



UNIVERSITAT DE  
BARCELONA

Facultat de Matemàtiques  
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

---

# Geometria multivista aplicada al problema minimal dels 5-punts

---

Autor: Ana María Cortés Hinojosa

Director: Dr. Simone Marchesi

Realitzat a: Departament  
de Matemàtiques i Informàtica

Barcelona, 13 de juny de 2023

## Abstract

Multiview geometry studies the geometric relationships between multiple views or images of a scene leveraging projective geometry. This allows us to do computer vision tasks such as the 3D reconstruction of the scene or the estimation of the position of the cameras that took the images. We will show the importance of the minimal problems and we will study in detail the 5-point algorithm, which offers an efficient way to find the relative positions of two cameras given only 5 corresponding points in two images.

## Resum

La geometria multivista estudia les relacions geomètriques entre diverses vistes o imatges d'una escena mitjançant geometria projectiva. Això ens permet realitzar tasques de *computer vision* com ara la reconstrucció en 3D de l'escena o l'estimació de la posició de les càmeres que han pres les imatges. Veurem la importància dels problemes minimalis i estudiarem en detall l'algoritme dels 5-punts, que ofereix una manera eficient de trobar les posicions relatives de dues càmeres donats només 5 punts corresponents en dues imatges.

## Agraïments

Vull donar les gràcies al Dr. Simone Marchesi per tota la seva ajuda i interès en aquest treball.

A la meva família i a les meves amigues de la universitat, en especial a l'Emma, la Carla i la Mercè, per tot el suport.

# Índex

Introducció . . . . .	1
<b>1 Preliminars</b>	<b>4</b>
1.1 Geometria afí en el marc projectiu . . . . .	4
1.2 Geometria en una vista . . . . .	6
1.3 Geometria en dues vistes . . . . .	12
1.3.1 Correspondència de punts . . . . .	12
1.3.2 Matriu fonamental . . . . .	13
1.3.3 Matriu essencial . . . . .	15
1.3.4 Propietats de la matriu essencial . . . . .	16
1.3.5 Recuperació de les matrius de càmera a partir de la matriu essencial	17
<b>2 Problemes minimal</b>	<b>22</b>
2.1 Problemes minimal . . . . .	22
2.2 El mètode RANSAC . . . . .	23
2.2.1 L'algoritme RANSAC . . . . .	24
<b>3 L'algoritme dels 5-punts</b>	<b>26</b>
3.1 Pas 0: Establiment de les bases de l'algoritme . . . . .	26
3.2 Pas 1: Extracció del subespai nul d'una matriu $5 \times 9$ . . . . .	27
3.2.1 Factorització QR . . . . .	28
3.2.2 Càlcul del subespai nul amb eliminació Gaussiana . . . . .	30
3.3 Pas 2: Expansió de les restriccions cúbiques . . . . .	31
3.3.1 Desenvolupament de les restriccions . . . . .	31
3.4 Pas 3: Eliminació amb Gauss-Jordan a la matriu $A$ $10 \times 20$ . . . . .	32
3.5 Pas 4: Expansió del determinant de la matriu $B$ i càlcul del polinomi de grau 10 . . . . .	33
3.6 Pas 5: Càlcul de les arrels del polinomi de grau 10 . . . . .	35
3.6.1 Successions de Sturm . . . . .	35
3.6.2 Mètode de la bisecció . . . . .	36

3.7	Pas 6; Recuperació de $R$ i $t$ a partir de cada arrel real del polinomi . . . .	37
3.7.1	Descomposició en valors singulars de $E$ . . . . .	37
3.7.2	Triangulació de punts . . . . .	38
3.7.3	Desambigüitat de la reconstrucció . . . . .	40
3.8	Aplicació de RANSAC a l'algorisme dels 5-punts . . . . .	41
<b>4</b>	<b>Conclusions</b>	<b>42</b>

## Introducció

La *Computer Vision* (visió per ordinador) és el camp que estudia com els ordinadors poden analitzar i entendre imatges i vídeos en 2D de manera similar a com ho faria una persona. Les tasques típiques de la computer vision [15] inclouen el reconeixement automàtic d'objectes en imatges, la detecció d'esdeveniments en vídeos i la reconstrucció d'escenes en 3D a partir de moltes imatges en 2D.

Per a poder entendre una escena a partir de les imatges preses ens cal analitzar les relacions geomètriques entre els diferents objectes de l'espai des dels punts de vista donats. Aquesta manera d'estudiar l'espai s'anomena geometria multivista.

En la geometria des d'una vista veurem els elements d'una càmera i els seus paràmetres interns i externs, que determinen la seva geometria. A continuació, ens centrarem en el cas on tenim dues vistes. La idea principal serà veure com podem sintetitzar la informació d'una escena vista per dues càmeres en una matriu que anomenarem matriu fonamental. Veurem a més, que si suposem coneguts alguns paràmetres de les nostres càmeres, en particular el calibratge, podem definir la matriu essencial, una versió de la matriu fonamental que ens permet simplificar els càlculs a l'hora de resoldre problemes de computer vision.

A partir dels conceptes de geometria en dues vistes sorgeixen molts problemes que ens podem plantejar:

- i)* Correspondència de punts: Donades dues imatges preses des de punts de vista diferents, puc identificar-hi els punts que corresponen al mateix punt del món?
- ii)* Geometria de les càmeres: Donat un conjunt de punts corresponents en les dues imatges  $\{x_i \leftrightarrow x'_i\}$  amb  $i = 1, \dots, n$ , quines són les càmeres  $P$  i  $P'$  en la primera i segona vista respectivament? Quants punts són suficients per trobar-les?

Aplicant a un conjunt de punts corresponents en dues vistes les restriccions i definicions de la geometria epipolar (és a dir, en dues vistes) obtenim un sistema de polinomis que anomenem problema minimal. Els anomenem així volent reflectir la idea de que només utilitzem la quantitat "mínima" d'informació per tal que la solució quedi determinada per a, com a molt, un nombre finit de solucions. En general, els problemes minimal són problemes de reconstrucció en 3D.

En el darrer capítol ens centrem en particular en el problema minimal dels 5-punts, on donats cinc punts corresponents d'una mateixa escena vistos des de dues càmeres calibrades diferents, recuperem les possibles posicions de les càmeres.

Els algorismes dels 8-punts (Hartley 1997 [11]), 7-punts [13, Capítol 11] i dels 6-punts [18] són problemes minimal, precursors a l'algoritme dels 5-punts. Es tracta d'un mètode per calcular la posició relativa entre dues càmeres, calibrades o no depenent de l'autor, on ens venen donats 8, 7 i 6 punts corresponents a les dues imatges respectivament. L'objectiu és, com en el problema dels 5-punts, trobar la matriu essencial, a partir de la qual podem deduir la posició relativa i l'orientació de les dues càmeres. A les conclusions, parlarem sobre l'eficiència i els resultats de l'algoritme dels 5-punts comparant-los amb aquests tres algorismes semblants.

## El problema dels 5-punts

La reconstrucció de les posicions de les càmeres i l'estructura de l'escena a partir de múltiples punts de vista ha estat objecte d'estudi al llarg dels últims dos segles, al principi pels experts en fotografia i recentment en computer vision.

En el problema clàssic, els paràmetres interns de la càmera s'assumeixen coneguts. En concret, suposarem que les càmeres són calibrades i trobarem la posició relativa dels punts i les càmeres. L'escala general de la configuració de l'escena no pot ser recuperada només a partir d'imatges. En altres paraules, només veient una foto d'una habitació no podem saber si és una habitació de mida normal o es tracta de l'habitació d'una casa de nines molt realista.

Al 1913 Kruppa [16] ja va demostrar que el problema dels 5-punts tenia com a molt onze solucions. Va utilitzar geometria projectiva i va afegir els epipols de les dues imatges com a sisè i setè punts. Després va introduir la cònica absoluta (que no tractarem en aquest treball, es pot trobar una descripció completa a [13, Capítol 3]) i va obtenir una equació polinòmica de grau 27, d'on 16 solucions es podien desestimar.

Més endavant, en 1988 Damazure [5], utilitzant la matriu essencial va demostrar que en general hi havia 10 solucions diferents, però no era possible calcular aquestes solucions tal i com va definir les equacions. En [7] i [14] es van proposar algunes millores respecte als casos anteriors però seguien sense poder-se calcular les solucions.

A finals de la dècada dels noranta van començar a sorgir alguns algorismes numèrics que podien implementar-se obtenint per fi solucions del problema. En el capítol 3 d'aquest treball veurem de manera detallada la primera solució eficient al problema dels 5-punts [20] de David Nistér. Utilitza una eliminació gaussiana millor que ens porta a tenir un polinomi de grau 10 del que podem extreure les deu solucions.

## Estructura de la Memòria

Aquest treball està dividit en tres parts:

- En el primer capítol trobem els conceptes fonamentals de la geometria multivista en la computer vision. Comencem tractant la geometria en una vista i com està relacionada amb la geometria projectiva, ja que les imatges es representen com a projeccions d'una escena tridimensional en un pla imatge 2D. A continuació, ens centrem en la geometria en dues vistes on definim la matriu fonamental i la matriu essencial que contenen la informació d'una escena en dues vistes en el cas en que tenim càmeres no calibrades i calibrades respectivament. A més, veiem com, donades dues imatges d'una mateixa escena preses des de punts de vista diferents, podem trobar punts corresponents que representen el mateix punt del món.
- En el segon capítol veiem com la geometria multivista juga un paper fonamental a l'hora de reconstruir una escena en 3D a partir d'imatges. Comentem el procés general que se segueix i la importància dels problemes minimalis.
- En el tercer capítol estudiem l'algorisme dels 5-punts de David Nistér [20]. Es tracta del primer algorisme eficient que permet recuperar la posició de dues càmeres a partir de només cinc punts corresponents en les dues vistes. Veurem detalladament l'algorisme utilitzant els conceptes introduïts al primer capítol i, a més, parlarem de com s'hauria d'implementar per tal d'aconseguir un programa eficient.

- Per últim, a les conclusions comentarem els resultats de l'algoritme dels 5-punts obtinguts en els experiments realitzats a [20]. A més, compararem aquest algoritme amb els algoritmes dels 8, 7 i 6-punts, fent èmfasi en quin algoritme és millor en cada cas depenent de les dades que tinguem.



# Capítol 1

## Preliminars

En aquest primer capítol introduïrem els conceptes i resultats bàsics sobre geometria en múltiples vistes aplicats a la computer vision. Començarem parlant de la geometria des de la perspectiva d'una única càmera i després veurem el cas en que hi ha dues càmeres.

**Definició 1.0.1.** Anomenem geometria en  $n$  – vistes (“n-view geometry”) o geometria multivista a la geometria des de la perspectiva de  $n$  càmeres o  $n$  imatges diferents.

### 1.1 Geometria afí en el marc projectiu

La geometria projectiva és una eina important a l'hora de treballar amb la geometria multivista. A continuació recordem alguns conceptes i justifiquem l'ús d'algunes idees en les següents seccions.

Sabem que en el pla projectiu dues rectes qualsevol intersequen en un punt: si són rectes paral·leles intersequen en un punt de l'infinit i si no ho són intersequen en un punt (finit) del pla afí. Veiem doncs que utilitzar l'espai projectiu ens ajuda a treballar amb més facilitat ja que ens permet generalitzar més els càlculs, per exemple, no havent de considerar diferents tipus de rectes donat que totes les rectes del pla projectiu intersequen.

En els següents apartats s'utilitzarà en algunes ocasions el concepte de paral·lelisme en l'espai projectiu fent un abús de llenguatge, donat que ens és útil per a visualitzar i entendre alguns dels conceptes. Entendrem en aquells casos que ens estem restringint al subespai afí contingut en l'espai projectiu corresponent, que és on té sentit l'ús d'aquest terme. De la mateixa manera, quan utilitzem la perpendicularitat, ens estarem restringint al subespai afí amb el producte vectorial habitual.

**Definició 1.1.1.** Un *espai projectiu* sobre un cos  $\mathbb{K}$  és un triple  $(\mathbb{P}, E, \pi)$  format per un conjunt  $\mathbb{P}$ , un  $\mathbb{K}$ -espai vectorial  $E$  de dimensió finita i una aplicació  $\pi : E \setminus \{0\} \rightarrow \mathbb{P}$  complint les següents condicions:

a) és exhaustiva;

b) qualsevol parella de vectors no nuls  $u, v \in E$  satisfà  $\pi(u) = \pi(v)$  si, i només si,  $u$  i  $v$  són proporcionals, és a dir,  $u = \lambda v$  amb  $\lambda \in \mathbb{K}$ .

**Definició 1.1.2.** Considerem un pla afí  $\mathbb{A}_n$  on es fixen coordenades (afins) que anomenem  $x_1, x_2, \dots, x_n$ . Per a cada punt  $p = (x_1, \dots, x_n) \in \mathbb{A}_n$ , anomenem *coordenades homogènies*

de  $p$  a qualsevol conjunt de  $n + 1$  nombres  $\alpha_1, \alpha_2, \dots, \alpha_{n+1}$  que compleixi  $x_1 = \alpha_1/\alpha_{n+1}, \dots, x_n = \alpha_n/\alpha_{n+1}$ .

D'ara en endavant estudiem el cas general de dimensió  $n$  i ens centrem en el cas de dimensió 3 donat que és el que més ens interessarà en les següents seccions, ja que el món en que ens movem és l'espai en tres dimensions.

Les coordenades homogènies de  $p = (x_1, x_2, x_3, x_4)$  estan determinades, mòdul escalar, de manera que podem dir que  $p$  té com a coordenades homogènies  $x_1, x_2, x_3, 1$ . Qualsevol quaterna  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  de coordenades homogènies d'un punt complint que  $\alpha_4 \neq 0$  és la quaterna de coordenades homogènies del punt de coordenades afins  $\alpha_1/\alpha_4, \alpha_2/\alpha_4$  i  $\alpha_3/\alpha_4$ .

**Definició 1.1.3.** Els punts amb coordenades homogènies  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  amb  $\alpha_1 \neq 0$  o  $\alpha_3 \neq 0$  o  $\alpha_2 \neq 0$  i  $\alpha_4 = 0$  s'anomenen *punts impropis*, *punts de l'infinit* o *punts ideals*.

**Observació 1.1.4.** Els punts de  $\mathbb{A}_3$ , on es pot suposar que  $\alpha_4 = 1$ , s'anomenen *punts propis*.

Observem que assignant a cada element  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \in \mathbb{K}^4 - \{0, 0, 0, 0\}$  el punt de coordenades homogènies  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ , aleshores el conjunt de tots els punts propis i impropis, adquireix estructura de pla projectiu i és, per definició, la clausura projectiva  $\bar{\mathbb{A}}_3$  del pla afí  $\mathbb{A}_3$ . En  $\bar{\mathbb{A}}_3$  els punts impropis formen el pla  $\alpha_4 = 0$ , un pla distingit anomenat pla impropis o pla de l'infinit.

**Observació 1.1.5.** Com a conseqüència de la completió de l'espai triada, considerem  $\alpha_{n+1} = 0$  l'hiperplà de l'infinit. En el cas de dimensió 3 en particular tenim que  $\alpha_4 = 0$  és el pla de l'infinit. Utilitzarem aquesta convenció en els següents apartats.

En general, donat un espai afí  $\mathbb{A}_n$  de dimensió  $n$ , amb espai de vectors lliures  $F$ , si considerem el projectivitzat  $\mathbb{P}(F)$ , que és un espai projectiu de dimensió  $n - 1$  els elements del qual, són les classes mòdul proporcionalitat  $[v]$  dels vector  $v \in F - \{0\}$ , és a dir, les direccions dels vectors lliures no nuls. Ens referirem a cada  $[v] \in \mathbb{P}(F)$  com la direcció de  $v$ .

Prenent les direccions com a nous punts, juntament amb els punts de  $\mathbb{A}_n$  tenim els punts de la clausura projectiva:

**Definició 1.1.6.** El conjunt  $\bar{\mathbb{A}}_n = \mathbb{A}_n \cup \mathbb{P}(F)$  és la clausura projectiva de  $\mathbb{A}_n$ .

La *clausura projectiva* és un espai projectiu de dimensió  $n$  que està format pels punts de  $\mathbb{A}_n$ , és a dir, els punts propis, i les direccions dels vectors lliures no nuls de  $\mathbb{A}_n$ , els punts de l'infinit.

**Observació 1.1.7.** En altres paraules, podem descriure l'espai projectiu com:

$$\mathbb{P}^2 = \mathbb{A}_2 \cup \{\text{recta de l'infinit}\}$$

$$\mathbb{P}^3 = \mathbb{A}_3 \cup \{\text{pla de l'infinit}\}$$

...

$$\mathbb{P}^n = \mathbb{A}_n \cup \{\text{hiperplà de l'infinit}\}$$

S'acostuma a anomenar al conjunt de punts impropis  $\mathbb{P}(F)$  hiperplà impropis o hiperplà de l'infinit  $H_\infty$ .

