

Uma Abordagem para a Estereoscopia via Funções de Green de Equações de Casamento

Leonardo Beltrão

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre em Computação. Área de concentração: Computação Visual e Interface.

Orientador: José Ricardo A. Torreão

Niterói, Março de 2007.

Uma Abordagem para a Estereoscopia via Funções de Green de Equações de Casamento

Leonardo Beltrão

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre em Computação. Área de concentração: Computação Visual e Interface.

Aprovada por:

Prof. José Ricardo de Almeida Torreão (Presidente) – UFF

Prof. Jacques Facon – PUC-PR

Prof. Anselmo Antunes Montenegro - UFF

Niterói, 16 de Abril de 2007.

B453 Beltrão, Leonardo.

Uma abordagem para a estereoscopia via Funções de Green de equações de casamento / Leonardo Beltrão – Niterói, RJ : [s.n.], 2007.

84f.

Orientador: José Ricardo A. Torreão.

Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, 2007.

1. Computação visual. 2. Estereoscopia. 3. Funções de Green.

Aos meus queridos filhos, Murilo, Heitor e Eduardo, e a minha esposa, Simone, pelo apoio e compreensão durante os momentos em que me ausentei para este trabalho.

Agradecimentos

Ao meu orientador, José Ricardo Torreão, pela sua paciência, compreensão, e apoio nas situações mais adversas que nos deparamos ao longo deste projeto.

Aos meus pais por tudo que eles me ensinaram nesta vida.

Ao meu amigo Abel Francisco pela sua colaboração com as figuras ilustrativas.

A todos os amigos e familiares que me ajudaram nesta jornada, em especial, a Deus que me deu luz e inspiração quando tudo parecia estar perdido.

Uma Abordagem para a Estereoscopia via Funções de Green de Equações de Casamento

Leonardo Beltrão

Resumo

A estereoscopia é o principal processo da visão responsável pela nossa percepção tridimensional do mundo. A diferença relativa de posição entre as projeções, em nossos dois olhos, de um mesmo ponto de uma cena 3D chama-se disparidade, e incorpora informação sobre a sua profundidade. Diversos processos computacionais têm sido propostos para a estimação de disparidades, mas poucos se preocuparam em buscar um embasamento consistente na neurofisiologia da visão. Dentre estes, os mais bem sucedidos parecem ser os baseados no chamado *modelo da energia* para as células do córtex visual. O objetivo do presente trabalho é o de propor uma nova abordagem para a estimação de disparidades, tentando preservar os elementos fisiologicamente plausíveis do modelo da energia. Para isto, nós utilizaremos as funções de Green de equações de casamento de imagens. Dada uma imagem de entrada, o método da função de Green permite gerar uma segunda imagem que corresponderia a uma versão deslocada da primeira. Assim, dado um par de imagens estereoscópicas, nós podemos filtrá-las, dentro de janelas apropriadas, por funções de Green que induzirão deslocamentos variados sobre as mesmas. Por outro lado, como as imagens estereoscópicas constituem visões deslocadas de uma mesma cena, podemos procurar por um deslocamento equivalente usando o filtro da função de Green, que permita inferir a disparidade original entre as duas imagens, e portanto, a profundidade de cada ponto na cena. Quando as janelas empregadas assumem a forma de funções de Gabor, nós demonstramos que a abordagem da função de Green efetivamente preserva os aspectos fisiologicamente plausíveis do modelo da energia, permitindo a estimação confiável de disparidades a partir tanto de imagens sintéticas quanto de imagens reais.

Palavras-chave: Funções de Green, Neurofisiologia da Visão, Estereoscopia, Visão Computacional.

An Approach for Disparity Estimation through Green's Functions of Matching Equations

Leonardo Beltrão

Abstract

Stereoscopy is the main vision process leading to our 3D perception of the world. The relative displacement between the projections, in our two eyes, of one specific point on a 3D scene is referred to as disparity, and incorporates information on the point's distance from the observer. Several computational models have been proposed for disparity estimation, but few have sought to find a solid grounding on biological vision. Among these, the most successful seem to be based on the so-called *energy model* of the cortical cells. The goal of the present work is to propose a new approach for disparity estimation, trying to preserve the main neurophysiologically plausible features of the energy model. For this purpose, we will use Green's functions of image matching equations. Given an input image, the Green's kernel produces a shifted version of it. Thus, given a pair of stereoscopic images, we can filter them by the Green's function, over an appropriate window, in order to induce different relative shifts between them. Since the stereoscopic images are themselves shifted visions of the same scene, we can search for an equivalent displacement, using the Green's function filter, which will allow us to infer the original disparity encoded in the stereo pair, and thus to recover a depth map of the scene. When the employed windows take the form of Gabor functions, we demonstrate that the Green's function approach effectively preserves the plausible physiological aspects of the energy model, thus allowing the reliable estimation of disparities both from synthetic and from real-world stereograms.

Keywords: Green's Functions, Biological Vision, Stereoscopy, Computer Vision.

Sumário

1	Introdução	5
2	Estereoscopia	7
2.1	Princípios Geométricos da Estereoscopia	9
2.2	Abordagens para Estereoscopia.....	13
2.2.1	Correspondência por Características.....	13
2.2.2	Correspondência por Área.....	15
2.2.3	Correspondência por Pixel.....	16
2.2.4	Correspondência por Fase.....	17
2.2.5	Correspondência por Energia	18
3	Neurofisiologia da Visão.....	19
3.1.1	As Células Nervosas	20
3.2	Retina.....	23
3.2.1	Campo Receptivo - CR.....	25
3.2.2	Fotorreceptores	27
3.2.3	Células Bipolares, Horizontais e Amacrinas	28
3.3	Núcleo Geniculado Lateral e Córtex Visual Primário	29
3.3.1	As Células do Córtex.....	31
3.3.2	Células Simples - CS.....	32
3.3.3	Células Complexas - CC	35
3.4	Células Binoculares Sensíveis à Disparidade.....	37
3.4.1	Células Simples Binoculares – CSB.....	37
3.4.2	Células Complexas Binoculares - CCB.....	40
3.4.3	Detecção de Disparidade pelo Modelo da Energia.....	41
4	Estereoscopia baseada na Função de Green	43
4.1	Campo de Disparidades Uniforme.....	43
4.2	Campo de Disparidades Linear	45
4.3	Cálculo de Disparidades a partir do Filtro de Green	47
4.3.1	Abordagem para o Cálculo de Disparidades	50
5	Resultados.....	53
5.1	Experimentos	58
5.1.1	Imagens Sintéticas.....	58
5.1.2	Imagens Reais.....	61
6	Conclusão e Trabalhos Futuros.....	67
	Apêndice I	69
	Apêndice II	71
	Referências Bibliográficas	74

Sumário de Figuras

Figura 2-1: Efeito da Duplicidade Visual.....	7
Figura 2-2: Geometria da Projeção Binocular.....	8
Figura 2-3: Estereograma de Pontos Aleatórios.	9
Figura 2-4: Triangulação Inerente à Estereoscopia.	11
Figura 2-5: Geometria Epipolar.	12
Figura 2-6: Geometria Epipolar para Projeção Retificada.	13
Figura 2-7: Correspondência por Características.	14
Figura 2-8: Correspondência por Área.	16
Figura 3-1: A Diferença de Percepção em Cada Olho.	19
Figura 3-2: Estrutura de uma Célula Nervosa.	21
Figura 3-3: O Caminho Visual e sua Representação em Blocos.....	22
Figura 3-4: Estrutura Interna do Olho Humano.....	23
Figura 3-5: Organização das Células na Retina.	24
Figura 3-6: Resposta das Células Centro-ativo e Centro-inativo.....	26
Figura 3-7: Estrutura dos Fotorreceptores para Células Centro-ativo e Centro-Inativo.	27
Figura 3-8: Os Tipos de Fotorreceptores da Retina Humana.	28
Figura 3-9: Mapeamento no Córtex Visual de um Objeto.	30
Figura 3-10: Tipos de Campos Receptivos para Células Simples.	33
Figura 3-11: Comportamento das Células Simples.....	33
Figura 3-12: Modelo de Interligação das Células Simples.....	34
Figura 3-13: Resposta das Células Complexas para Diferente Estímulos.....	36
Figura 3-14: Modelo de Interligação das Células Complexas a partir de Células Simples.....	37
Figura 3-15: Estrutura da Função de Gabor Unidimensional.....	38
Figura 3-16: Tipos de CR para Seletividade por Disparidade nas CS.....	39
Figura 4-1: Função de Green de Complexa.....	47
Figura 5-1: Processo de Deslocamento para Equilibrar as Disparidades.....	54

Figura 5-2a: Parâmetros para Pseudocódigo do Método de Green	55
Figura 5-2b: Pseudocódigo do Método de Green.....	55
Figura 5-3: Efeito de Deslocamento do Filtro de Green.....	57
Figura 5-4: Resultados Estereograma Random Dots.....	59
Figura 5-5: Resultados do Estereograma Zwei.....	60
Figura 5-6: Resultados do Estereograma Pentagon.....	62
Figura 5-7: Resultados do Estereograma Tree ou Baum.....	63
Figura 5-8: Resultados do Estereograma Shrub.....	65
Figura 5-9: Resultados do Estereograma Parkmeter.....	66

Sumário de Tabelas

Tabela 2-1: Principais Tipos de Características Utilizadas para Correspondência.	15
Tabela 5-1: Combinação dos Filtros de Green para Induzir Disparidades Positivas e Negativas	53
Tabela 5-2: Parâmetros de Entrada Preservados Constante nos Experimentos.	58

1 Introdução

A estereoscopia (do grego, "visão sólida") é o processo da visão que proporciona ao observador a capacidade de perceber a dimensão de profundidade em uma cena, a partir das imagens bidimensionais projetadas em um sistema binocular. Cada olho ou câmera vê a cena de dois pontos de vista diferentes, permitindo ao cérebro, ou a um sistema computacional, combinar as informações obtidas a partir do par de imagens, de modo a produzir uma representação tridimensional.

A visão estereoscópica computacional foi claramente inspirada na visão dos seres vivos, mas, dentre os inúmeros algoritmos propostos nesta área, poucos se preocuparam em buscar um embasamento neurofisiológico consistente. Os mais bem-sucedidos, dentre estes, parecem ser os algoritmos baseados no *modelo da energia* para os neurônios do córtex visual [42,18,43].

Duas classes principais de neurônios binoculares (neurônios que respondem à estimulação de qualquer dos dois olhos) foram identificadas no córtex visual primário dos mamíferos superiores: as *células simples* e as *células complexas* [12]. As células binoculares simples somam linearmente os impulsos retiniais recebidos dos dois olhos, enquanto as células complexas combinam de uma forma quadrática as respostas que lhes chegam das células simples, sendo por isso chamadas de *neurônios energia*. Com base neste modelo para a resposta dos neurônios complexos, alguns pesquisadores já conseguiram bons resultados com abordagens estereoscópicas fisiologicamente plausíveis [22,36].

O objetivo do presente trabalho será justamente o de apresentar uma nova abordagem para o clássico problema da estereoscopia, tentando preservar a maior parte das características neurofisiologicamente plausíveis do modelo da energia. O ponto de partida para isto serão as funções de Green das equações de casamento de sinais. As funções de Green constituem uma ferramenta matemática tradicional para a solução de equações diferenciais não-homogêneas, encontrando vastas aplicações tanto na ciência quanto na engenharia [44]. Por outro lado, as equações de casamento de imagem são comumente empregadas em visão computacional e processamento de imagens - por exemplo, na estimação do fluxo óptico [24] e na própria estereoscopia [23].

Quando empregado no contexto das equações de casamento, o método da função de Green permite determinar, dada uma imagem de entrada, uma segunda imagem que corresponderia a uma versão deslocada da primeira. Tal propriedade já possibilitou a sua aplicação em processos como a detecção de bordas [45] e a estimação de forma tridimensional via *shape-from-shading* [40], e a sua extensão para o problema da estereoscopia representa um passo natural.

Conforme será aqui demonstrado, dado um par de imagens estereoscópicas, nós podemos filtrá-las, dentro de janelas apropriadas, por funções de Green que induzirão deslocamentos variados sobre as mesmas. Como as imagens estereoscópicas constituem visões deslocadas de uma mesma cena, ao induzir os movimentos adequados através das funções de Green, nós podemos colocá-las localmente em registro, desta forma inferindo o deslocamento original de cada ponto nas duas imagens, e portanto a sua profundidade na cena. Quando as janelas empregadas assumem a forma de funções de Gabor, nós demonstramos aqui que a abordagem estereoscópica via função de Green preserva muitos dos aspectos neurofisiologicamente plausíveis do modelo da energia.

O restante deste trabalho está organizado como segue: No Capítulo 2, nós introduzimos os fundamentos geométricos da estereoscopia, e apresentamos algumas das abordagens mais comuns para este problema. No Capítulo 3, nós fazemos uma breve dissertação sobre as bases neurofisiológicas da visão, culminando com a apresentação do modelo da energia para a visão estereoscópica. No Capítulo 4 é apresentada a abordagem da função de Green para a estereoscopia, cuja implementação e resultados experimentais aparecem no Capítulo 5. O Capítulo 6 encerra este documento com as nossas conclusões e propostas para trabalhos futuros.

2 Estereoscopia

Para entender, na prática, a importância da visão binocular, basta ao leitor tentar utilizar apenas um dos seus olhos na realização das suas atividades cotidianas. Logo ficará claro que ações em que se precisa alcançar objetos à frente já não serão mais triviais, pois a visão monocular não permite a aferição precisa de distâncias. O sistema visual utiliza vários sinais para inferir a dimensão de profundidade, mas a principal contribuição para este processo vem mesmo da visão binocular ou estereoscópica.

Neste momento uma pergunta interessante deve surgir. Se o mundo possui três dimensões e as imagens projetadas em nossas retinas apenas duas, qual seria o processo pelo qual o nosso sistema visual recupera esta dimensão perdida? A resposta encontrada por Charles Wheatstone, em 1838, teve como base a identificação da diferença relativa de posição entre os objetos visualizados por cada um dos nossos olhos, e a verificação de que esta informação possui relação direta com a percepção da profundidade da cena. Para exemplificar esta diferença, basta fixar o olhar num objeto distante e depois colocar um polegar na direção do objeto, de modo a tentar ocultá-lo (Figura 2-1). Observar-se-á que surgem duas imagens do dedo, cada uma provindo de um dos dois olhos. Tal separação das imagens, chamada 'disparidade', está relacionada com a distância entre os olhos e o dedo. No caso limite em que o polegar se afasta até o ponto de encostar-se no objeto focado, a separação visual entre os dedos será nula, e o polegar também estará em foco.

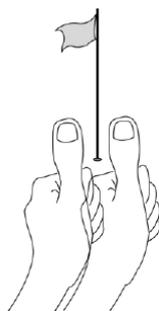


Figura 2-1: Efeito da Duplicidade Visual.

A duplicidade visual decorrente da visão binocular para os objetos que estão a uma distância inferior à distância focal. Neste caso os olhos estão focados no objeto ao fundo

[34].

O conceito de disparidade binocular foi introduzido para se quantificar esta diferença de posição entre os objetos da cena vistos por cada olho, resultante do fato de os dois olhos estarem lateralmente separados e verem o mundo de pontos de vista ligeiramente diferentes (Figura 2-2).

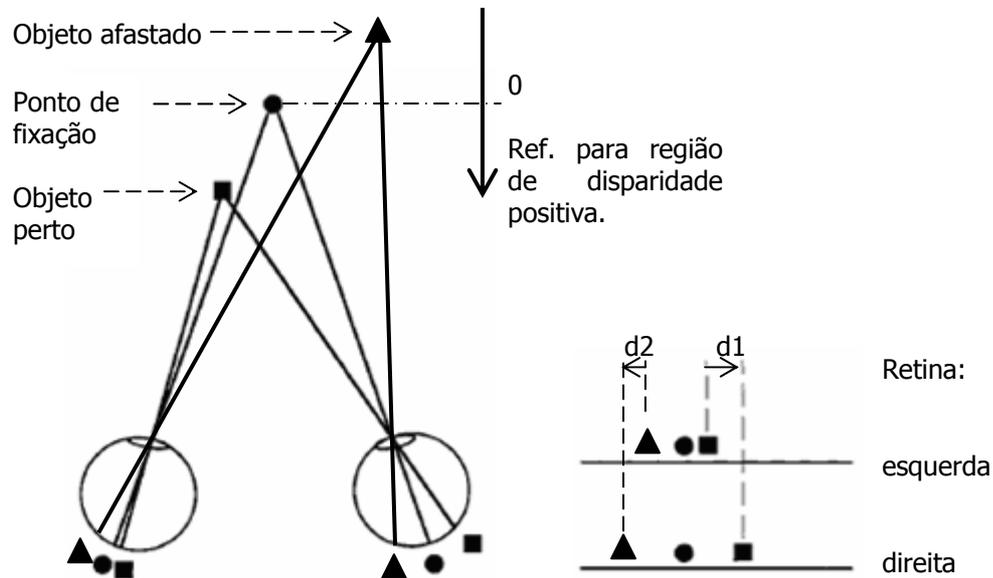


Figura 2-2: Geometria da Projeção Binocular

(Esq.) Geometria da projeção binocular nas retinas e referencial adotado para disparidades positivas. (Dir.) Disparidade binocular $d1$ para o objeto perto e $d2$ para o objeto afastado (adaptação da fig. 1 de [21]).

Apenas a projeção do ponto de fixação (ponto em que os olhos convergem) se dá nos centros das retinas (fóveas) dos dois olhos, regiões onde o objeto aparece em foco perfeito. Todos os demais pontos não se projetam em locais correspondentes nas duas retinas, e conseqüentemente possuem um deslocamento relativo à fóvea. Normalmente, esta separação é expressa em termos de ângulo visual.

Desde a descoberta de Wheatstone, métodos baseados em princípios matemáticos, psicofísicos e computacionais têm sido sugeridos na tentativa de desvendar os mecanismos utilizados pelo cérebro no cálculo das disparidades. Em 1960, Bela Julesz [10] deu uma importante contribuição neste sentido, ao demonstrar que a visão estereoscópica não depende de sinais monoculares, como os derivados da perspectiva, do sombreamento e do tamanho dos objetos,

para inferir a sua profundidade na cena. Ele conseguiu isso com a criação dos estereogramas de pontos aleatórios (Figura 2-3a), que são pares de imagens incorporando informação de disparidade, mas sem incluir dados de profundidade monoculares, já que se constituem de pontos distribuídos aleatoriamente. Para a simulação de uma região focada (disparidade nula), os pontos aleatórios se localizam de forma idêntica nas duas imagens, enquanto deslocamentos relativos, que podem ser positivos ou negativos, são efetuados nas demais regiões, simulando objetos a diferentes profundidades na cena. Quando cada imagem é vista individualmente, ela não traz informação tridimensional, aparecendo como um campo plano de pontos aleatórios. Porém, quando as duas imagens são vistas binocularmente - i.e., quando as imagens da esquerda e da direita são simultaneamente apresentadas, respectivamente aos olhos esquerdo e direito -, a informação de profundidade, codificada pelas disparidades, é então percebida, gerando uma representação tridimensional. Esta descoberta demonstrou que o cérebro pode computar disparidades binoculares independentemente de outros sinais visuais monoculares.

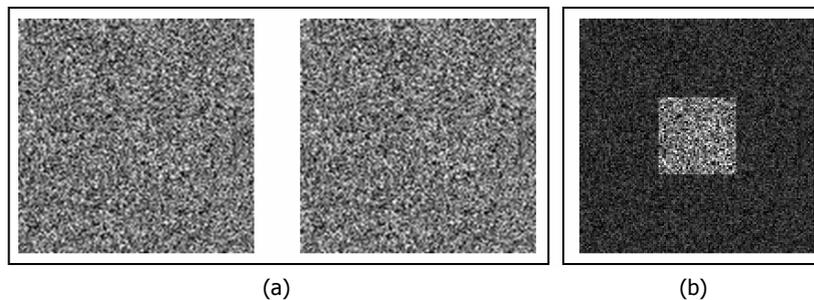


Figura 2-3: Estereograma de Pontos Aleatórios.

Em (a) tem-se um típico estereograma de pontos aleatórios, incorporando um mapa de disparidade de dois níveis, como representado em (b): um plano de fundo com um quadrado destacado na frente.

Antes de nos aprofundarmos no estudo da neurofisiologia da visão estéreo, nós iremos apresentar os conceitos geométricos básicos envolvidos neste processo.

2.1 Princípios Geométricos da Estereoscopia

A estereoscopia pode ser abordada sob diferentes perspectivas. Inicialmente, devemos salientar que, apesar de este processo haver sido originalmente inspirado na visão binocular humana,

nada impede que se faça uso de mais de duas imagens, em sua implementação computacional. Podem-se também assumir diferentes configurações de câmeras, como no caso das imagens capturadas por satélite – obtidas por uma única câmera em diferentes posições - ou das imagens obtidas simultaneamente por duas câmeras, simulando a visão humana, por exemplo, na navegação robótica. Nosso interesse aqui se concentrará na estereoscopia binocular, e segundo um modelo de projeção perspectiva em câmeras paralelas, conforme explicado a seguir.

O princípio que rege a geometria da visão estereoscópica é a triangulação. Conforme ilustrado na Figura 2-4, assumindo projeção perspectiva, cada ponto da cena estará registrado numa posição única em cada imagem capturada, pois para cada câmera há somente um centro de projeção. O ponto no objeto e o ponto na imagem direita ou esquerda estão ligados por uma reta que passa pelos centros projetivos direito e esquerdo, respectivamente. Com o conhecimento da localização dos pontos-imagem (**P1** e **P2**) e dos centros projetivos das câmeras (**C1** e **C2**) é possível determinar o ponto de intersecção do prolongamento das retas que unem estes pontos, ou seja, é possível determinar, por triangulação, a posição espacial do ponto-objeto (**P**) na cena.

