

Desenvolvimento de Algoritmos de Visão Estereoscópica para Aplicações em Robótica Móvel

JOSÉ ALEXANDRE DE FRANÇA[†]

Orientador:

MARCELO RICARDO STEMMER

Proposta de tese apresentada à coordenação do curso de pós-graduação em engenharia elétrica da Universidade Federal de Santa Catarina, como parte dos requisitos necessários a defesa do exame de qualificação do curso de doutorado.

UFSC, Florianópolis - SC, Brasil

[†]<http://research.eeol.org/>

Resumo

Na robótica móvel, os problemas fundamentais são: desviar de obstáculos e seguir pela melhor trajetória dentro do ambiente do robô até a sua meta. A identificação e desvio dos obstáculos é fundamental para a segurança do robô. Além disso, é através do conhecimento desses obstáculos que o veículo pode planejar a sua trajetória.

Neste cenário, a visão computacional está tornando-se a base para o desenvolvimento de sensores que fornecem informações para o mapeamento do ambiente do robô e auxílio no planejamento da sua trajetória.

Em oposição a trabalhos deste tipo que utilizam, além do sistema de visão, outros sensores, um pre-conhecimento ou alterações no ambiente do robô, propõe-se um sistema de visão que identifique os obstáculos perigosos para robô e monte um mapa 2D com a localização desses obstáculos dentro do seu ambiente de visão.

Para estimar-se a localização dos obstáculos é necessário o uso de visão estereoscópica. Em sistemas deste tipo, é necessário que o conjunto de câmeras esteja calibrado. Esta calibração envolve o conhecimento das matrizes de projeção das câmeras. Essas matrizes concentram os parâmetros intrínsecos e extrínsecos do par estereoscópico e para serem estimadas é necessário um conjunto de correspondências entre pontos das duas imagens.

Quando o sistema não está calibrado, estabelecer tais correspondências é uma tarefa extremamente complexa. Normalmente, para facilitar a tarefa de calibração, faz-se uso de gabaritos. Contudo, neste caso, a tarefa de calibração não pode ser realizada durante a operação normal do robô. Por isso, neste trabalho, propõe-se uma nova técnica de calibração de um par estereoscópico que utiliza um gabarito de apenas uma única dimensão. Assim, esse gabarito pode facilmente ser manipulado pelo próprio veículo autônomo.

Após a calibração das câmeras, correspondências de pontos devem ser estabelecidas para identificação e localização dos obstáculos. Contudo, essa tarefa pode ser facilitada se a matriz fundamental que relaciona as duas imagens for conhecida. Em vista disso, propõe-se aperfeiçoamentos no cálculo desta matriz e em algoritmos de correspondência de pontos específicos para a identificação de objetos no campo de visão do robô.

Por último, com os objetos identificados, o sistema de visão criará um mapa 2D com a localização desses objetos no ambiente do robô.

Abstract

Fundamentals problems in the mobile robotic are: to turn aside of the obstacles and to go away through the best trajectory inside of robot environment until its goal. The identification and bypass of obstacles is fundamental to the security of robot. Besides, is through the knowledge of these obstacles that the mobile can to plan it trajectory.

In this scenery, the computational vision is making the base to development of sensors that give information to the mapping of robot environment and aim in the planning of its trajectory.

In opposition of papers this kind that use, besides the vision system, others sensors (ultra-sound, lasers...), a pre-knowledge or alterations of robot environment, is propose a vision system that identify the dangers obstacles to robot and make the 2D map with localization of these obstacles inside of the environment of vision.

To estimate the localization of obstacles is necessary the use of stereoscopic vision. In systems that kind, is necessary that the set of cameras have been calibrated. That calibration involves the knowledge of matrixes of cameras projection. Those matrixes concentrate the intrinsic and extrinsic parameters of stereoscopic couple and to be estimated is necessary a set of correspondence between two images.

When the system is not calibrated, to establish this correspondence is a extremely complex task. Often, to facilitate the task of calibration, is necessary to use a known pattern. However, in this case, the task of calibration can not be done during the normal operation of robot. Therefore, in this work, a new stereo vision system calibration technique is proposed. In this technique, the calibration is accomplished through an apparatus of calibration of a single dimension. This way, the apparatus easily can be manipulated by the own vehicle autonomous.

After the calibration of cameras, points correspondences must be established to identification and localization of obstacles. However, that task can be facilitated if the fundamental matrix that relates two images will be known. In function of that, we propose improvement in the calcule of that matrix and algorithms of correspondence of specific points to identification of objects in campo de visão of the robot.

Finally, with the objects identified, the system of vision will make a 2D map with the location those objects in the robot environment.

Conteúdo

Lista de Figuras	ii
Lista de Símbolos e Abreviaturas	v
1 Introdução	1
Notação	7
2 Fundamentos da Estereoscopia	8
2.1 Modelagem da Câmera	8
2.1.1 Aperfeiçoamento do Modelo <i>Pinhole</i>	11
2.2 Calibração da Câmera	11
2.2.1 Calibração com oito ou mais mapeamentos	13
2.2.2 Gabarito de Calibração	14
2.3 Estereoscopia	15
2.3.1 Reconstrução 3D	16
2.4 Emparelhamento de Pontos	16
2.4.1 Detecção dos Pontos de Interesse	18
2.4.2 Restrições	22
2.5 Emparelhamento de pontos sem calibração	28
3 Estereoscopia Através da Matriz Fundamental	34
3.1 Matriz Fundamental	34
3.2 Determinação Algébrica da Matriz Fundamental	36
3.3 Estimação da Matriz Fundamental	37
3.3.1 Solução com apenas sete emparelhamentos	38
3.3.2 Método dos 8 pontos	39
3.3.3 Método Não-Linear	40
3.3.4 Método do Paralaxe Virtual	42

3.3.5	Métodos Robustos	45
3.4	Restrição epipolar para emparelhamento de pontos	48
3.5	Calibração Automática	49
3.6	Calibração relativa de câmeras	52
3.6.1	Por fatoração da matriz fundamental	52
3.6.2	Relativa a uma base projetiva	52
3.6.3	Estimação da transformação projetiva	56
3.7	Retificação das Imagens	57
3.8	Emparelhamento em imagens retificadas	61
3.9	Triangulação	65
3.9.1	Método linear	65
3.9.2	Método ótimo	66
4	Contribuições para Estereoscopia Aplicada à Robótica Móvel	69
4.1	Método rápido para emparelhamento de pontos sem calibração	69
4.2	Método robusto para o cálculo da matriz fundamental	72
4.3	Método flexível para calibração de um par estereoscópico	73
4.3.1	Fundamentação teórica do problema	74
4.3.2	Proposta para solução do problema	76
4.4	Estereoscopia aplicada a detecção de obstáculos	77
4.4.1	Calibração do sistema de visão do robô	78
4.4.2	Segmentação dos obstáculos	79
4.4.3	Geração do mapa de obstáculos	81
4.5	Metodologia	81
4.6	Cronograma	82
	Bibliografia	85

Lista de Figuras

2.1	Esboço de uma câmera <i>pinhole</i>	9
2.2	Geometria do modelo <i>pinhole</i> . Observe que, por simplicidade, o plano de imagem foi representado na frente do centro da câmera.	9
2.3	Gabarito utilizado na calibração de câmeras.	15
2.4	Projeção no plano de imagem de duas câmeras de um ponto M no espaço 3D.	17
2.5	As direções das variâncias calculadas pelo operador de Moravec. A menor dentre essas variâncias é o valor do operador de Moravec.	19
2.6	(a) e (b) são imagens capturadas por um par estereoscópico e com os pontos de interesse marcados com o operador de Moravec. (c) e (d) são as mesmas imagens com os pontos de interesse marcados através do operador de SUSAN.	21
2.7	Conjunto de pontos interessantes com o mapa de distâncias mínimas esboçado.	25
2.8	Par (m_i, m'_j) candidato à correspondência com vizinhos dentro de uma janela R	25
2.9	Falha na restrição de unicidade.	27
2.10	(a) e (b) possuem correspondências marcadas pelo método de SCOTT e LONGUET-HIGGINS.	30
2.11	(a) e (b) formam um par de imagens com correspondências marcadas por relaxação. O par de imagens (c) e (d) é outro exemplo de correspondência por relaxação.	33
3.1	Geometria Epipolar.	35
3.2	Geometria epipolar e o paralaxe virtual.	43

3.3	(a) e (b) formam um par de imagens com emparelhamentos marcados por relaxação e restrição epipolar. O par de imagens (c) e (d) é outro exemplo de correspondência pela mesma técnica.	50
3.4	Quatro pontos co-planares M_1, M_2, \dots, M_4 formam um quadrilátero, cujas diagonais se intersectam nos pontos B_1, B_2 e B_3	54
3.5	(a) as imagens originais de um par estereoscópico e (b) a reconstrução 3D das imagens através da calibração relativa.	56
3.6	Dois exemplos de imagens com linhas epipolares esboçadas.	58
3.7	(a) Imagens originais com linhas epipolares oblíquas e não-paralelas. (b) As mesmas imagens retificadas, mas com distorção.	60
3.8	Emparelhamento de pontos em imagens retificadas.	61
3.9	Restrição esquerda-direita. O ponto b não tem um correspondente na imagem I' , mas o algoritmo de emparelhamento pode formar dois pares (b, c') e (c, c') . Contudo, tomando I' como referência, apenas o par (c, c') é formado.	62
3.10	Configuração de múltiplas janelas para emparelhamento por semelhança. A correlação é calculada dentro das áreas escuras.	63
4.1	Imagens com emparelhamento encontrados pelo método de SCOTT e LONGUET-HIGGINS, mas com as modificações propostas neste trabalho.	71
4.2	Configuração proposta de múltiplas janelas para emparelhamento de pontos sem calibração.	72
4.3	Exemplo do gabarito de uma única dimensão utilizado na técnica de calibração proposta. Por simplicidade, apenas um dos planos de imagem foi esboçado.	75
4.4	Classificação das regiões pelo veículo móvel.	78
4.5	Parâmetros a serem estimados para classificação dos objetos no ambiente do robô móvel.	79
4.6	Um típico gráfico de $SAD \times$ disparidade na proximidade das bordas dos objetos de uma imagem.	81
4.7	(a) Uma das famosas imagens “ <i>Head and Lamp</i> ” da <i>University of Tsukuba</i> e (b) o seu mapa ideal de disparidades.	83

Lista de Símbolos e Abreviaturas

A	Matriz com os parâmetros intrínsecos da câmera 1
A'	Matriz com os parâmetros intrínsecos da câmera 2
C ou C	Coordenadas do ponto central da câmera 1
C' ou C'	Coordenadas do ponto central da câmera 2
\tilde{C} ou \tilde{C}	Coordenadas homogêneas do ponto central da câmera 1
\tilde{C}' ou \tilde{C}'	Coordenadas homogêneas do ponto central da câmera 2
$d(x, y)$	Distância euclidiana do ponto x ao ponto y ou do ponto x à reta y
e ou e	Epipolo $\mathbf{e} = [e_u, e_v]^T$ do plano I
e' ou e'	Epipolo $\mathbf{e}' = [e'_u, e'_v]^T$ do plano I'
\tilde{e} ou $\tilde{\mathbf{e}}$	Coordenadas homeogêneas do epipolo $\tilde{\mathbf{e}} = [e_u, e_v, e_t]^T$ do plano I
\tilde{e}' ou $\tilde{\mathbf{e}'}$	Coordenadas homeogêneas do epipolo $\tilde{\mathbf{e}'} = [e'_u, e'_v, e'_t]^T$ do plano I'
F	A matriz fundamental que relaciona pontos em I com retas em I'
I	Plano de imagem da câmera 1
I'	Plano de imagem da câmera 2
l ou l	Linha epipolar de m em I'
l' ou l'	Linha epipolar de m' em I
M ou M	Coord. de um ponto 3D no sistema de coord. do ambiente
\tilde{M} ou $\tilde{\mathbf{M}}$	Coord. homogêneas de um ponto 3D no sistema de coord. do ambiente
m ou m	Coord. de um ponto $\mathbf{m} = [u, v]^T$ no plano da imagem I
m' ou m'	Coord. de um ponto $\mathbf{m}' = [u', v']^T$ no plano da imagem I'
\tilde{m} ou $\tilde{\mathbf{m}}$	Coord. homogêneas de $\tilde{\mathbf{m}} = [u, v, t]^T$ no plano da imagem I
\tilde{m}' ou $\tilde{\mathbf{m}'}$	Coord. homogêneas de $\tilde{\mathbf{m}'} = [u', v', t']^T$ no plano da imagem I'
P	Matriz de projeção da câmera 1
P'	Matriz de projeção da câmera 2
R	Rotação de I em relação à origem do sist. de coord. do ambiente.
R'	Rotação de I' em relação à origem do sist. de coord. do ambiente.
-RC	Translação de I em relação à origem do sist. de coord. do ambiente.
-R'C'	Translação de I' em relação à origem do sist. de coord. do ambiente.
SVD	Decomposição do valor singular (<i>Singular Value Decomposition</i>)

Capítulo 1

Introdução

A visão talvez seja o sentido mais importante do ser humano. É através deste sentido que são realizadas diversas tarefas do dia-a-dia, como por exemplo, identificar pessoas, locais e objetos, estimar distâncias, velocidade e altura. Além disso, em diversas ocasiões, a visão pode perceber a presença de perigo bem antes dos outros sentidos. Todas essas tarefas são bem mais complexas sem o uso da visão. Em vista disto, para a solução de problemas em diversas áreas do conhecimento humano, é natural o uso de percepção visual, ou seja, câmeras e sistemas de visão computacional.

Pode-se definir visão computacional como sendo o conjunto de algoritmos através dos quais sistemas baseado em computadores podem extrair informações (dentro de um contexto específico) dos pixels que compõem uma imagem. A grande maioria das técnicas usadas em visão computacional ainda são objetos de pesquisa, pois esta encontra-se em seu estágio inicial, ou seja, o sistema que “vê” está muito distante de ser implementado. Por isso, os sistemas deste tipo que existem hoje são construídos para aplicações específicas, supõem condições especiais de funcionamento (iluminação, tipo de ambiente, faixa de cor etc.) e aplicam-se apenas a um conjunto limitado de situações (fornecendo resultados ruins para todos os outros).

Na prática, os algoritmos de visão computacional existentes são responsáveis apenas por uma pequena parte do processo de percepção visual. Desse modo, um conjunto desses algoritmos precisa ser aplicado em cascata para formar sistemas mais complexos. Mesmo assim, apesar das limitações, conhecendo-se bem as técnicas e o problema a ser solucionado, a visão computacional é aplicada com sucesso em diversas áreas do conhecimento (em alguns casos com resultados superiores aos obtidos com outras metodologias). Uma dessas área é a robótica.

A robótica é uma área fascinante que envolve grupos de trabalhos multidisciplinares.

Uma de suas vertentes é a robótica móvel, que lida com a concepção de robôs capazes de locomover-se. As aplicações de robôs móveis são as mais distintas possíveis. Dentre elas pode-se citar [8]:

- Exploração de ambientes desconhecidos ou hostis para o ser humano. Nesta linha de aplicação existem robôs especialmente projetados, por exemplo, para a exploração de vulcões. Outra aplicação importante é a manutenção de plantas nucleares, onde o veículo pode transitar em locais com elevados níveis de radiação para realizar limpeza, inspeção ou reparos em casos de acidente.
- Sistemas de Manufatura Flexível (SMF). O grau de automação das fábricas tem avançado a passos largos e o principal uso de robôs móveis tem sido no transporte de peças e produtos entre máquinas, células de manufatura ou setores de produção. As esteiras rolantes apropriadas para linhas de produção contínua, nas quais a seqüência de máquinas é sempre a mesma, têm sido substituídas por veículos automáticos. Isto permite uma rápida resposta do meio de produção às mudanças na seqüência de montagem, resultando em uma maior flexibilidade.
- Sistemas automáticos de transporte. Veículos automáticos ainda são objetos de pesquisa nesta área, mas se apresentam como uma alternativa de melhoria tanto para o tráfego caótico das grandes cidades como para a redução dos índices de acidentes de trânsito. Estão em funcionamento sistemas de informação de tráfego que permitem que o motorista obtenha a rota com menor tráfego e que o leva ao seu destino em menor tempo, através de um computador de bordo embutido no painel do veículo. E tem-se investido em pesquisas de técnicas de guiagem automática [7]. Tais sistemas estão detendo-se ao uso de marcas artificiais na estrada ou em suas laterais, utilizadas como referência pelo veículo para manter-se na pista, GPS (*Global Positioning System*) para a determinação da posição e instrumentos para a detecção de obstáculos.

Dois termos têm sido bastante empregados na classificação de robôs móveis: veículos automáticos e veículos autônomos.

Veículos Automáticos são veículos que utilizam algum método de guiagem que altera o ambiente de trabalho, tais como, guias, marcadores artificiais implantados no ambiente (códigos de barra, figuras geométricas incomuns ao ambiente, etc), sendo incapazes de tomar decisões sobre rotas que não tenham sido previamente planejadas.

O ambiente de operação do veículo deve ser modificado de acordo com o tipo de guia ou marcador empregado.

De uma forma geral, os veículos automáticos possuem uma locomoção bastante limitada. Em compensação, podem utilizar sensores simples que necessitam de pouco processamento. Como exemplo deste tipo de veículo, pode-se citar o trabalho desenvolvido por BORGES [8]. Neste trabalho, foi utilizado um sensor óptico constituído de uma matriz de receptores e transmissores infravermelho. Com esse sensor, determina-se a posição do veículo em relação a uma faixa-guia localizada no piso do ambiente do robô. Com isso, a ação de controle necessária para mantê-lo na sua rota é determinada.

Veículos Autônomos são capazes de realizar o planejamento de sua rota com base em análises de dados provenientes de seus sistemas sensoriais. Possuem uma arquitetura hierárquica de controle e na maioria das vezes são projetados seguindo conceitos da teoria de agentes autônomos. Entretanto outras ferramentas, como redes neurais, lógica nebulosa e algoritmos evolutivos, têm sido aplicadas também em sua arquitetura, principalmente no auxílio da tomada de decisões, fusão de dados multisensoriais, aquisição de conhecimento e planejamento de navegação autônoma.

Em algumas aplicações da robótica móvel, por exemplo, em sistemas multiagentes como a *RoboCup* [38], inicialmente, é necessário determinar-se, segundo um sistema de coordenadas qualquer, a posição do veículo, do seu “alvo” (local para o qual ele deve mover-se) e, em alguns casos, dos obstáculos inseridos no seu ambiente. Em seguida, ações devem ser realizadas para que o veículo chegue ao seu destino. Neste caso, a dificuldade está no fato do ambiente de um veículo autônomo, geralmente, ser desconhecido e/ou dinâmico. Por isso, para que o veículo possa atingir sua meta apesar das dificuldades deste ambiente, ele deve ser dotado de sensores robustos capazes de fornecer as informações de posição necessárias.

Em alguns trabalhos, como por exemplo, [64] e [16], sensores de baixo custo, como ultra-som, odômetro e bússola digital, são utilizados no controle de trajetórias de robôs móveis autônomos. Contudo, esses sistemas são incapazes de reconhecer obstáculos com precisão e apresentam elevados erros de posição [71].

Nos animais, o principal sentido é a visão. Ele é o principal responsável na estimação de posição e percepção dos obstáculos. Por isso, é natural que a visão computacional seja muito utilizada também com veículos autônomos. Exemplos dessa metodologia podem ser encontrados em [14] e [13]. Nestes trabalhos, são realizadas estimações de distâncias através da medição dos tamanhos relativos de marcas-guias em imagens capturadas de uma câmera. Em [66], um sistema de visão é utilizado para extrair do

ambiente do veículo, através da transformada de Hough e interpolação por mínimos quadrados, guias para sua navegação. Já em [22], um sistema de visão (em conjunto com sensores de ultra-som) é utilizado para reconhecer obstáculos que estão a uma distância de 1 m de um veículo.

No processo de formação da imagem, basicamente, a câmera projeta um ponto no espaço em uma plano imaginário chamado de plano de imagem da câmera. Todo esse processo é descrito por uma matriz chamada matriz de projeção da câmera.

A matriz de projeção é uma característica particular de cada câmera e depende do processo de fabricação. Além disso, essa matriz pode ser alterada durante a execução normal do sistema, por exemplo, após uma alteração na distância focal da câmera, e pode variar de acordo com as condições do ambiente (variações térmicas e mecânicas) [21]. Por isso, a única maneira de determinar a matriz de projeção corretamente é através de calibração.

No processo de calibração da câmera, as coordenadas de um conjunto de pontos no espaço, assim como as coordenadas das projeções no plano da imagem desses pontos, devem ser conhecidas. Uma maneira de obter-se essas coordenadas é através de um gabarito de calibração. Normalmente, este gabarito é composto por dois planos ortogonais com padrões contrastantes nas suas fases. Então, pode-se considerar a origem do sistema de coordenadas do ambiente como sendo um dos cantos do gabarito. Assim, as coordenadas dos padrões do gabarito e suas respectivas projeções no plano da imagem podem ser conhecidos com grande exatidão [86].

Os métodos baseados no gabarito de calibração são bastante exatos, mas possuem duas limitações importantes [21]: Primeiro, nem sempre um objeto que sirva como gabarito está disponível. Neste caso, é impossível realizar a calibração. Segundo, se os parâmetros da câmera forem alterados durante o funcionamento normal do sistema, a inspeção visual deve ser interrompida e o gabarito (novamente) deve ser posto no campo de visão da câmera.

Em vista disso, métodos de calibração automática de câmeras foram desenvolvidos. Neste caso, a calibração não necessita de gabarito e pode ser realizada durante a operação normal de inspeção visual, enquanto a câmera realiza uma série de movimentos em relação a uma cena rígida. Contudo, as primeiras técnicas de calibração automática eram muito susceptíveis à ruídos e não têm uso prático [99]. Felizmente, em 1997, FAUGERAS *et al* [21] propuseram modificações que aumentaram a exatidão destas e as tornaram mais estáveis.

O uso de visão monocular em veículos autônomos tem limitações. Por exemplo,

com a matriz de projeção conhecida, ou seja, com a câmera calibrada, à partir de um ponto no plano da imagem, é possível saber apenas a reta a qual pertence o ponto 3D correspondente [31]. Com isso, não se pode medir a distância deste ponto 3D à outro ponto no espaço e nem, caso este mova-se, a sua velocidade. Em tais aplicações, o uso de visão monocular faz necessária a obtenção de mais de uma imagem de uma mesma cena a partir de pontos de vistas diferentes [13, 14]. Mesmo assim, além de necessitar de mais tempo de processamento, essa técnica tende a produzir desvios na trajetória dos veículos autônomos [41]. Por isso, nas aplicações em que tais informações são importantes, utilizando-se duas câmeras e através de um processo chamado triangulação, pode-se utilizar estereoscopia (duas imagens de uma mesma cena fornecidas por um par de câmeras) para estimá-las [32].

Após os trabalhos independentes de FAUGERAS [20] e HARTLEY *et al* [30], uma técnica de calibração de um par estereoscópico conhecida como *weak calibration* tornou-se popular. Nesta, para calibração de um par estereoscópico, necessita-se apenas de uma imagem de cada câmera e, no mínimo, cinco mapeamentos de pontos quaisquer entre elas. Assim, a calibração pode ser feita de forma bem menos complexa. Contudo, perde-se algumas informações importantes, por exemplo, comprimentos e ângulos.

O processo de triangulação necessita de duas imagens de uma mesma cena, mas de pontos de vistas distintos. Assim, um ponto 3D tem projeções no plano de ambas as imagens. Então, conhecendo-se a localização dessas projeções no plano das imagens, pode-se estimar a localização do ponto 3D. Por isso, o processo de emparelhamento de pontos é um dos problemas fundamentais da estereoscopia [4, 33, 53, 54, 65, 67, 74, 89, 91, 96].

Se o par de câmeras não está calibrado, não existe nenhuma restrição geométrica para estabelecer os emparelhamentos, ou seja, de uma forma geral, dado um ponto no plano de imagem de uma câmera, todos os pontos pertencentes ao plano de imagem da outra câmera que compõe o par são candidatos à emparelhamento [96].

O emparelhamento de pontos é uma tarefa extremamente complexa e existem várias técnicas para realizá-la. Contudo, basicamente, para realizar emparelhamento de pontos em imagens sem calibração, determina-se, independentemente, pontos de interesse em ambas as imagens e, em seguida, tenta-se encontrar associações de pontos [65].

Com os pontos de interesse determinados e segmentados, para que o problema de emparelhamento de pontos tenha solução, devem existir restrições que sirvam de guia para a resolução da tarefa. Estas restrições são encontradas no sistema de visão humano, nas propriedades das câmeras e sensores ópticos, e na geometria. Para en-

contrar os possíveis emparelhamentos, essas restrições devem ser aplicadas ao conjunto de pontos de interesse de duas imagens de forma seqüencial ou paralela. Desta forma, espera-se que, ao final do processo, a maior quantidade possível de emparelhamentos tenha sido encontrada corretamente [53].

Com alguns emparelhamentos encontrados, a matriz fundamental pode ser calculada. Esta matriz relaciona um ponto no plano de uma imagem ao seu ponto correspondente na outra imagem [29, 31, 46, 48, 85, 93, 98]. Contudo, a matriz fundamental não indica a localização exata do ponto, mas apenas a reta (no plano de imagem) a qual este ponto pertence. A esta reta, dá-se o nome de reta epipolar.

O problema em calcular-se a matriz fundamental é que, normalmente, a sua estimação é bastante suscetível à ruídos. Por isso, devido a instabilidade da técnica utilizada para segmentar os pontos de interesse ou a erros cometidos no processo de emparelhamento, a matriz fundamental encontrada não é muito confiável [85, 93]. Em vista disso, vários métodos robustos, baseados no algoritmo RANSAC, LMedS, ou M-estimadores, são empregados na detecção e eliminação do ruído. A matriz fundamental calculada desta forma apresenta resultados bem melhores.

A matriz fundamental transforma o processo de emparelhamento de pontos de um problema 2D em um de apenas uma dimensão. Além disso, se as linhas epipolares forem paralelas entre si e ao eixo das abscissas, o correspondente de um ponto pode ser encontrado de forma menos complexa, por exemplo, simplesmente calculando uma função que forneça o grau de semelhança entre os dois pontos [23, 25, 26, 43].

Dessa forma, com as câmeras calibradas e a matriz fundamental conhecida, pode-se utilizar estereoscopia para resolver os problemas fundamentais da robótica móvel: desviar de obstáculos e seguir pela melhor trajetória dentro do ambiente do robô até a sua meta. A localização e desvio dos obstáculos é fundamental para a segurança do robô. Além disso, é através do conhecimento desses obstáculos que o veículo pode planejar a sua trajetória.

Em oposição a trabalhos que utilizam, além do sistema de visão, outros sensores (ultra-som, lasers etc.), um pre-conhecimento ou alterações no ambiente do robô, propõe-se um sistema de visão que identifique os obstáculos perigosos para robô e monte um mapa 2D com a localização desses obstáculos dentro do seu ambiente de visão.

Pretende-se examinar o uso de uma nova técnica de calibração de uma par estereoscópica que utiliza a matriz fundamental e um gabarito de apenas uma dimensão. Além disso, propõe-se aperfeiçoamentos no cálculo da matriz fundamental e em algoritmos

de emparelhamento de pontos específicos para a localização de obstáculos no campo de visão do robô. Dessa forma, com os obstáculos localizados, o sistema de visão criará um mapa 2D com a localização desses objetos no ambiente do robô.

Para expor o trabalho, o conteúdo do texto está dividido da seguinte forma:

No capítulo 2, têm-se uma visão geral dos problemas intrínsecos da estereoscopia e algumas idéias de como solucioná-los. Em particular, discute-se a necessidade de estabelecer-se emparelhamento entre pontos de duas imagens estereoscópicas e a complexidade desta tarefa quando o sistema não está calibrado.

No capítulo 3, apresenta-se a matriz fundamental e as dificuldades embutidas no seu processo de estimação. Esta matriz relaciona duas imagens estereoscópicas e torna o processo de visão bem menos complexo. Além disso, técnicas de calibração de câmeras e de retificação de imagens, além de ferramentas que permitem localizar pontos no ambiente 3D, são discutidas.

Por último, no capítulo 4 são apresentadas as contribuições pretendidas para este trabalho no âmbito da robótica móvel.

Notação

No decorrer do texto, utiliza-se uma notação, na qual 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, $\tilde{\mathbf{x}} = [x, y, \dots, 1]^T$. Também, é usada a prática notação $\mathbf{A}^{-T} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ para toda matriz quadrada inversível. Finalmente, uma reta l no plano da imagem que passa através do ponto $\mathbf{m} = [u, v]^T$ é descrita pela equação $au + bv + c = 0$. Então, se $\mathbf{l} = [a, b, c]^T$, pode-se escrever $\mathbf{l}^T \tilde{\mathbf{m}} = 0$ ou $\tilde{\mathbf{m}}^T \mathbf{l} = 0$.

Capítulo 2

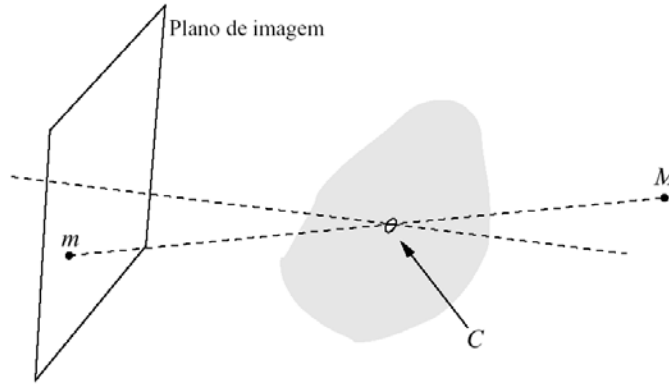
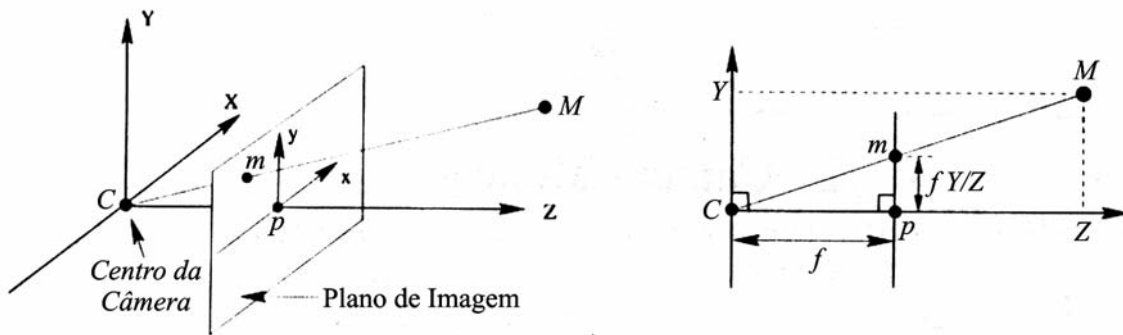
Fundamentos da Estereoscopia

A imagem capturada por uma câmera é uma representação em duas dimensões do espaço 3D. Essa representação é feita projetando-se pontos do espaço em um plano conhecido como plano de imagem. Neste capítulo, é dado um modelo matemático bem detalhado de uma câmera. Esse modelo relaciona um ponto no plano da câmera com uma reta no espaço 3D. Em seguida, mostra-se que a estereoscopia é baseada na localização de um ponto 3D no espaço através das projeções desde ponto em duas imagens e na intersecção das retas associadas a essas duas projeções. Além disso, a importância do emparelhamento entre pontos de duas imagens para a estereoscopia é apresentada, assim como a complexidade de estabelecer tais emparelhamentos. Por último, no final do capítulo, apresenta-se algumas técnicas de emparelhamento de pontos em imagens sem calibração.

