
ESTIMAÇÃO ROBUSTA DA MATRIZ FUNDAMENTAL

José A. de França*
josealexandre@eeol.org

Marcelo R. Stemmer†
marcelo@das.ufsc.br

*Departamento de Engenharia Elétrica
Universidade Estadual de Londrina
Caixa Postal 6025
86051-990, Londrina, PR, Brasil

†Departamento de Automação e Sistemas
Universidade Federal de Santa Catarina
Caixa Postal 476
88040-970, Florianópolis, SC, Brasil

ABSTRACT

In some applications of stereo vision, points correspondence between two cameras images is necessary. If the fundamental matrix is known, the complexity of this task is reduced. However, is necessary a initial correspondence set to estimate that matrix. Robust algorithms to detect and eliminate the noise and outliers in that initial set can be used. This way, they increase significantly the precision of stereo vision system. In this work, we analyzed some robust methods based on LMedS to estimate the fundamental matrix. The purpose is to use the fundamental matrix to accomplish points correspondence. The analyzed methods are compared in terms of speed, computational complexity and accuracy.

KEYWORDS: Epipolar Geometry, Camera Calibration, Robust Techniques, Fundamental Matrix, Stereo Vision.

RESUMO

Em algumas aplicações de visão estereoscópica, o emparelhamento de pontos entre as imagens das duas câmeras é necessário. Se a matriz fundamental é conhecida, esta tarefa pode ser realizada de forma menos complexa. Contudo, é necessário um conjunto inicial de emparelhamentos de pontos para que esta matriz seja calculada. Algoritmos robustos capazes de detectar e eliminar ruídos e falsos emparelhamentos existentes neste conjunto inicial podem ser utilizados. Estes algoritmos aumentam significativamente a exatidão do sistema de visão estereoscópica. Este trabalho analisa alguns métodos robustos baseados no LMedS para o cálculo da matriz fundamental. O objetivo é utilizar a matriz fundamental para realizar emparelhamento de pontos. Os métodos analisados são comparados em termos de velocidade, complexidade computacional e exatidão.

PALAVRAS-CHAVE: Geometria Epipolar, Calibração de Câmeras, Técnicas Robustas, Matriz Fundamental, Visão Estereoscópica.

1 INTRODUÇÃO

Em algumas aplicações da robótica móvel, por exemplo, em sistemas multiagentes como a *RoboCup* (Kitano et al., 1997), inicialmente é necessário determinar-se, com base em um sistema de coordenadas qualquer, a posição do robô, do seu “alvo” (local para o qual ele deve mover-se) e, em alguns casos, dos obstáculos inseridos no ambiente. Em seguida, ações devem ser realizadas para que o veículo (robô) chegue ao seu destino. Geralmente, o ambiente de um veículo autônomo é desconhecido e/ou dinâmico. Por isso, para que ele possa atingir a sua meta é necessário que o mesmo seja dotado de sensores robustos capazes de fornecer as informações de posição necessárias.

Diversos trabalhos na área de veículos autônomos são baseados em sensores, tais como sonares e lasers, que fornecem informações 3D diretamente (Shoval et al., 1994; Dudek et al., 1996). Em outros casos, marcas guias (*landmarks*) e sistemas de visão monoculares são combinados com um conhecimento prévio do ambiente para localizar os veículos e guiá-los em seu trajeto (Lorigo et al., 1997). Em oposição a projetos desse tipo, alguns pesquisadores (Burschka et al., 2002) propõem sistemas que obtêm informações apenas de um par estereoscópico, sem o uso de lasers, sensores de ultra-som e com (relativamente) poucas suposições sobre o ambiente do robô. Isso tem a vantagem de aumentar a autonomia dos veículos, pois um par estereoscópico consome bem menos energia que, por exemplo, sistemas à lasers. Além disso, como se sabe, sonares possuem uma baixa densidade de dados (Burschka et al., 2002).

Em algumas aplicações de visão estereoscópica, por exemplo,

reconstrução 3D (Debevec et al., 1996), calibração de câmeras (Luong and Faugeras, 1997) e algumas técnicas de detecção de obstáculos (Burschka and Hager, 2002), o emparelhamento de pontos entre as imagens das duas câmeras é necessário. O problema consiste em, dado um determinado ponto na imagem de uma das câmeras, determinar o ponto correspondente na imagem da outra câmera que compõe o par. Esta tarefa é extremamente complexa e dela depende a exatidão de qualquer sistema baseado em visão estereoscópica.

