

José A. de França

Universidade Estadual de Londrina
Departamento de Engenharia Elétrica
Caixa Postal 6025
86051-990, Londrina - PR, Brazil
<http://research.eeol.org/>
Email: josealexandre@eeol.org

Maria B. de M. França

Universidade Estadual de Londrina
Departamento de Engenharia Elétrica
Caixa Postal 6025
86051-990, Londrina - PR, Brazil
<http://bernadete.eeol.org/>
Email: bernadete@eeol.org

Marcelo R. Stemmer

Universidade Federal de Santa Catarina
Departamento de Automação e Sistemas
Caixa Postal 476
88040-970, Florianópolis-SC, Brazil
<http://www.das.ufsc.br/~marcelo/>
Email: marcelo@das.ufsc.br

Abstract—This paper address the fundamental problems on mobile robotics, that is, detection and localization of obstacles. In this work, we solved these problems with a binocular stereo system. First, the fundamental matrix is estimated and the system is self-calibrated. Soon afterwards, we use dense disparity images from a correlation-based stereo reconstruction algorithm to segment and to locate the objects in scene. Experiments in real images show that the new technique has a good performance.

Index Terms — Stereovision, Point Correspondences, Fundamental Matrix, self-calibration, Obstacle Avoidance.

I. INTRODUÇÃO

A robótica móvel é uma das áreas de pesquisa mais ativas em todo o mundo. Isto é devido a sua aplicação imediata em uma variedade de tarefas como, por exemplo, missões espaciais, operações em ambientes perigosos e/ou subaquáticos, segurança civil, exploração científica, indústria, e exploração de áreas de desastre. Dentro deste contexto, a identificação e desvio de obstáculos é uma tarefa fundamental. Desde que, nos animais, o principal sentido envolvido nesta tarefa é a visão, é natural que diversos trabalhos utilizem visão computacional na navegação de robôs móveis. Dentre estes, alguns são baseados em sensores, tais como sonares e lasers, que fornecem informações 3D diretamente [1]. Em outros casos, marcas guias e sistemas monoculares são combinados com um conhecimento prévio do ambiente para situar o veículo e guiá-lo em seu trajeto [2]. Em oposição a projetos desse tipo, descrevemos um sistema de visão que detecta e estima a localização de obstáculos apenas à partir de um par estéreo, 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 do veículo autônomo, pois um par estéreo consome bem menos energia que, por exemplo, sistemas à lasers.

Neste trabalho, após um estágio de calibração, o sistema adquire um par de imagens estereoscópicas, realiza uma retificação destas imagens e produz um mapa de disparidades da cena. Em seguida, utilizando este mapa, obstáculos que estão próximos ao robô são segmentados e têm a sua localização determinada. Isto torna possível conduzir um veículo móvel através do caminho mais seguro.

Para desenvolver o sistema, foram criados novos algoritmos para correspondência de pontos em imagens estereoscópicas e

recuperação da geometria epipolar [3]. Estes algoritmos foram desenvolvidos para aplicações com restrição de tempo. Por isso, possuem baixo custo computacional.

O restante deste artigo é estruturado como segue. Na seção II, apresenta-se a notação adotada. A seção III contém os passos necessários para calibrar o sistema. A abordagem utilizada para detectar e conhecer a localização dos obstáculos é discutida, respectivamente, nas seções IV e V. Por último, as conclusões e os trabalhos futuros estão na seção VII.

II. 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* [3], 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, $\tilde{\mathbf{x}} = [x, y, \dots, t]^T$. Além disso, uma reta $ax + by + ct = 0$ é representada por $\mathbf{l} = [a, b, c]^T$. Finalmente, usamos a prática notação $A^{-T} = (A^{-1})^T = (A^T)^{-1}$ para toda matriz quadrada inversível.

III. CONFIGURAÇÃO DO SISTEMA

Inicialmente, o sistema necessita de uma fase de configuração. Esta consiste em estimar a matriz fundamental e as matrizes de projeção das duas câmeras.