2.1 Modelagem da Câmera

O modelo de câmera mais utilizado é o *pinhole* [19]. Um esboço dele é apresentado na figura 2.1. Neste modelo, um raio de luz proveniente de um ponto M no espaço passa através de um orifício C e incide em um plano no ponto m . O plano de incidência dos raios é chamado de plano de imagem e o orifício C é chamado de centro da câmera. A distância do plano de imagem ao centro da câmera é um parâmetro intrínseco da câmera e é conhecido como distância focal, geralmente, representado por f .

Agora, considere a figura 2.2. Nesta, por simplicidade, o plano de imagem é apresentado na frente do centro da câmera. Além disso, a câmera é apresentada em um sistema de coordenadas. A origem deste sistema é o centro da câmera. O plano de imagem também tem um sistema de coordenadas 2D com origem no ponto p , geral-

Figura 2.1: Esboço de uma câmera *pinhole*.Figura 2.2: Geometria do modelo *pinhole*. Observe que, por simplicidade, o plano de imagem foi representado na frente do centro da câmera.

mente, chamado de ponto principal. Por semelhança de triângulos, um cálculo simples mostra que o ponto $M = (X, Y, Z)$ no espaço é mapeado no ponto $m = (fX/Z, fY/Z)$ no plano da imagem. Se for considerado as coordenadas homogêneas do ponto M e m , respectivamente, $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$ e $\tilde{\mathbf{m}} = [fX, fY, Z]^T$, pode-se escrever

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.1)$$

Como dito, a equação 2.1 supõe que a origem do sistema de coordenadas do plano de imagem é o ponto principal. Contudo, essa suposição nem sempre é correta. Por isso, é mais correto mapear o ponto 3D $M = (X, Y, Z)$ no ponto $m = (fX/Z + p_x, fY/Z + p_y)$ no plano da imagem, onde (p_x, p_y) são as coordenadas do ponto principal. Assim, a

equação 2.1 pode ser re-escrita como

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2.2)$$

Em geral, é mais conveniente expressar pontos no espaço em termos de um ponto de referência no ambiente. As coordenadas do ambiente estão relacionadas com as coordenadas do centro da câmera por uma rotação e uma translação. Assim, se $\tilde{\mathbf{X}}$ é um ponto no sistema de coordenadas do ambiente, este ponto é mapeado no ponto $\tilde{\mathbf{X}}_c$ no sistema de coordenadas da câmera através da equação

$$\tilde{\mathbf{X}}_c = \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ \mathbf{0}^T & 1 \end{bmatrix} \tilde{\mathbf{X}}, \quad (2.3)$$

onde \mathbf{C} são as coordenadas do centro da câmera no sistema de coordenadas do ambiente, \mathbf{R} é uma matriz 3x3 que representa a rotação da câmera em relação às coordenadas do ambiente e $\mathbf{0}$ é um vetor coluna nulo 3x1.

Finalmente, substituindo (2.3) em (2.2) e considerando $\tilde{\mathbf{m}} = [u, v]^T$, tem-se a equação geral do modelo *pinhole*, ou seja,

$$\begin{aligned} \tilde{\mathbf{m}} &= s_m \begin{bmatrix} \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ \mathbf{0}^T & 1 \end{bmatrix} \tilde{\mathbf{M}} \\ &= s_m \mathbf{A} [\mathbf{R} \mid -\mathbf{RC}] \tilde{\mathbf{M}}, \end{aligned} \quad (2.4)$$

onde

$$\mathbf{A} = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

e s_m é uma constante desconhecida, mas necessária para indicar que os dois lados da equação podem diferir por uma constante.

A matriz \mathbf{A} tem três graus de liberdade e concentra todos os parâmetros intrínsecos da câmera. Já as matrizes \mathbf{R} e $-\mathbf{RC}$ têm um total de seis graus de liberdade [31] e agrupam os parâmetros extrínsecos.

Frequentemente, define-se a matriz de projeção do modelo *pinhole* igual a

$$\mathbf{P} = s_m \mathbf{A} [\mathbf{R} \mid -\mathbf{RC}]. \quad (2.6)$$

Neste caso, a equação (2.4) torna-se

$$\tilde{\mathbf{m}} = s_m \mathbf{P} \tilde{\mathbf{M}}. \quad (2.7)$$

2.1.1 Aperfeiçoamento do Modelo *Pinhole*

O modelo *pinhole* assume que o sistema de coordenadas do plano da imagem tem escalas iguais nos eixos x e y. Contudo, isso nem sempre é verdade para as câmeras CCDs utilizadas na prática. Por isso, se forem utilizadas coordenadas em pixels, faz-se necessário o uso de fatores de escala diferentes para cada direção. Se, por exemplo, o número de pixels por unidades de distância for k_x e k_y para, respectivamente, as direções x e y, deve-se alterar a equação (2.5) para [31]

$$\mathbf{A} = \begin{bmatrix} f k_x & 0 & p_x \\ 0 & f k_y & p_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.8)$$

Além disso, em alguns casos especiais, a matriz de pixels de uma câmera pode ter as linhas e colunas não-perpendiculares. Neste caso, tem-se que introduzir um parâmetro extra para modelar essa característica. Assim, segundo [31], a equação (2.8) torna-se

$$\mathbf{A} = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.9)$$

onde s é o fator de não-perpendicularidade da matriz de pixels.

2.2 Calibração da Câmera

Como visto na seção 2.1, um ponto no espaço 3D pode ser mapeado em um ponto no plano da imagem através da equação (2.7). Contudo, antes de utilizar esta equação, a matriz de projeção da câmera [equação (2.6)] deve ser conhecida. Essa matriz é uma característica particular de cada câmera e depende do processo de fabricação. Além disso, essa matriz pode ser alterada durante a execução normal do sistema, por

exemplo, após uma alteração na distância focal da câmera, e pode variar de acordo com as condições do ambiente (variações térmicas e mecânicas) [21]. Por isso, a única maneira de determinar-se a matriz de projeção corretamente é através de calibração.

De (2.7), dado um mapeamento entre um ponto M_i no espaço e um ponto m_i no plano da imagem, tem-se a relação

$$\tilde{\mathbf{m}}_i = s_{m_i} \mathbf{P} \widetilde{\mathbf{M}}_i.$$

A equação anterior pode ser expressa em termos do produto cruzado, ou seja,

$$\tilde{\mathbf{m}}_i \times \mathbf{P} \widetilde{\mathbf{M}}_i = \mathbf{0}. \quad (2.10)$$

Agora, se a j -ésima linha da matriz \mathbf{P} for denotada por \mathbf{p}_j^T , pode-se escrever

$$\mathbf{P} \widetilde{\mathbf{M}}_i = \begin{bmatrix} \mathbf{p}_1^T \widetilde{\mathbf{M}}_i \\ \mathbf{p}_2^T \widetilde{\mathbf{M}}_i \\ \mathbf{p}_3^T \widetilde{\mathbf{M}}_i \end{bmatrix}. \quad (2.11)$$

Considerando, $\tilde{\mathbf{m}}_i = [u_i, v_i, t_i]^T$, de (2.10) e (2.11), tem-se

$$\tilde{\mathbf{m}}_i \times \mathbf{P} \widetilde{\mathbf{M}}_i = \begin{bmatrix} v_i \mathbf{p}_3^T \widetilde{\mathbf{M}}_i - t_i \mathbf{p}_2^T \widetilde{\mathbf{M}}_i \\ t_i \mathbf{p}_1^T \widetilde{\mathbf{M}}_i - u_i \mathbf{p}_3^T \widetilde{\mathbf{M}}_i \\ u_i \mathbf{p}_2^T \widetilde{\mathbf{M}}_i - v_i \mathbf{p}_1^T \widetilde{\mathbf{M}}_i \end{bmatrix} = \mathbf{0}. \quad (2.12)$$

Desde que $\mathbf{p}_j^T \widetilde{\mathbf{M}}_i = \widetilde{\mathbf{M}}_i^T \mathbf{p}_j$, $j = 1, \dots, 3$, podemos escrever (2.12) como

$$\begin{bmatrix} \mathbf{0}^T & -t_i \widetilde{\mathbf{M}}_i^T & v_i \widetilde{\mathbf{M}}_i^T \\ t_i \widetilde{\mathbf{M}}_i^T & \mathbf{0}^T & -u_i \widetilde{\mathbf{M}}_i^T \\ -v_i \widetilde{\mathbf{M}}_i^T & u_i \widetilde{\mathbf{M}}_i^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{bmatrix} = \mathbf{0}, \quad (2.13)$$

onde $\mathbf{0}$ é uma matriz coluna nula de três elementos.

A equação (2.13) fornece apenas duas equações independentes. Por isso, por exemplo, pode-se desprezar a última equação e re-escrever (2.13) como

$$\begin{bmatrix} \mathbf{0}^T & -t_i \widetilde{\mathbf{M}}_i^T & v_i \widetilde{\mathbf{M}}_i^T \\ t_i \widetilde{\mathbf{M}}_i^T & \mathbf{0}^T & -u_i \widetilde{\mathbf{M}}_i^T \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{bmatrix} = \mathbf{0}. \quad (2.14)$$

Desde que a matriz de projeção tem doze elementos e, desprezando o fator de

escala, onze graus de liberdade, são necessárias onze equações para calcular \mathbf{P} . Assim, precisa-se de seis mapeamentos $\tilde{\mathbf{m}}_i \leftrightarrow \tilde{\mathbf{M}}_i$ para resolver $\mathbf{U}_6 \mathbf{p} = 0$ (onde \mathbf{U}_6 é uma matriz 11×12) e calcular a matriz de projeção.

2.2.1 Calibração com oito ou mais mapeamentos

Na prática, dispõe-se de $i > 6$ mapeamentos $\tilde{\mathbf{m}}_i \leftrightarrow \tilde{\mathbf{M}}_i$. Além disso, deve haver ruído em alguns pares de mapeamentos. Esse ruído afeta significativamente a qualidade da matriz de projeção calculada. Assim, pode-se usar o método dos quadrados mínimos para estimar

$$\min_{\mathbf{p}} \|\mathbf{U}_i \mathbf{p}\|^2. \quad (2.15)$$

A solução trivial para a equação (2.15) é $\mathbf{p} = 0$. Evidentemente, essa solução não é interessante. Por isso, alguma restrição deve ser imposta. Já que não se conhece nada sobre os coeficientes de \mathbf{p} , pode-se impor uma restrição sobre a sua norma, por exemplo, $\|\mathbf{p}\| = 1$. Neste caso, a função a ser minimizada torna-se

$$(\mathbf{U}_i \mathbf{p})^T (\mathbf{U}_i \mathbf{p}) \equiv \mathbf{p}^T \mathbf{B} \mathbf{p} \text{ sujeito a } \|\mathbf{p}\| = 1, \quad (2.16)$$

onde $\mathbf{B} = \mathbf{U}_i^T \mathbf{U}_i$ é uma matriz simétrica. Como toda matriz simétrica, \mathbf{B} pode ser decomposta em $\mathbf{B} = \mathbf{U} \mathbf{E} \mathbf{U}^T$, com $\mathbf{E} = \text{diag}(v_1, \dots, v_m)$ e $\mathbf{U} = [\mathbf{e}_1, \dots, \mathbf{e}_m]$, onde v_i e \mathbf{e}_i são os autovalores e autovetores de \mathbf{B} , respectivamente.

Neste caso, o problema da equação (2.16) pode ser encarado como sendo descobrir q_1, \dots, q_m que minimize $\mathbf{p}^T \mathbf{B} \mathbf{p}$ sujeito a $q_1^2 + \dots + q_m^2 = 1$, onde $\mathbf{p} = q_1 \mathbf{e}_1 + \dots + q_m \mathbf{e}_m$, ou seja, $\mathbf{p}^T \mathbf{B} \mathbf{p} = q_1^2 v_1 + \dots + q_m^2 v_m$ [92].

Agora, a função sem restrição a ser minimizada torna-se

$$\mathcal{J} = \|\mathbf{U}_i \mathbf{p}\|^2 + \lambda(\|\mathbf{p}\|^2 - 1) = q_1^2 v_1 + \dots + q_m^2 v_m + \lambda(q_1^2 + \dots + q_m^2 - 1),$$

onde λ é o multiplicador de Lagrange. Derivando-se \mathcal{J} com respeito a q_1 até q_m e λ tem-se,

$$\begin{aligned} 2q_1 v_1 + 2q_1 \lambda &= 0 \\ 2q_2 v_2 + 2q_2 \lambda &= 0 \\ \dots & \\ 2q_m v_m + 2q_m \lambda &= 0 \\ q_1^2 + \dots + q_m^2 - 1 &= 0. \end{aligned}$$

Existem m soluções para este sistema de equações e a i -ésima solução é dada por

$$q_i = 1, q_j = 0 \text{ para } j = 1, \dots, m, \text{ exceto } i, \lambda = -v_i.$$

O valor de \mathcal{J} que corresponde a i -ésima solução é

$$\mathcal{J} = v_i.$$

Desde que se deseja minimizar \mathcal{J} , a solução para o problema original da equação (2.16) é o autovetor de \mathbf{B} que corresponde ao menor autovalor.

Como observado por HARTLEY [31], a exatidão da matriz de projeção estimada à partir da equação (2.16) é bastante melhorada se os pontos no plano da imagem são normalizados. Esta normalização deve ser tal que o centróide do conjunto de pontos fique na origem e os pontos $\tilde{\mathbf{m}}_i$ estão a uma distância média de $\sqrt{2}$ da origem. Além disso, os pontos no sistema de coordenadas também devem ser normalizados. Novamente, o centróide deve estar na origem do sistema de coordenadas e a distância média dos pontos $\tilde{\mathbf{M}}_i$ à origem deve ser $\sqrt{3}$. Com isso, na média, os pontos $\tilde{\mathbf{M}}_i$ teram coordenadas $\tilde{\mathbf{M}}_i = [1, 1, 1, 1]^T$.

2.2.2 Gabarito de Calibração

Na seção anterior, é suposta a existencia de um conjunto de mapeamentos $m_i \leftrightarrow M_i$ para realizar a estimação de \mathbf{P} . Uma maneira de obter-se esses mapeamentos é através de um gabarito de calibração. Normalmente, este gabarito é composto por dois planos ortogonais com padrões contrariantes nas suas fases. Um exemplo deste gabarito é apresentado na figura 2.3. Com este gabarito, pode-se considerar a origem do sistema de coordenadas do ambiente como sendo um dos cantos do gabarito. Assim, os vértices dos quadrados negros do gabarito (figura 2.3) são conhecidos com grande exatidão. Então, o erro cometido na medição dos pontos m_i pode ser expresso pela equação

$$\sum_i d(m_i, \hat{m}_i)^2,$$

onde m_i é o ponto medido e \hat{m}_i o ponto derivado de $\mathbf{P}\tilde{\mathbf{M}}_i$. Se for assumido que os erros de medição têm uma distribuição gaussiana, a solução de

$$\min_{\mathbf{P}} \sum_i d(m_i, \mathbf{P}\tilde{\mathbf{M}}_i)^2 \quad (2.17)$$

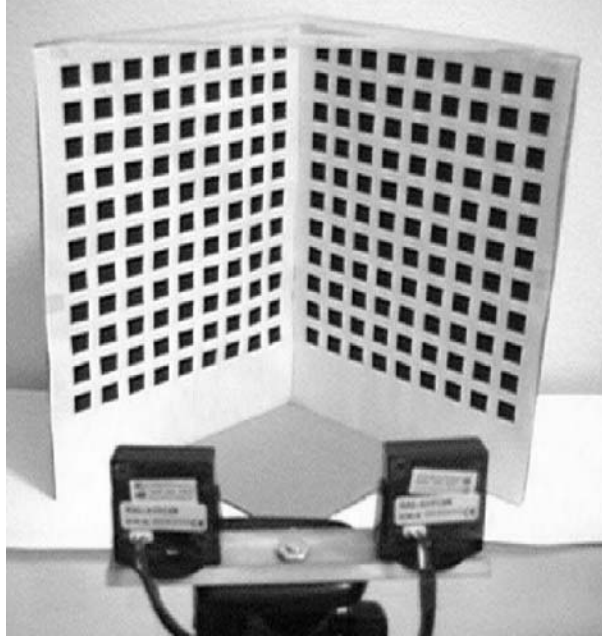


Figura 2.3: Gabarito utilizado na calibração de câmeras.

fornece uma estimaco melhor que a obtida atravs da equaco (2.15). Alm disso, pode-se impor restries a \mathbf{P} . Por exemplo, sabe-se que a matriz \mathbf{P} depende dos parmetros intrnicos da cmera representados pela equaco (2.8). Assim, pode-se normalizar $\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \end{bmatrix} \mathbf{W}$, onde \mathbf{W}  uma matriz 4×4 .

A desvantagem deste mtodo  que ele  computacionalmente mais pesado, pois necessita de uma estimaco inicial de \mathbf{P} [normalmente conseguida atravs da equaco (2.15)] e de um algoritmo de minimizao no-linear como, por exemplo, o Levenberg-Marquardt [68]. Neste trabalho, utiliza-se a implementaco da biblioteca Cephes [59] para o algoritmo Levenberg-Marquardt, disponvel *online* na internet.

2.3 Estereoscopia

Como apresentado na seo 2.1, a equaco geral do modelo *pinhole* [equaco (2.4)] relaciona um ponto M no espao com um ponto m no plano da imagem. Contudo, esta equaco est em coordenadas homogneas. Por isso, \tilde{m} pode diferir de \tilde{M} por uma constante. Geometricamente, isso significa que,  partir do ponto m , se conhece apenas a reta no espao a qual o ponto M pertence (figura 2.1). Por isso, atravs da equaco (2.4), alm de no se conhecer a localizao do ponto M , no se pode medir a distncia deste ponto a outro ponto no espao e nem, caso o ponto M mova-se, a sua velocidade. No entanto, nas aplicaes em que tais informaes so importantes,

em um processo chamado triangulação, pode-se utilizar estereoscopia (duas imagens fornecidas por um par de câmeras) para estimá-las.

2.3.1 Reconstrução 3D

No modelo *pinhole*, para cada ponto m no plano da imagem, existe uma reta no espaço 3D associada que contém o segmento \overline{MC} (figura 2.1). Esta reta é conhecida como raio óptico (do inglês *optical ray*). O ponto M pertencente a reta \overline{MC} pode ser expresso como sendo $\widetilde{\mathbf{M}} = \mathbf{P}^+ \widetilde{\mathbf{x}}$, onde \mathbf{P}^+ é a pseudo-inversa [31] da matrix \mathbf{P} , ou seja, $\mathbf{P}\mathbf{P}^+ = \mathbf{I}$, que implica $\mathbf{P}^+ = \mathbf{P}^T(\mathbf{P}\mathbf{P}^T)^{-1}$. Já o ponto C é encontrado pela equação $\mathbf{P}\widetilde{\mathbf{C}} = 0$. Dessa forma, um ponto W pertencente a reta \overline{MC} pode ser representado em forma paramétrica como [31]

$$\widetilde{\mathbf{W}} = \mathbf{P}^+ \widetilde{\mathbf{m}} + \alpha \widetilde{\mathbf{C}}, \quad (2.18)$$

onde $\alpha \in \mathbb{R}$.

Agora, considere o conjunto estereoscópico da figura 2.4. Neste, um ponto M no espaço projeta um ponto m no plano de imagem 1 e ponto m' no plano da imagem 2. Pela figura, vê-se que o ponto M pertence a ambos os raios ópticos associados aos pontos m e m' . Por isso, pode-se escrever

$$\mathbf{P}^+ \widetilde{\mathbf{m}} + \alpha \widetilde{\mathbf{C}} = \mathbf{P}'^+ \widetilde{\mathbf{m}}' + \alpha' \widetilde{\mathbf{C}}'. \quad (2.19)$$

A equação anterior fornece três equações e apenas duas incógnitas (α e α'). Dessa forma, através das imagens de um par estereoscópico, pode-se determinar as coordenadas 3D de um ponto M e assim calcular distâncias, comprimentos, velocidade etc.

2.4 Emparelhamento de Pontos

Se um objeto no espaço 3D é “visto” por duas câmeras diferentes, um conjunto de pontos pertencentes a este objeto tem projeções em ambos os planos de imagem. Por exemplo, considere M_i um ponto deste conjunto. Logo, M_i tem duas projeções m_i e m'_i , respectivamente, no plano da imagem 1 e no plano da imagem 2 (figura 2.4). Os pontos m_i e m'_i são ditos pontos correspondentes e formam um emparelhamento (m_i, m'_i) .

Na seção 2.3, mostra-se que as informações 3D podem ser recuperadas através de visão estereoscópica. Contudo, no processo de triangulação e mesmo na calibração das

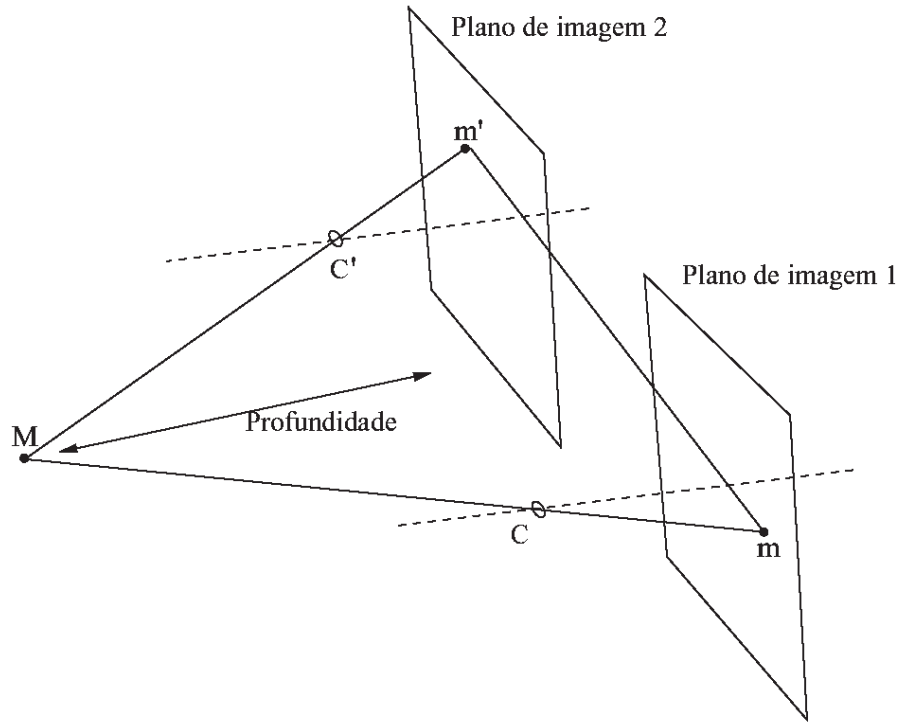


Figura 2.4: Projeção no plano de imagem de duas câmeras de um ponto M no espaço 3D.

câmeras é necessário conhecer-se alguns emparelhamentos. Além disso, muitos algoritmos utilizados em estereoscopia para, por exemplo, reconstrução 3D, reconhecimento de objetos e auto-calibração, assumem a existência de alguns emparelhamentos. Por isso, o processo de emparelhamento de pontos é um dos problemas fundamentais da estereoscopia.

Se o par de câmeras não está calibrado, não existe nenhuma restrição geométrica para estabelecer os emparelhamentos, ou seja, de uma forma geral, dado um ponto m no plano de imagem da câmara 1, todos os pontos pertencentes ao plano de imagem da câmara 2 são candidatos à emparelhamento. Além disso, o ponto m' que corresponde ao ponto m pode estar fora do plano de imagem da câmara 2 ou mesmo ocluído.

Uma ensaio completo sobre emparelhamento de pontos pode ser visto em [42, 89, 91]. Contudo, basicamente, existem duas abordagens para resolver-se o problema de emparelhamento de pontos [65]. Na primeira abordagem, determina-se pontos de interesse em uma imagem e, para cada um desses pontos, tenta-se encontrar o seu correspondente na segunda imagem. Na segunda abordagem, determina-se, independentemente, pontos de interesse em ambas as imagens e, em seguida, tenta-se encontrar associações de pontos. A primeira abordagem é muito complexa devido à dimensão do espaço

de busca combinacional intrínseco ao problema emparelhamento de pontos. Contudo, pode ser a única alternativa quando deseja-se encontrar o correspondente de um ponto específico.

Na segunda abordagem, para reduzir a complexidade, a maioria dos métodos cria uma medida de similaridade e assume que os pares candidatos à emparelhamento são independentes uns dos outros. Isto pode gerar ambiguidade. Em vista disso, alguns autores consideram o conjunto inteiro de pontos de interesse e todas as possibilidades possíveis de emparelhamento ao mesmo tempo [54, 65].

Na próxima seção, discute-se alguns métodos de identificar os pontos de interesse nas imagens e a influência da estabilidade e exatidão destes métodos no processo de estereoscopia.

2.4.1 Detecção dos Pontos de Interesse

Como visto, o processo de emparelhamento de pontos em imagens sem calibração é extremamente complexo. Para reduzir a complexidade, geralmente, é realizado um pré-processamento nas imagens para detectar e isolar caracteres primitivos que sejam de fácil identificação. Por serem bastante estáveis sob variações de iluminação e orientação das câmeras, os cantos (ou vértices) são muito utilizados em estereoscopia como caracteres primitivos [36].

Agora, suponha que um canto m_r na primeira imagem e o canto correspondente m'_r na segunda imagem tenha, respectivamente, coordenadas (u, v) e (u', v') . Na prática, um detector de cantos padrão irá detectar estes mesmos cantos em $m = (u + \Delta u, v + \Delta v)$ e $m' = (u' + \Delta u', v' + \Delta v')$, onde Δu , Δv , $\Delta u'$ e $\Delta v'$ representam deslocamentos em pixels e podem ser positivos ou negativos. Por isso, mesmo se o algoritmo de emparelhamento utilizado encontrar corretamente o par (m, m') , haverá erros de localização de alguns pixels no emparelhamento. Isto, por exemplo, faz com que as raios ópticos relacionados aos pontos m e m' não se toquem, ou seja, torna difícil a determinação das coordenadas 3D do vértice em questão [32]. Por isso, é desejável que o algoritmo de detecção de cantos tenha as seguintes características.

- Forneça Δu , Δv , $\Delta u'$ e $\Delta v'$ próximos de zero;
- as contribuições de Δu e $\Delta u'$ sejam ambas positivas ou ambas negativas, e;
- da mesma forma, as contribuições de Δv e $\Delta v'$ sejam ambas positivas ou ambas negativas.

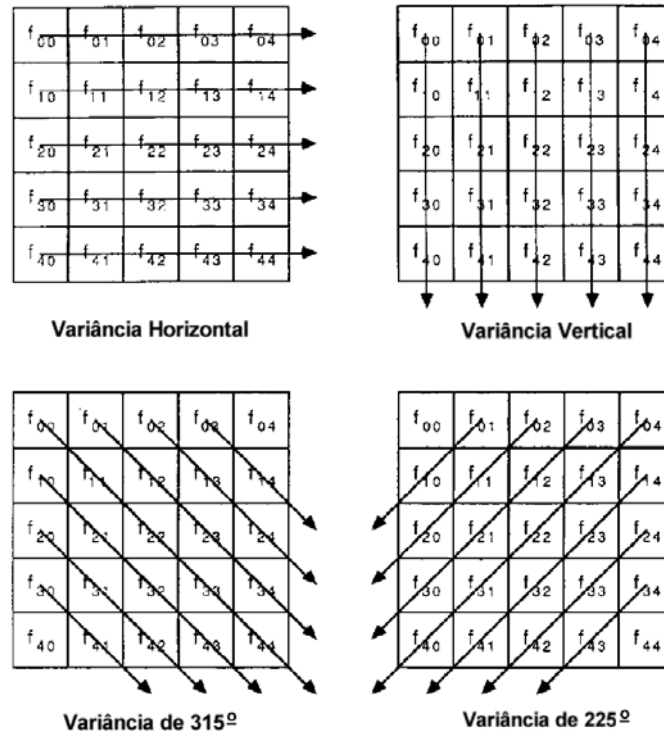


Figura 2.5: As direções das variâncias calculadas pelo operador de Moravec. A menor dentre essas variâncias é o valor do operador de Moravec.

Além disso, é importante que o operador detecte apenas uma coordenada para cada vértice da imagem. Isso diminui a incidência de emparelhamentos ambiguos ou com multiplicidade.

Á seguir, dois detectores de cantos são analisados.

Operador de Moravec

O operador de Moravec foi proposto em 1980 por MORAVEC [58]. Moravec definiu pontos de interesse como “os pontos da imagem que têm uma variância elevada em quatro direções”. Por isso, o operador de Moravec encontra o mínimo entre as variâncias horizontal, vertical e em ambas as diagonais (225° e 315°), para cada ponto da imagem. Este operador foi utilizado para aplicações em visão estereoscópica, por exemplo, por [56] e [61].

O operador de Moravec pode ser descrito pela equação (2.20) e figura 2.5.

$$\begin{aligned}
M(x, y) = \min\{ & \sum_{x-2 \leq x' \leq x+2} \sum_{y-2 \leq y' \leq x+2} (I(x', y') - I(x' + 1, y'))^2, \\
& \sum_{x-2 \leq x' \leq x+2} \sum_{y-2 \leq y' \leq x+2} (I(x', y') - I(x', y' + 1))^2, \\
& \sum_{x-2 \leq x' \leq x+2} \sum_{y-2 \leq y' \leq x+2} (I(x' + 1, y') - I(x', y' + 1))^2, \\
& \sum_{x-2 \leq x' \leq x+2} \sum_{y-2 \leq y' \leq x+2} (I(x', y') - I(x' + 1, y' + 1))^2 \}
\end{aligned} \tag{2.20}$$

Geralmente, nos vértices da imagem existem diversos pontos com variâncias elevadas agrupados. Por isso, para impedir uma alta concentração de pontos de interesse em uma pequena região, o ponto que possui maior variância dentro de uma determinada região, por exemplo, uma janela de 20 x 20 pixels, é escolhido. Em seguida, se este ponto tem variância maior ou igual a um limiar (determinado pela aplicação), ele é marcado como um ponto de interesse.

Nas figuras 2.6(a) e 2.6(b), são apresentadas duas imagens, nas quais os pontos de interesse (encontrados com o operador de Moravec) são ressaltados. As imagens foram capturadas através de um par estereoscópico. O operador encontrou 704 pontos na imagem da figura 2.6(a) e 637 pontos na imagem da figura 2.6(b). Evidentemente, alguns pontos de interesse da imagem da figura 2.6(a) não possuem um correspondente na figura 2.6(b) e vice-versa. Isto dificulta ainda mais o processo de emparelhamento de pontos. Contudo, isto é comum à todos os detectores de cantos e é algo com que o algoritmo de emparelhamento deve lidar.

