Algoritmo Robusto para Correspondência de Pontos em Imagens Estereoscópicas na Ausência de Calibração

Robust Algorithm for Point Matching in Uncalibrated Stereo Vision Systems

José A. de França[1]; Marcelo R. Stemmer[2]; Maria B. de M. França[3]

Resumo

Apresenta-se um novo algoritmo de correspondência de pontos em imagens estereoscópicas. As câmeras que capturam as imagens não precisam estar calibradas. O único pré-requisito é a existência de um conjunto de cantos segmentados em cada imagem. Para realizar a correspondência, inicialmente, técnicas não-paramétricas são aplicadas as imagens que compõem o par e um conjunto de candidatos à correspondência é formado. Em seguida, a confiança de cada candidato é calculada através de uma equação proposta. Por último, a matriz fundamental do sistema é estimada e a restrição epipolar utilizada para eliminar falsas correspondências. Testes em imagens reais demonstram a viabilidade do método proposto.
Palavras-chave: Transformada Censo. Correspondência de Pontos. Visão Binocular. Matriz Fundamental.

Abstract

This article introduces a new point matching algorithm for stereo images. The cameras used for capturing the image do not need to be calibrated. The only requirement is the existence of a set of segmented corners in each image. In order to execute the point matching, the algorithm starts by applying non-parametric techniques to the pair of images and a set of candidate matches is selected. After that, the reliability of each point is calculated based on a proposed equation. Finally, the fundamental matrix of the system is estimated and the epipolar restriction is used to eliminate outliers. Tests made on real images demonstrate the viability of the proposed method.
Key words: Census Transform. Point Correspondences. Stereo Vision. Fundamental Matrix.

[1] Universidade Estadual de Londrina, Departamento de Engenharia Elétrica, E-mail: josealexandre@eeol.org

[2] Universidade Federal de Santa Catarina, Departamento de Automação e Sistemas, E-mail: marcelo@das.ufsc.br

[3] Universidade Estadual de Londrina, Departamento de Engenharia Elétrica, E-mail: bernadete@eeol.org

Introdução

Estabelecer correspondência de pontos entre duas ou mais imagens é uma das tarefas mais comuns na visão computacional. De fato, diversos algoritmos encontrados na literatura em áreas como, por exemplo, estereoscopia (HIRSCHMÜLLER, 2002), calibração automática de câmeras (HABED; BOUFAMA, 2002), estimação do movimento (do inglês *motion estimation*) (DORNAIKA; CHUNG, Sept., 2000), rastreamento de objetos (do inglês *object tracking*) (BAE, jun., 2003), assumem a existência de um conjunto de correspondências de pontos entre duas ou mais imagens. Contudo, apesar dos esforços de pesquisadores em todo o mundo, o problema mostra-se extremamente complexo e ainda não existe uma solução automática que dê bons resultados na maioria dos casos.

Diversos fatores tornam a correspondência de pontos difícil: (i) a ambigüidade inerente ao problema requer a introdução de restrições físicas e geométricas; (ii) oclusões, i.e., pontos em uma imagem sem um correspondente na outra; (iii) distorções radiométricas que fazem a projeção de um mesmo ponto 3D ter tons de cinza diferentes, quando ele é capturado por câmeras distintas, e; (iv) distorções projetivas que tornam a forma de um objeto diferente, quando capturadas de pontos-de-vista distintos.

Recentemente, Brown, Burschka e Hager (2003) e Scharstein e Szeliski (2002) revisaram vários problemas e algumas soluções para o processo de correspondência de pontos. Contudo, em tais trabalhos, implicitamente, os parâmetros das câmeras que compunham o conjunto binocular eram supostos conhecidos. No presente trabalho, os esforços foram concentrados no problema de correspondência de pontos em imagens estereoscópicas na ausência de calibração. Nesse caso, a única restrição geométrica disponível é a matriz fundamental (ARMANGUÉ; SALVI, 2003). Infelizmente, tal matriz tem que ser estimada a partir de um conjunto inicial de correspondências. Para formar este conjunto inicial, uma das abordagens é segmentar pontos interessantes (i.e., contornos, linhas, cantos) nas duas imagens e, em seguida, estabelecer correspondência entre eles. Isso reduz a complexidade do problema, mas torna possível estabelecer apenas um conjunto esparso de correspondências. Contudo, tal conjunto é suficiente para diversas aplicações (MOISAN; STIVAL, 2004; CHEN, 2003; MENG; ZHUANG, 2001).

Na concepção do algoritmo proposto, o parâmetro de projeto principal foi, com uma "complexidade computacional baixa", estabelecer um conjunto de correspondência de pontos para cálculo da matriz fundamental. Por isso, em quase todas as suas fases, o algoritmo é não-iterativo e utiliza apenas operações com números inteiros. Basicamente, o método proposto funciona como segue. Após extrair um conjunto de cantos em duas imagens estereoscópicas, usa-se a transformada censo (ZABIH; WOODFILL, 1994) para reduzir a influência das oclusões e distorções, e estabelecer um conjunto inicial de candidatos à correspondência. Em seguida, em oposição a algoritmos de correspondência que utilizam técnicas iterativas como, por exemplo, relaxação, uma equação que indica o grau de confiança de um par candidato à correspondência é proposta e utilizada para eliminar a ambigüidade de forma não-iterativa. Em seguida, a matriz fundamental é estimada robustamente e a restrição epipolar empregada para eliminar falsas correspondências. Com isso, testes em imagens reais mostram que o algoritmo proposto estabelece uma quantidade maior de boas correspondências mesmo na presença de distorções radiométricas e projetivas, e oclusões.

Notação

No decorrer do texto, matrizes e vetores são representados por letras, números ou símbolos em ne-

grito. Constantes são expressas por letras, números ou símbolos em itálico.

Os planos de imagem das câmeras que compõem o conjunto binocular são expressos por I_1 e I_2 para, respectivamente, a câmera da esquerda e a câmera da direita.

As coordenadas da projeção de um ponto 3D no plano de imagem I_α, $\alpha \in \{1, 2\}$, são representadas como $\mathbf{m}_\alpha = [u_\alpha, v_\alpha]^T$. Também, $I_\alpha(u_\alpha, v_\alpha)$ representa o valor do pixel da imagem I_α que está na posição (u_α, v_α). Além disso, as coordenadas homogêneas de um ponto $\mathbf{m}_\alpha = [x_\alpha, y_\alpha, ...]^T$ são representadas por $\widetilde{\mathbf{m}}_\alpha$, isto é, $\widetilde{\mathbf{m}}_\alpha = [x_\alpha, y_\alpha, ..., 1]^T$. Um segundo índice, se houver, indica a posição de um ponto específico em um conjunto de pontos.

