Universidade do Estado do Rio de Janeiro
Instituto Politécnico

David da Costa de Pinho

Geometria trifocal em reconstrução 3D

Nova Friburgo
2016

David da Costa de Pinho

Geometria trifocal em reconstrução 3D

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, da Universidade do Estado do Rio de Janeiro. Área de concentração: Matemática Aplicada e Computação Científica.

Orientador: Prof. Dr. Ricardo Fabbri

Nova Friburgo

2016

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/D

D979 Pinho, David da Costa de
Geometria trifocal em reconstrução 3D / David da Costa de Pinho. – Nova Friburgo, 2016-
128 f.

Orientador: Prof. Dr. Ricardo Fabbri
Dissertação (Mestrado) – Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Programa de Pós-Graduação em Modelagem Computacional, 2016.

1. Visão computacional.. 2. Reconstrução 3D.. 3. Geometria trifocal.. I. Prof. Dr. Ricardo Fabbri. II. Universidade do Estado do Rio de Janeiro. III. Instituto Politécnico. IV. Título

CDU 02:141:005.7

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura Data

David da Costa de Pinho

Geometria trifocal em reconstrução 3D

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, da Universidade do Estado do Rio de Janeiro. Área de concentração: Matemática Aplicada e Computação Científica.

Aprovada em 19 de fevereiro de 2016.

Banca Examinadora:

Prof. Dr. Ricardo Fabbri (Orientador)
Instituto Politécnico – UERJ

Prof^a. Dr. Lis Ingrid Roque Lopes Custódio
Instituto Politécnico - UERJ

Prof. Dr. Nelson Violante de Carvalho
Universidade Federal do Rio de Janeiro

Nova Friburgo

2016

AGRADECIMENTOS

Agradeço aos senhores José Saulo Pereira de Pinho, Vilma Nadir Mozer, Pablo Mozer de Pinho, Patrick Mozer de Pinho e Jussara dos Santos por todo o apoio logístico como transporte, estadia e alimentação, muito importantes para a manutenção do curso de Pós-Graduação e confecção da dissertação, já que não recebi apoio financeiro de qualquer órgão de fomento a pesquisas acadêmicas.

Agradeço ao prof. Ricardo Fabbri por se disponibilizar como orientador, disponibilizar máquina de uso pessoal e espaço em seu laboratório. Por, várias vezes, priorizar minha pesquisa em detrimento de funções burocráticas no âmbito do IPRJ e até mesmo em detrimento de suas pesquisas pessoais. Por transmitir suas reflexões e filosofia acerca de estudos acadêmicos.

Agradeço ao prof. Ricardo Carvalho de Barros pelos ensinos em álgebra linear que foram cruciais em pelo menos três momentos durante minhas investigações teóricas.

Agradeço ainda ao prof. Luís Fernando de Oliveira, do Departamento de Física Aplicada e Termodinâmica do Instituto de Física Armando Dias Tavares - UERJ, por ter disponibilizado os códigos Latex para a formatação final da dissertação de acordo com as normas de elaboração de teses e dissertações da UERJ.

Na falta de uma interpretação geométrica, um grande conjunto de equações algébricas pode ser tão informativo quanto um bloco de códigos de máquina.

Stephen Maybank

RESUMO

PINHO, D. C. *Geometria trifocal em reconstrução 3D*. 2016. 128 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2016.

Neste trabalho nós investigamos alguns benefícios na utilização da geometria trifocal aplicada em sistemas de reconstrução 3D multifocal e estimação de modelos de múltiplas câmeras, em oposição à utilização da geometria epipolar com pares de câmeras. Apresentamos a interpretação e o detalhamento matemático dos pontos mais importantes de dois artigos recentes e essenciais, objetivando resolver problemas trifocais de estrutura de curvas no futuro. Equanto um deles mostra a aplicação do quatérnion de Hamilton e das bases de Gröbner para a cálculo da pose de uma câmera dadas estruturas correspondentes 3D–2D, o outro é a abordagem mais eficiente computacionalmente (até a presente data) na reconstrução das câmeras num sistema trifocal. Os primeiros capítulos e os apêndices apresentam as teorias básicas sobre geometria projetiva, álgebra linear e geometria algébrica necessárias ao avanço das pesquisas no campo da reconstrução 3D trifocal e multifocal e sistemas de determinação de estruturas a partir do movimento.

Palavras-chave: Visão computacional. Reconstrução 3D. Geometria trifocal. Bases de Gröbner.

ABSTRACT

PINHO, D. C. *Trifocal geometry in 3D reconstruction.* 2016. 128 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2016.

In this work we investigate some advantages of using trifocal geometry applied to 3D multiview reconstruction systems and multiple camera model estimation, as opposed to using pairwise epipolar geometry. We present a detailed exposition and interpretation of the most important mathematical tools in two key recent papers, aiming at tackling trifocal problems for general curvilinear structures in the future. While one of them shows Hamilton's quaternions and Gröbner bases applications for computing the pose of one camera given 3D–2D corresponding structures, the other one is the most computationally efficient approach (to date) for trifocal reconstruction of camera systems. The first chapters and appendices present the basic theory about projective geometry, linear algebra and algebraic geometry useful for advancing research in the field of 3D multiview reconstruction and structure from motion systems.

Keywords: Computer vision. 3D reconstruction. Trifocal geometry. Gröbner Bases.

SUMÁRIO

	INTRODUÇÃO	9
1	**GEOMETRIA DE UMA E DUAS CÂMERAS**	16
1.1	**O espaço projetivo em duas dimensões**	16
1.1.1	A reta	16
1.1.2	O ponto	16
1.1.3	A cônica	20
1.1.4	Cônica dual ou cônica reta	21
1.1.5	Cônicas degeneradas	22
1.1.6	A relação polo-polar	23
1.1.7	Transformações projetivas em $\mathbb{P}^2$	24
1.1.8	A geometria projetiva de uma dimensão	26
1.1.9	Subgrupo de transformações projetivas	28
1.2	**O espaço projetivo em três dimensões**	32
1.2.1	A transformação de pontos e planos em $\mathbb{P}^3$	34
1.2.2	Direções e o plano no infinito.	35
1.2.3	A cônica absoluta	36
1.2.4	A quádrica dual absoluta	37
1.3	**A câmera P**	38
1.4	**A ação projetiva de uma câmera P**	45
1.4.1	A ação projetiva de P em retas	46
1.4.2	A ação projetiva de P em quádricas	48
1.4.3	A geometria bifocal	50
1.4.4	A homografia induzida por um plano	52
1.4.5	A matriz fundamental F	54
1.4.6	Propriedades da matriz fundamental	56
1.4.7	Configuração canônica das câmeras	57
1.4.8	Ambiguidade projetiva das câmeras, dada F	59
2	**RECONSTRUÇÃO DE UMA CÂMERA P: KUANG E ÅSTRÖM**	61
2.1	**Formulação do problema**	61
2.2	**Sistema de equações polinomiais**	66
3	**GEOMETRIA TRIFOCAL**	68
3.1	**O Problema**	68
3.2	**Transferência epipolar**	70
3.3	**O tensor trifocal e as relações de incidência**	71
3.3.1	Homografia induzida por um plano	73

3.3.2 Relações de incidência entre pontos e retas 74
3.3.3 Relação de incidência para retas epipolares 76
3.4 **Transferências usando o tensor trifocal** 77
3.4.1 Transferência de pontos 77
3.4.2 Transferência de retas 79
3.5 **Reconstrução das câmeras a partir do tensor trifocal** 79
3.5.1 Extração da matriz fundamental 80
3.5.2 Extraindo as matrizes das câmeras 80
4 **GEOMETRIA TRIFOCAL CALIBRADA: NISTÉR E SCHAFFALITZKY** 83
4.1 **O teorema 1: homografia da reta epipolar** 83
4.1.1 Nova abordagem para a homografia da reta epipolar 85
4.2 **Dedução da IAC** 87
4.2.1 Calibração e a imagem da cônica absoluta 87
4.3 **O teorema 2: restrição Kruppa** 88
4.4 **Mudança de coordenadas projetivas** 91
4.4.1 Corregistro de pontos pertencentes a diferentes planos de imagem 91
4.5 **O teorema 3: pontos e epipolos co-cônicos** 92
4.6 **O teorema 4: interseção das tangentes** 93
4.7 **O teorema 5: projeção de ω em B** 94
4.7.1 Representação canônica de uma cônica 95
4.7.2 Representação canônica paramétrica 98
4.7.3 Demonstração do teorema 99
4.8 **O teorema 6: projeção de ω' em B** 100
4.9 **O teorema 7: a função que relaciona os epipolos** 100
4.9.1 Feixe de cônicas 101
4.9.2 Estabelecimento e demonstração do teorema 102
4.10 **O teorema 8: a curva C** 104
4.11 **Teoremas 9 a 20: a curva G e suas propriedades** 104
4.12 **Teoremas 21 a 25: orientação e convergência** 108
4.12.1 A restrição de orientação 108
4.12.2 A restrição de convergência 108
4.13 **Incluindo a terceira imagem** 111
5 **CONCLUSÃO** 113
REFERÊNCIAS 114
APÊNDICE A – Alguns conceitos e definições de álgebra linear 117
APÊNDICE B – Alguns conceitos e definições de geometria algébrica . 120

INTRODUÇÃO

A Visão Computacional é uma área de pesquisa que se ocupa em utilizar imagens em 2D de um cenário, obtidas de uma ou mais câmeras digitais para produzir uma construção virtual em 3D, ou extrair qualquer outro tipo de informação útil como o cruzamento de informações entre imagens. A captação da geometria 3D de um cenário a partir de imagens produzidas por câmeras, analisando a disparidade entre elementos correspondentes nessas imagens, é chamada de visão *estereoscópica*, da qual a visão computacional se utiliza para desenvolver modelos adequados à descrição geométrica tanto da cena 3D em questão como das câmeras que fazem a projeção da cena na imagem 2D.

A geometria projetiva é a principal teoria matemática usada em visão computacional tridimensional. Com a introdução de algumas definições e conceitos, a geometria projetiva pode ser estudada em termos de álgebra linear. Apesar de possuir algumas limitações, tal abordagem traz alguns benefícios de modelagem e é amplamente utilizada por pesquisadores dessa área. Em pesquisas recentes, Kuang e Åström (2013), verificou-se a eficácia do uso de técnicas de geometria algébrica em resoluções de sistemas de equações polinomiais, oriundos de problemas de visão computacional 3D. Fabbri e Kimia (2011) introduziram uma abordagem em termos de geometria diferencial, onde é possível aplicar o cálculo diferencial trabalhando diretamente com as equações algébricas. Essa abordagem permite um maior poder de modelagem matemática e computacional do espaço tridimensional através das imagens.

Motivação.

As motivações para a pesquisa estão ligadas às suas aplicações em diversas áreas. A extração de informação da geometria 3D de um ambiente pode, por exemplo, ser utilizada na determinação da orientação e posição de veículos autônomos, para que os mesmos identifiquem a trajetória para os seus objetivos considerando a possibilidade de obstáculos. Um exemplo famoso, segundo Matthies et al. (2007) , é a aplicação em pesquisas interplanetárias onde foram usados algoritmos de visão computacional em veículos de exploração de Marte (Mars Exploration Rover - MER, em inglês), observado na Figura 1.

Uma aplicação mais recente foi a reconstrução virtual de objetos, estátuas ou monumentos de valor histórico e cultural que foram destruídos pela ação do tempo ou do homem (possivelmente em guerras), através da aquisição de imagens em acervos ou na

Figura 1 - Visão computacional em Marte.

Legenda: Visão computacional em pesquisas interplanetárias na missão NASA/JPL Mars Exploration Rover (MER).
Fonte: Matthies et al. (2007).

internet. Um exemplo é o Museu de Mosul [1], no Iraque, atacado pelo Estado Islâmico em 2015, que teve algumas de suas obras reconstruídas como podemos visualizar na Figura 2.

As aplicações de visão computacional são bastante variadas, com pesquisas em outras áreas como astronomia, medicina e química. Para se ter uma ideia, Ballard e Brown (1982) já apresentavam imagens e tabelas com resumos de aplicações em oito áreas diferentes no início dos anos 1980.

Novos avanços em reconstrução 3D vêm sendo realizados constantemente como observamos em Fabbri e Kimia (2015), que apresentam uma abordagem para a reconstrução de um desenho 3D (3D *drawing* em inglês) de uma cena. Isto é, um conjunto de fragmentos de curvas em 3D interligadas de forma que preserve suas relações espaciais, capturadas em forma de gráficos a partir de um grande conjunto de dados multifocal. A ideia base é aprimorar uma abordagem anterior (3D curve *sketch*) divulgada pelos mesmos autores, Fabbri e Kimia (2010a), conforme exemplo visualizado na Figura 3. A maioria dos métodos de reconstrução não utilizam geometria diferencial de curvas, mas se baseiam em correlacionar pontos de interesse através das imagens e produzem uma desorganizada

[1] Outros exemplos podem ser encontrados em http://www.projectmosul.org/

Figura 2 - O leão de Mosul.

Legenda: A reconstrução do Leão de Mosul, um dos artefatos destruídos por extremistas religiosos em 2015 no Iraque.
Fonte: http://www.projectmosul.org/

nuvem de pontos em 3D, Figura 5. Tais métodos obtêm sucesso em ambientes controlados, com cenas em larga escala e imagens ricas em textura, mas não podem ser aplicados em configurações gerais. Não podem reconstruir superfícies suaves e homogêneas nem suas fronteiras devido a esparsidade de pontos de interesse, como também não podem reconstruir regiões que se alteram drasticamente com a mudança de luz ambiente. Exemplos de imagens sujeitas a essas limitações são dados na Figura 4. Daí são utilizadas outras técnicas de melhoramento dos resultados, como a interpolação de pontos e texturização, mas que não fornecem um resultado plenamente satisfatório. Podemos ver as fases desse processo em um exemplo no conjunto de imagens na Figura 5.

Portanto, ainda é necessário que se continuem as pesquisas em visão computacional, e o presente trabalho tem a finalidade de, primeiramente, detalhar as abordagens de artigos recentes na busca de novas combinações de ferramentas matemáticas que possam melhorar as atuais abordagens e atenuar o esforço computacional. Segundo, a verificação dos benefícios obtidos com o uso da geometria trifocal em transferência de pontos e reconstrução 3D em comparação com a geometria epipolar (mais usada atualmente), tudo com o objetivo de auxiliar na construção de uma nova abordagem usando geometria diferencial multifocal em condições realísticas.

Figura 3 - 3D *drawing*

Legenda: A abordagem transforma um conjunto de imagens calibradas no desenho 3D, na forma de um gráfico consistindo de segmentos de curvas que se ligam nas junções.

Fonte: Fabbri e Kimia (2015).

Figura 4 - Objetos com pouca textura.

Legenda: Exemplos de objetos que não podem ser reconstruídos através da abordagem tradicional usando pontos de interesse e a geometria epipolar.
Fonte: Fabbri et al. (2012).

Visão geral e objetivos.

A seção 1 engloba a teoria básica de geometria projetiva utilizada em visão computacional 3D, com abordagens em termos de álgebra linear e usando o tipo de notação mais difundido entre pesquisadores da área. Além dos conceitos básicos, também são apresentados alguns conceitos e definições um pouco mais avançados indispensáveis ao entendimento das demais seções da dissertação, tudo com a finalidade de facilitar a compreensão do leitor sem a necessidade de consulta em outras publicações. Em seguida, há a apresentação da geometria epipolar, num sistema de visão estéreo com duas câmeras, bem como extração da matriz fundamental e a reconstrução das câmeras, usando essa abordagem bifocal que é mais comum do que a abordagem trifocal.

Na seção 2 temos o detalhamento de um artigo recente e representativo para a reconstrução de uma câmera a partir de objetos geométricos em 3D e suas respectivas imagens em 2D, Kuang e Åström (2013). A importância dessa publicação se verifica na utilização dos quatérnions de Hamilton para a parametrização da matriz de rotação da câmera, pois desta forma a matriz que possui nove componentes é parametrizada com apenas quatro variáveis. Outra vantagem é utilização das bases de Gröbner e da matriz de ação para solução computacional de sistemas de equações polinomiais de grau elevado e com muitas variáveis.

A geometria trifocal e suas características são abordadas na seção 3, juntamente com a descrição dos benefícios do uso dessa geometria num sistema com três câmeras em comparação com uso da geometria epipolar entre cada par de câmeras. Após as com-

Figura 5 - Método tradicional de reconstrução.

Legenda: Em cima, duas imagens extraídas de um vídeo (utilizado em reconstrução 3D) da estátua da Medusa em Sagalassos, Turquia. No meio, a nuvem de pontos obtida após a reconstrução 3D juntamente com as posições da câmera. Em baixo, aplicação de interpolação de pontos e texturização.

Fonte: http://www.cs.unc.edu/ marc/research.html

parações são apresentados métodos para a extração das matrizes fundamentais e métodos de reconstrução das câmeras a partir do tensor trifocal.

Na seção 4 há o detalhamento matemático da abordagem mais eficiente, até a presente data, para a reconstrução 3D das câmeras num sistema trifocal calibrado, conforme Nistér e Schaffalitzky (2006). O artigo é bastante denso, com 25 teoremas, e apresenta a reconstrução de duas câmeras num sistema bifocal utilizando uma intrincada rede de conhecimentos de geometria projetiva. Depois da reconstrução das duas primeiras câmeras, é utilizada a reconstrução 3D de pontos para a reconstrução da terceira câmera, num procedimento similar ao apresentado na seção 2, ou seja, os dois artigos quase se complementam. Esse trabalho deixa bem clara a dificuldade de se realizar a reconstrução num sistema com três imagens de quatro pontos 3D numa cena.

No apêndice A são fornecidas ferramentas de álgebra linear acompanhadas de algumas definições mais restritas à assimilação da dissertação. No apêndice B há uma breve introdução aos conceitos básicos de geometria algébrica seguidos da apresentação (informal e em termos de exemplos) de uma teoria para resolução de sistemas de equações polinomiais com várias variáveis.

1 GEOMETRIA DE UMA E DUAS CÂMERAS

1.1 O espaço projetivo em duas dimensões

Para a compreensão mínima das técnicas mais importantes investigadas nessa dissertação, faz-se necessária a revisão condensada de um arcabouço notacional e conceitual que está disperso na literatura.

1.1.1 A reta

Sabe-se que uma reta no plano $\mathbb{R}^2$ pode ser representada pela equação $a\,x+b\,y+c=0$, onde a reta fica perfeitamente determinada pelos valores das constantes a, b e c. Desta forma, podemos representar retas através de vetores, e assim a reta $a\,x+b\,y+c=0$ seria representada por $\boldsymbol{l}=(a,b,c)^\top \in \mathbb{R}^3$, utilizando o símbolo em negrito $\boldsymbol{l}$ para indicar tal vetor escrito em coluna, por padrão. Nota-se que a relação entre uma dada reta e o seu respectivo vetor não é biunívoca, pois o vetor $k\,(a,b,c)^\top$, tal que $k \neq 0$, representa a reta $k\,a\,x+k\,b\,y+k\,c=0$ que é a mesma reta $a\,x+b\,y+c=0$. Temos, então, infinitos vetores que representam uma mesma reta e formam uma *classe de equivalência*, onde essa classe pode ser representada por qualquer um de seus vetores. Os vetores de uma classe de equivalência, definida pela multiplicação por um escalar, são conhecidos como vetores *homogêneos*. O conjunto de classes de equivalência de vetores em $\mathbb{R}^3-(0,0,0)^\top$ forma o *Espaço Projetivo* $\mathbb{P}^2$. O vetor $(0,0,0)^\top$ foi excluído por não representar reta alguma. Após essas considerações, dizemos que uma reta no plano projetivo é representada pelo vetor $(a,b,c)^\top$ em *coordenadas homogêneas*.

1.1.2 O ponto

Sabemos que em $\mathbb{R}^2$ os pontos são representados através de pares ordenados do tipo (x,y), e assim cada ponto pode ser identificado como um vetor $(x,y)^\top$. Os vetores em coordenadas homogêneas que se referem a pontos serão representados pelo símbolo em negrito $\mathbf{x}$, configurando uma terceira dimensão cuja coordenada é 1. Desse jeito, $\mathbf{x}=(x,y,1)^\top$. Sabemos também que um ponto (x,y) pertence a uma reta $(a,b,c)^\top$ se, e

Figura 6 - Plano projetivo $\mathbb{P}^2$.

Legenda: O plano π representa o espaço projetivo $\mathbb{P}^2$. Pontos e retas pertencentes a esse espaço são representados, respectivamente, por vetores e planos que passam pela origem do $\mathbb{R}^3$. O ponto ideal é representado por um vetor que pertence ao plano $x\,y$, o qual é paralelo ao plano π. O plano $x\,y$ representa a reta no infinito.

Fonte: Adaptado, Hartley e Zisserman (2004).

somente se, $a\,x + b\,y + c = 0$, ou, matricialmente:

$$(x, y, 1)^\top \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0 \qquad \text{ou} \qquad \mathbf{x}^\top \boldsymbol{l} = 0.$$

Ou seja, temos um ponto de $\mathbb{R}^2$ representado como um vetor com três coordenadas. Observe que para $k \neq 0$, temos:

$$(k\,x, k\,y, k)^\top \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0 \qquad \Leftrightarrow \qquad (x, y, 1)^\top \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0.$$

Portanto, variando k, podemos considerar os vetores em coordenadas homogêneas $(k\,x, k\,y, k)^\top \in \mathbb{P}^2$, como representantes do mesmo ponto $(x, y) \in \mathbb{R}^2$, e podemos resgatar nossa representação original aplicando o procedimento $(x/k, y/k)^\top$, pois $k \neq 0$. Geometricamente, podemos pensar no plano projetivo como um conjunto de retas passando pela origem do $\mathbb{R}^3$, onde cada reta representa um único ponto, que é a interseção dessa reta com o plano

$\mathbb{P}^2$. Analogamente, retas em $\mathbb{P}^2$ são formadas pela interseção de $\mathbb{P}^2$ com planos em $\mathbb{R}^3$. Na Figura 6, observamos como a interseção da reta com o plano define um ponto, assim como a interseção de dois planos definem uma reta. Incluímos na representação o ponto ideal $(x, y, 0)^\top$ e a reta no infinito que serão definidos mais adiante.

Como a representação de um ponto ou uma reta por vetores em coordenadas homogêneas não é alterada pela multiplicação por um escalar, para determinarmos um ponto ou uma reta precisamos apenas de duas componentes do vetor, pois a terceira componente pode ser configurada como escala. Assim, dizemos que um ponto ou uma reta tem dois graus de liberdade.

Retas determinadas por dois pontos e a interseção de duas retas.

A representação de retas em coordenadas homogêneas nos permite calcular a interseção de duas retas usando o produto vetorial. Sabemos que o produto vetorial de dois vetores, digamos $\boldsymbol{l} \times \boldsymbol{l}'$ em R^3 resulta num terceiro vetor que é ortogonal ao plano definido pelos vetores $\boldsymbol{l}$ e $\boldsymbol{l}'$. Em particular, esse terceiro vetor $\boldsymbol{l} \times \boldsymbol{l}'$ será ortogonal a qualquer dos vetores $\boldsymbol{l}$ e $\boldsymbol{l}'$, assim seu produto escalar é zero

$$\boldsymbol{l} \cdot (\boldsymbol{l} \times \boldsymbol{l}') = \boldsymbol{l}' \cdot (\boldsymbol{l} \times \boldsymbol{l}') = 0.$$

O produto escalar também pode ser representado por multiplicação matricial,

$$\boldsymbol{l}^\top (\boldsymbol{l} \times \boldsymbol{l}') = \boldsymbol{l}'^\top (\boldsymbol{l} \times \boldsymbol{l}') = 0, \tag{1}$$

e sendo $\mathbf{x}$ o ponto de interseção entre as duas retas temos

$$\boldsymbol{l}^\top \mathbf{x} = \boldsymbol{l}'^\top \mathbf{x} = 0 \tag{2}$$

onde, comparando 1 e 2,

$$\mathbf{x} = \boldsymbol{l} \times \boldsymbol{l}'.$$

Com argumento análogo podemos constatar que dois pontos definem uma reta através do produto vetorial entre eles, pois basta verificar que dois pontos $\mathbf{x}$ e $\mathbf{x}'$ pertencem a reta definida por

$$\boldsymbol{l} = \mathbf{x} \times \mathbf{x}'.$$

Pontos ideais e a reta no infinito.

Duas retas paralelas no espaço Euclidiano têm equações

$$a\,x + b\,y + c = 0 \qquad \text{e} \qquad a\,x + b\,y + c' = 0$$

com coordenadas homogêneas

$$\boldsymbol{l} = (a, b, c)^\top \qquad \text{e} \qquad \boldsymbol{l}' = (a, b, c')^\top.$$

Computando a interseção entre as duas retas e ignorando o fator de escala $(c' - c)$ temos

$$\begin{aligned} \boldsymbol{l} \times \boldsymbol{l}' &= (a, b, c)^\top \times (a, b, c')^\top \\ &= (c' - c)\,(b, -a, 0)^\top \\ &= (b, -a, 0)^\top, \end{aligned}$$

e se tentarmos encontrar as coordenadas não homogêneas do ponto de interseção temos

$$\left(\frac{b}{0}, \frac{-a}{0}\right)^\top,$$

o que não faz sentido matematicamente mas sugere que o ponto tem coordenadas que tendem ao infinito. Vetores homogêneos $(x, y, z)^\top$ com $z \neq 0$ correspondem a pontos finitos em R^2 e, se $z = 0$, os pontos são conhecidos como *pontos ideais* ou *pontos no infinito* no plano projetivo.

Se tomarmos um ponto no infinito genericamente $\mathbf{x} = (x, y, 0)^\top$ percebemos que ele pertence a uma determinada reta $\boldsymbol{l}_\infty = (0, 0, 1)^\top$, pois

$$\boldsymbol{l}_\infty^\top \mathbf{x} = (0, 0, 1) \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = 0,$$

e tal reta é conhecida como *reta no infinto*. Assim, no plano projetivo temos que retas paralelas se encontram em ponto no infinito e o conjunto de pontos no infinito constituem a reta no infinito.

Um fato interessante sobre pontos no infinito é que eles são as direções das retas no plano projetivo $\mathbb{P}^2$. Um feixe de retas paralelas, que se intersectam num mesmo ponto no infinito, têm a mesma direção. Observa-se que a interseção de uma reta qualquer $\boldsymbol{l} = (a, b, c)^\top$ com a reta no infinito $\boldsymbol{l}_\infty = (0, 0, 1)^\top$,

$$\boldsymbol{l} \times \boldsymbol{l}_\infty = (b, -a, 0),$$

é um vetor que, em coordenadas não homogêneas $(b, -a)$, é ortogonal ao vetor (a, b), que

é o vetor normal à reta dada $\boldsymbol{l} = (a, b, c)^\top$. Assim $(b, -a)$ constitui-se a direção da reta $\boldsymbol{l}$, e como a reta no infinito contém todos os pontos do tipo $(b, -a, 0)$, dizemos que a reta no infinito é o conjunto de direções das retas no plano projetivo $\mathbb{P}^2$.

1.1.3 A cônica

Em Geometria Euclidiana, as cônicas são de três tipos principais: elipse, hipérbole e parábola. São definidas, algebricamente, por uma equação do segundo grau em duas variáveis, considerando coordenadas não homogêneas:

$$a\,x^2 + b\,x\,y + c\,y^2 + d\,x + e\,y + f = 0.$$

Sabemos que um ponto pertence à cônica se ele é solução da equação acima, a qual pode ser representada utilizando multiplicação matricial e vetores em coordenadas homogêneas,

$$(x, y, 1) \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0.$$

Podemos generalizar essas coordenadas homogêneas multiplicando-se por um fator x_3 e definindo $x_1 = x_3\,x$ e $x_2 = x_3\,y$, e a equação fica

$$a\,x_1^2 + b\,x_1\,x_2 + c\,x_2^2 + d\,x_1\,x_3 + e\,x_2\,x_3 + f\,x_3^2 = 0.$$

Novamente, em notação matricial:

$$(x_1, x_2, x_3) \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \qquad \text{ou} \qquad \mathbf{x}^\top C\,\mathbf{x} = 0.$$

Já que um ponto pertence à cônica se, e somente se, satisfaz a última equação, temos que C fica definida como a matriz homogênea que representa uma cônica no plano projetivo $\mathbb{P}^2$,

$$C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}.$$

Percebemos que as cônicas são representadas por matrizes 3×3 simétricas e, portanto, possuem seis variáveis. Usando uma dessas variáveis para fixar a escala, temos que as cônicas possuem cinco graus de liberdade, pois podemos por exemplo, dividir todas as

coordenadas da matriz C por f. Portanto, são necessários cinco pontos para que a mesma fique determinada.

Retas tangentes a cônicas.

Um conceito utilizado mais à frente nas investigações da calibração e estimação da pose relativa de câmeras é o de cônicas e envelope de retas tangentes. Uma reta tangente a uma cônica num ponto $\mathbf{x}$ qualquer tem a simples forma

$$\boldsymbol{l} = C\mathbf{x}.$$

De fato, se a reta $\boldsymbol{l}$ passa por um ponto $\mathbf{x}$ temos que $\mathbf{x}^\top \boldsymbol{l} = 0$, e se esse ponto $\mathbf{x}$ pertence à cônica C então $\mathbf{x}^\top C\,\mathbf{x} = 0$. Se $\mathbf{x}$ é o único ponto de interseção da reta com a cônica, comparando as duas relações temos que $\boldsymbol{l} = C\mathbf{x}$ é a tangente procurada. Agora veremos que se a reta passa por mais um ponto na cônica, então toda a reta pertence à cônica e não é uma tangente. Se a reta passa também por um outro ponto $\mathbf{y}$ na cônica, temos que $\mathbf{y}^\top C\,\mathbf{y} = 0$ e $\boldsymbol{l}^\top \mathbf{y} = 0$. Mas como $\boldsymbol{l} = C\mathbf{x}$ e considerando que C é simétrica, temos

$$\boldsymbol{l}^\top \mathbf{y} = (C\mathbf{x})^\top \mathbf{y} = \mathbf{x}^\top C\,\mathbf{y} = 0.$$

Agora considerando todos os pontos em $\boldsymbol{l} = (\mathbf{x}+\alpha\mathbf{y})$ parametrizados por α, vamos verificar que todos eles pertencem à cônica C usando as relações anteriores:

$$(\mathbf{x}+\alpha\mathbf{y})^\top C\,(\mathbf{x}+\alpha\mathbf{y}) = \mathbf{x}^\top C\,\mathbf{x} + \alpha\mathbf{y}^\top C\,\mathbf{x} + \mathbf{x}^\top C\,\alpha\mathbf{y} + \alpha\mathbf{y}^\top C\,\alpha\mathbf{y} = 0,$$

já que cada uma das parcelas é zero. Assim toda a reta $\boldsymbol{l}$ ligando $\mathbf{x}$ e $\mathbf{y}$ está em C, que neste caso é dita ser uma cônica degenerada.

1.1.4 Cônica dual ou cônica reta

A cônica definida por $\mathbf{x}^\top C\,\mathbf{x} = 0$ é chamada cônica de pontos já que é definida por uma equação que envolve pontos. Mas como pontos e retas têm a mesma representação por vetores de três componentes no plano projetivo $\mathbb{P}^2$, podemos também definir uma cônica por retas, chamada cônica dual ou de retas denotada por C^*, onde $\boldsymbol{l}^\top C^*\,\boldsymbol{l} = 0$. Tal notação indica que C^* é a matriz adjunta de C e, sendo C simétrica e não singular, $C^* = C^{-1}$. Com efeito, sendo $\boldsymbol{l}$ tangente à cônica C temos $\boldsymbol{l} = C\,\mathbf{x}$, e sendo C não singular

Figura 7 - Cônica envelope.

Legenda: A cônica C é o envelope de todas as retas $\boldsymbol{l}_i$ que satisfazem $\boldsymbol{l}^\top C^* \boldsymbol{l} = 0$.

Fonte: O autor.

temos que o ponto de tangência é $\mathbf{x} = C^{-1}\boldsymbol{l}$. Como $\mathbf{x} \in C$ temos que $\mathbf{x}^\top C\,\mathbf{x} = 0$, assim

$$\begin{aligned}
\mathbf{x}^\top C\,\mathbf{x} &= (C^{-1}\boldsymbol{l})^\top C\,(C^{-1}\boldsymbol{l}) \\
&= \boldsymbol{l}^\top C^{-\top}\boldsymbol{l} \\
&= \boldsymbol{l}^\top C^{-1}\boldsymbol{l} \\
&= 0.
\end{aligned}$$

Lembrando que $C^{-\top} = C^{-1}$ pois C é simétrica. A relação $\boldsymbol{l}^\top C^* \boldsymbol{l} = 0$ indica que as retas $\boldsymbol{l}_i$ são tangentes à cônica C, e por isso C é conhecida como cônica envelope, como podemos visualizar na Figura 7.