Operador SUSAN

O operador SUSAN [80] segue o método usual de detecção de cantos, ou seja, dada uma imagem, usando uma janela pré-determinada centrada em cada pixel desta imagem, aplica-se localmente um conjunto de regras. O resultado dessas regras são processadas e fornecem um conjunto de cantos na saída. O operador SUSAN foi utilizado, por exemplo, por PILU [65] em 1997.

O operador SUSAN usa uma janela circular de 3,4 pixels de raio e, para cada pixel dentro desta janela, é calculado

$$n(m_0) = \sum_m c(m, m_0),$$

onde m é um vetor de localização de um pixels dentro da janela, m_0 é a localização do

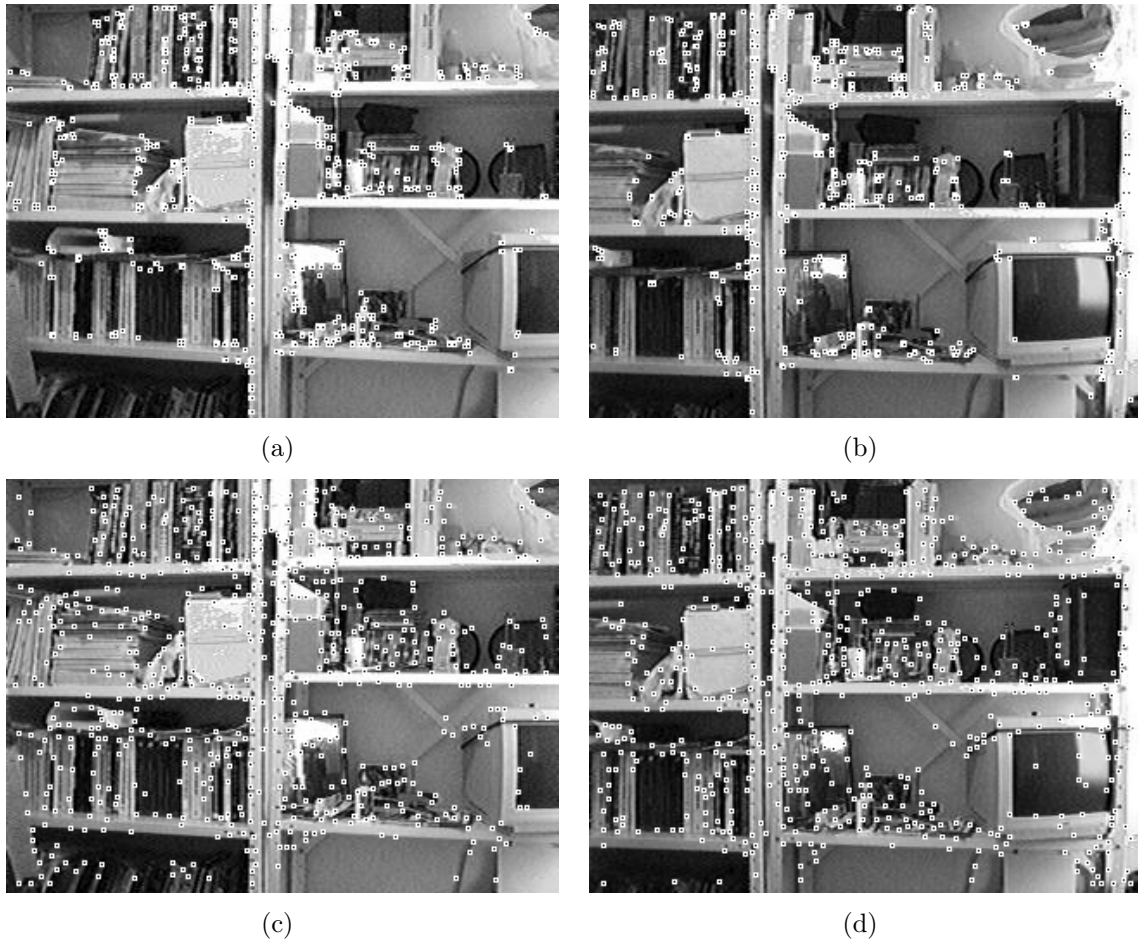


Figura 2.6: (a) e (b) são imagens capturadas por um par estereoscópico e com os pontos de interesse marcados com o operador de Moravec. (c) e (d) são as mesmas imagens com os pontos de interesse marcados através do operador de SUSAN.

ponto central da janela e

$$c(m, m_0) = e^{-[(I(m) - I(m_0))/t]^6},$$

onde $I(m)$ é a intensidade (tons de cinza) do pixel na localização m e t é um limiar (do inglês *threshold*). Por fim, se $n(m_0)$ for maior que um limiar g , o ponto na posição m_0 é considerado um canto. Pode-se dizer que o operador SUSAN considera um canto como o ponto que tem menos vizinhos similares [45].

Nas figuras 2.6(c) e 2.6(d), são apresentadas duas imagens com pontos de interesse (encontrados com o operador de SUSAN) marcados. Neste caso, o operador encontrou 658 pontos na imagem da figura 2.6(c) e 673 pontos na imagem da figura 2.6(d). Novamente, alguns pontos de interesse não tem o seu correspondente marcado na outra imagem. Contudo, este operador marcou apenas um vértice em cada canto da imagem.

Observando-se as figuras 2.6(a) e 2.6(b), nota-se que o operador de Moravec tende à marcar mais de um vértice em torno de um mesmo canto. Por isso, apesar de os operadores Moravec e SUSAN tenham marcado um número muito próximo de vértices, as imagens com os cantos marcados pelo operador SUSAN parecem possuir mais cantos, pois estes estão espalhados por toda a imagem. Em vista disso, o operador SUSAN parece ser bastante estável e mais indicado à aplicação de estereoscopia.

2.4.2 Restrições

Após um conjunto de pontos serem destacados nas duas imagens, pode-se iniciar o processo de emparelhamento. Como já discutido, se o conjunto estereoscópico não está calibrado, não existe restrição geométrica para o problema de emparelhamento. Contudo, para que o problema de emparelhamento de pontos tenha solução, deve existir restrições que sirvam de guia para a resolução da tarefa. Estas restrições são encontradas no sistema de visão humano, nas propriedades das câmeras e sensores ópticos, e na geometria. Para encontrar os possíveis emparelhamentos, essas restrições devem ser aplicadas ao conjunto de pontos de interesse de duas imagens de forma seqüencial ou paralela. Desta forma, espera-se que, ao final do processo, a maior quantidade possível de emparelhamentos tenha sido encontrada corretamente.

À seguir, algumas dessas restrições são discutidas.

Semelhança

A restrição mais básica ao emparelhamento de pontos é a restrição de semelhança. Ela é baseada no valor de intensidade da imagem na posição do ponto de interesse e obriga o correspondente de um ponto ser similar a ele.

A forma mais simples de expressar a semelhança é através do somatório das diferenças absolutas dado pela seguinte equação

$$SAD(m_i, m'_j) = \sum_{k=-n}^n \sum_{l=-m}^m |I_1(u_i + k, v_i + l) - I_2(u'_j + k, v'_j + l)|, \quad (2.21)$$

onde $I_x(u, v)$ é o valor da função de intensidade da imagem x no ponto (u, v) e n e m definem uma janela de $(2n + 1) \times (2m + 1)$ pixels.

A equação (2.21) é baseada na suposição de que os tons de cinza em torno da projeção de um mesmo ponto 3D é a mesma nas duas imagens de um par estereoscópico. Contudo, esta equação é muito sensível à distorções radiométricas, nas quais os tons de

cinza em uma imagem diferem dos tons de cinza na outra imagem por uma constante de desvio e um fator de ganho [1, 2]. Por isso, geralmente, a correlação cruzada normalizada de média zero é uma medida que fornece melhores resultados e pode ser definida como segue.

Dado um ponto de interesse $m_i = (u_i, v_i)$ na imagem 1 e um ponto $m'_j = (u'_j, v'_j)$ (imagem 2) a correlação cruzada normalizada de média zero entre estes pontos é dada por

$$ZNCC(m_i, m'_j) = \frac{\sum_{k=-n}^n \sum_{l=-m}^m [I_{d1}(u_i + k, v_i + l) \times I_{d2}(u'_j + k, v'_j + l)]}{(2n+1)(2m+1)\sqrt{\sigma^2(I_1) \times \sigma^2(I_2)}}, \quad (2.22)$$

onde

$$I_{dx}(u_y + k, v_y + l) = I_x(u_y + k, v_y + l) - \overline{I_x(u_y, v_y)},$$

com

$$\overline{I_x(u, v)} = \sum_{k=-n}^n \sum_{l=-m}^m \frac{I_x(u + k, v + l)}{(2n+1)(2m+1)} \quad (2.23)$$

igual ao nível de cinza médio em torno de $I_x(u, v)$, e

$$\sigma(I_x) = \sqrt{\frac{\sum_{k=-n}^n \sum_{l=-m}^m [I_x(u, v) - \overline{I_x(u, v)}]^2}{(2n+1)(2m+1)}} \quad (2.24)$$

é o desvio padrão da imagem I_x na vizinhança de uma janela de $(2n+1) \times (2m+1)$ centrada em (u, v) .

A correlação está na faixa de -1 , para dois pontos totalmente descorrelacionados, e 1 , para pontos totalmente correlacionados.

Pelo critério de semelhança, para realizar o emparelhamento entre os pontos interessantes de duas imagens, uma par de pontos (m_i, m'_j) é considerado correspondente se a correlação, calculada pela equação (2.22), for maior que um limiar C_{th} .

Na prática, devido as distorções radiométricas, se for usado sozinho, o critério de semelhança produz muita ambiguidade. Isto faz com que a semelhança forneça melhores resultados quando é aplicada em ambientes fechados e com iluminação controlada. Apesar disso, alguns autores [65, 96] a utilizaram como restrição inicial para resumir ainda mais o espaço inicial de busca. Além disso, por ser computacionalmente simples, a equação (2.21) é muito utilizada em sistemas que possuem restrições de tempo, por exemplo, BANKS *et al* [3] e HIRSCHMÜLLER [33] a utilizaram em seus sistemas

estereoscópicos de tempo-real.

Outras equações que expressão a similaridade podem ser encontradas em [2] e [79].

Mapa de Distâncias Mínimas

A teoria do mapa de distâncias mínimas (do inglês *minimal mapping*) foi proposta por ULLMAN [87] em 1979 e depois atualizada por ele mesmo em 1984 [88]. Basicamente, essa teoria garante que dado um conjunto de N pontos interessantes m_i em uma imagem de um par estereoscópico e M pontos interessantes m'_j na outra imagem, a função

$$J_d = A \sum_i^N \sum_j^M c_{ij} d(m_i, m'_j) + B \left(\sum_i^N \sum_j^M c_{ij} \right)^{-1} \quad (2.25)$$

atinge o mínimo quando todos os emparelhamentos são realizados corretamente. Na equação (2.25), A e B são constantes que dependem da aplicação e devem ser determinadas empiricamente, c_{ij} é uma variável que indica a existência ($c_{ij} = 1$) ou não ($c_{ij} = 0$) de um emparelhamento entre m_i e m'_j , e $d(m_i, m'_j)$ é uma medida de distância entre m_i e m'_j .

O segundo somatório na equação (2.25) garante que a solução trivial, na qual todos os c_{ij} são iguais à zero, não seja a solução do problema.

A restrição do mapa de distâncias mínimas é bastante intuitivo e é ilustrado na figura 2.7. Nesta figura, os pontos interessantes das duas imagens são esboçados na mesma figura e a distância entre os pontos correspondentes é ilustrada através de linhas contínuas. Fica claro que se, por exemplo, dois pontos tiverem os seus pares trocados, as distâncias associadas serão maiores e, portanto, a equação (2.25) não estará no seu estado mínimo.

Na prática, em que nem todo ponto m_i tem o seu correspondente pertencente ao conjunto de m'_j e $N \neq M$, a restrição das distâncias mínimas não é suficiente. Apesar disso, muitos autores [28, 65, 74] utilizaram esta restrição para resolver problemas de emparelhamento.

Vizinhança

Considerando um par de pontos candidato à emparelhamento (m_i, m'_j) , pode-se representar $N(m_i)$ e $N(m'_j)$ como sendo, respectivamente, o conjunto de vizinhos de m_i e m'_j dentro de uma janela circular de raio R (figura 2.8). Se (m_i, m'_j) for um bom candidato a emparelhamento, espera-se que existam muitos emparelhamentos (n_k, n'_l) ,

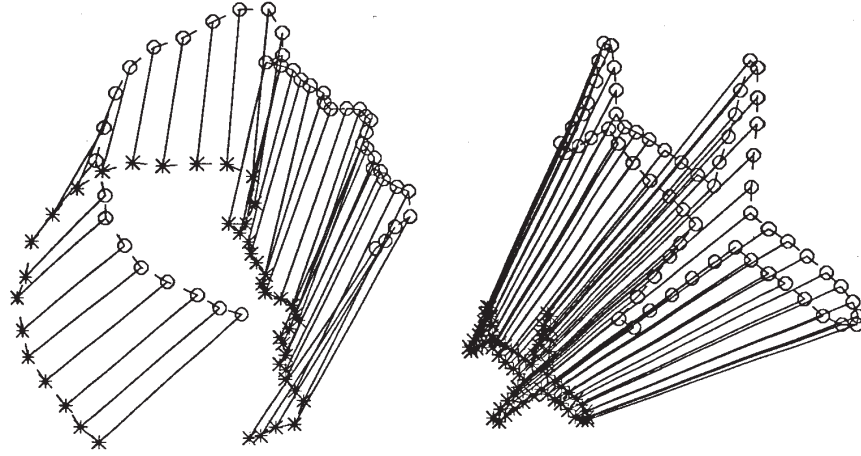


Figura 2.7: Conjunto de pontos interessantes com o mapa de distâncias mínimas esboçado.

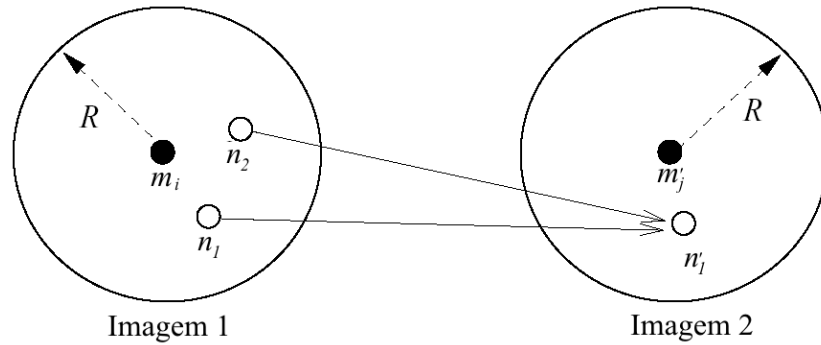


Figura 2.8: Par (m_i, m'_j) candidato à correspondência com vizinhos dentro de uma janela R .

onde $n_k \in N(m_i)$ e $n_l \in N(m'_j)$, tal que a posição relativa entre n_k e m_i é semelhante a posição relativa entre n'_l e m'_j . Por outro lado, se (m_i, m'_j) não é um bom emparelhamento, não se espera encontrar vizinhos com posições relativas semelhantes.

Formalmente, pode-se definir a “força” de uma par (m_i, m'_j) baseado nos pares candidatos a emparelhamento vizinhos através da equação

$$F(m_i, m'_j) = \sum_{n_k \in N(m_i)} \left[\max_{n'_l \in N(m'_j)} \frac{A_{kl} \delta(m_i, m'_j; n_k, n'_l)}{1 + D(m_i, m'_j; n_k, n'_l)} \right], \quad (2.26)$$

onde A_{kl} são constantes que dependem do grau de confiança no emparelhamento (n_k, n'_l) ,

$$D(m_i, m'_j; n_k, n'_l) = [d(m_i, m'_j) + d(n_k, n'_l)]/2,$$

com $d(m, n) = \|m - n\|$, a distância euclidiana entre os pontos m e n , e

$$\delta(m_i, m'_j; n_k, n'_l) = \begin{cases} e^{-r/\varepsilon_r} & \text{se é um candidato à emparelhamento e } r < \varepsilon_r \\ 0 & \text{caso contrário} \end{cases},$$

onde

$$r = \frac{|d(m_i, n_k) - d(m'_j, n'_l)|}{D(m_i, m'_j; n_k, n'_l)}$$

é uma medida do erro das posições relativas entre as duplas (m_i, n_k) e (m'_j, n'_l) , e ε_r é um limiar para está medida de erro.

É importante observar que, caso um ponto m_i tenha vários candidatos a emparelhamento na outra imagem, o operador “max” da equação (2.26) garante que apenas aquele que tem a posição relativa mais próxima contribui para a “força” do par (m_i, m'_j) .

Em 1985, POLLARD *et al* [67] utilizou o critério de vizinhança para emparelhamento de pontos e, em 1994, o seu trabalho foi aperfeiçoado por ZHANG *et al* [96]. Exemplos de outros trabalhos que utilizam a restrição da vizinhança são [56, 61, 62].

Disparidade

Como discutido na seção 2.3, em estereoscopia duas câmeras capturam um par de imagens do ambiente a partir de pontos de vista diferentes. Nestas imagens, projeções de um mesmo ponto no espaço presentes nas duas imagens apresentam um deslocamento relativo, chamado de disparidade, que é tanto maior quanto mais próximo o objeto estiver das câmeras. A informação de disparidade pode ser utilizada para estimar a profundidade do objeto, que é a distância do objeto até o conjunto de câmeras.

Considerando o caso de câmeras com os planos de imagem coplanares e os eixos ópticos alinhados, como no exemplo da figura 2.4, a profundidade p está relacionada à disparidade d por

$$p = (b \times f) \frac{1}{d}, \quad (2.27)$$

onde f é distância focal das duas câmeras e b é a distância entre os centros de projeções C_1 e C_2 .

Através da equação (2.27), mostra-se que quanto mais distante um ponto M estiver das câmeras, menor será a disparidade das suas projeções. Dentre outras coisas, isto demonstra que, na prática, será difícil determinar com exatidão a profundidade de pontos muito distâtes das câmeras. Contudo, limitando-se as distâncias mínima e máxima dos objetos de interesse em uma cena, pode-se usar a disparidade como uma

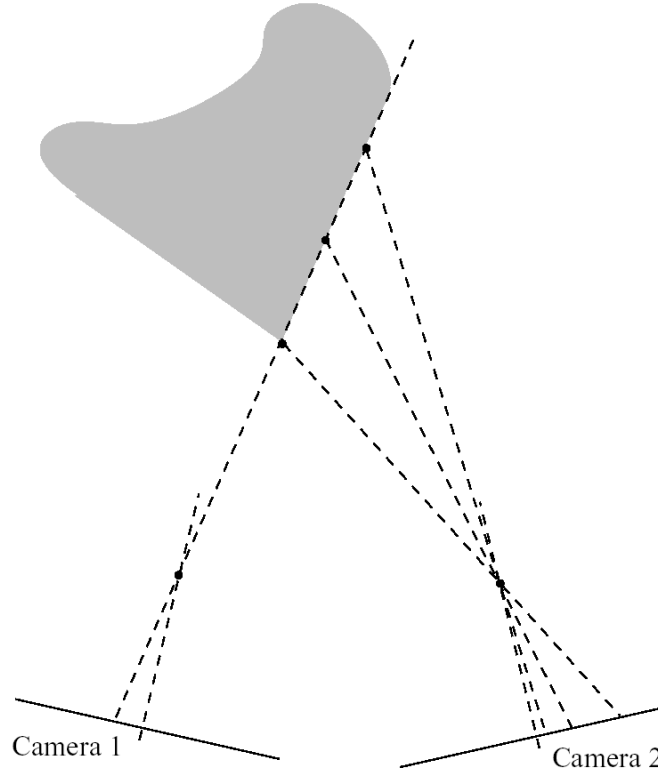


Figura 2.9: Falha na restrição de unicidade.

restrição que auxilie no processo de emparelhamento de pontos. Por exemplo, pode-se determinar que em emparelhamentos (m_i, m'_j) ,

$$d_{\min} \leq |d(m_i, m'_j)| \leq d_{\max}. \quad (2.28)$$

Unicidade

Dado um par de imagens, cada uma com conjunto de pontos de interesse marcados, mesmo após aplicar-se todas as restrições já discutidas, é natural que ainda exista alguma ambiguidade. Isto acontece quando um ponto na imagem 1 forma dois ou mais emparelhamentos com pontos de interesse da imagem 2 e todos os pares são igualmente plausíveis. Neste caso, deve-se escolher um dos pares ou, para reduzir a incidência de emparelhamentos errados, eliminar todos os pares ambíguos.

É importante observar que, em alguns casos especiais (como o da figura 2.9), um único ponto de interesse pode emparelhar com mais de um ponto da imagem. Além disso, a incidência de pares ambíguos aumenta se existirem objetos transparentes na cena [6].

2.5 Emparelhamento de pontos sem calibração

Em um sistema de visão estereoscópica sem calibração, as restrições da seção 2.4.2 são os únicos suportes à tarefa de emparelhamento de pontos.

Na maioria das aplicações, as imagens capturadas pelo par estereoscópico estão em pontos de vista ligeiramente diferentes. Isto reduz as transformações geométricas sofridas pelos objetos de uma imagem em relação a outra e, portanto, aumenta o grau de semelhança entre os pontos correspondentes e facilita o seu emparelhamento. Apesar disso, a um custo de maior complexidade, trabalhos mais recentes como o de BAUMBERG [4] realizam a tarefa de emparelhamento em imagens com pontos de vista bastante distintos.

Já que os pontos correspondentes apresentam uma grau de semelhança elevado, é natural que muitos trabalhos utilizem a restrição de semelhança para criar um conjunto inicial de candidatos à emparelhamento [33, 65, 79, 96]. Em seguida, outras restrições são utilizadas para eliminar a ambiguidade. Dentro desta abordagem, TOMASI e MANDUCHI [83] desenvolveram um método no qual, ao invés de utilizar a função valor de intensidade da imagem na posição do ponto m como sendo $I(m)$, utilizam uma função que associa um conjunto de pontos a um valor de intensidade, ou seja, $m(I)$. Em tese, após montada a função, esta abordagem fornece emparelhamento por correlação sem necessidade de busca.

Outra restrição bastante utilizada é o mapa de distâncias mínimas. Como essa restrição pode ser expressa através de uma equação semelhante a equação (2.25), muitas vezes a tarefa de emparelhamento é encarada como um problema de minimização em que se deseja encontrar o mínimo da função dada pela equação (2.25) [53, 57]. Nesta mesma metodologia, GRZYWACZ [28] utilizou redes neurais de Hopfield [34, 35] para minimizar o mapa de distâncias. Redes de Hopfield também foram utilizadas em [56, 61, 62] para otimizar o problema de emparelhamento de pontos, mas, nestes casos, foi utilizada a restrição de vizinhança.

Como discute-se na seção 2.4.2, mesmo após aplicar-se diversas restrições aos pares candidatos à emparelhamento, ainda deve haver ambiguidade em alguns desses. Contudo, trabalhos mais recentes como o de ŠÁRA [72], que através de heurística fornece o candidato à emparelhamento menos ambíguo, ajudam a solucionar este problema.

Emparelhamento por SVD

Em 1991, SCOTT e LONGUET-HIGGINS [74] propuseram um método para emparelhamento de pontos que utiliza algumas propriedades da Decomposição do Valor Singular (do inglês *Singular Value Decomposition* ou, simplesmente, SVD) para satisfazer as restrições de unicidade e o mapa de distâncias mínimas.

Considerando I uma imagem contendo m pontos de interesse, m_i ($i = 1, \dots, m$), e I' outra imagem com n pontos de interesse, m'_j ($j = 1, \dots, n$), o algoritmo de SCOTT e LONGUET-HIGGINS [74] realiza o emparelhamento entre os pontos m_i e m'_j em três passos.

1. Uma matriz de proximidade, \mathbf{G} , é gerada. Cada elemento \mathbf{G}_{ij} desta matriz é a distância entre os pontos m_i e m'_j com ponderação gaussiana, ou seja,

$$\mathbf{G}_{ij} = e^{-r_{ij}^2/(2\sigma^2)} \quad i = 1, \dots, m \text{ e } j = 1, \dots, n, \quad (2.29)$$

onde $r_{ij} = d(m_i, m'_j)$ é a distância euclidiana (considerando que os pontos estão no mesmo plano). \mathbf{G} é positivamente definida e G_{ij} decresce de 1 para 0 monotonicamente com a distância. O parâmetro σ controla a proximidade entre os dois conjuntos de pontos: um σ pequeno reforça pontos mais próximos, um σ grande permite que pontos correspondentes estejam mais distantes.

2. Calcula-se a Decomposição do Valor Singular (SVD) de $\mathbf{G} \in \mathbf{M}_{m,n}$:

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

Nesta equação, $\mathbf{U} \in \mathbf{M}_m$ e $\mathbf{V} \in \mathbf{M}_n$ são matrizes ortogonais e $\mathbf{\Sigma} \in \mathbf{M}_{m,n}$ é uma matriz diagonal com Σ_{ii} (sua diagonal principal) igual aos valores singulares da matriz em ordem decrescente.

3. Obtém-se a matriz

$$\hat{\mathbf{G}} = \mathbf{U}\mathbf{E}\mathbf{V}^T,$$

onde \mathbf{E} é uma matriz obtida substituindo-se todos os elementos Σ_{ii} da matriz $\mathbf{\Sigma}$ pela unidade. Através da matriz $\hat{\mathbf{G}} \in \mathbf{M}_{m,n}$ pode-se determinar se há emparelhamento entre os pontos m_i e m'_j como segue: se \hat{G}_{ij} é o maior elemento da linha i e da coluna j , (m_i, m'_j) é considerado um emparelhamento de 1:1 [74].

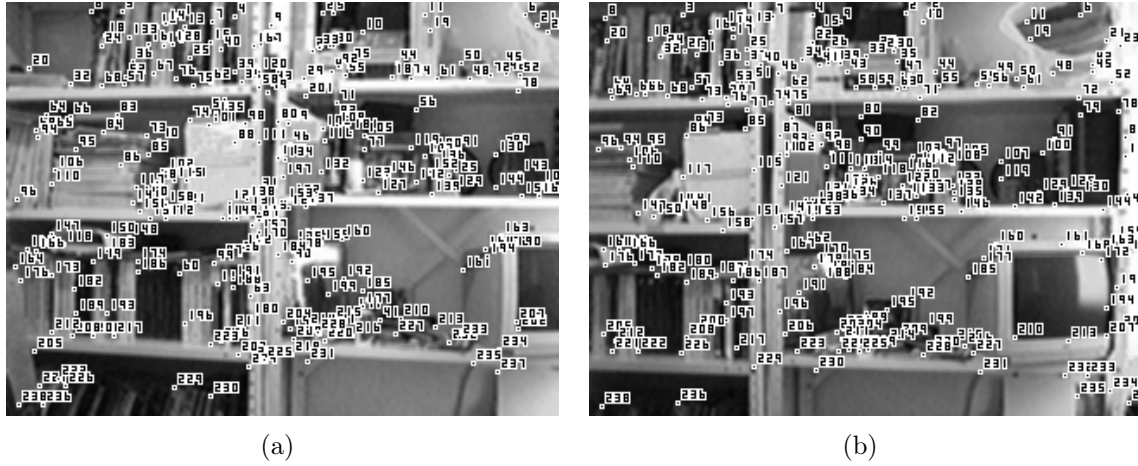


Figura 2.10: (a) e (b) possuem correspondências marcadas pelo método de SCOTT e LONGUET-HIGGINS.

Como pode ser observado, esse método é bastante simples, não-iterativo¹ e sempre produz um resultado com unicidade. Contudo, na prática, o algoritmo produz bons resultados apenas quando a maior parte dos pontos de interesse da imagem 1 tem o seu correspondente segmentado como um ponto interessante da imagem 2 e vice-versa. Infelizmente, como discutido na seção 2.4.1, isto não ocorre na maioria das vezes.

Nas figuras 2.10(a) e 2.10(b), apresenta-se o resultado do algoritmo para o conjunto de cantos interessantes das figuras 2.6(c) e 2.6(d). Como pode-se notar, o emparelhamento de pontos não foi estabelecido de forma correta, pois o número de emparelhamentos corretos foi praticamente nulo.

Para tentar aumentar a robustez do algoritmo, em 1997, PILU [65] propôs uma alteração no cálculo da matriz \mathbf{G} para que ela incorpore também o princípio de semelhança. A idéia foi calcular a matriz \mathbf{G} usando, além da distância euclidiana, a correlação calculada pela equação (2.22). Em seu artigo, PILU [65] propôs a matriz \mathbf{G} como sendo

$$G_{ij} = e^{-(C_{ij}-1)^2/(2\gamma^2)} e^{-r_{ij}^2/(2\sigma^2)}$$

ou

$$G_{ij} = e^{-0.5[(C_{ij}-1)^2/\gamma^2 + r_{ij}^2/\sigma^2]} \quad i = 1, \dots, m \text{ e } j = 1, \dots, n, \quad (2.30)$$

onde C_{ij} é a correlação entre os pontos m_i e m'_j , e o parâmetro γ controla o grau de correlação entre os dois conjuntos de pontos: um γ pequeno reforça apenas os pontos muito correlacionados, um γ grande permite que pontos correspondentes sejam menos

¹Evidentemente, alguma iteração pode ser necessaria no cálculo da decomposição SVD.

semelhantes.

As alterações propostas por PILU [65] não são suficientes para que o algoritmo produza bons resultados quando muitos pontos de interesse em uma imagem não correspondam a nenhum ponto de interesse na outra imagem. Além disso, não elimina outra grande desvantagem do algoritmo: para um conjunto grande de pontos interessantes, o tempo necessário para calcular-se a decomposição do valor singular da matriz \mathbf{G} . Com isso, apesar de não ser iterativo, o algoritmo leva muito tempo para realizar emparelhamento de pontos.

Emparelhamento por Relaxação

Como dito, dadas duas imagens sem calibração, mas com alguns pontos de interesse segmentados em cada uma, o critério de semelhança é o mais intuitivo na determinação de emparelhamentos de pontos. Neste caso, o emparelhamento é estabelecido com segue.

Como discutido, em imagens descalibradas, cada ponto de interesse na imagem 1 forma um possível emparelhamento com todos os pontos de interesse da imagem 2. Assim, a equação (2.22) é aplicada a todos esses possíveis emparelhamentos e aqueles que obtiverem uma semelhança maior ou igual a um limiar C_{th} são considerados emparelhamentos prováveis.

Como a semelhança não é condição suficiente para estabelecer o emparelhamento de pontos, com o emparelhamento por correlação, um ponto m_i na primeira imagem possui vários pontos m'_j ($j = 1, \dots, n$) candidatos a emparelhamento. Por isso, é necessário aplicar-se outra restrição, por exemplo, a restrição da vizinhança para eliminar a ambiguidade. Neste caso, pode-se calcular a força de cada provável emparelhamento através da equação (2.26) e elimina-se os emparelhamentos fracos.