A. Geometria Epipolar

Em um sistema de visão estéreo sem calibração, a única restrição disponível é a geometria epipolar. Esta geometria já foi descrita em numerosos artigos como, por exemplo, [4], [5]. Basicamente, dado um ponto \mathbf{M} no espaço 3D, a geometria epipolar relaciona as projeções \mathbf{m}_1 e \mathbf{m}_2 desse ponto, respectivamente, na primeira e segunda imagens, pela equação

$$\tilde{\mathbf{m}}_2^T \mathbf{F} \tilde{\mathbf{m}}_1 = 0, \quad (1)$$

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

Na visão estéreo, uma vez que a matriz fundamental é conhecida, todas as demais tarefas tornam-se menos complexas. Por isso, neste trabalho, o primeiro objetivo é estimar esta matriz. Para isto, um conjunto de correspondência de pontos $(\mathbf{m}_{1i}, \mathbf{m}_{2i})$ deve ser conhecido. Assim, desenvolvemos

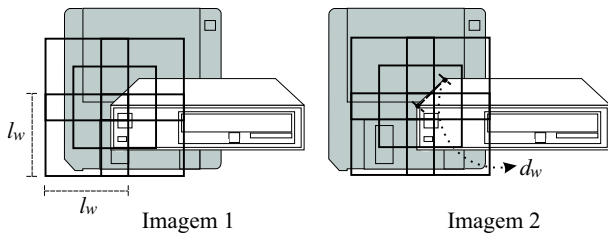


Fig. 1: Correlação com 5 janelas. Aqui, em pelo menos uma janela, a semelhança é conservada mesmo com oclusões.

um algoritmo de correspondência de pontos em imagens estereoscópicas na ausência de calibração. Este está descrito, de forma simplificada, na seção à seguir.

1) *Correspondência de Pontos*: Quando o sistema não está calibrado, o espaço de busca necessário à correspondência de pontos é muito grande. Por isso, neste trabalho, dado um par de imagens estereoscópicas, I_1 e I_2 , o algoritmo de detecção de cantos SUSAN [6] é aplicado em ambas as imagens. Em seguida, nosso algoritmo tenta encontrar cantos em I_1 que correspondam a cantos em I_2 . Isto é feito usando a restrição da semelhança, através de correlação por *SAD* [7], ou seja, um ponto m_{1i} forma um candidato a correspondência com o ponto m_{2i} mais correlacionado.

Desde que a correlação por *SAD* é muito sensível a distorções radiométricas, a transformada posto [8] é aplicada as imagens antes da correlação.

Além das distorções radiométricas, outras dificuldades inerentes ao problema da correspondência de pontos são as oclusões e as distorções projetivas. Estas fazem com que a vizinhança da projeção de um ponto 3D seja diferente em cada imagem do par estéreo. Uma maneira de reduzir este problema é diminuir a janela de correlação. Contudo, isto aumenta a influência do ruído e tende a reduzir o número de boas correspondências. Por isto, usamos a configuração de 5 janelas de correlação apresentada na Fig. 1. Nesta, a correlação *SAD* é calculada em uma janela central e em 4 periféricas (Fig. 1). A correlação final é dada pela seguinte equação

$$SAD = \frac{SAD_0 + SAD_{Min}}{2}, \quad (2)$$

onde SAD_0 é a correlação da janela central e SAD_{Min} é a correlação da janela periférica que apresenta a maior semelhança.

Após criar um conjunto de candidatos à correspondência $c_{ij} = (m_{1i}, m_{2j})$, definimos um critério para estimar a confiança de cada um destes pares. Considerando um par de pontos candidato à correspondência c_{ij} , podemos representar $N(m_{1i})$ e $N(m_{2j})$ como sendo, respectivamente, o conjunto de vizinhos de m_{1i} e m_{2j} dentro de uma janela quadrada de lado $2l_n$ (Fig. 2). Se c_{ij} for um bom candidato à correspondência, espera-se que existam muitos candidatos a correspondência $v_{kl} = (n_{1k}, n_{2l})$, onde $n_{1k} \in N(m_{1i})$ e $n_{2l} \in N(m_{2j})$. Além disso, a posição relativa entre n_{1k} e m_{1i} é semelhante a posição relativa entre n_{2l} e m_{2j} . Por

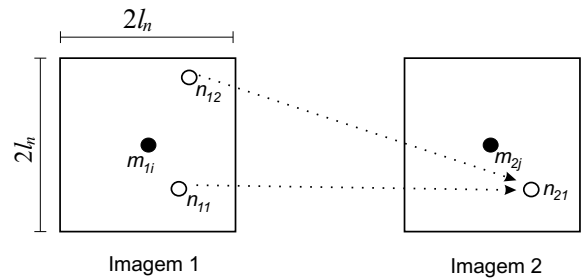


Fig. 2: Um par (m_{1i}, m_{2j}) candidato à correspondência com vizinhos dentro de uma janela de lado $2l_n$. O par (n_{11}, n_{21}) aumenta a confiança do par (m_{1i}, m_{2j}) .

outro lado, se c_{ij} não é um bom candidato, não se espera encontrar vizinhos com posições relativas semelhantes.

