FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

N689g	Nogueira, Fernando Marques de Almeida Geração automática de mapas de disparidade em visão estéreo. / Fernando Marques de Almeida Nogueira.--Campinas, SP: [s.n.], 1998. Orientador: Clésio Luis Tozzi Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação. 1. Visão binocular. 2. Visão de robô. 3. Visão por computador. 4. Teoria dos casamentos. I. Tozzi, Clésio Luis. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação

Departamento de Engenharia e Automação Industrial

Geração automática de mapas de disparidade em visão estéreo

Autor: **Fernando Marques de Almeida Nogueira**
Orientador: **Prof. Dr. Clésio Luis Tozzi**

Dissertação submetida à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, para preenchimento dos pré-requisito parciais para a obtenção do Título de Mestre em Engenharia Elétrica.

Abril 1998

Resumo

A informação tridimensional de uma cena pode ser obtida usando a técnica denominada "visão estéreo". A maior dificuldade desta técnica é a determinação automática da correspondência entre as imagens (pontos homólogos). A solução do problema de correspondência pode ser obtida a partir do matching baseado em áreas e/ou do matching baseado em feições e representada na forma de um mapa de disparidade. A maioria dos algoritmos contemporâneos encontrados na literatura, utilizam-se do matching baseado em feições para a solução do problema de correspondência. Esta classe de matching permite gerar mapas de disparidade esparsos, que são então interpolados para gerar mapas de disparidade densos. Mapas de disparidade gerados a partir de processos de interpolação são suaves e não representam corretamente as variações bruscas de profundidade da cena. Neste trabalho propõe-se uma metodologia na qual é gerado inicialmente um mapa de disparidade denso e suave, através da interpolação dos valores de disparidade obtidos a partir do matching baseado em feições. Em uma fase posterior do processo, os valores deste mapa são então corrigidos de forma iterativa através do matching baseado em áreas, obtendo-se um mapa de disparidade que suporta superfícies não suaves. Os conceitos principais utilizados para o desenvolvimento desta metodologia são baseados nas evidências biológicas de movimento de vergência dos olhos e gradiente de disparidade presentes nos sistemas de visão de seres vivos. Testes com imagens sintéticas e reais indicam que os mapas de disparidade gerados são coerentes com a estrutura tridimensional da cena.

Abstract

Tridimensional information of a scene can be obtained using the technique named "shape from stereo". The major difficulty of this technique is the automatic determination of correspondence between images (matching points). The stereo matching solution can be obtained from area-based matching and/or feature-based matching and represented as a disparity map. Most of the contemporaneous algorithms found in the literature employ feature-based matching techniques for the stereo matching. This class of matching produces a sparse disparity map, which is further interpolated resulting in a smooth dense disparity map that does not represent correctly the scene abrupt variations of depth. An approach for stereo matching that results in dense and non-smooth disparity map is presented in this work. In a first step, a smooth and dense disparity map is obtained through interpolation of the disparity values obtained from feature-based matching. In a second step, this map is iteratively adjusted, using area-based matching techniques, and a dense disparity map that supports non-smooth surfaces is obtained. The main ideas used in this approach are based on biologic evidence of eyes vergence movement and disparity gradient that are present in the living beings vision systems. Experiments using synthetic and real images were made and the results show the coherence between the 3-D geometry of the scene and the disparity maps generated.

Agradecimentos

-Ao Prof. Dr. Clésio Luis Tozzi pela orientação, amizade e paciência dedicada a mim durante a realização deste trabalho;

-A Xuxu pelo incentivo, dedicação e paciência.

-Ao meu amigo Maurício Galo pelas proveitosas discussões e sugestões, incentivo e amizade.

-Aos meus amigos Maurício Figueiredo, Ronaldo, José Alfredo e Marquinhos (Bacalhau) pelas sugestões propostas, discussões e apoio.

-Aos demais professores, e alunos do curso de Pós-Graduação da FEEC, pelos ensinamentos, apoio e agradável convívio que me proporcionaram nestes anos;

-Aos funcionários da FEEC pelo apoio.

-Ao Centro Tecnológico para Informática - CTI pela utilização das câmeras digitais que permitiram a captura do par de imagens reais;

-Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pela concessão de bolsa.

Dedico este trabalho aos meus pais Marco Aurélio e Lúcia, à minha irmã Débora, à minha namorada Xuxu e aos meus amigos (galera) que sempre me apoiaram com amor, carinho e compreensão em todos os momentos de minha vida e ao Clube de Regatas Flamengo, meu time de coração.

Sumário

Lista de tabelas

Lista de figuras

Capítulo 1

1. Introdução

A visão é o sentido que um grande número de espécies animais dispõem e que capacita-os a adquirir, processar, extrair e representar informações do mundo que os cerca a partir de imagens.

O desenvolvimento computacional, das câmeras de vídeo analógicas e mais recentemente das câmeras de vídeo digitais tornaram possível desenvolver sistemas de visão artificial que realizam certas tarefas similares as dos sistemas de visão biológicos. Tais sistemas estão ainda muito distantes de alcançarem a robustez e capacidade de generalização dos sistemas biológicos, haja visto a complexidade desse sentido, mas em algumas aplicações já é possível realizar as tarefas de aquisição, processamento, extração e representação da informação automaticamente com relativo sucesso.

Em certas aplicações não é necessária a informação tridimensional dos objetos que estão sendo analisados, como por exemplo, a inspeção de placas de circuitos impresso, contagem de células em tecidos humanos e várias outras aplicações que envolvem processos de análise qualitativo e quantitativo. Porém quando a informação tridimensional da cena em questão é necessária para a obtenção de outras informações de mais alto nível, a complexidade dos sistemas aumenta significantemente. Como exemplos de aplicação de sistemas de visão artificiais que necessitam da informação tridimensional da cena pode-se citar a navegação autônoma, geração de mapas topográficos, inspeção tridimensional de produtos industrializados e robótica em geral.

A reconstrução tridimensional de uma cena é uma tarefa extremamente rápida, precisa e eficiente em sistemas de visão biológicos e infelizmente ainda obscura do ponto de vista do conhecimento de seu funcionamento. Várias teorias buscam explicar esse funcionamento, mas poucas soluções já foram concretizadas e consagradas pela comunidade científica.

1.1 Visão humana

A visão humana é o mais nobre e complexo sentido que os seres humanos possuem. Ela é uma fonte de grande parte das informações que os seres humanos necessitam diariamente para realizar tarefas tais como, localização e deslocamento no ambiente que os cercam, reconhecer pessoas e objetos, interpretar textos, etc. A visão é devida a diversos processos biológicos que ocorrem desde a aquisição da luz pelos olhos até a representação abstrata da informação no cérebro.

Segundo GUYTON (1989), a determinação da profundidade pelo sistema visual humano deve ocorrer devido a três principais meios que são:

Percepção da profundidade através das dimensões das imagens retinianas de objetos conhecidos.

A distância focal (f) do olho está para a profundidade (p) do objeto, assim como, metade da dimensão da imagem retiniana do objeto (I) está para metade da dimensão do objeto (O). Este meio é monocular. A figura 1.1 ilustra esta situação.

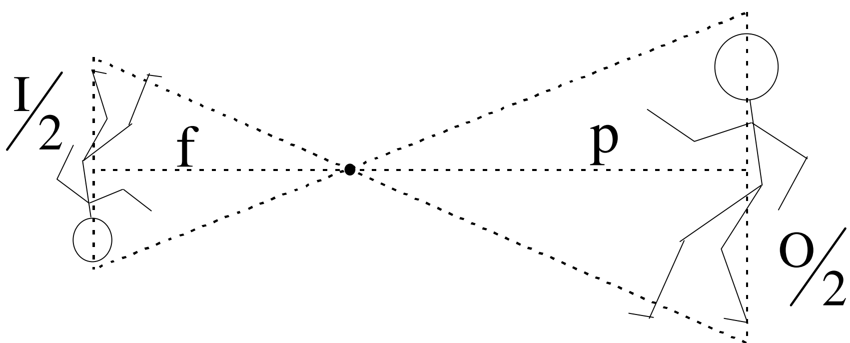


Fig. 1.1 - Percepção da profundidade através das dimensões das imagens retinianas de objetos conhecidos.

Percepção da profundidade através da disparidade de movimento.

A disparidade de movimento é percebida quando os olhos estão em movimento e a cena está fixa ou quando a cena, ou parte dela, está em movimento e os olhos estão fixos ou ainda mesmo quando ambos estão em movimento. Um exemplo típico desse tipo de disparidade é quando uma pessoa viaja em um trem e observa um poste próximo à linha férrea e uma montanha distante desta mesma linha férrea. Nestas condições, o poste passará rapidamente pelo campo visual do observador enquanto a montanha, ao fundo, passará mais lentamente. Este meio de percepção de profundidade envolve o conceito de velocidade em imagens e, a exemplo do meio anterior, este também é monocular.

Percepção da profundidade por visão estéreo.

Este tipo de percepção de profundidade utiliza-se dos dois olhos para obter a informação de profundidade da cena.

Apesar do conhecimento das áreas do sistema visual humano onde são realizados alguns processamentos da informação visual, a fusão das imagens provenientes do olho esquerdo e do olho direito, ainda são desconhecidas. Há indícios que a fusão das imagens ocorra em camadas adjacentes ao par de corpos geniculados laterais, situados na via visual e também no próprio córtex visual. O indício que permite suspeitar que a fusão das imagens ocorra nos corpos geniculados laterais, é que os sinais provenientes das retinas esquerda e direita se encontram em cada corpo geniculado lateral, ou seja, um mesmo corpo geniculado lateral recebe sinais tanto da retina da esquerda quanto da direita. O indício que a fusão das imagens ocorra no córtex visual é devido a células específicas do córtex visual tornarem-se excitadas em áreas do campo visual onde os pontos de maior luminosidade estão fora de registro. Este assunto é muito abragente e as referências MARR, 1980; GUYTON, 1989; ANGELIS *et al.*, 1991; 1991; COSTA, 1993; HAYKIN, 1994; ANDERSON *et al.*, 1995; ARNDT *et al.*, 1995; FRISBY, 1995; TITTLE & TODD, 1995; e KREMER, 1997 são recomendadas para maiores informações.

Na figura 1.2 nota-se três cubos em cada imagem de mesmo tamanhos, representados em tons de cinza, porém em profundidades diferentes.

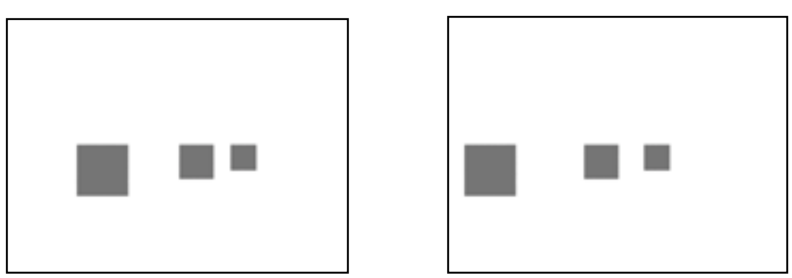


Fig. 1.2 - Imagens estereoscópicas

Estas imagens, a primeira vista, não contém nenhuma informação tridimensional, mas com algum tempo observando-as e com certa prática, a estrutura tridimensional é revelada. Neste momento, ocorre a visualização de apenas uma imagem com a sensação de tridimensionalidade. A capacidade do sistema visual humano "fundir" o par de imagens em apenas uma única imagem é devida a capacidade desse sistema em resolver com sucesso, o problema de correspondência (*matching*) entre os pontos homólogos.

A figura 1.3 foi extraída de MARR, (1980) e representa um estereograma de pontos randômicos, que a exemplo da figura 1.1, também podem ser "fundidos", apesar de não possuírem nenhuma informação a respeito de formas, objetos ou outras feições de mais alto nível quando vistos monocularmente. Isto confirma a habilidade do sistema visual humano em obter informações tridimensionais de qualquer tipo de cena.

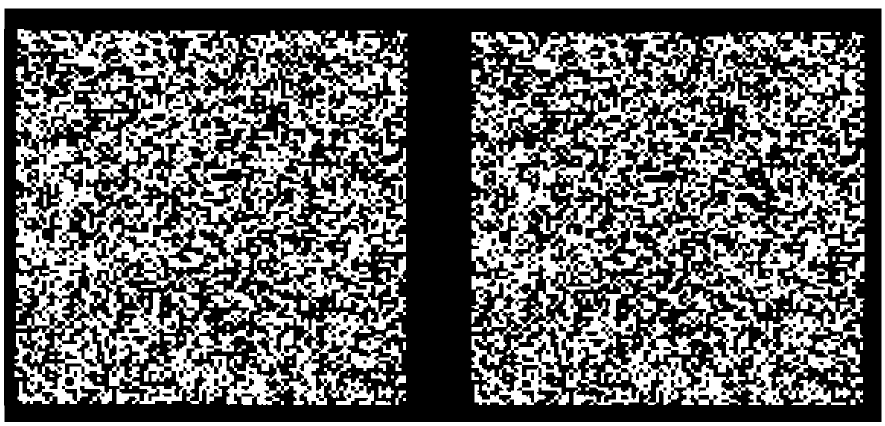


Fig. 1.3- Estereograma de pontos randômicos.

1.2 Visão artificial

Existem várias denominações similares para o termo visão artificial. Entre as mais comuns e utilizadas estão os termos: visão computacional, visão robótica, visão de máquina, visão cibernética e a própria denominação visão artificial.

Existem várias técnicas de reconstrução tridimensional empregadas em visão artificial; essas técnicas são denominadas de *Shape from X*, onde X são as técnicas utilizadas, podendo ser ativas ou passivas. As principais técnicas são:

Shape from shading

O *shape from shading* é uma técnica ativa de reconstrução tridimensional de uma cena baseada em fotometria, ou seja, na informação radiométrica contida em uma imagem. Esta técnica consiste em determinar a relação existente entre o vetor normal de um ponto de coordenadas X, Y e Z, sobre uma superfície no espaço-objeto e a irradiância de sua imagem, de coordenadas x e y, no espaço-imagem. Esta relação é dada por um modelo de iluminação.

Os modelos de iluminação mais utilizados são os modelos de Lambert e de Phong, que postulam basicamente (o modelo de Phong é mais complexo) que a irradiância é proporcional ao cosseno do ângulo formado entre o vetor normal à superfície e o vetor direção da fonte de iluminação, multiplicado pela intensidade da fonte de iluminação e o coeficiente de reflexão difuso (albedo). A equação 1.1 é o modelo de iluminação proposto por Lambert.

$$E = \rho.I.\cos(\theta) \tag{1.1}$$

onde:

E é a irradiância;

ρ é o coeficiente de reflexão difuso; e

θ é o ângulo formado entre o vetor normal a superfície e o vetor direção da fonte de iluminação.

A principal vantagem dessa técnica é que ela é monocular, evitando assim, o problema de correspondência presente em algumas outras técnicas. Porém, o modelo de iluminação de Lambert, assume que a superfície é perfeitamente difusa (Lambertiana), o que não condiz com a natureza dos objetos e superfícies reais. As desvantagens para esta abordagem são: o coeficiente de reflexão difusa é de difícil determinação, e a necessidade de, no mínimo, uma fonte de iluminação com posições conhecidas. As referências HORN, 1977; WOODHAM, 1978; GALO & TOZZI, 1996; e HASEGAWA, 1997 são recomendadas sobre esse assunto.

Shape from motion

A técnica denominada *Shape from motion* utiliza parâmetros geométricos para a determinação da estrutura tridimensional da cena, além da variável tempo. Os modelos matemáticos empregados nesta técnica, consideram o deslocamento em função do tempo, da imagem, de pontos 3-D no plano imagem, portanto, neste caso existe a necessidade de se resolver o problema de correspondência, a exemplo do *shape from stereo*, mas de duas imagens tomadas em tempos distintos e não em posições distintas.

Como se pode deduzir, esta técnica é monocular e passiva e há a necessidade que a câmera, ou os objetos, ou ambos, estejam em movimento. O fato destes estarem em movimento e a necessidade de grandes seqüências de imagens para obter resultados precisos, pode ser uma desvantagem dessa técnica em algumas aplicações.

Esta técnica está habilitada à reconstrução tridimensional apenas de objetos rígidos 3-D (suas dimensões tridimensionais não variam em função do tempo) o que pode ser vantajoso ou desvantajoso conforme a aplicação.

As referências AGGARWAL *et al.*, 1981; BALLARD & BROWN, 1982; SCHALKOFF, 1989; BIGUN *et al.*, 1991; HO & PONG, 1996; BIMBO *et al.*, 1996 e POELMAN & KANADE, 1997 trazem informações mais detalhadas sobre esse assunto.

Shape from focus

O *Shape from focus* é uma técnica de reconstrução tridimensional que permite determinar as coordenadas 3-D de uma cena, a partir da modelagem dos efeitos que os parâmetros intrínsecos de uma câmera exercem sobre as imagens capturadas, com pequena profundidade de campo[1] . Os parâmetros intrínsecos utilizados são: a distância focal e a abertura do diafragma.

O fato primordial desta técnica está em tomar várias imagens de uma mesma cena variando a distância focal ou movimentando a câmera ou a cena na direção do eixo ótico da câmera, mantendo a profundidade de campo pequena. Com isso, obtém-se várias imagens com pontos focalizados e desfocalizados. Um ponto de uma certa imagem nitidamente focalizado, determina a distância focal ou a profundidade (se a distância focal for mantida fixa) correta para a formação daquela imagem para aquele ponto.

[1] Região do espaço-objeto onde as imagens de um objeto são nítidas dentro de um certo critério de aceitação.

A desfocalização de uma imagem pode ser visto como um processo de filtragem passa-baixa, onde as componentes de altas freqüências diminuem com o aumento da desfocalização, portanto, o operador de medida de foco deve ser um detetor de altas freqüências, ou seja, um filtro passa-alta. Uma vez determinado os pontos focalizados em cada imagem, pode-se associar a profundidade com esses pontos ou calculá-la através da equação de lentes, no caso em que a distância focal variou e a câmera e a cena permaneceram fixos durante as tomadas.

As principais vantagens desta técnica está em ser passiva e monocular, evitando o problema de correspondência de outras técnicas.

Devido a necessidade de pequena profundidade de campo, a abertura do diafragma da câmera deve ser grande, impossibilitando a utilização do modelo de câmera de orifício (fácil utilização). Outras desvantagens desta técnica são: ser extremamente sensível as condições de iluminação, e necessitar de várias imagens, aumentando o custo computacional.

São sugeridas as referências ENS & LAWRENCE, 1993; NAYAR & NAKAGAWA, 1994; NAYAR *et al.*, 1996 sobre este assunto.

Shape from texture

A técnica de *Shape from texture* consiste basicamente em projetar um padrão de textura sobre uma cena de interesse e capturar a imagem desta. Nesta técnica, a posição e orientação da câmera, bem como as do projetor, devem ser conhecidos. Esta é obviamente uma técnica ativa e monocular, porém, o problema de correspondência entre a imagem capturada e o padrão projetado existe e é semelhante ao problema de correspondência do *Shape from stereo.* A principal vantagem desta em relação ao *Shape from stereo*, é que o padrão a ser projetado pode ser elaborado com inferências geométricas, que facilitem a correspondência com a imagem. A maior desvantagem desta técnica é ser uma técnica ativa, limitando a sua aplicação a ambientes onde se pode controlar as condições de iluminação.

A reconstrução tridimensional é obtida a partir das relações geométricas entre o padrão de textura projetado e a imagem deste padrão.

As referências BAJCSY & LIEBERMAN, 1976; SHRIKHANDE & STOCKMAN, 1989; VUYLSTEKE & OOSTERLINCK, 1990; BALLARD & BROWN, 1982; e TOMMASELLI *et al.*, 1996 tratam este assunto mais detalhadamente.

Shape from stereo

A primeira técnica de reconstrução tridimensional (a partir de imagens) e, ainda hoje, a mais difundida, é a "visão estéreo" (*Shape from stereo*), que consiste basicamente em utilizar duas ou mais câmeras em posições distintas para obter a informação tridimensional da cena, em relação ao sistema de coordenadas adotado.

A visão estéreo utiliza-se da disparidade entre as imagens como a fonte de informação para a recuperação da estrutura tridimensional, uma vez que esta se relaciona de maneira inversamente proporcional à profundidade da cena.

Em um sistema típico de visão estéreo composto por duas câmeras, a disparidade associada a um ponto qualquer em uma imagem pode ser entendido como o deslocamento aparente entre este ponto e o seu ponto homólogo na outra imagem.

A maior dificuldade desta técnica de reconstrução é a determinação dos pontos homólogos entre as duas imagens. Devido a sua importância, esta dificuldade é denominada de "problema de correspondência".

A solução do problema de correspondência pode ser representado na forma de um mapa de disparidade. Para o caso discreto, este mapa pode ser representado por uma matriz cujo seus elementos são os valores de disparidade. Quando esta matriz não possui valores de disparidade para todos os seus elementos gera-se um mapa de disparidade esparso.

O mapa de disparidade esparso é uma solução parcial para o problema de correspondência, pois, a partir deste, não é possível reconstruir a estrutura tridimensional da cena como um todo.

O mapa de disparidade denso é uma matriz que possui valores de disparidade para todos os seus elementos, tornando possível a reconstrução completa da estrutura tridimensional da cena. Entretanto, este tipo de mapa de disparidade, apesar de denso, é gerado não permitindo variações abruptas entre valores de disparidade vizinhos, na maioria das publicações relacionadas, limitando com isso, as aplicações desta técnica para apenas a reconstrução de superfícies suaves.

Portanto, em aplicações gerais de reconstrução tridimensional, o mapa de disparidade gerado deve ser denso e permitir variações abruptas entre valores de disparidade vizinhos.

1.3 Objetivo do trabalho

O objetivo desse trabalho é desenvolver uma metodologia para o problema de correspondência estéreo aplicável em cenas não específicas, visando a obtenção de uma mapa de disparidade denso que permita variações abruptas nos valores de disparidade.

1.4 Organização do trabalho

Este trabalho é composto de seis capítulos, incluindo esta introdução.

No capítulo 2 são feitas considerações a respeito das relações geométricas no processo estéreo, revisando alguns conceitos sobre transformações geométricas e relações polinomiais, formação da imagem, erros sistemáticos, calibração de câmeras, disparidade, geometria epipolar e retificação estéreo.

O capítulo 3 trata-se o estabelecimento da correspondência de pontos homólogos, as estratégias de *matching*, premissas, evidências biológicas e trabalhos relacionados.

O capítulo 4 trata da solução do problema de correspondência através da metodologia proposta.

No capítulo 5 são apresentados resultados dos testes realizados em dados simulados e reais.

O capítulo 6 apresenta as conclusões sobre este trabalho e algumas sugestões para trabalhos futuros.

Capítulo 2

Aspectos geométricos da visão estéreo

2. Introdução

Em visão estéreo, as coordenadas tridimensionais de um ponto no referencial do espaço-objeto são calculadas através de uma triangulação. O triângulo utilizado é formado por um ponto P no referencial do espaço-objeto, a imagem p_E deste ponto no plano sensor p_E da câmera da esquerda e a imagem p_D deste ponto no plano sensor p_D da câmera da direita (figura 2.1).

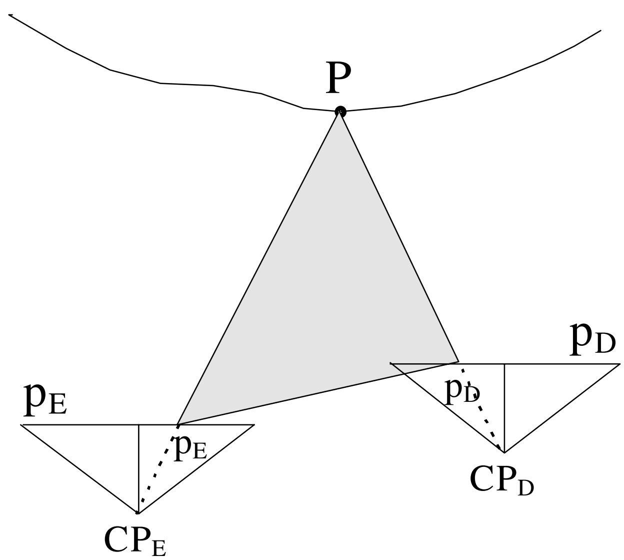


Fig. 2.1 - Triângulo da geometria estéreo.

A reconstrução das coordenadas tridimensionais em visão estéreo é, portanto, um problema baseado apenas em relações geométricas entre o par de imagens estereoscópicas.

Este capítulo apresenta uma breve revisão sobre as principais transformações geométricas e relações polinomiais que possibilitam melhor compreensão dos demais tópicos deste capítulo: formação geométrica da imagem por meio de equações de colinearidade, erros sistemáticos, calibração de câmeras, disparidade, geometria epipolar, retificação estéreo e modelos para reconstrução tridimensional.

As transformações geométricas e relações polinomiais são ainda, um importante requisito para compreensão da dimensão e complexidade do problema de correspondência estéreo, tratado em detalhes no capítulo 3.

2.1 Transformações geométricas e relações polinomiais

Transformação geométrica é uma relação funcional entre objetos geométricos de dois espaços.

Segundo KLEIN (1939) , transformação geométrica é uma generalização da noção de função de movimento.

As transformações geométricas podem ser: ativas ou passivas. Nas transformações ativas, o espaço é transformado e o sistema de referência é mantido fixo, já nas transformações passivas, o espaço é estacionário e o sistema de referência se movimenta (LUGNANI, 1987).

As relações polinomiais, a exemplo das transformações geométricas, também relacionam objetos entre dois espaços, porém, quando estas são utilizadas, a invariância a uma ou mais propriedades geométricas não são garantidas necessariamente.

2.1.1 Transformação ortogonal

A transformação ortogonal é também denominada de transformação de corpo rígido.

O modelo matemático da transformação ortogonal no plano é representado por:

$$\begin{bmatrix} X^* \\ Y^* \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & \text{sen}(\alpha) \\ -\text{sen}(\alpha) & \cos(\alpha) \end{bmatrix} . \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \qquad (2.1)$$

onde:

X e Y são coordenadas no espaço a ser transformado;

X* e Y* são coordenadas no espaço transformado;

X_0 e Y_0 são parâmetros de translação nas direções X e Y, respectivamente; e

α é o ângulo de rotação no plano.

No espaço, a transformação ortogonal é representada por:

$$\begin{bmatrix} X^* \\ Y^* \\ Z^* \end{bmatrix} = R.\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \tag{2.2}$$

onde:

$$R = \begin{bmatrix} \cos\Phi\cos K & \cos\Omega \,\mathrm{sen} K + \mathrm{sen}\Omega\,\mathrm{sen}\Phi\cos K & \mathrm{sen}\Omega\,\mathrm{sen}K - \cos\Omega\,\mathrm{sen}\Phi\cos K \\ -\cos\Phi\,\mathrm{sen}K & \cos\Omega\cos K - \mathrm{sen}\Omega\,\mathrm{sen}\Phi\,\mathrm{sen}K & \mathrm{sen}\Omega\cos K + \cos\Omega\,\mathrm{sen}\Phi\,\mathrm{sen}K \\ \mathrm{sen}\Phi & -\mathrm{sen}\Omega\cos\Phi & \cos\Omega\cos\Phi \end{bmatrix} \tag{2.3}$$

X, Y, e Z são coordenadas no espaço a ser transformado;

X*, Y* e Z* são coordenadas no espaço transformado;

X_0, Y_0 e Z_0 são parâmetros de translação nas direções X, Y e Z, respectivamente; e

K, Φ e Ω são ângulos de rotação em tornos dos eixos X, Y e Z, respectivamente.

Este tipo de transformação geométrica afeta apenas a posição de objetos. A escala é preservada após a transformação.

2.1.2 Transformação isogonal

A transformação isogonal é também conhecida como transformação de similaridade, conforme ou de Helmert.

Esta transformação no plano é representada por:

$$\begin{bmatrix} X^* \\ Y^* \end{bmatrix} = \lambda.\begin{bmatrix} \cos(\alpha) & \mathrm{sen}(\alpha) \\ -\mathrm{sen}(\alpha) & \cos(\alpha) \end{bmatrix}.\begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \tag{2.4}$$

onde :

λ é o parâmetro de fator de escala; e

demais parâmetros como definido para a expressão 2.1.

Esta transformação no espaço é representada por:

$$\begin{bmatrix} X^* \\ Y^* \\ Z^* \end{bmatrix} = \lambda.R.\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \tag{2.5}$$

onde:

λ é o parâmetro de fator de escala; e

demais parâmetros como definido para a expressão 2.2.

Este tipo de transformação afeta a posição e a escala do objeto. A forma é preservada.

2.1.3 Transformação afim

A transformação afim no plano é definida analiticamente por:

$$\begin{bmatrix} X^* \\ Y^* \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} . \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \tag{2.6}$$

onde:

a_{ij} são parâmetros de difícil interpretação geométrica, com i, j = 1 e 2; e

demais parâmetros como definido para a expressão 2.1.

No espaço, a transformação afim é dada por:

$$\begin{bmatrix} X^* \\ Y^* \\ Z^* \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \tag{2.7}$$

onde:

a_{ij} são parâmetros de difícil interpretação geométrica, com i, j = 1, 2 e 3; e

demais parâmetros como definido para a expressão 2.2.

Os parâmetros a_{ij} englobam rotações, escalas e ângulos de não ortogonalidade entre os eixos. No caso bidimensional os parâmetros a_{ij} englobam uma rotação, dois fatores de escalas (uma na direção x e outra na direção y) e um ângulo de não ortogonalidade entre os eixos X e Y. No caso tridimensional os parâmetros a_{ij} englobam três rotações, três fatores de escalas (na direção X, na direção Y e outra na direção Z) e três ângulos de não ortogonalidade entre os eixos X, Y e Z.

Esta transformação afeta a posição, a escala e a forma do objeto. O paralelismo é preservado.

2.1.4 Transformação projetiva

A transformação projetiva no plano é modelada matematicamente por:

$$X^* = \frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1} \tag{2.8}$$

$$Y^* = \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1} \tag{2.9}$$

onde:

a_i são parâmetros de difícil interpretação geométrica, para i = 1,...8; e

demais parâmetros como definido para a expressão 2.1.

A transformação projetiva no espaço é representada por:

$$X^* = \frac{b_1X + b_2Y + b_3Z + b_4}{b_{13}X + b_{14}Y + b_{15}Z + 1} \tag{2.10}$$

$$Y^* = \frac{b_5X + b_6Y + b_7Z + b_8}{b_{13}X + b_{14}Y + b_{15}Z + 1} \tag{2.11}$$

$$Z^* = \frac{b_9X + b_{10}Y + b_{11}Z + b_{12}}{b_{13}X + b_{14}Y + b_{15}Z + 1} \tag{2.12}$$

onde:

b_i são parâmetros de difícil interpretação geométrica, para i = 1,...15; e

demais parâmetros como definido para a expressão 2.2.

Para representar a transformação projetiva em notação matricial, como nas demais transformações já citadas, é necessário utilizar o espaço de coordenadas homogêneas, para isto faz-se:

$$X = \frac{X_h}{W_h} \quad e \quad X^* = \frac{X_h^*}{W_h^*} \tag{2.13}$$

$$Y = \frac{Y_h}{W_h} \quad e \quad Y^* = \frac{Y_h^*}{W_h^*} \tag{2.14}$$

$$Z = \frac{Z_h}{W_h} \quad e \quad Z^* = \frac{Z_h^*}{W_h^*} \tag{2.15}$$

onde:

X_h , Y_h , Z_h e W_h são coordenadas homogêneas no espaço a ser transformado;

X, Y e Z são coordenadas heterogêneas no espaço a ser transformado;

X_h^*, Y_h^*, Z_h^* e W_h^* são coordenadas homogêneas no espaço transformado;

e

X^*, Y^* e Z^* são coordenadas heterogêneas no espaço transformado.

Com isto tem-se a transformação projetiva em coordenadas homogêneas no plano dada por:

$$\begin{bmatrix} X_h * \\ Y_h * \\ W_h * \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} . \begin{bmatrix} X_h \\ Y_h \\ W_h \end{bmatrix} \tag{2.16}$$

onde:

c_{ij} são parâmetros de difícil interpretação geométrica, para i = j = 1, 2 e 3.

A transformação projetiva em coordenadas homogêneas no espaço é representada por:

$$\begin{bmatrix} X_h * \\ Y_h * \\ Z_h * \\ W_h * \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{bmatrix} . \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ W_h \end{bmatrix} \tag{2.17}$$

onde:

d_{ij} são parâmetros de difícil interpretação geométrica, para i = j = 1,..., 4.

A transformação projetiva em coordenadas homogêneas no plano possui nove parâmetros, mas apenas oito são independentes, no espaço tem-se dezesseis parâmetros, mas somente quinze deles são independentes.

Esta transformação afeta a posição, a escala, a forma e o paralelismo do objeto. A colinearidade é preservada.

2.1.5 Relações polinomiais

As relações polinomiais são muito utilizadas em problemas de interpolação, modelagem de distorções e em outras aplicações que geralmente utilizam modelos empíricos.

Os polinômios empregados podem ser lineares ou não lineares em relação aos parâmetros e de várias ordens. Um exemplo de uma relação polinomial pode ser representado por:

$$X^* = p_0 + p_1X + p_2X^2 \tag{2.18}$$

$$Y^* = q_0 + q_1Y + q_2Y^2 \tag{2.19}$$

$$Z^* = r_0 + r_1Z + r_2Z^2 \tag{2.20}$$

onde:

p_i, q_i e r_i são parâmetros dos polinômios.

As relações polinomiais podem afetar a posição, a escala, a forma, o paralelismo e a colinearidade de objetos, dependendo do polinômio.

A tabela 2.1 mostra um resumo das transformações geométricas e relações polinomiais, dos efeitos antes e após a transformação, as propriedades geométricas afetadas após a transformação e a quantidade de parâmetros independentes, empregados nas transformações geométricas.

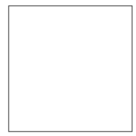
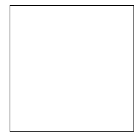
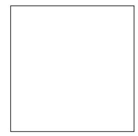
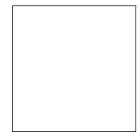
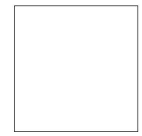
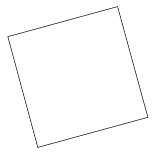
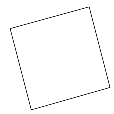
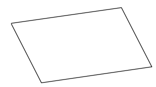
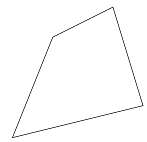
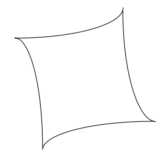
	Transf. Ortogonal	**Transf. Isogonal**	**Transf. Afim**	**Transf. Projetiva**	**Relação Polinomial**
Antes transf.					
Após transf.					
Posição	sim	sim	sim	sim	sim
Escala	não	sim	sim	sim	sim
Forma	não	não	sim	sim	sim
Paralelismo	não	não	não	sim	sim
Colinearidade	não	não	não	não	sim
Parâmetros[1] 2-D	3	4	6	8	dependente polinômio
Parâmetros[1] 3-D	6	7	12	15	dependente polinômio

Tabela 2.1 - Resumo das transformações geométricas e relações polinomiais.

2.2 Formação da imagem

A luz refletida pelos objetos presentes em uma cena é capturada pelo sistema sensor da câmera proporcionalmente a sua intensidade. A posição que cada raio luminoso irá ocupar no plano sensor da câmera, é modelado matematicamente pelas equações de colinearidade.

As equações de colinearidade são também denominadas erroneamente de equações projetivas. De acordo com LUGNANI (1987), as equações projetivas são transformações geométricas que possuem uma importante propriedade, entre várias outras, a de inversibilidade, que não se verifica nas equações de colinearidade. As equações de colinearidade diretas modelam um mapeamento de um espaço

[1] Parâmetros 2 e 3-D referem-se ao número de parâmetros independentes empregados.

tridimensional para um espaço bidimensional, logo a sua inversibilidade é impossível, sem que informações adicionais sejam fornecidas. Assim, as equações de colinearidade são relações funcionais entre os sistemas referenciais do espaço-imagem e espaço-objeto e vice-versa.

Antes de formalizar estas equações, faz-se necessário que um sistema referencial do espaço-objeto e um sistema referencial do espaço-imagem, bem como o sentido de contagem dos ângulos de rotação dos eixos entre os dois sistemas, seja adotado arbitrariamente. Nas figuras 2.2 e 2.3, as letras maiúsculas representam os eixos no espaço-objeto, as letras minúsculas, representam os eixos no espaço-imagem e as letras do alfabeto grego, representam os ângulos de rotação entre os eixos dos dois sistemas: K (kapa) representa a rotação em torno do eixo Z, Φ (fi) representa a rotação em torno do eixo Y e Ω (omega) representa a rotação em torno do eixo X.

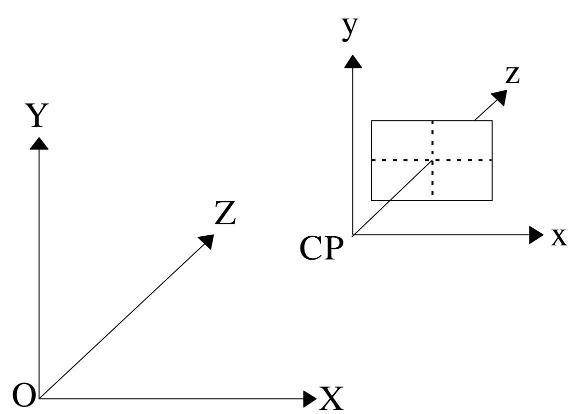


Fig. 2.2 - Sistema global e sistema imagem de coordenadas.

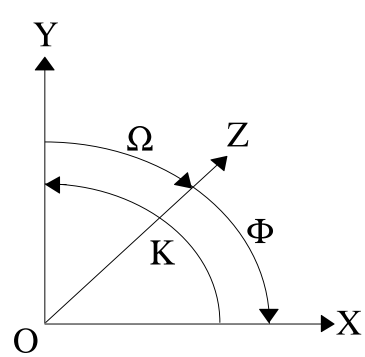


Fig. 2.3 - Sentido de contagem dos ângulos.

Ambos os sistemas adotados são levógiros (obedecem a regra da mão esquerda). Nos sistemas levógiros, as rotações positivas são contadas no sentido horário quando o sistema é visto a partir de um eixo positivo em direção a origem (FOLEY *et al*., 1982).

Essa configuração para o sistema de coordenadas foi adotada porque permite interpretação natural de que o sistema está superposto à tela (*display*) do computador e que grandes valores em Z estão mais distantes do observador de que pequenos valores neste eixo.

2.2.1 Equações de colinearidade diretas

As equações de colinearidade diretas podem ser expressas por uma simples semelhança de triângulos, como mostra a figura 2.4.

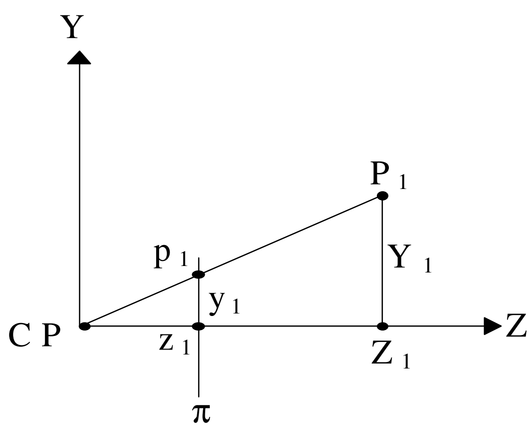


Fig. 2.4 - Formação da imagem.

Na figura 2.4, P_1 é um ponto no referencial do espaço-objeto de coordenadas (0, Y_1, Z_1) e p_1 é a imagem deste ponto no plano sensor π de coordenadas (0, y_1, z_1). CP é o centro perspectivo e portanto z_1 é o lugar geométrico onde o eixo ótico intercepta o plano sensor (denominado ponto principal), logo a distância $\overline{CP\ z_1}$ é a distância focal da câmera. Por semelhança dos triângulos $\overline{P_1Z_1CP}$ e $\overline{p_1z_1CP}$ obtém-se que :

$$\frac{Y_1}{y_1} = \frac{Z_1}{z_1} \tag{2.21}$$

e portanto

$$y_1 = \frac{z_1 . Y_1}{Z_1} \tag{2.22}$$

por analogia a expressão (2.22), tem-se:

$$x_1 = \frac{z_1 . X_1}{Z_1} \tag{2.23}$$

onde:

x_1 , y_1 e z_1 são coordenadas no referencial do espaço-imagem;

X_1, Y_1 e Z_1 são coordenadas no referencial do espaço-objeto.

Admitindo que o CP está na origem do sistema referencial do espaço-imagem, z_1 é igual a distância focal f, assim as coordenadas no referencial do espaço-imagem serão representadas por (x, y, f), onde f é constante para qualquer ponto neste espaço. Portanto as equações 2.22 e 2.23 ficam:

$$x = f \frac{X}{Z} \tag{2.24}$$

$$y = f \frac{Y}{Z} \tag{2.25}$$

As equações 2.24 e 2.25 permitem dizer que um ponto P_1 no referencial do espaço-objeto, a imagem deste ponto P_1 no referencial do espaço-imagem (p_1) e o centro perspectivo CP são colineares, possibilitando determinar as coordenadas de um ponto no referencial do espaço-imagem a partir das coordenadas deste ponto no referencial do espaço-objeto. Estas equações são denominadas equações de colinearidade diretas. Porém as equações 2.24 e 2.25 foram definidas desta maneira porque a origem do sistema global de coordenadas coincide com a origem do sistema imagem de coordenadas (CP) e também porque não existe rotações entre esses dois sistema.