Se a matriz fundamental é conhecida, o emparelhamento de pontos podem ser realizado de forma menos complexa. Por exemplo, através da retificação das duas imagens (Loop and Zhang, 1999). Contudo, é necessário um conjunto inicial de emparelhamentos para que a matriz fundamental seja estimada. Geralmente, este conjunto inicial possui ruído e falsos emparelhamentos, pois o sistema não está calibrado e a única restrição geométrica, isto é, a matriz fundamental, ainda não está disponível. Assim, os emparelhamentos iniciais são estabelecidos utilizando-se apenas heurística. Ruído e falsos emparelhamentos prejudicam a estimação da matriz fundamental e degradam a exatidão de todas as medições do sistema de visão. Em tais situações, algoritmos robustos para a estimação da matriz fundamental devem ser utilizados. Estes algoritmos são capazes de detectar e eliminar os ruídos e falsos emparelhamentos existentes no conjunto de emparelhamentos utilizado no cálculo da matriz fundamental. Dessa forma, os algoritmos robustos aumentam significativamente a exatidão do sistema de visão estereoscópica como um todo.

Este trabalho analisa alguns métodos robustos baseados no LMedS para o cálculo da matriz fundamental. Os métodos analisados são comparados em termos de velocidade, complexidade computacional e exatidão. Um *script* do Matlab que implementa os métodos abordados neste artigo, assim como, mais resultados experimentais estão disponíveis na internet na url <http://research.eeol.org/papers/sbai/2003/>.

Na seção 2 é apresentada a notação utilizada neste artigo. Em seguida, na seção 3 discute-se os métodos utilizados para calcular a matriz fundamental. Os resultados experimentais são apresentados na seção 4 e as conclusões na seção 5.

2 NOTAÇÃO

Matrizes são representadas por letras, números ou símbolos em negrito e constantes por letras, números ou símbolos em itálico. Assim, considerando o modelo de câmera *pinhole*, as coordenadas de um ponto 3D no sistema de coordenadas do ambiente é apresentado como $\mathbf{M} = [x, y, z]^T$ e o ponto correspondente na retina da imagem como $\mathbf{m} = [u, v]^T$. As coordenadas homogêneas de um ponto $\mathbf{x} = [x, y, \dots]^T$ são representadas por $\tilde{\mathbf{x}}$, isto é, $\tilde{\mathbf{x}} = [x, y, \dots, 1]^T$. Finalmente, usamos a prática notação $A^{-T} = (A^{-1})^T = (A^T)^{-1}$ para toda matriz quadrada inversível.

3 ESTIMAÇÃO DA MATRIZ F

Em um sistema de visão estereoscópica sem calibração, a única informação disponível é a geometria epipolar (Zhang et al., 1995). Essa geometria já foi descrita em numerosos artigos, como por exemplo, (Zhang, 1998) e (Torr and Murray, 1997).

Basicamente, dado um ponto \mathbf{M} no espaço 3D, a geometria epipolar relaciona as projeções \mathbf{m} e \mathbf{m}' desse ponto, respectivamente, na primeira e segunda imagens, pela equação

$$\tilde{\mathbf{m}}'^T \mathbf{F} \tilde{\mathbf{m}} = 0, \quad (1)$$

onde \mathbf{F} é uma matriz 3x3 chamada matriz fundamental.

A matriz fundamental deve ser estimada a partir de um conjunto de emparelhamentos de pontos $(\mathbf{m}_i, \mathbf{m}'_i)$. Desde que \mathbf{F} tem posto 2 e é definida apenas por um fator de escala, são necessárias, no mínimo, sete emparelhamentos para a sua estimação. Contudo, na prática, dispõe-se de mais emparelhamentos e pode-se utilizar essas pares extras para aumentar a exatidão dos resultados.

A seguir, algumas das principais técnicas para estimação da matriz fundamental são discutidas.

3.1 Método dos 8 pontos

A matriz fundamental deve ser estimada a partir de um conjunto de emparelhamentos $(\mathbf{m}_i, \mathbf{m}'_i)$. Para isto, considere um ponto $\mathbf{m}_i = [u_i, v_i]^T$ e seu o correspondente na outra imagem $\mathbf{m}'_i = [u'_i, v'_i]^T$. Neste caso, eles devem satisfazer a equação (1). Esta equação pode ser re-escrita como uma equação linear e homogênea em termos dos nove coeficientes da matriz \mathbf{F} :

$$\mathbf{u}_i^T \mathbf{f} = 0, \quad (2)$$

na qual $\mathbf{u}_i = [u_i u'_i, v_i u'_i, u'_i, u_i v'_i, v_i v'_i, v'_i, u_i, v_i, 1]^T$ e $\mathbf{f} = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T$. F_{ij} é o elemento de \mathbf{F} da linha i e coluna j .

Com n emparelhamentos, e considerando a equação (2), tem-se o seguinte sistemas de equações lineares

$$\mathbf{U}_n \mathbf{f} = 0, \quad (3)$$

com $\mathbf{U}_n = [\mathbf{u}_1, \dots, \mathbf{u}_n]^T$. Assim, o método dos quadrados-mínimos pode ser utilizado para resolver

$$\min_{\mathbf{f}} \sum_i \|\mathbf{U}_n \mathbf{f}\|^2, \quad \text{sujeito a } \|\mathbf{f}\| = 1. \quad (4)$$

Apesar de poder ser utilizado mais de oito pontos para resolver a equação (4), este método geralmente é referenciado como método dos oito pontos.