Formalmente, definimos a confiança, Cf , de um par c_{ij} candidato a correspondência pela equação

$$Cf(c_{ij}) = \sum_{n_{1k} \in N(m_{1i})} \sum_{n_{2l} \in N(m_{2j})} \Phi(c_{ij}, v_{kl}), \quad (3)$$

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

$$r(c_{ij}, v_{kl}) = \frac{|d(m_{1i}, n_{1k}) - d(m_{2j}, n_{2l})|}{[d(m_{1i}, n_{1k}) + d(m_{2j}, n_{2l})]/2} \quad (4)$$

é uma medida do erro das posições relativas entre os pares (m_{1i}, n_{1k}) e (m_{2j}, n_{2l}) , e ε_r é um limiar para esta medida. Além disso, $d(\cdot, \cdot)$ é a distância euclidiana.

Com a equação (3), podemos descartar correspondências pouco confiáveis e eliminar a ambigüidade (conservando apenas os pares mais confiáveis). Contudo, ainda é provável que algumas poucas falsas correspondências não sejam descartadas. Por isso, utilizamos o algoritmo LMedS (*Least Median of Squares*) [4] para elimina-las. O LMedS estima a matriz fundamental robustamente e, em seguida, elimina os pares de pontos que não satisfazem a equação (1).

A implementação do algoritmo LMedS utilizada neste trabalho é baseada no algoritmo de 8 pontos normalizado [9] e está descrita em [10].

Como observado por Rousseeuw [11], o algoritmo LMedS é sensível ao ruído Gaussiano. Assim, a matriz fundamental obtida por este método não é ótima. Por isso, após a eliminação das falsas correspondências, utilizamos um método não-linear para estimar tal matriz.

2) *Estimação da Matriz Fundamental*: Se duas imagens são projeções de um plano P no espaço projetivo 3D, elas estão relacionadas por uma homografia, \mathbf{H} . Para pontos que não pertencem a P , $\tilde{\mathbf{m}}_2 \neq \lambda \mathbf{H} \tilde{\mathbf{m}}_1$, onde λ é fator de escala desconhecido. Esse fato (que representa o paralaxe) permite estimar um epipolo se \mathbf{H} for conhecida [12]. Assim, a matriz fundamental pode ser conhecida através da relação [12]

$$\mathbf{F} = [\mathbf{e}_2]_{\times} \mathbf{H}, \quad (5)$$

onde $[\mathbf{e}_2]_{\times}$ é a matriz anti-simétrica que representa o produto cruzado com o epipolo $\tilde{\mathbf{e}}_2$ da imagem I_2 .

O método do paralaxe virtual proposto por [13] é baseado na equação (5), mas dispensa o conhecimento prévio de um plano. Para isto, são necessárias inicialmente três correspondências de pontos $(\tilde{\mathbf{m}}_{1i}, \tilde{\mathbf{m}}_{2i})$, $i = 1, 2$ e 3 . Com essas correspondências, pode-se formar duas bases projetivas, ou seja, $B_1 = \{\tilde{\mathbf{m}}_{11}, \dots, \tilde{\mathbf{m}}_{13}, \tilde{\mathbf{m}}_{10}\}$ e $B_2 = \{\tilde{\mathbf{m}}_{21}, \dots, \tilde{\mathbf{m}}_{23}, \tilde{\mathbf{m}}_{20}\}$, onde $\tilde{\mathbf{m}}_{10} = \tilde{\mathbf{m}}_{11} + \tilde{\mathbf{m}}_{12} + \tilde{\mathbf{m}}_{13}$ e $\tilde{\mathbf{m}}_{20} = \tilde{\mathbf{m}}_{21} + \tilde{\mathbf{m}}_{22} + \tilde{\mathbf{m}}_{23}$.

No sistema de coordenadas definido por B_1 e B_2 , \mathbf{H} tem a forma

$$\mathbf{H} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (6)$$

onde α e β são constantes desconhecidas.

Assim, considerando a equação (5) e $\tilde{\mathbf{e}}_2 = [e_{2u}, e_{2v}, e_{2t}]^T$, tem-se

$$\mathbf{F} = \begin{bmatrix} 0 & -\beta e_{2t} & e_{2y} \\ \alpha e_{2t} & 0 & -e_{2x} \\ -\alpha e_{2y} & -\beta e_{2x} & 0 \end{bmatrix}. \quad (7)$$

Agora, \mathbf{F} pode ser estimada através de um método linear (veja [13], [10] para mais detalhes), semelhante ao método dos oito pontos, mas que necessita de apenas cinco correspondências de pontos. Contudo, o método não tem aplicação prática, pois a sua exatidão é muito dependente dos pontos que compõem as bases B_1 e B_2 . Assim, dependendo da escolha dessas bases, o método fornece péssimos resultados.