Para eliminar-se os emparelhamentos fracos, diversas técnicas podem ser utilizadas. POLLARD *et al* [67] utilizaram, por exemplo, a abordagem “o-vencedor-ganha-tudo”, ou seja, dado um ponto m_i na primeira imagem com diversos pares de prováveis correspondentes (m_i, m'_j) associados, onde $j = 1, \dots, n$, considera-se o par vencedor aquele que tiver a maior força. Nesta abordagem, a eliminação dos emparelhamentos fracos é realizada em um único passo. Contudo, ela não produz bons resultados, pois os emparelhamentos são tratados individualmente e de forma independente.

Uma abordagem que produz melhores resultados, ao custo de uma menor velocidade, é “o-perdedor-não-ganha-nada”. Agora, a eliminação dos emparelhamentos fracos

é realizada em várias iterações, em cada uma delas, dado um ponto m_i na primeira imagem, apenas o par associado (m_i, m'_j) mais fraco é eliminado. Em seguida, a equação (2.26) é aplicada novamente aos pares (m_i, m'_j) remanescentes. Então, novamente os pares mais fracos são eliminados até que não ache mais ambiguidade.

A técnica “o-perdedor-não-ganha-nada” é ineficiente se um ponto m_i possui muitos prováveis correspondentes. Por isso, ZHANG *et al* [96] desenvolveram um método intermediário: “alguns-vencedores-ganham-tudo”. Nesta técnica, para cada ponto m_i que possui dois ou mais prováveis emparelhamentos, seleciona-se o emparelhamento mais forte. Assim, tem-se um conjunto com N prováveis emparelhamentos, onde N é o número de pontos da imagem 1 com emparelhamentos ambíguos. Em seguida, são formadas duas tabelas. Na primeira, os N emparelhamentos são organizados em ordem decrescente de força. Na outra, os emparelhamentos são organizadas em ordem decrescente de certeza. Para montar a segunda tabela, ZHANG *et al* [96] definiram o grau de certeza do emparelhamento mais provável de um ponto m_i como sendo

$$U_A(m_i) = 1 - F^{(2)}/F^{(1)},$$

onde $F^{(1)}$ é a força do par de emparelhamento mais provável de m_i e $F^{(2)}$ é a força do segundo par mais provável de m_i . Então, $U_A(m_i)$ é um valor entre 1 (certeza) e 0 (ambiguidade).

Após a construção das duas tabelas, considera-se como corretos os emparelhamentos que pertencerem aos q por cento dos primeiros elementos de ambas as tabelas. Assim, emparelhamentos fortes não serão escolhidos se forem ambíguos e emparelhamentos ambíguos não serão escolhidos mesmo se possuam uma força elevada.

Após considerar os emparelhamentos ganhadores como corretos, o algoritmo recalcula a força de todos os emparelhamentos que ainda possuam ambiguidade e volta a montar as duas tabelas e escolher mais pares ganhadores. Esse processo é repetido até que não reste nenhum emparelhamento ambíguo.

A figura 2.11 apresenta dois pares de imagens com emparelhamentos marcados pelo algoritmo de relaxação proposto por ZHANG *et al* [96]. Em ambos os casos, o algoritmo consegue estabelecer muitos emparelhamentos com sucesso, mas também comete alguns erros de localização e falsos emparelhamentos.

Os objetos das imagens das figuras 2.11(a) e 2.11(b) estão em planos muito próximos. Por isso, a disparidade entre os pontos correspondentes é muito próxima para todos os pares das imagens. Por isso, o algoritmo de relaxação funcionou melhor neste

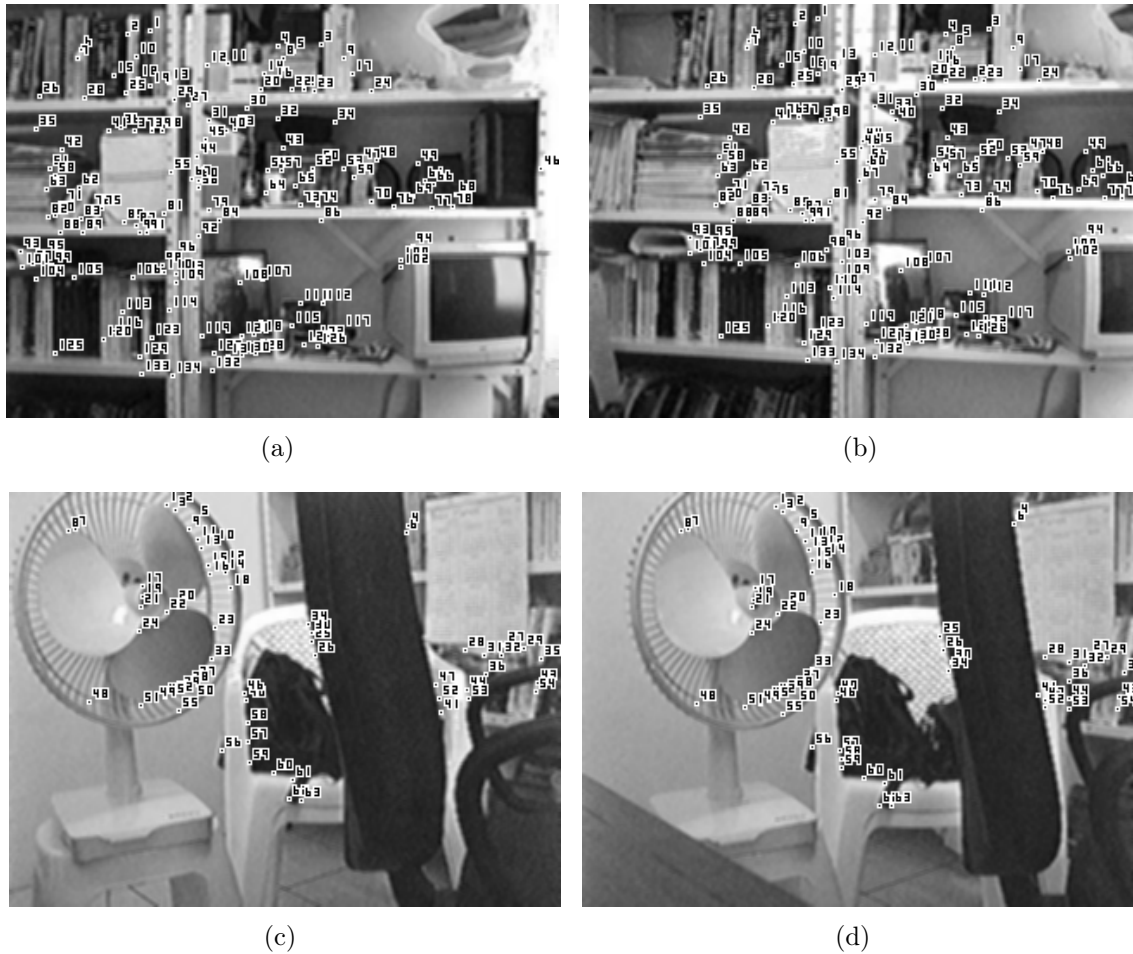


Figura 2.11: (a) e (b) formam um par de imagens com correspondências marcadas por relaxação. O par de imagens (c) e (d) é outro exemplo de correspondência por relaxação.

caso que, por exemplo, no caso das imagens das figuras 2.11(c) e 2.11(d), que possuem um conjunto de emparelhamentos com disparidade muito distintas.

Dos resultados obtidos com os dois métodos de emparelhamentos de pontos em imagens descalibradas discutidos, observa-se que a existência (em ambas as imagens) de pontos de interesse sem um correspondente dificulta significativamente a tarefa de emparelhamento de pontos. Por isso, os algoritmos de emparelhamento, normalmente, fornecem alguns falsos emparelhamentos. Mesmo assim, como discute-se no próximo capítulo, o conjunto de emparelhamentos obtido é suficiente para estimar-se a matriz fundamental. Essa matriz relaciona as duas imagens de um par estereoscópico e torna a maior parte das tarefas bem menos complexa.

Capítulo 3

Estereoscopia Através da Matriz Fundamental

Na estereoscopia, uma vez que um conjunto de emparelhamentos tenha sido estabelecido, é possível estimar-se a matriz fundamental. Essa matriz concentra a única restrição existente em um par estereoscópico descalibrado, ou seja, a geometria epipolar. Como discute-se no decorrer deste capítulo, com a geometria epipolar conhecida, tarefas como emparelhamento de pontos, calibração de câmeras e até mesmo reconstrução 3D, tornam-se bem menos complexas.

3.1 Matriz Fundamental

Em um sistema com câmeras sem calibração, a única informação disponível é a geometria epipolar. Essa geometria já foi descrita em numerosos artigos, por exemplo, [93] e [85]. Basicamente, a geometria epipolar pode ser entendida se for considerado o caso de duas câmeras como apresentado na figura 3.1. Nesta figura, C e C' são, respectivamente, os centros ópticos da primeira e segunda câmeras. Então, dado um ponto \mathbf{m} na primeira imagem, o ponto correspondente, \mathbf{m}' , na segunda imagem está restrito a uma linha chamada linha epipolar de \mathbf{m} , representada na figura por l' . A linha l' é a intersecção do plano Π , definido por \mathbf{m} , C e C' (chamado de plano epipolar), com o plano da segunda imagem, I' . Isto acontece porque o ponto \mathbf{m} da imagem pode corresponder a um ponto qualquer da reta CM (M pode estar no infinito) e a projeção de CM em I' é a linha l' . Além disso, observa-se que todas as linhas epipolares dos pontos na primeira imagem passam através de um ponto comum, \mathbf{e}' , que é conhecido como epipolo. O ponto \mathbf{e}' é a intersecção da linha CC' com o plano de imagem I' .

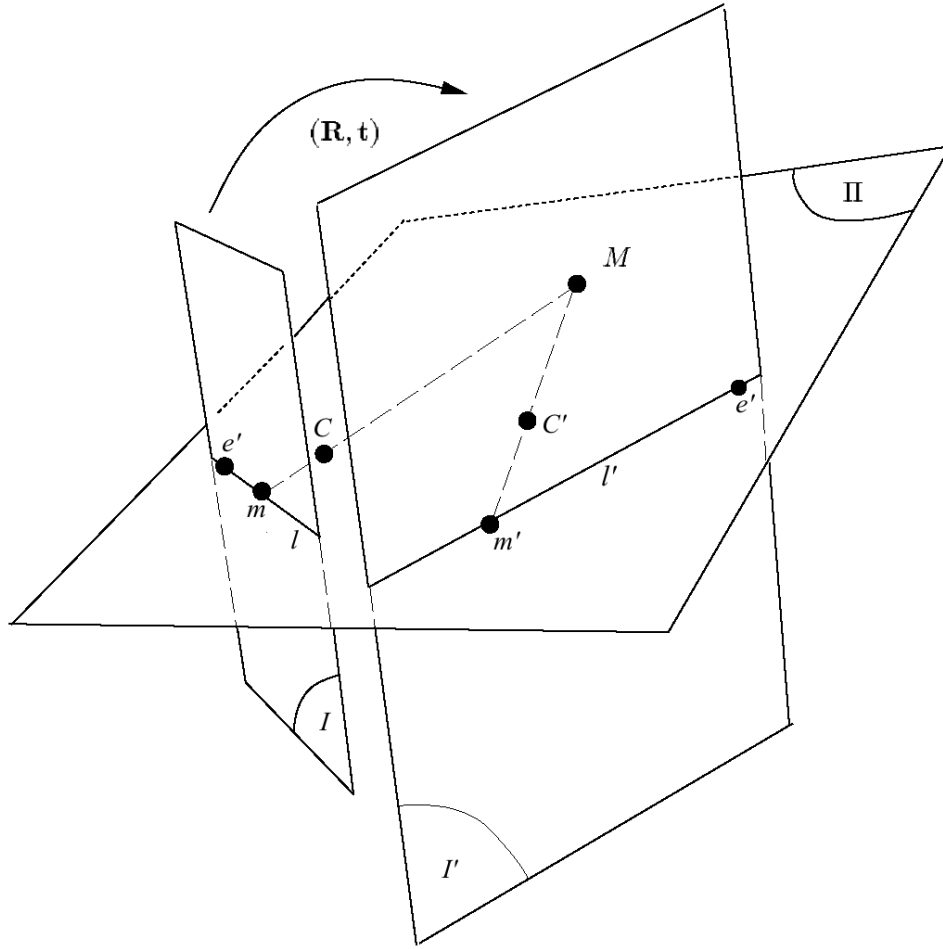


Figura 3.1: Geometria Epipolar.

Isto pode ser entendido como segue. Para cada ponto \mathbf{m}_i na imagem I , a sua linha epipolar \mathbf{l}'_i em I' é a intersecção do plano Π^i , definido por \mathbf{m}_i , C e C' , com o plano de imagem I' . Dessa forma., todos os planos epipolares Π^i formam um conjunto de planos contendo a linha CC' . Logo, eles precisam intersectar I' em um ponto comum, que é \mathbf{e}' . Finalmente, pode-se facilmente observar a simetria da geometria epipolar. O ponto correspondente na primeira imagem de cada ponto \mathbf{m}'_i , sobre a linha \mathbf{l}'_i , precisa pertencer a linha epipolar \mathbf{l}_i , que é a intersecção do mesmo plano Π^i com o plano da primeira imagem I . Todas as linhas epipolares formam um conjunto contendo o epipolo \mathbf{e} , que é a intersecção da linha CC' com o plano de imagem I . A simetria também leva a seguinte observação. Se \mathbf{m} (um ponto em I) e \mathbf{m}' (um ponto em I') correspondem a um único ponto \mathbf{M} no espaço, então \mathbf{m} e \mathbf{m}' , C e C' precisam pertencer a um único plano. Isto é conhecido como restrição co-planar ou restrição epipolar.

Agora, sem perda de generalidade, assumindo que a câmera 1 está na origem do sistema de coordenadas do ambiente e considerando a equação (2.4), tem-se

$$\tilde{\mathbf{m}} = s_m \mathbf{A} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}' = s_{m'} \mathbf{A}' [\mathbf{R} \mid -\mathbf{RC}'] \tilde{\mathbf{M}}. \quad (3.1)$$

Eliminando-se $\tilde{\mathbf{M}}$ da equação anterior, obtém-se a seguinte equação fundamental

$$\tilde{\mathbf{m}}'^T \mathbf{A}'^{-T} \mathbf{T} \mathbf{R} \mathbf{A}^{-1} \tilde{\mathbf{m}} = 0, \quad (3.2)$$

onde $\mathbf{T} = -\mathbf{RC}'$ é uma matriz antisimétrica, tal que $\mathbf{T}\mathbf{x} = \mathbf{T} \times \mathbf{x}$ para todo vetor 3-D \mathbf{x} (\times representa o produto cruzado).

A equação (3.2) é uma restrição fundamental por trás de quaisquer duas imagens se elas são projeções em perspectiva de uma mesma cena. Fazendo $\mathbf{F} = \mathbf{A}'^{-T} \mathbf{T} \mathbf{R} \mathbf{A}^{-1}$, a equação (3.2) pode ser re-escrita como

$$\tilde{\mathbf{m}}'^T \mathbf{F} \tilde{\mathbf{m}} = 0, \quad (3.3)$$

onde \mathbf{F} é conhecida como a matriz fundamental das duas imagens. Geometricamente, $\mathbf{F}\tilde{\mathbf{m}}$ define a linha epipolar do ponto \mathbf{m} na segunda imagem. A equação (3.3) não diz nada além de que o ponto correspondente na imagem da direita de \mathbf{m} está sobre a linha epipolar correspondente.

Transpondo a equação (3.3), encontra-se a relação simétrica que relaciona a segunda imagem com a primeira, ou seja,

$$\tilde{\mathbf{m}}^T \mathbf{F}^T \tilde{\mathbf{m}}' = 0. \quad (3.4)$$

Do trabalho de LUONG e FAUGERAS [48], sabe-se que a matriz fundamental tem posto menor ou igual a 2. Por isso, desde que a equação (3.3) é homogênea, são necessários, no mínimo, sete emparelhamentos (m, m') para calcular-se esse matriz. Por último, como $\det(\mathbf{T}\mathbf{x}) = 0$,

$$\det(\mathbf{F}) = 0.$$

3.2 Determinação Algébrica da Matriz Fundamental

Se o par estereoscópico está calibrado, a partir das matrizes de projeção \mathbf{P} e \mathbf{P}' , é possível encontrar a matriz fundamental. Neste trabalho, utiliza-se a metodologia

proposta por XU e ZHANG [100] em 1997. Como pode-se ver na seção 2.3.1, dado um ponto m no plano da imagem I , o raio óptico associado é

$$\widetilde{\mathbf{M}}(\alpha) = \mathbf{P}^+ \widetilde{\mathbf{m}} + \alpha \widetilde{\mathbf{C}}.$$

Além disso, os pontos $\widetilde{\mathbf{M}} = \mathbf{P}^+ \widetilde{\mathbf{m}}$ e C (o centro da câmera) pertencem a está raio óptico e são projetados na imagem I' , respectivamente, em $\widetilde{\mathbf{m}}' = \mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}$ e $\widetilde{\mathbf{e}}' = \mathbf{P}' \widetilde{\mathbf{C}}$ (o epipolo da imagem I'). Assim, a linha epipolar de m pode ser expressa como sendo

$$\begin{aligned} \mathbf{l}' &= (\mathbf{P}' \widetilde{\mathbf{C}}) \times (\mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}) \\ &= \widetilde{\mathbf{e}}' \times (\mathbf{P}' \mathbf{P}^+ \widetilde{\mathbf{m}}) \\ &= [\widetilde{\mathbf{e}}']_{\times} (\mathbf{P}' \mathbf{P}^+) \widetilde{\mathbf{m}}. \end{aligned} \quad (3.5)$$

Contudo, $\mathbf{l}' = \mathbf{F} \widetilde{\mathbf{m}}$. Então, de (3.5)

$$\mathbf{F} = [\widetilde{\mathbf{e}}']_{\times} (\mathbf{P}' \mathbf{P}^+) \quad (3.6)$$

ou

$$\mathbf{F} = [\widetilde{\mathbf{e}}']_{\times} \mathbf{Q}, \quad (3.7)$$

onde $\mathbf{Q} = \mathbf{P}' \mathbf{P}^+$.

3.3 Estimação da Matriz Fundamental

Da equação (3.3), vê-se que a matriz fundamental deve ser estimada a partir de um conjunto de emparelhamentos $(\mathbf{m}_i, \mathbf{m}'_i)$. Para isto, considere um ponto $\mathbf{m}_i = [u_i, v_i]^T$ na imagem I o correspondente do ponto $\mathbf{m}'_i = [u'_i, v'_i]^T$ da imagem I' . Neste caso, eles devem satisfazer a equação (3.3). Esta equação pode ser re-escrita como uma equação linear e homogênea em termos dos nove coeficientes da matriz \mathbf{F} :

$$\mathbf{u}_i^T \mathbf{f} = 0, \quad (3.8)$$

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

$$\mathbf{f} = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T,$$

onde F_{ij} é o elemento de \mathbf{F} da linha i e coluna j .

Com n emparelhamentos entre I e I' , e considerando a equação (3.8), tem-se o seguinte sistemas de equações lineares

$$\mathbf{U}_n \mathbf{f} = \mathbf{0}, \quad (3.9)$$

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

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

3.3.1 Solução com apenas sete emparelhamentos

Dispondo-se de apenas o mínimo de sete emparelhamentos, na equação (3.9), $n = 7$ e $\text{rank}(\mathbf{U}_7) = 7$. Dessa forma, para encontrar uma solução para este sistema de equações, pode-se considerar a decomposição do valor singular de \mathbf{U}_7 , ou seja,

$$\mathbf{U}_7 = \mathbf{U} \Sigma \mathbf{V}^T.$$

Assim, se for definido r como

$$\sigma_1 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = 0,$$

onde σ_i são os valores singulares de \mathbf{U}_7 . Isto é, σ_r é o menor valor singular de \mathbf{U}_7 diferente de zero. Então, $B = \{v_{r+1}, \dots, v_m\}$ é uma base que representa todo os espaço nulo \mathbf{U}_7 , onde v_i são os autovetores de \mathbf{U}_7 correspondentes aos seus valores singulares nulos.

Como $\mathbf{U}_7 \in \mathbf{M}_{7 \times 9}$, $B = \{v_8, v_9\}$. Além disso, devido a sua homogeniedade, a matriz fundamental pertence a família de matrizes $\alpha \mathbf{F}_1 + (1 - \alpha) \mathbf{F}_2$, onde \mathbf{F}_1 e \mathbf{F}_2 são as matrizes correspondentes à v_8 e v_9 . Desde que, $\det(\mathbf{F}) = 0$, para encontrar a matriz fundamental, deve ser feito

$$\det[\alpha \mathbf{F}_1 + (1 - \alpha) \mathbf{F}_2] = 0. \quad (3.10)$$

A equação (3.10) fornece um polinômio de grau três. Por isso, esta equação possui pelo menos uma solução real. Para cada solução real, a matriz fundamental é dada por

$$\mathbf{F} = \alpha \mathbf{F}_1 + (1 - \alpha) \mathbf{F}_2.$$

Como visto na seção 2.4, os algoritmos de emparelhamento podem fornecer emparelhamentos com ruído e até mesmo falsos emparelhamentos. Isto degrada totalmente a matriz fundamental calculada pelo método dos sete emparelhamentos [85, 93]. Por isso, este método não tem aplicação prática [9]. Felizmente, normalmente dispõe-se de mais de sete emparelhamentos. Assim, um algoritmo de minimização pode ser utilizado para estimar-se a matriz fundamental de forma mais confiável.

3.3.2 Método dos 8 pontos

O método dos quadrados mínimos pode ser utilizado para resolver

$$\min_{\mathbf{F}} \sum_i (\tilde{\mathbf{m}}_i'^T \mathbf{F} \tilde{\mathbf{m}}_i)^2. \quad (3.11)$$

Que também pode ser escrita como

$$\min_{\mathbf{f}} \|\mathbf{U}_n \mathbf{f}\|^2, \text{ sujeito a } \|\mathbf{f}\| = 1. \quad (3.12)$$

Apesar de poderem ser utilizados mais de oito pontos para resolver a equação (3.11), este método geralmente é referenciado como método dos oito pontos.

Em 1995, analisando a estabilidade deste método através dos autovalores e autovetores da matriz \mathbf{U}_n , HARTLEY [29] mostrou que se obtém melhores resultados se as coordenadas homogêneas dos pares $(\mathbf{m}_i, \mathbf{m}_i')$, na média, forem igual a unidade. Além disso, resultados experimentais mostram que uma maior estabilidade é obtida se os centróides dos pontos nas duas imagens estiverem na origem [29]. Por isso, antes da minimização:

1. As coordenadas dos pontos no plano da imagem devem ser transladadas de forma que o centróide dos pontos fique na origem, e;
2. Os pontos devem sofrer uma mudança de escala para que, na média, a distância do ponto à origem seja igual a $\sqrt{2}$.

Essas transformações devem ocorrer independentemente em cada imagem e podem ser resumidas em uma matriz $\mathbf{T}_{3 \times 3}$. Neste caso, o algoritmo dos oito pontos pode ser resumido como segue [29].

1. Realiza-se uma transformação nas coordenadas dos pontos da forma $\tilde{\tilde{\mathbf{m}}}_i = \mathbf{T} \tilde{\mathbf{m}}_i$ e $\tilde{\tilde{\mathbf{m}}}_i' = \mathbf{T}' \tilde{\mathbf{m}}_i'$;

2. Estima-se a matriz $\hat{\mathbf{F}}$ através da equação

$$\min_{\mathbf{F}} \sum_i (\tilde{\mathbf{m}}_i'^T \hat{\mathbf{F}} \tilde{\mathbf{m}}_i)^2, \text{ e;} \quad (3.13)$$

3. obtém-se $\mathbf{F} = \mathbf{T}'^T \hat{\mathbf{F}} \mathbf{T}$.

O método dos oito pontos fornece resultados extremamente exatos se for utilizado um conjunto grande de emparelhamentos. Contudo, a estimativa torna-se pobre à medida que o número de emparelhamentos utilizados aproxima-se de oito [29].

3.3.3 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 de 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 e FAUGERAS [48] em 1996. Nesta, \mathbf{F} é expressa em termos de quatro constantes a , b , c e d , e das coordenadas dos epipolos $e = (e_u, e_v)$ e $e' = (e'_u, e'_v)$, respectivamente, da primeira e segunda imagem:

$$\mathbf{F} = \begin{bmatrix} a & b & -ae_u - be_v \\ c & d & -ce_u - de_v \\ -ae'_u - ce'_v & -be'_u - de'_v & F_{33} \end{bmatrix}, \quad (3.14)$$

onde $F_{33} = (ae_u + be_v)e'_u + (ce_u + de_v)e'_v$.

Esta parametrização vem do fato dos epipolos poderem ser expressos como sendo [46]

$$\begin{aligned} e_u &= \frac{F_{23}F_{12} - F_{22}F_{13}}{F_{22}F_{11} - F_{21}F_{12}} \\ e_v &= \frac{F_{13}F_{21} - F_{11}F_{23}}{F_{22}F_{11} - F_{21}F_{12}} \\ e'_u &= \frac{F_{32}F_{21} - F_{22}F_{31}}{F_{22}F_{11} - F_{21}F_{12}} \\ e'_v &= \frac{F_{31}F_{12} - F_{11}F_{32}}{F_{22}F_{11} - F_{21}F_{12}} \end{aligned}, \quad (3.15)$$

caso eles não estejam no infinito.

Além da parametrização, LUONG e FAUGERAS [48] comprovaram experimentalmente que se obtém melhores resultados quando a , b , c e d são divididos pelo maior em valor absoluto. Esta divisão é possível porque \mathbf{F} é definida por um fator de escala.

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 e_u , e_v , e'_u e e'_v tem um valor infinito.

Para contornar este problema, ZHANG e LOOP [98] propuseram um novo método, no qual uma transformação projetiva é aplicada aos pontos m_i e m'_i , ou seja, $\tilde{\mathbf{m}}_i = \mathbf{T}_n \tilde{\mathbf{m}}_i$ e $\tilde{\mathbf{m}}'_i = \mathbf{T}'_n \tilde{\mathbf{m}}'_i$. As matrizes \mathbf{T}_n e \mathbf{T}'_n são tais que, no novo sistema de coordenadas, o primeiro elemento da matriz fundamental do novo sistema de coordenadas, $\hat{\mathbf{F}}$, é igual a 1 e os epipolos não estão no infinito. Isto é realizado da seguinte forma [98]

1. Inicializa-se \mathbf{T}_n e \mathbf{T}'_n igual a matriz identidade;
2. Dado um conjunto inicial de emparelhamento, obtém-se uma estimativa inicial \mathbf{F}_0 da matriz fundamental e dos epipolos $\tilde{\mathbf{e}}_0 = [e_u, e_v, e_t]^T$ e $\tilde{\mathbf{e}}'_0 = [e'_u, e'_v, e'_t]^T$;
3. Encontra-se o maior elemento de \mathbf{F}_0 em valor absoluto, denotado por (r_{\max}, c_{\max}) ;
4. Se $c_{\max} \neq 1$, permuta-se as linhas 1 e c_{\max} da matriz \mathbf{T}_n e permuta-se os elementos 1 e c_{\max} do epipolo $\tilde{\mathbf{e}}_0$;
5. Se $r_{\max} \neq 1$, permuta-se as linhas 1 e r_{\max} da matriz \mathbf{T}'_n e permuta-se os elementos 1 e r_{\max} do epipolo $\tilde{\mathbf{e}}'_0$;
6. Se $|e_v| > |e_t|$, permuta-se os elementos 2 e 3 do epipolo $\tilde{\mathbf{e}}_0$ e permuta-se as linhas 2 e 3 de \mathbf{T}_n , e;
7. Se $|e'_v| > |e'_t|$, permuta-se os elementos 2 e 3 do epipolo $\tilde{\mathbf{e}}'_0$ e permuta-se as linhas 2 e 3 de \mathbf{T}'_n .

Como todo critério de otimização não-linear, este método é dependente da distribuição do ruído dos pontos utilizados na minimização [93]. Normalmente, para a estimação da matriz fundamental, considera-se que o ruído tem uma distribuição gaussiana com média zero e matriz de covariância dada por [15]

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

onde σ é o, normalmente desconhecido, nível de ruído. Evidentemente, após a transformação proposta em [98], essa consideração não é mais válida. Por isso, ZHANG e LOOP [98] demonstraram como estimar \mathbf{F} conservando a distribuição original do ruído através de uma minimização ponderada por gradiente, ou seja, como descrito em [98],

$$\min_{\mathbf{F}} \sum_i \frac{(\tilde{\mathbf{m}}_i'^T \hat{\mathbf{F}} \tilde{\mathbf{m}}_i)^2}{\hat{\mathbf{l}}_i^T \mathbf{T}_n \mathbf{Z} \mathbf{T}_n^T \hat{\mathbf{l}}_i + \hat{\mathbf{l}}_i' \mathbf{T}_n' \mathbf{Z} \mathbf{T}_n'^T \hat{\mathbf{l}}_i'}, \quad (3.16)$$

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

Agora, com a parametrização da equação (3.14) e o critério de minimização da equação (3.16), 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 (3.13).

Além de ser computacionalmente muito complexo, este método ainda tem a limitação de ser muito dependente de \mathbf{F}_0 . Por isso, geralmente o algoritmo converge para um mínimo local da equação (3.16). Além disso, para um conjunto de pontos muito maior que oito, o método não-linear e o método da equação (3.11) têm resultados muito próximos [29].

3.3.4 Método do Paralaxe Virtual

Recentemente, BOUFAMA e MOHR [9] propuseram um novo método para calcular a matriz fundamental. O método é 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.

Como ilustrado na figura 3.2, considere um ponto M no espaço 3D e faça o ponto virtual Q ser a interseção do plano Π_{123} (definido pelos pontos M_1 , M_2 e M_3) com a linha \overline{CM} . Se o epipolo e' na segunda imagem é conhecido, então, um ponto m na primeira imagem corresponde a uma linha epipolar na segunda imagem dada por $\overline{e'q'}$, onde q' é a projeção do ponto Q na segunda imagem. Desde que a linha epipolar correspondente ao ponto m também é dada por $\mathbf{F}\tilde{\mathbf{m}}$, então, conhecendo-se $\overline{e'q'}$, o cálculo de \mathbf{F} é imediato.