Em 1995, Hartley (Hartley, 1995) sugeriu uma mudança no sistema de coordenadas dos pares $(\mathbf{m}_i, \mathbf{m}'_i)$ para aumentar a exatidão da estimação de \mathbf{F} . No novo sistema de coordenadas, os pontos \mathbf{m}_i e \mathbf{m}'_i devem estar a uma distância média da origem igual a $\sqrt{2}$. Além disso, o centróide do conjunto de pontos deve estar na origem do novo sistema de coordenadas. Hartley provou que, com essas alterações, para um conjunto de emparelhamentos muito maior que oito, o método baseado nos quadrados-mínimos fornece uma estimativa bastante exata de \mathbf{F} . Contudo, a estimativa torna-se pobre à medida que o número de emparelhamentos utilizadas nos cálculos aproxima-se de oito.

3.2 Método Não-Linear

Um dos motivos para o método dos oito pontos não funcionar para um conjunto pequeno de pontos é o negligenciamento da restrição do posto da matriz \mathbf{F} . Para levar em

consideração o posto de \mathbf{F} , pode-se utilizar parametrização. Uma parametrização possível foi proposta por Luong (Luong et al., 1993) em 1993. Nesta, \mathbf{F} é expressa em termos de quatro constantes a, b, c e d , e das coordenadas dos epipolos $e = (x, y)$ e $e' = (x', y')$, respectivamente, da primeira e segunda imagem:

$$\mathbf{F} = \begin{bmatrix} a & b & -ax - by \\ c & d & -cx - dy \\ -ax' - cy' & -by' - dy' & F_{33} \end{bmatrix}, \quad (5)$$

onde $F_{33} = (ax + by)x' + (cx + dy)y'$.

Além disso, como \mathbf{F} é definida por um fator de escala, a, b, c e d são divididos pelo maior em valor absoluto.

Apesar de apresentar resultados mais exatos, esta parametrização não funciona bem se, pelo menos um, dos epipolos estiver no infinito. Isto ocorre porque, neste caso, ao menos um dentre x, y, x' e y' tem um valor infinito.

Para contornar este problema, Zhang e Loop (Zhang and Loop, 2001) propuseram um novo método, no qual uma transformação projetiva é aplicada aos pontos \mathbf{m}_i e \mathbf{m}'_i , ou seja, $\tilde{\mathbf{m}}_i = \mathbf{T}\mathbf{m}_i$ e $\tilde{\mathbf{m}}'_i = \mathbf{T}'\mathbf{m}'_i$. As matrizes \mathbf{T} e \mathbf{T}' são tais que, no novo sistema de coordenadas, o primeiro elemento da matriz fundamental do novo sistema de coordenadas $\tilde{\mathbf{F}}$ é igual a 1 e os epipolos não estão no infinito. Contudo, todo critério de otimização não-linear é dependente da distribuição do ruído dos pontos utilizados na minimização (Zhang, 1998). Normalmente, considera-se que o ruído tem uma distribuição gaussiana com média zero e matriz de covariância dada por

$$\Lambda_{m_i} = \Lambda_{m'_i} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

onde σ é o, normalmente desconhecido, nível de ruído. Evidentemente, após a transformação proposta em (Zhang and Loop, 2001), essa consideração não é mais válida. Por isso, Zhang e Loop demonstraram como estimar \mathbf{F} conservando a distribuição original do ruído através de uma minimização ponderada por gradiente, ou seja, como descrito em (Zhang and Loop, 2001),

$$\min_{\tilde{\mathbf{F}}} \sum_i \frac{(\tilde{\mathbf{m}}_i^T \tilde{\mathbf{F}} \tilde{\mathbf{m}}_i)^2}{\hat{\mathbf{l}}_i^T \mathbf{Z} \mathbf{T} \mathbf{Z}^T \hat{\mathbf{l}}_i + \hat{\mathbf{l}}_i'^T \mathbf{T}' \mathbf{Z}' \mathbf{T}'^T \hat{\mathbf{l}}_i'}, \quad (6)$$

onde $\tilde{\mathbf{F}} = \mathbf{T}'\mathbf{F}\mathbf{T}$, $\hat{\mathbf{l}}_i = \tilde{\mathbf{F}}\tilde{\mathbf{m}}_i$, $\hat{\mathbf{l}}_i' = \tilde{\mathbf{F}}^T\tilde{\mathbf{m}}_i'$ e $\mathbf{Z} = \text{diag}(1, 1, 0)$.

Agora, com a parametrização da equação (5) e o critério de minimização da equação (6), têm-se um método bastante exato para calcular a matriz fundamental. Contudo, este método requer uma minimização não-linear que, geralmente, necessita de uma estimativa inicial da matriz \mathbf{F} . Esta estimativa inicial pode ser conseguida, por exemplo, pela equação (4).

