

**Dottorato di ricerca in Informatica e Matematica**  
**XXIX ciclo**

**Progetto di ricerca**

**Dottorando:** Dott. Michele Giuliano Fiorentino

**Tutor:** Dott.ssa Eleonora Faggiano

**Coordinatore**

Prof. Donato Malerba

Firma del dottorando \_\_\_\_\_

Firma del tutor \_\_\_\_\_

## **1) Titolo della ricerca:**

Sviluppo e confronto di tecniche geometriche di calibrazione di una camera: verso una generalizzazione.

## **2) Area nella quale si inquadra la ricerca:**

Geometria

## **3) Obiettivi della ricerca**

La ricerca si sviluppa nell'ambito degli studi riguardanti le tecniche di ricostruzione delle immagini per le quali è fondamentale una buona calibrazione dello strumento di acquisizione (camera).

In tale settore sono state sviluppate varie tecniche di calibrazione che sfruttano le proprietà dei punti all'infinito (vanishing points). Gli approcci per la calibrazione di una camera sono generalmente classificati in diverse categorie, che dipendono dalla conoscenza iniziale della scena registrata (oggetti geometrici conosciuti o meno) e dal tipo di oggetti, ossia oggetti 3D, modelli piani o oggetti uno-dimensionali.

L'obiettivo della ricerca è quello di cercare una generalizzazione delle tecniche di calibrazione di una camera, che possa consentire di scegliere opportunamente la tecnica più adeguata a seconda delle tipologie di oggetti che ci si trova ad acquisire (3D, 2D o 1D). Inoltre si vuole ampliare l'analisi delle tecniche di calibrazione che utilizzano 2 o 3 vanishing points da una singola vista o da una doppia vista, al caso in cui le viste siano superiori e generalizzarle al caso n-dimensionale.

La geometria, in particolar modo la geometria proiettiva, fornisce una base teorica consistente per affrontare queste problematiche.

## **4) Motivazioni della ricerca**

Negli ultimi quarant'anni, nell'ambito della Computer Vision sono stati fatti enormi progressi per quanto riguarda tutte quelle tecniche di visione artificiale o computazionale volte a riprodurre le abilità della visione umana in dispositivi elettronici per l'acquisizione e la comprensione delle immagini. I campi di applicazione sono soprattutto ingegneristici e trovano particolare importanza nelle interazioni uomo-macchina e modellazione di oggetti.

Sin dagli anni '80 c'è stato un crescente interesse nella comunità che si occupava del riconoscimento degli oggetti, per muoversi al di là del semplice approccio tramite modellazione manuale, a favore di una sempre più invadente automatizzazione dell'acquisizione. In questo si è sempre cercato di puntare notevolmente alla ricerca degli invarianti geometrici. È proprio in parallelo a questo sviluppo che nel 1992 Richard Hartley e Oliver Faugueras iniziarono ad applicare la teoria della geometria proiettiva alle relazioni tra le visioni prospettiche multiple.

In letteratura (sezione 5.5) esistono generalizzazioni del numero di punti di vista che intervengono durante l'acquisizione di immagini. Questa ricerca nasce invece dall'esigenza di generalizzare le tecniche di calibrazione utilizzate e di ampliare i punti di vista al caso n-dimensionale, cogliendo e analizzando parallelamente aspetti positivi e negativi delle tecniche di calibrazione differenti su citate.

## **5) Stato dell'arte**

## 5.1 Geometria e Computer Vision

I temi di ricerca in computer vision legati strettamente alla geometria sono molteplici: storicamente sono legati a calibrazione di una o più camere, ed hanno portato allo sviluppo di tecniche per l'acquisizione di dati 3D da camera come structure-from-motion, visione stereo e multi-view reconstruction.

Una camera è uno strumento di misura estremamente utile, che non produce solo una immagine realistica della scena, ma fornisce anche delle informazioni dalle quali possono essere ricavate importanti proprietà geometriche. Una delle aree più attive nella Computer Vision è la ricostruzione della scena. Quando le camere vengono utilizzate in applicazioni che richiedono una ricostruzione degli oggetti o un'interazione col mondo esterno, inoltre, è importante che queste siano opportunamente calibrate. La problematica della calibrazione di una camera, rappresenta uno degli aspetti notevoli e fondamentali, nel campo della ricerca in Computer Vision. Una delle problematiche maggiormente affrontate è quella che riguarda la combinazione di vincoli derivanti dalla calibrazione di una camera, per la ricostruzione della scena geometrica da una vista singola o multipla [1]. I tipi di vincoli riguardanti la calibrazione sono essenzialmente di tre tipi: vincoli della scena (ad esempio il parallelismo e la perpendicolarità tra rette), vincoli derivanti dalla parziale conoscenza delle caratteristiche di una camera e vincoli derivanti dal movimento reciproco tra viste differenti.