1.1.5 Cônicas degeneradas

Quando a matriz que representa a cônica não tem posto completo ocorrem cônicas degeneradas, daí a cônica será duas retas caso a matriz tenha posto 2 ou uma reta repetida caso a matriz tenha posto 1. Apresentamos a seguir dois exemplos de cônicas degeneradas, uma cônica ponto e uma cônica dual. A cônica

$$C = \boldsymbol{l}\,\mathbf{m}^\top + \mathbf{m}\,\boldsymbol{l}^\top$$

é uma cônica ponto composta por duas retas $\boldsymbol{l}$ e $\mathbf{m}$. Assim, se um ponto $\mathbf{x} \in \boldsymbol{l}$ então $\boldsymbol{l}^\top \mathbf{x} = 0$ e esse ponto deve satisfazer $\mathbf{x}^\top C\,\mathbf{x} = 0$. De fato,

$$\begin{aligned}
\mathbf{x}^\top C\,\mathbf{x} &= \mathbf{x}^\top(\boldsymbol{l}\,\mathbf{m}^\top + \mathbf{m}\,\boldsymbol{l}^\top)\,\mathbf{x} \\
&= \mathbf{x}^\top \boldsymbol{l}\,\mathbf{m}^\top\mathbf{x} + \mathbf{x}^\top\mathbf{m}\,\boldsymbol{l}^\top\mathbf{x} \\
&= 0\,\mathbf{m}^\top\mathbf{x} + \mathbf{x}^\top\mathbf{m}\,0 \\
&= 0.
\end{aligned}$$

A argumentação é analoga caso $\mathbf{x} \in \mathbf{m}$. A equação de uma cônica dual degenerada é composta por pontos e a matriz correspondente tem posto 2 caso seja composta por dois pontos, ou posto 1 caso seja composta por um ponto repetido,

$$C^* = \mathbf{x}\,\mathbf{y}^\top + \mathbf{y}\,\mathbf{x}^\top.$$

De forma análoga à argumentação anterior podemos demosntrar que C^* define um equação em retas,

$$\boldsymbol{l}^\top\, C^*\boldsymbol{l} = 0.$$

1.1.6 A relação polo-polar

Vimos anteriormente que uma reta $\boldsymbol{l}$ é tangente a uma cônica num ponto $\mathbf{x}$ se

$$\boldsymbol{l} = C\,\mathbf{x}.$$

Mas se o ponto $\mathbf{x}$ não pertence à cônica C então temos uma relação denominada *polo-polar*, onde a reta $\boldsymbol{l}$ é chamada reta *polar* de $\mathbf{x}$, e $\mathbf{x}$ é chamado o *polo* da reta $\boldsymbol{l}$, conforme a ilustração na Figura 8.

A reta polar $\boldsymbol{l} = C\,\mathbf{x}$ intercepta a cônica C em dois pontos $\mathbf{y}_1$ e $\mathbf{y}_2$, e as retas tangentes à cônica C nesses dois pontos se interceptam em $\mathbf{x}$. De fato, considere o ponto $\mathbf{y}_1 \in C$ e a reta tangente a esse ponto $\mathbf{m} = C\,\mathbf{y}_1$. Essa tangente contém o ponto $\mathbf{x}$ se

$$\mathbf{x}^\top\mathbf{m} = 0 \qquad \text{ou} \qquad \mathbf{x}^\top C\,\mathbf{y}_1 = 0,$$

e, usando a simetria da cônica C e aplicando a transposição em $\mathbf{x}^\top C$, temos

$$\mathbf{x}^\top C\,\mathbf{y}_1 = (C\,\mathbf{x})^\top\mathbf{y}_1 = 0,$$

mostrando que o ponto $\mathbf{y}_1$ pertence à reta $\boldsymbol{l} = C\,\mathbf{x}$. Assim, a reta $\boldsymbol{l}$ intercepta a cônica C no ponto $\mathbf{y}_1$, no qual a tangente a C contém o ponto $\mathbf{x}$. Observa-se que se o ponto

Figura 8 - Relação polo-polar.

(a) (b)

Legenda: (a) A reta $\boldsymbol{l} = C\,\mathbf{x}$ é a reta polar de $\mathbf{x}$, e o ponto $\mathbf{x} = C^{-1}\boldsymbol{l}$ é o polo da reta $\boldsymbol{l}$. Os pontos $\mathbf{x}$ e $\mathbf{y}$ são chamados conjugados, pois obedecem à relação $\mathbf{y}^\top C\,\mathbf{x} = 0$. (b) Duplicação da figura para visualização da simetria da relação polo-polar. O ponto $\mathbf{x}$ pertence à reta polar de $\mathbf{y}$.

Fonte: O autor.

$\mathbf{x}$ se aproxima da cônica C, os dois pontos $\mathbf{y}_1$ e $\mathbf{y}_2$ vão se aproximando um do outro, e quando ponto $\mathbf{x}$ pertence à conica C a reta polar se torna a reta tangente. O argumento é análogo para $\mathbf{y}_2$.

Pontos conjugados.

Quando um ponto $\mathbf{y}$ qualquer pertence à reta polar $\boldsymbol{l} = C\,\mathbf{x}$ temos que

$$\mathbf{y}^\top \boldsymbol{l} = \mathbf{y}^\top C\,\mathbf{x} = 0,$$

e quaisquer dois pontos que satisfazem a relação $\mathbf{y}^\top C\,\mathbf{x} = 0$ são chamados *conjugados* com relação à cônica C. Repare que o ponto y não precisa necessariamente pertencer à cônica. Mais ainda, se um ponto $\mathbf{y}$ está na reta polar de $\mathbf{x}$, então $\mathbf{x}$ está na reta polar de $\mathbf{y}$. Como vimos, se $\mathbf{y}$ está na reta polar de $\mathbf{x}$ vale a relação $\mathbf{y}^\top C\,\mathbf{x} = 0$, e tomando a transposta temos que $\mathbf{x}^\top C\,\mathbf{y} = 0$, pois C é simétrica.

1.1.7 Transformações projetivas em $\mathbb{P}^2$

A tranformação projetiva também conhecida como projetividade, colineação ou homografia, é uma função (invertível) que leva pontos no plano $\mathbb{P}^2$ em pontos no plano $\mathbb{P}^2$ preservando a colinearidade desses pontos. Mais formalmente, uma função $h : \mathbb{P}^2 \to \mathbb{P}^2$

é uma transformação projetiva se, e somente se, existe uma matriz $H_{3\times 3}$ onde para cada ponto $\mathbf{x} \in \mathbb{P}^2$ temos que $h(\mathbf{x}) = H\,\mathbf{x}$ pertencente a $\mathbf{x} \in \mathbb{P}^2$. De fato, dados três pontos $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$ na mesma reta $\boldsymbol{l}$, temos que $\boldsymbol{l}^\top \mathbf{x}_i = 0$ para $i = 1, 2$ e 3. Seja dada ainda uma matriz invertível $H_{3\times 3}$. Definindo

$$\boldsymbol{l}' = H^{-\top}\boldsymbol{l} \qquad \text{e} \qquad \mathbf{x}'_i = H\,\mathbf{x}_i,$$

vemos que todos os pontos $\mathbf{x}'_i$ pertencem à reta $\boldsymbol{l}'$, pois

$$\begin{aligned} \boldsymbol{l}'^\top \mathbf{x}'_i &= (H^{-\top}\boldsymbol{l})^\top H\,\mathbf{x}_i \\ &= \boldsymbol{l}^\top H^{-1} H\,\mathbf{x}_i \\ &= \boldsymbol{l}^\top \mathbf{x}_i = 0 \end{aligned}$$

preservando assim, a colinearidade dos pontos. A implicação direta é demasiadamente grande. A argumentação mostra que qualquer transformação linear aplicada a coordenadas homogêneas é uma transformação projetiva em $\mathbb{P}^2$, ou seja, a transformação projetiva é simplesmente uma transformação linear em $\mathbb{R}^3$. Podemos ver que o efeito da matriz H não é alterado pela multiplicação por um escalar diferente de zero na equação

$$\mathbf{x}' = H\,\mathbf{x},$$

pois os pontos permanecem colineares após a transformação, alterando apenas a escala. Assim, a matriz H é homogênea, e apenas a razão entre os seus elementos é significante, ou seja, são nove elementos mas oito graus de liberdade. Os pontos $\mathbf{x} \leftrightarrow \mathbf{x}'$ são chamados *pontos correspondentes*.

Transformação de retas.

Na argumentação anterior vimos que a reta $\boldsymbol{l}'$ foi definida na forma $\boldsymbol{l}' = H^{-\top}\boldsymbol{l}$. Tal definição não foi em vão, pois a definição nesta forma garante que os pontos depois de transformados permaneçam colineares. Assim, a transformação projetiva de pontos nos dá essa regra para transformação projetiva de retas.

Transformação de cônicas.

Invertendo a função $\mathbf{x}' = H\,\mathbf{x}$ temos que $\mathbf{x} = H^{-1}\mathbf{x}'$, e substituindo $\mathbf{x}$ na equação matricial de uma cônica $\mathbf{x}^\top C\,\mathbf{x} = 0$ temos

$$\begin{aligned} \mathbf{x}^\top C\,\mathbf{x} &= (H^{-1}\mathbf{x}')^\top C\,H^{-1}\mathbf{x}' \\ &= \mathbf{x}'^\top H^{-\top} C\,H^{-1}\mathbf{x}' \\ &= \mathbf{x}'^\top C'\mathbf{x}' = 0 \end{aligned}$$

onde $\mathbf{x}'^\top C'\mathbf{x}' = 0$ é a equação matricial da cônica depois de aplicada a transformação

Figura 9 - Razão cruzada.

Legenda: (a) Cada grupo de quatro pontos está relacionado aos outros grupos por uma transformação projetiva em $\mathbb{P}$. Como a razão cruzada é uma invariante, tais grupos possuem a mesma razão cruzada. (b) A razão cruzada das quatro retas é uma invariante projetiva do plano que as contém (ver Subseção 1.1.9). As retas contendo os quatro pontos podem ser vistas em $\mathbb{P}^2$ como planos de imagens, objetos planares, relacionados por centros de projeção.

Fonte: Hartley e Zisserman (2004).

projetiva, com $C' = H^{-\top} C\, H^{-1}$ sendo a regra para a transformação de cônicas.

1.1.8 A geometria projetiva de uma dimensão

Vamos abrir um parênteses na discussão sobre o plano $\mathbb{P}^2$ para discorrer brevemente sobre a *reta projetiva* $\mathbb{P}$. A reta projetiva tem muita similaridade com o plano projetivo, pois pontos na reta são representados em coordenadas homogêneas (vetores com duas componentes) e são projetivamente transformados de acordo com uma homografia $H_{2\times 2}$. Na reta projetiva $\mathbb{P}$ os pontos têm as seguintes notação e transformação projetiva:

$$\overline{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \text{e} \qquad \overline{\mathbf{x}}' = H\,\overline{\mathbf{x}}. \tag{3}$$

Pontos no infinito têm a componente $x_2 = 0$.

A razão cruzada.

A definição a seguir é uma *invariante* (ver Subseção 1.1.9) básica da geometria projetiva. Dados quatro pontos numa reta, conforme a Figura 9, a razão cruzada é

definida por

$$\text{cross}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4) = \frac{|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_2||\overline{\mathbf{x}}_3\, \overline{\mathbf{x}}_4|}{|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_3||\overline{\mathbf{x}}_2\, \overline{\mathbf{x}}_4|},$$

onde

$$|\overline{\mathbf{x}}_i\, \overline{\mathbf{x}}_j| = \det \begin{bmatrix} x_{i1} & x_{j1} \\ x_{i2} & x_{j2} \end{bmatrix}.$$

Vamos verificar que a razão cruzada é uma invariante sob transformação projetiva, ou seja, dada uma homografia como a Equação 3 temos que

$$\text{cross}(\overline{\mathbf{x}}'_1, \overline{\mathbf{x}}'_2, \overline{\mathbf{x}}'_3, \overline{\mathbf{x}}'_4) = \text{cross}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4). \tag{4}$$

De fato,

$$\begin{aligned}
\text{cross}(\overline{\mathbf{x}}'_1, \overline{\mathbf{x}}'_2, \overline{\mathbf{x}}'_3, \overline{\mathbf{x}}'_4) &= \frac{|\overline{\mathbf{x}}'_1\, \overline{\mathbf{x}}'_2||\overline{\mathbf{x}}'_3\, \overline{\mathbf{x}}'_4|}{|\overline{\mathbf{x}}'_1\, \overline{\mathbf{x}}'_3||\overline{\mathbf{x}}'_2\, \overline{\mathbf{x}}'_4|} \qquad \text{como} \qquad \overline{\mathbf{x}}' = H\,\overline{\mathbf{x}} \\
&= \frac{|H\,\overline{\mathbf{x}}_1\, H\,\overline{\mathbf{x}}_2||H\,\overline{\mathbf{x}}_3\, H\,\overline{\mathbf{x}}_4|}{|H\,\overline{\mathbf{x}}_1\, H\,\overline{\mathbf{x}}_3||H\,\overline{\mathbf{x}}_2\, H\,\overline{\mathbf{x}}_4|} \\
&= \frac{|H\,[\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_2]||H\,[\overline{\mathbf{x}}_3\, \overline{\mathbf{x}}_4]|}{|H\,[\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_3]||H\,[\overline{\mathbf{x}}_2\, \overline{\mathbf{x}}_4|]} \qquad [\overline{\mathbf{x}}_i\, \overline{\mathbf{x}}_j] \quad \text{é matriz} \\
&= \frac{|H|\,|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_2|\,|H|\,|\overline{\mathbf{x}}_3\, \overline{\mathbf{x}}_4|}{|H|\,|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_3|\,|H|\,|\overline{\mathbf{x}}_2\, \overline{\mathbf{x}}_4|} \\
&= \frac{|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_2||\overline{\mathbf{x}}_3\, \overline{\mathbf{x}}_4|}{|\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_3||\overline{\mathbf{x}}_2\, \overline{\mathbf{x}}_4|} \\
&= \text{cross}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4),
\end{aligned}$$

e portanto a Equação 4 segue.

Como corolário, segundo Semple e Kneebone (1952), existe homografia H onde $\overline{\mathbf{x}}'_i = H\,\overline{\mathbf{x}}_i$ se, e somente se, $\text{cross}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4) = \text{cross}(\overline{\mathbf{x}}'_1, \overline{\mathbf{x}}'_2, \overline{\mathbf{x}}'_3, \overline{\mathbf{x}}'_4)$.

Retas concorrentes.

Vimos que pontos e retas têm representações similares em $\mathbb{P}^2$ usando vetores homogêneos com três componentes. Daí, fazendo as devidas alterações, todo teorema relacionado a pontos são válidos também para retas, e dizemos que existe a *dualidade* entre

pontos e retas no plano $\mathbb{P}^2$. Analogamente, quatro retas concorrentes em um único ponto têm geometria em $\mathbb{P}$ dual com quatro pontos colineares, assim a razão cruzada entre essas retas também é uma invariante sob projetividade. Conforme a Figura 9, o valor da razão cruzada entre as quatro retas é dado pelos pontos de interseção de uma quinta reta transversal a essas quatro, ou seja

$$\text{cross}(\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4) = \text{cross}(\overline{\mathbf{x}}_1, \overline{\mathbf{x}}_2, \overline{\mathbf{x}}_3, \overline{\mathbf{x}}_4).$$

Tal asserção é devida ao teorema de *Pappus* conforme Springer (1964).

1.1.9 Subgrupo de transformações projetivas

As transformações projetivas se dividem em subgrupos de transformações com características mais específicas. Esses subgrupos são as *isometrias*, *similaridades*, *afinidades* e a própria *transformação projetiva*. Cada subgrupo mantém *invariante* algumas propriedades do objeto ao qual a transformação está sendo aplicada. Por exemplo, a distância entre dois pontos num quadrado se mantém a mesma após a aplicação de uma transformação isométrica (translação, rotação e possível reflexão), mas pode variar após a aplicação de uma transformação de similaridade, a qual pode gerar um quadrado maior ou menor que o original. Portanto, distância entre dois pontos é uma invariante sob isometria mas não é sob similaridade.

Transformação isométrica.

É representada pela matriz

$$\begin{bmatrix} \epsilon\, cos\,\theta & -sen\,\theta & t_x \\ \epsilon\, sen\,\theta & cos\,\theta & t_y \\ 0 & 0 & 1 \end{bmatrix},$$

onde $\epsilon = \pm 1$. Se $\epsilon = 1$ a isometria preserva a orientação e é chamada *Transformação Euclidiana* (composição de translação e rotação). Se $\epsilon = -1$ a isometria reverte a orientação. A transformação Euclidiana é o tipo de isometria mais importante, denotada por H_E e representada pela matriz em forma de blocos

$$H_E = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

Sendo $R_{2\times 2}$ uma matriz de rotação, $\mathbf{t}$ um vetor com duas componentes e $\mathbf{0} = (0,0)^\top$ um vetor nulo, essa transformação tem três graus de liberdade, um para o ângulo de rotação em R e dois para o vetor de translação. Assim, são necessários no mínimo dois pontos correspondentes $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ para determinarmos a matriz da transformação Euclidiana. As invariantes são distância entre dois pontos, ângulo entre duas retas e área.

Tranformação de similaridade.

A transformação de similaridade é denotada por H_S e é representada pela matriz

$$H_S = \begin{bmatrix} s\cos\theta & -s\,sen\,\theta & t_x \\ s\,sen\,\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix},$$

que neste caso é a composição de uma transformação Euclidiana com uma escala (sem reflexão) s. A transformação de similaridade preserva a forma dos objetos e tem quatro graus de liberdade, com o escalar s contando para um grau a mais que a transformação Euclidiana. A similaridade também pode ser determinada usando dois pontos correspondentes e tem a matriz em bloco

$$H_S = \begin{bmatrix} s\,R & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

Com a mudança na escala das figuras, a área e a distância entre dois pontos não é mais a mesma após a transformação, mas os ângulos entre as retas não mudam, e pontanto, ângulo é uma invariante sob similaridade. Em particular, retas paralelas permanecem paralelas. A razão entre as distâncias antes e após a transformação permanece a mesma por conta do cancelamento da escala, e analogamente, a razão entre as áreas também permanece a mesma.

Transformação afim.

A transformação afim, ou afinidade, é uma transformação linear seguida de uma translação, denotada por H_A e representada pela matriz

$$H_A = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}.$$

Figura 10 - Transformação afim.

(a) (b) (c) (d) (e)

Legenda: (a) Um quadrado. (b) Rotação $R(50°)$. (c) Escala não uniforme D com $\lambda_1 = 0,5$ e $\lambda_2 = 1,1$. (d) Rotação $R(-50°)$. (e) Rotação final $R(-70°)$.

Fonte: O autor.

A matriz possui seis graus de liberdade correspondentes ao seis elementos da matriz, e são necessários três pontos correspondentes para determinarmos a matriz de afinidade que pode ser escrita em blocos como

$$H_A = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

Como uma forma de ajudar a entender os efeitos da matriz A vamos decompô-la em transformações mais fundamentais

$$A = R(\theta)R(-\phi)DR(\phi),$$

com uma matriz diagonal $D = diag(\lambda_1, \lambda_2)$ onde as componentes são escalares não-isotrópicos que alteram a escala nos eixos x e y respectivamente. Uma rotação ϕ que define o sentido de aplicação dos escalares da matriz D, uma rotação contrária $-\phi$ e uma última rotação θ. Podemos visualizar os efeitos das rotações e dos escalares na Figura 10. Os dois graus de liberdade a mais em relação a uma similaridade vêm do ângulo ϕ e da razão entre os escalares aplicados por D, $\frac{\lambda_1}{\lambda_2}$.

Temos três importantes invariantes sob afinidade: retas paralelas se matêm paralelas, razão entre seguimentos de retas paralelas e razão entre as áreas. Mesmo que retas paralelas sejam invariantes, a afinidade não mantém, em geral, o ângulo entre retas nem a razão entre as distâncias correspondentes como acontece com a similaridade.

Figura 11 - Subgrupo de transformações projetivas.

Legenda: Respectivamente, transformações: euclidiana, similaridade, afinidade, e projetividade.
Fonte: Adaptado, Hartley e Zisserman (2004).

Transformações projetivas.

É uma transformação linear mais geral de coordenadas homogêneas (já definida na Subseção 1.1.7), pois generaliza a transformação de afinidade. Na forma de blocos temos a representação

$$H_P = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{v}^\top & v \end{bmatrix},$$

onde $\mathbf{v} = (v_1, v_2)$. Como a matriz é 3×3 temos nove variáveis a serem determinadas, mas que se reduzem a oito tomando uma das variáveis para escala. Assim, precisamos de quatro pontos correspondentes para determinarmos os oito parâmetros da matriz. Além de manter a colinearidade dos pontos, uma invariante fundamental na transformação projetiva é a razão cruzada entre quatro pontos colineares, ou seja, a razão das razões de segmentos contidos numa mesma reta. Na Figura 11 podemos ver um resumo das características geométricas preservadas (em um quadrado, por exemplo) por cada subgrupo de transformação.

1.2 O espaço projetivo em três dimensões

O ponto.

Analogamente à representação de um ponto no plano $\mathbb{P}^2$, um ponto no espaço $\mathbb{P}^3$ é repesentado através de coordenadas homogêneas, acrescentando-se uma quarta coordenada ao vetor que representa esse ponto. Desta forma, $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ e $X_4 \neq 0$, onde $\mathbf{X}$ é a representação em coordenadas homogêneas do ponto $(X, Y, Z)^\top \in \mathbb{R}^3$, e portanto um ponto nesse espaço tem três graus de liberdade. Para realizar tal mudança basta tomar

$$X = X_1/X_4 \ , \ Y = X_2/X_4 \ \text{ e } \ Z = X_3/X_4.$$

O plano.

Temos que a representação algébrica de um plano π no espaço $\mathbb{R}^3$ é dada pela equação

$$\pi_1 X + \pi_2 Y + \pi_3 Z + \pi_4 = 0,$$

onde π_i são os coeficientes da equação. Um ponto em $\mathbb{R}^3$ pertence ao plano se, e somente se, satifaz a equação acima que na forma matricial, usando a representação do ponto em coordenadas homogêneas, é dada por

$$(\pi_1, \pi_2, \pi_3, \pi_4)^\top \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = 0.$$

Fazendo as substituições

$$X = X_1/X_4 \ , \ Y = X_2/X_4 \ \text{ e } \ Z = X_3/X_4,$$

generalizamos a representação do ponto e a equação se torna

$$(\pi_1, \pi_2, \pi_3, \pi_4)^\top \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = 0 \qquad \text{ou} \qquad \pi^\top \mathbf{X} = 0.$$

Desta forma, verificamos que, assim como os pontos, um plano $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^\top$ fica inteiramente determinado por um vetor com quatro coordenadas em $\mathbb{P}^3$. Aqui temos uma analogia com o espaço $\mathbb{P}^2$, já que pontos e retas têm a mesma representação vetorial com três componentes naquele espaço, e a dualidade entre pontos e retas em $\mathbb{P}^2$ também se verifica entre pontos e planos em $\mathbb{P}^3$. Como multiplicar a equação algébrica de um plano por um escalar diferente de zero não altera o plano, temos que os planos possuem três graus de liberdade. Podemos dividir as três primeiras coordenadas pela última, por exemplo, e fixar uma escala.

Obs: As três primeiras coordenadas do vetor que representa o plano corresponde ao vetor normal ao plano, conforme definido em Álgebra Linear.

A reta.

Uma reta pode ser definida passando por dois pontos. Em $\mathbb{P}^2$, como os dois pontos estão no mesmo plano, uma reta passando por esses dois pontos tem apenas dois graus de liberdade, conforme visto anteriormente. Mas em $\mathbb{P}^3$, como os dois pontos podem estar em planos diferentes, temos que uma reta apresenta quatro graus de liberdade, dois graus para cada ponto. Assim, uma reta deveria ser representada por um vetor com cinco coordenadas em $\mathbb{P}^3$, mas vetores desse tipo não podem ser usados, facilmente, em expressões matemáticas que envolvem vetores com quatro coordenadas representando pontos e planos. Portanto é necessário encontrar uma outra representação, e uma das formas mais simples, de uso corrente em algoritmos, é definir a reta através de dois pontos não coincidentes ((**??**)). Outras representações de uma reta no espaço projetivo $\mathbb{P}^3$, diferentes da apresentada aqui, podem ser encontradas em (**??**).

Uma reta $\mathbf{L}$ passando por dois pontos $\mathbf{A}$ e $\mathbf{B}$ é representada pelo espaço linha gerado pela matriz W composta pelos pontos $\mathbf{A}^\top$ e $\mathbf{B}^\top$ em linha,

$$W = \begin{bmatrix} A^\top \\ B^\top \end{bmatrix},$$

onde o espaço gerado por $W^\top$ é o conjunto de pontos do tipo $a\,\mathbf{A} + b\,\mathbf{B}$ pertencentes à reta $\mathbf{L}$ procurada.

A quádrica.

Similarmente à cônica em $\mathbb{P}^2$, uma quádrica Q em $\mathbb{P}^3$ é definida pela equação

$\mathbf{X}^\top Q\,\mathbf{X} = 0,$

Figura 12 - Exemplos de quádricas.

Legenda: Parabolóide à esquerda e hiperbolóide de uma folha à direita.
Fonte: Hartley e Zisserman (2004).

onde Q é uma matriz simétrica 4×4 com nove graus de liberdade.

Mais detalhes sobre a dedução da definição de uma quádrica podem ser encontrados em Hartley e Zisserman (2004), bem como outros exemplos além daqueles observados na Figura 12.

1.2.1 A transformação de pontos e planos em $\mathbb{P}^3$

A transformação de pontos e planos em $\mathbb{P}^3$ é inteiramente análoga àquela argumentada na Subseção 1.1.7 para pontos e retas em $\mathbb{P}^2$, pois assim como os pontos têm uma relação dual com retas em $\mathbb{P}^2$, os pontos também têm uma relação dual com planos em $\mathbb{P}^3$ e, como vimos, ambos têm a mesma representação em vetores homogêneos com quatro componentes. Assim, sob uma tranformação projetiva de pontos $\mathbf{X}' = H_{4\times4}\mathbf{X}$ temos que planos se transformam de acordo com:

$$\pi' = H^{-\top}\pi.$$

1.2.2 Direções e o plano no infinito.

O plano no infinito é denotado por π_∞, tem a forma canônica $\pi_\infty = (0,0,0,1)^\top$ e é composto por pontos no infinito $D = (X_1, X_2, X_3, 0)^\top$ que são as direções das retas contidas no espaço. No plano projetivo $\mathbb{P}^2$ duas retas paralelas se encontram num ponto no infinito, e analogamente, em $\mathbb{P}^3$ dois planos paralelos se encontram numa reta que está contida no plano no infinito, e também, uma reta é paralela a outra reta (ou a um plano) se o ponto de interseção está no π_∞. O plano no infinito é invariável sob uma transformação afim.

De acordo com a Subseção 1.2.1 temos que a transformação projetiva de planos no espaço $\mathbb{P}^3$ se dá pela fórmula $\pi' = H^{-\top}\pi$, assim podemos verificar porque um plano no infinito é invariante sob a tranformação afim. Conforme a Subseção 1.1.9 vimos que a transformação afim é dada por:

$$H_A = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix},$$

então

$$H_A^{-\top} = \begin{bmatrix} A^{-\top} & \mathbf{0} \\ -\mathbf{t}^\top A^{-\top} & 1 \end{bmatrix},$$

e aplicando ao plano no infinito temos:

$$\begin{aligned} \pi' &= H_A^{-\top}\pi_\infty \\ &= \begin{bmatrix} A^{-\top} & \mathbf{0} \\ -\mathbf{t}^\top A^{-\top} & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^\top \\ &= \pi_\infty. \end{aligned}$$

Existem outros planos que podem ser invariantes sob uma transformação afim menos específica, mas só o plano no infinito é invariante em uma transformação afim geral. Cabe resaltar que π_∞ é invariável apenas quando a transformação é aplicada ao conjunto, e não a pontos isolados pertencentes a esse plano.

1.2.3 A cônica absoluta

A cônica absoluta é uma cônica (ponto) denotada por Ω_∞ e está contida no plano no infinito π_∞. Em coordenadas Euclidianas o plano no infinito tem definição $\pi_\infty = (0,0,0,1)^\top$ (Subseção 1.2.2) e a cônica absoluta é definida através da equação matricial

$$\mathbf{X}^\top \Omega_\infty \mathbf{X} = 0,$$

onde as componentes de $\mathbf{X}$ obedecem às restrições

$$X_1^2 + X_2^2 + X_3^2 = 0 \qquad \text{e} \qquad X_4 = 0.$$

De acordo com essas restrições e com esse sistema de coordenadas, e para satisfazer a equação matrical da cônica, podemos definir $\Omega_\infty = I_{3\times 3}$, pois

$$X_1^2 + X_2^2 + X_3^2 = 0 \Leftrightarrow (X_1, X_2, X_3)\, I\, (X_1, X_2, X_3)^\top = 0.$$

Da mesma forma que o plano no infinito é invariante sob uma transformação afim, a cônica absoluta é invariante sob uma transformação de similaridade. Já que a cônica absoluta está contida no plano no infinito, a transformação afim que mantem fixo esse plano também deverá manter invariável as coordenadas da cônica, sendo então a cônica absoluta invariável sob uma transformação afim:

$$H_A = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

Aplicando a homografia afim H_A aos pontos da cônica absoluta temos que as três primeiras componentes desses pontos são afetadas apenas pela submatriz $A_{3\times 3}$,

$$\begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 0 \end{pmatrix} = \begin{pmatrix} A \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \\ 0 \end{pmatrix}.$$

Como vimos, essas três primeiras componentes são utilizadas na definição da cônica absoluta, e portanto Ω_∞ vai sofrer influência apenas de A em tal transformação. De acordo com a tranformação de cônicas na Subseção 1.1.7, e com o fato de Ω_∞ ser invariável sob H_A, temos que $A^\top I A^{-1} = I$, assim:

$$\begin{aligned} I &= A^{-\top} I\, A^{-1} \\ &= A^{-\top} A^{-1} \qquad \text{(tomando a inversa)} \\ &= A\, A^\top. \end{aligned}$$

Desta forma percebemos que A é uma matriz ortogonal (possivelmente com aplicação de escala), o que está de acordo com a definição de tranformação de similaridade H_S conforme a Subseção 1.1.9.

1.2.4 A quádrica dual absoluta

A quádrica dual absoluta é a dual degenerada da cônica absoluta Ω_∞, e é denotada por $\mathtt{Q}^*_\infty$. Analogamente ao fato de que C é o envelope das retas tangentes à cônica C no $\mathbb{P}^2$, temos que, geometricamente, Ω_∞ é o envelope de todos os planos que são tangentes à cônica absoluta e satisfazem à equação da quádrica dual absoluta. Algebricamente, $\mathtt{Q}^*_\infty$ é representada por uma matriz homogênea 4×4 e num sistema de coordenadas métrico tem a forma canônica

$$\mathtt{Q}^*_\infty = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}. \tag{5}$$

Para ver que os planos envelopados por Ω_∞ são tangentes à cônica absoluta Ω_∞, considere um plano $\pi = (\mathbf{v}, k)^\top$. Esse plano está no envelope definido por Ω_∞ se, por definição, $\pi^\top \mathtt{Q}^*_\infty \pi = 0$. Usando a matriz canônica na Equação 5 temos

$$\begin{aligned} \pi^\top \mathtt{Q}^*_\infty \pi &= \begin{pmatrix} \mathbf{v}^\top & k \end{pmatrix} \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} \begin{pmatrix} \mathbf{v} \\ k \end{pmatrix} \\ &= \mathbf{v}^\top \mathbf{v} \\ &= 0, \end{aligned} \tag{6}$$

onde $\mathbf{v}$ representa a reta de interseção do plano π com o plano no infinito π_∞. De acordo com as Subseções 1.1.4 e 1.2.3, a reta $\mathbf{v}$ que pertence ao plano no infinito, é tangente à cônica absoluta se, e somente se, $\mathbf{v}^\top I\, \mathbf{v} = 0$, o que está de acordo com a Equação 6. Portanto, o envelope definido por Ω_∞ contém os planos que são tangentes à Ω_∞.

Para constatar que $\mathbf{v}$ é a reta de interseção de π com π_∞, considere o sistema abaixo onde um ponto $\mathbf{X}$ pertence a esses dois planos. Esse ponto pertence a cada um

dos planos se satisfaz o sistema

$$\begin{cases} \begin{pmatrix} \mathbf{v}^\top & k \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = 0 \\ \pi_\infty^\top \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = 0. \end{cases}$$

Pela segunda equação, como $\pi_\infty = (0, 0, 0, 1)^\top$ (Subseção 1.2.2) temos que $X_4 = 0$ (o que já não era novidade). Substituindo $X_4 = 0$ na primeira equação e efetuando a multiplicação matricial, a constante k desaparece multiplicada por zero e restará

$$\mathbf{v}^\top \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = 0.$$

A última equação representa a equação de pertinência de um ponto a uma reta contida num plano conforme as Subseções 1.1.1 e 1.1.2.