Por último, uma reta $\mathbf{l}_\alpha$, no plano de imagem I_α e que passa pelo ponto $\mathbf{m}_\alpha = [u_\alpha, v_\alpha]^T$, deve satisfazer $a_\alpha u_\alpha + b_\alpha v_\alpha + c_\alpha = 0$. Essa mesma reta é representada no texto como sendo $\mathbf{l}_\alpha = [a_\alpha, b_\alpha, c_\alpha]^T$. Assim, é utilizada uma prática notação para a equação da mesma, ou seja, $\mathbf{l}_\alpha^T \mathbf{m}_\alpha = 0$ ou $\mathbf{m}_\alpha^T \mathbf{l}_\alpha = 0$. Novamente, um segundo índice, se houver, indica a posição da reta em um conjunto de retas.

Método Proposto

Para reduzir o espaço de busca, o algoritmo proposto pressupõe a existência de um conjunto de cantos segmentados em cada imagem do par estereoscópico. Em seguida, um conjunto de pares de cantos candidatos à correspondência é formado baseado na correlação entre tais cantos. Logo após, a ambigüidade é eliminada atribuindo-se um grau de confiança a cada par candidato e eliminando-se correspondências pouco confiáveis. Por último, a matriz fundamental é estimada robustamente e a restrição epipolar utilizada para eliminar as falsas correspondências.

A seguir, as etapas do algoritmo são descritas em pormenores.

Escolha dos candidatos à correspondência

Para formação do conjunto de candidatos à correspondência, é utilizada a restrição da semelhança. Esta é baseada no valor de intensidade dos pixels da imagem na posição do ponto de interesse e obriga o correspondente de um ponto a ser similar a ele.

Considerando um ponto $\mathbf{m}_{1i}$ no plano I_1 e $\mathbf{m}_{2j}$ no plano I_2. A forma mais simples de expressar a semelhança entre $\mathbf{m}_{1i}$ e $\mathbf{m}_{2j}$ é por meio do somatório das diferenças absolutas dado pela seguinte equação

$$SAD(c_{ij}) = \sum_{(u,v)\in W} |I_1(u, v) - I_2(u + x, v + y)|, \quad (1)$$

onde $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ é um par candidato à correspondência, $I_z(k, l)$ é o valor da função de intensidade da imagem z na posição (k, l), W é uma janela de correlação centrada em $\mathbf{m}_{1i}$, x e y representam o deslocamento de $\mathbf{m}_{2j}$ em relação a $\mathbf{m}_{1i}$, e $(u, v) \in W$ representa todos os pontos dentro de W.

A equação (1) é baseada na suposição de que os tons de cinza em torno da projeção de um mesmo ponto 3D são os mesmos nas duas imagens de um conjunto binocular. Contudo, tal suposição nem sempre é válida, porque os tons de cinza em torno de um ponto diferem dos tons de cinza ao redor do seu correspondente na outra imagem por uma constante de desvio e um fator de ganho. Por isso, geralmente, a correlação cruzada normalizada e de média zero é uma medida que fornece melhores resultados. Tal medida é definida por

$$ZNCC(c_{ij}) = \frac{\sum_{(u,v)\in W} \delta_1(u, v)\delta_2(u + x, v + y)}{\sqrt{\sigma_1^2(u, v)\sigma_2^2(u + x, v + y)}}, \quad (2)$$

onde $\delta_z(k,l) = I_z(k,l) - \overline{I}_z$ e

$$\sigma_z^2(k,l) = \sum_{(u,v)\in W} [\delta_z(k,l)]^2. \quad (3)$$

Apesar de ser mais robusta, a equação (2) aumenta a complexidade computacional do algoritmo. Em vista disso, Zabih e Woodfill (1994) propuseram a aplicação de técnicas não-paramétricas ao problema de correspondência de pontos. Estas técnicas, ao invés de considerar a intensidade de um pixel em si, baseiam-se na ordem relativa das intensidades dos pixels dentro de uma janela W_n. Dessa forma, elas são robustas com respeito as distorções radiométricas.

Uma das técnicas não-paramétricas conhecidas é a transformada posto (ZABIH; WOODFILL, 1994). Nesta, um pixel na posição (u,v), é substituído pelo número de pixels dentro de uma janela W_n, centrada em (u,v), que possuem intensidade menor que $I(u,v)$. Assim, a imagem original é transformada em uma matriz de números. Além disso, o número de bits necessário para armazenar a imagem após a transformação é, geralmente, menor. Por exemplo, para W_n de dimensão 5×5, são necessários apenas 5 bits para representar cada elemento da imagem transformada.

Já que reduz o efeito das distorções radiométricas, a transformada posto permite o uso de correlação por SAD para expressar a semelhança entre pontos. Contudo, esta transformada não retém a localização dos pixels dentro da janela W_n. Para isso, pode-se utilizar a transformada censo (ZABIH; WOODFILL, 1994).

Na transformada censo, os pixels dentro da janela W_n, centrada em (u,v), são mapeados em uma seqüência de bits. Assim, se um pixel tem intensidade maior que $I(u,v)$, o bit correspondente na seqüência é feito igual a "1", mas, em caso contrário ele é igual a "0". Com isso, para W_n de dimensão 5×5, são necessários 24 bits para representar cada elemento da imagem transformada. Um aumento significativo!

Após a transformada censo, para comparar a semelhança entre elementos de duas imagens, deve-se utilizar distância Hamming (HAMMING, 1950), por meio do somatório das distâncias Hamming dentro de uma janela W, ou seja,

$$SHD(c_{ij}) = \sum_{(u,v)\in W} I_1'(u,v) \oplus I_2'(u+x, y+v), \quad (4)$$

onde $\oplus$ é a distância Hamming entre I_1' e I_2'. Aqui, I_1' e I_2' representam, respectivamente, a transformada censo das imagens 1 e 2. Assim, quanto menor o SHD, maior a semelhança entre os pontos que compõem o par c_{ij}.

A robustez da transformada censo pode ser demonstrada observando a figura 1. Nesta, considere que o pixel central das janelas 3×3 são correspondentes. Contudo, as intensidades desses pixels são diferentes entre si. Obviamente, isso dificulta a correspondência por correlação. No entando, aplicando-se a transformada censo, os pixels em questão são transformados em duas seqüências de bits idênticas, logo a distância Hamming entre eles é nula. Pelos critérios descritos nesta seção, isso significa dizer que os pixels têm a mesma "assinatura". Além disso, a aplicação da transformada censo a todos os pixels dentro de W, seguido do cálculo da distância Hamming, deve revelar que tais pontos são correspondentes com uma maior segurança.

Como as transformadas posto e censo envolvem apenas comparações entre número inteiros, elas podem ser aplicadas a um baixo custo computacional. Contudo, como envolve a manipulação de uma quantidade menor de dados, a transformada posto tende a ser de implementação mais fácil.

255	43	78
89	123	199
200	23	12

10001100

240	25	85
100	150	199
200	35	18

10001100

Figura 1. Duas janelas de correlação 3×3, cujos tons de cinza são diferentes. Contudo, a transformada censo revela que a distância Hamming dos pontos centrais é nula.