Para isto, é preciso encontrar as coordenadas dos pontos projetados nas imagens 1 e 2. Este processo é conhecido como processo de correspondência ou de casamento, pois necessita-se associar os pontos de uma imagem com os pontos da outra. Esta etapa geralmente é a que requer maior processamento na estereoscopia, pois um casamento errado pode levar a inconsistência na geometria dos objetos da cena reconstruída. Toda a robustez da visão estéreo depende, portanto, da eficiência do processo de correspondência.

As principais dificuldades, no casamento estereoscópico, advêm, por um lado, da fotometria, e, por outro, da disposição geométrica das câmeras. A falta de controle sobre a iluminação da cena pode criar condições que impeçam a diferenciação dos pontos capturados nas imagens, principalmente devido aos erros e ruídos que são introduzidos durante o seu processo de digitalização. A outra dificuldade intrínseca relaciona-se à obtenção das imagens a partir de pontos de vistas diferentes, o que acaba por criar regiões visíveis somente em uma das imagens. Estas regiões, chamadas regiões de oclusão, não podem ser mapeadas quanto a sua profundidade na cena, através da técnica de triangulação.

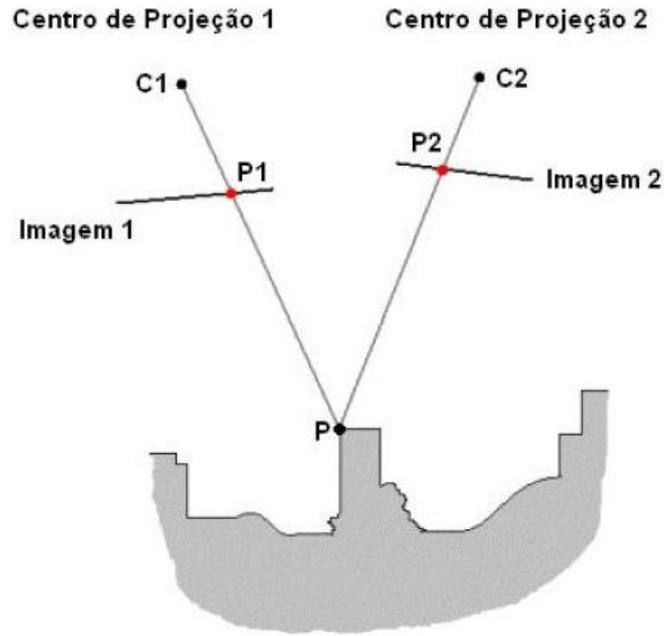


Figura 2-4: Triangulação Inerente à Estereoscopia.

No tocante à disposição das câmeras, nada se pode fazer de modo a minimizar o problema da oclusão quando se usa apenas duas câmeras, porém pode-se facilitar o processo de correspondência quando são tomados certos cuidados no posicionamento e no ajuste do par de câmeras.

Ao contrário do que pode parecer, o processo de correspondência para cada ponto-objeto não precisa ser uma busca bidimensional até se encontrar o par perfeito de pontos-imagem. A geometria epipolar permite simplificar muito este processo. Observe-se, na Figura 2-5, que os centros projetivos **C1** e **C2** das duas câmeras são coplanares aos pontos **P1**, **P2** e **P**. Este plano é chamado de plano epipolar (Π). A intersecção de plano epipolar com as imagens **I1** e **I2** produz as linhas epipolares **e1** e **e2**, respectivamente. Com base nesta geometria, pode-se observar que a projeção do ponto **P** cairá exatamente sobre as retas **e1** e **e2** das projeções **I1** e **I2**, respectivamente. Esta condição transforma o processo de correspondência num processo de identificar os pares de retas epipolares, e em seguida tentar encontrar a melhor associação entre os pontos destas duas retas.

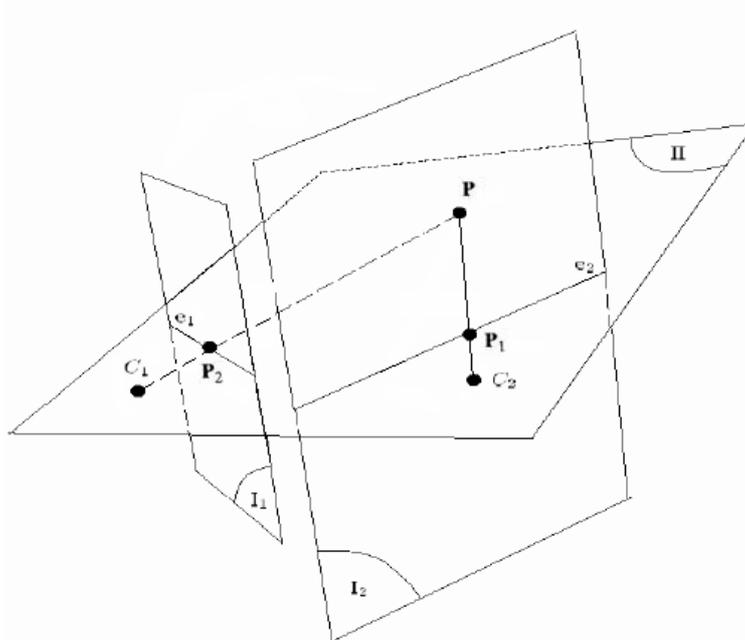


Figura 2-5: Geometria Epipolar.

Por uma questão de conveniência computacional, os planos ***I1*** e ***I2*** são geralmente escolhidos de modo a serem coplanares e paralelos à linha que liga os centros projetivos ***C1*** e ***C2*** (linha base). Tal configuração projetiva pode ser obtida fisicamente ou, mais facilmente, através de uma transformada analítica [29]. As imagens projetadas segundo esta condição são chamadas de imagens retificadas ou configuração de câmeras paralelas (Figura 2-6). A conveniência desta configuração é que as linhas epipolares para o estéreo geométrico retificado são colineares, o que simplifica o processo de correspondência a uma mera questão de comparar os vetores de pontos de ambas as projeções para cada valor da ordenada y .

Assim, extrair a componente de profundidade a partir de um par de imagens estéreo envolve basicamente três etapas: I) preparar as imagens adquiridas, isto é, filtrar os ruídos e retificar as imagens, II) realizar o processo de correspondência entre os pontos das imagens, e por fim III) construir o mapa de disparidades da cena. Destas etapas, fica evidente que a dificuldade maior advém do problema de correspondência. A seguir, nós iremos analisar os prós e os contras das diferentes técnicas já desenvolvidas para atacar este problema.

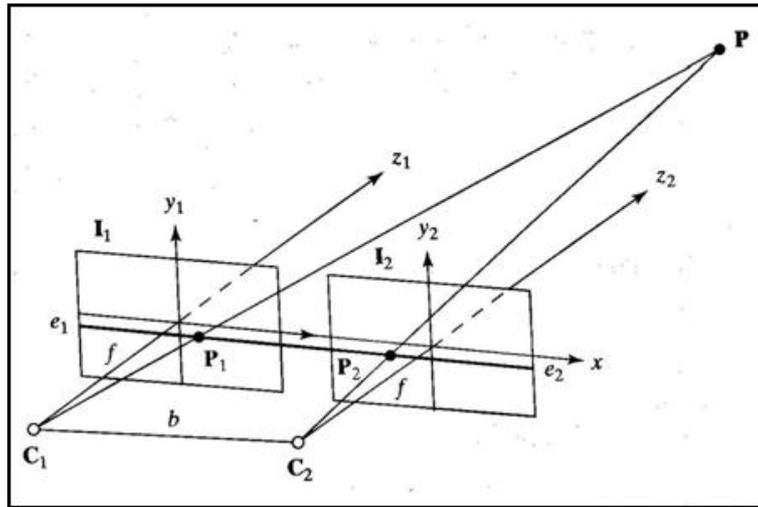


Figura 2-6: Geometria Epipolar para Projeção Retificada.

As imagens retificadas são obtidas geralmente por meio de transformações analíticas produzidas a partir das imagens projetadas.

2.2 Abordagens para Estereoscopia

A etapa que requer maior cuidado, nas diferentes abordagens para estereoscopia, está no processo de correspondência. Com base nas referências [5] e [6], nós podemos classificar os algoritmos de correspondência em cinco tipos: Baseados em Características, Baseados em Área, Baseados no Pixel, Baseados na Fase, e Baseados na Energia. A seguir, nós apresentaremos as principais características de cada uma destas abordagens.

2.2.1 Correspondência por Características

Várias propriedades podem ser usadas para descrever os objetos da cena, e algumas são facilmente identificadas em ambas as imagens projetadas. Esta é a idéia por trás dos algoritmos de correspondência baseados em características: tentar identificar características nas imagens que possam ser utilizadas no processo de correspondência. Picos no gradiente de intensidade, o formato dos cantos e bordas dos objetos, e segmentações de regiões e texturas nas imagens são algumas das variações utilizadas dentro desta técnica (veja Tabela 2-1). O problema de

correspondência passa então a ser o de procurar o melhor casamento entre estas características identificadas nas duas imagens.

Uma vantagem observada neste método está na redução da quantidade de informações utilizadas para a correspondência. Por outro lado, uma vez que as características identificadas estão distribuídas aleatoriamente ao longo da cena (veja Figura 2-7), os resultados produzidos por este processo são mapas de disparidades esparsas, fazendo-se necessário o uso de alguma técnica de interpolação, ou de algum outro critério complementar, para preencher as lacunas de disparidade sem introduzir muito erro.

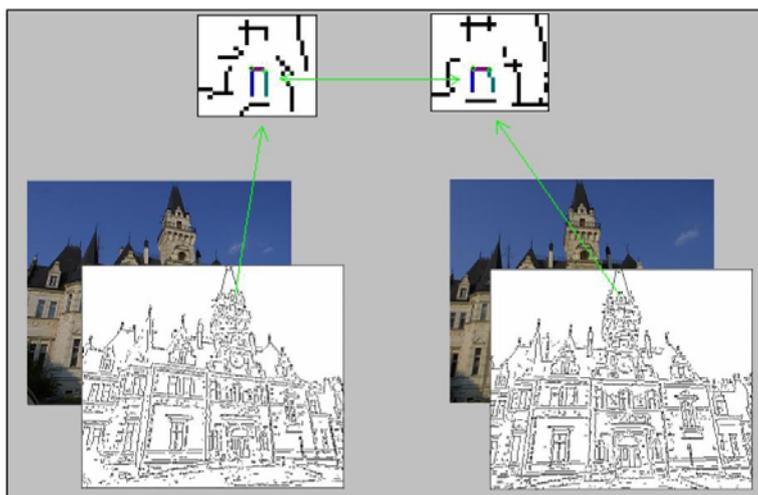


Figura 2-7: Correspondência por Características.

Exemplo de segmentos de retas utilizados no processo de estereoscopia por características.

Um dos problemas que podem ocorrer com este tipo de técnica está na possível ambigüidade das características segmentadas. Para se superar isto, algumas restrições podem se tornar necessárias, como, por exemplo, fazer uso do critério de continuidade no valor da disparidade, com base na vizinhança. Por outro lado, nesta abordagem o processo de casamento é significativamente mais rápido, quando comparado com as técnicas baseadas em área ou em pixel, conforme veremos a seguir.

Características	Detalhamento
Bordas	Orientação e transição de intensidade positiva ou negativa.
Cantos	Coordenadas dos cantos, tipo de junção nos cantos (tipo Y, L, A, etc).
Segmentos de linhas	Coordenadas dos pontos extremos, pontos médio, e orientação das linhas.
Regiões	Áreas internas, linhas de contorno, e centróides.

Tabela 2-1: Principais Tipos de Características Utilizadas para Correspondência.

2.2.2 Correspondência por Área

Esta é uma das técnicas mais utilizadas, e se baseia na premissa de que as disparidades na vizinhança de um pixel são aproximadamente constantes. Geralmente se utilizam as próprias intensidades ao redor do pixel para o processo de casamento. Para se quantificar o grau de casamento entre as janelas, são geralmente utilizados os métodos de máxima correlação ou de mínima diferença quadrática entre as intensidades.

Com um tamanho apropriado de janela, podem-se reduzir as ocorrências de falsos positivos (ambigüidades de intensidades) durante a busca por correspondência entre as regiões de cada imagem do par (**Figura 2-8**). Porém, um novo parâmetro é introduzido no processo: o tamanho da janela. Escolher o tamanho adequado e o formato da janela não é um processo simples, mas depende de se conhecerem informações *a priori* sobre as imagens (por exemplo, a sua resolução) e a cena (a disposição dos objetos), ou então de se empregarem técnicas dinâmicas de ajuste da janela ao longo do processo. No final das contas, o resultado produzido acaba trazendo alguns erros, pelo não atendimento da premissa de disparidade constante dentro das janelas.

Por outro lado, uma característica positiva deste tipo de abordagem é a obtenção de mapas de disparidades com densidade de média para alta, não havendo em geral a necessidade de se aplicar uma etapa de interpolação para preencher lacunas nos mapas.

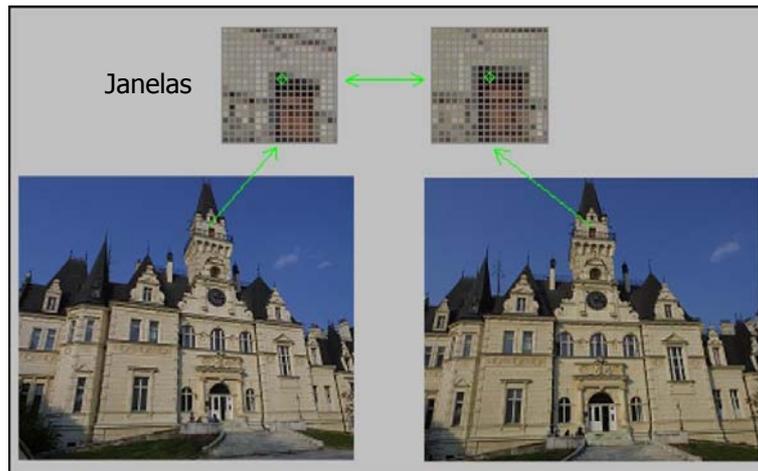


Figura 2-8: Correspondência por Área.

Exemplo de par de janelas usado no processo de casamento por área.

2.2.3 Correspondência por Pixel

Esta abordagem pode ser vista como o caso limite da abordagem por área, quando a janela reduz-se a apenas um pixel. O casamento do par de imagens é feito então ponto a ponto, e com base em informações (em geral intensidades) armazenadas em cada ponto. As primeiras propostas para este tipo de abordagem buscavam minimizar a diferença de intensidade das imagens da esquerda e direita, porém sujeitas à restrição de continuidade [3] e [4]. Recentemente, novas técnicas mais elaboradas foram incorporadas, para se lidar com os problemas de oclusões e mínimos locais. Basicamente, todas as melhorias foram alcançadas a partir de novas restrições incorporadas à função de erro, como também fazendo uso de abordagens inspiradas em técnicas de otimização [6].

Duas conseqüências diretas podem ser associadas a este tipo de abordagem: a obtenção de mapas de disparidades com densidade alta, e o elevado custo computacional, devido ao processo de correspondência ser proporcional ao número de pixels nas imagens.

2.2.4 Correspondência por Fase

O método de correspondência por fase [27,28] baseia-se no resultado matemático de que o deslocamento espacial de um sinal produz um deslocamento proporcional na sua fase, quando o consideramos no domínio da frequência. Dado o sinal $f(x)$, tem-se, pela teoria de Fourier, que este pode ser caracterizado por um conjunto de diferentes componentes de frequência ω , cada uma possuindo uma amplitude $|F(\omega)|$ e uma fase $\phi(\omega)$. Simbolicamente, esta relação pode ser representada como

$$f(x) \mapsto |F(\omega)| \cdot \exp[i\phi(\omega)] \quad (2.1)$$

onde o lado direito é a transformada de Fourier de $f(x)$.

A transformada de Fourier apresenta a seguinte propriedade, conhecida como *propriedade do deslocamento*: Se a relação (2.1) é válida, vale também

$$f(x-d) \mapsto |F(\omega)| \cdot \exp[i(\phi(\omega) - \Delta\phi(\omega))] \quad (2.2)$$

onde $\Delta\phi(\omega) = \omega d$. Isto significa que, ao se deslocar espacialmente um sinal, o módulo da sua transformada de Fourier não se altera, mas sim a sua fase; e esta alteração de fase é proporcional ao deslocamento espacial do sinal. Como as representações de uma dada cena 3D nas imagens direita e esquerda de um sistema estereoscópico diferem localmente por um deslocamento espacial, pode-se utilizar este resultado para obter uma estimativa para o mapa de disparidades da cena: sendo $\phi_r(\omega)$ e $\phi_l(\omega)$ as fases em uma dada região das imagens direita e esquerda, respectivamente, tal estimativa seria dada por

$$d = \frac{\phi_r - \phi_l}{\omega} \quad (2.3)$$

A principal vantagem do processo de correspondência por fase está na rapidez dos cálculos. Alguns pesquisadores chegaram a propor uma abordagem que faz uso da transformada rápida de Fourier (FFT-1D) para o cômputo de disparidades em sistemas de tempo-real [26]. Por outro lado, a dificuldade do método consiste essencialmente em obter a informação de fase local de modo confiável. Neste sentido, a maioria das abordagens acaba por lançar mão do uso de técnicas que ponderam a confiabilidade da(s) fase(s) encontrada(s) por meio de restrições de continuidade da disparidade numa dada vizinhança, tanto no domínio espacial quanto no domínio da frequência (neste último caso, considerando-se continuidade das fases).

2.2.5 Correspondência por Energia

O chamado *modelo da energia*, para o cálculo das disparidades binoculares, foi inspirado por descobertas relativas à neurofisiologia da visão estereoscópica em animais superiores, e será abordado no Capítulo 3.

3 Neurofisiologia da Visão

Durante o processo de evolução das espécies, houve um momento em que um grupo de animais passou a apresentar os dois olhos posicionados à frente da cabeça. Essa alteração foi sendo desenvolvida e aperfeiçoada com o passar dos anos, pois se tratava de uma característica vantajosa para a sobrevivência daqueles animais, possibilitando-lhes uma nova capacidade visual: a estereoscopia. [35]. O efeito estereoscópico é produzido através da fusão, pelo cérebro, das imagens observadas pelos dois olhos, resultando na noção de profundidade, como é ilustrado na Figura 3-1. Aqui, nós pretendemos apresentar, de forma bastante resumida, as estruturas e os mecanismos cerebrais envolvidos no processamento estereoscópico.

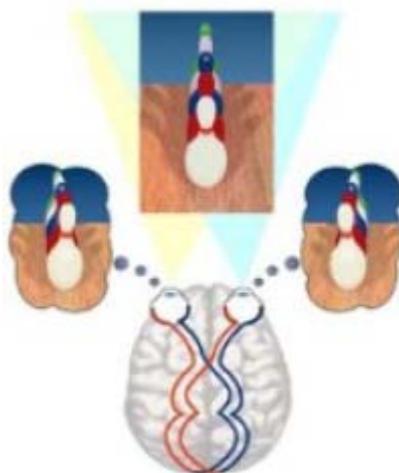


Figura 3-1: A Diferença de Percepção em Cada Olho.
Visão em cada olho para uma mesma cena [33].

Os estudos realizados sobre o cérebro humano ainda não atingiram a sua plenitude, e conseqüentemente não se pode determinar com precisão todas as áreas relacionadas com o processamento visual. Mesmo sendo o sistema perceptivo aquele em que os estudos de neurofisiologia mais evoluíram, ainda há muitas perguntas sem respostas.

Nós concentraremos a nossa atenção nas faculdades do cérebro voltadas ao processamento visual, mas, antes de iniciarmos propriamente o estudo da visão, nós precisamos adquirir um mínimo de conhecimento sobre as células nervosas.

3.1.1 As Células Nervosas

As células nervosas são as estruturas elementares para se criar outras estruturas cerebrais maiores e mais específicas a um determinado propósito. Elas não são as únicas estruturas identificáveis no cérebro, porém são aquelas que estamos interessados em conhecer melhor, por serem as estruturas elementares do sistema visual.

Uma célula nervosa (veja Figura 3-2) é composta basicamente por um corpo globular onde está localizado o núcleo da célula. A partir do corpo globular projetam-se dois tipos de ramificações: os axônios, que são fibras nervosas que levam o sinal até outras células nervosas; e os dendritos, que trazem sinais de outras células até o corpo globular. Além destes elementos, existe uma estrutura que envolve todo o corpo, os axônios e os dendritos, denominada membrana celular.

As regiões em que os dendritos e axônios de células distintas se interligam são chamadas de sinapses. Nestas junções nervosas, não chega a haver contato físico entre as duas células; a informação é transmitida de uma célula para outra por meio de agentes químicos específicos, os neurotransmissores.

As células nervosas estão longe de serem todas iguais, existindo tantos tipos diferentes que não se sabe ao certo o seu número (estima-se que este possa ser até superior a mil [11]), mas as células nervosas que desempenham um papel similar estão geralmente interconectadas, compartilhando de alguma forma as suas informações.

Quando uma estrutura precisa trocar informação com outra, faz uso dos axônios, que, quando muitos longos, geralmente surgem também agrupados, tomando a forma de condutores. Estas ligações entre diferentes estruturas podem ser modeladas como o percurso de um caminho com vários estágios em série.