1.3 A câmera P

A câmera é um dispositivo que transforma um cenário 3D, referenciado por um sistema de coordenadas do mundo, em uma imagem 2D no sistema de coordenadas da câmera. Algebricamente, a *câmera* é modelada como uma transformação linear entre um ponto 3D no espaço e um ponto 2D no plano da imagem, representada por uma matriz $P_{3\times 4}$ com algumas propriedades que realizam o mapeamento entre os pontos. Existem vários tipos de câmeras, mas para tratar das características básicas vamos utilizar o modelo buraco de alfinete, do inglês *pinhole*.
Esse tipo de câmera é uma composição de três mudanças de coordenadas, onde a matriz P pode ser decomposta em:

$$P = K\,[I|\mathbf{0}] \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} = K\,[R|\mathbf{t}],$$

tais que

- a matriz $\begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$ contendo os parâmetros extrínsecos transforma os pontos em coordenadas 3D no mundo para as coordenadas 3D da câmera.
- a matriz $[I|\mathbf{0}]$ aplica a projeção perspectiva passando um ponto na coordenada 3D da câmera para um ponto 2D em coordenadas normalizadas no plano da imagem.
- a matriz K, matriz de calibração contendo os parâmetros intrínsecos, transforma pontos em coordenadas normalizadas no plano da imagem para as coordenadas na imagem "final"(considerando mudanças como medida em pixels, por exemplo).

Mais a frente vamos abordar cada um desses tópicos com mais detalhes.

Acompanhando pela Figura 13 para fazermos a construção da câmera, consideramos (a princípio) o *centro de projeção*, ou *centro da câmera*, como a origem do espaço tridimensional Euclidiano, com o *plano da imagem* ou *plano focal* sendo $Z = f$, onde f é a *distancia focal* entre o plano da imagem e o centro de projeção. Assim, o plano da imagem é definido como $(x, y, f)^\top$, mas seu sistema de coordenadas é medido em pixels e é centralizado no canto inferior esquerdo da imagem. Por *coordenadas normalizadas da imagem* nos referimos aos pontos $(x, y)^\top$ expressos como $(x, y, 1)^\top$, (considerando que a câmera tenha distância focal unitária) no plano da imagem com relação ao sistema de coordenadas da câmera. O *plano detector* fica em $(x, y, -f)^\top$. Como podemos obeservar na Figura 13, um ponto $\mathbf{X}$ no espaço 3D é mapeado a um ponto $\mathbf{x}$ no plano da imagem por uma reta que liga $\mathbf{X}$ ao centro de projeção, e intersecta o plano da imagem. O plano que passa pelo centro da câmera e é paralelo ao plano da imagem é chamado *plano principal*, o plano xy na Figura 13. O *eixo principal* passa pelo centro da câmera e é perpendicular ao plano da imagem, onde a interseção desse eixo com o plano da imagem forma o *ponto principal*, o qual pode ser pensado como a projeção do centro da câmera no plano da imagem. Com vetores representados em coordenadas homogêneas, podemos expressar o mapeamento através de um operador linear, onde realizamos a multiplicação da matriz P, que representa a câmera, por um ponto no espaço, resultando num ponto na imagem conforme o esquema abaixo:

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

Nessa matriz temos que f é a distância focal (entre o centro da câmera e o plano da imagem), e o vetor de coluna de zeros à direita representa as coordenadas do centro da câmera, que nesse caso está na origem do sistema. Essa matriz pode ser desmembrada da seguinte maneria: uma matriz diagonal $diag(f, f, 1)$ multiplicada pela matriz $[I|\mathbf{0}]$, onde I é uma matriz identidade 3×3 e $\mathbf{0}$ é um vetor coluna de zeros. Compactamente, podemos escrever a última relaçao como:

Figura 13 - A câmera P.

Legenda: Visualização das características básicas de uma câmera como distância focal, eixo principal, plano da imagem, centro de projeção e o ponto principal $\mathbf{P}$.
Fonte: Adaptado, Hartley e Zisserman (2004).

$$\mathbf{x} = diag(f, f, 1)\,[I|\mathbf{0}]\,\mathbf{X},$$

onde $diag(f, f, 1)\,[I|\mathbf{0}]$ é a matriz da câmera para o modelo *pinhole*.

Nessa dedução, consideramos que o ponto principal coincide com a origem do plano da imagem, mas na prática isso pode não acontecer. Dessa forma, devemos acrescentar as coordenadas do ponto na construção da matriz da câmera. Sendo (p_x, p_y) as coordenadas do ponto principal, desejamos o seguinte mapeamento (em coordenadas não homogêneas):

$$(X, Y, Z)^\top \to (\tfrac{f\,X}{Z} + p_x\,,\ \tfrac{f\,Y}{Z} + p_y),$$

o qual nos fornece a seguinte relação de projeção em coordenadas homogêneas:

$$\begin{pmatrix} f\,X + Z\,p_x \\ f\,Y + Z\,p_y \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Definindo a notação

$$K = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix},$$

a relação de projeção assume a forma

$$\mathbf{x} = K\,[I|\mathbf{0}]\,\mathbf{X}_{cam}, \tag{7}$$

onde K é denominada matriz de calibração da câmera, a qual resume os parâmetros internos da mesma, e $\mathbf{X}_{cam}$ enfatiza que o ponto 3D está escrito no sistema de coordenadas da câmera.

A matriz de calibração K ainda não está completa, pois o sistema de coordenadas do plano da imagem é medido em pixels, que podem ser não quadrados, e além disso os eixos cartesianos desse sistema podem ser não ortogonais. Assim, vamos inserir mais dois parâmetros para definir a medição em pixels e mais um parâmetro para definir o ângulo de inclinação entre os eixos cartesianos.

Sendo m_x e m_y a quantidade de pixels nas direções x e y, respectivamente, a transformação para a coordenadas em pixels se dá pela multiplicação da matriz $diag(m_x, m_y, 1)$ à esquerda da matriz K,

$$\begin{bmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

onde $\alpha_x = m_x f$, $\alpha_y = m_y f$ representa a distância focal em termos das dimensões dos pixels. Similarmente, $x_0 = m_x p_x$ e $y_0 = m_y p_y$ são as coordenadas do ponto principal medidas em pixels. Para completar a generalidade, acrescentamos o parâmetro s, referente à inclinação entre os eixos, na matriz

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

Até aqui temos expressado o ponto $\mathbf{X}$ no espaço em relação às coordenadas da câmera, assumindo que o centro da câmera está situado na origem do sistema de eixos, o qual recebe o nome de sistema de coordenadas da câmera. Mas na maioria das vezes esse não é o caso, portanto desejamos fazer o mapeamento de um ponto 3D cujo as coordenadas estejam expressas com relação a um sistema de coordenadas qualquer, chamado sistema de coordenadas do mundo. Vamos continuar completando a matriz P para mapear pontos no sistema de coordenadas do mundo.

Considerando o sistema de coordenadas do mundo, denotaremos um ponto nesse sistema, em coordenadas não homogêneas, por $\overline{\mathbf{X}}$, e o mesmo ponto no sistema de coordenadas da câmera será denotado por $\overline{\mathbf{X}}_{cam}$. A transferência de sistemas de coordenadas é feita através da relação $\overline{\mathbf{X}}_{cam} = R\,(\overline{\mathbf{X}} - \overline{\mathbf{C}})$, onde $\overline{\mathbf{C}}$ representa o centro da câmera no sistema de coordenadas do mundo e R representa uma matriz de rotação 3×3. Passando os vetores para coordenadas homogêneas, podemos aplicar a rotação no vetor $\overline{\mathbf{C}}$, do centro da câmera, e inseri-lo na quarta coluna da matriz escrevendo a relação

$$\mathbf{X}_{cam} = \begin{bmatrix} R & -R\,\overline{\mathbf{C}} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

onde podemos escrever

$$R = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \qquad \text{e} \qquad -R\,\overline{\mathbf{C}} = (t_1, t_2, t_3)^\top,$$

e a relação acima fica:

$$\mathbf{X}_{cam} = \begin{bmatrix} a & b & c & t_1 \\ d & e & f & t_2 \\ g & h & i & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

Substituindo $\mathbf{X}_{cam}$ na Equação 7, obtemos

$$\mathbf{x} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a & b & c & t_1 \\ d & e & f & t_2 \\ g & h & i & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

multiplicando a segunda e terceira matrizes,

$$\mathbf{x} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c & t_1 \\ d & e & f & t_2 \\ g & h & i & t_3 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

que pode ser escrita de forma resumida como

$$\mathbf{x} = K\,[R|\mathbf{t}]\,\mathbf{X},$$

com $\mathbf{t} = (t_1, t_2, t_3)^\top$ denominado vetor de translação.

Acabamos de incluir todos os parâmetros na matriz $P = K\,[R|\mathbf{t}]$, para o mapeamento de um ponto 3D homogêneo no sistema de coordenadas do mundo em seu respectivo ponto 2D no plano da imagem, no sistema de coordenadas da imagem. Temos cinco parâmetros internos na matriz de calibração K, $(\alpha_x, \alpha_y, s, p_x, p_y)$, três parâmetros externos na matriz de rotação R, apesar de suas nove entradas, e mais três parâmetros externos no vetor de translação totalizando 11 graus de liberdade. Repare que, tendo a matriz P dimensão 3×4, ela possui 12 componentes e, usando uma para determinar a escala, sobram exatamente os 11 graus de liberdade, e a câmera P é chamada câmera de *projeção finita*.

A matriz de calibração K.

A intenção nesta seção é mostrar qual é a natureza de cada um dos parâmetros internos da matriz de calibração. Para isso, vamos resolver o problema de transferir um

ponto qualquer no espaço (em coordenadas não homogêneas) para o plano da imagem na câmera, conforme encontrado em (**??**).

Problema:

Dado um ponto 3D $p^w = (x^w, y^w, z^w)^\top$ em coordenadas do mundo e sendo (ξ, η) as coordenadas do ponto na imagem da câmera, quais os valores de (ξ, η)?

Primeiramente, vamos transformar o ponto em coordenadas do mundo para as coordenadas da câmera, $p = (x, y, z)^\top$ e

$$p = Rp^w + T$$

onde R e T são parâmetros extrínsecos. Agora, vamos projetar o ponto para o plano imagem normalizada usando similaridade de triângulos

$$\begin{cases} \tilde{x} = \dfrac{x}{z} \\ \tilde{y} = \dfrac{y}{z} \end{cases}$$

Depois dessa projeção, vamos aplicar os parâmetros internos para obtermos as coordenadas finais da imagem. O primeiro passo é incluir a distância focal.

$$\begin{cases} \tilde{x} = f\dfrac{x}{z} \\ \tilde{y} = f\dfrac{y}{z} \end{cases}$$

Então converter para pixels:

$$\begin{cases} \tilde{x}_{pix} = s_\xi f\dfrac{x}{z} \\ \tilde{y}_{pix} = s_\eta f\dfrac{y}{z}, \end{cases}$$

onde s_ξ e s_η é a quantidade de pixels por unidade de comprimento nas direções ξ e η, respectivamente. Como podemos ver, f tem efeito similar a s_ξ e s_η. Seria suficiente mantermos apenas dois parâmetros, digamos f e a razão de aspecto $\frac{s_\xi}{s_\eta}$, mas vamos manter os três parâmetros por enquanto. Em seguida, fazemos a translação para o ponto inferior esquerdo da imagem usando $(t_{\tilde{\xi}}, t_{\tilde{\eta}})$, que são as coordenadas do ponto principal em relação

Figura 14 - Mudança de coordenadas.

Legenda: Transformação das coordenadas de um ponto no plano com eixos perpendiculares para um plano com eixos relacionados por um ângulo θ qualquer.

Fonte: O autor.

ao inferior esquerdo da imagem:

$$\begin{cases} \tilde{\xi} = s_\xi f \dfrac{x}{z} + t_{\tilde{\xi}} \\ \tilde{\eta} = s_\eta f \dfrac{y}{z} + t_{\tilde{\eta}}, \end{cases}$$

e convertemos para o sistema de coordenadas em função de um ângulo θ entre os eixos, pois os eixos podem ser não ortogonais. Observe na Figura 14 a conversão das coordenadas de um ponto Q qualquer:

$$\begin{cases} \tilde{\xi} = \xi + \eta \cos \theta \\ \tilde{\eta} = \eta \sin \theta \end{cases}$$

Isolando as variáveis do sistema de coordenadas com ângulo genérico temos:

$$\begin{cases} \xi = \tilde{\xi} - \tilde{\eta} \cot \theta \\ \eta = \tilde{\eta} \csc \theta \end{cases}$$

Tabela 1 - Resumo dos parâmetros da câmera.

Variáveis Livres	**Descrição**
3	rotação
3	translação
2	mudança de escala em x, y
1	distância focal (pode ser incorporada como escala)
2	ponto principal
1	ângulo entre os eixos
6	total extrinsecos
6	total intrinsecos
12	total
11	total único (distância focal para fixar escala)

Fonte: O autor.

os quais fornecem

$$\begin{cases} \xi = \overbrace{s_\xi f}^{\alpha_\xi} \dfrac{x}{z} + \overbrace{(-\cot\theta\, s_\eta f)}^{\sigma_\theta} \dfrac{y}{z} + \overbrace{t_{\tilde{\xi}} - \cot\theta\, t_{\tilde{\eta}}}^{\xi_0} \\ \eta = \underbrace{s_\eta f \csc\theta}_{\alpha_\eta} \dfrac{y}{z} + \underbrace{\csc\theta\, t_{\tilde{\eta}}}_{\eta_0} \end{cases}$$

Assim, a transformação de um ponto expresso nas coordenadas da câmera para a imagem é feita usando cinco parâmetros da matriz de calibração

$$K = \begin{bmatrix} \alpha_\xi & \sigma & \xi_o \\ 0 & \alpha_\eta & \eta_o \\ 0 & 0 & 1 \end{bmatrix}.$$

A fórmula acima explicita a interpretação de cada um desses parâmetros.

1.4 A ação projetiva de uma câmera P

Nesta subseção iremos verificar como se dá a projeção aplicada por uma câmera P em planos, retas, cônicas e quádricas.

Ação projetiva de P em planos.

A equação de projeção $\mathbf{x} = P\,\mathbf{X}$ é uma transformação de um ponto em coordenadas 3D no sistema de coordenadas do mundo para pontos 2D no sistema de coordenadas

Figura 15 - Transformação plano a plano.

Legenda: (a)Transformação projetiva de plano a plano induzida pela câmera P. (b)Plano π retroprojetado por uma reta $\boldsymbol{l}$ no plano da imagem através da câmera P.
Fonte: Hartley e Zisserman (2004).

do plano da imagem na câmera e, por conta das características invariantes sob transformação projetiva, temos a liberdade para escolher o sistema de coordenadas do mundo. Assim, suponha que tal sistema de coordenadas seja posicionado de forma que o plano xy corresponda ao plano π, conforme ilustrado na Figura 15. Desta forma, pontos no espaço que pertençam ao plano π terão a componente $X_3 = 0$, e a ação de uma câmera P nesses pontos é dada por

$$\begin{aligned}
\mathbf{x} &= P\,\mathbf{X} \\
&= \begin{pmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} & \mathbf{p_4} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_4} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ 1 \end{pmatrix},
\end{aligned}$$

onde $\mathbf{p_i}$ são as colunas da matriz P. Assim, a transformação de um ponto $\mathbf{X}_\pi = (X_1, X_2, 1)^\top \in \pi$ para um ponto na imagem é, em geral, uma homografia planar, ou uma transformação projetiva de plano a plano $\mathbf{x} = H\,\mathbf{X}_\pi$.

1.4.1 A ação projetiva de P em retas

Geometricamente, como podemos ver na Figura 15, uma reta $\mathbf{L}$ no espaço 3D junto com o centro $\mathbf{C}$ da câmera definem um plano. A interseção desse plano com o plano da

imagem é uma reta $\boldsymbol{l}$, que é a imagem da reta $\mathbf{L}$ no espaço. Algebricamente, considere uma reta $\mathbf{L}$ parametrizada por λ passando pelos pontos $\mathbf{A}$ e $\mathbf{B}$, onde cada ponto nessa reta é dado por $\mathbf{X}(\lambda) = \mathbf{A} + (\lambda)\,\mathbf{B}$. Considere ainda, $\mathbf{a}$ e $\mathbf{b}$ sendo as imagens dos pontos $\mathbf{A}$ e $\mathbf{B}$ sob a ação da câmera P. Aplicando a projeção da câmera P aos pontos da reta $\mathbf{L}$ temos

$$\begin{aligned}
\mathbf{x}(\lambda) &= P\,\mathbf{X}(\lambda) \\
&= P(\mathbf{A} + (\lambda)\,\mathbf{B}) \\
&= P\,\mathbf{A} + (\lambda)\,P\,\mathbf{B} \\
&= \mathbf{a} + (\lambda)\,\mathbf{b},
\end{aligned}$$

onde cada $\mathbf{x}$ em função de λ pertence à reta passando por $\mathbf{a}$ e $\mathbf{b}$ no plano da imagem.

Retroprojeção de retas.

Geometricamente, a retroprojeção de retas é o conjunto de pontos no espaço pertencentes a um plano, o qual é definido pelo centro da câmera e uma reta na imagem, como na Figura 15. Algebricamente, sendo a reta na imagem denotada por $\boldsymbol{l}$ e a câmera por P, o plano retroprojetado é $P^\top \boldsymbol{l}$. De fato, um ponto $\mathbf{X}$ no espaço é projetado na imagem como $\mathbf{x} = P\,\mathbf{X}$, e esse ponto pertence a reta $\boldsymbol{l}$ na imagem se $\mathbf{x}^\top \boldsymbol{l} = 0$ ou, substituindo, $(P\,\mathbf{X})^\top \boldsymbol{l} = 0$. Então, aplicando a transposta, $\mathbf{X}^\top P^\top \boldsymbol{l} = 0$ e $P^\top \boldsymbol{l}$ é tomado como o plano que contém o ponto $\mathbf{X}$ no espaço. Assim, tal plano é a retroprojeção da reta $\boldsymbol{l}$.

A ação projetiva de P em cônicas.

Um cone é uma quádrica degenerada, representada por uma matriz 4×4 que não tem posto completo, e é denotada por $\mathtt{Q}_{co}$.

Retroprojeção de cônicas.

Uma cônica retroprojeta um cone $\mathtt{Q}_{co}$ que, neste caso, tem o vértice coincidindo com o centro da câmera, onde esse vértice é o vetor nulo da matriz 4×4 que representa o cone. Uma cônica retroprojeta um cone de acordo com

$$\mathtt{Q}_{co} = P^\top C\,P$$

pois, um ponto $\mathbf{X}$ no espaço é projetado na imagem na forma $\mathbf{x} = P\,\mathbf{X}$, e $\mathbf{x} \in C$ se satisfaz a equação $\mathbf{x}^\top C\,\mathbf{x} = 0$. Substituindo $\mathbf{x}$ na equação da cônica temos $(P\,\mathbf{X})^\top C\,P\,\mathbf{X} = 0$, e aplicando a transposição, $\mathbf{X}^\top P^\top C\,P\,\mathbf{X} = 0$. O ponto $\mathbf{X}$ é projetado na cônica C se , se e somente se, $\mathbf{X} \in \mathtt{Q}_{co}$, que deve ser definida como $\mathtt{Q}_{co} = P^\top C\,P$. Assim, um cone é a retroprojeção de uma cônica, e repare que não há a projeção direta de uma cônica, já que uma cônica é definida por uma matriz 3×3 e a matriz da câmera é 3×4.

1.4.2 A ação projetiva de P em quádricas

Para tratarmos desse assunto, primeiramente precisamos definir alguns conceitos.

Contorno gerador e contorno aparente.

Na formação da imagem de uma superfície, os raios de luz passando pelo centro da câmera são tangentes à essa superfície em 3D. O contorno dessa superfície nesses pontos de tangência é transformado num contorno no plano da imagem conforme a Figura 16. O contorno da superfície é denominado *contorno gerador* e denotado por $\mathbf{\Gamma}$, o contorno formado na imagem é chamado *contorno aparente* e denotado por γ. Vale ressaltar que o contorno gerador depende das posições do centro da câmera e da própria superfície, e não depende do plano da imagem.

Projeção de quádricas.

A quádrica é uma superfície lisa e o seu contorno gerador produz um contorno aparente no plano da imagem, através dos raios retroprojetados pelos pontos do contorno aparente e passando pelo centro da câmera, conforme um exemplo na Figura 16. Assim, o contorno gerador é uma cônica e sob uma transformação projetiva o contorno aparente também é uma cônica na imagem. Como estamos usando relações de tangência aos contornos gerador e aparente, será necessária a utilização da quádrica dual $\mathtt{Q}^*$, já que por definição a equação que representa a quádrica dual fica definida pelos planos π que são tangentes ao contorno gerador na quádrica $\mathtt{Q}$. Portanto, a equação da quádrica dual é dada por

$$\pi^\top \mathtt{Q}^* \pi = 0.$$

Figura 16 - Contornos de quádricas.

Legenda: (a) O contorno aparente é a imagem do contorno gerador, onde essa imagem fica definida pelas retas através do centro da câmera que são tangentes aos pontos no contorno gerador. (b) Uma quádrica tangenciada por um cone que é o formado pelo conjunto de retas passando pelo centro da câmera, onde tais retas são tangentes ao contorno gerador $\mathbf{\Gamma}$.

Fonte: Hartley e Zisserman (2004).

Sejam $\boldsymbol{l}$ as retas tangentes à conica C na imagem (contorno aparente), onde a cônica dual define a relação $\boldsymbol{l}^\top C^* \boldsymbol{l} = 0$. Essas retas $\boldsymbol{l}$ retroprojetam planos definidos por $\pi = P^\top \boldsymbol{l}$ (Subseção 1.4.1), que são tangentes à quádrica $\mathtt{Q}$ e satisfazem a relação $\pi^\top \mathtt{Q}^* \pi = 0$. Substituindo o plano nessa última relação temos

$$\begin{aligned}
\pi^\top \mathtt{Q}^* \pi &= (P^\top \boldsymbol{l})^\top \mathtt{Q}^* P^\top \boldsymbol{l} \\
&= \boldsymbol{l}^\top P \, \mathtt{Q}^* P^\top \boldsymbol{l} \\
&= \boldsymbol{l}^\top C^* \boldsymbol{l} \\
&= 0,
\end{aligned}$$

onde tomamos $C^* = P \, \mathtt{Q}^* P^\top$, a imagem da quádrica sob a projeção P.

1.4.3 A geometria bifocal

A geometria epipolar.

Considere um cenário com um ponto $\mathbf{X}$ no espaço 3D e dois planos que contêm as imagens $\mathbf{x}$ e $\mathbf{x}'$ desse ponto $\mathbf{X}$. Considere ainda um plano π que contém esse ponto $\mathbf{X}$ e os dois centros de projeção (ou centro da câmera) das duas câmeras, denotados por $\mathbf{C}$ e $\mathbf{C}'$. A geometria epipolar se constitui nas relações existentes entre as interseções desse plano π com os dois planos das imagens observado na Figura 17. Sendo $\mathbf{x}$ a imagem 2D de $\mathbf{X}$ no primeiro plano de imagem, a construção desse cenário é motivada pela busca de $\mathbf{x}'$, também 2D, que seja correspondente a $\mathbf{x}$, onde $\mathbf{x}'$ é a imagem de $\mathbf{X}$ no segundo plano de imagem. Ou seja, existem as matrizes que promovem a ação de cada câmera onde $\mathbf{x} = P\mathbf{X}$ e $\mathbf{x}' = P'\mathbf{X}$, e quando isso acontece dizemos que os pontos são *pontos correspondentes*, e denotamos por $\mathbf{x} \leftrightarrow \mathbf{x}'$ (já definidos anteriormente). Podemos observar que $\mathbf{x}$ e $\mathbf{x}'$ retroprojetam duas retas definidas por $\mathbf{C}$ e $\mathbf{C}'$, onde esses retas pertencem a π e se interceptam em $\mathbf{X}$. Essa propriedade é essencial para nos auxiliar a definir as correspondências entre os pontos nas duas imagens.

A *reta base* é a reta que passa pelo centro de cada câmera $\mathbf{C}$ e $\mathbf{C}'$. O plano π fica determinado pela reta base e pela reta definida por $\mathbf{x}$ e $\mathbf{C}$. A interseção de π com o segundo plano de imagem define uma reta chamada *reta epipolar*, que é a imagem na segunda visão da reta retroprojetada por $\mathbf{x}$. Ou seja, dadas duas imagens de uma cena, para cada ponto numa imagem vai existir uma reta epipolar correspondente na outra imagem. Pela construção, sabemos que $\mathbf{x}' \in \pi$ e pertence ao plano da segunda imagem, portanto $\mathbf{x}'$ pertence à reta epipolar que será denotada por $\boldsymbol{l}'$. Assim, nossa busca pela correspondência de $\mathbf{x}$ se restringirá a uma busca numa reta e não a uma busca em todo

Figura 17 - O plano epipolar.

Legenda: O plano epipolar contém o ponto $\mathbf{X}$, os centros das câmeras, e define as retas epipolares em cada imagem, o que é a base para correspondência entre $\mathbf{x}$ e $\mathbf{x}'$.
Fonte: Adaptado, Hartley e Zisserman (2004).

o plano de imagem da segunda câmera.

$\mathbf{x} \rightarrow \boldsymbol{l}'$

A imagem do centro de projeção da primeira câmera na segunda fotografia é denominada *epipolo* e denotada por $\boldsymbol{e}'$. Analogamente, a imagem de $\mathbf{C}'$ na primeira fotografia também é chamada epipolo e é denotada por $\boldsymbol{e}$. O plano π é denominado *plano epipolar* e num sistema com duas vistas temos vários planos epipolares definidos por um único parâmetro, e que formam o feixe de planos girando em torno da reta base.

1.4.4 A homografia induzida por um plano

Considere um plano π no espaço 3D com coordenadas expressas em um referencial no mundo, com duas câmeras, mas que não passa pelo centro de nenhuma das duas câmeras. Portanto esse plano não é um plano epipolar e é dito estar em posição geral. Assim, vamos determinar uma homografia definida em função desse plano π.

Dadas as duas matrizes de projeção canônicas

$$P = [I|\mathbf{0}] \qquad \text{e} \qquad P' = [R|\mathbf{t}],$$

(as câmeras não são calibradas pois não possuem a matriz K e a origem do sistema cartesiano do mundo coincide com o centro da primeira câmera) um plano $\pi = (\mathbf{v}, 1)^\top$, e um ponto $\mathbf{X} \in \pi$, então a homografia induzida por π é

$$\mathbf{x}' = H\,\mathbf{x}, \qquad \text{onde} \qquad H = R - \mathbf{t}\,\mathbf{v}^\top.$$

Podemos tomar a última coordenada de π igual a 1, pois nos interessa apenas que o plano não passe pelo centro da primeira câmera $\mathbf{C} = (0, 0, 0, 1)^\top$. Observe o esquema na Figura 18. Na primeira câmera temos a projeção

$$\mathbf{x} = P\,\mathbf{X} = [I|\mathbf{0}]\,\mathbf{X},$$

e qualquer ponto 3D do tipo $\mathbf{X} = (\mathbf{x}^\top, \rho)^\top$ vai satisfazer a projeção acima. De fato

$$P\,\mathbf{X} = [I|\mathbf{0}]\,(\mathbf{x}^\top, \rho)^\top = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \rho \end{pmatrix} = \mathbf{x},$$

e assim qualquer ponto na reta retroprojetada por $\mathbf{x}$ pode ser parametrizado por ρ. Como $\mathbf{X} \in \pi$, $\pi^\top \mathbf{X} = 0$ e ρ fica determinado

$$\begin{aligned} \pi^\top \mathbf{X} &= 0 \\ (\mathbf{v}^\top, 1)\,(\mathbf{x}^\top, \rho)^\top &= 0 \\ \mathbf{v}^\top \mathbf{x} + \rho &= 0 \\ \rho &= -\mathbf{v}^\top \mathbf{x}, \end{aligned}$$

e dessa forma o ponto $\mathbf{X} = (\mathbf{x}^\top, -\mathbf{v}^\top \mathbf{x})^\top$. Projetando o ponto 3D na segunda imagem

Figura 18 - Homografia planar.

Legenda: O mapeamento induzido por um plano no espaco 3D. O ponto $\mathbf{x}$ retroprojeta uma reta que intersecta o plano π no ponto $\mathbf{X}_\pi$, então $\mathbf{X}_\pi$ é projetado na outra imagem em $\mathbf{x}'$.

Fonte: Hartley e Zisserman (2004).

temos

$$\begin{aligned}
\mathbf{x}' &= P'\mathbf{X} \\
&= [R|\mathbf{t}]\,(\mathbf{x}^\top, -\mathbf{v}^\top\mathbf{x})^\top \\
&= R\,\mathbf{x} - \mathbf{t}\,\mathbf{v}^\top\mathbf{x} \\
&= (R - \mathbf{t}\,\mathbf{v}^\top)\,\mathbf{x}.
\end{aligned}$$

Portanto

$$\mathbf{x}' = H\,\mathbf{x}, \qquad \text{onde} \qquad H = R - \mathbf{t}\,\mathbf{v}^\top.$$

1.4.5 A matriz fundamental F

A *matriz fundamental* reúne todas as informações da geometria epipolar, é a representação algébrica dessa geometria espacial com duas câmeras e, por isso, também é conhecida como *tensor bifocal*.

Geometricamente, a matriz fundamental pode ser obtida através do mapeamento de $\mathbf{x}$ na primeira imagem a um ponto $\mathbf{x}'$ na segunda imagem utilizando um plano π qualquer no espaço. Após esse mapeamento, identificamos a reta epipolar definida pelos pontos $\mathbf{x}'$ e $\boldsymbol{e}'$. Assim, considere um plano π que não passe pelos centros das câmeras, pois desse jeito uma reta retroprojetada por $\mathbf{x}$ deve interceptar o plano π em algum ponto $\mathbf{X}$. Usando a transferência através do plano π, o ponto $\mathbf{X}$ deve ser projetado a um ponto $\mathbf{x}'$ na segunda imagem e, como $\mathbf{X}$ pertence à reta retroprojetada por $\mathbf{x}$, o ponto $\mathbf{x}'$ deverá pertencer à reta epipolar $\boldsymbol{l}'$, já que $\boldsymbol{l}'$ é a imagem na segunda câmera da reta retroprojetada por $\mathbf{x}$ conforme a Figura 17. Assim, pela Subseção 1.4.4 existe uma homografia 2D H em função de π onde cada ponto $\mathbf{x}_i$ é mapeado a um ponto $\mathbf{x}'_i$,

$$\mathbf{x}'_i = H_\pi\,\mathbf{x}_i. \tag{8}$$

Tais pontos são ditos projetivamente equivalentes, já que são projetivamente equivalentes a um mesmo ponto $\mathbf{X}_i$ no plano π. Obtido o ponto $\mathbf{x}'$, a reta epipolar passando por $\mathbf{x}'$ e $\boldsymbol{e}'$ pode ser calculada fazendo

$$\begin{aligned}
\boldsymbol{l}' &= \boldsymbol{e}' \times \mathbf{x}' \\
&= [\boldsymbol{e}']_\times \mathbf{x}'.
\end{aligned} \tag{9}$$

Substituindo 8 em 9 temos

$$\begin{aligned} \boldsymbol{l}' &= [\boldsymbol{e}']_\times \mathbf{x}' \\ &= [\boldsymbol{e}']_\times H_\pi \mathbf{x} \\ &= F\,\mathbf{x}, \end{aligned}$$

onde definimos $F = [\boldsymbol{e}']_\times H_\pi$ é a matriz fundamental. Como $[\boldsymbol{e}']_\times$ é uma matriz anti-simétrica 3×3 obtida de um vetor, então $[\boldsymbol{e}']_\times$ tem posto 2 e como H_π obtida de um plano tem posto 3, temos que a matriz fundamental tem posto 2. F é o mapeamento de um plano projetivo a um feixe de retas epipolares.

Algebricamente, a matriz fundamental pode ser determinada a partir das matrizes das duas câmeras P e P'. Vamos primeiramente definir a equação que representa a reta retroprojetada por $\mathbf{x}$. Dado um ponto $\mathbf{x}$ na imagem, precisamos determinar o conjunto de pontos no espaço 3D que são mapeados nesse ponto $\mathbf{x}$, e este raio de luz (modelado por uma reta) será definido pela junção de dois de seus pontos. E conhecemos dois pontos pertencentes a esta reta, $\mathbf{C}$ é o centro da câmera tal que $P\,\mathbf{C} = \mathbf{0}$, e o ponto $P^+\mathbf{x}$ onde P^+ é a pseudo-inversa de P definida na Subseção A. O ponto $P^+\mathbf{x}$ pertence à reta retroprojetada por $\mathbf{x}$ pois aplicando-lhe a projeção efetuada pela câmera temos o próprio ponto $\mathbf{x}$ como retorno:

$$P^+\mathbf{x} \to P\,P^+\mathbf{x} = I\,\mathbf{x} = \mathbf{x}.$$

Portanto, podemos formar a reta pela junção dos pontos $\mathbf{C}$ e $P^+\mathbf{x}$, parametrizada por λ

$$\mathbf{X}(\lambda) = P^+\mathbf{x} + \lambda\,\mathbf{C}.$$

Computando a imagem desses dois pontos na segunda câmera temos

$$\begin{aligned} P^+\mathbf{x} &\to P'P^+\mathbf{x} \\ \mathbf{C} &\to P'\mathbf{C}, \end{aligned}$$

onde $P'\,\mathbf{C}$ é o epipolo na segunda imagem, e por isso, a reta epipolar pode ser definida em termos do produto vetorial desses dois pontos

$$\boldsymbol{l}' = P'\mathbf{C} \times P'P^+\mathbf{x}. \tag{10}$$

Mas como $P'\mathbf{C} = \boldsymbol{e}'$, podemos substitui-lo na equação anterior

$$\begin{aligned} \boldsymbol{l}' &= P'\mathbf{C} \times P'P^+\mathbf{x} \\ &= [\boldsymbol{e}']_\times P'P^+\mathbf{x} \\ &= F\,\mathbf{x}, \end{aligned}$$

definindo $F = [\boldsymbol{e}']_\times P' P^+$. Essa relação para F é a mesma fórmula derivada anteriormente, com a homografia planar calculada em termos das duas câmeras, $H_\pi = P' P^+$.