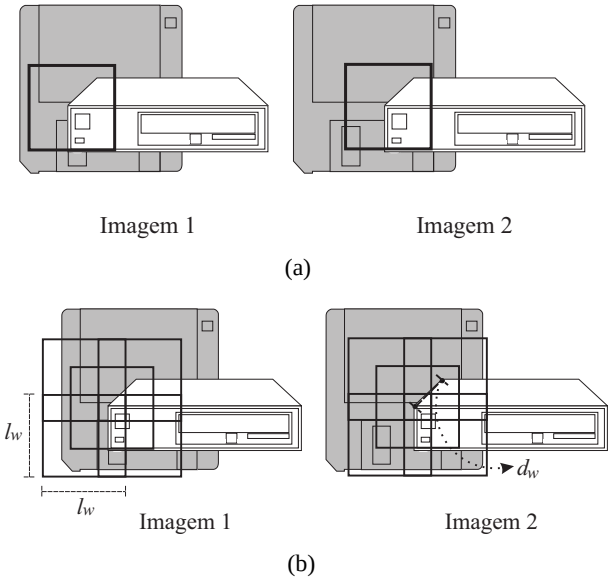

Figura 2. (a) influência das oclusões na correlação com apenas uma janela de correlação. (b) correlação com 5 janelas. Aqui, em pelo menos uma janela, a semelhança é conservada.

Múltiplas janelas de correlação

Oclusões e distorções projetivas fazem com que a vizinhança de um ponto interessante seja diferente nas duas imagens de um conjunto binocular. Na figura 2(a), ilustra-se um exemplo típico de oclusão. Nela, objetos a distâncias diferentes das câmeras parecem mover-se quando são observados de perspectivas diferentes. Assim, a imagem em torno do ponto de interesse é diferente nas duas imagens e isso, evidentemente, dificulta a correspondência por semelhança. Uma maneira de reduzir este problema é diminuir a largura, l_w, da janela de correlação, contudo, isto aumenta a influência do ruído e tende a reduzir o número de boas correspondências. Em vista disso, Kanade e Okutomi (1994) propuseram um método no qual o tamanho e a forma da janela de correlação adaptam-se ao local do ponto de interesse, porém esse método tem um custo computacional muito elevado. Por isso, vários autores como, por exemplo, Hirschmüller, Innocent e Garibaldi (2002), Fusiello, Roberto e Trucco (), propuseram uma alternativa à ténica de Kanade e Okutomi (1994). Tal alternativa é baseada em múltiplas janelas de correlação. Por exemplo, Hirschmüller, Innocent e Garibaldi (2002) utilizaram um sistema com cinco janelas de correlação, como ilustrado na figura 2(b). Assim, a correlação entre pontos utilizada era um valor baseado na correlação devido a cada uma das cinco janelas. Pode-se dizer que esse método é uma simplificação da técnica de Kanade e Okutomi (1994).

Técnicas de correlação com janelas mais bem elaboradas aumentam significativamente o custo computacional do algoritmo. Por isso, neste trabalho, utilizou-se a configuração padrão, ou seja, uma única janela com tamanho fixo, o que significa que se estar depositando extrema confiança na transformada censo.

Abordagem proposta

No método proposto, para escolher os pares candidatos à correspondência, inicialmente se aplica a transformada censo às duas imagens. Em seguida, para cada canto segmentado, $\mathbf{m}_{1i}$, de I_1, calcula-se a semelhança entre este e todos os cantos segmentados, $\mathbf{m}_{2j}$, de I_2 que estão dentro de uma janela de dimensão $2l_s \times 2l_s$, centrada em $\mathbf{m}_{1i}$ (figura 3). Então, o par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ que apresentar a maior semelhança, é considerado um candidato à

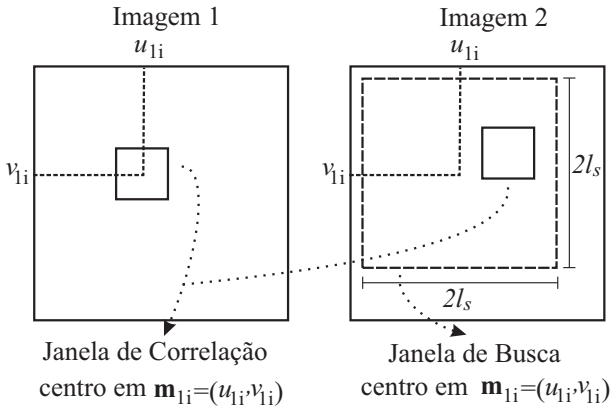

Figura 3. Correspondência de pontos por correlação.

correspondência.

Devido a instabilidade dos algoritmos disponíveis para segmentar os cantos em ambas as imagens, na prática, diversos cantos de uma imagem não correspondem a nenhum canto na outra imagem. Isso dificulta o processo de seleção dos candidatos à correspondência e pode fazer com que exista ambigüidade entre o conjunto de candidatos à correspondência, ou seja, um canto de I_2 forme um par com mais de um ponto de I_1. Assim, uma outra etapa para eliminar a ambigüidade é necessária e discutida a seguir.

Eliminação da Ambigüidade

Considerando o par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ um candidato à correspondência, representa-se o conjunto de vizinhos de $\mathbf{m}_{1i}$ e de $\mathbf{m}_{2j}$ dentro de uma janela de dimensão $2l_n \times 2l_n$ (figura 4) por, respectivamente, $N(\mathbf{m}_{1i})$ e $N(\mathbf{m}_{2j})$. Assim, se c_{ij} for um bom candidato à correspondência, espera-se que existam muitos candidatos à correspondência $V_{kl} = (\mathbf{n}_{1k}, \mathbf{n}_{2l})$, onde $\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})$ e $\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})$, tal que a "posição relativa" entre $\mathbf{n}_{1k}$ e $\mathbf{m}_{1i}$ é semelhante à posição relativa entre $\mathbf{n}_{2l}$ e $\mathbf{m}_{2j}$. Por outro lado, se c_{ij} não é um bom candidato, não se espera encontrar a mesma relação entre esse pontos. Baseado nesta propriedade, definiu-se uma medida da confiança de um candidato à correspondência. Tal medida foi proposta tendo como principal critério a simplicidade.

Formalmente, definimos a confiança, $Reab$, de um par $c_{ij} = (\mathbf{m}_{1i}, \mathbf{m}_{2j})$ candidato à correspondência, baseada nos candidatos à correspondência vizinhos, pela equação

$$Reab(c_{ij}) = \sum_{\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})} \left(\sum_{\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})} \Phi(c_{ij}, v_{kl}) \right), \quad (5)$$

onde $\Phi(c_{ij}, v_{kl})$ é igual a 1 se $(\mathbf{n}_{1k}, \mathbf{n}_{2l})$ é um candidato à correspondência e $r(c_{ij}, v_{kl}) < \varepsilon_r$; caso contrário, é igual a zero. Aqui,

$$r(c_{ij}, v_{kl}) = \frac{|d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) - d(\mathbf{m}_{2j}, \mathbf{n}_{2l})|}{[d(\mathbf{m}_{1i}, \mathbf{n}_{1k}) + d(\mathbf{m}_{2j}, \mathbf{n}_{2l})]/2} \quad (6)$$

é uma medida do erro das posições relativas entre os pares $(\mathbf{m}_{1i}, \mathbf{n}_{1k})$ e $(\mathbf{m}_{2j}, \mathbf{n}_{2l})$, e ε_r é um limiar para esta medida.

