di coppie di segmenti prese a caso, quello dei nuovi punti di fuga stimati, mediante l'approccio EM schematizzato in figura 2.18.

L'iterazione può essere portata avanti un numero fissato di volte, oppure basarsi sui nuovi punti di fuga ottenuti.

L'operazione è facoltativa, ma permette di filtrare ulteriormente la clusterizzazione, andando a scegliere solo i segmenti più consistenti e ripescandone potenzialmente alcuni importanti che prima potevano essere stati scartati.

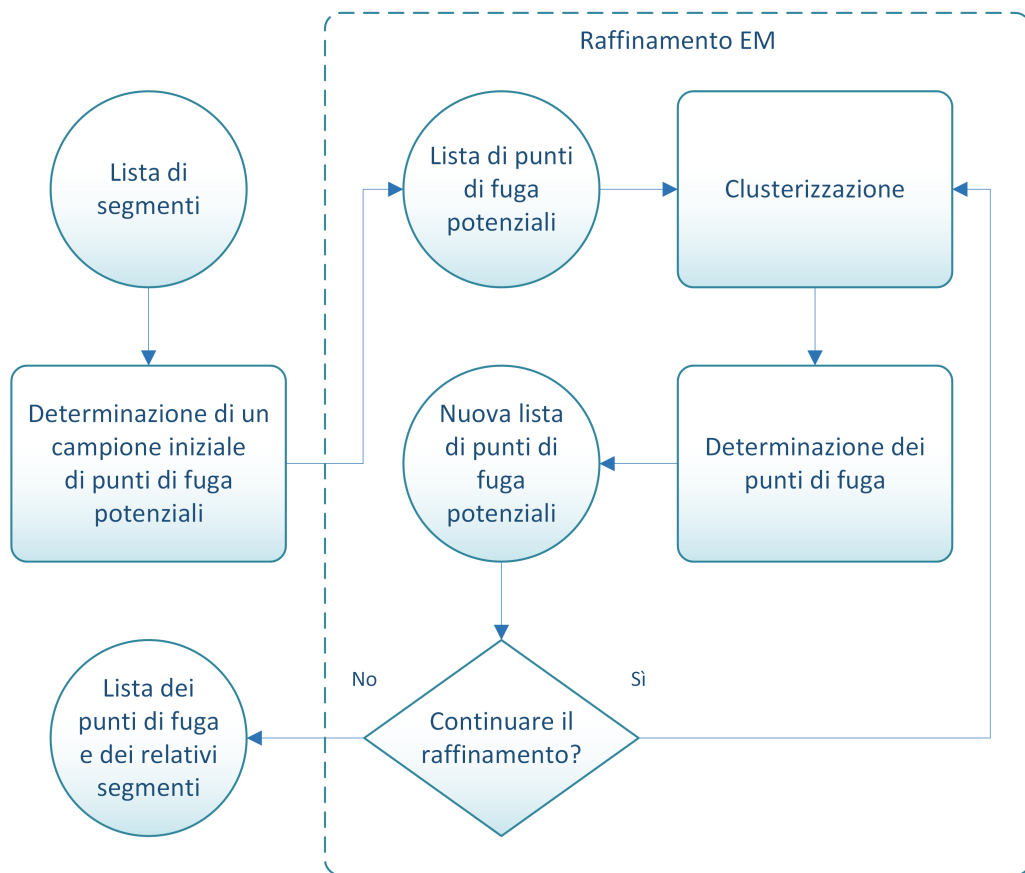


Figura 2.18: Raffinamento EM: l'iterazione viene compiuta sostituendo ai punti di fuga potenziali iniziali, quelli ottenuti dalla fase di stima dei punti di fuga dai cluster

2.5 Determinazione dei punti di fuga ortogonali

I punti di fuga individuati possono essere sfruttati per calibrare la fotocamera. È sufficiente individuare tre punti di fuga ortogonali dai quali poter estrapolare le informazioni necessarie. In questo modo si fornisce un metodo di calibrazione alternativo, rispetto al metodo classico basato su più foto e un'immagine di calibrazione.

Lo schema può essere sintetizzato come in figura 2.19.

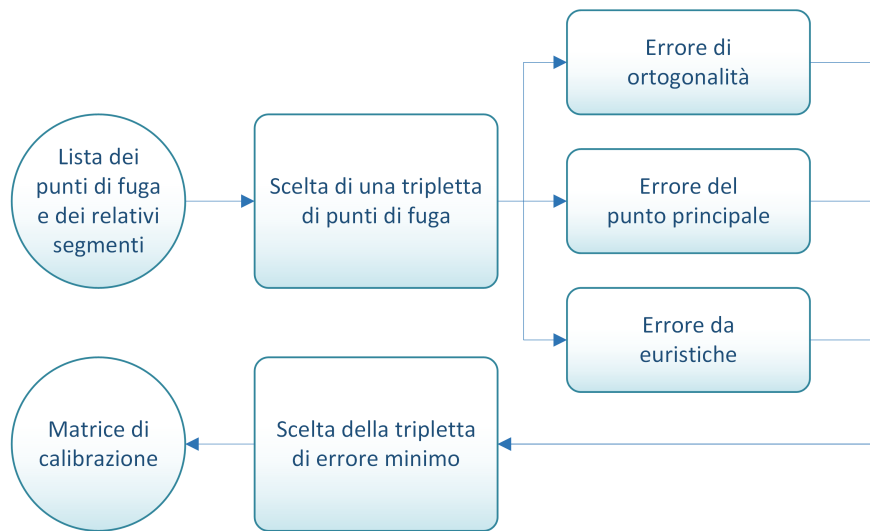


Figura 2.19: Determinazione della matrice di calibrazione: vengono calcolati gli errori relativi a ogni possibile tripletta scelta e quindi viene scelta la matrice di calibrazione generata dalla tripletta che restituisce l'errore minore

Se il numero dei punti di fuga trovati è almeno 3, è possibile verificarne l'ortogonalità. Due punti di fuga \mathbf{v}_1 e \mathbf{v}_2 sono tra loro ortogonali se:

$$\mathbf{v}_1^T \omega \mathbf{v}_2 = 0$$

dove $\omega = \mathbf{K}^{-T} \mathbf{K}^{-1}$ rappresenta l'immagine della conica assoluta (*IAC*) e \mathbf{K} è la matrice 3×3 dei parametri interni della fotocamera.

Il problema è che questa operazione è possibile solo conoscendo ω e quindi \mathbf{K} , che però è proprio la matrice che si vuole determinare. Per fare ciò

è necessario stimarne i valori in base alla tripletta considerata e quindi scegliere infine la tripletta che con quei parametri restituisce l'errore più contenuto. Un esempio è dato dalla figura 2.20.

2.5.1 Calcolo della matrice di calibrazione potenziale

Nel primo capitolo abbiamo visto come la matrice di calibrazione possa essere semplificata in:

$$\mathbf{K} = \begin{bmatrix} \alpha & 0 & x_0 \\ 0 & \alpha & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \alpha = f$$

Per definirla occorre ricavare il punto principale e la lunghezza focale, nel modo che è già stato descritto.

Una volta calcolata la matrice di calibrazione per la terna scelta, occorre stimare la bontà di questa terna e questo può essere fatto valutando una funzione di errore che prenda in considerazione diversi aspetti. Alla fine la terna scelta (e quindi la relativa matrice di calibrazione) sarà quella che restituirà l'errore più contenuto.

2.5.2 Errore di ortogonalità

Questo errore si basa sull'equazione precedentemente indicata per la quale due punti di fuga \mathbf{v}_1 e \mathbf{v}_2 sono tra loro ortogonali se:

$$\mathbf{v}_1^T \omega \mathbf{v}_2 = 0$$

Poiché abbiamo una tripletta di punti di fuga da considerare, la ricerca di quelli più adatti significa individuare i punti che rispettano l'equazione:

$$(\mathbf{v}_1^T \omega \mathbf{v}_2)^2 + (\mathbf{v}_1^T \omega \mathbf{v}_3)^2 + (\mathbf{v}_2^T \omega \mathbf{v}_3)^2 = 0$$

In pratica è molto difficile trovare una tripletta che rispetti esattamente l'equazione precedente, quindi il problema si riduce nel individuare i tre punti che la minimizzano:

$$\mathbf{v}_{1,2,3} = \min [(\mathbf{v}_1^T \omega \mathbf{v}_2)^2 + (\mathbf{v}_1^T \omega \mathbf{v}_3)^2 + (\mathbf{v}_2^T \omega \mathbf{v}_3)^2]$$

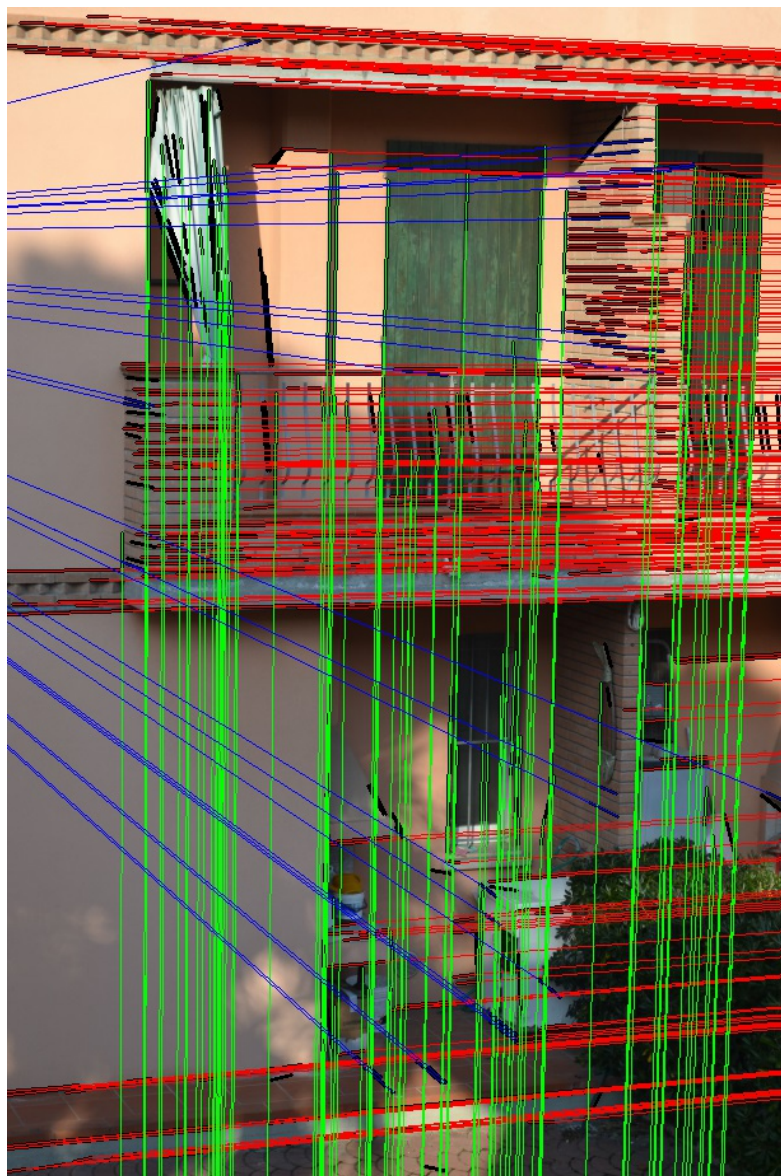


Figura 2.20: Determinazione dei punti di fuga ortogonali: sono stati mantenuti solo i punti di fuga e le linee di fuga relativi alla terna scelta

L'errore di ortogonalità si può quindi esprimere con:

$$\text{orthErr}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = (\mathbf{v}_1^T \omega \mathbf{v}_2)^2 + (\mathbf{v}_1^T \omega \mathbf{v}_3)^2 + (\mathbf{v}_2^T \omega \mathbf{v}_3)^2$$

2.5.3 Errore del punto principale

Poiché è noto che nella maggiorparte delle telecamere il punto principale si trova sempre nel punto centrale dell'immagine, è possibile impiegare questa informazione per avere una stima della bontà della tripletta considerata.

Se i tre punti di fuga scelti sono quelli corretti, il valore calcolato del punto principale dovrebbe trovarsi il più vicino possibile al centro dell'immagine, mentre una distanza maggiore dovrebbe essere indizio di un errore maggiore. Per questo motivo l'errore può essere stimato come la distanza euclidea tra il punto principale calcolato e il centro dell'immagine:

$$\text{ppErr}(\mathbf{img}, \mathbf{pp}) = \sqrt{\left(\frac{\mathbf{img}_{width}}{2} - \mathbf{pp}_x\right)^2 + \left(\frac{\mathbf{img}_{height}}{2} - \mathbf{pp}_y\right)^2}$$

Purtroppo non è possibile ricorrere alla formula inversa, usata per determinare il punto principale, per ottenere la posizione corretta dei punti di fuga, perché il numero delle variabili incognite sovrasta quello delle equazioni che è possibile utilizzare (sarebbe necessario sfruttare anche il valore richiesto di lunghezza focale, ma è proprio quello che si sta cercando di ottenere).

Perciò, partendo da questo errore si può pensare di minimizzarlo, modificando la posizione dei punti di fuga per cercare di convergere ad un punto di minimo locale. Purtroppo questo approccio si scontra con l'impossibilità di gestire tutti e tre i punti in modo armonico, poiché la determinazione del punto principale coinvolge tutti e tre i punti di fuga in modo complementare.

Spostando lievemente anche solamente un punto di fuga si può notare come punto principale vari in modo sensibile, perciò si tratta di un'operazione da compiere.

2.5.4 Errore delle euristiche

Purtroppo spesso la sola posizione dei punti di fuga non fornisce un'informazione sufficiente per potere valutare correttamente la bontà di una tripletta scelta. Per questo è importante definire altre strategie che possano

indirizzare l'algoritmo a scartare triplette che verosimilmente non possono contribuire alla terna di punti di fuga ortogonali che si stanno cercando. Poiché queste euristiche cercano di individuare le terne da scartare, l'errore che determineranno (in caso del non soddisfacimento dei vincoli imposti) dovrà essere massimo, in modo da escludere automaticamente la terna da una potenziale candidatura.

Triangolo dei punti di fuga senza angoli ottusi

La prima strategia si basa su un'idea già descritta in alcuni lavori[6] precedenti e consiste nel verificare che il triangolo formato dai punti di fuga non contenga angoli ottusi.

Verificarlo è molto semplice. Dato il triangolo $\triangle ABC$ formato dai punti di fuga **A**, **B** e **C**, si possono determinare i lati del triangolo \overline{AB} , \overline{BC} e \overline{AC} . A questo punto è possibile determinare gli angoli α , β e γ relativi ai punti **A**, **B** e **C**, grazie al teorema di Carnot:

$$\alpha = \arccos \left(\frac{\overline{AC}^2 + \overline{AB}^2 - \overline{BC}^2}{2 \cdot \overline{AC} \cdot \overline{AB}} \right)$$

$$\beta = \arccos \left(\frac{\overline{BC}^2 + \overline{AB}^2 - \overline{AC}^2}{2 \cdot \overline{BC} \cdot \overline{AB}} \right)$$

$$\gamma = \pi - \alpha - \beta$$

Rosa dei venti

Questa strategia si concentra sulla *direzione* dei punti di fuga. La semplice posizione dei punti di fuga infatti non permette di apprendere se le linee di fuga lo raggiungano da una direzione particolare o da tutte le direzioni. I casi sono esemplificati dalla figura 2.21.

Questo può essere dirimente, perché nella terna di punti di fuga considerati non possono trovarsi punti di fuga con linee di fuga che provengono dalle medesime direzioni (figura 2.22).

Quando l'angolo misurato è minore di 180° , si può calcolare una direzione specifica del fascio di linee di fuga. L'angolo di questa direzione viene quindi discretizzato in 8 direzioni corrispondenti ai punti cardinali. Se invece l'angolo è maggiore di 180° , si assume il punto di fuga come centrale, con

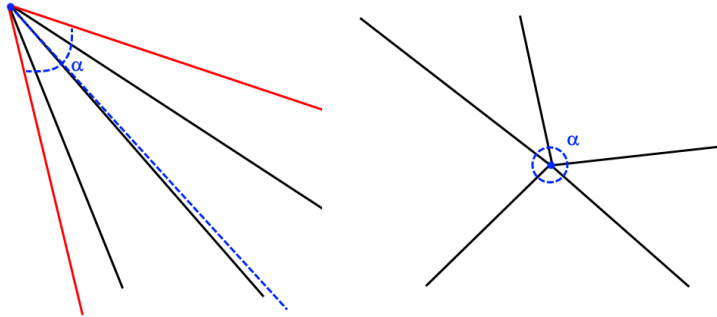


Figura 2.21: Direzione dei punti di fuga: nella prima immagine 6 linee di fuga identificano un fascio convergente di ampiezza α , calcolato considerando le linee di fuga più esterne (in rosso). La direzione è ottenuta dividendo l'angolo in due ed è rappresentata dalla linea blu tratteggiata. La seconda immagine mostra un esempio di punto di fuga centrale, dove le linee di fuga provengono da tutte le direzioni. In questo caso non vi sono linee di fuga esterne e nemmeno una direzione specifica.

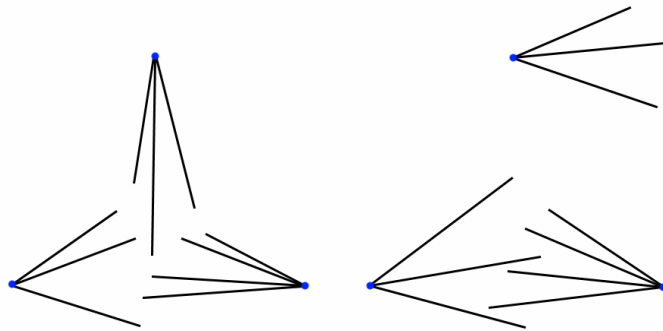


Figura 2.22: Esempi di direzioni di punti di fuga: a sinistra un esempio diterna candidabile, con i fasci dei punti di fuga diretti tutti in direzioni diverse; a destra un esempio diterna non candidabile

linee di fuga da tutte le direzioni.