Em problemas de reconstrução 3D podemos considerar a posição da primeira câmera como a origem do sistema de coordenadas do mundo, assim a matriz P, sua pseudo-inversa e a matriz da segunda câmera são, respectivamente

$$P = K\,[I|\mathbf{0}], \qquad P^+ = \begin{bmatrix} K^{-1} \\ \mathbf{0}^\top \end{bmatrix} \qquad \text{e} \qquad P' = K'[R|\mathbf{t}].$$

Substituindo na Equação 10 temos

$$\begin{aligned} \boldsymbol{l}' &= P'\mathbf{C} \times P'P^+\mathbf{x} \\ &= K'[R|\mathbf{t}] \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \times K'[R|\mathbf{t}] \begin{bmatrix} K^{-1} \\ \mathbf{0}^\top \end{bmatrix} \mathbf{x} \\ &= [K'\mathbf{t}]_\times K'\,R\,K^{-1}\mathbf{x}, \end{aligned}$$

onde teremos uma nova definição para a matriz fundamental

$$F = [K'\mathbf{t}]_\times K'\,R\,K^{-1}. \tag{11}$$

Essa definição para F é importante para falarmos sobre a matriz *essencial*.

1.4.6 Propriedades da matriz fundamental

A partir da função $\mathbf{x} \to \boldsymbol{l}'$ dada pela relação $\boldsymbol{l}' = F\,\mathbf{x}$, podemos derivar a propriedade mais básica da matriz fundamental. Observe que se $\mathbf{x}' \in \boldsymbol{l}' \Rightarrow \mathbf{x}'^\top \boldsymbol{l}' = 0$, e temos

$$\begin{aligned} \boldsymbol{l}' &= F\,\mathbf{x} \\ \mathbf{x}'^\top \boldsymbol{l}' &= \mathbf{x}'^\top F\,\mathbf{x} \\ 0 &= \mathbf{x}'^\top F\,\mathbf{x}. \end{aligned}$$

A relação acima é denominda *condição de correspondência*, pois pontos que a satisfazem devem ser coplanares e tal relação se mostra uma ferramenta para a determinação da matriz fundamental a partir dos pontos correspondentes $\mathbf{x} \leftrightarrow \mathbf{x}'$. Observe que essa forma de determinar F não requer o conhecimento das câmeras P e P'.

Se F é a matriz para as câmeras (P, P') então $F^\top$ é a matriz para (P', P), pois

basta aplicar a transposição em

$$\begin{aligned} \mathbf{x}'^\top F\,\mathbf{x} &= 0 \\ \mathbf{x}^\top F^\top \mathbf{x}' &= 0, \quad \text{onde} \quad F^\top \mathbf{x}' = \boldsymbol{l}. \end{aligned}$$

Os epipolos $\boldsymbol{e}$ e $\boldsymbol{e}'$ são os espaços nulos à direita e à esquerda de F, respectivamente. Vimos que $\boldsymbol{e}' \in \boldsymbol{l}' \Rightarrow \boldsymbol{e}'^\top \boldsymbol{l}' = 0$, e $\boldsymbol{l}' = F\,\mathbf{x}$ para todo ponto $\mathbf{x} \in \boldsymbol{l}$ (exceto $\boldsymbol{e}$), daí

$$\begin{aligned} \boldsymbol{l}' &= F\,\mathbf{x} \\ \boldsymbol{e}'^\top \boldsymbol{l}' &= \boldsymbol{e}'^\top F\,\mathbf{x} \\ 0 &= \boldsymbol{e}'^\top F\,\mathbf{x}, \quad \text{para todo } \mathbf{x} \\ \mathbf{0}^\top &= \boldsymbol{e}'^\top F. \end{aligned}$$

A dedução é análoga para $F\,\boldsymbol{e} = \mathbf{0}$.

1.4.7 Configuração canônica das câmeras

A relação $\boldsymbol{l}' = F\,\mathbf{x}$ é uma relação projetiva, já que envolve apenas relacionamentos projetivos entre os objetos geométricos, como interseção de retas e planos, e no desenvolvimento algébrico envolve apenas a projeção realizada pela câmera entre pontos no mundo e pontos na imagem. Com isso, a relação depende apenas das coordenadas projetivas da imagem, ou seja, é projetivamente invariante. Assim, dadas transformações projetivas das imagens $\hat{\mathbf{x}} = H\,\mathbf{x}$ e $\hat{\mathbf{x}}' = H'\mathbf{x}'$ podemos deduzir a correspondente relação $\hat{\boldsymbol{l}}' = \hat{F}\hat{\mathbf{x}}$ com $\hat{F} = H'^{-\top} F\,H^{-1}$. Analogamente, F depende somente das propriedades projetivas das câmeras P e P', já que as imagens são geradas pela ação dessas duas câmeras. As câmeras dependem tanto do sistema de coordenadas das imagens como do sistema de coordenadas do mundo, mas a matriz fundamental é inalterável por uma transformação projetiva do espaço 3D.

Observe que se $H_{4\times4}$ é uma transformação projetiva do espaço 3D, então as relações de projeção podem ser escritas como

$$\begin{aligned} \mathbf{x} &= P\,\mathbf{X} \\ \mathbf{x} &= P\,H\,H^{-1}\mathbf{X} \end{aligned} \quad \text{e} \quad \begin{aligned} \mathbf{x}' &= P'\mathbf{X} \\ \mathbf{x}' &= P'H\,H^{-1}\mathbf{X} \end{aligned}.$$

Se $\mathbf{x} \leftrightarrow \mathbf{x}'$ é uma correspondência para o par de câmeras (P, P') em relação ao ponto $\mathbf{X}$, então essas imagens também são correspondentes para o par de câmeras $(P\,H, P'H)$ em relação ao ponto $H^{-1}\mathbf{X}$. Como a matriz fundamental depende apenas das coordenadas das imagens (Subseção 1.4.6), temos que a matriz fundamental para (P, P') será a mesma para $(P\,H, P'H)$. Ou seja, um par de câmeras unicamente define uma matriz fundamental, mas uma matriz fundamental é consistente com vários pares de câmeras.

Já que temos vários pares de câmeras consistentes com a mesma matriz fundamental, podemos escolher um par de câmeras, através de uma transformação projetiva, de forma que algumas futuras abordagens e deduções sejam facilitadas sem prejuízo das características projetivas. Dadas câmeras calibradas $\hat{P}$ e $\hat{P}'$ podemos acrescentar uma linha em $\hat{P}$ de forma que se torne uma matriz 4×4 invertível denotada por $\hat{P}^*$. Aplicando a inversa de $\hat{P}^*$ em $\hat{P}$ e $\hat{P}'$ temos que

$$\begin{array}{rcl} P & = & \hat{P}\,\hat{P}^{*-1} \\ P & = & [I|\mathbf{0}] \end{array} \quad \text{e} \quad \begin{array}{rcl} P' & = & \hat{P}'\,\hat{P}^{*-1} \\ P' & = & [M|\mathbf{m}]. \end{array}$$

As câmeras P e P' definidas acima estão em sua forma *canônica*. Observe, pela dedução da Equação 11, que a matriz fundamental para câmeras canônicas é simplesmente

$$F = [\mathbf{m}]_\times M.$$

Extraindo as câmeras canônicas a partir de F.

Nesta seção será apresentada uma das formas de se obter as matrizes das câmeras para um sistema com duas visões. Como vimos na Subseção 1.4.7, é possível aplicar uma transformação projetiva de forma a se obter o par de câmeras canônicas para um sistema bifocal, e assim facilitar o algebrismo e a computação sem prejuízo das características projetivas do sistema. Desta forma, dada a matriz F computada através da codição de correspondência, vamos mostrar como extrair o par de câmeras em sua forma canônica.

Primeiramente, observe pelo Apêndice A o caso 3×3 em que uma matriz M é anti-simétrica se satisfaz a condição $\mathbf{v}^\top M\,\mathbf{v} = 0$. Considerando a matriz $P'^\top F\,P$, a mesma será anti-simétrica se $\mathbf{X}^\top P'^\top F\,P\,\mathbf{X} = 0$. Mas $\mathbf{x} = P\,\mathbf{X}$ e $\mathbf{x}'^\top = \mathbf{X}^\top P'^\top$, e substituindo temos que $\mathbf{x}'^\top F\,\mathbf{x} = 0$. Pela Subseção 1.4.6, essa condição de correspondência mostra que F é a matriz fundamental consistente com o par de câmeras (P, P') já que $P'^\top F\,P$ seja anti-simétrica.

Agora, tomando a configuração canônica $P = [I|\mathbf{0}]$ e $P' = [SF\,|\boldsymbol{e}']$, onde S é uma matriz anti-simétrica e $\boldsymbol{e}'$ é o epipolo que satisfaz $\boldsymbol{e}'^\top F = \mathbf{0}^\top$ (Subseção 1.4.6), temos que $P'^\top F\,P$ é anti-simétrica, pois

$$P'^\top F\,P = [SF\,|\boldsymbol{e}']^\top F\,[I|\mathbf{0}] = \begin{bmatrix} F^\top S^\top F & \mathbf{0} \\ \boldsymbol{e}'^\top F & 0 \end{bmatrix} = \begin{bmatrix} F^\top S^\top F & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix}.$$

Assim, dada F temos que um par de câmeras consistente é simplesmente $P = [I|\mathbf{0}]$ e $P' = [SF\,|\boldsymbol{e}']$. Repare que a escolha da matriz S é arbitrária mas com a exigência de que a matriz P' deve ter posto 3 para ser uma câmera válida. Isso não é problema pois

segundo (**??**) uma boa escolha é $S = [\boldsymbol{e}']_\times$ pois $\boldsymbol{e}'$ não pertence ao espaço gerado pelas colunas de $[\boldsymbol{e}']_\times F$. Finalmente, o par de câmeras canônicas pode ser obtido tomando

$$P = [I|\mathbf{0}] \qquad \text{e} \qquad P' = [[\boldsymbol{e}']_\times F\,|\boldsymbol{e}']. \tag{12}$$

1.4.8 Ambiguidade projetiva das câmeras, dada F

De uma matriz fundamental podemos extrair diferentes pares de câmeras consistentes com a mesma geometria epipolar, em contraste com o fato de, dado um par de câmeras, a matriz fundamental ser única. Esses pares de câmeras extraídas de F diferem apenas por uma transformação projetiva.

Inicialmente, podemos aplicar uma transformação projetiva e trabalhar com a forma canônica das câmeras $P = \tilde{P} = [I|\mathbf{0}]$, $P' = [A|\mathbf{a}]$ e $\tilde{P}' = [\tilde{A}|\tilde{\mathbf{a}}]$. Evocando o resultado em 1.4.7 sabemos que a matriz fundamental pode ser obtida fazendo $F = [\mathbf{a}]_\times A = [\tilde{\mathbf{a}}]_\times \tilde{A}$. Como $[\mathbf{a}]_\times$ e $[\tilde{\mathbf{a}}]_\times$ são anti-simétricas temos que

$$\mathbf{a}^\top F = \mathbf{a}^\top [\mathbf{a}]_\times A = \mathbf{0}^\top \quad \text{e} \quad \tilde{\mathbf{a}}^\top F = \tilde{\mathbf{a}}^\top [\tilde{\mathbf{a}}]_\times \tilde{A} = \mathbf{0}^\top,$$

e como F tem posto 2 e não pode ser invertida, temos que $\mathbf{a}^\top$ e $\tilde{\mathbf{a}}^\top$ são iguais a menos de escala, portanto

$$\tilde{\mathbf{a}}^\top = k\,\mathbf{a}^\top. \tag{13}$$

Considerando $[\mathbf{a}]_\times A = [\tilde{\mathbf{a}}]_\times \tilde{A}$ e substituindo a Equação 13 temos

$$\begin{aligned} [\mathbf{a}]_\times A &= k\,[\mathbf{a}]_\times \tilde{A} \\ 0_{3\times 3} &= k\,[\mathbf{a}]_\times \tilde{A} - [\mathbf{a}]_\times A \\ 0_{3\times 3} &= [\mathbf{a}]_\times (k\,\tilde{A} - A). \end{aligned}$$

Já que $\mathbf{a}$ é um vetor nulo à direita de $[\mathbf{a}]_\times$ podemos tomar, para um vetor $\mathbf{v}$ qualquer,

$$k\,\tilde{A} - A = \mathbf{a}\mathbf{v}^\top \quad \text{e daí} \quad \tilde{A} = \frac{A + \mathbf{a}\mathbf{v}^\top}{k}. \tag{14}$$

Substituindo as Equações 13 e 14 na câmera $\tilde{P}'$ temos que

$$P' = [A|\mathbf{a}] \quad \text{e} \quad \tilde{P}' = [k^{-1}(A + \mathbf{a}\mathbf{v}^\top)|k\,\mathbf{a}], \tag{15}$$

e vamos usar a homografia

$$H = \begin{bmatrix} k^{-1}I & \mathbf{0} \\ k^{-1}\mathbf{v}^\top & k \end{bmatrix}$$

para mostrar que as câmeras estão projetivamente relacionadas.

$$PH = [I|\mathbf{0}] \begin{bmatrix} k^{-1}I & \mathbf{0} \\ k^{-1}\mathbf{v}^\top & k \end{bmatrix} = k^{-1}[I|\mathbf{0}] = k^{-1}\tilde{P}, \quad \text{e}$$

$$P'H = [A|\mathbf{a}] \begin{bmatrix} k^{-1}I & \mathbf{0} \\ k^{-1}\mathbf{v}^\top & k \end{bmatrix} = [k^{-1}(A + \mathbf{a}\mathbf{v}^\top)|k\,\mathbf{a}] = [\tilde{A}|\tilde{\mathbf{a}}] = \tilde{P}'.$$

Juntando a extração das câmeras dada em 12 com a equação geral para câmeras dada em 15 temos a fórmula geral para câmeras canônicas dada a matriz fundamental

$$P = [I|\mathbf{0}] \quad \text{e} \quad P' = [[\boldsymbol{e}']_\times F + \boldsymbol{e}'\mathbf{v}^\top|\lambda\,\boldsymbol{e}'],$$

para um vetor $\mathbf{v}$ qualquer e um escalar $\lambda \neq 0$.

A matriz essencial E.

Historicamente, a matriz essencial foi introduzida antes da matriz fundamental e pode ser pensada como um caso particular da matriz fundamental (**??**), onde os efeitos da matriz de calibração K foram removidos da câmera P. Considere uma câmera $P = K\,[R|\mathbf{t}]$ onde conhecemos sua matriz de calibração. Podemos inverter a matriz K e aplicá-la na câmera de modo que a projeção se torne $P = [R|\mathbf{t}]$, chamada *câmera normalizada*. Aplicando essa projeção a um ponto no espaço 3D $\mathbf{X}$ temos

$$\hat{\mathbf{x}} = P\,\mathbf{X} = [R|\mathbf{t}]\,\mathbf{X}$$

onde o ponto $\hat{\mathbf{x}}$ é dito estar em *coordenadas normalizadas*. Um par de câmeras normalizadas

$$P = [I|\mathbf{0}] \qquad \text{e} \qquad P' = [R|\mathbf{t}]$$

tem a matriz *essencial* como matriz correspondente que reúne todas as características relativas à geometria epipolar para essas duas câmeras. Usando a derivação da matriz fundamental na Equação 11, podemos definir a matriz essencial

$$E = [\mathbf{t}]_\times R.$$

2 RECONSTRUÇÃO DE UMA CÂMERA P: KUANG E ÅSTRÖM

Nesta seção apresentaremos o problema da determinação da pose de uma câmera, que tem sido estudado extensivamente pela comunidade da área de visão computacional. Por exemplo, o caso mínimo da determinação da pose de uma câmera usando três pontos foi estudado inicialmente por Fischler e Bolles (1981) e outras formulações para problemas mínimos desse tipo foram comparadas e revisadas por Nöle et al. (1994). Usando retas, foi desenvolvida uma solução mínima com três retas e suas correspondências por Chen (1991), e outro desenvolvimento envolvendo retas pode ser encontrado em Dhome et al. (1989). Recentemente, um caso mínimo usando combinação de pontos e retas foi publicado por Ramalingam et al. (2011). E mais recentemente ainda, visando aplicação em correspondência entre curvas Fabbri et al. (2012) desenvolveram uma solução mínima usando dois pontos-tangentes (definidos mais a frente).

Uma câmera pode ser fatorada em $P = K[R|\mathbf{t}]$, e o artigo de Kuang e Åström (2013) trata da determinação dos parâmetros externos R e $\mathbf{t}$, considerando que são conhecidos quase todos os parâmetros internos da matriz K, menos a distância focal f. Assim, o artigo busca determinar sete parâmetros no total, três na matriz de rotação R, três no vetor de translação $\mathbf{t}$ e um na matriz de calibração. Alguns trabalhos anteriores já resolveram problemas de determinação desses sete parâmetros mas todos eles usaram somente correspondência entre pontos 3D e pontos 2D, $\mathbf{X} \leftrightarrow \mathbf{x}$, e as principais contribuições do artigo são o uso de correspondências entre outras entidades geométricas como retas e pontos-tangentes, e a apresentação de um procedimento para resolução de sistemas de equações polinomiais utilizando as bases de Gröbner e a matriz de ação. Temos interesse nessa tecnologia computacional para o trabalho futuro de auto-calibração trifocal usando curvas e superfícies.

Como já vimos, os parâmetros externos R e $\mathbf{t}$ tem a finalidade de trazer um ponto 3D escrito no sistema de coordenadas do mundo para o sistema de coordenadas da câmera, e na Figura 19 podemos observar a ação de $[R|\mathbf{t}]$ bem como a correspondência entre entidades geométricas.

2.1 Formulação do problema

O artigo utiliza o modelo *pinhole* de câmera para aplicar a projeção de um ponto $\mathbf{X}$ em 3D em sua imagem $\mathbf{x}$ em 2D dada pela equação de projeção

$$\lambda\,\mathbf{x} = P\,\mathbf{X}, \tag{16}$$

Figura 19 - Projeção de retas, pontos e tangentes.

Legenda: A matriz $[R|\mathbf{t}]$ transforma as estruturas 3D em coordenadas do mundo para as coordenadas da câmera, deixando-as em coordenadas normalizadas. Depois que as estruturas estão em coordenadas normalizadas, a matriz K faz a projeção no plano da imagem final, aplicando características como tamanho dos pixels, distância focal, ponto principal e inclinação entre eixos.

Fonte: Kuang e Åstrom (2013).

Figura 20 - Objetos geométricos.

Legenda: Objetos utilizados para gerar equações na determinação dos parâmetros de P.
Fonte: Kuang e Åstrom (2013).

onde a matriz $P_{3\times4}$ pode ser fatorada conforme descrito acima, e λ é um fator de escala descrito na Subseção 1.1.2.

Para obter estabilidade numérica e uma configuração prática da câmera, geralmente se assume que o plano da imagem tem o ponto principal centrado na origem, pixels quadrados e eixos perpendiculares (*skew* zero). Como a única incógnita na matriz de calibração é a distância focal, a mesma pode ser descrita como

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & w \end{bmatrix}, \tag{17}$$

onde $w = \frac{1}{f}$.

Quantidade de equações.

Cada objeto geométrico nos fornece uma determinada quantidade de equações para determinarmos os parâmetros de P, e vamos analisar de que forma esses objetos providenciam tais equações. Na Figura 20 podemos visualizar um exemplo de cada uma das entidades geométricas.

Equações obtidas usando pontos.

Dado um ponto 3D $\mathbf{X}$ e sua respectiva imagem em 2D $\mathbf{x} = (x, y, 1)^\top$, a equação de projeção $\lambda\,\mathbf{x} = P\,\mathbf{X}$ nos fornece três equações. Mas, pela Subseção 1.1.2, um ponto 2D tem apenas dois graus de liberdade apesar de suas três componentes, então dessas três

equações apenas duas são linearmente independentes. Para eliminarmos o fator de escala λ aplicamos o produto vetorial em ambos os lados da equação de projeção:

$$\begin{aligned} \lambda\,\mathbf{x} &= P\,\mathbf{X} \\ \lambda\,[\mathbf{x}]_\times \mathbf{x} &= [\mathbf{x}]_\times P\,\mathbf{X} \\ \mathbf{0} &= [\mathbf{x}]_\times P\,\mathbf{X}. \end{aligned} \tag{18}$$

Equações obtidas usando retas.

Uma reta 3D $\mathbf{L}$ pode ser representada usando um ponto $\mathbf{X}$ e uma direção $\mathbf{D}$, na forma $\mathbf{L} = \mathbf{X} + k\,\mathbf{D}$. Dadas a reta $\mathbf{L}$ e sua imagem $\boldsymbol{l}$ obtemos duas equações usando os dois pontos $\mathbf{X}$ e $\mathbf{D}$. Um ponto $\mathbf{x}$ pertence a uma reta $\boldsymbol{l}$ se $\boldsymbol{l}^\top \mathbf{x} = 0$, e substituindo essa relação na equação de projeção 16 temos:

$$\begin{aligned} \boldsymbol{l}^\top P\,\mathbf{X} &= 0 \quad \text{se } k = 0 \\ \boldsymbol{l}^\top P(\mathbf{X} + k\,\mathbf{D}) &= 0 \quad \text{se } k \neq 0. \end{aligned} \tag{19}$$

Equações obtidas usando pontos-tangentes.

É possível obter mais três equações se conhecemos o ponto $\mathbf{X}$, uma direção $\mathbf{D}$ através $\mathbf{X}$ bem como suas respectivas imagens $\mathbf{x}$ e $\mathbf{d}$. Com a correspondência $\mathbf{X} \leftrightarrow \mathbf{x}$ podemos usar a relação apresentada na Equação 18 e conseguir duas equações. Para a terceira equação, usamos a reta definida pelos pontos na imagem, $\boldsymbol{l} = \mathbf{x} \times \mathbf{d}$, juntamente com as Equações 19, pois tomando a diferença entre tais equações e usando a reta $\boldsymbol{l}$ temos

$$\boldsymbol{l}^\top P\,\mathbf{D} = 0. \tag{20}$$

Note que essas três equações são obtidas usando um ponto-tangente com uma direção apenas. Se usarmos ponto-tangente com mais direções, conseguimos uma equação a mais para cada direção, pois para cada $\mathbf{D}_i$ das n direções, teremos $\boldsymbol{l}_i = \mathbf{x} \times \mathbf{d}_i$ e podemos formular as equações

$$[\mathbf{x}]_\times P\,\mathbf{X} = \mathbf{0} \quad \text{e} \quad \boldsymbol{l}_i^\top P\,\mathbf{D}_i = 0 \quad \text{para } i = 1, ..., n. \tag{21}$$

Na tabela abaixo podemos verificar um resumo da quantidade de equações conseguidas para correspondências entre cada objeto geométrico.

ponto	reta	ponto com 1 tangente	ponto com 2 tangentes
2	2	3	4

Casos úteis.

Com as quantidades de equações fornecidas por cada correspondência podemos formar várias combinações entre pontos, retas e ponto-tangentes para obtermos as sete equações necessárias para calcular os sete graus de liberdade de P. Vamos ver dois exemplos de problema mínimo.

- **Dois pontos e um ponto-tangente (P2T1)**: com essa combinação temos três pontos e uma direção passando por um dos pontos, então podemos formar seis equações usando a Relação 18 e mais uma equação usando a Relação 20.
- **Um ponto-tangente com uma direção mais um ponto-tangente com duas direções (T1T2)**: temos uma reta passando por um dos pontos e duas retas passando pelo outro ponto, então podemos formar quatro equações usando 18 e mais três equações usando 20.

Podemos formar outros problemas mínimos conforme os apresentados aqui e vamos mostrar agora um caso supra-restringido, onde temos mais equações do que incógnitas a determinar.

- **Quatro retas (P4L)**: Dadas quatro correspondências entre retas temos oito equações usando as Relações 19. Como precisamos de sete equações podemos usar uma delas para verificar a solução dada pelas demais.

Parametrização de P.

Uma das ideias teóricas mais importantes do método e que a matriz de rotação da câmera é parametrizada com quaternions, pois desse modo produz sistemas de equações polinomiais relativamente fáceis de serem resolvidos. Utilizando quaternions, a matriz de rotação é dada por

$$R = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd \\ 2ad+2bc & a^2-b^2+c^2-d^2 & 2cd-2ab \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 \end{bmatrix}$$

Se tratando de uma representação em coordenadas homogêneas podemos usar uma das variáveis para fixar a escala, assim tomando $a = 1$ reduzimos o número de variáveis e

facilitamos a solução do sistema de equações polinomiais.

O vetor de translação é parametrizado normalmente como $\mathbf{t} = (t_x, t_y, t_z)^\top$ e a matriz $[R|\mathbf{t}]$ se torna

$$[R|\mathbf{t}] = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd & t_x \\ 2ad+2bc & a^2-b^2+c^2-d^2 & 2cd-2ab & t_y \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 & t_z \end{bmatrix}$$

Para completar a matriz P basta aplicar, à esquerda de $[R|\mathbf{t}]$ a matriz de calibração K conforme 17.

$$\begin{aligned} P &= K\,[R|\mathbf{t}] \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & w \end{bmatrix} \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd & t_x \\ 2ad+2bc & a^2-b^2+c^2-d^2 & 2cd-2ab & t_y \\ 2bd-2ac & 2ab+2cd & a^2-b^2-c^2+d^2 & t_z \end{bmatrix} \\ &= \begin{bmatrix} a^2+b^2-c^2-d^2 & 2bc-2ad & 2ac+2bd & t_x \\ 2ad+2bc & a^2-b^2+c^2-d^2 & 2cd-2ab & t_y \\ w\,(2bd-2ac) & w\,(2ab+2cd) & w\,(a^2-b^2-c^2+d^2) & w\,t_z \end{bmatrix} \end{aligned}$$

Substituindo $t'_z = w\,t_z$, temos que determinar as sete variáveis $\{b, c, d, t_x, t_y, t'_z, w\}$. Mas observe que as variáveis $\{t_x, t_y, t'_z\}$ são todas lineares usando as Relações 18, 19 e 20, e portanto podem ser eliminadas do sistemas de equações restando apenas as quatro variáveis $\{b, c, d, w\}$. Depois de determinadas essas quatro variáveis, o vetor de translação pode ser recuperado por substituição.

2.2 Sistema de equações polinomiais

Para resolver computacionalmente os sistemas de equações polinomiais decorrentes do emprego dos casos úteis na determinação de P, Kuang e Åstr om (2013) utilizam métodos baseados em geometria algébrica. Primeiramente, utilizando a ferramenta *Macaulay2*, Grayson e Stillman (1993-2002), é verificada a existência de 20 soluções para o caso mínimo (P2T1). Para sistemas com pouca quantidade de variáveis, usar apenas o método das bases de Gröbner pode ser rápido e numericamente estável, mas aqui é empregado o método exposto em Byröd et al. (2009), que consiste basicamente em extrair a base linear a partir das bases de Gröbner e empregar a matriz de ação, da qual as soluções podem ser obtidas pela decomposição em autovalores. No apêndice B é fornecido um resumo sobre a teoria básica de geometria algébrica, bem como exemplos da

utilização do método das bases de Gröbner e da matriz de ação para resolver sistemas computacionalmente.

3 GEOMETRIA TRIFOCAL

Assim como num sistema com duas imagens, podemos extrair informações sobre o cenário em 3D através da correspondência entre pontos, retas, tangentes e segmentos de curvas com três ou mais imagens. No entanto, grande parte das pesquisas em visão computacional não tem utilizado a geometria trifocal em problemas de reconstrução 3D e transferência de pontos de uma imagem para outra. Menor ainda é a utilização da geometria diferencial para abordar esses problemas para curvas gerais, tais como cristas de ondas. Um dos nossos objetivos é fazer um levantamento das características e conceitos encontrados num sistema trifocal em comparação com a geometria epipolar, e relatar alguns dos benefícios conseguidos através da inclusão de uma terceira imagem.

3.1 O Problema

Considere um cenário em 3D com três imagens em 2D desse cenário conforme a Figura 22, onde podemos visualizar as três imagens $\mathbf{x}$, $\mathbf{x}'$ e $\mathbf{x}''$ de um ponto 3D $\mathbf{X}$ (dois sinais de apóstrofo indicam o objeto pertencente à terceira câmera) e os pontos correpondentes $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$ são previamente conhecidos. O desafio da reconstrução 3D é determinar, a partir das correspondências nas imagens, as matrizes das câmeras P, P' e P'' que realizam as projeções $\mathbf{x} = P\,\mathbf{X}$, $\mathbf{x}' = P'\mathbf{X}$ e $\mathbf{x}'' = P''\mathbf{X}$ em cada imagem, bem como a determinação do ponto $\mathbf{X}$. Nesse trabalho está sendo dada prioridade à determinação das matrizes das câmeras, por ser um dos maiores problemas atuais em visão computacional e que torna-se primordial no caso da correspondência entre curvas Fabbri e Kimia (2011).

É mais comum a utilização da geometria epipolar bifocal para problemas de reconstrução 3D, fazendo a extração das câmeras usando a geometria epipolar e aplicando um algoritmo para determinação do ponto 3D a partir da triangulação $(\mathbf{x}, \mathbf{x}', \mathbf{X})$. Mas quando aumentamos para três a quantidade de imagens, aumentam-se também as possibilidades de degenerações (que depende da posição do ponto 3D $\mathbf{X}$) se continuarmos utilizando as ferramentas da geometria epipolar a cada par de imagens. Mostraremos que alguns casos de degenerações podem ser removidos usando a teoria da geometria trifocal.

Vamos aplicar o procedimento desenvolvido por Faugeras et al. (2001) que consiste em utilizar a geometria epipolar bifocal num problema de transferência de pontos num sistema trifocal, avaliar as deficiências dessa abordagem e em seguida mostrar como a geometria trifocal pode erradicar algumas dessas deficiências. O plano definido pelos pontos $\mathbf{C}, \mathbf{C}'$ e $\mathbf{X}$ forma o plano epipolar para as câmeras 1 e 2. Analogamente, temos o plano epipolar para as câmeras 1 e 3 e para as câmeras 2 e 3. A interseção de cada

Figura 21 - Plataforma espanhola.

Legenda: Três imagens da plataforma petrolífera espanhola Casablanca. Exemplo de três pontos correspondentes em vermelho representando um canto do heliporto.
Fonte: http://www.repsol.com/

Figura 22 - Geometria trifocal.

Fonte: Adaptado, Forsyth e Ponce (2003).

plano epipolar com os planos das imagens geram as retas epipolares. Por exemplo, a reta epipolar $\boldsymbol{l}_1''$ definida pelos pontos $\boldsymbol{e}_{31}$ e $\mathbf{x}''$ na terceira imagem, é a reta epipolar com relação ao ponto $\mathbf{x}$ na primeira imagem. Ou seja, $\boldsymbol{l}_1''$ é a imagem, na câmera 3, da reta retroprojetada por $\mathbf{x}$ através do centro de projeção $\mathbf{C}$ na câmera 1. Essa reta é definida por $\boldsymbol{l}_1'' = F_{31}\,\mathbf{x}$ ou $\boldsymbol{l}_1'' = \boldsymbol{e}_{31} \times \mathbf{x}''$, onde F_{ij} é a matriz fundamental para as câmeras i e j tomando como base a câmera j. O plano formado pelos centros de projeção $\mathbf{C}, \mathbf{C}'$ e $\mathbf{C}''$ é chamado *plano trifocal* e as interseções desse plano com os planos das imagens geram as *retas trifocais* denotadas por $\mathbf{t}_i$.

3.2 Transferência epipolar

Supondo que temos os pontos correspondentes $\mathbf{x} \leftrightarrow \mathbf{x}'$ para as câmeras 1 e 2, bem como as matrizes fundamentais para os três planos epipolares da Figura 22, F_{21}, F_{31} e F_{32}, desejamos determinar as coordenadas do ponto $\mathbf{x}''$ no plano de imagem da câmera 3. Como vimos, os pontos $\mathbf{x}$, $\mathbf{x}'$ e $\mathbf{x}''$ são imagens de um mesmo ponto 3D $\mathbf{X}$, e por isso $\mathbf{x}''$ é um ponto correspondente ao ponto $\mathbf{x}$ e, portanto, $\mathbf{x}'' \in \boldsymbol{l}_1''$, onde $\boldsymbol{l}_1'' = F_{31}\mathbf{x}$ é a reta epipolar para as câmeras 1 e 3. Por um argumento similar, constatamos que $\mathbf{x}''$ é correspondente ao ponto $\mathbf{x}'$ e por isso $\mathbf{x}'' \in \boldsymbol{l}_2'' = F_{32}\mathbf{x}'$. Como temos os pontos e as matrizes fundamentais, podemos calcular as retas epipolares na terceira imagem. O ponto $\mathbf{x}''$ será calculado como a interseção entre essas duas retas:

$$\begin{aligned} \mathbf{x}'' &= \boldsymbol{l}_1'' \times \boldsymbol{l}_2'' \\ \mathbf{x}'' &= F_{31}\mathbf{x} \times F_{32}\mathbf{x}'. \end{aligned} \tag{22}$$

Esse metodo é chamado de *transferência epipolar* e não podemos determinar $\mathbf{x}''$ nos seguintes casos:

- acompanhando ainda pela Figura 22, observa-se que se o ponto 3D $\mathbf{X}$ pertence à reta base definida por $\mathbf{C}$ e $\mathbf{C}''$, então a projeção desse ponto pela câmera 1 será $\mathbf{x} = P\,\mathbf{X} = \boldsymbol{e}_{13}$, já que $\mathbf{X}$ está alinhado com $\mathbf{C}''$ e $\boldsymbol{e}_{13} = P\,\mathbf{C}''$. Sendo $\mathbf{x} = \boldsymbol{e}_{13}$ temos que, pela Subseção 1.4.6 $F_{31}\boldsymbol{e}_{13} = \mathbf{0}$, e a reta epipolar $\boldsymbol{l}_1'' = \mathbf{0}$. Geometricamente, não podemos determinar um plano epipolar para as câmeras 1 e 3;

- analogamente ao caso anterior, se $\mathbf{X}$ pertence à reta base definida por $\mathbf{C}'$ e $\mathbf{C}''$, não podemos definir o plano epipolar para as câmeras 2 e 3;

Já que o ponto procurado $\mathbf{x}''$ é calculado como interseção das duas retas epipolares $\boldsymbol{l}_1''$ e $\boldsymbol{l}_2''$, o procedimento de tranferência falha em mais dois casos quando essas retas são iguais ou muito próximas.