Agora, alguns comentários devem ser feitos:

1. A equação (5) conta o número de candidatos à correspondências que são vizinhos do candidato $(\mathbf{m}_{1i}, \mathbf{m}_{2j})$ e possuem posições relativas semelhantes.

2. Para que um par v_{kl} aumente a confiança de c_{ij}, pode-se impor que o ângulo, θ, entre $\overrightarrow{\mathbf{m}_{1i}\mathbf{n}_{1k}}$ e $\overrightarrow{\mathbf{m}_{2j}\mathbf{n}_{2l}}$ seja menor que um limiar θ_{th}. Isso aumenta o custo computacional, mas reduz o número de falsas correspondências.

3. Se mais de um ponto $\mathbf{n}_{1k} \in N(\mathbf{m}_{1i})$ forma um par candidato com um mesmo ponto $\mathbf{n}_{2l} \in N(\mathbf{m}_{2j})$ (figura 4), o somatório da equação (5) deve contar todos os pares ambíguos como se fossem apenas um único par.

4. Na equação (6), pode-se fazer $d(\mathbf{m}_x, \mathbf{n}_y)$ igual à distância euclidiana. Contudo, neste trabalho utilizou-se

$$d(\mathbf{m}_x, \mathbf{n}_y) = |u_m - u_n| + |v_m - v_n|, \quad (7)$$

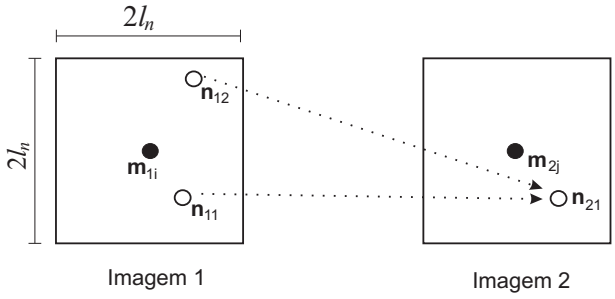

Figura 4. Um par (m_{1i}, m_{2j}) candidato à correspondência com vizinhos dentro de uma janela de lado $2l_n$. O par (n_{11}, n_{21}) aumenta a confiança do par (m_{1i}, m_{2j}).

onde considera-se $\mathbf{m}_x = [u_m, v_m]^T$ e $\mathbf{n}_y = [u_n, v_n]^T$. A equação (7) tem um custo computacional menor que o da distância euclidiana e o seu uso na equação (6) não prejudica o cálculo da confiança de $(\mathbf{m}_{1i}, \mathbf{m}_{2j})$.

Para eliminar a ambigüidade, deve-se aplicar a equação (5) em todos os candidatos à correspondência e:

1. Candidatos com $Reab(c_{ij}) \leq rb_{th}$ são descartados.
2. Caso haja ambigüidade, prevalece o candidato mais confiável.
3. Se dois pares ambíguos possuem a mesma confiança, prevalece aquele que tem a maior semelhança.

Seguindo as orientações acima, a ambigüidade é eliminada em apenas um único passo. Contudo, espera-se que algumas falsas correspondências persistam. Por isso, a matriz fundamental deve ser estimada e a restrição epipolar utilizada para eliminar as falsas correspondências.

Eliminação das falsas correspondências

Quando o conjunto binocular não está calibrado, a geometria epipolar é a única restrição geométrica disponível. Tal geometria já foi descrita em numerosos artigos, por exemplo, os trabalhos de Torr e Murray (1997), Zhang (1998), Armangué e Salvi (2003). Basicamente, a geometria epipolar pode ser entendida se for considerado o caso de duas câmeras como apresentado na figura 5. Nesta, $\mathbf{C}_1$ e $\mathbf{C}_2$ são, respectivamente, os centros ópticos da primeira e segunda câmeras. Então, dado um ponto $\mathbf{m}_1$ na primeira imagem, I_1, o ponto correspondente, $\mathbf{m}_2$, na segunda imagem, I_2, está restrito a uma reta chamada "reta epipolar" de $\mathbf{m}_1$, representada na figura por $\mathbf{l}_2$. A reta $\mathbf{l}_2$ é a intersecção do plano Π, definido por $\mathbf{M}$, $\mathbf{C}_1$ e $\mathbf{C}_2$ (chamado de plano epipolar), com o plano I_2. Isto acontece porque o ponto $\mathbf{m}_1$ pode corresponder a qualquer ponto da reta $\overline{\mathbf{C}_1\mathbf{M}}$ e a projeção de $\overline{\mathbf{C}_1\mathbf{M}}$ em I_2 é a reta $\mathbf{l}_2$. Além disso, observa-se que todas as retas epipolares dos pontos de I_1 passam através de um ponto comum, $\mathbf{e}_2$, em I_2. Este ponto é conhecido como "epipolo". O ponto $\mathbf{e}_1$ é a intersecção da reta $\overline{\mathbf{C}_1\mathbf{C}_2}$ com o plano I_2. Finalmente, pode-se facilmente observar a simetria da geometria epipolar. O correspondente em I_1 de cada ponto $\mathbf{m}_{2i}$, sobre a reta $\mathbf{l}_{2i}$, precisa pertencer a reta epipolar $\mathbf{l}_{1i}$, que é a intersecção do mesmo plano Π_i com o plano I_1. Todas as retas epipolares formam um conjunto contendo o epipolo $\mathbf{e}_1$, que é a intersecção da reta $\overline{\mathbf{C}_1\mathbf{C}_2}$ com o plano I_1.

Normalmente, todas as restrições impostas pela geometria epipolar são resumidas na seguinte equação

$$\widetilde{\mathbf{m}}_2^T \mathbf{F} \widetilde{\mathbf{m}}_1 = 0, \qquad (8)$$

onde $\mathbf{F}$ é conhecida como a "matriz fundamental" do conjunto binocular.