Quindi, in base alle direzioni ottenute dei tre punti di fuga (in termini di punti cardinali), si fanno delle considerazioni sulla configurazione della terna:

- Nel caso un punto di fuga sia centrale, le direzioni degli altri due devono essere ortogonali
- Nel caso non vi siano punti di fuga centrali, si controlla che non vi siano due punti di fuga con la stessa direzione

Bisettrici

Questa strategia si preoccupa di controllare la bisettrice di ogni fascio di linee di fuga, che corrisponde alla loro direzione.

Dati tre punti di fuga, si considera il relativo triangolo. Per ogni punto di fuga si considera il lato opposto del triangolo, e quindi si considera la direzione di quel punto di fuga.

Per ogni punto di fuga deve verificarsi che ogni bisettrice intersechi il lato opposto del triangolo. In caso contrario, come in figura 2.23, la terna dei punti di fuga viene considerata non candidabile.

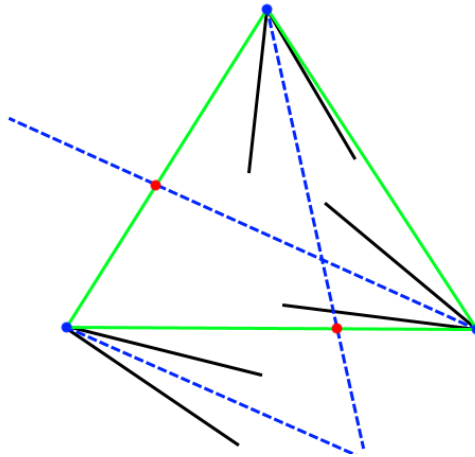


Figura 2.23: Esempio di bisettrici di punti di fuga: la terna non è candidabile poiché solo due bisettrici si intersecano con il lato opposto

Capitolo 3

Risultati sperimentali

In questo capitolo viene messo alla prova il metodo proposto e valutato in confronto a metodologie alternative. Viene presentato il set di immagini utilizzate e i risultati ottenuti vengono discussi.

Infine le prestazioni dell'approccio proposto vengono confrontate con quelle di metodologie alternative.

3.1 Testbed

Per eseguire i test col metodo proposto è stato impiegato un notebook con CPU Intel Core i7-3635QM a 2.40GHz e 8 GB di memoria RAM, su piattaforma Windows 8.1.

Le immagini sono fotografie scattate con una macchina fotografica digitale Nikon D3100, con un obiettivo AF-S Nikkor 18-55mm 1:3.5:5.6G, focus automatico e sistema di riduzione delle vibrazioni attivato. Per ogni scatto l'obiettivo è stato mantenuto ad una lunghezza focale di 55mm.

Prima di procedere con l'elaborazione, le immagini sono state ridimensionate a un quarto della loro dimensione originale. Oltre ad alleggerire le immagini, questo porta implicitamente ad approssimare l'immagine. Ne consegue che il numero di segmenti estratti è minore e la lista ottenuta dal riconoscitore conterrà solamente i segmenti più importanti.

Per la calibrazione classica è stato realizzato un pannello su cui è stato applicato un pattern di calibrazione a forma scacchiera.

La figura 3.1 rappresenta la strumentazione utilizzata per la calibrazione col metodo classico.



Figura 3.1: Strumenti usati per la calibrazione tramite pattern a scacchiera

3.2 Obiettivi dei test

Con questi risultati si intende verificare le prestazioni dell'algoritmo a fronte di diverse situazioni, a cui corrispondono le diverse immagini, e confermare la bontà delle scelte intraprese in fase di progettazione.

Al termine di ogni test si ottiene una matrice di calibrazione. Poiché il punto principale è centrale nella quasi totalità delle telecamere e si tratta di una misura molto suscettibile al rumore, è stato assunto al centro dell'immagine e quindi l'unico parametro da analizzare è la lunghezza focale.

Per valutare la qualità del risultato, verrà confrontata la lunghezza focale ottenuta con l'approccio presentata con quella misurata attraverso la calibrazione classica, che verrà considerata il valore corretto. Inoltre il confronto sarà effettuato anche con altri metodi di autocalibrazione, per poterli valutare direttamente.

Poiché l'algoritmo presenta al suo interno delle componenti randomiche, verranno effettuate delle ripetizioni dei test, per poter valutare una tendenza che abbia maggior significato statistico.

3.3 Determinazione della ground truth

Per poter procedere con l'analisi dei risultati è necessario determinare il valore di riferimento che permetta di discriminare da un valore buono e un valore cattivo. Poiché la calibrazione classica è una metodologia ormai consolidata per la calibrazione affidabile di una telecamera, si è deciso di considerare come ground truth il valore di lunghezza focale così stimato.

Per questo motivo si è provveduto a calibrare la macchina fotografica nel settaggio proposto, attraverso il metodo classico offerto dalla libreria OpenCV, utilizzando come pattern una scacchiera di cui sono state effettuate alcune fotografie.

La matrice di calibrazione standard così ottenuta è:

$$\mathbf{K}_{\text{std}} = \begin{bmatrix} 2116.53 & 0 & 306.5 \\ 0 & 2116.53 & 460.5 \\ 0 & 0 & 1 \end{bmatrix}$$

perciò $f_{\text{std}} = 2116.53$.

3.4 Immagini dei test

Vengono enumerate e presentate le immagini utilizzate per il test:

Immagine n°1 : Si tratta dell'immagine campione già presentata nel capitolo precedente. Le direzioni ortogonali sono facilmente intuibili, ma l'immagine presenta oggetti di disturbo come ombre, e linee curve (come la ringhiera del balcone).

Immagine n°2 : L'immagine presenta direzioni ortogonali molto nette, ma la profondità di campo è molto ridotta per la vicinanza con l'oggetto al momento dello scatto. Ciò enfatizza gli effetti distorsivi della lente.

Immagine n°3 : Le direzioni ortogonali sono nette, ma la scalinata obliqua può creare difficoltà al momento della selezione della terna di punti di fuga ortogonali.

Immagine n°4 : Le direzioni ortogonali sono nette, ma le travi di legno possono creare difficoltà nella selezione dei punti di fuga ortogonali.

Immagine n°5 : Due direzioni ortogonali sono piuttosto evidenti, ma le persone in strada determinano una notevole fonte di disturbo.

Immagine n°6 : Due direzioni ortogonali sono piuttosto evidenti, ma la terza è completamente assente.

Immagine n°7 : Le direzioni ortogonali sono evidenti, ma l'immagine è molto scura.

Immagine n°8 : Le direzioni ortogonali sono evidenti, ma la vegetazione non fornisce profili netti, complicando l'analisi dell'immagine.

Immagine n°9 : La presenza di linee oblique rende difficoltosa la scelta corretta dei tre punti di fuga ortogonali.

Immagine n°10 : La mancanza di linee ortogonali dominanti rende difficile l'analisi dell'immagine.



Figura 3.2: Immagine di test n°1

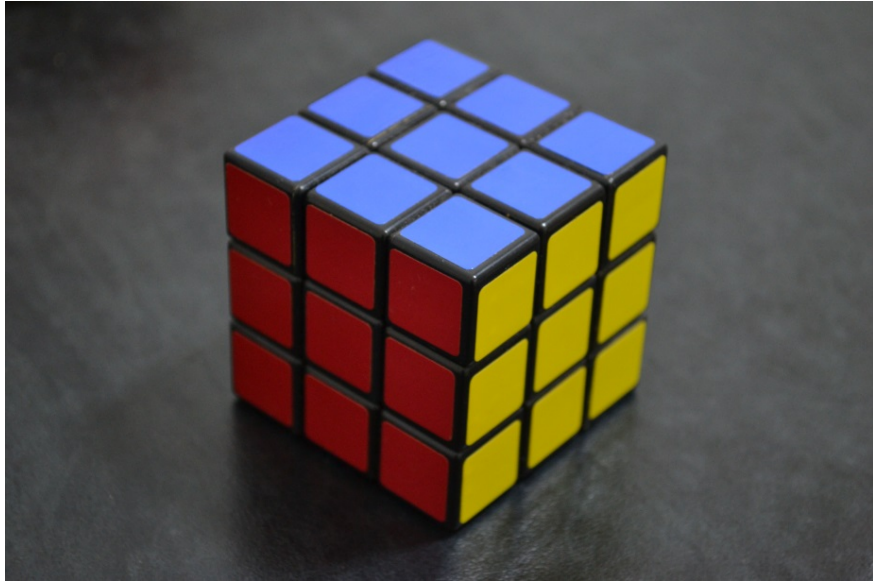


Figura 3.3: Immagine di test n°2

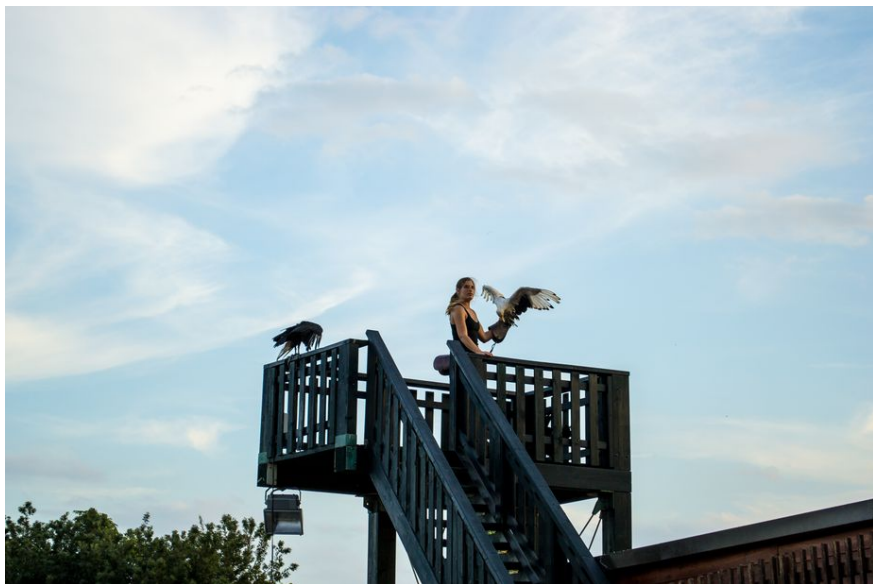


Figura 3.4: Immagine di test n°3



Figura 3.5: Immagine di test n°4



Figura 3.6: Immagine di test n°5



Figura 3.7: Immagine di test n°6

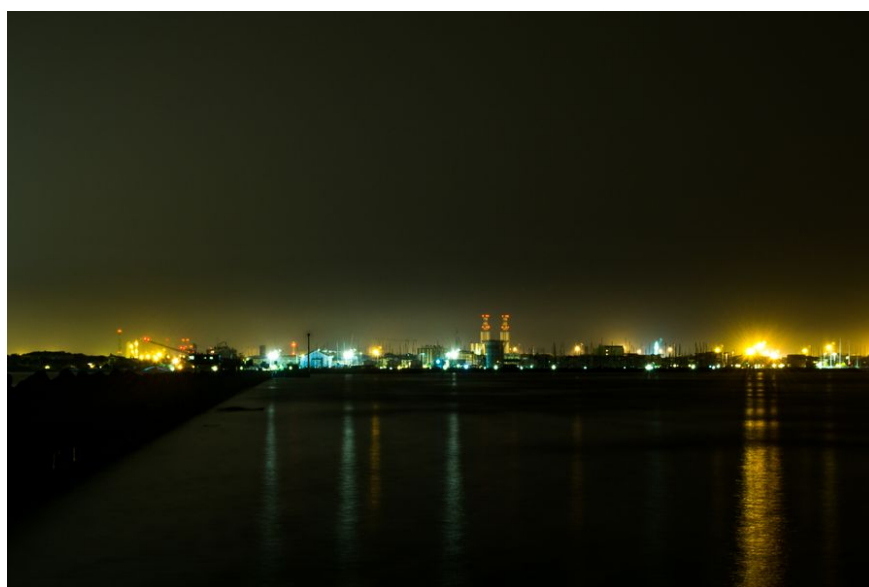


Figura 3.8: Immagine di test n°7

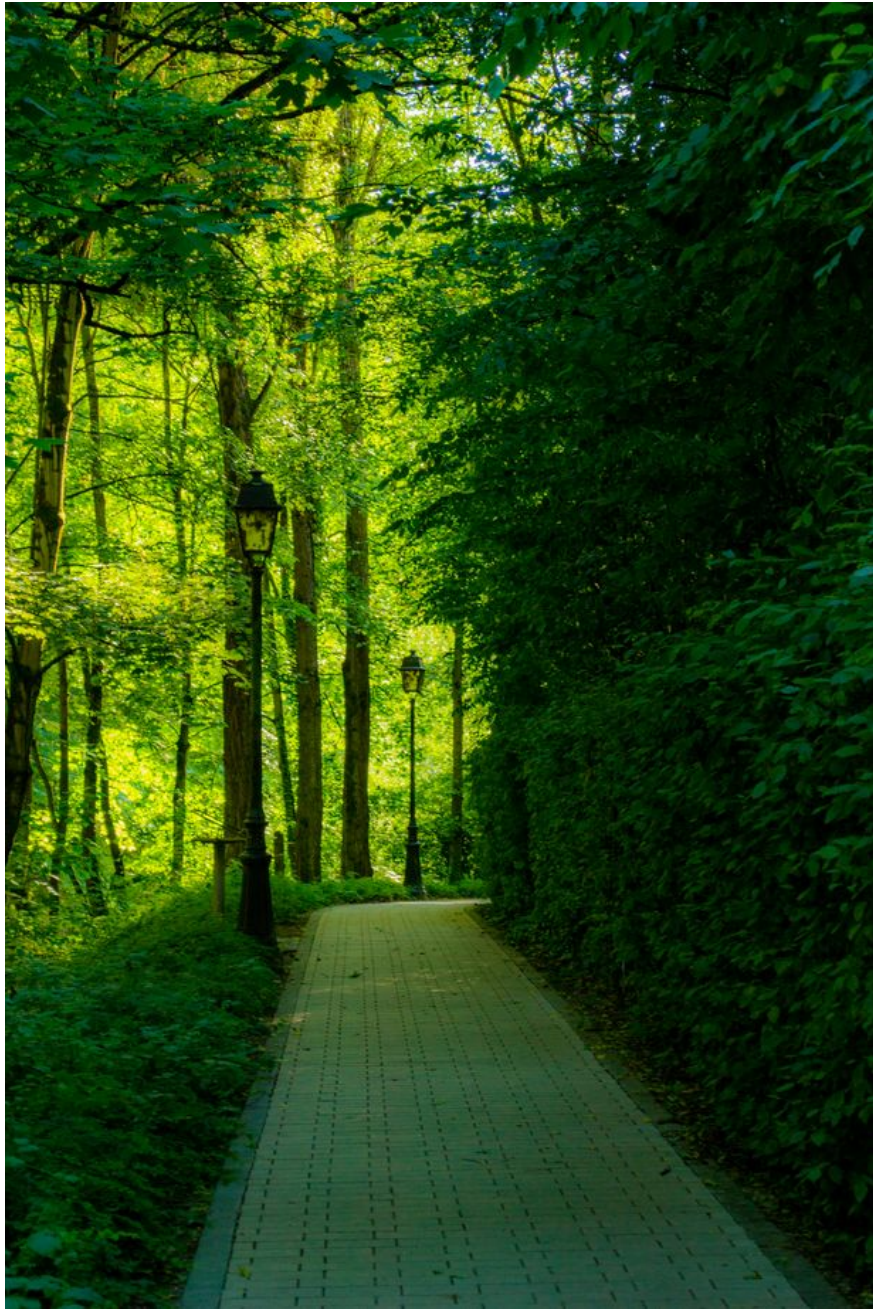


Figura 3.9: Immagine di test n°8

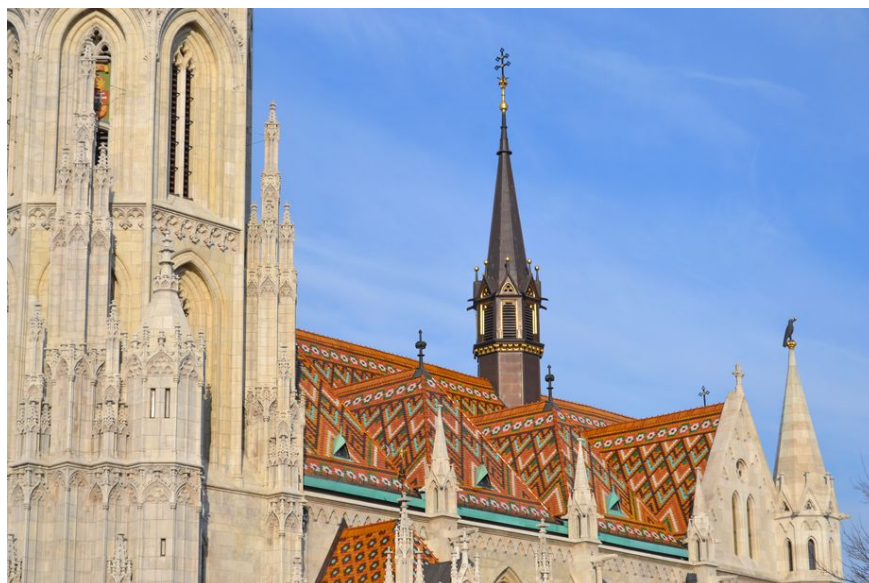


Figura 3.10: Immagine di test n°9

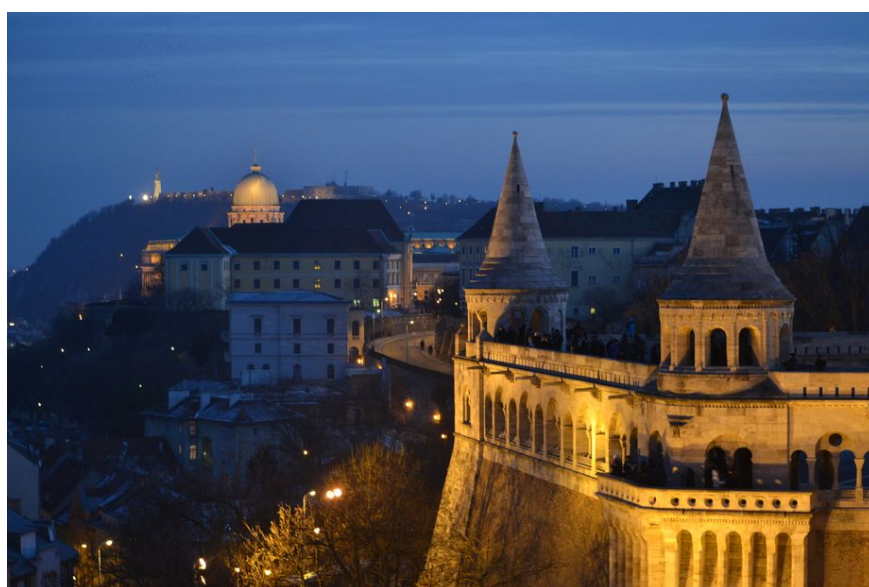


Figura 3.11: Immagine di test n°10

3.5 Risultati ottenuti

In questa sezione verranno presentati e discussi i risultati ottenuti. In ogni caso si farà affidamento al valore di lunghezza focale misurato attraverso la calibrazione classica, che verrà assunto come *ground truth*.