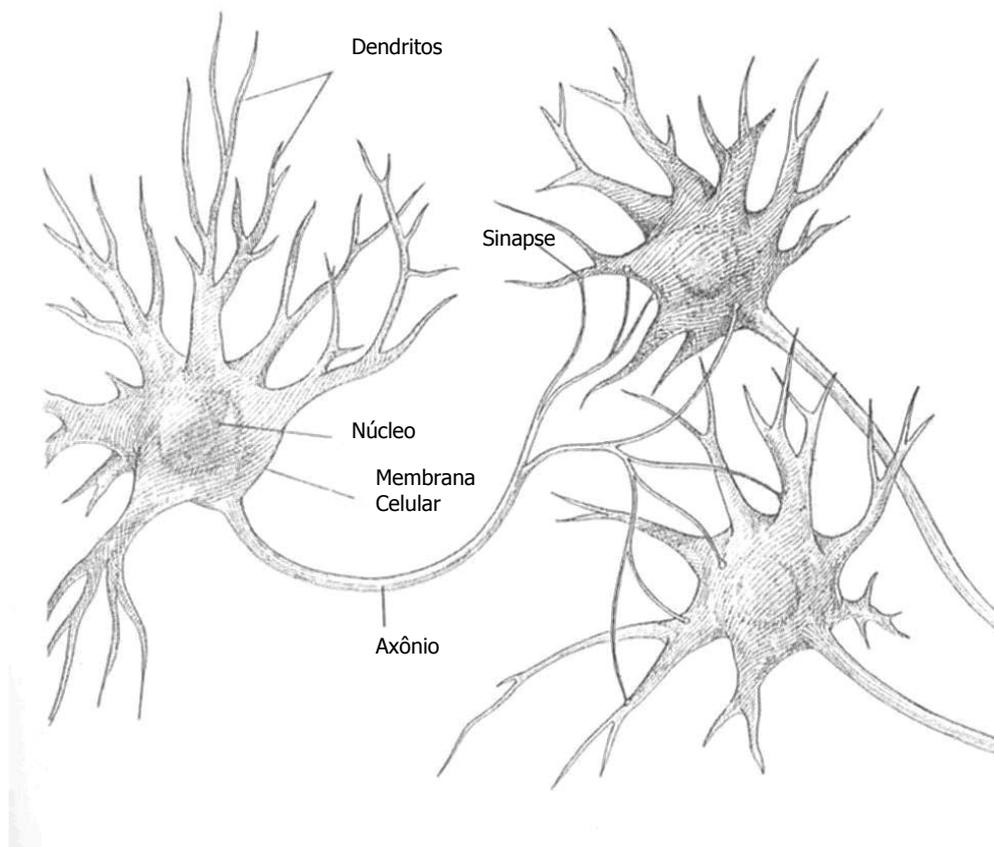


Figura 3-2: Estrutura de uma Célula Nervosa.

Estas são as principais partes que compõem a célula nervosa [11]

Um bom exemplo deste processo de conexão serial é o próprio *caminho visual*, isto é, a seqüência de estágios percorridos pela informação visual no nosso cérebro (veja Figura 3-3). A luz que incide na *retina* segue pelo *nervo óptico* (conjunto de axônios) até o *corpo lateral geniculado* que, por sua vez, envia os sinais nervosos até o *córtex estriado*. Vale ressaltar que, ao longo do caminho visual, outras áreas do cérebro também podem ser alimentadas, desde que o tipo de informação disponível num estágio seja útil para outras funções cerebrais.

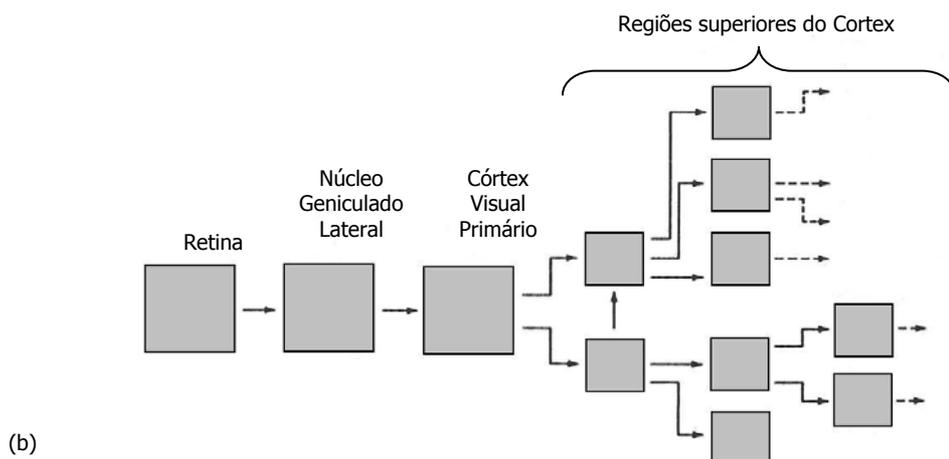
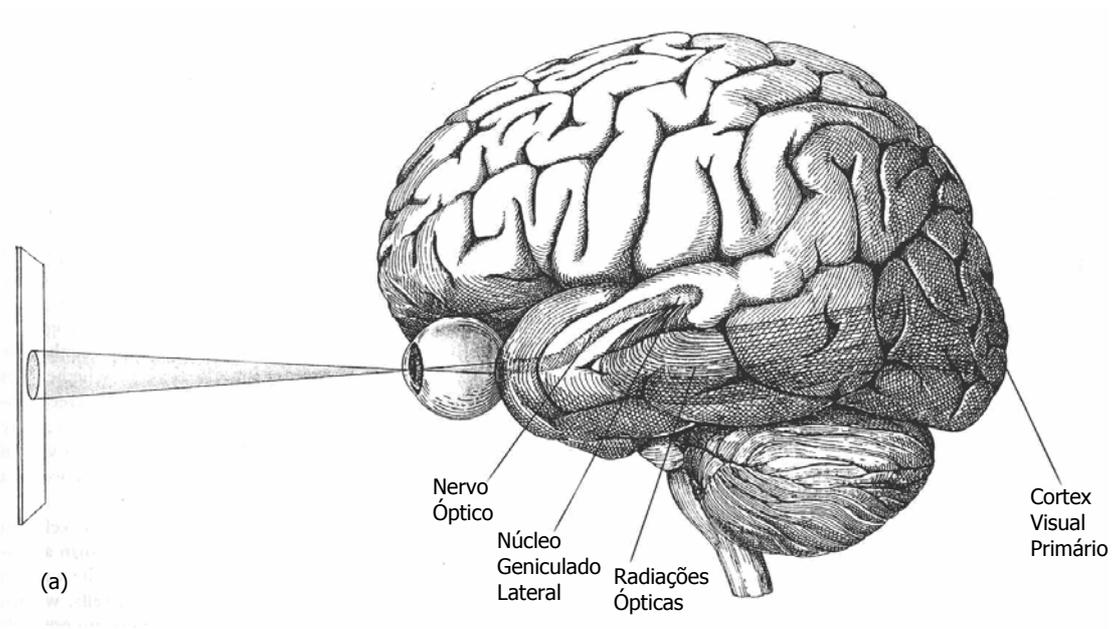


Figura 3-3: O Caminho Visual e sua Representação em Blocos.
 (a) o Caminho percorrido pelos sinais visuais dentro do cérebro e (b) sua representação em blocos para o fluxo de informação [11].

Percorrendo o caminho visual, a primeira estrutura que devemos analisar é a retina, local onde a luz recebida pelos olhos é transformada em sinal nervoso. Depois analisaremos como este

sinal é transformado em cada estágio seguinte do caminho visual, até chegarmos às células simples e complexas do córtex visual primário, que têm papel fundamental na estereoscopia.

3.2 Retina

A retina é a parte neural dos olhos (veja Figura 3-4), onde ocorre a tradução da luz em sinais nervosos. A retina já faz parte do cérebro, constituindo uma extensão deste para dentro dos olhos. Pode-se dizer que ela é a antena que capta os sinais luminosos do mundo externo e os prepara para que outras regiões do cérebro possam processá-los. A retina tem a forma de uma membrana com espessura da ordem de um quarto de milímetro, sendo composta por três camadas de células nervosas sobrepostas: os fotorreceptores, as células bipolares e as células ganglionares.

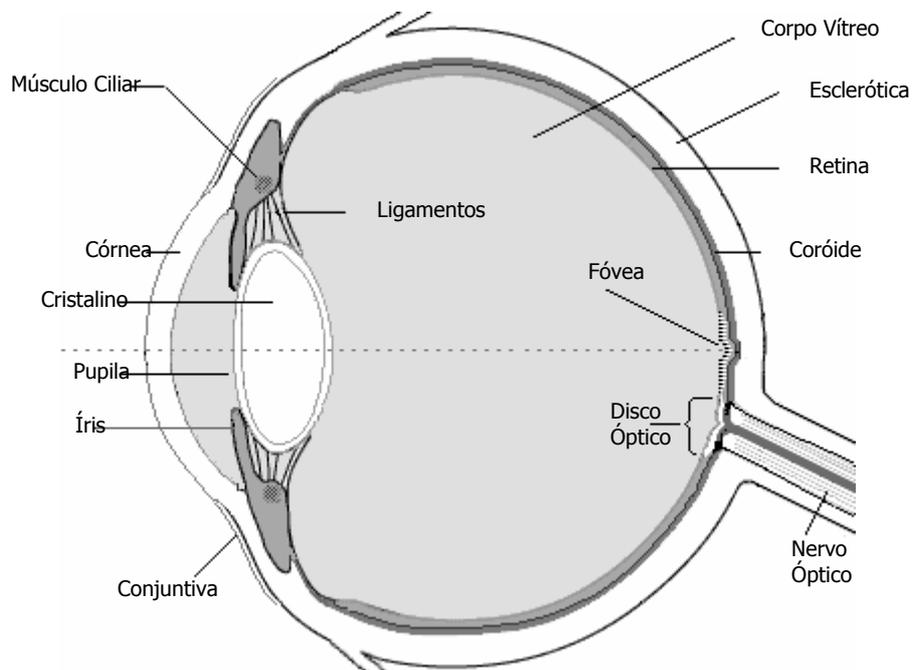


Figura 3-4: Estrutura Interna do Olho Humano.

A primeira camada fica no fundo da retina, e é composta por receptores fotossensíveis. Existem dois tipos de fotorreceptores na retina, diferenciáveis facilmente segundo a sua forma: os cones

e os bastonetes. Na retina humana existem cerca de 3 milhões de cones, pouco sensíveis à luz, mas que produzem imagens em cores e com definição elevada; e 100 milhões de bastonetes, mais sensíveis à luz, mas que produzem imagens em preto-e-branco. Existe uma pequena região no centro da retina onde a nossa acuidade visual é maior. Nesta região, identificada como fóvea, todos os fotorreceptores são cones.

As camadas na frente dos fotorreceptores são translúcidas e não geram um espalhamento significativo da luz que passa por elas. Na região da fóvea, mesmo pequenos desvios poderiam causar distorções desastrosas. Para evitar este problema, nesta região houve um rearranjo dos cones, de modo que eles ficassem diretamente expostos à luz, empurrando para os lados as células das outras camadas e formando um vale no meio da região da fóvea.

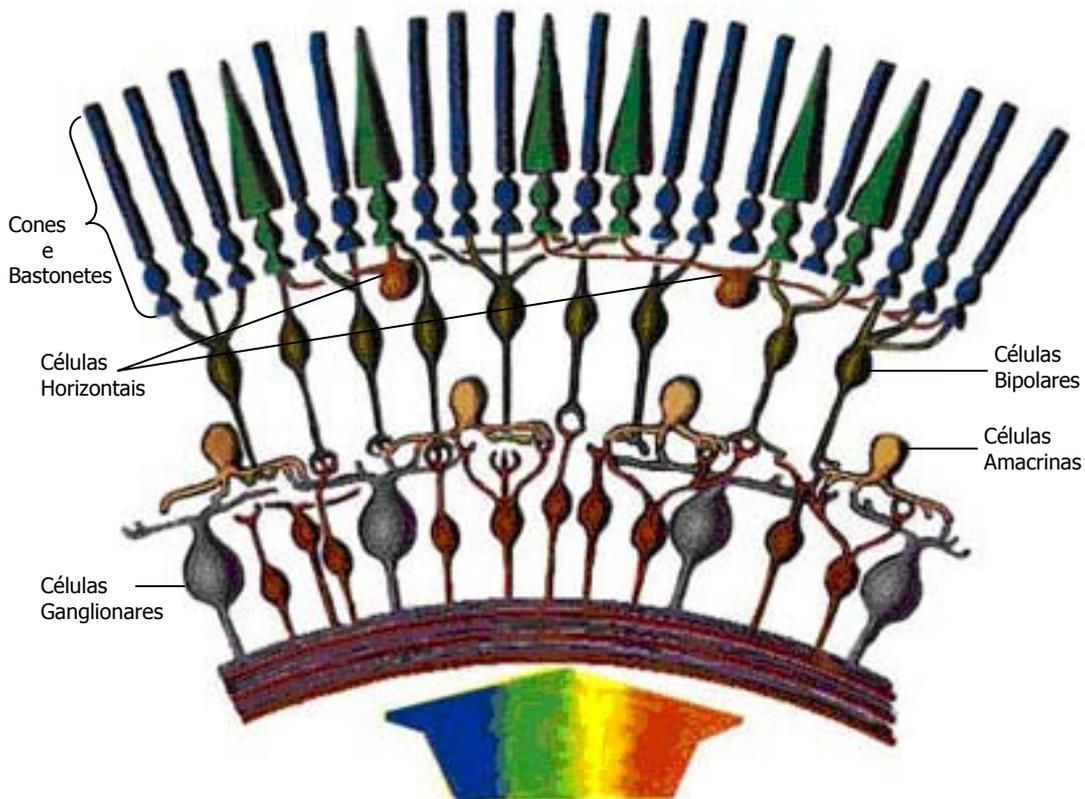


Figura 3-5: Organização das Células na Retina.

Secção da retina mostrando suas camadas e a organização de cada tipo de célula.

A segunda camada de células nervosas é composta principalmente pelas células bipolares. Estas células servem basicamente para ligar os receptores às células ganglionares. Geralmente as células bipolares estão ligadas a dois ou mais fotorreceptores, fazendo assim uma convergência e superposição dos sinais luminosos. Além das células bipolares, nesta camada intermediária existem mais dois tipos de células: as células horizontais e as células amacrinas. As horizontais ficam próximas das sinapses dos fotorreceptores com as células bipolares, e se ligam a um grande número de fotorreceptores, criando assim um caminho alternativo para o fluxo dos sinais. De forma similar, as células amacrinas também se interligam a várias células ganglionares, mas ficam mais próximas das sinapses das células bipolares com as células ganglionares, conforme a Figura 3-5 ilustra. As células ganglionares formam a terceira camada da retina. Estas células possuem longos axônios que caminham paralelos à retina em direção a uma cavidade próxima ao centro da retina, por onde saem formando o chamado nervo óptico.

3.2.1 Campo Receptivo - CR

Antes de estudarmos mais detalhes sobre os fotorreceptores e as demais células que formam a retina, vamos examinar qual é o sinal que sai da retina pelas as células ganglionares.

Nos estudos sobre a retina, foi identificado que, mesmo em condições de ausência completa de luz incidente, ainda se detectam sinais nervosos, com taxa quase constante, em todas as células ganglionares. As taxas destes impulsos variam de um único impulso até vinte impulsos por segundo. Iluminando-se uma determinada região da retina com um pequeno feixe de luz e medindo-se a atividade das células ganglionares associadas com esta região iluminada, observou-se que havia uma relação entre o tamanho e a forma do feixe luminoso e a taxa de impulsos emitidos. Para algumas células ganglionares, o máximo de emissão de impulsos ocorria quando se aumentava o diâmetro de um feixe luminoso de forma circular até um determinado limite. Depois deste limite, a atividade ia decaindo até estabilizar-se na mesma taxa da ausência de luz (veja Figura 3-6). Verificou-se que várias células ganglionares apresentavam este mesmo comportamento, e estas foram denominadas células centro-ativo.

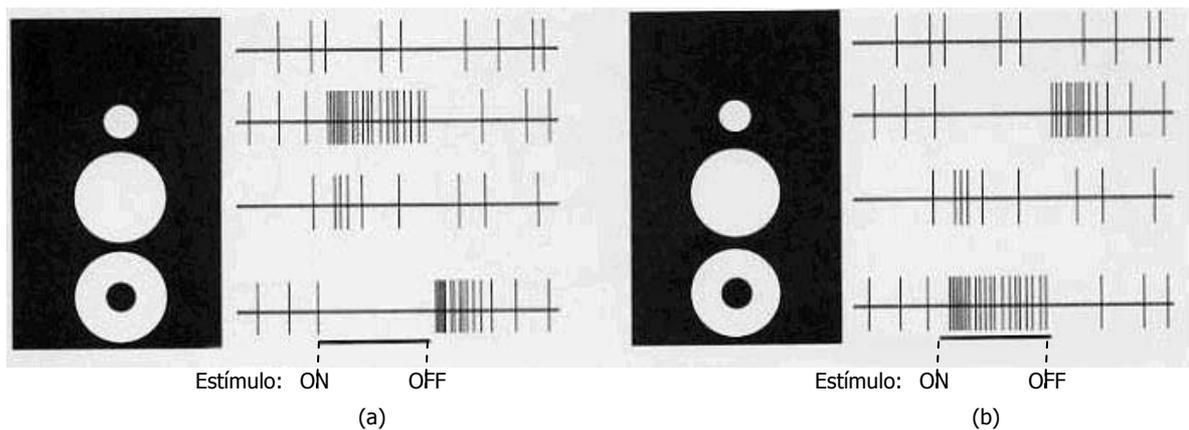


Figura 3-6: Resposta das Células Centro-ativo e Centro-inativo.
 Ilustração da resposta das células centro-ativo (a) e centro-inativo (b) a estímulos luminosos de diferentes formas.

Um segundo grupo de células ganglionares foi identificado durante os experimentos. Neste grupo observou-se que, quando o formato do feixe luminoso era circular (conforme usado no experimento das células centro-ativo), não existia alteração na taxa de impulsos emitidos pelas células. Porém, quando se alterava o formato do feixe para um círculo grande com o centro escuro, como num anel, estas células aumentavam sua atividade proporcionalmente à espessura do anel de luz até certo limite. Este grupo de células foi denominado como células centro-inativo (Figura 3-6).

A combinação das duas regiões da retina que estão relacionadas com a resposta de uma célula é identificada como o *campo receptivo* desta célula (explicar melhor o que é campo receptivo). A Figura 3-7 ilustra a metade do campo receptivo de células ganglionares, com sua região central e de vizinhança, que passaremos a referenciar como campo centro-vizinhança. É importante, porém, ter em mente que o conceito de campo receptivo se aplica a qualquer tipo de célula do caminho visual.

O tamanho dos campos receptivos muda ao longo da retina. Eles são menores na região da fóvea e progressivamente aumentam à medida que nos afastamos dela. Isto significa dizer que a nossa precisão visual é maior na fóvea e menor nas regiões periféricas. De onde se conclui que, em termos de precisão visual, o que importa é o diâmetro do campo receptivo.

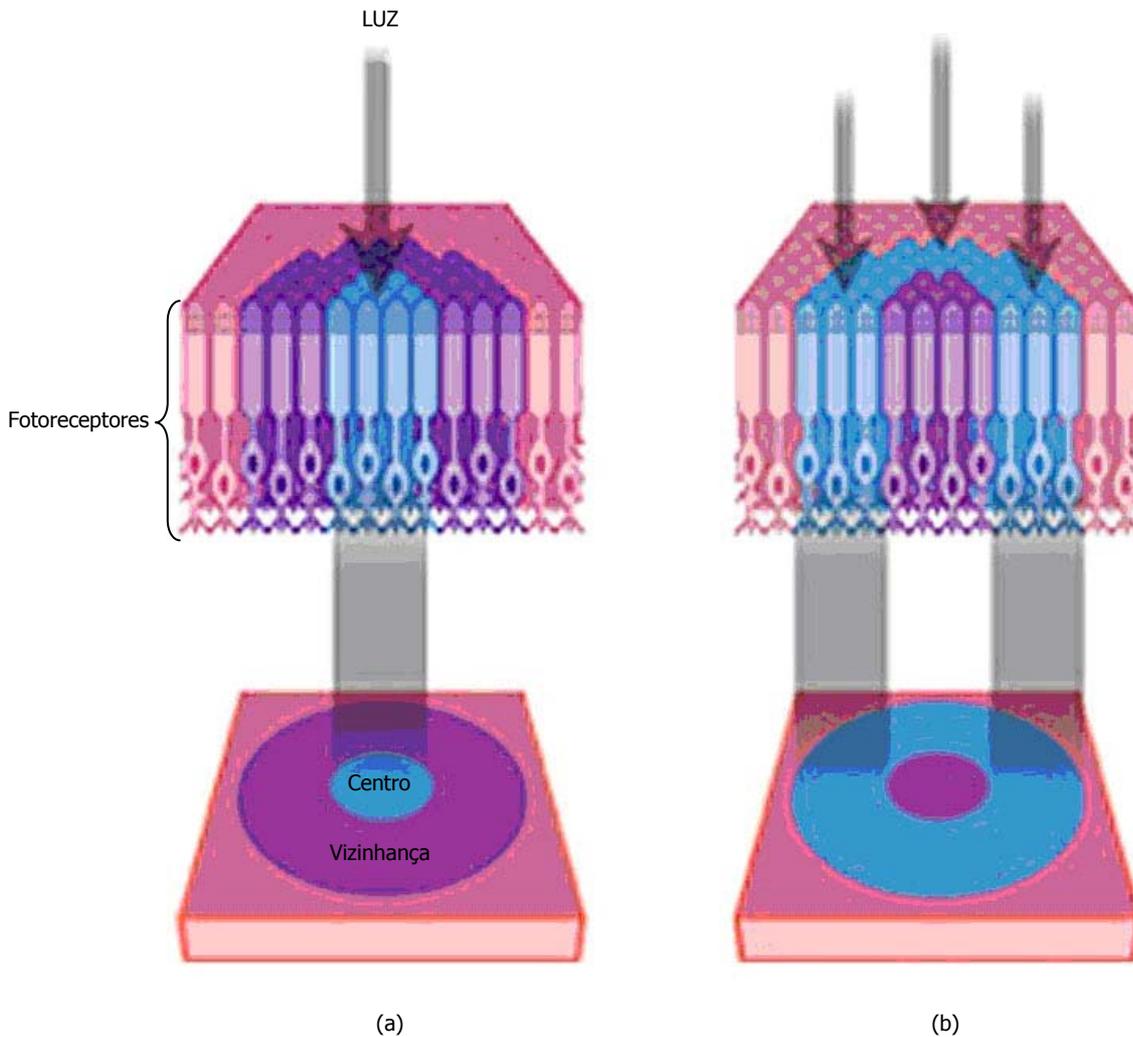


Figura 3-7: Estrutura dos Fotorreceptores para Células Centro-ativo e Centro-Inativo. Ilustração da estrutura dos fotorreceptores para células ganglionares do tipo: (a) centro-ativo (on-center) e (b) centro-inativo (off-center).