Quando estas limitações não existem, faz-se necessário primeiro, transformar as coordenadas do sistema referencial do espaço-objeto para o sistema referencial do espaço-imagem e a partir destas, utilizar as equações 2.24 e 2.25. Esta transformação entre os dois sistemas de coordenadas é modelada por uma transformação isogonal no espaço. Formalizando, tem-se:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \lambda.R.\begin{bmatrix} X_G - X_{CP} \\ Y_G - Y_{CP} \\ Z_G - Z_{CP} \end{bmatrix} \quad (2.26)$$

onde:

X_i, Y_i e Z_i são coordenadas de um ponto P no sistema referencial do espaço-imagem;

X_G, Y_G e Z_G são coordenadas de um ponto P no sistema referencial do espaço-objeto;

X_{CP}, Y_{CP} e Z_{CP} são coordenadas do centro perspectivo;

λ é um fator de escala; e

R é a matriz de rotação, como definida para expressão 2.3.

Fazendo o produto matricial da expressão 2.26, resulta:

$$X_i = \lambda.[r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP}) \quad (2.27)$$

$$Y_i = \lambda.[r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP}) \quad (2.28)$$

$$Z_i = \lambda.[r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP}) \quad (2.29)$$

Portanto as coordenadas X_i, Y_i e Z_i são as coordenadas X_G, Y_G e Z_G transformadas para o sistema referencial do espaço-imagem, e agora estas coordenadas estão no mesmo sistema com que foram definidas as equações 2.24 e 2.25, podendo serem substituídas nestas. Como resultado, obtém-se:

$$x = f\frac{\lambda[r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP})]}{\lambda[r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})]} \tag{2.30}$$

$$y = f\frac{\lambda[r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP})]}{\lambda[r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})]} \tag{2.31}$$

Cancelando λ nas expressões 2.30 e 2.31, tem-se:

$$x = f\frac{r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \tag{2.32}$$

$$y = f\frac{r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \tag{2.33}$$

onde :

x e y são coordenadas de um ponto no referencial do espaço-imagem;

X_G,Y_G e Z_G são coordenadas de um ponto no referencial do espaço-objeto;

f é a distância focal;

X_{CP}, Y_{CP}, Z_{CP} são coordenadas do centro perspectivo da câmera; e

r_{ij} são os elementos da matriz de rotação R, com (i, j) =1,...,3, como definida para a expressão 2.3.

As expressões 2.32 e 2.33 são as equações de colinearidade diretas para casos genéricos. Considerando o caso particular onde $[X_{CP}\ Y_{CP}\ Z_{CP}] = [0\ 0\ 0]$ e $[K\ \Phi\ \Omega] = [0\ 0\ 0]$ estas expressões degeneram-se para as expressões 2.24 e 2.25.

Outra maneira de expressar as equações de colinearidade diretas consiste em utilizar uma representação linear. Porém, para isto, é necessário utilizar o conceito de coordenadas homogêneas. Nesta forma, as equações 2.32 e 2.33 são representadas por um produto matricial da forma:

$$\begin{bmatrix} x_h \\ y_h \\ w_h \end{bmatrix} = M.\begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix} \tag{2.34}$$

onde:

x_h, y_h e w_h são coordenadas homogêneas no sistema referencial do espaço-imagem;

X_h, Y_h, Z_h e 1 são coordenadas homogêneas no sistema referencial do espaço-objeto; e

M é uma matriz 3x4 composta pelos parâmetros de orientação (K, Φ e Ω), parâmetros de posição (X_{CP}, Y_{CP} e Z_{CP}) e a distância focal f.

A matriz M é comumente conhecida como matriz de calibração. Esta matriz pode ser escrita da seguinte forma:

$$M = \begin{bmatrix} f.r_{11} & f.r_{12} & f.r_{13} & -f.(r_{11}.X_{CP} + r_{12}.Y_{CP} + r_{13}.Z_{CP}) \\ f.r_{21} & f.r_{22} & f.r_{23} & -f.(r_{21}.X_{CP} + r_{22}.Y_{CP} + r_{23}.Z_{CP}) \\ r_{31} & r_{32} & r_{33} & -r_{31}.X_{CP} - r_{32}.Y_{CP} - r_{33}.Z_{CP} \end{bmatrix} \tag{2.35}$$

Reescrevendo a expressão 2.35 da seguinte forma:

$$\begin{bmatrix} x_h \\ y_h \\ w_h \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} . \begin{bmatrix} X_h \\ Y_h \\ Z_h \\ 1 \end{bmatrix} \tag{2.36}$$

onde:

m_{ij} são elementos da matriz de calibração.

e fazendo:

$$x = \frac{x_h}{w_h}; \text{ e} \tag{2.37}$$

$$y = \frac{y_h}{w_h} \tag{2.38}$$

obtém-se as coordenadas no sistema referencial do espaço-imagem, da mesma forma que nas expressões 2.32 e 2.33. A referência GANAPATHY (1984) é recomendada sobre este assunto.

2.2.2 Equações de colinearidade inversas

Utilizando as equações 2.24 e 2.25, pode-se escrever estas equações de forma inversa tal que:

$$X = \frac{x.Z}{f} \tag{2.39}$$

$$Y = \frac{y.Z}{f} \tag{2.40}$$

Pode-se notar que as coordenadas X e Y no sistema referencial do espaço-objeto dependem, das coordenadas x e y no sistema referencial do espaço-imagem, da distância focal f e, também de Z, que é uma variável no sistema referencial do espaço-objeto. Portanto, conclui-se que a partir de uma única imagem, não é possível determinar as coordenadas tridimensionais no referencial do espaço-objeto, sem que ao menos uma dessas componentes seja conhecida, caracterizando a não inversibilidade citada em 2.2. Porém, as equações de colinearidade diretas, podem ser escritas inversamente como em 2.39 e 2.40, com a ressalva de dependerem de uma coordenada do sistema referencial do espaço-objeto (tridimensional).

Em certas aplicações, as equações de colinearidade inversas podem ser utilizadas, como por exemplo, na reconstrução de uma superfície que a priori, possua alguma informação tridimensional, tal como, a profundidade desta em relação à câmera.

Como nas expressões 2.24 e 2.25, as expressões 2.39 e 2.40 estão definidas para um caso muito específico, onde a origem do sistema referencial do espaço-imagem (CP) coincide com a origem do sistema referencial do espaço-objeto e não há rotações entre os eixos dos dois sistemas. Para tornar tais expressões genéricas pode-se escrever a seguinte equação simétrica da reta:

$$\frac{X_G - X_{CP}}{X_i - X_{CP}} = \frac{Y_G - Y_{CP}}{Y_i - Y_{CP}} = \frac{Z_G - Z_{CP}}{Z_i - Z_{CP}} = t \qquad (2.41)$$

onde:

t é uma constante para esta reta.

Fazendo simples manipulações algébricas obtém-se:

$$X_G - X_{CP} = (Z_G - Z_{CP}).\frac{X_i - X_{CP}}{Z_i - Z_{CP}} \qquad (2.42)$$

$$Y_G - Y_{CP} = (Z_G - Z_{CP}).\frac{Y_i - Y_{CP}}{Z_i - Z_{CP}} \qquad (2.43)$$

De forma análoga a expressão 2.26, tem-se que:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda.R.\begin{bmatrix} X_i - X_{CP} \\ Yi - Y_{CP} \\ Zi - Z_{CP} \end{bmatrix} \qquad (2.44)$$

Fazendo as seguintes manipulações matriciais, obtém-se:

$$\begin{bmatrix} X_i - X_{CP} \\ Y_i - Y_{CP} \\ Z_i - Z_{CP} \end{bmatrix} = \frac{1}{\lambda}.R^{-1}.\begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (2.45)$$

Passando as coordenadas do centro perspectivo (CP) para o segundo termo da equação, resulta:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \frac{1}{\lambda}.R^{-1}.\begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} X_{CP} \\ Y_{CP} \\ Z_{CP} \end{bmatrix} \qquad (2.46)$$

Sendo que R é ortogonal (propriedade das matrizes de rotação), tem-se que:

$$R^{-1} = R^{T} \tag{2.47}$$

Substituindo a expressão 2.47 em 2.46 e realizando o produto matricial, obtém-se:

$$X_i = \frac{1}{\lambda}.(r_{11}.x + r_{21}.y + r_{31}.z) + X_{CP} \tag{2.48}$$

$$Y_i = \frac{1}{\lambda}.(r_{12}.x + r_{22}.y + r_{32}.z) + Y_{CP} \tag{2.49}$$

$$Z_i = \frac{1}{\lambda}.(r_{13}.x + r_{23}.y + r_{33}.z) + Z_{CP} \tag{2.50}$$

Substituindo as expressões 2.48 e 2.50 em 2.42, e 2.49 e 2.50 em 2.43, tem-se:

$$X_G - X_{CP} = (Z_G - Z_{CP}).\frac{\left(\frac{1}{\lambda}.(r_{11}.x + r_{21}.y + r_{31}.z) + X_{CP}\right) - X_{CP}}{\left(\frac{1}{\lambda}.(r_{13}.x + r_{23}.y + r_{33}.z) + Z_{CP}\right) - Z_{CP}} \tag{2.51}$$

$$Y_G - Y_{CP} = (Z_G - Z_{CP}).\frac{\left(\frac{1}{\lambda}.(r_{12}.x + r_{22}.y + r_{32}.z) + Y_{CP}\right) - Y_{CP}}{\left(\frac{1}{\lambda}.(r_{13}.x + r_{23}.y + r_{33}.z) + Z_{CP}\right) - Z_{CP}} \tag{2.52}$$

Fazendo simples manipulações matemáticas, obtém-se:

$$X_G = X_{CP} + (Z_G - Z_{CP}).\frac{r_{11}.x + r_{21}.y + r_{31}.z}{r_{13}.x + r_{23}.y + r_{33}.z} \tag{2.53}$$

$$Y_G = Y_{CP} + (Z_G - Z_{CP}).\frac{r_{12}.x + r_{22}.y + r_{32}.z}{r_{13}.x + r_{23}.y + r_{33}.z} \tag{2.54}$$

onde:

$z = f.$

As expressões 2.53 e 2.54 são as equações de colinearidade inversas para casos genéricos. No caso específico de $[X_{CP}\ Y_{CP}\ Z_{CP}] = [0\ 0\ 0]$ e $[K\ \Phi\ \Omega] = [0\ 0\ 0]$, as expressões 2.53 e 2.54 degeneram-se para as expressões 2.39 e 2.40.

2.3 Erros sistemáticos

Os erros sistemáticos presentes em um sistema sensor são ocasionados por três fatores que são:

1) Deslocamento do ponto principal;
2) Distorções do sistema de lentes; e
3) Escalas diferentes do *pixel* nas direções x e y.

O deslocamento do ponto principal é dado pela distância entre este ponto e o centro da imagem.

As distorções do sistema de lentes ocorrem devido a impossibilidade prática de construir uma superfície ideal para as lentes. Esta superfície deveria ser um perfeito parabolóide de revolução. Esta dificuldade origina a distorção radial simétrica. Outro problema associado às distorções do sistema de lentes, é devido também a impossibilidade prática em alinhar perfeitamente os eixos óticos das lentes, que compõem o sistema de lentes. Devido a esta segunda dificuldade origina-se a distorção descentrada. Porém a grande maioria das aplicações de visão artificial desprezam a distorção descentrada devido a sua pequena magnitude. A distorção radial simétrica é considerada em algumas aplicações que requerem alta precisão.

Escalas diferentes do *pixel* nas direções x e y resultam em um *pixel* não mais quadrado, mas sim retangular. Este problema está associado à freqüência de transmissão dos dados da câmera não ser necessariamente igual à freqüência de transmissão do conversor A/D (GALO, 1993), e/ou devido a própria construção do *pixel* não ser quadrado em algumas câmeras, sendo que neste caso, esta diferença de escala é nominal, sendo fornecida pelo fabricante.

Estes erros sistemáticos podem ser eliminados reduzindo-se as coordenadas de tela ao referencial imagem (centro da imagem).

A obtenção dos parâmetros de correção para estes erros sistemáticos será tratada em 2.4.

Assumindo que os parâmetros de correção dos erros sistemáticos são conhecidos, o seguinte modelo, descrito em TOMMASELLI & TOZZI (1993), pode ser utilizado para reduzir as coordenadas de tela ao referencial imagem.

$$x = x_R - c_x + (x_R - c_x).(k_1 r^2) + (x_R - c_x).ds_x \qquad (2.55)$$

$$y = y_R - c_y + (y_R - c_y).(k_1 r^2) \qquad (2.56)$$

onde:

x e y são coordenadas no referencial do espaço-imagem, isentas de erros sistemáticos e reduzidas ao referencial imagem;

x_R e y_R são coordenadas reduzidas ao centro aproximado da imagem;

c_x e c_y são coordenadas do centro de distorção radial das lentes;

k_1 é o coeficiente de primeira ordem de distorção radial simétrica;

ds_x é o fator de escala entre as dimensões horizontais e verticais do *pixel*; e

$$r^2 = x_R^2 + y_R^2 \qquad (2.57)$$

2.4 Calibração de câmeras

A calibração de câmeras é uma fase do processo estéreo na qual são determinados os parâmetros intrínsecos e extrínsecos de uma câmera. Os parâmetros intrínsecos são:

-Distância focal;

-Posição do ponto principal;

-Intervalos de amostragem (fator de escala entre as dimensões horizontais e verticais do *pixel*); e

-Coeficientes de distorções do sistema de lentes.

Os parâmetros extrínsecos são:

-Posição do centro perspectivo (X_{CP}, Y_{CP}, Z_{CP}); e

-Orientação do sistema imagem (K, Φ, Ω).

O processo de calibração possibilita determinar todos esses parâmetros, porém os parâmetros intrínsecos estão relacionados com a construção física da câmera, sendo que uma vez determinados, podem ser considerados fixos. Os parâmetros extrínsecos exigem suas determinações toda vez que a câmera sofre algum movimento de rotação ou translação.

A formalização da calibração de câmeras se dá por resolver um sistema de equações. Para isto, é necessário conhecer as coordenadas de alguns pontos no referencial do espaço-imagem e no referencial do espaço-objeto.

A resolução do sistema de equações pode ser realizada de forma linear ou não-linear. Existem também modelos baseados na coplanaridade e outros na não coplanaridade dos pontos no referencial do espaço-objeto. As referências GALO (1993); TSAI (1986); e TSAI (1987) tratam este assunto mais detalhadamente. Um exemplo de modelo matemático, baseado no modelo de correção de erros sistemáticos (descrito em 2.3), que pode ser utilizado é o seguinte:

$$x_R - c_x + (x_R - c_x).(k_1 r^2) + (x_R - c_x).ds_x = f\,\frac{r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \quad (2.58)$$

$$y_R - c_y + (y_R - c_y).(k_1 r^2) = f\,\frac{r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP})}{r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})} \quad (2.59)$$

onde:

X_{CP}, Y_{CP} e Z_{CP} são coordenadas do centro perspectivo no sistema referencial do espaço-objeto;

X_G, Y_G e Z_G são coordenadas de um ponto no sistema referencial do espaço-objeto;

r_{ij} são elementos da matriz de rotação entre os sistemas referenciais do espaço-objeto e espaço-imagem, os quais são função dos ângulos K, Φ e Ω (definidos em 2.2), para i e j = 1,2 e 3);

f é a distância focal; e

demais parâmetros como definidos para as expressões 2.55 e 2.56.

Este modelo é não-linear e necessita que os pontos no espaço-objeto sejam não coplanares. Esta não coplanaridade dos pontos no espaço-objeto é necessária para este modelo para evitar a correlação de Z_G e f, que ocasionaria singularidade na matriz que é invertida.

Como cada ponto fornece duas equações, um total de seis pontos de coordenadas conhecidas é suficiente para resolver este sistema de maneira sobre-determinada, através, por exemplo, do método dos mínimos quadrados.

2.5 Disparidade

A partir de uma única imagem, a reconstrução tridimensional de uma cena através de relações geométricas é impossível devido ao problema de não inversibilidade citado em 2.2.2. Já em um sistema composto por no mínimo duas câmeras (estéreo), a informação tridimensional pode ser obtida pela diferença entre as coordenadas de um ponto na imagem da esquerda e as coordenadas de seu homólogo na imagem da direita. Esta diferença é denominada paralaxe ou disparidade.

Adotando-se um sistema referencial de câmera tal que a sua origem seja o canto superior esquerdo da imagem, o eixo x paralelo às linhas da imagem e o eixo y paralelo às colunas da imagem, sendo contados positivamente da esquerda para direita e de cima para baixo, respectivamente como na figura 2.5, a disparidade pode existir tanto na direção x quanto na direção y em uma imagem.

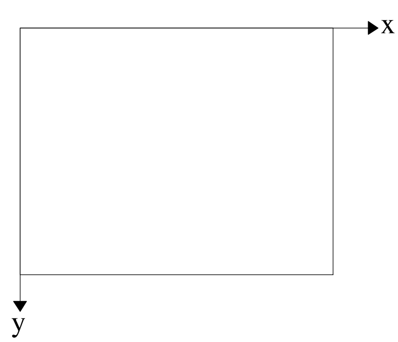


Fig. 2.5 - Sistema de tela de uma imagem.

A disparidade na direção y não é útil no cálculo da profundidade de uma cena e sim na determinação da orientação relativa entre o par de câmeras. Portanto a disparidade na direção x é a verdadeira informação tridimensional em um sistema estéreo. O termo disparidade será utilizado, a partir de agora, apenas para disparidade na direção x.

Em uma configuração ideal de câmeras para fins estéreo, onde os eixos óticos são paralelos e alinhados e as câmeras separadas por uma linha de base b, a profundidade pode ser calculada por:

$$Z = \frac{b.f}{d} \tag{2.60}$$

com

$$d = x_E - x_D \tag{2.61}$$

onde:

Z é a profundidade;

b é o comprimento da base;

f é a distância focal;

d é a disparidade; e

x_E e x_D são coordenadas de pontos homólogos na direção x nas imagens da esquerda e da direita, respectivamente.

2.6 Geometria epipolar e retificação estéreo

A geometria epipolar presente no par de imagens estereoscópicas é baseada no fato que um ponto no espaço-objeto (P), a projeção deste ponto na câmera da esquerda (p_E) e na câmera da direita (p_D), bem como os centros perspectivos de ambas as câmeras (CP_E e CP_D) são coplanares. A seguinte figura ilustra este fato:

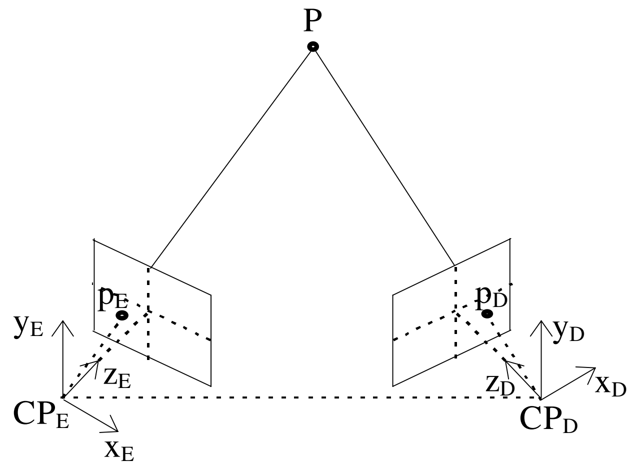


Fig. 2.6 - Geometria epipolar.

Esta injunção de coplanaridade dos pontos P, p_E, p_D, CP_E, CP_D faz com que dado um ponto em uma imagem (p_E), o seu ponto homólogo (p_D) esteja necessariamente sobre uma linha na outra imagem e vice-versa. Esta linha é denominada linha epipolar e o plano formado por tais pontos é denominado plano epipolar. A interseção do plano epipolar com os planos imagens de cada câmera geram as linhas epipolares homólogas. A figura 2.7 ilustra a situação em que apenas uma parte (para efeito de ilustração) do plano epipolar, intercepta os planos imagens das câmeras.

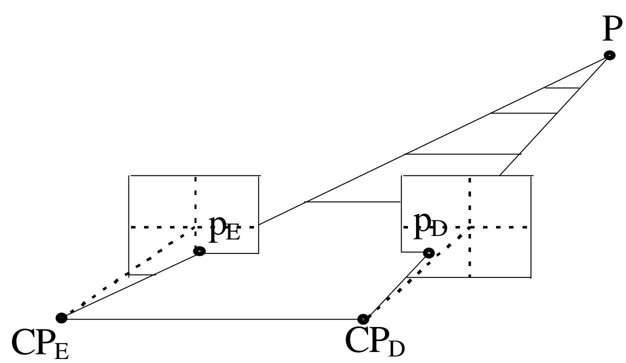


Fig. 2.7 - Plano epipolar e linhas epipolares homólogas.

As linhas epipolares homólogas podem ser determinadas através de vários modelos matemáticos diferentes, entretanto, todos os modelos matemáticos impõem a condição de coplanaridade entre os pontos homólogos, os centros perspectivos e o ponto objeto. Uma possível maneira para determinar as linhas epipolares homólogas, utiliza a equação de coplanaridade como o modelo matemático que impõe a injunção de coplanaridade. Uma outra possível maneira para determinar as linhas epipolares homólogas consiste em utilizar as equações de colinearidade diretas e inversas como modelo matemático para impor a injunção de coplanaridade.

A determinação das linhas epipolares homólogas utilizando a equação de coplanaridade é a maneira mais elegante (matematicamente) e comum para tal finalidade.

A utilização das equações de colinearidade diretas e inversas para determinar as linhas epipolares homólogas, é um caminho alternativo de mais fácil interpretação, embora seja um método mais caro computacionalmente.

2.6.1 Determinação das linhas epipolares homólogas através da equação de coplanaridade

A equação de coplanaridade é definida para os parâmetros relativos de orientação entre a câmera da esquerda e a câmera da direita. Se a calibração das câmeras é conhecida de forma absoluta, faz-se necessário calcular tais parâmetros de calibração de forma relativa. Uma prática muito comum que é utilizada para tal tarefa, é adotar o sistema imagem de coordenadas da câmera da esquerda como sendo o referencial do sistema do espaço-objeto e calcular os parâmetros de orientação relativa entre as câmeras da seguinte forma:

$$bx = X_{CPD} - X_{CPE} \tag{2.62}$$

$$by = Y_{CPD} - Y_{CPE} \tag{2.63}$$

$$bz = Z_{CPD} - Z_{CPE} \tag{2.64}$$

$$K_R = K_D - K_E \tag{2.65}$$

$$\Phi_R = \Phi_D - \Phi_E \tag{2.66}$$

$$\Omega_R = \Omega_D - \Omega_E \tag{2.67}$$

onde:

K_R, Φ_R e Ω_R são os ângulos de rotações relativos entre os sistemas imagens das câmeras da esquerda e da direita em torno dos eixos Z, Y e X, respectivamente;

K_E, Φ_E, Ω_E e K_D, Φ_D, Ω_D são os ângulos de rotações dos sistemas referenciais da câmera da esquerda e da direita, respectivamente, em relação ao sistema referencial do espaço-objeto, como definidos em 2.2;

bx, by e bz são componentes do vetor base na direção X, Y e Z, respectivamente; e

X_{CPE}, Y_{CPE}, Z_{CPE} e X_{CPD}, Y_{CPD}, Z_{CPD} são as coordenadas dos centros perspectivos das câmeras da esquerda e direita, respectivamente, no sistema referencial do espaço-objeto.

Admitindo que M_E e M_D são matrizes de rotação entre os sistemas imagens de coordenadas da câmera da esquerda e da direita, respectivamente, em relação ao sistema referencial do espaço-objeto, B é uma matriz composta pelos componentes do vetor base bx, by e bz e I_E é uma matriz composta pelos parâmetros de orientação interior x_{0E}, y_{0E} e f_E da câmera da esquerda e I_D é uma matriz composta pelos parâmetros de orientação interior x_{0D}, y_{0D} e f_D da câmera da direita, a equação de coplanaridade é definida como:

$$\begin{bmatrix} x_E & y_E & 1 \end{bmatrix} . I_E^T . M_E . B . M_D^T . I_D . \begin{bmatrix} x_D \\ y_D \\ 1 \end{bmatrix} = 0 \tag{2.68}$$

onde:

x_E e y_E são coordenadas no espaço-imagem da câmera da esquerda; e

x_D e y_D são coordenadas no espaço-imagem da câmera da direita.

$$B = \begin{bmatrix} 0 & bz & -by \\ -bz & 0 & bx \\ by & -bx & 0 \end{bmatrix}; \tag{2.69}$$

$$M_E = \begin{bmatrix} m_{E11} & m_{E12} & m_{E13} \\ m_{E21} & m_{E22} & m_{E23} \\ m_{E31} & m_{E32} & m_{E33} \end{bmatrix}; \tag{2.70}$$

$$M_D = \begin{bmatrix} m_{D11} & m_{D12} & m_{D13} \\ m_{D21} & m_{D22} & m_{D23} \\ m_{D31} & m_{D32} & m_{D33} \end{bmatrix}; \tag{2.71}$$

$$I_E = \begin{bmatrix} 1 & 0 & -x_{0E} \\ 0 & 1 & -y_{0E} \\ 0 & 0 & -f_E \end{bmatrix}; \tag{2.72}$$

$$I_D = \begin{bmatrix} 1 & 0 & -x_{0D} \\ 0 & 1 & -y_{0D} \\ 0 & 0 & -f_D \end{bmatrix}; \tag{2.73}$$

Porém, se a orientação relativa for determinada adotando-se o sistema referencial do espaço-imagem da câmera da esquerda como sendo o sistema referencial do espaço-objeto, como já citado, a matriz M_E degenera-se para a matriz identidade, ficando :

$$\begin{bmatrix} x_E & y_E & 1 \end{bmatrix} . I_E^T . B . M_D^T . I_D . \begin{bmatrix} x_D \\ y_D \\ 1 \end{bmatrix} = 0 \tag{2.74}$$

Fazendo:

$$F = I_E^T . M_E . B . M_D^T . I_D \tag{2.75}$$

onde:

F é a matriz fundamental.

A equação de coplanaridade pode ser escrita como:

$$\begin{bmatrix} x_E & y_E & 1 \end{bmatrix} F \begin{bmatrix} x_D \\ y_D \\ 1 \end{bmatrix} = 0 \qquad (2.76)$$

Fazendo este produto matricial, tem-se:

$$\begin{aligned} & x_E x_D f_{11} + y_E x_D f_{21} + x_D f_{31} + \\ & x_E y_D f_{12} + y_E y_D f_{22} + y_D f_{32} + \\ & x_E f_{13} + y_E f_{23} + f_{33} = 0 \end{aligned} \qquad (2.77)$$

onde:

f_{ij} são os elementos da matriz fundamental.

Como na expressão 2.77 os elementos f_{ij} são conhecidos, dado um ponto em uma imagem é possível determinar a equação da reta epipolar na outra imagem.

Como exemplo, assumindo que um ponto na imagem esquerda é conhecido, a equação da reta epipolar na imagem da direita fica:

$$(x_E f_{11} + y_E f_{21} + f_{31}) x_D + (x_E f_{12} + y_E f_{22} + f_{32}) y_D + (x_E f_{13} + y_E f_{23} + f_{33}) = 0 \qquad (2.78)$$

A expressão 2.78 é a equação geral da reta, onde os elementos entre os parênteses são os coeficientes desta equação.

Uma vez determinada a equação da reta epipolar na imagem direita, pode-se tomar qualquer ponto sobre esta e através da expressão 2.77, determinar a equação da reta epipolar na imagem da esquerda, obtendo assim, as retas epipolares homólogas.

2.6.2 Determinação das linhas epipolares homólogas através das equações de colinearidade diretas e inversas

A determinação das linhas epipolares homólogas utilizando as equações de colinearidade diretas e inversas é de fácil interpretação geométrica, pois consiste basicamente, em seguir uma metodologia composta por apenas três passos que são:

<u>Passo 1</u>

De posse dos parâmetros de calibração da câmera da esquerda, projeta-se um ponto no referencial do espaço-imagem, no referencial do espaço-objeto para duas coordenadas Z (espaço-objeto) arbitrárias e distintas, através das equações de colinearidade inversas (equações 2.53 e 2.54). Ao se projetar um mesmo ponto imagem p_{E1} no espaço-objeto com coordenadas Z's distintas, geram-se dois pontos P_1 e P_2 no espaço-objeto que estão sobre uma mesma reta (princípio da colinearidade), a figura 2.8 ilustra este fato:

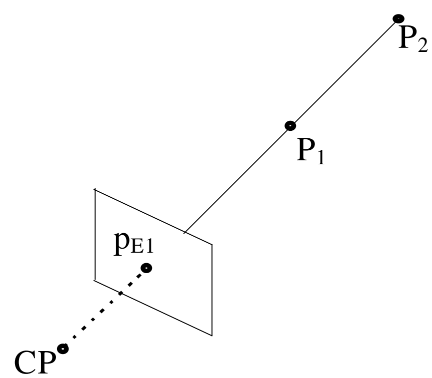


Fig. 2.8 - Projeção de um ponto no espaço-imagem para o espaço-objeto com Z's distintos.

<u>Passo 2</u>

De posse dos parâmetros de calibração da câmera da direita, projeta-se os pontos P_1 e P_2 nesta câmera utilizando as equações de colinearidade diretas (equações 2.32 e 2.33). Desde que a câmera da direita está em posição distinta da câmera da esquerda, a projeção de P_1 e P_2 resultará em dois pontos imagens p_{D1} e p_{D2} na câmera da direita, sendo que estes dois pontos imagens definem a equação da reta epipolar nesta câmera. A figura 2.9 ilustra este fato:

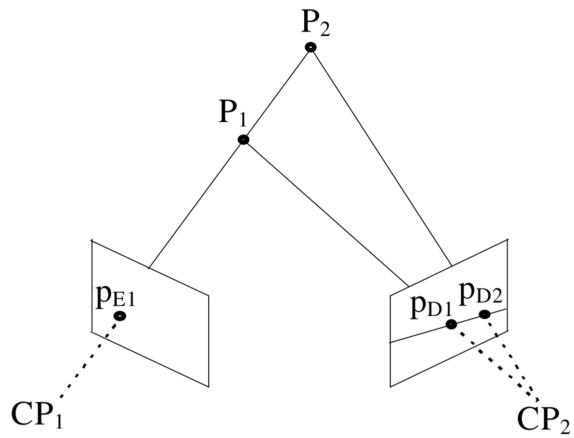


Fig. 2.9 - Geração da reta epipolar.

Passo 3

Para determinar a equação da reta epipolar na câmera da esquerda, faz-se necessário determinar apenas um ponto nesta câmera, já que p_{E1} é conhecido. Para isto pode-se tomar p_{D1} ou p_{D2} e da mesma forma que no passo 1; projeta-se este ponto no referencial do espaço-objeto com uma nova coordenada Z (espaço-objeto) arbitrária, dado os parâmetros de calibração da câmera da direita, utilizando as equações de colinearidade inversa, gerando um ponto P_3 no referencial do espaço-objeto. De posse de P_3 e os parâmetros de calibração da câmera da esquerda, projeta-se este ponto na câmera da esquerda, utilizando as equações de colinearidade diretas, gerando p_{E2}, sendo que p_{E1} e p_{E2} determinam a equação da reta epipolar na câmera da esquerda. A figura 2.10 mostra a determinação das linhas epipolares homólogas.

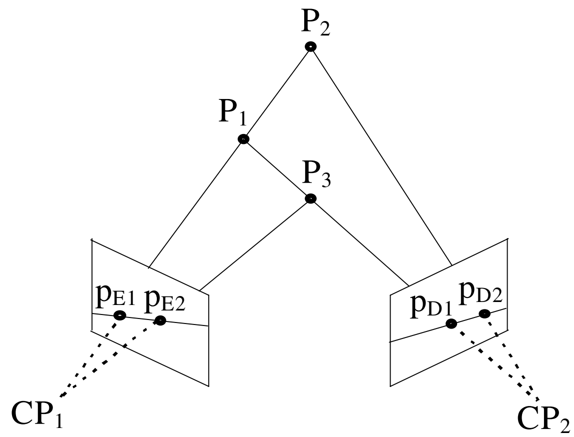
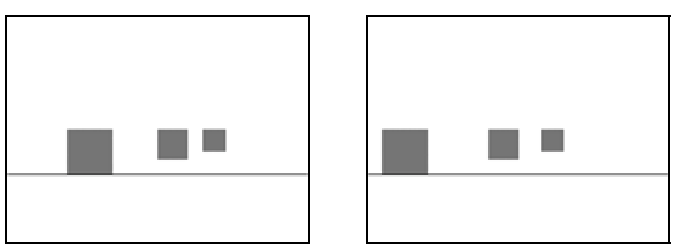


Fig. 2.10 - Determinação das linhas epipolares homólogas.

Este método de determinação das linhas epipolares homólogas é descrito detalhadamente em POZ & GALO (1992).

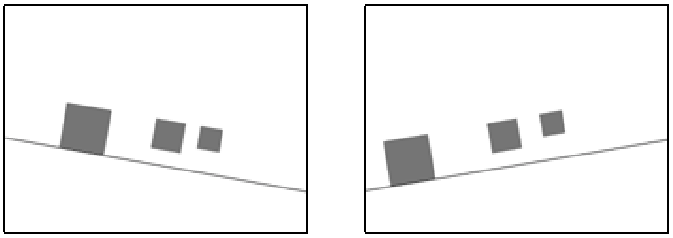
As figuras 2.11 a 2.19, mostram para cada par de imagens estereoscópicas um par de linhas epipolares homólogas para diferentes parâmetros de calibração de câmeras. As imagens são compostas por três cubos de mesmas dimensões no espaço-objeto, porém em diferentes profundidades. Estas imagens foram geradas utilizando o programa POV-RAY. No lado direito das figuras encontram-se os parâmetros de calibração das câmeras da esquerda (índice E) e da direita (índice D). A distância focal é igual a unidade para todas as imagens.



$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

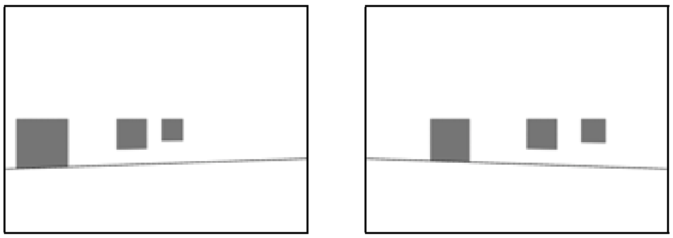
Fig. 2.11 - Linhas epipolares homólogas para câmeras deslocadas na direção X.

$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 10$	$K_D = -10$
$\Phi_E = 0$	$\Phi_D = 0$



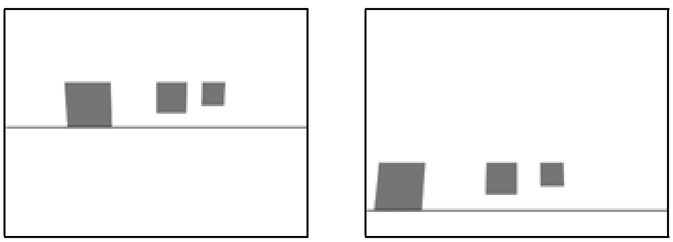
$\Omega_E = 0$ $\Omega_D = 0$

Fig. 2.12 - Efeito da rotação em torno do eixo Z.



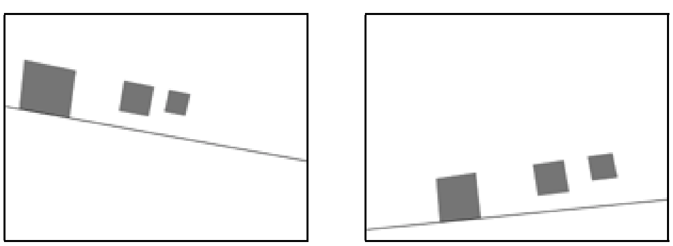
$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0$
$CP_{ZE} = 0$ $CP_{ZD} = 0$
$K_E = 0$ $K_D = 0$
$\Phi_E = 10$ $\Phi_D = -10$
$\Omega_E = 0$ $\Omega_D = 0$

Fig. 2.13 - Efeito da rotação em torno do eixo Y.



$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0$
$CP_{ZE} = 0$ $CP_{ZD} = 0$
$K_E = 0$ $K_D = 0$
$\Phi_E = 0$ $\Phi_D = 0$
$\Omega_E = 10$ $\Omega_D = -10$

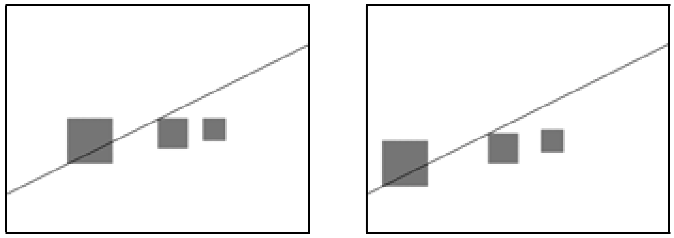
Fig. 2.14 - Efeito da rotação em torno do eixo X.



$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0$
$CP_{ZE} = 0$ $CP_{ZD} = 0$
$K_E = 10$ $K_D = -10$
$\Phi_E = 10$ $\Phi_D = -10$
$\Omega_E = 10$ $\Omega_D = -10$

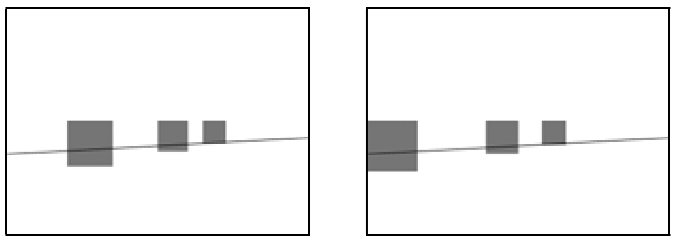
Fig. 2.15 - Efeito conjunto das rotações em torno dos eixos Z, Y e X

$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0.2$
$CP_{ZE} = 0$ $CP_{ZD} = 0$
$K_E = 0$ $K_D = 0$



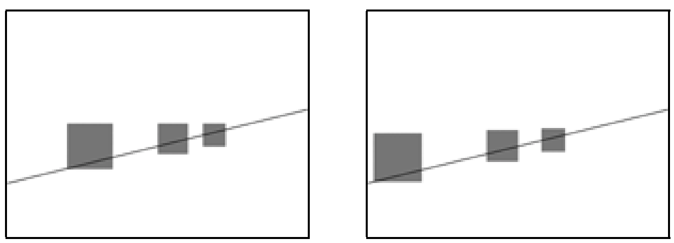
$\Phi_E = 0$ $\Phi_D = 0$
$\Omega_E = 0$ $\Omega_D = 0$

Fig. 2.16 - Efeito da translação na direção Y.



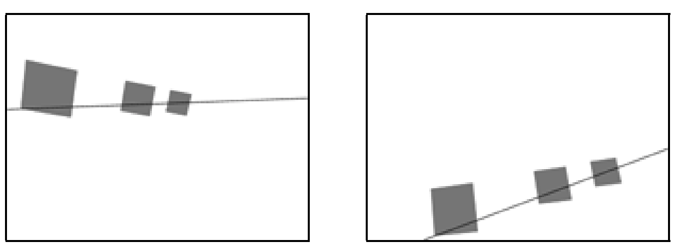
$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0$
$CP_{ZE} = 0$ $CP_{ZD} = 0.3$
$K_E = 0$ $K_D = 0$
$\Phi_E = 0$ $\Phi_D = 0$
$\Omega_E = 0$ $\Omega_D = 0$

Fig. 2.17 - Efeito da translação na direção Z.



$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0.2$
$CP_{ZE} = 0$ $CP_{ZD} = 0.3$
$K_E = 0$ $K_D = 0$
$\Phi_E = 0$ $\Phi_D = 0$
$\Omega_E = 0$ $\Omega_D = 0$

Fig. 2.18 - Efeito conjunto das translações nas direções Y e Z.



$CP_{XE} = 0$ $CP_{XD} = 1$
$CP_{YE} = 0$ $CP_{YD} = 0.2$
$CP_{ZE} = 0$ $CP_{ZD} = 0.3$
$K_E = 10$ $K_D = -10$
$\Phi_E = 10$ $\Phi_D = -10$
$\Omega_E = 10$ $\Omega_D = -10$

Fig. 2.19 - Efeito conjunto das translações em Y e Z e das rotações em torno dos eixos Z, Y e X.

Como pode-se notar nas figuras 2.11 a 2.19, qualquer ponto em uma imagem sobre uma linha epipolar estará obrigatoriamente sobre a linha epipolar homóloga na outra imagem, tornado o processo de correspondência unidimensional.

2.6.3 Retificação estéreo

Apesar das linhas epipolares homólogas facilitarem a tarefa de *matching,* ainda há a dificuldade da pesquisa do ponto homólogo ser realizada sobre a linha epipolar, que não é necessariamente coincidente com às linhas da imagem (*scanlines*). As linhas epipolares homólogas estarão paralelas às linhas das matrizes imagens se às rotações e translações sofridas na câmera na esquerda forem iguais às rotações e translações sofridas na câmera da direita, com exceção da translação na direção da base (no caso, a direção do eixo X), como por exemplo, a figura 2.11.

Para contornar o problema das linhas epipolares não estarem paralelas às linhas da imagem, pode-se gerar novas imagens retificadas de tal forma que as linhas epipolares estejam paralelas às linhas das matrizes imagens. Este processo é denominado retificação estéreo e para que este seja realizado corretamente, é necessário que:

1) As distâncias focais das imagens retificadas sejam iguais;

2) O plano sobre o qual serão geradas as imagens retificadas seja paralelo à base (eixo epipolar); e

3) As coordenadas de um ponto imagem e de seu ponto homólogo, só difiram na direção x, ou seja, que a disparidade em y seja nula.

Para realizar a tarefa de retificação estéreo é necessário portanto, reamostrar o par de imagens estereoscópicas ao longo das linhas epipolares homólogas, com isso, esta reamostragem implica em utilizar algum critério de interpolação nos níveis de cinza das imagens.