Per prima cosa viene valutata la qualità del metodo di autocalibrazione proposto, verificando la lunghezza focale stimata automaticamente e quella migliore che potenzialmente si può ottenere scegliendo una terna diversa di punti di fuga.

Quindi viene discussa la distribuzione dei risultati a fronte di scelte randomiche effettuate nel corso dell'esecuzione.

Infine vengono confrontati i risultati ottenuti con altre soluzioni proposte dalla letteratura, per valutare le prestazioni in relazione a soluzioni alternative proposte.

3.5.1 Misurazione intrinseca della qualità

In questo test viene valutata la capacità del programma di stimare correttamente la lunghezza focale, determinando la terna corretta. Per fare ciò è stato eliminato qualsiasi elemento randomico, perciò sono stati mantenuti tutti i segmenti rilevati e come insieme iniziale dei punti di fuga potenziali (in fase di clusterizzazione) sono state considerate tutte le intersezioni possibili tra i vari segmenti.

Il valore ottenuto della lunghezza focale viene confrontato con quello ottenuto scegliendo manualmente la terna da considerare. Questo permette di valutare la massima efficacia della soluzione proposta a fronte dei punti di fuga determinati.

Per ogni test, la prima immagine mostra la terna selezionata automaticamente, mentre la seconda (quando presente) mostra la terna migliore scelta manualmente. Nel caso di una sola immagine, la terna scelta è stata la migliore. Ogni punto di fuga è identificato da un colore diverso, lo stesso usato per identificare i segmenti e le rispettive linee di fuga appartenenti a quel punto di fuga.

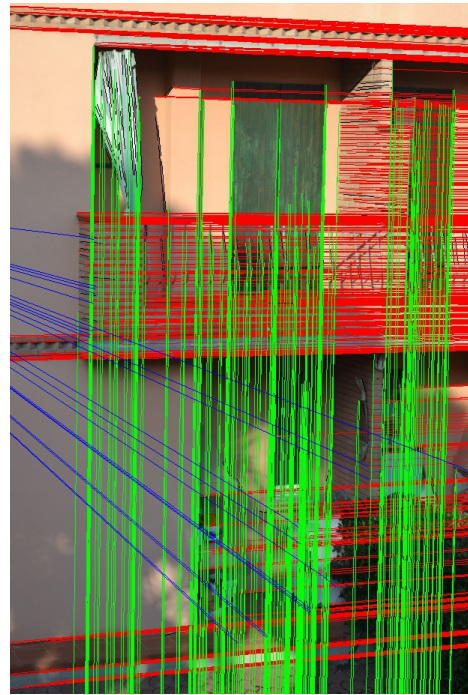
Per ogni immagine viene riportata la lunghezza focale determinata dalla terna selezionata.

Immagine n°1

La terna migliore non viene scelta in virtù di un maggiore errore del punto principale. Ciò è evidenziato dal fatto che la terna migliore produca un risultato più lontano dal valore corretto, rispetto a quella scelta automaticamente. Ciononostante i risultati sono comunque molto buoni.



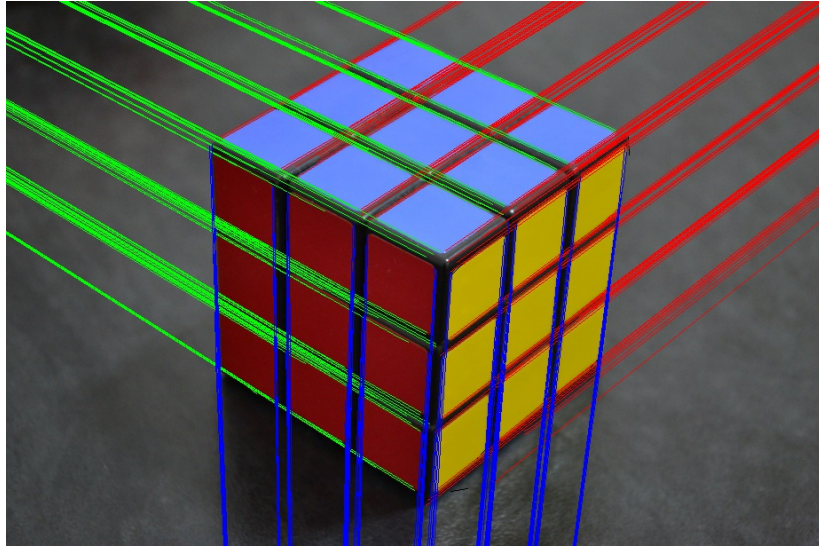
$$f = 2034.78$$



$$f = 1523.6$$

Immagine n°2

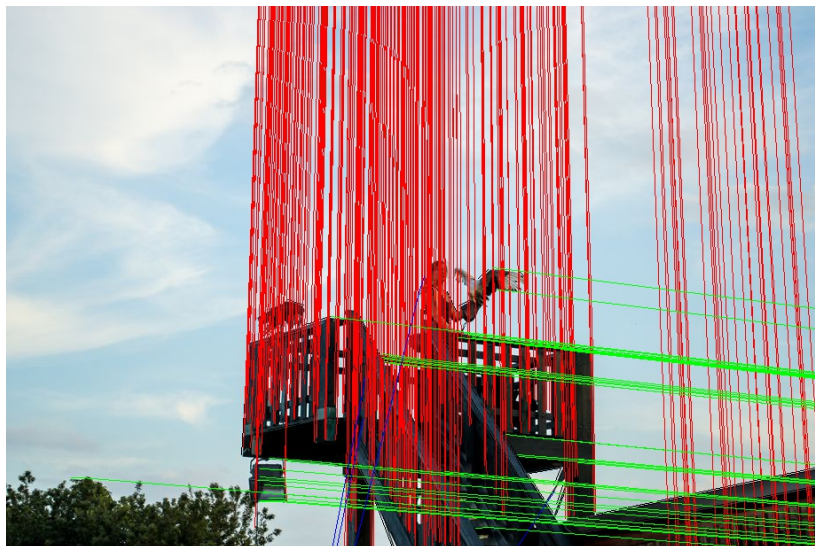
L'immagine viene segmentata correttamente e i punti di fuga determinati chiaramente, ma gli effetti distorsivi dovuti alla distanza ravvicinata dello scatto non permettono di avere una stima corretta della lunghezza focale.



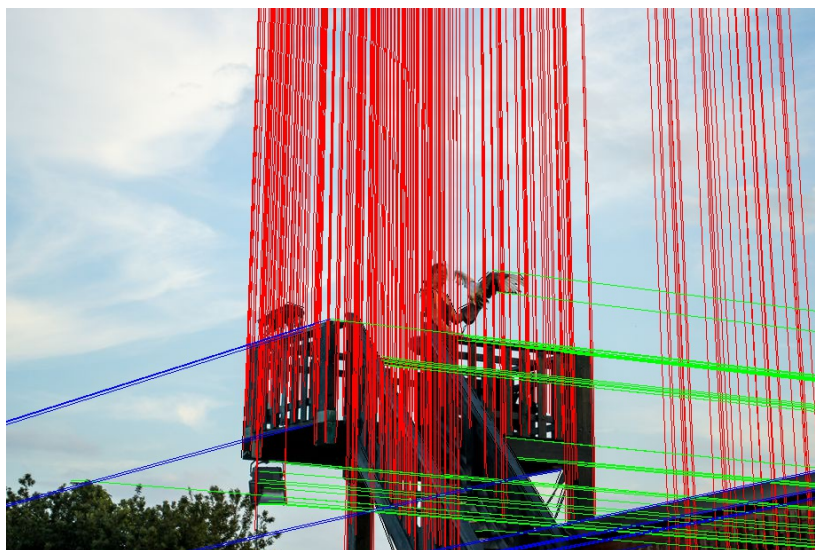
$$f = 1114.05$$

Immagine n°3

La presenza di linee ortogonali segmentate con precisione, permette di ottenere risultati molto buoni in entrambi i casi.



$$f = 2243.99$$



$$f = 2508$$

Immagine n°4

Nonostante la presenza di altre linee parallele disturbi la determinazione della terna corretta, grazie alle linee dominanti i valori ottenuti risultano molto buoni in entrambi i casi.



$$f = 2635.03$$



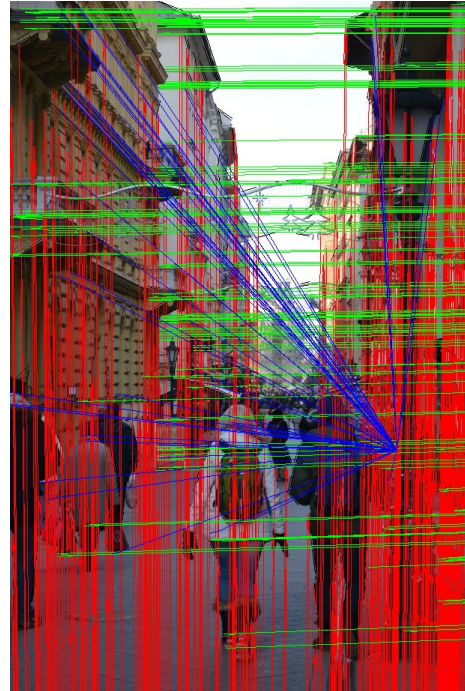
$$f = 2033.39$$

Immagine n°5

Come era stato previsto in fase di analisi, la presenza di un elevato numero di elementi di disturbo impedisce di isolare correttamente le direzioni ortogonali e di conseguenza il valore ottenuto di lunghezza focale risulta ben lontano dal valore misurato con calibrazione classica.



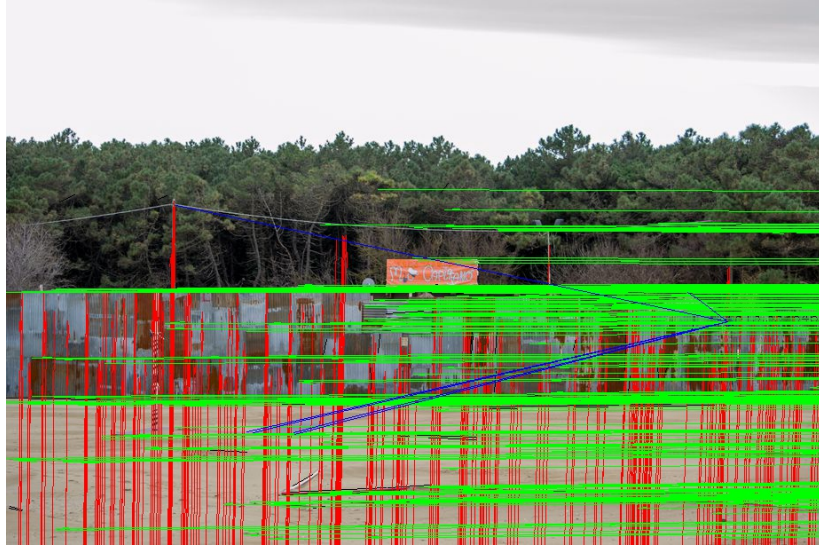
$$f = 4586.84$$



$$f = 5571.47$$

Immagine n°6

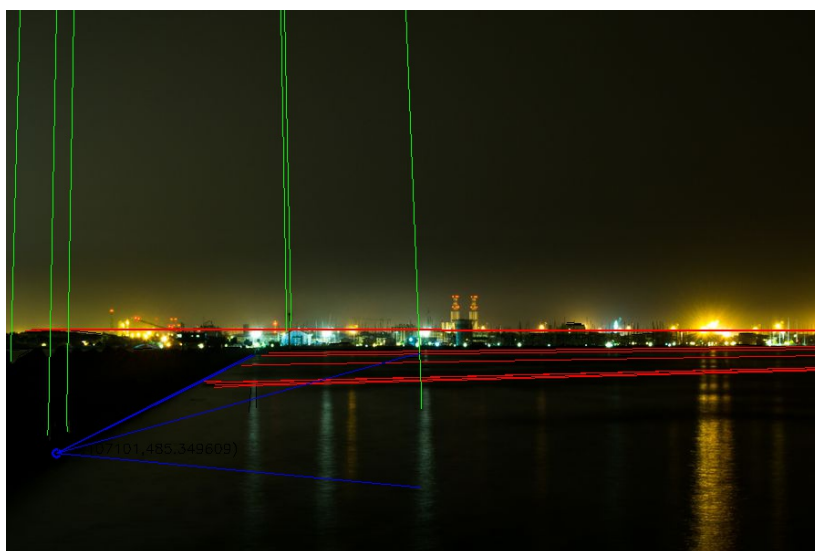
Come previsto, la presenza di sole due direzioni dominanti rende impossibile ottenere misurazioni affidabili.



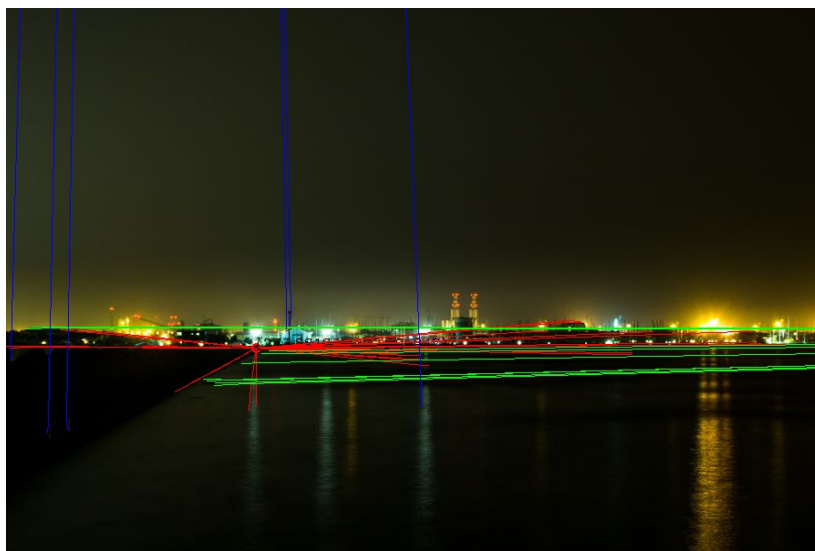
$$f = 8158.66$$

Immagine n°7

L'immagine scura non permette di avere bordi ben contrastati e ciò rende complessa l'estrazione dei segmenti. In condizioni così difficili il risultato è comunque piuttosto buono.



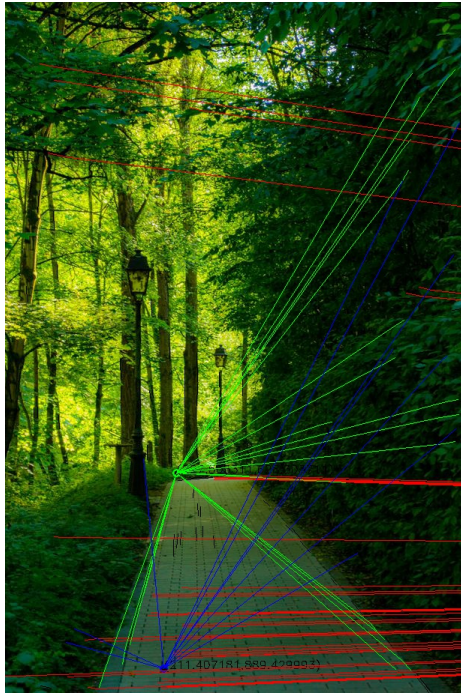
$$f = 1137.19$$



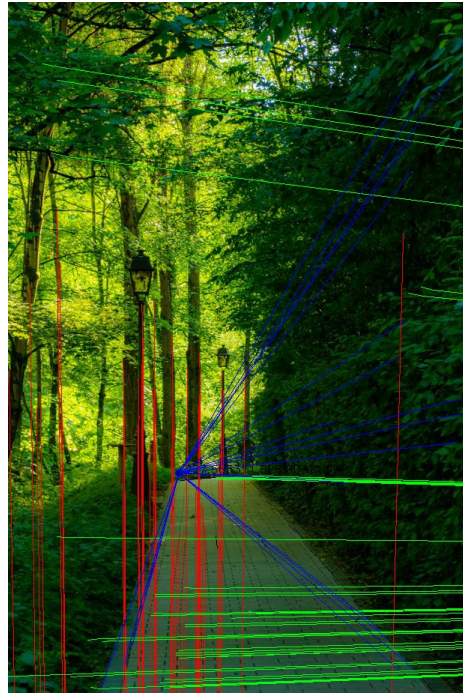
$$f = 2592.37$$

Immagine n°8

La presenza di fogliame porta all'estrazione di molti segmenti irregolari, che forniscono elementi di disturbo nella clusterizzazione. Questo ovviamente porta a determinare punti di fuga non precisi, che determinano un errore apprezzabile nella stima della lunghezza focale.