- repare que se os centros de projeção não estão alinhados então os epipolos $\boldsymbol{e}_{31}$ e $\boldsymbol{e}_{32}$ são diferentes. As retas epipolares $\boldsymbol{l}''_1$ e $\boldsymbol{l}''_2$ são iguais se $\boldsymbol{e}_{31} \in \boldsymbol{l}''_2$ e $\boldsymbol{e}_{32} \in \boldsymbol{l}''_1$. Geometricamente, isso significa que o ponto $\mathbf{X}$ está alojado no plano trifocal e $\mathbf{x}'' \in \mathbf{t}_3$.

- se os centros de projeção estão alinhados então os epipolos $\boldsymbol{e}_{31}$ e $\boldsymbol{e}_{32}$ são iguais, e daí $\boldsymbol{l}''_1 = \boldsymbol{l}''_2$. Geometricamente, temos infinitos planos passando pela reta definida por $\mathbf{C}$, $\mathbf{C}'$ e $\mathbf{C}''$, e o ponto $\mathbf{x}''$ será indeterminado para qualquer ponto $\mathbf{X}$ no espaço 3D.

Todos os casos se resumem ao fato de o ponto $\mathbf{X}$ estar alojado no plano trifocal, pois desse jeito os planos epipolares coincidem com o plano trifocal. Mais ainda, a acurácia do ponto $\mathbf{x}''$ fica bastante prejudicada se $\mathbf{X}$ está próximo do plano trifocal, pois as retas epipolares se tornam menos transversas.

Mostramos a determinação do ponto $\mathbf{x}''$ especificamente, mas esse desenvolvimento através da transferência epipolar é análogo para a predição de $\mathbf{x}$ ou $\mathbf{x}'$. Quase todos os casos de degeneração são resolvidos através da introdução de uma outra ferramenta algébrica, o *tensor trifocal*.

3.3 O tensor trifocal e as relações de incidência

Para a derivação do tensor trifocal usaremos a correspondência entre retas $\boldsymbol{l} \leftrightarrow \boldsymbol{l}' \leftrightarrow \boldsymbol{l}''$ ao longo de três imagens, onde todas essas retas são imagens de uma mesma reta $\mathbf{L}$ no espaço 3D. Esse é o procedimento padrão desenvolvido originalmente por Spetsakis e Aloimonos (1990) e reproduzido, por exemplo, por Faugeras et al. (2001) e Forsyth e Ponce (2003). A abordagem algébrica aqui utilizada pode ser encontrada em Hartley e Zisserman (2004).

Suponha que temos três imagens em 2D $\boldsymbol{l}$, $\boldsymbol{l}'$ e $\boldsymbol{l}''$ de uma reta $\mathbf{L}$ em 3D, conforme a Figura 23. Pela Subseção 1.4.1, essas três retas nas imagens retroprojetam planos que, pela própria construção, devem se interceptar em $\mathbf{L}$. Esta imposição geométrica gera uma restrição algébrica para as três retas correspondentes, já que em geral três planos no espaço não se interceptam numa única reta. Pela Subseção 1.4.7, através de uma transformação projetiva podemos utilizar o conjunto de matrizes canônicas para as câmeras $P = [I|\mathbf{0}]$, $P' = [A|\mathbf{a}_4]$ e $P'' = [B|\mathbf{b}_4]$, onde A e B são matrizes 3×3 e os vetores $\mathbf{a}_i$ e $\mathbf{b}_i$ representam a i-ésima coluna das câmeras P' e P'' respectivamente, para $i = 1, 2, 3$ e 4.

Observa-se que tomando $P = [I|\mathbf{0}]$ estamos considerando o centro da câmera 1 como a origem de coordenadas do espaço 3D, e assim as coordenadas do centro da câmera 1 são $\mathbf{C} = (0, 0, 0, 1)^\top$. Os epipolos na segunda e terceira imagens em relação à câmera 1 são, respectivamente, dados pela projeção $\boldsymbol{e}' = P'\mathbf{C}$ e $\boldsymbol{e}'' = P''\mathbf{C}$, e assim temos que

Figura 23 - Três imagens de uma reta.

Legenda: O desenvolvimento algébrico do tensor trifocal realizado a partir das três imagens de uma reta no espaço 3D.
Fonte: Hartley e Zisserman (2004).

$\boldsymbol{e}' = \mathbf{a}_4$ e $\boldsymbol{e}'' = \mathbf{b}_4$ (já que não vamos mais nos referir aos demais epipolos de um sistema trifocal, usaremos a notação simplificada $\boldsymbol{e}'$ e $\boldsymbol{e}''$). Como retas retroprojetam planos, esses planos são dados, de acordo com a Subseção 1.4.1, por

$$\begin{array}{rclcrclcrcl} \pi & = & P^\top \boldsymbol{l} & & \pi' & = & P'^\top \boldsymbol{l}' & & \pi'' & = & P''^\top \boldsymbol{l}'' \\ \pi & = & \begin{pmatrix} \boldsymbol{l} \\ 0 \end{pmatrix}, & & \pi' & = & \begin{pmatrix} A^\top \boldsymbol{l}' \\ \mathbf{a}_4^\top \boldsymbol{l}' \end{pmatrix} & \text{e} & \pi'' & = & \begin{pmatrix} B^\top \boldsymbol{l}'' \\ \mathbf{b}_4^\top \boldsymbol{l}'' \end{pmatrix}. \end{array}$$

Como estamos supondo que essas três retas são imagens de uma única reta no espaço, temos que esses três planos devem se interceptar em $\mathbf{L}$, e essa restrição geométrica pode ser expressa algebricamente pelo fato de que a matriz formada pelos vetores dos três planos $M = [\pi\ \pi'\ \pi'']$ deve ter posto dois. De fato, um ponto $\mathbf{X} \in \mathbf{L}$ pode ser escrito como combinação linear de outros dois pontos pertencentes a $\mathbf{L}$ que sejam linearmente independentes, digamos $\mathbf{X} = \lambda_1 \mathbf{X}_1 + \lambda_2 \mathbf{X}_2$, com λ_i escalares. Como $\mathbf{X}$ pertence a cada um dos planos, temos que $\pi^\top \mathbf{X} = 0$, $\pi'^\top \mathbf{X} = 0$ e $\pi''^\top \mathbf{X} = 0$, e daí $M^\top \mathbf{X} = \mathbf{0}$. Encarando $M^\top$ como um operador linear, desde que $\mathbf{X} = \lambda_1 \mathbf{X}_1 + \lambda_2 \mathbf{X}_2$, temos que o núcleo de $M^\top$ tem dimensão dois, e como $M^\top$ tem quatro colunas o domínio tem dimensão quatro.

Pelo teorema do núcleo e imagem de uma transformação linear, a dimensão do domínio é a soma da dimensão do núcleo com a dimensão da imagem, e daí a dimensão da imagem é dois. A dimensão da imagem é o posto do operador linear $M^\top$, que é igual ao posto de M. Considerando α, β e k escalares, o posto dois de M pode ser interpretado

com uma dependência linear entre suas colunas, e como

$$M = \begin{bmatrix} \boldsymbol{l} & A^\top \boldsymbol{l}' & B^\top \boldsymbol{l}'' \\ 0 & \mathbf{a}_4^\top \boldsymbol{l}' & \mathbf{b}_4^\top \boldsymbol{l}'' \end{bmatrix},$$

podemos definir o sistema:

$$\begin{pmatrix} \boldsymbol{l} \\ 0 \end{pmatrix} = \alpha \begin{pmatrix} A^\top \boldsymbol{l}' \\ \mathbf{a}_4^\top \boldsymbol{l}' \end{pmatrix} + \beta \begin{pmatrix} B^\top \boldsymbol{l}'' \\ \mathbf{b}_4^\top \boldsymbol{l}'' \end{pmatrix}. \tag{23}$$

Pela quarta equação do sistema 23 temos

$$\alpha = k\, \mathbf{b}_4^\top \boldsymbol{l}'' \quad \text{e} \quad \beta = -k\, \mathbf{a}_4^\top \boldsymbol{l}' \quad \Rightarrow \quad 0 = \alpha\, \mathbf{a}_4^\top \boldsymbol{l}' + \beta\, \mathbf{b}_4^\top \boldsymbol{l}''.$$

Substituindo os valores de α e β nas três primeiras equações do sistema 23 e desconsiderando o fator de escala k, temos

$$\begin{aligned} \boldsymbol{l} &= \alpha\, A^\top \boldsymbol{l}' + \beta\, B^\top \boldsymbol{l}'' \\ \boldsymbol{l} &= (\mathbf{b}_4^\top \boldsymbol{l}'') A^\top \boldsymbol{l}' - (\mathbf{a}_4^\top \boldsymbol{l}') B^\top \boldsymbol{l}'' \\ \boldsymbol{l} &= (\boldsymbol{l}''^\top \mathbf{b}_4) A^\top \boldsymbol{l}' - (\boldsymbol{l}'^\top \mathbf{a}_4) B^\top \boldsymbol{l}''. \end{aligned}$$

A i-ésima coordenada de $\boldsymbol{l}$ pode ser dada por

$$\begin{aligned} l_i &= \boldsymbol{l}''^\top (\mathbf{b}_4 \mathbf{a}_i^\top) \boldsymbol{l}' - \boldsymbol{l}'^\top (\mathbf{a}_4 \mathbf{b}_i^\top) \boldsymbol{l}'' \\ l_i &= \boldsymbol{l}'^\top (\mathbf{a}_i \mathbf{b}_4^\top) \boldsymbol{l}'' - \boldsymbol{l}'^\top (\mathbf{a}_4 \mathbf{b}_i^\top) \boldsymbol{l}'' \\ l_i &= \boldsymbol{l}'^\top (\mathbf{a}_i \mathbf{b}_4^\top - \mathbf{a}_4 \mathbf{b}_i^\top) \boldsymbol{l}'' \\ l_i &= \boldsymbol{l}'^\top \mathrm{T}_i \boldsymbol{l}'', \end{aligned}$$

onde definimos $\mathrm{T}_i = \mathbf{a}_i \mathbf{b}_4^\top - \mathbf{a}_4 \mathbf{b}_i^\top$. O conjunto das três matrizes $\{\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3\}$ é denominado *tensor trifocal* e a reta $\boldsymbol{l}$ é dada por

$$\boldsymbol{l} = \begin{pmatrix} \boldsymbol{l}'^\top \mathrm{T}_1 \boldsymbol{l}'' \\ \boldsymbol{l}'^\top \mathrm{T}_2 \boldsymbol{l}'' \\ \boldsymbol{l}'^\top \mathrm{T}_3 \boldsymbol{l}'' \end{pmatrix}. \tag{24}$$

3.3.1 Homografia induzida por um plano

Vimos na subseção anterior que podemos determinar o tensor trifocal através da relação entre três retas correspondentes, mas existem mais quatro relações entre retas e pontos correspondentes envolvendo o tensor trifocal. Para derivarmos essas outras relações e voltarmos ao problema da transferência de pontos, precisamos definir a homo-

Figura 24 - Homografia planar e o tensor trifocal.

Legenda: O plano π' retroprojetado por $\boldsymbol{l}'$ induz uma homografia que relaciona $\boldsymbol{l}$ com $\boldsymbol{l}''$.

Fonte: Adaptado, Hartley e Zisserman (2004).

grafia entre a primeira e terceira imagens induzida por um plano π', retroprojetado por uma reta $\boldsymbol{l}'$ na segunda imagem, conforme a Figura 24. As homografias entre pontos e retas em dois planos são dadas, respectivamente, por $\mathbf{x}'' = H\,\mathbf{x}$ e $\boldsymbol{l}'' = H^{-\top}\boldsymbol{l}$ (Subseção 1.1.7). Invertendo $H^{-\top}$ temos que $\boldsymbol{l} = H^{\top}\boldsymbol{l}''$. Como temos três retas correspondentes que satizfazem a Relação 24, podemos comparar essa relação com $\boldsymbol{l} = H^{\top}\boldsymbol{l}''$ e deduzir que $H = [\mathrm{T}_1^{\top}\boldsymbol{l}', \mathrm{T}_2^{\top}\boldsymbol{l}', \mathrm{T}_3^{\top}\boldsymbol{l}']$, pois

$$\boldsymbol{l} = \begin{pmatrix} \boldsymbol{l}'^{\top}\mathrm{T}_1\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_2\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_3\boldsymbol{l}'' \end{pmatrix} = \begin{pmatrix} (\mathrm{T}_1^{\top}\boldsymbol{l}')^{\top}\boldsymbol{l}'' \\ (\mathrm{T}_2^{\top}\boldsymbol{l}')^{\top}\boldsymbol{l}'' \\ (\mathrm{T}_3^{\top}\boldsymbol{l}')^{\top}\boldsymbol{l}'' \end{pmatrix} = \begin{pmatrix} \mathbf{h}_1^{\top}\boldsymbol{l}'' \\ \mathbf{h}_2^{\top}\boldsymbol{l}'' \\ \mathbf{h}_3^{\top}\boldsymbol{l}'' \end{pmatrix} = \begin{bmatrix} \mathbf{h}_1^{\top} \\ \mathbf{h}_2^{\top} \\ \mathbf{h}_3^{\top} \end{bmatrix} \boldsymbol{l}'' = H^{\top}\boldsymbol{l}'',$$

onde tomamos $\mathbf{h}_i = \mathrm{T}_i^{\top}\boldsymbol{l}'$. A homografia deduzida acima (denotada por H_{13}) representa a transformação de um ponto da primeira imagem para a terceira através de uma reta na segunda imagem, $\mathbf{x}'' = H_{13}\mathbf{x}$. Analogamente, podemos deduzir a homografia da primeira imagem para a segunda através de uma reta na terceira imagem, $\mathbf{x}' = H_{12}\mathbf{x}$.

3.3.2 Relações de incidência entre pontos e retas

Uma das relações de incidência já foi deduzida anteriormente, chamada reta-reta-reta, e agora vamos deduzir mais quatro relações. Supondo ainda $\boldsymbol{l}$, $\boldsymbol{l}'$ e $\boldsymbol{l}''$ três imagens de uma reta $\mathbf{L}$, sabemos que um ponto $\mathbf{x}$ pertence à uma reta $\boldsymbol{l}$ na câmera 1 se $\mathbf{x}^{\top}\boldsymbol{l} = 0$, onde

a multiplicação matricial pode ser escrita sob a notação de somatório como $\sum_{i=1}^{3} x^i l_i = 0$. O índice contravariante em x é uma convenção para facilitar o algebrismo na notação padrão de tensor trifocal. Essa convensão se deve ao fato de o tensor trifocal possuir três índices e não apenas dois como ocorre com as matrizes. A Equação 24 pode ser dada como $\boldsymbol{l}'^\top \mathrm{T}_i \boldsymbol{l}'' = l_i$ para $i = 1, 2, 3$, e usando o somatório temos que

$$\begin{aligned}
\boldsymbol{l}'^\top \mathrm{T}_i \boldsymbol{l}'' &= l_i \\
\textstyle\sum_{i=1}^{3} x^i (\boldsymbol{l}'^\top \mathrm{T}_i \boldsymbol{l}'') &= \textstyle\sum_{i=1}^{3} x^i l_i \\
\boldsymbol{l}'^\top (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i) \boldsymbol{l}'' &= 0.
\end{aligned}$$

Desta forma, temos a segunda relação de incidência chamada ponto-reta-reta, para um ponto $\mathbf{x}$ na primeira imagem e as retas $\boldsymbol{l}'$ e $\boldsymbol{l}''$ na segunda e terceira imagens respectivamente.

Pela Subseção 1.4.1 sabemos da existência de um ponto $\mathbf{X} \in \mathbf{L}$ que projeta $\mathbf{x} \in \boldsymbol{l}$ na primeira imagem e projeta a outros dois pontos $\mathbf{x}' \in \boldsymbol{l}'$ e $\mathbf{x}'' \in \boldsymbol{l}''$, ou seja, temos a correspondência $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$. Assim, podemos deduzir as relações para pontos na segunda e terceira imagens utilizando a homografia induzida por plano, e vamos tratar da terceira relação de incidência, o caso ponto-reta-ponto. Usando o resultado da Subseção 3.3.1 temos que

$$\begin{aligned}
\mathbf{x}'' &= H_{13}\mathbf{x} \\
&= [\mathrm{T}_1^\top \boldsymbol{l}', \mathrm{T}_2^\top \boldsymbol{l}', \mathrm{T}_3^\top \boldsymbol{l}']\mathbf{x} \\
&= \left(x_1 \mathrm{T}_1^\top \boldsymbol{l}' + x_2 \mathrm{T}_2^\top \boldsymbol{l}' + x_3 \mathrm{T}_3^\top \boldsymbol{l}' \right) \\
&= (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i^\top)\, \boldsymbol{l}'
\end{aligned}$$

Como estamos expressando os vetores em coordenadas homogêneas, essas equações são invariantes para a aplicação de um fator de escala. De qualquer maneira, podemos eliminar o problema das diferenças de escala aplicando a transposição seguida do produto vetorial à direita em ambos os lados da equação

$$\begin{aligned}
\mathbf{x}'' &= (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i^\top)\, \boldsymbol{l}' \\
\mathbf{x}''^\top &= \boldsymbol{l}'^\top (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i) \\
\mathbf{x}''^\top [\mathbf{x}'']_\times &= \boldsymbol{l}'^\top (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i)[\mathbf{x}'']_\times \\
\mathbf{0}^\top &= \boldsymbol{l}'^\top (\textstyle\sum_{i=1}^{3} x^i \mathrm{T}_i)[\mathbf{x}'']_\times .
\end{aligned} \tag{25}$$

O quarto caso, ponto-ponto-reta, é análogo ao apresentado aqui substituindo $\mathbf{x}''$ por $\mathbf{x}'$, e usando a homografia induzida pelo plano retroprojetado por $\boldsymbol{l}''$, $\mathbf{x}' = H_{12}\mathbf{x}$.

Por último, temos a relação de incidência para três pontos. Observe que a reta $\boldsymbol{l}'$ na Relação 25 pode ser definida como $\boldsymbol{l}' = \mathbf{x}' \times \mathbf{y}' = [\mathbf{x}']_\times \mathbf{y}'$, já que $\mathbf{x}' \in \boldsymbol{l}'$ e $\mathbf{y}'$ é um ponto

qualquer no plano da segunda imagem. Substituindo $\boldsymbol{l}'$, a Relação 25 se torna

$$\mathbf{y}'^{\top}[\mathbf{x}']_{\times}(\sum_{i=1}^{3} x^i \mathrm{T}_i)[\mathbf{x}'']_{\times} = \mathbf{0}^{\top},$$

mas observa-se que tal relação é válida para qualquer reta $\boldsymbol{l}'$ que seja imagem de $\mathbf{L}$, e portanto não depende da escolha do ponto $\mathbf{y}'$ que pode ser descartado. A relação ponto-ponto-ponto é

$$[\mathbf{x}']_{\times}(\sum_{i=1}^{3} x^i \mathrm{T}_i)[\mathbf{x}'']_{\times} = 0_{3\times 3}.$$

Cabe resaltar que apenas satisfazer qualquer das relações de incidência não garante a incidência no espaço 3D, ou seja, o tensor trifocal obtido pode não ter consistência com a geometria trifocal de abordagem do problema.

3.3.3 Relação de incidência para retas epipolares

Suponha que os planos retroprojetados por $\boldsymbol{l}$ e $\boldsymbol{l}'$ sejam coincidentes e que o ponto 3D $\mathbf{X}$ pertença a esse plano denotado por π', de acordo com a Figura 25. Neste caso, π' é o plano epipolar para as câmeras 1 e 2, e a reta retroprojetada por $\mathbf{x}$, imagem de $\mathbf{X} \in \mathbf{L}$, está inteiramente contida em π'. Como dois planos não coincidentes sempre definem uma reta, o plano retroprojetado por $\boldsymbol{l}''$ intercepta o plano π' em $\mathbf{L}$. Já que a reta retroprojetada por $\mathbf{x}$ está inteiramente contida em π', essa reta deverá interceptar a reta $\mathbf{L}$ no ponto $\mathbf{X}$. Desta forma, pela Subseção 3.3.2, temos a correspondência $\mathbf{x} \leftrightarrow \boldsymbol{l}' \leftrightarrow \boldsymbol{l}''$ que satisfaz a relação

$$\boldsymbol{l}'^{\top}(\sum_{i=1}^{3} x^i \mathrm{T}_i)\boldsymbol{l}'' = 0.$$

Observe que, como dois planos sempre definem uma reta, a incidência não depende da escolha do plano retroprojetado pela câmera 3, ou seja, essa incidência vai ocorrer qualquer que seja a reta $\boldsymbol{l}''$. Consequentemente, a relação anterior não depende de $\boldsymbol{l}''$, e pode ser reduzida a

$$\boldsymbol{l}'^{\top}(\sum_{i=1}^{3} x^i \mathrm{T}_i) = \mathbf{0}^{\top} \quad \text{ou} \quad (\sum_{i=1}^{3} x^i \mathrm{T}_i)\boldsymbol{l}'' = \mathbf{0}. \tag{26}$$

A argumentação é análoga considerando o plano epipolar definido pelas retas $\boldsymbol{l}$ e $\boldsymbol{l}''$, e neste caso a relação será independente da posição da reta $\boldsymbol{l}'$. Pelas Relações 26 notamos que é

Figura 25 - Transferências com retas epipolares.

Legenda: Quando **L** pertence a um plano epipolar definido para duas imagens, a relação de incidência $\mathbf{x} \leftrightarrow \boldsymbol{l}' \leftrightarrow \boldsymbol{l}''$ não depende da reta numa terceira imagem.

Fonte: Adaptado, Hartley e Zisserman (2004).

possível calcular as retas epipolares como o espaço nulo da matriz $\sum_{i=1}^{3} x^i \mathrm{T}_i$. Variando o ponto $\mathbf{x}$ é possível obter várias retas epipolares, e todas passam pelo mesmo epipolo. Assim, podemos calcular os epipolos $\boldsymbol{e}'$ e $\boldsymbol{e}''$ como interseção dessas retas. Tal resultado será importante na abordagem sobre reconstrução 3D por ocasião da extração das matrizes das câmeras a partir do tensor trifocal.

3.4 Transferências usando o tensor trifocal

3.4.1 Transferência de pontos

A primeira vantagem no uso do tensor trifocal é que, dos casos de degeneração apresentados na Subseção 3.2, apenas um deles não pode ser resolvido pelo uso do tensor, como veremos a seguir.

Considere novamente que se deseja, dada a correspondência entre pontos nas duas primeiras imagens $\mathbf{x} \leftrightarrow \mathbf{x}'$, determinar as coordenadas do ponto correspondente na terceira imagem, $\mathbf{x}''$. Escolhendo uma reta $\boldsymbol{l}'$ passando por $\mathbf{x}'$ na segunda imagem e de posse do tensor trifocal, podemos utilizar a relação de incidência

$$\mathbf{x}'' = (\sum_{i=1}^{3} x^i \mathrm{T}_i^\top)\, \boldsymbol{l}' \tag{27}$$

Figura 26 - Redução de degenerações.

Legenda: Desde de que **L** não pertença ao plano trifocal, não há motivos para degenerações à medida que variamos **X** em **L**.
Fonte: Hartley e Zisserman (2004).

deduzida na Subseção 3.3.2 para computar diretamente $\mathbf{x}''$. Devemos apenas ter um pouco de cuidado na escoha da reta $\boldsymbol{l}'$, pois no caso de ser a reta epipolar correspondente a $\mathbf{x}$ temos, pela Subseção 3.3.3, que $(\sum_{i=1}^{3} x^i \mathrm{T}_i^\top)\,\boldsymbol{l}' = \mathbf{0}$, e o ponto $\mathbf{x}''$ não pode ser determinado. Uma das formas de resolver o problema seria escolher duas ou três retas passando por $\mathbf{x}'$, computar o ponto $\mathbf{x}''$ e escolher aquele que apresenta a maior norma. Outra opção, caso tenha extraído a matriz fundamental F_{21} para as câmeras 1 e 2, é calcular a reta epipolar $\boldsymbol{l}'_{\boldsymbol{e}} = F_{21}\mathbf{x}$ e escolher $\boldsymbol{l}'$ que seja perpendicular a $\boldsymbol{l}'_{\boldsymbol{e}}$, ou seja, $\boldsymbol{l}'^\top \boldsymbol{l}'_{\boldsymbol{e}} = 0$.

Considere, pela Figura 22, o caso onde o ponto $\mathbf{X}$ está alojado na reta base definida por $\mathbf{C}$ e $\mathbf{C}'$. Nessa posição, a imagem de $\mathbf{X}$ na câmera 2 é o epipolo $\boldsymbol{e}_{21} = \mathbf{x}'$. Para escolha de qualquer reta que passe por $\mathbf{x}'$ para aplicarmos a Relação 27, estaremos escolhendo uma reta epipolar, e como vimos anteriormente, não poderemos definir a posição do ponto $\mathbf{x}''$. Tal reta base é o único lugar que $\mathbf{X}$ pode se alojar no plano trifocal que não permite que o ponto $\mathbf{x}''$ seja calculado usando o tensor trifocal, em contraste com a transferência epipolar que se degenera para qualquer $\mathbf{X}$ no plano trifocal. Observe pela Figura 26 que, desde que a reta $\mathbf{L}$ não pertença ao plano trifocal, não há motivos para a degeneração da transferência usando o tensor trifocal à medida que variamos a posição de $\mathbf{X}$ em $\mathbf{L}$. No caso onde os centros das câmeras estão alinhados, não poderemos determinar a posição de $\mathbf{x}''$ caso $\mathbf{X}$ esteja em quaisquer das três retas base, já que assim essas três retas são uma reta só. Lembramos que a transferência epipolar falha para centros de projeção alinhados qualquer que seja a posição de $\mathbf{X}$.

3.4.2 Transferência de retas

Podemos realizar a transferência de retas utilizando a relação

$$\boldsymbol{l} = \begin{pmatrix} \boldsymbol{l}'^{\top}\mathrm{T}_1\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_2\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_3\boldsymbol{l}'' \end{pmatrix},$$

que dá uma maneira explícita de calcularmos a reta na primeira imagem dadas as retas nas câmeras 2 e 3. Mas podemos calcular retas em quaisquer das três imagens dadas as outras duas, aplicando o produto vetorial por $\boldsymbol{l}$ em ambos os lados da equação

$$\begin{aligned} \boldsymbol{l} \times \boldsymbol{l} &= \boldsymbol{l} \times \begin{pmatrix} \boldsymbol{l}'^{\top}\mathrm{T}_1\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_2\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_3\boldsymbol{l}'' \end{pmatrix} \\ \mathbf{0} &= \boldsymbol{l} \times \begin{pmatrix} \boldsymbol{l}'^{\top}\mathrm{T}_1\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_2\boldsymbol{l}'' \\ \boldsymbol{l}'^{\top}\mathrm{T}_3\boldsymbol{l}'' \end{pmatrix}. \end{aligned}$$

O sistema fornece três equações, sendo duas independentes, para determinar os dois graus de liberdade da reta procurada. Outra vantagem no uso de tensor trifocal é que retas explícitas não podem ser determinadas usando apenas matrizes fundamentais.

Repare, pela Figura 25, que se as retas $\boldsymbol{l}$ e $\boldsymbol{l}'$ retroprojetam o mesmo plano, então não teremos uma única reta definida em 3D que possa ser projetada na terceira imagem, e essas duas retas são retas epipolares correspondentes. Similarmente, não podemos determinar a reta na segunda imagem caso as retas na primeira e terceira imagens sejam retas epipolares correspondentes. Em posições gerais, não há motivo para a geometria epipolar entre as câmeras 1 e 2, e câmeras 1 e 3 serem iguais, ou seja, os dois tipos de transferência não se degeneram simultaneamente com frequência. No entanto, assim como pontos na tranferência epipolar, não é possível aplicar a transferência através tensor trifocal se a reta 3D estiver contida no plano trifocal.

3.5 Reconstrução das câmeras a partir do tensor trifocal

Assim como foi mostrado no caso bifocal, mostraremos agora como extrair as matrizes fundamentais e as matrizes das câmeras a partir do tensor trifocal.

3.5.1 Extração da matriz fundamental

Considerando um ponto $\mathbf{x}$ na primeira imagem, como vimos na Subseção 3.3.1, temos uma homografia que relaciona a primeira com a terceira imagem através de um plano retroprojetado por uma reta $\boldsymbol{l}'$ na segunda imagem, conforme a Figura 24. Com isso, podemos definir a reta epipolar $\boldsymbol{l}''$ na terceira imagem relativa ao ponto $\mathbf{x}$ na primeira imagem. O ponto $\mathbf{x}''$ na terceira imagem é dado por

$$\mathbf{x}'' = H_{13}\mathbf{x} \quad \text{ou} \quad \mathbf{x}'' = [\mathrm{T}_1^\top \boldsymbol{l}', \mathrm{T}_2^\top \boldsymbol{l}', \mathrm{T}_3^\top \boldsymbol{l}']\,\mathbf{x},$$

e a reta epipolar na terceira imagem pode ser calculada como $\boldsymbol{l}'' = \boldsymbol{e}'' \times \mathbf{x}''$. Substituindo a expressão para $\mathbf{x}''$ temos

$$\begin{aligned} \boldsymbol{l}'' &= \boldsymbol{e}'' \times \mathbf{x}'' \\ \boldsymbol{l}'' &= [\boldsymbol{e}'']_\times [\mathrm{T}_1^\top \boldsymbol{l}', \mathrm{T}_2^\top \boldsymbol{l}', \mathrm{T}_3^\top \boldsymbol{l}']\,\mathbf{x} \\ \boldsymbol{l}'' &= F_{31}\,\mathbf{x}, \end{aligned}$$

onde a matriz fundamental pode ser obtida por $F_{31} = [\boldsymbol{e}'']_\times [\mathrm{T}_1^\top \boldsymbol{l}', \mathrm{T}_2^\top \boldsymbol{l}', \mathrm{T}_3^\top \boldsymbol{l}']$.

Algebricamente esta expressão se mantém para qualquer reta $\boldsymbol{l}'$, mas devemos escolher uma reta para evitar a condição de degeneração onde $\mathrm{T}_i^\top \boldsymbol{l}' = \mathbf{0}$. Uma boa escolha é tomar $\boldsymbol{l}' = \boldsymbol{e}'$ pois $\boldsymbol{e}'$ é perpendicular ao espaço nulo à direita de cada $\mathrm{T}_i^\top$. Assim , a matriz fundamental é obtida simplesmente por

$$F_{31} = [\boldsymbol{e}'']_\times [\mathrm{T}_1^\top \boldsymbol{e}', \mathrm{T}_2^\top \boldsymbol{e}', \mathrm{T}_3^\top \boldsymbol{e}'].$$

Uma análise similar permite verificar que podemos obter uma relação que envolve a segunda e primeira imagens,

$$F_{21} = [\boldsymbol{e}']_\times [\mathrm{T}_1 \boldsymbol{e}'', \mathrm{T}_2 \boldsymbol{e}'', \mathrm{T}_3 \boldsymbol{e}''].$$

3.5.2 Extraindo as matrizes das câmeras

Da mesma forma que acontece no caso bifocal, por conta da ambiguidade projetiva, podemos determinar as matrizes das câmeras tomando a câmera 1 como $P = [I|\mathbf{0}]$ aplicando uma transformação projetiva. Seguindo a Sseçãoubseção 1.4.7, dadas duas câmeras $P = [I|\mathbf{0}]$ e $P' = [M|\mathbf{m}]$, a matriz fundamental é extraída como $F = [\mathbf{m}]_\times M$. Usando a matriz fundamental $F_{21} = [\boldsymbol{e}']_\times [\mathrm{T}_1 \boldsymbol{e}'', \mathrm{T}_2 \boldsymbol{e}'', \mathrm{T}_3 \boldsymbol{e}'']$ extraída do tensor trifocal na Subseção

3.5.1, temos que o par de câmeras, dado o tensor trifocal é

$$P = [I|\mathbf{0}] \quad \text{e} \quad P' = [[\mathrm{T}_1\boldsymbol{e}'', \mathrm{T}_2\boldsymbol{e}'', \mathrm{T}_3\boldsymbol{e}'']|\boldsymbol{e}']. \tag{28}$$

Poderíamos pensar que as câmeras 1 e 3 poderiam ser obtidas usando a matriz fundamental $F_{31} = [\boldsymbol{e}'']_\times[\mathrm{T}_1^\top\boldsymbol{e}', \mathrm{T}_2^\top\boldsymbol{e}', \mathrm{T}_3^\top\boldsymbol{e}']$, e daí a câmera 3 seria $P'' = [[\mathrm{T}_1^\top\boldsymbol{e}', \mathrm{T}_2^\top\boldsymbol{e}', \mathrm{T}_3^\top\boldsymbol{e}']|\boldsymbol{e}'']$, mas está *incorreto*. A câmera 3 não pode ser escolhida independentemente das características projetivas das câmeras 1 e 2. Pois, observe que podemos reconstruir o ponto 3D $\mathbf{X}$ a partir das câmeras 1 e 2, e usá-lo para determinar a câmera 3 usando as correspondências $\mathbf{X} \leftrightarrow \mathbf{x}''$, conforme o artigo na seção 2. Assim, P'' depende das características projetivas do par (P, P').