A interpolação pode ser realizada por vários métodos, tais como interpolação pelo vizinho mais próximo, linear, bilinear, polinomial cúbica, splines, etc.

A figura 2.20 mostra o par de imagens originais em linhas cheias e as imagens retificadas em linhas pontilhadas sobre um corte no plano Y. Nesta figura os parâmetros de orientação relativa da câmera da esquerda e da direita diferem apenas na direção Z (bz) e X (bx).

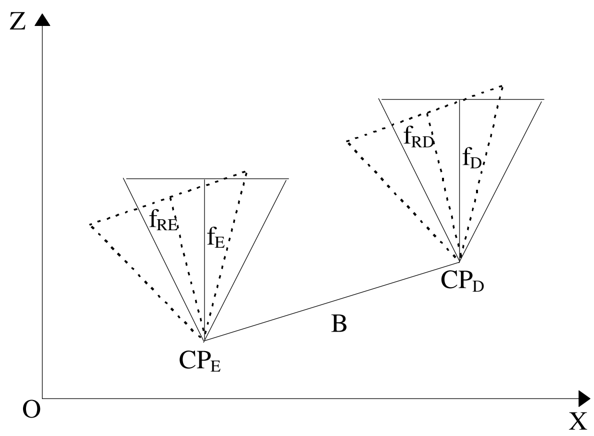


Fig. 2.20 - Imagens originais e imagens retificadas.

Ainda sobre a figura 2.20, se os parâmetros de calibração das câmeras são conhecido de forma absoluta ou relativa, é possível determinar quais serão os parâmetros para as imagens retificadas e então realizar uma transformação projetiva no plano entre a imagem original e a imagem retificada.

Outra maneira de realizar a retificação estéreo consiste em projetar os pontos das imagens originais em um plano π_o no referencial do espaço-objeto através das equações de colinearidade inversas para uma coordenada Z arbitrária e parâmetros de calibração das câmeras originais, análogo ao item 2.2.2; e então, projetar nos planos imagens retificados os pontos sobre π_o, através das equações de colinearidade diretas utilizando os parâmetros de calibração retificados. Este procedimento surtirá o mesmo efeito da transformação projetiva no plano entre as imagens originais e as imagens retificadas, pois as posições dos CP's não são alteradas. A figura 2.21 ilustra esta situação apenas para uma das câmeras onde π é o plano original, π_R e o plano retificado e π_o é o plano no sistema referencial do espaço-objeto.

Portanto, o processo de retificação estéreo consiste basicamente em determinar os parâmetros de calibração para as imagens retificadas, exceto as posições dos CP's, que são iguais as das imagens originais, e interpolar os níveis de cinza para as imagens retificadas.

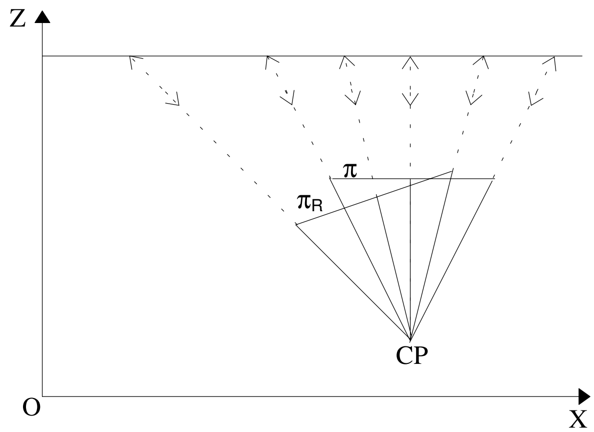


Fig. 2.21 - Projeção no espaço-objeto e no espaço-imagem.

2.6.4 Determinação dos parâmetros de calibração retificados

A determinação dos parâmetros de calibração retificados pode ser feita através de relações trigonométricas simples entre as componentes do vetor base nas direções X, Y e Z (B na fig. 2.20) e os eixos do sistema referencial do espaço-objeto, porém, primeiramente pode-se definir a distância focal retificada, que na maioria dos casos é adotada como a média das distâncias focais originais:

$$f_R = \frac{f_E + f_D}{2} \tag{2.79}$$

onde:

f_R é a distância focal retificada; e

f_E e f_D são as distâncias focais das câmeras da esquerda e da direita, respectivamente.

De acordo com a figura 2.3, que mostra o sentido positivo de contagem dos ângulos de rotação para o sistema levógiro adotado, pode-se determinar o ângulo de rotação K_R retificado através da expressão 2.80 onde:

$$K_R = \operatorname{arctg}\left(\frac{CP_{DY} - CP_{EY}}{CP_{DX} - CP_{EX}}\right) \tag{2.80}$$

A seguinte figura mostra esta relação trigonométrica sobre um corte no plano Z.

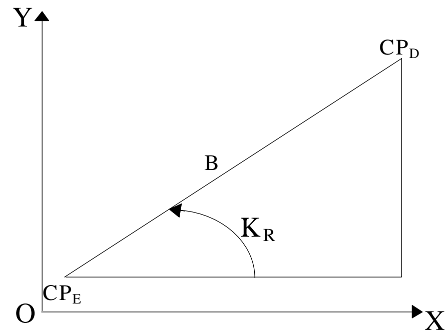


Fig. 2.22 - Determinação de K_R retificado.

De modo análogo a expressão 2.80, pode-se determinar o ângulo de rotação Φ_R retificado como:

$$\Phi_R = \operatorname{arctg}\left(\frac{CP_{EZ} - CP_{DZ}}{CP_{DX} - CP_{EX}}\right) \tag{2.81}$$

O sentido de contagem do ângulo Φ não coincide com o sentido de contagem para o sistema trigonométrico. Devido a este fato, o numerador na expressão 2.81 teve seus elementos invertidos em relação à expressão 2.80. A figura 2.23 mostra a determinação de Φ_R retificado sobre um corte no plano Y.

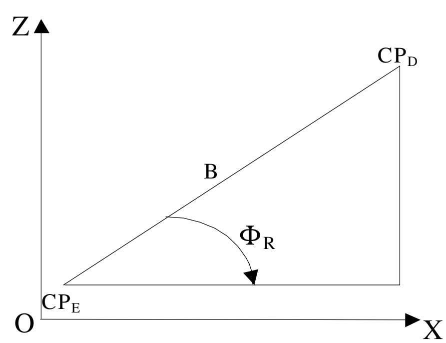


Fig. 2.23 - Determinação de Φ_R retificado.

O ângulo de rotação Ω_R retificado em torno do eixo X não pode ser determinado de maneira análoga à K_R e Φ_R, pois há uma indeterminação deste ângulo, devido que por uma reta (base) no espaço tridimensional, poderão passar infinitos planos sobre esta. Porém o ângulo Ω_R retificado pode ser calculado como a média dos ângulos Ω_E e Ω_D das calibrações das câmeras originais, ficando:

$$\Omega_R = \frac{\Omega_E + \Omega_D}{2} \tag{2.82}$$

Resumindo, os parâmetros de calibração para as imagens retificadas ficam:

$$f_{ER} = f_{DR} = \frac{f_E + f_D}{2}$$

$$CP_{E_R} = CP_E \qquad\qquad CP_{D_R} = CP_D$$

$$K_{ER} = K_{DR} = \operatorname{arctg}\left(\frac{CP_{DY} - CP_{EY}}{CP_{DX} - CP_{EX}}\right)$$

$$\Phi_{ER} = \Phi_{DR} = \operatorname{arctg}\left(\frac{CP_{EZ} - CP_{DZ}}{CP_{DX} - CP_{EX}}\right)$$

$$\Omega_{ER} = \Omega_{DR} = \frac{\Omega_E + \Omega_D}{2}$$

onde os índices:

E e D representam parâmetros de calibração para a câmera da esquerda e da direita, respectivamente; e

R representa parâmetros retificados.

Um método de interpolação muito utilizado é a interpolação bilinear, devido ao seu relativo baixo custo computacional e por fornecer bons resultados para este tipo de problema. Esta interpolação consiste em aplicar duas interpolações lineares na direção x, e uma na direção y aos quatro *pixels* vizinhos ao ponto de interesse.

Pode-se seguir o seguinte procedimento para a interpolação bilinear, proposto por STRAUCH (1991):

1) Interpolar o nível de cinza ao longo do segmento $\overline{(x_i, y_j)(x_{i+1}, y_j)}$, obtendo-se o nível de cinza na posição (x_i',y_j);

2) Interpolar o nível de cinza ao longo do segmento $\overline{(x_i, y_{j+1})(x_{i+1}, y_{j+1})}$, determinando o nível de cinza na posição (x_i',y_{j+1});

3) Interpolar o nível de cinza ao longo do segmento $\overline{(x_i', y_j)(x_i', y_{j+1})}$, obtendo-se o nível de cinza na nova posição (x_i',y_j').

A interpolação bilinear pode ser expressa pelo seguinte modelo matemático:

$$g(x_i',y_j') = g(x_i,y_j) + \Delta x\ [g(x_{i+1},y_j)-g(x_i,y_j)] + \Delta y\ [g(x_i,y_{j+1})-g(x_i,y_j)] + \Delta x \Delta y[g(x_i,y_j)-g(x_{i+1},y_j)-g(x_i,y_{j+1})+g(x_{i+1},y_{j+1})] \tag{2.83}$$

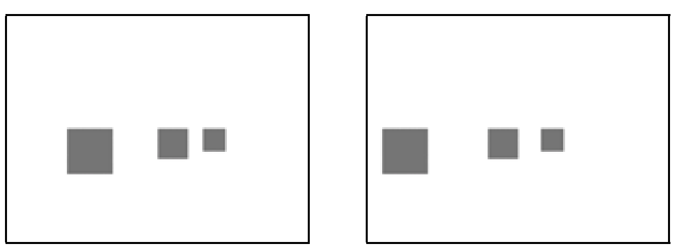
onde:

$g(x_i,y_j)$ é a função nível de cinza na posição (x_i,y_j),

$$\Delta x = x_i - x_i' \tag{2.84}$$

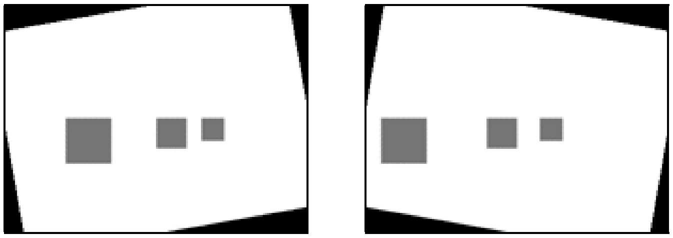
$$\Delta y = y_i - y_i' \tag{2.85}$$

A exemplo das figuras 2.11 a 2.19, as figuras 2.24 a 2.32 mostram os pares de imagens estereoscópicas agora retificadas para os mesmos parâmetros de calibração das câmeras originais das figuras anteriores.



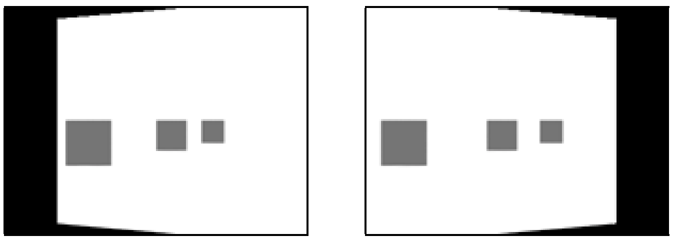
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.24 -Imagens retificadas para câmeras deslocadas na direção X.



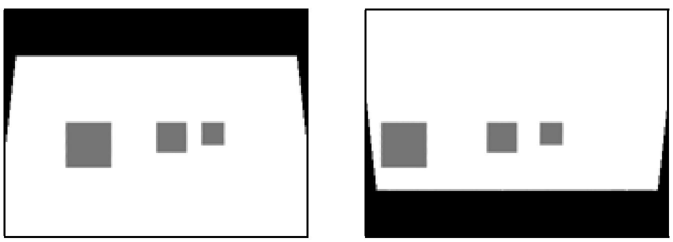
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 10$	$K_D = -10$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.25 - Imagens retificadas sob efeito da rotação em torno do eixo Z.



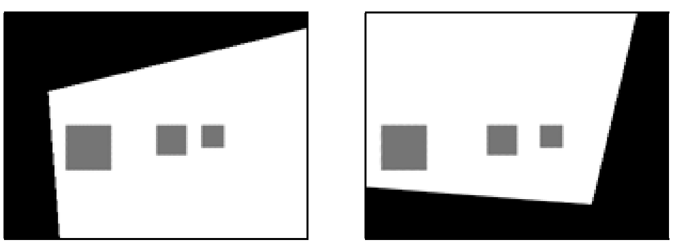
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 0$	$K_D = 0$
$\Phi_E = 10$	$\Phi_D = -10$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.26 - Imagens retificadas sob efeito da rotação em torno do eixo Y.



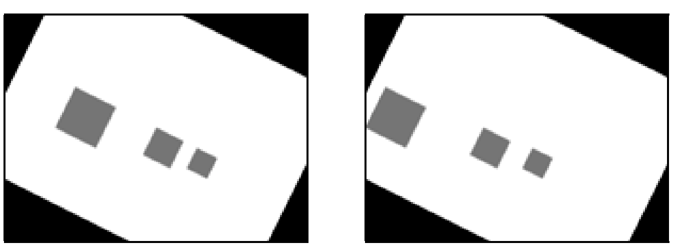
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 10$	$\Omega_D = -10$

Fig. 2.27 - Imagens retificadas sob efeito da rotação em torno do eixo X.



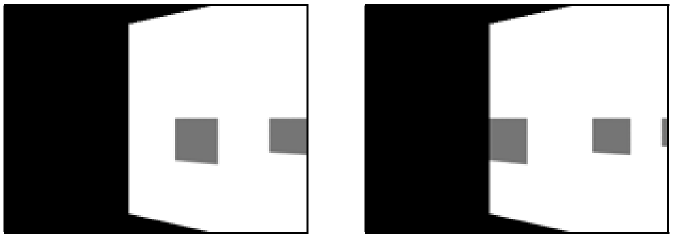
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 10$	$K_D = -10$
$\Phi_E = 10$	$\Phi_D = -10$
$\Omega_E = 10$	$\Omega_D = -10$

Fig. 2.28 - Imagens retificadas sob efeito conjunto das rotações em torno dos eixos Z, Y e X.



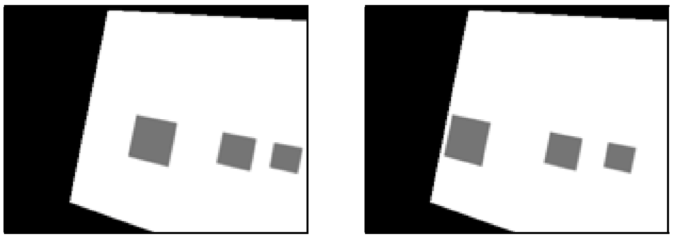
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0.2$
$CP_{ZE} = 0$	$CP_{ZD} = 0$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.29 - Imagens retificadas sob efeito da translação na direção Y.



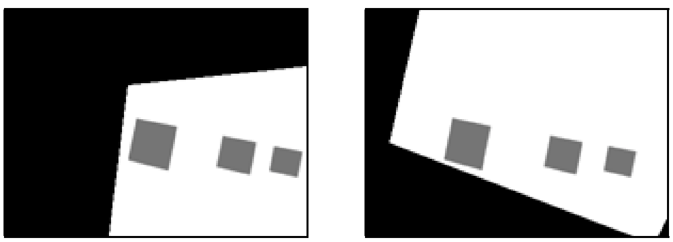
$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0$
$CP_{ZE} = 0$	$CP_{ZD} = 0.3$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.30 - Imagens retificadas sob efeito da translação na direção Z.



$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0.2$
$CP_{ZE} = 0$	$CP_{ZD} = 0.3$
$K_E = 0$	$K_D = 0$
$\Phi_E = 0$	$\Phi_D = 0$
$\Omega_E = 0$	$\Omega_D = 0$

Fig. 2.31 - Imagens retificadas sob efeito conjunto das translações nas direções Y e Z.



$CP_{XE} = 0$	$CP_{XD} = 1$
$CP_{YE} = 0$	$CP_{YD} = 0.2$
$CP_{ZE} = 0$	$CP_{ZD} = 0.3$
$K_E = 10$	$K_D = -10$
$\Phi_E = 10$	$\Phi_D = -10$
$\Omega_E = 10$	$\Omega_D = -10$

Fig. 2.32 – Imagens retificadas sob efeito conjunto das translações em Y e Z e das rotações em torno dos eixos Z, Y e X.

Nas figuras 2.24 a 2.32, nota-se também que as translações causam grandes distorções nas imagens retificadas em relação às imagens originais (figuras 2.29 a 2.31), porém para todas as imagens retificadas, as três imposições citadas em 2.6.3 são válidas, ou seja, qualquer ponto sobre uma linha em uma imagem, estará obrigatoriamente sobre a mesma linha na outra imagem, reduzindo significativamente o processo de correspondência ou *matching*.

Um problema que ocorre no processo de retificação estéreo é se as imagens retificadas forem geradas com o mesmo tamanho das imagens originais, haverá “cortes” nestas imagens em relação às imagens originais, como nas áreas escuras das imagens nas figuras 2.25 a 2.32, exceto nas imagens da figura 2.24 que na verdade não necessitam de retificação, pois suas linhas epipolares homólogas já são paralelas às linhas das matrizes imagens. A solução para este problema é aumentar o tamanho das imagens retificadas de tal forma que a imagem original “caiba” inteiramente na imagem retificada. Porém, para grandes variações entre os parâmetros de calibração das imagens originais e os parâmetros de calibração retificados, resultarão tamanhos muito superiores das imagens retificadas em relação às imagens originais, podendo assim, inviabilizar a utilização da retificação estéreo.

2.7 Reconstrução tridimensional

Existem várias maneiras de interpretar e demonstrar os modelos matemáticos de reconstrução tridimensional da visão estéreo. Todos os modelos, porém, são apenas formas diferentes de representação matemática, já que a obtenção das coordenadas tridimensionais no espaço-objeto são oriundas da triangulação citada em 2**.** e portanto totalmente dependente da disparidade associada ao par de pontos homólogos. Antes de apresentar os modelos de reconstrução tridimensional, será mostrado em 2.7.1 uma restrição do intervalo de correspondência que pode ajudar o processo de *matching*, reduzindo o espaço de pesquisa do ponto homólogo. No item 2.7.2 será tratado a reconstrução tridimensional propriamente dita.

2.7.1 Restrição do intervalo de correspondência

Como já visto em 2.6, a geometria epipolar possibilita que, dado um ponto em uma imagem, a pesquisa do ponto homólogo seja restrita à apenas uma linha na outra imagem. A restrição do intervalo de correspondência que será tratada neste item, restringe ainda mais a pesquisa do ponto homólogo, pois define um intervalo de busca ao longo da linha epipolar através de simples verificação geométrica.

Admitindo que as câmeras estão em uma configuração ideal, a profundidade mínima dos pontos no referencial do espaço-objeto pode ser calculado como:

$$P_{min} = \frac{b.\cot g(\alpha)}{2} \tag{2.86}$$

onde:

P_{min} é a profundidade mínima;

b é a base (eixo epipolar); e

α é o ângulo de cobertura.

A seguinte figura ilustra esta situação:

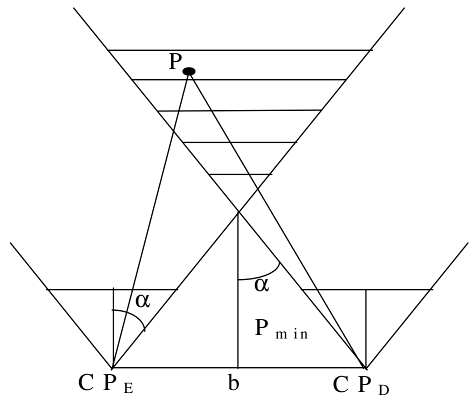


Fig. 2.33 – Profundidade mínima.

De acordo com a expressão 2.86, pode-se calcular a profundidade mínima para uma base de comprimento b com configuração das câmeras dita ideal. Porém a profundidade máxima não pode ser determinada, já que um ponto no infinito poderia ser imageado pelas câmeras. Portanto a profundidade máxima dependerá da precisão dos sistemas sensores, mais especificamente da resolução espacial (e implicitamente da resolução radiométrica) e da base b. Desta forma, um ponto em uma imagem poderá estar situado sobre qualquer posição ao longo da linha epipolar na outra imagem, porém se a configuração das câmeras é ideal, ou as imagens sofreram o processo de retificação estéreo, dado um ponto p_{E1} na imagem da esquerda, o seu ponto homólogo, p_{D1} estará obrigatoriamente em uma posição igual ou à esquerda, na imagem da direita da posição de p_{E1} na imagem da esquerda, independente de sua profundidade. O mesmo é válido para um ponto p_{D2} situado na imagem da direita, que terá seu ponto homólogo, p_{E2} em uma posição obrigatoriamente igual ou à direita da posição de p_{D2}, na imagem da direita. A figura 2.34 deixa claro esta situação:

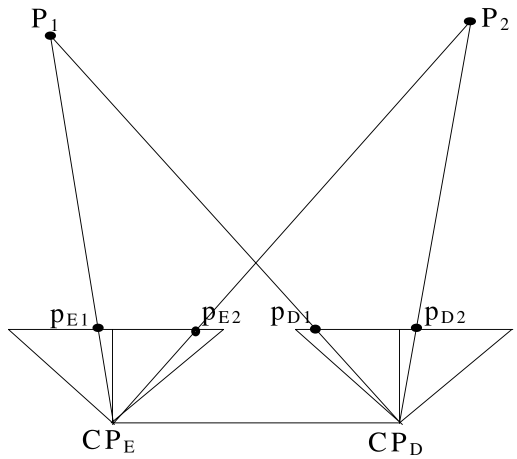


Fig. 2.34 - Restrição do intervalo de correspondência.

Esta restrição é mais simples de ser compreendida matematicamente. Para uma configuração de câmeras ideal, ou se as imagens são retificadas, a profundidade pode ser calculada pela expressão 2.60 e 2.61 que são reapresentadas abaixo. Como nesta expressão b e f são sempre positivos, a disparidade d deve ser sempre positiva ou nula, caso contrário a profundidade seria negativa e portanto a cena estaria atrás das câmeras, o que é impossível.

$$Z = \frac{b.f}{d}$$

$$d = x_E - x_D$$

Portanto esta simples, mas importante restrição, implica que:

$$x_E \geq x_D, \ \forall \ (x_E, x_D) \tag{2.87}$$

Esta restrição, apesar de ser geométrica, é menos aproveitada em configurações de câmeras com grande base b, pois a área de superposição das imagens diminui com o aumento da base b. Já para bases curtas, esta simples restrição pode ser bem utilizada, podendo restringir o espaço de busca do ponto homólogo significantemente.

2.7.2 Modelos para Reconstrução Tridimensional

A reconstrução tridimensional resulta da resolução de um sistema composto por três incógnitas e quatro equações. O modelo matemático é dado pelas equações de colinearidade diretas (expressões 2.32 e 2.33) já citadas. Este modelo é não-linear, o que torna a resolução do sistema dependente de valores iniciais aproximados e processos iterativos. Uma maneira de contornar este problema está em utilizar as equações de colinearidade escritas de forma linear (expressão 2.34), que implica em utilizar coordenadas homogêneas.

A resolução deste sistema (linear ou não-linear) determina a interseção das duas retas formadas pelos pontos homólogos e os centros perspectivos das duas câmeras (como ilustra a figura 2.1). Quando estas duas retas são quase paralelas, o ponto de interseção destas é de difícil determinação, tornando o sistema de equações mal condicionado para esta situação. Nestes casos, o modelo linear mostra sua vulnerabilidade, já que este é uma aproximação do modelo não-linear. Esta situação está relacionada diretamente com o comprimento da base e consequentemente com a profundidade da cena, portanto, a utilização de grandes comprimentos de base favorecem a reconstrução tridimensional, porém dificultam significantemente o processo de correspondência, que será tratado em detalhes no capítulo 3.

Em casos que se queira ou necessite utilizar o modelo não-linear, uma solução possível é utilizar o modelo linear para gerar valores aproximados e então utilizar o modelo não-linear para a solução mais refinada, porém, aumentando consideravelmente o custo computacional. As coordenadas imagem (x,y) que serão tratadas neste item são coordenadas reduzidas ao centro da imagem isentas de erros sistemáticos como tratado no item 2.3.

Recordando o item 2.2.1, as expressões 2.37 e 2.38 possibilitam realizar simples manipulações matemáticas podendo serem reescritas na seguinte forma:

$$X_h(m_{11} - m_{31}.\,x_h) + Y_h(m_{12} - m_{32}.\,x_h) + Z_h(m_{13} - m_{33}.\,x_h) = (m_{34}.\,x_h) - m_{14} \quad (2.88)$$

$$X_h(m_{21} - m_{31}.\,y_h) + Y_h(m_{22} - m_{32}.\,y_h) + Z_h(m_{23} - m_{33}.\,y_h) = (m_{34}.\,y_h) - m_{24} \quad (2.89)$$

Através das expressões 2.88 e 2.89, obtém-se um sistema de três incógnitas e duas equações, que é linear em relação às coordenadas X_h, Y_h e Z_h. Com a utilização de duas câmeras, pode-se escrever quatro equações para este sistema, que em notação matricial fica:

$$\begin{bmatrix} m_{E11} - m_{E31}x_E & m_{E12} - m_{E32}x_E & m_{E13} - m_{E33}x_E \\ m_{E21} - m_{E31}y_E & m_{E22} - m_{E32}y_E & m_{E23} - m_{E33}y_E \\ m_{D11} - m_{D31}x_D & m_{D12} - m_{D32}x_D & m_{D13} - m_{D233}x_D \\ m_{D21} - m_{D31}y_D & m_{D22} - m_{D32}y_{D2} & m_{D23} - m_{D33}y_D \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} m_{E34}x_E - m_{E14} \\ m_{E34}y_E - m_{E24} \\ m_{D34}x_D - m_{D14} \\ m_{D34}y_D - m_{D24} \end{bmatrix} \quad (2.90)$$

onde:

m_{cij} são os elementos ij da matriz de calibração. O índice c representa as matrizes de calibração para as duas câmeras (E esquerda e D direita);

x_E, y_E, x_D e y_D são coordenadas nos sistemas referenciais do espaço imagens das câmeras da esquerda e da direita (E esquerda e D direita), sem o índice h da expressões 2.88 e 2.89 para simplificação da notação; e

X, Y e Z são coordenadas tridimensionais no espaço-objeto sem o índice h da expressões 2.88 e 2.89 para simplificação da notação.

A expressão 2.90 pode ser resolvida simplesmente pelo método dos mínimos quadrados ou qualquer outra forma de resolução de sistemas lineares sobre-determinados.

Escrevendo a expressão 2.90 como :

$$A.V = B \quad (2.91)$$

onde:

A é a matriz 4x3 da expressão 2.90;

V é o vetor composto pelas coordenadas 3-D da expressão 2.90; e

B é o vetor 4x1 da expressão 2.90.

Pode-se resolver este sistema de uma maneira muito simples calculando a matriz pseudo-inversa A^+ de A, ficando:

$$V = A^{+}.B \tag{2.92}$$

onde:

$$A^{+} = (A^{T}.A)^{-1}.A^{T} \tag{2.93}$$

A resolução pelo método não-linear para a reconstrução tridimensional consiste em utilizar as equações de colinearidade diretas na sua forma original implicando portanto em processos iterativos e dependentes de valores iniciais aproximados. De acordo com as expressões 2.32 e 2.33, as equações de colinearidade diretas são dadas por:

$$x = f\,\frac{\lambda[r_{11}(X_G - X_{CP}) + r_{12}(Y_G - Y_{CP}) + r_{13}(Z_G - Z_{CP})]}{\lambda[r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})]}$$

$$y = f\,\frac{\lambda[r_{21}(X_G - X_{CP}) + r_{22}(Y_G - Y_{CP}) + r_{23}(Z_G - Z_{CP})]}{\lambda[r_{31}(X_G - X_{CP}) + r_{32}(Y_G - Y_{CP}) + r_{33}(Z_G - Z_{CP})]}$$

Considerando que a distância focal f, os elementos da matriz de rotação r_{ij} e as coordenadas do centro perspectivo X_{CP}, Y_{CP} e Z_{CP} são isentos de erros, e portanto, podem ser tratadas como constantes do ponto de vista estatístico, pode-se escrever as observações ajustadas $(x,y)^a$ em função dos parâmetros ajustados $(X,Y,Z)^a$ de acordo com o modelo matemático do método paramétrico (GEMAEL, 1994) dado por:

$$L_a = f(X_a) \tag{2.94}$$

onde:

L_a é o vetor das observações ajustadas; e

X_a é o vetor dos parâmetros ajustados.

Se f, r_{ij} e (X_{CP},Y_{CP},Z_{CP}) não podem ser considerados como constantes do ponto de vista estatístico, o método combinado é mais apropriado para tal resolução. Este método de ajustamento de observações é tratado em detalhes em GEMAEL (1994).

Considerando o método paramétrico como satisfatório para a resolução do problema de reconstrução tridimensional, faz-se:

$$A = \begin{bmatrix} \frac{\partial x_1}{\partial X} & \frac{\partial x_1}{\partial Y} & \frac{\partial x_1}{\partial Z} \\ \frac{\partial y_1}{\partial X} & \frac{\partial y_1}{\partial Y} & \frac{\partial y_1}{\partial Z} \\ \frac{\partial x_2}{\partial X} & \frac{\partial x_2}{\partial Y} & \frac{\partial x_2}{\partial Z} \\ \frac{\partial y_2}{\partial X} & \frac{\partial y_2}{\partial Y} & \frac{\partial y_2}{\partial Z} \end{bmatrix} \tag{2.95}$$

$$L_b = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} \tag{2.96}$$

$$X_0 = \begin{bmatrix} Xo \\ Yo \\ Zo \end{bmatrix} \tag{2.97}$$

$$L_0 = f(X_0) \tag{2.98}$$

onde:

x_E e y_E são as equações de colinearidade diretas escritas para a câmera da esquerda;

x_D e y_D são as equações de colinearidade diretas escritas para a câmera da direita;

A é a matriz das derivadas parciais;

L_b é o vetor das observações;

X_0 é o vetor dos parâmetros aproximados; e

L_0 é um vetor cujo seus elementos são função dos parâmetros aproximados, que são obtidos substituindo os parâmetros aproximados nas equações de colinearidade diretas.

De posse desses vetores e da matriz A citada, pode-se calcular o vetor de correções aos parâmetros aproximados como:

$$X = -\left(A^T.P.A\right)^{-1}.\left(A^T.P.L\right) \tag{2.99}$$

onde:

X é o vetor de correções aos parâmetros aproximados;

P é a matriz peso; e

$$L = L_0 - Lb. \tag{2.100}$$

Como X é o vetor de correções aos parâmetros aproximados, este deve ser somado ao vetor dos parâmetros aproximados como:

$$X_a = X_0 + X \tag{2.101}$$

onde:

X_a é o vetor dos parâmetros ajustados.

Este processo é iterativo, e portanto, após o cálculo de X_a , este deve ser assumido como o vetor dos parâmetros aproximados, reiniciando todo o processo. As iterações devem ser realizadas até que alcancem a precisão almejada ou um certo número de iterações. A matriz P é a matriz de peso das observações, portanto, observações mais confiáveis devem receber um peso maior de que as observações menos confiáveis. Porém, assumindo que todas as observações possuem a mesma precisão e que estas não são correlacionadas, a matriz peso degenera-se na matriz identidade.

Como se pode notar, este processo de resolução da reconstrução estéreo pelo método não-linear é de alto custo computacional, pois envolve a formação de diversas matrizes e vetores, além de somas, multiplicações e inversão de matrizes a cada iteração.

A tabela 2.2 mostra alguns resultados da reconstrução tridimensional das coordenadas de um ponto no sistema referencial do espaço-objeto para diferentes comprimentos de base, utilizando o modelo linear e o modelo não-linear. As coordenadas deste ponto foram projetadas no sistema referencial do espaço-imagem analiticamente, gerando os respectivos pontos homólogos e, portanto, não sujeitos a erros sistemáticos e a erros no processo de correspondência. As coordenadas tridimensionais escolhidas deste ponto foram:

$$X = 1000 \qquad Y = 1 \qquad Z = 1$$

Estas coordenadas foram escolhidas propositadamente, porque dificultam a interseção citada para as posições dos centros perspectivos das câmeras. A posição do centro perspectivo da câmera da esquerda, manteve-se fixo e coincidente com a origem do sistema referencial do espaço-objeto, ou seja, suas coordenadas são todas iguais a 0 (zero). A posição do centro perspectivo da câmera da direita, variou apenas na direção X de uma grandeza igual ao tamanho da base, dado na tabela 2.2.

Os ângulos de orientação dos sistemas referenciais das câmeras em relação ao sistema referencial do espaço-objeto são todos iguais a 0 (zero). A distância focal de ambas as câmeras é igual a 1 (um). O erro da reconstrução tridimensional, para o método linear e não-linear, foi determinado pela a diferença entre as coordenadas originais e as coordenadas reconstruídas, sendo dado apenas a ordem de grandeza destes para cada coordenada, devido às suas pequenas magnitudes.

Tamanho da base	**Ordem do erro (Linear)**	**Ordem do erro (Não linear)**
100	$(10^{-11}, 10^{-14}, 10^{-14})$	$(10^{-12}, 10^{-15}, 10^{-15})$
10	$(10^{-8}, 10^{-11}, 10^{-11})$	$(10^{-11}, 10^{-14}, 10^{-14})$
1	$(10^{-6}, 10^{-9}, 10^{-9})$	$(10^{-10}, 10^{-13}, 10^{-13})$
0.1	$(10^{-6}, 10^{-9}, 10^{-9})$	$(10^{-9}, 10^{-12}, 10^{-12})$
0.01	$(10^{-3}, 10^{-6}, 10^{-6})$	$(10^{-8}, 10^{-11}, 10^{-11})$
0.001	$(10^{-2}, 10^{-5}, 10^{-5})$	$(10^{-7}, 10^{-10}, 10^{-10})$

Tabela 2.2 - Comparações dos modelos de reconstrução tridimensional para diferentes tamanhos de base.

A tabela 2.2 mostra que os erros nas coordenadas X, Y e Z são de pequenas magnitudes para ambos os métodos de reconstrução tridimensional, porém, o método não-linear apresentou os menores erros em todos os casos.

Nas maiorias das aplicações da visão estéreo, o método linear pode ser utilizado sem grandes problemas, uma vez que os erros sistemáticos e principalmente, os erros no processo de estabelecimento de correspondência (que se tratado no capítulo 3), geralmente possuem magnitudes muito maiores de que as magnitudes dos erros, devido somente à utilização do modelo de reconstrução tridimensional (linear ou não-linear).

As referências KRAUS (1992) e HARTLEY (1997) tratam mais detalhamente os modelos de reconstrução tridimensional.

Capítulo 3

Estabelecimento da correspondência

3. Introdução

O estabelecimento da correspondência ou *matching* é a tarefa principal e mais difícil em visão estéreo. É esta tarefa que permite que seres humanos e demais seres vivos dotados de sistemas de visão binocular percebam as estruturas tridimensionais de uma cena.

Este assunto tem recebido grande atenção, tornando-se a principal linha de pesquisa por parte de pesquisadores da área de visão artificial, porém, este problema encontra-se ainda em aberto, já que a maioria das soluções encontradas resolvem este problema em casos muito específicos e em cenas bem controladas.

O problema de correspondência (*matching*) consiste em, dado um ponto em uma imagem, determinar o ponto homólogo na outra imagem. Devido a efeitos de oclusão, ruídos, superfícies transparentes e principalmente distorções geométricas e radiométricas entre o par de imagens, esta definição aparentemente simples do problema de correspondência (*matching*) torna-se um problema extremamente complexo. Todos estes problemas citados são conseqüência direta das posições distintas do par de câmeras, exceto os problemas relacionados a ruídos. Estes problemas serão denominados de “problemas da visão estéreo” neste trabalho.

A distorção radiométrica entre um par de pontos homólogos, ou seja, um ponto em uma imagem não ter necessariamente a mesma intensidade de seu ponto homólogo na outra imagem, é ocasionada devido às superfícies dos objetos em cenas reais não serem perfeitamente Lambertianas, e portanto, existir uma componente de reflexão especular[1] (que depende da direção de visualização) que é, teoricamente, o fator causador desta distorção.

[1] Veja modelo de iluminação de Phong em (FOLEY *et al.*, 1982).

Uma maneira para tornar um algoritmo de *matching* invariante a essa distorção, seria utilizar o modelo de iluminação da cena (quem modela esta distorção), porém para isso seria necessário conhecer a posição da fonte de luz e os coeficientes de reflexão difuso e especular das superfícies dos objetos.

As distorções geométricas entre o par de imagens são modeladas matematicamente por uma relação polinomial (citada em 2.1.5) cujo polinômio é formado por uma composição de transformações projetivas (citada em 2.1.4), sendo que para cada valor de disparidade, existe uma transformação projetiva no plano que transforma as coordenadas dos pontos em uma imagem para as coordenadas dos pontos homólogos na outra imagem, assim o número de transformações projetivas no plano que compõem esta relação polinomial existente em um sistema estéreo é igual ao número de valores de disparidades presente na cena. Formalizando, o problema de correspondência ou *matching* pode ser representado pela equação:

$$p_E = f_n(p_D) \tag{3.1}$$

onde:

p_E e p_D são pontos na imagem da esquerda e da direita, respectivamente e

f_n é uma relação polinomial composta por n transformações projetivas.

A figura 3.1 ilustra esta relação polinomial:

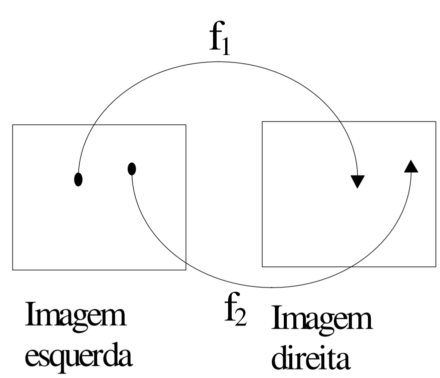


Fig. 3.1 - Função de *matching*.

O algoritmo de *matching* deveria ser portanto invariante a esta relação polinomial, porém não existe ferramentas matemáticas que consigam esta invariância. Por exemplo, momentos invariantes (HU, 1962), são invariantes apenas a transformação isogonal, muito menos complexa que esta relação polinomial.

Os efeitos de oclusão ocasionam que pontos em uma imagem não tenham correspondência na outra imagem ou como nos casos em que há cenas que possuem superfícies transparentes, ocasionando que um mesmo ponto em uma imagem possua dois ou mais pontos homólogos na outra imagem. O algoritmo de *matching* para ser invariante a esses dois problemas deveria possuir alguma forma de inteligência para detectar quais pontos do par de imagens são pontos de oclusão ou representam superfícies transparentes.

Além de todos esses problemas, ainda há o problema de ruídos inerentes nas imagens que podem ser de natureza sistemática ou aleatória, dificultando ainda mais o estabelecimento da correspondência.

Estes problemas citados são fundamentais para compreensão da dimensão e complexidade do problema de *matching* em visão estéreo.

Existem algumas premissas formuladas que auxiliam a resolução do problema de correspondência, mas muitas dessas são válidas sobre certas restrições na cena, ou que na maioria das vezes podem ocorrer, mas não em sua totalidade. Estas premissas serão tratadas em maiores detalhes em 3.3.

As estratégias de *matching* podem ser agrupadas em duas classes principais em respeito ao tipo de primitivas utilizadas que são:

-*Matching* baseado em áreas; e

-*Matching* baseado em feições.

As duas classes citadas serão tratadas em 3.1e 3.2 respectivamente, o item 3.4 apresenta algumas evidências biológicas da visão estéreo, o item 3.5 traz um resumo de algumas teorias e algoritmos contemporâneos para o estabelecimento da correspondência.

3.1 *Matching* baseado em áreas

O *matching* baseado em áreas utiliza-se da correlação de imagens digitais com funções específicas de correlação. Estas funções de correlação são funções matemáticas que determinam o grau de interdependência entre duas funções de mesma natureza (YANNIRIS, 1974).

O método de correlação de imagens digitais por funções de correlação, é aplicado selecionando-se uma janela (área) de referência em uma imagem e uma janela de pesquisa na outra imagem. Para tornar genérico as próximas demonstrações, será assumido que todas as janelas são bidimensionais, porém, com a injunção fornecida pela geometria epipolar, estas janelas podem ser unidimensionais diminuindo consideravelmente o custo computacional.

A janela de pesquisa deve ser obrigatoriamente maior que a janela de referência, podendo chegar ao tamanho da imagem inteira quando nenhuma informação é conhecida a priori sobre o tamanho desta. A janela de pesquisa é "cortada" em sub-janelas do tamanho da janela de referência e então aplicada a função de correlação escolhida entre a janela de referência e as sub-janelas.

O número de sub-janelas com dimensões da janela de referência existentes na janela de pesquisa, é função da dimensão das duas janelas e pode ser calculado por:

$$nsj = (p_x - r_x + 1).(p_y - r_y + 1) \tag{3.2}$$

onde:

nsj é o número de sub-janelas;
p_x é a dimensão da janela de pesquisa na direção x;
r_x é a dimensão da janela de referência na direção x;
p_y é a dimensão da janela de pesquisa na direção y; e
r_y é a dimensão da janela de referência na direção y.

Para cada uma destas $(p_x - r_x + 1).(p_y - r_y + 1)$ combinações, deve-se aplicar uma função de correlação e determinar (de acordo com o critério da função utilizada) a janela homóloga (GALO, 1994).