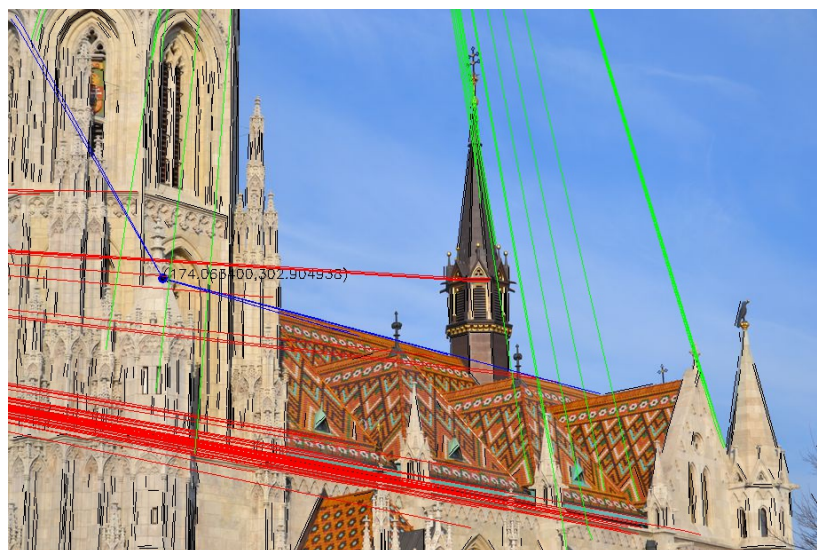
$$f = 1498.84$$



$$f = 3511.88$$

Immagine n°9

Un valore di sogliatura non perfetto, non permette di determinare correttamente il bordo, di conseguenza i segmenti estratti non consentono una perfetta determinazione dei punti di fuga e quindi i valori delle lunghezze focali risultano comunque sottostimati.



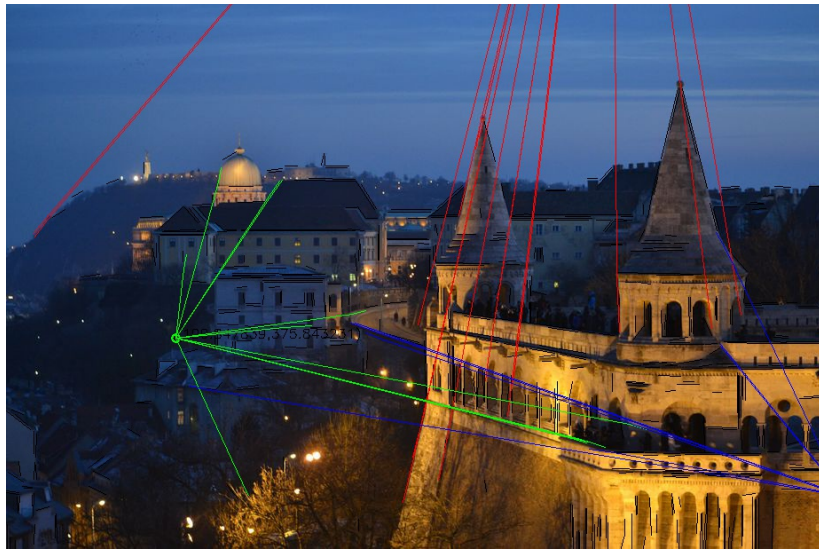
$$f = 859.174$$



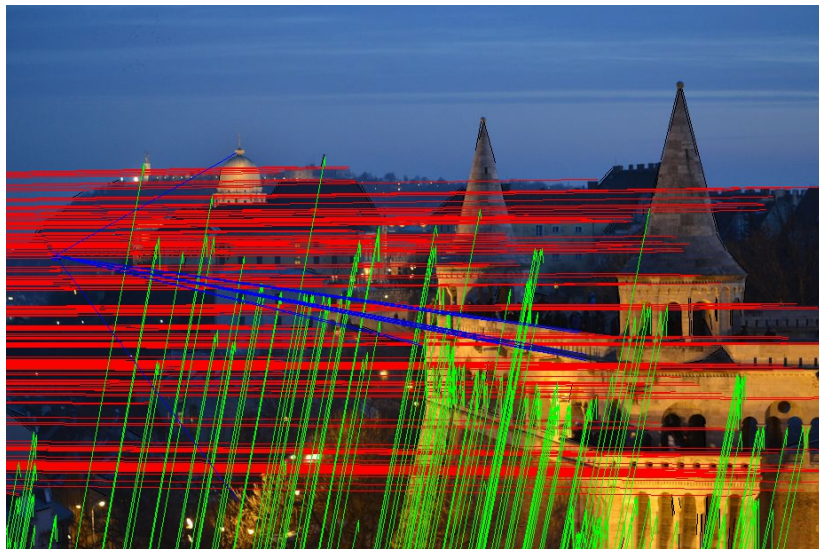
$$f = 902.655$$

Immagine n°10

La mancanza di linee ortogonali chiare non permette una stima precisa della lunghezza focale e ciò determina un errore apprezzabile del valore ottenuto.



$$f = 691.317$$



$$f = 3773.3$$

3.5.2 Misurazione statistica della qualità

Con questo test si vuole valutare la distribuzione dei risultati a fronte degli elementi randomici presenti, che consentono di velocizzare l'esecuzione.

È stato fissato un massimo di 500 segmenti scelti e un massimo di 2000 punti di fuga potenziali iniziali.

Ogni test è stato ripetuto 100 volte per poter valutare una tendenza che abbia maggior significato statistico. Dopo ognuno di questi test, viene considerata la lunghezza focale determinata automaticamente.

I risultati verranno rappresentati su un'istogramma che, a fronte dei diversi valori di lunghezza focale, presenterà le frequenze con cui tali valori si sono presentati.

I valori di lunghezza focale riscontrati sono stati aggregati in intervalli di 50pixel, per evitare che minime dispersioni (insignificanti all'atto pratico) potessero ostacolare l'osservazione della tendenza.

Immagine n°1

Come attesta il grafico, la stima è buona. La maggiorparte dei valori di lunghezza focale si concentra tutta attorno a 1915 pixel, compiendo un errore del 9.5% rispetto al valore misurato mediante calibrazione standard.

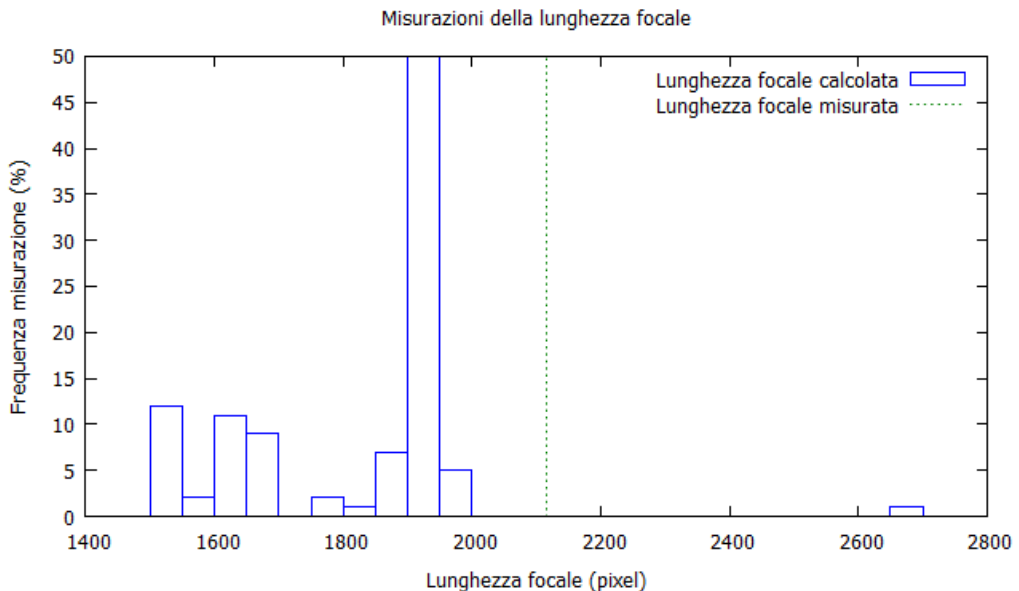


Immagine $n^{\circ}2$

La linee di fuga ben definite per l'assenza di oggetti di disturbo, consentono alla quasi totalità dei casi di concentrarsi sul medesimo valore, 1114 con un errore del 47.3% dovuto a gli effetti distorsivi.

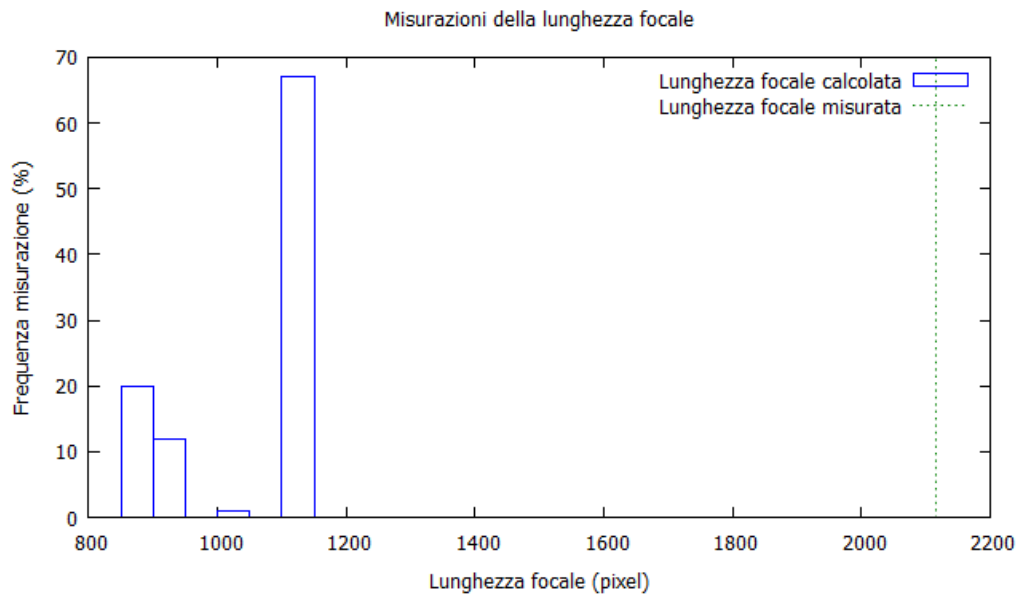


Immagine n°3

La presenza di linee oblique parallele deviano il valore della lunghezza focale da quello atteso. Cionostante la maggioranza dei casi si concentra attorno al valore 2243, con un errore del 6%.

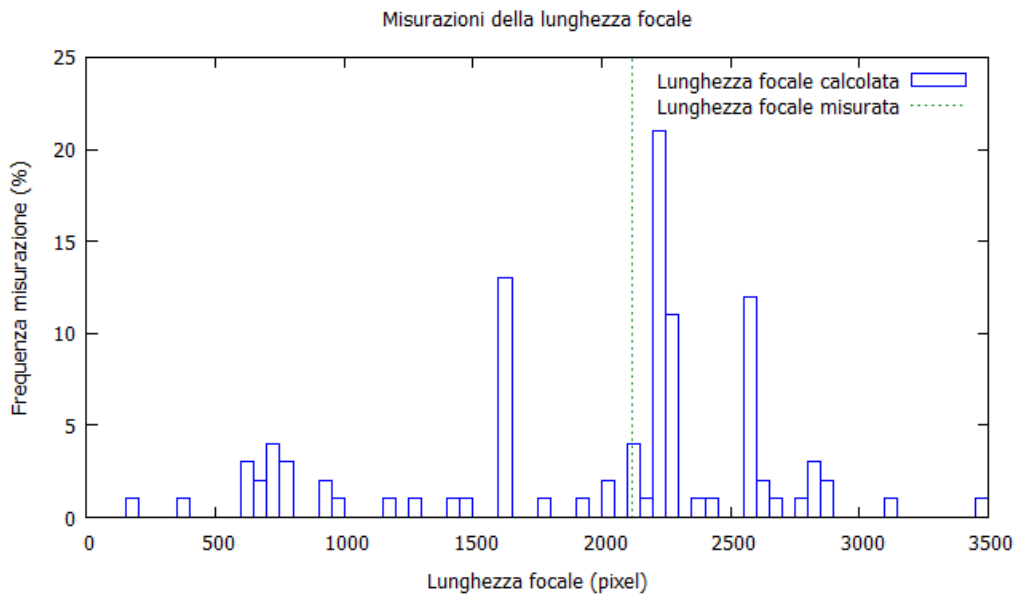


Immagine n°4

Anche in questo caso le linee oblique parallele disturbano la corretta selezione dei punti di fuga ortogonali. Poiché le direzioni ortogonali sono meno dominanti rispetto all'immagine precedente, in questo caso la lunghezza focale si orienta attorno al valore di 1618 pixel, con un errore del 23.5%.

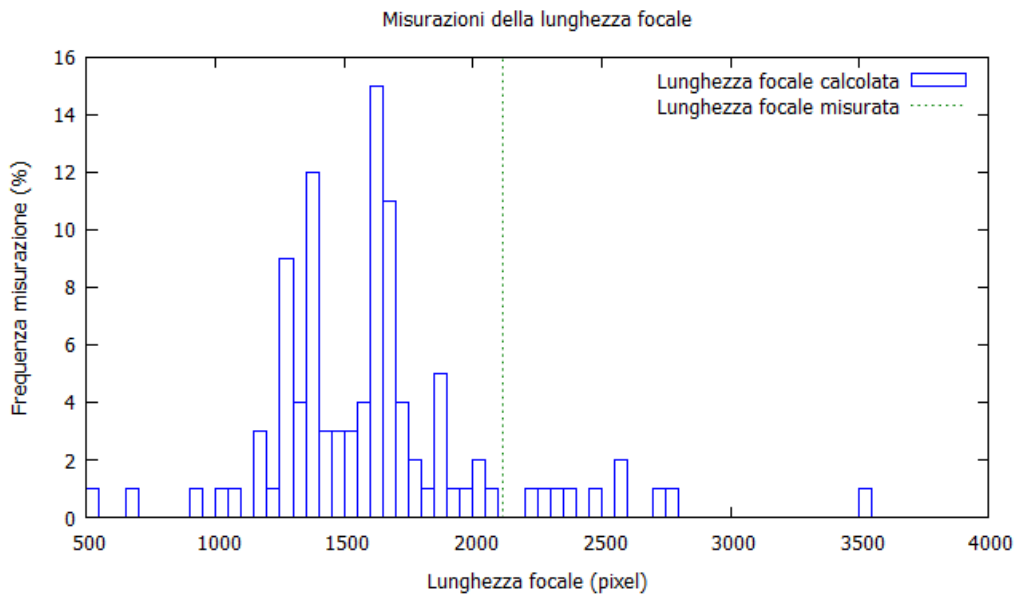


Immagine n°5

La presenza di un elevato numero di elementi di disturbo nell'immagine, impediscono una stima corretta della lunghezza focale e ciò è reso evidente dall'elevata dispersione dei valori. La concentrazione maggiore avviene attorno a 554 pixel, a cui corrisponde un errore del 73.8%.

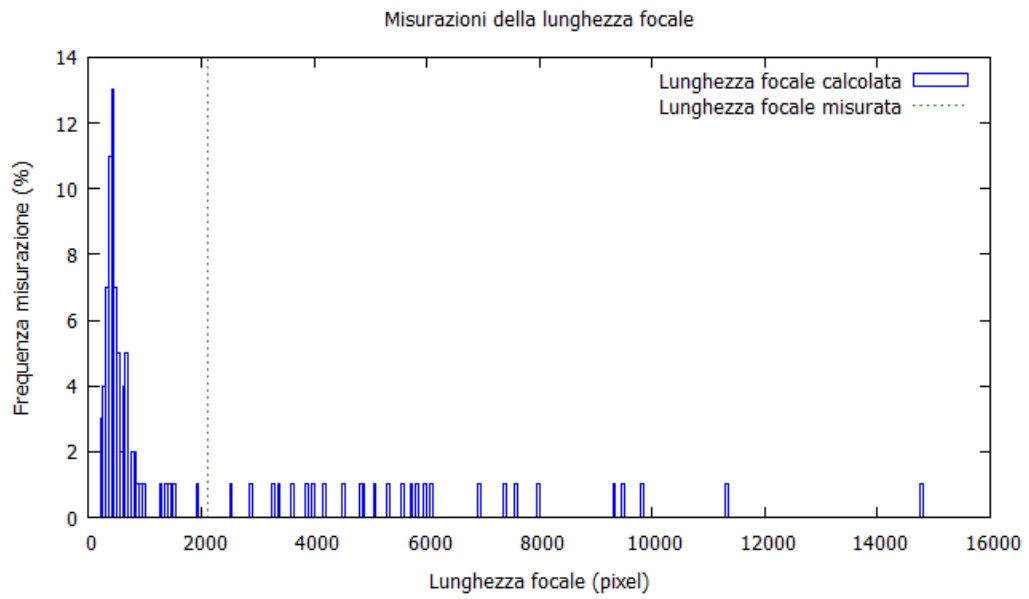


Immagine n°6

In questo caso l'assenza nell'immagine di una terza direzione ortogonale ben definita porta alla determinazione di valori inaffidabili, che si concentrano attorno al valore 7687, pari ad un errore del 263%. Più del doppio rispetto alla lunghezza focale reale. La scarsa affidabilità del dato è resa evidente anche dalla notevole dispersione dei dati.