Agora, considere H e H' duas homografias correspondentes a projeção do plano Π_{123} nos planos Π e Π' , na primeira e segunda imagens respectivamente, ou seja, $\Pi = H(\Pi_{123})$ e $\Pi' = H'(\Pi_{123})$. Desse modo, existe uma homografia $K = H'H^{-1}$, onde

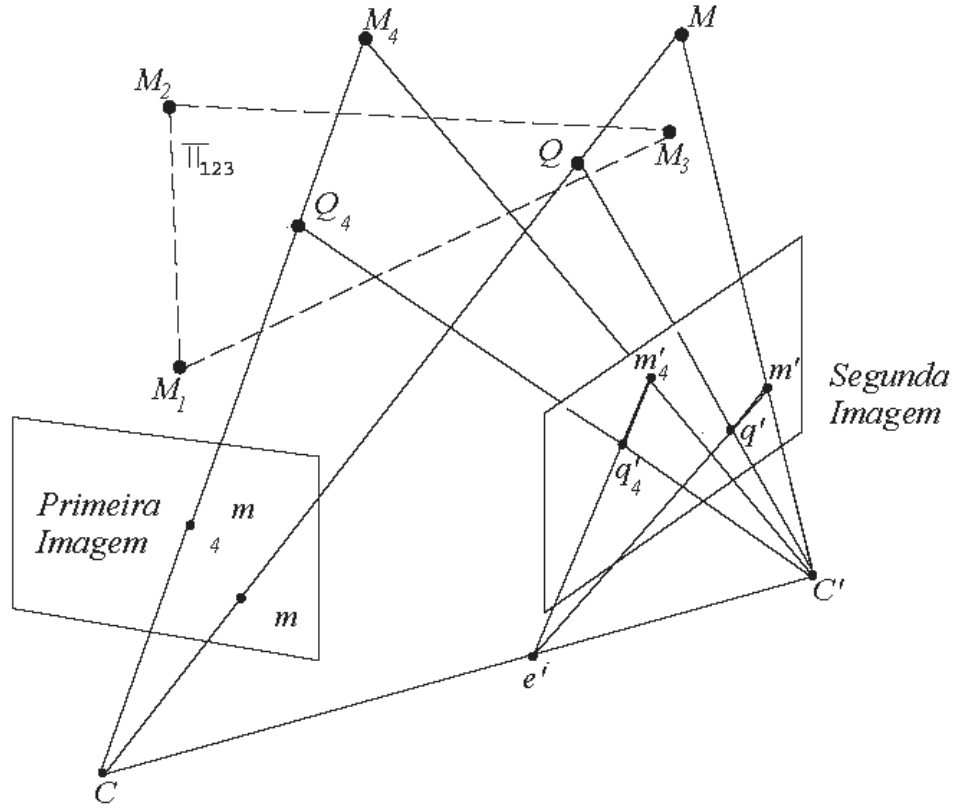


Figura 3.2: Geometria epipolar e o paralaxe virtual.

H^{-1} é a homografia inversa de H . Logo,

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

Se considerarmos a homografia K descrita por uma matriz $\mathbf{C}_{3 \times 3}$, tem-se

$$\forall m \in \Pi, \mathbf{C}\tilde{\mathbf{m}} = \lambda\tilde{\mathbf{q}}', \quad (3.17)$$

onde λ é um fator de escala desconhecido.

Desde que a matriz \mathbf{C} é homogênea, são necessárias oito equações independentes para defini-la. Contudo, para cada par $(\tilde{\mathbf{m}}_i, \tilde{\mathbf{m}}'_i)$, $i = 1, 2$ e 3 , igual as projeções nas duas imagens dos pontos M_1 , M_2 e M_3 , a equação (3.17) fornece duas equações independentes. Por isso, sem uma outra restrição, a matriz \mathbf{C} não pode ser calculada. Dessa forma, para o cálculo de \mathbf{C} , BOUFAMA e MOHR [9] propuseram uma mudança

no sistema de coordenadas de modo que, no novo sistema de coordenadas

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

onde $(\tilde{\mathbf{m}}_0, \tilde{\mathbf{m}}'_0)$ é um emparelhamento qualquer usado apenas para completar a base de \mathcal{P}^3 .

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

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

Agora, considerando a equação (3.17) e um emparelhamento entre $\tilde{\mathbf{m}} = [u, v, t]^T$ e $\tilde{\mathbf{m}}' = [u', v', t']^T$, tem-se

$$\mathbf{C}\tilde{\mathbf{m}} = \lambda[\alpha u, \beta v, t]^T. \quad (3.18)$$

Com α e β desconhecidos. Além disso, desde que, segundo a figura 3.2, q' pertence a linha epipolar $\overline{e'q'}$, podemos escrever

$$(\tilde{\mathbf{e}}' \times \tilde{\mathbf{m}}) \cdot \tilde{\mathbf{q}}' = 0, \quad (3.19)$$

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

De (3.18), a equação (3.19) pode ser escrita da forma [9]

$$a_i V_1 + b_i V_2 + c_i V_3 + d_i V_4 + e_i V_5 + f_i V_6 = 0, \quad (3.20)$$

onde $V_1 = \alpha e'_t$, $V_2 = \alpha e'_t$, $V_3 = \beta e'_u$, $V_4 = \beta e'_t$, $V_5 = e'_v$ e $V_6 = e'_u$, com e'_u , e'_v e e'_t sendo as coordenadas homogêneas do epipolo e' , ou seja, $\tilde{\mathbf{e}}' = [e'_u, e'_v, e'_t]^T$.

A equação (3.20) é linear e possui apenas cinco parâmetros independentes. Considerando os três pontos P_1 , P_2 e P_3 , utilizados para gerar o novo sistema de coordenadas (figura 3.2), é necessário um total de oito emparelhamentos de pontos para resolver a equação (3.20) e encontrar α , β e o epipolo $\tilde{\mathbf{e}}'$. Em seguida, a matriz funda-

mental é calculada através da equação

$$\mathbf{F} = \begin{bmatrix} 0 & -\beta e'_t & e'_v \\ \alpha e'_t & 0 & -e'_u \\ -\alpha e'_v & -\beta e'_u & 0 \end{bmatrix}. \quad (3.21)$$

Em nossos experimentos, observamos que, quando o número de emparelhamentos utilizados nos cálculos é próximo a 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 [17]. Isso acontece porque depositamos extrema confiança nos emparelhamentos (m_1, m'_1) , (m_2, m'_2) e (m_3, m'_3) , utilizados na formação do novo sistema de coordenadas. De fato, a exatidão do algoritmo está baseada na exatidão destes três pares iniciais e qualquer erro em suas coordenadas é transferido para as coordenadas dos outros pares (m_i, 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.

3.3.5 Métodos Robustos

Como discutido na seção 2.4, o conjunto de emparelhamentos utilizados na estimação de \mathbf{F} possui ruído e falsos emparelhamentos. O ruído é devido à grandes erros (maiores que três pixels) na localização 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.

O ruído e os falsos emparelhamento 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 com ruído. Exemplos desses métodos podem ser encontrados em [63], [76] e [84].

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 , utiliza-se uma técnica que forneça uma

boa estimação de \mathbf{F}_j com n pares (m_i, m'_i) .

3. Para cada \mathbf{F}_j , determina-se 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 sua linha epipolar $\mathbf{F}_j \tilde{\mathbf{m}}_i$ e $d(\tilde{\mathbf{m}}'_i, \mathbf{F}_j^T \tilde{\mathbf{m}}_i)$ é a distância do ponto \mathbf{m}'_i à sua linha epipolar $\mathbf{F}_j^T \tilde{\mathbf{m}}_i$ [46].

4. Retem-se a estimação \mathbf{F}_j que tem o menor M_j (denotado por M_{\min}) dentre todos os m M_j 's.
5. Elimina-se os pares (m_i, m'_i) que possuírem o resíduo

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

6. Finalmente, refina-se a matriz fundamental estimando-a sem os emparelhamentos com ruído.

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,

$$m = \frac{\log(1 - P)}{\log[1 - (1 - \varepsilon)^n]}. \quad (3.22)$$

No presente trabalho, no passo 5 do algoritmo anterior, $r_{th} = 1,99\hat{\sigma}$, onde $\hat{\sigma}$ é a estimativa do desvio padrão, como proposto por ROUSSEEUW [70], dada por

$$\hat{\sigma} = 1,4826 [1 + 5/(N - n)] \sqrt{M_{\min}}.$$

No passo 1, os emparelhamentos que compõem cada subconjunto devem estar bem espalhados pela área da imagem. Caso contrário, a estimação de \mathbf{F}_j é instável e inútil. Em vista disto, ZHANG *et al* [96] propuseram o método de “seleção aleatória regularmente distribuída”. O método pode ser descrito como segue.

1. Escolhe-se uma das imagens e calcula-se as coordenadas máxima e mínima dos pontos interessantes;

2. A imagem é dividida em $b \times b$ regiões. A cada região é associado o conjunto de pontos, e, indiretamente, o conjunto de emparelhamentos delimitado pela mesma. As regiões que não comportam nenhum ponto são descartadas;
3. Para escolher-se os emparelhamentos. Inicialmente, p regiões são escolhidas aleatoriamente. Em seguida, de cada região, um emparelhamento é escolhido também de forma aleatória.

Se for assumido que os emparelhamentos ruins, ou seja, os falsos emparelhamentos e o emparelhamentos com ruído, estão uniformemente distribuídos no espaço, as regiões têm o mesmo número de emparelhamentos e a seleção aleatória é uniforme, a equação (3.22) ainda vale [96].

Para evitar que um emparelhamento dentro de uma região com poucos pontos tenha uma probabilidade elevada de ser escolhida, é desejável que uma região com muitos emparelhamentos tenha probabilidade maior de ser selecionada do que uma com poucos emparelhamentos. Em vista disso, considerando o número total de regiões com emparelhamentos igual a l , divide-se o intervalo $[0 \ 1]$ em l intervalos, tal que, a largura, L_i , do i -ésimo intervalo seja

$$L_i = \frac{n_i}{\sum_{i=1, \dots, l} n_i},$$

onde n_i é o número de emparelhamentos dentro do i -ésimo intervalo. Durante a seleção da região, um número, produzido por um gerador aleatório uniforme no intervalo $[0 \ 1]$, que pertença ao i -ésimo intervalo implica que a i -ésima região foi selecionada.

O número de pontos n que compõem cada subconjunto (passo 1 do algoritmo) deve ser o menor possível. Caso contrário, pela equação (3.22), será necessário um número grande de subconjuntos. Por isso, é costume estimar-se $\mathbf{F}j$ pela equação (3.13), seguido de uma minimização não-linear da equação (3.16). 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 $n = 8$, examinou-se também a sua utilização na estimação de $\mathbf{F}j$.

Neste trabalho, nos experimentos realizados, as imagens foram divididas em 16 regiões, ou seja, $b = 4$. Observou-se que o método do paralaxe virtual torna-se bem mais preciso quando os emparelhamentos utilizadas no cálculo 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 posição 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, obtém-se ótimos resultados no passo 6 do algoritmo LMedS se for utilizado apenas 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 ele fornece uma matriz fundamental com uma distância média de até 10 vezes menor. Contudo, o seu tempo de computação é muito 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 \mathbf{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 [17].

3.4 Restrição epipolar para emparelhamento de pontos

Como discute-se na seção 2.4, se a matriz fundamental não for conhecida, não existe nenhuma restrição geométrica para os emparelhamentos. Por isso, pelo que está demonstrado na seção 2.5, todo algoritmo de emparelhamento sem a matriz fundamental é bastante susceptível à ruído. Mesmo assim, os emparelhamentos encontrados sem a restrição epipolar ainda podem ser utilizados para calcular a matriz fundamental desde que seja utilizado algum algoritmo robusto como, por exemplo, o da seção 3.3.5.

Dadas duas imagens com um conjunto de pontos interessantes marcados, os algoritmos para emparelhamento que não utilizam a restrição epipolar, além de considerarem erroneamente alguns falsos emparelhamentos como corretos, tendem a eliminar alguns bons emparelhamentos. Por isso, ZHANG *et al* [96], após estimar a matriz fundamen-

tal, utilizou a restrição epipolar sobre o conjunto inicial de pontos de interesse. Dessa forma, muitas falsos emparelhamentos são eliminadas no estágio inicial do algoritmo. Isso reduz bastante a ambiguidade e um algoritmo de emparelhamento aplicado aos pontos interessantes restantes tende a produzir um número maior de emparelhamentos e uma percentagem menor de pares com ruído [96].

Em resumo, a restrição epipolar pode ser utilizada para auxílio ao emparelhamento da seguinte forma

1. Utiliza-se um operador para selecionar pontos interessantes nas duas imagens;
2. Aplica-se um algoritmo de emparelhamento que não seja baseado na restrição epipolar, por exemplo, um dos discutidos na seção 2.5;
3. Calcula-se a matriz fundamental utilizando um algoritmo robusto (seção 3.3.5);
4. Aplica-se a restrição epipolar ao conjunto de pontos interessantes do passo 1, e;
5. Finalmente, aplica-se ao conjunto de pontos interessantes restantes, o mesmo algoritmo utilizado no passo 2.

Os passos 3, 4 e 5 foram aplicados as imagens da figura 2.11 e os resultados são apresentados na figura 3.3. Como vê-se, apesar do algoritmo não estar livre de cometer erros, no caso da figura 3.3, nenhum emparelhamento errado permaneceu.

3.5 Calibração Automática

As coordenadas 3D de um ponto M no espaço estão associadas as coordenadas 2D de sua projeção m no plano da câmera através da equação (2.7). Em vista disso, em estereoscopia, para determinar-se as coordenadas de um ponto no espaço, é necessário estimar a matriz de projeção da câmera através de um processo conhecido como calibração. Neste processo, é necessário conhecer-se com exatidão as coordenadas 3D de um conjunto de pontos M_i e as coordenadas 2D m_i e m'_i de suas projeções, respectivamente, na primeira e segunda imagens.

Na seção 2.2, apresenta-se uma classe de métodos de calibração de câmeras que fazem uso de um gabarito semelhante ao da figura 2.3 (página 15). Neste caso, desde que é realizada através de um objeto, cuja geometria no espaço 3D é conhecida com exatidão, a calibração é realizada de forma muito eficiente. Dentre estes métodos, o mais conhecido talvez seja o proposto por TSAI [86] em 1986.

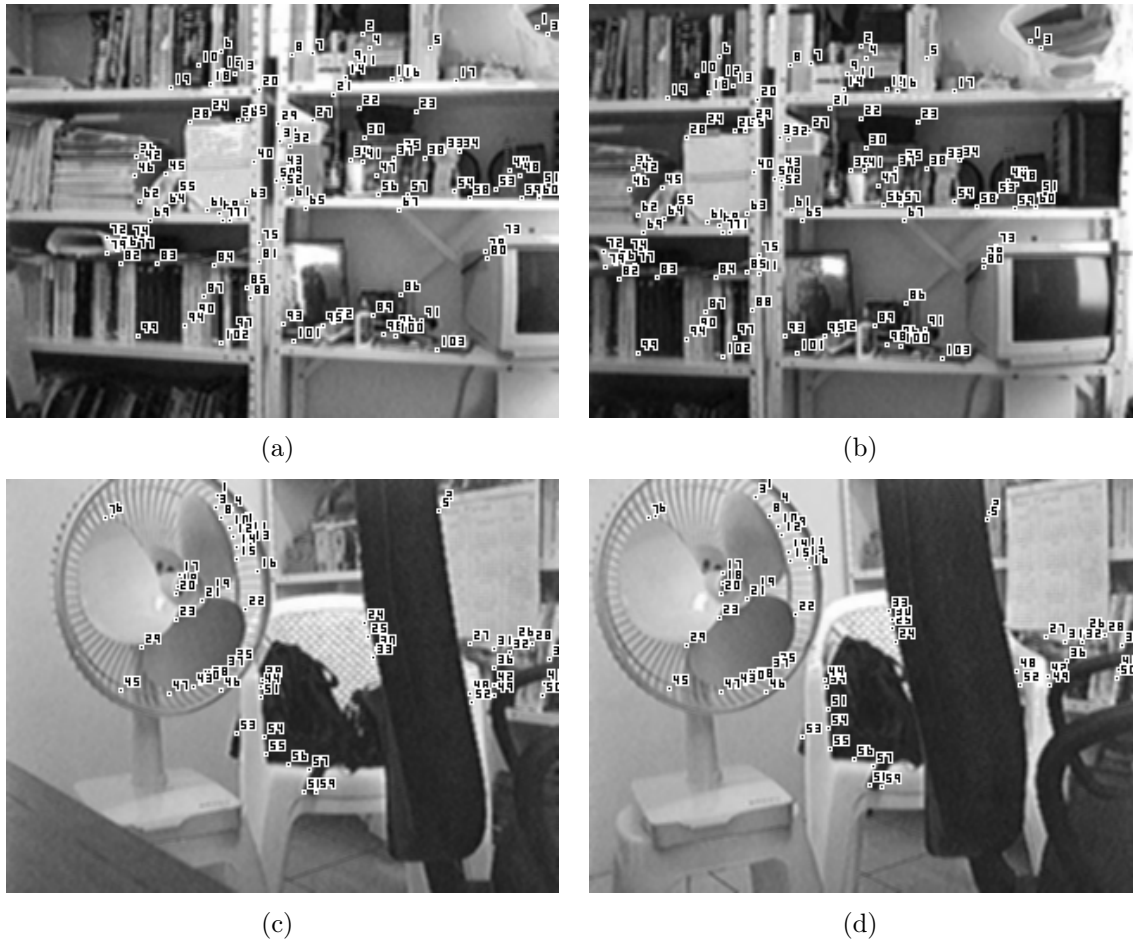


Figura 3.3: (a) e (b) formam um par de imagens com emparelhamentos marcados por relaxação e restrição epipolar. O par de imagens (c) e (d) é outro exemplo de correspondência pela mesma técnica.

Recentemente, foram propostas novas técnicas de calibração baseada em um gabarito composto de um único plano [81, 94, 101]. Nesta categoria, o gabarito precisa ser movido em frente as câmeras e várias imagens deste (em diferentes posições) devem ser capturadas. Contudo, o conhecimento do movimento realizado não é necessário. Este tipo de calibração é bem mais flexível do que as técnicas que utilizam um ou mais planos ortogonais [94].

Os métodos baseados em gabaritos são bastante exatos, mas possuem duas limitações importantes [21]:

- Nem sempre um objeto como o da figura 2.3 está disponível. Neste caso, é impossível realizar a calibração, e;
- Caso, durante o funcionamento normal do sistema, os parâmetros da câmera

forem alterados¹, a inspeção visual deve ser interrompida e o gabarito deve novamente ser posto no campo de visão da câmera.

Em vista disso, em 1992, MAYBANK e FAUGERAS [55] desenvolveram um novo método para calibração de câmeras. O método desenvolvido não necessita de gabarito e pode ser realizado durante a operação normal de inspeção visual. A câmera realiza três deslocamentos através de uma cena estática e uma imagem deve ser capturada em cada um desses deslocamentos. Se os parâmetros internos da câmera forem considerados constantes durante os três deslocamentos, através de um conjunto de emparelhamentos mapeados entre cada um dos três pares de imagens possíveis, pode-se estimar os parâmetros internos e externos da câmera [55], através das equações de KRUPPA [40] aplicadas à visão computacional. Contudo, a técnica proposta é muito susceptível à ruídos e não tem uso prático [99]. Felizmente, em 1997, FAUGERAS *et al* [21] propuseram modificações que aumentaram a exatidão da técnica e a tornou estável.

Apesar da facilidade de não necessitar de gabarito, a autocalibração exige a estimação de um grande número de parâmetros. Em vista disso, a técnica é matematicamente muito complexa [94]. Contudo, os algoritmos envolvidos são, em sua essência, paralelos. Assim, pode-se utilizar *clusters* de computadores para acelerar o cálculo. Entretanto, mesmo neste caso, o tempo de processamento ainda é elevado [47]. Felizmente, ZHANG *et al* [99] propôs um novo método de calibração automática de um par estereoscópico baseado no trabalho de FAUGERAS *et al* [21]. Neste método, o modelo da câmera *pinhole* é simplificado e um número maior de deslocamentos são realizados. Assim, as equações de KRUPPA [40] são substituídas por uma minimização não-linear um pouco menos complexa.

Apesar de algumas pesquisas ainda serem feitas na área de calibração automática, por exemplo, o trabalho de Yi Ma e seus associados [50–52], este campo perdeu um pouco o interesse após a consolidação dos trabalhos independentes de FAUGERAS [20] e HARTLEY *et al* [30].

¹Os parâmetros da câmera podem ser alterados por fatores como mudança de temperatura e trepidação do sistema. Outras vezes, eles podem ser modificados intencionalmente, por exemplo, no ajuste da distância focal (*zoom*).

3.6 Calibração relativa de câmeras

FAUGERAS [20] e HARTLEY *et al* [30] desenvolveram o que é hoje conhecido como *Weak Calibration*. Neste trabalho, este método é referenciado como “Calibração Relativa”.

A calibração relativa pode ser realizada por fatoração da matriz fundamental ou em relação a uma base projetiva. A seguir, ambos os métodos são discutidos. Contudo, se desejado, outros métodos podem ser vistos em [69].

3.6.1 Por fatoração da matriz fundamental

Uma possível fatoração para a matriz fundamental é a da equação (3.7). Neste caso, \mathbf{P} e \mathbf{P}' podem ser quaisquer matrizes 3×4 que satisfaçam a equação (3.6). Uma solução viável é considerar a representação canônicas [31] de \mathbf{P} e \mathbf{P}' , ou seja,

$$\mathbf{P}_w = \begin{bmatrix} \mathbf{I} & \mathbf{0}_3 \end{bmatrix} \quad \text{e} \quad \mathbf{P}'_w = \begin{bmatrix} \mathbf{Q} & \mathbf{e}' \end{bmatrix}.$$

Do trabalho de LUONG e VIÉVILLE [49], sabe-se que

$$\mathbf{Q} = -\frac{[\mathbf{e}']_{\times}}{\|\mathbf{e}'\|^2} \mathbf{F}.$$

A equação anterior está em função do epipolo \mathbf{e}' , logo ele deve ser estimado. Como discute-se na seção 3.3, deve haver emparelhamentos ruins na estimação de \mathbf{F} . Por isso, não se deve utilizar a equação (3.15) para estimar-se os epipolos. Um método que trás melhores resultados [21] é através da equação

$$\min_{\tilde{\mathbf{e}}} \|\mathbf{F}\tilde{\mathbf{e}}\|^2 \quad \text{sujeito à} \quad \|\tilde{\mathbf{e}}\|^2 = 1. \quad (3.23)$$

Neste caso, $\tilde{\mathbf{e}}$ é igual ao autovetor da matriz $\mathbf{F}^T \mathbf{F}$ que corresponde ao menor autovalor [92].

3.6.2 Relativa a uma base projetiva

Neste caso, para realizar a calibração necessita-se apenas de uma imagem de cada câmera (I e I') e, no mínimo, cinco emparelhamentos entre elas. Com estes emparelhamentos, define-se três sistemas de coordenadas distintos, ou seja, considerando que $\tilde{\mathbf{M}}_1, \tilde{\mathbf{M}}_2, \dots, \tilde{\mathbf{M}}_5$ sejam as coordenadas de cinco pontos no espaço 3D, $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2, \dots, \tilde{\mathbf{m}}_5$

sejam as coordenadas das projeções desses pontos na imagem I e $\tilde{\mathbf{m}}'_1, \tilde{\mathbf{m}}'_2 \dots \tilde{\mathbf{m}}'_5$ sejam as projeções dos mesmos pontos na imagem I' , pode-se fazer arbitrariamente:

1. $\widetilde{\mathbf{M}}_1 = \tilde{\mathbf{e}}_1 = [1, 0, 0, 0]^T$, $\widetilde{\mathbf{M}}_2 = \tilde{\mathbf{e}}_2 = [0, 1, 0, 0]^T \dots \widetilde{\mathbf{M}}_5 = \tilde{\mathbf{e}}_5 = [1, 1, 1, 1]^T$, onde $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2 \dots \tilde{\mathbf{e}}_5$ formam a base canônica do espaço \mathcal{P}^4 ;
2. $\tilde{\mathbf{m}}_1 = [1, 0, 0]^T$, $\tilde{\mathbf{m}}_2 = [0, 1, 0]^T \dots \tilde{\mathbf{m}}_4 = [1, 1, 1]^T$, ou seja, $\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2 \dots \tilde{\mathbf{m}}_4$ formam a base canônica do espaço \mathcal{P}^3 , e;
3. $\tilde{\mathbf{m}}'_1 = [1, 0, 0]^T$, $\tilde{\mathbf{m}}'_2 = [0, 1, 0]^T \dots \tilde{\mathbf{m}}'_4 = [1, 1, 1]^T$, ou seja, a base canônica do espaço \mathcal{P}^3 .

Segundo TOMASI [82], a única exigência para que os pontos $\widetilde{\mathbf{M}}_1, \widetilde{\mathbf{M}}_2 \dots \widetilde{\mathbf{M}}_5$ forme uma base do espaço \mathcal{P}^4 é que, dentre eles, não haja um grupo de quatro pontos co-planares. Assim, supondo-se que $M_1, M_2 \dots M_4$ (figura 3.4) são co-planares, eles formam um quadrilátero, cujas diagonais se intersectam nos pontos B_1, B_2 e B_3 . Neste caso, como o modelo *pinhole* mapeia linhas em \mathcal{P}^4 em linhas em \mathcal{P}^3 , as imagens desses pontos formam os quadriláteros $m_1, m_2 \dots m_4$ e $m'_1, m'_2 \dots m'_4$. Além disso, as diagonais desses quadriláteros se intersectam, respectivamente, em b_1, b_2, b_3 e b'_1, b'_2, b'_3 . Neste caso, segundo FAUGERAS [20], a condição de que b'_1 pertença a linha epipolar de b_1 é necessária e suficiente para que os pontos $M_1, M_2 \dots M_4$ sejam co-planares.

No novo sistema de coordenadas formado pelos pontos $\widetilde{\mathbf{M}}_1, \widetilde{\mathbf{M}}_2 \dots \widetilde{\mathbf{M}}_5$ e suas projeções, pode-se calcular a matriz de projeção da primeira câmera através da equação (2.7), ou seja,

$$\mathbf{P}_w \widetilde{\mathbf{M}}_i = \rho_i \tilde{\mathbf{m}}_i \quad i = 1, 2, \dots, 4. \quad (3.24)$$

Nesta equação, ρ_i ($i = 1, 2, \dots, 4$) são constante que servem para explicitar que, como trata-se de uma equação homogênia, $\mathbf{P}_w \widetilde{\mathbf{M}}_i$ pode diferir de $\tilde{\mathbf{m}}_i$ por uma constante.

Da equação (3.24), encontra-se a matriz de projeção como sendo

$$\mathbf{P}_w = \begin{bmatrix} \rho_1 & 0 & 0 & \rho_4 \\ 0 & \rho_2 & 0 & \rho_4 \\ 0 & 0 & \rho_3 & \rho_4 \end{bmatrix}. \quad (3.25)$$

Agora, considerando $\tilde{\mathbf{m}}_5 = [\alpha, \beta, \gamma]^T$, $\mu = \rho_5$ e $\nu = \rho_4$, tem-se [20]

$$\mathbf{P}_w = \mu \mathbf{X} + \nu \mathbf{Y}, \quad (3.26)$$

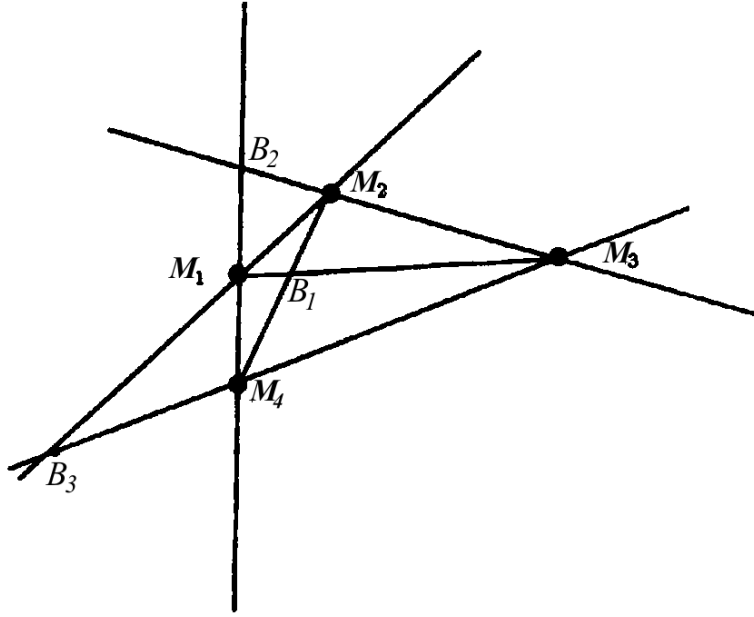


Figura 3.4: Quatro pontos co-planares M_1, M_2, \dots, M_4 formam um quadrilátero, cujas diagonais se intersectam nos pontos B_1, B_2 e B_3 .

onde

$$\mathbf{X} = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \gamma & 0 \end{bmatrix} \text{ e } \mathbf{Y} = \begin{bmatrix} -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

De forma similar,

$$\mathbf{P}'_w = \mu' \mathbf{X}' + \nu' \mathbf{Y}. \quad (3.27)$$

Desde que \mathbf{P}_w e \mathbf{P}'_w são homogêneas, pode-se considerar $\lambda = \mu/\nu$ e $\lambda' = \mu'/\nu'$. Assim, das equações (3.26) e (3.27), tem-se as matrizes de projeção das duas câmeras em função de apenas duas constantes, λ e λ' . Para encontrar λ e λ' , deve-se usar as equações² $\mathbf{P}_w \tilde{\mathbf{C}} = 0$ e $\mathbf{P}'_w \tilde{\mathbf{C}} = \tilde{\mathbf{e}}'$, onde $\tilde{\mathbf{C}}$ é o centro da primeira câmera e $\tilde{\mathbf{e}}'$ o epipolo da imagem I' . Estas equações fornecem [20]

$$\tilde{\mathbf{e}}' = \begin{bmatrix} \frac{\lambda' \alpha' - \lambda \alpha}{1 - \lambda \alpha} & \frac{\lambda' \beta' - \lambda \beta}{1 - \lambda \beta} & \frac{\lambda' \gamma' - \lambda \gamma}{1 - \lambda \gamma} \end{bmatrix}^T. \quad (3.28)$$

²Estas equações são bastantes intuitivas e podem ser comprovadas observando-se a figura 3.1.

De forma semelhante, pode-se conseguir

$$\tilde{\mathbf{e}} = \begin{bmatrix} \frac{\lambda\alpha - \lambda'\alpha'}{1 - \lambda'\alpha'} & \frac{\lambda\beta - \lambda'\beta'}{1 - \lambda'\beta'} & \frac{\lambda\gamma - \lambda'\gamma'}{1 - \lambda'\gamma'} \end{bmatrix}^T. \quad (3.29)$$

Novamente, as equações (3.28) e (3.29) estão em função de λ e λ' . Contudo, se forem conseguidas oito ou mais emparelhamentos de pontos, pode-se calcular a matriz fundamental das duas imagens. Assim, os epipolos podem ser estimados à partir da matriz \mathbf{F} .

Agora, com o epipolo $\tilde{\mathbf{e}}' = [e'_u, e'_v, e'_t]^T$ e as equações (3.28) e (3.29), obtém-se [20]

$$\lambda = \frac{\tilde{\mathbf{e}}' \cdot (\tilde{\mathbf{m}}_5 \times \tilde{\mathbf{m}}'_5)}{\tilde{\mathbf{e}}'_5 \cdot (\tilde{\mathbf{m}}_5 \times \tilde{\mathbf{m}}'_5)} \quad \text{e} \quad \lambda' = \lambda \frac{\gamma e'_u e'_v (\beta - \alpha) + \alpha e'_v e'_t (\gamma - \beta) + \beta e'_t e'_u (\alpha - \gamma)}{\gamma' e'_u e'_v (\beta - \alpha) + \alpha' e'_v e'_t (\gamma - \beta) + \beta' e'_t e'_u (\alpha - \gamma)}, \quad (3.30)$$

onde $\tilde{\mathbf{e}}'_5 = [\alpha e'_u, \beta e'_v, \gamma e'_t]^T$.