Além de ser computacionalmente muito complexo, este método ainda tem a limitação de ser muito dependente da estimativa inicial de \mathbf{F} . Por isso, geralmente o algoritmo converge para um mínimo local da equação (6). Além disso, para um conjunto de pontos muito maior que oito, o método não-linear e o método da equação (4) têm resultados muito próximos (Hartley, 1995).

3.3 Método do Paralaxe Virtual

Recentemente, Boufama (Boufama and Mohr, 1998) propôs um novo método para calcular a matriz fundamental. O método é

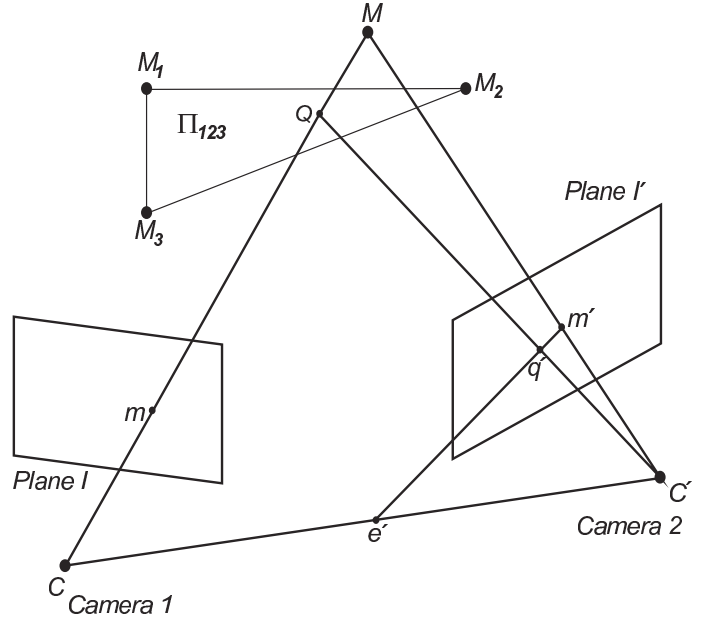


Figura 1: Geometria epipolar e o paralaxe virtual.

baseado na computação da homografia entre duas imagens de um plano virtual e na detecção do paralaxe de pontos não pertencentes a este plano. Este método é descrito à seguir.

Três pontos não-colineares, M_1, M_2 , e M_3 , em \mathcal{P}^3 definem um plano Π_{123} (figura 1). Se esse plano é observado por duas câmeras denominadas 1 e 2, existe uma projeção de Π_{123} em I e I' , respectivamente, o plano de imagem da câmera 1 e da câmera 2. Então, pode-se definir duas homografias, H e H' , correspondentes a projeção do plano Π_{123} nos planos de imagem I e I' respectivamente, ou seja, $I = H(\Pi_{123})$ e $I' = H'(\Pi_{123})$. Desse modo, existe uma homografia $K = H'H^{-1}$, onde H^{-1} é a homografia inversa de H . Logo,

$$\forall \mathbf{m} \in I, K(\mathbf{m}) = \mathbf{q}', \text{ com } \mathbf{q}' \in I'.$$

Se considerarmos a homografia K descrita por uma matriz \mathbf{H}_p 3×3 , tem-se

$$\forall \mathbf{m} \in I, \mathbf{H}_p \tilde{\mathbf{m}} = \lambda \tilde{\mathbf{q}}', \quad (7)$$

onde λ é um fator de escala desconhecido.

Desde que a matriz \mathbf{H}_p é homogênia, são necessárias oito equações independentes para defini-la. Contudo, para cada par de correspondência $(\tilde{\mathbf{m}}_i, \tilde{\mathbf{m}}'_i)$, $i = 1, 2$ e 3 , igual as projeções nos dois planos de imagens dos pontos P_1, P_2 e P_3 , a equação (7) fornece duas equações independentes. Por isso, sem uma outra restrição, a matriz \mathbf{H}_p não pode ser calculada. Contudo, $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2, \tilde{\mathbf{m}}_3$ e $\tilde{\mathbf{m}}_0 = \tilde{\mathbf{m}}_1 + \tilde{\mathbf{m}}_2 + \tilde{\mathbf{m}}_3$ definem um novo sistema de coordenadas no plano de imagem I . De forma semelhante, $\tilde{\mathbf{m}}'_1, \tilde{\mathbf{m}}'_2, \tilde{\mathbf{m}}'_3$ e $\tilde{\mathbf{m}}'_0 = \tilde{\mathbf{m}}'_1 + \tilde{\mathbf{m}}'_2 + \tilde{\mathbf{m}}'_3$ definem um sistema de coordenadas em I' . Agora, nos sistemas de coordenadas definidos, os pares $(\tilde{\mathbf{m}}_i, \tilde{\mathbf{m}}'_i)$, $i = 1, 2$ e 3 , são iguais a

$$\begin{aligned} \tilde{\mathbf{m}}_1 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T, & \tilde{\mathbf{m}}'_1 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T, \\ \tilde{\mathbf{m}}_2 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T, & \tilde{\mathbf{m}}'_2 &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T, \\ \tilde{\mathbf{m}}_3 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T, & \tilde{\mathbf{m}}'_3 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T, \\ \tilde{\mathbf{m}}_0 &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T \text{ e } & \tilde{\mathbf{m}}'_0 &= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T. \end{aligned} \quad (8)$$