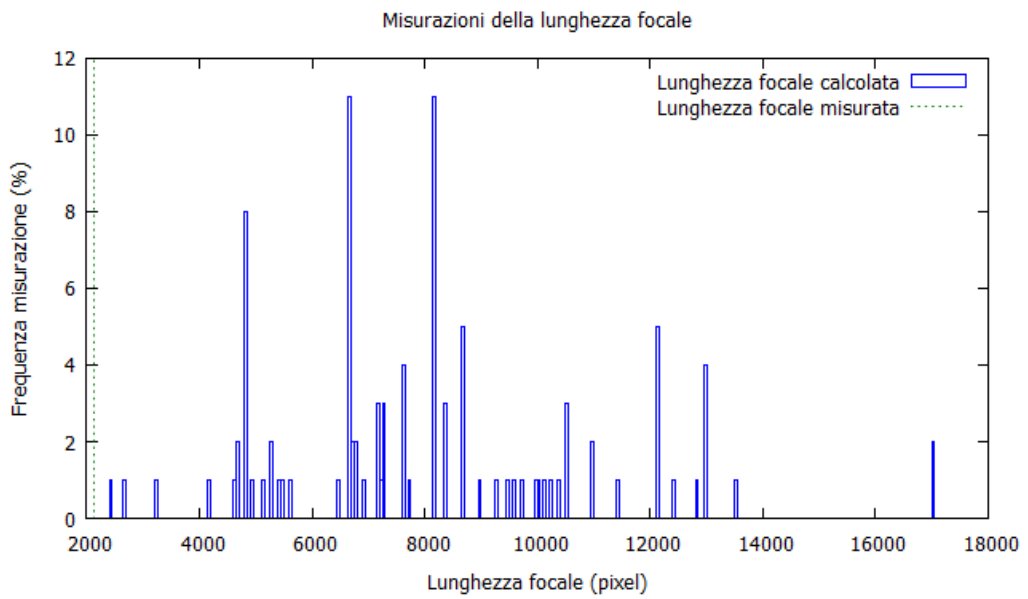


Immagine n°7

La presenza di pochi segmenti, dovuti all'immagine scura, determina un'influenza minore degli elementi randomici dell'algoritmo e quindi causa una concentrazione spiccata verso un unico risultato: 1137 pixel pari ad un errore del 46.3%. La mancanza di contrasto sufficiente impedisce al riconoscitore di estrarre un maggior numero di segmenti che avrebbero reso la misura più precisa.

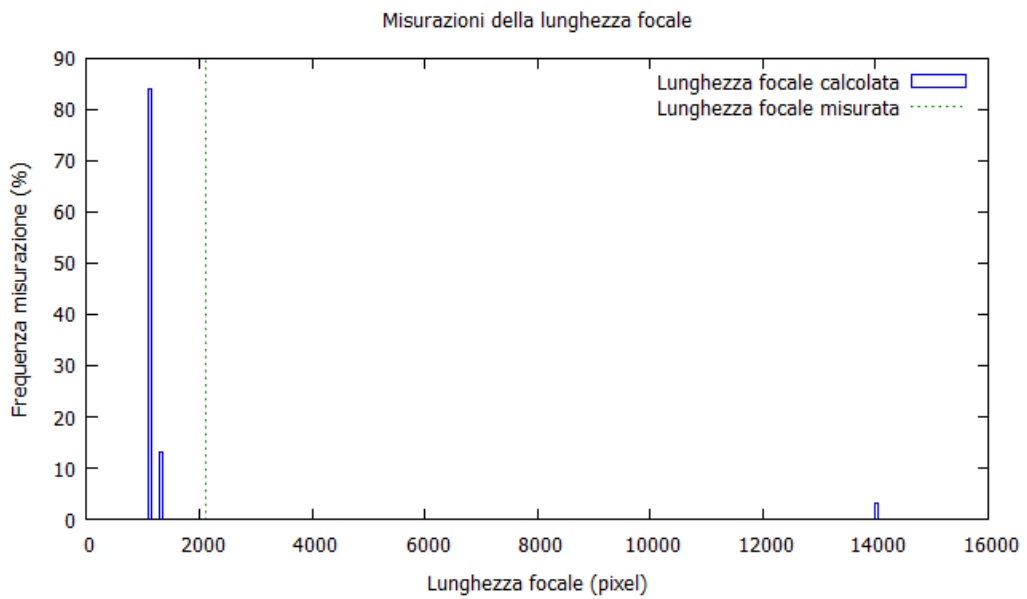


Immagine n°8

I profili delle foglie rendono difficile ottenere una stima corretta dei punti di fuga e ciò determina una concentrazione meno decisa, che si orienta attorno a due valori: 300 pixel (85.8%) e 1331 pixel (37.1%).

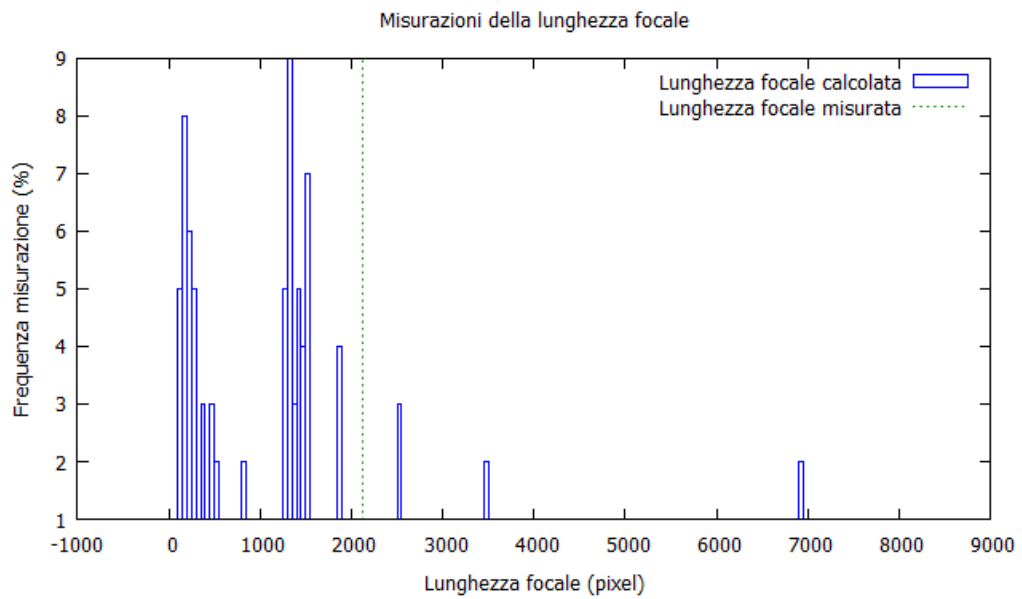


Immagine n°9

L'alto numero di direzioni non ortogonali, unito ad una segmentazione difficile a causa di una sogliatura non perfetta, non permette spesso di ottenere un valore di lunghezza focale vicino a quello atteso. I dati si concentrano maggiormente attorno al valore 543, pari ad un errore del 74.3%. Tuttavia i dati sono molto distribuiti e si orientano attorno al valore 1827: un errore del 13.7%.

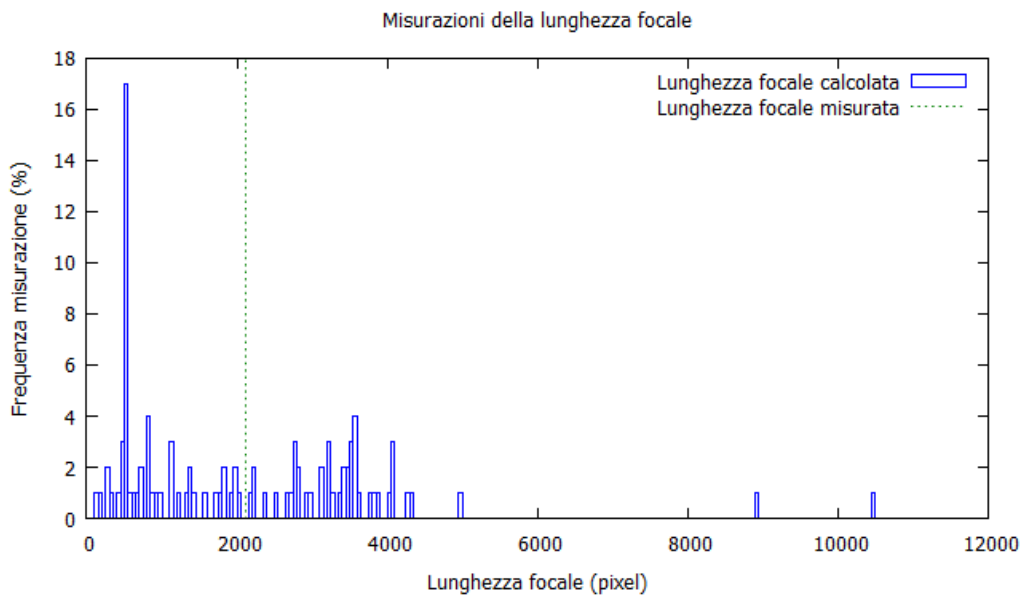
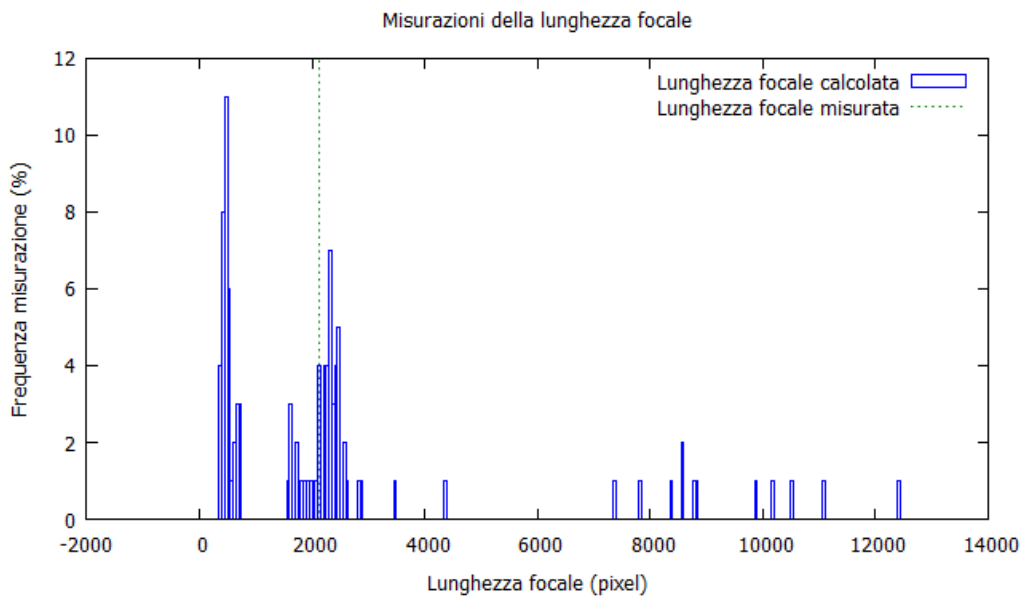


Immagine n°10

Anche in questo caso la mancanza di linee ortogonali chiare non permette una stima precisa della lunghezza focale, che si concentra attorno a due valori. Il primo è 520 pixel, a cui corrisponde un errore del 75,4%. Il secondo invece è 2036 pixel e corrisponde alla scelta della terna corretta di punti di fuga. È il valore attorno a cui si orienta l'intera distribuzione e in questo caso l'errore è limitato al 3.8%.

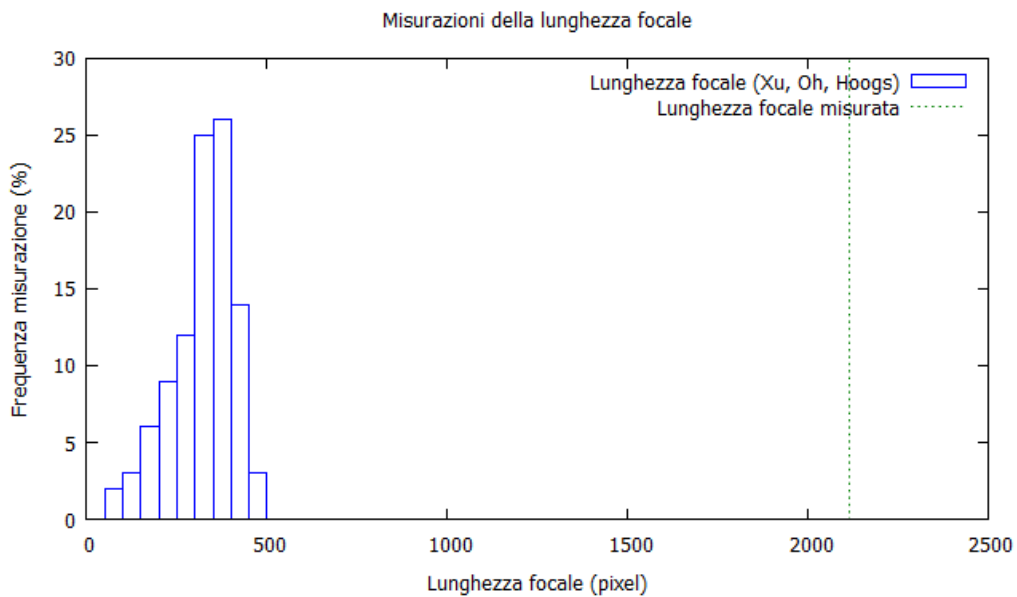


3.6 Confronto con altre metodologie

In questa sezione viene confrontato l'approccio proposto con soluzioni alternative. In particolare verrà paragonato con la soluzione di Xu, Oh e Hoogs[9] (che è un miglioramento della soluzione di Tardif[7]), per la determinazione dei punti di fuga, e con la proposta *MSAC* di Nieto e Salgado[5], per avere una valutazione qualitativa.

3.6.1 Confronto col metodo proposto da Xu, Oh e Hoogs

Poiché la differenza tra l'approccio proposto e quello di Xu, Oh e Hoogs consiste essenzialmente nella diversa metodologia di stima dei punti di fuga, sono state mantenute le medesime condizioni variando solo quel singolo aspetto. Poiché le euristiche erano troppo punitive nei confronti dei punti di fuga identificati, sono state disabilitate nella fase di determinazione della terna ortogonale.



In questo caso il confronto verrà effettuato solamente con la prima immagine di test, in quanto è del medesimo tipo di quelle su cui sono state progettate

le due metodologie ed è sufficiente per poter già bene apprezzare differenze qualitative tra i due approcci.

Come si può notare, i risultati si discostano dal valore corretto molto più rispetto all'approccio proposto. Questo è imputabile alla tecnica di determinazione dei punti di fuga di Xu, Oh e Hoogs che, pur essendo molto sofisticata, non riesce a dimostrarsi efficace a fronte di outlier che inevitabilmente sono presenti in immagini reali, con segmenti estratti attraverso modalità automatiche.

3.6.2 Confronto col metodo proposto da Nieto e Salgado

Il metodo proposto da Nieto e Salgado è più competitivo grazie alla tecnica basata su un approccio Ransac, che riesce a sopportare meglio gli outlier. Per questo motivo può essere interessante valutare le prestazioni dei due approcci in ogni immagine. Viene poi definito un punteggio per valutare le prestazioni misurate in ogni test e quindi viene effettuata una comparazione prestazionale tra l'MSAC e l'approccio proposto. Anche in questo caso sono state disabilitate le euristiche poichè troppo punitive nei confronti dei punti di fuga identificati.

Nella pagina successiva seguono gli istogrammi ottenuti per ogni immagine. Il primo istogramma riguarda il test effettuato con l'approccio proposto, mentre il secondo corrisponde allo stesso test eseguito col metodo di Nieto e Salgado.

A seguire vi sarà la comparazione delle due metodologie sulla base dei risultati dei test.