3.2.2 Fotorreceptores

Os fotorreceptores respondem à luz através de um processo onde uma molécula de pigmento visual absorve um fóton de luz visível e se transforma quimicamente num novo componente que não consegue mais absorver a luz tão bem. Esta nova substância produz uma alteração na membrana celular dos cones e bastonetes, em resposta à luz, que vai alterar a taxa de

liberação de neurotransmissores (glutamato) nas sinapses com as células bipolares. Estas por sua vez irão retransmitir o sinal de acordo com a taxa de liberação do neurotransmissor.

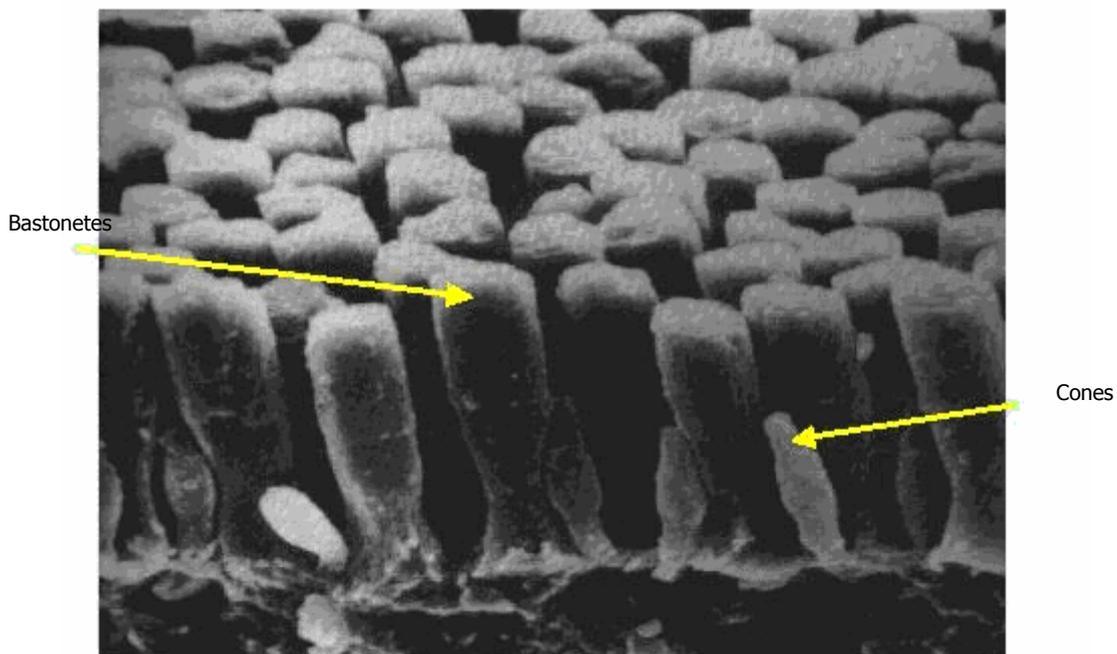


Figura 3-8: Os Tipos de Fotorreceptores da Retina Humana.

Há dois tipos de fotorreceptores: cones e bastonetes; e eles diferem em vários pontos, principalmente no que se refere à sensibilidade à luz. Os cones são menos sensíveis, quase não tendo utilidade em baixas intensidades luminosas, motivo pela qual a nossa visão noturna é pouco definida e basicamente em tons de cinza. Na retina humana existem três tipos de cones, cada um contendo um tipo ligeiramente diferente de pigmento (opsina). A diferença reside no comprimento de onda luminosa que cada um dos tipos de pigmento absorve: azul, verde ou amarelo/vermelho. Comprimentos de onda intermediários excitam os pigmentos de forma proporcional à distribuição espectral de energia sobre os comprimentos de onda em que os fotorreceptores são sensíveis.

3.2.3 Células Bipolares, Horizontais e Amacrinas

As células bipolares ocupam uma posição estratégica na retina, pois todos os sinais que se originam nos fotorreceptores têm que passar por elas para chegar às células ganglionares.

Todas as características de células centro-ativo e centro-inativo identificadas nas células ganglionares também foram identificadas nas células bipolares, o que significa que as células bipolares é que determinam as regiões de campo receptivo centro-vizinhança.

As duas camadas de neurônios restantes nesta zona intermediária da retina são as células horizontais e amacrinas, que se comunicam lateralmente através da retina.

As células horizontais ligam pares de fotorreceptores vizinhos com as células bipolares. Assim, a comunicação entre um receptor e uma célula bipolar é influenciada pela quantidade de luz recebida pelos receptores vizinhos. Este fluxo lateral de informação permite à retina regular a acuidade visual e o contraste entre luz e sombra. Além disso, elas são provavelmente responsáveis, pelo menos em parte, pela formação da vizinhança em forma de anel no campo receptivo observado nas células ganglionares e bipolares. Em suma, as células horizontais recebem informação dos fotorreceptores, porém não há uma saída bem determinada para esta informação.

As células amacrinas ligam pares vizinhos de células bipolares com as células ganglionares. Não é claro, até agora, o papel destas, embora algumas células amacrinas sejam altamente sensíveis a alterações de iluminação ou movimento.

3.3 Núcleo Geniculado Lateral e Córtex Visual Primário

Embora não se tenha ainda um completo conhecimento do sistema visual, os primeiros estágios do caminho visual já foram mapeados. Este caminho possui uma característica muito importante para o nosso estudo: sua estrutura está organizada, o que significa dizer que pontos vizinhos na retina continuam sendo pontos vizinhos no núcleo geniculado lateral e também nos primeiros estágios do córtex visual primário, o que significa dizer que todos compartilham o campo receptivo das células retiniais.

Após os sinais nervosos deixarem a retina, eles seguem pelo nervo óptico até o estágio seguinte, denominado núcleo geniculado lateral. No meio do caminho entre a retina e o núcleo geniculado lateral existe um cruzamento parcial dos nervos ópticos que saem de cada olho.

Este ponto é denominado quiasma. A Figura 3.9 ilustra este cruzamento. Os sinais provenientes do hemisfério direito da retina de ambos os olhos são encaminhado para o núcleo geniculado lateral do lado direito. Similarmemente, os sinais do hemisfério esquerdo da retina vão para o núcleo geniculado lateral esquerdo. Após a passagem pelo núcleo geniculado lateral, os sinais visuais seguem através de uma longa via pelo interior do cérebro até chegarem à sua parte posterior, onde ocorrem as sinapses com as células nervosas do córtex visual primário.

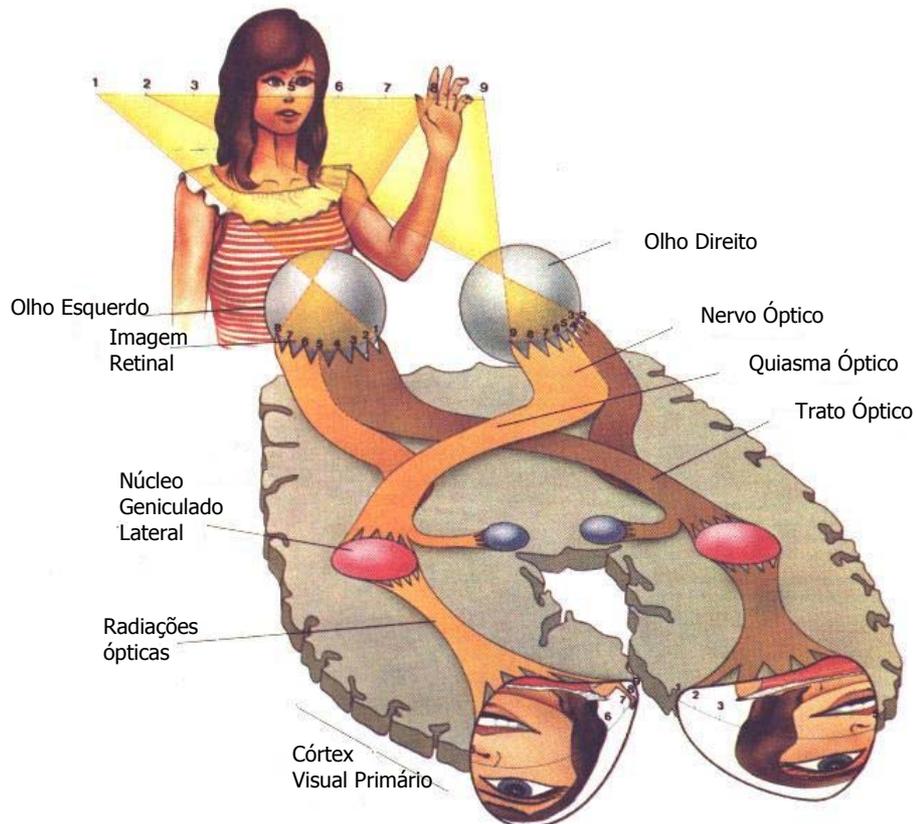


Figura 3-9: Mapeamento no Córtex Visual de um Objeto.

Na Figura 3.9, os olhos estão focando o nariz da pessoa, ponto referenciado com o número 5. Este ponto central é mapeado na fóvea, centro projetivo da retina. O termo campo visual refere-se ao campo de visão do ambiente externo que estamos olhando. Mais precisamente, o campo visual esquerdo corresponde à região do campo visual à esquerda do ponto central do que estamos observando e o campo visual direito fica à direita deste ponto. O referencial para se determinar se a região externa pertence ao campo visual esquerdo ou direito é o ponto

central do que estamos olhando. É importante distinguir que campo visual é o que nós vemos no mundo exterior, enquanto que campo receptivo corresponde a uma pequena parcela do mundo externo visto por uma única célula, conforme discutido anteriormente.

Desde o núcleo geniculado lateral, a informação visual ascende ao córtex cerebral, onde será novamente processada, interpretada e memorizada.

3.3.1 As Células do Córtex

O nosso principal objetivo é entender como as células do córtex respondem aos estímulos visuais, em particular aos binoculares. Inicialmente, necessitamos saber se as características centro-ativo e centro-inativo ainda persistem nestas células, e qual a sua contribuição dentro do sistema visual.

O córtex visual primário, ou córtex estriado, tem a forma de uma camada muito fina (cerca de 2 mm de espessura), porém com uma longa extensão, chegando a abrigar cerca de 200 milhões de células.

Devido a esta grandeza, entender o fluxo de informação no córtex não é uma tarefa trivial. Sabe-se que o córtex possui algumas conexões com outras áreas do cérebro que ainda não estão bem mapeadas. Entretanto, o nosso objetivo neste estudo é entender o funcionamento das estruturas iniciais, a parte mais simples do córtex, de cujo funcionamento já se tem uma boa compreensão.

As células dos primeiros estágios do córtex respondem de forma bem similar às células geniculadas. O campo receptivo delas tem simetria circular. Estas células alimentam uma segunda região de células que possuem características novas, ainda não identificadas anteriormente no caminho visual.

Descobrir os estímulos necessários para se ter alguma resposta neste novo grupo de células foi um passo fundamental para se entender o seu papel. Quando estimuladas com retângulos de luz de espessura desprezível em relação ao comprimento (vamos referenciá-los como linhas), verificou-se que estas células levavam em consideração a orientação dos estímulos e o sentido

em que eles se movimentavam. As linhas usadas eram de três tipos: I) uma linha iluminada sobre um fundo escuro; II) uma linha escura sobre um fundo claro; ou III) uma linha de borda entre uma região clara e outra escura. Durante os experimentos, observou-se que uma célula típica que responde bem a estímulos orientados para uma determinada direção não produz resposta alguma na direção ortogonal. Além disso, identificou-se que estas células nos primeiros estágios do córtex respondem melhor quando a linha está em movimento e não parada.

Foram identificadas surpreendentes diferenças entre essas células, sensíveis a orientação, do córtex visual. Descobriam-se células sintonizadas para diferentes direções do estímulo, como também para diferentes regiões da retina. As diferenças mais significativas encontradas, porém, diziam respeito ao modo como as diferentes células se comportavam, ou seja, ao modo com que elas deveriam ser estimuladas para gerar uma resposta. De acordo com este comportamento, as células foram classificadas em *células simples* e em *células complexas*.

3.3.2 Células Simples - CS

As células simples receberam esta denominação por apresentarem um comportamento mais simples do que as células complexas. Além disso, estas células simples são encontradas bem próximas aos primeiros estágios do córtex, logo após as células corticais com simetria circular.

Do mesmo modo que as células ganglionares da retina, cada célula simples possui um pequeno campo receptivo bem definido, o qual, dependendo de onde o estímulo luminoso incide, pode gerar uma resposta excitatória ou inibitória. Uma das diferenças entre as células simples e as demais reside na geometria do campo receptivo. As células simples possuem um campo receptivo com simetria em uma determinada direção. Foram identificados três tipos básicos de campo receptivo para as células simples: I) uma linha fina com resposta excitatória centrada numa região inibitória; II) uma linha fina com resposta inibitória centrada numa região excitatória; e III) uma linha de transição entre uma região inibitória e outra excitatória, conforme ilustra a Figura 3-10. Dentre estes tipos, o mais comumente encontrado é o primeiro.

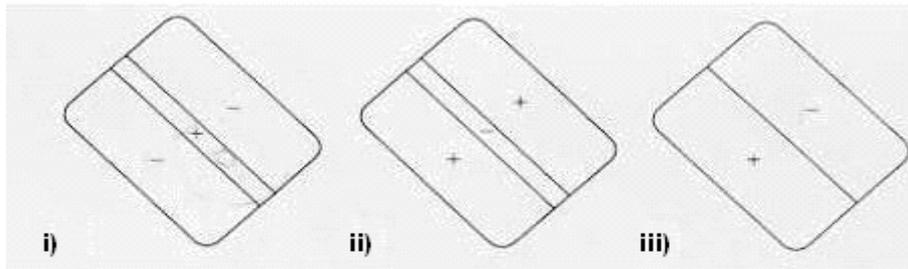


Figura 3-10: Tipos de Campos Receptivos para Células Simples.

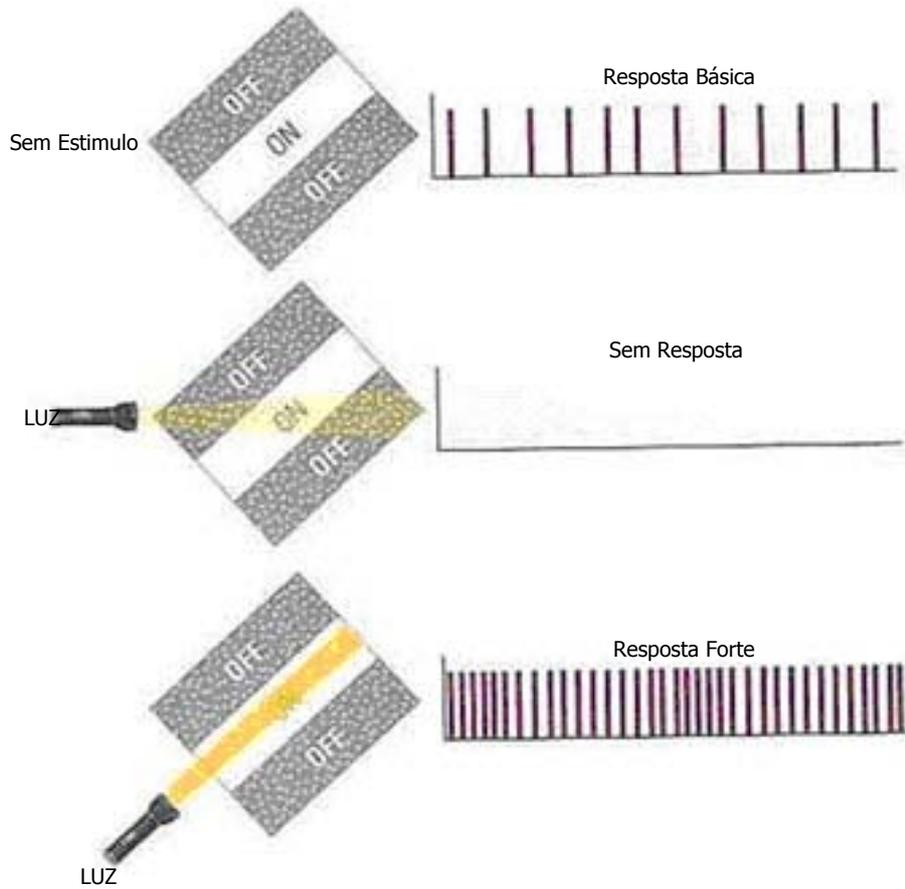


Figura 3-11: Comportamento das Células Simples.
Resposta das células simples com CR tipo I para diferentes direções do estímulo luminoso.

Durante os experimentos para se determinar o campo receptivo das células simples, observou-se que quanto mais ajustado ao campo receptivo da parte excitatória for o estímulo, maior será a excitação produzida pela célula. Qualquer desvio neste casamento do estímulo luminoso com

o campo receptivo pode produzir um efeito de cancelamento mútuo, por exemplo, quando o estímulo luminoso cai tanto na região inibitória quanto na região excitatória (veja Figura 3-11). Observou-se ainda que o tamanho do campo receptivo das células simples também está relacionado com a localização da região da retina. Menores campos receptivos ocorrem para regiões próximas à fóvea.

Não se sabe exatamente como as células simples se interligam com outras células dos estágios anteriores, para produzir este comportamento no campo receptivo. Várias propostas de organização foram apresentadas, e a que tem maior aceitação devido à sua simplicidade, está ilustrada na Figura 3-12. Nesta estrutura, a célula simples recebe sinais das células corticais com campo receptivo circular do primeiro estágio do córtex. As três condições básicas para estas quatro células são: a) possuem campos receptivos vizinhos ou bem próximos, b) terem os campos receptivos alinhados no campo visual, e c) serem todas do mesmo tipo, no caso da figura, do tipo centro-ativo. Uma vez satisfazendo-se estas condições, é possível produzir campos receptivos maiores para as células simples, e o tamanho será proporcional à quantidade de células que vai alimentar a célula simples.

Vale ressaltar que, dentre todos os estímulos experimentados com as células simples, eram os estímulos luminosos fixos, parados, sem movimento que geravam o melhor resultado. Porém, existiam células que não apresentavam uma boa resposta para estímulos estacionários. Esta foi uma das razões práticas de se classificar as células em células simples e células complexas.

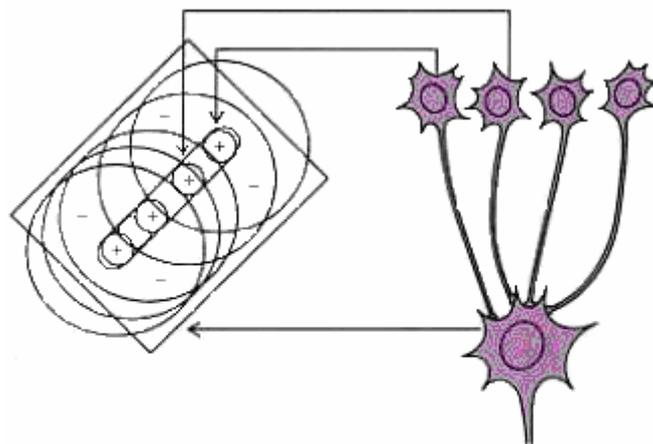


Figura 3-12: Modelo de Interligação das Células Simples.

3.3.3 Células Complexas - CC

As células complexas são as células mais comuns no córtex cerebral, correspondendo a uma proporção de aproximadamente 75% da população total.

O comportamento das células complexas as diferencia das células simples, uma vez que estas são sensíveis a estímulos em movimento, enquanto aquelas eram sensíveis a estímulos estacionários. As características de seletividade quanto à direção do estímulo, observadas nas células simples, também se repetem com as células complexas (veja Figura 3.13). No entanto, para um estímulo excitar uma célula simples é preciso este esteja bem localizado dentro do campo receptivo da célula, enquanto que, numa célula complexa, basta o estímulo mover-se dentro do seu campo receptivo.