A equação (8) é uma restrição por trás de quaisquer duas imagens se estas são projeções em perspectiva de uma mesma cena. Geometricamente, $\mathbf{F}\widetilde{\mathbf{m}}_1$ define a reta epipolar do ponto $\mathbf{m}_1$ no plano I_2. Assim, a equação (8) não diz nada além de que o ponto correspondente de $\mathbf{m}_1$ (em I_2), ou seja, $\mathbf{m}_2$,

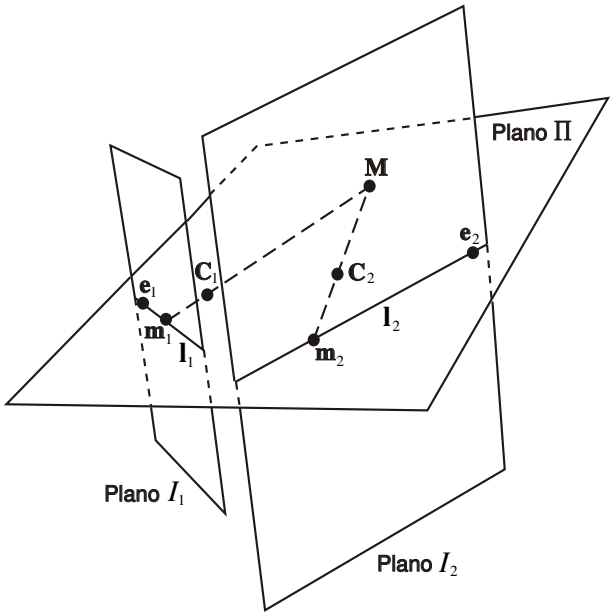

Figura 5. Geometria epipolar.

está sobre a sua reta epipolar $\mathbf{F}\widetilde{\mathbf{m}}_1$.

Encontram-se na literatura diversos métodos para estimar-se a matriz fundamental. Veja-se, por exemplo, os trabalhos de Torr e Murray (1997), Zhang (1998), Armangué e Salvi (2003) para três análises críticas diferentes de tais métodos. Contudo, o método mais conhecido é o algoritmo de 8 pontos (HARTLEY, 1997). Tal método, dado um conjunto com $n \geq 8$ correspondências, estima a matriz fundamental de forma linear, resolvendo

$$\sum_{i=1}^{n} \left\| \widetilde{\mathbf{m}}_{2i}^T \mathbf{F} \widetilde{\mathbf{m}}_{1i} \right\|^2, \quad \text{sujeito a } \|\mathbf{F}\| = 1. \qquad (9)$$

Evidentemente, se a matriz fundamental do conjunto binocular for conhecida, a equação (8) pode ser utilizada na busca por correspondências de pontos. Infelizmente, o algoritmo de 8 pontos já pressupõe a existência de um conjunto de correspondências. Assim, a solução adotada nesta trabalho é utilizar o conjunto de candidatos à correspondência formado pelo método discutido na seção anterior e estimar a matriz fundamental de forma robusta, ou seja, eliminando-se as falsas correspondências enquanto se realiza a estimação.

Um dos métodos robustos mais conhecidos na literatura é o RANSAC (do inglês: *Random Sample Consensus*), introduzido na visão computacional por Fischler e Bolles (1981). Basicamente, o RANSAC pode ser resumido nos passos a seguir.

Dado um conjunto de N correspondências $c_i = (\mathbf{m}_{1i}, \mathbf{m}_{2i})$, escolhe-se, aleatoriamente, N_{sc} subconjuntos de N_{nc} correspondências. Para cada subconjunto (indexado por j), estima-se a matriz fundamental, $\mathbf{F}_j$, e calcula-se o resíduo, $r_{ji}^2(\mathbf{F}_j, c_i)$, de todas as N correspondências. Cada resíduo é comparado com um limiar r_{th}^2. Se $r_{ji}^2 < r_{th}^2$, a correspondência c_i é considerada uma boa correspondência. Após N_{sc} iterações, retém-se a matriz fundamental, $\mathbf{F}_{win}$, que ajustou-se ao maior número de boas correspondências. Por último, estima-se $\mathbf{F}$ com apenas as boas correspondências (identificadas por $r_{(win)i}^2 < r_{th}^2$).

O número de subconjuntos, N_{sc}, utilizado no RANSAC deve ser tal que, supondo, dentre todas as N correspondências, uma porcentagem ϵ de falsas correspondências, exista uma probabilidade p de que (ao menos) um subconjunto j seja composto apenas por boas correspondências. Neste caso, N_{sc} é igual a

$$N_{sc} = \log(1-p) / \log\left(1-(1-\epsilon)^{N_{nc}}\right). \qquad (10)$$

Neste trabalho, a tática utilizada é iniciar N_{sc} usando a equação anterior e atualizá-lo a cada iteração j, ou seja, uma vez que a percentagem, ϵ_j, de falsas correspondências tenha sido determinada, N_{sc} é atualizado por (10).

Da equação anterior, vê-se que N_{sc} aumenta exponencialmente com N_{nc} e ϵ. Por exemplo, considerando uma probabilidade $p = 99\%$, se $\epsilon = 25\%$ e $N_{nc} = 7$, então $N_{sc} = 33$. Contudo, se $\epsilon = 40\%$

e $N_{nc} = 8$, temos $N_{sc} = 272$. Assim, desde que quanto maior N_{sc}, maior o custo computacional, o ideal é ter-se N_{nc} e ϵ tão pequenos quanto possível.

O valor de ϵ depende de como o conjunto total de correspondências foi estabelecido, ou seja, ele depende do grau de confiança do algoritmo de correspondência de pontos.

Como $\mathbf{F}$ tem sete graus de liberdade, o valor mínimo para N_{nc} é sete. Contudo, a solução com apenas sete correspondências não é estável. Por isso, neste trabalho, é utilizado o método de 8 pontos com $N_{nc} = 8$.

No RANSAC, o parâmetro mais crítico a ser escolhido é o limiar r_{th}^2, pois dele depende o critério que diz se uma correspondência é boa ou ruim. Se r_{th}^2 é muito pequeno, boas correspondências podem ser consideradas ruins. Por outro lado, um r_{th}^2 grande faz com que algumas falsas correspondências não sejam detectadas.

Normalmente, o resíduo r_{ji}^2 é dado por

$$r_{ji}^2 = d^2(\widetilde{\mathbf{m}}_{2i}, \mathbf{F}_j\widetilde{\mathbf{m}}_{1i}) + d^2(\widetilde{\mathbf{m}}_{1i}, \mathbf{F}_j^T\widetilde{\mathbf{m}}_{2i}), \quad (11)$$

onde $d^2(\cdot)$ é o quadrado da distância euclidiana.

Além disso, se for considerado um ruído com uma distribuição gaussiana de média zero e desvio padrão σ_d, o resíduo r_{th}^2 é definido como

$$r_{th}^2 = d_{th}^2\sigma_d^2, \quad (12)$$

onde d_{th}^2 deve ser escolhido tal que exista uma probabilidade p_d de uma boa correspondência ser erroneamente considerada uma falsa correspondência. Contudo, muitas vezes, d_{th}^2 é escolhido empiricamente. Por exemplo, Hartley e Zisserman (2000) utilizaram $d_{th}^2 = 3,84$, Zhang (1998) utilizou $d_{th}^2 = 2,5$ e Torr e Murray (1997) consideraram $d_{th}^2 = 1,99$.

O método RANSAC pode obter uma boa estimação de $\mathbf{F}$, mesmo que mais de 50% das correspondências sejam falsas. A desvantagem evidente é que ele necessita de uma estimação do desvio padrão do ruído, σ_d.