Immagine n°1

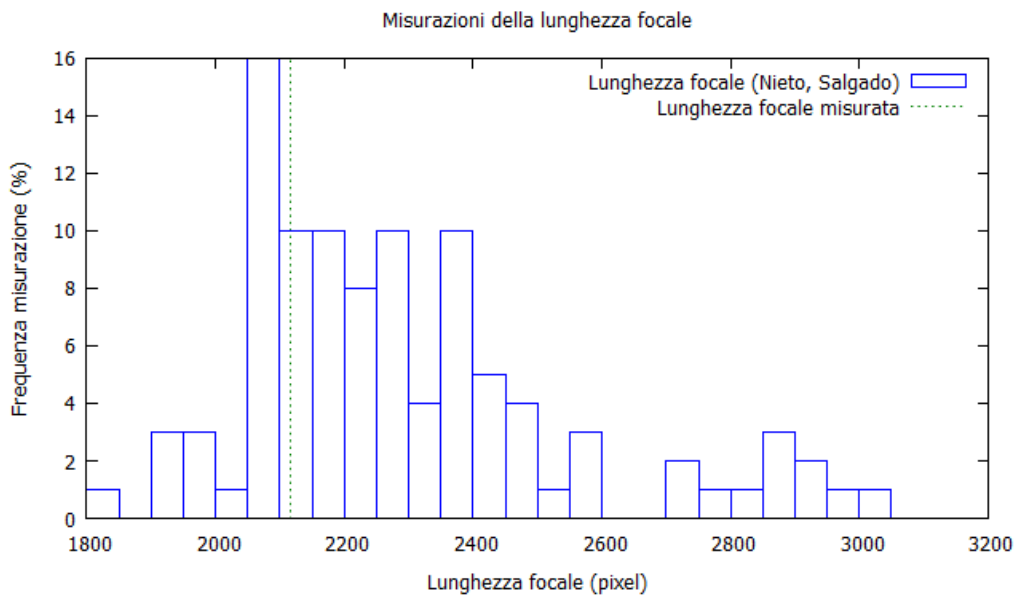
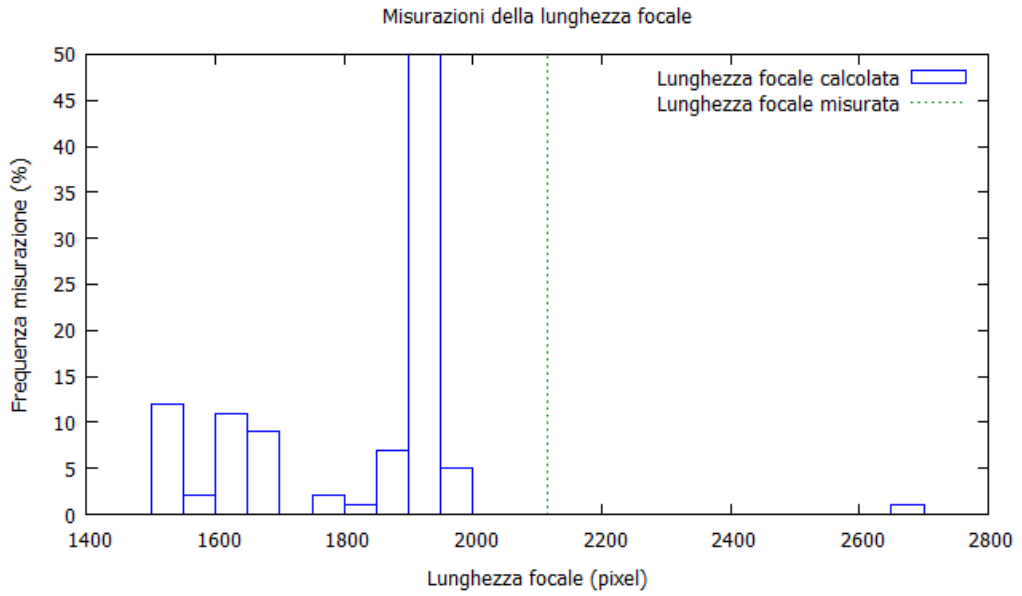


Immagine n°2

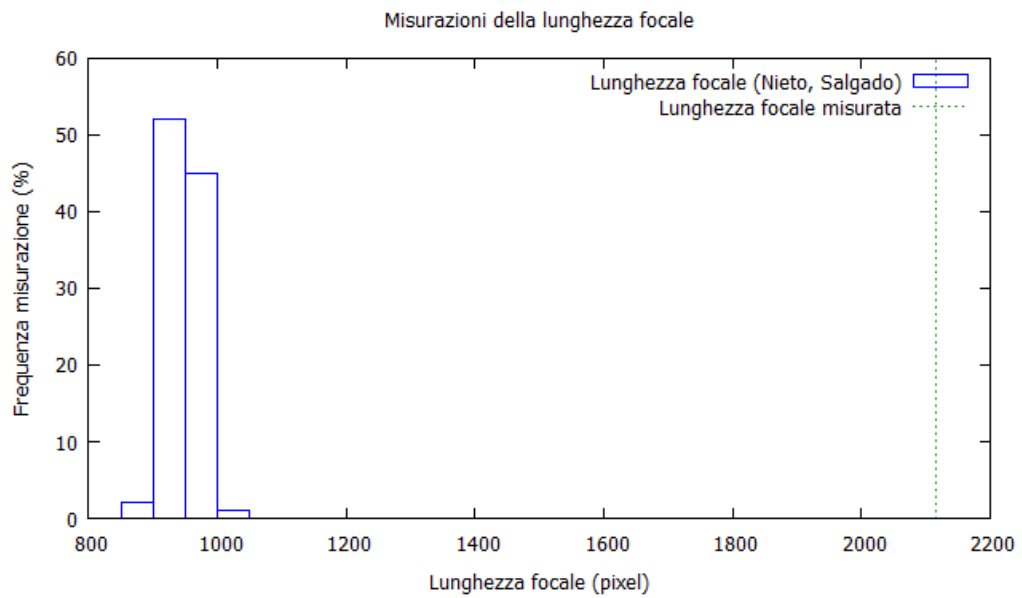
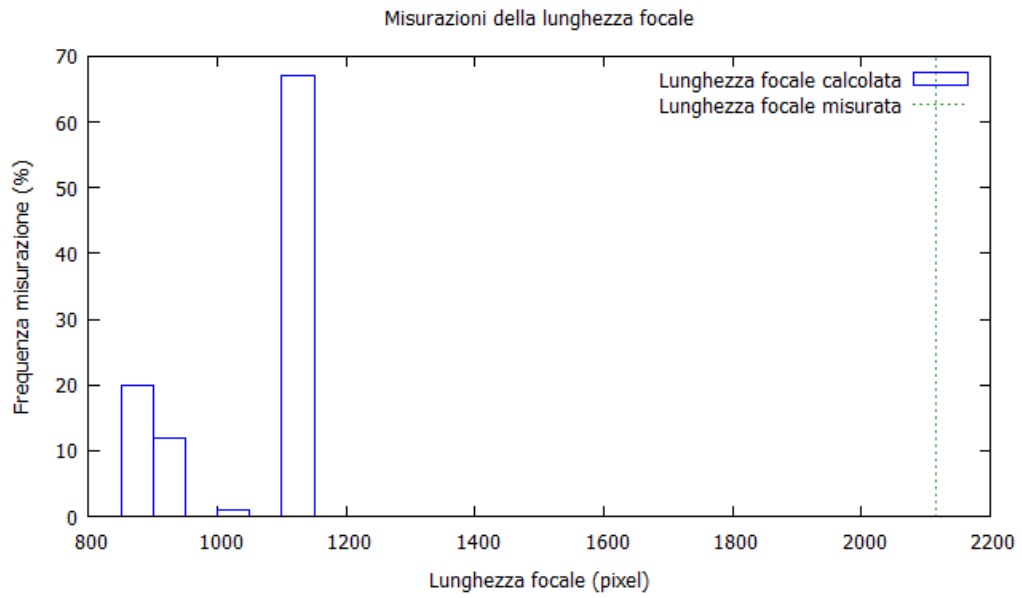


Immagine n°3

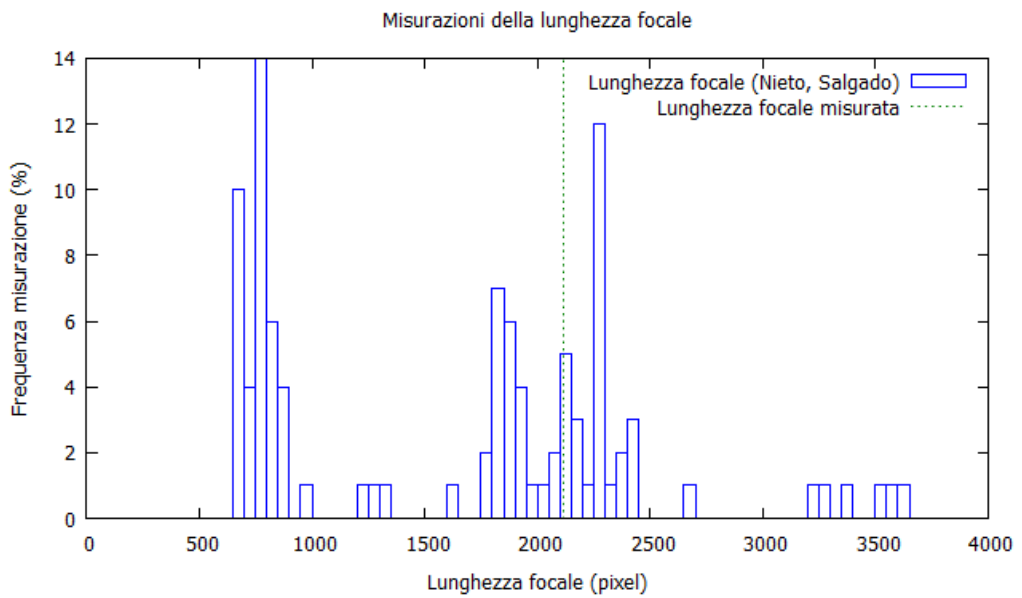
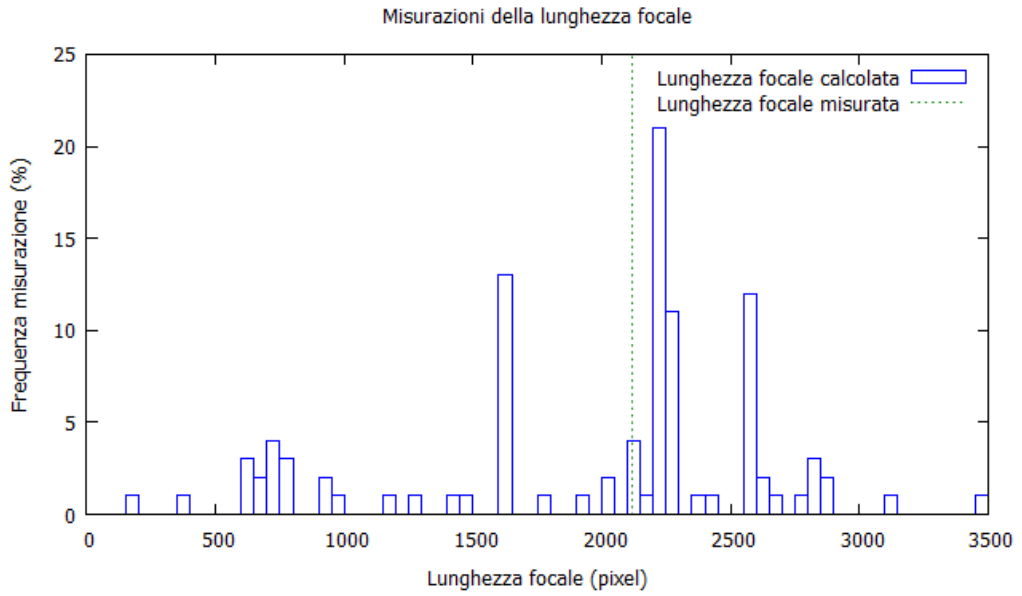


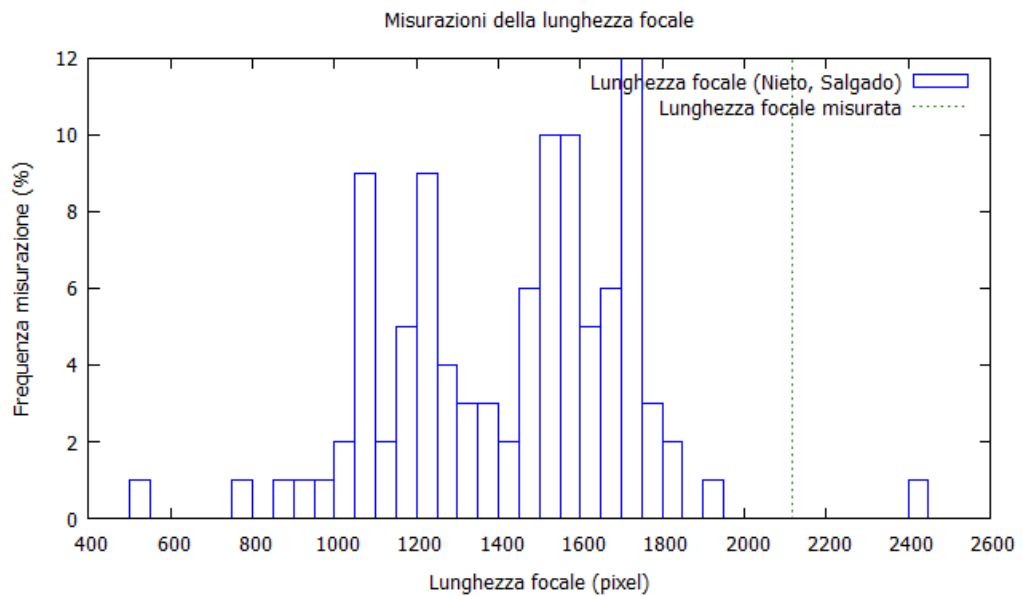
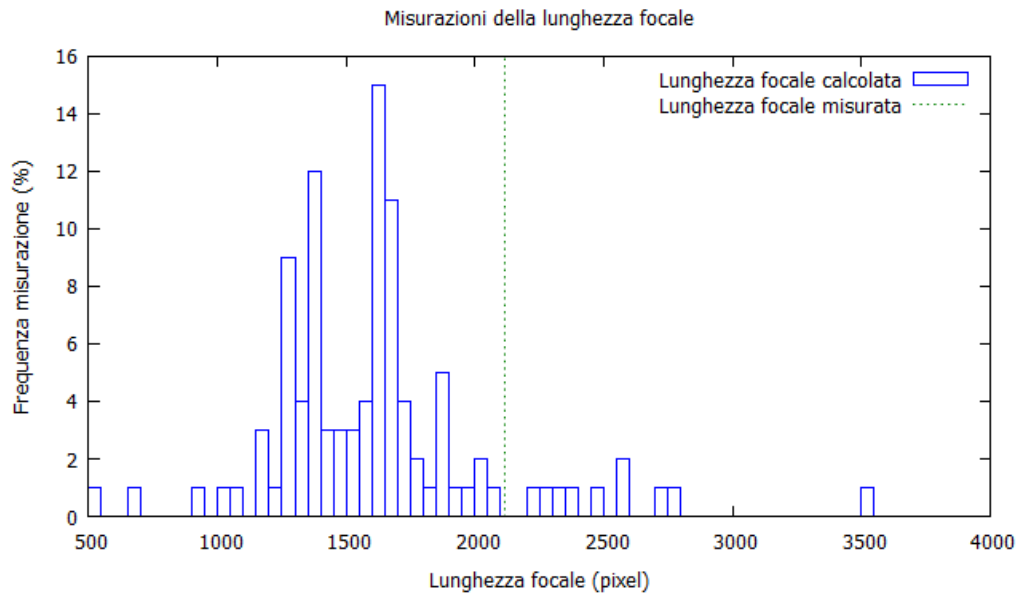
Immagine $n^{\circ}4$ 

Immagine n°5

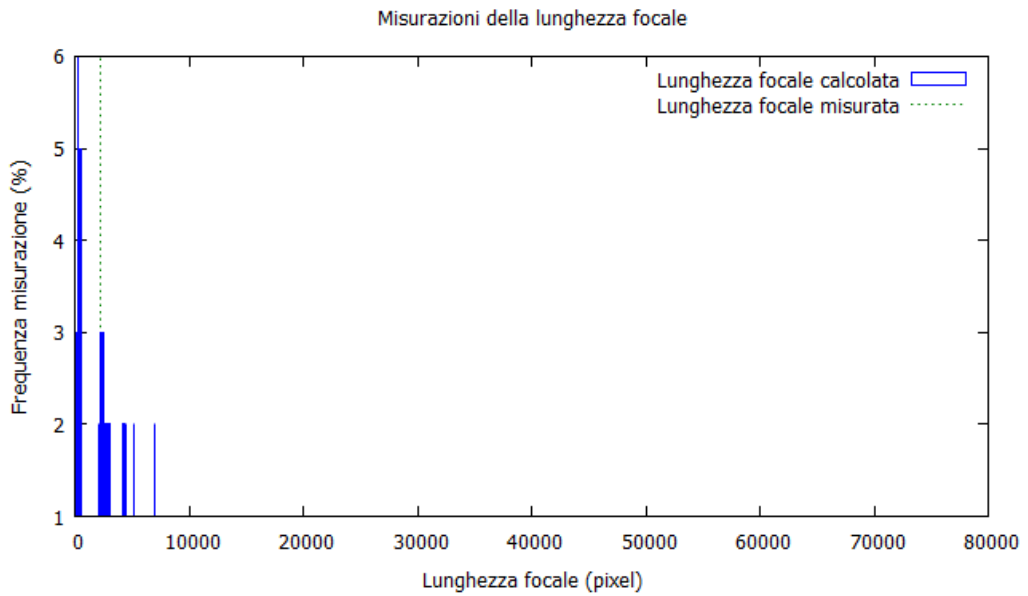
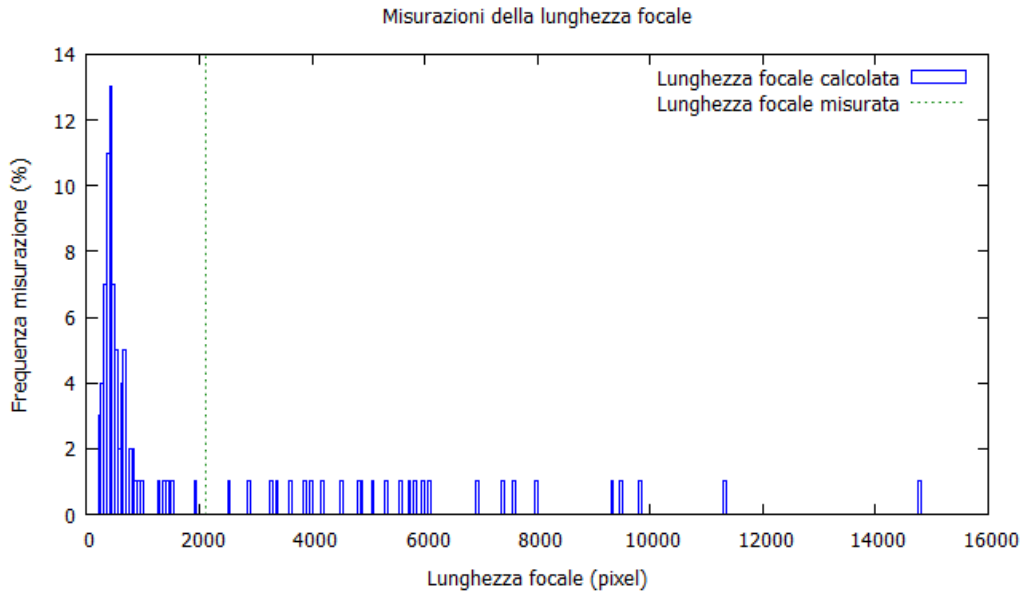