Ignorando por um momento os problemas relacionados as bases B_1 e B_2 , observamos que a equação (7) fornece uma parametrização muito simples para a matriz fundamental. Assim, podemos usá-la para resolver

$$\sum_{i=1}^n g_i (\tilde{\mathbf{m}}_{2i}^T \mathbf{F} \tilde{\mathbf{m}}_{1i})^2 = 0, \quad \text{sujeito a } \|\mathbf{F}\| = 1, \quad (8)$$

onde \mathbf{F} é dada pela equação (7) e

$$g_i = \left(\frac{1}{\mathbf{l}_{1i}^T \mathbf{T}_{1B} \mathbf{Z} \mathbf{T}_{1B}^T \mathbf{l}_{1i} + \mathbf{l}_{2i}^T \mathbf{T}_{2B} \mathbf{Z} \mathbf{T}_{2B}^T \mathbf{l}_{2i}} \right), \quad (9)$$

onde consideramos que B_1 e B_2 são caracterizadas pelas transformações \mathbf{T}_{1B} e \mathbf{T}_{2B} respectivamente, $\mathbf{Z} = \text{diag}(1, 1, 0)$, $\mathbf{l}_{2i} = \mathbf{F} \tilde{\mathbf{m}}_{1i}$ e $\mathbf{l}_{1i} = \mathbf{F}^T \tilde{\mathbf{m}}_{2i}$.

Experimentalmente, concluímos que o método do paralaxe virtual fornece os melhores resultados quando B_1 e B_2 têm o esboço semelhante a figura 3. Desde que, como todo método não-linear, o método proposto necessita de uma estimativa inicial, \mathbf{F}_0 , da matriz fundamental, podemos usar \mathbf{F}_0 para conseguir pares de correspondências “virtuais” que formem bases B_1 e B_2 ideais. Isto pode ser realizado como segue:

- 1) Considerando apenas a imagem I_1 , calcula-se a altura h e a largura L do retângulo que envolve o conjunto de correspondências desta imagem. Em seguida, para que o triângulo formado pelos pontos de B_1 e este retângulo tenham mesma área e centróide, podemos colocar seus vértices nas coordenadas $\mathbf{m}_{11} = [\bar{u}_1 - 2\sqrt{2}L/3, \bar{v}_1]^T$, $\mathbf{m}_{12} = [\bar{u}_1 + \sqrt{2}L/3, \bar{v}_1 + h]^T$ e $\mathbf{m}_{13} = [\bar{u}_1 +$

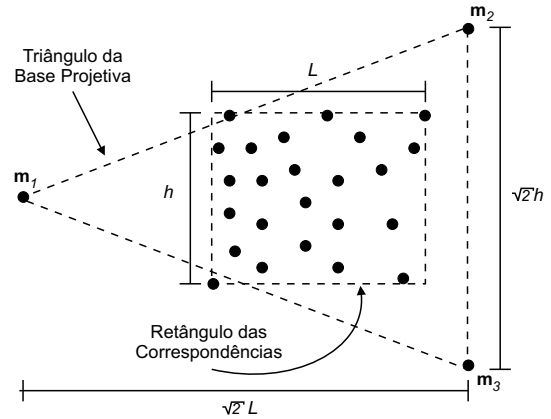


Fig. 3: Bases ideais para o algoritmo do paralaxe virtual.

$\sqrt{2}L/3, \bar{v}_1 - h]^T$, onde consideramos o centróide do retângulo igual a $\mathbf{c}_{1r} = [\bar{u}_1, \bar{v}_1]^T$;

- 2) Assim, os pontos que formam a base projetiva na imagem I_2 podem ser: \mathbf{m}_{21} igual a intersecção da reta $[1, 0, -\bar{u}_2 + 2\sqrt{2}L'/3]^T$ com a reta epipolar $\mathbf{l}_{21} = \mathbf{F}_0 \tilde{\mathbf{m}}_{21}$, \mathbf{m}_{22} igual a intersecção da reta $[1, 0, -\bar{u}_2 - \sqrt{2}L'/3]^T$ com a reta $\mathbf{l}_{22} = \mathbf{F}_0 \tilde{\mathbf{m}}_{22}$, e \mathbf{m}_{23} igual a intersecção da reta $[1, 0, -\bar{u}_2 + \sqrt{2}L'/3]^T$ com a reta $\mathbf{l}_{23} = \mathbf{F}_0 \tilde{\mathbf{m}}_{23}$. Aqui, L' é a largura do retângulo que envolve as correspondências em I_2 e \bar{u}_2 é a coordenada x do centróide do conjunto de correspondências em I_2 .