## 1.2 Geometria en una vista

En aquesta secció definirem el concepte de càmera així com les seves propietats i paràmetres intrínsecs. A més, estudiarem alguns dels diferents tipus de càmeres. De manera intuïtiva podem pensar simplement que una càmera és una aplicació que projecta un punt del món 3D a una imatge 2D.

**Definició 1.2.1.** Sigui  $C$  un punt i  $\pi \subset \mathbb{P}^3$  un pla. Podem definir una aplicació

$$\begin{aligned} g : \mathbb{P}^3 \setminus C &\rightarrow \mathbb{P}^2 \\ X &\mapsto x = \pi \cap CX \end{aligned}$$

que projecta una escena en un espai 3D a una imatge en el pla 2D. Anomenem *centre de càmera* al punt  $C$  i pla imatge o pla focal a  $\pi$ . Donat un punt  $X \in \mathbb{P}^3 \setminus C$  considerem la recta  $l$  que uneix  $x$  amb el centre de la càmera. Aleshores la imatge del punt  $X$  per  $g$  és el punt d'intersecció de la recta  $l$  amb el pla imatge  $\pi$ . Anomenem a aquesta aplicació *càmera*.

Representem l'aplicació  $g$  mitjançant una matriu  $P$  de dimensions  $3 \times 4$  que anomenem *matriu de càmera*. Relaciona coordenades homogènies d'un punt del món en el 3-espai a coordenades homogènies del punt imatge en la imatge plana.

**Definició 1.2.2.** Els *punts d'imatge*  $x$  (“image points”) són els punts definits per la intersecció del pla imatge amb les rectes que uneixen els punts del món amb el centre de la càmera. Són els punts que apareixen a la imatge 2D. Anomenem *punts del món* (“world points”) als punts d'una escena en un espai 3D.

**Observació 1.2.3.** Acostumem a anomenar als punts del món  $X$  o  $Q$  (en majúscula) i estan representats per un 4-vector de coordenades homogènies de la forma  $(X_1, Y_1, Z_1, 1)^T$  i  $x$  o  $q$  (en minúscula) als punts de la imatge representats per un 3-vector de coordenades homogènies.

Considerem la projecció central de punts de l'espai a un pla. Definim alguns conceptes importants de les càmeres:

**Definició 1.2.4.** El centre de projecció de la càmera s'anomena *centre de la càmera*.

**Definició 1.2.5.** El pla en que es projecten els punts s'anomena *pla imatge* o *pla focal* (“focal plane”). La distància del centre de la càmera al pla imatge s'anomena *distància focal*  $f$  (“focal length”).

**Definició 1.2.6.** La recta des del centre de la càmera perpendicular al pla imatge s'anomena l'*eix principal* (“principal axis”) de la càmera i el punt en que l'eix principal es troba amb el pla imatge s'anomena *punt principal* (“principal point”).

**Definició 1.2.7.** El pla a través del centre de la càmera paral·lel al pla imatge es diu *pla principal de la càmera* (“principal plane”).

**Proposició 1.2.8.** Sigui  $P$  la matriu d'una càmera qualsevol. Aleshores aquesta matriu es pot descompondre com  $P = K[R|t]$ , on  $K$  és la matriu de calibratge,  $R$  la matriu de rotació i  $t$  el vector de translació.

*Demostració.* Es tracta d'una descomposició de matrius. Per a una demostració completa consultar [13, Secció 6.2.4].  $\square$

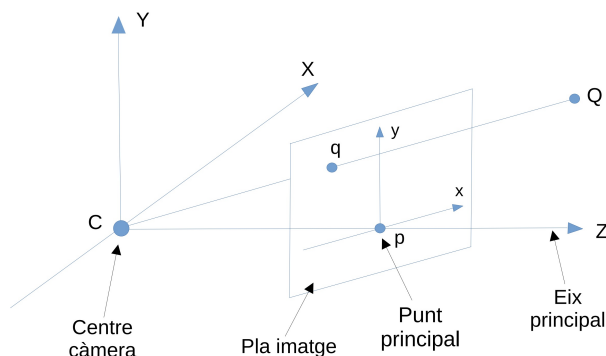


Figura 1.1: Exemple de càmera (de tipus pinhole) i les seves parts. En aquest cas tenim  $C$  el centre de la càmera situat a l'origen de coordenades i  $Q$  el punt del món projectat al punt  $q$  del pla imatge. Observem que el pla imatge està situat davant del centre de la càmera.

La matriu de càmera  $P$  conté totes les propietats de la càmera com ara el seu centre o la seva distància focal. En concret, els paràmetres interns de la càmera, com ara la distància focal, estan continguts en una submatriu  $K$  de dimensions  $3 \times 3$  que es pot obtenir a partir de la matriu  $P$  mitjançant la descomposició  $QR$ . Estudiarem tots aquests conceptes de manera detallada en les següents seccions.

A continuació, considerem un exemple senzill de càmera per a familiaritzar-nos amb els conceptes i després veurem formalment com s'expressa la matriu de càmera i quins són els seus elements i propietats.

**Exemple 1.2.9.** El tipus de càmera més bàsic que podem trobar és el model *Pinhole*. Es tracta d'una capsula amb una petita obertura (d'aquí el nom "pinhole") en un dels costats. La llum entra per aquesta obertura i forma la imatge invertida al costat oposat de la caixa. Es considera el model més bàsic donat que les seves característiques, com ara que és finita o que té els píxels quadrats, conceptes que veurem més endavant, ens restringeixen més la descripció de la matriu de la càmera. Podem veure-la simplement com una projecció de punts.

Considerem la projecció central dels punts de l'espai a un pla. Suposem que tenim el centre de projecció, és a dir, el centre de la càmera en l'origen de coordenades d'un sistema de coordenades Euclidi mirant cap a l'eix de les  $Z$ . Suposem que el pla imatge es troba a una distància  $f$  del centre de la càmera. Tindrem doncs que el pla d'imatge és  $Z = f$ .

Suposem que els punts del món i els punts imatge estan representats per vectors homogenis. Sigui  $(X_1, Y_1, Z_1, 1)^T$  un punt del món que està situat davant de la càmera. Considerem la recta que va del centre de projecció a aquest punt del món. El punt on aquesta recta talla el pla imatge, és la imatge del punt del món.

**Observació 1.2.10.** En general, sigui  $A$  una matriu de dimensions  $n \times m$  qualsevol i  $v$  un vector de dimensió  $m \times 1$ , utilitzarem la notació  $[A|v]$  per a referir-nos a la matriu  $A$  afegint el vector  $v$  com a última columna, de manera que queda una matriu de dimensions  $n \times m + 1$ .

Considerem  $X$  un punt del món representat per un 4-vector de coordenades homogènies i  $x$  un punt imatge representat per un 3-vector de coordenades homogènies.

Aleshores, seguint amb la idea de l'exemple anterior, podem expressar la projecció central com una aplicació lineal entre les seves coordenades homogènies, en forma d'una multiplicació de matrius:

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX_1 \\ fY_1 \\ Z_1 \end{pmatrix} = \begin{pmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}$$

La matriu de càmera d'aquesta expressió es pot escriure com  $diag(f, f, 1)[I|0]$  amb  $diag(f, f, 1)$  una matriu diagonal i  $[I|0]$  una matriu dividida en un bloc  $3 \times 3$ , la matriu identitat en aquest cas, afegint un vector columna més a la dreta, el vector nul.

Observem que abans hem suposat que l'origen de coordenades en el pla imatge estava en el punt principal. Suposem ara que el punt principal és un altre punt  $(p_x, p_y)^T$ .

**Definició 1.2.11.** Un *marc de coordenades* defineix la posició i l'orientació dels objectes de l'espai en relació a un origen i uns eixos de coordenades fixats.

**Observació 1.2.12.** Recordem que el punt principal és un punt del pla imatge i dins d'aquest pla tenim un marc de coordenades propi. Es tracta per tant d'un punt en dues dimensions que depèn del marc de coordenades escollit a la imatge.

Llavors podem expressar l'aplicació com,

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Xp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{pmatrix} f & p_x & 0 \\ & f & p_y \\ & & 1 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix}$$

Prenent  $K = \begin{pmatrix} f & p_x \\ & f & p_y \\ & & 1 \end{pmatrix}$  podem reescriure el producte de matrius com  $x = K[I|0]X$ .

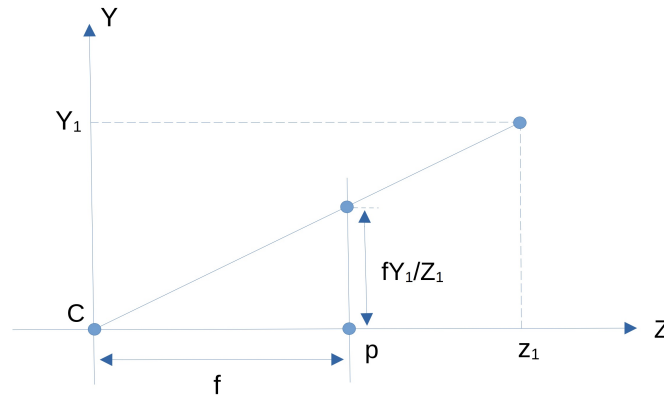


Figura 1.2: Esquema representatiu de la geometria del model de càmera pinhole. Tenim  $C$  el centre de càmera,  $f$  la distància focal i  $p$  el punt principal.

A més, cal remarcar que ens aquest cas hem assumit que la càmera es troba a l'origen del sistema Euclidi de coordenades de l'eix principal de la càmera apuntant recte cap a l'eix  $Z$ . Aquest sistema de coordenades s'anomena el *marc de coordenades de la càmera* (“camera coordinate frame”).

**Definició 1.2.13.** Un *píxel* és l'element més petit d'una imatge digital que pot ser processat o analitzat. Normalment, es representa com un quadrat petit que conté un únic color.

Els píxels es poden utilitzar per representar les coordenades d'imatge de les càmeres. Per exemple, al conèixer les coordenades de píxels d'un punt en diferents vistes i els paràmetres corresponents de la càmera, es possible calcular la posició tridimensional de la càmera respecte de l'escena.

Veiem que fins ara hem suposat que les coordenades de la imatge eren coordenades Euclídiades que tenen la mateixa escala en les dues direccions dels eixos. Per a les càmeres CCD, és a dir, les càmeres amb un cert tipus de sensor per a detectar la imatge, hi ha la possibilitat de tenir píxels que no són quadrats. Cal introduir llavors els factors d'escala que poden ser diferents en cada direcció. Siguin  $m_x$  i  $m_y$  el nombre de píxels per unitat de distància en la direcció dels eixos  $x$  i  $y$  respectivament, aleshores la matriu de calibratge es pot definir de la forma següent.

**Definició 1.2.14.** Anomenem *matriu de calibratge* a la matriu de la forma  $K = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$

amb:

$\alpha_x = fm_x$  és el factor d'escala en la direcció de l'eix de coordenades  $x$ .

$\alpha_y = fm_y$  és el factor d'escala en la direcció de l'eix de coordenades  $y$ .

$s$  és el paràmetre antisimètric (“skew”)

$(x_0, y_0)^T = (m_x p_x, m_y p_y)$  són les coordenades del punt principal.

La *relació d'aspecte* ve donada per  $\alpha_y/\alpha_x$ .

**Observació 1.2.15.** La definició de matriu de calibratge més habitual i amb la que es treballa normalment degut als models de càmera més utilitzats és la següent:

$$K = \begin{pmatrix} f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}$$

on  $f$  la distància focal,  $s$  és el paràmetre antisimètric i  $(p_x, p_y)^T$  són les coordenades del punt principal.

Veiem doncs que conèixer la matriu de calibratge equival a conèixer els paràmetres interns de la càmera. Cal tenir en compte que el paràmetre  $s$  serà igual a 0 en la majoria de càmeres d'ús habituals. El cas  $s \neq 0$  és molt inusual, es pot trobar un cas on apareix descrit a [13, Secció 6.2.4].

**Definició 1.2.16.** Diem que una càmera és *calibrada* si coneixem la seva matriu de calibratge  $K$ , és a dir, si coneixem els seus paràmetres interns. En cas contrari diem que és una càmera no calibrada.

Més endavant veurem la importància de tenir (o suposar) que les càmeres són calibrades. Això ens permetrà simplificar molt els càlculs en els problemes en geometria multivista, aconseguint d'aquesta manera fer-los resolubles en alguns casos.

Quan tenim una càmera que no satisfà cap restricció específica tenim un tipus de càmera més general que anomenem *càmera projectiva general*. És una càmera que es pot representar com una matriu homogènia  $3 \times 4$  qualsevol de rang 3. Observem que ha de tenir rang 3 per tal que l'espai d'arribada de la projecció sigui un pla. Una matriu  $3 \times 4$  té 12 coeficients, mòdul escalar, tenim doncs que en general té 11 graus de llibertat.

Podem descompondre la matriu d'una càmera general projectiva com  $P = [M|p_4]$ , amb  $M$  una matriu  $3 \times 3$  i  $p_4$  un vector columna. A partir d'aquesta descomposició podem distingir dos tipus principals de càmeres.

**Definició 1.2.17.** Diem que una càmera  $P = [M|p_4]$  és *finita* si el seu centre de càmera és un punt finit. Diem que una càmera és *infinita* si el centre de la càmera es troba en un punt de l'infinit. Si  $M$  és singular aleshores la càmera és infinita i si no ho és, llavors és finita.

Recordem que diem que una matriu és *singular* si té determinant igual a zero, és a dir, si no té rang màxim. Ens centrarem en estudiar el cas de les càmeres finites, tot i això moltes de les propietats i resultats que veurem també són vàlids per a càmeres infinites, per a més detalls consultar [13, Capítol 6].

## Translació i rotació de càmeres

En general, els punts de l'espai s'expressaran en termes de marcs de sistemes de coordenades Euclídiades diferents, conegut com marc de coordenades del món (“world coordinate frame”). Dos marcs de coordenades estan relacionats per una rotació i una translació. Veiem primer una definició alternativa de la matriu de càmera, equivalent a la ja donada anteriorment a la Proposició (1.2.8).

**Definició 1.2.18.** Anomenem matriu de càmera a la matriu de la forma  $P = KR[I - \tilde{C}]$  on  $K$  és la matriu de calibratge,  $R$  la matriu de rotació i  $\tilde{C}$  és el centre de la càmera.

Els paràmetres de les matrius de rotació  $R$  i centre de la càmera  $\tilde{C}$  relacionen l'orientació i la posició de la càmera a un sistema de coordenades del món.

**Observació 1.2.19.** Ens fixem que aquesta manera de definir la matriu de càmera ens permet saber si el centre de càmera és finit només mirant l'últim vector columna.

**Exemple 1.2.20.** Sigui  $\tilde{x}$  un 3-vector no homogeni que representa les coordenades d'un punt en un sistema de coordenades del món i sigui  $\tilde{x}_{cam}$  la representació d'aquest mateix punt en el marc del sistema de coordenades de la càmera. Aleshores podem escriure  $\tilde{x}_{cam} = R(\tilde{x} - \tilde{C})$  on  $\tilde{C}$  representa les coordenades del centre de la càmera en el marc de coordenades del món i  $R$  és una matriu de rotació que representa la orientació del marc de coordenades de la càmera.

Ens fixem ara en el cas de la càmera de tipus “pinhole”. Tindrà una matriu de càmera de la forma  $P = KR[I - \tilde{C}]$  que ens donarà 9 graus de llibertat: 3 de la matriu  $K$  pels elements  $f, p_x, p_y$ , 3 de la matriu  $R$  i 3 del vector  $\tilde{C}$ . Com ja hem comentat abans, al tractar-se d'un tipus de càmera molt bàsic, té menys graus de llibertat ja que alguns dels seus paràmetres estan prefixats.

**Definició 1.2.21.** Els paràmetres continguts a la matriu de calibratge  $K$  són els *paràmetres interns* (o *orientació interna*) de la càmera. Els paràmetres de les matrius  $R$  i  $\tilde{C}$  s'anomenen *paràmetres externs* (o *orientació externa*) de la càmera.

**Observació 1.2.22.** Normalment no ens interessa haver de donar el centre de la càmera de manera explícita. Aleshores representem la transformació del món a la imatge com  $\tilde{X} = R(\tilde{X} + t)$ . Anomenarem  $t$  el *vector de translació*. La matriu de càmera corresponent serà doncs  $P = K[R|t]$  amb  $t = -R\tilde{C}$ . Aquesta serà la representació de matriu de càmera que usarem habitualment.

## Propietats de les càmeres

Sigui  $P = [M|p_4]$  la matriu d'una càmera projectiva qualsevol, veurem a continuació algunes de les seves propietats principals. Abans recordarem el concepte de subespai nul, que serà important en les següents seccions.