La geometria differenziale, invece, viene ampiamente utilizzata per tematiche di computer vision 3D, ovvero per le superfici: alcune tematiche in questo caso sono legate a matching tra superfici 3D, segmentazione e classificazione di nuvole di punti e mesh, 3D retrieval.

Notevole importanza ricopre la geometria anche nel caso in cui la computer vision venga applicata alla robotica: si utilizzano tecniche geometriche per camera pose estimation in real-time (dove la camera è, in questo caso, il sensore di un agente autonomo che naviga un ambiente). Parliamo, in tal caso, di SLAM (Semantic Localization and Mapping), Kinect Fusion, etc..

Infine tecniche geometriche sono adoperate in risvolti più applicativi, ad esempio per il fitting di primitive geometriche su nuvole di punti, per segmentazione di nuvole di punti (uno dei metodi più adoperati al momento è basato su NURBS, Non-uniform rational B-spline).

## 5.2 Le basi storiche della Computer Vision

La geometria proiettiva è stata utilizzata nella computer vision sin dal principio. Per esempio nel 1965 Roberts [2] nella sua tesi di dottorato ha usato la matrice di trasformazione proiettiva dal 3D al 2D per rappresentare la visualizzazione 3-dimensionale dei poliedri. Ha sviluppato un algoritmo lineare per determinare i parametri matriciali ed è stato in grado di visualizzare oggetti immersi nello spazio 3-dimensionale su piani (2-dimensionali) con l'uso della geometria proiettiva.

Nel corso degli anni, la proiezione prospettica è stata analizzata in molti modi differenti: la maggior parte degli studi parte dalle equazioni della proiezione centrale e quindi continua con alcune semplificazioni e nuovi gruppi di parametri che costituiscono il risultato teorico principale. Un esempio esplicativo è il problema della determinazione del minimo numero di punti di corrispondenza necessari per calcolare le trasformazioni tra due viste prospettiche dello stesso insieme di punti contenuti nello spazio 3-dimensionale.

## 5.3 Il background: Geometria proiettiva e trasformazioni

Un'interessante introduzione al piano geometrico proiettivo per la computer vision e quindi utile ai nostri scopi è in Mundy e Zisserman [3]. Qui vengono sviluppati i concetti di base e fondamentali per i passi successivi. Un approccio più formale è invece quello di Semple e Kneebone [4].

Una descrizione delle proprietà ed entità dello spazio proiettivo  $P_3$  si deve a Hartley e Zisserman [5]: alcune di queste proprietà non sono altro che una generalizzazione di quelle del piano proiettivo (ad esempio le rette parallele che si incontrano ad  $l_\infty$  qui sono i piani paralleli che si intersecano a  $\pi_\infty$ ). Anche per questa parte è rilevante il lavoro di Mundy e Zisserman, nonché quello precedente di Hilbert e Cohn-Vossen [6] per alcune proprietà dettagliate di curve e superfici.

#### 5.4 Camera Geometry e Single View Geometry

Uno strumento di acquisizione (ad esempio una fotocamera) è una funzione tra il mondo reale 3D (spazio degli oggetti) e un'immagine 2D: esso può essere pensato come una funzione,

$$X = (X, Y, Z)^T \mapsto x = (x, y)^T$$

Ci sono diverse tipologie di strumenti e quelli probabilmente più interessanti sono quelli a proiezione centrale. Le caratteristiche di questa tipologia di strumenti sono esaminate con i risultati della geometria proiettiva: le entità geometriche (ad esempio il centro di proiezione e il piano dell'immagine) possono essere dedotte molto facilmente a partire dalla matrice di trasformazione.

Una gerarchia di queste tipologie è delineata in Aloimonos [7], invece Mundy e Zisserman [3] generalizzano alcune proprietà negli strumenti di acquisizione di tipo affine (dove la matrice di trasformazione ha delle proprietà particolari). Faugeras invece sviluppa le proprietà degli strumenti di acquisizione di tipo proiettivo in [8].