Agora, se tivermos uma boa estimativa inicial \mathbf{F}_0 , com essas bases B_1 e B_2 ideais e as correspondências $(\mathbf{T}_{1B} \tilde{\mathbf{m}}_{1i}, \mathbf{T}_{2B} \tilde{\mathbf{m}}_{2i})$ obtidas pelo nosso método, obteremos sempre bons resultados com o método não-linear do paralaxe virtual.

B. Calibração do Par Estéreo

Com a matriz fundamental estimada, Hartley *et al.* [14] demonstraram que é possível encontrar matrizes de projeção \mathbf{P}_1^P e \mathbf{P}_2^P , respectivamente, da câmera 1 e 2, e realizar uma reconstrução dentro do espaço projetivo 3D. Assim, considerando o ponto \mathbf{M}_i^E no espaço euclidiano 3D e as suas projeções \mathbf{m}_{1i} e \mathbf{m}_{2i} , é possível encontrar o ponto \mathbf{M}_i^P no espaço projetivo 3D. Este processo é conhecido como *weak calibration* [14], [15].

Movendo o par estéreo por um deslocamento euclidiano (rotação e translação não-nulas) e utilizando as mesmas matrizes \mathbf{P}_1^P e \mathbf{P}_2^P , obtemos uma reconstrução \mathbf{M}_i^{2P} para o mesmo ponto \mathbf{M}_i^E . Com quatro correspondências $(\mathbf{M}_i^{1P}, \mathbf{M}_i^{2P})$ é possível encontrar uma transformação projetiva, \mathbf{T}^{21} , tal que

$$\mathbf{M}_i^{2P} = \mathbf{T}^{21} \mathbf{M}_i^{1P}. \quad (10)$$

Em 1995, Zisserman *et al.* [16] demonstraram que o plano no infinito, $\mathbf{\Pi}_\infty$, está relacionado com a transformação \mathbf{T}^{21} , ou seja, $\mathbf{\Pi}_\infty$ é o autovetor que corresponde ao autovalor real de $(\mathbf{T}^{21})^{-T}$.

Já é bem conhecido que, conhecendo-se $\mathbf{\Pi}_\infty$, pode-se facilmente encontrar uma transformação projetiva que transforme

as matrizes \mathbf{P}_1^p e \mathbf{P}_2^p , respectivamente, nas matrizes \mathbf{P}_1 e \mathbf{P}_2 no espaço euclidiano. Isto é feito através de um algoritmo linear (veja [3], [17] para mais detalhes).

Como o procedimento descrito nessa seção, nosso sistema está completamente calibrado, através de um único deslocamento euclidiano do par estéreo e do uso de algoritmos lineares e não-iterativos.

IV. SEGMENTAÇÃO DOS OBSTÁCULOS

Para identificar os obstáculos que estão no campo de visão do robô, nossa abordagem é utilizar correspondência de pontos por correlação para formar um mapa de disparidade. Desde que estamos aplicando a transformada posto as imagens, a correlação utilizada pode ser do tipo *SAD*. Com o mapa de disparidades, podemos desprezar todos os pontos da imagem que estiverem a uma profundidade maior que um limiar, d_p . Assim, a imagem resultante será constituída por agrupamentos de pixels que representam os obstáculos próximos ao veículo.

Já que a matriz fundamental do sistema é conhecida, a tarefa de correspondência de pontos consiste em, dado um ponto m_1 na imagem I_1 , procurar o ponto m_2 mais correlacionado na imagem I_2 , sobre a reta epipolar correspondente. Esta tarefa é menos complexa se as imagens estiverem retificadas. Em imagens retificadas, as retas epipolares nas duas imagens são horizontais e pontos correspondentes possuem a mesma coordenada y .

Para realizar a retificação, utilizamos o método proposto por Pollefeys *et al.* [18] e, recentemente, atualizado por Chen *et al.* [19]. Este método tem um baixo custo computacional, pois não envolve o cálculo de matrizes de homografia. Ao invés disso, as linhas epipolares, em ambas as imagens, são extraídas (utilizando o algoritmo de Bresenham [20]), uma após a outra, para formar as novas imagens retificadas. Assim, a distorção é mínima e o tamanho das imagens retificadas não é muito maior que as originais. Por isso, o método é indicado para aplicações que possuem restrição de tempo. A sua desvantagem é que a correspondência entre as coordenadas dos pontos das imagens retificadas e das imagens originais são armazenadas em uma tabela, ou seja, uma grande quantidade de memória é necessária (veja [19] para mais detalhes).