De acordo com a Subseção 1.4.8 e usando a dedução dada pela Equação 28, uma forma mais geral de se obter P e P' é dada por

$$P = [I|\mathbf{0}] \quad \text{e} \quad P' = [[\mathrm{T}_1\boldsymbol{e}'', \mathrm{T}_2\boldsymbol{e}'', \mathrm{T}_3\boldsymbol{e}''] + \boldsymbol{e}'^\top\mathbf{v}|\lambda\,\boldsymbol{e}'], \tag{29}$$

para algum vetor $\mathbf{v}$ e um escalar λ. E para determinarmos P'' devemos obter um trio de câmeras dessa família que seja compatível com a dedução do tensor trifocal dada na Subseção 3.3,

$$\mathrm{T}_i = \mathbf{a}_i\mathbf{b}_4^\top - \mathbf{a}_4\mathbf{b}_i^\top, \tag{30}$$

onde $\mathbf{a}_i$ são a colunas de P' e $\mathbf{b}_i$ são as colunas de P''. Por conta da ambiguidade projetiva, dentre essa família de câmeras em 29, podemos escolher $P' = [[\mathrm{T}_1\boldsymbol{e}'', \mathrm{T}_2\boldsymbol{e}'', \mathrm{T}_3\boldsymbol{e}'']|\boldsymbol{e}']$ para desta forma termos $\mathbf{a}_i = \mathrm{T}_i\boldsymbol{e}''$ para i =1,2 e 3. Notando que $\mathbf{a}_4 = \boldsymbol{e}'$ e $\mathbf{b}_4 = \boldsymbol{e}''$, podemos substituir as três últimas relações na Equação 30 e obter

$$\mathrm{T}_i = \mathrm{T}_i\boldsymbol{e}''\boldsymbol{e}''^\top - \boldsymbol{e}'\mathbf{b}_i^\top,$$

onde, isolando $\mathbf{b}_i^\top$ temos

$$\boldsymbol{e}'\mathbf{b}_i^\top = \mathrm{T}_i(\boldsymbol{e}''\boldsymbol{e}''^\top - I) \tag{31}$$

Como o fator de escala é irrelevante, podemos escolher $\boldsymbol{e}'$ de forma que $\boldsymbol{e}'^\top\boldsymbol{e}' = 1$, multiplicar $\boldsymbol{e}'^\top$ pela esquerda na Relação 31 e aplicar a transposta para chegarmos a

$$\mathbf{b}_i = (\boldsymbol{e}''\boldsymbol{e}''^\top - I)\mathrm{T}_i^\top\boldsymbol{e}'.$$

Agora podemos definir a câmera 3 considerando as características projetivas das câmeras

1 e 2,

$$P'' = [(\boldsymbol{e}''\boldsymbol{e}''^{\top} - I)[\mathrm{T}_1^{\top}\boldsymbol{e}', \mathrm{T}_2^{\top}\boldsymbol{e}', \mathrm{T}_3^{\top}\boldsymbol{e}']|\boldsymbol{e}''],$$

e assim verificarmos que o tensor trifocal determina completamente as matrizes das câmeras a menos de uma ambiguidade projetiva.

4 GEOMETRIA TRIFOCAL CALIBRADA: NISTÉR E SCHAFFALITZKY

Para determinar as poses de três câmeras num problema trifocal, a partir de três imagens com quatro pontos correspondentes em cada imagem, no caso onde K é conhecida, primeiro Nistér e Schaffalitzky (2006) desenvolvem uma abordagem num sistema bifocal para determinação da rotação e da translação da segunda câmera em relação à primeira, utilizando duas imagens com quatro pontos correspondentes.

Eles demonstram que, dados os pontos corrrespondentes, os epipolos de cada imagem estão restritos a se alojarem numa curva de grau 10, a chamada curva décica, bem como desenvolvem um método para a obtenção dessa curva. De posse dos epipolos, definem a geometria epipolar, estimam a segunda câmera em relação à primeira e fazem a reconstrução 3D dos quatro pontos. Com os pontos em 3D e a terceira imagem, estimam a terceira câmera consistente com a estrutura 3D definida pelas duas primeiras câmeras. Utilizam essa teoria para desenvolver o que é ainda a solução mais eficiente para o problema (até onde pudemos averiguar), o notório desafio de estabelecer as poses das câmeras num sistema com três imagens e quatro pontos. O artigo possui vários teoremas e alguns deles são interessantes em termos de pesquisas em geometria projetiva. Por isso, vamos reproduzir o enunciado dos teoremas e abriremos subseções explicando conceitos e definições que são importantes ao entendimento dos teoremas. Depois de realizadas essas explicações será fornecida a demonstração de alguns teoremas, as quais não se encontram no artigo.

4.1 O teorema 1: homografia da reta epipolar

Uma das primeiras afirmações do artigo é que as retas epipolares numa imagem estão homograficamente relacionadas com as retas epipolares na segunda imagem. Dados n correspondências de pontos em duas imagens, o teorema trata da *restrição epipolar*, Faugeras (1993). Ou seja, temos uma homografia 1D que relaciona o feixe de retas através $\mathbf{e}$ com o feixe de retas através $\mathbf{e}'$ chamada *homografia da reta epipolar*. Podemos visualizar essa situação na Figura 27.

Teorema 4.1.1. *As n retas passando pelos pontos $\boldsymbol{e}$ e $\mathbf{x}_i$ na primeira imagem estão homograficamente relacionadas com as n retas passando por $\boldsymbol{e}'$ e $\mathbf{x}'_i$ na segunda imagem.*

Demonstração. O feixe de retas epipolares em cada imagem estão perspectivamente relacionados através do centro de projeção Q na reta base, conforme a Figura 28. Suponha que $\boldsymbol{l}$ e $\boldsymbol{l}'$ sejam duas retas epipolares correspondentes, e $\mathbf{m}$ é uma reta qualquer que não

Figura 27 - Restrição epipolar.

Legenda: Feixe de retas passando pelo epipolo em cada imagem, relacionadas duas a duas por uma homografia.

Fonte: O autor.

Figura 28 - Perspectiva entre feixes de retas epipolares.

Legenda: Retas epipolares relacionadas por uma homografia e com perspectiva centrada em **Q**.

Fonte: Adaptado, Hartley e Zisserman (2004).

passa pelo epipolo $\boldsymbol{e}$. O produto vetorial

$$\mathbf{m} \times \boldsymbol{l} = \mathbf{x} \qquad \text{ou} \qquad [\mathbf{m}]_\times \, \boldsymbol{l} = \mathbf{x}, \tag{32}$$

nos fornece o ponto $\mathbf{x}$ de interseção entre essas retas. Pela subseção 1.4.5 temos que

$$\boldsymbol{l}' = F\,\mathbf{x},$$

assim, $\boldsymbol{l}$ está relacionada a $\boldsymbol{l}'$ através da equação

$$\boldsymbol{l}' = F\,[\mathbf{m}]_\times \, \boldsymbol{l}.$$

Uma boa escolha para $\mathbf{m}$ é a reta $\boldsymbol{e}$, já que

$$\mathbf{m}^\top \boldsymbol{e} = \boldsymbol{e}^\top \boldsymbol{e} \neq 0,$$

então $\boldsymbol{e}$ é uma reta que não passa pelo epipolo $\boldsymbol{e}$ como desejado. Assim, a homografia da reta epipolar pode ser escrita como

$$\boldsymbol{l}' = F\,[\boldsymbol{e}]_\times \, \boldsymbol{l}.$$

O argumento é análogo para o mapeamento da segunda imagem para a primeira,

$$\boldsymbol{l} = F^\top [\boldsymbol{e}']_\times \, \boldsymbol{l}'.$$

□

4.1.1 Nova abordagem para a homografia da reta epipolar

Outra forma de pensar sobre essa homografia é realizando uma translação de forma que os epipolos estejam na origem do plano de cada imagem, e com as retas epipolares passando pelo epipolo, cada reta pode ser parametrizada através do ângulo que ela forma com o eixo das abscissas, denominados α_1 e α_2 na primeira e segunda imagens respectivamente. A homografia da reta epipolar pode ser representada pela equação de uma hipérbole alinhada aos eixos perpendiculares, conforme descrito por Fabbri e Kimia (2010b),

$$a\,x\,y + b\,x + c\,y + d = 0 \qquad \text{ou} \qquad y = \frac{-b\,x - d}{a\,x + c}, \tag{33}$$

onde $x = tan\,\alpha_1$ e $y = tan\,\alpha_2$.

Figura 29 - Reta epipolar.

Legenda: Plano da imagem centrado em **e** com a reta epipolar parametrizada por α_1.

Fonte: O autor.

Vemos na Figura 29 que cada reta epipolar fica definida por um ponto no eixo da tangente, $(1, tan\,\alpha_1)$ na primeira imagem e $(1, tan\,\alpha_2)$ na segunda, onde consideramos o círculo trigonométrico de raio 1. Usando esses pontos, a equação da hipérbole pode ser escrita na forma

$$(1\,,\,tan\,\alpha_1)\overbrace{\begin{bmatrix} d & c \\ b & a \end{bmatrix}}^{H}\begin{pmatrix} 1 \\ tan\,\alpha_2 \end{pmatrix} = 0, \tag{34}$$

sendo $\begin{bmatrix} d & c \\ b & a \end{bmatrix}$ a homografia procurada.

O problema é que a Equação 34 não está definida para retas epipolares verticais, pois utilizamos a tangente, e uma saída é parametrizar cada reta através da própria definição de tangente, pois assim temos uma representação isotrópica para as retas:

$$x = tan\,\alpha_1 = \frac{sen\,\alpha_1}{cos\,\alpha_1} \qquad \text{e} \qquad y = tan\,\alpha_2 = \frac{sen\,\alpha_2}{cos\,\alpha_2}. \tag{35}$$

Desta forma, substituindo 35 em 33 e multiplicando por $cos\,\alpha_1\,cos\,\alpha_2$ temos

$$a\,sen\,\alpha_1\,sen\,\alpha_2 + b\,sen\,\alpha_1\,cos\,\alpha_2 + c\,cos\,\alpha_1\,sen\,\alpha_2 + d\,cos\,\alpha_1\,cos\,\alpha_2 = 0.$$

Com essa conversão, deixamos de parametrizar as retas epipolares pelo eixo das tangentes

e passamos a parametrizar pelo círculo trigonométrico. Aqui temos um argumento para contagem dos graus de liberdade para fixarmos a relação entre retas epipolares, pois devemos determinar os dois epipolos (duas variáveis cada totalizando quatro graus de liberdade), e mais a matriz que representa a homografia (quatro variáveis mas apenas três graus de liberdade utilizando uma variável para fixar a escala). Portanto, precisamos de no mínimo sete restrições para determinarmos a relação entre duas imagens na geometria epipolar.

Na nossa abordagem, consideramos o círculo trigonométrico com raio 1 mas, segundo Fabbri e Kimia (2010b), podemos estipular raios arbitrários r_1 e r_2 em cada uma das imagens que ajuda a garantir estabilidade numérica para a determinação dos coeficientes. Assim, obtemos a equação:

$$a\, r_1\, sen\, \alpha_1\, r_2\, sen\, \alpha_2 + b\, r_1\, sen\, \alpha_1\, r_2\, cos\, \alpha_2 + c\, r_1\, cos\, \alpha_1\, r_2\, sen\, \alpha_2 + d\, r_1\, cos\, \alpha_1\, r_2\, cos\, \alpha_2 = 0.$$

Em resumo, as vantagens dessa última representação é ser isotrópica e ter boa estabilidade numérica para escolhas convenientes de r_1 e r_2, e novamente na representação matricial:

$$(r_1\, cos\, \alpha_1\, ,\, r_1\, sen\, \alpha_1) \begin{bmatrix} d & c \\ b & a \end{bmatrix} \begin{pmatrix} r_2\, cos\, \alpha_2 \\ r_2\, sen\, \alpha_2 \end{pmatrix} = 0$$

4.2 Dedução da IAC

Uma ferramenta algébrica bastante utilizada nessa abordagem de Nistér e Schaffalitzky (2006) é a imagem da cônica absoluta (IAC em inglês) denotada por ω. Como a abordagem assume que as câmeras são calibradas, vamos mostrar que é possível obter a IAC em função da matriz de calibração da câmera, K.

4.2.1 Calibração e a imagem da cônica absoluta

Inicialmente, vamos determinar a homografia que faz o mapeamento do plano no infinito π_∞ para o plano da imagem na câmera. Se $\mathbf{X}_\infty \in \pi_\infty$ então pode ser escrito na forma

$$\mathbf{X}_\infty = (\mathbf{d}^\top, 0)^\top$$

e é projetado por uma câmera geral $P = K\,[R|\mathbf{t}]$ como

$$\begin{aligned}\mathbf{x} &= P\,\mathbf{X}_\infty \\ &= K\,[R|\mathbf{t}]\,(\mathbf{d}^\top, 0)^\top \\ &= K\,R\,\mathbf{d}.\end{aligned}$$

Portanto temos a relação

$$\mathbf{x} = H\,\mathbf{d}$$

onde $H = K\,R$. Como a cônica absoluta Ω_∞ está contida no plano no infinito (contém pontos no infinito), ela pode ser projetada sob a homografia H para computarmos sua imagem. Sob uma homografia de ponto $\mathbf{x} \to H\,\mathbf{x}$, uma cônica se transforama de acordo com $C \to H^{-\top} C\,H^{-1}$, conforme a subseção 1.1.7. De acordo com a subseção 1.2.3, $C = \Omega_\infty = I$, daí temos

$$\begin{aligned}\omega &= H^{-\top}\Omega_\infty\,H^{-1} \\ &= (K\,R)^{-\top}\,I\,(K\,R)^{-1} \\ &= K^{-\top}R^{-\top}R^{-1}K^{-1} \\ &= K^{-\top}K^{-1}\end{aligned}$$

pois R é uma matriz de rotação e portanto satisfaz $R^{-\top}R^{-1} = I$. Assim, de posse de uma câmera calibrada é simples obter a IAC.

Analogamente ao que é feito na subseção 1.1.4, podemos definir a imagem dual da cônica absoluta (sigla DIAC em inglês), denotada por ω^*, a partir de ω como

$$\omega^* = \omega^{-1} = K\,K^\top.$$

Esta é uma cônica definida por retas e é a imagem da quádrica dual absoluta $\mathtt{Q}^*_\infty$, definida na subseção 1.2.4. A imagem é dada por

$$\omega^* = P\,\mathtt{Q}^*_\infty P^\top,$$

conforme a subseção 1.4.2. Nota-se que tais objetos geométricos são apenas artifícios para se trabalhar com a informada matriz K.

4.3 O teorema 2: restrição Kruppa

A correspondência entre retas tangentes às duas cônicas ω e ω' é conhecida como a restrição *Kruppa*. Na Figura 30 podemos observar as restrições epipolar e Kruppa simul-

Figura 30 - Restrições Kruppa e epipolar.

Legenda: Retas epipolares tangentes às cônicas ω e ω'.
Fonte: O autor.

taneamente. A restrição Kruppa foi originalmente introduzida por Faugeras e Maybank (1992) e está relacionada com uma cônica contida num plano no espaco 3D e suas imagens nas duas câmeras, e é válida para cônicas gerais, mas aqui no nosso caso vamos tratar especificamente da cônica absoluta Ω_∞ contida no plano infinito π_∞ no espaço, e as imagens da cônica absoluta (IAC) ω e ω'.

Teorema 4.3.1. *As duas tangentes à IAC ω passando pelo epipolo $\boldsymbol{e}$ estão relacionadas por uma homografia da reta epipolar às duas tangentes à IAC ω' passando pelo epipolo $\boldsymbol{e}'$.*

Demonstração. Sendo ω^* e $\omega^{*\prime}$ as respectivas cônicas duais (DIAC), considere um plano passando pelo centro das duas câmeras e tangenciando essa cônica Ω_∞ no espaço. Por construção, esse plano vai projetar uma reta epipolar em cada imagem onde cada reta será tangente às imagens ω e ω' da cônica Ω_∞, observado na Figura 31. E suponha ainda que $\boldsymbol{l}_1$ e $\boldsymbol{l}_2$ sejam as retas tangentes à cônica ω na primeira imagem conforme a Figura 30. Com essas duas retas tangentes podemos construir uma cônica degenerada D (posto 2) na forma

$$D = [\boldsymbol{e}]_\times \, \omega^* \, [\boldsymbol{e}]_\times. \tag{36}$$

Para mostrar que D é uma cônica (ponto) temos que verificar que é válida a relação

Figura 31 - Homografia planar e a restricão Kruppa.

Legenda: Um plano epipolar tangenciando uma cônica no espaco 3D projeta retas epipolares que tangenciam as cônicas nos planos das imagens.
Fonte: Hartley e Zisserman (2004).

$\mathbf{y}_1^\top D\,\mathbf{y}_1 = 0$, onde $\mathbf{y}_1 \in \boldsymbol{l}_1$.

$$\begin{aligned}
\mathbf{y}_1^\top\, D\,\mathbf{y}_1 &= \mathbf{y}_1^\top [\boldsymbol{e}]_\times\, \omega^*\, [\boldsymbol{e}]_\times\, \mathbf{y}_1 \\
&= ([\boldsymbol{e}]_\times^\top\, \mathbf{y}_1)^\top \omega^*\, [\boldsymbol{e}]_\times\, \mathbf{y}_1 \\
&= \boldsymbol{l}_1^\top \omega^*\, \boldsymbol{l}_1 \qquad\qquad \text{pois} \qquad \boldsymbol{l}_1 = \boldsymbol{e} \times \mathbf{y}_1 \\
&= 0
\end{aligned}$$

onde a última passagem segue do fato de ω^* ser dual. Observe que usamos $[\boldsymbol{e}]_\times^\top\, \mathbf{y}_1 = [\boldsymbol{e}]_\times\, \mathbf{y}_1$ quando na verdade $[\boldsymbol{e}]_\times^\top = -[\boldsymbol{e}]_\times$, mas aqui a diferença de escala é irrelevante. Observamos que o desenvolvimento anterior é análogo para $\boldsymbol{l}_2$.

Na segunda imagem, analogamente à primeira, $D' = [\boldsymbol{e}']_\times\, \omega^{*\prime}\, [\boldsymbol{e}']_\times$ e se transforma de acordo com a regra $D' = H_\infty^{-\top} D\, H_\infty^{-1}$ na subseção 1.1.7. O símbolo ∞ se deve ao fato de que a homografia foi deduzida de maneria similar àquela da subseção 1.4.4, apenas

com a diferença de ter sido com base no plano no infinito. Assim:

$$\begin{aligned} [\boldsymbol{e}']_\times \, \omega^{*\prime} \, [\boldsymbol{e}']_\times &= D' \\ &= H_\infty^{-\top} D \, H_\infty^{-1} \\ &= H_\infty^{-\top} [\boldsymbol{e}]_\times \, \omega^* \, [\boldsymbol{e}]_\times \, H_\infty^{-1} \qquad \text{pois} \qquad D = [\boldsymbol{e}]_\times \, \omega^* \, [\boldsymbol{e}]_\times \end{aligned} \tag{37}$$

Como ω é uma cônica envelope, as retas que são tangentes à ω satisfazem a equação de ω^*, e temos que essas tangentes estão projetivamente relacionadas, em conjunto, pela Equação 37. Como as tangentes passam pelos epipolos, são retas epipolares e o teorema segue. □

Observa-se que, como conjunto, as tangentes não vêm em qualquer ordem particular, pois a relação não explicita uma mapeamento individual entre as retas, há apenas a relação entre produtos simétricos. Mas veremos na Subseção 4.7 que esse fato não causará problemas.

4.4 Mudança de coordenadas projetivas

De acordo com Semple e Kneebone (1952), podemos escolher as coordenadas projetivas de forma que os quatro pontos tenham as mesmas coordenadas em cada imagem. Ou seja, podemos assumir que $\mathbf{x}_i = \mathbf{x}'_i$ pensando nos dois planos de imagem como registrados no mesmo sistema de coordenadas. Vejamos como isso pode ser feito.

4.4.1 Corregistro de pontos pertencentes a diferentes planos de imagem

Suponha a existência de pontos correspondêntes em dois planos de imagem relacionados por uma transformação projetiva $\mathbf{x}' = H\,\mathbf{x}$, onde $\mathbf{x} \in \mathbb{P}^2$ e $\mathbf{x}' \in \mathbb{P}'^2$. Nosso objetivo é fazer o corregistro de pontos da segunda imagem no sistema de coordenadas da primeira imagem, mas também poderíamos fazer o corregistro dos pontos das duas imagens para um terceiro plano projetivo qualquer. Fazendo as seguintes mudanças de coordenadas projetivas

$$\begin{aligned} \overline{\mathbf{x}} &= I\,\mathbf{x} \\ \overline{\mathbf{x}}' &= H^{-1}\mathbf{x}', \end{aligned} \tag{38}$$

temos que $\mathbf{x} = \overline{\mathbf{x}}$ e $\mathbf{x}' = H\,\overline{\mathbf{x}}'$. Substituindo essas duas últimas igualdades em $\mathbf{x}' = H\,\mathbf{x}$ temos

$$\begin{aligned} \mathbf{x}' &= H\,\mathbf{x} && \Rightarrow \\ H\,\overline{\mathbf{x}}' &= H\,\overline{\mathbf{x}} && \Rightarrow \\ \overline{\mathbf{x}}' &= H^{-1}H\,\overline{\mathbf{x}} && \Rightarrow \\ \overline{\mathbf{x}}' &= \overline{\mathbf{x}}. \end{aligned}$$

Cabe ressaltar que as mudanças de coordenadas em 38 não altera nosso desenvolvimento, já que aqui nos interessa propriedades que são invariantes sob transformações projetivas, como tangências, interseções e razão cruzada. Portanto, onde foi utilizada a matriz identidade poderia ter sido usada qualquer outra homografia no caso de transferência dos pontos para um terceiro plano qualquer. A homografia H é determinada a partir dos quatro pontos correspondentes supostos inicialmente pelo artigo.

4.5 O teorema 3: pontos e epipolos co-cônicos

Aplicando as mudanças de coordenadas descritas na Subsecão 4.4.1 nos quatro pontos e nos epipolos, temos que $\mathbf{x} = \mathbf{x}'$ e todos os seis pontos em questão estarão expressos no mesmo sistema de coordenadas, no mesmo plano de imagem. Observe a Figura 32.

Teorema 4.5.1. *Os epipolos $\boldsymbol{e}$, $\boldsymbol{e}'$ e os quatro pontos $\mathbf{x}_i$ na imagem devem pertencer a uma cônica B. Reciprocamente, dois epipolos que são co-cônicos com os quatro pontos na imagem devem satisfazer à restricão epipolar.*

Demonstração. Pela subseção 1.1.3, uma cônica fica determinada por cinco pontos, como temos quatro pontos na imagem vamos deixar a cônica B na dependência de $\boldsymbol{e}$ (o quinto ponto necessário). Analogamente, o epipolo $\boldsymbol{e}'$ também define uma cônica com os quatro pontos na imagem, digamos B'. Pela subsecão 4.1, existe a homografia da reta epipolar que relaciona as quatro retas do feixe passando por $\boldsymbol{e}$ com as quatro retas do feixe passando por $\boldsymbol{e}'$. Fazendo o corregistro dos pontos e pelo teorema de Steiner, em Semple e Kneebone (1952), se dois feixes de retas passando por $\boldsymbol{e}$ e $\boldsymbol{e}'$ estão homograficamente relacionados, o ponto de interseção entre retas correspondentes desses feixes descreve uma cônica que passa por $\boldsymbol{e}$ e $\boldsymbol{e}'$. Portanto $B = B'$ e todos os seis pontos em questão são co-cônicos.

Reciprocamente, pelo teorema de Chasles em Semple e Kneebone (1952), se quatro pontos $\mathbf{x}_i \in B$ formam um feixe de retas concorrentes em um quinto ponto $\boldsymbol{e} \in B$, a razão cruzada entre as quatro retas não depende da posição de $\boldsymbol{e}$. Assim, a razão cruzada do feixe concorrente em $\boldsymbol{e}$ é igual a do feixe concorrente em $\boldsymbol{e}'$. Pela subseção 1.1.8, a razão cruzada num feixe de retas é igual a razão cruzada entre os quatro pontos de interseção

Figura 32 - Pontos co-cônicos.

Legenda: Epipolos co-cônicos com os quatro pontos na imagem produzem homografia única.

Fonte: O autor.

de uma reta qualquer com o feixe. Pelo corolário ainda na subseção 1.1.8, dois conjuntos de quatro pontos colineares têm a mesma razão cruzada se, e somente se, correspondem sob uma homografia. Pela dualidade da representação entre pontos e retas, temos a homografia entre retas epipolares. □

Esta cônica $B = B(\boldsymbol{e})$ será bastante importante durante a abordagem pois, se os dois epipolos são co-cônicos com os quatro pontos na imagem, Figura 32, existe uma *única* homografia de reta epipolar que faz a relação das retas através $\mathbf{e}$ com as retas através $\mathbf{e}'$.

4.6 O teorema 4: interseção das tangentes

Nota-se que podemos parametrizar o feixe de retas passando por $\mathbf{e}'$ (ou $\mathbf{e}$) por pontos pertencentes à cônica B, pois retas correspondentes nos dois feixes interceptam B no mesmo ponto.

Teorema 4.6.1. *As restrições de calibração são equivalentes à condição de que duas retas tangentes a ω que passam por $\boldsymbol{e}$ intersectam a cônica B nos mesmos dois pontos adicionais que as duas retas tangentes a ω' passando por $\boldsymbol{e}'$.*

Demonstração. Na subseção 4.3, vimos que as duas retas tangentes à conica ω estão

Figura 33 - Interseção entre B e as tangentes.

Legenda: Os pontos de interseção das tangentes com a cônica B coincidem, ocasionando a coincidência das projeções das cônicas ω e ω' na cônica B.

Fonte: O autor.

projetivamente relacionadas com as duas tangentes à cônica ω'. Pelo teorema de Steiner, o ponto de interseção entre duas retas correspondentes, em dois feixes homograficamente relacionados, descrevem uma cônica. Assim, as retas tangentes à cônica ω se interceptam às retas tangentes à cônica ω' em dois pontos adicionais sobre a cônica B. □

Isto é, as projeções de ω e ω' em B, através dos respectivos epipolos, devem coincidir. Essa construção geométrica pode ser visualizada na Figura 33 e será a fundação para o resto da abordagem.

4.7 O teorema 5: projeção de ω em B

A demonstração do próximo teorema é dada por Nistér e Schaffalitsky (2006), mas vamos incluir alguns detalhes e posteriormente dar uma demonstração extendida. Acompanhando pela Figura 33, podemos pensar na representação da projeção de ω e ω' em B como a reta que liga os dois pontos de interseção das tangentes com B, pois tal fato vai resolver o problema de que as tangentes não têm ordenação específica. Essa reta que liga os dois pontos de interseção, de acordo com a projeção de ω através do epipolo $\mathbf{e}$ é representada por $(\omega \diamond B)\,\mathbf{e}$.

A Equação 43 é uma fórmula projetivamente invariante de uma cônica. Para ver

isso, basta tomar a configuração para as cônicas B e ω^* (mais a frente veremos o porquê dessas configurações)

$$B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \text{e} \qquad \omega^* = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix},$$

e efetuar as multiplicações constantes na Equação 43. Lembrando que o traço de uma matriz é definido pela soma dos elementos da diagonal principal, o resultado das multiplicações é uma matriz simétrica, que pela subseção 1.1.3, representa uma cônica. Pela subseção 1.1.7, uma homografia leva uma cônica em outra cônica, a qual mantém as características invariantes sob transformação projetiva.

4.7.1 Representação canônica de uma cônica

Como vimos na subseção 1.1, o plano projetivo $\mathbb{P}^2$ tem seus pontos representados por vetores homogêneos com três componentes. Da álgebra linear sabemos que o espaço $\mathbb{R}^3$ também tem seus pontos representados por vetores com três componentes. Portanto, existe uma analogia entre as representações mesmo que no plano projetivo a relação entre vetores e pontos não seja biunívoca como em $\mathbb{R}^3$. Da mesma forma que podemos escolher a base canônica para representar vetores em $\mathbb{R}^3$, fazendo a devida transformação projetiva do sistema de coordenadas, podemos escolher uma base para representação do plano projetivo com a inclusão de um *vetor unitário* $\mathbf{u} = (1,1,1)^\top$. Assim, cada ponto em $\mathbb{P}^2$ pode ser representado em termos da base $\{\mathbf{x}_1 = (1,0,0)^\top, \mathbf{x}_2 = (0,1,0)^\top, \mathbf{x}_3 = (0,0,1)^\top, \mathbf{u} = (1,1,1)^\top\}$ onde os três primeiros vetores são chamados *pontos de referência*. Para detalhes ver teorema 1 no apêndice de Semple e Kneebone (1952).

Os pontos de referência, que são linearmente independentes, formam os vértices de um triângulo chamado *triângulo de referência*, com o ponto unitário no interior do triângulo. Esse triângulo vai servir como uma base para a representação do plano projetivo, conforme a Figura 34. Por exemplo, representando um ponto qualquer por $\mathbf{x} = (x,y,z)^\top$ e uma reta qualquer por $\boldsymbol{l} = (a,b,c)^\top$ a reta determinada pelos pontos $\mathbf{x}_1$ e $\mathbf{u}$ tem equação $z - y = 0$. De fato, se $\mathbf{x}_1 \in \boldsymbol{l}$ e $\mathbf{u} \in \boldsymbol{l}$ então

$$\boldsymbol{l}^\top \mathbf{u} = 0 \Rightarrow a + b + c = 0 \qquad \text{e} \qquad \boldsymbol{l}^\top \mathbf{x}_1 = 0 \Rightarrow a = 0. \qquad \text{Assim,} \quad b = -c. \tag{39}$$

Dado um ponto qualquer $\mathbf{x} \in \boldsymbol{l}$ a equação da reta fica

$$\boldsymbol{l}^\top \mathbf{x} = 0 \Rightarrow a\,x + b\,y + c\,z = 0,$$

Figura 34 - Triângulo de referência.

Legenda: Triângulo de referência com vértices determinados por três vetores que, juntamente com o vetor unitário, servem de base para representação do plano projetivo.

Fonte: O autor.

e usando as Relações 39 temos que

$$-c\,y + c\,z = 0 \Rightarrow z - y = 0,$$

já que $c \neq 0$ pois do contrário $\boldsymbol{l} = (0,0,0)^\top$ e não representaria reta alguma, conforme a subseção 1.1.1.

Existem algumas formas padronizadas pelas quais a representação da cônica B pode ser simplificada através da escolha conveniente de uma estrutura de referência. Uma dessas escolhas é tomar o triângulo de referência como tendo dois de seus lados tangentes à cônica, e o terceiro lado como reta polar em relação ao seu vértice oposto, e escolher um ponto da cônica como ponto unitário. Ver Figura 35. Como vimos na subseção 1.1.3, a cônica B pode ser representada na forma matricial com coordenadas homogêneas como

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{bmatrix} a & b/2 & c/2 \\ b/2 & d & e/2 \\ c/2 & e/2 & f \end{bmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

ou através da equação

$$a\,x^2 + b\,x\,y + d\,y^2 + c\,x\,z + e\,y\,z + f\,z^2 = 0.$$

Figura 35 - Relação polo-polar.

Legenda: A cônica B numa relação polo-polar com um triângulo de referência que representa o plano projetivo $\mathbb{P}^2$.
Fonte: O autor.