Neste novo sistema de coordenadas, \mathbf{H}_p tem a forma

$$\mathbf{H}_p = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Logo, considerando a equação (7), a figura 1, $\tilde{\mathbf{m}} = [u, v, t]^T$ e $\tilde{\mathbf{m}}' = [u', v', t']^T$, tem-se

$$\mathbf{H}_p \tilde{\mathbf{m}} = [\alpha u, \beta v, t]^T = \lambda \tilde{\mathbf{q}}', \quad (9)$$

com λ , α e β desconhecidos.

Agora, como ilustrado na figura 1, considere um ponto M no espaço 3D e faça o ponto virtual Q ser a interseção do plano Π_{123} com a linha \overline{CM} . Se o epipolo e' na segunda imagem é conhecido, então, um ponto m na primeira imagem possui uma linha epipolar na segunda imagem dada por $\overline{e'm'}$, onde m' é a projeção do ponto M na segunda imagem. Agora, desde que, segundo a figura 1, q' pertence a linha epipolar $\overline{e'm'}$, podemos escrever

$$(\tilde{\mathbf{e}}' \times \tilde{\mathbf{m}}') \cdot \tilde{\mathbf{q}}' = 0, \quad (10)$$

onde “ \times ” é o produto cruzado e “ \cdot ” é o produto escalar.

Substituindo-se (9) em (10) e considerando $\tilde{\mathbf{e}}' = [e'_u, e'_v, e'_t]^T$, após um tratamento algébrico, tem-se

$$\mathbf{a}_i^T \mathbf{V} = 0, \quad (11)$$

na qual $\mathbf{a}_i = [u_i v'_i, -u_i t'_i, v_i t_i, -v_i u'_i, t_i u'_i, -t_i v'_i]^T$ e $\mathbf{V} = [V_1, V_2, V_3, V_4, V_5, V_6]^T$, com $V_1 = \alpha e'_t$, $V_2 = \alpha e'_t$, $V_3 = \beta e'_u$, $V_4 = \beta e'_t$, $V_5 = e'_v$ e $V_6 = e'_u$.

Com n correspondências de pontos, e considerando a equação (11), tem-se o seguinte sistemas de equações lineares

$$\mathbf{A}_n \mathbf{V} = 0, \quad (12)$$

com $\mathbf{A}_n = [\mathbf{a}_1, \dots, \mathbf{a}_n]^T$.

Resolvendo-se a equação (12), β , α e $\tilde{\mathbf{e}}'$ podem ser encontrados observando-se que (Boufama and Mohr, 1998)

- $e'_u = V_6$ e $e'_v = V_5$;
- Se $V_4 = 0$, então $e'_t = 0$, $\alpha = V_2/e'_v$ e $\beta = V_3/e'_u$;
- Se $V_4 \neq 0$, então, se $e'_u \neq 0$, $\beta = V_3/e'_u$, $e'_t = V_4/\beta$ e $\alpha = V_1/e'_t$, caso contrário $\alpha = V_2/e'_v$, $e'_t = V_1/\alpha$ e $\beta = V_4/e'_t$.

A equação (12) é linear e possui apenas cinco parâmetros independentes. Considerando os três pontos M_1 , M_2 e M_3 , utilizados para gerar o novo sistema de coordenadas (figura 1), é necessário um total de oito correspondências de pontos para resolver a equação (12) e encontrar α , β e o epipolo e' . Assim, a matriz fundamental pode ser calculada, pois esta está relacionada com a homografia \mathbf{H}_p e o epipolo e' pela relação (Luong and Faugeras, 1996)

$$\begin{aligned} \hat{\mathbf{F}} &= [\mathbf{e}']_{\times} \mathbf{H}_p \\ &= \begin{bmatrix} 0 & -e'_t & e'_y \\ e'_t & 0 & -e'_x \\ -e'_y & -e'_x & 0 \end{bmatrix} \begin{bmatrix} \beta & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\beta e'_t & e'_y \\ \alpha e'_t & 0 & -e'_x \\ -\alpha e'_y & -\beta e'_x & 0 \end{bmatrix}, \quad (13) \end{aligned}$$

onde $\hat{\mathbf{F}}$ é a matriz fundamental para o sistema de coordenadas definido pelas bases $B = \{\tilde{m}_1, \dots, \tilde{m}_3, \tilde{m}_0\}$ e $B' = \{\tilde{m}'_1, \dots, \tilde{m}'_3, \tilde{m}'_0\}$.