Si sono studiate le caratteristiche nella proiezione di punti e rette: Faugeras e Mourrain [9], e Faugeras e Papadopoulo [10] sviluppano la proiezione delle rette utilizzando le coordinate di Plücker; Koenderink [11], [12] e Giblin e Weiss [13] forniscono molte proprietà sulla generazione del contorno e sul contorno apparente, e la loro relazione con la geometria differenziale delle superfici.

#### 5.5 Two-View Geometry

Nella two-view geometry si ampliano i punti di vista (l'obiettivo è quello di raggiungere l'n-view geometry): le due viste possono essere ottenute simultaneamente come in un impianto stereo o acquisite sequenzialmente, per esempio con uno strumento di acquisizione che si muove relativamente alla scena. Queste due situazioni sono geometricamente equivalenti. La geometria epipolare è la geometria intrinseca tra le due view, è indipendente dalla struttura della scena e dipende soltanto dai parametri interni dello strumento di acquisizione: la geometria epipolare è la cosiddetta geometria della visione stereoscopica. Essa si preoccupa di analizzare i vincoli e le relazioni geometriche che collegano due immagini 2-dimensionali della stessa scena 3-dimensionale, catturata da due strumenti di acquisizione con posizione e orientamento differenti. Anche questa geometria intrinseca può essere sviluppata con l'utilizzo di opportune matrici di trasformazione [14] [15] [16] [17].

#### 5.6 Three-view ed n-view geometry

Con l'introduzione dei tensori trifocali (aventi proprietà analoghe a quelle delle matrici di trasformazione della single view e della 2-view geometry) si dà l'avvio ad una nuova geometria. La scoperta dei tensori trifocali può essere attribuita a [18] e [19] i quali li utilizzano per la ricostruzione della scena in caso di strumenti di acquisizione calibrati. Più tardi è stato mostrato che queste caratteristiche potevano essere applicate anche nel caso di strumenti non calibrati [20].

Il passaggio all'n-view geometry è naturale ed è una generalizzazione dei casi precedenti.

Uno dei risultati più importanti della n-view geometry non calibrata è quello della ricostruzione della struttura della scena esclusivamente a partire dalle corrispondenze punto-immagine. I metodi per la ricostruzione affine a partire dalla traslazione sono apparsi in Moons et al. [21].

## 6) Approccio al problema

Si studieranno dapprima a fondo le varie tecniche di calibrazione di una camera, cogliendone gli aspetti caratterizzanti, analizzandone le differenze e mettendo in risalto aspetti positivi o negativi delle diverse tecniche tra loro. In seguito verranno analizzati esempi di studi ottenuti con le varie tecniche di calibrazione, cogliendone le principali criticità. Infine si cercherà un filo comune che leghi le varie tecniche.

In parallelo a questo studio verranno analizzate le piattaforme, i software e gli strumenti che vengono adoperati per l'acquisizione delle immagini (CVLab dell'Università degli Studi di Bologna).

## 7) Ricadute applicative

Alcuni degli obiettivi della computer vision sono quelli di costruire sistemi capaci di prendere decisioni a partire da una descrizione della scena estrapolata da immagini/video, inferire il mondo 3D a partire da immagini digitali, riconoscere oggetti, scene e contesto a partire da immagini digitali. È per questo motivo che le tecniche di computer vision sono estremamente utilizzate in vari settori industriali:

- Assistenza al guidatore: sistemi che informano i guidatori sui pericoli, forniscono un cruise control adattivo alla situazione del traffico e forniscono assistenza alla guida;
- Gestione del traffico: Sistemi di lettura targhe, controllo velocità e stato del traffico;
- Film e TV: Sistemi per tracciare oggetti in video per fornire trasmissioni avanzate;
- Recupero di immagini;
- Industria dell'automazione: Sistemi per robot vision-guided nell'industria automobilistica;
- Cibo e agricoltura: Sistemi di visione per monitorare i raccolti;
- Medicina e biomedica: individuazione e tracciamento di marcatori;
- Controllo della sicurezza: sistemi di visione per la videosorveglianza, per il tracciamento degli oggetti in movimento;
- Biometrica: sistemi per il riconoscimento di impronte e del viso;
- Modellizzazione 3-dimensionale;
- Video-Games.

In questo panorama, la ricerca può velocizzare e rendere più adattive le tecniche di computer vision adoperate.

## 8) Riferimenti bibliografici

[1] D. Liebowitz. *Camera calibration and Reconstruction of Geometry from images*. PhD Thesis, University of Oxford, Dept. Engineering Science, June 2001.