Supondo que a cônica passe pelos pontos $\mathbf{x}_1 = (1,0,0)^\top$ e $\mathbf{x}_3 = (0,0,1)^\top$ temos que

$$\mathbf{x}_1 \in B \Rightarrow a = 0 \quad \text{e} \quad \mathbf{x}_3 \in B \Rightarrow f = 0,$$

daí a equação da cônica será

$$d\,y^2 + b\,x\,y + c\,x\,z + e\,y\,z = 0. \tag{40}$$

Consirede a reta $\boldsymbol{l} = (l_1, l_2, l_3)^\top$ que passa pelos pontos $\mathbf{x}_1$ e $\mathbf{x}_3$,

$$\boldsymbol{l}^\top \mathbf{x}_1 = \begin{pmatrix} l_1 & l_2 & l_3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 0 \Rightarrow l_1 = 0,$$

$$\boldsymbol{l}^\top \mathbf{x}_3 = \begin{pmatrix} l_1 & l_2 & l_3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0 \Rightarrow l_3 = 0,$$

e portanto $\boldsymbol{l} = (0, l_2, 0)^\top$. Observando que $\boldsymbol{l}$ está numa relação polo-polar com o ponto $\mathbf{x}_2$, pela subseção 1.1.6 temos que $B\,\mathbf{x}_2 = \boldsymbol{l}$, daí

$$\begin{bmatrix} a & b/2 & c/2 \\ b/2 & d & e/2 \\ c/2 & e/2 & f \end{bmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} b/2 \\ d \\ e/2 \end{pmatrix} = \begin{pmatrix} 0 \\ l_2 \\ 0 \end{pmatrix},$$

e teremos $b = e = 0$ e $d = l_2$. E como $l_2 \neq 0$ a Equação 40 toma a forma

$$l_2\, y^2 + c\, x\, z = 0, \quad \text{ou} \quad y^2 = k\, x\, z \quad \text{com} \quad k = -\frac{c}{l_2}. \tag{41}$$

A cônica B passa também pelo ponto unitário $\mathbf{u} = (1, 1, 1)^\top$, assim substituindo na Equação 41 temos que $k = 1$, e a equação da cônica se reduz finalmente a

$$y^2 = x\, z, \tag{42}$$

que é a equação conônica da cônica em $\mathbb{P}^2$. Na forma matricial a cônica é representada por

$$B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

4.7.2 Representação canônica paramétrica

A Equação 42 pode ser escrita na forma

$$\left(\frac{y}{z}\right)^2 = \frac{x}{z},$$

e fazendo a substituição $\dfrac{y}{z} = \theta$ temos que $\dfrac{x}{z} = \theta^2$ e $x : y : z = \theta^2 : \theta : 1$. Assim, todo ponto pertencente à cônica B terá a forma $(\theta^2, \theta, 1)^\top$, que é realmente uma parametrização da curva, no sentido de que existe uma relação biunívoca entre cada ponto da curva e o valor do parâmetro θ. No caso do artigo de Nistér e Schaffalitzky (2006) esta formalização é bastante útil pois, como a cônica B está em função do quinto ponto, o epipolo $\boldsymbol{e}$, temos que esse ponto, e consequentemente a cônica, pode ser definida apenas pelo parâmetro θ. Podemos observar que é válida a relação

$$\begin{pmatrix} 1 & \theta & \theta^2 \end{pmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{pmatrix} 1 \\ \theta \\ \theta^2 \end{pmatrix} = 0.$$

4.7.3 Demonstração do teorema

Após a apresentação dos conceitos necessários, vamos estabelecer o teorema e demonstrá-lo em detalhes.

Teorema 4.7.1. *A projeção da IAC ω na cônica B, usando o epipolo $\boldsymbol{e}$ como centro de projeção, é dada pelos pontos de interseção entre a reta $(\omega \diamond B)\,\boldsymbol{e}$ e B, onde a cônica $(\omega \diamond B)$ é definida como*

$$(\omega \diamond B) \equiv 2\,B\,\omega^* B - tr(\omega^*\,B)\,B. \tag{43}$$

Demonstração. Pela subseção 4.6, a reta epipolar $\boldsymbol{l}$ tangente à IAC intercepta a cônica B em um ponto $\mathbf{q}$, conforme a Figura 33. Como $\mathbf{q} \in B$ e B tem a forma definida na subseção 4.7.1, temos que $\mathbf{q}$ pode ser parametrizado por $\mathbf{q} = (1, \lambda, \lambda^2)^\top$, conforme a subseção 4.7.2. Desta forma, precisamos mostrar que $\mathbf{q} \in (\omega \diamond B)\,\boldsymbol{e}$, ou seja, $[(\omega \diamond B)\,\boldsymbol{e}]^\top \mathbf{q} = 0$ onde $\boldsymbol{e} = (1, \theta, \theta^2)^\top$. A reta $\boldsymbol{l}$ pode ser definida pelos pontos $\boldsymbol{e}$ e $\mathbf{q}$ como $\boldsymbol{l} = \boldsymbol{e} \times \mathbf{q}$ e, fazendo as substituições, vem

$$\begin{aligned}
\boldsymbol{l} &= (1, \theta, \theta^2) \times (1, \lambda, \lambda^2) \\
&= (\lambda - \theta)(\theta\lambda, -(\theta + \lambda), 1) \\
&= (\theta\lambda, -(\theta + \lambda), 1)
\end{aligned}$$

ignorando o fator de escala $(\lambda - \theta)$. Como $\boldsymbol{l}$ é tangente à cônica ω, temos que a reta satisfaz a relação definida na Subseção 1.1.4, $\boldsymbol{l}^\top \omega^* \boldsymbol{l} = 0$. Parametrizando ω^* por $[c_{ij}]$ temos que a relação $\boldsymbol{l}^\top \omega^* \boldsymbol{l} = 0$ se torna

$$\begin{pmatrix} \theta\lambda & -(\theta + \lambda) & 1 \end{pmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \begin{pmatrix} \theta\lambda \\ -(\theta + \lambda) \\ 1 \end{pmatrix} = 0,$$

que é equivalente à equação

$$\theta^2\lambda^2 c_{11} + (-2\,\theta^2\,\lambda - 2\,\theta\lambda^2)\,c_{12} + 2\,\theta\lambda\,c_{13} + (\theta + \lambda)^2\,c_{22} + (-2\,\theta - 2\,\lambda)\,c_{23} + c_{33} = 0. \tag{44}$$

Usando a definição de $(\omega \diamond B)$ em 43, a parametrização de ω^* por $[c_{ij}]$ e a forma canônica de B (subseção 4.7.2), temos que

$$(\omega \diamond B) = \begin{bmatrix} 2\,c_{33} & -4\,c_{23} & 2\,c_{22} \\ -4\,c_{23} & 4\,(c_{13}+c_{22}) & -4\,c_{12} \\ 2\,c_{22} & -4\,c_{12} & 2\,c_{11} \end{bmatrix},$$

e a relação $[(\omega \diamond B)\,\boldsymbol{e}]^\top \mathbf{q} = \boldsymbol{e}^\top (\omega \diamond B)\,\mathbf{q} = 0$ se torna

$$\begin{pmatrix} 1 & \theta & \theta^2 \end{pmatrix} \begin{bmatrix} 2\,c_{33} & -4\,c_{23} & 2\,c_{22} \\ -4\,c_{23} & 4\,(c_{13}+c_{22}) & -4\,c_{12} \\ 2\,c_{22} & -4\,c_{12} & 2\,c_{11} \end{bmatrix} \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \end{pmatrix} = 0,$$

que é equivalente à equação

$$\theta^2\lambda^2 c_{11} + (-2\,\theta^2\,\lambda - 2\,\theta\lambda^2)\,c_{12} + 2\,\theta\lambda\,c_{13} + (\theta+\lambda)^2\,c_{22} + (-2\,\theta - 2\,\lambda)\,c_{23} + c_{33} = 0, \quad (45)$$

a menos do fator de escala 2. A Equação 44 é igual a Equação 45, e como 44 é uma relação válida, também é a Equação 45. O teorema segue. □

O desenvolvimento é análogo para o ponto de interseção das demais tangentes com B.

4.8 O teorema 6: projeção de ω' em B

Teorema 4.8.1. *Dado que os epipolos $\boldsymbol{e}$ e $\boldsymbol{e}'$ são co-cônicos com quatro pontos na imagem, a restrição Kruppa é equivalente à restricao de que as retas $(\omega \diamond B)\,\boldsymbol{e}$ e $(\omega' \diamond B)\,\boldsymbol{e}'$ coincidem.*

Demonstração. Pela Subseção 4.6, as retas tangentes à ω' intersectam B nos mesmos pontos adicionais, onde um deles foi denotado por $\mathbf{q}$. Como os pontos são co-cônicos podemos usar a forma canônica de B e a parametrização canônica para $\boldsymbol{e}'$. O desenvolvimento é análogo ao da subseção 4.7. □

4.9 O teorema 7: a função que relaciona os epipolos

Primeiramente nota-se que $(\omega \diamond B) = B\,(\omega \cdot B)$ se definirmos a homografia

$$(\omega \cdot B) \equiv 2\,\omega^*\,B - tr(\omega^*\,B)\,I, \quad (46)$$

onde B está sendo "cancelado"na definição de $(\omega \diamond B)$. Antes de passarmos ao teorema vamos fazer um esclarecimento sobre alguns conceitos.

4.9.1 Feixe de cônicas

Dadas duas cônicas B_1 e B_2 pertencentes a $\mathbb{P}^2$, cujo a forma quadrática é definida por $\mathbf{x}^\top B_1 \mathbf{x} = 0$ e $\mathbf{x}^\top B_2 \mathbf{x} = 0$, chamamos de *feixe de cônicas* a família de cônicas definidas por

$$B = \lambda\, B_1 + \mu\, B_2 \quad \text{onde} \quad \mathbf{x} \in \mathbb{P}^2. \tag{47}$$

Observe que as cônicas B_1 e B_2 também pertencem ao feixe, onde basta tomarmos os valores $\lambda = 1$ e $\mu = 0$ e vice-versa para verificarmos tal fato. Assim, quaisquer duas cônicas pertencentes ao feixe podem ser escolhidas como uma base para a representação do feixe. Resolvendo o sistema de equações

$$\begin{cases} \mathbf{x}^\top B_1 \mathbf{x} = 0 \\ \mathbf{x}^\top B_2 \mathbf{x} = 0 \end{cases}$$

temos vários casos para as soluções mas trataremos aqui somente do caso com quatro soluções distintas, já que Nistér e Schaffalitzky (2006) pressupõe que sejam fornecidos quatro pontos em posições gerais no plano da imagem, os quais são utilizados na definição do feixe de cônicas.

Chamamos de *pontos base* de um feixe de cônicas, os pontos que são comuns a todas as cônicas do feixe, isto é, eles são os pontos de interseção das duas cônicas B_1 e B_2 que são a base do feixe. Assim, o feixe de cônicas fica definido pelos quatro pontos de interseção dessas duas cônicas.

Como os quatro pontos $\mathbf{x}_i$ que definem o feixe são todos diferentes, temos que o feixe contém exatamente três cônicas degeneradas que são constituídas por pares de retas. Definindo as retas por dois pontos temos um total de combinações de $C_{4,2} = 6$ retas dadas por

$$\boldsymbol{l}_k = \mathbf{x}_i \times \mathbf{x}_j \quad \text{com} \quad k = 1, ..., 6,\ i = 1, 2, 3,\ j = 2, 3, 4, \quad \text{e} \quad i < j.$$

Pela subseção 1.1.5, cada cônica degenerada do feixe pode ser expressa como

$$\begin{aligned} B_1 &= \boldsymbol{l}_1 \boldsymbol{l}_2^\top + \boldsymbol{l}_2 \boldsymbol{l}_1^\top \\ B_2 &= \boldsymbol{l}_3 \boldsymbol{l}_4^\top + \boldsymbol{l}_4 \boldsymbol{l}_3^\top \\ B_3 &= \boldsymbol{l}_5 \boldsymbol{l}_6^\top + \boldsymbol{l}_6 \boldsymbol{l}_5^\top. \end{aligned} \tag{48}$$

Observa-se que com seis retas temos um total de combinações de $C_{6,2} = 15$ cônicas, no entanto duas retas escolhidas devem passar pelos quatro pontos base, ou seja, duas retas não podem compartilhar um mesmo ponto base. Portanto temos apenas três cônicas degeneradas, as quais são normalmente usadas como base do feixe de cônicas por serem mais facilmente obtidas.

4.9.2 Estabelecimento e demonstração do teorema

O esquema para o próximo teorema está representado na Figura 36 e o símbolo $\sim$ significa igualdade a menos da escala.

Teorema 4.9.1. *Os epipolos estão ralacionados por*

$$(\omega' \cdot B)\, \boldsymbol{e}' \sim (\omega \cdot B)\, \boldsymbol{e} \sim \mathbf{p} \tag{49}$$

o que implica numa função de sétimo grau

$$\mathbf{e} \to \mathbf{e}' \sim (\omega' \cdot B)^* \, (\omega \cdot B)\, \mathbf{e}. \tag{50}$$

A restrição Kruppa disponibiliza quatro soluções para o epipolo $\mathbf{e}$ *na cônica* B*, e as soluções são as interseções de* B *com a curva*

$$C \equiv (\omega \cdot B)^\top \, (\omega' \cdot B)^{*\top} \, B \, (\omega' \cdot B)^* (\omega \cdot B). \tag{51}$$

Demonstração. Pela subseção 4.8 $(\omega \diamond B)\, \boldsymbol{e} \sim (\omega' \diamond B)\, \boldsymbol{e}'$, e como a cônica B é própria (posto completo) podemos invertê-la e eliminá-la na definição de $(\omega \diamond B)$ e de $(\omega' \diamond B)$ para chegar à Relação 49. Assumindo que $|(\omega' \cdot B)| \neq 0$, podemos invertê-la e chegar na função em 50. Como $\boldsymbol{e}' \in B \Rightarrow \boldsymbol{e}'^\top B\, \boldsymbol{e}' = 0$, e substituindo a função em 50 vem

$$\begin{aligned}
0 &= \boldsymbol{e}'^\top B\, \boldsymbol{e}' \\
&= ((\omega' \cdot B)^* \, (\omega \cdot B)\, \mathbf{e})^\top B \, (\omega' \cdot B)^* \, (\omega \cdot B)\, \mathbf{e} \\
&= \boldsymbol{e}^\top (\omega \cdot B)^\top (\omega' \cdot B)^{*\top} B \, (\omega' \cdot B)^* \, (\omega \cdot B)\, \boldsymbol{e} \\
&= \boldsymbol{e}^\top C\, \boldsymbol{e}.
\end{aligned}$$

Como $\boldsymbol{e} \in C$ e $\boldsymbol{e} \in B$, então $\boldsymbol{e}$ está na interseção entre as duas curvas. No caso em que $|(\omega' \cdot B)| = 0$, devemos observar que, para câmeras e pontos em posições gerais (que excluem degenerações por serem casos raros), não há motivos para que $(\omega \cdot B)$ e $(\omega' \cdot B)$ se degenerem ao mesmo tempo. Assim, sendo $|(\omega' \cdot B)| = 0$ podemos reescrever as equações 50 e 51 nas formas

$$\boldsymbol{e} \sim (\omega \cdot B)^* (\omega' \cdot B)\, \boldsymbol{e}' \quad \text{e} \quad C_1 \equiv (\omega' \cdot B)^\top (\omega \cdot B)^{*\top} \, B \, (\omega \cdot B)^* \, (\omega' \cdot B).$$

Figura 36 - Interseção entre curvas.

Legenda: Epipolos que satisfazem as restrições de projeção e de calibração podem ser encontrados como as interseções da cônica B com a curva C.

Analogamente à dedução anterior, $\boldsymbol{e}' \in C_1$ e a interseção de B e C_1 vai gerar quatro soluções distintas para $\boldsymbol{e}'$. Podemos perceber, pela Equação 50, a relação polo-polar entre o ponto $\mathbf{p} \sim (\omega' \cdot B)\,\boldsymbol{e}'$ e a reta $\boldsymbol{l}$ dada por $\boldsymbol{l} = B\,(\omega' \cdot B)\,\boldsymbol{e}'$ ou $\boldsymbol{l} = B\,\mathbf{p}$.

Vimos que $\mathbf{p} \sim (\omega \cdot B)\,\boldsymbol{e}$ e, considerando novamente a curva

$$C \equiv (\omega \cdot B)^\top\,(\omega' \cdot B)^{*\top}\,B\,(\omega' \cdot B)^*(\omega \cdot B)$$

temos, mesmo com $|(\omega' \cdot B)| = 0$, que

$$\boldsymbol{e}^\top C\,\boldsymbol{e} = 0 \quad \text{ou} \quad \boldsymbol{e}^\top(\omega \cdot B)^\top\,(\omega' \cdot B)^{*\top}\,B\,(\omega' \cdot B)^*(\omega \cdot B)\boldsymbol{e} = 0,$$

pois

$$\begin{aligned}
(\omega' \cdot B)^*(\omega \cdot B)\boldsymbol{e} &\sim (\omega' \cdot B)^*\mathbf{p} \\
&\sim (\omega' \cdot B)^*(\omega' \cdot B)\,\boldsymbol{e}' \\
&\sim |(\omega' \cdot B)|\,I\,\boldsymbol{e}' \\
&\sim 0\,I\,\boldsymbol{e}' \\
&\sim \mathbf{0}.
\end{aligned}$$

Assim, mesmo sendo $(\omega' \cdot B)$ degenerada temos que $\boldsymbol{e} \in C$. $\square$

Pela subseção 4.9.1, temos que quaisquer duas cônicas de um feixe de cônicas podem ser usadas como base do feixe, assim a cônica $B(\boldsymbol{e})$ pode ser expressa na forma

$$B(\boldsymbol{e}) = (\boldsymbol{e}^\top B_2 \boldsymbol{e})\, B_1 - (\boldsymbol{e}^\top B_1 \boldsymbol{e})\, B_2,$$

onde B está sendo expressa em termos de $\boldsymbol{e}$, usando as cônicas B_1 e B_2 como base do feixe. Repare que os escalares $\boldsymbol{e}^\top B_2 \boldsymbol{e}$ e $\boldsymbol{e}^\top B_1 \boldsymbol{e}$ são sempre diferentes de zero pois, dados os quatro pontos $\mathbf{x}_i$ na imagem, a cônica B fica unicamente determinada pelo quinto ponto $\boldsymbol{e}$, e portanto $\boldsymbol{e} \notin B_1$ e $\boldsymbol{e} \notin B_2$.

4.10 O teorema 8: a curva C

Como a cônica B está sendo expressa em termos de $\boldsymbol{e}$, para qualquer escolha de $\boldsymbol{e}$ teremos que $\boldsymbol{e}^\top B\, \boldsymbol{e} = 0$, mas esse não é o caso para a curva C definida na subseção 4.9.2, e assim temos o próximo teorema.

Teorema 4.10.1. *Um epipolo hipotético $\boldsymbol{e}$ que define uma cônica B satisfaz a equação de 16º grau*

$$\boldsymbol{e}^\top C\, \boldsymbol{e} = 0 \tag{52}$$

se, e somente se, satisfaz as restrições de projetividade e calibração.

Demonstração. Esse teorema é uma consequência do teorema na subseção 4.9.2 onde vimos que os epipolos estão na interseção da cônica B com a curva C, e portanto devem satisfazer à equação 52. O teorema da subseção 4.9.2 é dependente da restrição de calibração na subseção 4.6 e da restrição de projeção na subseção 4.7. □

4.11 Teoremas 9 a 20: a curva G e suas propriedades

Teoremas 9, 10 e 11.

Esses teoremas têm demonstração completa no artigo e não necessitam de consulta em bibliografia extra, apesar de conterem bastante desenvolvimento algébrico. A utilidade deles se resume em verificar que os possíveis epipolos pertencem à curva de equação $\boldsymbol{e}^\top C\, \boldsymbol{e} = 0$ mesmo no caso onde $B(\boldsymbol{e})$ é uma cônica degenerada, ou seja, a equação $\boldsymbol{e}^\top C\, \boldsymbol{e} = 0$ contém $|B|$ como fator. Nesse caso, os epipolos pertencem às seis retas que passam pelos quatro pontos $\mathbf{x}_i$ na imagem. Em seguida, é demonstrado como o fator $|B|$ pode ser eliminado da equação de 16º grau $\boldsymbol{e}^\top C\, \boldsymbol{e} = 0$ para se chegar numa equação de

Figura 37 - Redução de C a G.

Legenda: Podemos observar, à direita, a curva dos possíveis epipolos definida pela equação de décimo grau. À esquerda, a mesma curva com as seis retas, antes de removido o fator $|B|$ da equação de décimo-sexto grau.
Fonte: Nistér e Schaffalitsky (2006).

10º grau definida por

$$e^\top G\, e = 0, \tag{53}$$

onde a curva G é dada por

$$G = 4\, U^\top D'^*\, B^*\, D'^*\, U - 4\, t'^2\, B\, D\, D'^*\, (D\, B - t\, I) + s\, D^* - t^2\, t'^2\, D'^*, \tag{54}$$

com as definições:

$$\begin{aligned} D &\equiv \omega^*, \\ t &\equiv tr(D\, B), \\ U &\equiv (\omega \cdot B) \text{ e} \\ s &= (16\, |B|\, |D'|\, t' + t'^4 - 4\, t'^2\, tr(B^*\, D'^*)). \end{aligned}$$

Na Figura 37 podemos visualizar dois exemplos de curvas. À esquerda, uma curva com as seis retas passando pelos quatro pontos na imagem, definida pela Equação 52 de 16º grau que contém $|B|$ como fator. À direita, a mesma curva sem as seis retas, depois de removido o fator $|B|$, definida pela Equação 53 de 10º grau.

Teorema 12.

Teorema 4.11.1. *Dados três pontos correspondentes e o epipolo* $\boldsymbol{e}$ *em uma imagem, as restrições de projeção e calibração possibilitam quatro soluções para o epipolo* $\boldsymbol{e}'$ *na outra imagem.*

A demonstração desse teorema é construtiva e se trata de um algoritmo fornecido no apêndice A do artigo. Temos quatro possibilidades para o epipolo $\boldsymbol{e}$ obtido como interseção da cônica B com a curva de 10º grau G. E para cada um deles, temos mais quatro possibilidades para o epipolo $\boldsymbol{e}'$ de acordo com o algoritmo B publicado no apêndice do artigo. Ou seja, como necessitamos dos pares de epipolos para obter a matriz esencial E, teremos um total de 16 possibilidades para E para cada cônica B do feixe de cônicas passando pelos quatro pontos na imagem dados inicialmente.

Teoremas 13 a 17.

Os teoremas 13 a 17 argumentam que a curva de grau 10 é exatamente a curva dos possíveis epipolos, e servem de base para o estabelecimento do principal resultado do artigo. Esses teoremas também possuem suas demonstrações no artigo.

Teorema 18.

É o principal resultado do artigo e sua demonstração é uma consequência dos teoremas 7, 10 e 17. O teorema é colocado como segue.

Teorema 4.11.2. *Um ponto é um possivel epipolo de acordo com as restrições de projetividade e calibração se, e somente se, o ponto pertence à curva de grau 10 definida pela equação* $e^\top G e = 0$ *conforme 53. Para pontos* $\boldsymbol{e}$ *em geral pertencentes a essa curva, o outro epipolo* $\boldsymbol{e}'$ *pode ser calculado usando a função de sétimo grau,* $\mathbf{e} \to \mathbf{e}' \sim (\omega' \cdot B)^* (\omega \cdot B)\, \mathbf{e}$*, dada em 50.*

Na Figura 38 observamos alguns exemplos de curvas de décimo grau com seus possíveis epipolos.

Teoremas 19 e 20.

Os teoremas 19 e 20 fornecem mais propriedades da curva de grau dez. O teorema 19 é uma versão mais completa do teorema 7 que se aplica mesmo quando a cônica B,

Figura 38 - Curvas de grau 10.

Legenda: Quatro exemplos de curvas de décimo grau definidas pela Equação 53, juntamente com seus respectivos epipolos.

Fonte: Nistér e Schaffalitsky (2006).

definida em 47, é degenerada.

4.12 Teoremas 21 a 25: orientação e convergência

Esses teoremas tratam das restrições de orientação, para duas câmeras, e se resumem em dois fatos que podem ser visualizados na Figura 39. Primeiro, a *restrição de orientação epipolar*, onde as duas semirretas retroprojetadas por pontos correspondentes nas imagens devem estar no mesmo semiplano do plano epipolar (já que a reta base divide o plano epipolar em dois semiplanos). Ou seja, os pontos correspondentes devem estar alojados no mesmo semiplano. Segundo, a *restrição de convergência*, onde essas duas semirretas devem convergir para um ponto comum no espaço projetivo 3D.

4.12.1 A restrição de orientação

Após determinados os epipolos, podemos determinar as retas epipolares através do produto vetorial entre os quatro pontos $\mathbf{x}_i$ e os epipolos. Com a retas epipolares podemos determinar a homografia da reta epipolar conforme a subseção 4.1, para em seguida determinarmos a matriz essencial E e a duas câmeras correspondentes. Mas existem quatro configurações para a determinação das duas câmeras, que são obtidas permutando a posição das duas câmeras (revertendo a orientação da reta base), ou rotacionando uma das câmeras 180° em torno da reta base, ou os dois. As quatro configurações podem ser observadas na figura 39. A restrição de orientação é exatamente o fato de o ponto 3D deve estar na parte da frente das duas câmeras. Nistér e Schaffalitsky (2006) apresentam um método onde a restrição é satizfeita dependendo das posições relativas entre os epipolos fornecidos pela curva de grau 10 e os quatro pontos $\mathbf{x}_i$.

4.12.2 A restrição de convergência

Após a aplicação da restrição de orientação precisamos saber se as retas retroprojetadas pelos pontos correspondentes se intersectam em um ponto no plano epipolar. Tal interseção depende da relação entre o ângulo formado pelas retroprojeções de $\mathbf{x}_i$ e $\boldsymbol{e}$, e o ângulo formado pelas retroprojeções de $\mathbf{x}'_i$ e $\boldsymbol{e}'$. Um ponto 3D no sistema de coordenadas da câmera pode ser escrito como $\mathbf{X} = (\lambda\mathbf{d}^\top, 1)^\top$, e é projetado na imagem através da

Figura 39 - Possíveis configurações para as câmeras.

Fonte: Hartley e Zisserman (2004).

equação de projeção

$$\begin{aligned}
\mathbf{x} &= K[I|\mathbf{0}]\mathbf{X} \\
\mathbf{x} &= K[I|\mathbf{0}](\lambda\mathbf{d}^\top, 1)^\top \\
\mathbf{x} &= K\mathbf{d} \quad \text{a menos da escala.}
\end{aligned}$$

O vetor $\mathbf{d}$ é a direção da reta retroprojetada por $\mathbf{x}$ e pode ser obtido pela expressão $\mathbf{d} = K^{-1}\mathbf{x}$, medido no sistema de coordenadas da câmera. Sabemos da álgebra linear que, dados dois vetores $\mathbf{d}_1$ e $\mathbf{d}_2$, a desigualdade de Cauchy-Schuarz é dada pela relação

$$|\langle \mathbf{d}_1, \mathbf{d}_2 \rangle| \leq \|\mathbf{d}_1\| \, \|\mathbf{d}_2\| \quad \text{ou} \quad -1 \leq \frac{\langle \mathbf{d}_1, \mathbf{d}_2 \rangle}{\|\mathbf{d}_1\| \, \|\mathbf{d}_2\|} \leq 1,$$

onde, usando a multiplicação matricial podemos derivar uma relação para o cosseno do ângulo entre dois vetores

$$\cos\theta = \frac{\mathbf{d}_1^\top \mathbf{d}_2}{\sqrt{\mathbf{d}_1^\top \mathbf{d}_1}\sqrt{\mathbf{d}_2^\top \mathbf{d}_2}},$$

com os vetores $\mathbf{d}_1 = K^{-1}\mathbf{x}$ e $\mathbf{d}_2 = K^{-1}\boldsymbol{e}$. Aplicando a substituição temos

$$\cos\theta = \frac{(K^{-1}\mathbf{x})^\top K^{-1}\boldsymbol{e}}{\sqrt{(K^{-1}\mathbf{x})^\top K^{-1}\mathbf{x}}\sqrt{(K^{-1}\boldsymbol{e})^\top K^{-1}\boldsymbol{e}}}$$

Figura 40 - Ângulo de convergência.

Legenda: O ângulo θ formado pelas retas retroprojetadas pelos pontos $\mathbf{x}_i$ e $\boldsymbol{e}$ através do centro de projeção $\mathbf{C}_1$ da primeira câmera. O procedimento é análogo para $\mathbf{x}'_i$ e $\boldsymbol{e}'$ na segunda câmera.

Fonte: Adaptado, Hartley e Zisserman (2004).

$$\cos\theta = \frac{\mathbf{x}^\top K^{-\top} K^{-1} \boldsymbol{e}}{\sqrt{\mathbf{x}^\top K^{-\top} K^{-1} \mathbf{x}} \sqrt{\boldsymbol{e}^\top K^{-\top} K^{-1} \boldsymbol{e}}}. \tag{55}$$

Como Nistér e Schaffalitsky (2006) supõem que a câmera é calibrada, ou seja, conhecemos a matriz K, podemos determinar o ângulo entre duas retas retroprojetadas por pontos no plano da imagem. Por isso, uma câmera calibrada é chamada *sensor de direção*.

Observa-se que, pela Subseção 4.2.1, podemos determinar a imagem da cônica absoluta IAC a partir da matriz de calibração, e a Equação 55 se torna

$$\cos\theta = \frac{\mathbf{x}^\top \omega\, \boldsymbol{e}}{\sqrt{\mathbf{x}^\top \omega\, \mathbf{x}} \sqrt{\boldsymbol{e}^\top \omega\, \boldsymbol{e}}}. \tag{56}$$

O desenvolvimento é análogo para $\mathbf{x}'_i$ e $\boldsymbol{e}'$ escritos em outro sistema de coordenadas, pois a Equação 56 é projetivamente invariante. De fato, considerando uma transformação projetiva H onde, pela Subseção 1.1.7, $\mathbf{x}' = H\,\mathbf{x}$, a Equação 56 fica

$$\cos\theta = \frac{(H^{-1}\mathbf{x}')^\top \omega\, H^{-1}\boldsymbol{e}'}{\sqrt{(H^{-1}\mathbf{x}')^\top \omega\, H^{-1}\mathbf{x}'} \sqrt{(H^{-1}\boldsymbol{e}')^\top \omega\, H^{-1}\boldsymbol{e}'}}$$

$$\cos\theta = \frac{\mathbf{x}'^\top H^{-\top} \omega\, H^{-1}\boldsymbol{e}'}{\sqrt{\mathbf{x}'^\top H^{-\top} \omega\, H^{-1}\mathbf{x}'} \sqrt{\boldsymbol{e}'^\top H^{-\top} \omega\, H^{-1}\boldsymbol{e}'}}$$

$$\cos\theta = \frac{\mathbf{x}'^\top \omega' \, \boldsymbol{e}'}{\sqrt{\mathbf{x}'^\top \omega' \, \mathbf{x}'}\sqrt{\boldsymbol{e}'^\top \omega' \, \boldsymbol{e}'}}$$

onde $\omega' = H^{-\top} \omega \, H^{-1}$ (como qualquer cônica se transforma) de acordo com a subseção 1.1.7.

4.13 Incluindo a terceira imagem

As novidades do artigo se encontram na abordagem da reconstrução 3D para duas imagens e o procedimento seguinte segue o modelo padrão de projetar as reconstruções dos quatro pontos 3D na terceira imagem e minimizar o erro em relação aos dados já disponíveis. Assim, dados quatro pontos numa imagem e suas respectivas correspondências em outras duas imagens (todas com as câmeras calibradas), podemos escolher duas dessas imagens para traçarmos a curva de décimo grau.

Com os quatro pontos correspondentes $x \leftrightarrow x'$ podemos determinar os oito graus de liberdade de uma homografia H que é usada para corregistrar os dois planos de imagens num único plano. Os pontos co-registrados juntamente com os epipolos formam uma cônica que fica na dependência de $\boldsymbol{e}$ que pode ser parametrizado por θ, $\boldsymbol{e} = (1, \theta, \theta^2)^\top$. Para cada estimativa de θ teremos uma cônica

$$B(e) = (\boldsymbol{e}^\top B_2 \boldsymbol{e}) \, B_1 - (\boldsymbol{e}^\top B_1 \boldsymbol{e}) \, B_2,$$

onde as cônicas B_1 e B_2 são duas cônicas quaisquer usadas para parametrizar o feixe de cônicas ao qual pertencem. Podemos usar os quatro pontos da imagem para determinar B_1 e B_2 que nesse caso serão degeneradas. As cônicas ω e ω' são obtidas usando

$$\omega = K^{-\top} K^{-1} \quad \text{e} \quad \omega' = K'^{-\top} K'^{-1},$$

as quais são necessárias para obter, junto com a cônica B, a função que relaciona os epipolos

$$\mathbf{e}' \sim (\omega' \cdot B)^* \, (\omega \cdot B) \, \mathbf{e}.$$

As cônicas B, ω e ω^* são utilizadas para calcular a curva G conforme a Relação 54. Os quatro possíveis epipolos $\boldsymbol{e}$ são calculados como a interseção da curvas B e G conforme o algoritmo exposto no apêndice A do artigo, mas também podemos usar as bases de Gröbner e a matriz de ação para resolver numericamente tal sistema.

Como vimos, temos quatro soluções para $\mathbf{e}$ e cada solução pode ser usada para calcular $\mathbf{e}'$ na segunda imagem de acordo com a Equação 50 e com o algoritmo no apêndice B do artigo, o qual fornece quatro soluções para $\boldsymbol{e}'$. Com os epipolos determinamos a ho-

Figura 41 - Complexidade da reprojeção.