Em nossos experimentos, observamos que, quando o número de emparelhamentos utilizados nos cálculos é próximo de oito e os pontos M_1 , M_2 e M_3 estão bem distribuídos pela imagem, este método apresenta resultados muito bons. Contudo, ele fornece um resultado bastante inexato quando o número de emparelhamentos é muito maior que oito. Isso acontece porque depositamos extrema confiança nos emparelhamentos $(\mathbf{m}_1, \mathbf{m}'_1)$, $(\mathbf{m}_2, \mathbf{m}'_2)$ e $(\mathbf{m}_3, \mathbf{m}'_3)$, utilizados na formação do novo sistema de coordenadas. De fato, a exatidão do algoritmo está baseada na precisão destes três pares iniciais e qualquer erro em suas coordenadas é transferido para as coordenadas dos outros pares $(\mathbf{m}_i, \mathbf{m}'_i)$. Contudo, o fato deste método ser computacionalmente simples e, para oito emparelhamentos, fornecer resultados exatos, torna-o indicado no uso de algoritmos robustos para o cálculo da matriz fundamental.

3.4 Métodos Robustos

Geralmente, os emparelhamentos utilizados na estimação de \mathbf{F} possuem ruídos e falsos emparelhamentos. Ruídos são devidos à grandes erros na localização (maiores que três pixels) dos pontos de interesse que compõem os emparelhamentos (m_i, m'_i) . Já os falsos emparelhamentos ocorrem porque, em imagens sem calibração, a única restrição geométrica, isto é, a geometria epipolar, ainda não está disponível e os emparelhamentos são estabelecidos utilizando-se apenas heurística.

Ruídos e as falsos emparelhamentos afetam significativamente o cálculo da matriz fundamental se aplicarmos diretamente os métodos discutidos. Em vista disso, diversos autores já propuseram métodos robustos que identificam e eliminam os emparelhamentos ruins, ou seja, aqueles que possuem ruído ou são falsos emparelhamentos. Exemplos desses métodos podem ser encontrados em (Olsen, 1992), (Shapiro and Brady, 1994) e (Torr, 1995).

O método LMedS é um dos métodos robustos mais utilizados no cálculo da matriz fundamental. A seguir, listamos os passos que compõem esse algoritmo.

1. Dado N emparelhamentos, uma técnica Monte Carlo é utilizada para escolher m subconjuntos de n diferentes pares (m_i, m'_i) .
2. Para cada subconjunto, indexado por j , usamos uma técnica que forneça uma boa estimação de \mathbf{F}_j com n pares (m_i, m'_i) .
3. Para cada \mathbf{F}_j , determinamos a média do resíduo ao quadrado, denotado por M_j , com respeito ao conjunto completo de emparelhamentos, isto é,

$$M_j = \frac{1}{N} \sum_{i=1 \dots N} [d^2(\tilde{\mathbf{m}}'_i, \mathbf{F}_j \tilde{\mathbf{m}}_i) + d^2(\tilde{\mathbf{m}}_i, \mathbf{F}_j^T \tilde{\mathbf{m}}'_i)],$$

na qual $d(\tilde{\mathbf{m}}'_i, \mathbf{F}_j \tilde{\mathbf{m}}_i)$ é a distância do ponto \mathbf{m}'_i à linha $\mathbf{F}_j \tilde{\mathbf{m}}_i$ e $d(\tilde{\mathbf{m}}_i, \mathbf{F}_j^T \tilde{\mathbf{m}}'_i)$ é a distância do ponto \mathbf{m}_i à linha $\mathbf{F}_j^T \tilde{\mathbf{m}}'_i$.

4. Retemos a estimação \mathbf{F}_j que tem o menor M_j (denotado por M_{\min}) dentre todos os m M_j 's.

5. Eliminamos os pares (m_i, m'_i) que possuírem o resíduo

$$r_i^2 = [d^2(\tilde{m}'_i, \mathbf{F}_j \tilde{m}_i) + d^2(\tilde{m}_i, \mathbf{F}_j^T \tilde{m}'_i)] \geq r_{th}^2.$$

6. Refinamos a matriz fundamental estimando-a sem os emparelhamentos ruins.

Três comentários podem ser feitos a respeito do algoritmo descrito.

- Em nossa implementação, $r_{th} = 1,99\hat{\sigma}$, onde $\hat{\sigma}$ é a estimativa do desvio padrão dada por $\hat{\sigma} = 1,4826 [1 + 5/(N - n)] \sqrt{M_{\min}}$ (creditado a Rousseeuw (Rousseeuw, 1987)).
- No passo 1, os emparelhamentos que compõem cada subconjunto devem estar bem espalhados pela área da imagem. Caso contrário, a matriz \mathbf{F}_j não se ajustará bem aos outros emparelhamentos. Por isso, neste trabalho foi utilizado a técnica de seleção de subconjuntos descrita em (Zhang et al., 1995), na Seção 6.3.
- Assumindo que o conjunto completo de emparelhamentos possa ter no máximo uma fração ε de ruído, a probabilidade, P , de pelo menos um subconjunto dos m subconjuntos seja livre de ruídos é dada por $P = 1 - [1 - (1 - \varepsilon)^n]^m$, ou seja,

$$m = \frac{\log(1 - P)}{\log[1 - (1 - \varepsilon)^n]}. \quad (14)$$

O número de pontos n que compõem cada subconjunto deve ser o menor possível. Caso contrário, pela equação (14), será necessário um número grande de subconjuntos. Por isso, é costume estimar-se \mathbf{F}_j pela equação (4), seguido de uma minimização não-linear da equação (6). Essa prática fornece resultados exatos à um custo computacional elevado. Neste trabalho, já que o método do paralaxe virtual fornece uma ótima estimativa da matriz fundamental para $n = 8$, examinamos também a sua utilização na estimação de \mathbf{F}_j .