**Definició 1.2.23.** Considerem una aplicació lineal representada per una matriu  $A$  de dimensions  $m \times n$  amb coeficients en un cos  $\mathbb{K}$  (normalment  $\mathbb{R}$  o  $\mathbb{C}$ ). Anomenarem *subespai nul* (“nullspace”) dret o  $\ker(A)$  als vectors  $x \in \mathbb{K}^n$  tal que  $Ax = 0$ , on  $0$  és el vector nul. El subespai nul esquerre està format pels vectors  $x \in \mathbb{K}^n$  tal que  $x^T A = 0^T$ .

1. **Centre de la càmera:** El centre de la càmera és el subespai nul dret  $C$  de  $P$ , és a dir,  $PC = 0$ .

La matriu  $P$  és de rang 3 i té 4 columnes, per tant el seu subespai nul dret tindrà dimensió 1.

Considerem una recta que conté  $C$  i qualsevol altre punt  $Q$  del 3-espai. Llavors els punts d'aquesta recta es poden representar com el “join”

$$X(\lambda) = \lambda Q + (1 - \lambda)C$$

Aplicant a aquesta recta la càmera  $x = PX$  tenim que els punts de la recta es projecten a

$$x = PX(\lambda) = \lambda PQ + (1 - \lambda)PC = \lambda PQ$$

ja que  $PC = 0$ . Això implica que tots els punts de la recta van a parar al mateix punt imatge  $PQ$ , i aleshores la recta ha de ser un raig que passa pel centre de la càmera. Com que per qualsevol elecció del punt  $Q$  la recta  $X(\lambda)$  passarà pel centre de càmera, obtenim que  $C$  ha de ser el centre de càmera. Es pot calcular que de fet el centre de càmera serà de la forma:

- Si la càmera és finita ( $M$  és no singular):  $C = \begin{pmatrix} -M^{-1}p_4 \\ 1 \end{pmatrix}$
- Si la càmera és infinita ( $M$  és singular): Sigui  $d$  el 3-vector nul de  $M$ , complint que  $Md = 0$ , aleshores  $C = \begin{pmatrix} d \\ 0 \end{pmatrix}$

2. **Punts columna:** Els 3-vectors columna de la càmera projectiva tenen un sentit geomètric com a punts imatge. Per a  $i = 1, 2, 3$  els vectors columna  $p_i$  són els punts de l'infinit en la imatge corresponents als eixos  $X, Y$  i  $Z$  respectivament. Això passa donat que aquests punts  $p_i$  són les imatges de les direccions dels eixos. La columna  $p_4$  és la imatge de l'origen de coordenades.



## 1.3 Geometria en dues vistes

En aquesta secció parlarem de la geometria des de la perspectiva de dues vistes. Podria ser el cas de dues càmeres que fan una fotografia en el mateix moment des de dues posicions diferents o podria tractar-se d'una càmera que és mou respecte a l'escena. Aquests dos exemples de situacions són geomètricament equivalents i per a nosaltres seran iguals.

Comencem establint una notació. Suposem que tenim dues vistes, cadascuna d'elles tindrà associada una matriu de càmera,  $P$  i  $P'$  respectivament. Aleshores un punt  $X$  del 3-espai es representa a les imatges com  $x = PX$  a la primera vista i  $x' = P'X$  a la segona.

**Definició 1.3.1.** Diem que dos punts  $x$  i  $x'$  de la primera i segona vista respectivament són *corresponents* si són la imatge del mateix punt del món  $X$  del 3-espai.

**Observació 1.3.2.** Un feix de plans és el conjunt de plans que intersequen amb una recta en el 3-espai, anomenada eix del feix.

**Definició 1.3.3.** Donades dues càmeres  $P$  i  $P'$ . Anomenem *línia de base* (“baseline”) a la recta que uneix els centres de les dues càmeres.

En la *geometria epipolar*, és a dir, la geometria entre dues vistes, estudiarem el comportament de la intersecció dels plans imatge amb el feix de plans que té la línia de base com a eix.

**Definició 1.3.4.** Sigui  $x$  la imatge en una vista del punt del món  $X$ . Considerem el pla  $\pi$  determinat per la línia de base i la recta que passa per  $x$  i  $X$ . Anomenem *recta epipolar*, la recta d'intersecció de  $\pi$  amb el segon pla imatge.

**Definició 1.3.5.** Anomenem *epipol* al punt d'intersecció de la recta que uneix els centres de les càmeres (línia de base) amb el pla imatge.

**Observació 1.3.6.** Podem veure l'epipol com la imatge del centre de la càmera d'una vista en l'altre vista. De la mateixa manera, la recta epipolar és la imatge en la segona vista de la recta que uneix el punt  $x$  de la primera imatge amb el punt del món  $X$ .

**Definició 1.3.7.** El *pla epipolar* és un pla qualsevol que conté la línia de base.

**Observació 1.3.8.** Observem com es relacionen entre ells els elements definits: la recta epipolar és la intersecció del pla epipolar amb el pla imatge. A més, totes les rectes epipolars intersequen amb l'epipol. Ens fixem també que, l'antiimatge d'un punt  $x$  d'una imatge és el raig que uneix el centre de la càmera amb  $X$  el punt del món que representa  $x$ .

### 1.3.1 Correspondència de punts

Donat un punt imatge  $x$  a la primera vista, ens podem preguntar com aquesta informació ens limita la posició del punt  $x'$  corresponent a la segona vista. L'algoritme dels 5-punts que treballarem en els següents capítols parteix de 5 punts corresponents en dues vistes, ens interessa doncs saber com arribar a tenir aquests punts corresponents.

Siguin  $P$  i  $P'$  dues matrius de càmera de la primera i segona vista respectivament amb centres  $C$  i  $C'$  i  $X$  un punt del 3-espai que correspon a  $x = PX$  i  $x' = P'X$ . Sigui  $\pi_1$

i  $\pi_2$  els plans d'imatge. Resolent l'equació  $PX = x$  aplicant  $P^+$  la matriu pseudoinversa de  $P$  tal que  $PP^+ = I$ , obtenim la família d'infinites solucions

$$X(\lambda) = P^+x + \lambda C \quad (1.3.1)$$

Amb  $C$  el centre de càmera que recordem que és  $\ker(P) = C$ .

En particular, dos punts d'aquesta recta són:

- Si  $\lambda = 0$  aleshores  $X = P^+x \Rightarrow X = P'(P^+x)$ , vist des de la segona càmera.
- Si  $\lambda = \infty$  aleshores  $X = C \Rightarrow X = P'(C)$ , vist des de la segona càmera.

Llavors, la recta epipolar és el join d'aquests dos punts  $l' = (P'C) \vee P'P^+x$ . Equivalentment, podem escriure la recta epipolar com  $l' = (x \vee C \vee C') \cap \pi_2$ . Ens fixem que  $C \vee C'$  és la línia de base. Veiem que  $P'C \vee P'P^+x = (x \vee C \vee C') \cap \pi_2$ . Prenent la recta  $P^+x \vee C'$ .

Aleshores els punts  $P'C$  i  $P'P^+x \in l'$  ja que  $(P'X \vee C') \vee P'P^+x = \pi_2$  i  $C \in P'C \vee C'$ ,  $(P^+x) \in P'P^+x$ .

Considerem el pla epipolar  $\pi = C \vee P^+x \vee C' = C \vee x \vee C'$ . Ens fixem que també podem definir aquest pla com  $\pi = C \vee x \vee P'P^+x$ .

És a dir, per a trobar la recta epipolar  $l'$  al pla imatge de la segona vista on es troba el punt corresponent a  $x$ ,  $x'$ , només ens cal conèixer les matrius de càmera  $P$  i  $P'$ ,  $C$  el centre de la primera càmera, a més del punt  $x$ .

Tenim doncs que donat un parell d'imatges sabem que per a tot punt  $x$  en una imatge, existeix una recta epipolar  $l'$  corresponent en l'altre imatge. Aleshores qualsevol punt  $x'$  de la segona imatge corresponent al punt  $x$  ha d'estar a la recta epipolar  $l'$ . Això, fa que per a trobar el punt  $x'$  corresponent no haguem de buscar a tota la imatge sinó que només cal mirar a la recta  $l'$ .

La recta epipolar és la projecció en la segona imatge de la recta que passa pel punt  $x$  i a través del centre de la càmera  $c$  a la primera càmera. Tenim llavors l'aplicació  $x \mapsto l'$  d'un punt en una imatge a la seva recta epipolar corresponent en l'altre imatge.

És a dir, un punt a la primera vista defineix una recta epipolar en l'altre vista on es troba el punt corresponent. Ens adonem doncs que la geometria epipolar només depèn de les càmeres, en concret de la seva posició relativa i els seus paràmetres interns.

### 1.3.2 Matriu fonamental

La matriu fonamental  $F$  conté la geometria intrínseca d'una escena en dues vistes. Aquesta matriu és independent de l'estructura de l'escena i es pot calcular a partir de correspondències de punts de les imatges, sense conèixer els paràmetres interns ni la posició relativa de les càmeres.

**Definició 1.3.9.** Suposem que tenim dues imatges preses per dues càmeres amb centres no coincidents. Aleshores la *matriu fonamental*  $F$  és l'única matriu homogènia de dimensions  $3 \times 3$  i  $\text{rang}(F) = 2$  que satisfà  $x'^T F x = 0$  per a tots els punts corresponents en les dues vistes  $\{x \leftrightarrow x'\}$ .

La matriu fonamental és la representació algebraica de la geometria epipolar. Donat un parell d'imatges, sabem pel que hem vist a la secció anterior que per a cada punt  $x$  d'una imatge existeix una recta epipolar  $l'$  a l'altre imatge. Dit d'una altra manera, qualsevol punt  $x'$  de la segona imatge corresponent al punt  $x$  ha de pertànyer a la recta epipolar  $l'$ .

Com hem vist abans, la recta epipolar és la projecció en la segona imatge de la recta que travessa el punt  $x$  i l'uneix amb el centre de la primera càmera  $C$ . Aleshores podem definir l'aplicació  $F : x \rightarrow l'$ , que va d'un punt en la imatge a la recta epipolar de l'altre imatge.

A continuació deduïm la matriu fonamental a partir de les correspondències de punts.

Sabem, per la definició de matriu fonamental que donat un punt  $x = (x_0, x_1, x_2) \in \pi_1$  de la primera imatge i el seu punt corresponent  $x' = (x'_0, x'_1, x'_2) \in \pi_2$  de la segona imatge aleshores s'ha de complir que  $x'^T F x = 0$ . Veiem-ho,

Per definició sabem que la recta epipolar és

$$F x = l' \quad (1.3.2)$$

Havíem vist, a partir de la correspondència de punt en dues imatges, que la recta epipolar es podia expressar com

$$l' = (P' C) \vee (P' P^+ x) = [e']_x (P' P^+) x = F x$$

Observem que  $(P' C) = P'(C)$  és el centre de la primera càmera vist des de la segona càmera (en  $\pi_2$ ), que és per definició l'epipol  $e'$ .

Veiem que  $x' \in l'$  ja que  $[e']_x (P' P^+) = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$  que és la recta epipolar.

$$l' = [e']_x (P' P^+) x = F x$$

Ens fixem en la darrera igualtat, per definició  $F$  és la matriu de la transformació projectiva

$$F : x \mapsto l' \quad (1.3.3)$$

D'aquesta manera, hem trobat la matriu fonamental  $F = [e']_x (P' P^+)$ .

## Propietats de la matriu fonamental

Comentem a continuació algunes de les propietats més importants de la matriu fonamental:

- (i) **Transposada:** Si  $F$  és la matriu fonamental del parell de càmeres  $(P', P)$ , aleshores  $F^T$  és la matriu fonamental del parell en l'ordre contrari  $(P, P')$ .
- (ii) **Recta epipolar:** Per a qualsevol punt  $x$  de la primera imatge, la seva recta epipolar corresponent és  $l' = F x$ . Anàlogament, la recta epipolar d'un punt  $x'$  de la segona imatge és  $l = F^T x'$ .

(iii) **L'epipol:** Per a qualsevol punt  $x \neq e$  la recta epipolar corresponent és  $l' = Fx$ . Sabem que aquesta recta ha de contenir l'epipol  $e'$ . Llavors, se satisfà que  $e'^T(Fx) = (e'^T F)x = 0$  per a tot  $x$ . Per tant,  $e'^T F = 0$ , és a dir,  $e'$  és el vector del subespai nul esquerra de  $F$ . Anàlogament,  $F e = 0$  i és el vector subespai nul dret de  $F$ .

(iv) **F té 7 graus de llibertat:** Una matriu homogènia de dimensions  $3 \times 3$  qualsevol té 8 graus de llibertat, hi ha 9 elements però multiplicant per un escalar adequat tenim un grau de llibertat menys. A més, sabem que la matriu  $F$  satisfà que  $\det(F) = 0$ , amb això tenim un grau de llibertat menys.

**Teorema 1.3.10.** *Una matriu de coeficients reals  $F$  de dimensions  $3 \times 3$  no nul·la és una matriu fonamental si, i només si, satisfà l'equació*

$$\det(F) = 0 \tag{1.3.4}$$

*Demostració.* Veiem les dues implicacions.

$\implies$  Suposem que  $F$  és la matriu fonamental aleshores tenim, per definició, que

$$\begin{aligned} F : \mathbb{P}^2 &\longrightarrow \mathbb{P}^{2V} \\ x &\longmapsto Fx \end{aligned}$$

On  $\mathbb{P}^{2V}$  és l'espai dual de  $\mathbb{P}^2$ , que és l'espai projectiu que parametriza les rectes de  $\mathbb{P}^2$ . Aquesta aplicació té associada una aplicació entre els espais vectorials  $V$  i  $W$  amb  $\dim(V) = 3$ ,  $\dim(W) = 3$  i  $Z$  el subespai d'arribada de  $\dim(Z) = 2$ .

$$F : V \longrightarrow W \supset Z$$

Siguin  $\pi_1$  i  $\pi_2$  els plans imatge de la primera imatge i segona imatge respectivament. Observem que  $\mathbb{P}^2 \simeq \pi_1$  i  $\mathbb{P}^{2V} \simeq \{\text{rectes } l \in \pi_2 \text{ tal que l'epipol } \in l\}$ . És a dir,  $\mathbb{P}^{2V} \simeq \mathbb{P}^1$ . Tenim doncs, que l'aplicació no és injectiva i això implica que el rang de la matriu de l'aplicació no és de rang màxim. Per tant,  $\det(F) = 0$ .

$\impliedby$  Suposem que  $\det(F) = 0$ , aleshores  $F$  no té rang màxim, tenim doncs, dos possibles casos:

i) Si  $\text{rang}(F) = 2$  aleshores tenim l'aplicació entre espais vectorials  $F : V^3 \longrightarrow W^2$  amb  $\dim(V^3) = 3$  i  $\dim(W^2) = 2$ , que té associada l'aplicació entre espais projectius  $F : \mathbb{P}^2 \longrightarrow \mathbb{P}^1$ , és a dir, el conjunt d'arribada de l'aplicació són totes les rectes que passen per un punt, en concret són les rectes que contenen l'epipol.

ii) Si  $\text{rang}(F) = 1$  llavors tenim l'aplicació entre espais vectorials  $F : V^3 \longrightarrow W^1$  amb  $\dim(V) = 3$  i  $\dim(W) = 1$ , on tots els punts de  $V$  s'envien a la mateixa recta  $l$  de  $W$ , que és la recta epipolar.  $\square$

### 1.3.3 Matriu essencial

La matriu essencial es pot considerar un cas particular de la matriu fonamental en que les coordenades d'imatge estan normalitzades, és a dir, les càmeres són calibrades. De fet, podem pensar la matriu fonamental com la generalització de la matriu essencial en que eliminem la suposició (no essencial) de que la càmera està calibrada.

En resum la matriu essencial té menys graus de llibertat i algunes propietats afegides respecte la matriu fonamental. Això fa que treballar amb la matriu essencial sigui més

fàcil que fer-ho amb la fonamental. De fet la millora en els càlculs es tan significativa que sovint trobarem el calibratge de les càmeres prèviament o suposarem que el coneixem, ja que d'altre manera molts problemes no serien resolubles. A continuació comentem aquestes propietats.

## Coordenades normalitzades

Considerem una matriu càmera descomposta com  $P = K[R|t]$  i sigui  $x = PX$  un punt de la imatge. Si coneixem la matriu de calibratge  $K$  aleshores podem aplicar la seva inversa  $K^{-1}$  al punt  $x$  per obtenir el punt  $\hat{x} = K^{-1}x = K^{-1}K[R|t]X = [R|t]X$  on  $\hat{x}$  és el punt imatge expressat en coordenades normalitzades.

**Definició 1.3.11.** Sigui  $P = K[R|t]$  una matriu de càmera. A la matriu càmera  $P' = K^{-1}P = [R|t]$  l'anomenem *matriu de càmera normalitzada*.