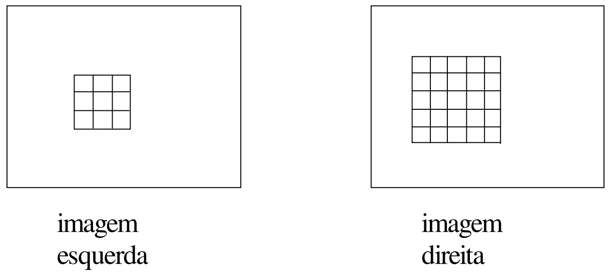


Figura 3.2 - Janela de referência 3x3 (imagem esquerda) e janela de pesquisa 5x5 (imagem direita)

Esta classe de *matching* tem a desvantagem de ser obtida diretamente dos níveis de cinza das imagens, que por sua vez, são muito sensíveis aos problemas da visão estéreo, citado na introdução deste capítulo, por outro lado nenhuma espécie de processamento monocular (extração de feições, por exemplo) se faz necessário, o que torna os algoritmos desta classe mais simples de que os algoritmos da classe de *matching* baseado em feições.

Grandes áreas são mais estáveis de que pequenas áreas nos processos de correlação porque estas possuem mais informações da função em questão, e portanto, maior poder de redução de ambigüidades, porém estes processos são sensíveis a transformações radiométricas e geométricas (exceto translação). Quanto aos problemas referentes as transformações geométricas, mais especificamente com relação as rotações relativas entre uma imagem e outra, estes podem ser eliminados realizando a correlação ao longo das linhas epipolares homólogas ou em imagens retificadas. Entretanto, para janelas muito grandes, os problemas referentes a escala, não ortogonalidade entre os eixos, a não preservação de paralelismo e a não preservação da colinearidade entre as duas imagens são acentuados, prejudicando o *matching*.

O *matching* baseado em áreas pode atingir precisões ao nível de *sub-pixel* quando um refinamento da correlação é utilizado, minimizando o erro médio quadrático gerado por um modelo composto por uma transformação geométrica e uma transformação radiométrica entre duas áreas homólogas. Sobre este assunto sugere-se as referências ACKERMANN (1984); e PERTL (1984).

3.1.1 Funções de correlação

Neste item são descritas as funções de correlação mais utilizadas. É considerado somente o caso discreto, já que as imagens digitais são discretas.

As variáveis utilizadas nas expressões serão representadas da seguinte forma:

p_x é a dimensão da janela de pesquisa na direção x;

r_x é a dimensão da janela de referência na direção x;

p_y é a dimensão da janela de pesquisa na direção y;

r_y é a dimensão da janela de referência na direção y;

$g_r(i,j)$ é o valor numérico da imagem (função de tom de cinza) na posição (i,j) no sistema da janela de referência;

$g_p(i,j)$ é o valor numérico da imagem (função de tom de cinza) na posição (i,j) no sistema da janela de pesquisa; e

$C_n(a,b)$ é a função n-ésima para o par (a,b), estando este limitado pelo seguinte intervalo:

$$0<a<p_x - r_x + 1 \tag{3.3}$$

$$0<b<p_y - r_y + 1 \tag{3.4}$$

Função Quociente

A função quociente é representada matematicamente por:

$$C_1(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} \frac{g_p(i+a, j+b)}{g_r(i, j)} \tag{3.5}$$

No caso de perfeita correlação, a função assume o valor 1. Se a função não assumir este valor o melhor *matching* será dado na posição (a,b) que tiver o valor mais próximo da unidade.

Função Quociente Modificada (Modified Quotient Function)

Esta função pode ser representada por:

$$C_2(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} \frac{\max[g_r(i,j), g_p(i+a, j+b)]}{\min[g_r(i,j), g_p(i+a, j+b)]} \tag{3.6}$$

O menor valor que a função assumir será a posição de melhor *matching*.

Função Erro

A função erro é expressa por:

$$C_3(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_r(i,j)\text{-}g_p(i+a,j+b) \tag{3.7}$$

A posição (a,b) em que C_3 assumir o menor valor será a de melhor *matching*. Se a função assumir o valor 0, a correlação é dita perfeita.

Função Erro Quadrático

Esta função é definida por:

$$C_4(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} [g_r(i,j)\text{-}g_p(i+a,j+b)]^2 \tag{3.8}$$

Como na Função Erro, a posição (a,b) de melhor *matching* é encontrada quando C_4 assumir o menor valor.

A principal vantagem desta função de correlação, em relação a função erro, reside no fato desta amplificar mais os pequenos erros, fornecendo, consequentemente, com melhor exatidão a posição de melhor correlação (STRAUCH, 1991).

Função Covariância Cruzada (Cross-Correlation Function)

Esta função é definida por:

$$C_5(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} [g_r(i,j) - \overline{g_r}].[g_p(i+a,j+b) - \overline{g_{p_{a,b}}}] \qquad (3.9)$$

onde:

$$\overline{g_r} = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_r(i,j) \qquad (3.10)$$

$$\overline{g_{p_{a,b}}} = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_p(i+a,j+b) \qquad (3.11)$$

A equação 3.10 representa a média dos tons de cinza da janela de referência.

A equação 3.11 representa a média dos tons de cinza das sub-janelas da janela de pesquisa.

A posição (a,b) onde C_5 assumir o maior valor, será a posição de melhor *matching*.

Função Covariância Cruzada Modificada

A Função Covariância Cruzada Modificada é expressa por:

$$C_6(a,b) = \frac{1}{r_x . r_y} \sum_{i=0}^{r_y-1} \sum_{j=0}^{r_x-1} g_r(i,j).g_p(i+a,j+b) \qquad (3.12)$$

Como na Função Covariância Cruzada, esta função fornece a posição de melhor *matching* quando C_6 for máximo.

Todas estas funções de correlação possuem suas vantagens e desvantagens. Alguns exemplos destas vantagens e desvantagens são apresentados em YANNIRIS (1974): por exemplo, a função covariância cruzada pode fornecer resultados errados para sinais monotônicos. A função erro não falha para esse tipo de sinal, mas em compensação, se houver uma forte diferença de brilho, a janela correlacionada pode ser falsa.

Coeficiente de Correlação

Todas as funções de correlação definidas acima possuem a desvantagem de serem sensitivas as mudanças na amplitude de $g_r(i,j)$ e $g_p(i+a,j+b)$. Um caminho para contornar esta dificuldade está em realizar o *matching* via o coeficiente de correlação (GONZALEZ & WOODS, 1993).

Quando as janelas de referência e de pesquisa são tratadas como vetores, o coeficiente de correlação pode ser visto como o cosseno do ângulo α formado entre dois vetores:

$$\cos(\alpha) = \frac{g_r \cdot g_p}{|g_r| \cdot |g_p|} = \frac{g_r \cdot g_p}{\sqrt{g_r \cdot g_r}\sqrt{g_p \cdot g_p}} \tag{3.13}$$

O coeficiente de correlação pode ser calculado utilizando todas as funções de correlação citadas. Para a função covariância cruzada definida em 3.9, o coeficiente de correlação é dado por:

$$CC_5(a,b) = \frac{\sum_{i=0}^{ry-1}\sum_{j=0}^{rx-1}[g_r(i,j) - \overline{g_r}].[g_p(i+a,j+b) - \overline{g_{p_{a,b}}}]}{\sqrt{\sum_{i=0}^{ry-1}\sum_{j=0}^{rx-1}\left[g_r(i,j) - \overline{g_r}\right]^2 . \sum_{i=0}^{ry-1}\sum_{j=0}^{rx-1}\left[g_p(i+a,j+b) - \overline{g_{p_{a,b}}}\right]^2}} \tag{3.14}$$

onde:

$CC_5(a,b)$ é o coeficiente de correlação da função covariância cruzada para o par (a,b).

Para as demais funções de correlação o procedimento para obter o coeficiente de correlação é análogo à expressão 3.14.

O coeficiente de correlação fornece a posição de melhor *matching* quando for igual a unidade. O coeficiente de correlação tem a vantagem de assumir valores em um espaço normalizado entre -1 e 1, ao contrário das funções de correlação simplesmente, e ser invariante a escalas radiométricas diferentes entre as duas funções que serão correlacionadas.

A utilização do coeficiente de correlação é o método mais utilizado para estabelecer a correspondência nos algoritmos mais comuns de *matching* baseado em áreas.

3.2 *Matching* baseado em feições

O *matching* baseado em feições utiliza feições simbólicas extraídas a partir dos níveis de cinza das imagens, ao invés destes níveis de cinza diretamente, para estabelecer a correspondência. Como as feições extraídas das imagens são menos sensíveis à distorções radiométricas e geométricas, é de se esperar que o *matching* realizado com tais primitivas seja mais consistente.

Visando aplicações de visão estéreo, feição pode ser definida segundo KREMER (1997) como uma propriedade associada a um *pixel* e seus vizinhos em uma imagem capturada que não varia significantemente se o ângulo de visualização da câmera é mudado.

Cada feição extraída em cada imagem gera um vetor de atributos, sendo que o *matching* é realizado neste espaço de atributos através de alguma medida de similaridade. Estes atributos são representações de mais alto nível das informações de uma cena.

Uma medida de similaridade que pode ser empregada é, por exemplo o próprio coeficiente de correlação citado em 3.1.1, porém, calculado no espaço de atributos, que representam as feições detectadas no par de imagens estereoscópicas e não mais entre áreas. Portanto o *matching* baseado em feições pode ser realizado em dois passos principais que são:

1) Extração das feições de interesse em ambas as imagens.

2) Estabelecimento da correspondência das feições entre as duas imagens através de algum critério de similaridade.

Embora simples, esta metodologia consegue bons resultados podendo alcançar precisão ao nível de *sub-pixel*, dependendo da maneira que as feições forem extraídas.

A principal vantagem desta classe de *matching* está em que simples comparações entre as feições candidatas a feições homólogas podem ser feitas, tornando o estabelecimento da correspondência mais rápido de que o *matching* baseado em áreas.

A grande desvantagem desta classe de *matching* está em que a correspondência é estabelecida apenas para poucos pontos do par de imagens, possibilitando determinar pontos 3-D apenas esparsos, ao contrário do *matching* baseado em área, que possibilita determinar um denso *grid* tridimensional.

Existem diversas técnicas de extração de feições que de forma geral são abordadas em uma grande área do processamento de imagens denominado segmentação de imagens. A extração de feições pode ser uma tarefa desde trivial até uma tarefa de extrema complexidade, dependendo principalmente do tipo de feições que são de interesse.

A tarefa de extração de feições visando o processo estéreo deve ser realizada com um conhecimento a priori de quais feições serão de interesse para estabelecer a correspondência com sucesso, assim é de se esperar que aplicações de inspeção industrial devam ter suas feições segmentadas nas imagens com critérios diferentes de feições segmentadas em imagens médicas, por exemplo. De forma mais técnica, para este exemplo citado, a extração de feições retas para aplicações de inspeção industrial seria um critério muito mais coerente a ser adotado para extração das feições de que este mesmo critério em aplicações médicas.

3.2.1 Caracterização de feições

As feições mais utilizadas como primitivas de *matching* são bordas e segmentos de bordas (pontos de borda conectados) que podem ser determinados com precisão *sub-pixel*. Outra maneira de alcançar precisão *sub-pixel* é determinar representações matemáticas das feições, como por exemplo, através da transformada de Hough ou com ajuste de polinômios. Outra possibilidade é determinar os momentos de uma região obtida, por exemplo, através de técnicas de crescimento de regiões.

3.2.2 Bordas

Quando não é conhecida a natureza das feições de interesse, as variações abruptas nas intensidades podem ser utilizadas como feições. Estas feições são conhecidas como bordas de intensidade e a grande maioria dos algoritmos estéreos baseado em feições utilizam-se destas para o processo de *matching*, possibilitando a sua aplicação em cenas mais gerais (algoritmos de aplicações gerais).

Existem vários detetores de bordas, como por exemplo os operadores de Sobel, Roberts e Prewitt (GONZALEZ & WOODS, 1993), porém em algoritmos estéreos o detetor de bordas mais utilizado é o Laplaciano da Gaussiana.

O Laplaciano de uma função bidimensional f(x,y) é a soma da derivada segunda de f(x,y) em relação a x e em relação a y como:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.15}$$

onde:

$\nabla^2 f$ é o gradiente de f(x,y).

A função Gaussiana é dada por:

$$h(x,y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{3.16}$$

onde:

h(x,y) é a função Gaussiana; e

σ é o desvio padrão.

Fazendo:

$$r^2 = x^2 + y^2 \tag{3.17}$$

o Laplaciano da Gaussiana pode ser expresso portanto pela derivada segunda de h em relação a r, ficando:

$$\nabla^2 h = \left(\frac{r^2 - \sigma^2}{\sigma^4}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{3.18}$$

onde:

$\nabla^2 h$ é o Laplaciano da Gaussiana.

O gráfico desta função assume a forma de um "chapéu mexicano", sendo que ele corta o eixo x (*zero crossings*) em $r = \pm\sigma$.

O Laplaciano da Gaussiana possui algumas vantagens sobre os demais operados de detecção de bordas que são:

-Os *zero crossings* são ótimo estimadores da localização das bordas;

-É um operador robusto, mesmo na presença de bordas borradas e grandes quantidades de ruídos presente nas imagens.

A figura 3.3 ilustra uma imagem unidimensional com 256 níveis de cinza (8 *bits*) e as localizações de seus *zero crossings* representados pelos sinais "+" e "-" indicando a transição entre valores positivos e negativos, respectivamente.

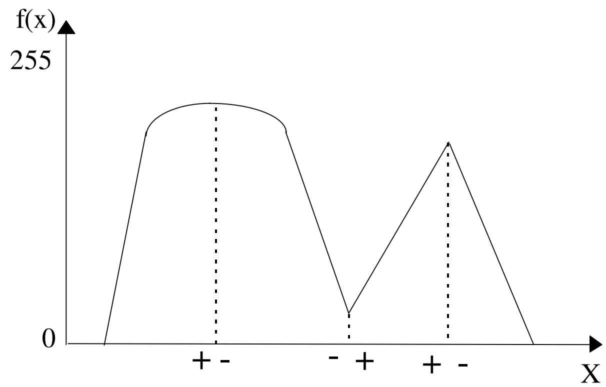


Fig. 3.3 - Z*ero crossings* em uma imagem unidimensional.

Ainda, de acordo com a figura 3.3, pode-se notar que além das localizações das bordas, pode-se associar a cada borda o seu sinal (positivo ou negativo), ou seja, a polaridade da borda, diminuindo sensivelmente a ambigüidade entre as bordas no par de imagens estereoscópicas, que implica que bordas positivas em uma imagem só deveram se correlacionar com bordas positivas na outra imagem. O mesmo é válido para bordas negativas. Existem evidências biológicas que o *matching* em sistema de visão biológicos respeitam a polaridade das bordas.

3.2.3 Feições retas

A utilização de feições retas em algoritmos de *matching* tem despertado grande interesse em pesquisas sobre visão estéreo devido às várias vantagens sobre os demais tipos de feições. Segundo TOMMASELLI & TOZZI (1993) as principais vantagens deste tipo de feição são:

-Abundância destas em imagens de ambientes modificados pelo homem e de produtos industrializados;

-Facilidade de detecção em imagens digitais;

-Possibilidade de obtenção de precisão *sub-pixel*; e

-Menor probabilidade de erros grosseiros no processo de *matching*.

AYACHE (1993) cita outras vantagens deste tipo de feição para o estabelecimento da correspondência estéreo que são:

-Reduz a complexidade por reduzir o número de possíveis *matches*, já que sempre existe menos feições retas do que pontos em uma imagem;

-Atributos geométricos extraídos das feições retas são ricos em informações e portanto mais discriminantes no processo de *matching*; e

-A posição e orientação de uma curva são geralmente obtidos com maior precisão que a posição de um ponto isolado, consequentemente a reconstrução tridimensional é mais precisa.

Além dessas vantagens citadas, existe a possibilidade de associar a uma feição reta a sua polaridade, a intensidade média, tamanho, etc, aumentando assim, o espaço de atributos destas feições e consequentemente melhorando a performance do *matching*.

Dentre as técnicas mais comum para se detectar feições retas em uma imagem pode-se citar a transformada de Hough (HOUGH, 1962), que será tratada em 3.2.4. Outro técnica possível para esta tarefa é um método semelhante a transformada de Hough denominado agrupamento $\theta-\rho$ descrito em TOMMASELLI & TOZZI (1993); e NOGUEIRA & TOMMASELLI (1995).

3.2.4 Transformada de Hough

A transformada de Hough (HOUGH, 1962) é geralmente utilizada para detectar linhas retas, porém esta pode ser utilizada para detectar qualquer curva que possa ser representada na forma paramétrica, como por exemplo, circunferências e elipses.

Para linhas retas, a transformada de Hough consiste em determinar um espaço de parâmetros ou espaço de Hough no qual são selecionadas os parâmetros das retas de maior ocorrência. Este espaço de parâmetros é também conhecido como células de acumulação.

Dado um ponto i de coordenada (x_i, y_i) e a equação reduzida da reta de acordo com a expressão 3.19, como:

$$y_i = ax_i + b \qquad (3.19)$$

onde:

a e b são os coeficientes angular e linear da reta.

Existem infinitas retas que passam pelo ponto (x_i, y_i), ou formalmente, existem infinitos valores para o par de parâmetros (a,b) dado o ponto i que satisfazem esta equação. Entretanto, escrevendo a expressão 3.19 como:

$$b = -x_i a + y_i \qquad (3.20)$$

e considerando o espaço de parâmetros (a,b), produz-se a equação de uma única reta. Considerando agora um ponto j de coordenadas (x_j, y_j), existe também no espaço de parâmetros (a,b) uma reta associada a este ponto.

Supondo que i e j são dois pontos que pertencem a uma mesma reta presente em uma imagem e dado um valor de a é possível calcular b, sendo que este só será igual para os dois pontos quando a for coerente com a reta a que estes pontos pertencem, já que i e j são distintos. Assim atribuindo valores de a é possível calcular valores de b e assinalar com um "voto" a ocorrência daquele par (a,b) no espaço de parâmetros ou células de acumulação para cada ponto n de coordenada (x_n, y_n).

As seguintes figuras ilustram uma reta no espaço xy que está representada pela interseção de duas retas no espaço de parâmetros (a,b).

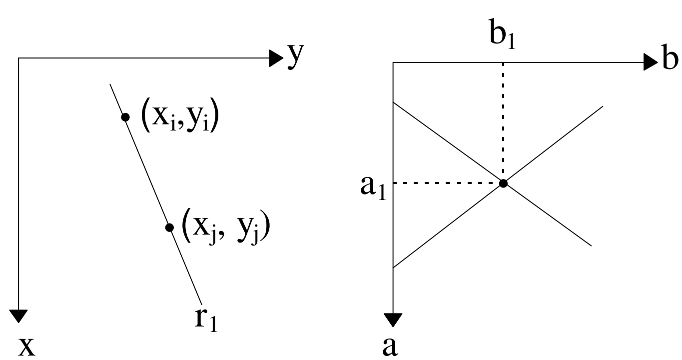


Fig. 3.4 - Espaço xy e espaço ab.

De acordo com a figura 3.4, é possível perceber que se houvesse outro ponto qualquer sobre r_1 no espaço xy, este seria representado por uma reta no espaço de parâmetros (a,b) que interceptaria as outras duas retas também no ponto de coordenada (a_1,b_1) no espaço de parâmetros (a,b). Desta forma o ponto de coordenadas (a_1, b_1) receberia o número de "votos" igual ao número de pontos da reta r_1 no espaço xy e as demais coordenadas (a,b) receberiam apenas um ou nenhum voto, possibilitando assim, segmentar facilmente as retas no espaço de parâmetros.

Existe um problema em utilizar a equação da reta na forma descrita na expressão 3.20, porque os parâmetros a e b tendem ao infinito, quando a reta no espaço xy se aproxima da vertical. Uma maneira de contornar este problema está em escrever a equação da reta na sua forma paramétrica como:

$$x.\cos(\theta) + y.\sin(\theta) = \rho \qquad (3.21)$$

onde:

θ e ρ são a inclinação e a menor distância da reta a origem do sistema, respectivamente.

As seguintes figuras ilustram o espaço xy e o espaço θ-ρ.

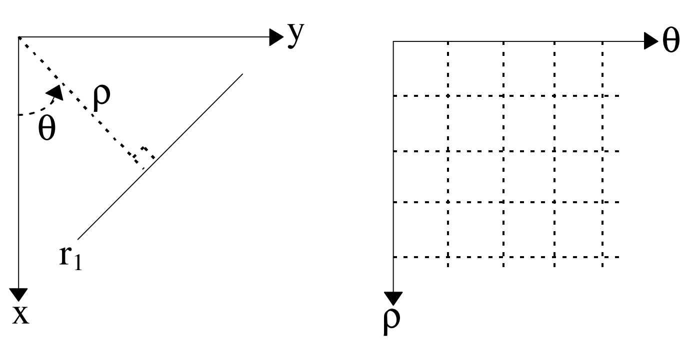


Fig. 3.5 - Espaço xy e espaço $\theta-\rho$ discretizado.

A vantagem da representação da reta na sua forma paramétrica é que os intervalos de θ e de ρ são limitados. O intervalo de θ varia de -90^0 a $+90^0$ e o intervalo de ρ varia de $-\sqrt{W^2+H^2}$ a $\sqrt{W^2+H^2}$, sendo que W é o comprimento e H é a largura da imagem.

Apesar de simples, a transformada de Hough tem um custo computacional relativamente alto. A obtenção de bons resultados exige que o espaço de parâmetros (θ,ρ) tenha uma variação pequena entre as células de acumulação (aumento da discretização), aumentando assim, as dimensões dessa matriz.

Portanto a detecção de feições retas através da transformada de Hough pode ser realizada pela seguinte seqüência:

1) Detectar bordas nas imagens (espaço xy);

2) Aplicar a transformada de Hough para os pontos de borda;

3) Selecionar no espaço de Hough (espaço $\theta\rho$) as retas de maior ocorrência através, por exemplo, de um simples *threshold*; e

4) Verificar no espaço xy, de posse dos parâmetros das retas selecionadas, as coordenadas iniciais e finais das retas, bem como conectar ou não falhas nestas.

As referências BALLARD & BROWN (1982); SCHALKOFF (1989); COSTA (1993); GONZALEZ & WOODS (1993); JOO *et al.* (1993); e YUEN & MA (1997) são sugeridas para maiores informações sobre este assunto.

3.2.5 Regiões

Regiões segmentadas no par de imagens estereoscópicas também podem ser utilizadas como feições visando estabelecer correspondência entre estas. Existem diversas técnicas para representação de regiões em imagens digitais, porém quase todas as técnicas existentes utilizam um princípio básico denominado agregação de *pixels*. Esta agregação de *pixels* é realizada sobre algum predicado lógico que decide se um dado *pixel* deve ou não ser agregado a uma dada região.

Em regiões das imagens onde a intensidade comporta-se de forma relativamente homogênea, o processo de *matching* baseado em áreas falha justamente devido a esta homogeneidade, já se estas regiões homogêneas forem detectadas e segmentadas é possível associar um vetor de atributos a estas e realizar o *matching* neste espaço.

Dentre as principais técnicas de segmentação de regiões em imagens digitais pode-se citar o crescimento de regiões por agregação de *pixels* (BALLARD & BROWN, 1982; SCHALKOFF, 1989; GONZALEZ & WOODS, 1993) e *watershed* (MEYER & BEUCHER, 1990; LOTUFO & TRETTEL, 1996; FACON, 1996). Estas técnicas são tratadas nas referências citadas e não serão detalhadas neste trabalho. O intuito deste item está em mostrar como gerar atributos uma vez segmentadas as regiões.

3.2.6 Simples atributos de regiões

Os atributos mais simples de uma região podem ser expressos pela a média, mediana e mínimo e máximo valores das intensidades dos *pixels* pertencentes a uma região. Estes atributos podem ser utilizados no processo de *matching* estéreo, porém são muito sensíveis aos problemas da visão estéreo citado em 3.

A área e o perímetro de uma região podem ser utilizados como atributos associados a uma região, porém estes tipos de atributos são utilizados em aplicações em que o tamanho de uma região é relevante. Devido a geometria estéreo, um par de regiões ditas homólogas não terão o mesmo tamanho, consequentemente a mesma área e o mesmo perímetro. Para contornar este problema é possível determinar a compacticidade (*compactness*) de uma região através da expressão 3.22 como:

$$c = p^2 / a \tag{3.22}$$

onde:

c é a compacticidade;

p é o perímetro; e

a é a área.

Este tipo de atributo de uma região é invariante a translação, escala e rotação, sendo assim, um atributo não invariante a todas as transformações geométricas ocorridas entre o par de imagens, como citado em 3., mas de certa forma, interessante para o processo estéreo, principalmente em situações de fraca perspectiva (como imagens áreas, por exemplo) e por ter baixo custo computacional.

Outro atributo que pode ser associado a uma região é sua direção e seu espalhamento principal, descritos pelo maior autovalor e seu correspondente autovetor, obtidos a partir da matriz covariância, gerada utilizando os valores de intensidade dos *pixels* contidos em uma região como variáveis randômicas. Estes atributos são invariantes a rotação, mas não à mudanças de escala. Para minimizar este problema, pode ser utilizado a razão entre o maior e o menor autovalor como um atributo para o processo estéreo.

3.2.7 Atributos topológicos de regiões

Atributos topológicos são quase que ideais para o processo estéreo. Segundo uma simples definição de GONZALEZ & WOODS (1993), topologia é o estudo de propriedades de uma figura que não são afetados por qualquer deformação, uma vez que não haja rompimento ou união de figuras.

Estes atributos são quase ideais para o processo estéreo e só não são ideais porque nestes existem a possibilidade de rompimento e união de regiões em uma imagem e na outra não. Este fato pode ocorrer nos casos de oclusão e de superfícies transparentes, como citado em 3..

Os principais atributos topológicos para regiões gerais (não importando a forma) são o número de "buracos" e componentes conectados em uma região. Um componente conectado de um conjunto é um subconjunto de tamanho máximo tal que qualquer dois de seus pontos possam ser unidos por uma curva conectada que circunda inteiramente o subconjunto. O número de Euler trata estes dois atributos como:

$$E = C - H \tag{3.23}$$

onde:

E é o número de Euler;

C é o número de componentes conectados; e

H é o número de "buracos".

Em regiões representadas por linhas retas, o número de Euler é dado por:

$$W - Q + F = C - H = E \tag{3.24}$$

onde:

W é o número de vértices;

Q é o número de bordas; e

F é o número de faces.

O grande problema em aplicar essa classe de atributos é justamente determinar as regiões coerentemente.

3.2.8 Momentos invariantes de regiões

A forma de uma região pode ser descrita quantitativamente utilizando-se momentos.

Momentos (HU, 1962) são muito utilizados em aplicações de estatística e física, por possuírem interpretação matemática. Segundo GONZALEZ & WOODS (1993), dado uma função unidimensional g(r) de uma variável arbitrária r, e tratando a amplitude de g como uma variável randômica v, pode-se formar um histograma cuja amplitude é $p(v_i)$, para i = 1,2,...k, onde k é o número de incrementos discretos da amplitude, o n-ésimo momento υ de v em torno de sua média pode ser dado por:

$$\upsilon_n(v) = \sum_{i=1}^{k} (v_i - m)^n p(v_i) \tag{3.25}$$

onde:

$$m = \sum_{i=1}^{K} v_i p(v_i) \tag{3.26}$$

Se g(r) for normalizado para uma unidade de área tem-se que $\upsilon_0 = 1$ $\upsilon_1 = 0$. O segundo momento pode ser interpretado como a variância, o terceiro momento é uma medida da simetria e o quarto momento é uma medida do achatamento relativo desse gráfico. Os demais momentos, apesar de possuírem significado matemático, são de difícil interpretação de sua natureza.

Para uma função bidimensional f(x,y), o momento m de ordem (p+q) é definido para o caso discreto como:

$$m_{pq} = \sum_{x}\sum_{y} x^p y^q f(x,y) \tag{3.27}$$

onde:

p, q = 1,2,....

Estes momentos não são invariantes a translação, para solucionar este problema determina-se os momentos centrais μ como:

$$\mu_{pq} = \sum_{x}\sum_{y} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{3.28}$$

onde:

$$\bar{x} = \frac{m_{10}}{m_{00}} \tag{3.29}$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \tag{3.30}$$

Para tornar estes momentos centrais μ invariantes a escala, determina-se os momentos centrais normalizados η dados por:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{3.31}$$

onde:

$$\gamma = \frac{p+q}{2} + 1 \tag{3.32}$$

Ainda assim, estes momentos centrais normalizados não são invariantes a rotações e para contornar este problema, determina-se um conjunto de sete momentos denominados de momentos invariantes ϕ que são dados por:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{3.33}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{3.34}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{3.35}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{3.36}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{3.37}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{3.38}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})\end{aligned} \tag{3.39}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{3.40}$$

Este conjunto de sete momentos são invariantes a translação, rotação e escala.

O principal problema associado aos momentos invariantes está no aumento de problemas numéricos com o aumento da ordem destes.

As referências FLUSSER & SUK (1994); e COSTA (1996) trazem aplicações de momentos invariantes.

3.3 Premissas

Embora existam inúmeras técnicas para realizar o estabelecimento da correspondência, a grande maioria utilizam propriedades locais presentes no par de imagens estereoscópicas para tal tarefa, devido justamente à natureza do problema ser essencialmente local.

As principais premissas utilizadas em algoritmos de *matching* estéreo podem ser divididas em duas categorias: premissas fundamentadas em propriedades geométricas e premissas empíricas da visão estéreo, que na maioria das vezes podem ocorrer, porém, não são válidas em todas as situações possíveis e dependem basicamente da estrutura 3-D da cena.

As premissas fundamentadas em propriedades geométricas são as já citadas geometria epipolar (em 2.6) e restrição do intervalo de correspondência (em 2.7.1). A primeira diz que dado um ponto em uma imagem, o seu homólogo está obrigatoriamente sobre a linha epipolar na outra imagem, tornando o *matching* um problema unidimensional. A segunda implica que para configurações de câmeras ditas ideais, a disparidade é obrigatoriamente nula ou positiva, restringindo assim o espaço de busca do ponto homólogo ao longo da linha epipolar. As premissas tratadas em 3.3.1, 3.3.2 e 3.3.4 são empíricas.

3.3.1 Ordem

Esta premissa empírica da visão estéreo assume que a ordem de pontos homólogos ao longo das linhas epipolares homólogas não é alterada. Apesar de esta premissa seja verdadeira para a grande maioria dos pontos das imagens estereoscópicas, ela falha em regiões onde a disparidade varia bruscamente. A figura 3.6 ilustra uma possível situação na qual esta premissa falha.

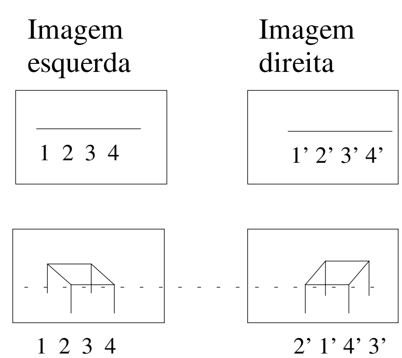


Fig. 3.6 - Premissa de Ordem.

Na figura 3.6, o par de imagens acima mostra uma seqüência de pontos na imagem da esquerda e a mesma seqüência de pontos na imagem da direita ao longo das linhas epipolares homólogas. No par de imagens abaixo aparece uma situação de uma cena na qual esta seqüência não foi preservada.

3.3.2 Continuidade

A premissa empírica da visão estéreo de continuidade implica em que a disparidade varia suavemente, exceto em limites de objetos.

Uma dificuldade em implementar esta premissa é justamente determinar os limites dos objetos em uma cena. Outra dificuldade está em assumir que a disparidade varia suavemente sobre objetos, que não necessariamente é verdade.

3.3.3 Unicidade

Esta premissa empírica da visão estéreo consiste em que um ponto em uma imagem só poderá ter um único ponto homólogo correspondente na outra imagem.

Dado um ponto P e outro Q distintos no espaço-objeto e, por exemplo, o CP_E (centro perspectivo da câmera da esquerda) serem colineares, o princípio da unicidade falha, pois estes dois pontos no espaço-objeto serão projetados sobre apenas um ponto p_{e1} no espaço-imagem da câmera da esquerda e dois pontos p_{d1} e p_{d2} na câmera da direita. Porém esta situação só pode ocorrer em dois casos:

O primeiro caso ocorre se um objeto da cena tiver superfície transparente, por exemplo se o ponto P na figura 3.7, for transparente.

A segunda situação, muito mais comum, ocorre mesmo quando P e Q são opacos, mas devido ao chamado "problema de oclusão" esta situação pode ocorrer, ou seja, a câmera da esquerda não pode "enxergar" o ponto Q devido a este estar ocluído pelo ponto P, mas a câmera da direita, por estar em posição distinta da câmera da esquerda, pode "enxergar" o ponto P e o ponto Q.

A figura 3.7 mostra estas situações em que a premissa de unicidade falha.

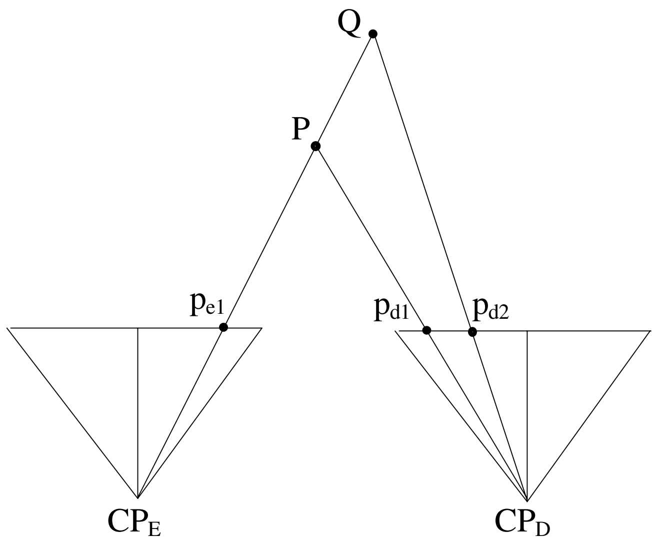


Fig. 3.7 - Situação na qual falha a premissa de Unicidade.

3.4 Evidências biológicas da *stereopis*

Existem evidências biológicas que permitem realizar algumas deduções e sugestões a respeito da *stereopis* (processo estéreo em seres vivos). As principais evidências biológicas existentes são:

Compatibilidade

Esta evidência biológica está relacionada com a polaridade de bordas de intensidade, impondo que uma borda em uma imagem cuja polaridade é positiva somente deverá se corresponder com bordas na outra imagem cujas polaridades também são positivas. Para bordas negativas, o raciocínio é análogo.

Seletividade direcional

A exemplo da evidência de compatibilidade, esta evidência também está relacionada a bordas de intensidade, sugerindo que a diferença entre as direções de duas bordas ditas homólogas não ultrapassem um certo limite angular. Um valor muito utilizado para este limite é de 30^0, embora este valor não seja comprovado cientificamente.

Processos cooperativos

Esta evidência é baseada em que fortes *matches* vizinhos devam influenciar positivamente fracos *matches* e fracos *matches* vizinhos devam inibir fortes *matches.*

Área fusional de Panum

A área fusional de Panum é uma área de tamanho fixo que delimita o intervalo permitido de busca do ponto homólogo. Assim, dado um ponto em uma imagem, o seu ponto homólogo deverá encontrar-se na outra imagem em uma posição que esteja contida neste intervalo. No caso do ponto homólogo não estar em uma posição contida neste intervalo, a correspondência (e portanto a fusão dos pontos homólogos), não é possível. OGLE (1950) encontrou que a fusão em seres humanos é obtida entre um intervalo ± 7' de arco próximo à fóvea (parte central da retina onde a visão é mais desenvolvida), assim o comprimento total da área fusional de Panum é de 14' de arco.

Gradiente de disparidade

O gradiente de disparidade pode ser determinado para objetos que ocorrem próximos no campo visual como a taxa de variação da disparidade entre esses objetos. Segundo BURT & JULESZ (1980), a fusão em seres humanos de ao menos um objeto falha quando este gradiente de disparidade excede um valor crítico aproximadamente igual a unidade. Este assunto será tratado mais detalhadamente em 3.4.1.

Movimento de vergência dos olhos

O movimento de vergência dos olhos pode ser entendido como o grau de estrabismo que os olhos assumem para concentrar atenção em uma região da cena. A exemplo do gradiente de disparidade, este assunto será tratado mais detalhadamente em 3.4.2.

Estas evidências biológicas citadas são oriundas de pesquisas e testes em seres vivos e ainda são contestadas por vários pesquisadores da área de biologia e visão artificial. As referências MARR (1980); e BURT & JULESZ (1980), tratam estas evidências mais detalhadamente e são fortemente recomendadas para maiores informações.

3.4.1 Gradiente de disparidade

O gradiente de disparidade é uma evidência biológica bastante simples de ser vivenciada. Colocando-se, por exemplo, o dedo indicador da mão esquerda a aproximadamente 20cm a frente dos olhos e o dedo indicador da mão direita a aproximadamente 40cm a frente dos olhos, não será possível enxergar os dois dedos perfeitamente fundidos num mesmo instante. Se a visão é concentrada no dedo esquerdo, o dedo direito aparecerá duplo devido ao fenômeno de diplopia[1] e se o inverso for realizado, ou seja, concentrar atenção no dedo direito, o dedo esquerdo aparecerá duplo. Porém, colocando-se os dois dedos próximos um ao outro, mesmo que ainda separados por uma pequena diferença entre suas profundidades em relação aos olhos, estes poderão ser fundidos e o fenômeno de diplopia desaparece.

Através desta simples experiência é possível deduzir que o sistema visual humano se adapta localmente para um intervalo de disparidade na qual esta pode variar e que ainda assim a fusão será realizada. Quando este intervalo, determinado localmente, excede o limite determinado por BURT & JULESZ (1980), a fusão não é mais possível e o fenômeno de diplopia ocorre.

O conceito de gradiente de disparidade pode ocasionar momentaneamente alguma confusão quando comparado ao conceito da área fusional de Panum, pois ambos sugerem um intervalo permitido no qual a fusão dos pontos homólogos é possível, porém, o intervalo sugerido através do conceito da área fusional de Panum é fixo para todo o campo visual, enquanto o intervalo sugerido através do conceito de gradiente de disparidade é variável dentro do campo visual. Portanto, o gradiente de disparidade também determina uma área fusional, embora seja diferente da área fusional de Panum.

Segundo AYACHE (1991), o gradiente de disparidade pode ser definido como a derivada da função disparidade δ em relação a profundidade ρ. A função δ é expressa por:

$$\delta = f(\rho, x, y) \tag{3.41}$$

e a sua derivada é expressa por:

[1] Fenômeno de enxergar duplo um mesmo objeto.

$$\gamma(\rho, x, y) = \frac{\partial \delta}{\partial \rho}(\rho, x, y) \tag{3.42}$$

onde:

δ é a disparidade;

ρ é a profundidade; e

γ é o gradiente de disparidade.

Segundo BURT & JULESZ (1980), o gradiente de disparidade é definido entre dois objetos, que ocorrem próximos no campo visual, como a diferença entre suas disparidades dividida por sua separação angular ou por sua separação no espaço ciclopeano.

O espaço ciclopeano recebe esta denominação porque o substantivo ciclope é o termo da mitologia utilizado para dar nome a um gigante com um só olho na testa.

O espaço ciclopeano pode ser definido matematicamente tal que a sua origem $OC(x_{OC}, y_{OC}, z_{OC})$ é dada por:

$$x_{OC} = \frac{CP_{XE} + CP_{XD}}{2} \tag{3.44}$$

$$y_{OC} = CP_{YE} = CP_{YD} \tag{3.45}$$

$$z_{OC} = CP_{ZE} = CP_{ZD} \tag{3.46}$$

onde:

CP_{XE}, CP_{YE} e CP_{ZE} são coordenadas do centro perspectivo da câmera da esquerda; e

CP_{XD}, CP_{YD} e CP_{ZD} são coordenadas do centro perspectivo da câmera da direita.

Uma vez definido o espaço ciclopeano, é possível transformar as coordenadas dos pontos homólogos para o espaço ciclopeano e vice-versa.

A transformação do espaço das imagens estereoscópicas para o espaço ciclopeano é dado por:

$$x_C = \frac{x_E + x_D}{2} \tag{3.47}$$

$$y_C = y_E = y_D \tag{3.48}$$

$$d(x_C, y_C) = x_E - x_D \tag{3.49}$$

onde:

x_E e y_E são coordenadas de um ponto na imagem da esquerda;

x_D e y_D são coordenadas de um ponto na imagem da direita; e

x_C, y_C e $d(x_C,y_C)$ são coordenadas de um ponto no espaço ciclopeano.

A transformação do espaço ciclopeano para o espaço das imagens estereoscópicas é dado por:

$$x_E = x_C + \frac{d(x_C, y_C)}{2} \tag{3.50}$$

$$y_E = y_C \tag{3.51}$$

$$x_D = x_C - \frac{d(x_C, y_C)}{2} \tag{3.52}$$

$$y_D = y_C \tag{3.53}$$

onde:

x_E, y_E, x_D, y_D, x_C, y_C, $d(x_C,y_C)$ como definido para as expressões 3.47, 3.48 e 3.49.

Considerando que as câmeras estão em uma configuração ideal e dado $A_E(x_{AE},y_A)$ e $B_E(x_{BE},y_B)$ serem pontos na imagem da esquerda e $A_D(x_{AD},y_A)$ e $B_D(x_{BD},y_B)$ serem pontos na imagem da direita e estes por sua vez, serem projeções de dois pontos A_P e B_P no sistema referencial do espaço-objeto, respectivamente, então a disparidade d_A de A_P e d_B de B_P é dado por:

$$d_A = x_{AE} - x_{AD} \tag{3.54}$$

$$d_B = x_{BE} - x_{BD} \tag{3.55}$$

Dado A_E e A_D, e B_E e B_D transformados para o espaço ciclopeano ficando $A_C(x_{AC},y_A,d_A)$ e $B_C(x_{BC},y_B,d_B)$, respectivamente, o gradiente de disparidade γ é então dado por:

$$\gamma = \frac{d_A - d_B}{d(A_C B_C)} \tag{3.49}$$

onde:

$d(A_C B_C)$ é a separação entre A_C e B_C no espaço ciclopeano.