Na próxima seção, os resultados do método LMedS são comparados quando são utilizadas as técnicas já discutidas.

4 RESULTADOS EXPERIMENTAIS

O algoritmo LMedS para o cálculo da matriz fundamental foi testado em diversos pares de imagens. Na Figura 2, um exemplo das imagens utilizadas nos experimentos é apresentado. O resultado do cálculo da matriz fundamental para as imagens da Figura 2 é apresentado na Tabela 1.

A matriz \mathbf{F}_j (seção 3.4) foi estimada através de quatro métodos: 8 pontos (segunda linha da Tabela 1), paralaxe virtual (primeira linha da Tabela 1), 8 pontos seguido de uma minimização não-linear (terceira linha da Tabela 1) e paralaxe virtual seguido por uma estimação não-linear (última linha da Tabela 1).

A metodologia utilizada nos experimentos foi a seguinte.

1. Emparelhamentos foram computados utilizando-se uma técnica automática desenvolvida pelos próprios autores e disponível na internet na url



Figura 2: Um par de imagens com 150 emparelhamentos marcados.

<http://research.eol.org/visao/match/>. Os emparelhamentos ruins não foram retirados, pois o algoritmo LMedS deve ser capaz de identificá-los e eliminá-los.

2. A matriz fundamental foi estimada utilizando-se o algoritmo LMedS com todos os métodos discutidos: virtual paralaxe, algoritmo de oito pontos e o método não-linear. Nesta estimação, apenas metade do conjunto de emparelhamentos foi utilizado. Os subconjuntos de emparelhamentos utilizados no cálculo foram os mesmos para todos os métodos.
3. Os emparelhamentos ruins foram retirados e, para cada ponto m_i , a linha epipolar correspondente $\mathbf{F} \tilde{m}_i$ e a distância do ponto correspondente m'_i à esta linha foram calculadas. A distância média dos pontos à sua linha epipolar foi utilizada como medida da qualidade da matriz fundamental calculada.

Essa metodologia de testes foi aplicada em diferentes pares de imagens. Contudo, todos os pares forneceram resultados (proporcionais entre os quatro métodos) muito semelhantes aos da Tabela 1.

Em nossos testes, observou-se que o método do paralaxe virtual torna-se bem mais exato quando os emparelhamentos utilizados

Método	Tempo	Dist. Média
Paralaxe	1,00	1,3529
8 Pontos	1,07	0,8676
N-Linear 1	33,98	0,7220
N-Linear 2	30,72	0,5705

Tabela 1: Comparação dos métodos analisados. O tempo é relativo ao algoritmo mais rápido e a distância média é dada em pixels.

na estimação estão bem espalhados pela imagem. Neste caso, a sua exatidão fica próxima dos métodos não-lineares testados. Neste método, a exatidão é mais dependente da localização dos emparelhamentos do que nos outros.

O método robusto baseado no algoritmo dos 8 pontos é rápido e fornece bons resultados. Pode ser utilizado em sistemas com pouco poder de processamento, pois é bastante simples de ser implementado. Além disso, caso o número de emparelhamentos obtidos robustamente for muito maior que 8, o passo 6 do algoritmo da seção 3.4 tem ótimos resultados se este for estimado com o algoritmo dos 8 pontos (sem o estágio de minimização não-linear).

O algoritmo não-linear fornece resultados bastante exatos. Em relação ao método dos 8 pontos e do paralaxe virtual, em alguns casos, com um mesmo subconjunto j de emparelhamentos (algoritmo da seção 3.4) ele fornece uma matriz fundamental com uma distância média de até 10 vezes menor. Contudo, o seu tempo de computação é, em média, quase 34 vezes maior que, por exemplo, o método dos 8 pontos. Além disso, desde que o método não-linear fornece apenas um mínimo local em torno do ponto inicial, é de se esperar que o subconjunto j que gere a matriz F_j com a menor resíduo é o mesmo para, por exemplo, o método dos 8 pontos e o método não-linear. Por isso, pode-se utilizar o algoritmo de 8 pontos ou o do paralaxe virtual para determinar o subconjunto de menor distância média e aplicar o algoritmo não-linear apenas para melhorar a matriz fundamental calculada por este subconjunto. Isto produz o melhor resultado em termos de exatidão e rapidez.