Outra característica importante a ser obervada é que, de forma geral, o algoritmo RANSAC procura a matriz $\mathbf{F}_j$ que minimiza a função a seguir

$$\min_{\mathbf{F}_j} \sum_{i=1}^{N} \mathcal{J}(r_{ji}^2),$$

onde

$$\mathcal{J}(r_{ji}^2) = \begin{cases} 0, & \text{se } r_{ji}^2 \leq r_{th}^2 \\ 1, & \text{se } r_{ji}^2 > r_{th}^2 \end{cases}$$

Na equação anterior, é evidente que, se r_{th}^2 tiver um valor muito elevado, todas as correspondências serão consideradas boas. Nesee caso, qualquer matriz $\mathbf{F}_j$ teria a mesma pontuação, ou seja, $\sum_{i=1}^{N} \mathcal{J}(r_{ji}^2)$ seria sempre igual a N. Por isso, Torr e Zisserman (1998) sugeriram uma discreta alteração na função $\mathcal{J}(r_{ji}^2)$, ou seja,

$$\mathcal{J}(r_{ji}^2) = \begin{cases} r_{ji}^2, & \text{se } r_{ji}^2 \leq r_{th}^2 \\ r_{th}^2, & \text{se } r_{ji}^2 > r_{th}^2 \end{cases}$$

Agora, cada boa correspondência contribui com um valor diferente e proporcional ao seu grau de ajuste a $\mathbf{F}_j$. Assim, mesmo considerando um número igual de boas correspondências, a função $\mathcal{J}(r_{ji}^2)$ deve ter valores diferentes para matrizes fundamentais diferentes.

Torr e Zisserman (1998) demonstraram que essa pequena modificação produz uma sensível melhoria no desempenho do algoritmo RANSAC. Assim, desde que o custo computacional adicionado é desprezível, este algoritmo (conhecido como MSAC, do inglês: *M-Estimator Sample Consensus*) é utilizado no presente trabalho.

A etapa de eliminação das falsas correspondências é a única fase iterativa do algoritmo proposto. Con-

tudo, o uso do algoritmo MSAC aumenta significativamente a robustez do método. Como demonstrado empiricamente na próxima seção, ele garante a obtenção de um conjunto confiável de boas correspondências em situações bem distintas.

Resultados Experimentais

Nesta seção, o método de correspondência de pontos proposto é avaliado. Para isto, são utilizados os três pares de imagens das figuras 6, 7 e 8. O par da figura 6 possui o epipolo próximo ao centro das imagens. Este foi escolhido para contrastar com o par da figura 7, cujos epipolos tendem ao infinito. Por último, o par da figura 8 foi utilizado como imagem-teste devido à distorção projetiva elevada. Tal distorção ocorreu porque as câmeras que formavam o conjunto binocular possuiam distâncias focais bem distintas. Este fato dificulta ainda mais o processo de correspondência de pontos.

Cada imagem-teste tem dimensão de 640 $\times$ 480 pixels. Por comodidade, o par de imagens da figura 6 é referenciado no texto por **MESA** e os das figuras 7 e 8, respectivamente, por **PLANTA** e **DESKTOP**.

O algoritmo SUSAN (SMITH, 1992) foi utilizado para extrair cantos de cada uma das imagens. São com esses cantos que os métodos avaliados devem tentar estabelecer correspondência de pontos.

O desempenho do algoritmo proposto é comparado ao do método de Zhang, Deriche e FAUGERAS O; LUONG (1995). Tal método é um dos mais citados na literatura e é tido como um dos mais robustos. Outro fator importante é que o método de Zhang, Deriche e FAUGERAS O; LUONG (1995) também é dividido em três etapas: formação do conjunto de candidatos à correspondência; eliminação da ambigüidade, e; identificação das falsas correspondências. Contudo, ao contrário do método proposto, este método não utiliza transformações não-paramétricas. Por isso, a correlação dada pela equação (2) tem que ser usada para formar o conjunto de candidatos à correspondência. Outro diferencial é o uso de técnicas iterativas de relaxação para eliminar a ambigüidade. Tal técnica utiliza uma equação bastante complexa para calcular o grau de confiança dos candidatos à correspondência. Isso contribui para que esta seja a fase do algoritmo que requer mais tempo de processamento.

Ambos os métodos testados foram implementados na forma de um programa para o MATLAB. Ao contrário do proposto por Zhang, Deriche e FAUGERAS O; LUONG (1995), a implementação do algoritmo **ZHANG** utilizada neste trabalho usa o método MSAC para identificar as falsas correspondências. Isso foi necessário por dois motivos: (a) o método LMedS, usado no artigo original de Zhang, Deriche e FAUGERAS O; LUONG (1995), não fornece bons resultados se a quantidade de falsas correspondências é maior que 50 %, e; (b) já que o método proposto também usa o MSAC, isto facilita a comparação dos resultados.

O algoritimo **ZHANG** foi implementado com os parâmetros recomendados no artigo de Zhang, Deriche e FAUGERAS O; LUONG (1995). Para o método proposto, a transformada censo foi aplicada às imagens com uma janela W_n de 5×5 pixels. Com isso, cada pixel da imagem é transformado em uma seqüência de 24 bits. Para formar o conjunto de candidados a correspondência, utilizou-se uma janela de correlação de 11×11 pixels. Além disso, dado um determinado ponto, $\mathbf{m}_{1i} \in I_1$, a correlação deste é calculada com todos os pontos pertencentes a I_2 e que estão dentro de uma janela de busca centrada em $\mathbf{m}_{1i}$, de diâmetro igual a $1/4$ da largura da imagem. Estes são valores empíricos, mas que funcionaram bem para todos os casos testados.

Para a fase de eliminação da ambigüidade, o método proposto foi implementado considerando $\varepsilon_r = 0,04$, $\theta_{th} = \pi/2$ e $rb_{th} = 2,5\%$ do número

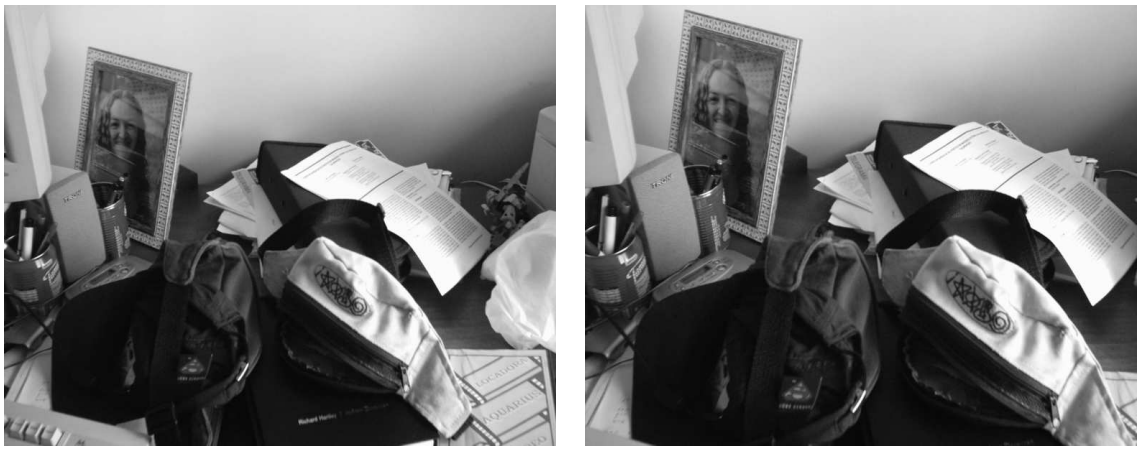

Figura 6. O par de imagens chamado **MESA**, utilizado na avaliação dos métodos de correspondência de pontos.