Um limite de gradiente de disparidade igual a unidade define uma região com uma forma de um cone no espaço ciclopeano para o ponto A_C e outra para o ponto B_C, possibilitando assim, reduzir o espaço de busca do ponto homólogo. A figura 3.8 mostra o espaço ciclopeano.

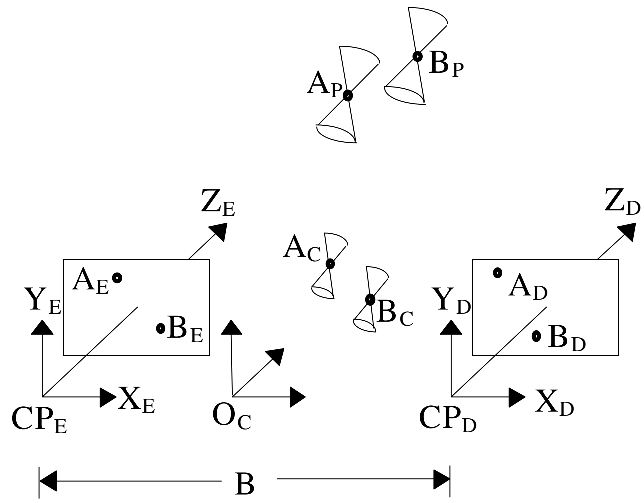


Fig. 3.8 - Espaço ciclopeano O_C.

As referências TRIVEDI & LLOYD (1985); LLOYD (1985) trazem mais informações sobre este assunto.

3.4.2 Movimento de vergência dos olhos

As imagens capturadas pelos olhos representam todo o campo visual, entretanto, a visão é concentrada, com maior intensidade, em apenas uma área pequena do campo visual. Nas demais áreas do campo visual, a visão é apenas superficial, do ponto de vista das informações extraídas destas áreas.

A área do campo visual onde a visão é mais concentrada para fins estereoscópicos é a área fusional (não necessariamente a de Panum), cujo seu centro (aproximadamente) é dado pela interseção dos eixos óticos dos olhos. Portanto, os olhos se movimentam, um em direção ao outro, para concentrar atenção em apenas uma região da cena. Este movimento é denominado de movimento de vergência dos olhos. A figura 3.9 ilustra este movimento para um ponto próximo e um ponto distante dos olhos.

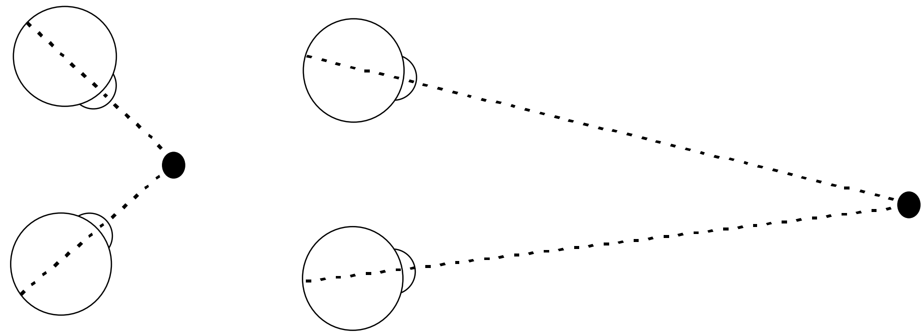


Fig. 3.9 - Movimentos de vergência dos olhos

O movimento de vergência dos olhos é o mecanismo que fornece grosseiramente um valor de disparidade entre áreas homólogas de um par de imagens estereoscópicas, porém, não se conhece ao certo qual o mecanismo que o sistema visual dos seres vivos utiliza para controlar este movimento.

Ainda a respeito sobre a experiência citada em 3.4.1, quando a fusão dos dois dedos simultaneamente não é possível, os olhos se movimentam (movimento de vergência dos olhos) e a visão é concentrada em apenas um dos dedos, determinando assim, um valor de disparidade aproximado para aquela região da cena e então um novo intervalo de disparidade é determinado em torno deste valor aproximado. Este fato sugere que o mecanismo de controle do movimento de vergência dos olhos seja controlado por parâmetros correspondentes a regiões da cena que ultrapassem o limite de gradiente de disparidade, indicando que olhos devam mover-se para concentrar atenção em outras áreas das imagens.

Os olhos podem mover-se com valores de ângulos diferentes. Por exemplo, se uma região de atenção da cena está em frente ao olho esquerdo, este permanecerá estático, enquanto o olho direito irá mover-se em direção ao olho esquerdo de um certo valor angular.

3.4.3 Dificuldades em implementar algoritmos inspirados nas evidências biológicas

As evidências biológicas citadas da *stereopis* são importantes informações do processo estéreo em seres vivos, porém, para descrever algumas destas evidências na forma de um algoritmo não é uma tarefa simples nem tão pouco direta.

As evidências biológicas de compatibilidade e seletividade direcional são relativamente simples de serem descritas na forma de um algoritmo, uma vez que qualquer tipo de filtro passa-alta, aplicado ao par de imagens estereoscópicas, poderá detectar bordas de intensidade, e então a estas, associar suas polaridades e direções.

A polaridade de uma borda de intensidade é obtida através do gradiente de intensidade (na direção x e na direção y) e a sua direção através da função arco-tangente do gradiente na direção y dividido pelo gradiente na direção x.

Outra evidência biológica relativamente simples de ser descrita na forma de um algoritmo é a área fusional de Panum, uma vez que, dado um ponto em uma imagem, está evidência biológica delimita um intervalo fixo permitido de busca sobre a linha epipolar do ponto homólogo na outra imagem.

Uma maneira de descrever a evidência biológica de processos cooperativos na forma de um algoritmo é utilizar técnicas de rotulação por relaxação (ROSENFELD *et al.*, 1976; HUMMEL & ZUCKER, 1983; PRICE, 1985; HAKEN, 1995; WU, 1995).

A rotulação por relaxação é razoavelmente um modelo geral proposto por ROSENFELD *et al.*, (1976) para rotulação de imagens. No processo estéreo, a relaxação consiste basicamente em identificar pontos ou feições de interesse em cada uma das imagem e a partir destes (ou destas), assinalar rótulos únicos para cada par de pontos ou feições ditas homólogas.

Para cada par de *matches*, uma probabilidade de *matching* é determinada iterativamente dependendo da probabilidade dos *matches* vizinhos, de tal maneira que fortes *matches* vizinhos aumentam a probabilidade de um fraco *match* e fracos *matches* vizinhos diminuem a probabilidade de um forte *match*. O item 3.5.4 traz um resumo do algoritmo cooperativo de MARR & POGGIO (1976) baseado em rotulação por relaxação. Outra referência sugerida sobre este assunto é BARNARD & THOMPSON, 1985.

A evidência biológica de gradiente de disparidade pode ser descrita na forma de um algoritmo utilizando as definições para o gradiente de disparidade citadas em 3.4.1, porém, a utilização destas definições, apesar de corretas, é inviável e contraditória (se algumas considerações não forem feitas).

Segundo a definição de AYACHE (1991) para o gradiente de disparidade (citada em 3.4.1), este é a derivada parcial da função disparidade em relação a profundidade. A contradição desta definição está justamente em calcular a derivada da função disparidade em relação a profundidade, uma vez que a função disparidade e consequentemente a profundidade, são as incógnitas deste problema, e portanto, ainda desconhecidas.

A segunda definição citada em 3.4.1 para o gradiente de disparidade, diz que este é determinado entre dois objetos próximos no campo visual como a diferença entre suas disparidades dividida por sua separação no espaço ciclopeano. O principal problema para implementar está definição é que o gradiente de disparidade seria determinado apenas entre objetos próximos, e portanto, a grande maioria da cena continuaria sem ter estes valores de gradiente determinados.

Outra implicação desta definição está em determinar quando dois objetos estão aceitavelmente próximos, ou seja, um conceito de proximidade deve ser definido. No caso de nenhum par de objetos estar relativamente próximo, o gradiente de disparidade não poderia ser determinado em nenhuma região da imagem no espaço ciclopeano. A figura 3.10 ilustra algumas direções para as quais seria possível calcular o gradiente de disparidade.

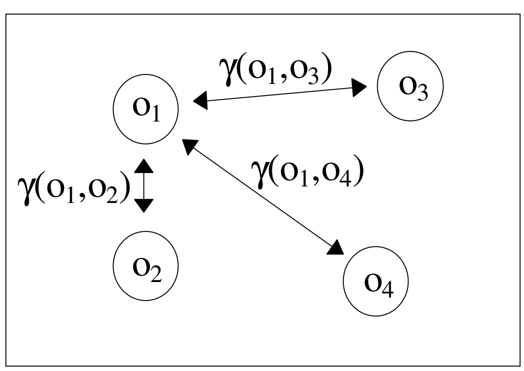


Fig. 3.10 - Dificuldades em implementar o gradiente de disparidade.

Na figura 3.10, o_1, o_2, o_3 e o_4 representam objetos no espaço ciclopeano, $\gamma(o_1,o_2)$, $\gamma(o_1,o_3)$ e $\gamma(o_1,o_4)$ representam as direções possíveis para calcular os gradientes de disparidade a partir de o_1 em relação a o_2, o_3 e o_4, respectivamente, desde que estes objetos atendessem o conceito de proximidade estipulado.

As dificuldades encontradas para descrever a evidência biológica de movimento de vergência dos olhos na forma de um algoritmo também são várias e complexas, mesmo porque, o mecanismo que o sistema visual dos seres vivos utiliza para controlar este movimento é ainda desconhecido.

A possível explicação que poderia ser dada sobre o mecanismo de controle do movimento de vergência dos olhos (citado em 3.4.2) também é contraditório para ser descrito na forma de um algoritmo, pois, a região para a qual o limite de gradiente de disparidade fosse excedido não poderia ser determinada devido a função de disparidade (e portanto o gradiente de disparidade) ser ainda desconhecida.

Uma possível solução para descrever o movimento de vergência dos olhos na forma de um algoritmo sugerida por MARR & POGGIO (1979) consiste em utilizar a correlação de grandes janelas no domínio espacial para controlar este movimento, uma vez que, grandes janelas tendem a possuir performances mais estáveis (via o teorema de limite central) do que pequenas janelas em processos de correlação.

O problema inerente a esta sugestão de MARR & POGGIO (1979) é devido a grandes janelas serem fortemente afetadas pelos problemas da visão estéreo citado em 3., principalmente em relação aos efeitos de oclusão e transformações geométrica ocorridas entre o par de imagens estereoscópicas. A figura 3.11 mostra um par de imagens estereoscópicas com um par de linhas epipolares homólogas destacadas em preto.

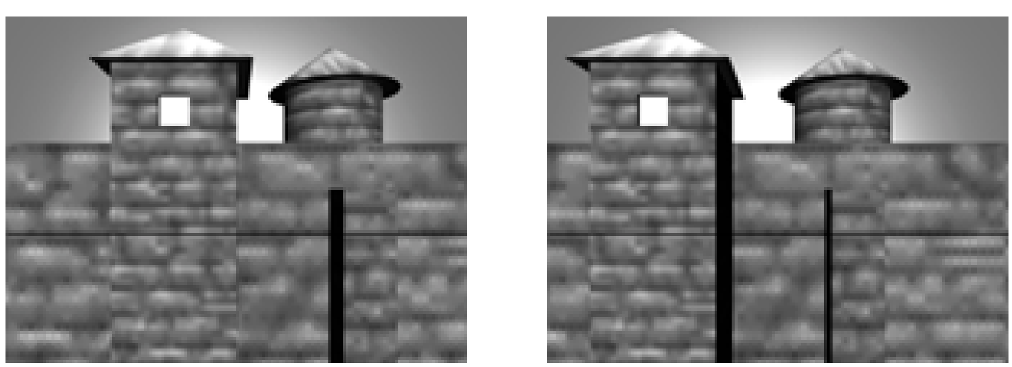


Fig. 3.11 - Par de imagens estereoscópicas.

A figura 3.12 mostra o gráfico das intensidades das linhas epipolares homólogas ao longo do eixo x das imagens.

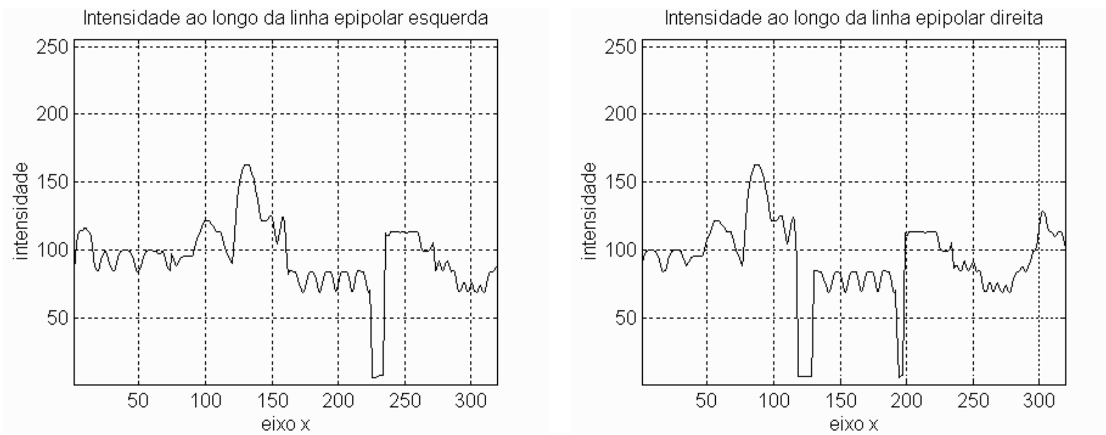


Fig. 3.12 - Gráfico de intensidades ao longo das linhas epipolares homólogas.

Na figura 3.12 é possível visualizar claramente os efeitos de oclusão e com mais dificuldade a variação de escala entre os dois gráficos. Estes efeitos podem resultar em falsos *matches* mesmo para grandes janelas.

3.5 Teorias e algoritmos sobre *stereopsis*

A teoria computacional da *stereopsis* é hoje a área mais estudada no processo estéreo, envolvendo outras técnicas e ciências como relaxação, otimização, programação dinâmica e inteligência artificial.

Este item foi extraído basicamente de DHOND & AGGARWAL (1989) e portanto, está referência é fortemente recomendada para maiores esclarecimentos.

3.5.1 Teoria de Marr & Poggio

A teoria de MARR & POGGIO (1979) propõe que o processador visual humano resolve o problema de correspondência em cinco passos principais que são:

1) As imagens da direita e da esquerda são filtradas em doze diferentes orientações de máscaras cada uma aproximada pela diferença de duas funções Gaussianas com desvios padrão constantes na razão de 1:1.75;

2) *Zero crossings* nas imagens filtradas são encontrados ao longo de linhas perpendiculares a orientação das máscaras;

3) Para cada tamanho de máscara, o *matching* toma lugar entre os segmentos de *zero crossings* extraídos a partir de cada imagem filtrada que são de mesmo sinal e aproximadamente a mesma orientação. Ambigüidades de *matchings* locais são resolvidas considerando o sinal das disparidades de *matches* não ambíguos próximos;

4) Grandes máscaras controlam o movimento de vergência; e

5) Os resultados da correspondência são armazenados em um *buffer* dinâmico chamado esboço 2.5-D.

MARR & POGGIO (1979) formularam duas regras básicas para o *matching*. A primeira regra consiste que cada ponto em uma imagem pode ser correspondido com apenas um ponto na outra imagem, e portanto, somente associar um valor de disparidade para um par de pontos homólogos, ou seja, unicidade citada em 3.3.3. A segunda regra consiste em que a matéria é coesa, portanto a disparidade varia suavemente, exceto em descontinuidades de profundidade que ocorrem em fronteiras de superfície, ou seja, continuidade, citada em 3.3.2.

3.5.2 Implementação da Teoria de Marr & Poggio por Grimson

GRIMSON (1980) implementou a teoria de MARR & POGGIO (1979) e revelou certos detalhes de sua implementação que não foram descritos anteriormente por MARR & POGGIO (1979). Ele mostrou que mudanças de intensidade ocorrendo em uma escala particular podem ser detectadas pela localização dos *zero crossings* na saída do filtro Laplaciano da Gaussiana. Ao invés de convoluir cada imagem com doze máscaras direcionais, GRIMSON (1980) convoluiu o filtro Laplaciano da Gaussiana e agrupou os pontos de *zero crossings* em doze grupos direcionais. A forma precisa do operador é dada em coordenadas polar (ρ,θ) por:

$$\nabla^2 G(r,\theta) = \left(\frac{r^2 - 2\sigma^2}{\sigma^4}\right)\exp\left(\frac{-r^2}{2\sigma^2}\right) \tag{3.50}$$

onde:

σ é o desvio-padrão.

Este operador tem a forma de um chapéu mexicano invertido. O comprimento da região central negativa é dado por:

$$w_{2\text{-}D} = 2\sqrt{2}\sigma \tag{3.51}$$

onde:

$w_{2\text{-}D}$ é o comprimento da região central negativa do operador.

GRIMSON (1980) utilizou três ou quatro tamanhos de filtros em sua implementação.

Na fase de *matching*, GRIMSON (1980) utilizou a estratégia assim denominada de *coarse-to-fine*, ou seja, disparidades obtidas em imagens filtradas com grandes filtros eram usadas para guiar a pesquisa dos pontos homólogos em imagens filtradas com filtros pequenos.

MARR (1980) estudou a distribuição de probabilidade do intervalo entre *zero crossings* de mesmo sinal obtidas a partir da convolução de um estereograma de pontos randômicos com o filtro Laplaciano da Gaussiana. Os resultados indicaram que se a disparidade entre as imagens é menor que $\pm\ w/2$, a pesquisa para pontos homólogos em um intervalo de $\pm\ w/2$ produziria *matches* corretos com probabilidade de 0.95. Porém, GRIMSON (1980) utilizou um intervalo de $\pm\ w$ e, ainda assim, obteve a mesma probabilidade determinada por MARR (1980).

Para cada *zero crossing* na imagem da esquerda, os possíveis pontos candidatos a ponto homólogo na imagem da direita são pesquisados ao longo da linha epipolar tal que:

$$x_1 + d_i - w \leq x_2 \leq x_1 + d_i + w \tag{3.52}$$

onde:

x_1 e x_2 é a coordenada na direção x de um ponto na imagem da esquerda e da direita, respectivamente;

d_i é a disparidade estimada; e

$$w = 2\sqrt{2}\sigma . \tag{3.53}$$

Os pontos de *zero crossings* com mesma polaridade e aproximadamente a mesma orientação (dentro de um intervalo de $\pm\ 30^\circ$) são determinados como pontos homólogos. Se somente um ponto é determinado dentro do intervalo $\pm\ w$, então este ponto é aceito como correto e a disparidade associada a este é armazenada no *buffer* dinâmico esboço 2.5-D.

Se mais de um ponto é encontrado dentro do intervalo $\pm\ w$, então aquele que tiver o mesmo tipo de disparidade (divergente, convergente ou nula) da disparidade dominante naquela região, é aceito como correto.

Mais recentemente, GRIMSON (1980) publicou algumas modificações em sua implementação que foram a Continuidade Figural e a disparidade vertical.

Em sua primeira implementação da teoria de MARR & POGGIO (1979), GRIMSON (1980) utilizou o teste de continuidade regional da disparidade para validar os *matches*. Ele observou que isto causava dificuldades na propagação da disparidade em fronteiras ocluídas entre objetos e ao longo de superfícies finas e alongadas, assim os pontos homólogos das feições tendiam para formar contornos estendidos. Desta forma a injunção de Continuidade Figural proposta por MAYHEW & FRISBY (1981) que requer, continuidade de disparidade ao longo de contornos, foi julgada mais apropriada para esta tarefa.

Outra modificação proposta por GRIMSON (1980) para a sua implementação anterior, consiste em realizar a pesquisa do ponto homólogo não mais ao longo da linha epipolar delimitada pelo comprimento w, mas também nas linhas epipolares vizinhas, ou seja, permitindo também disparidade na direção y (vertical). Há evidências psicológicas que o sistema de visão humano movimenta os olhos para corrigir os desalinhamentos das linhas epipolares homólogas. Estes desalinhamentos das linhas epipolares homólogas são causados principalmente pelos efeitos da discretização nas imagens digitais. Assim a expressão 3.52 que só permitia a pesquisa horizontal foi modificada para as seguintes expressões:

$$x_1 + d_i - w \leq x_2 \leq x_1 + d_i + w \tag{3.54}$$

$$y_1 - \varepsilon \leq y_2 \leq y_1 + \varepsilon \tag{3.55}$$

onde:

ε é a altura do espaço de pesquisa na direção vertical.

A altura ε pode ser calculada em função da precisão da calibração das câmeras, já que esta precisão é responsável pelo desalinhamento das linhas epipolares.

3.5.3 Teoria de Prazdny - Princípio da Coerência

PRAZDNY (1985) formulou um algoritmo extremamente simples, não iterativo, paralelo, local, com poucas injunções e apto a resolver o problema de correspondência até mesmo em estereogramas de pontos randômicos portando superfícies transparentes. Ele baseou-se no princípio que "o mundo real (físico) não é feito de pontos cujas profundidades variam caoticamente, mas de objetos que ocupam um bem definido volume tridimensional".

A regra de continuidade proposta por MARR & POGGIO (1979), estipula que pontos próximos no espaço-imagem são projeções de pontos próximos no espaço-objeto, isto implica na suavidade da superfície, e portanto, na suavidade do mapa de disparidade resultante. Entretanto, a continuidade é esperada somente para superfícies opacas e regiões sem descontinuidades de profundidade.

O princípio da coerência é muito mais geral, pois tem como regra que um mapa de disparidade descontínuo pode ser uma superposição de vários mapas de disparidade contínuos entrelaçados, sendo que cada um corresponde a um pedaço suave da superfície.

O princípio da coerência exige que as disparidades vizinhas de elementos correspondentes ao mesmo objeto tridimensional sejam similares, sugerindo que o principal mecanismo de redução de ambigüidades deva ser a ocorrência de disparidades similares. Disparidades dissimilares não inibiriam cada outra porque em situações de superfícies transparentes, a disparidade poderia ser escolhida por uma série de feições correspondentes para outras superfícies. Portanto, duas disparidades são similares no caso em que uma pondere com maior intensidade a probabilidade de ocorrência da outra devido a estas possivelmente conterem informações sobre a mesma superfície, ou dissimilares, no caso em que elas contenham informações distintas, e não deveriam interagir como um todo, devido a estas possivelmente conterem informações sobre superfícies distintas.

Para traduzir este princípio na forma de um algoritmo, PRAZDNY (1985) buscou uma forma de quantificar uma medida de similaridade entre disparidades vizinhas. Como uma primeira aproximação, esta medida de similaridade deveria ser uma função que englobasse os três requisitos seguintes:

1) A função de similaridade de disparidade deve ser inversamente proporcional a diferença de disparidade de pontos correspondidos.

2) Pontos mais distantes devem exercer menor influência, enquanto *matches* próximos devem ter maior poder de redução de *matches* ambíguos.

3) Grandes separações devem originar um função de similaridade de disparidade do tipo *flat*, pequenas separações devem originar uma função de similaridade cujo máximo global seja no centro desta (simétrica).

Uma função que engloba estes requisitos é uma função de distribuição Gaussiana como:

$$s(i,j) = \frac{1}{c|i-j|\sqrt{2\pi}}\exp\left(-\frac{|d_j - d_i|^2}{2c^2|i-j|^2}\right) \tag{3.56}$$

onde:

s(i,j) expressa a quantidade suportada de disparidade d_i em um ponto i recebida a partir da disparidade d_j em um outro ponto j;

$|i-j|$ é a distância entre dois pontos;

d_i e d_j é a disparidade associada ao ponto i e ao ponto j, respectivamente; e

c é o fator de escala que controla o espalhamento da curva.

A quantidade $\left|d_i - d_j\right| / \left|i - j\right|$ no expoente da Gaussiana na expressão 3.56 é o gradiente de disparidade relatado em BURT & JULESZ (1980).

Como pode-se notar, não há inibição neste algoritmo, possibilitando a correspondência para imagens de uma cena que contenha superfícies transparentes. Nota-se também que a primitiva de *matching* não foi citada em momento algum, sendo este algoritmo independente desta primitiva.

3.5.4 Algoritmo cooperativo de Marr & Poggio

MARR & POGGIO (1979) usaram a informação da vizinhança de *matches* em um simples esquema iterativo. Para cada par de linhas epipolares homólogas, uma rede de neurônios conectados é elaborada. As conexões horizontais e verticais são tomadas como conexões inibitórias significando que todos os neurônios ao longo de cada linha horizontal ou vertical inibem cada outro, assim que, finalmente somente um *match* permanece sobre cada linha (horizontal ou vertical) de acordo com o princípio da unicidade, já citado em 3.3.3.

As conexões diagonais são tomadas como conexões excitatórias, significando que estas favorecem *matches* adjacentes diagonalmente em terem a mesma disparidade, de acordo com o princípio da continuidade, citado em 3.3.2. A figura 3.13 ilustra este modelo de rede neural sendo E_X e D_X as linhas epipolares homólogas da esquerda e da direita, respectivamente, e a figura 3.14 mostra as conexões inibitórias (-) e excitatórias (+) e a conseqüente região em forma de disco formada por estas conexões.

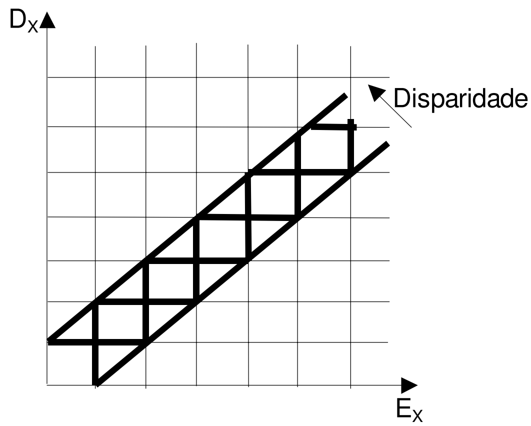


Fig. 3.13 - Rede neural proposta por Marr & Poggio.

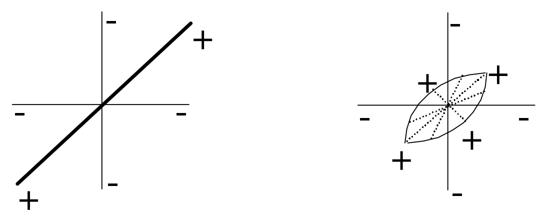


Fig. 3.14 - Conexões excitatórias e inibitórias e disco excitatório.

Para este modelo de rede, $C^t_{x,y;d}$ denota o estado de um neurônio em um tempo t para a coordenada (x,y) na imagem da esquerda correspondente para a coordenada (x+d,y) na imagem da direita.

Inicialmente, todos os neurônios que representam possíveis pontos homólogos são carregados com 1's e todos os demais são carregados com 0's. Desta forma quando a iteração começa, cada neurônio adiciona o estado da vizinhança excitatória em S(x,y,d) e subtrai de uma soma ponderada dos estados dos neurônios na vizinhança inibitória em O(x,y,d). Esta iteração é representada pela seguinte relação:

$$C^{t+1}_{X,Y;d} = \sigma\left(\sum_{X',Y';d' \in S(X,Y,d)} C^t_{X',Y';d'} - \varepsilon \sum_{X',Y';d' \in O(X,Y,d)} C^t_{X',Y';d'} + C^0_{X,Y;d} \right) \qquad (3.57)$$

onde:

ε é o peso para o efeito de inibição;

σ é a função limiar;

S é o conjunto de pontos x', y', d', tais que $|x - x'| \leq 1$ e d = d' ; e

O é o conjunto de pontos x', y', d', tais que $|x - x'| \leq 1$ e $|d - d'| = 1$.

Este algoritmo resolveu o problema de correspondência com sucesso para estereogramas de pontos randômicos. Ele representa um mecanismo muito simples para a propagação de unicidade e continuidade entre pontos homólogos vizinhos, resolvendo o problema de ambigüidade de maneira iterativa.

3.5.5 Relaxação multi-níveis de Terzopoulos (processos hierárquicos)

Os processos hierárquicos em algoritmos estéreos são caracterizados principalmente, pela troca da correspondência estéreo obtida em diferentes níveis de informações.

TERZOPOULOS (1983) desenvolveu um eficiente modelo de relaxação multi-níveis para o problema de correspondência. Os métodos convencionais hierárquicos empregam coordenação computacional recursiva e fluxo de resultados intermediários a partir de baixos níveis em direção a níveis mais altos de informações. Esta estratégia é conhecida como *coarse to fine* e foi proposta por MARR & POGGIO (1979).

Na estratégia *coarse to fine*, resultados em níveis mais baixos são utilizados como aproximações para o próximo nível. Tal estratégia resulta em algoritmos tipicamente seqüenciais, tornando ineficiente o desempenho computacional, já que o maior tempo gasto durante o processamento se faz durante os processos de relaxação sobre cada nível, enquanto processos em outros níveis permanecem ociosos. A estratégia de TERZOPOULOS (1983) mantém processos ativos em todos os níveis executando operações de relaxação simultâneas, procurando otimizar um objetivo funcional multi-nível, com cada termo possuindo três componentes que são:

1) Uma versão discreta de uma dada funcional em cada nível de um *multigrid* hierárquico;

2) Uma funcional aditiva, unindo cada nível para um próximo nível mais alto, com exceção do último nível mais alto.

3) Uma funcional aditiva, unindo cada nível para um próximo nível mais baixo, com exceção do último nível mais baixo.

O algoritmo de TERZOPOULOS (1983) pode ser entendido como uma otimização da estratégia *coarse to fine*.

Capítulo 4

Solução do problema de correspondência

4. Introdução

O problema de correspondência estéreo é um problema essencialmente local, e portanto, soluções globais não fornecem bons resultados nem tão pouco robustez para a solução. Esta afirmação é um indicador de que uma possível solução para o problema de correspondência seja obtida através de um algoritmo baseado em derivadas de funções.

As soluções para o problema de correspondência existentes que não envolvem derivadas geralmente obtêm bons resultados em cenas cuja a superfície a ser reconstruída é relativamente suave, fazendo com que a disparidade também varie suavemente para toda cena. O termo "relativamente suave" foi utilizado porque há casos em que a superfície no espaço-objeto não é necessariamente suave, mas mesmo assim, o mapa de disparidade referente a cena pode ser considerado suave. Este fato ocorre quando a relação base - profundidade é pequena.

Quando o mapa de disparidade associado a uma cena não é suave, o par de imagens estereoscópicas não deve ser tratado integralmente, mas sim decomposto em regiões. Estas regiões são áreas de atenção da cena que são denominadas áreas fusionais.

A principal vantagem de solucionar o problema de correspondência localmente é que a maioria dos problemas da visão estéreo citado em 3., tais como transformações geométricas e efeitos de oclusão entre o par de imagens estereoscópicas, são fortemente minimizados nestas condições.

A metodologia proposta para solucionar o problema de correspondência estéreo apresentada neste capítulo utiliza as premissas geométricas da visão estéreo (citadas em 2.6 e 2.7.1) e todas as evidências biológicas (como inspiração para formalizar um algoritmo) com exceção das evidências biológicas de processos cooperativos e área fusional de Panum. As evidências biológicas da visão estéreo são tratadas em 3.4.

As premissas empíricas da visão estéreo (citadas em 3.3.1, 3.3.2 e 3.3.3) não serão utilizadas na metodologia proposta devido a estas não poderem ser generalizadas para qualquer tipo de cena.

A evidência biológica de área fusional de Panum não será utilizada devido a esta ser conflitante com o conceito de gradiente de disparidade. Segundo BURT & JULESZ (1980), o gradiente de disparidade é o fator limitante para a fusão quando dois ou mais objetos ocorrem próximos no campo visual e não a área fusional de Panum. Esta afirmação diz, em termos mais simples, que o intervalo de disparidade permitido na busca do ponto homólogo deve ser variável e diretamente proporcional a magnitude deste gradiente (desde que este não exceda o limite de gradiente de disparidade), ao contrário da área fusional de Panum, que assume um intervalo de disparidade fixo permitido na busca do ponto homólogo para todo o campo visual.

O gradiente de disparidade portanto, será empregado como um preditor do intervalo de disparidade permitido na determinação do ponto homólogo, adaptando-se continuamente para cada ponto do par de imagens estereoscópicas. A solução encontrada para descrever esta evidência biológica na forma de um algoritmo é tratada em 4.3.

A razão da não utilização dos processos cooperativos se deve ao fato de que estes propagam a própria disparidade associada a pontos vizinhos, fazendo com que o mapa de disparidade se torne suave. O gradiente de disparidade não propaga a disparidade propriamente dita de pontos vizinhos, e sim, um intervalo de disparidade adaptativo.

As soluções encontradas para descrever as evidências biológicas de movimento de vergência dos olhos, compatibilidade e seletividade direcional na forma de um algoritmo são tratadas em 4.1.

A tabela 4.1 mostra as premissas e evidências biológicas que serão empregadas e as que não serão empregadas na metodologia proposta para a solução do problema de correspondência.

Empregadas	Não empregadas
Geometria epipolar	Ordem
Restrição do intervalo de correspondência	Continuidade
Movimento de vergência dos olhos	Unicidade
Gradiente de disparidade	Processos cooperativos
Seletividade direcional	Área fusional de Panum
Compatibilidade	

Tabela 4.1 - Premissas e evidências biológicas empregadas na metodologia proposta.

4.1 Controle do movimento de vergência dos olhos através do *matching* baseado em feições

A imagem de uma cena capturada por uma câmera digital, não deve ser comparada a uma imagem capturada pelo olho devido as várias diferenças entre as mesmas, mas principalmente, devido a imagem digital representar toda a cena e a imagem capturada pelo olho (onde a visão é realmente efetiva) representar apenas uma região desta. Um analogia mais coerente, seria entender a imagem digital como a imagem do campo visual capturada pelo olho, onde não necessariamente a visão está concentrada.

A imagem digital deve portanto conter algumas áreas onde esta atenção é concentrada. Em aplicações de visão estéreo, as áreas de atenção de uma imagem devem ser correspondidas com áreas de atenção da outra imagem, resultando em áreas de atenção homólogas (áreas fusionais).

A determinação das áreas de atenção homólogas serão realizadas através do *matching* baseado em feições.

O *matching* baseado em feições foi escolhido, ao invés do *matching* baseado em áreas utilizando grandes janelas como sugerido por MARR & POGGIO (1979), devido a este ser mais robusto e ser realizado em um nível mais alto de informação (atributos associados as feições).

É natural que feições controlem o movimento de vergência porque em geral, estas possuem características distintas que as diferem das demais áreas das imagens estereoscópicas na fase de segmentação. Esta distinção faz com os olhos concentrem atenção nestas áreas inicialmente, justamente por representarem singularidades da cena como um todo.

Um problema desta classe de *matching* é justamente determinar quais tipos de feições são mais interessantes para uma dada aplicação. Como um dos objetivos deste trabalho visa a obtenção de uma solução para o problema de correspondência estéreo sem que um tipo de cena específica seja definida, faz-se necessário escolher tipos de feições que são comuns na maioria das aplicações.

Em aplicações que utilizam imagens de objetos industrializados ou de ambientes modificados pelo homem é comum encontrar feições retas. Já em imagens de satélites artificiais e imagens médicas, por exemplo, a linearidade das feições são raras, porém é muito comum regiões que possuem texturas razoavelmente homogêneas, tais como florestas, rios e culturas em imagens orbitais e órgãos de seres vivos em imagens médicas, portanto, as feições que serão utilizadas para o *matching* e consequentemente para controlar o movimento de vergência, serão feições retas e regiões homogêneas.

Para a tarefa de segmentação de feições retas utilizou-se a transformada de Hough (citada em 3.2.4) e para a segmentação de regiões homogêneas utilizou-se algumas ferramentas tradicionais de processamento de imagens, tais como *threshold*, fechamento de buracos (*close holes)*, abertura morfológica e rotulação. A segmentação de feições retas é tratada no item 4.1.1, a segmentação de regiões homogêneas é tratada no item 4.1.2 e o *matching* de feições nos itens 4.1.3 e 4.1.4.

4.1.1 Segmentação de feições retas

A segmentação de feições retas nas imagens estereoscópicas foi realizada através da transformada de Hough (HOUGH, 1962) citada em 3.2.4.

Os parâmetros θ e ρ das feições retas são determinados no espaço de parâmetros através de um simples *threshold*, que seleciona os pares (θ,ρ) com maiores ocorrências, sem considerar se existem pares (θ,ρ) muito similares. Este fato pode ser prejudicial para a fase de *matching* de feições retas (tratado no item 4.1.3) porque dificulta a redução de ambiguidades entre estas feições nas imagens estereoscópicas.

Para solucionar o problema de proximidade das feições retas utilizou-se um *threshold* para θ e outro *threshold* para ρ, limitando assim, a seleção de duas ou mais feições retas de aproximadamente mesma inclinação θ e aproximadamente mesma distância ρ. Desta maneira, o resultado desta segmentação foram as feições retas de maiores ocorrências no espaço de parâmetros cujos os parâmetros θ e ρ são relativamente distintos.

4.1.2 Segmentação de regiões homogêneas

Ao contrário dos métodos tradicionais de segmentação de regiões conhecidos como crescimento de regiões que são baseados em agregação de *pixels*, utilizou-se para esta tarefa outras ferramentas de processamento de imagens.

O crescimento de regiões inicializa a agregação de *pixels* a partir de alguns *pixels* ditos como "sementes" e um predicado lógico decide se um dado *pixel* nas vizinhanças de uma semente em questão, deve ou não ser agregado à aquela região. Este método não foi empregado devido a dificuldade em encontrar critérios (que visem a visão estéreo) para a determinação das sementes. Este método é muito comum e tradicional em processamento de imagens e as referências BALLARD & BROWN (1982); e GONZALEZ & WOODS (1993) são sugeridas para melhores informações.

As regiões foram então segmentadas seguindo os seguintes passos:

1) Binarização das imagens estereoscópicas através de um valor de *threshold* superior e um valor de *threshold* inferior;

2) Fechamento de buracos através de operadores morfológicos;

3) Abertura morfológica para a remoção de pequenas áreas e saliências externas e

4) Rotulação das regiões.

A binarização das imagens estereoscópicas é uma tarefa simples e consiste em determinar quais *pixels* das imagens possuem valores de intensidade contidos no intervalo de *threshold* superior e inferior. Os *pixels* que estão contidos neste intervalo recebem o valor um e os demais o valor zero.

O segundo passo para segmentar as regiões foi utilizar um conjunto de operadores morfológicos, tais como negação e abertura morfológica, para eliminar os "buracos" nas regiões.

Após a operação de fechamento de buracos ainda restaram pequenas regiões que não seriam interessantes no processo de *matching*. Para eliminar estas pequenas regiões e saliências externas utilizou-se o operador morfológico de abertura.

Após estas operações, as imagens estereoscópicas apresentam regiões com boa segmentação para o processo de *matching*, podendo então serem rotuladas.

As referências HARALICK *et al.* (1987); MEYER & BEUCHER (1990), FACON (1996); e LOTUFO & TRETTEL (1996), tratam detalhadamente os conceitos de morfologia matemática utilizados nesta etapa e são sugeridas para melhores informações.

4.1.3 *Matching* de feições retas

O *matching* de feições retas foi implementado utilizando a função de correlação erro quadrático (citada em 3.1.1) no espaço de atributos associado a cada reta como a medida de similaridade entre as feições da imagem da esquerda e as feições da imagem da direita.

Utilizando-se da transformada de Hough, determinou-se a inclinação θ e a distância ρ para cada feição reta selecionada nas imagens estereoscópicas. O tamanho das feições retas pode ser calculado através das coordenadas iniciais e finais de cada feição reta, porém este valor está presente no próprio espaço de parâmetros (células de acumulação) e sua obtenção é direta.

De posse dos parâmetros θ e ρ de cada feição reta selecionada nas imagens estereoscópicas, verifica-se no espaço xy, as coordenadas iniciais e finais, o sinal do gradiente de intensidade (não confundir com gradiente de disparidade) e a intensidade média, bem como a existência ou não de falhas, podendo assim gerar novas feições retas de coordenadas iniciais e finais distintas, apesar de possuírem os mesmos parâmetros θ e ρ.

A figura 4.1 ilustra um exemplo de duas feições retas que possuem os mesmos parâmetros θ e ρ, mas coordenadas iniciais e finais distintas.

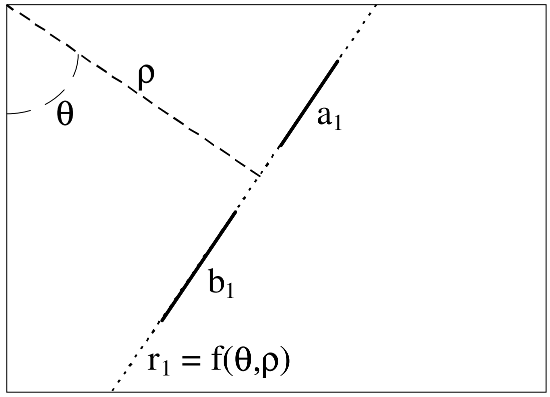


Fig. 4.1 - Feições retas de mesmo parâmetros θ e ρ.

Nesta figura 4.1, a linha pontilhada mostra a reta r_1 que é função dos parâmetros θ e ρ, e as feições retas a_1 e b_1 que possuem os mesmos parâmetros θ e ρ, mas suas coordenadas iniciais e finais são distintas devido a ocorrência de uma falha.