Definim la següent notació, que utilitzarem d'ara en endavant. Sigui  $v = (a, b, c)^T$  un 3-vector, aleshores podem definir la seva matriu antisimètrica corresponent de la següent manera:

$$[a]_x = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \quad (1.3.5)$$

Considerem ara un parell de matrius de càmera normalitzades  $P = [I|0]$  i  $P' = [R|t]$ . La matriu fonamental corresponent a aquest parell de càmeres normalitzades es diu *matriu essencial* i és de la forma

$$E = [t]_x R = R[R^T t]_x \quad (1.3.6)$$

**Definició 1.3.12.** La *matriu essencial* es defineix a partir de la següent equació

$$(\hat{x}')^T E \hat{x} = 0 \quad (1.3.7)$$

on  $\hat{x}$  i  $\hat{x}'$  són coordenades d'imatge normalitzades per a punts corresponents  $\{x \leftrightarrow x'\}$ .

**Observació 1.3.13.** Si substituïm  $\hat{x} = K^{-1}x$  i  $\hat{x}'$  en la Definició (1.3.12) obtenim  $x'^T K'^{-T} E K^{-1}x = 0$ . Comparant-ho amb l'expressió que caracteritza la matriu fonamental  $x'^T F x = 0$  Podem relacionar la matriu fonamental amb la matriu essencial mitjançant la següent expressió

$$E = K'^T F K \quad (1.3.8)$$

### 1.3.4 Propietats de la matriu essencial

Considerem la matriu essencial  $E = [t]_x R$ . Aquesta serà la manera en la que denotarem en general la matriu essencial. Veiem les seves propietats.

**Proposició 1.3.14.** Sigui  $E$  la matriu essencial d'una escena, aleshores se satisfà

$$\det(E) = 0 \quad (1.3.9)$$

Anomenem a aquesta condició la *restricció cúbica* o “single cubic constrain”.

*Demostració.* Donat que  $E$  és un cas particular de matriu fonamental, només cal veure la demostració del Teorema (1.3.10).  $\square$

**Teorema 1.3.15.** *Una matriu real no nul·la  $3 \times 3$   $E$  és una matriu essencial si, i només si, satisfà l'equació*

$$EE^T E - \frac{1}{2} \text{traça}(EE^T)E = 0 \quad (1.3.10)$$

*Demostració.* Es pot trobar la demostració completa a [5]. □

La matriu essencial es pot calcular a partir de l'equació que la defineix (Definició 1.3.12). Una vegada coneixem la matriu essencial, podem calcular les matrius de càmera. Podem assumir sempre que la primera matriu de càmera és  $P = [I|0]$ , és a dir, una matriu de càmera canònica amb la posició centrada en l'origen de coordenades. Per a calcular la segona matriu de càmera serà necessari factoritzar  $E = SR$ , amb  $S$  una matriu antisimètrica i  $R$  una matriu de rotació. A continuació ho veiem amb més detall.

### 1.3.5 Recuperació de les matrius de càmera a partir de la matriu essencial

La descomposició en valors singulars (DVS) és un tipus de factorització matricial, que ens servirà per a poder descompondre la matriu essencial  $E$ .

**Teorema 1.3.16.** *Tota matriu  $A \in \mathbb{R}^{m \times n}$  de rang  $r$  es pot escriure com*

$$A = UDV^T = (U_1, U_2) \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} \quad (1.3.11)$$

on  $U \in \mathbb{R}^{m \times m}$ ,  $V \in \mathbb{R}^{n \times n}$  són matrius ortogonals,  $U_1 \in \mathbb{R}^{m \times r}$ ,  $V_1 \in \mathbb{R}^{n \times r}$  i  $D_1 = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{n \times r}$  és una matriu diagonal positiva. Els  $\sigma_1 \geq \dots \geq \sigma_r > 0$  s'anomenen valors singulars de  $A$ . Si escrivim  $U = (u_1, \dots, u_m)$  i  $V = (v_1, \dots, v_n)$  aleshores  $u_i$  amb  $i = 1, \dots, m$ ,  $v_i$  amb  $i = 1, \dots, n$  s'anomenen vectors singulars per l'esquerra i per la dreta respectivament.

*Demostració.* Es pot trobar la demostració completa a [21, Part 1]. □

Utilitzant la descomposició en valors singulars obtenim una altre manera de caracteritzar la matriu essencial que ve donada pel següent teorema.

**Teorema 1.3.17.** *Una matriu  $3 \times 3$  és una matriu essencial si, i només si, dos dels seus valors singulars són iguals i el tercer és zero.*

*Demostració.* Es dedueix directament de la demostració del Teorema (1.3.19). □

Considerem un resultat previ d'àlgebra lineal que utilitzarem per demostrar el següent teorema. Recordem que una matriu  $A$  és simètrica si  $A = A^T$ , antisimètrica si  $A = -A^T$  i ortogonal si  $A^T = A^{-1}$ .

**Proposició 1.3.18.** *Sigui  $A$  una matriu real*

(i) *Si  $A$  és una matriu simètrica, aleshores podem descompondre  $A$  com  $A = UDU^T$ , amb  $U$  una matriu ortogonal i  $D \in \mathbb{R}$  una matriu diagonal. A més, els seus valors propis són reals i els seus vectors propis són ortogonals.*

(ii) Si  $A$  és una matriu antisimètrica, aleshores  $A = UBU^T$  amb  $B$  una matriu diagonal per blocs de la forma  $\text{diag}(a_1Z, a_2Z, \dots, a_mZ, 0, \dots, 0)$  amb  $Z = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ . A més, els valors propis d'una matriu antisimètrica són imaginaris purs.

*Demostració.* Per a una demostració completa consultar [10]. □

El següent teorema és un dels més importants que tractarem i el que ens permet recuperar les matrius de càmera a partir de la matriu essencial.

**Teorema 1.3.19.** *Sigui  $E = U\text{diag}(1, 1, 0)V^T$  la descomposició en valors singulars de la matriu essencial  $E$ , amb  $U$  i  $V^T$  complint que  $\det(U) > 0$  i  $\det(V) > 0$ . Aleshores  $t = t_u \equiv (u_{13} \ u_{23} \ u_{33})$  i  $R$  és igual a  $R_a \equiv UWV^T$  o  $R_b \equiv UW^TV^T$ , on  $W$  és la matriu  $W = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ .*

*Demostració.* Considerem la matriu  $Z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ .

Observem que  $Z$  és antisimètrica, és a dir,  $Z^T = -Z$  i la matriu  $W$  és ortogonal,  $W^T = W^{-1}$ .

Veurem primer que  $E = SR$  amb  $S = UZU^T$ ,  $R = UWV^T$  o  $R = UW^TV^T$  i després veurem que  $S = [t]_x = (u_{13} \ u_{23} \ u_{33})^T$ .

Observem que podem factoritzar  $E$  com:

$$E = SR = (UZU^T)(UWV^T) = (UZ)Id(WV^T) = UZWV^T = U\text{diag}(1, 1, 0)V^T.$$

$$\text{ja que } ZW = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{diag}(1, 1, 0).$$

Veiem ara que no hi ha cap altre factorització possible: Suposem que  $E = SR$ . La matriu  $S = UZU^T$  (per la Proposició 1.3.18) queda determinada pel fet que té el mateix subespai nul esquerra que la matriu  $E$ . La matriu de rotació es pot escriure com  $R = UXV^T$  amb  $X$  una certa matriu de rotació. Aleshores,

$$U\text{diag}(1, 1, 0)V^T = E = SR = (UZU^T)(UXV^T) = U(ZX)V^T$$

$\Rightarrow ZX = \text{diag}(1, 1, 0)$ . Com que  $X$  és una matriu de rotació aleshores ha de ser  $X = W$  o  $X = W^T$ .

Per últim, comprovem que  $S = [t]_x$  amb  $t = (u_{13} \ u_{23} \ u_{33})$ .

La factorització  $E = SR$  determina la part  $t$  de la matriu de càmera  $P'$ , mòdul escalar, de  $S = [t]_x$ .

Recordem que la norma de Frobenius d'una matriu  $A$  es defineix com  $\|A\|_F = \sqrt{\text{tr}(\text{tr} A^T A)}$ . La norma de Frobenius de  $S = UZU^T$  és

$$\|S\|_F = \sqrt{\text{tr}(\text{tr} S^T S)} = \sqrt{\sum \sigma_i^2}$$

amb  $\sigma_i$  els valors singulars de  $S$ . Sigui  $S = \begin{pmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{pmatrix}$ . Aleshores,

$$SS^T = \begin{pmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{pmatrix} \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & -bc & -ac \\ -bc & a^2 + c^2 & -ab \\ -ac & -ab & b^2 + c^2 \end{pmatrix}$$

$$\Rightarrow \|S\|_F = \sqrt{\text{tr}a\text{ç}a(SS^T)} = \sqrt{2(a^2 + b^2 + c^2)} = \sqrt{2}$$

Això vol dir que si  $S = [t]_x$  aleshores  $\|t\|_F = 1$  ja que  $a^2 + b^2 + c^2 = 1$  i  $t = [a \ b \ c]$ . Observem que és una normalització convenient per a la línia de base de les dues matrius de càmera.

Ara, com que  $St = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$ , utilitzant la descomposició de  $S = UZU^T$ , tenim  $UZU^T \cdot t = 0$ . Aïllant, obtenim que  $t = U(0, 0, 1)^T = U_3$ , on  $U_3 = (u_{13} \ u_{23} \ u_{33})^T$  és la darrera columna de la matriu  $U$ . □

Veiem al teorema anterior que el signe de la matriu essencial  $E$  (i per tant de  $t$ ) no es pot determinar. Aleshores donada una matriu essencial hi ha 4 possibles eleccions de la matriu de càmera  $P'$ , basades en les dues possibles eleccions de  $R$ , la matriu de rotació, i els 2 possibles signes (positiu o negatiu) de  $t$ .

Qualsevol combinació de  $R$  i  $t$  complint el Teorema (1.3.19) satisfà la restricció epipolar, és a dir, compleix  $q'^T E q = 0$  amb  $q$  i  $q'$  punts corresponents. Com ja hem dit abans, es considera que la primera matriu de càmera és  $[I|0]$ , una matriu de càmera canònica amb la posició centrada en l'origen de coordenades. A més, suposem que el vector  $t$  és unitari, l'anomenarem  $t_u$ . Llavors tenim quatre possibilitats per a la segona matriu de càmera:

$$\begin{aligned} P_A &\equiv [R_a | t_u] \\ P_B &\equiv [R_a | -t_u] \\ P_C &\equiv [R_b | t_u] \\ P_D &\equiv [R_b | -t_u] \end{aligned}$$

Només una d'aquestes quatre matrius correspon a la configuració real. Observem que  $P_B$  té la mateixa rotació que  $P_A$  però té el vector de translació en negatiu. Correspon al *reflex* de  $P_A$ , que s'obté girant una de les vistes 180 graus al voltant de la línia base.

Veiem que  $P_C$  té diferent matriu de rotació que  $P_A$  i el mateix vector de translació. Ens fixem que  $R_a = UDV^T$  i  $R_b = UD^T V^T$ . És a dir, la configuració amb  $P_C$  representa la *parella invertida* ("twited pair") de  $P_A$ .

Finalment, tenim  $P_D$  que té la mateixa matriu de rotació que  $P_C$  i té el vector de translació de signe contrari  $-t$ . Per tant, és la parella invertida de  $P_B$  i el reflex de  $P_C$ .

En conclusió, una de les configuracions és la correcta, l'altre és la parella invertida rotant 180 graus respecte la línia de base i les altres dues són el reflex i el reflex invertit.

Si  $P_A$  dona una configuració i  $P_B$  correspon al seu reflex, observem que podem anar de l'una a l'altre aplicant una transformació  $H_r$ ,



$$P_A H_t = P_B \Rightarrow [R_a | t_u] H_t = [R_a | -t_u]$$

Observem que les quatre primeres columnes es mantenen iguals, ja que tenen la mateixa rotació  $R_a$ , i l'última columna és la mateixa però amb signe canviat. Aleshores deduïm que la transformació és

$$H_t = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \text{diag}(1, 1, 1, -1)$$

De la mateixa manera, si  $P_A$  dona una configuració i  $P_C$  correspon a la seva parella invertida, podem anar de l'una a l'altre aplicant la transformació  $H_t$ ,

$$P_A H_t = P_C \Rightarrow [R_a | t_u] H_t = [R_b | t_u]$$

$$\Rightarrow [UDV^T | t_u] H_t = [UD^T V^T | t_u]$$

$$\Rightarrow H_t = \begin{pmatrix} Id_{3 \times 3} & 0 \\ -2v_{13} & -2v_{23} & -2v_{33} & -1 \end{pmatrix}$$

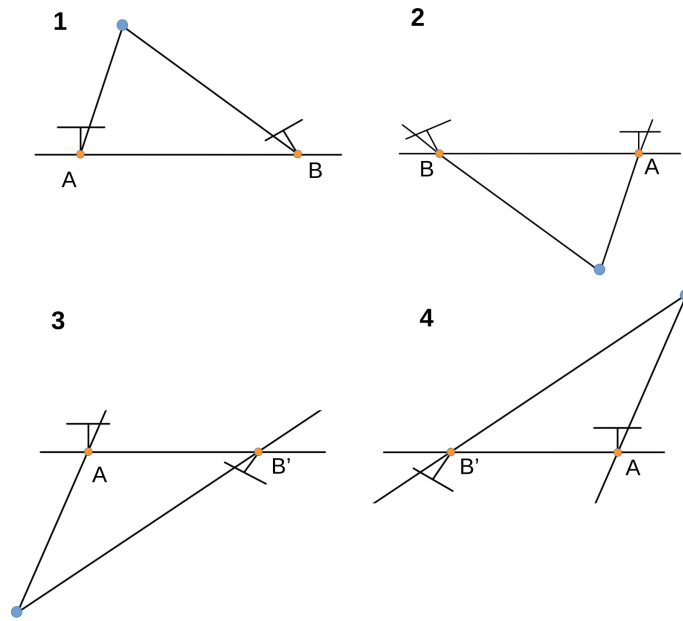


Figura 1.3: Exemple de quatre possibles solucions per a una reconstrucció a partir de la matriu  $E$  de dues càmeres calibrades. Observem que entre les de la fila de dalt i les de la fila de sota la càmera  $B$  rota 180 graus al voltant de la línia de base. Les escenes de la columna dreta i esquerra estan relacionades per un reflex. En aquest cas només **1** és un punt reconstruït que es troba davant de les dues càmeres.

Més en general, tenim el següent teorema de reconstrucció projectiva.

**Teorema 1.3.20.** *Suposem que  $\{x_i \leftrightarrow x'_i\}$  és un conjunt de correspondències entre punts de dues imatges i que la matriu fonamental  $F$  queda únicament determinada per la condició  $x_i'^T F x_i = 0$  per a tot  $i$ . Siguin  $(P_1, P'_1, X_{1i})$  i  $(P_2, P'_2, X_{2i})$  dues reconstruccions de les correspondències  $x_i \leftrightarrow x'_i$ . Aleshores existeix una matriu no singular  $H$  tal que  $P_2 H = P_1$ ,  $P'_2 H = P'_1$  i  $X_{2i} = H X_{1i}$  per a tot  $i$ , excepte per als punts complint  $F x_i = x_i'^T F = 0$*

*Demostració.* Per a una demostració completa veure [13, Secció 10.3]. □

**Observació 1.3.21.** Aquest teorema és vàlid per a la matriu fonamental i com que la matriu essencial és un cas particular de matriu fonamental, aleshores també és vàlid per a la matriu essencial.

Aquest teorema ens diu que si un conjunt de punts corresponents en dues vistes determina una matriu fonamental, aleshores l'escena i les càmeres es poden reconstruir a partir d'aquestes correspondències només. A més, qualsevol reconstrucció d'aquestes correspondències és projectivament equivalent.

Els punts que es troben a la recta que uneix els dos centres de càmera (línia de base) queden exclosos donat que no es poden determinar de manera única encara que puguem determinar les matrius de càmera. Aquest fet té relació amb la triangulació de punts, que es veurà en detall al tercer capítol.

## Capítol 2

# Problemes minimalis

### 2.1 Problemes minimalis

La geometria multivista juga un paper fonamental a l'hora de reconstruir una escena en 3D a partir de varies imatges. A continuació, comentem el procés general que se segueix pas a pas i com en aquest procediment apareixen polinomis que cal resoldre, i per tant sorgixen els anomenats problemes minimalis, de gran importància matemàtica dins d'aquest camp.