As células complexas também não possuem o mapeamento de suas ligações bem determinado. Uma suposição é a de que ela receba sinais de um grupo de células simples com campo receptivo do tipo III conforme a Figura 3.10. As premissas sobre as células simples, neste modelo, são: a) todas possuírem a mesma direção de seletividade, b) estarem organizadas de modo a repetir o lado das regiões excitatórias e inibitórias dos campos receptivos, e c) os seus campos receptivos alinharem-se paralelamente, e lado a lado no campo visual. Outra conclusão a que se chega é que o tamanho do campo receptivo das células complexas é um pouco maior que o das células simples que a compõem, uma vez que o campo receptivo resultante é formado pela superposição de vários campos receptivos de células simples. A Figura 3-14 ilustra este modelo.

Neurofisiologicamente, a visão estereoscópica tem início no córtex visual primário. Nesta região foram identificadas células, tanto simples como complexas, com campos receptivos binoculares [11], isto é, que recebem sinais de ambos os olhos. Algumas destas células mostraram-se sensíveis à disparidade, apresentando respostas distintas de acordo com o posicionamento relativo do estímulo nas duas retinas.

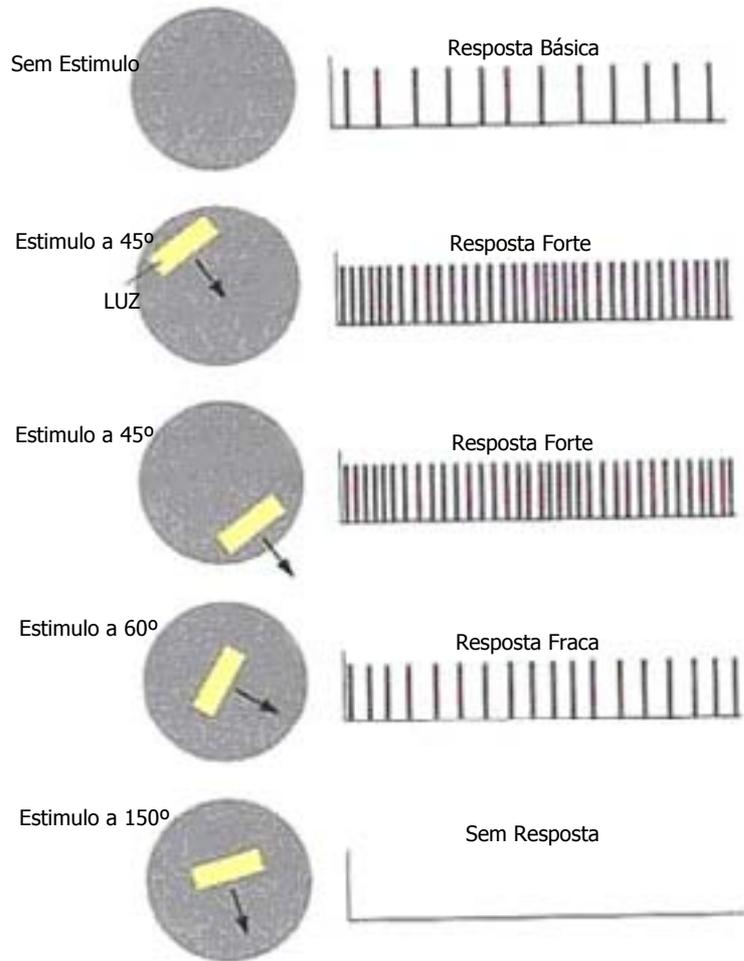


Figura 3-13: Resposta das Células Complexas para Diferente Estímulos.

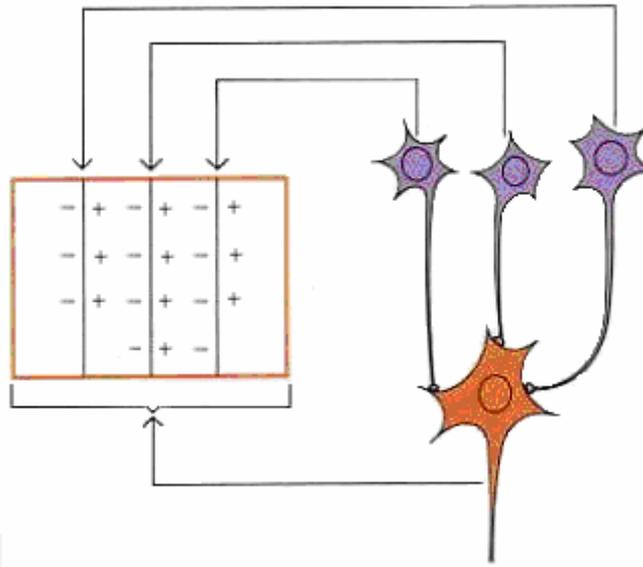


Figura 3-14: Modelo de Interligação das Células Complexas a partir de Células Simples.

3.4 Células Binoculares Sensíveis à Disparidade

3.4.1 Células Simples Binoculares – CSB

As células binoculares sensíveis à disparidade desempenham evidentemente um papel fundamental na visão estereoscópica, e têm sido objeto de intensos estudos. Os modelos atualmente aceitos para a sua descrição [7] representam o campo receptivo das células binoculares *simples* por meio de funções de Gabor, que são senóides moduladas por envoltórias gaussianas (veja Figura 3-15). Uma função de Gabor unidimensional (aqui só estaremos interessados em disparidades horizontais) assume a forma geral

$$g(x) = \exp\left[-\frac{(x + \alpha)^2}{2\sigma^2}\right] \sin(\omega x + \phi) \quad (3.1)$$

onde ω e ϕ denotam a frequência angular e a fase da função senoidal, respectivamente, enquanto $x = \alpha$ denota o centro da envoltória gaussiana, e σ a sua largura.

Intuitivamente, a componente gaussiana da função de Gabor determina a localização e a largura do campo receptivo na retina. Por outro lado, a componente oscilatória determina a organização espacial deste campo, expressa através da frequência espacial e da fase, que vão definir a distância entre as regiões excitatórias e inibitórias (os picos e vales da Figura 3-15 (c)) dentro do campo receptivo.

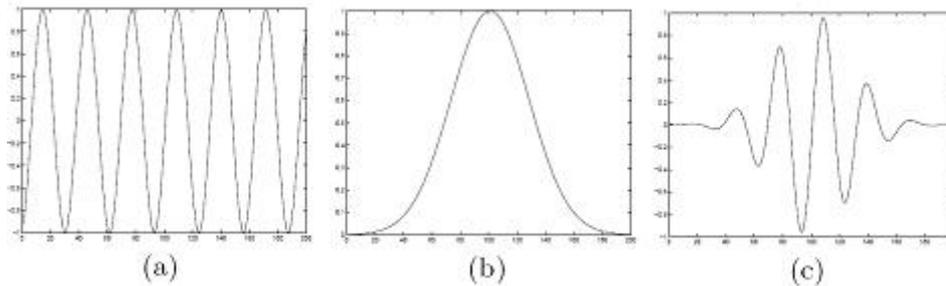


Figura 3-15: Estrutura da Função de Gabor Unidimensional.

Em a) componente oscilatória, b) componente gaussiana, c) produto das duas componentes resulta na função de Gabor.

No modelo mais geral para uma célula binocular simples – chamado de *modelo de deslocamento em posição e em fase* -, os campos receptivos da célula, relativos ao olho esquerdo e ao olho direito, são representados por funções de Gabor de mesma frequência angular e mesma largura, mas podendo diferir tanto em sua fase como em seu centro. Neste caso, também referido como *modelo híbrido*, os campos receptivos esquerdo e direito da célula simples seriam representados por funções da forma

$$g_l(x) = \exp\left[-\frac{(x)^2}{2\sigma^2}\right] \sin(\omega x + \phi_l) \quad (3.2)$$

$$g_r(x) = \exp\left[-\frac{(x-d)^2}{2\sigma^2}\right] \sin(\omega x + \phi_r) \quad (3.3)$$

onde os campos receptivos estão deslocados por d e defasados por $\Delta\phi = |\phi_l - \phi_r|$, com todos os demais parâmetros conservando-se idênticos.

Outros dois modelos para a organização dos campos receptivos das células simples - o *modelo de deslocamento em posição* e o *modelo de deslocamento em fase* - constituem casos particulares do modelo híbrido. No primeiro deles, assume-se que a seletividade à disparidade da célula simples resulta simplesmente da diferença posicional entre os seus campos receptivos, a célula apresentando uma resposta máxima quando o estímulo incorpora uma disparidade correspondente ao deslocamento relativo entre os dois campos [12-17] (um tal tipo de disparidade, que produz resposta máxima, é chamado de *disparidade característica* da célula).

Já no modelo de deslocamento em fase [9,15,16], os campos receptivos direito e esquerdo da célula simples são assumidos coincidentes, mas com organização espacial distinta, representada por sua diferença de fase. Estes dois tipos de modelo aparecem ilustrados na Figura 3-16.

Não existe ainda um consenso quanto à real forma de organização dos campos receptivos das células simples, mas o modelo híbrido parece ser o que fornece o maior grau de flexibilidade para explicar os resultados fornecidos pela neurofisiologia.

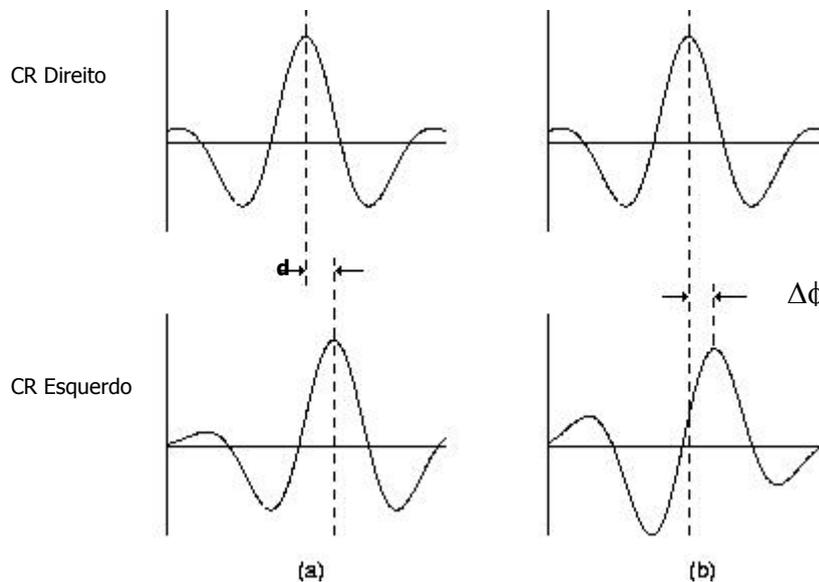


Figura 3-16: Tipos de CR para Seletividade por Disparidade nas CS.
Seletiva por: a) deslocamento posicional, b) deslocamento por fase.

Resposta das Células Simples Binoculares

Qualquer que seja o modelo assumido para o seu campo receptivo binocular, os estudos neurofisiológicos [9,8,15,16] parecem indicar que a resposta das células simples pode ser modelada como a soma dos estímulos apresentados aos dois olhos, cada um deles ponderado pelo respectivo campo receptivo. Matematicamente, representa-se este resultado por

$$r_s = \int_{-\infty}^{\infty} [g_l(x)I_l(x) + g_r(x)I_r(x)] dx \quad (3.4)$$

onde I_l e I_r representam, respectivamente, as intensidades das imagens nas retinas esquerda e direita. Isto significa que as células simples são *neurônios lineares*.

3.4.2 Células Complexas Binoculares - CCB

A célula complexa, ao contrário da simples, não possui um campo receptivo claramente dividido em sub-regiões excitatórias e inibitórias, respondendo igualmente a estímulos em qualquer parte do seu campo receptivo. Em decorrência disto, e do fato de o seu campo receptivo ser maior do que o das células simples, as células complexas não possuem uma dependência significativa com relação à fase dos estímulos, em geral apresentando um valor da disparidade característica mais bem definido do que as células simples [21]. Isto tende a indicar que as células complexas seriam as principais responsáveis pela detecção de disparidades.

Ohzawa e outros pesquisadores [18] descobriram que as células complexas do córtex visual podiam ser modeladas como *neurônios energia*, isto é, que a sua resposta podia ser obtida pela soma do quadrado das respostas de *duas células simples em quadratura*, ou seja, com uma diferença de fase de 90 graus entre os seus campos receptivos. A maneira mais simples de expressar uma tal resposta é utilizando a função de Gabor complexa, para representar o par de células simples em quadratura. Ou seja, usando representações do tipo

$$W_l(x) = \exp \left[i(\omega x + \phi_l) \right] \exp \left[-\frac{x^2}{2\sigma^2} \right] \quad (3.5)$$

$$W_r(x) = \exp \left[i(\omega x + \phi_r) \right] \exp \left[-\frac{(x-d)^2}{2\sigma^2} \right] \quad (3.6)$$

para os campos receptivos destas células, a resposta da célula complexa alimentada pelas mesmas seria dada por

$$r_q = \left| \int_{-\infty}^{\infty} [W_r(x) I_r(x) + W_l(x) I_l(x)] dx \right|^2 \quad (3.7)$$

Com base em fundamentos neurofisiológicos e computacionais, verificou-se ser necessário adicionar um novo elemento à definição da resposta das células complexas, de modo a garantir que o seu campo receptivo fosse maior que o das células simples [12], e formado pela sobreposição de campos receptivos vizinhos. A função de ponderação espacial $s(\sigma_c)$, correspondente a uma gaussiana bidimensional simétrica, com desvio padrão σ_c , foi introduzida de modo a ponderar a contribuição da vizinhança em função da sua distância ao centro da célula complexa. Experimentos computacionais identificaram que este elemento extra também contribui para se obter uma curva de resposta complexa mais robusta para diferentes tipos de estímulos [7]. Assim, em geral nós definimos

$$r_c = r_q * s(\sigma_c) \quad (3.8)$$

onde r_q indica a resposta em (3.7), e r_c indica a resposta da célula complexa.

3.4.3 Detecção de Disparidade pelo Modelo da Energia

Utilizando uma população de células complexas, modeladas, como acima, a partir de pares de células simples em quadratura - cada um deles com diferentes parâmetros de fase -, Qian e

Zhu [22] conseguiram extrair disparidades a partir de estereogramas de pontos aleatórios. A abordagem utilizada consistiu essencialmente em determinar, para cada ponto das imagens, o par de células em quadratura que fornecia o máximo valor para a medida na equação (3.7), que pode ser reescrita, no *modelo do deslocamento em fase*, como

$$r_q = \left| \int_{-\infty}^{\infty} W_r(x) I_r(x) dx + \exp(i\Delta\phi) \int_{-\infty}^{\infty} W_r(x) I_l(x) dx \right|^2 \quad (3.9)$$

Posteriormente, incorporando uma estratégia multi-escala, além de um *pooling* espacial (conforme o $s(\sigma_c)$ introduzido acima) e em orientações, foi possível também extrair disparidades a partir de cenas reais [36].

Este modelo para a estereoscopia permanece sendo o mais bem fundamentado nos estudos neurofisiológicos sobre o sistema visual dos mamíferos. A seguir, nós introduziremos uma nova abordagem para a estimação de disparidades, tentando preservar as principais características fisiologicamente plausíveis do modelo da energia.

4 Estereoscopia baseada na Função de Green

A abordagem aqui proposta para o cálculo das disparidades tenta preservar boa parte das características neurofisiológicas do modelo da energia, mas baseando-se no uso da função de Green de equações de casamento, conforme descrito a seguir.

Um par de imagens retiniais projetadas a partir de uma mesma cena 3D podem ser relacionadas como:

$$I_l(x + U) = I_r(x) \quad (4.1)$$

onde x indica o eixo de referência ao longo da linha epipolar (aqui considerada horizontal), e U indica o campo de disparidades binoculares. Equações como a acima, interpretadas como uma condição de conservação da irradiância, têm sido comumente consideradas no contexto da visão computacional, não apenas para a estimação de disparidades [23], mas também para o cálculo do fluxo óptico [24], a partir de um par de imagens de entrada. Aqui nós faremos um uso diferente da equação (4.1), considerando-a como uma condição para se encontrar o par estereoscópico perfeito de uma determinada imagem de entrada, quando o campo U assume uma forma conhecida.

4.1 Campo de Disparidades Uniforme

Vamos começar assumindo o caso de um campo de disparidades uniforme, $U(x) = u$, onde u é uma constante. Calculando a série de Taylor para o lado esquerdo da equação (4.1), até a segunda ordem, obtemos:

$$\frac{u^2}{2} I_l'' + u I_l' + I_l = I_r \quad (4.2)$$

onde as plicas indicam derivada com relação a x .

Sobre um domínio infinito, $-\infty < x < \infty$, uma solução limitada para equação (4.2), tendendo a zero quando x tende para $\pm \infty$ pode ser obtida como:

$$I_l(x) \equiv I_r(x-u) = \int_{-\infty}^{\infty} G_u(x, x_0) I_r(x_0) dx_0 \quad (4.3)$$

onde $G_u(x, x_0)$, chamada *função de Green* para a equação (4.2), é a solução para tal equação quando o seu lado direito é igual à função impulso unitário, $\delta(x-x_0)$ [46]. Uma vez que a equação (4.2) é linear, e do tipo invariante sob translação, G_u também será um operador invariante sob translação, assumindo a forma:

$$G_u(x, x_0) \equiv G_u(x-x_0) = \frac{2}{u} \sin\left(\frac{x-x_0}{u}\right) \exp^{-\left(\frac{x-x_0}{u}\right)} \quad (4.4)$$

para $x > x_0$, com $G_u(x, x_0) = 0$, caso contrário.

Portanto, dentro da aproximação considerada, o deslocamento de uma dada imagem pela distância fixa u , para a *direita*, é obtido através da convolução desta imagem com o filtro $G_u(x)$. É fácil observar que um deslocamento semelhante para a *esquerda* pode ser conseguido apenas invertendo a direção do eixo de coordenadas x , isto é, com $G_u(-x)$, considerando-se o parâmetro u sempre como positivo. Esta propriedade de deslocamento do filtro G_u já foi explorada anteriormente para diferenciação de sinais [45], e para estimação de forma 3D em visão computacional [40,47].

É importante observar que nós podemos definir uma forma mais geral para a imagem $I_r(x-u)$, ao introduzir a solução homogênea da equação (4.2), obtida quando seu termo do lado direito é igual a zero. Esta pode então ser expressa como:

$$\int_{-\infty}^{\infty} H_u(x, x_0) I_r(x_0) dx_0 \quad (4.5)$$

onde $H_u(x, x_0)$ tem uma forma similar ao $G_u(x, x_0)$ da equação (4.4), exceto que a função seno deve ser substituída pelo cosseno. Deste modo, a forma mais geral para se expressar o par estereoscópico perfeito de $I_r(x)$ é filtrando esta imagem por uma combinação linear dos filtros $G_u(x, x_0)$ e $H_u(x, x_0)$, representados, respectivamente, como as partes real e imaginária do filtro complexo:

$$K_u(x, x_0) \equiv K_u(x - x_0) = \frac{2}{u} \exp \left[\frac{-(1-i)(x - x_0)}{u} \right] \quad (4.6)$$

para $x > x_0$, com $K_u(x, x_0) = 0$, em caso contrário.

Usando isto, nós podemos introduzir a versão complexa da equação (4.3):

$$I_r(x - u) = \int_{-\infty}^{\infty} K_u(x, x_0) I_r(x_0) dx_0 \quad (4.7)$$

cuja parte imaginária corresponde à solução da equação de casamento (4.2), enquanto a parte real é a solução para a versão homogênea da mesma equação. Para se obter um deslocamento equivalente para a esquerda, deve-se usar o filtro $K_u(-x)$.

4.2 Campo de Disparidades Linear

O filtro complexo $K_u(x, x_0)$ tem características semelhantes às de uma função de Gabor, e pode-se verificar que, para valores pequenos de x e x_0 , tais que $|x + x_0| \ll 2a$, ele pode ser obtido como o limite da seguinte forma, ainda mais próxima a uma Gabor:

$$K(x, x_0) = 2k \exp [ik(x - x_0)] \exp \left[-\frac{(x + a)^2 - (x_0 + a)^2}{2\sigma^2} \right] \quad (4.8)$$

quando $u \equiv 1/k \equiv \sigma^2/a$. Neste caso, a parte imaginária de $K(x, x_0)$ é a função de Green da equação diferencial:

$$\frac{1}{2} \left(\frac{\sigma^2}{a} \right)^2 I_l'' + \left(\frac{\sigma^2}{a^2} \right) (x+a) I_l' + \left[1 + \frac{(x+a)^2 + (\sigma^2 - a^2)}{2a^2} \right] I_l = I_r \quad (4.9)$$

que pode ser considerada como uma versão generalizada da equação original de casamento (4.2): para pequenos valores de x em comparação com a (nomeadamente, quando $|x+a| \ll a \left(1 + \sqrt{3 - (\sigma^2/a^2)} \right)$), esta pode ser aproximada como:

$$\frac{1}{2} \left(\frac{\sigma^2}{a} \right)^2 I_l'' + \left(\frac{\sigma^2}{a} + \frac{\sigma^2}{a^2} x \right) I_l' + I_l = I_r \quad (4.10)$$

que, por sua vez, tende para a equação (4.2) quando σ^2/a^2 tende para zero. Até primeira ordem em σ^2/a , a equação 4.10 corresponde a uma relação de conservação da irradiância, tal como a equação (4.1), para um campo de disparidades linear:

$$U(x) = u + u' x \quad (4.11)$$

onde $u \equiv \sigma^2/a$ e $u' \equiv \sigma^2/a^2$ são constantes, representando, respectivamente, a disparidade em $x=0$ e sua correspondente derivada.