Figura 7. Imagens **PLANTA** utilizadas na avaliação dos métodos propostos.

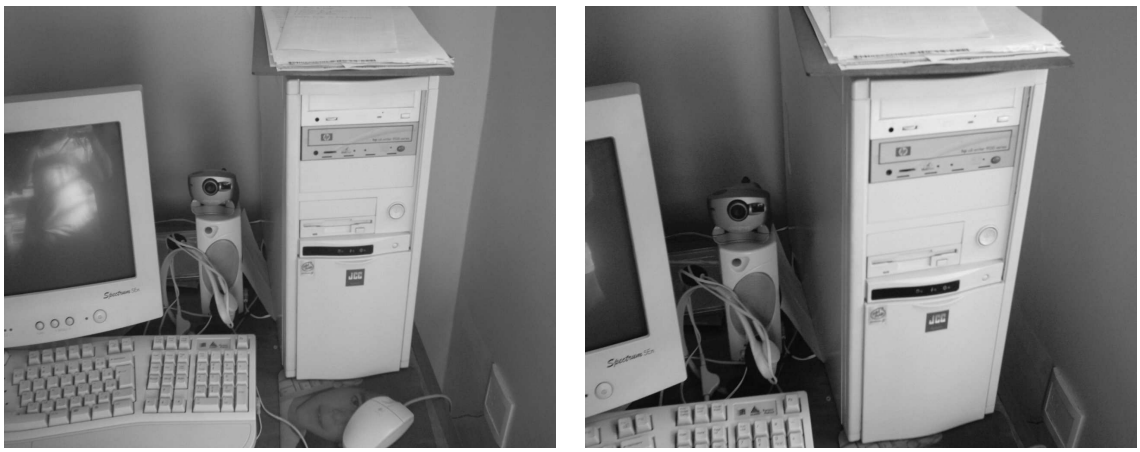

Figura 8. Imagens referenciadas como **DESKTOP** e utilizadas na avaliação dos métodos de correspondência de pontos.

de cantos extraídos da imagem 1. O parâmetro ε_r depende da "semelhança" entre as câmeras que compõem o conjunto binocular, ou seja, quanto mais distintos (entre si) forem os parâmetros das duas câmeras, maior deve ser o valor de ε_r. Por exemplo, no par **DESKTOP**, as câmeras possuem distâncias focais bem distintas. Por isso, para este par, ε_r foi feito igual a 0,10. Outro importante fator que influencia a escolha de ε_r é a separação entre as câmeras. Uma separação maior implica um ε_r também maior. Por isso, esse parâmetro deve ser escolhido experimentalmente para cada conjunto binocular. Contudo, uma vez que as câmeras permanecem fixas durante toda a operação normal do conjunto binocular, um ε_r bem escolhido servirá adequadamente durante todo o tempo que o sistema for utilizado.

Na tabela 1 são apresentados, para cada par de imagens, a quantidade de correspondências estabelecidas em cada fase pelos métodos testados. Além disso, é apresentado também o ajuste da matriz fundamental (estimada pelo algoritmo MSAC) ao conjunto final de correspondências, dado por

$$r^2(\mathbf{F}_j) = \frac{1}{2N}\sum_{i=0}^{N} r_{ji}^2, \qquad (13)$$

onde r_{ji}^2 é dado por (11).

Este ajuste fornece uma indicação da qualidade do conjunto de correspondências encontrado. Um ajuste pequeno (menor que um) indica um bom conjunto de correspondências.

Observando a tabela 1, exceto na primeira etapa, nota-se que os resultados de ambos os métodos são bem semelhantes. Em particular, ambos têm dificuldade em estabelecer correspondências de pontos com o par **DESKTOP**. Contudo, devido à distorção projetiva elevada, como os métodos utilizam semelhança para formar o conjunto inicial de correspondências, isto já é esperado. Mesmo assim, os métodos conseguem estabelecer, em todos os casos, um conjunto de correspondências que ajusta-se bem à geometria epipolar do conjunto binocular e que tem mais que 100 elementos.

A tabela 2 mostra o tempo de execução dos algoritmos. Observa-se que o algoritmo **ZHANG** necessita de um tempo muito maior para concluir as etapas 1 e 2, respectivamente, formação do conjunto de correspondências e eliminação da ambigüidade. Por isso, o algoritmo proposto é executado, em média, em um tempo total dez vezes menor. Isso acontece, principalmente, devido ao grande número de candidatos à correspondência encontrados na primeira etapa do processo e, evidentemente, esse fato sobrecarrega a etapa de eliminação da ambigüidade. Por outro lado, no método proposto, este processo é feito de forma bastante simples, não-iterativa e utilizando apenas operações com números inteiros.

Conclusão

Talvez devido à popularidade das técnicas de calibração clássica, maior ênfase é dada ao processo de correspondência de pontos em sistemas já calibrados. Para o caso não-calibrado, os algoritmos disponíveis são, na maioria das vezes, pouco eficientes, pois utilizam técnicas complexas e de custo computacional elevado, ou, ao contrário, usam uma medida de semelhança simples, que forma um conjunto de correspondência inicial pouco confiável, e, em seguida, aplicam a restrição epipolar, através de algoritmos como o RANSAC ou LMedS, para identificar os candidatos confiáveis. Nesse último caso, devido à grande ambigüidade, o algoritmo de correspondência torna-se bastante iterativo. Em oposição, o algoritmo de correspondência proposto utiliza técnicas simples, mas muito eficientes, pois consegue estabelecer um conjunto de correspondência inicial reduzido, mas confiável. Tal confiança é aumentada ao longo de três fases distintas. O resultado é um algoritmo simples, eficiente e rápido.

Tabela 1. Quantidade de correspondências encontradas em cada etapa do algoritmo pelos métodos analisados. Na primeira coluna, a quantidade de cantos encontrada em cada imagem do par é apresentada entre parenteses.