Com as imagens retificadas I_1^r e I_2^r , a correspondência de pontos é feita como segue. Dado um ponto $\mathbf{m}_1^r = [u_1, v_1]^T$ na imagem I_1^r , o seu correspondente $\mathbf{m}_2^r = [u_2 + d, v_2]^T$ é aquele que apresenta a maior correlação. Aqui, d é uma constante positiva que representa a disparidade entre m_1^r e m_2^r . As coordenadas dos pontos nas imagens originais podem ser conseguidas rapidamente com ajuda da tabela gerada durante a retificação [19], [18].

O maior problema na elaboração do mapa de disparidade são as oclusões. Estas dificultam a correspondência de pontos nas regiões próximas as bordas dos obstáculos. Para amenizar este problema, ao invés de utilizar o sistema de múltiplas janelas descrito na seção III-A.1, utilizamos a consistência esquerda-direita [21]. Desde que a correlação entre pontos

de ambas as imagens, pertencentes a uma mesma linha horizontal, seja retida em memória, o aumento da complexidade computacional requerido para checar a consistência esquerda-direita é muito baixo.

Com os obstáculos segmentados, como discutimos na próxima seção, triangulação pode ser utilizada para estimar a sua localização.

V. LOCALIZAÇÃO DOS OBSTÁCULOS

Uma vez que os obstáculos tenham sido segmentados, as correspondências estabelecidas para elaboração do mapa de disparidade e as matrizes de projeção das câmeras (estimadas durante a calibração) podem ser utilizadas para estimar a localização dos obstáculos. Isto é feito como segue. Dado duas matrizes de projeção \mathbf{P}_1 e \mathbf{P}_2 , respectivamente, da imagem I_1 e I_2 , e uma correspondência (m_1, m_2) , tem-se as seguintes equações [3]

$$\tilde{\mathbf{m}}_1 = s\mathbf{P}_1\tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}_2 = s'\mathbf{P}_2\tilde{\mathbf{M}}, \quad (11)$$

onde, $\tilde{\mathbf{M}}$ são as coordenadas homogêneas de um ponto no espaço e, sem perda de generalidade, consideramos $\tilde{\mathbf{m}}_1 = [u_1, v_1, 1]^T$ e $\tilde{\mathbf{m}}_2 = [u_2, v_2, 1]^T$. Além disso, em (11), s e s' são constantes desconhecidas e necessárias devido a equação (11) ser homogênea. Agora, considerando $\mathbf{P}_1 = [\mathbf{p}_{11}, \mathbf{p}_{12}, \mathbf{p}_{13}]^T$ e $\mathbf{P}_2 = [\mathbf{p}_{21}, \mathbf{p}_{22}, \mathbf{p}_{23}]^T$, temos

$$1/s = \mathbf{p}_{13}^T\tilde{\mathbf{M}} \quad \text{e} \quad 1/s' = \mathbf{p}_{23}^T\tilde{\mathbf{M}}. \quad (12)$$

Substituindo (12) em (11), podemos escrever

$$\mathbf{U}\tilde{\mathbf{M}} = \mathbf{0}, \quad (13)$$

onde \mathbf{U} é uma matriz 4×4 dada por

$$\mathbf{U} = [\mathbf{p}_{11} - u_1\mathbf{p}_{13}, \mathbf{p}_{12} - v_1\mathbf{p}_{13}, \mathbf{p}_{21} - u_2\mathbf{p}_{23}, \mathbf{p}_{22} - v_2\mathbf{p}_{23}]^T. \quad (14)$$

Dessa forma, uma estimativa da localização dos obstáculos presente no campo visual do veículo móvel é conhecida. Com essas informações, é possível utilizar algoritmos como, por exemplo, o proposto por Davison e Murray [22], que criam um mapa 2D do ambiente e indicam o melhor caminho a ser seguido.

VI. RESULTADOS EXPERIMENTAIS

Desde que abordamos apenas o sistema de visão de um veículo móvel, os testes foram realizados em computador *desktop* equipado com um processador Pentium IV. O par estéreo foi constituído por duas câmeras que geram imagens em tons de cinza.

Como descrito na seção III, inicialmente, um conjunto de correspondências é estabelecido pelo algoritmo da seção III-A.1 e a matriz fundamental é estimada utilizando o algoritmo descrito na seção III-A.2.