Donat un conjunt d'imatges d'una mateixa escena:

1. Extracció de característiques: Es comença detectant els punts o rectes distingides en cadascuna de les imatges donades. Depenent del nostre problema i algoritme en concret ens interessarà trobar uns objectes en particular. Per exemple, l'algoritme dels 5-punts que treballam en el següent capítol necessita cinc punts.
2. Correspondència de característiques: Les característiques detectades es relacionen en les diferents imatges. És a dir, s'identifica quines imatges mostren una part comú de l'escena.
3. Verificació geomètrica: Normalment no totes les correspondències que haguem trobat en el pas anterior seran correctes. Una manera habitual de verificar-les és mitjançant l'algoritme RANSAC, del que parlarem a la següent secció. També es pot fer utilitzant transformacions geomètriques. Si es pot trobar una transformació correcta que compleixi unes certes propietats de la imatge, aleshores es consideren verificades les dades. Es pot trobar més informació sobre aquesta darrera manera a [15].
4. Inicialització de la reconstrucció: Escollim una imatge de la que haguem verificat les dades i comencem a reconstruir punts.
5. Registre d'imatge: Calculem les posicions relatives de les càmeres a partir de les posicions de les càmeres ja reconstruïdes prèviament.
6. Triangulació: És el procés de determinar la posició tridimensional d'un punt del món a partir de la seva projecció en les imatges. Es fa mitjançant la intersecció dels raigs antiimatge dels punts de les imatges preses des de múltiples punts de vista. Parlarem d'aquest procés més en detall al següent capítol.

7. “Bundle ajustement”: És el procés de perfeccionament que optimitza simultàniament els punts 3D reconstruïts i els paràmetres de la càmera. Minimitza l’error de re-projecció, assegurant una reconstrucció més precisa i consistent. Es pot trobar més informació sobre aquest procés a [23].

Ens centrem en el procediment per a trobar la posició relativa de les càmeres. Recordem que no podem recuperar l’escala general de la configuració de l’escena només a partir d’imatges. Per exemple, si tenim una foto d’un vaixell no podem saber si es tracta d’un vaixell de molts metres d’eslora o de la foto d’una maqueta molt realista que només ocupa uns centímetres.

Cal que resolguem el problema d’àlgebra i geometria següent:

Siguin  $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{P}_{\mathbb{R}}^2 \times \mathbb{P}_{\mathbb{R}}^2$  el conjunt de punts corresponents entre dues imatges. Considerem  $E$  una matriu essencial que satisfà les equacions

$$\det(E) = 0 \tag{2.1.1}$$

$$EE^T E - \frac{1}{2} \text{traça}(EE^T)E = 0 \tag{2.1.2}$$

Si  $(x_i, y_i)$  és un parell de punts corresponents correcte i  $E$  és la veritable matriu essencial, aleshores per definició complirà

$$y_i^T E x_i = 0 \tag{2.1.3}$$

Les equacions (2.1.1), (2.1.2) i (2.1.3) ens defineixen un sistema de polinomis anomenat *problema minimal*, reflectint el fet que s’utilitza la quantitat mínima de dades per tal de poder determinar la solució de manera única o com a màxim obtenir un nombre finit de solucions.

Centrant-nos en la part de verificació geomètrica, el fet que moltes de les correspondències de punts (“matches”) que trobem siguin total o parcialment incorrectes ens dona greus problemes en els següents passos de l’algoritme. En general, els processos que s’utilitzen per a trobar aquestes relacions donen un nombre elevat de correspondències errònies i si es tracta d’imatges amb textures similars o repetitives els errors són encara majors.

Necessitem llavors que el nostre sistema sigui robust davant de les dades incorrectes, els anomenats punts “outliers”.

**Definició 2.1.1.** Anomenem *llindar de distància* o *distància límit* a la distància màxima del nostre model a la que es pot trobar un punt per a que considerem que s’ajusta bé al model i hi pertany. Diem que un punt és “*inlier*” si es troba dins del llindar de distància al model que volem estimar. Altrament, direm que és un punt “*outlier*”.

## 2.2 El mètode RANSAC

El *RANSAC* (*RANdom SAmples Consensus*) és un mètode iteratiu per a estimar els paràmetres d’un model matemàtic a partir d’un conjunt de dades que poden contenir alguns outliers, o dades errònies. S’utilitza en computer vision perquè se solen tenir dades amb molts outliers. A continuació, detallem l’algoritme general que es segueix.

### 2.2.1 L'algorithm RANSAC

Donats els següents paràmetres d'entrada:

*data*: un conjunt de punts observats que poden contenir dades errònies.

*model*: el model al que volem que s'ajustin les dades.

*n*: el nombre mínim de punts per a estimar els paràmetres del model.

*k*: màxim nombre d'iteracions permeses.

*t*: valor (distància) límit per a determinar si els punts de dades s'ajusten bé al model.

*d*: nombre de punts de dades propers al model necessaris per a acceptar que un model s'ajusta bé amb les dades.

#### Algoritme:

1. Seleccionem aleatòriament un subconjunt de  $s$  punts de *data*.
2. Estimem els paràmetres del *model* utilitzant el subconjunt de dades seleccionat. Usem el *model* per predir els valor de la resta de punts de dades.
3. Determinem el nombre de punts de dades que es troben a certa distància límit  $t$  del *model*.
4. Si el nombre de punts inliers és més gran que  $d$  (o que un cert percentatge del nombre total de punts de dades), reestimem el *model* utilitzant tots els punts inliers. Hem acabat.
5. Si el nombre de punts inliers és menor que  $d$ , seleccionem aleatòriament un nou subconjunt de punts i tornem al punt 2.
6. Repetim el procés  $k$  vegades o fins que el model obtingut no millori. Un cop acabat, s'escull el subconjunt de punts que més inliers tenia i es reestima el *model* utilitzant tots els punts d'aquest subconjunt.

Aquest mètode té com a benefici que serveix per estimar molts tipus de models diferents incloent lineals i no lineals i que és molt eficient per a grans conjunts de dades. A continuació, veiem les seves limitacions i com es resolen.

#### Distància llindar

A la pràctica la distància llindar s'escull empíricament. Veiem-ho rigorosament.

Suposem que volem escollir una distància llindar  $t$  (= "threshold") tal que la probabilitat de que un punt  $p$  sigui inlier és  $\alpha$ , és a dir,  $P(\text{punt } p \text{ inlier}) = \alpha$ .

Per a calcular aquesta probabilitat necessitem la distribució de probabilitat associada a la distància d'un inlier al model. Si suposem que l'error de mesura és Gaussià amb mitjana igual a 0 i desviació estàndard  $\sigma$ , aleshores podem calcular un valor de  $t$ .

En aquest cas el quadrat de la distància al punt, que anomenarem  $d_{\perp}^2$ , és la suma de variables gaussianes al quadrat, que segueix una distribució  $\chi_m^2$  amb  $m$  graus de llibertat on  $m$  és la codimensió del model.

La probabilitat que el valor d'una variable aleatòria de  $\chi_m^2$  sigui menor que  $k^2$ , és a dir,  $P(Y) \leq k^2$  amb  $Y \sim \chi_m^2$ , ve donada per la distribució acumulativa chi-quadrat

$$F_m(k^2) = \int_0^{k^2} \chi_m^2(\xi) d\xi$$

De la distribució acumulativa,

$$\begin{cases} \text{punt inlier} & \text{Si } d_{\perp}^2 < t^2 \\ \text{punt outlier} & \text{Si } d_{\perp}^2 \geq t^2 \end{cases} \quad (2.2.1)$$

amb  $t^2 = F_m^{-1}(\alpha)\sigma^2$ .

Normalment s'escull  $\alpha = 0,95$ , això vol que dir que hi ha un 95% de probabilitats que el punt sigui inlier. En altres paraules, només rebutjarem incorrectament un punt inlier el 5% de les vegades. Per a més informació consultar [13] o [8].

## Nombre de punts necessaris

Volem trobar el nombre  $N$  de mostres suficientment gran per assegurar amb probabilitat  $p$  que al menys una de les mostres aleatòries de  $s$  punts està lliure de punts outliers. S'acostuma a escollir  $p = 0,99$ .

Segui  $w$  la probabilitat de que algun punt seleccionat sigui un inlier aleshores  $1 - w = \epsilon$  és la probabilitat que sigui outlier.

Llavors seran necessàries al menys  $N$  seleccions (cadascuna de  $s$  punts) amb

$$(1 - w^s)^N = 1 - p \Rightarrow N \cdot \log(1 - (1 - \epsilon)^s) = \log(1 - p)$$

$$\Rightarrow N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)}$$

## Capítol 3

# L'algoritme dels 5-punts

En aquesta secció es presenta una solució eficient al problema clàssic dels 5 punts. Aquest problema minimal consisteix a trobar les possibles solucions per a la posició relativa de dues càmeres calibrades, donats cinc punts corresponents en les dues vistes.

L'algoritme consisteix en calcular de manera exacta els coeficients d'un polinomi de grau deu i a continuació trobar les seves arrels. Es tracta del primer algoritme adequat per a la implementació numèrica i que comparat amb el rendiment dels algorismes dels 8-punts i 7-punts dona uns bons resultats. Serveix per estimar l'estructura i el moviment en temps real amb baixa latència. El sistema en temps real utilitza només dades visuals i s'ha demostrat en importants conferències.

Suposarem que les càmeres estan calibrades, aquesta suposició ens permet simplificar el problema i reduir el nombre de paràmetres desconeguts en els càlculs de les posicions relatives de les càmeres. Recordem que diem que una càmera està calibrada si coneixem els seus paràmetres interns com ara la distància focal o el punt principal. D'aquesta manera, evitem haver de trobar i aproximar correctament aquests paràmetres i ens centrem només en trobar les posicions de les càmeres utilitzant les correspondències.

A més, al fer aquesta suposició obtenim estabilitat i unicitat de solucions. Aplicar les restriccions intrínseques de calibratge sovint ofereix una millora crucial tant de la precisió com de la robustesa de les estimacions de l'estructura i del moviment. Actualment, la manera estàndard d'aconseguir-ho és a través d'una estimació inicial sense calibrar seguida d'un refinament iteratiu per posar l'estimació d'acord amb les restriccions de calibratge.

Un cop coneguts els paràmetres interns, l'algoritme dels 5-punts és una manera més directa d'aplicar exactament les restriccions de calibratge i d'obtenir una reconstrucció. Clarament, només podem trobar la posició relativa dels punts i les càmeres, l'escala general de la configuració mai es pot recuperar únicament a partir de les imatges.

A continuació, detallem l'algoritme pas a pas, explicant també com s'implementaria per tal d'obtenir un programa eficient.

### 3.1 Pas 0: Establiment de les bases de l'algoritme

Considerem dues imatges d'una mateixa escena preses per dues càmeres des de punts de vista diferents. Siguin  $q_i = (q_1, q_2, q_3)$  i  $q'_i = (q'_1, q'_2, q'_3)$ , amb  $i = 1, \dots, 5$  els punts de la imatge de la primera i segona vista respectivament, representats per 3-vectors homogenis

i siguin  $Q_i$  els punts del món corresponents als punts de les imatges representats per 4-vectors homogenis.

Sigui  $F$  la matriu fonamental d'aquesta escena, sabem per la Definició (1.3.9) que se satisfà la igualtat  $q'^T F q = 0$ . Com que estem suposant que les càmeres són calibrades podem utilitzar que  $q'^T E q = 0$  (Definició 1.3.12), amb  $E$  la matriu essencial associada a l'escena.

Aleshores desenvolupant  $q'^T E q = 0$ ,

$$(q'_1 \quad q'_2 \quad q'_3) \begin{pmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} =$$

$$(q'_1 E_{11} + q'_2 E_{21} + q_3 E_{31} \quad q'_1 E_{12} + q'_2 E_{22} + q_3 E_{32} \quad q'_1 E_{13} + q'_2 E_{23} + q_3 E_{33}) \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} =$$

$$(q_1 q'_1 E_{11} + q_1 q'_2 E_{21} + q_1 q_3 E_{31}) + (q_2 q'_1 E_{12} + q_2 q'_2 E_{22} + q_2 q_3 E_{32}) + (q_3 q'_1 E_{13} + q_3 q'_2 E_{23} + q_3 q_3 E_{33}) = 0$$

Reagrupant termes i expressant la matriu  $E$  com un vector  $9 \times 1$ , podem definir,

$$\tilde{q} = (q_1 q'_1 \quad q_2 q'_1 \quad q_3 q'_1 \quad q_1 q'_2 \quad q_2 q'_2 \quad q_3 q'_2 \quad q_1 q'_3 \quad q_2 q'_3 \quad q_3 q'_3)^T$$

$$\tilde{E} = (E_{11} \quad E_{12} \quad E_{13} \quad E_{21} \quad E_{22} \quad E_{23} \quad E_{31} \quad E_{32} \quad E_{33})^T$$

Aleshores podem reescriure  $q'^T E q = \tilde{q}^T \tilde{E} = 0$ .

Aplicant els 5 punts corresponents a l'expressió anterior, obtenim 5 vectors  $\tilde{q}_1^T, \tilde{q}_2^T, \tilde{q}_3^T, \tilde{q}_4^T, \tilde{q}_5^T$ . Observem que cadascun d'aquests vectors és de dimensió  $1 \times 9$ .

Si col·loquem aquests vectors un a sobre de l'altre obtenim una matriu  $M$  de dimensions  $5 \times 9$  de la següent forma:

$$M = \begin{pmatrix} \tilde{q}_1^T \\ \tilde{q}_2^T \\ \tilde{q}_3^T \\ \tilde{q}_4^T \\ \tilde{q}_5^T \end{pmatrix}$$

## 3.2 Pas 1: Extracció del subespai nul d'una matriu $5 \times 9$

Recordem que el nostre objectiu és trobar la matriu  $E$  a partir dels punts corresponents en les dues vistes. Llavors sabem que s'ha de complir que  $q_i'^T E q_i = 0$  per  $i = 1, \dots, 5$  i això és el mateix que

$$\tilde{q}_1^T \tilde{E} = 0, \quad \tilde{q}_2^T \tilde{E} = 0, \quad \tilde{q}_3^T \tilde{E} = 0, \quad \tilde{q}_4^T \tilde{E} = 0, \quad \tilde{q}_5^T \tilde{E} = 0$$

Volem doncs que se satisfaci que  $M \tilde{E} = 0$ . Observem que  $\tilde{E}$  és el vector que anul·la  $M$  al fer la multiplicació i això és equivalent a trobar el subespai nul de la matriu  $M$ .



Sabem aleshores que la matriu essencial  $E$  es pot expressar com a combinació lineal de la base del subespai nul de  $M$ , és a dir,

$$E = xX + yY + zZ + wW \quad (3.2.1)$$

amb  $x, y, z, w \in \mathbb{R}$  paràmetres i  $X, Y, Z, W$  les matrius  $3 \times 3$  corresponents als generadors  $\tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{W}$  (vectors  $1 \times 9$ ) del subespai nul dret de la matriu  $M$ . Ens cal llavors calcular el subespai nul dret de  $M$ , ho farem mitjançant Gauss. Per àlgebra lineal sabem que  $\dim(\ker M) + \dim(\text{Im} M) = \dim(M)$  i, aleshores,  $\dim(\ker M) = 9 - 5 = 4$  vectors formen el subespai nul.

Per a poder realitzar el càlcul del subespai nul de manera més eficient i més estable numèricament (per a més detalls consultar [13]), primer ortogonaltzarem els vectors  $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4$  i  $\tilde{q}_5$  utilitzant la factorització QR.

### 3.2.1 Factorització QR

Sigui  $A$  una matriu de dimensions  $m \times n$ , amb  $m \geq n$ , aleshores podem factoritzar-la com  $A = QR$  amb  $Q \in \mathbb{R}^{m \times m}$  una matriu ortogonal i  $R \in \mathbb{R}^{m \times n}$  una matriu triangular superior. Anomenem a aquesta descomposició de matrius *factorització QR*.

Aquesta factorització ens proporciona una manera de calcular una base ortonormal per un conjunt de vectors. De manera general, tenim que si  $A$  té  $n$  columnes linealment independents, aleshores les primeres  $k$  columnes de la matriu  $Q$  formen una base ortonormal per a l'espai generat per les primeres  $k$  columnes de  $A$  amb  $1 \leq k \leq n$ . Es pot trobar una demostració d'aquest resultat a [21, Part 2].

Per a fer-ho utilitzarem el mètode de les matrius de Householder, també es pot fer amb rotacions de Givens però, degut a les dimensions de la nostra matriu, Householder és més eficient i per tant només veurem aquest mètode. Per a més informació sobre la descomposició utilitzant rotacions de Givens consultar [1, Secció 2.3.5] o [13, Part V].

**Definició 3.2.1.** Una matriu  $H_v \in \mathbb{R}^{m \times m}$ , s'anomena *matriu de Householder* (o transformació) si és de la forma

$$H = I - \frac{2vv^T}{v^T v}, \quad v \neq 0$$

on  $v \in \mathbb{R}^m$  i s'anomena *vector de Householder*.