Novamente, a equação (3.30) está em função do epipolo \mathbf{e}' . Estimando-o pela equação (3.23), λ e λ' podem ser conhecidos através de (3.30). Em seguida, \mathbf{P}_w e \mathbf{P}'_w podem ser calculadas diretamente pelas equações (3.26) e (3.27).

Como pode-se notar, na calibração relativa a uma base projetiva, as coordenadas dos pontos no ambiente do robô são conhecidas em relação a um sistema de coordenadas arbitrário, mas relacionado ao sistema de coordenadas do ambiente do robô por uma transformação projetiva \mathbf{T}_w desconhecida. De forma semelhante, na calibração relativa por fatoração da matriz fundamental, existem diversas matrizes que satisfazem a equação (3.6). De fato, se $\mathbf{Q} = \mathbf{P}'_w \mathbf{P}_w^+$ é solução, então $\mathbf{Q}_v = \mathbf{P}'_w \mathbf{P}_w^+ + \mathbf{e}' \mathbf{v}^T$, onde \mathbf{v} é um vetor qualquer, também é solução [97]. Por isso, as matrizes de projeção reais \mathbf{P} e \mathbf{P}' não são conhecidas. Contudo, HARTLEY *et al* [30] provaram que estas estão relacionadas por uma transformação projetiva \mathbf{T}_w desconhecida.

Teoricamente, os métodos propostos por FAUGERAS [20] e HARTLEY *et al* [30] devem fornecer o mesmo par de matrizes de projeção. Contudo, o método de HARTLEY *et al* [30] é mais robusto com relação ao ruído presente nos emparelhamentos utilizados. Além disso, o método de FAUGERAS [20] pode ter sua exatidão comprometida se os pontos que formam a base projetiva sejam escolhidos de forma errada. De fato, o ideal é que os pontos que compõem a base de \mathcal{P}^4 estejam bem espalhados pelo plano da imagem.

A calibração relativa é um método simples, intuitivo e não-iterativo, que permite calibrar um par estereoscópico e estimar a localização de um ponto no espaço \mathcal{P}^4 . Contudo, como discute-se na próxima seção, se as coordenadas 3D de um conjunto de

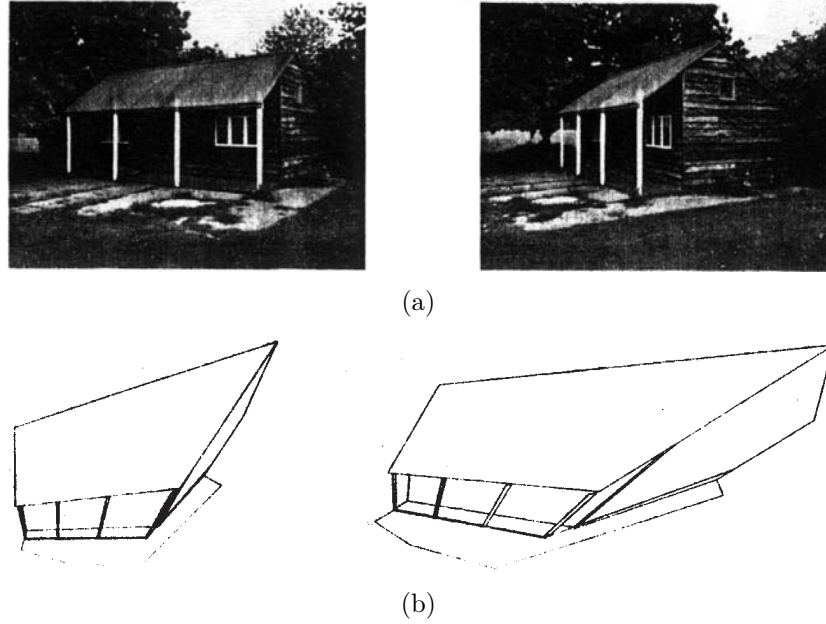


Figura 3.5: (a) as imagens originais de um par estereoscópico e (b) a reconstrução 3D das imagens através da calibração relativa.

pontos do ambiente da câmera forem conhecidas, a transformação projetiva pode ser conhecida [30]. Assim, as coordenadas dos pontos podem ser conhecidas no espaço euclidiano 3D.

Como ilustração, apresenta-se na figura 3.5(a) um par de imagens estereoscópicas de uma casa de campo. A sua reconstrução 3D, através da calibração relativa, é apresentada na figura 3.5(b). Como pode-se observar, a reconstrução é deformada. Contudo, caso sejam conhecidas as coordenadas 3D (em relação ao sistema de coordenadas do ambiente) de alguns pontos da figura 3.5(a), a transformação \mathbf{T}_w pode ser calculada e a deformação da figura 3.5(b) pode ser removida.

3.6.3 Estimação da transformação projetiva

Dado um emparelhamento (m_i, m'_i) , através da calibração relativa é possível estimar um ponto M_{wi} no espaço com relação a uma base de \mathcal{P}^4 . Como dicute-se na seção 3.6.2, as coordenadas deste ponto no sistema de coordenadas do ambiente da câmera podem ser encontradas através da relação

$$\lambda_i \widetilde{\mathbf{M}}_{wi} = \mathbf{T}_w \widetilde{\mathbf{M}}_i, \quad (3.31)$$

onde \mathbf{T}_w é uma matriz 4×4 que representa uma transformação projetiva desconhecida e λ_i é uma constante também desconhecida.

Considerando $\mathbf{v} = \mathbf{T}_w \widetilde{\mathbf{M}}_i = [v_1, v_2, v_3, v_4]^T$ e $\widetilde{\mathbf{M}}_{wi} = [X_{wi}, Y_{wi}, Z_{wi}, T_{wi}]^T$, a equação (3.31) fornece

$$\begin{aligned} Y_{wi}v_1 - X_{wi}v_2 &= 0 \\ Z_{wi}v_1 - X_{wi}v_3 &= 0 \\ T_{wi}v_1 - X_{wi}v_4 &= 0 \end{aligned} \quad (3.32)$$

Agora, fazendo $\mathbf{x} = [\mathbf{T}_{w11}, \mathbf{T}_{w12}, \dots, \mathbf{T}_{w44}]^T$, onde \mathbf{T}_{wij} é o elemento da linha i e coluna j da matriz \mathbf{T}_w , pode-se demonstrar que a equação (3.32) torna-se

$$\mathbf{B}_i \mathbf{x} = \mathbf{0}, \quad (3.33)$$

onde \mathbf{B}_i é uma matriz 3×16 dada por

$$\mathbf{B}_i = \begin{bmatrix} Y_{wi}\widetilde{\mathbf{M}}_i^T & -X_{wi}\widetilde{\mathbf{M}}_i^T & \mathbf{0}_3^T & \mathbf{0}_3^T \\ Z_{wi}\widetilde{\mathbf{M}}_i^T & \mathbf{0}_3^T & -X_{wi}\widetilde{\mathbf{M}}_i^T & \mathbf{0}_3^T \\ T_{wi}\widetilde{\mathbf{M}}_i^T & \mathbf{0}_3^T & \mathbf{0}_3^T & -X_{wi}\widetilde{\mathbf{M}}_i^T \end{bmatrix}.$$

Com n emparelhamentos (m_i, m'_i) , tem-se n equação da forma (3.33). Assim, desde que \mathbf{T}_w é homogênea, pode-se encontrar \mathbf{x} resolvendo

$$\min_{\mathbf{x}} \|\mathbf{B}_n \mathbf{x}\|^2, \quad \text{sujeito a } \|\mathbf{x}\| = 1. \quad (3.34)$$

Pelo que discute-se na seção 2.2.1, a solução desse problema é o autovetor de $\mathbf{B} = \mathbf{B}_n^T \mathbf{B}_n$ que corresponde ao menor autovalor.

Na estimação de \mathbf{T}_w , pode-se utilizar métodos não-lineares. Contudo, se as coordenadas dos mapeamentos $\widetilde{\mathbf{M}}_{wi} \leftrightarrow \widetilde{\mathbf{M}}_i$ forem conhecidas com exatidão, por exemplo, através do uso de um gabarito de calibração, o método da equação (3.34) fornece resultados muito próximos aos conseguidos através de abordagens não-lineares [94].

3.7 Retificação das Imagens

Uma vez que as matrizes de projeção são conhecidas o sistema de visão estereoscópico está pronto para ser utilizado. Contudo, este só é útil se, dado um ponto m no plano da imagem I , for possível identificar o seu correspondente m' no plano de imagem



Figura 3.6: Dois exemplos de imagens com linhas epipolares esboçadas.

I' . Felizmente, a matriz fundamental torna a tarefa de emparelhamento bem menos complexa, pois reduz o espaço de busca de um plano para uma linha. Em outras palavras, dado um ponto m no plano da imagem I , a restrição epipolar garante que o seu correspondente m' está sobre a linha epipolar $l' = F\tilde{m}$ no plano de imagem I' . Contudo, em geral, as linhas epipolares não são paralelas e nem alinhadas com os eixos de coordenadas do plano de imagem. Para ilustrar este fato, apresenta-se na figura 3.6 duas imagens com algumas de suas linhas epipolares esboçadas.

Se as linhas epipolares não forem horizontais e paralelas, durante o processo de emparelhamento, consome-se um tempo desnecessário realizando comparações entre *pixels* e retas inclinadas no plano da imagem [23, 25, 26, 43]. Como se vê, a complexidade do processo de emparelhamento de pontos reduz-se significativamente se as linhas epipolares são paralelas e alinhadas com o eixo horizontal da imagem. Isto pode ser feito através de uma transformação projetiva 2D, ou seja, uma homografia, em ambas as imagens. Normalmente, este processo é chamado de retificação.

O processo de retificação pode ser entendido como o processo de transformar a matriz fundamental de um par de imagens na sua forma canônica. Isto é feito aplicando-se uma homografia que transfere os epipolos das duas imagens para o infinito, ou seja, $e_r = e'_r = [1, 0, 0]^T$, onde e_r e e'_r são os epipolos das imagens retificadas. Neste caso, a

matriz fundamental transforma-se em

$$\mathbf{F}_r = [\tilde{\mathbf{e}}_r]_{\times} = [\tilde{\mathbf{e}}'_r]_{\times} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (3.35)$$

onde $[\mathbf{x}]_{\times}$ é uma matriz que representa o produto cruzado com o vetor linha \mathbf{x} [43].

Agora, considerando que as homografias aplicadas as imagens I e I' para realizar a retificação são, respectivamente, \mathbf{H}_r e \mathbf{H}'_r , para um ponto m em I e um ponto m' em I' , tem-se

$$\tilde{\mathbf{m}}_r = \mathbf{H}_r \tilde{\mathbf{m}} \quad \text{e} \quad \tilde{\mathbf{m}}'_r = \mathbf{H}'_r \tilde{\mathbf{m}}'. \quad (3.36)$$

Contudo, da equação (3.3), tem-se

$$\tilde{\mathbf{m}}'^T \mathbf{F}_r \tilde{\mathbf{m}}_r = 0. \quad (3.37)$$

Então, de (3.36) e (3.37), tem-se

$$\tilde{\mathbf{m}}'^T \mathbf{H}_r'^T \mathbf{F}_r \mathbf{H}_r \tilde{\mathbf{m}} = 0,$$

ou seja,

$$\mathbf{F} = \mathbf{H}_r'^T \mathbf{F}_r \mathbf{H}_r = \mathbf{H}_r'^T [\tilde{\mathbf{e}}_r]_{\times} \mathbf{H}_r. \quad (3.38)$$

Existem infinitas homografias \mathbf{H}_r e \mathbf{H}'_r que satisfazem a equação (3.38). Contudo, normalmente escolhe-se \mathbf{H}_r e \mathbf{H}'_r que minimizam a distorção entre as imagens retificadas. Tal distorção ocorre quando os epípolos da imagem original estão perto do centro da imagem ou, pelo menos, estão na sua região visível. Um exemplo de distorção causado pela retificação da imagem pode ser visto na figura 3.7.

LOOP e ZHANG [43] desenvolveram um método para minimizar a distorção. Para isto, cada homografia \mathbf{H}_r e \mathbf{H}'_r é decomposta em três homografias. A primeira é responsável pela retificação em si, e as outras duas são calculadas de forma a minimizar a distorção. Contudo, o método envolve minimizações não-lineares e o cálculo das raízes de um polinômio de grau sete.

No presente trabalho, as câmeras utilizadas no par estereoscópico estão posicionadas em planos verticais quase paralelos. Neste caso, os epípolos estão muito distantes do centro das imagens e a distorção pode ser negligenciada [23, 25, 26]. Por isso, métodos mais simples de retificação podem ser utilizados. Um método possível é o proposto recentemente por FUSIELLO [26].

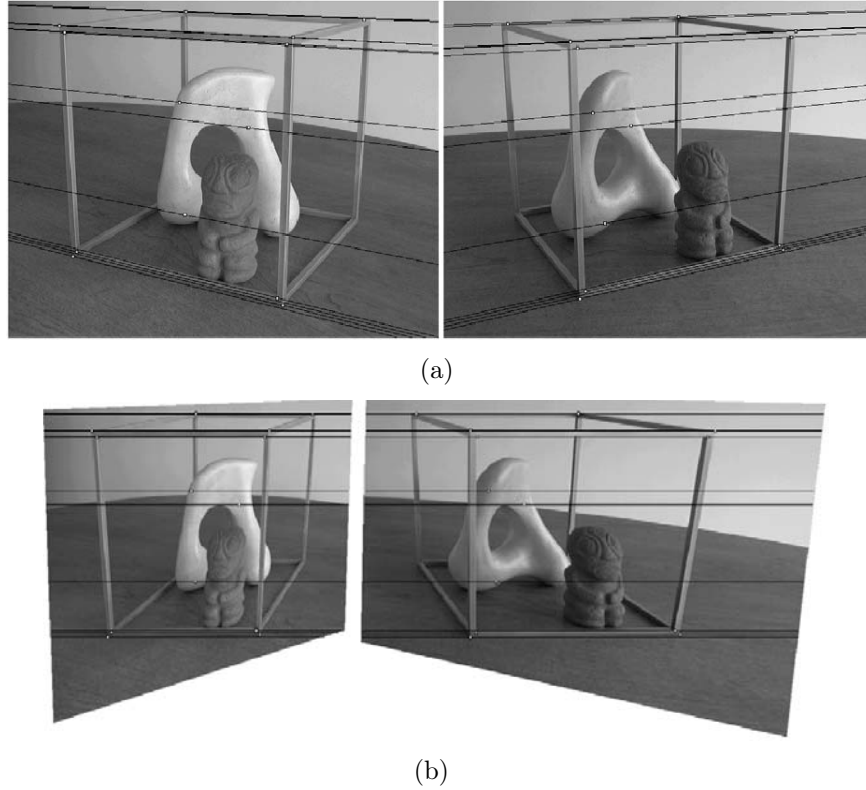


Figura 3.7: (a) Imagens originais com linhas epipolares oblíquas e não-paralelas. (b) As mesmas imagens retificadas, mas com distorção.

O algoritmo de FUSIELLO [26] calcula \mathbf{H}_r e \mathbf{H}'_r através das matrizes de projeção originais (\mathbf{P} e \mathbf{P}') e retificadas (\mathbf{P}_r e \mathbf{P}'_r) das duas câmeras. As matrizes \mathbf{P}_r e \mathbf{P}'_r são da forma

$$\mathbf{P}_r = \mathbf{A}_r[\mathbf{R}_r | -\mathbf{R}_r\tilde{\mathbf{C}}] \quad \mathbf{P}'_r = \mathbf{A}_r[\mathbf{R}_r | -\mathbf{R}_r\tilde{\mathbf{C}}'],$$

onde $\tilde{\mathbf{C}}$ e $\tilde{\mathbf{C}}'$ são os centros de projeção³ das duas imagens, \mathbf{A}_r é a matriz 3×3 que corresponde aos parâmetros intrínsecos de \mathbf{P}_r e \mathbf{P}'_r e pode ser feita igual a identidade e

$$\mathbf{R}_r = \begin{bmatrix} \mathbf{r}_{r1}^T \\ \mathbf{r}_{r2}^T \\ \mathbf{r}_{r3}^T \end{bmatrix}.$$

³Conhecidas as matrizes de projeção \mathbf{P} e \mathbf{P}' , os centros de projeção podem ser estimados pelas equações $\mathbf{P}\tilde{\mathbf{C}} = 0$ e $\mathbf{P}'\tilde{\mathbf{C}}' = 0$.

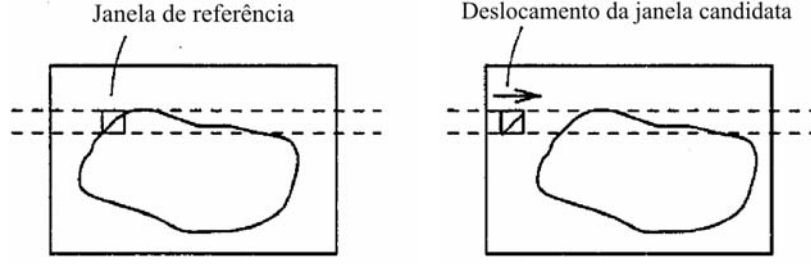


Figura 3.8: Emparelhamento de pontos em imagens retificadas.

Por sua vez (de [26]),

$$\begin{aligned}\mathbf{r}_{r1} &= (\tilde{\mathbf{C}} - \tilde{\mathbf{C}}') / \|\tilde{\mathbf{C}} - \tilde{\mathbf{C}}'\|; \\ \mathbf{r}_{r2} &= \mathbf{z} \times \mathbf{r}_{r1}, \text{ onde } \mathbf{z}^T = [0, 0, 1], \text{ e;} \\ \mathbf{r}_{r3} &= \mathbf{r}_{r1} \times \mathbf{r}_{r2}.\end{aligned}$$

Por último, \mathbf{H}_r e \mathbf{H}'_r podem ser obtidas através de

$$\mathbf{H}_r = \mathbf{A}_r \mathbf{R}_r (\mathbf{A} \mathbf{R})^{-1} \quad \text{e} \quad \mathbf{H}'_r = \mathbf{A}_r \mathbf{R}_r (\mathbf{A}' \mathbf{R}')^{-1}.$$

Uma vez que as imagens estejam retificadas, dado um ponto m , o seu correspondente m' tem a mesma coordenada y . Desse forma, m' pode ser encontrada maximizando algum critério de semelhança que deve ser calculado dentro de janelas centralizadas em m e m' [27].

3.8 Emparelhamento em imagens retificadas

Em imagens retificadas, o correspondente de um ponto $m = (u, v)$ tem coordenada $y = v$. Além disso, tendo como referência a imagem I (imagem da esquerda), a coordenada x do ponto correspondente $m' = (u', v)$ está sempre à direita de u , ou seja, $u' > u$. Essas restrições reduzem bastante o espaço de busca. Por isso, após a retificação, uma medida de semelhança é suficiente para realizar o emparelhamento de pontos.

Normalmente, como ilustrado na figura 3.8, a imagem I é tomada como referência. Então, para encontrar o correspondente de um ponto m na outra imagem, uma medida de semelhança é aplicada entre uma janela de correlação de referência centrada em m (imagem I) e uma janela de correlação que desloca-se horizontalmente da esquerda para a direita na imagem I' .

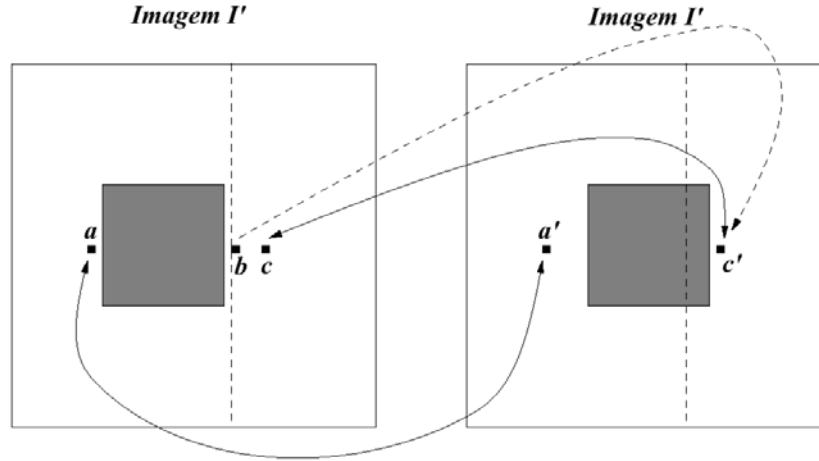


Figura 3.9: Restrição esquerda-direita. O ponto b não tem um correspondente na imagem I' , mas o algoritmo de emparelhamento pode formar dois pares (b, c') e (c, c') . Contudo, tomando I' como referência, apenas o par (c, c') é formado.

Na seção 2.4.2, apresenta-se alguns medidas que quantificam o grau de semelhança entre dois pontos de imagens diferentes. Infelizmente, além do problema de oclusão, as distorções radiométricas e projetivas podem fazer com que o emparelhamento por semelhança forneça resultados falsos.

Na oclusão, devido a diferenças na disparidade entre objetos que estão a distâncias distintas da câmera, a imagem de um ponto 3D é encoberta em uma das imagens por objetos mais próximos. Neste caso, o ponto não possui um correspondente na outra imagem. Contudo, o algoritmo de emparelhamento pode produzir um resultado errado baseado no ponto da outra imagem que apresenta a maior semelhança com o ponto em questão.

Para reduzir a influência da oclusão no emparelhamento de pontos, FUSIELLO [24] utilizou a restrição esquerda-direita. A figura 3.9 ilustra a aplicação desta restrição. Nesta, o ponto b não tem um correspondente na imagem I' , mas, tomando a imagem da esquerda I como referência, o algoritmo de emparelhamento pode formar dois pares (b, c') e (c, c') . Contudo, considerando I' como referência, apenas o par (c, c') é formado. Isto indica que o par (b, c') é falso e deve ser descartado.

Em seu sistema, FUSIELLO [24] efetua o emparelhamento por semelhança aplicando a restrição esquerda-direita em todos os pontos da imagem. Isso minimiza a formação de falsos emparelhamentos e reforça a restrição de unicidade. Contudo, essa abordagem pode não ser aceitável em sistemas que possuem restrição de tempo.

Uma suposição inerente ao processo de emparelhamento por semelhança é que a

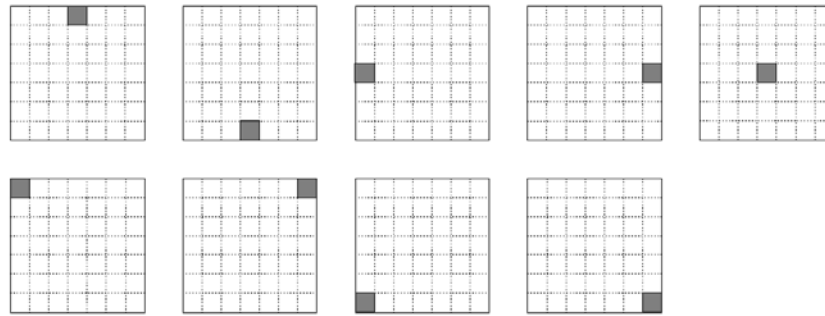


Figura 3.10: Configuração de múltiplas janelas para emparelhamento por semelhança. A correlação é calculada dentro das áreas escuras.

imagem possui disparidade contínua [33]. Contudo, regiões próximas as bordas dos objetos que compõem uma imagem podem ser regiões de descontinuidade. Quando a janela de correlação abrange uma região sem disparidade constante, o emparelhamento por semelhança falha e obtém-se um erro na estimação da profundidade. Esse problema aumenta com o tamanho da janela de correlação. Por outro lado, uma janela de correlação pequena possui pouca variação de tons de cinza e, portanto, uma baixa relação sinal/ruído. KANADE e OKUTOMI [37] abordaram esse problema através de uma janela adaptativa, cujo tamanho e forma variam de acordo com a região da imagem. Contudo, essa técnica necessita de uma estimação inicial do mapa de disparidade da cena e é muito lenta para uma implementação em tempo-real quando não utiliza *hardware* dedicado.

O método de múltiplas janelas é computacionalmente mais eficiente para minimizar a influência de quebra na disparidade. Uma configuração comum deste método é calcular a semelhança dentro de nove janelas de tamanho diferentes (figura 3.10), mas centradas em diferentes posições, para um mesmo ponto de interesse. Após o cálculo das nove correlações, considera-se apenas aquele que fornece a maior semelhança. Este método foi utilizado, por exemplo, por FUSIELLO [24] e pode ser entendido como uma simplificação do método proposto por KANADE e OKUTOMI [37], com possibilidade de ser utilizado em aplicações de tempo-real [33].

Além de reduzir a influência da descontinuidade na disparidade, o método de múltiplas janelas ameniza o problema das distorções projetivas. Contudo, como discute-se na seção 2.4.2, o critério de semelhança também é sensível a distorções radiométricas. Para contornar esse problema, pode-se utilizar medidas de correlação normalizadas como, por exemplo, $ZNCC$ [equação (2.22)]. No entanto, essas normalizações exigem muito processamento e acabam tendo muita influência no tempo total de processamento.

Além disso, isso não elimina completamente a influência das distorções radiométricas. Em vista disso, ZABIH e WOODFILL [33] introduziram o conceito de grandezas não-paramétricas no emparelhamento de pontos. Para isto, eles criaram as transformadas rank e census.

Na transformada rank, uma janela é deslocada por todos os pontos da imagem e, em cada ponto, conta-se o número de pixels dentro da janela que possui intensidade menor que o ponto central da mesma. Por isso, a transformada rank transforma uma imagem em tons de cinza em uma matriz de inteiros entre 0 e $N - 1$, onde N é o número de pixels dentro da janela.

A transformada rank pode ser expressa por

$$R(u, v) = N - \sum_{(i,j) \in W} U[I(u + i, v + j) - I(u, v)],$$

onde

$$U[t] = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

e $(i, j) \in W$ são todos os vizinhos do ponto central $m = (u, v)$ que estão dentro da janela W da transformada ranque.

Desde que a transformação rank efetua uma normalização na imagem, o emparelhamento de pontos pode ser realizado, simplesmente, através de semelhança por *SAD* [equação (2.21)] de forma semelhante como ilustrado na figura 3.8.

A transformada census transforma a imagem em uma matriz de seqüências de bits. Nesta transformada, o valor de cada pixel dentro da janela de transformação é comparado com o valor do ponto central. Se este for maior, um “1” é inserido na posição correspondente da seqüência de bits. Caso contrário, o bit correspondente é “0”. Após a transformação census, o emparelhamento deve ser realizado através da distância hamming, ou seja, são considerados correspondentes os pontos, cujas seqüências de bits é a mais semelhante.

O trabalho de ZABIH e WOODFILL [33] mostra que a transformada rank fornece melhores resultados que a transformada census. De fato, vários trabalhos comprovam o sucesso da transformada na redução da influência das distorções radiométricas [1–3]. Contudo, ainda existem poucos trabalhos sobre o assunto.

Por ser baseada apenas em comparações de valores dos pixels que compõem as imagens, o emparelhamento de pontos por transformada rank e *SAD* é ideal para aplicações de tempo-real e implementações por *hardware* [39, 90].

Recentemente, BROWN *et al* [10] e SCHARSTEIN e SZELISKI [73] publicaram excelentes trabalhos sobre pesquisas na área de emparelhamento por correlação.

3.9 Triangulação

O método de triangulação descrito pela equação (2.19) é conhecido como algoritmo do ponto médio. Apesar de simples, como demonstrado pelo trabalho de BEARDSLEY *et al* [5], este método não fornece bons resultados. Isto ocorre porque, devido a erros na digitalização da imagem ou ruídos presentes no algoritmo de emparelhamento, dado um emparelhamento (m, m') no plano das imagens do par estereoscópico, os seus raios ópticos não se intersectam em nenhum ponto do espaço 3D. Em outras palavras, o par (m, m') não satisfaz a restrição epipolar resumida pela equação (3.3). Nesta seção, dois outros métodos de triangulação são discutidos.

3.9.1 Método linear

Dado duas matrizes de projeção \mathbf{P} e \mathbf{P}' , respectivamente, da imagem I e I' , e um emparelhamento (m, m') , tem-se as seguintes equações

$$\tilde{\mathbf{m}} = s\mathbf{P}\tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}' = s'\mathbf{P}'\tilde{\mathbf{M}}, \quad (3.39)$$

onde, sem perda de generalidade, considerou-se $\tilde{\mathbf{m}} = [u, v, 1]^T$ e $\tilde{\mathbf{m}}' = [u', v', 1]^T$. Além disso, em (3.39), s e s' são constantes desconhecidas e necessárias por a equação (3.39) ser homogênea. Agora, considerando $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]^T$ e $\mathbf{P}' = [\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{p}'_3]^T$, tem-se

$$1/s = \mathbf{p}_3^T \tilde{\mathbf{M}} \quad \text{e} \quad 1/s' = \mathbf{p}_3'^T \tilde{\mathbf{M}}. \quad (3.40)$$

Substituindo (3.40) em (3.39), pode-se escrever

$$\mathbf{U}\tilde{\mathbf{M}} = \mathbf{0}, \quad (3.41)$$

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

$$\mathbf{U} = [\mathbf{p}_1 - u\mathbf{p}_3, \mathbf{p}_2 - v\mathbf{p}_3, \mathbf{p}'_1 - u'\mathbf{p}'_3, \mathbf{p}'_2 - v'\mathbf{p}'_3]^T. \quad (3.42)$$

Como bem sabe-se, desde que $\tilde{\mathbf{M}}$ está em coordenadas homogêneas, pode-se impor $\|\tilde{\mathbf{M}}\| = 1$. Assim, a solução de (3.42) é um autovetor de $\mathbf{U}^T\mathbf{U}$ que corresponde ao

menor autovalor.

A equação (3.41) fornece um resultado exato apenas se o emparelhamento (m, m') satisfaz a equação $\tilde{\mathbf{m}}' \mathbf{F} \tilde{\mathbf{m}} = 0$. Contudo, ruídos nos emparelhamentos podem comprometer a exatidão da estimação de $\tilde{\mathbf{M}}$. Mesmo assim, equação (3.41) pode ser utilizada, por exemplo, quando emparelhamentos forem conseguidos com auxílio de um gabarito de calibração [97].

3.9.2 Método ótimo

O processo de triangulação pode ser formulado como sendo encontrar o par de pontos (\hat{m}, \hat{m}') que minimize

$$d(m, \hat{m})^2 + d(m', \hat{m}')^2, \quad \text{sujeito à } \tilde{\mathbf{m}}'^T \mathbf{F} \tilde{\mathbf{m}} = 0, \quad (3.43)$$

onde $d(x, y)$ é a distância euclidiana do ponto x ao ponto y ou do ponto x à reta y .

Ou ainda, a restrição anterior pode ser re-escrita como

$$d(m, \hat{\mathbf{l}})^2 + d(m', \hat{\mathbf{l}}')^2, \quad (3.44)$$

onde $\hat{\mathbf{l}} = \mathbf{F}^T \tilde{\mathbf{m}}'$ e $\hat{\mathbf{l}}' = \mathbf{F} \tilde{\mathbf{m}}$.