Nas Figs. 4(a) e 4(b), apresentamos um par de imagens estéreo de uma cena que contém uma porta aberta. Nesta figuras, as correspondências de pontos e algumas linhas epipolares estão esboçadas. Para o caso destas figuras, o algoritmo de



(a)



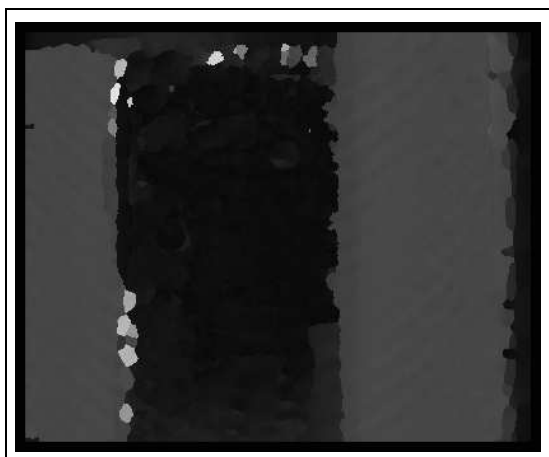
(b)



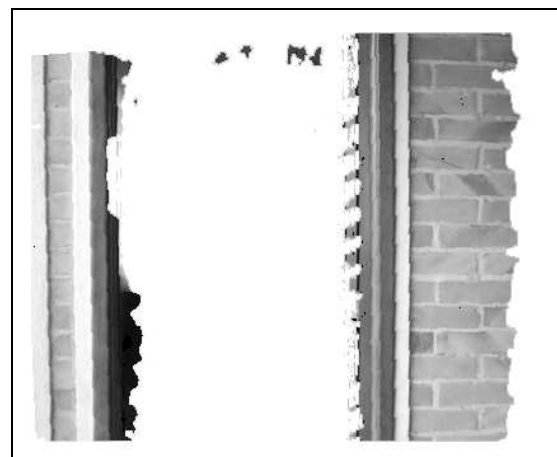
(c)



(d)



(e)



(f)

Fig. 4: (a) e (b) são imagens de um par estéreo com correspondência de pontos e linhas epipolares esboçadas; (c) e (d) é o mesmo par após a retificação (linhas epipolares horizontais); (e) é o mapa de disparidades das imagens retificadas, e; (f) são os obstáculos (paredes) identificados.

correspondência estabeleceu $n = 105$ boas correspondências e nenhuma falsa correspondência.

Com as correspondências obtidas, o método não-linear do paralaxe virtual (seção III-A.2) foi utilizado para estimar a matriz fundamental. A exatidão desta estimação pode ser medida através do resíduo

$$r^2(\mathbf{F}) = \sum_{i=1}^n [d^2(\tilde{\mathbf{m}}_{2i}, \mathbf{F}\tilde{\mathbf{m}}_{1i}) + d^2(\tilde{\mathbf{m}}_{1i}, \mathbf{F}^T\tilde{\mathbf{m}}_{2i})], \quad (15)$$

onde $d(\tilde{\mathbf{x}}_i, \mathbf{l}_i)$ é a distância euclidiana do ponto $\tilde{\mathbf{x}}_i$ à reta \mathbf{l}_i .

Para o exemplo da figura 4, $r^2(\mathbf{F}) = 200 \times 10^{-3}$, ou seja, a matriz foi estimada com bastante exatidão.

Uma vez que a matriz fundamental foi estimada, as matrizes de projeção das câmeras são estimadas pelo método da seção III-B.

A pesar do sistema estar calibrado, as linhas epipolares são oblíquas em relação ao plano das imagens [Figs. 4(a) e 4(b)]. Por isso, o algoritmo de retificação da seção IV é utilizado. Nas Figs. 4(c) e 4(d), o resultado desta operação é apresentado. Agora, as linhas epipolares são horizontais e linhas correspondentes possuem mesma coordenada y . Assim, utilizando correlação e o teste de consistência esquerda direita, computamos o mapa de disparidades apresentado na Fig. 4(e).

Com o mapa de disparidade, podemos eliminar todos os pixels que correspondem a objetos distante das câmeras. Assim, como apresentado na Fig. 4(f), obtemos uma imagem com as paredes identificadas como obstáculos e o vão da porta como caminho livre. Desse modo, como o processo de triangulação descrito na seção V, pode-se utilizar algoritmos para que um robô autônomo entre pela porta de forma segura.