5 CONCLUSÃO

A matriz fundamental é de grande importância em qualquer sistema de visão estereoscópica. Ela transforma o emparelhamento de pontos de um problema de duas dimensões para um problema de uma única dimensão, pois sabe-se que o ponto correspondente está sobre uma linha epipolar.

O conjunto inicial de emparelhamentos, necessário ao cálculo da matriz fundamental, contém uma porcentagem desconhecida de emparelhamentos ruins. Isto pede a utilização de um algoritmo robusto para o cálculo desta matriz. Dentre os algoritmos robustos, o LMedS é um dos que melhor se ajusta ao cálculo da matriz fundamental. Este trabalho demonstra de forma detalhada que o algoritmo LMedS baseado nos algoritmos de 8 pontos e no não-linear fornece o melhor resultado em termos de velocidade e exatidão. Além disso, os testes demonstram que o pouco conhecido método do paralaxe virtual pode dar ótimos resultados se os pontos que formam os emparelhamentos estiverem bem distribuídos pela imagem.

6 AGRADECIMENTOS

Os autores agradecem aos revisores pelos comentários que ajudaram a elevar o nível do trabalho, à CAPES e à CPG-Uel pelo financiamento das pesquisas e ao professor Eduardo Camponogara do DAS-UFSC pelos comentários esclarecedores.

REFERÊNCIAS

- Boufama, B. S. and Mohr, R. (1998). A stable and accurate algorithm for computing epipolar, *Intern. Journal of Pattern Recognition and Artificial Intelligence* **12**(6): 817–840.
- Burschka, D. and Hager, G. (2002). Scene classification from dense disparity maps in indoor environments, *ICPR*.
- Burschka, D., Lee, S. and Hager, G. (2002). Stereo-based obstacle avoidance in indoor environments with active sensor re-calibration, *IEEE International Conference on Robotics and Automation*, pp. 2066–2072.
- Debevec, P. E., Taylor, C. J. and Malik, J. (1996). Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach, *Computer Graphics* **30**(Annual Conference Series): 11–20.
- Dudek, G., Freedman, P. and Rekleitis, I. (1996). Just-in-time sensing: efficiently combining sonar and laser range data for exploring unknown worlds, *IEEE Conf. on Robotics and Automation*, pp. 667–672.
- Hartley, R. (1995). In defence of the 8-point algorithm, *5th International Conference on Computer Vision*, IEEE Computer Society, pp. 1064–1070.
- Kitano, H., Tambe, M., Stone, P., Veloso, M., Coradeschi, S., Osawa, E., Matsubara, H., Noda, I. and Asada, M. (1997). The RoboCup synthetic agent challenge, 97, *International Joint Conference on Artificial Intelligence*.
- Loop, C. and Zhang, Z. (1999). Computing rectifying homographies for stereo vision, *Conference on Computer Vision and Pattern Recognition*, pp. 125–131.
- Lorigo, L., Brooks, R. and Grimson, W. (1997). Visually-guided obstacle avoidance in unstructured environments.
- Luong, Q. T., Deriche, R., Faugeras, O. and Papadopoulos, T. (1993). On determining the fundamental matrix: analysis of different methods and experimental results, *Technical Report RR-1894*, INRIA.
- Luong, Q.-T. and Faugeras, O. (1996). The fundamental matrix: theory, algorithms, and stability analysis, *The International Journal of Computer Vision* **17**(1): 43–76.
- Luong, Q.-T. and Faugeras, O. (1997). Self-calibration of a moving camera from point correspondences and fundamental matrices, *IJCV* **22**(3): 261–289.
- Olsen, S. L. (1992). Epipolar line estimation, *European Conference on Computer Vision*, Vol. 588, G. Sandini, pp. 307–311.
- Rousseeuw, P. J. (1987). *Robust Regression and Outlier Detection*, Wiley.

- Shapiro, L. and Brady, J. (1994). Rejecting outliers and estimating errors in an orthogonal regression framework, *Royal A-350*: 403–439.
- Shoval, S., Borenstein, J. and Koren, Y. (1994). Mobile robot obstacle avoidance in a computerized travel aid for the blind, *IEEE International Conference on Robotics and Automation*, pp. 2023–2029.
- Torr, P. (1995). *Outlier Detection and Motion Segmentation*, PhD thesis, Dept. of Engineering Science, University of Oxford.
- Torr, P. and Murray, D. (1997). The development and comparison of robust methods for estimating the fundamental matrix, *International Journal of Computer Vision* **24**(3): 271–300.
- Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: a review, *The International Journal of Computer Vision* **27**(2): 161–195.
- Zhang, Z., Deriche, R., Faugeras, O. D. and Luong, Q. T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artificial Intelligence* **78**(1-2): 87–119.
- Zhang, Z. and Loop, C. (2001). Estimating the fundamental matrix by transforming image points in projective space, *Computer Vision and Image Understanding* **82**(2): 174–180.