Deve-se ressaltar que o filtro complexo de Green, $K(x, x_0)$, pode ser expresso como:

$$K(x, x_0) = N(x_0) W(x) \bar{W}(x_0) \quad (4.12)$$

onde $\bar{W}(x)$ corresponde à conjugada de $W(x)$, que é a função de Gabor complexa:

$$W(x) = \exp(ikx) \exp\left[-\frac{(x+a)^2}{2\sigma^2}\right] \quad (4.13)$$

Esta função, por sua vez, é a solução complexa para a forma homogênea da equação (4.9), satisfazendo à condição de que $W(x) \rightarrow 0$ quando $x \rightarrow \infty$ (veja Apêndice I). Desta forma, $K(x, x_0)$ consiste essencialmente no produto de duas funções de Gabor. O fator N , que depende somente de x_0 , garante que a parte imaginária de (4.12) seja realmente uma função de Green (ver Apêndice I). As representações gráficas das partes imaginária e real de $K(x, x_0)$ são apresentadas na Figura 4-1.

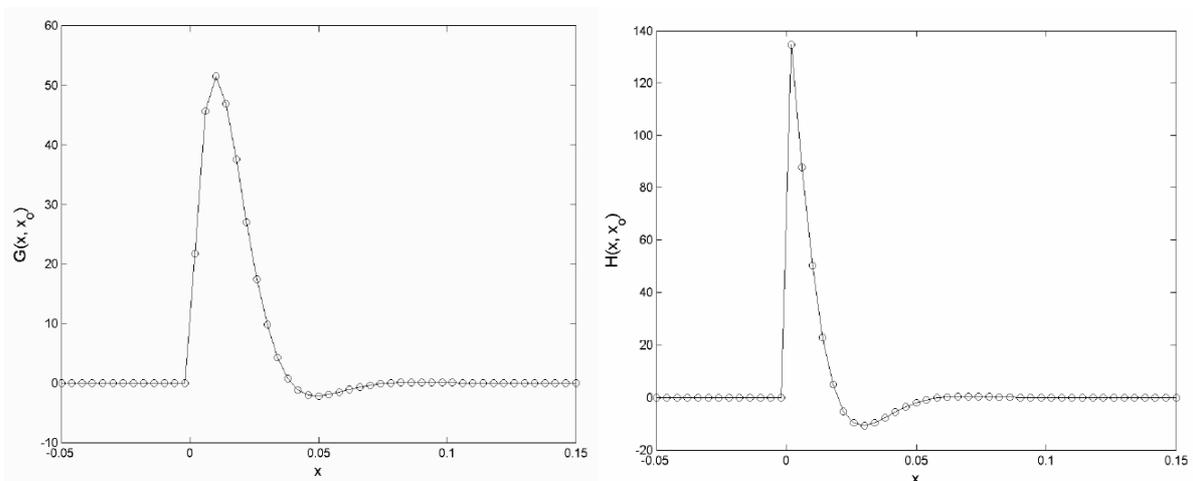


Figura 4-1: Função de Green de Complexa.

Gráfico das partes imaginária (acima) e real (abaixo) do filtro complexo de Green, $K(x, x_0)$, da equação 4.9, quando se assume $x_0 = 0$ e os parâmetros $a = 20$,

$$\sigma = 0.5, \quad k \equiv a/\sigma^2 = 80.$$

4.3 Cálculo de Disparidades a partir do Filtro de Green

Aqui consideramos o filtro de Green $K(x, x_0)$ no contexto da estimação de disparidades estereoscópicas. Conforme provado no Apêndice II, $K(x, x_0)$ apresenta a seguinte propriedade importante: quando aplicado sobre um sinal modulado por uma função de Gabor, o resultado será um sinal modulado de forma similar, entretanto para uma versão espacialmente deslocada

do sinal de entrada. Sendo mais específicos, vamos assumir que o sinal de entrada seja equivalente ao observado por uma célula simples do tipo híbrida, proveniente, por exemplo, da retina direita (veja equação 3.6):

$$I_{r1}(x) = \exp \left[i(\omega x + \phi_r) \right] \exp \left[-\frac{(x + \alpha_r)^2}{2\sigma^2} \right] I_r(x) \quad (4.14)$$

Vamos filtrar então este sinal pelo $K(x, x_0)$ da equação 4.8, para obter:

$$I_{r2}(x) = \int_{-\infty}^{\infty} K(x, x_0) I_{r1}(x_0) dx_0 \quad (4.15)$$

Conforme provado no Apêndice II, o resultado será:

$$I_{r2} = k_r u_r \exp \left[i(\omega x + \psi_r) \right] \exp \left[-\frac{(x + \alpha_r)^2}{2\sigma^2} \right] I_r(x - u_r) \quad (4.16)$$

onde $\psi_r = \phi_r + \Delta\phi_r$, para:

$$\Delta\phi_r = 1 - (\kappa_r - \omega) u_r \quad (4.17)$$

e onde (veja equação (4.7))

$$I_r(x - u_r) \equiv \int_{-\infty}^{\infty} K(x - x_0) I_r(x_0) dx_0 \quad (4.18)$$

é a versão deslocada para a direita do sinal recebido da retina direita considerando termos até a segunda ordem para o parâmetro de deslocamento:

$$u_r = \frac{\sigma^2}{a - \alpha_r} > 0 \quad (4.19)$$

Similarmente, se considerarmos como entrada o sinal retinal esquerdo modulado por uma função de Gabor, poderemos obter a sua versão deslocada à esquerda, usando o filtro de Green apropriado. Por exemplo, vamos assumir que o sinal de entrada agora seja:

$$I_{l1}(x) = \exp \left[i(\omega x + \phi_l) \right] \exp \left[-\frac{(x + \alpha_l)^2}{2\sigma^2} \right] I_l(x) \quad (4.20)$$

A função de Green para um deslocamento à esquerda assumiria a forma:

$$K^-(x, x_0) = 2k_l \exp \left[-ik_l(x - x_0) \right] \exp \left[-\frac{(x - a)^2 - (x_0 - a)^2}{2\sigma^2} \right] \quad (4.21)$$

para $x < x_0$, com $K^-(x, x_0) = 0$, em caso contrário. Aqui, os parâmetros a e σ são considerados idênticos aos parâmetros usados no deslocamento à direita, mas, de modo geral, podemos ter $k_l \neq k_r$, isto é, a frequência espacial do filtro de Green para deslocamento à esquerda pode ser diferente daquela para o deslocamento à direita.

Filtrando I_{l1} por $K^-(x, x_0)$, chegaremos ao resultado:

$$I_{l2} = \int_{-\infty}^{\infty} K^-(x, x_0) I_{l1}(x_0) dx_0 = k_l u_l \exp \left[i(\omega x + \psi_l) \right] \exp \left[-\frac{(x + \alpha_l)^2}{2\sigma^2} \right] I_l(x + u_l) \quad (4.22)$$

onde :

$$u_l = \frac{\sigma^2}{a + \alpha_l} > 0 \quad (4.23)$$

e $\psi_l = \phi_l + \Delta\phi_l$, para

$$\Delta\phi_l = 1 - (\kappa_l + \omega) u_l \quad (4.24)$$

4.3.1 Abordagem para o Cálculo de Disparidades

Nossa abordagem para cálculo de disparidades é motivada pelos resultados da equação (4.16) e (4.22). Dada a imagem projetada na retina direita, $I_r(x)$, fazemos o seu produto com a função de Gabor complexa, formando o sinal $I_{r1}(x)$, conforme a equação (4.14). Em seguida, aplicamos o filtro de Green $K(x, x_0)$, induzindo um deslocamento espacial sobre $I_r(x)$ (como também um deslocamento de fase sobre a componente oscilatória da função de Gabor), para produzir o sinal $I_{r2}(x)$, conforme a equação (4.16). Procedemos de forma similar com o sinal da retina esquerda, $I_l(x)$, modulado por uma função de Gabor: agora fazendo uso do filtro de Green $K^-(x, x_0)$, induzimos um deslocamento espacial para a esquerda, e também um deslocamento de fase, de forma a produzir o sinal $I_{l2}(x)$, conforme (4.22). Uma vez que as imagens $I_r(x)$ e $I_l(x)$ representam, num sentido amplo, visões deslocadas da mesma cena, podemos estimar as disparidades nelas registradas, efetuando uma busca pelo melhor casamento entre os sinais $I_{r1}(x)$ e $I_{l1}(x)$ - correspondentes àquelas imagens de entrada, no nível das células simples - quando submetidos a diferentes deslocamentos (u_l e u_r) por meio dos filtros de Green.

A fim de facilitar a análise subsequente, vamos assumir o caso em que as funções de Gabor são escolhidas com $\alpha_r = \alpha_l = 0$ e $\phi_r = \phi_l = 0$, e os parâmetros k_r e k_l dos filtros de Green satisfazem a relação $k_r u_r = k_l u_l = 1$, de modo eliminar os fatores multiplicativos globais em (4.16) e (4.22). Neste caso específico, é fácil verificar que os deslocamentos espaciais para a esquerda e a direita, u_l e u_r , serão ambos dados por $u = \sigma^2 / a$, e o deslocamento de fase será simplesmente $\psi_r = -\psi_l = \omega u$. Deste modo, se variarmos o parâmetro a e mantivermos todos os demais parâmetros do filtro de Green e da função de Gabor fixos, nós poderemos simular diferentes deslocamentos relativos entre as imagens das retinas. Definindo a função disparidade como $d \equiv 2u = 2\sigma^2 / a$, o valor da disparidade local pode ser estimado, em cada ponto x , como, por exemplo:

$$\bar{d} = 2 \frac{\sigma^2}{\bar{a}} \quad (4.25)$$

onde $\bar{a} \equiv \bar{a}(x)$ denota o valor de a encontrado para o melhor casamento entre $I_{r2}(x)$ e $I_{l2}(x)$.

Várias medidas de casamento poderiam ser consideradas. Aqui, nós adotamos $R(x) = |I_{r2}(x) - I_{l2}(x)|^2$, que, como veremos, permite uma comparação com a resposta das células complexas, de acordo com o modelo da energia. Nós temos:

$$R(x) = |I_{r2}(x) - I_{l2}(x)|^2 = \exp\left(-\frac{x^2}{\sigma^2}\right) \left| I_r(x - u_r) - \exp(i\Delta\psi) I_l(x + u_l) \right|^2 \quad (4.26)$$

ou

$$R(x) = \exp\left(-\frac{x^2}{\sigma^2}\right) \left\{ \left| I_r(x - u_r) \right|^2 + \left| I_l(x + u_l) \right|^2 - 2 \operatorname{Re} \left[\exp(i\Delta\psi) \overline{I_r(x - u_r)} I_l(x + u_l) \right] \right\} \quad (4.27)$$

onde definimos:

$$\Delta\psi = \psi_l - \psi_r = -2\omega u \quad (4.28)$$

e onde $\operatorname{Re}[\]$ denota a parte real do seu argumento.

Deve-se destacar que, para uma comparação adequada com o modelo da energia, conforme [48,22,49], a equação (4.27) deve ser considerada para $x = 0$, já que, ao empregar uma integração simples, em vez de uma convolução, no cômputo das respostas das células simples (conforme 3.4), esses autores efetivamente consideraram a resposta da célula complexa na origem [38].

Comparando com a equação (3.9), e lembrando que

$$I_r(x - u_r) = \int_{-\infty}^{\infty} K_{u_r}(x - x_0) I_r(x_0) dx_0 \quad (4.29)$$

e

$$I_l(x + u_l) = \int_{-\infty}^{\infty} K_{u_l}(x_0 - x) I_l(x_0) dx_0 \quad (4.30)$$

nós podemos ver que, desconsiderando-se o sinal do terceiro termo entre chaves da equação (4.27), o qual pode ser facilmente acomodado como uma defasagem de π no fator $\Delta\psi$ (por exemplo, se nós assumíssemos $\phi_r - \phi_l = \pi$, ao invés de $\phi_r - \phi_l = 0$), a medida $R(x)$ é similar à resposta prevista, para uma célula complexa, pelo modelo da energia, mas incorporando deslocamento em fase e em posição. O deslocamento posicional, neste caso, seria equivalente à medida de disparidade $d = 2u$, enquanto o deslocamento em fase é dado simplesmente por $\Delta\psi = -\omega d$; deve ser notado que uma expressão similar a esta última, exceto quanto ao sinal, foi usada para definir a medida de disparidade no modelo da energia baseado em fase empregado por [22], no qual se assumiu uma população de células simples, cada uma com um diferente valor de fase.

Uma melhor compreensão sobre a relação entre o método de Green e o modelo da energia pode talvez ser obtida ao se reescrever $R(x)$ como:

$$R(x) = \left| \int_{-\infty}^{\infty} [K(x, x_0) W_r(x_0) I_r(x_0) - K^-(x, x_0) W_l(x_0) I_l(x_0)] dx_0 \right|^2 \quad (4.31)$$

onde usamos as equações (4.15) e (4.22), e também utilizamos W_r e W_l como notação para as funções de Gabor, nas definições de $I_{r1}(x)$ e $I_{l1}(x)$.

5 Resultados

Com base no modelo da função de Green para a equação de casamento com campo de disparidades linear, descrito na Seção 4.3.1, nós implementamos um algoritmo de casamento estereoscópico por meio da ferramenta computacional MATLAB. Para dar uma visão geral da seqüência dos passos seguidos neste algoritmo, desde a entrada das imagens até a construção do mapa de disparidades, nós incluímos o seu pseudocódigo na Figura a e 5-2b. Para melhor compreensão do mesmo, nós estabelecemos, abaixo, as convenções seguidas para obtemos disparidades positivas e negativas em nossa abordagem. A Tabela 5-1 mostra o efeito esperado ao se aplicar os filtros de Green, $K(x, x_0)$ e $K^-(x, x_0)$, sobre as imagens de entrada.

$I_{11}(x)$	$I_{r1}(x)$	Efeito Produzido
Deslocamento para Direita: $K(x, x_0)$	Deslocamento para Esquerda: $K^-(x, x_0)$	Favorece casamento de objetos próximos, que em nosso referencial (Figura 2-2) significa <i>disparidades positivas</i> .
Deslocamento para Esquerda: $K^-(x, x_0)$	Deslocamento para Direita: $K(x, x_0)$	Favorece casamento de objetos afastados, que em nosso referencial (Figura 2-2) significa <i>disparidades negativas</i> .

Tabela 5-1: Combinação dos Filtros de Green para Induzir Disparidades Positivas e Negativas

A implementação efetuada segue basicamente os passos descritos na Seção 4.3.1: as imagens de entrada são multiplicadas por funções de Gabor e depois filtradas pelo núcleo de Green para que sejam gerados diferentes deslocamentos espaciais. A resposta $R(x)$ é computada para os diversos deslocamentos, e ponderada espacialmente por uma função gaussiana semelhante à $s(\sigma_c)$ introduzida no modelo da energia. Aquele deslocamento que produzir o máximo valor para o inverso desta resposta espacialmente ponderada vai indicar a estimativa local para a disparidade. Como foi observado que, após deste processo, o mapa de disparidades resultante apresentava correspondências espúrias, decorrentes da impossibilidade de casamento de certas regiões, acrescentamos, como etapa final, a aplicação de um filtro de mediana. Tal filtro, como qualquer outro filtro passa-baixa, tem a característica de suavizar o sinal, porém com a vantagem de manter a definição das bordas [37].

A faixa de disparidades de cada par de imagens de entrada vai desde um valor mínimo, que pode ser negativo, até um valor máximo. Quando um par de imagens de entrada possui disparidade máxima $D_{\max} = +c$ e mínima $D_{\min} = -c$, dizemos que este par contém disparidades em equilíbrio. Quando esta condição é satisfeita, identificamos experimentalmente que o nosso método produz resultados de melhor qualidade, porém é necessário que c seja igual ou inferior a 5 pixels.

Existem inúmeras situações em que as disparidades não estão em equilíbrio. Para que o nosso método suporte tais situações, faz-se necessário introduzir uma etapa adicional, em que se produzem deslocamentos incrementais entre os eixos de referência da imagem esquerda e da direita, até que a disparidade máxima torne-se numericamente igual à disparidade mínima ($-D_{\min} \cong D_{\max}$). Veja diagrama na Figura .

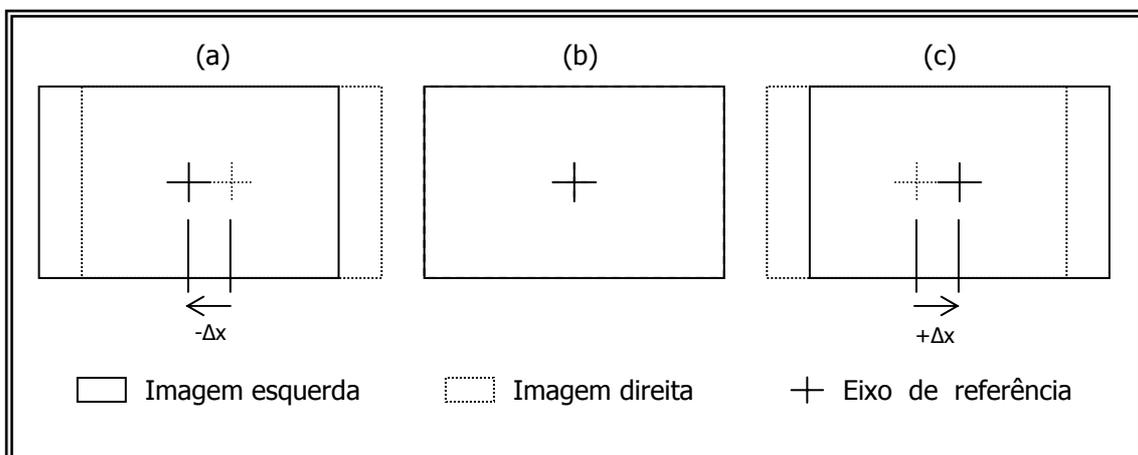


Figura 5-1: Processo de Deslocamento para Equilibrar as Disparidades.

Em (b) as imagens de entrada encontram-se com seus eixos de coordenadas sem qualquer defasagem, isto é, com deslocamento nulo entre os eixos de referência. Esta é a situação original das imagens de entrada. Já em (a) e (c) exemplificam-se as situações em que um deslocamento para a esquerda ou para a direita, são necessários para compensar a discrepância entre as disparidades máxima e mínima, deixando-as próximas ao equilíbrio.

O problema deste processo é que nem sempre sabemos de antemão a faixa de disparidades do par de imagens, e sem contar com esta informação, não é possível saber quando o equilíbrio é atingido. Nestes casos, uma alternativa razoável, como verificamos experimentalmente, foi criar uma métrica que totaliza a diferença quadrática entre as intensidades dos pontos das imagens de entrada, assumindo-se então que o equilíbrio é atingido quando tal diferença é mínima.

Parâmetros	Descrição
ω	Frequência espacial usada tanto na função de Gabor como no filtro de Green. Unidade de medida: <i>rad/pixel</i> .
ϕ_l e ϕ_r	Fases das células simples modeladas pela função de Gabor para os campos receptivos nas retinas esquerda e direita, respectivamente. Unidade de medida: <i>radiano</i>
σ	Largura do campo receptivo das células simples, modeladas pela função de Gabor. Unidade de medida: <i>pixel</i>
α_l e α_r	Deslocamento posicional das células simples, modeladas pela função de Gabor, em relação à origem das coordenadas abscissas. Unidade de medida: <i>pixel</i>
σ_c	Desvio padrão da função da gaussiana 2D de ponderação espacial. Unidade de medida: <i>pixel</i>
<i>raio</i>	Largura horizontal e vertical do filtro Mediana usado para eliminar valores espúrios do mapa de disparidade. Unidade de medida: <i>pixel</i>
D_{\min} , D_{\max} e D_{passo}	Faixa de disparidades a ser varrida pelo método, e o espaçamento entre duas disparidades consecutivas. Unidade de medida: <i>pixel</i>

Figura 5-2a: Parâmetros para Pseudocódigo do Método de Green.

Uma consequência direta do processo de equilibrar as disparidades é a necessidade de criarmos regiões de borda no início e no final de cada linha das imagens, como margem de manobra para o seu deslocamento. Para fim dos experimentos aqui reportados, estas regiões de borda se restringiram a 5 pixels em cada lado.