- [2] L.G. Roberts, Machine Perception of three-dimensional solids, *Optical and Electro-optical Information Processing*, editor Tippett, J. Et al., MIT Press, pagg. 159-197, 1965
- [3] J. Mundy, A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992
- [4] J. G. Semple, G. T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press. 1979
- [5] R. Hartley, A. Zisserman, Multiple view Geometry in computer Vision second edition, Cambridge University Press, 2003
- [6] D. Hilbert, S. Cohn-Vossen. *Geometry and the Imagination*. Chelsea, NY, 1956.
- [7] J. Y. Aloimonos. Perspective Approximations. *Image and Vision Computing*, 8(3): pagg. 177-192. Agosto 1990
- [8] O. D. Faugeras. *Three-Dimensional Computer Vision: a geometric Viewpoint*. MIT Press, 1993.
- [9] O. D. Faugeras, B. Mourrain. *On the geometry and algebra of point and line correspondences between  $N$  images*. In Proc. International Conference on Computer Vision. Pagg:485-492. Springer-verlag, 1994.
- [10] O.D. Faugeras, T. Papadopoulo. *Grassman-Cayley algebra for modelling systems of cameras and the algebraic equations of the manifold of trifocal tensors*. Technical Report 3225, INRIA, Sophia-Antipolis, Francia, 1997.
- [11] J.J. Koenderink. *What does the occluding contour tell us about solid shape?* *Perception*, 13:321-330, 1984.
- [12] J.J. Koenderink. *Solid Shape*. MIT Press, 1990.
- [13] P. Giblin, R. Weiss. *Reconstruction of surfaces from profiles*. In Proc. 1st International Conference on Computer Vision, London, pagg:136-144. Londra, 1987.
- [14] H.C. Longuet-Higgins. *A computer algorithm, for reconstructing a scene from two projections*. *Nature*, 293:133-135. Settembre 1981.
- [15] T.S. Huang, O.D. Faugeras. *Some properties of the E-matrix in two view motion estimation*. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11, pagg:1310-1312, 1989.
- [16] S. J. Maybank, *Theory of reconstruction from image motion*. Springer-Verlag, Berlin, 1993.
- [17] B. K. P. Horn. *Relative orientation*. *International Journal of Computer Vision*, 4, pagg:59-78, 1990.
- [18] M.E. Spetsakis, J. Aloimonos. *A multi-frame approach to visual motion perception*. *International Journal of Computer Vision*, 16(3), pagg:245-255, 1991.
- [19] J. Weng, N. Ahuja, T.S. Huang. *Closed-form solution and maximum likelihood: A robust approach to motion and structure estimation*. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1988.

[20] R. I. Hartley. *Projective reconstruction from line correspondence*. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1994.

[21] T. Moons, L. Van Gool, M Van Diest, E. Pauwels. *Affine reconstruction from perspective image pairs*. In J. Mundy, A. Zisserman, and D. Forsyth, editors, Application of Invariance in Computer Vision, LNCS 825, Springer-Verlag, 1994.

[22] L. Grammatikopoulos, G. Karrasa, E. Petsa. *Camera calibration combining images with two vanishing points*, Int. Archives of the Photogrammetry, Remote Sensing and Partial Information Sciences, XXXV-5, pagg:99-104

[23] R. Orghidan, J. Salvi, M. Gordan and B. Orza, *Camera calibration using two or three vanishing points*, Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on Computer Science and Information Systems, 2012, pagg:123-130

## **9) Fasi del progetto**

Il progetto si suddivide nelle seguenti fasi:

FASE 0: Studio approfondito della letteratura esistente

FASE 1: Analisi delle tecniche di calibrazione di una camera

FASE 2: Analisi delle strumentazioni, software e piattaforme

FASE 3: Raccolta di dati dagli strumenti

FASE 4: Confronto delle tecniche di calibrazione

FASE 5: Generalizzazione dei risultati e conclusione della ricerca





## **10) Valutazione dei risultati.**

I risultati della ricerca potranno essere dichiarati raggiunti nel momento in cui si sia trovata una generalizzazione totale o parziale delle varie tecniche di calibrazione. Inoltre i risultati raggiunti durante la ricerca verranno discussi tramite colloqui diretti con gli esperti del settore, anche nelle occasioni di soggiorni presso altre sedi universitarie. Verranno anche contattati referee che si esprimeranno sul percorso della ricerca.

## **11) Eventuali referenti esterni al Dipartimento**

Prof. Federico Tombari, Assistant Professor presso l'Università di Bologna, CVLAB

Prof.ssa Francesca Odone, Assistant Professor presso l'Università degli Studi di Genova