Sigui  $A$  la matriu que volem factoritzar,  $e_1 = (1 \ 0 \ \dots \ 0)^T$  i  $x_i$  el vector columna  $i$ -èssim de la matriu  $A$  prenent només les darreres  $n - (i - 1)$  files de la matriu de la iteració anterior (en la primera iteració, la matriu  $A$ ). Aleshores tenim  $v_i = x_i - e_1 \|x_i\|$  i calculem la matriu de Householder  $H_i = I - \frac{2v_i v_i^T}{v_i^T v_i}$ . Llavors calculem la matriu de la iteració multiplicant  $H_i(H_{i-1} \dots A)$ . Si aquesta matriu de la iteració és triangular superior, és igual a  $R$  i hem acabat. Si no ho és, seguim iterant. Trobem  $Q$  a partir de  $A = QR$ , és a dir,  $Q = H_1^T \dots H_n^T$ , amb  $n$  l'última iteració necessària. Es pot demostrar que si es tracta d'una matriu  $m \times n$  seran necessàries com a molt  $n$  iteracions i si és una matriu quadrada  $n - 1$ .

A continuació, veiem un petit exemple amb una matriu  $3 \times 3$ , de manera semblant es realitzaria per a la nostra matriu  $M$ .

**Exemple 3.2.2.** Considerem la matriu  $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$ . Aleshores prenem  $x_1 = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix}^T$ , el primer vector columna de la matriu i tenim  $x'_1 = e_1 \|x_1\| = (\|x_1\| \ 0 \ 0)^T = (\sqrt{6} \ 0 \ 0)^T$ . Definim  $v_1 = x_1 - x'_1 = (1 - \sqrt{6} \ 2 \ 1)^T$ . Aleshores tenim la primera matriu de Householder,

$$H_1 = I - \frac{2v_1v_1^T}{v_1^Tv_1} = \begin{pmatrix} 0,41 & 0,82 & 0,41 \\ 0,82 & -0,13 & -0,56 \\ 0,41 & -0,56 & 0,72 \end{pmatrix}$$

Llavors multiplicant-la per la matriu  $A$  obtenim la primera columna reduïda,

$$H_1A = \begin{pmatrix} \sqrt{6} & 4,08 & 3,27 \\ 0 & 0,13 & -1,13 \\ 0 & 0,56 & 1,44 \end{pmatrix}$$

Prenem  $x_2 = (0,13 \ 0,56)^T$ ,  $x'_2 = (\sqrt{0,13^2 + 0,56^2} \ 0)^T$

$$\Rightarrow v_2 = x_2 - x'_2 = (0,13 - \sqrt{0,13^2 + 0,56^2} \ 0,56)^T$$

Calculem la segona matriu de Householder,

$$H'_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2v_2v_2^T}{v_2^Tv_2} = \begin{pmatrix} 0,22 & 0,98 \\ 0,98 & -0,23 \end{pmatrix} \Rightarrow H_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0,22 & 0,98 \\ 0 & 0,98 & -0,23 \end{pmatrix}$$

Aleshores multipliquem aquesta nova matriu de Householder i obtenim una matriu triangular superior,

$$H_2(H_1A) = \begin{pmatrix} \sqrt{6} & 4,08 & 3,27 \\ 0 & 0,58 & 1,16 \\ 0 & 0 & -1,43 \end{pmatrix} = R$$

Finalment, trobem la matriu  $Q$ ,

$$\begin{aligned} A = QR &\Rightarrow Q^T A = R \Rightarrow Q^T = H_2 H_1 \Rightarrow Q = (H_2 H_1)^T = H_1^T H_2^T \\ &\Rightarrow Q = \begin{pmatrix} 0,41 & 0,58 & 0,71 \\ 0,82 & -0,58 & 0 \\ 0,41 & 0,58 & -0,71 \end{pmatrix} \end{aligned}$$

En el nostre cas, volem ortogonalitzar els vectors  $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4$  i  $\tilde{q}_5$ . Considerem doncs la matriu  $A_1 = (\tilde{q}_1 \ \tilde{q}_2 \ \tilde{q}_3 \ \tilde{q}_4 \ \tilde{q}_5)$  de dimensions  $9 \times 5$  formada per aquests cinc vectors com a columnes. Aplicant la factorització  $QR$  obtenim  $A_1 = Q_1 R_1$  amb  $Q_1$  una matriu de dimensions  $9 \times 9$  i  $R_1$  de dimensions  $9 \times 5$ . Aleshores prenem els cinc primers vectors columna de la matriu  $Q_1$ , que anomenarem  $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4$  i  $\tilde{q}_5$ , que formen una base ortogonal de  $\tilde{q}_i$   $i = 1, \dots, 5$ .

### 3.2.2 Càlcul del subespai nul amb eliminació Gaussiana

Un cop trobats  $\tilde{q}_i$   $i = 1, \dots, 5$ , els vectors que formen una base ortogonal de  $\tilde{q}_i$   $i = 1, \dots, 5$ , recordem que volíem calcular una base del subespai nul de la matriu  $M$ , on  $M$  estava formada pels 5 vectors  $\tilde{q}_i$   $i = 1, \dots, 5$  posats un a sobre de l'altre.

Com hem dit abans, per motius d'eficiència, calculem equivalentment una base del subespai nul de  $M'$  formada pels vectors ortogonals  $\tilde{q}_i$  amb  $i = 1, \dots, 5$ , escrits un a sobre l'altre en forma de files. Trobarem aquesta base mitjançant eliminació Gaussiana. A continuació, veiem el mètode en detall.

Sigui  $A$  una matriu  $m \times n$ , primer construïm la matriu augmentada per files

$$\left[ \begin{array}{c} A \\ I \end{array} \right]$$

amb  $I$  la matriu identitat de dimensions  $n \times n$ . Calculant Gauss per columnes obtindrem  $\left[ \begin{array}{c} B \\ C \end{array} \right]$ , amb  $B$  una matriu escalonada. La base del subespai nul de  $A$  estarà formada pels vectors columna de  $C$  tals que la columna corresponent de la matriu  $B$  és una columna de zeros. A continuació, veiem un petit exemple.

**Exemple 3.2.3.** Sigui

$$A = \begin{bmatrix} 1 & 0 & -3 & 0 & 2 & -8 \\ 0 & 1 & 5 & 0 & -1 & 4 \\ 0 & 0 & 0 & 1 & 7 & -9 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \left[ \begin{array}{c} A \\ I \end{array} \right] = \left[ \begin{array}{cccccc|cccc} 1 & 0 & -3 & 0 & 2 & -8 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 5 & 0 & -1 & 4 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 7 & -9 & 0 & 0 & 0 & 1 & 7 & -9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

Apliquem Gauss per columnes fins a obtenir la matriu identitat a la part superior esquerra i zeros a la part superior dreta. Ens queda,

$$\left[ \begin{array}{cccccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 3 & -2 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -5 & 1 & -4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -7 & 9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

$$\Rightarrow v_1 = [3 \quad -5 \quad 1 \quad 0 \quad 0 \quad 0], v_2 = [-2 \quad 1 \quad 0 \quad -7 \quad 1 \quad 0], v_3 = [8 \quad -4 \quad 0 \quad 9 \quad 0 \quad 1]$$

Són els vectors de la base del subespai nul de la matriu  $A$ .

En el nostre cas, col·loquem la base ortogonal formada pels 5 vectors en columna, en forma de matriu

$$[\tilde{q}_1 \quad \tilde{q}_2 \quad \tilde{q}_3 \quad \tilde{q}_4 \quad \tilde{q}_5 \quad | \quad I]^T$$

amb  $I$  la matriu identitat de dimensions  $9 \times 9$ . Aleshores apliquem Gauss amb pivotatge per columnes fins que la part superior estigui escalonada.

Una vegada aplicat aquest procés tenim 4 vectors  $\tilde{X}, \tilde{Y}, \tilde{Z}, \tilde{W}$  de dimensió  $1 \times 9$  que formen una base ortogonal del subespai nul i, per tant, que generen el subespai nul. Els posem en forma de matriu  $3 \times 3$  i anomenem a aquestes matrius  $X, Y, Z, W$ .

Tenim que  $E = xX + yY + zZ + wW$  i necessitem calcular el valor de les constants  $x, y, z, w \in \mathbb{R}$ , per a poder calcular la matriu  $E$ . Observem que podem suposar sense pèrdua de generalitat que  $w = 1$ .

### 3.3 Pas 2: Expansió de les restriccions cúbiques

Considerem els vectors que formen el subespai nul calculat en el pas anterior, ara col·locats en forma de matriu,

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}, Y = \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{pmatrix}, Z = \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{pmatrix}, W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix},$$

Sabem que  $E = xX + yY + zZ + W$ , llavors sumant les matrius i multiplicant pels paràmetres obtenim que,

$$E = \begin{pmatrix} xx_{11} + yy_{11} + zz_{11} + w_{11} & \dots & xx_{13} + yy_{13} + zz_{13} + w_{13} \\ \dots & \dots & \dots \\ xx_{31} + yy_{31} + zz_{31} + w_{31} & \dots & xx_{33} + yy_{33} + zz_{33} + w_{33} \end{pmatrix}$$

Volem trobar els valors de  $x, y, z \in \mathbb{R}$ . Per a fer-ho aplicarem les anomenades restriccions cúbiques ("cubic constrains"), vistes a la Proposició (1.3.14) i el Teorema (1.3.15). Recordem que eren les següents:

$$\det(E) = 0$$

$$EE^T E - \frac{1}{2} \text{tr}a\text{ç}a(EE^T)E = 0$$

Aplicant la primera restricció aconseguim una equació de grau 3. Utilitzant la segona obtenim nou equacions de grau 3. En total 10 equacions de grau 3 en  $x, y$  i  $z$ .

#### 3.3.1 Desenvolupament de les restriccions

Per a implementar aquest pas en una màquina ens interessa calcular el determinant de  $E$  expandint els menors. Per fer-ho podem crear una funció  $o_1(p_i, p_j)$  que multipliqui dos polinomis de grau un i una altra funció  $o_2(p_i, p_j)$  que multipliqui dos polinomis amb  $\text{grau}(p_i) = 2$  i  $\text{grau}(p_j) = 1$ . Aleshores el desenvolupament en menors quedaria així,

$$\det(E) = o_2(o_1(E_{12}, E_{23}) - o_1(E_{13}, E_{22}), E_{31}) + o_2(o_1(E_{13}, E_{21}) - o_1(E_{11}, E_{23}), E_{32}) + o_2(o_1(E_{11}, E_{22}) - o_1(E_{12}, E_{21}), E_{33}) \quad (3.3.1)$$

Per a expandir l'altre restricció tindrem en compte que si tenim matrius simètriques només cal que calculem la part triangular superior. D'aquesta manera evitem repetir càlculs i a la vegada utilitzem només la memòria necessària a l'hora d'implementar aquest càlcul en una màquina.

Podem calcular la part triangular superior de la matriu simètrica  $EE^T$  amb

$$(EE^T)_{ij} = \sum_{k=1}^3 = o_1(E_{ik}, E_{jk}), \quad (3.3.2)$$

Aleshores calculem la part triangular superior de la següent matriu,

$$D_{ij} = \begin{cases} (EE^T)_{ij} - \frac{1}{2} \sum_{k=1}^3 (EE^T)_{kk} & \text{Si } i = j \\ (EE^T)_{ij} & \text{Si } i \neq j \end{cases} \quad (3.3.3)$$

Per últim, multipliquem  $D \cdot E = 0$  i com que  $D$  és simètrica només utilitzem la part triangular superior d'aquesta. Obtenim que

$$(DE)_{ij} = \sum_{k=1}^3 o_2(D_{ik}, E_{kj}) \quad (3.3.4)$$

### 3.4 Pas 3: Eliminació amb Gauss-Jordan a la matriu $A$ $10 \times 20$

Fins ara, hem calculat 10 equacions en  $x, y, z$ . Ens fixem que tenint en compte les multiplicacions de polinomis que hem fet, tenim que cada terme de les equacions polinòmiques és  $p_i p_j$  tal que  $\text{grau}(p_i) + \text{grau}(p_j) \leq 3$ .

Tenim polinomis sobre 3 variables  $x, y$  i  $z$ , on cada terme satisfà que la suma dels graus és inferior o igual a 3. Això vol dir que tenim com a màxim  $\binom{3+3}{3} = \binom{6}{3} = 20$  termes a cada equació  $(1, x, y, z, x^2, y^2, z^2, xy, xz, yz, x^3, y^3, z^3, x^2y, x^2z, y^2x, y^2z, z^2x, z^2y, xyz)$ .

Aleshores posant les 10 equacions una a sobre de l'altre ens queda una matriu de dimensions  $10 \times 20$ . Anomenarem a aquesta matriu  $A$ .

Aplicant Gauss-Jordan amb pivotatge parcial sobre la matriu  $A$ , ens queden les equacions de la forma següent:

A	$x^3$	$y^3$	$x^2y$	$xy^2$	$x^2z$	$x^2$	$y^2z$	$y^2$	$xyz$	$xy$	$x$	$y$	1
$\langle a \rangle$	1	.	.	.	.	.	.	.	.	.	[2]	[2]	[3]
$\langle b \rangle$		1	.	.	.	.	.	.	.	.	[2]	[2]	[3]
$\langle c \rangle$			1	.	.	.	.	.	.	.	[2]	[2]	[3]
$\langle d \rangle$				1	.	.	.	.	.	.	[2]	[2]	[3]
$\langle e \rangle$					1	.	.	.	.	.	[2]	[2]	[3]
$\langle f \rangle$						1	.	.	.	.	[2]	[2]	[3]
$\langle g \rangle$							1	.	.	.	[2]	[2]	[3]
$\langle h \rangle$								1	.	.	[2]	[2]	[3]
$\langle i \rangle$									1	.	[2]	[2]	[3]
$\langle j \rangle$										1	[2]	[2]	[3]

On  $[N]$  denota un polinomi de grau  $N$  en la variable  $z$ ,  $\cdot$  és un nombre real i els espais en blanc són zeros. És a dir, tenim les equacions de la forma següent,

$$\begin{aligned}
\langle a \rangle &= x^3 + \dots \\
\langle b \rangle &= y^3 + \dots \\
\langle c \rangle &= x^2y + \dots \\
\langle d \rangle &= xy^2 + \dots \\
\langle e \rangle &= x^2z + x \cdot \alpha_e(z) + y \cdot \beta_e(z) + \gamma_e(z) \\
\langle f \rangle &= x^2 + x \cdot \alpha_f(z) + y \cdot \beta_f(z) + \gamma_f(z) \\
\langle g \rangle &= y^2z + x \cdot \alpha_g(z) + y \cdot \beta_g(z) + \gamma_g(z) \\
\langle h \rangle &= y^2 + x \cdot \alpha_h(z) + y \cdot \beta_h(z) + \gamma_h(z) \\
\langle i \rangle &= xyz + x \cdot \alpha_i(z) + y \cdot \beta_i(z) + \gamma_i(z) \\
\langle j \rangle &= xy + x \cdot \alpha_j(z) + y \cdot \beta_j(z) + \gamma_j(z)
\end{aligned}$$

Amb  $\alpha_\bullet, \beta_\bullet, \gamma_\bullet$  polinomis en  $z$  de graus  $\text{grau}(\alpha_\bullet) = 2$ ,  $\text{grau}(\beta_\bullet) = 2$ ,  $\text{grau}(\gamma_\bullet) = 3$ .

Observem que podem aturar-nos quatre files abans al fer Gauss-Jordan donat que ara podem definir les equacions  $\langle k \rangle, \langle l \rangle$  i  $\langle m \rangle$  utilitzant les cancel·lacions que es produeixen al fer combinació de les equacions que hem reduït. Definim llavors,

$$\begin{aligned}
\langle k \rangle &= \langle e \rangle - z \langle f \rangle = [x^2z + x\alpha_e(z) + y\beta_e(z) + \gamma_e(z)] - z[x^2 + x\alpha_f(z) + y\beta_f(z) + \gamma_f(z)] \\
&= x(\alpha_e(z) - z\alpha_f(z)) + y(\beta_e(z) - \beta_f(z)) + (\gamma_e(z) - z\gamma_f(z))
\end{aligned}$$

$$\begin{aligned}
\langle l \rangle &= \langle g \rangle - z \langle h \rangle = [y^2z + x\alpha_g(z) + y\beta_g(z) + \gamma_g(z)] - z[y^2 + x\alpha_h(z) + y\beta_h(z) + \gamma_h(z)] \\
&= x(\alpha_g(z) - z\alpha_h(z)) + y(\beta_g(z) - \beta_h(z)) + (\gamma_g(z) - z\gamma_h(z))
\end{aligned}$$

$$\begin{aligned}
\langle m \rangle &= \langle i \rangle - z \langle j \rangle = [xyz + x\alpha_i(z) + y\beta_i(z) + \gamma_i(z)] - z[xy + x\alpha_j(z) + y\beta_j(z) + \gamma_j(z)] \\
&= x(\alpha_i(z) - z\alpha_j(z)) + y(\beta_i(z) - \beta_j(z)) + (\gamma_i(z) - z\gamma_j(z))
\end{aligned}$$

Es pot trobar una bona implementació de la reducció de Gauss-Jordan amb pivotatge (i pivotatge parcial) a [19, Capítol 2].