Immagine n°6

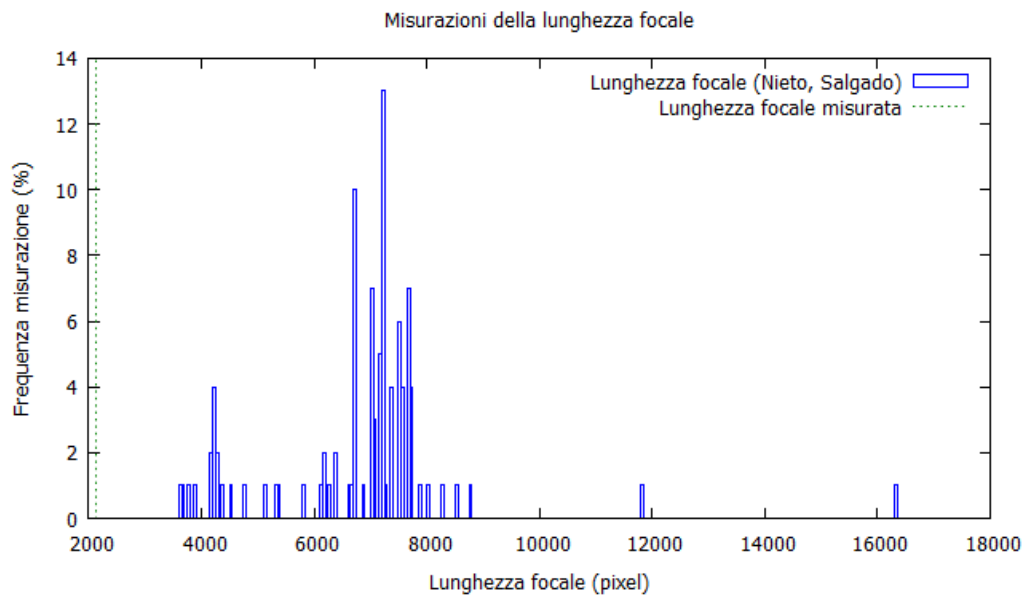
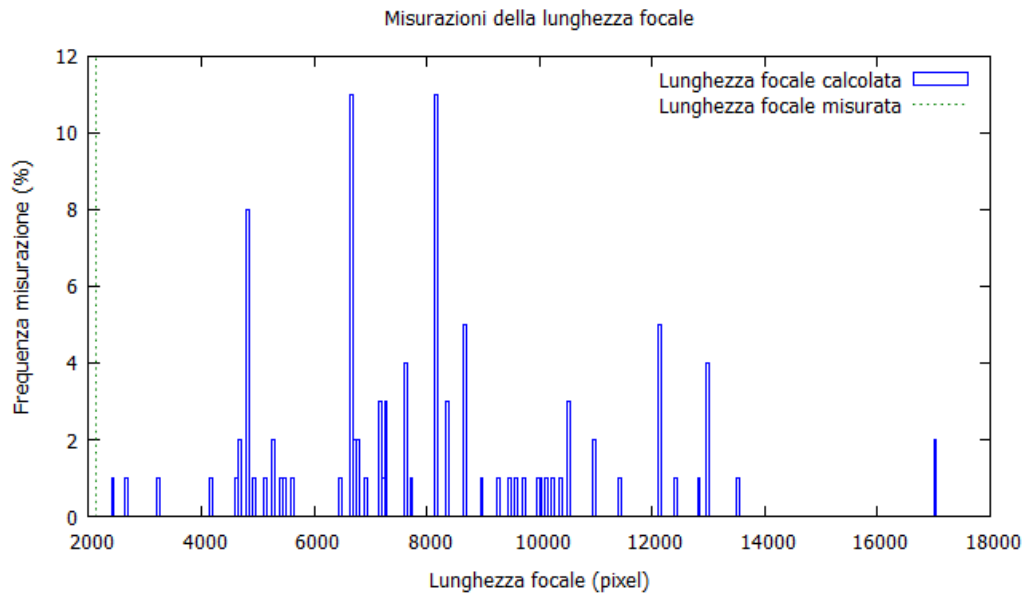


Immagine n°7

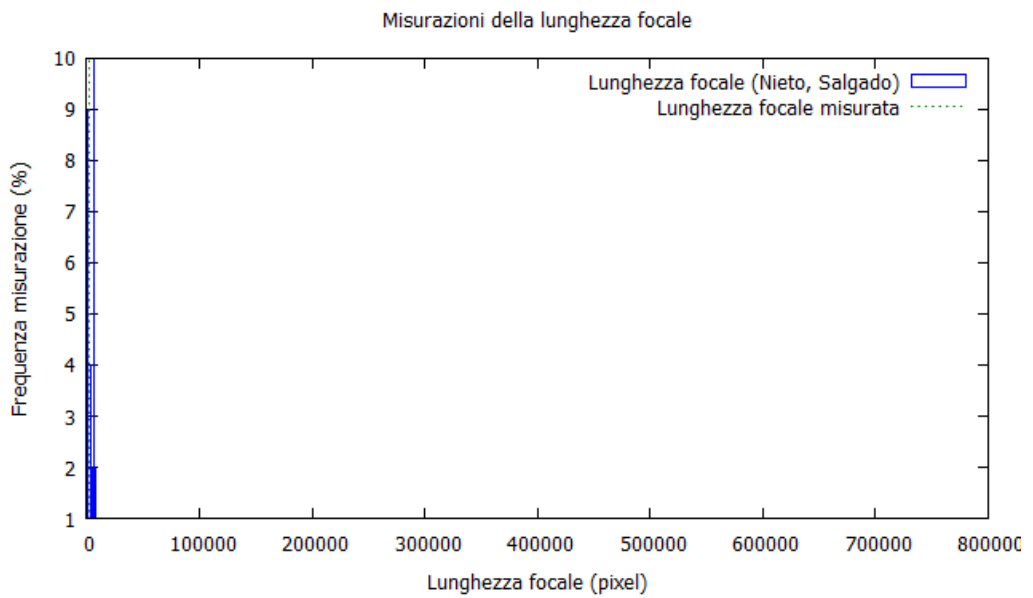
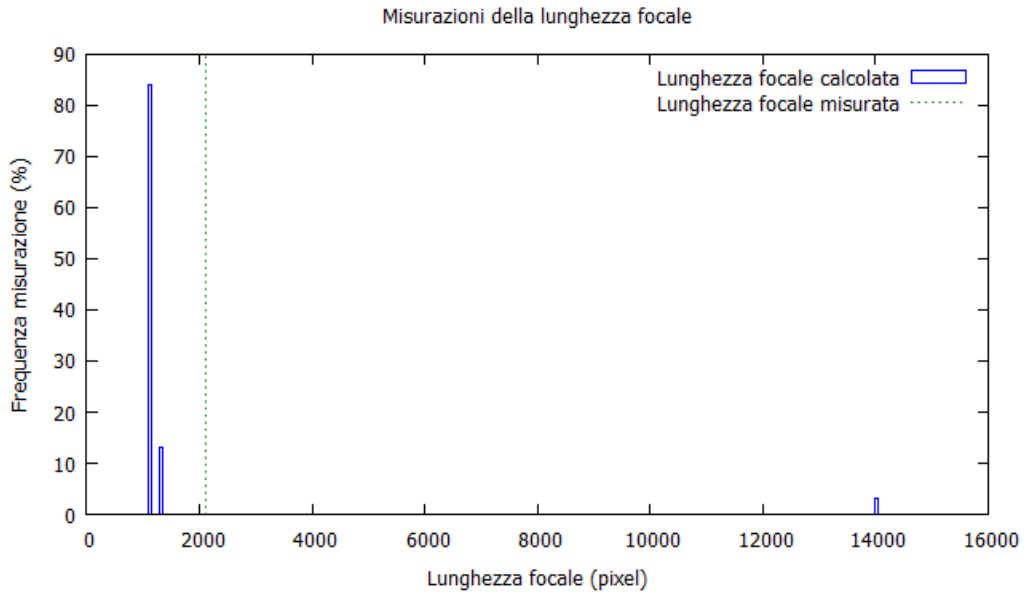


Immagine n°8

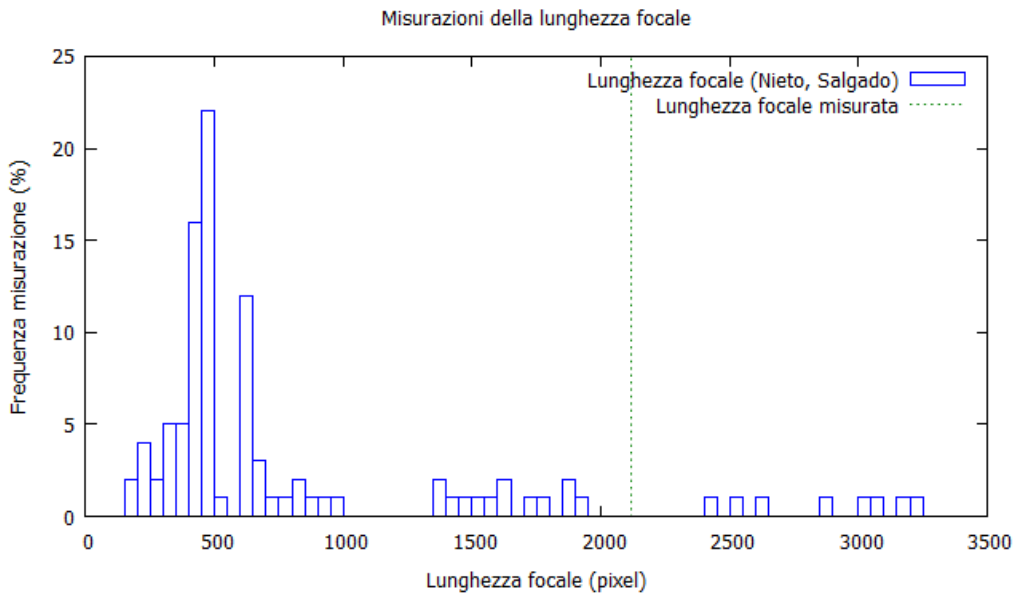
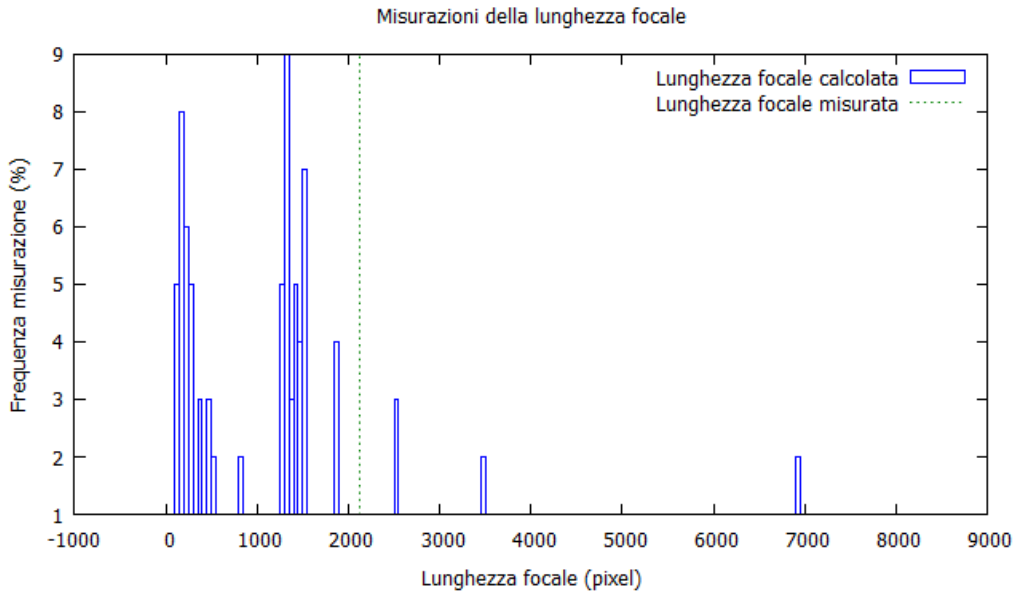


Immagine n°9

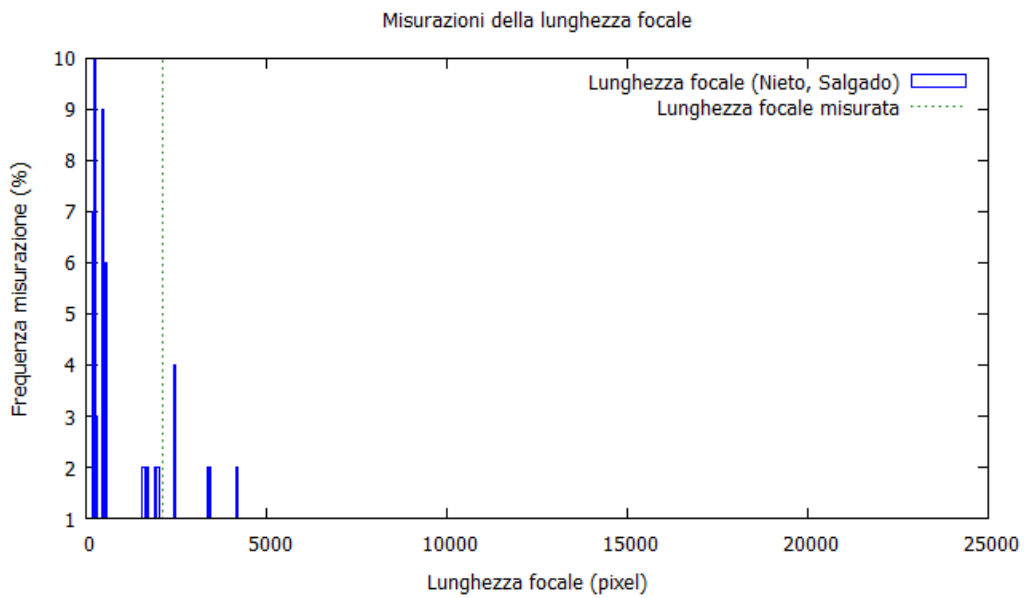
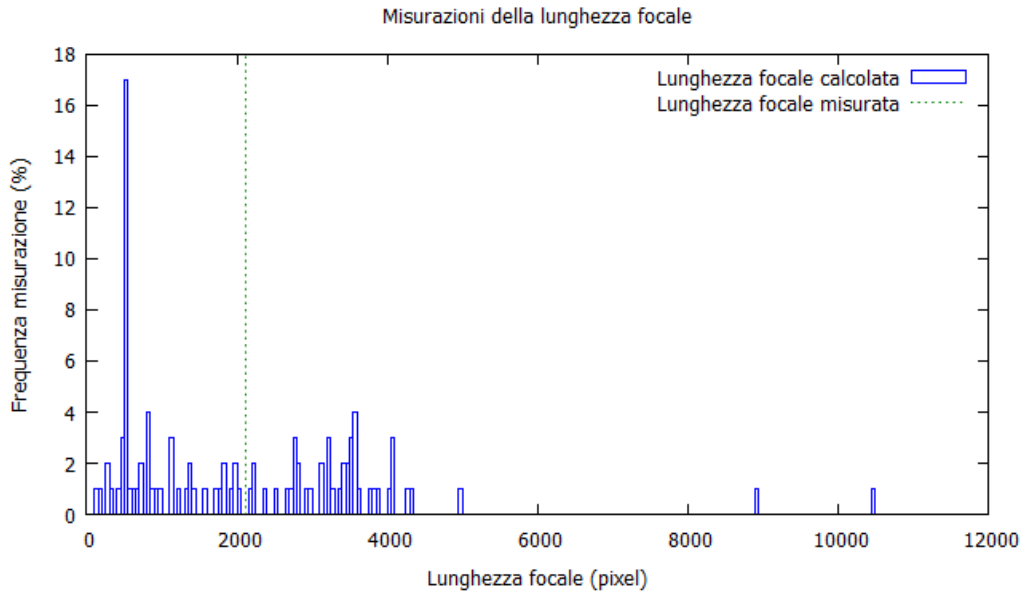
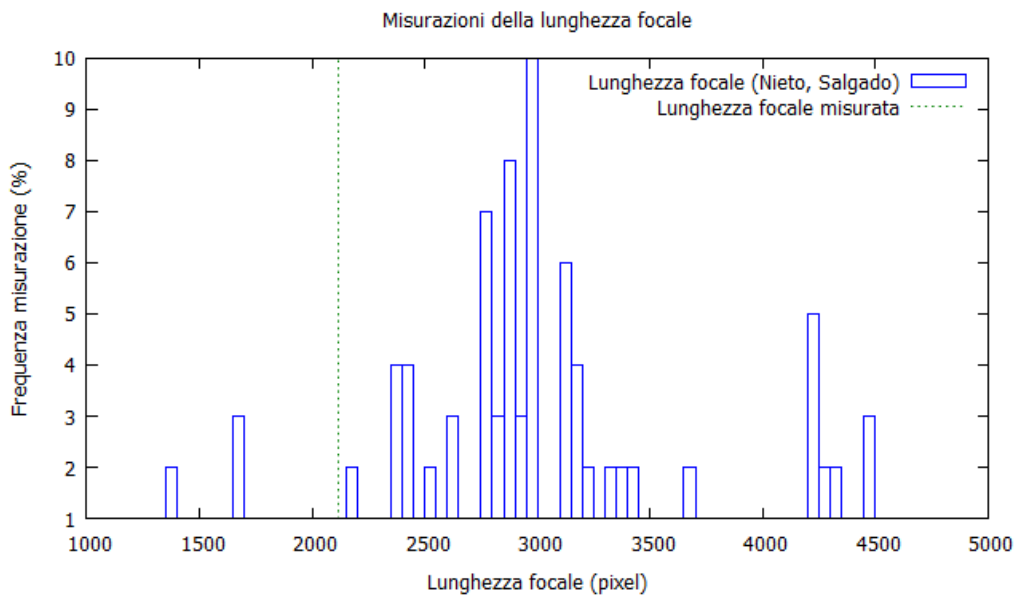
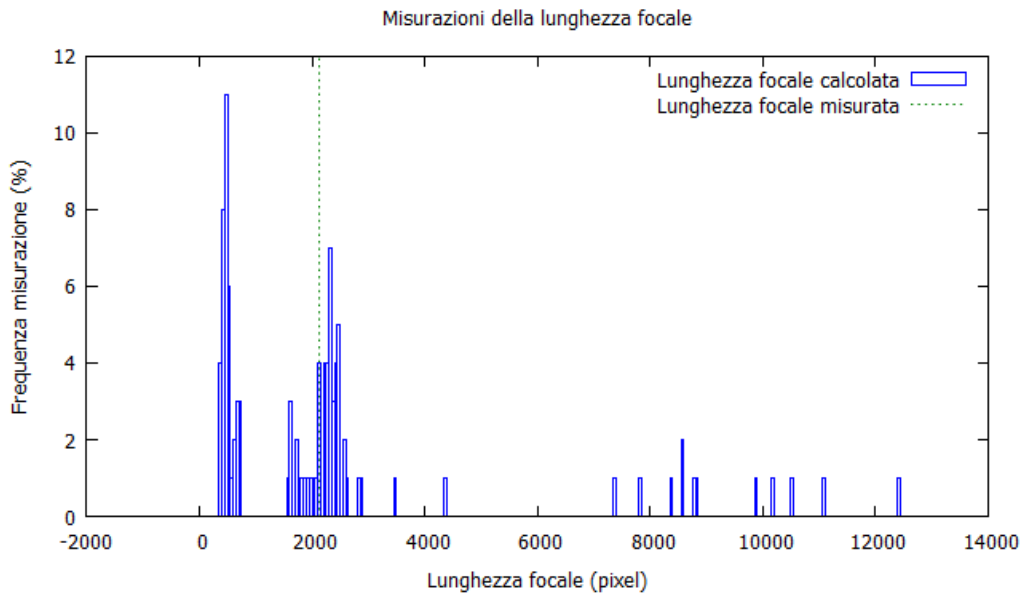