PSEUDOCÓDIGO DO MÉTODO DE GREEN

```
% Entrada de Dados
 $I_l[x,y]$  = Lê Imagem (Esquerda);
 $I_r[x,y]$  = Lê Imagem (Direita);
Entrada Parâmetros Modelo ( $\omega, \phi_r, \phi_l, \sigma, \alpha_r, \alpha_l, \sigma_c$ );
Entrada Parâmetros Controle ( $D_{\min}, D_{\max}, D_{\text{passo}}, \text{raio}$ );
% Preparação das Imagens
Equilibra faixa disparidade ( $I_l, I_r, D_{\min}, D_{\max}$ );
Elimina componente DC imagens ( $I_l, I_r$ );
 $I_{l1}$  = Aplica Filtro de Gabor ( $I_l$ );
 $I_{r1}$  = Aplica Filtro de Gabor ( $I_r$ );
% Simula Diferentes Deslocamentos nas Imagens
Loop ( $d = D_{\min}$  até  $D_{\max}$  em  $D_{\text{passo}}$ )
     $u_l$  = Define Deslocamento Imagem Esquerda ( $d, \sigma, \alpha_l$ );
     $u_r$  = Define Deslocamento Imagem Direita ( $d, \sigma, \alpha_r$ );
     $I_{l2}$  = Aplica Filtro de Green Imagem Esquerda ( $I_{l1}, u_l$ );
     $I_{r2}$  = Aplica Filtro de Green Imagem Direita ( $I_{r1}, u_r$ );
    calcula vetor  $R_1[d] = |I_{r2} - I_{l2}|^2$  em cada ponto;
Fim Loop ( $d$ );
% Expande Campo Receptivo conforme Célula Complexa
calcula vetor  $R_2[d] = R_1[d] * s(\sigma_c)$  em todos pontos;
% Constrói Mapa de Disparidade
constrói o vetor Peso [ $d$ ] em cada ponto =  $\frac{1}{R_2[d]}$ ;
 $\bar{D}$  em cada ponto = disparidade que produziu o maior Peso [ $d$ ] naquele ponto;
% Elimina Espúrios do Mapa de Disparidade
Mapa Disparidade = Filtro Mediana ( $\bar{D}(x,y), \text{raio}$ );
```

Figura 5-2b: Pseudocódigo do Método de Green.

O método de casamento estereoscópico via funções de Green preserva características da técnica de correspondência por pixel (veja Seção 2.2.3). O principal diferencial com relação àquela, está no fato de só utilizarmos a informação de intensidade em cada ponto, para efeitos de comparação, após multiplicar as imagens por Gabor e filtrá-las por Green. A propriedade de deslocamento espacial do filtro de Green faz com que as intensidades nas imagens filtradas, $I_{l2}(x)$ e $I_{r2}(x)$, sofram uma influência de outros pontos colineares na sua vizinhança (veja Figura 5-3). Assim, ao empregarmos estas intensidades, influenciadas pela sua vizinhança, como medida pontual para o processo de correspondência, automaticamente ganhamos em robustez por reduzirmos os falsos positivos, já que o processo passa a se assemelhar a uma correspondência por área. A contrapartida desta técnica é introduzir um maior custo computacional.

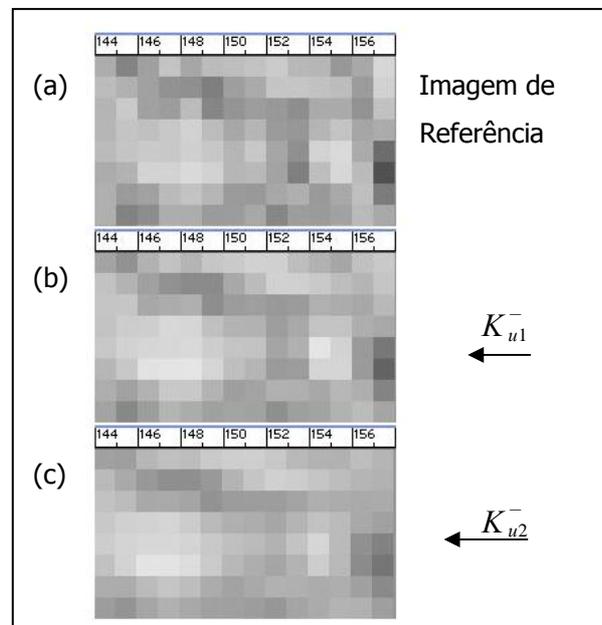


Figura 5-3: Efeito de Deslocamento do Filtro de Green.

Em (a) imagem de referência, sem qualquer deslocamento; (b) e (c) resultado dos efeitos de deslocamento, para esquerda, produzidos pelo filtro Green quando aplicado sobre a imagem de referência, sendo $u_1 < u_2$.

A seguir mostraremos como a nossa abordagem se mostrou consistente e estável para diferentes parâmetros de entrada, bem como, nos casos em que mantivemos estes parâmetros constantes ao longo de diferentes experimentos.

5.1 Experimentos

Aqui apresentaremos alguns resultados obtidos com a nossa abordagem. Algumas simplificações, já apresentadas na Seção 4.3.1, foram adotadas, de modo a facilitar nossa análise e também reduzir o tempo de investigação dos efeitos de cada parâmetro sobre os resultados produzidos. Em todos os experimentos os valores adotados na Tabela 5-2 foram mantidos constantes. Os valores assumidos pelos demais parâmetros serão informados em cada experimento.

ω	2π rads/pixel
ϕ_l e ϕ_r	0 rad e 0 rad
σ	<Largura da Imagem em Pixels> \div 2
α_l e α_r	0 pixel e 0 pixel
D_{passo}	0.25 pixel

Tabela 5-2: Parâmetros de Entrada Preservados Constante nos Experimentos.

5.1.1 Imagens Sintéticas

As

Figura 5-4 e 5-5 apresentam os resultados obtidos com estereogramas sintéticos. Para este tipo de imagens, o mapa real de disparidades é geralmente conhecido *a priori*. No caso Figura 5-4, o estereograma apresentado consiste de pontos aleatórios com densidade igual a 50%, e resolução 128x128 pixels. A faixa de disparidades presente neste estereograma abrange apenas dois valores -2 a -5. Na implementação do nosso algoritmo, após equilibrarmos as disparidades na faixa de -2 a +1, a busca pelo casamento foi realizada desde -5 até +5, em passos de 0.25 pixels. Observa-se que, pela equação (4.25), a variável a , e consequentemente a variável u , podem assumir valores fracionários, o que permite que obtenhamos precisão de deslocamento ao nível de sub-pixel. O método utilizou o equivalente a 41 pares de células simples em quadratura, para cobrir esta faixa de disparidade com precisão de um quarto de pixel.

Na Figura 5-5, nós apresentamos os resultados para um estereograma sintético de disparidades contínuas (Zwei). As imagens de entrada possuem disparidades de -4 a +4 pixels, e resolução de 256x256 pixels. Neste caso, a faixa de disparidades já estava equilibrada. A busca pelo casamento foi feita entre -5 a +5 pixels, com o correspondente a 41 pares de células simples em quadratura.

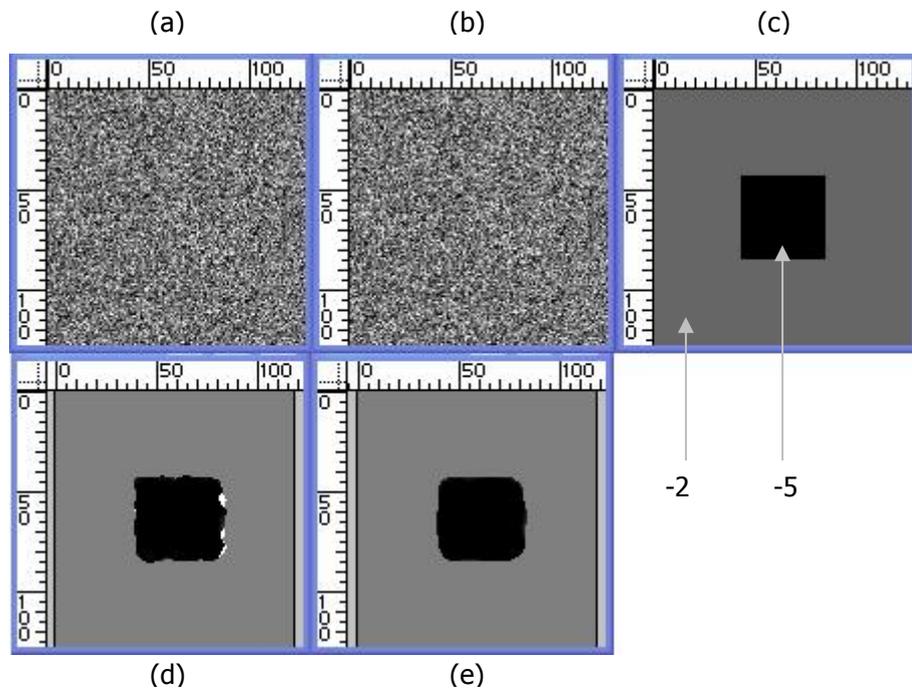


Figura 5-4: Resultados Estereograma Random Dots.

Em (a) e (b) mostramos o par de imagens de entrada com 128x128 pixels e, em (c), seu respectivo mapa de disparidades reais [30]. Em (d), aparece o resultado do casamento para faixa de busca de +5 à -5 com passo de 0.25 pixels e $\sigma_c = 2$ pixels, e em (e), o mesmo experimento com aplicação de um filtro de mediana 5x5.

Nós identificamos que a nossa abordagem, exceto nas regiões de borda dos objetos, onde ocorre oclusão, forneceu resultados bastante fiéis aos mapas verdadeiros. No caso do segundo experimento, verificou-se uma nítida limitação no que concerne à sensibilidade para pequenas variações na disparidade. Suspeitamos que isto seja decorrência do comportamento de alguns parâmetros do nosso método, que ainda não controlamos suficientemente, e/ou de questões intrínsecas à natureza das próprias imagens.

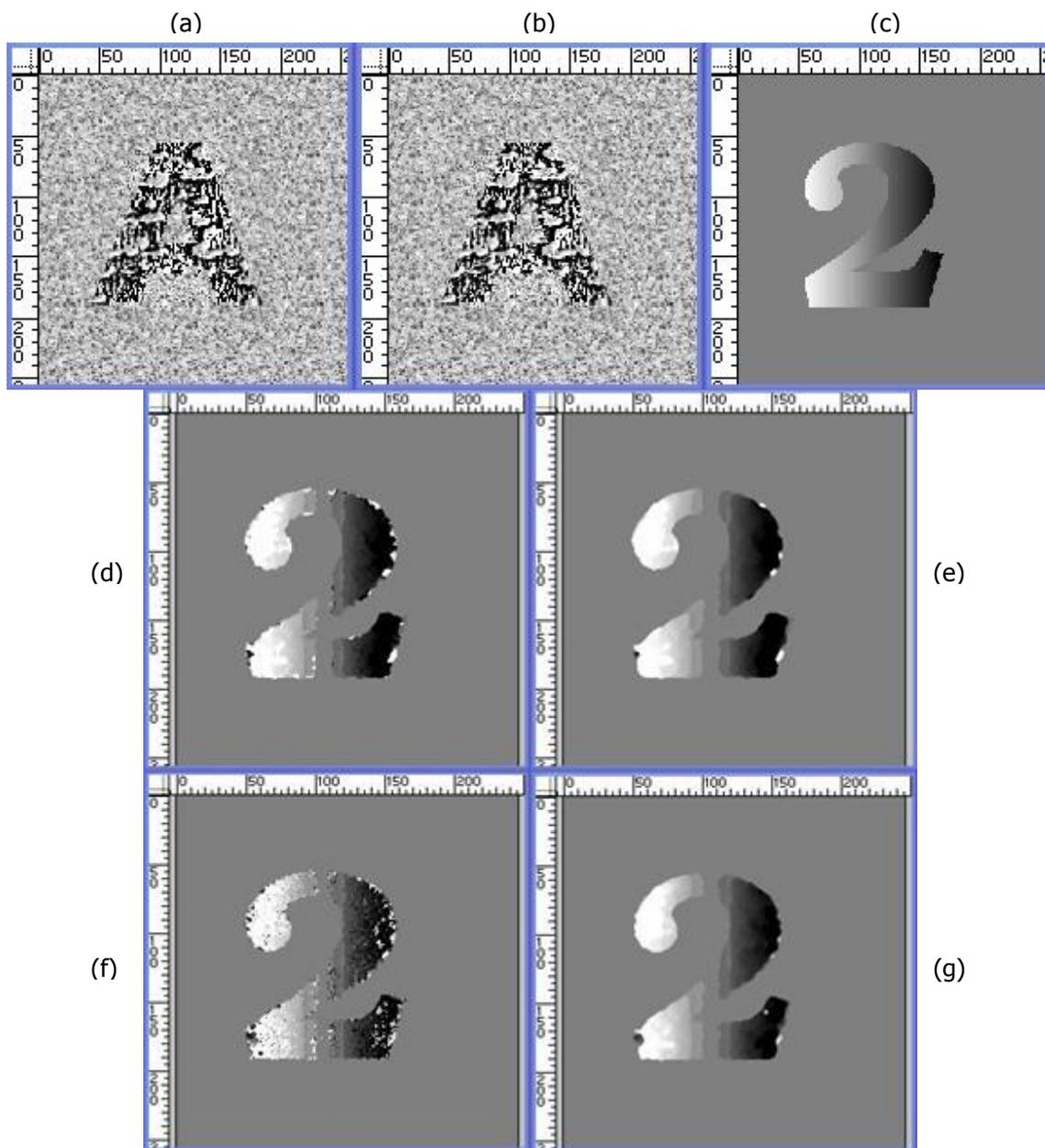


Figura 5-5: Resultados do Estereograma Zwei.

Em (a) e (b) mostramos as imagens de entrada e, em (c), o mapa real de disparidades, todos com 256x256 pixels [32]. Em (d) e (e) – primeiro experimento – aparece o resultado do casamento para faixa de busca entre +5 e -5, com passo de 0.25 pixels e $\sigma_c = 1$ pixel (d), e o mesmo experimento aplicando-se um filtro de mediana 2x2 (e).

Em (f) e (g) – segundo experimento -, temos o resultado obtido com os mesmos parâmetros do experimento anterior, exceto que com $\sigma_c = 0.5$ pixel (f), e o mesmo experimento aplicando-se um filtro de mediana 1x1 (g).

Uma outra característica do método que pode ser observada nos experimentos com o Zwei é a função do parâmetro σ_c . Quanto maior o seu valor, maior o efeito de suavização produzido pela função de ponderação espacial (compare Figura 5-5 (f) em relação à (d)). No primeiro experimento o desvio padrão corresponde a 1 pixel, enquanto no segundo o seu valor é a metade, o que concorda com a visivelmente maior suavização do primeiro resultado, sem o uso do filtro de mediana. O efeito de suavização do σ_c também permite eliminar resultados espúrios, porém ao custo de se perder a definição das bordas dos objetos. Nestes casos, observamos que o melhor é não lançar mão de valores para σ_c superiores a 2 pixels, deixando o ruído remanescente para ser eliminado pelo do filtro mediana.

5.1.2 Imagens Reais

Nos quatro experimentos tratados a seguir, da Figura 5-6 até a Figura 5-9, lidamos com pares de imagens do mundo real. Para estes casos, o mapa de disparidades real não é conhecido *a priori*, embora seja possível analisar o deslocamento para pontos-chaves no par de imagens a fim de identificar a faixa de disparidade presente neste.

No caso da Figura 5-6 o estereograma apresentado consiste em duas imagens aéreas retificadas segundo um modelo de câmeras paralelas. Como as imagens não foram adquiridas simultaneamente, existem pequenas variações entre elas, devido aos deslocamentos de objetos na cena. Utilizamos uma versão deste estereograma com resolução de 256x256 pixels. A faixa de disparidades deste vai de +4 à -4, ou seja, está equilibrada. Nossa busca pelo casamento foi realizada na faixa de -5 até +5 em passos de 0.25 pixels. O método utilizou o equivalente a 41 pares de células simples em quadratura para cobrir esta faixa de disparidades com precisão de um quarto de pixel. O resultado, sem aplicação do filtro de mediana, apresenta várias estimativas espúrias, em decorrência de oclusões, superfícies inclinadas e objetos em movimento. Motivo pelo qual nós tivemos de utilizar um raio relativamente grande para o filtro de mediana, de modo a eliminar estas falhas.

O estereograma da Figura 5-7 caracteriza uma situação onde existe uma grande variação de disparidades na região do solo. As imagens utilizadas apresentam uma resolução de 256x256 pixels e uma faixa de disparidades aproximadamente equilibrada, de +3 à -3. Nossa busca por

casamento novamente se concentrou na faixa de -5 até +5 pixels, em passos de 0.25, o equivalente a 41 pares de células simples em quadratura. O resultado, sem o uso do filtro de mediana, apresenta algumas estimativas espúrias, principalmente em decorrência de oclusões e superfícies inclinadas. Por outro lado, a suavidade na transição entre as disparidades é bem visível, tanto antes quanto depois da filtragem por mediana.

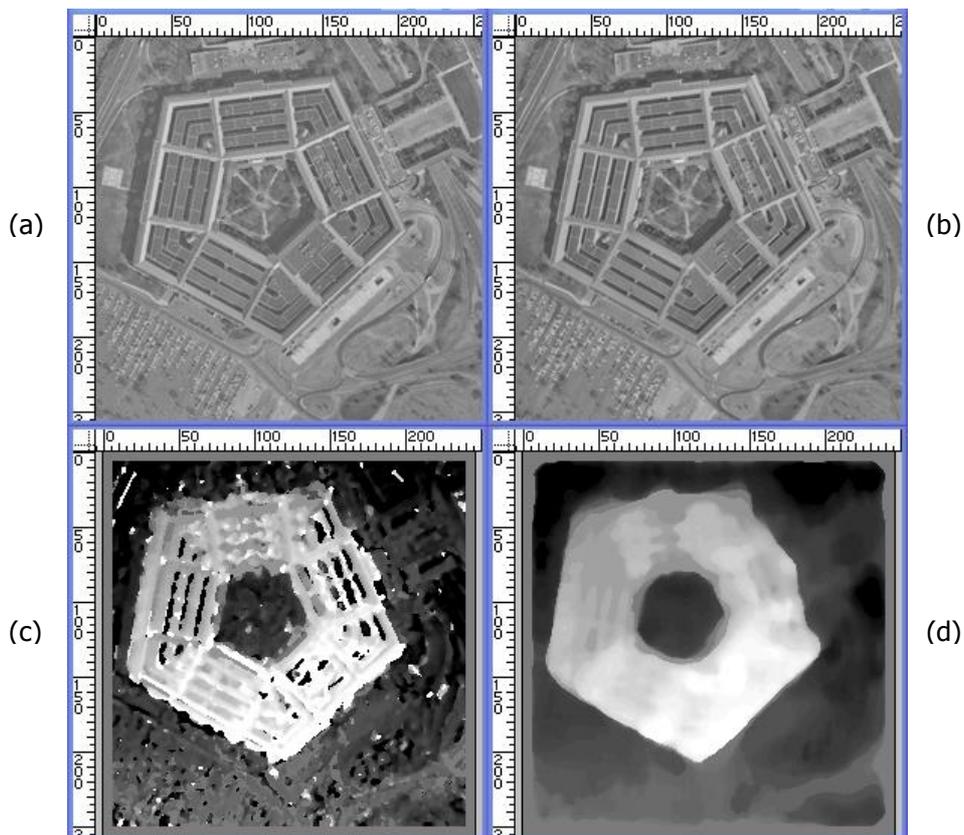


Figura 5-6: Resultados do Estereograma Pentagon.

Em (a) e (b), temos o par de imagens de entrada [31]. Em (c), o resultado para faixa de busca de +5 a -5 com passo de 0.25 pixels e $\sigma_c = 1$ pixel, e em (d), o mesmo experimento, após aplicação do filtro de mediana 8x8.

O estereograma da Figura 5-8 é geralmente considerado difícil, por apresentar regiões com pouca textura, tanto no chão quanto no vão da janela, além de possuir uma grande superfície inclinada ao fundo. As imagens utilizadas apresentam uma resolução de 256x256 pixels e uma faixa de disparidades de aproximadamente +9 a -1. Após equilibrar as disparidades na faixa +4 a -5, num primeiro experimento, a busca pelo casamento se concentrou nesta mesma faixa.

No intuito de analisar a robustez do algoritmo, nós a seguir ampliamos a faixa de procura para de +5 a -6. Ambos os resultados mostraram-se equivalentes, exceto nos pontos problemáticos: região escura e sem textura na janela, e em alguns pontos da parede ao fundo, onde ocorre um efeito de expansão e compressão nas projeções em cada imagem, em decorrência da sua inclinação. Estas disparidades espúrias foram bem eliminadas, no primeiro caso, por um filtro de mediana com raio 4. Sob o ponto de vista neurofisiológico, no primeiro experimento precisou-se do equivalente a 37 pares de células simples em quadratura, enquanto que, no segundo, utilizamos o equivalente a 45 destes pares, para poder atender à maior faixa de disparidades.

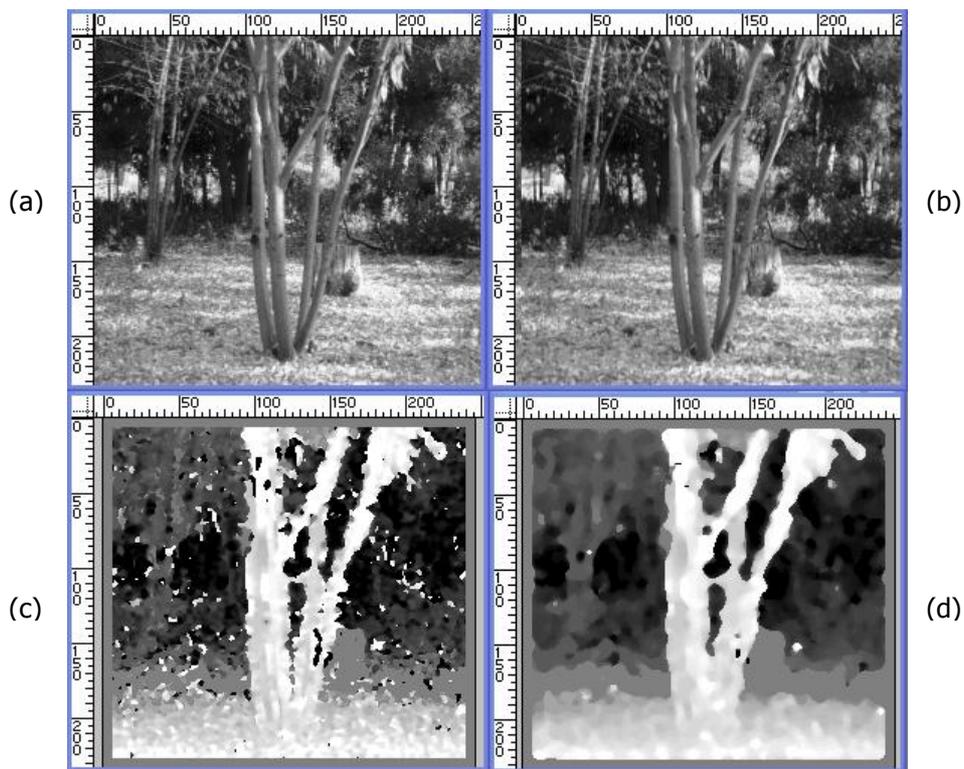


Figura 5-7: Resultados do Estereograma Tree ou Baum.