Para resolver o problema da equação (3.44), em 1996, HARTLEY e STURM [32] propuseram uma mudança no sistema de coordenadas das duas imagens de maneira que, neste novo sistema, os pontos $\tilde{\mathbf{m}}$ e $\tilde{\mathbf{m}}'$ são iguais a $\tilde{\mathbf{m}}_t = \tilde{\mathbf{m}}'_t = [0, 0, 1]^T$ e os epipolos da imagem tornam-se $\tilde{\mathbf{e}}_t = \tilde{\mathbf{e}}'_t = [1, 0, t]^T$. A mudança no sistema de coordenadas é realizada através de homografias compostas por duas matrizes 3×3 : \mathbf{T}_t e \mathbf{R}_t . Por exemplo, para a primeira imagem e considerando $\tilde{\mathbf{m}} = [u, v, 1]^T$, tem-se

$$\mathbf{T}_t = \begin{bmatrix} 1 & 0 & -u \\ 0 & 1 & -v \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.45)$$

que é responsável pela translação $\tilde{\mathbf{m}}$ e

$$\mathbf{R}_t = \begin{bmatrix} \cos \Theta & -\sin \Theta & 0 \\ \sin \Theta & \cos \Theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.46)$$

responsável pela rotação do epipolo $\tilde{\mathbf{e}} = [e_u, e_v, e_t]$ em torno da origem.

Para calcular \mathbf{R}_t , pode-se utilizar a igualdade

$$\mathbf{R}_t \mathbf{T}_t \tilde{\mathbf{e}} = [1, 0, t]^T.$$

Isto fornece a restrição

$$\sin \Theta(e_u - e_t u) + \cos \Theta(e_v - e_t v) = 0. \quad (3.47)$$

As matrizes \mathbf{R}'_t e \mathbf{T}'_t (relacionadas à segunda imagem) podem ser calculadas por equações semelhantes as (3.45), (3.46) e (3.47) [32].

A matriz fundamental \mathbf{F}_t do novo sistema de coordenadas relaciona-se com a matriz fundamental original \mathbf{F} pela equação

$$\mathbf{F}_t = \mathbf{R}'_t \mathbf{T}'_t{}^{-T} \mathbf{F} \mathbf{T}_t^{-1} \mathbf{R}_t^T.$$

Contudo, desde que $\mathbf{F}_t [1, 0, t]^T = [1, 0, t']^T \mathbf{F}_t = 0$, sabe-se que \mathbf{F}_t é da forma

$$\mathbf{F}_t = \begin{bmatrix} t'td & -t'c & -t'd \\ -tb & a & b \\ -td & c & d \end{bmatrix}.$$

Agora, para resolver a equação (3.44), HARTLEY e STURM [32] consideraram $\hat{\mathbf{I}}$ como a reta que passa pelo epipolo $\tilde{\mathbf{e}}_t = [1, 0, t]^T$ e o ponto $[0, \varphi, 1]^T$, ou seja,

$$\hat{\mathbf{I}}(\varphi) = [0, \varphi, 1]^T \times [1, 0, t]^T = [t\varphi, 1, -\varphi]^T. \quad (3.48)$$

Assim,

$$d(u_t, \hat{\mathbf{I}}(\varphi))^2 = \frac{\varphi^2}{1 + (t\varphi)^2}. \quad (3.49)$$

Além disso, desde que

$$\hat{\mathbf{I}}'(\varphi) = \mathbf{F}_t [0, \varphi, 1]^T = [-t'(c\varphi + d), a\varphi + b, c\varphi + d]^T. \quad (3.50)$$

Então,

$$d(u'_t, \hat{\mathbf{I}}'(\varphi))^2 = \frac{(c\varphi + d)^2}{(a\varphi + b)^2 + t'^2(c\varphi + d)^2}. \quad (3.51)$$

De (3.49) e (3.51), a equação (3.44) pode ser re-escrita como

$$T(\varphi) = \frac{\varphi^2}{1 + (t\varphi)^2} + \frac{(c\varphi + d)^2}{(a\varphi + b)^2 + t'^2(c\varphi + d)^2}. \quad (3.52)$$

Por sua vez, o mínimo da equação (3.52) é encontrado calculando-se as raízes da sua derivada.

Uma vez encontrado φ que minimiza a equação (3.52), através de (3.48) e (3.50), pode-se encontrar as retas que minimizam (3.44). Segundo HARTLEY e STURM [32], os pontos \hat{m} e \hat{m}' que minimizam a equação (3.43) são, respectivamente, as projeções ortogonais de m em $\hat{\Gamma}$ e de m' em $\hat{\Gamma}$.

O método descrito tem a vantagem de ser não-iterativo e fornecer o resultado ótimo. Contudo, tem a desvantagem de ser (computacionalmente) complexo, pois envolve a busca das raízes de um polinômio de grau seis. Outros métodos de triangulação podem ser encontrados em [69].

Capítulo 4

Contribuições para Estereoscopia Aplicada à Robótica Móvel

Neste trabalho, pretende-se aplicar à robótica móvel o que está discutido nos capítulos anteriores. São apresentadas as características do sistema de visão estereoscópica proposto, os objetivos a serem alcançados e as contribuições pretendidas.

4.1 Método rápido para emparelhamento de pontos sem calibração

Neste trabalho, foram testadas diversas técnicas de emparelhamento de pontos [28, 65, 74, 75, 96], mas apenas a técnica baseada em correlação e relaxação proposta por ZHANG *et al* [96] (seção 2.5) apresentou bons resultados para o caso geral, no qual o sistema não está calibrado e diversos pontos interessantes de uma imagem não formam par com nenhum ponto da outra imagem. Contudo, a técnica proposta em [96] é iterativa, complexa e, dependendo da ambigüidade inicial gerada pelo algoritmo de emparelhamento por correlação, consome muito tempo de processamento [65].

Em particular, como discute-se na seção 2.5, o método de emparelhamento baseado em SVD, mesmo com as alterações propostas por PILU [65], não apresenta bons resultados para o caso geral. Contudo, no presente trabalho, propõe-se alterações que fornecem melhores resultados tanto na exatidão dos emparelhamentos, como no tempo de computação da SVD:

1. com um conjunto de pontos interessantes marcados nas duas imagens, realiza-se o emparelhamento por correlação;

2. elimina-se do conjunto de pontos de interesse aqueles que não possuem nenhum possível correspondente, ou seja, a correlação com nenhum ponto da outra imagem é maior que um limiar C_{th} ;
3. com os pontos interessantes que restarem, na montagem da matriz \mathbf{G} , os pares (m_i, m'_j) que possuem uma correlação menor que um limiar C_{th} têm $G_{ij} = NC$, onde NC é uma constante que atenua ao máximo a contribuição do par (m_i, m'_j) na matriz \mathbf{G} , e;
4. calcula-se a SVD de \mathbf{G} e monta-se a matrix $\hat{\mathbf{G}}$.

Os testes mostram que, na maioria das vezes, a matriz \mathbf{G} calculada dessa forma tem a metade da dimensão da matriz calculada da forma proposta em [74] e [65]. Isso torna o cálculo da decomposição SVD bem mais rápido.

Nas figuras 4.1(a) e 4.1(b) são apresentados resultados deste novo método para os pontos interessantes das figuras 2.6(c) e 2.6(d). Já nas figuras 4.1(c) e 4.1(d), apresenta-se os resultados para outro par de imagens. Como pode-se observar, agora muitos emparelhamentos foram marcados corretamente, mas ainda existem muitos falsos emparelhamentos.

Aplicando o algoritmo de ZHANG *et al.* [96] (seção 2.5) as mesmas imagens, também obtém-se alguns falsos emparelhamentos. Contudo, tem-se a força de cada emparelhamento dado pela equação 2.26. Assim, pode-se calcular a força média dos emparelhamentos e eliminar aquelas que estão muito abaixo da média. Por isso, nos casos da figura 2.11, nenhum falso emparelhamento foi marcado.

No caso do algoritmo proposto neste trabalho, a força dos emparelhamentos encontrados também pode ser calculada. Assim, os emparelhamentos mais fracos podem ser eliminados acrescentando-se um quinto passo ao algoritmo proposto. Dessa forma, acredita-se que ele produzirá tão poucos falsos emparelhamentos quanto os produzidos pelo algoritmo de ZHANG *et al.* [96], mas requerendo bem menos poder de processamento.

Como discute-se na seção 3.8, a transformada rank reduz o efeito das distorções radiométricas e é ideal para aplicações de tempo-real, pois possibilita o uso da correlação baseada em *SAD*. Por isso, nesse trabalho pretende-se aplicar a transformada rank em uma janela em torno dos pontos de interesse. Dessa forma, em conjunto com a matrix $\hat{\mathbf{G}}$ e eliminando-se os emparelhamentos “fracos”, espera-se obter um algoritmo rápido e que produza um grande número de emparelhamentos.

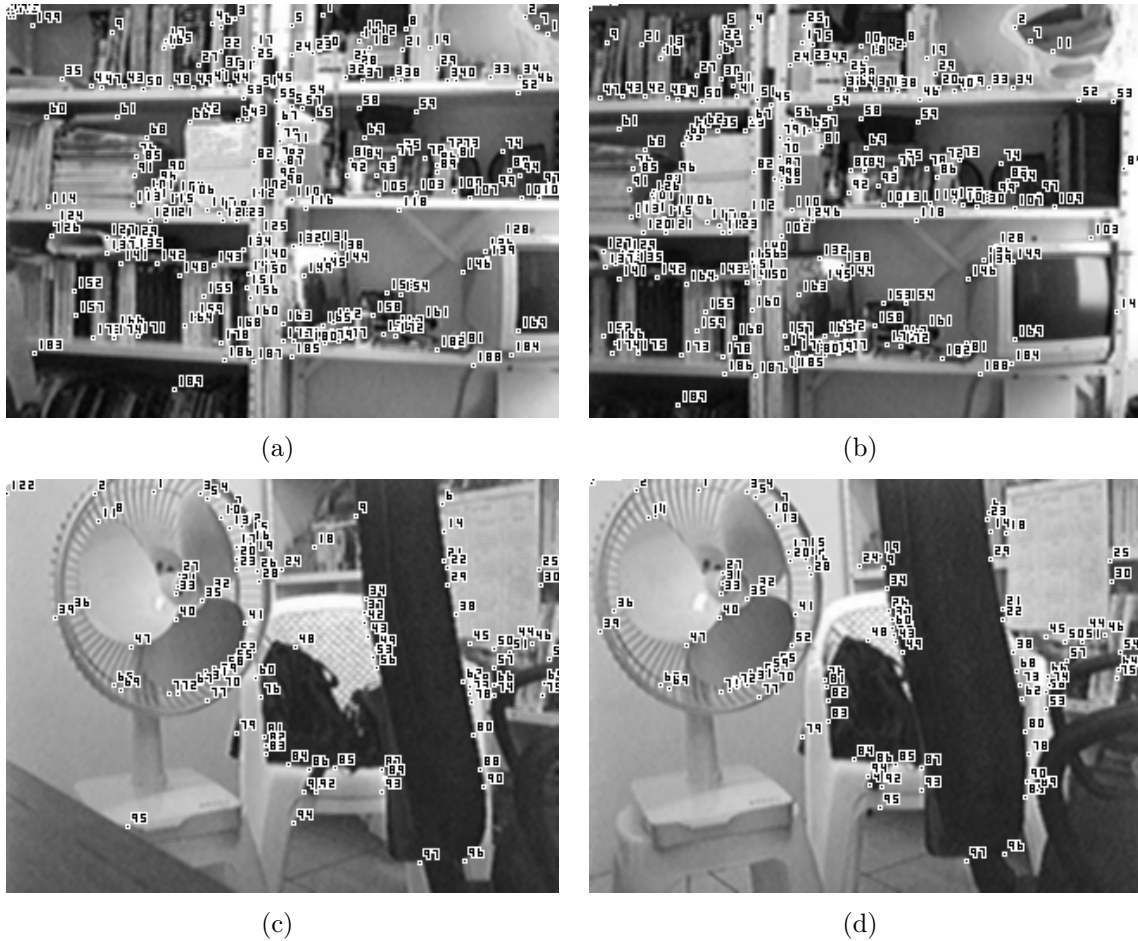


Figura 4.1: Imagens com emparelhamento encontrados pelo método de SCOTT e LONGUET-HIGGINS, mas com as modificações propostas neste trabalho.

Por último, desde que a transformada rank não elimina o problema causado pelas descontinuidades na disparidade da imagem, pretende-se utilizar a técnica de múltiplas janelas no cálculo do *SAD*. Para isto, serão utilizadas cinco janelas que se sobrepõem de acordo com a figura 4.2. Então, o *SAD* considerado será composto pela soma do *SAD* da janela central, SAD_0 , com os das duas janelas periféricas que possuírem a maior correlação, SAD_1 e SAD_2 , ou seja

$$SAD = SAD_0 + SAD_1 + SAD_2.$$

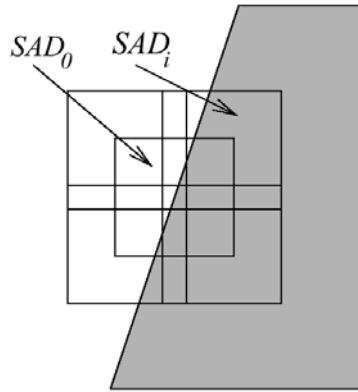


Figura 4.2: Configuração proposta de múltiplas janelas para emparelhamento de pontos sem calibração.

4.2 Método robusto para o cálculo da matriz fundamental

Com alguns emparelhamentos de pontos encontrados pelo método da seção anterior, pode-se estimar a matriz fundamental. Como visto na seção 3.3.5, devido à presença de ruído e falsos emparelhamentos, deve-se utilizar um método robusto para esta estimação.

Ao contrário do que foi feito em [96], os resultados da seção 3.3.5 (que também estão disponíveis em [17]) mostram que não é necessário utilizar métodos não-lineares para a estimação de cada \mathbf{F}_j (passo 2 do algoritmo da seção 3.3.5). Por isso, neste trabalho pretende-se utilizar o método robusto LMedS baseado no paralaxe virtual para a estimação da matriz fundamental, e não em algum método não-linear.

Nos testes realizados, observou-se que o método do paralaxe virtual fornece resultados muito bons se os emparelhamentos estiverem bem espalhados pelas imagens. Por isso, pretende-se adequar o método de seleção aleatória de emparelhamentos proposto em [96] para a técnica do paralaxe virtual. Com isso, espera-se que o método dê melhores resultados e, portanto, tenha-se um método rápido e robusto para cálculo da matriz fundamental.

Evidentemente, uma vez que a matriz fundamental tenha sido estimada robustamente, como foi descrito na seção 3.4, ela pode ser utilizada para garantir que não haja emparelhamentos com ruído.

Este método robusto para o cálculo da matriz fundamental, em conjunto com o algoritmo de emparelhamento da seção anterior, deverá fornecer um excelente método

robusto para emparelhamento de pontos em imagens sem calibração.

4.3 Método flexível para calibração de um par estereoscópico

As técnicas existentes de calibração automática ainda não são viáveis [94, 95]. Os algoritmos que utilizam essa abordagem são computacionalmente muito complexos, pois necessitam estimar muitos parâmetros. Por isso, nem sempre os resultados são aceitáveis [95].

A calibração com gabarito é simples, mas pouco flexível. Ela tem que ser realizada antes da operação do robô e, caso haja alguma mudança nos parâmetros do par estereoscópico, a operação do sistema deve ser interrompida para que o gabarito seja posto novamente a frente das câmeras. Não é necessário dizer que isto reduz bastante a autonomia do robô.

Pela figura 3.5, observa-se que com a calibração relativa é possível saber onde as retas se encontram e o paralelismo entre elas é conservado. Contudo, os ângulos entre as retas são alterados e perde-se informações de comprimento. Se estas informações forem necessárias, um gabarito deve ser utilizado.

Em vista disso, trabalhos recentes, como o de ZHANG *et al* [97], utilizam um gabarito embarcado no robô e que ocupa uma parte do campo de visão da par estereoscópico. Esta abordagem resolve os dois problemas, ou seja, torna a calibração simples e aumenta a autonomia do robô. Contudo, um último problema permanece. Na calibração com gabarito, o modelo da câmera é ajustado a um conjunto de dados. Além dos dados utilizados no ajuste possuírem ruído, o modelo ajustado descreve bem apenas a região na qual os pontos coletados se encontram [97]. Por isso, a validade do modelo do par estereoscópico estimado torna-se menor a medida que se afasta do gabarito. Isto afeta principalmente a exatidão da matriz fundamental (se está for calculada a partir das matrizes de projeção estimadas). Normalmente, isso não é observado porque a maioria dos trabalhos sobre calibração de câmeras com gabarito limitam-se a analisar o quanto o modelo da câmera ajusta-se aos pontos do gabarito [77].

Como discute-se na seção 3.3, a matriz fundamental pode ser estimada a partir de um conjunto de emparelhamentos qualquer. Por sua vez, estes emparelhamentos podem ser estabelecidos em todo o campo de visão da câmera, mesmo com o sistema descalibrado e, normalmente, os emparelhamentos obtidos dessa forma estão a

diferentes distâncias da câmera. Em vista disso, a matriz fundamental e o modelo da câmera estimados na calibração relativa deveria ser válido para uma distância maior. Contudo, falsos emparelhamentos e ruídos podem comprometer a qualidade da matriz fundamental estimada e, por conseguinte, da calibração relativa. Por isso, desde que os emparelhamentos mapeados através do gabarito são bastante exatos, uma abordagem que trás melhores resultados é utilizar emparelhamentos do ambiente da câmera e do gabarito no cálculo da matriz fundamental. Assim, a matriz fundamental e o modelo da câmera obtido através de calibração relativa representa com exatidão uma faixa maior de distância.

Embarca-se um gabarito no veículo móvel de forma que este ocupe parte do campo de visão do robô nem sempre é possível. O gabarito precisa estar a uma distância considerável da câmera para que está não ocupe todo o campo de visão. Por exemplo, no trabalho de ZHANG *et al* [97], o gabarito está a 1,5 m das câmeras. Assim, está abordagem poder ser inadmissível em veículos autônomos de pequeno e médio porte. Por isso, o ideal é que o gabarito seja compacto o suficiente para que seja inserido a frente das câmeras pelo próprio sistema de visão e apenas no momento da calibração. Em vista disso, no presente trabalho pretende-se investigar a utilização de objetos de uma dimensão na estimação da transformação \mathbf{T}_w . Na figura 4.3, o esboço do gabarito proposto é apresentado.

Por ser leve e compacto, um gabarito de uma dimensão pode ser facilmente manuseado pelo próprio sistema de visão e, mesmo assim, fornecer um conjunto de pontos, cujas coordenadas são conhecidas com grande exatidão.

4.3.1 Fundamentação teórica do problema

Como discute-se na seção 3.6.3, a matriz \mathbf{T}_w é uma matriz 4×4 . Contudo, evidentemente trata-se de uma matriz homogênea. Por isso, existem 15 parâmetros desconhecidos e que necessitam ser estimados.

Cada ponto de um gabarito utilizado para calibração fornece duas equações na forma de matrizes, ou seja,

$$\tilde{\mathbf{m}} = s_m \mathbf{P}_w \mathbf{T}_w \tilde{\mathbf{M}} \quad \text{e} \quad \tilde{\mathbf{m}}' = s_{m'} \mathbf{P}'_w \mathbf{T}_w \tilde{\mathbf{M}}, \quad (4.1)$$

onde s_m e $s_{m'}$ são duas constantes desconhecidas. Contudo, considerando que o ponto $\tilde{\mathbf{M}}$ não está no infinito, devido a restrição co-planar (seção 3.1), a equação (4.1) fornece apenas três equações independentes. Assim, por exemplo, pode-se utilizar um gabarito

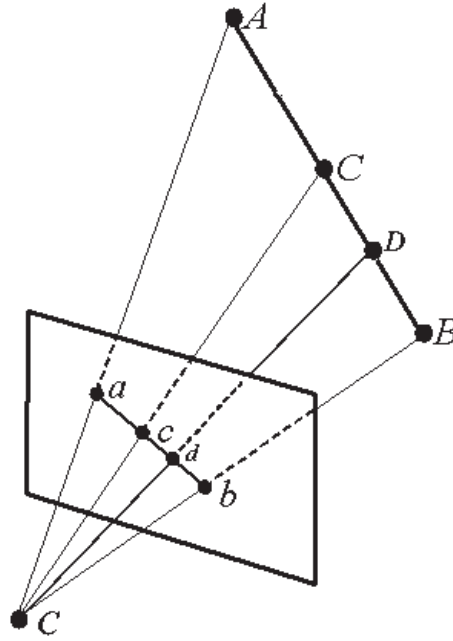


Figura 4.3: Exemplo do gabarito de uma única dimensão utilizado na técnica de calibração proposta. Por simplicidade, apenas um dos planos de imagem foi esboçado.

composto por dois pontos (A e B) separados por uma distância L conhecida. Nesta caso, necessita-se de três parâmetros para especificar o ponto A , mas apenas dois para o ponto B . Com isso, incluindo a matriz \mathbf{T}_w , tem-se $15 + 3 + 2 = 20$ parâmetros desconhecidos, mas apenas $3 + 3 = 6$ equações independentes. Contudo, com N imagens, tem-se $15 + 3N + 2N = 15 + 5N$ parâmetros desconhecidos e $3N + 3N = 6N$ equações. Assim, se o gabarito for movido a frente das câmeras e 15 imagens de cada câmera forem capturadas, é possível realizar a calibração. Contudo, dessa forma, a técnica não é flexível devido ao grande número de imagens necessárias.

Adicionando mais pontos ao gabarito, o número de parâmetros a serem estimados continua o mesmo, pois estes pontos são combinações lineares de A e B . Contudo, adicionando mais pontos, tem-se mais equações independentes. Por exemplo, como um terceiro ponto C tem a restrição da colinearidade entre A e B , ele fornece $3N - 1N = 2N$ equações independentes. Um quarto ponto D possui as restrições de colinearidade e razão cruzada¹ entre A , B e C . Assim, D fornece apenas $3N - 1N - 1N = N$ equações independentes. Dessa forma, com um gabarito de quatro pontos e N imagens, tem-se

¹Ao contrário da geometria euclidiana, na geometria projetiva a razão de comprimentos sobre uma reta varia sobre transformações. Contudo, a razão de razões (ou razão cruzada) é invariante a transformações projetivas. Assim, se um ponto m_i em uma reta for representado em coordenadas homogêneas por $\tilde{m}_i = [m_{i1} \ m_{i2}]^T$, dado quatro pontos m_1, m_2, m_3 e m_4 , pertencentes a mesma reta,

$15 + 5N$ incógnitas e $3N + 3N + 2N + N = 9N$ equações independentes. Então, com quatro imagens é possível efetuar a calibração.

4.3.2 Proposta para solução do problema

Considerando uma reta em \mathcal{P}^3 com quatros pontos A , B , C e D , as projeções desses pontos em I como sendo a , b , c , d , e em I' como sendo a' , b' , c' , d' , \mathbf{T}_w pode ser estimada resolvendo-se a seguinte equação não-linear

$$\sum_{i=1}^N \left[d(a_i, \hat{\mathbf{l}}_{a_i})^2 + d(b_i, \hat{\mathbf{l}}_{b_i})^2 + d(c_i, \hat{\mathbf{l}}_{c_i})^2 + d(d_i, \hat{\mathbf{l}}_{d_i})^2 + \dots \right. \\ \left. \dots + d(a'_i, \hat{\mathbf{l}}'_{a_i})^2 + d(b'_i, \hat{\mathbf{l}}'_{b_i})^2 + d(c'_i, \hat{\mathbf{l}}'_{c_i})^2 + d(d'_i, \hat{\mathbf{l}}'_{d_i})^2 \right] = 0, \quad (4.2)$$

onde $d(m_i, \hat{\mathbf{l}}_{m_i})$ é a distância do ponto m_i (na imagem I) a sua reta epipolar $\hat{\mathbf{l}}_{m_i} = \mathbf{F}^T \tilde{\mathbf{m}}'_i$

O ponto $\tilde{\mathbf{m}}'_i$ é dado pela equação

$$\tilde{\mathbf{m}}'_i = s_{\hat{m}'} \mathbf{P}' \tilde{\mathbf{M}}, \quad (4.3)$$

onde $\mathbf{P}' = \mathbf{P}'_w \mathbf{T}_w$ e $s_{\hat{m}'}$ é uma constante desconhecida e necessária por (4.3) ser homogênea. Assim, considerando $\tilde{\mathbf{m}}'_i = [\hat{u}', \hat{v}', 1]^T$ e $\mathbf{P}' = [\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{p}'_3]^T$, da equação (4.3), tem-se

$$\begin{aligned} \hat{u}' &= s_{\hat{m}'} \mathbf{p}'_1{}^T \tilde{\mathbf{M}}, \\ \hat{v}' &= s_{\hat{m}'} \mathbf{p}'_2{}^T \tilde{\mathbf{M}}, \\ 1 &= s_{\hat{m}'} \mathbf{p}'_3{}^T \tilde{\mathbf{M}}. \end{aligned}$$

Logo,

$$\hat{u}' = \frac{\mathbf{p}'_1{}^T \tilde{\mathbf{M}}}{\mathbf{p}'_3{}^T \tilde{\mathbf{M}}} \quad \text{e} \quad \hat{v}' = \frac{\mathbf{p}'_2{}^T \tilde{\mathbf{M}}}{\mathbf{p}'_3{}^T \tilde{\mathbf{M}}}.$$

Equações semelhantes podem ser derivadas para $d(m'_i, \hat{\mathbf{l}}'_{m_i})$, ou seja, a distância do ponto m'_i (na imagem I') a sua reta epipolar $\hat{\mathbf{l}}'_{m_i} = \mathbf{F} \tilde{\mathbf{m}}_i$.

Se for considerado as coordenadas esféricas, na equação (4.2), o ponto B pode ser define-se a razão cruzada entre eles como sendo

$$Cross(m_1, m_2, m_3, m_4) = \frac{|\tilde{m}_1 \tilde{m}_2| |\tilde{m}_3 \tilde{m}_4|}{|\tilde{m}_1 \tilde{m}_3| |\tilde{m}_2 \tilde{m}_4|},$$

onde

$$|\tilde{m}_i \tilde{m}_j| = \det \begin{bmatrix} \tilde{m}_{i1} & \tilde{m}_{j1} \\ \tilde{m}_{i2} & \tilde{m}_{j2} \end{bmatrix}.$$

dado por

$$B = A + L \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix}.$$

Por último,

$$C = \lambda_{AC}A + \lambda_{BC}B \quad \text{e} \quad D = \lambda_{AD}A + \lambda_{BD}B,$$

onde λ_{AC} , λ_{BC} , λ_{AD} e λ_{BD} são constantes conhecidas, mas que dependem da localização dos pontos C e D no gabarito.

Pode-se usar o algoritmo Levenberg-Marquardt [68] para resolver a equação (4.2). Infelizmente, esse algoritmo precisa de um valor inicial para \mathbf{T}_w , para as coordenadas do ponto A , θ e para ϕ . Esses valores devem ser conseguidos através de um método linear que deverá ser desenvolvido.

4.4 Estereoscopia aplicada a detecção de obstáculos

Detecção automática de obstáculos é um problema fundamental em robótica móvel [5]. Tal tarefa é essencial para proteger os robôs móveis de colisões e para mantê-los distantes de escadas e saliências no piso, enquanto trafegam por ambientes desconhecidos. Diversos sistemas de detecção de obstáculos são baseados em sensores, tais como sonares e lasers, que fornecem informações 3D diretamente [18, 78]. Em outros casos, marcas guias (*landmarks*) e sistemas monoculares são combinados com um conhecimento prévio do ambiente para localizar os veículos e guiá-los em seu trajeto [13, 14, 44]. Em oposição a projetos desse tipo, propõe-se um sistema que obtenha 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 [12].

O sistema deverá identificar os obstáculos e formar um mapa, no qual a localização desses obstáculos serão identificadas. Isso facilitará o planejamento do melhor caminho a ser percorrido pelo robô. Para desempenhar tais tarefas, o sistema deverá ser capaz de calibrar o par estereoscópico, segmentar os potenciais obstáculos e estimar a localização destes.

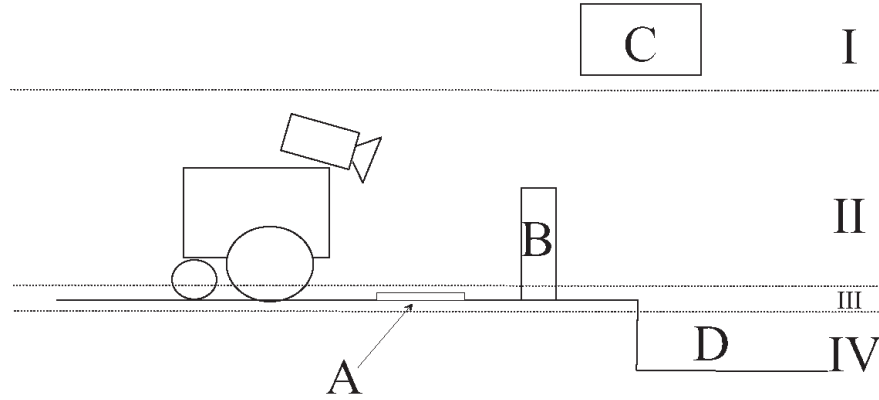


Figura 4.4: Classificação das regiões pelo veículo móvel.

4.4.1 Calibração do sistema de visão do robô

Como dito, o sistema será composto por um par estereoscópico. A calibração do sistema de câmeras será automática e realizada através de calibração proposta na seção 4.3. Além de estimar a matriz de projeção das câmeras, o estágio de calibração dará suporte a identificação das regiões, cujos objetos representem potenciais obstáculos.

Na robótica móvel, os objetos do ambiente do robô podem ser encontrados em quatro regiões [12]. Na figura 4.4, essas regiões são nomeadas como I, II, III e IV. Evidentemente, o veículo móvel deve ser capaz de identificar e desviar dos objetos pertencentes as regiões II e IV, por exemplo, o objeto B e a saliência D (figura 4.4). Por outro lado, objetos da região I (objeto C na figura 4.4) não são perigosos e o veículo pode perfeitamente ultrapassar objetos da região III.

Para classificar um objeto como estando dentro de uma das quatro regiões da figura 4.4, uma segunda calibração deve ser realizada. Esta calibração consiste em estimar a distância H que a câmera está do piso e o ângulo de inclinação Θ da câmera em relação ao eixo vertical (figura 4.5). Assim, conhecendo-se as coordenadas de um ponto P (figura 4.5) no piso do ambiente, pode-se determinar o ângulo β e a profundidade Z_p [11].