VII. CONCLUSÃO E TRABALHOS FUTUROS

Apresentamos um sistema de visão estéreo de detecção e localização de obstáculos para uso em robótica móvel. Em oposição a trabalhos deste tipo que utilizam, além do sistema de visão, outros sensores, um pre-conhecimento ou alterações no ambiente do robô, propomos um sistema de visão que identifica obstáculos no campo de visão do robô e torna possível a elaboração de um mapa 2D com a localização desses obstáculos.

O sistema ainda possui algumas falhas. Em alguns casos, o piso do ambiente é confundido com obstáculos próximos. Para evitar este problema, a altura das câmeras e a sua inclinação em relação a horizontal devem ser estimados durante a fase de calibração. Dessa forma, esperamos que seja possível conhecer a distância das câmeras ao piso e distinguir dentre este e os obstáculos.

Um outro problema surge quando obstáculos possuem texturas "pobres". Neste caso, existe pouca informação para calcular a disparidade na região desses objetos. Assim, penas as suas bordas são detectadas. Contudo, este problema pode ser corrigido se o sistema for dotado de fonte de iluminação estruturada que forneça a textura necessária.

AGRADECIMENTOS

Os autores agradecem à CAPES e à PROPPG-UEL pelo financiamento das pesquisas.

REFERENCES

- [1] G. A. Borges and M.-J. Aldon, "Motion estimation by iterative 2-D features matching in range images," in *IEEE International Conference on Robotics and Automation*, 2000, pp. 3197–3202.
- [2] W.-H. Lee, K.-S. Roh, and I.-S. Kweon, "Self-localization of a mobile robot without camera calibration using projective invariants," *Pattern Recognition Letters*, vol. 21, no. 1, pp. 45–60, 2002.
- [3] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. MIT Press, Cambridge, MA, 2001.
- [4] X. Armangué and J. Salvi, "Overall view regarding fundamental matrix estimation," *Image and Vision Computing*, vol. 21, pp. 205–220, 2003.
- [5] Z. Zhang, "Determining the epipolar geometry and its uncertainty: a review," *The International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–195, August 1998.
- [6] S. Smith, "A new class of corner finder," in *3rd British Machine Vision Conference*, 1992, pp. 139–148.
- [7] J. Banks and P. Corke, "Quantitative evaluation of matching methods and validity measures for stereo vision," *The International Journal of Robotics Research*, vol. 20, no. 7, pp. 521–532, July 2001.
- [8] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *3rd European Conference on Computer Vision*, vol. 2, 1994, pp. 151–158.
- [9] R. Hartley, "In defence of the eight point algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 1997.
- [10] J. A. de França and M. R. Stemmer, "Estimação robusta da matriz fundamental," in *VI Simpósio Brasileiro de Automação Inteligente*, Bauru, Brazil, september 2003, pp. 253–258. [Online]. Available: <http://research.eeol.org/publications/>
- [11] P. J. Rousseeuw, *Robust Regression and Outlier Detection*. Wiley, 1987.
- [12] Q.-T. Luong and O. Faugeras, "The fundamental matrix: theory, algorithms, and stability analysis," *The International Journal of Computer Vision*, vol. 17, no. 1, pp. 43–76, 1996.
- [13] B. S. Boufama and R. Mohr, "A stable and accurate algorithm for computing epipolar," *Intern. Journal of Pattern Recognition and Artificial Intelligence*, vol. 12, no. 6, pp. 817–840, 1998.
- [14] R. Hartley, R. Gupta, and T. Chang, "Stereo from uncalibrated cameras," in *II European Conference on Computer Vision*, 1992, pp. 579–587.
- [15] O. D. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?" in *II European Conference on Computer Vision*, 1992, pp. 563–578.
- [16] A. Zisserman, P. Beardsley, and I. Reid, "Metric calibration of a stereo rig," in *IEEE Workshop on Representations of Visual Scenes*, Boston, EUA, 1995, pp. 93–100.
- [17] R. Horaud, G. Csurka, and D. Demirdjian, "Stereo calibration from rigid motions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1446–1452, 2000.
- [18] M. Pollefeys, R. Koch, and L. J. V. Gool, "A simple and efficient rectification method for general motion," in *ICCV (I)*, 1999, pp. 496–501.
- [19] Z. Chen, C. Wu, and H. T. Tsui, "A new image rectification algorithm," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 251–260, 2003.
- [20] D. Hearn and M. Baker, *Computer Graphics, 2nd ed.* Prentice-Hall International Inc Press, New Jersey, 1998.
- [21] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.
- [22] A. J. Davison and D. W. Murray, "Simultaneous localization and map-building using active vision," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 865–880, 2002.