Os sinais do gradiente de intensidade (que indicam as polaridades na direção x e na direção y) foram determinados através dos operadores de Sobel. Estes operadores são máscaras 3 x 3 e são muito comuns em processamento de imagens. Utilizando os operadores de Sobel na direção x, o gradiente G_x nesta direção é obtido pela seguinte máscara:

-1	0	1
-2	0	2
-1	0	1

Utilizando os operadores de Sobel na direção y, o gradiente G_y nesta direção é obtido pela seguinte máscara:

-1	-2	-1
0	0	0
1	2	1

Com o intuito de restringir o *matching*, utilizou-se as premissas geométricas da visão estéreo de geometria epipolar e restrição do intervalo de correspondência, citadas em 2.6 e 2.7.1, e as evidências biológicas de seletividade direcional e compatibilidade, citadas em 3.4.

Devido a problemas inerentes no processo de segmentação, adotou-se um pequeno intervalo de variação aceitável na direção y, para que feições retas que tenham pequenas discrepâncias nas coordenadas iniciais e finais nesta direção, possam ainda ser correspondidas, relaxando assim, a geometria epipolar.

Uma feição reta em uma imagem só pode corresponder-se com uma ou mais feições retas na outra imagem se todas essas restrições forem satisfeitas. O par de feições candidatas a feições homólogas que possuem a maior medida de similaridade é adotada como feições homólogas, no caso de empate de duas ou mais maiores medidas de similaridades, esses pares então são tidos como homólogos, evitando assim, a premissa empírica da visão estéreo de unicidade (citada em 3.3.2).

Os atributos utilizados para computar o erro quadrático no espaço de atributos entre duas feições retas são:

-Tamanho da feição reta;

- θ; e

- Intensidade média da feição reta.

Apesar do espaço de atributos possuir uma dimensão relativamente pequena (dimensão 3), as restrições impostas resolvem a grande maioria das ambigüidades, tornando quase sempre simples a determinação das feições homólogas através da medida de similaridade, quando está se faz necessária.

Ao utilizar o tamanho da feição reta como um atributo de medida de similaridade, está-se na verdade, violando o princípio de que as dimensões não são preservadas entre as imagens estereoscópicas (devido aos problemas da visão estéreo, citado em 3.), mas a exemplo da seletividade direcional, que só permite que duas feições retas correspondam-se para os casos em que suas orientações são semelhantes (dentro de um certo intervalo de aceitação), é de se esperar que duas feições retas também possuam tamanhos semelhantes para corresponderem-se.

Entretanto, cabe ressaltar que o tamanho das feições retas não foi utilizado como um fator decisivo se duas feições retas devem ou não corresponderem-se, como nas restrições citadas, e sim, apenas com uma ponderação na medida de similaridade caso está seja necessária.

Uma versão simplificada do algoritmo para o *matching* de feições retas pode ser escrito como:

SE (geometria epipolar é coerente E restrição do intervalo de correspondência é respeitado E compatibilidade é respeitada E seletividade direcional é coerente) ENTÃO (calcular erro quadrático no espaço de atributos).

4.1.4 *Matching* de regiões

O *matching* de regiões foi implementado, a exemplo do *matching* de feições retas, utilizando como medida de similaridade entre as regiões da imagem da esquerda e as regiões da imagem da direita, a função de correlação erro quadrático (citada em 3.1.1) no espaço de atributos associado a cada região.

Os atributos topológicos de uma região seriam quase que ideais para o processo de *matching*, como citado em 3.2.6, porém para regiões que não são representadas por linhas retas, o espaço de atributos que é gerado utilizando esta classe de atributos é de dimensão 2, quando considera-se como atributos o número de componentes conectados e o número de "buracos" de uma região, ou de dimensão 1, quando considera-se apenas o número de Euler.

Outra desvantagem dos atributos topológicos de uma região é que estes são sempre representados por números inteiros e na maioria das vezes, com valores baixos, aumentando assim, a ambigüidade de regiões, e portanto, dificultando o processo de *matching* estéreo, além de serem difíceis de determiná-los em um processo automático.

Os atributos associados a cada região segmentada nas imagens estereoscópicas foram os sete momentos invariantes descritos em 3.2.7.

Os sete momentos invariantes foram escolhidos para o processo de *matching* de regiões porque estes sintetizam em um espaço de dimensão 7 informações a respeito da forma e intensidade de uma região.

Os sete momentos invariantes são invariantes a translação, rotação e escala, porém como citado em 3., a relação que modela as distorções geométricas entre o par de imagens estereoscópicas é uma relação polinomial cujo polinômio é uma composição de transformações projetivas, e portanto, a invariância quanto a translação, rotação e escala é insuficiente para obter invariância total no processo de *matching* estéreo; porém devido a não existir ferramentas matemáticas que consigam invariância para este tipo de relação polinomial, os sete momentos invariantes foram utilizados, não como geradores de atributos ideais para o processo de *matching* estéreo, mas sim, como os melhores geradores de atributos possíveis para esta tarefa.

Quando f(x,y) é considerado unitário para todo (x,y), os momentos de ordem p+q calculados representam apenas informações em respeito a forma de uma região plana, portanto este fato ocorre quando utiliza-se imagens binárias.

Quando f(x,y) possui valores diferentes da unidade, os momentos de ordem p+q calculados representam informações de uma região tridimensional, que no caso de imagens em níveis de cinza, englobam em medidas únicas, informações em respeito a forma e a intensidade de uma região.

Uma vez de posse das imagens estereoscópicas binarizadas segmentadas em regiões, a obtenção das imagens com suas respectivas intensidades nas regiões segmentas é simples e esta tarefa é realizada multiplicando a imagem original pela imagem binarizada.

Através ainda do momento m de ordem (p+q) é possível determinar o centróide de cada região através das expressões 3.29 e 3.30.

Para restringir o *matching* de regiões, utilizou-se as premissas geométricas de geometria epipolar e restrição do intervalo de correspondência, citadas em 2.6 e 2.7.1. A exemplo do *matching* de feições retas, adotou-se também um pequeno intervalo de variação aceitável na direção y devido aos problemas inerentes no processo de segmentação, fazendo com regiões que não estejam nas mesmas linhas epipolares possam ainda ser correspondidas.

Uma região em uma imagem só pode corresponder-se com uma ou mais regiões na outra imagem se todas estas restrições forem satisfeitas. O par de regiões candidatas a regiões homólogas que possuem a maior medida de similaridade é adotada como regiões homólogas, no caso de empate de duas ou mais maiores medidas de similaridades, esses pares então são tidos como homólogos.

Como resultado deste processo de *matching*, tem-se regiões homólogas, porém todos os *pixels* das regiões ditas homólogas, não devem ser considerados homólogos devido aos problemas da visão estéreo citada em 3., e portanto, utilizou-se apenas como resultado deste processo, os centróides das regiões homólogas, tendo portanto, centróides homólogos.

Os centróides homólogos não representam a disparidade real devido a estes serem influenciados por todos os pontos das regiões que podem e são na maioria das vezes, distorcidas uma em relação a outra, mesmo que as regiões sejam realmente homólogas; porém estes são os *pixels* com maiores chances de serem realmente homólogos, dentre os demais das regiões. Este fato faz com que a disparidade resultante seja um valor aproximado da disparidade real, porém com grandes de chances de estar próximo do valor real.

O intuito dessa fase do processo estéreo é controlar o movimento de vergência e portanto resultados aproximados são aceitáveis, uma vez que as etapas seguintes do processo refinarão os resultados.

Uma versão simplificada do algoritmo para o *matching* regiões pode ser escrito como:

SE (geometria epipolar é coerente E restrição do intervalo de correspondência é respeitada) ENTÃO (calcular erro quadrático no espaço de atributos).

4.2 Interpolação do mapa de disparidade

Em uma primeira fase do processo estéreo, o *matching* baseado em feições gerou uma lista de pontos de coordenadas (x_E,y_E) referentes a imagem da esquerda e seus respectivos pontos homólogos de coordenadas (x_D,y_D) referentes a imagem da direita.

O *matching* de feições retas gerou, para cada par de feições retas homólogas, todos os pontos homólogos sobre estas ao contrário do *matching* de regiões, que apenas determinou os centróides homólogos.

Essa lista de coordenadas de pontos homólogos permite determinar a disparidade para apenas estes pontos, gerando assim, um mapa de disparidade para a cena apenas esparso e, portanto, o gradiente de disparidade não pode ser calculado, a menos que estes pontos ocorram relativamente próximos no espaço ciclopeano.

A solução encontrada para calcular o gradiente de disparidade para toda a imagem ciclopeana foi interpolar os valores de disparidade através de algum método de interpolação, gerando com isso, um mapa de disparidade denso para a imagem ciclopeana.

A interpolação dos valores de disparidade gera um mapa de disparidade denso, porém suave. Entretanto o gradiente de disparidade é utilizado como um preditor do intervalo de disparidade permitido na determinação do ponto homólogo, como citado em 4.1 e, portanto, está suavidade poderá ser minimizada ou mesmo eliminada em uma fase posterior no processo estéreo que será tratada em 4.6.

A dificuldade em interpolar os valores de disparidade está em que estes não são regularmente espaçados, aumentando consideravelmente a complexidade do algoritmo desenvolvido para esta tarefa.

Existem diversos métodos de interpolação tais como o de Lagrange, de Lagrange quadrático, de Newton e *splines*. Estes tipos de interpolação são tratados detalhadamente em KREYSZIG (1993); e FOLEY *et al* (1982). Entretanto, TOMMASELLI, *et al* (1996) sugerem realizar a interpolação sob algum critério a partir da triangulação de Delanay. A triangulação de Delanay é um método de interpolação de alto custo computacional e a exemplo de TOMMASELLI, *et al* (1996), não será utilizado para esta tarefa.

TOMMASELLI, *et al* (1996) utilizaram um polinômio linear em relação aos parâmetros para ajustar uma superfície sobre cada conjunto de nove pontos. O modelo matemático deste polinômio é representado por:

$$Z = a + b.X + c.Y + d.X^2 + e.X.Y + f.Y^2 \tag{4.1}$$

onde:

X, Y e Z são coordenadas no espaço tridimensional; e

a, b, c, d, e e f são parâmetros do polinômio.

Este polinômio mostrou-se eficiente para poucos pontos e relativamente próximos, porém no caso de interpolação dos valores de disparidade no espaço ciclopeano nenhuma dessas duas considerações podem ser utilizadas generalizadamente.

O método de interpolação utilizado foi o de interpolação de *spline* bi-harmônica (SANDWELL, 1987). Para esta tarefa, utilizou-se a função "GRIDDATA" do *software* MATLAB, versão 4.2c.1.

A interpolação *spline* bi-harmônica foi utilizada para a tarefa de interpolação dos pontos homólogos oriundos do *matching* baseado em feições porque este método consiste basicamente em utilizar funções de Green de operadores bi-harmônicos para minimizar a curvatura de interpolação de dados não regularmente espaçados. Assim, as curvaturas de interpolação entre os pontos oriundos do *matching* baseado em feições são mínimas (sob este critério), gerando um mapa de disparidade fortemente suave.

Outros métodos de interpolação poderiam ser utilizados (por exemplo, os métodos de interpolação já citados neste item), entretanto, optou-se em utilizar este método de interpolação (que gera um mapa de disparidade fortemente suave) porque nenhuma suposição à respeito da superfície do mapa de disparidade pode ser feita prematuramente. Esta metodologia, visa a obtenção de um mapa de disparidade denso que permita a ocorrência de superfícies não suaves, e portanto, a ocorrência de superfícies suaves também devem ser consideradas.

Cabe ressaltar, que a lista de coordenadas dos pontos homólogos não está referenciada no espaço ciclopeano e sim, no espaço-imagem da câmera da esquerda e no espaço-imagem da câmera da direita. Portanto existe a necessidade de transformar estas coordenadas nos espaços imagem para o espaço ciclopeano. Esta transformação é realizada através das expressões 3.47, 3.48 e 3.49.

A utilização do espaço ciclopeano facilita a implementação dos programas, pois pode-se associar a disparidade de um par de pontos homólogos a um único ponto na forma $d(x_c,y_c)$ e, portanto, utilizar apenas uma única imagem.

4.3 Cálculo do gradiente de disparidade no mapa de disparidade

O mapa de disparidade interpolado é um mapa de disparidade denso em que todos os *pixels* da imagem ciclopeana possuem, embora interpolado, um valor de disparidade $d(x_c,y_c)$. Este fato permite que o gradiente de disparidade possa ser calculado para todos os *pixels* (exceto nas bordas da imagem) da imagem ciclopeana, já que estes estão o mais próximos possíveis (uns dos outros) considerando que a imagem ciclopeana é também uma imagem discreta.

Devido a imagem ciclopeana possuir valores de disparidade $d(x_c,y_c)$ para todos os seus *pixels*, pode-se tratar esta imagem como uma imagem comum de intensidade e, portanto, o gradiente calculado na imagem ciclopeana é o próprio gradiente de disparidade.

A operação de derivada de uma função no caso discreto é aproximada utilizando-se operadores cuja a somatória de seus pesos é nula.

O operador que será utilizado para calcular o gradiente de disparidade será o operador de Prewitt. Este operador foi escolhido porque este é um operador resultante da convolução de um filtro digital unidimensional passa-baixa por um filtro digital unidimensional passa-alta. Assim, este operador realiza em uma única operação de convolução, as tarefas de atenuar ruídos e detectar bordas.

O gradiente na direção x, utilizando o operador de Prewitt é obtido pela convolução da seguinte máscara pela imagem:

-1	0	1
-1	0	1
-1	0	1

O gradiente na direção y, utilizando o operador de Prewitt é obtido pela convolução da seguinte máscara pela imagem:

-1	-1	-1
0	0	0
1	1	1

A disparidade é determinada através da expressão 2.61 para imagens retificadas, assim, o gradiente de disparidade na direção y não deverá ser utilizado, pois variações nesta direção não são permitidas nestas circunstâncias.

O operador de Prewitt do gradiente na direção x é então convoluído com a imagem ciclopeana determinando o gradiente de disparidade.

O módulo do gradiente de disparidade $\gamma(x_c,y_c)$ fornece a informação de quanto a disparidade está variando no *pixel* de coordenada (x_c,y_c) e o sinal deste gradiente fornece o sentido desta variação, indicando se a disparidade está variando de valores menores para valores maiores ou vice-versa.

Quando a disparidade varia no sentido de valores menores para valores maiores, o sinal do gradiente de disparidade é positivo, quando a disparidade varia no sentido de valores maiores para valores menores o sinal do gradiente de disparidade é negativo. Quando a disparidade não varia, o gradiente de disparidade é nulo. A figura 4.2 mostra um mapa de disparidade discreto de tamanho 3 x 9, o operador G_x de Prewitt de gradiente na direção x e o mapa de gradiente de disparidade resultante da convolução deste operador pelo mapa de disparidade.

mapa de disparidade

1	1	1	3	3	3	1	1	1
1	1	1	3	3	3	1	1	1
1	1	1	3	3	3	1	1	1

G_x

-1	0	1
-1	0	1
-1	0	1

mapa de gradiente de disparidade

	0	6	6	0	-6	-6	0	

Fig. 4.2 - Exemplo de mapa de disparidade, G_x e mapa de gradiente de disparidade.

Na figura 4.2, a primeira e a última linha e a primeira e a última coluna não receberam valores de gradiente de disparidade devido aos problemas tradicionais de processamentos nas bordas da imagem.

4.4 Intervalo de busca do *pixel* homólogo

O gradiente de disparidade G_X é um preditor do tamanho do espaço de busca do ponto homólogo, assim, dado um *pixel* de coordenadas (x_E,y_E) na imagem da esquerda, o *pixel* homólogo (x_D,y_D) na imagem da direita poderá estar em alguma posição do seguinte intervalo:

$$x_E - d_i(x_C, y_C) - \lambda.|G_X| \leq \quad x_D \quad \leq x_E - d_i(x_C, y_C) + \lambda.|G_X| \tag{4.2}$$

$$y_D = y_E \tag{4.3}$$

onde:

$d_i(x_C,y_C)$ é o valor de disparidade interpolado na posição (x_C,y_C) do espaço ciclopeano; e

λ é um fator de escala para o gradiente de disparidade.

Uma vez que

$$x_D = x_E - d(x_C, y_C) \tag{4.4}$$

e portanto

$$x_{Da} = x_E - d_i(x_C, y_C) \tag{4.5}$$

onde:

x_{Da} é a coordenada aproximada x de um *pixel* na imagem da direita, calculada a partir do valor de disparidade interpolado $d_i(x_C,y_C)$ na posição (x_C,y_C) do espaço ciclopeano.

a expressão 4.2, pode ser escrita diretamente como:

$$x_{Da} - \lambda.|G_X| \leq \quad x_D \quad \leq x_{Da} + \lambda.|G_X| \tag{4.6}$$

O fator de escala λ foi introduzido para dar maior versatilidade ao gradiente de disparidade. Este por sua vez, é uma medida relativamente empírica e está relacionado com a suavidade da cena e a qualidade da interpolação. Como estes parâmetros são difíceis de serem determinados, sua omissão é possível e para isto basta com este seja igual a unidade.

Outra interpretação possível que pode ser dada ao fator de escala λ é para o limite do gradiente de disparidade determinado por BURT & JULESZ (1980). Como citado em 3.4 e 3.4.1, o gradiente de disparidade não excede a unidade para o sistema de visão humano, porém, computacionalmente este limite é perfeitamente possível de ser restringido ou relaxado, permitindo assim, um intervalo menor ou maior para a fusão de dois pontos homólogos.

Ainda a respeito sobre este fator de escala, este pode ser utilizado para dividir o valor do gradiente de disparidade resultante do operador de Prewitt na direção x pela a dimensão da máscara deste operador, já que esta operação utiliza a informação da vizinhança e não somente a do *pixel* em questão. Para esta tarefa, este fator deve ser igual a unidade dividida pela dimensão da máscara (no caso, $1/3$).

A expressão 4.2 e a sua outra forma de representação, a expressão 4.6, não levam em consideração o sentido do gradiente de disparidade, uma vez que estes utilizam o módulo do valor deste gradiente.

Quando o sentido do gradiente de disparidade é levado em consideração, o intervalo de busca do *pixel* homólogo na direção x para o sinal do gradiente positivo fica:

$$x_E - d_i(x_C, y_C) + \lambda.G_X \leq \quad x_D \quad \leq x_E - d_i(x_C, y_C) \tag{4.7}$$

ou

$$x_{Da} - \lambda.G_X \leq \quad x_D \quad \leq x_{Da} \tag{4.8}$$

Para o sinal do gradiente de disparidade negativo, o intervalo de busca do *pixel* homólogo fica:

$$x_E - d_i(x_C, y_C) \leq \quad x_D \quad \leq x_E - d_i(x_C, y_C) - \lambda.G_X \tag{4.9}$$

ou

$$x_{Da} \leq \quad x_D \quad \leq x_{Da} - \lambda.G_X \tag{4.10}$$

É importante ressaltar que o gradiente de disparidade é calculado para cada *pixel* da imagem ciclopeana, gerando assim, um intervalo permitido de busca do *pixel* homólogo para cada *pixel* desta imagem. Esta variação do intervalo de busca é o que difere em utilizar o conceito de gradiente de disparidade ao invés do conceito da área fusional de Panum, onde este intervalo é fixo para toda a imagem ciclopeana.

4.5 Determinação do *pixel* homólogo no intervalo de busca através do *matching* baseado em áreas

A busca do *pixel* homólogo no intervalo de busca permitido determinado através do gradiente de disparidade foi implementado utilizando o *matching* baseado em áreas descrito em 3.1.

O coeficiente de correlação da função covariância cruzada foi a medida de similaridade empregada no processo de correlação, também citado em 3.1.

Esta classe de *matching* pode ser utilizada para esta tarefa sem maiores problemas, pois o intervalo de busca determinado é geralmente pequeno, possibilitando que as dimensões das janelas de referência e pesquisa tenham dimensões também pequenas. Este fato minimiza fortemente os problemas quanto as transformações geométricas (composição de transformações projetivas) entre as imagens estereoscópicas.

Este processo de *matching* também utilizou as premissas geométricas de geometria epipolar e restrição do intervalo de correspondência. A premissa de geometria epipolar foi empregada implicitamente neste processo, pois o intervalo de busca permitido já utiliza esta premissa. A premissa de restrição do intervalo de correspondência pode eventualmente ser necessária, se o intervalo de busca determinado exceder a restrição de não permitir disparidades negativas. Este fato ocorre se a coordenada máxima do intervalo de busca na direção x (x máximo) na imagem da direita exceder o valor da coordenada na direção x de um dado *pixel* na imagem da esquerda.

Para minimizar o problema de oclusão citado em 3., utilizou-se um *threshold* para a diferença das médias das intensidades nas janelas de referência e de pesquisa, sendo que, se esta diferença exceder o valor de *threshold* permitido, os *pixels* associados a estas médias não podem ser correspondidos.

Este *threshold* foi denominado de *threshold* de oclusão e pode ser considerado uma premissa empírica da visão estéreo, pois na maioria das vezes, as áreas de oclusão possuem valores bem distintos de intensidade, entretanto, este fato não pode ser generalizado para todos os casos de oclusão.

Uma versão simplificada do algoritmo para o *matching* baseado em áreas no intervalo de busca permitido pode ser escrito como:

SE (restrição do intervalo de correspondência é respeitada E *threshold* de oclusão é coerente) ENTÃO (calcular o coeficiente de correlação entre a janela de referência e as janelas de pesquisa ao longo do intervalo de busca).

4.6 Correção do mapa de disparidade interpolado

A correção do mapa de disparidade interpolado é a fase final e mais importante desta metodologia. É nesta fase que a suavidade do mapa de disparidade é minimizada ou mesmo eliminada em regiões onde a disparidade varia bruscamente.

A lista de coordenadas de pontos homólogos oriunda do *matching* baseado em feições (*matching* de feições retas e *matching* de regiões) foi utilizada para o controle do movimento de vergência nas áreas das imagens estereoscópicas onde estas feições ocorreram. A partir destas coordenadas determinou-se as coordenadas no espaço ciclopeano e suas respectivas disparidades, sendo que estas foram interpoladas para as demais áreas da imagem ciclopeana, gerando com isso, um mapa de disparidade denso para toda esta imagem.

O processo de interpolação, obviamente, é uma aproximação do mapa de disparidade "real", porém, nas coordenadas dos pontos homólogos (provenientes do *matching* baseado em feições) transformadas para o espaço ciclopeano e nas vizinhanças destes, é de se esperar que o processo de interpolação gere um mapa de disparidade onde o erro entre este mapa interpolado e o mapa de disparidade "real", seja mínimo para estes pontos. A figura 4.3 mostra em traço cheio uma curva R representando os valores "reais" de disparidade e em pontilhado, uma curva I interpolada com base em algum critério de interpolação dos valores de disparidades (representados por pontos).

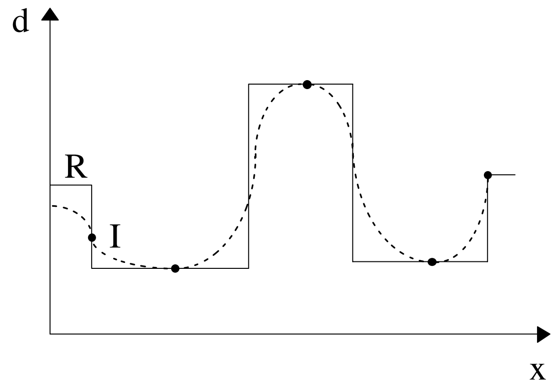


Fig. 4.3 - Curva real R e curva interpolada I.

Na figura 4.3 percebe-se que o erro entre a curva “real” R e a curva interpolada I é nulo para os pontos fornecidos para o processo de interpolação e mínimo em suas vizinhanças. Cabe lembrar que, apenas para simplificação, esta figura ilustra este conceito para um espaço bidimensional, e portanto, este deve ser expandido para um espaço tridimensional que é a verdadeira dimensão deste problema.

Utilizando-se deste conceito, é natural que o início da correção do mapa de disparidade interpolado seja nos pontos oriundos do *matching* baseado em feições e em suas vizinhanças. Desta forma o processo de correção do mapa de disparidade interpolado se dá através dos seguintes passos:

1) Calcular o gradiente de disparidade através do mapa de disparidade interpolado nos pontos oriundos do *matching* baseado em feições;

2) Determinar o intervalo permitido de busca do *pixel* homólogo;

3) Realizar o *matching* baseado em áreas no intervalo de busca permitido e adotar como *pixel* homólogo o de maior coeficiente de correlação; e

4) Calcular o novo valor de disparidade e corrigir o mapa de disparidade interpolado com este novo valor.

Esta seqüência de passos corrige, em uma primeira fase desse processo, os valores interpolados do mapa de disparidade apenas para os pontos no espaço ciclopeano oriundos do *matching* baseado em feições.

A segunda fase do processo de correção do mapa de disparidade é feita de modo semelhante seguindo os mesmos passos da primeira fase, apenas com a diferença que o gradiente de disparidade agora é calculado para os pontos nas vizinhanças dos pontos iniciais no mapa de disparidade parcialmente corrigido.

Este processo é iterativo e cada valor de disparidade corrigido para um ponto do mapa de disparidade parcialmente corrigido determina uma vizinhança e portanto, os próximos pontos cujas disparidades serão corrigidas. Entretanto, como geralmente existe mais de um ponto inicial é necessário determinar uma ordem de prioridade para quais pontos devam ser corrigidos primeiramente.

A ordem de prioridade de correção dos pontos do mapa de disparidade é baseada no conceito de distância. Esta distância é determinada entre um dado ponto do mapa de disparidade e o ponto mais próximo pertencente ao conjunto dos pontos iniciais oriundos do *matching* baseado em feições contada sobre alguma métrica adotada.

Para implementar este processo iterativo utilizou-se um procedimento análogo ao de dilatação geodésica.

A dilatação geodésica foi desenvolvido a partir dos conceitos de morfologia matemática e as referências HARALICK *et al.* (1987); MEYER & BEUCHER (1990); FACON (1996); e LOTUFO & TRETTEL (1996) são sugeridas sobre este assunto.

O cálculo desta distância, entre um dado ponto e o ponto mais próximo pertencente ao conjunto dos pontos iniciais, é uma tarefa extremamente complexa se esta fosse implementada para o caso contínuo, entretanto, no caso discreto esta tarefa é simples.

Inicialmente gera-se uma imagem com o mesmo tamanho da imagem ciclopeana como todos seus valores nulos. Esta imagem foi denominada de imagem de controle.

Para as mesmas coordenadas dos pontos iniciais no espaço ciclopeano, a imagem de controle recebeu valores iguais a 1 (um). A imagem de controle é então "varrida" ao longo de suas linhas da esquerda para direita e de cima para baixo. A cada posição que for encontrado o valor 1, os vizinhos diretos desta posição recebem o valor 2. Em uma segunda vez que a imagem de controle é varrida, agora para cada posição que tiver o valor 2, os vizinhos diretos desta posição recebem o valor 3 e este processo continua enquanto não houver mais nenhuma posição da imagem de controle com valor zero.

A figura 4.4 mostra este processo para uma imagem de controle de tamanho 7 x 7, cuja a métrica utilizada foi a vizinhança quatro (forma de cruz).

A imagem superior esquerda da figura 4.4 mostra a primeira iteração do processo, onde as posições que receberam o valor 1, são as posições dos pontos iniciais. A imagem superior direita desta figura mostra a segunda iteração e a imagem inferior esquerda mostra a terceira iteração deste processo. Por fim, a imagem inferior direita da figura 4.4, mostra o resultado final. Ainda nestas imagens nota-se o conceito de vizinhança em torno de um ponto que pode ser adotado utilizando esta metodologia.

0	0	0	0	0	0	1
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	1	0
0	0	0	0	0	0	0

0	0	0	0	0	2	1
0	0	2	0	0	0	2
0	2	1	2	0	0	0
0	0	2	0	0	0	0
0	0	0	0	0	2	0
0	0	0	0	2	1	2
0	0	0	0	0	2	0

0	0	3	0	3	2	1
0	3	2	3	0	3	2
3	2	1	2	3	0	3
0	3	2	3	0	3	0
0	0	3	0	3	2	3
0	0	0	3	2	1	2
0	0	0	0	3	2	0

5	4	3	4	3	2	1
4	3	2	3	4	3	2
3	2	1	2	3	4	3
4	3	2	3	4	3	4
5	4	3	4	3	2	3
6	5	4	3	2	1	2
7	6	5	4	3	2	3

Fig. 4.4 - Evolução do processo iterativo para a imagem de controle.

Esta imagem de controle foi gerada utilizando uma métrica de vizinhança quatro, mas outras métricas podem ser utilizadas, tais como vizinhança oito, disco digital de um raio r qualquer ou outra métrica que seja de interesse para um problema específico.

Com a imagem de controle determinada, o processo de correção do mapa de disparidade parcialmente corrigido em uma iteração n se dá através dos seguintes passos:

1) Calcular o gradiente de disparidade através do mapa de disparidade parcialmente corrigido para os pontos de distância n na imagem de controle;

2) Determinar o intervalo permitido de busca do *pixel* homólogo;

3) Realizar o *matching* baseado em áreas no intervalo de busca permitido e adotar como *pixel* homólogo o de maior coeficiente de correlação; e

4) Calcular o novo valor de disparidade e corrigir o mapa de disparidade parcialmente corrigido com este novo valor.

5) Voltar ao passo 1 se n é menor ou igual ao maior valor da imagem de controle.

Um aspecto interessante desta metodologia, é que o gradiente de disparidade é determinado no mapa de disparidade interpolado apenas na primeira iteração do processo de correção para os pontos iniciais, onde a interpolação é bem realizada. Para as demais iterações, o gradiente de disparidade é calculado no mapa de disparidade parcialmente corrigido, assim o gradiente de disparidade é determinado não mais em um mapa de disparidade suave, mas em um mapa de disparidade que permite variações abruptas de disparidade.

É importante ressaltar que esta metodologia não propaga a disparidade associada a um *pixel* na imagem ciclopeana, como sugere a evidência biológica de processos cooperativos proposta por MARR & POGGIO (1976) e citada em 3.4.3 e 3.5.4, e sim um intervalo permitido de busca, que se adapta no decorrer do processo, permitindo variações abruptas nos valores de disparidade dentro de um certo limite, estando de acordo com a evidência biológica do gradiente de disparidade proposta por BURT & JULESZ (1980).

Este limite é o limite de gradiente de disparidade, e quando este é excedido o fenômeno de diplopia ocorre e a fusão não é mais possível, sendo necessário que uma nova área de atenção seja determinada através do movimento de vergência para reiniciar este processo.

O fenômeno de diplopia é representado computacionalmente não por “enxergar” duplo um mesmo objeto, mas sim pela determinação incorreta do ponto homólogo.

Através do mapa de disparidade corrigido, a determinação das coordenadas tridimensionais no espaço-objeto é uma tarefa trivial e consiste apenas em utilizar os modelos de reconstrução tridimensional citados em 2.7.2.

Cabe ainda ressaltar, que esta metodologia não utiliza nenhuma das premissas empíricas da visão estéreo, citadas em 3.3.1, 3.3.2 e 3.3.3, com exceção do *threshold* de oclusão, citada em 4.5, que pode ser entendido como uma premissa empírica.

Capítulo 5

Testes e resultados

5. Introdução

Este capítulo apresenta os testes realizados e resultados obtidos para três pares de imagens estereoscópicas retificadas, sendo dois desses pares de imagens sintéticas, e um par de imagens reais.

Os testes foram realizados com o objetivo de determinar o desempenho do *matching* de regiões, *matching* de feições retas e geração do mapa de disparidade. Todas as imagens utilizadas são de dimensões de 320 colunas por 240 linhas (320 x 240).

Os testes realizados com o objetivo de determinar o desempenho da geração do mapa de disparidade utilizaram o módulo do gradiente de disparidade (sem informação do sentido da variação da disparidade), e o valor direto deste gradiente (considerando o sentido da variação da disparidade).

É importante lembrar que todas as imagens utilizadas são retificadas para o processo estéreo, e portanto, a disparidade se relaciona com a profundidade de maneira inversamente proporcional (expressão 2.60). Este fato possibilita que as análises sejam realizadas apenas em relação a disparidade, não sendo necessário a reconstrução dos pontos no referencial do espaço-objeto.

5.1 Imagens estereoscópicas

Este item apresenta as imagens estereoscópicas utilizadas nos testes. As imagens sintéticas foram geradas a partir do programa de renderização POV-RAY. Este programa é baseado em arquiteturas *ray tracing* para renderização de imagens sintéticas.

As imagens reais foram adquiridas no Laboratório de Visão Computacional do Centro Tecnológico para Informática - CTI.

As imagens sintéticas constituem a mesma cena, porém, o par de imagens 1 foi gerado com comprimento da base do dobro do comprimento da base do par de imagens 2, tendo em visto analisar a influência deste parâmetro de configuração de câmeras no processo estéreo.

As imagens sintéticas foram geradas com configurações de câmeras ditas ideais para o processo estéreo, e portanto, não necessitaram ser retificadas.

As imagens reais foram tomadas com os parâmetros de calibração das câmeras próximos dos parâmetros de calibração ditos ideais. Estes parâmetros foram determinados utilizando o programa denominado CCESC, desenvolvido pelo engenheiro Maurício Galo. Estes parâmetros mostraram-se bastante próximos dos parâmetros que resultariam em uma configuração ideal ao ponto de não ser necessária a realização da retificação estéreo. Para confirmar este fato, estas imagens foram checadas manualmente e observou-se que as linhas epipolares eram paralelas as linhas das imagens.

As operações de segmentação de regiões homogêneas foram feitas utilizando o programa KHOROS 2.0.

Par de imagens 1

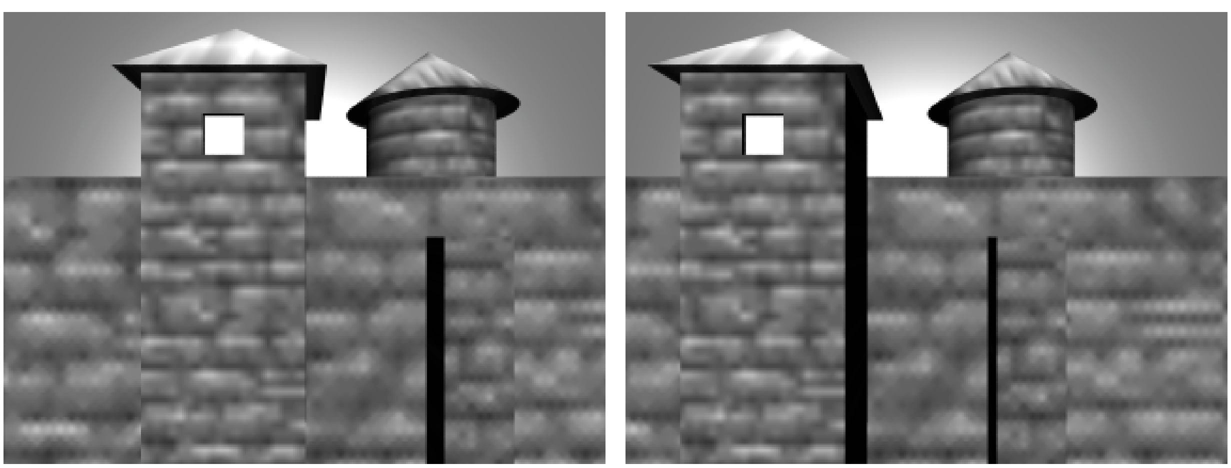


Fig. 5.1 - Par de imagens 1.

Estas imagens foram modeladas matematicamente e então renderizadas. A textura das paredes das torres, do muro, do pilar à frente do muro e dos telhados das torres foram extraídas de imagens reais, assim, estas imagens possuem modelos geométricos rígidos e conhecidos e texturas reais, que tornam a cena mais realista.

A região clara que escure gradativamente em direção aos limites destas imagens foi introduzida propositadamente para simular efeitos de iluminação que ocorrem em imagens reais.

A imagem seguinte ilustra a mesma cena das imagens 1, porém de um ponto de vista acima das torres, para melhor visualização das profundidades dos objetos presentes na cena.

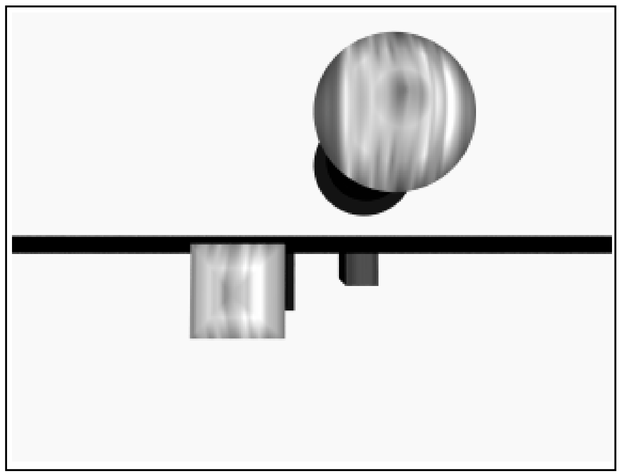


Fig. 5.2 - Cena das imagens 1 a partir de um ponto de vista acima das torres.

Esta cena possui algumas regiões de variações abruptas de profundidade e consequentemente, variações abruptas de disparidade.

Par de imagens 2

A exemplo do par de imagens 1, o par de imagens 2 representam a mesma cena, porém o comprimento da base (distância entre as câmeras) utilizado para a geração destas imagens é metade do comprimento da base utilizada para gerar as imagens 1.

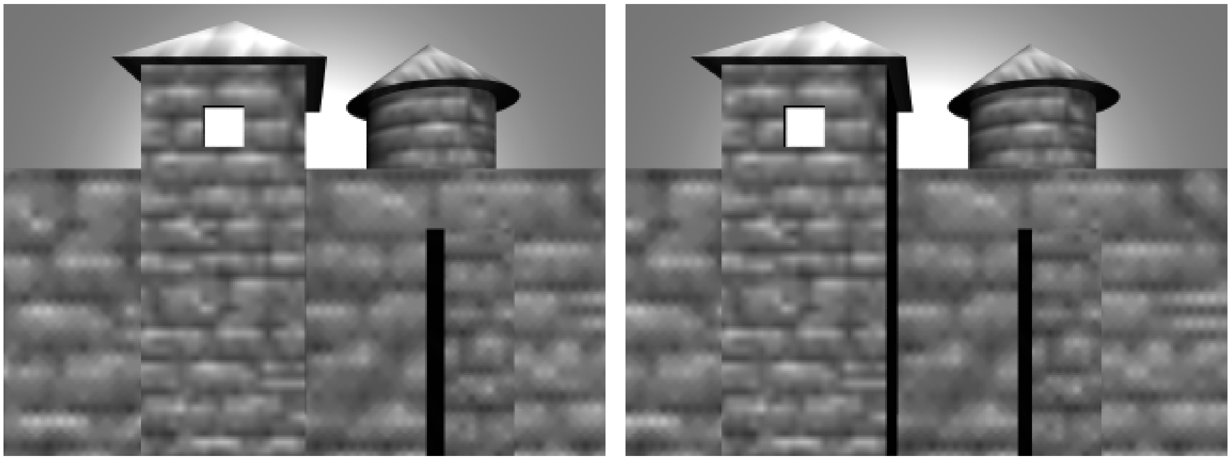


Fig. 5.3 - Par de imagens 2.

Para a geração destas imagens, a posição da câmera da imagem da esquerda manteve-se fixa em relação à imagem da esquerda do par de imagens 1, porém a posição da câmera da imagem da direita sofreu uma translação na direção X de metade do comprimento da base utilizada para gerar o par de imagens 1.

Analisando os pares de imagens 1 e 2, percebe-se claramente que as áreas de oclusão do par de imagens 2 são menores em relação as áreas de oclusão do par de imagens 1. Com maior dificuldade, é possível perceber que as distorções geométricas entre as imagens estereoscópicas do par de imagens 2 são também menores que as distorções geométricas entre as imagens estereoscópicas do par de imagens 1.

Par de imagens 3

As imagens 3 constituem uma cena de uma garrafa e uma jarra. Devido a textura desses objetos ser muito homogênea, colocou-se fitas adesivas branca para alterar a homogeneidade da textura destes.

Fig. 5.4 - Par de imagens 3.

Para a garrafa colocou-se pequenos pedaços de fita adesiva branca e para a jarra pedaços de fita branca maiores formando um reticulado.

Existem nestas imagens, fortes distorções radiométricas e geométricas. Facilmente é possível perceber as distorções radiométricas entre estas imagem reparando por exemplo na garrafa, que é mais escura na imagem da esquerda de que na imagem da direita.

5.2 Testes e resultados do *matching* de feições retas

As feições retas foram determinadas através da transformada de Hough e então realizado o *matching* de feições retas no espaço de atributos para as feições que atendiam as restrições impostas, citadas em 4.1.3.

Par de imagens 1

A figura 5.5 mostra as imagens de bordas de intensidade já binarizada referentes ao par de imagens 1. As imagens de bordas foram geradas usando os operadores de Sobel e então calculado a magnitude do gradiente de intensidade para *pixel*. Estes valores foram normalizados no intervalo de 0 a 255 e então aplicado um *threshold* cujo valor de corte adotado foi de 128.