Par de imagens	Método utilizado	Número de correspondências			Ajuste
		Etapa 1	Etapa 2	Etapa 3	
MESA	**ZHANG**	32.579	1.130	572	0,48
(2.079 × 1.982)	**PROPOSTO**	2.079	1.083	524	0,53
PLANTA	**ZHANG**	102.972	1.543	777	0,41
(2.731 × 2.833)	**PROPOSTO**	2.731	1.625	770	0,36
DESKTOP	**ZHANG**	158.532	1.499	230	0,57
(3.000 × 2.817)	**PROPOSTO**	2.290	1.034	141	0,48

Tabela 2. Tempo (em segundos) necessário para os métodos analisados concluirem cada um das etapas dos algoritmos.

Par de imagens	Método utilizado	Tempo de execução (s)			
		Etapa 1	Etapa 2	Etapa 3	Total
MESA	**ZHANG**	18,03	79,01	3,96	107,02
	PROPOSTO	7,44	2,56	3,90	13,90
PLANTA	**ZHANG**	24,94	481,75	3,39	515,08
	PROPOSTO	11,35	15,03	8,75	35,13
DESKTOP	**ZHANG**	29,34	1.644,21	194,98	1.868,53
	PROPOSTO	16,48	21,18	167,85	205,51

A análise dos resultados da seção anterior mostra que a transformada censo, em conjunto com a distância Hamming, constitui uma excelente medida de semelhança entre pontos. Além disso, a equação (5) possibilita uma prática ferramenta para redução da ambigüidade no processo de correspondência. Contudo, o método ainda pode ser melhorado. Possivelmente, uma alteração que aumentaria significativamente a confiança das correspondências seria estabalecer um limiar máximo para a distância Hamming, ou seja, uma correspondência fora desse limiar seria desconsiderada. Contudo, a escolha de tal limiar deve ser feita dinamicamente, de acordo com o conjunto total de possíveis correspondências. Um estudo deste tipo foi realizado recentemente por Kanatani e Kanazawa (2004), para o caso da equação (2). Um estudo semelhante para o caso da transformada censo reduziria o número de iterações necessárias à terceira fase do algoritmo proposto. Contudo, evidentente, aumentaria o seu custo computacional.

Agradecimentos

Os autores agradeçem aos revisores anônimos pelas contribuições para melhoria do texto e a PROPPG-UEL e ao CNPq pelo financiamento das pesquisas.

Referências

ARMANGUÉ, X.; SALVI, J. Overall view regarding fundamental matrix estimation. *Image and Vision Computing*, Guildford, v. 21, p. 205–220, 2003.

BAE, K.-H.; KOO, J.-S.; KIM, E.-S. A new stereo object tracking system using disparity motion vector. *Optics Communications*, Amsterdam, v. 221, n. 113, p. 23–35, jun., 2003.

BROWN, M.; BURSCHKA, D.; HAGER, G. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, New York, v. 25, n. 8,

p. 993–1008, 2003. Disponível em: <http://jobber.cs.jhu.edu/~burschka/>. Acesso em: set. 2005.

CHEN, Z.; WU, C.; TSUI, H. T. A new image rectification algorithm. *Pattern Recognition Letters*, Amsterdam, v. 24, n. 1-3, p. 251–260, 2003.

DORNAIKA, F.; CHUNG, R. Cooperative Stereo-Motion: Matching and Reconstruction. *Computer Vision and Image Understanding*, Guildford, v. 79, n. 3, p. 408–427, Sept., 2000.

FISCHLER, M.; BOLLES, R. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, New York, v. 24, n. 6, p. 381–385, 1981.

FUSIELLO, A.; ROBERTO, V.; TRUCCO, E. Efficient stereo with multiple windowing. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 1997, Porto Rico. Proceedings... New York, 1997. Disponível em: <http://citeseer.ist.psu.edu/fusiello97efficient.html>. Acesso em: set. 2005.

HABED, A.; BOUFAMA, B. S. Camera self-calibration from two views. *IEEE International Systems, Man and Cybernetics*, v. 4, p. 5, 2002.

HAMMING, R. Error-detecting and error-correcting codes. *Bell System Technical Journal*, New York, v. 29, n. 2, p. 147–160, 1950.

HARTLEY, R. In defence of the eight point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, New York, v. 19, n. 6, p. 580–593, 1997. Disponível em: <http://rsise.anu.edu.au/~hartley/My-Papers.html>. Acesso em: set. 2005.

HARTLEY, R.; ZISSERMAN, A. Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press, 2000.

HIRSCHMÜLLER, H.; INNOCENT, P.; GARIBALDI, J. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, Dordrecht, v. 47, n. 1, p. 229–246, 2002. Disponível em: <http://citeseer.ist.psu.edu/hirschmuller02realtime.html>. Acesso em: set. 2005.

KANADE, T.; OKUTOMI, M. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, New York, v. 16, n. 9, p. 920–932, 1994.

KANATANI, K.; KANAZAWA, Y. Automatic thresholding for correspondence detection. *International Journal of Image and Graphics*, v. 4, n. 1, p. 21–33, 2004. Disponível em: <http://www.suri.it.okayama-u.ac.jp/~kanatani/data/ejournal.html>. Acesso em: set. 2005.

MENG, Y.; ZHUANG, H. Self-calibration of camera-equipped robot manipulators. *The International Journal of Robotics Research*, Cambridge, v. 20, n. 11, p. 909–921, 2001.

MOISAN, L.; STIVAL, B. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *Pattern Recognition Letters*, Amsterdam, v. 57, n. 3, p. 201–218, 2004.

SCHARSTEIN, D.; SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, Dordrecht, v. 47, n. 1, p. 7–42, 2002. Disponível em: <http://citeseer.ist.psu.edu/scharstein01taxonomy.html>. Acesso em: set. 2005.

SMITH, S. A new class of corner finder. In: *BRITISH MACHINE VISION CONFERENCE*, 3, Washington, 1992. Proceedings... Washington: [Spinger Verlab], 1992, p. 139-148. 1992. Disponível em: <http://www.fmrib.ox.ac.uk/~steve/susan/>. Acesso em: set. 2005.

TORR, P.; MURRAY, D. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, Dordrecht, v. 24, n. 3, p. 271–300, 1997.

TORR, P. H. S.; ZISSERMAN, A. Robust computation and parametrization of multiple view relations. In: DESAI, U. (Ed.). *ICCV6*. New Delhi: Narosa Publishing House, 1998. p. 727–732. Disponível em: <http://wwwcms.brookes.ac.uk/~philiptorr/>. Acesso em: set. 2005.

ZABIH, R.; WOODFILL, J. Non-parametric local transforms for computing visual correspondence. In: *European Conference on Computer Vision*, 3. Stockholm, v. 2, p. 151–158, 1994. Disponível em: <http://citeseer.ist.psu.edu/article/zabih94nonparametric.html>. Acesso em: set. 2005.

ZHANG, Z. Determining the epipolar geometry and its uncertainty: A review. *The International Journal of Computer Vision*, v. 27, n. 2, p. 161–195, August 1998.

ZHANG, Z.; DERICHE, R.; FAUGERAS O; LUONG, Q. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, Amsterdam, v. 78, n. 1-2, p. 87–119, 1995.