### 3.5 Pas 4: Expansió del determinant de la matriu B i càlcul del polinomi de grau 10

Prenem les tres equacions  $\langle k \rangle, \langle l \rangle$  i  $\langle m \rangle$  que hem definit abans i les col·loquem en una matriu  $B$  de dimensions  $3 \times 3$  que conté polinomis en  $z$ .

B	x	y	1
$\langle k \rangle$	$\alpha_e(z) - z\alpha_f(z)$	$\beta_e(z) - z\beta_f(z)$	$\gamma_e(z) - z\gamma_f(z)$
$\langle l \rangle$	$\alpha_g(z) - z\alpha_h(z)$	$\beta_g(z) - z\beta_h(z)$	$\gamma_g(z) - z\gamma_h(z)$
$\langle m \rangle$	$\alpha_i(z) - z\alpha_j(z)$	$\beta_i(z) - z\beta_j(z)$	$\gamma_i(z) - z\gamma_j(z)$

Ens fixem que ens queda una matriu amb polinomis de grau  $[N]$  en la variable  $z$  tal i com es representa a la següent taula.

B	x	y	1
$\langle k \rangle$	$[3]$	$[3]$	$[4]$
$\langle l \rangle$	$[3]$	$[3]$	$[4]$
$\langle m \rangle$	$[3]$	$[3]$	$[4]$

Observem que el vector  $[x \ y \ 1]^T$  és un vector del nucli de  $B$  i, per tant, el determinant de  $B$  s'anul·la. Calculem doncs el determinant de  $B$ . Sigui  $B$  la matriu de polinomis de la forma,

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}$$

Calculem el determinant de  $B$  desenvolupant per menors,

$$p_1 \equiv B_{12}B_{23} - B_{13}B_{22}$$

$$p_2 \equiv B_{13}B_{21} - B_{11}B_{23}$$

$$p_3 \equiv B_{11}B_{22} - B_{12}B_{21}$$

Aleshores, igualant el determinant a 0 obtenim el polinomi en  $z$ ,

$$\langle n \rangle \equiv \det(B) = p_1B_{31} + p_2B_{32} + p_3B_{33} \quad (3.5.1)$$

Voldrem llavors trobar les arrels  $z$  d'aquest polinomi per a poder calcular després  $x$  i  $y$ , i així aconseguir trobar una expressió de la matriu  $E$ .

Els polinomis  $p_1, p_2$  i  $p_3$  ens serviran, a més, per a poder calcular el valor de  $x$  i  $y$  a partir de cada arrel que trobem de  $z$  del polinomi de grau 10, en concret tenim que,

$$x = \frac{p_1(z)}{p_3(z)}, \quad y = \frac{p_2(z)}{p_3(z)} \quad (3.5.2)$$

També podem veure  $B$  com la matriu associada al sistema

$$\begin{aligned} xB_{11} + yB_{12} + B_{13} &= 0 \\ xB_{21} + yB_{22} + B_{23} &= 0 \\ xB_{31} + yB_{32} + B_{33} &= 0 \end{aligned}$$

Aleshores substituint els valors de  $x = \frac{p_1(z)}{p_3(z)}$  i  $y = \frac{p_2(z)}{p_3(z)}$  a la primera equació del sistema veiem que se satisfà correctament,

$$B_{11}B_{12}B_{23} - B_{11}B_{13}B_{22} + B_{12}B_{13}B_{21} - B_{12}B_{11}B_{23} + B_{13}B_{11}B_{22} - B_{13}B_{12}B_{21} = 0$$

Anàlogament es pot comprovar que se satisfan les altres dues equacions.

## 3.6 Pas 5: Càlcul de les arrels del polinomi de grau 10

Ara volem calcular les arrels reals del polinomi  $\langle n \rangle$ . Per fer-ho utilitzarem les successions de Sturm per a trobar els intervals en que es troben les arrels i el mètode de la bisecció per a trobar les arrels dins d'aquests intervals.

### 3.6.1 Successions de Sturm

Sigui  $p(z)$  un polinomi de coeficients reals de grau  $n \geq 2$  i suposem que  $p(z)$  no té arrels múltiples. La successió de Sturm és una successió de polinomis  $f_0, \dots, f_n$  de grau  $0, \dots, n$  respectivament, amb  $f_n$  el propi polinomi i  $f_{n-1}$  la seva derivada, és a dir,

$$\begin{aligned}f_n(z) &\equiv p(z) \\f_{n-1}(z) &\equiv p'(z)\end{aligned}$$

Per a la resta de termes de la successió, calculant la divisió Euclidiana  $\frac{f_i}{f_{i-1}}$ , obtenim el quocient  $q_i(z) = k_i z + m_i$  i el residu  $r_i(z)$ . Aleshores definim recursivament la successió de Sturm com,

$$f_{i-2}(z) = -r_i(z) \tag{3.6.1}$$

$$f_i(z) = (q_i(z))f_{i-1}(z) + r_i(z) = (k_i z + m_i)f_{i-1}(z) - f_{i-2}(z) \tag{3.6.2}$$

Com que  $\text{grau}(f_n) > \text{grau}(f_{n-1}) > \dots > \text{grau}(f_0)$  donat que anem dividint els polinomis successivament, veiem clarament que la successió acaba sempre.

**Teorema 3.6.1.** *Sigui  $p(t) \in \mathbb{R}[t]$  i  $a, b \in \mathbb{R}$  tal que  $a < b$ . Suposem que  $p(a) \neq 0$  i  $p(b) \neq 0$ . Anomenem  $s(t)$  el nombre de canvis de signe de la successió. Aleshores, el nombre d'arrels reals de  $p$  en  $(a, b)$  comptades sense multiplicitat és igual a*

$$s(f_0(a), f_1(a), \dots, f_n(a)) - s(f_0(b), f_1(b), \dots, f_n(b))$$

*Demostració.* Es pot trobar la demostració completa a [4, Secció 7.3]. □

**Observació 3.6.2.** Estem considerant un cas simplificat de la regla de Sturm. Més en general, el teorema de Sturm ens diu que si  $P$  és un polinomi lliure de quadrats aleshores el nombre d'arrels reals diferents de  $P$  en l'interval  $(a, b]$  és  $s(a) - s(b)$  amb  $a, b \in \mathbb{R}$ ,  $a < b$ .

**Exemple 3.6.3.** Considerem el polinomi  $p(x) = x^3 + 3x^2 - 1$ . Primer de tot, calculem el polinomi derivat  $p'(x) = 3x^2 + 6x$  i els polinomis resultants són  $f_3(x) = p(x)$  i  $f_2(x) = p'(x)$ .

A continuació, dividim  $f_3(x)$  entre  $f_2(x)$  i trobem el residu. Per tant,  $f_1(x) = 2x + 1$ .

Continuem així fins que obtenim una seqüència de polinomis sense residu o un polinomi constant. En aquest cas, la seqüència completa de polinomis és:

$$\begin{aligned}f_3(x) &= x^3 + 3x^2 - 1 \\f_2(x) &= 3x^2 + 6x \\f_1(x) &= 2x + 1 \\f_0(x) &= \frac{9}{4}\end{aligned}$$



Finalment, comptem el nombre de canvis de signe en la seqüència per a  $x = -\infty$  i  $x = +\infty$  per a trobar el nombre d'arrels reals del polinomi en l'interval  $(-\infty, +\infty)$ .

$$s(f_0(-\infty), f_1(-\infty), f_2(-\infty), f_3(-\infty)) = s(f_0(-A), f_1(-A), f_2(-A), f_3(-A)) = 3$$

$$s(f_0(\infty), f_1(\infty), f_2(\infty), f_3(\infty)) = s(f_0(A), f_1(A), f_2(A), f_3(A)) = 0$$

Per a un cert nombre  $A \in \mathbb{R}$ . Deduïm doncs que el polinomi té tres arrels reals, a continuació determinem uns intervals on es troben aquestes arrels.

$$s(f_0(-3), f_1(-3), f_2(-3), f_3(-3)) = s(+, -, +, -) = 3$$

$$s(f_0(-2), f_1(-2), f_2(-2), f_3(-2)) = s(+, -, 0, +) = 2$$

$$s(f_0(-1), f_1(-1), f_2(-1), f_3(-1)) = s(+, -, -, +) = 2$$

$$s(f_0(0), f_1(0), f_2(0), f_3(0)) = s(+, +, 0, -) = 1$$

$$s(f_0(1), f_1(1), f_2(1), f_3(1)) = s(+, +, +, +) = 0$$

Per tant, veiem que el polinomi té una arrel a  $(-3, -2)$ , una altre a  $(-1, 0)$  i la tercera a  $(0, 1)$ .

En el nostre cas, el polinomi  $< n >$  de grau 10 té associada una successió de Sturm formada per 11 polinomis de grau deu fins a zero. Trobem el nombre d'arrels reals comptant el nombre de canvis de signe de la successió als dos extrems de l'interval. Aleshores trobem els intervals que contenen cadascuna d'aquestes arrels. A continuació, amb el mètode de la bisecció trobem una aproximació de les arrels.

### 3.6.2 Mètode de la bisecció

Un cop trobats uns valors  $a$  i  $b$  on la funció  $f$  canvia de signe, el Teorema de Bolzano ens assegura que si  $f$  és continua en  $[a, b]$  i  $f(a)f(b) < 0$ , aleshores existeix  $\alpha \in (a, b)$  tal que  $f(\alpha) = 0$ . De fet, pot existir més d'una arrel, però suposem que mitjançant el Teorema de Sturm (3.6.1) hem trobat un interval prou petit per assegurar que només hi ha continguda una arrel.

El mètode de la bisecció construeix llavors una successió d'intervals encaixats

$$(a_0, b_0) \supset (a_1, b_1) \supset \dots \supset (a_k, b_k) \supset \dots,$$

de manera que sempre contingui l'arrel buscada i fent que l'amplitud de cada interval sigui la meitat de l'anterior. Quan l'amplitud de l'interval sigui prou petita d'acord amb la precisió desitjada per a l'arrel, aleshores podrem considerar una bona aproximació de l'arrel qualsevol dels dos extrems de l'interval.

Per a la construcció de la successió d'intervals partim de  $a_0$  i  $b_0$  tal que  $f(a_0)f(b_0) < 0$  i es considera l'abscissa mitja de  $(a_0, b_0)$ ,  $c = \frac{1}{2}(a_0, b_0)$ . Si  $f(c) = 0$ ,  $c$  és l'arrel buscada.

Si no, escollim com  $(a_1, b_1)$  l'interval  $(a_0, c)$  o el  $(c, b_0)$  segons sigui  $f(a_0)f(c) < 0$  o  $f(c)f(b_0) < 0$ , respectivament. El procés iteratiu segueix successivament.

Aquest mètode convergeix sempre a l'arrel buscada i amb precisió fixada. Com a inconvenient tenim que no aprofita cap característica de la funció  $f$ , excepte el signe, cosa que el fa força lent.

En el nostre cas, es suficient amb 30 iteracions d'aquest mètode ja que això ens garanteix trobar l'arrel amb una certa precisió en un temps fixe, tal i com s'explica a [20, Secció 3.2.5]. A més, una vegada hem trobat les arrels del polinomi en  $z$ , calculem els valors de  $x$  i  $y$  aplicant les fórmules (3.5.2).

### 3.7 Pas 6; Recuperació de $R$ i $t$ a partir de cada arrel real del polinomi

Un cop trobades les arrels del polinomi  $\langle n \rangle$  ja podem calcular la matriu essencial com  $E = xX + yY + zZ + W$  amb els vectors (en forma de matriu) que generen el subespai nul i els paràmetres  $x, y$  i  $z$ .

Volem trobar la configuració de les dues càmeres dins de l'escena a partir de la matriu  $E$ . Per fer-ho, aplicarem el Teorema (1.3.19) per a obtenir les 4 possibles configuracions de càmeres i la triangulació de punts per a determinar quina d'elles és la veritable.

Cal tenir en compte que una vegada aplicats de manera eficient tots els altres passos de l'algorítme, les operacions que s'han de dur a terme en aquest pas poden ocupar una part important del temps d'execució ja que s'han de fer per a cada arrel real trobada del polinomi  $\langle n \rangle$ .

Per a descompondre la matriu essencial en la matriu rotació i el vector de translació, utilitzarem una versió de la descomposició en valors singulars especial més adequada al nostre problema, la veiem en detall a continuació.

#### 3.7.1 Descomposició en valors singulars de $E$

Considerem la matriu essencial  $E = [e_a \ e_b \ e_c]^T$  amb  $e_a, e_b$  i  $e_c$  els vector columna de la matriu. Suposem que tenim la veritable matriu  $E$  i que, per tant, satisfà les propietats de la matriu essencial, és a dir, suposem que té  $\text{rang}(E) = 2$  i té dos valors singulars iguals diferents de zero.

En primer lloc, calculem els productes vectorials

$$e_{ab} = e_a \times e_b$$

$$e_{ac} = e_a \times e_c$$

$$e_{bc} = e_b \times e_c$$

Recordem que ho podem fer donat que estem en el subespai afí dins de l'espai projectiu i utilitzem el producte vectorial habitual.

Després, prenem els vectors resultants, calculem la seva norma  $\|e_{ab}\|, \|e_{ac}\|$  i  $\|e_{bc}\|$  i triem el que té valor més gran. Podem suposar sense pèrdua de generalitat que  $e_{ab}$  és el

més gran. Aleshores definim els vectors,

$$v_a = \frac{e_a}{\|e_a\|}, \quad v_c = \frac{e_a \times e_b}{\|e_a \times e_b\|}, \quad v_b = v_c \times v_a$$

$$u_a = \frac{Ev_a}{\|Ev_a\|}, \quad u_b = \frac{Ev_b}{\|Ev_b\|}, \quad u_c = u_a \times u_b$$

Finalment, obtenim que les matrius de la descomposició en valors singulars de  $E$  són

$$U = [u_a \quad u_b \quad u_c]$$

$$V = [v_a \quad v_b \quad v_c]$$

Com ja vam veure al primer capítol (Subsecció 1.3.4), podem suposar que la primera càmera és  $[I|0]$  i està situada a l'origen de coordenades. Aleshores, un cop trobada la descomposició en valors singulars, aplicant el Teorema (1.3.19) sabem que la segona càmera serà una de les següents quatre possibles,  $P_A \equiv [R_a|t_u]$ ,  $P_B \equiv [R_a|-t_u]$ ,  $P_C \equiv [R_b|t_u]$  o  $P_D \equiv [R_b|-t_u]$ . Recordem a més (veure Subsecció 1.3.5), que podem anar d'una d'aquestes matriu de càmera a una altre aplicant una certa transformació  $H$ . En particular, podem relacionar  $P_A$  i  $P_C$  amb  $H_t$  i  $P_B$  i  $P_D$  amb  $H_r$ ,

$$H_t = \begin{pmatrix} Id_{3 \times 3} & 0 \\ -2v_{13} - 2v_{23} - 2v_{33} & -1 \end{pmatrix} \quad \text{i} \quad H_r = \text{diag}(1, 1, 1, -1)$$

### 3.7.2 Triangulació de punts

Considerem dues matrius de càmera  $P$  i  $P'$  i siguin  $x$  i  $x'$  dos punts de les dues imatges respectivament que satisfan la restricció epipolar  $x'^T F x = 0$  (i en particular  $x'^T E x = 0$ ). Aquesta restricció es pot interpretar de manera geomètrica en termes dels raigs a l'espai corresponents a l'antiimatge dels punts de la imatge. Vam veure que aquesta igualtat ens deia que  $x'$  pertany a la recta epipolar  $Fx$  (1.3.3). Això vol dir que l'antiimatge de les rectes que passen pels punts d'imatge  $x$  i  $x'$  es troben en un pla epipolar comú. Recordem que el pla epipolar conté els centres de les dues càmeres.

Aleshores tenim dues rectes contingudes en un pla i, per tant, intersequen en un cert punt  $X$  d'aquest pla epipolar. Aquest punt  $X$  es projecta via les dues càmeres als punts  $x$  i  $x'$  en les dues imatges.

Els únics punts en l'espai tridimensional que no podem determinar a partir de les seves imatges són els punts de la línia de base entre els centres de les dues càmeres. En el cas dels punts que pertanyen a la línia de base, les antiimatges dels punts són la mateixa recta, que és igual a la línia de base. Com que són la mateixa recta, intersequen en tota la seva longitud i per tant no poden determinar el punt  $X$  de manera única. Els punts de la línia de base es projecten als epipols de les dues imatges.

Diem que les *correspondències* de punts són *ideals* si la correspondència entre els punts de les dues imatges és perfecta, sense errors. És a dir, cada punt d'una de les imatges té un punt corresponent correcte a l'altre imatge. Això ens permet trobar de manera més precisa el punt del món del que són imatge. Suposar que ens trobem en una correspondència ideal simplifica el procés de triangulació donat que ens assegura que els raigs antiimatge dels punts corresponents intersequen exactament al veritable punt tridimensional de l'espai.