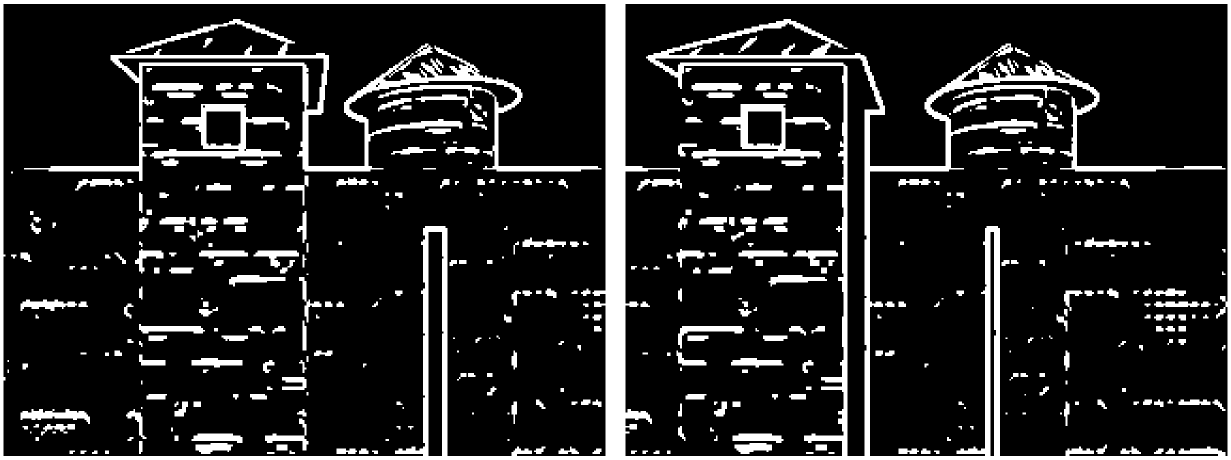


Fig. 5.5 - Imagens de bordas de intensidade binarizadas referentes ao par de imagens 1.

A transformada de Hough determinou 16 retas na imagem da esquerda e 14 retas na imagem da direita. O *threshold* utilizado no espaço de Hough foi de 10^0 em θ e/ou 10 *pixels* em ρ, para duas ou mais retas na mesma imagem e de comprimento maior que 10 *pixels* (ocorrência no espaço de Hough), assim as retas determinadas apresentaram-se mais separadas, evitando duas ou mais retas muito próximas.

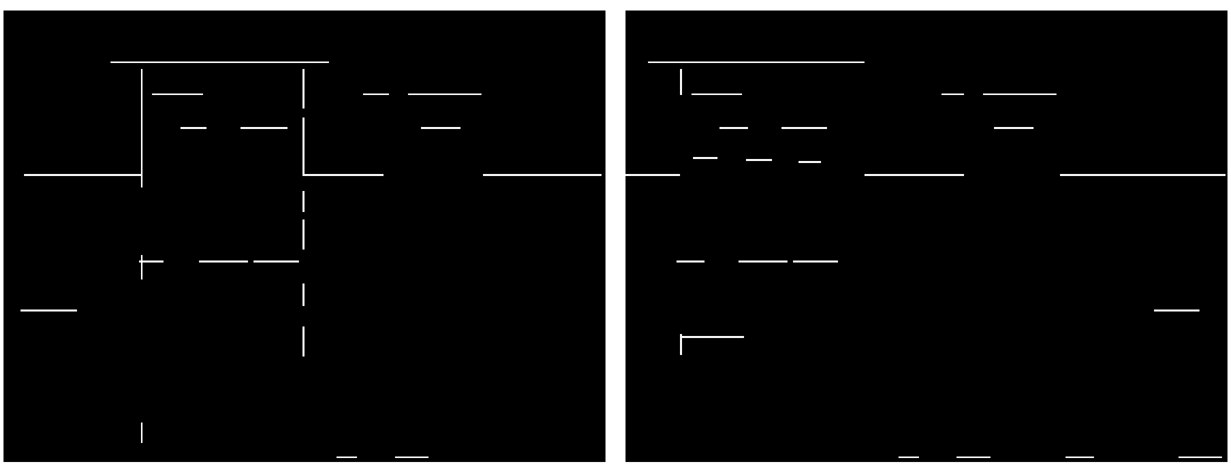


Fig. 5.6 - Imagens de feições retas referentes ao par de imagens 1.

O número de feições retas determinadas é relativamente pequeno, mas observando as imagens binarizadas da figura 5.5, percebe-se que as feições retas apresentam quase sempre falhas, dificultando a determinação destas através da transformada de Hough.

A figura 5.7 mostra as imagens de feições retas homólogas determinadas através do *matching* de feições retas.

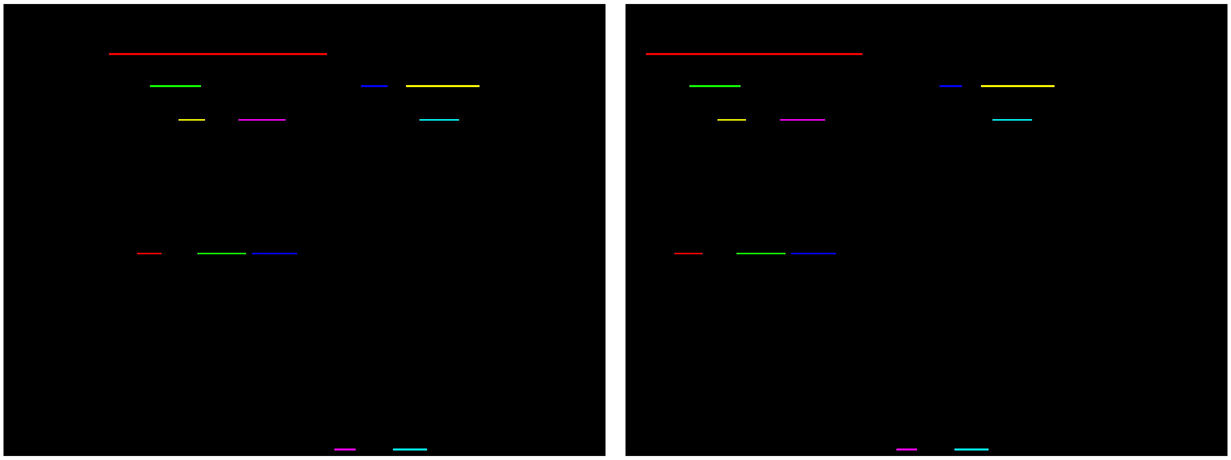


Fig. 5.7 - Imagens de retas homólogas referentes ao par de imagens 1.

Para melhor visualização, utilizou-se uma representação colorida, sendo que cada par de feições retas entre as imagens são de mesma cor e ocupam aproximadamente a mesma linha epipolar.

Na realização do *matching* de feições retas, utilizou-se um valor de *threshold* de 30^0 em relação a seletividade direcional e um intervalo de 3 *pixels* de discrepância em relação a geometria epipolar, assim uma reta em uma imagem pode se corresponder com uma outra reta na outra imagem mesmo que a diferença entre suas coordenadas na direção y não exceda o valor de 3 *pixels*.

Com o intuito de restringir ainda mais o *matching*, utilizou-se a evidência biológica de compatibilidade (polaridade).

Como resultado final desse processo, determinou-se apenas 9 feições retas homólogas, porém, com a grande vantagem que nenhuma feição reta foi correspondida erroneamente.

Par de imagens 2

Para as imagens 2, as imagens de bordas de intensidade binarizadas foram geradas utilizando os mesmo critérios para esta mesma tarefa das imagens 1, porém com um valor de *threshold* igual a 100 (para testar diferentes valores de *threshold*). A figura 5.8 mostras estas imagens.

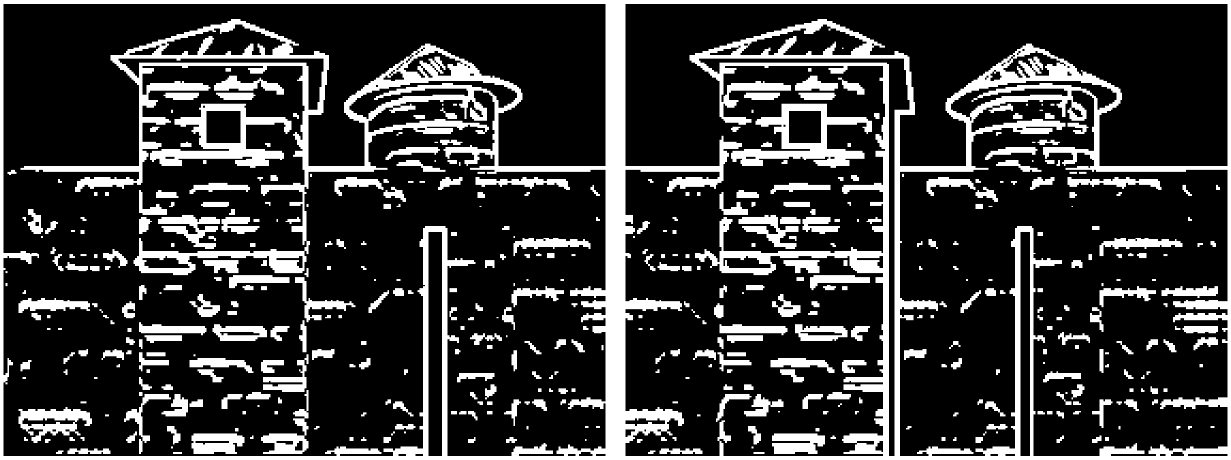


Fig. 5.8 - Imagens de bordas de intensidade binarizadas referentes ao par de imagens 2.

Para estas imagens, a transformada de Hough determinou 49 retas na imagem da esquerda e 27 retas na imagem da direita. Os *thresholds* utilizados no espaço de Hough foram os mesmos do par de imagens 1, ou seja, 10^0 em θ e/ou 10 *pixels* em ρ, para duas ou mais retas na mesma imagem e de comprimento maior que 10 *pixels*.

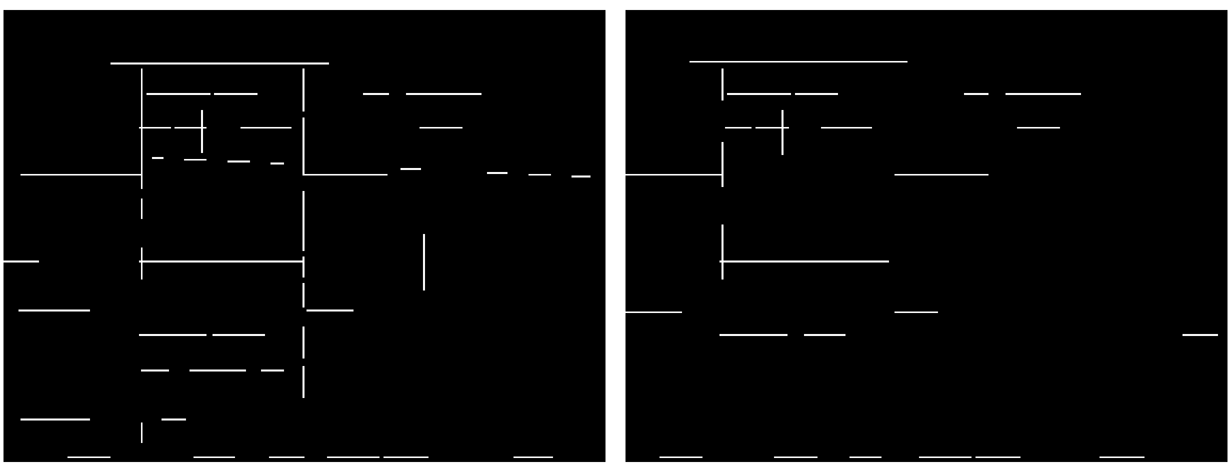


Fig. 5.9 - Imagens de feições retas referentes ao par de imagens 2.

A figura 5.10 mostra as imagens de feições retas homólogas determinadas através do *matching* de feições retas para as imagens 2.

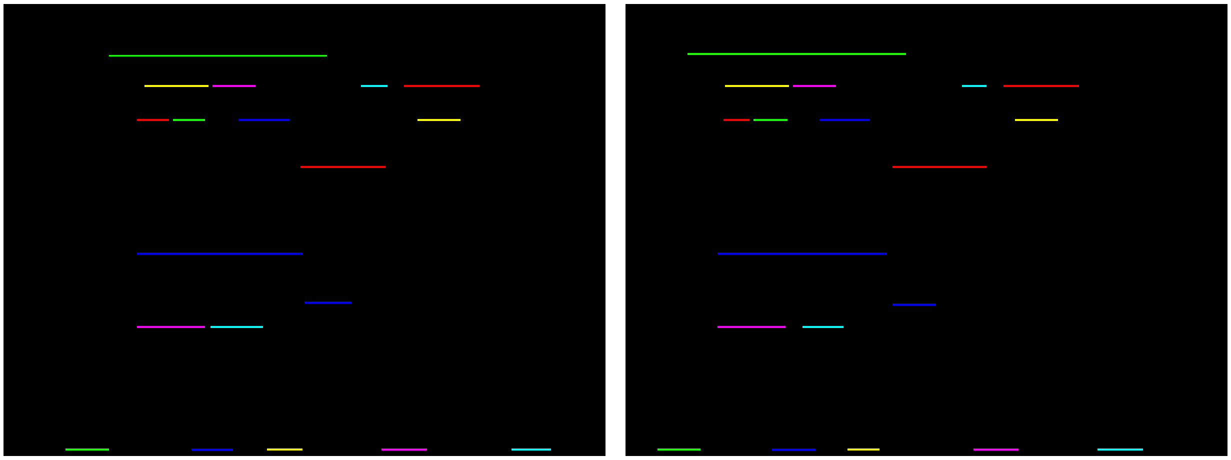


Fig. 5.10 - Imagens de retas homólogas referentes ao par de imagens 2.

Na realização do *matching* entre as retas utilizou-se os mesmos valores de *threshold* em relação a seletividade direcional e intervalo de discrepância em relação a geometria epipolar utilizados para as imagens 1. Utilizou-se também a restrição de compatibilidade.

Como resultado final desse processo, determinou-se apenas 19 feições retas homólogas, porém, nenhuma feição reta foi correspondida erroneamente.

Par de imagens 3

Nas imagens 3, as imagens de bordas de intensidade binarizadas foram geradas utilizando valores distintos para o *threshold*. A imagem de bordas de intensidade da esquerda foi binarizada com um valor de *threshold* igual a 112 e a da direita com um valor de *threshold* igual a 84. A figura 5.11 mostras estas imagens.

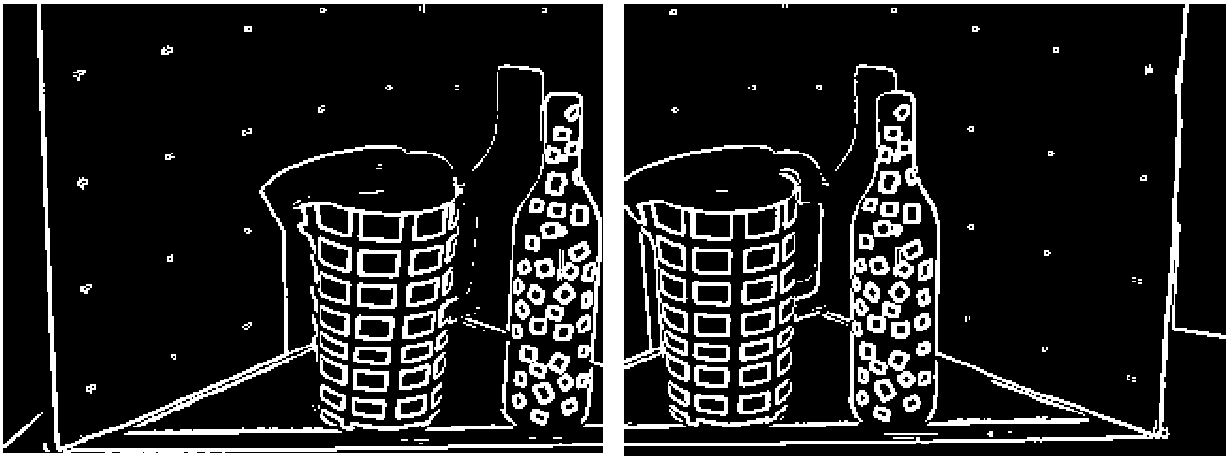


Fig. 5.11 - Imagens de bordas de intensidade binarizadas referentes ao par de imagens 3.

Nestas imagens, a transformada de Hough determinou 98 retas na imagem da esquerda e 132 retas na imagem da direita. Os *thresholds* utilizados no espaço de Hough foram 3^0 em θ e/ou 2 *pixels* em ρ, para duas ou mais retas na mesma imagem e de comprimento maior que 2 *pixels*.

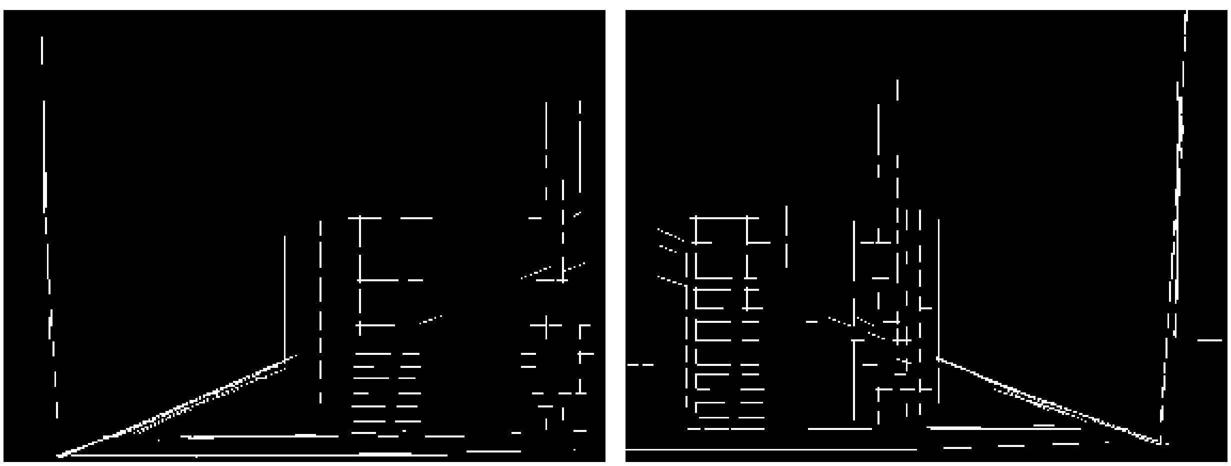


Fig. 12 - Imagens de feições retas referentes ao par de imagens 3.

A figura 5.13 mostra as imagens de feições retas homólogas determinadas através do *matching* de feições retas para as imagens 3.

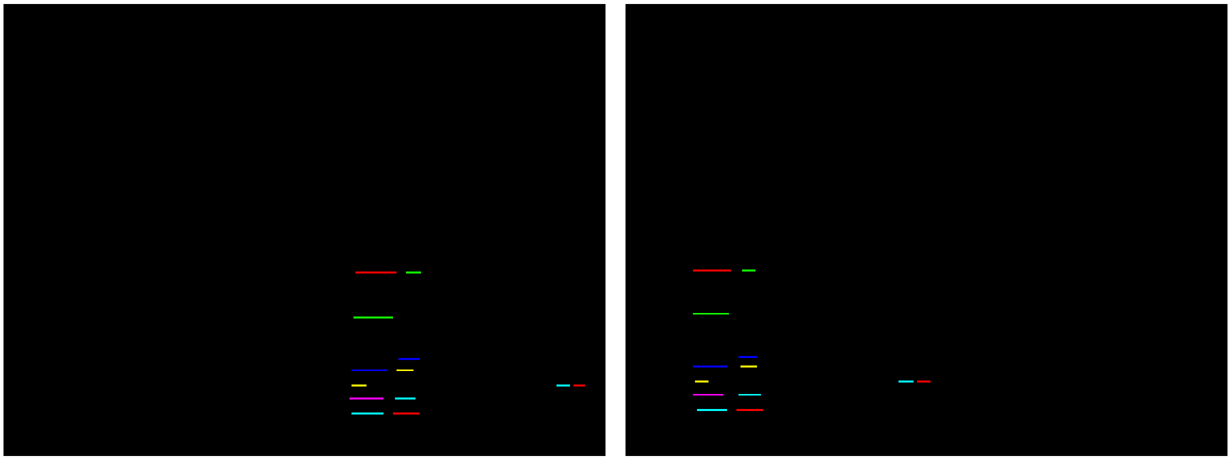


Fig. 5.13 - Imagens de retas homólogas referentes ao par de imagens 3.

Na realização do *matching* entre as retas utilizou-se um valor de *threshold* de 30^0 em relação a seletividade direcional e um intervalo de 3 *pixels* de discrepância em relação a geometria epipolar. A restrição de compatibilidade foi também utilizada.

Como resultado final desse processo, determinou-se apenas 13 feições retas homólogas, porém nenhuma feição reta foi correspondida erroneamente.

Todas as restrições impostas ao processo de *matching* de feições retas, fez com que o número de feições retas homólogas fosse pequeno, porém, a probabilidade dessas feições serem realmente homólogas é grande, haja visto que para os três pares de imagens estereoscópicas, nenhuma feição reta homóloga foi determinada erroneamente.

5.3 Testes e resultados do *matching* de regiões

As regiões relativamente homogêneas nas imagens estereoscópicas foram segmentadas seguindo a metodologia descrita em 4.1.2 e então calculado os sete momentos invariantes para cada região segmentada referente as imagens estereoscópicas. Os sete momentos invariantes foram calculados levando em consideração os valores de intensidade das imagens.

Para as regiões que atendiam as restrições impostas citadas em 4.1.4, o *matching* de regiões foi então realizado no espaço de atributos.

Par de imagens 1

A figura 5.14 mostra as imagens binarizadas após sofrerem a operação de *threshold* no intervalo de 128 a 160.

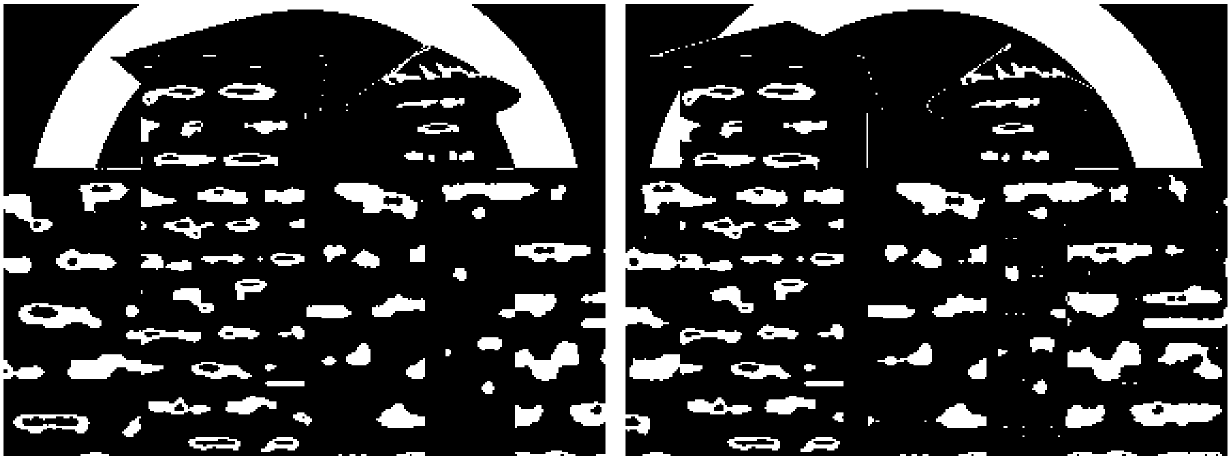


Fig. 5.14 - Imagens binarizadas referentes ao par de imagens 1.

A figura 5.15 mostra as imagens da figura 5.14 após a operação morfológica de fechamento de buracos (*close holes*). O elemento estruturante planar utilizado foi um disco digital de raio 3 sob a métrica euclidiana.

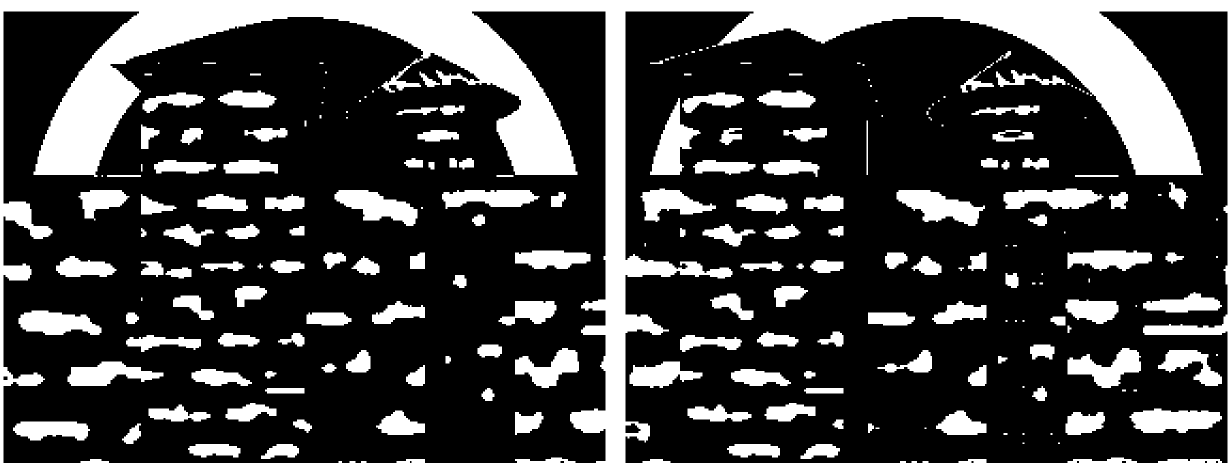


Fig. 5.15 - Imagens binarizadas após fechamento de buracos referentes ao par de imagens 1.

A figura 5.16 ilustra as imagens da figura 5.15 após a operação morfológica de abertura. O elemento estruturante planar utilizado nesta operação também foi um disco digital de raio 3 sob a métrica euclidiana.

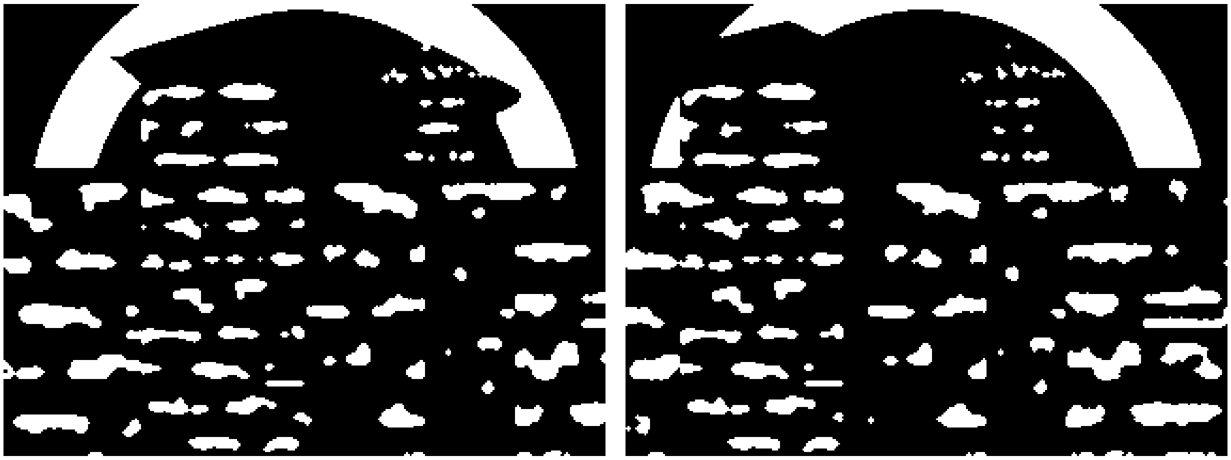


Fig. 5.16 - Imagens binarizadas após abertura morfológica referentes ao par de imagens 1.

As regiões das imagens da figura 5.16 foram então rotuladas. Devido a rotulação ser um processo de contagem de regiões, o número de regiões segmentadas em cada imagem é o valor mais alto de intensidade das imagens rotuladas. Neste processo gerou-se 89 regiões para a imagem esquerda e 94 regiões para a imagem da direita.

Os momentos foram calculados levando em consideração também os valores de intensidade das regiões. A figura 5.17 mostra as imagens das figuras 5.16 com suas respectivas intensidades originais.

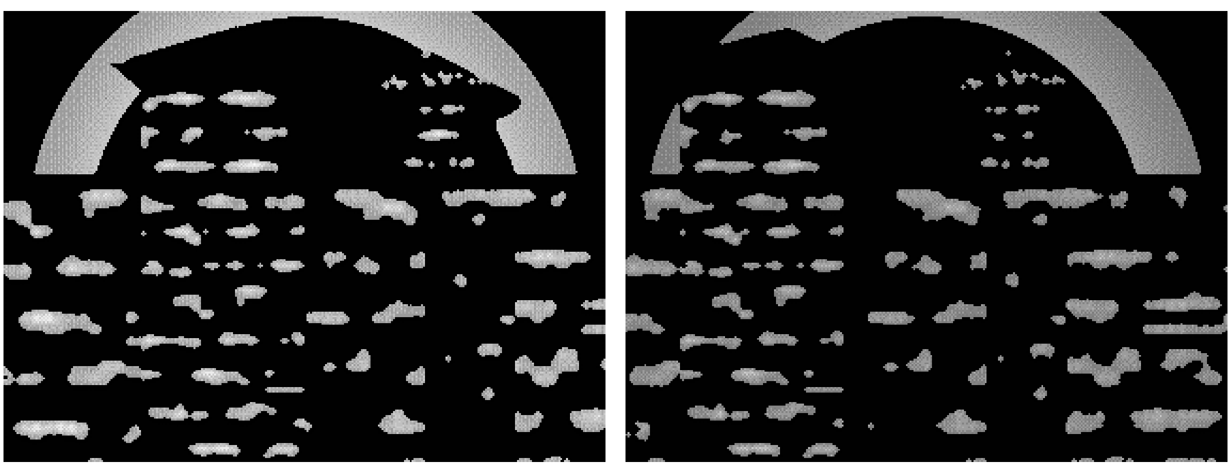


Fig. 5.17 - Imagens segmentadas em regiões com suas intensidades originais referentes ao par de imagens 1.

Após todos esses passos, os sete momentos invariantes foram determinados para cada uma das regiões das imagens da figura 5.17 e então realizado o *matching* de regiões no espaço de atributos associado a cada região, respeitando as restrições impostas a este tipo de *matching*, como citado em 4.1.4.

As regiões segmentadas que se situavam nas bordas das imagens não foram levadas em consideração no processo de *matching* devido as suas áreas serem afetadas por esses limites.

Na realização do *matching* entre as regiões, utilizou-se um valor de intervalo de 3 *pixels* de discrepância em relação a geometria epipolar.

As regiões coloridas das imagens da figura 5.18 são regiões homólogas para a mesma cor e aproximadamente mesma linha epipolar. As regiões brancas não foram correspondidas. O processo de *matching* determinou 59 regiões homólogas.

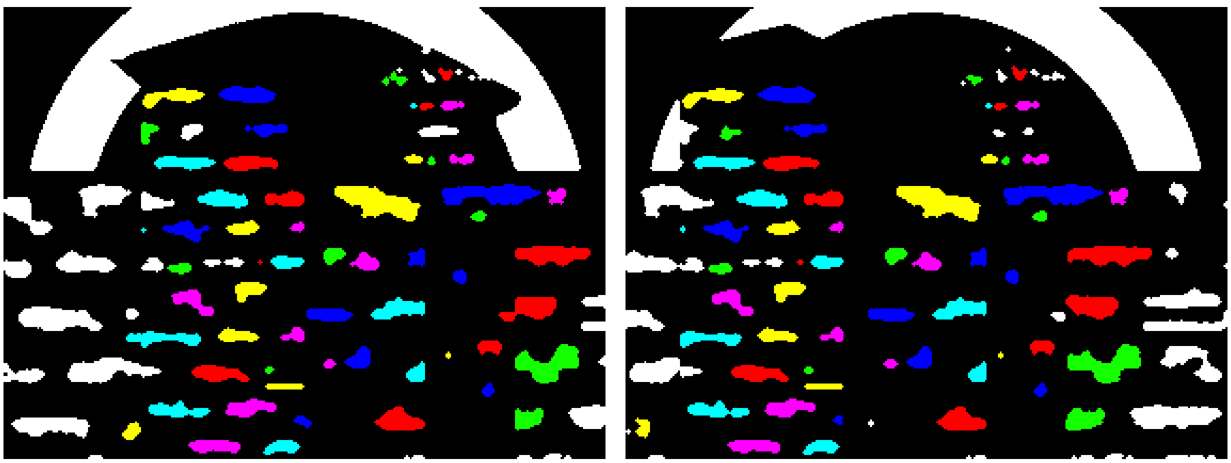


Fig. 5.18 - Imagens de regiões homólogas referentes ao par de imagens 1.

Das 59 regiões determinadas pelo processo como regiões homólogas, apenas duas regiões foram correspondidas erroneamente. A primeira destas regiões está representada pela cor verde, próximo ao canto esquerdo superior e a segunda destas, está representada pela cor amarelo, próximo ao canto esquerdo inferior das imagens.

Outras regiões em que houveram problemas foram as regiões representadas pela cor vermelho no centro direito das imagens. Na imagem da esquerda esta região é única e na imagem da direita esta região aparece separada de uma pequena região representada em branco. Porém, o processo de *matching* determinou mesmo assim essas duas regiões como homólogas. Este fato foi benéfico do ponto de vista que essa fase de *matching* de regiões visa obter alguns pontos homólogos para controlar o movimento de vergência para uma mesma área de atenção, e portanto esta tarefa foi atendida.

Par de imagens 2

Devido as imagens 2 serem parecidas ao par de imagens 1, todas as imagens geradas no processo de segmentação não serão apresentadas, mesmo porque, os mesmo parâmetros utilizados para a segmentação de regiões no par de imagens 1 foram utilizados para estas imagens.

A figura 5.19 ilustra as imagens já segmentadas em regiões com suas respectivas intensidades.

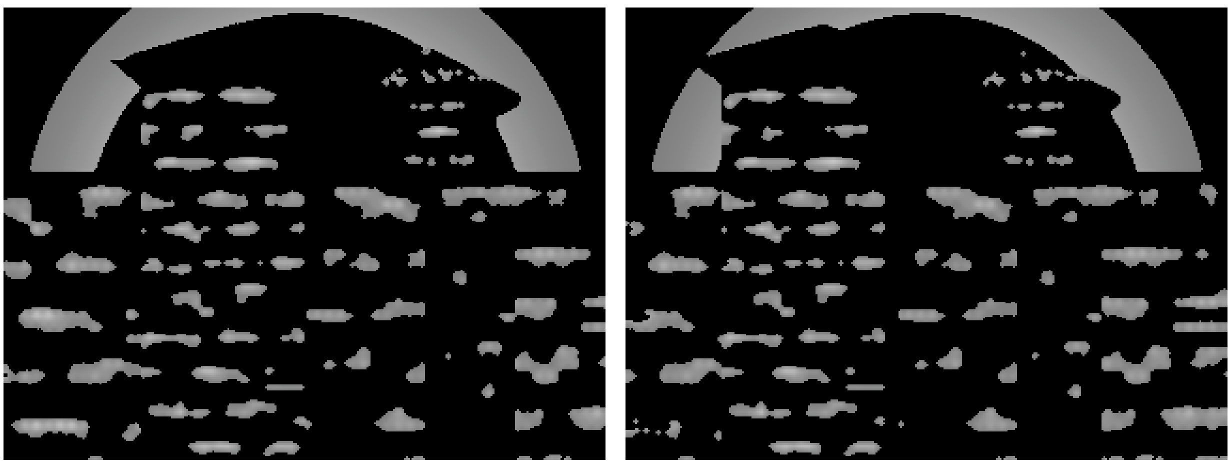


Fig. 5.19 - Imagens segmentadas em regiões com suas intensidades originais referentes ao par de imagens 2.

A realização do *matching* de regiões também utilizou um valor de intervalo de 3 *pixels* de discrepância em relação a geometria epipolar.

As regiões coloridas das imagens da figura 5.20 são regiões homólogas para a mesma cor e aproximadamente mesma linha epipolar. As regiões brancas não foram correspondidas. O processo de *matching* determinou 63 regiões homólogas a partir de 89 regiões na imagem da esquerda e 96 regiões na imagem da direita.

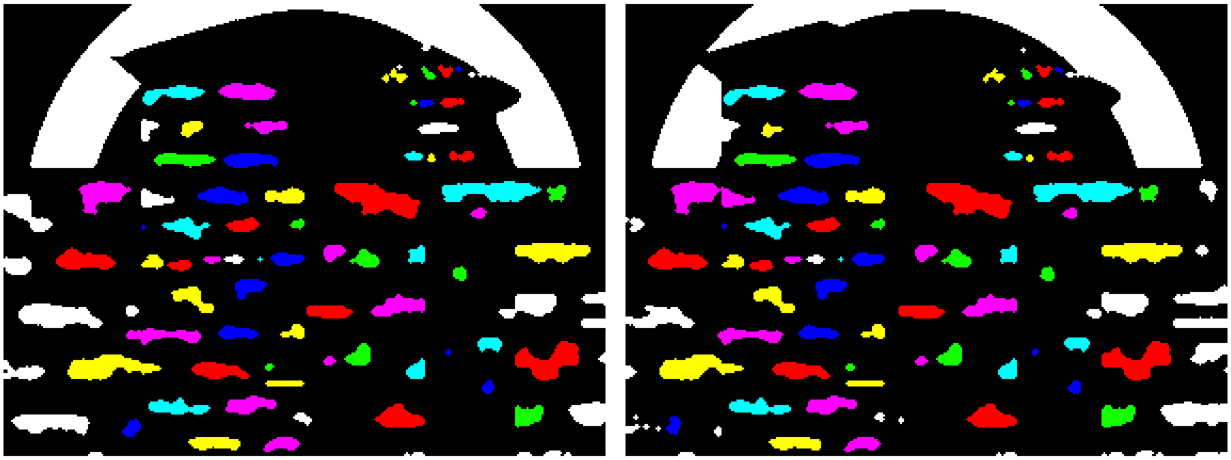


Fig. 5.20 - Imagens de regiões homólogas referentes ao par de imagens 2.

Das 63 regiões determinadas como regiões homólogas pelo *matching* de regiões, apenas um par foi erroneamente determinado. Estas regiões estão representadas pela cor azul no canto inferior esquerdo. Um região em que a determinação de sua homóloga foi parcialmente errada está representada em rosa no centro esquerdo das imagens. Na imagem da esquerda esta região é única e na imagem da direita esta região aparece unida a outra região.

Par de imagens 3

A figura 5.21 mostra as imagens binarizadas após sofrerem a operação de *threshold* cujo o valor de corte para a imagem da esquerda foi de 150 e para a imagem da direita de 182.

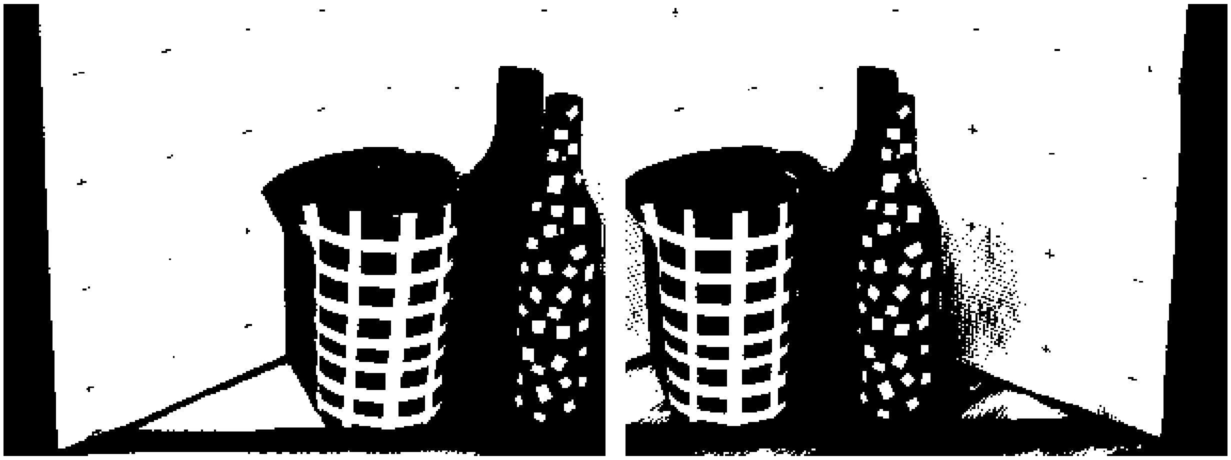


Fig. 5.21 - Imagens binarizadas referentes ao par de imagens 3.

A figura 5.22 mostra as imagens da figura 5.21 após a operação morfológica de fechamento de buracos (*close holes*). O elemento estruturante planar utilizado também foi um disco digital de raio 3 sobre a métrica euclidiana.

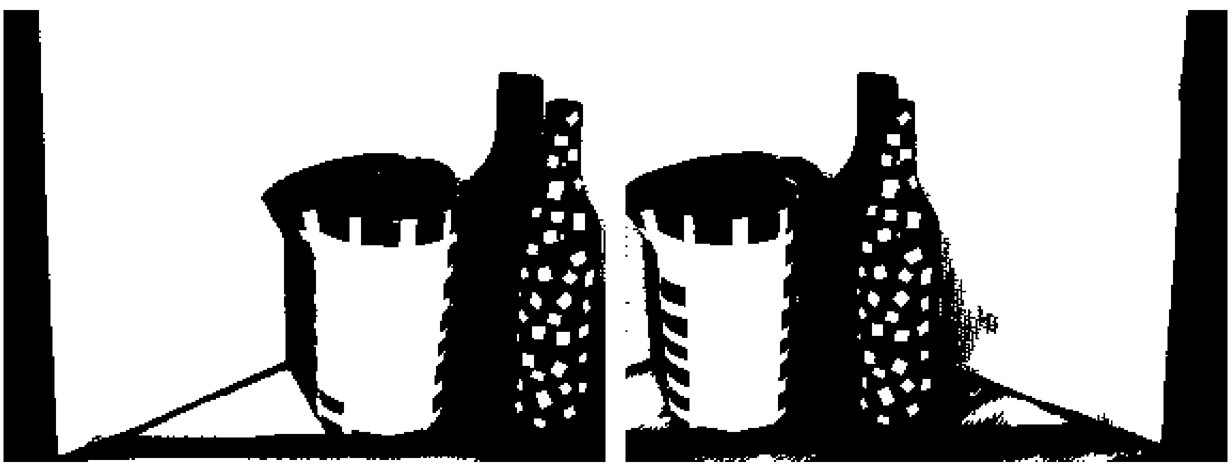


Fig. 5.22 - Imagens binarizadas após fechamento de buracos referentes ao par de imagens 3.