Neste trabalho, os pontos cujas projeções estão nas últimas linhas das imagens das duas câmeras serão considerados pontos no piso do ambiente. Esta é uma suposição aceitável, desde que o sistema visão será desenvolvido justamente para manter o robô distante de obstáculos. Assim, conhecidas as matrizes de projeção das duas câmeras, sempre será possível estimar-se, em relação ao sistema de coordenadas da câmera, as coordenadas de dois pontos $\mathbf{M}_1 = [X_1, Y_1, Z_1]^T$ e $\mathbf{M}_2 = [X_2, Y_2, Z_2]^T$ localizados no

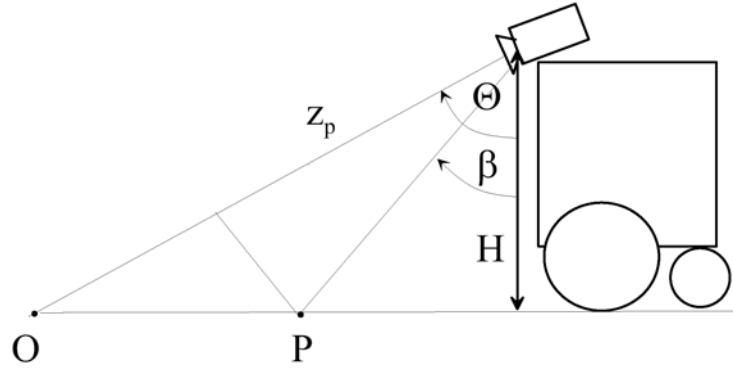


Figura 4.5: Parâmetros a serem estimados para classificação dos objetos no ambiente do robô móvel.

piso. Com isso, o ângulo Θ pode ser conhecido por

$$\Theta = \frac{\pi}{2} - \arccos \frac{|Z_1 - Z_2|}{d(M_1, M_2)}, \quad (4.4)$$

onde $d(M_1, M_2)$ é a distância euclidiana entre M_1 e M_2 .

Conhecendo-se Θ , H pode ser encontrado por [12]

$$H = M_i \begin{bmatrix} 0 \\ \cos(\frac{\pi}{2} - \Theta) \\ \sin(\frac{\pi}{2} - \Theta) \end{bmatrix}. \quad (4.5)$$

Evidentemente, para obter-se melhores resultados, métodos robustos baseados nas equações (4.4) e (4.5) deveram ser analisados.

Por último, com o sistema de visão do robô calibrado, como é discutido na próxima seção, espera-se ser possível classificar os objetos dentro das regiões da figura 4.4m, através do mapa de disparidades.

4.4.2 Segmentação dos obstáculos

Analisando-se a figura 4.5, pode-se demonstrar que a profundidade Z_p de um ponto O no sistema de coordenadas da câmera é dada por

$$Z_p = H \frac{\cos \gamma}{\cos \beta},$$

onde $\beta = \Theta + \gamma$ e γ é um ângulo dado em função da projeção de O no plano das imagens [12].

Desde que a profundidade Z_p é proporcional a disparidade, a identificação dos obstáculos será feita com o auxílio de um mapa de disparidades. Um mapa de disparidade consiste simplesmente de uma matriz que comporta a disparidade (seção 2.4.2) entre todos os pontos de duas imagens de um par estereoscópico [31]. Assim, os objetos serão áreas dentro das imagens com disparidades constantes.

Para calcular a disparidade entre dois pontos, inicialmente, emparelhamentos de pontos devem ser estabelecidos nas imagens das duas câmeras. Como o objetivo é apenas identificar a presença e a localização de objetos no campo de visão do robô, propõe-se o algoritmo a seguir que simplificará bastante esta tarefa.

1. retifica-se o par de imagens;
2. aplica-se a transformada rank nas duas imagens, e;
3. para cada ponto $m = (u, v)$ da imagem I , encontra-se o seu correspondente $m' = (u' > u, v)$ na imagem I' , calculando-se o SAD dentro do sistema de múltiplas janelas da figura 4.2.

Um dos maiores problemas da visão estereoscópica é a oclusão [24]. E, como discute-se na seção 3.8, a oclusão provoca a formação de falsos emparelhamentos nos pontos próximos as bordas dos objetos da imagem. Neste caso, considerando um ponto m da imagem I sem correspondente na imagem I' , o seu gráfico $SAD \times$ Disparidade é algo semelhante a figura 4.6. Assim, o algoritmo de emparelhamento tende a formar o emparelhamento baseado no par de pontos (m, m') , cuja correlação é $SAD(m, m') = C_{SAD2}$ (figura 4.6).

Para contornar este problema, neste trabalho, propõe-se que sejam retidos os dois candidatos a emparelhamento que apresentarem o menor SAD , ou seja, os pares cujas correlações são, respectivamente, C_{SAD1} e C_{SAD2} (figura 4.6). Assim, define-se a força da correlação como sendo

$$F_{SAD}(m, m') = \frac{C_{SAD1} - C_{SAD2}}{C_{SAD2}},$$

onde (m, m') é o par com $SAD(m, m') = C_{SAD2}$ e C_{SAD1} é a segunda menor correlação encontrada para o ponto m (figura 4.6).

Dessa forma, pode-se estabelecer que os candidatos a emparelhamento (m, m') que possuírem um baixo F_{SAD} são considerados possíveis falhas de emparelhamento. Então, a restrição esquerda-direita (seção 3.8) pode ser aplicada apenas a estes pares. Com

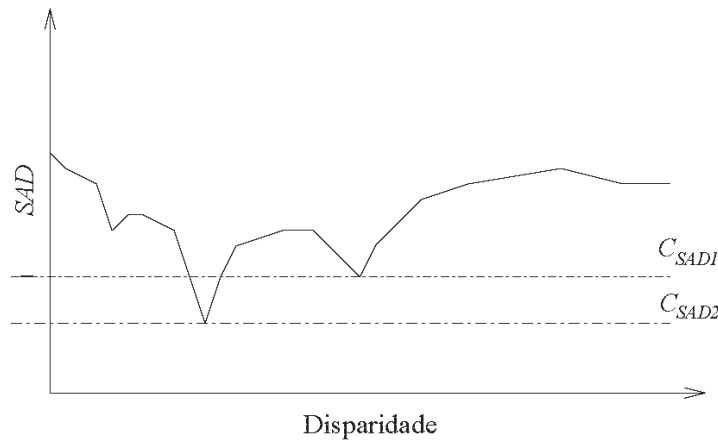


Figura 4.6: Um típico gráfico de $SAD \times$ disparidade na proximidade das bordas dos objetos de uma imagem.

isso, espera-se eliminar uma quantidade maior de falsos emparelhamentos, utilizando um algoritmo simples o suficiente para ser empregado em aplicações de tempo-real sem a necessidade de *hardware* específico.

Após a segmentação dos objetos, pode-se classificar em que região eles se encontram e determinar quais devem ser considerados obstáculos.

4.4.3 Geração do mapa de obstáculos

Com a segmentação de obstáculos da seção anterior, pode-se descrever um obstáculo pela localização do seu ponto médio e por sua extensão. Desse modo, é possível montar um mapa de obstáculos que poderá servir de base para determinar qual o melhor caminho que o veículo tem a seguir.

4.5 Metodologia

Desde que este trabalho restringe-se aos tópicos de calibração de câmeras, correspondência de pontos e mapa de disparidades, não serão utilizados robôs para realização de testes. Será montado um par estereoscópico, objetos serão inseridos em seu campo de visão e imagens serão capturadas. Todo o processamento das imagens será *offline*. Contudo, os tempos de processamento e a complexidade dos algoritmos serão estimados para avaliar a viabilidade destes em sistemas de tempo-real.

Para avaliar a performance do algoritmo de emparelhamento de pontos em imagens descalibradas proposto na seção 4.1, a matriz fundamental do par estereoscópico

será calculada através do uso de um gabarito composto por dois planos ortogonais. Em seguida, a distância média entre os pontos que compõem os emparelhamentos estabelecidos pelo algoritmo e as respectivas linhas epipolares serão calculadas. Esta distância média será usada como critério de performance. Além disso, desde que o algoritmo de emparelhamento de pontos em imagens descalibradas testado que obteve os melhores resultados foi o proposto por ZHANG *et al* [96], este será comparado com o algoritmo proposto.

Para avaliar a técnica de calibração proposta na seção 4.3, inicialmente, pretende-se conhecer o quanto a exatidão do modelo da câmera encontrado pelo método tradicional de TSAI [86] depende da distância do ponto 3D ao gabarito utilizado na calibração. Para isto, a matriz fundamental será obtida à partir das matrizes de projeção calculadas pelo método de TSAI [86]. Em seguida, o gabarito será transladado a diferentes distâncias da câmera e, para cada nova posição, o emparelhamento de pontos do gabarito será realizado. Assim, medindo-se a distância desses pontos a sua linha epipolar, ter-se-á uma medida do ajuste do modelo da câmera aos pontos 3D de regiões à diferentes distâncias da câmera. Repetindo esse procedimento para a técnica de calibração proposta na seção 4.3, a sua performance pode ser avaliada.

Com o algoritmo de emparelhamento de pontos baseado em múltiplas janelas proposto na seção 4.4.2, pretende-se calcular o mapa de disparidades das imagens “*Head and Lamb*” [60] da *University of Tsukuba* (figura 4.7). As imagens “*Head and Lamb*” formam um conjunto de 25 imagens capturadas por uma matriz de câmeras de dimensão 5×5 . Por terem seus mapas de disparidades feitos de forma iterativa, ou seja, sem uso de algoritmos automáticos, essas imagens são muito utilizadas para testar a performance de algoritmos de emparelhamento de pontos. Assim, pretende-se comparar o mapa de disparidades gerado pelo algoritmo com o mapa ideal da figura 4.7(b). Dessa forma, ter-se-á informações sobre a performance do algoritmo, sobre tudo, nas bordas dos objetos, onde o problema da oclusão ocorre.

4.6 Cronograma

Os trabalhos serão desenvolvidos em 18 meses e será dividido nas seguintes etapas:

- A. Implementação e teste do algoritmo de emparelhamento de pontos em imagens descalibradas;
- B. Implementação e teste do algoritmo robusto proposto para cálculo da matriz

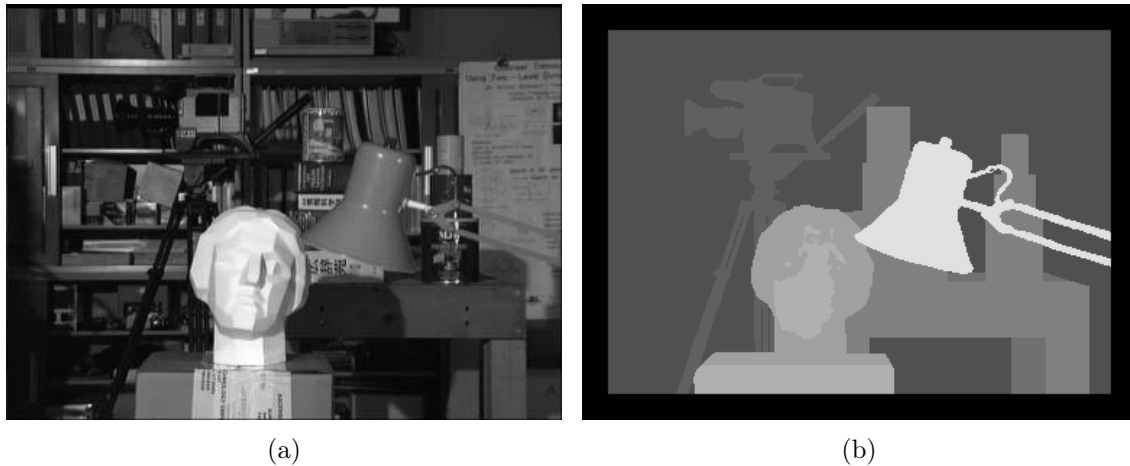


Figura 4.7: (a) Uma das famosas imagens “*Head and Lamp*” da *University of Tsukuba* e (b) o seu mapa ideal de disparidades.

fundamental;

- C. Inclusão da restrição epipolar no algoritmo de emparelhamento de pontos implementado na fase A;
- D. Equacionamento de um método linear para inicialização do algoritmo de calibração de câmeras proposto;
- E. Implementação e teste do algoritmo não-linear proposto para calibração de câmeras;
- F. Avaliação, através da metodologia da seção anterior, do modelo da câmera estimando pelo método proposto e pelo método de TSAI [86];
- G. Implementação e teste do algoritmo de emparelhamento de pontos em imagens calibradas proposto;
- H. Criação do mapa de disparidade para localização dos obstáculos da cena;
- I. Divulgação dos resultados, e;
- J. Elaboração do documento da tese.

Por sua vez, as fases serão desenvolvidas considerando o seguinte cronograma.

	Mês																	
Fase	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	X	X	X															
B		X	X	X														
C				X	X													
D					X	X												
E						X	X	X	X									
F									X	X	X							
G										X	X	X	X					
H												X	X	X	X			
I													X	X	X	X	X	X
J			X	X	X					X	X		X	X			X	X

Tabela 4.1: Cronograma para desenvolvimento das fases do trabalho.

Bibliografia

- [1] J. Banks and M. Bennamoun. Reliability analysis of the rank transform for stereo matching. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 31(6):870–880, 2001.
- [2] J. Banks, M. Bennamoun, and P. Corke. Non-parametric techniques for fast and robust stereo matching. In *IEEE Region 10 Annual Conference*, pages 365–368, 1997.
- [3] J. Banks, M. Bennamoun, and P. Corke. A hybrid stereo matching algorithm incorporating the rank constraint. In *7th International Conference on Image Processing And Its Applications*, pages 33–37, 1999.
- [4] A. Baumberg. Reliable feature matching across widely separated views. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, pages 774–781, 2000. URL <http://citeseer.nj.nec.com/baumberg00reliable.html>.
- [5] P.A. Beardsley, A. Zisserman, and D.W. Murray. Navigation using affine structure from motion. In *European Conference on Computer Vision*, pages 85–96, 1994.
- [6] F. Benevenuti. Calibração de um conjunto binocular para uso em estereoscopia e robótica móvel. Master’s thesis, UFSC, Brasil, 2002.
- [7] M. Bertozzi and A. Broggi. GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 1997. URL <http://citeseer.nj.nec.com/bertozzi98gold.html>.
- [8] G. Borges. Um sistema óptico de reconhecimento de trajetórias para veículos automáticos. Master’s thesis, Universidade Federal da Paraíba, Brasil, 1998. URL <http://www.lirmm.fr/~borges/Publications/Publications.html>.

- [9] B. S. Boufama and R. Mohr. A stable and accurate algorithm for computing epipolar. *Intern. Journal of Pattern Recognition and Artificial Intelligence*, 12 (6):817–840, 1998. URL <http://citeseer.nj.nec.com/524287.html>.
- [10] M.Z. Brown, D. Burschka, and G.D. Hager. Advances in computational stereo. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, to appear. URL <http://jobber.cs.jhu.edu/~burschka/pub-burschka.html>.
- [11] D. Burschka and G. Hager. Dynamic composition of tracking primitives for interactive vision-guided navigation. In *SPIE*, pages 114–125, 2001. URL <http://jobber.cs.jhu.edu/~burschka/pub-burschka.html>.
- [12] D. Burschka, S. Lee, and G. Hager. Stereo-based obstacle avoidance in indoor environments with active sensor re-calibration. In *IEEE International Conference on Robotics and Automation*, pages 2066–2072, 2002. URL <http://jobber.cs.jhu.edu/~burschka/pub-burschka.html>.
- [13] M. F. M. Campos and L. Chaimowicz. Visual measurement estimation for autonomous vehicle navigation. In *XIII Annual International Symposium on Aerospace/Defense Sensing Simulation and Controls*, Orlando, Florida, USA, 1999. URL <http://www.dcc.ufmg.br/~chaimo/public.html>.
- [14] L. Chaimowicz and M. F. M. Campos. Estimaco visual de distncias na navegaco de robs mveis. *IV SBAl*, 1999. URL <http://www.dcc.ufmg.br/~chaimo/public.html>.
- [15] G. Csurka, C. Zeller, Z. Zhang, and O. Faugeras. Characterizing the uncertainty of the fundamental matrix. Technical Report RR-2560, INRIA, 1995. URL <http://citeseer.nj.nec.com/csurka95characterizing.html>.
- [16] J. A. D. de Carvalho and J. S. S. Souza. Sistema inteligente de navegaco de robs mveis autnuomos. In *XII Brazilian Automatic Control Conference*, pages 575–600, 1998.
- [17] J.A. de Frana and M.R. Stemmer. Estimaco robusta da matriz fundamental. In *VI Simpsio Brasileiro de Automao Inteligente*, 2003. URL <http://research.eeol.org/papers/sbai/2003/>.

- [18] G. Dudek, P. Freedman, and I. Rekleitis. Just-in-time sensing: efficiently combining sonar and laser range data for exploring unknown worlds. In *IEEE Conf. on Robotics and Automation*, pages 667–672, 1996. URL <http://citeseer.nj.nec.com/74135.html>.
- [19] O. Faugeras and Q.-T. Luong. *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.
- [20] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *II European Conference on Computer Vision*, pages 563–578, 1992. URL <http://research.eeol.org/visao/>.
- [21] Olivier D. Faugeras, Quang-Tuan Luong, and Stephen J. Maybank. Camera self-calibration: Theory and experiments. In *European Conference on Computer Vision*, pages 321–334, 1992. URL <http://citeseer.nj.nec.com/faugeras92camera.html>.
- [22] R. Freitas. Sistema de visão para robôs móveis: Uma aplicação ao reconhecimento de referências geométricas. Master’s thesis, UFES, Brasil, 1999. URL <http://www.ele.ufes.br/pesq/robotica.html>.
- [23] A. Fusiello. Tutorial on rectification of stereo images. In *On-Line Compendium of Computer Vision [Online]*. R. Fisher (ed). Available: “<http://www.dai.ed.ac.uk/CVonline/>”, 1998. URL http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/FUSIELLO/tutorial.html.
- [24] A. Fusiello, V. Roberto, and E. Trucco. Efficient stereo with multiple windowing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 858–863, 1997. URL <http://citeseer.nj.nec.com/fusiello97efficient.html>.
- [25] A. Fusiello, E. Trucco, and A. Verri. Rectification with unconstrained stereo geometry. In *ICCV*, 1998. URL <http://citeseer.nj.nec.com/article/fusiello98rectification.html>.
- [26] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. URL <http://citeseer.nj.nec.com/fusiello00compact.html>.
- [27] C. Gauclin and T. Papadopoulos. Fundamental matrix estimation driven by stereo correlation. In *ACCV*, 2000. URL <http://citeseer.nj.nec.com/383044.html>.

-
- [28] N.M. Grzywacz and A.L. Yuille. Massively parallel implementations of theories for apparent motion. *Spatial Vision*, 3(1):15–44, 1988. URL <http://vp1.usc.edu/papers/15-44.pdf>.
- [29] R. Hartley. In defence of the 8-point algorithm. In *5th International Conference on Computer Vision*, pages 1064–1070. IEEE Computer Society, 1995. URL <http://www.syseng.anu.edu.au/%7Ehartley/My-Papers.html>.
- [30] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *II European Conference on Computer Vision*, pages 579–587, 1992. URL <http://www.syseng.anu.edu.au/%7Ehartley/My-Papers.html>.
- [31] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.
- [32] Richard I. Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding: CVIU*, 68(2):146–157, 1997. URL citeseer.nj.nec.com/hartley94triangulation.html.
- [33] H. Hirschmüller. Improvements in real-time correlation-based stereo vision. In *Proceedings of IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 141–148, 2001. URL <http://www.cse.dmu.ac.uk/~hbm/publications.html>.
- [34] J. J. Hopfield. Neural networks and systems with emergent selective computational abilities. In *Proc. Natl. Acad. Sci.*, number 79, pages 2554–2558, 1982. URL <http://research.eeol.org/visao/>.
- [35] J. J. Hopfield and D. W. Tank. “Neural” computation of decisions in optimization problems. *Biological Cybernetics*, 52:141–152, 1985. URL <http://research.eeol.org/visao/>.
- [36] D. Huynh. *Feature-based Stereo Vision on a Mobile Platform*. PhD thesis, University of Western Australia, 1994. URL <http://citeseer.nj.nec.com/huynh94featurebased.html>.
- [37] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.

- [38] H. Kitano, M. Tambe, P. Stone, M. Veloso, S. Coradeschi, E. Osawa, H. Matsubara, I. Noda, and M. Asada. The RoboCup synthetic agent challenge, 97. In *International Joint Conference on Artificial Intelligence*, 1997. URL <http://citeseer.nj.nec.com/kitano97robocup.html>.
- [39] K. Konolige. Small vision system: Hardware and implementation. In *8th International Symposium on Robotics Research*, 1997. URL <http://citeseer.nj.nec.com/konolige97small.html>.
- [40] E. Kruppa. Zur ermittlung eines objektes aus zwei perpesktiven mitinnerer orientierung. *Abhandlungen der Akademie der Wissenschaften, Wien*, 122:1939–1948, 1913.
- [41] B. Lamiroy, B. Espiau, N. Andreff, and R. Horaud. Controlling robots with two cameras: How to do it properly. In *2000 IEEE International Conference on Robotics e Automation*, pages 2100–2105, EUA, 2000.
- [42] R.A. Lane and N.A. Thacker. Stereo vision research: An algorithm survey, 1996. URL <http://citeseer.nj.nec.com/lane96stereo.html>.
- [43] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Conference on Computer Vision and Pattern Recognition*, pages 125–131, 1999. URL <http://citeseer.nj.nec.com/loop99computing.html>.
- [44] L. Lorigo, R. Brooks, and W. Grimson. Visually-guided obstacle avoidance in unstructured environments, 1997. URL <http://citeseer.nj.nec.com/lorigo97visuallyguided.html>.
- [45] E. Louprias, N. Sebe, S. Bres, and J.-M. Jolion. Wavelet-based salient points for image retrieval. In *International Conference on Image Processing*, 2000. URL <http://www.liacs.nl/home/nicu/publications/ICIPLouprias.pdf>.
- [46] Q. T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulos. On determining the fundamental matrix : analysis of different methods and experimental results. Technical Report RR-1894, INRIA, 1993. URL <http://citeseer.nj.nec.com/article/luong93determining.html>.
- [47] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *IJCV*, 22(3):261–289, 1997. URL <http://citeseer.nj.nec.com/luong97selfcalibration.html>.

- [48] Q.-T. Luong and O.D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *The International Journal of Computer Vision*, 17(1): 43–76, 1996. URL <http://citeseer.nj.nec.com/luong95fundamental.html>.
- [49] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. In *ECCV (1)*, pages 589–599, 1994. URL <http://citeseer.nj.nec.com/luong94canonic.html>.
- [50] Y. Ma, J. Kosecka, and S. S. Sastry. A differential geometric approach to camera self-calibration. In *CVPR*, page pp, 1999. URL <http://citeseer.nj.nec.com/123839.html>.
- [51] Y. Ma, R. Vidal, J. Koseck’a, and S. Sastry. Camera self-calibration: geometry and algorithms. Technical Report UCB/ERL M99/32, UC Berkeley, 1999. URL <http://citeseer.nj.nec.com/ma99camera.html>.
- [52] Yi Ma, Rene Vidal, Jana Kosecka, and Shankar Sastry. Kruppa equation revisited: Its renormalization and degeneracy. In *ECCV (2)*, pages 561–577, 2000. URL <http://citeseer.nj.nec.com/ma00kruppa.html>.
- [53] J. Maciel. *Global Matching: optimal solution to correspondence problems*. PhD thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal, 2001. URL <http://citeseer.nj.nec.com/563272.html>.
- [54] J. Maciel and J. Costeira. Robust point correspondence by concave minimization. In *11th BMVC*, pages 626–635, 2000. URL <http://citeseer.nj.nec.com/article/maciel00robust.html>.
- [55] S. Maybank and O.D. Faugeras. A theory of self-calibration of a moving camera. *IJCV*, 8(2):123–151, August 1992. URL <http://research.eeol.org/visao/classicos/maybank/>.
- [56] R. Mayoral and M.J. Perez-Ilzarbe. Competitive hopfield neural network for stereo vision correspondence. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on IJCNN 2000*, volume 5, pages 464–469, 2000.
- [57] Daniel P. McReynolds and David G. Lowe. Rigidity checking of 3d point correspondences under perspective projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1174–1185, 1996. URL <http://citeseer.nj.nec.com/mcreynolds96rigidity.html>.

- [58] H.P. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Stanford University, Setembro 1980.
- [59] Moshier. Cephes mathematical library. In "<http://sal.kachinatech.com/B/0/CEPHES.html>", 1989. URL <http://www.netlib.org/cephes/>.
- [60] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta. Occlusion detectable stereo-occlusion patterns in camera matrix. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 371–378, 1996.
- [61] N. M. Nasrabadi and C. Y. Choo. Hopfield network for stereo vision correspondence. *IEEE Transactions on Neural Network*, 3(1):5–13, 1992.
- [62] N.M. Nasrabadi, W. Li, B.G. Epranian, and C.A. Butkus. Use of hopfield network for stereo vision correspondence. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 429–432, 1989.
- [63] S. L. Olsen. Epipolar line estimation. In *European Conference on Computer Vision*, volume 588, pages 307–311. G. Sandini, 1992.
- [64] G. Ottoni. Planejamento de trajetórias para robôs móveis em ambientes desconhecidos. In *XIII CBA*, pages 2239–2244, 2000. URL <http://www.delet.ufrgs.br/~fetter/papers.html>.
- [65] M. Pilu. A direct method for stereo correspondence based on singular value decomposition. In *Computer Vision & Pattern Recognition*, pages 261–266, 1997. URL <http://citeseer.nj.nec.com/pilu97direct.html>.
- [66] J. Pimentel and E. Hemerly. An algorithm to detect lanemarks for mobile robots navigation. In *INDUSCON 2000*, pages 114–119, 2000.
- [67] S. Pollard, J. Mayhew, and J. Frisby. PFM: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.
- [68] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.

- [69] C. Rothwell, O. Faugeras, and G. Csurka. A comparison of projective reconstruction methods for pairs of views. *Computer Vision and Image Understanding*, 68(1):37–58, 1997. URL <http://citeseer.nj.nec.com/article/rothwell195comparison.html>.
- [70] P. J. Rousseeuw. *Robust Regression and Outlier Detection*. Wiley, 1987.
- [71] F. A. Sandi L., E. M. Hemerly, and W. F. Lages. Sistema para navegação e guiagem de robôs móveis autônomos. *SBS Controle e Automação*, 9:107–118, 1998. URL http://www.fee.unicamp.br/revista_sba/vol19/.
- [72] R. Sára. Finding the largest unambiguous component of stereo matching. In *ECCV (3)*, pages 900–914, 2002. URL <http://citeseer.nj.nec.com/sara02finding.html>.
- [73] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Technical Report MSR-TR-2001-81, Microsoft Research, 2001. URL <http://citeseer.nj.nec.com/scharstein01taxonomy.html>.
- [74] G. L. Scott and H. C. Longuet-Higgins. An algorithm for associating the features of two images. *Royal Society of London*, B(244):21–26, 1991. URL <http://research.eeol.org/visao/>.
- [75] L.S. Shapiro and J.M. Brady. Featurebased correspondence: an eigenvector approach. *Image and Vision Computing*, 10(5):283–288, 1992.
- [76] L.S. Shapiro and J.M. Brady. Rejecting outliers and estimating errors in an orthogonal regression framework. *Royal*, A-350:403–439, 1994.
- [77] S. Shih, Y. Hung, and W. Lin. Accuracy analysis on the estimation of camera parameters for active vision systems. Technical Report TR-IIS-96-006, Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan, 1996. URL <http://citeseer.nj.nec.com/shih96accuracy.html>.
- [78] S. Shoval, J. Borenstein, and Y. Koren. Mobile robot obstacle avoidance in a computerized travel aid for the blind. In *IEEE International Conference on Robotics and Automation*, pages 2023–2029, 1994. URL <http://citeseer.nj.nec.com/473279.html>.

-
- [79] P. Smith, D. Sinclair, R. Cipolla, and K. Wood. Effective corner matching. In *British Machine Vision Conference*, 1998. URL <http://citeseer.nj.nec.com/304109.html>.
- [80] S.M. Smith. A new class of corner finder. In *3rd British Machine Vision Conference*, pages 139–148, 1992. URL <http://www.fmrib.ox.ac.uk/~steve/susan/>.
- [81] P.F. Sturm and S.J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.
- [82] C. Tomasi. Mathematical methods for robotics and vision. Technical Report CS 205, Stanford University, 2000. URL <http://research.eeol.org/visao/>.
- [83] C. Tomasi and R. Manduchi. Stereo without search. In *ECCV (1)*, pages 452–465, 1996. URL <http://citeseer.nj.nec.com/tomasi96stereo.html>.
- [84] P. Torr. *Outlier Detection and Motion Segmentation*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1995. URL <http://research.microsoft.com/%7Ephiltorr/thesis.html>.
- [85] P. Torr and D. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997. URL <http://citeseer.nj.nec.com/torr97development.html>.
- [86] R. Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *IEEE Computer vision and pattern recognition*, pages 364–374, 1986. URL <http://research.eeol.org/visao/>.
- [87] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press, 1979.
- [88] S. Ullman. Maximizing rigidity: the incremental recovery of 3D structure from rigid and rubbery motion. *Perception*, 13:255–274, 1984.
- [89] J. Weng, N. Ahuja, and T.S. Huang. Matching two perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):806–825, 1992.

-
- [90] J. Woodfill and B. Von Herzen. Real-time stereo vision on the PARTS reconfigurable computer. In *IEEE Symposium on FPGAs for Custom Computing Machines*, pages 201–210, 1997. URL <http://citeseer.nj.nec.com/woodfill97realtime.html>.
- [91] Z. Zhang. Le probleme de la mise en correspondance: l’etat de l’art. Technical Report RR-2146, INRIA, 1993. URL <http://www.inria.fr/rrrt/rr-2146.html>.
- [92] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. Technical Report RR-2676, INRIA, 1997. URL <http://citeseer.nj.nec.com/zhang97parameter.html>.
- [93] Z. Zhang. Determining the epipolar geometry and its uncertainty: a review. *The International Journal of Computer Vision*, 27(2):161–195, August 1998. URL <http://citeseer.nj.nec.com/zhang98determining.html>.
- [94] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *7th IEEE International Conference on Computer Vision*, 1999. URL <http://research.microsoft.com/~zhang/>.
- [95] Z. Zhang. Camera calibration with one-dimensional objects. Technical Report MSR-TR-2001-120, Microsoft Research, 2002. URL <http://research.microsoft.com/~zhang/Papers/TR01-120.pdf>.
- [96] Z. Zhang, R. Deriche, O. D. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1995. URL <http://citeseer.nj.nec.com/article/zhang94robust.html>.
- [97] Z. Zhang, O. Faugeras, and R. Deriche. An effective technique for calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. *Videre: Journal of Computer Vision Research*, 1(1):58–68, 1997. URL <http://citeseer.nj.nec.com/300330.html>.
- [98] Z. Zhang and C. Loop. Estimating the fundamental matrix by transforming image points in projective space. *Computer Vision and Image Understanding*, 82(2):174–180, 2001. URL <http://citeseer.nj.nec.com/zhang01estimating.html>.

-
- [99] Z. Zhang, Q.-T. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. *IEEE Trans. Robotics and Automation*, 12(1):103–113, 1996. URL <http://citeseer.nj.nec.com/article/zhang93motion.html>.
- [100] Z. Zhang and G. Xu. A general expression of the fundamental matrix for both perspective and affine cameras. In *IJCAI*, pages 1502–1510, 1997. URL <http://citeseer.nj.nec.com/zhang97general.html>.
- [101] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. URL <http://citeseer.nj.nec.com/zhang98flexible.html>.