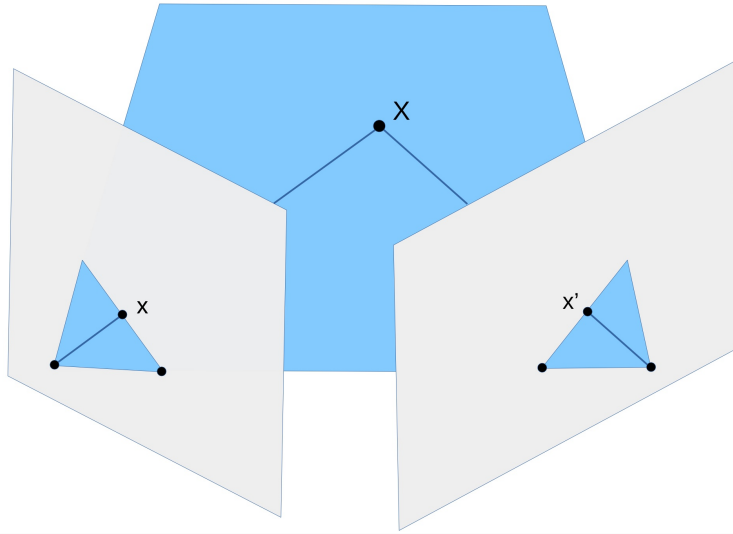


Figura 3.1: Els punts d'imatge  $x$  i  $x'$  tenen com a antiimatge un raig a l'espai. Si se satisfà la restricció epipolar  $x'^T F x = 0$ , aleshores aquests dos raigs es troben en un pla i intersequen en un punt  $X$  del 3-espai.

Observem que podem utilitzar un procediment que assumeix que les correspondències de punts són ideals donat que per a les veritables solucions, les matrius essencials trobades asseguruen la intersecció pels cinc parells de raigs.

A continuació, veiem un mètode eficient per a la triangulació d'una correspondència de punts ideal.

### Triangulació eficient d'una correspondència de punts ideal

Fins ara mitjançant l'algoritme hem trobat la matriu essencial  $E$  i les dues matrius de càmera  $[I|0]$  i  $P$ . Donat que per a una solució veritable els raigs antiimatge dels punts de les correspondències  $q \leftrightarrow q'$  segur que intersequen, aleshores no cal que minimitzem cap error mètric.

Per a correspondències de punts no ideals és recomanable realitzar una correcció prèvia per garantir la intersecció dels raigs mentre es minimitza l'error mètric. Es poden trobar algunes maneres de minimitzar l'error a les següents referències [13], [17].

Ens centrem doncs en el cas de les correspondències ideals. Per a trobar el punt triangulat intersecarem tres plans que són l'antiimatge de tres rectes de les imatges. Per trobar els plans escollim les rectes de les imatges següents:

- La recta  $a$ , és la recta epipolar corresponent al punt imatge  $q'$ , és a dir,  $a = E^T q'$
- La recta  $b$ , és la recta perpendicular a  $a$  que passa pel punt  $q$ , és a dir,

$$b = q \times (\text{diag}(1, 1, 0)a)$$

- La recta  $c$ , és la recta perpendicular a  $Eq$  que passa per  $q'$ , és a dir,

$$c = q' \times (\text{diag}(1, 1, 0)Eq)$$

Aleshores, les antiimatges d'aquestes rectes són els plans  $A = [a^T \ 0]$ ,  $B = [b^T \ 0]$  i  $C = P^T c$ , respectivament per a les rectes  $a, b$  i  $c$ .

Busquem ara la intersecció d'aquests tres plans. De manera formal, tenim que la intersecció és la contracció  $Q_l = \epsilon_{ijkl} A^i B^j C^k$  entre el tensor èpsilon i els tres plans, on el tensor èpsilon és un tensor tal que  $\epsilon_{ijkl} A^i B^j C^k D^l = \det([A \ B \ C \ D])$ . Per a més detalls sobre càlculs amb tensors consultar [13, Part V].

Llavors, podem definir la direcció del raig antiimatge de la intersecció entre les rectes  $a$  i  $b$ , o sigui,  $d \equiv a \times b$ . El punt de l'espai és la intersecció entre aquesta recta  $d$  i el pla  $C = [C_1 \ C_2 \ C_3 \ C_4]$ . Obtenim doncs que el punt triangulat és,

$$Q = [d^T C_4 \ -(d_1 C_1 + d_2 C_2 + d_3 C_3)]^T \quad (3.7.1)$$

**Observació 3.7.1.** Ens fixem que en el cas de les correspondències de punts ideals tenim que  $d = q$  i per tant no cal calcular  $a, b$  ni tampoc  $A, B$ .

En el cas de les correspondències no ideals, aquesta idea de prendre les rectes  $a, b$  i  $c$ , troba el punt del món en el raig antiimatge de  $q'$  que minimitza l'error de reprojecció en la primera imatge. Triangula punts del món a l'infinít (ja que els punts del món estan en  $\mathbb{P}^3$ ) correctament i és invariant a transformacions projectives de l'espai del món,  $\mathbb{P}^3$ .

### 3.7.3 Desambigüitat de la reconstrucció

Per a determinar quina de les 4 possibles càmeres  $P_A, P_B, P_C$  o  $P_D$  pertany a la veritable configuració de càmeres aplicarem la següent restricció.

**Definició 3.7.2.** Anomenem “*cheirality constrain*” a la restricció que considera que els punts d'escena, és a dir, els punts del món que apareixen representats a les imatges, han d'estar situats davant de les càmeres.

En primer lloc, triem un punt qualsevol dels nostres 5 punts corresponents inicials. Aleshores, utilitzant el parell de càmeres ( $[I|0], P_A$ ) triangulem el punt i obtenim un punt de l'espai que anomenarem  $Q = (Q_1 \ Q_2 \ Q_3 \ Q_4)$ .

Llavors apliquem la *cheirality constrain* per a comprovar si la configuració ( $[I|0], P_A$ ) és la correcta. Definim doncs,

$$\begin{aligned} c_1 &\equiv Q_3 Q_4 \\ c_2 &\equiv (P_A Q)_3 Q_4 \end{aligned}$$

Aleshores, segons el signe d'aquests productes tenim que,

Si  $c_1 \equiv Q_3 Q_4 < 0$  aleshores el punt es troba darrera de la primera càmera.

Si  $c_2 \equiv (P_A Q)_3 Q_4 < 0$  aleshores el punt es troba darrera de la segona càmera.

**Observació 3.7.3.** Utilitzem la notació  $(P_A Q)_3$  per a referir-nos a la tercera component del punt resultant d'aplicar  $P_A$  a  $Q$ .

Llavors, tenint en compte que volem que el punt estigui davant de les dues càmeres, tenim els següents quatre casos:

- i)* Si  $c_1 > 0$  i  $c_2 > 0$  llavors  $P_A$  i  $Q$  corresponen a la configuració real.

- ii)* Si  $c_1 < 0$  i  $c_2 < 0$  aleshores apliquem  $H_r$  i obtenim que  $P_B$  i  $H_r Q$  corresponen a la configuració real.
- iii)* Si  $c_1 c_2 < 0$  tenim dos possibles subcasos:
  - a)* Si  $Q_3(H_t Q)_4 > 0$  apliquem  $H_t$  i obtenim que  $P_C$  i el punt  $H_t Q$  corresponen a la veritable configuració.
  - b)* Si  $Q_3(H_t Q)_4 < 0$  apliquem  $H_r$  i  $H_t$  obtenim que  $P_D$  i el punt  $H_r H_t Q$  són de la configuració real.

Ens fixem que en el primer cas *i)* tenim que el punt es troba davant de les dues càmeres i per tant aquestes són  $([I|0], P_A)$ . En el segon cas *ii)*, com que  $c_1$  i  $c_2$  són negatius el punt es troba darrera de les dues càmeres i per tant cal aplicar  $H_r$  donat que volem el reflex de  $P_A$ , que és  $P_B$ .

En el cas *iii)* tenim que  $c_1 < 0$  i  $c_2 > 0$  o  $c_1 > 0$  i  $c_2 < 0$  i per tant el punt es troba darrera d'una de les dues càmeres, per tant, cal trobar la parella invertida aplicant  $H_t$ . Tenim els subcasos *a)* i *b)* donat que si  $Q_3(H_t Q)_4 > 0$  aleshores ara el punt ja es troba davant de les dues càmeres i hem acabat, però si  $Q_3(H_t Q)_4 < 0$  tenim que el punt encara es troba darrera d'una de les càmeres. En aquest cas, necessitem trobar a més el reflex de la parella invertida aplicant també  $H_r$ .

### 3.8 Aplicació de RANSAC a l'algorisme dels 5-punts

Utilitzem l'algorisme dels 5-punts juntament amb l'aplicació del mètode RANSAC en el cas en que tenim dues o tres vistes. Prenem un cert nombre de mostres aleatòries, cadascuna d'aquestes conté cinc parells de punts corresponents en les dues vistes. L'algorisme dels cinc punts s'aplica llavors a cada mostra i així es generen una sèrie d'hipòtesis.

Aleshores s'escull la millor hipòtesi de totes segons una mesura robusta. Tal i com es descriu a [8], aquesta manera d'aplicar el RANSAC s'utilitza per raons d'eficiència. Es pot trobar un exemple del tipus de test que s'utilitza amb RANSAC a [3], a més, en aquesta referència es pot veure com augmenta l'eficàcia i la rapidesa del mètode RANSAC al fer-ho d'aquesta forma a través de diversos experiments amb dades generades sintèticament i dades reals. Finalment, la millor hipòtesi es poleix mitjançant un refinament iteratiu, per a més detalls consultar [23].

## Capítol 4

# Conclusions

En aquest treball hem tractat diversos aspectes fonamentals de la geometria multivista en la computer vision. Hem vist els conceptes bàsics de la geometria en una única vista, la seva relació amb la geometria projectiva i la geometria en dues vistes en els casos de càmeres calibrades i no calibrades. També hem estudiat el problema de les correspondències de punts en dues imatges.

Ens hem centrat en l'estudi de l'algoritme dels 5-punts, que dona una solució efectiva per a trobar la matriu essencial i recuperar la configuració de les càmeres a partir de només 5-punts corresponents. Gràcies a les millores afegides que hem comentat a cada pas de l'algoritme s'aconsegueix que tingui una major precisió numèrica i sigui més eficient que els seus mètodes predecessors.

El temps de computació depèn parcialment del nombre de solucions reals, ja que un cop trobades les arrels del polinomi (3.5.1) cal calcular les possibles càmeres reconstruïdes amb DVS per a les matrius essencials calculades a partir de cadascuna de les arrels. A partir d'un estudi portat a terme amb aquest algoritme basat en  $10^5$  correspondències de punts i configuracions de càmeres aleatòries (es pot consultar a [20, Secció 6]) podem dir que de mitjana tenen 4.55 arrels reals. A més, es va trobar un exemple en que cinc punts corresponents donaven lloc a 10 solucions diferents físicament vàlides.

El fet de que l'algoritme sigui tant ràpid permet que sigui utilitzat per a la reconstrucció de l'estructura i el moviment d'un video en temps real amb una baixa latència.

En comparació amb els algoritmes dels 8, 7 i 6-punts, precursors al nostre, l'algoritme dels 5-punts ens dona en general molt millors resultats. Cal tenir en compte que en el cas dels algoritmes de 5 i 6-punts necessitem conèixer els paràmetres interns de les càmeres, mentre que en els altres dos casos no és necessari.

Recordem que trobàvem la matriu essencial com  $E = xX + yY + zZ + wW$ , (veure 3.2.1). L'algoritme dels 8-punts [11], té una única solució ja que podem suposar que  $x = y = z = 0$  i calcular directament  $W$ . En l'algoritme dels 7-punts [13, Capítol 11], es suposa que  $x = y = 0$  i es calculen  $Z$  i  $W$  introduint (3.2.1) a (1.3.10) obtenint una equació en  $z$  senzilla, on trobarem com a molt tres solucions. Finalment, en el cas de l'algoritme dels 6-punts [18], apliquem a (3.2.1) la restricció (1.3.15) i aconseguim un sistema d'equacions que resollem.

A partir dels experiments realitzats comparant els diferents algoritmes a [20, Secció 6] amb dades creades artificialment per a que hi hagués un nivell d'error realista i a la vegada desafiant, es va trobar que l'algoritme dels 5-punts era força més sensible amb la

translació que amb la rotació. Tot i així els resultats eren més que acceptables, els errors eren de l'ordre de 0,01 graus en la rotació i de menys d'un píxel en la translació. És a dir, els resultats per a aquest algoritme són molt bons numèricament i amb menys error que en els altres algoritmes.

Una altre pregunta que ens podem fer és com de precises han de ser les dades del calibratge de la càmera per a que el mètode funcioni correctament. Es va veure que si, per exemple, la distància focal era massa imprecisa aleshores els mètodes de 5 i 6-punts, que són els que suposen conegut el calibratge, no eren bons. En canvi els mètodes de 7 i 8-punts al no necessitar aquesta informació donaven millors resultats. Per tant, en cas de tenir dades errònies en el calibratge, és millor escollir un mètode que no sigui calibrat. En canvi si coneixem de manera precisa el calibratge, els resultats de l'algoritme dels 5-punts són molt millors.

Hem vist doncs que l'algoritme dels 5-punts és un algoritme molt eficient per a resoldre el problema minimal de la posició relativa de dues càmeres donat un conjunt de 5 punts corresponents en les dues vistes. A través de l'estudi d'aquest problema hem vist la importància dels problemes minimal dins la geometria multivista.



# Bibliografia

- [1] A. Aubanell, A. Benseny, A. Delshams. *Útiles Básicos de Cálculo Numérico*. Editorial Labor, 1993.
- [2] E. Casas-Alvero. *Analytic Projective Geometry*. Editorial European Mathematical Society, 2014.
- [3] O. Chum, J. Matas. *Randomized RANSAC with  $T_{d,d}$  Test*. BMVC, 448-457, 2002.
- [4] C. D'Andrea. *Cálculo de Raíces Reales de Polinomios*. Agost 2006. <http://www.ub.edu/arcades/realroots.pdf>
- [5] M. Demazure. *Sur deux problemes de reconstruction*. RR-0882, INRIA. ffinria-00075672, 1988.
- [6] L. Eldén, L. Wittmeyer-Koch. *Numerical Analysis, an introduction*. Academic Press, 1990.
- [7] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [8] M. Fischler, R. Bolles. *Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography*. Commun. Assoc. Comp. Mach., 24:381-395, 1981.
- [9] G. Gallego, E. Mueggler, P. Sturm. *Translation of "Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung" by Erwin Kruppa*. December 2017.
- [10] G. H. Golub, C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, second edition, 1989.
- [11] R. Hartley. *In Defense of the Eight-Point Algorithm*. PAMI, 19(6):580-593, 1997.
- [12] R. Hartley, P. Sturm. *Triangulation*. Computer vision and image understanding, 68(2): 146-157, 1997.
- [13] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*. Editorial Cambridge University Press, 2a edició, 2004.
- [14] A. Heyden, G. Sparr. *Reconstruction from Calibrated Cameras - a New Proof of the Kruppa-Demazure Theorem*. Jour. of Mathematical Imaging and Vision, 10-1 20, 1999.
- [15] J. Kileel, K. Kohn. *Snapshot of Algebraic Vision*. arXiv:2210.11443v1, October 2022.

- [16] E. Kruppa. *Zur Ermittlung eines Objektes aus Zwei Perspektiven mit Innerer Orientierung*. Sitz.-Ber. Akad. Wiss., Wien, Math. Naturw. Kl., Abt. IIa., 122:1939-1948, 1913
- [17] J. Oliensis, Y. Genc. *New Algorithms for Two-Frame Structure from Motion*. ICCV, 737-744 ,1999.
- [18] J. Philip. *A Non-Iterative Algorithm for Determining all Essential Matrices Corresponding to Five Point Pairs*. Photogrammetric Record, 15(88):589-599, 1996.
- [19] W. Press, S. Teukolsky, W. Vetterling, B. Flannery. *Numerical Recipes in C*. Cambridge Univ. Press, 1988.
- [20] D. Nistér. *An efficient solution to the five-point relative pose problem*. IEEE transactions on Pattern Analysis and Machine Intelligence, 26(6): 756-770, 2004.
- [21] L. N. Trefethen, D. Bau III. *Numerical linear algebra*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1997.
- [22] P. Torr, A. Fitzgibbon, A. Zisserman. *The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences*. IJCV, 32(1):27-44, 1999.
- [23] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon. *Bundle Adjustment - a Modern Synthesis*. LNCS, Springer Verlag. 1883:298-375, 2000.