Immagine n°10



Comparazione dei due approcci

Per poter confrontare i due metodi concretamente, nei confronti di ogni immagine, è necessario stabilire una forma di *punteggio* da assegnare ad ogni test, ossia un valore che riassume l'intero campione di dati ottenuti. Per avere un numero significativo, viene adottato il valore mediano registrato nei test.

In figura 3.43 si può vedere l'esito di relativo confronto. Calcolando lo scarto di ciascun test (figura 3.44) dal valore di lunghezza focale ottenuto con la calibrazione standard, si ottiene lo scarto totale di entrambi. Questo risultato viene sintetizzato nella figura 3.45. Confrontando tali valori, risulta che nei test l'approccio proposto si dimostrato più preciso del 29.8% rispetto all'approccio MSAC di Nieto e Salgado.

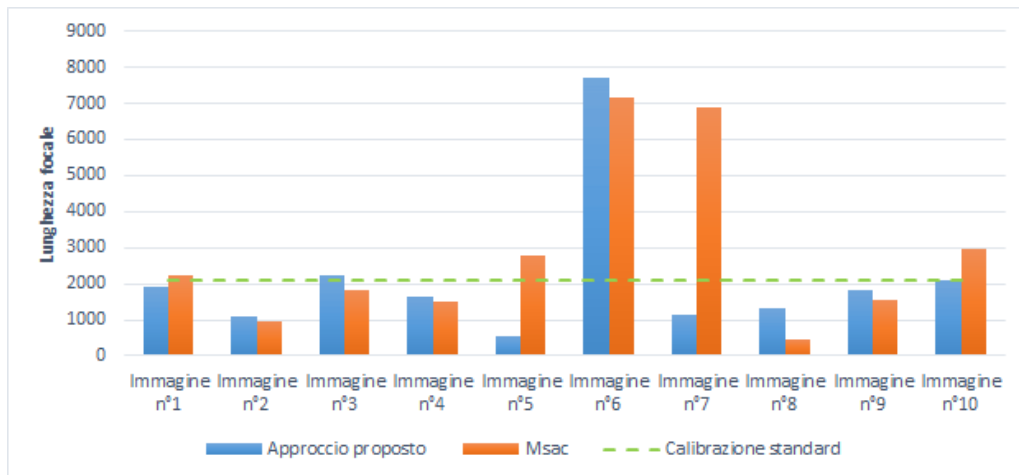


Figura 3.43: Confronto tra l'approccio proposto e MSAC: per ogni test viene vengono confrontate le due metodologie in relazione al valore di lunghezza focale ottenuto con la calibrazione standard

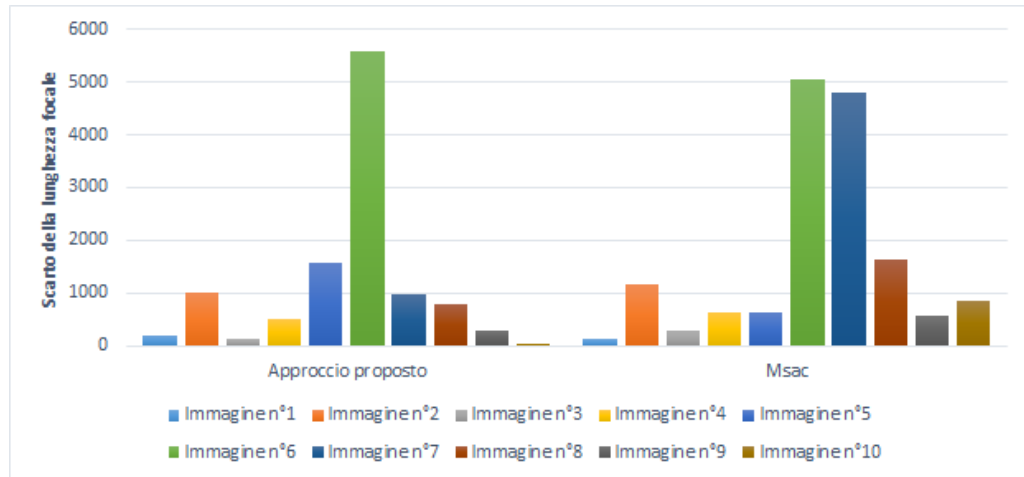


Figura 3.44: Scarto della lunghezza focale per ogni test, rispetto al valore di riferimento, tra l'approccio proposto e MSAC

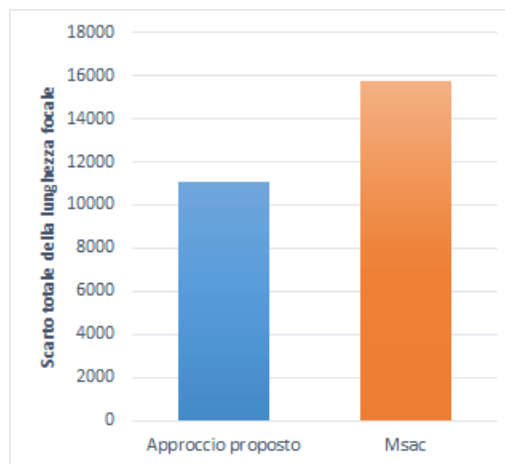


Figura 3.45: Scarto totale della lunghezza focale rispetto al valore di riferimento, tra l'approccio proposto e MSAC: minore è lo scarto, migliori sono complessivamente le prestazioni dell'approccio.

Capitolo 4

Conclusioni

È stata presentata una nuova metodologia che, a fronte di alcune deboli assunzioni sui probabili parametri interni della telecamera e l'ortogonalità di alcuni elementi della scena, permette l'autocalibrazione delle telecamere utilizzando una singola vista. Per consentire l'impiego in contesti di tempo reale è stata progettata in modo completamente modulare, per permettere maggior flessibilità e consentire maggiore scalabilità verso approcci computazionalmente meno onerosi. La tecnica proposta permette di migliorare le prestazioni dei lavori antecedenti, attraverso un nuovo approccio per stimare i punti di fuga e delle euristiche per determinare la terna di punti di fuga ortogonali.

L'algoritmo può essere applicato senza l'intervento di un operatore, usando parametri fissati ed è in grado di funzionare su una vasta varietà di immagini e geometrie della telecamera. I risultati sperimentali sono incoraggianti e dimostrano la bontà dell'approccio scelto. Anche con fotografie scattate in condizioni difficili o che presentano criticità di diversa natura (ragionevolmente più ostiche di quelle incontrate nella maggioranza dei casi di riferimento), i risultati tendono al valore di riferimento, restituito dagli approcci tradizionali tramite oggetto di calibrazione.

Risultati migliori possono essere ottenuti focalizzandosi sugli aspetti che si è scelto di non approfondire, quali la rilevazione e correzione di effetti distorsivi (come nella pubblicazione di Melo[4]) e la correzione dei segmenti rilevati. Un'altro aspetto centrale è la stima del valore di sogliatura che può condizionare tutto l'approccio. Si può pensare di anteporre alla fase di sogliatura una fase di *pre-analisi* in grado di identificare il sogliatore migliore

da usare per l'immagine in esame. Per la fase di determinazione dei punti di fuga si può pensare di introdurre un approccio più sofisticato che penalizzi i segmenti più corti, in quanto più sensibili al rumore, e valorizzi quelli più lunghi. Un'altra strada da perseguire è lo studio di nuove euristiche più efficaci per favorire o meno le terne di punti di fuga. Per fare questo si può pensare di introdurre un concetto di pattern, in modo da favorire (sulla base dell'esperienza) certe configurazioni di punti di fuga. Inoltre può essere utile identificare il concetto di *orizzonte* nell'immagine e utilizzare questa informazione per scartare i punti di fuga che non sono compatibili con questa informazione.

L'obiettivo può considerarsi raggiunto e l'algoritmo può considerarsi pronto per essere inserito all'interno della libreria opensource di destinazione. Può essere interessante cercare di estendere l'algoritmo di calibrazione a scene ancora più generali. L'assunzione che vi siano tre direzioni ortogonali principali può essere rimossa, o ridefinita[1], per coprire un numero arbitrario di direzioni e riuscire così a rendere adattabile la metodologia ad una casistica più vasta, che comprenda anche oggetti non rettangolari o orientati diversamente.

Bibliografia

- [1] M. Antunes and J. P. Barreto. A global approach for the detection of vanishing points and mutually orthogonal vanishing directions. In *CVPR*, pages 1336–1343. IEEE, 2013.
- [2] J. Deutscher, M. Isard, and J. MacCormick. Automatic camera calibration from a single manhattan image. In *Proceedings of the 7th European Conference on Computer Vision-Part IV, ECCV '02*, pages 175–205, London, UK, UK, 2002. Springer-Verlag.
- [3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [4] R. Melo, M. Antunes, J. Barreto, G. Falcao, and N. Goncalves. Un-supervised intrinsic calibration from a single frame using a plumb-line approach. *Computer Vision, IEEE International Conference on*, 0:537–544, 2013.
- [5] M. Nieto and L. Salgado. Non-linear optimization for robust estimation of vanishing points. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1885–1888, Sept 2010.
- [6] C. Rother. A new approach for vanishing point detection in architectural environments. In *In Proc. 11th British Machine Vision Conference*, pages 382–391, 2000.
- [7] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1250–1257, Sept 2009.

- [8] R. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(4):722–732, April 2010.
- [9] Y. Xu, S. Oh, and A. Hoogs. A minimum error vanishing point detection approach for uncalibrated monocular images of man-made environments. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1376–1383, June 2013.

Elenco delle figure

1.1	Rappresentazione schematica del modello Pinhole	2
1.2	Geometria della telecamera pinhole	3
1.3	Relazione tra coordinate immagine coordinate della telecamera	5
1.4	Conversione da coordinate globali a coordinate della telecamera	6
1.5	Effetto della distorsione radiale	8
1.6	Corrispondenza tra punti della scena e punti immagine . . .	10
1.7	Formazione di un punto di fuga	11
1.8	Immagine di punti della scena equidistanti	12
1.9	Punto di fuga di una linea	13
1.10	I tre punti di fuga e le linee di fuga di un cubo	14
1.11	Immagine di linee della scena equidistanti	14
1.12	Formazione della linea di fuga	15
1.13	L'angolo θ tra i due raggi	17
1.14	Costruzione geometrica del punto principale	19
1.15	Costruzione geometrica della lunghezza focale	20
2.1	Le fasi dell'algoritmo di autocalibrazione	22
2.2	Immagine utilizzata per il caso di studio	23
2.3	Schema dell'estrazione dei segmenti	24
2.4	Rilevamento dei bordi	26
2.5	Estrazione dei segmenti	27
2.6	Vantaggi del raffinamento dei segmenti	28
2.7	Effetto della distorsione radiale nel riconoscimento dei segmenti	30
2.8	Schema della clusterizzazione dei segmenti	32
2.9	Esempi di misure si consistenza	33
2.10	Modello di misura di consistenza usato	34
2.11	Modello di misura di consistenza approssimato	34

2.12	Clusterizzazione dei segmenti	36
2.13	Relazione tra il numero di segmenti e punti di fuga potenziali	38
2.14	Determinazione dei punti di fuga	40
2.15	Individuazione dei punti di fuga	41
2.16	Errore dell'intersezione dei segmenti	42
2.17	Perturbazione causata dai segmenti mal riconosciuti	45
2.18	Raffinamento EM	48
2.19	Determinazione della matrice di calibrazione	49
2.20	Determinazione dei punti di fuga ortogonali	51
2.21	Direzione dei punti di fuga	54
2.22	Esempi di direzioni di punti di fuga	54
2.23	Esempio di bisettrici di punti di fuga	55
3.1	Strumenti usati per la calibrazione con scacchiera	58
3.2	Immagine di test n°1	61
3.3	Immagine di test n°2	62
3.4	Immagine di test n°3	62
3.5	Immagine di test n°4	63
3.6	Immagine di test n°5	64
3.7	Immagine di test n°6	65
3.8	Immagine di test n°7	65
3.9	Immagine di test n°8	66
3.10	Immagine di test n°9	67
3.11	Immagine di test n°10	67
3.12	Misurazione intrinseca della qualità con immagine n°1	69
3.13	Misurazione intrinseca della qualità con immagine n°2	70
3.14	Misurazione intrinseca della qualità con immagine n°3	71
3.15	Misurazione intrinseca della qualità con immagine n°4	72
3.16	Misurazione intrinseca della qualità con immagine n°5	73
3.17	Misurazione intrinseca della qualità con immagine n°6	74
3.18	Misurazione intrinseca della qualità con immagine n°7	75
3.19	Misurazione intrinseca della qualità con immagine n°8	76
3.20	Misurazione intrinseca della qualità con immagine n°9	77
3.21	Misurazione intrinseca della qualità con immagine n°10	78
3.22	Misurazione statistica della qualità con immagine n°1	79
3.23	Misurazione statistica della qualità con immagine n°2	80
3.24	Misurazione statistica della qualità con immagine n°3	81

3.25	Misurazione statistica della qualità con immagine $n^{\circ}4$	82
3.26	Misurazione statistica della qualità con immagine $n^{\circ}5$	83
3.27	Misurazione statistica della qualità con immagine $n^{\circ}6$	84
3.28	Misurazione statistica della qualità con immagine $n^{\circ}7$	85
3.29	Misurazione statistica della qualità con immagine $n^{\circ}8$	86
3.30	Misurazione statistica della qualità con immagine $n^{\circ}9$	87
3.31	Misurazione statistica della qualità con immagine $n^{\circ}10$	88
3.32	Confronto con approccio di Xu, Oh e Hoogs	89
3.33	Confronto con approccio di Nieto e Salgado $n^{\circ}1$	91
3.34	Confronto con approccio di Nieto e Salgado $n^{\circ}2$	92
3.35	Confronto con approccio di Nieto e Salgado $n^{\circ}3$	93
3.36	Confronto con approccio di Nieto e Salgado $n^{\circ}4$	94
3.37	Confronto con approccio di Nieto e Salgado $n^{\circ}5$	95
3.38	Confronto con approccio di Nieto e Salgado $n^{\circ}6$	96
3.39	Confronto con approccio di Nieto e Salgado $n^{\circ}7$	97
3.40	Confronto con approccio di Nieto e Salgado $n^{\circ}8$	98
3.41	Confronto con approccio di Nieto e Salgado $n^{\circ}9$	99
3.42	Confronto con approccio di Nieto e Salgado $n^{\circ}10$	100
3.43	Confronto tra l'approccio proposto e MSAC	101
3.44	Scarto tra l'approccio proposto e MSAC	102
3.45	Scarto totale tra l'approccio proposto e MSAC	102