Legenda: A reprojeção do feixe de curvas na terceira imagem e do quarto ponto reconstruído nos dá uma ideia da dificuldade do problema de três imagens com quatro pontos.
Fonte: Nistér e Schaffalitsky (2006).

mografia da reta epipolar (subseção 4.1) e daí, a matriz essencial relativa às duas imagens (subseção 1.4.8). Em seguida, podemos extrair as câmeras e usá-las para reconstrução dos pontos 3D a partir da triangulação usando duas imagens. Três pontos 3D juntamente com suas projeções na terceira imagem (já conhecidas) são suficientes para determinar a terceira câmera, e podemos usar o quarto ponto para ser projetado na terceira imagem e comparar a projeção com o dado já obtido para refinar o resultado. Ou seja, o procedimento de determinar a terceira câmera consiste em minimizar um erro na imagem. Na subseção 2 são citados mais alguns trabalhos que permitem realizar essa tarefa.

Como o problema é supra restringido por um grau de liberdade, com os ruídos das correspondências, os valores encontrados para a reprojeção do quarto ponto não coincidirá com o dado observado mesmo que usada a correta solução. Como para cada θ escolhido teremos dezesseis configurações de câmeras e dezesseis reprojeções, todas essas reprojeções definem uma complexidade enorme de curvas. Na Figura 41 é dada uma ideia da dificuldade extrema do problema envolvendo três imagens e quatro pontos.

5 CONCLUSÃO

Existe toda uma engenhosidade no emprego de coordenadas homogêneas para fazer com que uma transformação projetiva funcione como uma transformação linear, mesmo que exista (disfarçadamente) a aplicação de uma translação e normalização por divisão. Além disso, a representação de pontos em coordenadas homogêneas favorece a representação de parâmetros e restrições através de objetos geométrico-algébricos abstratos, como pontos no infinito e a cônica absoluta, e a dedução consistente de propriedades como a interseção de retas paralelas.

As pesquisas em visão computacional tridimensional são auxiliadas por outras áreas da matemática aparentemente desconexas da geometria projetiva, pois na seção 2 verificamos a utilização do quatérnion de Hamilton para parametrizar uma matriz de rotação e das bases de Gröbner e matriz de ação na solução de sistemas de equações polinomiais, ambos empregados na determinação de uma câmera. A própria utilização de sistemas de equações polinomiais com muitas variáveis na modelagem de problemas atuais pode se constituir uma novidade para vários pesquisadores.

Constatamos na seção 3 alguns benefícios do uso do tensor trifocal num sistema com três imagens em detrimento do uso puramente da geometria epipolar. Há a diminuição de alguns casos de degeneração para a correspondência de pontos em três imagens, e há também o benefício da determinação direta de uma reta numa imagem dadas as retas correspondentes nas outras duas imagens. Não há equação para a transferência direta de retas usando apenas geometria epipolar.

O estudo da geometria trifocal envolve uma quantidade grande de conhecimento da teoria básica de geometria projetiva conforme vimos na seção 4. Mais ainda, estudar geometria trifocal em problemas de reconstrução requer o conhecimento também da geometria epipolar, não somente em termos de comparação sobre qual das abordagens funciona melhor para reconstrução, mas também para entendermos de que forma a geometria epipolar pode ser útil num sistema com três imagens. Ficou clara a complexidade da tarefa de se determinar as três câmeras num sistema trifocal.

Fazer essa determinação em termos de geometria diferencial é uma das fronteiras em visão computacional. Futuramente, temos a tarefa de tentar resolver esse problema usando geometria diferencial trifocal de primeira ordem com pontos-tangentes. Ademais, pretende-se verificar as vantagens dessa abordagem em determinação de estrutura a partir dos movimentos das câmeras e em sistemas de autocalibração estéreo.

REFERÊNCIAS

BALLARD, D. H. ; BROWN, C. M. *Computer Vision*. Prentice-Hall, 1982.

BYRÖD, M. ; JOSEPHSON, K. ; ÅSTRÖM, K. Fast and stable polynomial equation solving and its application to computer vision. *International Journal of Computer Vision*, 84(3):237–255, 2009.

CHEN, H. Pose determination from line-to-plane correspondences: Existence condition and closed-form solutions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(6):530–541, 1991.

LITTLE, J. ; COX, D. ; O'SHEA, D. *Using Algebraic Geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer-Verlag, 2005.

LITTLE, J. ; COX, D. ; O'SHEA, D. *Ideals, Varieties, and Algorithms -An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics. Springer, third edition, 2007.

DHOME, M. ; RICHETIN, M. ; LAPRESTÉ, J. T. ; RIVES, G. Determination of the attitude of 3d objects from a single perspective view. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(12):1265–1278, 1989.

FABBRI, R. ; KIMIA, B. B. 3D curve sketch: Flexible curve- based stereo reconstruction and calibration. *CVPR*, 2010a.

FABBRI, R. ; KIMIA, B. B. *Multiview Differential Geometry in Application to Computer Vision*. PhD thesis, Brown University, may 2011.

FABBRI, R. ; KIMIA, B. B. From multiview image curves to 3D drawings. *CVPR*, 2015.

FABBRI, R. ; KIMIA, B. B. 3D curve sketch: Flexible curve-based stereo reconstruction and calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, California, USA, 2010b. IEEE Computer Society Press.

FABBRI, R. ; GIBLIN, P. J. ; KIMIA, B. B. Camera pose estimation using first-order curve differential geometry. In *Proceedings of the IEEE European Conference in Computer Vision*, Lecture Notes in Computer Science. Springer, 2012.

FAUGERAS, O. *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993.

FAUGERAS, O. ; LUONG, Q. T. ; PAPADOPOULO, T. *The geometry of multiple images - the laws that govern the formation of multiple images of a scene and some of their applications,*. MIT Press, 2001.

FISCHLER, M. ; BOLLES, R. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

FORSYTH, D. A. ; PONCE, J. *Computer Vision, A Modern Approach.* Prentice Hall, 2003.

GRAYSON, D. ; STILLMAN, M. Macaulay 2, an open source computer algebra software., 1993-2002. URL ⟨www.math.uiuc.edu/Macaulay2/⟩.

HARTLEY, R. I. ; ZISSERMAN, A. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, second edition, 2004.

HARTLEY, R. I. ; Li, H. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 34(23):2303–2314, dezembro 2012.

KUANG, Y. ; ÅSTRÖM, K. Pose estimation with unknown focal length using points, directions and lines. In *Proceedings of IEEE International Conference on Computer Vision (ICCV2013)*, pages 529–536, 2013.

KUKELOVA, Z. ; BUJNAK, M. ; PAJDLA, T. Automatic generator of minimal problem solvers. In *ECCV (3)*, pages 302–315, 2008a.

KUKELOVA, Z. ; BUJNAK, M. ; PAJDLA, T. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *BMVC*, 2008b.

KÖSER, K. ; KOCH, R. Differential spatial resection - pose estimation using a single local image feature. In David A. Forsyth, Philip H. S. Torr, and Andrew Zisserman, editors, *ECCV (4)*, pages 312–325, 2008.

LONGUET-HIGGINS, H. C. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Setembro 1981.

LUONG, Q. T. ; VIÉVILLE, T. Canonical representations for the geometries of multiple pro- jective views. *Computer Vision and Image Understanding*, pages 193–229, setembro 1996.

MATTHIES, L. et al. Computer vision on mars. *International Journal of Computer Vision*, 2007.

NISTÉR, D. ; SCHAFFALITZKY, F. Four Points in Two or Three Calibrated Views: Theory and Practice. *International Journal of Computer Vision*, 67:211–231, 2006.

LUONG, Q. ; FAUGERAS, O. D. ; MAYBANK, S. Camera self-calibration: Theory and experiments. In *European Conference on Computer Vision*, volume LNCS 588, page 321–334. Springer-Verlag, 1992.

NÖLLE, M. et al. Review and analysis of solutions of the three point perspective pose estimation problem. In *Int. J. of Computer Vision.*, page 331–356, 1994.

RAMALINGAM, S. ; BOUAZIZ, S. ; STURM, P. F. Pose estimation using both points and lines for geo-localization. In *ICRA*, pages 4716–4723. IEEE, 2011.

SCHMID, C. ; ZISSERMAN, A. The geometry and matching of lines and curves over multiple views. *International Journal of Computer Vision*, 40(3):199–233, 2000.

SEMPLE, J. G. ; KNEEBONE, G. T. *Algebraic Projective Geometry.* Oxford University Press, 1952.

SPETSAKIS, M. E. ; ALOIMONOS, J. Structure from motion using line correspondences. *The International Journal of Computer Vision*, 1990.

SPRINGER, C. E. *Geometry and Analysis in Projective Space.* Freeman, 1964.

APÊNDICE A – Alguns conceitos e definições de álgebra linear

Esta seção do apêndice se destina a apresentar algumas ferramentas da álgebra linear necessárias ao entendimento de algumas seções da dissertação, funcionando como complementação dessas seções.

A.1 A matriz anti-simétrica $[\mathbf{v}]_\times$

Dado um vetor $\mathbf{v} = (v_1, v_2, v_3)^\top$, é possível construir uma matriz anti-simétrica com as componentes de $\mathbf{v}$ na forma

$$[\mathbf{v}]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix},$$

onde a notação $[\mathbf{v}]_\times$ indica uma relação com o produto vetorial entre um vetor qualquer $\mathbf{u}$ e $\mathbf{v}$.

$$\mathbf{v} \times \mathbf{u} = [\mathbf{v}]_\times \mathbf{u} = (\mathbf{v}^\top [\mathbf{u}]_\times)^\top$$

Sabemos que o produto vetorial entre um vetor e ele mesmo é sempre o vetor nulo, portanto o vetor $\mathbf{v}$ é o vetor nulo à direita e à esquerda de $[\mathbf{v}]_\times$. Desta forma, esta matriz anti-simétrica será sempre definida por seu vetor nulo.

Observe, ainda, que uma matriz anti-simétrica qualquer M satisfaz a relação $\mathbf{v}^\top M\, \mathbf{v} = 0$. Pois,

$$\begin{aligned}
\mathbf{v}^\top M\, \mathbf{v} &= \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{bmatrix} 0 & m_{12} & m_{13} \\ -m_{12} & 0 & m_{23} \\ -m_{13} & -m_{23} & 0 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\
&= \begin{pmatrix} -v_2 m_{12} - v_3 m_{13} & v_1 m_{12} - v_3 m_{23} & v_1 m_{13} + v_2 m_{23} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \\
&= -v_1 v_2 m_{12} - v_1 v_3 m_{13} + v_1 v_2 m_{12} - v_2 v_3 m_{23} + v_1 v_3 m_{13} + v_2 v_3 m_{23} \\
&= 0.
\end{aligned}$$

A.2 A decomposição SVD

Uma das formas de se obter a decomposição em valores singulares de uma matriz A é aplicando a decomposição em autovalores na matriz $A^\top A$. Assim, queremos determinar a decomposição

$$A = U\,D\,V^\top$$

onde as matrizes U e V são ortogonais. Observe que calculando $A^\top A$ temos

$$\begin{aligned} A^\top A &= V\,D\,U^\top U\,D\,V^\top \\ &= V\,D^2 V^{-1}, \end{aligned}$$

onde $V\,D^2V^{-1}$ é uma decomposição em autovalores.

Desta forma, aplicando a decomposição em autovalores na matriz $A^\top A$, conseguimos diretamente a matriz $V^\top$ da decomposição SVD, e basta extrair a raiz quadrada das componentes da diagonal de D para obtermos os valores singulares. Com uma inversão de matrizes podemos obter a matriz U:

$$\begin{aligned} A &= U\,D\,V^\top && \Rightarrow \\ U &= A\,(D\,V^\top)^{-1} && \Rightarrow \\ U &= A\,V^{-\top}D^{-1}. \end{aligned}$$

Não são todos os problemas que envolvem a decomposição SVD que necessitam que sejam calculadas todas a matrizes. Existem problemas onde não é necessário o cálculo da matriz U por exemplo. Então a complexidade computacional vai depender também daquilo que é necessário ao problema.

A.3 A pseudo-inversa de P

Primeiramente precisamos definir a inversa de uma matriz quadrada diagonal. Sendo D essa matriz, sua inversa será

$$D^+_{ii} = \begin{cases} 0 \ , \ \text{se} \ D_{ii} = 0 \\ \frac{1}{D_{ii}} \ , \ \text{se} \ D_{ii} \neq 0. \end{cases}$$

A matriz $P_{3\times 4}$ tem menor quantidade de linhas do que colunas, e nesse caso devemos extender a matriz P acrescentando uma linha de zeros para que a mesma se torne uma matriz quadrada. Não poderemos inverter tal matriz quadrada pois com uma linha de zeros ela será singular. Assim, aplicamos a decomposição SVD e em seguida definimos a

pseudo-inversa. Se

$$P = U\,D\,V^\top$$

é a decomposição SVD, podemos definir a pseudo-inversa como

$$P^+ = V\,D^+U^\top.$$

De fato, podemos verificar essa definição usando D^+ e o fato de que as matrizes U e V são ortogonais (suas inversas são suas transpostas).

$$\begin{aligned} P^+P &= V\,D^+U^\top\,U\,D\,V^\top \\ &= V\,D^+I\,D\,V^\top \\ &= V\,D^+D\,V^\top \\ &= V\,I\,V^\top \\ &= V\,V^\top \\ &= I \end{aligned}$$

Uma das maneiras de realizar a decomposição SVD é aplicando a decomposição em autovalores, a qual fornece a matriz V e fornece quase diretamente os valores singulares. Em seguida, fica relativamente fácil obter a matriz U.

APÊNDICE B – Alguns conceitos e definições de geometria algébrica

Esta seção se destina a apresentar um resumo das definições básicas em Geometria Algébrica, bem como alguns conceitos e definições mais avançados necessários ao entendimento do procedimento de solução de sistemas de equações polinomiais. Para acessar um detalhamento maior da teoria juntamente com as demonstrações dos teoremas, o leitor pode consultar os livros Cox et al. (2005) e Cox et al. (2007).

B.1 Definições básicas

A Geometria Algébrica tem como objetivo principal de estudos as variedades algébricas, os objetos geométricos formados pelos pontos que são soluções de um sistema de equações polinomiais. Para realizar esse estudo, utiliza técnicas de álgebra comutativa em conjunto com a linguagem geométrica.

Monômios são o produto de n variáveis da forma $x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot ... \cdot x_n^{\alpha_n}$ onde os expoentes são números inteiros não-negativos. O grau total de um monômio é a soma dos expoentes das variáveis e é denotado por $|\alpha| = \alpha_1 + \alpha_2 + ... + \alpha_n$, e para simplificar a notação é adotada a convenção onde $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$ é o vetor dos expoentes e $x^\alpha = x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot ... \cdot x_n^{\alpha_n}$ é a representação do monômio.

Polinômios são combinações lineares dos monômios denotados por

$$f = \sum_\alpha a_\alpha x^\alpha, \quad \text{onde} \quad a_\alpha \in k.$$

Os coeficientes desses polinômios pertencem a um conjunto numérico denominado *corpo*, denotado por k, que pode ser o conjunto dos números complexos ou reais, por exemplo. Informalmente, corpo é um conjunto fechado para adição, subtração, multiplicação e divisão com suas propriedades usuais. O conjunto de todos os polinômios nas n variáveis definidas num corpo k é denotado por $k[x_1, ..., x_n]$.

Se estivermos trabalhando com pouca variáveis podemos utilizar x, y e z e como exemplo temos

$$f = 2x^3y^2z + \frac{2}{3}y^3z^3 - 3xyz + y^2 \quad \in \quad \mathbb{Q}[x, y, z],$$

onde

- os *coeficientes* são 2, $\frac{2}{3}$, -3 e 1,
- os *termos* são cada monômio junto com o respectivo coeficiente

- o *grau total* do polinômio é o grau do monômio de maior grau; no caso temos dois monômios com grau 6.

A soma e o produto de dois polinômios também são polinômios, mas um polinômio f divide um polinômio g apenas se existe um polinômio h onde $g = f\,h$. Assim, entre polinômios ficam satisfeitos os axiomas de adição e multiplicação em um corpo, mas não a existência de um inverso multiplicativo, pois por exemplo, $f = \frac{1}{x}$ não é um polinômio. Os elementos que obedecem essa estrutura formam um *anel comutativo*, com estudos feitos com base na *álgebra comutativa*, e por isso o conjunto $k[x_1, ..., x_n]$ é chamado *anel de polinômios*.

B.2 Ideal e variedade

Um espaço n-dimensional num corpo k é definido como

$$k^n = \{(a_1, ..., a_n) \,/\, a_1, ..., a_n \in k\},$$

e é possível relacionar polinômios com um espaço n-dimensional observando que um polinômio nos dá uma função

$$f : k^n \to k \quad \text{com} \quad f = \sum_{\alpha} a_\alpha x^\alpha \quad \in \quad k[x_1, ..., x_n].$$

A capacidade em tratar polinômios como funções é o que possibilita ligar álgebra e geometria através das chamadas *variedades*, definida a seguir e denotada por $\mathbf{V}$.

$$\mathbf{V}(f_1, ..., f_s) = \{(a_1, ..., a_n) \in k^n / \, f_i(a_1, ..., a_n) = 0, \, \forall \; 1 \leq i \leq s\}.$$

Portanto, variedades podem ser entendidas como o conjunto de vetores que são solução de um sistema de equações polinomiais, e que fornecem uma representação geométrica para a solução do sistema sempre que possível.

Como exemplo, temos na Figura 42 a representação geométrica das variedades $\mathbf{V}(xy - x^3 + 1)$ e $\mathbf{V}(-x^2 - y + z)$, onde cada ponto no plano e no espaço é representado pelas soluções de $xy - x^3 + 1 = 0$ e $-x^2 - y + z = 0$, respectivamente.

Ideal é um subconjunto de $k[x_1, ..., x_n]$ denotado por I que satisfaz as condições:

- $0 \in I$;
- $f, g \in I \Rightarrow (f + g) \in I$;
- $f \in I$ e $h \in k[x_1, ..., x_n] \Rightarrow h\,f \in I$.

Figura 42 - Variedades.

(a) (b)

Legenda: As representações geométricas podem ser obtidas usando as funções (a) $y = \frac{x^3-1}{x}$ e (b) $z = x^2 + y$.

Fonte: (a) O autor. (b) Cox et al. (2005).

Se $f_1, ..., f_s$ são polinômios em $k[x_1, ..., x_n]$ então podemos definir um ideal através desses polinômios que são chamados de *base* do ideal, usando a notação

$$I = \langle f_1, ..., f_s \rangle = \{\sum_{i=1}^{s} h_i f_i \ : \ h_1, ..., h_s \in k[x_1, ..., x_n]\}.$$

Por exemplo, para sabermos se o polinômio $x^2-2x+2-y$ pertence ao ideal $\langle x-1-t\,,\, y-1-t^2\rangle$ devemos verificar se é possível escrevê-lo como combinação das bases desse ideal,

$$x^2 - 2x + 2 - y = (x - 1 + t)(x - 1 - t) + (-1)(y - 1 - t^2).$$

Um mesmo ideal pode ser formado por bases diferentes e, quando isso acontece, as variedades definidas por cada base também são iguais, ou seja

$$\langle f_1, ..., f_s \rangle = \langle g_1, ..., g_s \rangle \Rightarrow \mathbf{V}(f_1, ..., f_s) = \mathbf{V}(g_1, ..., g_s).$$

Por exemplo, temos a seguir um ideal formado por bases diferentes (podemos verificar que se trata do mesmo ideal escrevendo os polinômios da esquerda como combinação dos polinômios da direita e reciprocamente) e que possuem variedades iguais:

$$\left\langle 2x^2 + 3y^2 - 11\,,\, x^2 - y^2 - 3 \right\rangle = \left\langle x^2 - 4\,,\, y^2 - 1 \right\rangle, \text{ logo}$$

$$\mathbf{V}(2x^2 + 3y^2 - 11\,,\, x^2 - y^2 - 3) = \mathbf{V}(x^2 - 4\,,\, y^2 - 1) = \{(\pm 2, \pm 1)\}. \tag{57}$$

Observe que todos os polinômios de um ideal terão o mesmo conjunto variedade gerado pelos polinômios da base. Observe ainda, que a base de um ideal pode ser pensada como um sistema de equações polinomiais, e que para encontrarmos o conjunto solução do sistema (variedade) podemos converter a base dada numa base mais simples conforme a relação 57. Mais à frente veremos que é exatamente esse o procedimento realizado pelas bases de Gröbner.

B.3 As bases de Gröbner

Para sabermos se um polinômio f pertence a um ideal devemos dividir f pelas bases do ideal e verificar se o resto é zero, pois assim f pode ser escrito como combinação dos polinômios da base. Na divisão de polinômios de uma variável o resto é único, mas para polinômios de mais de uma variável, tanto o resto como os quocientes podem variar dependendo da ordem dos divisores e da ordenação de monômios escolhida (*lexicográfica, lexicográfica graduada e lexicográfica graduada reversa*). Para detalhes sobre divisão de polinômios de várias variáveis consultar Cox et al. (2005).

Por exemplo, vamos verificar se o polinômio $f = xy^2 - x$ pertence ao ideal $I = \langle xy + 1\,,\, y^2 - 1\rangle$. Efetuando a divisão de f por $\{xy + 1\,,\, y^2 - 1\}$ nesta ordem, obtemos

$$xy^2 - x = y(xy + 1) + 0(y^2 - 1) + (-x - y)$$

com resto $-x - y$. Mas efetuando a divisão na ordem $\{y^2 - 1\,,\, xy + 1\}$ teremos

$$xy^2 - x = x(y^2 - 1) + 0(xy + 1) + 0$$

com resto zero. Assim, $f \in I$ e vemos que, dada uma base qualquer de I, o resto zero é uma condição suficiente para que $f \in I$ mas não é uma condição necessária.

Precisamos de uma base que facilite verificar se um polinômio pertence a um ideal, e as bases de Gröbner tem essa propriedade, pois um polinômio f quando dividido pelas bases de Gröbner deixa resto único, não importanto a ordem dos divisores (base do ideal). Assim, sendo $G = \{g_1, ..., g_t\}$ uma base de Gröbner, então

$f \in I \Leftrightarrow$ o resto da divisão de f por G é zero.

Essa propriedade é tão importante que às vezes é tomada como a própria definição para as

bases de Gröbner. O resto da divisão de um polinômio f por $G = \{g_1, ..., g_t\}$ é denotado por $\overline{f}^G$. Por exemplo, dividindo $f = x^5y$ por $G = \{x^2y - y^2\,,\, x^4y^2 - y^2\}$ é $\overline{x^5y}^G = xy^3$, pois

$$x^5y = (x^3 + xy)(x^2y - y^2) + (0)(x^4y^2 - y^2) + xy^3.$$

Para verificarmos se uma base é uma base de Gröbner utilizamos o *critério de Buchberger*, e caso não seja, toda base de um ideal pode ser convertida numa base de Gröbner utilizando o *algoritmo de Buchberger*.

B.4 Resolvendo sistemas com as bases de Gröbner

Considere o sistema

$$\begin{cases} x^2 + y^2 + z^2 = 1 \\ x^2 + z^2 - y = 0 \\ x - z = 0, \end{cases}$$

cujas soluções podem ser obtidas como a extração do conjunto variedade do ideal

$$I = \left\langle x^2 + y^2 + z^2 - 1\,,\, x^2 + z^2 - y\,,\, x - z \right\rangle.$$

Como toda base de um ideal pode ser convertida numa base de Gröbner, aplicando o algoritmo de Buchberger calculamos uma nova base para o ideal

$$I = \left\langle x - z\,,\, -y + 2z^2\,,\, z^4 + \frac{1}{2}z^2 - \frac{1}{4} \right\rangle.$$

Como um ideal com bases diferentes possui o mesmo conjunto variedade, podemos converter as bases de Gröbner num novo sistema que terá o mesmo conjunto solução do anterior,

$$\begin{cases} x - z = 0 \\ -y + 2z^2 = 0 \\ z^4 + \frac{1}{2}z^2 = \frac{1}{4}. \end{cases}$$

Note que a última equação do sistema depende apenas de uma variável, que pode ser obtida através de um método numérico. O *teorema da eliminação* garante que, definida uma ordenação de monômios, as variáveis de um sistema são eliminadas de acordo com a precedência dessa ordem. Depois de calculada a primeira variável, por retro-substituição

podemos determinar o valor das variáveis restantes, onde tal retro-substituição é garantida, sob determinadas condições, pelo *teorema da extensão*. Alguns softwares de computação algébrica (como Maple) possuem pacotes para computação das bases de Gröbner. O procedimento apresentado aqui para solução de sistemas possui algumas deficiências que serão comparadas com o procedimento a seguir no final do apêndice.

B.5 Álgebra de um espaço dimensional finito

Na divisão de um polinômio $f \in k[x_1, ..., x_n]$ por uma base de Gröbner G obtemos a expansão

$$f = h_1 g_1 + ... + h_t g_t + \overline{f}^G,$$

onde $\overline{f}^G$ é uma combinação de monômios que não são divisíveis pelos termos líderes dos polinômios de G. Mas existem outros polinômios em $k[x_1, ..., x_n]$ que deixam o mesmo resto que f na divisão por G. Assim f definde uma *classe de equivalência* dada por

$$[f] = \{p \in k[x_1, ..., x_n] \,/\, p \equiv f \bmod I\}.$$

Ou seja, assim como temos os polinômios que pertencem ao ideal I, temos também o conjunto dos polinômios que não pertencem ao ideal, e cada um desses polinômios pertence à sua classe de equivalência. O conjunto de todas as classes de equivalência de um ideal forma um anel denominado *anel quociente*, denotado por

$$k[x_1, ..., x_n]/I = \{[f] \,/\, f \in k[x_1, ..., x_n]\}.$$

Quando multiplicamos dois restos de um ideal, $\overline{f}^G \cdot \overline{p}^G$, podemos obter um polinômio que pertence ao ideal, mas quando somamos dois restos ou multiplicamos um resto por um escalar, o polinômio gerado continua pertencendo ao anel quociente $k[x_1, ..., x_n]/I$. Por essa e outras características, temos que esse anel tem a estrutura de um *espaço vetorial* e é denominado *Álgebra*. Os monômios que não são divisíveis pelos termos líderes de G formam uma base para essa Álgebra, denominada *base linear*. Considere como exemplo o ideal gerado por

$$G = \{x^2 + \frac{3xy}{2} + \frac{y^2}{2} - \frac{3x}{2} - \frac{3y}{2}\,,\, xy^2 - x\,,\, y^3 - y\}, \tag{58}$$

cujo polinômios têm os termos líderes dados por $\{x^2\,,\, xy^2\,,\, y^3\}$. Extraindo os monômios

que não são divisíveis pelos termos líderes formamos a base linear

$B = \{1\,,\, x\,,\, y\,,\, xy\,,\, y^2\},$

onde os cinco monômios são uma base do espaço vetorial para a álgebra $A = \mathbb{C}[x, y]/I$. Assim, todo polinômio que não pertence a I pode ser escrito como uma *combinação linear* dos monômios de B.

A base linear B tem dimensão finita e sempre terá desde que sejam satisfeitas as condições do *teorema da finitude*, para o qual a principal condição é que sendo G uma base de Gröbner para um ideal $I \subset k[x_1, ..., x_n]$, então para cada $1 \leq i \leq n$, existe um $m_i \geq 0$ tal que $x_i^{m_i}$ é o termo líder de algum $g \in G$. Ou seja, considerando a base G na relação 58, se por exemplo y^{m_y} não pertencesse ao conjunto dos termos líderes para algum m_y, não teríamos um expoente que pudesse restringir y a pertencer à base B, e esta seria infinita.

B.6 Resolvendo sistemas com autovalores e autovetores

A ideia principal é explorar a estrutura de $A = k[x_1, ..., x_n]/I$ para estimar o valor de alguma funcão f nos valores de V. Observe que nossa tarefa é encontrar os valores de $x_i \in (x_1, ..., x_n)$ onde $(x_1, ..., x_n) \in V$, assim podemos tomar a função $f = x_i$, pois estimar o valor para f cumprirá a tarefa de encontrar os valores de cada componente do vetor solução.

Para determinar o valor de cada variável é definido um operador linear m_{x_i} como a multiplicação de cada variável x_i pelos polinômios b da base B.

$m_{x_i} : A \to A \quad \text{onde} \quad m_{x_i}([b]) = x_i \cdot [b].$

Desta forma, os possíveis valores para a variável x_i são computados através da decomposição em autovalores da matriz m_{x_i}. Por definição em Álgebra Linear, para encontrarmos a matriz m_{x_i}, devemos expandir $p_j = x_i \cdot b_j$ em termos da base B, e tomar a transposta da matriz dos coeficientes dessa expansão.

Contudo, multiplicando dois elementos de A, não podemos garantir que o resultado p_j seja um elemento de A que possa ser escrito como combinação linear dos elementos de B. Assim, devemos escrever a expansão de p_j em termos dos elementos da base G para obter o resto $\overline{p}^G$. Como todo polinômio

$$p \equiv \overline{p}^G \bmod I, \tag{59}$$

podemos expandir $\overline{p}^G$ ao invés de p.

Como exemplo, vamos calcular as soluções do sistema formado pelos polinômios da base G em 58,

$$\begin{cases} x^2 + \dfrac{3xy}{2} + \dfrac{y^2}{2} - \dfrac{3x}{2} - \dfrac{3y}{2} = 0 \\ xy^2 - x = 0 \\ y^3 - y = 0, \end{cases}$$

com termos lideres $\{x^2\,,\, xy^2\,,\, y^3\}$ e base $B = \{1\,,\, x\,,\, y\,,\, xy\,,\, y^2\}$. Determinando primeiramente os valores para x temos que $m_x(b_j) = x \cdot b_j$, portanto

$$\begin{aligned} x \cdot 1 &= x &&= 0 \cdot 1 + 1 \cdot x + 0 \cdot y + 0 \cdot xy + 0 \cdot y^2 \\ x \cdot x &= x^2 &&= 0 \cdot 1 + \tfrac{3}{2} \cdot x + \tfrac{3}{2} \cdot y - \tfrac{3}{2} \cdot xy - \tfrac{1}{2} \cdot y^2 \\ x \cdot y &= xy &&= 0 \cdot 1 + 0 \cdot x + 0 \cdot y + 1 \cdot xy + 0 \cdot y^2 \\ x \cdot xy &= x^2y &&= 0 \cdot 1 - \tfrac{3}{2} \cdot x - \tfrac{1}{2} \cdot y + \tfrac{3}{2} \cdot xy + \tfrac{3}{2} \cdot y^2 \\ x \cdot y^2 &= xy^2 &&= 0 \cdot 1 + 1 \cdot x + 0 \cdot y + 0 \cdot xy + 0 \cdot y^2. \end{aligned}$$

Tomando a transposta dos coeficientes temos a denominada *matriz de ação*,

$$m_x = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{3}{2} & 0 & -\frac{3}{2} & 1 \\ 0 & \frac{3}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & -\frac{3}{2} & 1 & \frac{3}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{3}{2} & 0 \end{bmatrix},$$

e os valores para x são dados pelos autovalores $\{-1, 2, 1, 1, 0\}$, que podem ser calculados usando o método em **??**. De forma análoga, podemos determinar os valores de y através dos autovalores $\{1, -1, 1, -1, 0\}$ de m_y. As soluções do sistema é o conjunto variedade do ideal I que tem como base G,

$$V(I) = \{(-1, 1), (2, -1), (1, 1), (1, -1), (0, 0)\}.$$

Podemos ver que na expansão de $x \cdot b_j$ acima, os monômios $\{x^2\,,\, x^2y\,,\, xy^2\}$ não

podem ser escritos como combinação linear da base B. Por isso, usando a propriedade em 59, fizemos a expansão do resto da divisão de cada um desses monômios pela base G,

$$\begin{aligned}
\overline{x^2}^G &= -\tfrac{3xy}{2} - \tfrac{y^2}{2} + \tfrac{3x}{2} + \tfrac{3y}{2} \\
\overline{x^2y}^G &= \tfrac{3xy}{2} + \tfrac{3y^2}{2} - \tfrac{3x}{2} - \tfrac{y}{2} \\
\overline{xy^2}^G &= x.
\end{aligned}$$

É possível determinar o conjunto variedade usando a retro-substituição ou a matriz de ação. Vejamos as principais diferenças:

- O método da eliminação impõe o uso da ordenação lexicográfica, a qual requer muita computação e, mesmo usando um algoritmo para mudança de ordenação, não é eficiente. A estrutura algébrica de $A = \mathbb{C}[x_1, ..., x_n]/I$ é independente de quaisquer das três ordenações de monômios escolhida, assim qualquer tipo de ordenação pode ser utilizada no cálculo dos operadores m_{x_i}.

- Computação numérica versus computação simbólica, e instabilidade numérica. As entradas das matrizes m_{x_i} sempre são números racionais e podem ser determinadas, exatamente, por computação simbólica. Assim, a computação numérica fica restrita somente ao cálculo dos autovalores.

- Podemos computar as matrizes m_{x_i} e seus autovalores separadamente e determinar os vetores $(x_1, ..., x_n)$ que dão as soluções aproximadas. Ou seja, o cálculo da variável x_j não depende do cálculo da variável x_i como na eliminação. Isso resolve um dos maiores problemas na retro-substituição, o problema dos *erros acumulados*.

- É possível reduzir o esforço computacional mais ainda calculando os autovalores de uma única matriz de ação $m_{c_1x_1+...+c_nx_n}$.