A figura 5.23 ilustra as imagens da figura 5.22 após a operação morfológica de abertura. O elemento estruturante planar utilizado nesta operação também foi um disco digital de raio 3 sob a métrica euclidiana.

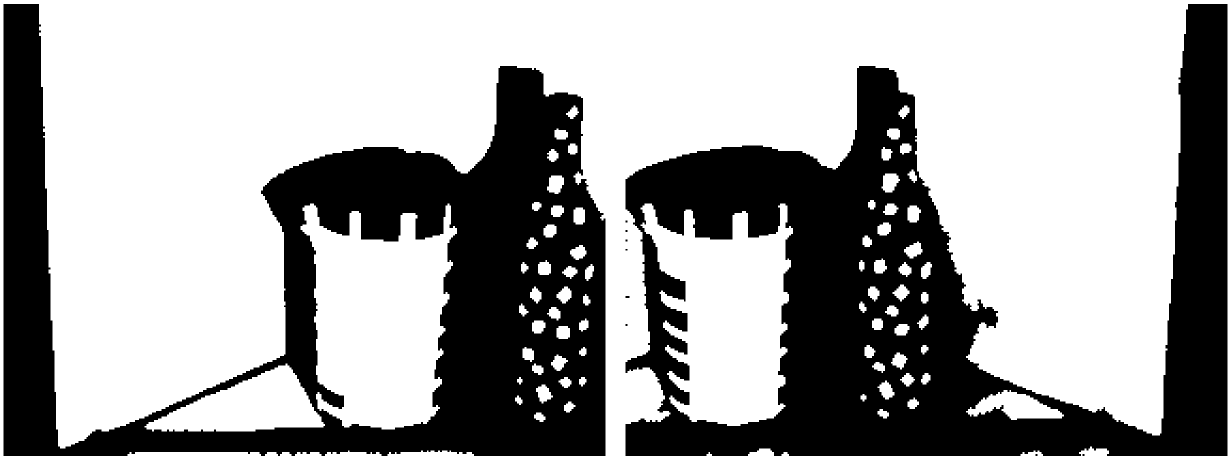


Fig. 5.23 - Imagens binarizadas após abertura morfológica referentes ao par de imagens 3.

As regiões das imagens da figura 5.23 foram então rotuladas e determinou-se 43 regiões na imagem da esquerda e 44 regiões na imagem da direita.

A figura 5.24 mostra as imagens da figura 5.23 com suas respectivas intensidades.

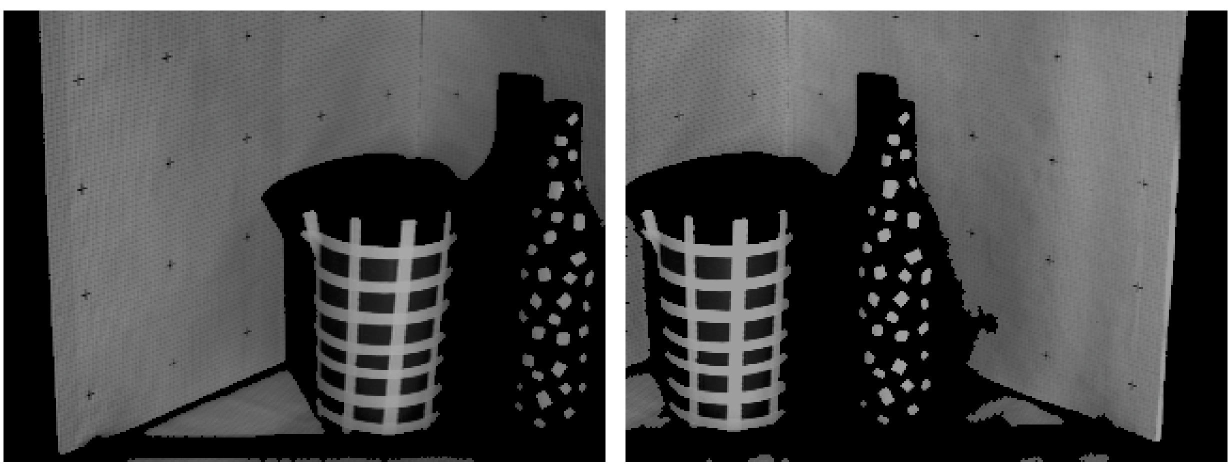


Fig. 5.24 - Imagens segmentadas em regiões com suas intensidades originais referentes ao par de imagens 3.

Finalmente, a figura 5.25 mostra as regiões homólogas determinadas pelo *matching* de regiões.

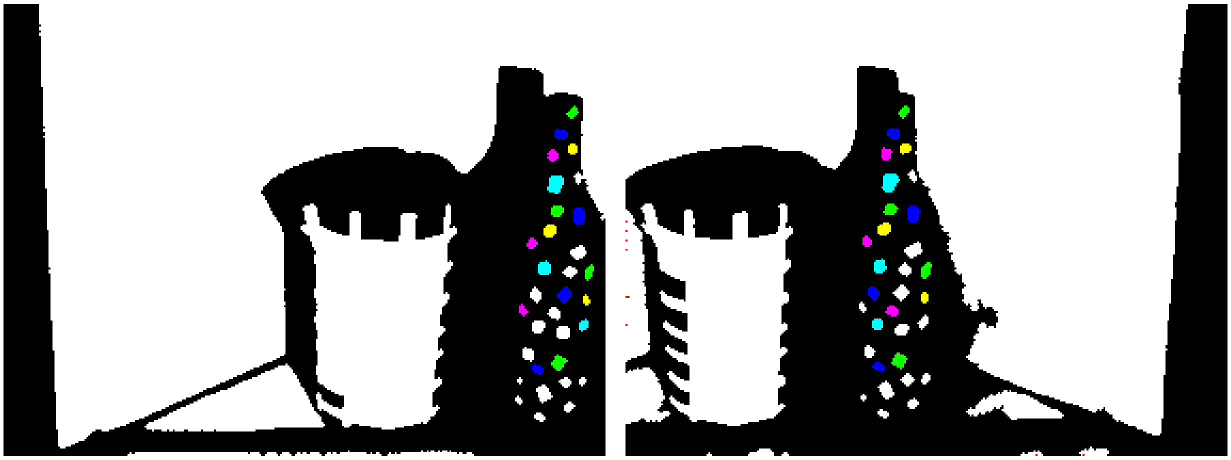


Fig. 5.25 - Imagens de regiões homólogas referentes ao par de imagens 3.

A realização do *matching* de regiões também utilizou um valor de intervalo de 3 *pixels* de discrepância em relação a geometria epipolar.

As regiões coloridas das imagens da figura 5.25 são regiões homólogas para a mesma cor e aproximadamente mesma linha epipolar. As regiões brancas não foram correspondidas. O processo de *matching* determinou 19 regiões homólogas.

Das 19 regiões determinadas como regiões homólogas pelo *matching* de regiões, 3 destas foram determinadas erroneamente. Estas regiões estão representadas pela cor azul no centro da garrafa, pela cor rosa logo abaixo desta região em azul e pela cor azul claro, logo abaixo desta regiões em rosa.

5.4 Testes e resultados da geração do mapa de disparidade

Uma lista de coordenadas de *pixels* homólogos foi gerada a partir do *matching* baseado em feições, possibilitando o controle do movimento de vergência para as áreas de atenção da cena.

O primeiro teste deste item, utiliza o módulo do gradiente de disparidade para a determinação do intervalo de busca do *pixel* homólogo. O segundo teste deste item, utiliza a informação do sentido do gradiente de disparidade para a determinação do intervalo de busca do *pixel* homólogo.

5.4.1 Teste 1 (módulo do gradiente de disparidade para determinação do intervalo de busca)

Par de imagens 1

Os parâmetros utilizados neste teste, para o par de imagens 1 foram:

-Dimensão na direção x da janela de correlação = 7 *pixels*;
-Dimensão na direção y da janela de correlação = 3 *pixels*;
-*Threshold* de oclusão = 10 níveis de intensidade; e
-Fator de escala para o gradiente de disparidade = 1.5.

A figura 5.26, mostra o mapa de disparidade interpolado a partir da lista de *pixels* homólogos oriundos do *matching* de feições, cujas coordenadas foram transformadas para o espaço ciclopeano.

Fig. 5.26 - Mapa de disparidade interpolado referente ao par de imagens 1 para o teste 1.

As regiões claras representam valores de disparidade maiores de que valores de disparidade das regiões escuras.

Alguns comentários podem ser feitos a respeito do mapa de disparidade interpolado na figura 5.26.

As regiões distintas das demais nos cantos esquerdo e direito do mapa de disparidade ocorreram porque este está referenciado no espaço ciclopeano, e portanto, essas áreas são áreas em que não houve superposição da imagem da esquerda com a imagem da direita e vice-versa. O fato do canto esquerdo ser totalmente escuro e o canto direito ser representado em níveis de cinza em sua parte inferior é devido a replicação de *pixels* utilizada para "preencher" estas áreas.

Ainda sobre esta figura, pode-se notar no canto esquerdo inferior, uma área nitidamente mais clara, representando maiores valores de disparidade. Este fato ocorreu devido ao processo de *matching* de regiões ter correspondido erroneamente uma região nesta área.

A figura 5.27 mostra o mapa de disparidade corrigido após o processo iterativo.

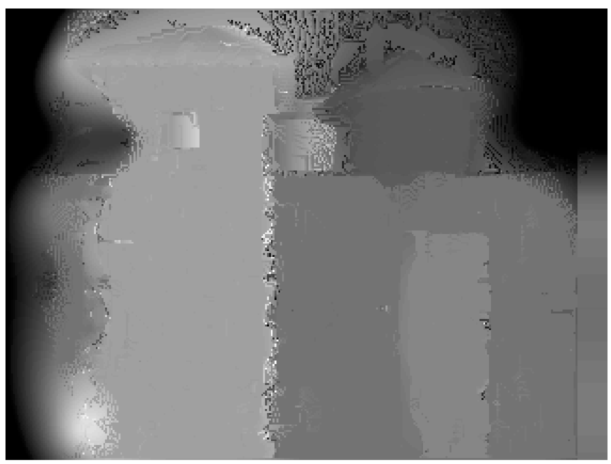


Fig. 5.27 - Mapa de disparidade corrigido referente ao par de imagens 1 para o teste 1.

O mapa de disparidade corrigido da figura 5.27 mostra que os valores de disparidades da maioria dos *pixels* no espaço ciclopeano foram determinados corretamente.

Devido a origem do sistema referencial do espaço-objeto ser o centro perspectivo da imagem ciclopeana e este ser um sistema levógiro, a profundidade da cena aumenta positivamente da origem para a cena, e portanto, altos valores de disparidade (áreas claras no mapa de disparidade) representam baixos valores de profundidade de acordo com a expressão 2.60. Exemplificando, a torre retangular está mais próxima da origem do sistema de que as outras regiões que compõem a cena, portanto, sua profundidade é menor e consequentemente, sua disparidade é maior.

Neste mapa de disparidade ainda, pode-se notar que variações abruptas nos valores de disparidade foram determinadas com sucesso em algumas áreas, ao contrário do mapa de disparidade interpolado.

O canto esquerdo da torre retangular foi relativamente mal reconstruído principalmente devido a presença de poucos *pixels* homólogos oriundos do *matching* baseado em feições, que controlariam melhor o movimento de vergência para esta região. Outro motivo desta má reconstrução é o fato desta região estar em uma área de oclusão, na qual o procedimento de *threshold* de oclusão falhou devido as intensidades serem semelhantes, ao contrário do canto direito desta mesma torre.

A área clara entre a torre retangular e a torre cilíndrica, bem como algumas áreas acima das torres, também foram mal reconstruídas, uma vez que esta áreas deveriam possuir valores baixos de disparidade, e portanto, representados em tons escuros. Este fato ocorreu devido também não existir nenhum *pixel* homólogo oriundo do *matching* baseado em feições, para o controle do movimento de vergência para estas áreas.

A figura 5.28 ilustra a imagem de controle que ordena o processo iterativo.

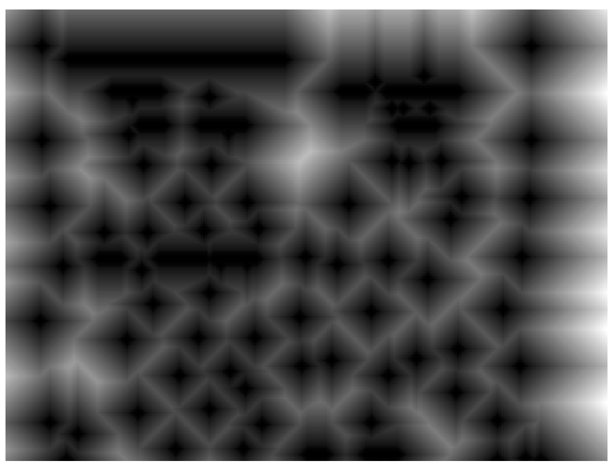


Fig. 5.28 - Imagem de controle referente ao par de imagens 1 para o teste 1.

Par de imagens 2

Os parâmetros utilizados neste teste, para o par de imagens 2 foram os mesmos adotados para as imagens 1.

A figura 5.29, mostra o mapa de disparidade interpolado.

Fig. 5.29 - Mapa de disparidade interpolado referente ao par de imagens 2 para o teste 1.

No mapa de disparidade interpolado da figura 5.29, nota-se a mesma área clara no canto inferior esquerdo, oriunda da correspondência errada de uma região no processo de *matching* de regiões.

As regiões que representam as áreas em que não houve superposição da imagem da esquerda com a imagem da direita e vice-versa são mais estreitas e difíceis de serem identificadas. Este fato é coerente, uma vez que o comprimento da base é menor, e portanto, a superposição das imagens estereoscópicas é maior.

A figura 5.30 mostra o mapa de disparidade corrigido após o processo iterativo.

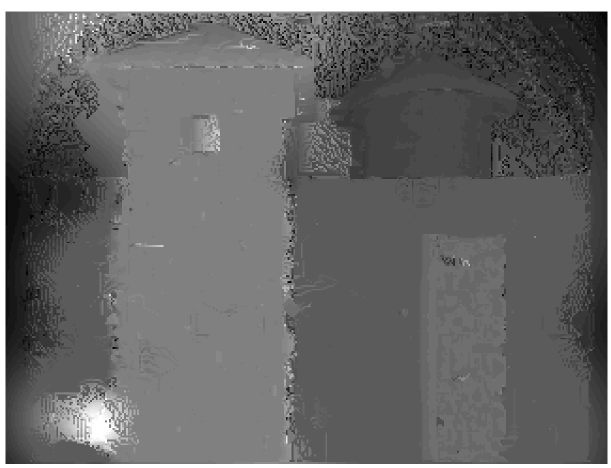


Fig. 5.30 - Mapa de disparidade corrigido referente ao par de imagens 2 para o teste 1.

O mapa de disparidade corrigido da figura 5.30 mostra que a grande maioria dos valores de disparidades dos *pixels* no espaço ciclopeano foram determinados corretamente.

Neste mapa de disparidade pode-se notar que variações abruptas nos valores de disparidade foram determinadas com sucesso na maioria das áreas em que este fato ocorre.

As áreas de não superposição das imagens da esquerda com a imagem da direita e vice-versa são melhores distinguidas neste mapa de disparidade corrigido.

Como era de se esperar, o canto esquerdo da torre retangular foi melhor reconstruído de que para o par de imagens 1, principalmente devido às áreas de oclusão serem menores entre o par de imagens estereoscópicas.

A área entre a torre retangular e a torre cilíndrica, bem como algumas áreas acima das torres, também foram relativamente mal reconstruídas para o mapa de disparidade corrigido referente ao par de imagens 2 devido aos mesmos problemas que ocasionaram esta má reconstrução para o par de imagens 1, porém, neste mapa de disparidade corrigido, estas áreas apresentam-se menos concentradas.

A figura 5.31 ilustra a imagem de controle que ordena este processo iterativo.

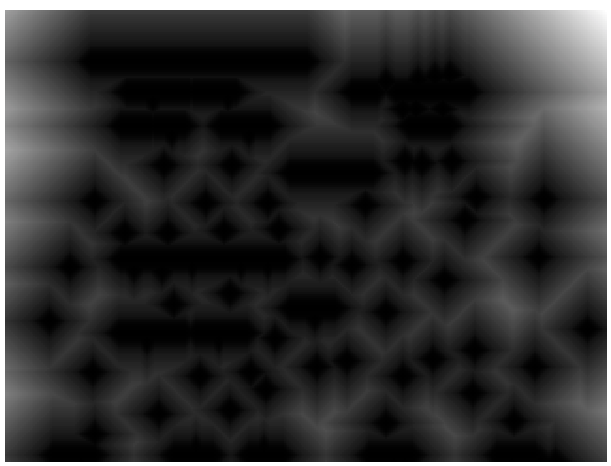


Fig. 5.31 - Imagem de controle referente ao par de imagens 2 para o teste 1.

Par de imagens 3

Os parâmetros utilizados neste teste, para o par de imagens 3 foram:

-Dimensão na direção x da janela de correlação = 7 *pixels*;

-Dimensão na direção y da janela de correlação = 3 *pixels*;

-*Threshold* de oclusão = 30 níveis de intensidade; e

-Fator de escala para o gradiente de disparidade = 1.5.

Estes parâmetros diferem dos anteriores apenas para o *threshold* de oclusão, devido as grandes distorções radiométricas entre estas imagens.

A figura 5.32 ilustra o mapa de disparidade interpolado referente ao par de imagens 3.

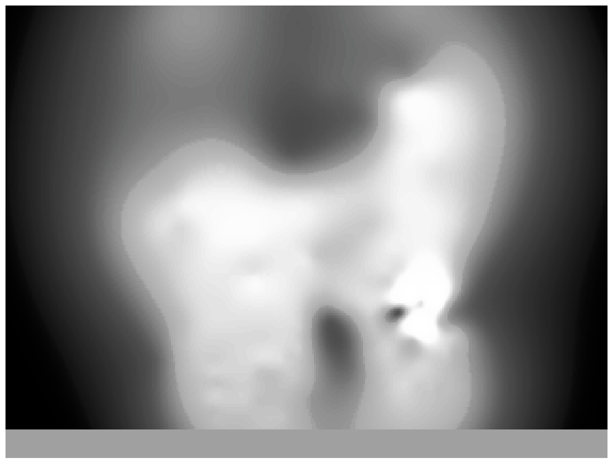


Fig. 5.32 - Mapa de disparidade interpolado referente ao par de imagens 3 para o teste 1.

Neste mapa de disparidade interpolado nota-se nos cantos esquerdo e direito, grandes áreas escuras que representam as áreas em que não houve superposição da imagem da esquerda com a imagem da direita e vice-versa.

A região mais clara no centro esquerdo (centro da garrafa) ocorreu devido as três regiões que foram correspondidas erroneamente pelo processo de *matching* de regiões. Nota-se ainda duas grandes áreas claras que representam a jarra e a garrafa.

A figura 5.33 ilustra o mapa de disparidade corrigido referentes ao par de imagens 3.

Fig. 5.33 - Mapa de disparidade corrigido referente ao par de imagens 3 para o teste 1.

Neste mapa de disparidade percebe-se com mais facilidade as áreas de não superposição entre as imagens estereoscópicas. Esta figura 5.33 ilustra bem o conceito de espaço ciclopeano, uma vez que os objetos que compõem a cena estão no extremo direito da imagem da esquerda e no extremo esquerdo da imagem da direita, e no mapa de disparidade corrigido (que está referenciado no espaço ciclopeano), estes objetos aparecem ao centro.

O mapa de disparidade corrigido da figura 5.33 apresentou a pior reconstrução em relação aos mapas referentes ao par de imagens 1 e 2. Isto ocorreu, principalmente, devido a pouca quantidade de *pixels* homólogos oriundos do *matching* baseado em feições que controlam o movimento de vergência para áreas de atenção da cena. Entretanto, alguns aspectos interessantes podem ser analisados neste mapa de disparidade.

O procedimento iterativo de correção do mapa de disparidade interpolado conseguiu corrigir parcialmente à área no centro da garrafa que havia sido interpolada indevidamente devido às três regiões que foram correspondidas erroneamente pelo processo de *matching* de regiões.

A área clara entre a jarra e a garrafa representa o cabo da jarra, apesar de não haver nenhum *pixel* oriundo do *matching* baseado em feições para controlar o movimento de vergência para concentrar atenção a esta área da cena.

A área logo acima da jarra e da garrafa representa o fundo da cena, sendo que este é formado por dois planos cuja a interseção destes é a região mais profunda da cena, e portanto, representado em tons mais escuros. As demais grandes áreas escuras são áreas de não superposição das imagens estereoscópicas, como já citado.

A figura 5.34 mostra a imagem de controle.

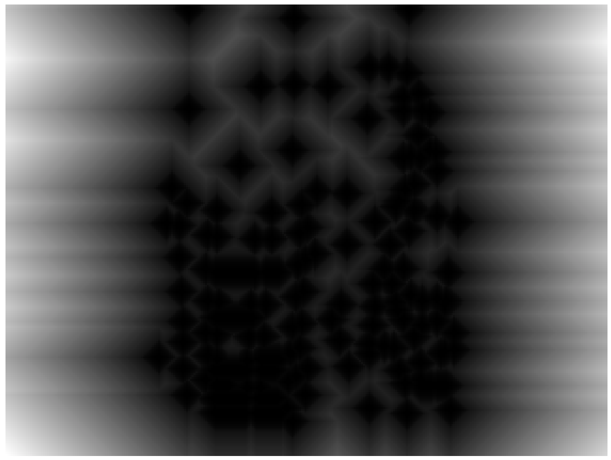


Fig. 5.34 - Imagem de controle referente ao par de imagens 3 para o teste 1.

5.4.2 Teste 2 (sentido do gradiente de disparidade para determinação do intervalo de busca)

O mapa de disparidade interpolado e a imagem de controle não serão mostrados neste sub-ítem para todas as imagens, pois estes são funções apenas dos *pixels* homólogos oriundos do *matching* baseado em feições, sendo que estes não foram alterados para esses testes, e portanto, o mapa de disparidade interpolado e a imagem de controle também não o foram.

Par de imagens 1

Os parâmetros utilizados neste teste, para as imagens 1 foram os mesmos utilizados no teste 1 para as imagens 1.

A figura 5.35, mostra o mapa de disparidade corrigido.

Fig. 5.35 – Mapa de disparidade corrigido referente ao par de imagens 1 para o teste 2.

A figura 5.35 é semelhante a figura 5.27, na qual o mapa de disparidade foi gerado utilizando o módulo do gradiente de disparidade como preditor do intervalo permitido do *pixel* homólogo.

Analisando o mapa de disparidade da figura 5.27 e o mapa de disparidade da figura 5.35, nota-se que este último é mais "suave" e ocorreu maiores problemas nas áreas de variação brusca dos valores de disparidade. Este fato ocorreu, principalmente, devido a propagação do sentido do gradiente de disparidade no momento de correção do mapa de disparidade. Assim, o sentido do gradiente de disparidade associado a um *pixel* no espaço ciclopeano é propagado ao seus vizinhos, fazendo com que o mapa de disparidade corrigido torne-se mais suave de que o mapa de disparidade corrigido a partir do módulo do gradiente de disparidade.

Par de imagens 2

Os parâmetros utilizados neste teste, para o par de imagens 2, foram os mesmos utilizados no teste 1 para o par de imagens 1.

A figura 5.36 ilustra o mapa de disparidade corrigido para estas imagens.

Fig. 5.36 – Mapa de disparidade corrigido referente ao par de imagens 2 para o teste 2.

Os mesmos problemas ocorridos para o mapa de disparidade referente ao par de imagem 1 do teste 2, ocorreram para este mapa de disparidade corrigido, ou seja, o mapa gerado durante o processo iterativo é mais “suave” que o mapa de disparidade gerado quando utilizado o módulo do gradiente de disparidade.

Par de imagens 3

Os parâmetros utilizados para este teste são os mesmos utilizados para o par de imagens 3 no teste 1.

A figura 5.37 ilustra o mapa de disparidade corrigido para estas imagens.

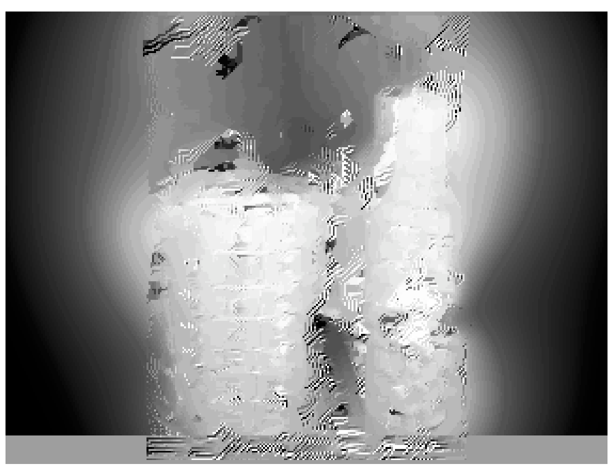


Fig. 5.37 - Mapa de disparidade corrigido referente ao par de imagens 3 para o teste 2.

Os mesmos problemas encontrados para os pares de imagens 1 e 2 do teste 2 foram presenciados para essas imagens.

Capítulo 6

Conclusões e trabalhos futuros

Os testes e resultados apresentados no capítulo 5 sugerem que a metodologia adotada é eficiente para a obtenção de um mapa de disparidade denso que permite variações bruscas nos valores de disparidade.

As premissas geométricas da visão estéreo utilizadas (geometria epipolar e restrição do intervalo de correspondência) impõem restrições globais e absolutas ao sistema, porém, existe ainda muitas ambiguidades ao longo das linhas epipolares, sendo que estas premissas não são suficientes para resolver o problema de correspondência estéreo com sucesso.

As premissas empíricas (ordem, unicidade e continuidade) impõem restrições locais e relativas, mas estas não garantem soluções para o problema de correspondência estéreo em muitas situações que eventualmente possam ocorrer, tais como superfícies transparentes e/ou variações abruptas nos valores de profundidade da cena e oclusões.

As evidências biológicas de seletividade direcional e compatibilidade (polaridade das bordas de intensidade) descritas na forma de algoritmo, apesar de importantes, foram empregadas como coadjuvantes na solução do problema de correspondência estéreo no processo de *matching* baseado em feições.

As evidências biológicas de movimento de vergência dos olhos e gradiente de disparidade descritos na forma de algoritmo, foram as principais ferramentas utilizadas para a eliminação de ambigüidades empregadas no processo de correspondência estéreo.

O *threshold* de oclusão, apesar de poder ser interpretado como uma premissa empírica da visão estéreo, mostrou-se eficiente em detectar áreas de oclusão, exceto quando os níveis de intensidade são semelhantes nestas áreas.

O *matching* baseado em feições mostrou-se robusto na determinação das feições homólogas, principalmente o *matching* de feições retas. A utilização de feições para controlar o movimento de vergência ao invés de correlacionar grandes áreas é do ponto de vista biológico, uma interpretação natural e mesmo espontânea, uma vez que feições constituem geralmente singularidades (características relevantes) da cena, e portanto, a atenção deve-se concentrar para essas regiões da cena.

Outro aspecto interessante em utilizar feições para controlar o movimento de vergência é que estas são mais estáveis para o processo de correspondência estéreo de que apenas os valores de intensidade devido a serem menos sensíveis aos problemas da visão estéreo.

A metodologia empregada para segmentar as feições de interesse (feições retas e regiões relativamente homogêneas) mostrou-se apenas regular para esta tarefa, determinando relativamente poucas feições.

O *matching* baseado em áreas deve utilizar janelas de dimensões pequenas para minimizar o problema causado pelas distorções geométricas e oclusões entre o par de imagens estereoscópicas.

O gradiente de disparidade mostrou-se um ótimo preditor do intervalo permitido de busca do *pixel* homólogo devido a se adaptar constantemente para cada ponto do mapa de disparidade, não propagando os valores de disparidade diretamente, o que ocasionaria em um mapa de disparidade suave, mas sim um intervalo permitido no qual o valor de disparidade possa variar.

A informação do sentido do gradiente de disparidade não melhorou os resultados como se esperava, ao contrário, o mapa de disparidade tornou-se mais suave de que o mapa de disparidade gerado quando utilizado o módulo do gradiente de disparidade para definir o intervalo de busca do *pixel* homólogo.

O sentido do gradiente de disparidade deve ser utilizado em situações nas quais o mapa de disparidade interpolado retrata quase que fielmente a disparidade "real". Estas situações ocorrem quando as superfícies da cena são suaves, e portanto, o processo de interpolação fornece uma boa aproximação dos valores de disparidade para todo o mapa de disparidade.

Nos testes do capítulo 5, os mapas de disparidade interpolados eram grosseiras aproximações do mapa de disparidade "real" e devido a isso, o módulo do gradiente de disparidade obteve melhores resultados, mostrando-se mais robusto.

O comprimento da base é um parâmetro importante para a reconstrução tridimensional e para o processo de correspondência estéreo. Os testes utilizando o par de imagens 2 obtiveram melhores resultados de que os testes utilizando o par de imagens 1, tanto para o teste 1 quanto para o teste 2, justamente devido ao comprimento da base do par de imagens 2 ser a metade do comprimento da base do par de imagens 1, fazendo com que as distorções geométricas e radiométricas e as áreas de oclusão entre as imagens estereoscópicas, fossem menores nestas imagens de que no par de imagens 1.

Como trabalhos futuros, sugere-se desenvolver outras técnicas de segmentação de feições que possibilitem a determinação de um maior número de feições, a determinação de um modelo matemático para o fator de escala do gradiente de disparidade em função das possíveis interpretações que este pode assumir e finalmente, realizar testes para outras imagens estereoscópicas.

Referências Bibliográficas

F. ACKERMANN. High Precision Digital Image Correlation. *Proceedings*, 39th Photogrammetric Week, Stuttgart, 1:231-244, 1984.

J. K. AGGARWAL, L. S. DAVIS and W. N. MARTIN. Correspondence Processes in Dynamic Scene Analysis. *Proceedings*, of IEEE, 69(5):562-571, May 1981.

C. H. ANDERSON, B. A. OLSHAUSEN and D. V. ESSEN. Routing Networks in Visual Cortex. *In: Handbook of Brain Theory*, M. A. Arbib (ed) 1:823-826. MIT Press, Cambridge, 1995.

G. C. D. ANGELIS, I. OHZAWA and R. D. FREEMAN. Depth is encoded in the Visual Cortex by a Specialized Receptive Field Structure. *Nature*, 11:156-159, 1991.

P. A. ARNDT, H. A. MALLOT, H.H. BULTHOFF. Human stereovision without localized image features. *Biological Cybernetics*, 72:279-293, 1995.

N. AYACHE. *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*. MIT Press, Cambridge, 1991.

R. BAJCSY & L. LIEBERMAN. Texture gradient as a depth cue. *Computer Graphics and Image Processing,* 5(1):52-67, 1976.

D. H. BALLARD and C. M. BROWN. *Computer Vision*. Prentice-Hall, Inc., Englewood Cliffs, 1982.

S. T. BARNARD and W. B. THOMPSON. Disparity analysis of image. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 8(6):679-698, 1985.

J. BIGUN, G. H. GRANLUND and J. WIKLUND. Multidimensional Orientation Estimation with Applications to Texture Analysis and Optical Flow. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 13:775-790, 1991.

A. D. BIMBO, P. NESI and J. L. C. SANZ. Optical Flow Computation Using Extended Constraints. *IEEE Transactions on Image Processing,* 5(5):206-218, 1996.

P. BURT & B. JULESZ. A Disparity Gradient Limit for Binocular Fusion. *Science*, 208:615-617, 1980.

J. A. F. COSTA. *Sistema de reconhecimento de padrões visuais invariante a transformações gométricas utilizando redes neurais artificiais de múltiplas camadas.* Dissertação de mestrado, E.E.S.C/USP, São Carlos, 1996.

L. F. COSTA. On the Most Advanced Signal Processing System: The Visual Cortex, its Modelling, and Implications for Computer Vision. *Anais do Workshop sobre Computação de Alto Desempenho para Processamento de Sinais*, 1:122-141, 1993.

U. R. DHOND and J. K. AGGARWAL. Structure from Stereo - A Review. *IEEE Transactions on Systems, Man and Cybernetics,* 19(6):1489-1510, 1989.

J. ENS and P. LAWRENCE. An Investigation of Methods for Determining Depth from Focus. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 15(2):97-108, 1993.

J. FACON. *Morfologia Matemática: teoria e exemplos.* Universitária Champagnat, Curitiba, 1996.

J. FLUSSER and T. SUK. A Moment-Based Approach to Registration of Imagens with Affine Geometric Distortion. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):382-387, 1994.

J. D. FOLEY, A. V. DAM., S. K. FEINER and, J. F. HUGHES. *Computer Graphics: principles and practice.* Addilson Wesley Publishing Company, Reading, 1982.

J. P. FRISBY. Stereo Correspondence and Neural Networks. *In: Handbook of Brain Theory*, M. A. Arbib (ed), 1:937-941. MIT Press, Cambridge, 1995.

M. GALO. *Calibração e aplicação de câmaras digitais*. Dissertação de mestrado, UFPR, Curitiba, 1993.

M. GALO. Determinação de Pontos Homólogos em Imagens Digitais. Departamento de Cartografia-UNESP, Câmpus de Presidente Prudente, 1994.

M. GALO and C. L. TOZZI. Surface reconstruction using multiple light sources and perspective projection. *Proceedings*, IEEE - International Conference on Image Processing*, Lausanne, Switzerland*, 1:309-312, 1996.

S. GANAPATHY. Decomposition of Transformation Matrices for Robot Vision. *Pattern recognition Letters* , 2:401-412, 1984.

C. GEMAEL. *Introdução ao Ajustamento de Observações: aplicações geodésicas*. Editora UFPR, Curitiba, 1994.

R. C. GONZALEZ and R. E. WOODS. *Digital Image Processing*. Addison-Wesley Publishing Company, Reading, 1993.

W. E. L. GRIMSON. A computer implementation of a theory of human stereo vision. *Phil. Trans. Roy. Soc. Lond.*, B292:217-253, 1980.

A. C. GUYTON, *Tratado de Fisiologia Médica*. 7a. edição, Guanabara Koogan, Rio de Janeiro, 1989.

H. HAKEN. Cooperative Phenomena. *Handbook of Brain Theory*, 1:261-266. MIT Press, Cambridge, 1995.

R. M. HARALICK, S. R. STERNBERG and X. ZHUANG. Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 9(4):532-550, 1987.

R. I. HARTLEY. Triangulation. *Computer Vision and Image Understanding*, 68(2):146-157, 1997.

J. K. HASEGAWA. *Shape from shading com projeção perspectiva e calibração de câmara.* Tese de doutorado, FEEC/UNICAMP, Campinas, 1997.

S. HAYKIN. *Neural Networks: a comprehensive foundation.* Mac Millan College Publishing Company, Inc, New York, 1994.

A. Y. K. HO and T. C. PONG. Cooperative Fusion of Stereo and Motion. *Pattern Recognition,* 29(1):121-130, 1996.

B.K.P. HORN. Understanding image intensities. *Artificial Intelligence*, 8:201-231, 1997.

P.V.C. HOUGH. Methods and Means for Recognizing Complex Patterns. U.S. Patent 3,069,654, 1962.

M. K. HU. Visual Pattern Recognition by Moment Invariants. *Transactions on Information Theory*, 8:179-187, 1962.

R. A. HUMMEL and S. W. ZUCKER. ON the Foundations of Relaxation Labeling Processes. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 5(3):267-287, 1983.

J. M. JOO, L. F. COSTA and R. KOBERLE. Geometric-Transformation-Invariant Pattern Recognition with Hough Transforms and Distance-Discriminator Neural Networks. *Anais do Workshop sobre Computação de Alto Desempenho para Processamento de Sinais,* 1:103-108, São Carlos, 1993.

F. KLEIN. *Elementary Mathematics from and Advanced Standpoint.* Dover Publications, Inc., New York, 1939.

K. KRAUS. *Photogrammetry*. Ferd Dummler Verlag,, Bonn, 1992.

S. C. KREMER. Natural Proprerties in an Artificial Neural Network. *Proceedings*, International Conference on Neural Networks*, Houston,*1: 612-616, 1997.

E. KREYSZIG. *Advanced engineering mathematics*. John Wiley & Sons, Inc, New York, 1993.

S. A. LLOYD. Binocular Stereo Algorithm Based on the Disparity-Gradient Limit and Using Optimization Theory. *Image and Vision Computing*, 3:177-181, 1985.

R. A. LOTUFO and E. TRETTEL. Image Segmentation by Mathematical Morphology. *Anais do Brazilian Workshop'96 on Mathematical Morphology, IME/USP, São Paulo,* 1996.

J. B. LUGNANI. *Introdução a fototriangulação.* Imprensa Universitária da UFPR, Curitiba, 1987.

D. MARR. *Vision.* W. H. Freeman and Company, New York, 1980.

D. MARR and T. POGGIO. Cooperative computation of stereo disparity. *Science*, 194:301-328, 1976.

D. MARR and T. POGGIO. A computational theory of human stereo vision. *Proceedings*, Royal Society of London, B204:301-328, 1979.

J. F.W. MAYHEW and J. P. FRISBY. Psychophysical and computational studies towards a theory of human stereopis. *Artificial intelligence,* 17:349-385, 1981.

F. MEYER and S. BEUCHER. Morphological Segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21-46, 1990.

S. K. NAYAR and Y. NAKAGAWA. Shape from Focus. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 16(8):824-831, 1994.

S. K. NAYAR, M. WATANABE and M. NOGUCHI. Real - Time Focus Range Sensor. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 18(12):1186-1197, 1996.

F. M. A. NOGUEIRA e A. M. G. TOMMASELLI. Desenvolvimento de um Módulo de Aquisição de Imagens para um Sistema Fotogramétrico Digital. *I Congresso de Iniciação Científica da Unesp, Guaratinguetá,* 1:55, 1995.

K. N. OGLE. *Researches in binocular vision*. Saunders, 1950.

A. PERTL. Digital Image Correlation with the Analytical Plotter Planicomp C 100. *International Archives of Photogrammetry and Remote Sensing,* Hamburg, 25(3b):874-882, 1984.

C. J. POELMAN and T. KANADE. A Paraperspective Factorization Method for Shape and Motion Recovery. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 19(3):206-218, 1997.

A. P. D. POZ e M. GALO. A Utilização do Conceito de Geometria Epipolar em Correlação de Imagens Digitais. *Pesquisas*, 19(2):137-142, 1992.

K. PRAZDNY. Detection of Binocular Disparities. *Biological Cybernetics*, 52:93-99, 1985.

K. E. PRICE. Relaxation Matching Techniques - A Comparison. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 7(5):617-623, 1985.

A. ROSENFELD, R. A. HUMMEL & S. W. ZUCKER. Scene labeling by relaxation operation. *IEEE Transactions on Systems, Man and Cybernetics*, 6:420-423, 1976.

D. T. SANDWELL. Biharmonic spline interpolation of GEOS-3 and seasat altimeter data. *Geophysical Research Letters*, 14(2):139-142, 1987.

R.J. SCHALKOFF. *Digital Image Processing and Computer Vision*. John Wiley & Sons, New York, 1989.

N. SHRIKHANDE and G. STOCKMAN. Surface Orientation from a Projected Grid. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 11(6):650-655, 1989.

J.C.M. STRAUCH. *Correlação de Imagens Digitais.* Dissertação de Mestrado, UFPR, 1991.

D. TERZOPOULOS. Multilevel computation processes for visual surface reconstruction. *Computer. Vision and Graphics Image Processing*, 24:52-96, 1983.

J. S. TITTLE and J. T. TODD. Perception of Three-Dimensional Structure. *In: Handbook of Brain Theory*, M. A. Arbib (ed), 1:715-718. MIT Press, Cambridge, 1995.

A.M.G. TOMMASELLI e C. L. TOZZI. Extração de Linhas Retas em Imagens Digitais para Aplicações Fotogramétricas. *Anais do XVI Congresso Brasileiro de Cartografia, Rio de Janeiro,* 1993.

A. M. G. TOMMASELLI, M. H. SHIMABUKURO, P. A. P. SCALCO and F. M. A. NOGUEIRA. Implementation of a Photogrammetric Range System. *International Archives of Photogrammetry and Remote Sensing, Vienna*. 31(B2):368-373, 1996.

H. P. TRIVEDI and S. A. LLOYD. The Role of Disparity Gradient in Stereo Vision. *Perception*, 14:685-690, 1985.

R. Y. TSAI. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *Proceedings*, IEEE conference on Computer Vision and Pattern Recognition, 1:364-374, 1986.

R. Y. TSAI. A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, 3(4):323-344, 1987.

P. VUYLSTEKE and A. OOSTERLINCK. Range Image Acquisition with a Single Binary-Encoded Light Pattern. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 12(2):148-163, February 1990.

R. J. WOODHAM. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. *Proceedings*, 22nd International Symposium, Society of Photo-optical Instrumentation Engineers, 136:143, 1978.

Q. X. WU. A Correlation-Relaxation-Labeling Framework for Computing Optical Flow-Templete Matching from a New Perspective. *IEEE Transactions on Pattern Analysis and Machine intelligence,* 17(8):843-853, 1995.

Y. YANNIRIS. *Improvements to the off-line epipolar correlation*. Msc thesis, The University of New Brunswick, Fredericton, Canadá, 1974.

S. Y. YUEN and C. H. MA. An Investigation of the Nature of Parameterization for the Hough Transform. *Pattern Recognition*, 30(6):1009-1040, 1997.