Em (a) e (b) temos o par de imagens de entrada [31 ou 32], em (c), o resultado para faixa de busca +5 a -5 com passo de 0.25 pixels e $\sigma_c = 1$ pixel, e em (d), o mesmo experimento acrescido do filtro mediana 3x3.

O estereograma da Figura 5-9 também é geralmente considerado difícil, por apresentar grandes regiões com pouca textura, além de possuir uma grande quantidade de superfícies inclinadas. As imagens utilizadas apresentam uma resolução de 256x256 pixels, e uma faixa de

disparidades aproximada de +8 a +1. No primeiro experimento, efetuamos o equilíbrio das disparidades para a faixa de +4 a -3, e uma busca de correspondência de +5 a -3.5 pixels. O resultado é aceitável para as regiões que não são problemáticas (sem textura e sem inclinação). Já, no restante da imagem, ocorrem erros nítidos, produzidos principalmente nas partes com sombra e sem textura das paredes.

A fim de avaliar o efeito do equilíbrio das disparidades nas imagens de entrada, no segundo experimento as imagens não foram equilibradas, mas apenas produzimos um pré-deslocamento negativo de 2 unidades na faixa de disparidades, isto é, as disparidades passaram a ser de +6 a -1. A faixa de busca do método se concentrou entre +7 e -1 pixels, ou no equivalente a de +9 a +1 pixel, se medida para as imagens não equilibradas (conforme apresentado nas figuras dos estereogramas).

Ambos os resultados são, de um modo geral, bem próximos entre si, no entanto, devido à falta de equilíbrio nas disparidades, fica visível a falta de sensibilidade para as disparidades positivas, por algumas destas estarem além da faixa confiável do método. Vale observar que as intensidades representadas nos resultados podem sofrer algum pequeno desvio no brilho em decorrência das faixas de busca não serem exatamente do mesmo tamanho. Sob o ponto de vista neurofisiológico, no primeiro experimento precisou-se do equivalente a 35 pares de células simples em quadratura, enquanto que, no segundo, utilizamos o equivalente a 33 pares.

De acordo com os resultados gerais aqui considerados, ficou evidente que o nosso método apresentou um comportamento bem robusto para boa parte dos experimentos. Nos experimentos das Figura 5-5, Figura 5-6 e Figura 5-7, em especial, todos os parâmetros de entrada mantiveram-se idênticos, ou seja, não seria nem necessário conhecer de antemão a faixa de disparidades presente em cada estereograma.

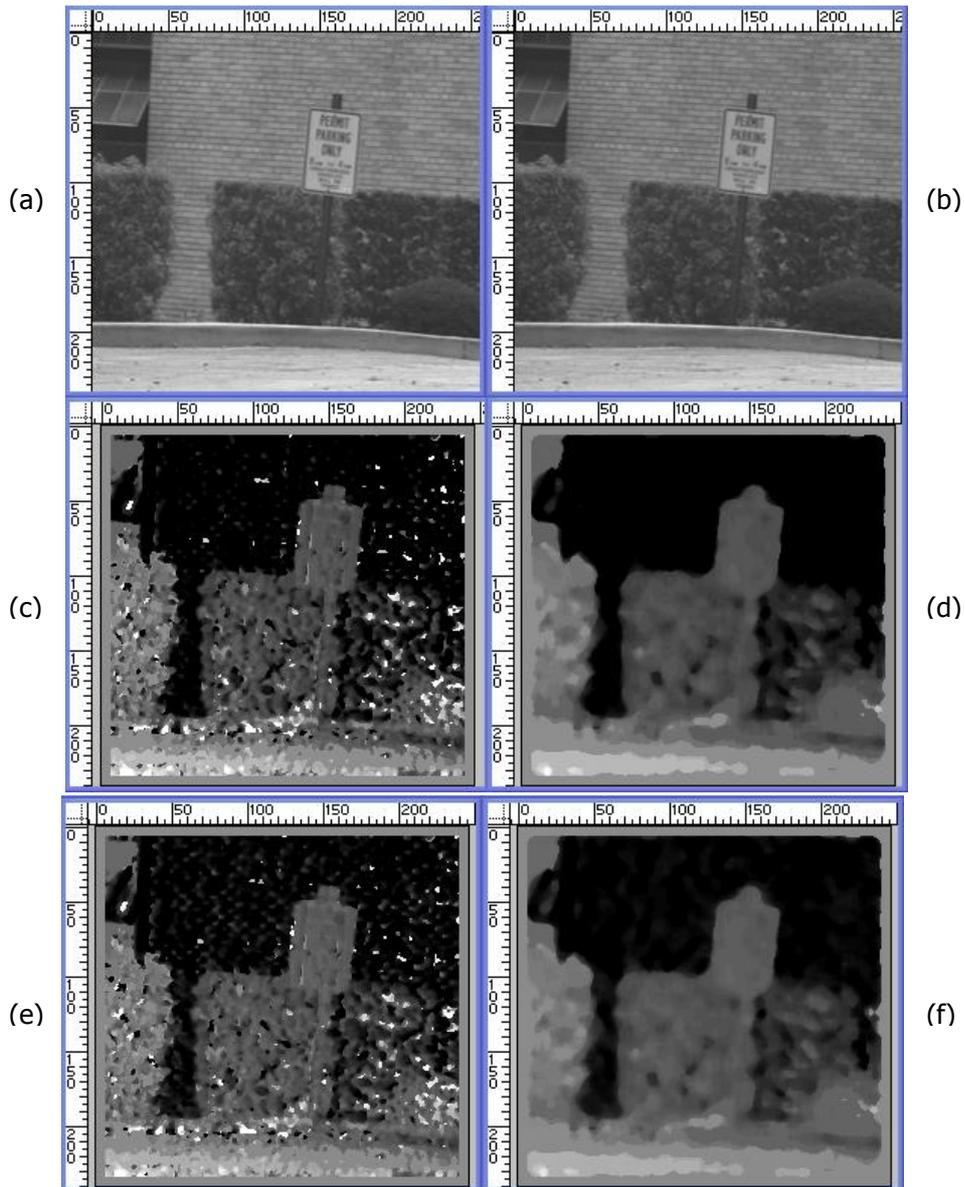


Figura 5-8: Resultados do Estereograma Shrub.

(a) e (b) correspondem às imagens 09 e 21, respectivamente, do banco de imagens em [31]. Em (c) e (d) têm-se os resultados do primeiro experimento, onde (c) foi obtido para faixa de busca entre +9 e +1* com passo de 0.25 pixels e $\sigma_c = 1$ pixel, e em (d) acrescentou-se uma filtragem por mediana 4x4. No segundo experimento, (e) é o resultado para faixa de busca entre +10 e 0*, mantendo-se os demais parâmetros como no experimento anterior, e em (f) acrescentou-se uma filtragem por mediana 4x4.

(*) Faixa de disparidades relativa às imagens de entrada, sem qualquer alteração pelo passo de equilíbrio.

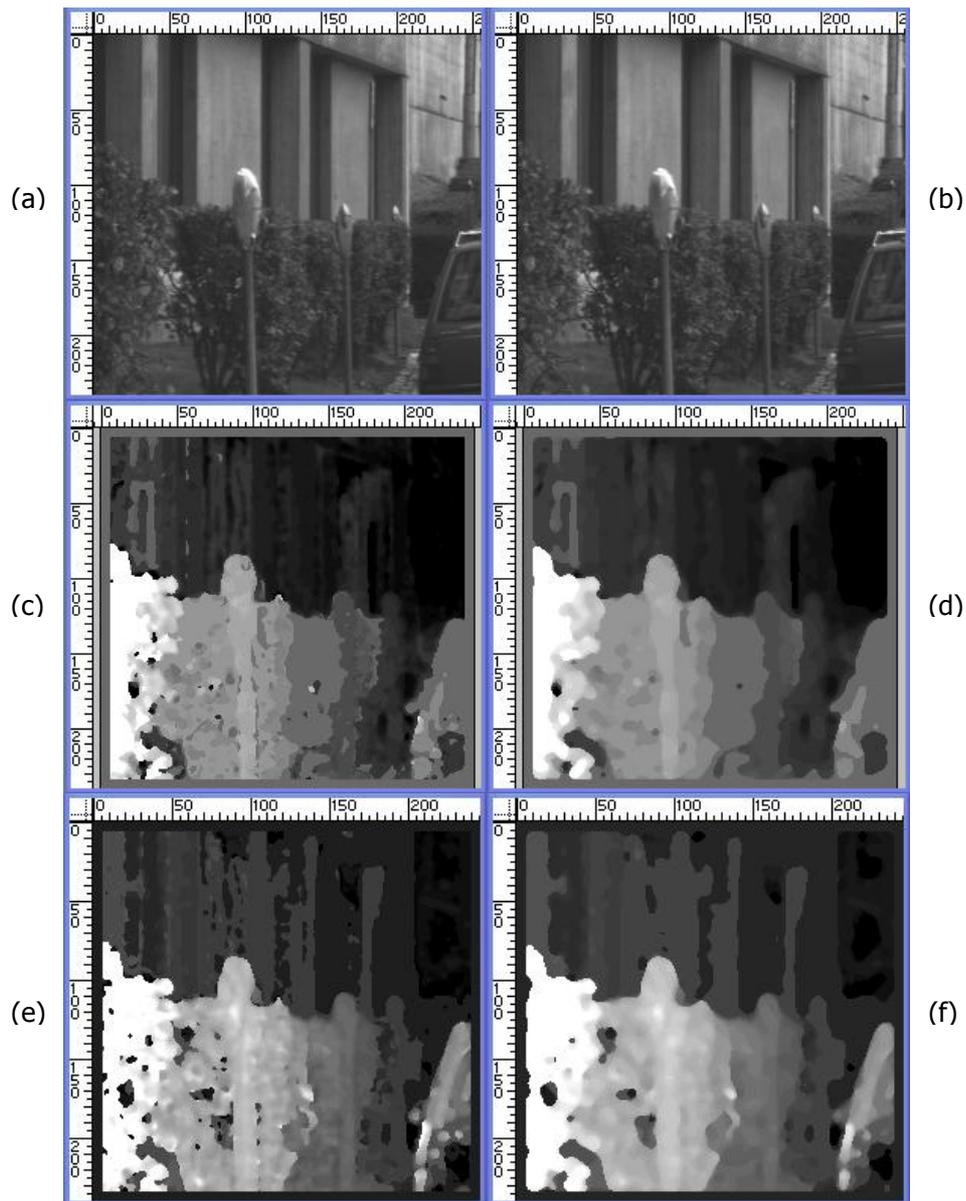


Figura 5-9: Resultados do Estereograma Parkmeter.

Em (a) e (b) temos o par de imagens de entrada [31]. Em (c) o resultado para o primeiro experimento com faixa de busca por disparidade entre $+9$ à $+0.5^*$ com passo de 0.25 pixels e $\sigma_c = 2$ pixel, e em (d) o mesmo experimento acrescido do filtro mediana 3x3. Em (e) o resultado para o segundo experimento com faixa de busca por disparidade entre $+9$ à $+1^*$ e mantendo os demais parâmetros iguais ao do experimento anterior, e em (f) o mesmo experimento acrescido do filtro mediana 3x3.

(*) Faixa de disparidades relativa às imagens de entrada, sem qualquer alteração pelo passo de equilíbrio.

6 Conclusão e Trabalhos Futuros

Apresentamos aqui uma nova abordagem para o processo de estimação de disparidades através do uso da função Green para uma equação de casamento. Quando um sinal modulado por uma função de Gabor é filtrado pelo núcleo de Green, nós conseguimos produzir um sinal similarmente modulado, mas deslocado em relação ao sinal original. Uma vez que as funções de Gabor são utilizadas para modelar o campo receptivo das células simples do córtex visual, nós podemos, por meio do filtro de Green, simular o equivalente a diferentes deslocamentos do sinal retinal apresentado a ambos os olhos, de modo a gerar diferentes entradas para as células simples binoculares. O campo de disparidades pode então ser estimado com base naquele deslocamento simulado que produz a melhor correspondência local entre as duas imagens. A medida de correspondência aqui utilizada, equação (4.27), mostrou-se similar ao que é previsto, no modelo da energia, para a resposta das células corticais complexas, e, com base nesta medida, nós conseguimos estimar mapas de disparidades confiáveis a partir de estereogramas tanto de imagens sintéticas como de imagens reais.

A principal contribuição do nosso trabalho foi, portanto, a de formular, de uma maneira neurofisiologicamente plausível, uma abordagem estereoscópica baseada no casamento de imagens.

Os filtros de Green aqui considerados, que apresentam a propriedade de deslocar um sinal modulado por Gabor, são essencialmente também constituídos por funções de Gabor (veja equações (4.12) e (4.13) e Apêndice I), que, por sua vez, são as soluções homogêneas para a equação de casamento considerada (4.9). Cabe ressaltar, daí, que as funções de Gabor aparecem, em nossa abordagem, de uma maneira menos *ad hoc* do que geralmente acontece no contexto de estimação de disparidades. As funções de Gabor são em geral atrativas, como filtros passa-faixa, pela sua localização ótima, tanto no domínio espacial quanto no da frequência, quando medida em termos do princípio da incerteza [50]. Porém, outras alternativas para filtros deste tipo também já foram consideradas no contexto da estereoscopia [26,38], e, ademais, o uso dos filtros de Gabor para modelar o campo receptivo das células simples tem sido questionado por alguns autores [39].

Por outro lado, embora não seja um pré-requisito dentro da nossa abordagem (veja, por exemplo, [2]), o emprego das funções de Gabor, aqui, surge de uma forma mais natural, já que elas são as soluções homogêneas da equação de casamento considerada. De um modo geral, a abordagem da função de Green independe do uso das funções de Gabor para modelar o campo receptivo das células corticais. É possível explorar esta mesma técnica com versões alternativas para as equações de casamento, e então analisar se os filtros passa-faixa encontrados seriam apropriados para a modelagem das células simples.

Do ponto de vista computacional, infelizmente, a abordagem aqui apresentada não pôde ser exaustivamente explorada, pois o tempo gasto no amadurecimento do algoritmo, até começarmos a obter resultados de qualidade, foi excessivamente longo. Podemos, sem sombra de dúvida, dizer que mais de 70% do tempo dedicado a este trabalho se resumiu à modelagem computacional da abordagem. Cada experimento aqui apresentado consumiu, em média, de duas a três horas em um computador Pentium-4 de 2.0 Ghz e 1 Gb RAM, e qualquer alteração nos dados de entrada nos forçava a repetir os cálculos praticamente do início. O pseudocódigo da Figura 5-2(b) descreve a primeira abordagem que chegou a produzir resultados com uma confiabilidade aceitável, o que deixa claro que nada impede que formas alternativas mais eficientes venham a ser propostas, no futuro, com base nos mesmos filtros de Green.

Futuramente, nós planejamos também adaptar a nossa atual implementação para incorporar uma estratégia *coarse-to-fine* com base no parâmetro σ_c , de modo a permitir que o método trabalhe de forma mais precisa, e também otimizada em termos de recursos computacionais. Outra possibilidade que também planejamos explorar são os efeitos produzidos pelo filtro de Green no domínio da frequência. Entender o comportamento de cada parâmetro de entrada, e a sua influência sobre o processo de deslocamento, permitir-nos-á reduzir o número de experimentos necessários para a produção de bons resultados, assim como dará subsídio para gerarmos um modelo computacional mais robusto da nossa abordagem.

Apêndice I

Aqui, vamos deduzir a função de Green da equação diferencial (4.9). Identificando $k \equiv a / \sigma^2$, vamos escrever essa equação como

$$\frac{1}{2} I_l'' + \left(\frac{x+a}{\sigma^2} \right) I_l' + \frac{1}{2} \left[k^2 + \frac{(x+a)^2 + \sigma^2}{\sigma^4} \right] I_l = k^2 I_r \quad (\text{I.1})$$

A função de Green, $G(x, x_0)$, deverá satisfazer

$$\frac{1}{2} G''(x, x_0) + \left(\frac{x+a}{\sigma^2} \right) G'(x, x_0) + \frac{1}{2} \left[k^2 + \frac{(x+a)^2 + \sigma^2}{\sigma^4} \right] G(x, x_0) = k^2 \delta(x - x_0) \quad (\text{I.2})$$

Considerando que a função impulso $\delta(x - x_0)$ se anula identicamente para $x > x_0$ ou $x < x_0$, $G(x, x_0)$ será a solução homogênea para (I.2) desde que $x \neq x_0$. Por outro lado, em $x = x_0$, ela deverá ser de tal forma que o lado esquerdo de (I.2) produza a descontinuidade associada à função δ . Integrando os dois lados da equação com relação a x , em torno de $x = x_0$, e assumindo que $G(x, x_0)$ é contínua, concluímos que deve existir uma descontinuidade na derivada da função em x_0 , ou seja,

$$G'(x_{0+}, x_0) - G'(x_{0-}, x_0) = 2k^2 \quad (\text{I.3})$$

onde $x_{0+} = x_0 + \varepsilon$, e $x_{0-} = x_0 - \varepsilon$, para um ε infinitesimal.

Verificamos então que uma função $G(x, x_0)$, satisfazendo tais condições e tendendo a zero no infinito, pode ser obtida como a parte imaginária do núcleo complexo:

$$K(x, x_0) = \begin{cases} N(x_0) W(x) \overline{W}(x_0) & x > x_0 \\ 0 & x < x_0 \end{cases} \quad (\text{I.4})$$

onde $W(x)$ é a solução homogênea complexa para (I.2), tendendo a zero quando $x \rightarrow \infty$. $K(x, x_0)$ vai satisfazer, portanto, a equação (I.2) para $x > x_0$ ou $x < x_0$, faltando apenas determinar o fator $N(x_0)$, a partir da condição de descontinuidade de (I.3). É fácil verificar que este será dado por

$$N(x_0) = \frac{2k^2}{\text{Im}[W'(x_0)\overline{W}(x_0)]} \quad \text{(I.5)}$$

e vamos finalmente obter

$$G(x, x_0) = 2k^2 \frac{\text{Im}[W(x)\overline{W}(x_0)]}{\text{Im}[W'(x_0)\overline{W}(x_0)]} \quad \text{(I.6)}$$

onde $W(x)$, a solução homogênea para (I.2), é a função de Gabor

$$W(x) = \exp[ikx] \exp\left[-\frac{(x+a)^2}{2\sigma^2}\right] \quad \text{(I.7)}$$

Substituindo (I.7) em (I.6), vamos então obter

$$G(x, x_0) = 2k \sin[k(x - x_0)] \exp\left[-\frac{(x+a)^2 - (x_0+a)^2}{2\sigma^2}\right] \quad \text{(I.8)}$$

que é parte imaginária do núcleo complexo em (4.8).

Seguindo os mesmos passos anteriores, a função de Green $G_u(x, x_0)$ para a equação de casamento de um campo uniforme de disparidade, equação (4.2), pode ser obtida de modo similar através da parte imaginária do filtro complexo $K_u(x, x_0)$, definido na equação (4.6).

Apêndice II

Aqui, vamos deduzir a equação (4.16), da Seção 4.3. Por simplicidade, omitiremos o subscrito r nas variáveis envolvidas.

O sinal de entrada:

$$I_1(x) = \exp \left[i(\omega x + \phi) \right] \exp \left[-\frac{(x + \alpha)^2}{2\sigma^2} \right] I(x) \quad \text{(II.1)}$$

é filtrado pelo filtro complexo de Green $K(x, x_0)$, onde

$$K(x, x_0) = 2k \exp \left[ik(x - x_0) \right] \exp \left[-\frac{(x + a)^2 - (x_0 + a)^2}{2\sigma^2} \right] \quad \text{(II.2)}$$

para $x > x_0$, com $K_u(x, x_0) = 0$, em caso contrário. O sinal resultante será:

$$I_2(x) = \int_{-\infty}^{\infty} K(x, x_0) I_1(x_0) dx_0 \quad \text{(OII.3)}$$

ou

$$I_2(x) = 2k \exp \left[i(kx + \phi) \right] \exp \left[-\frac{(x + a)^2}{2\sigma^2} \right] \times \quad \text{(II.4)}$$

$$\times \int_{-\infty}^x \exp(-ikx_0) \exp(i\omega x_0) \exp \left[\frac{-(x_0 + \alpha)^2 + (x_0 + a)^2}{2\sigma^2} \right] I(x_0) dx_0$$

a partir do qual, se substituimos a variável de integração por $\xi = x - x_0$, resulta:

$$I_2(x) = 2k \exp \left[i(kx + \phi) \right] \exp \left[-\frac{(x + \alpha)^2}{2\sigma^2} \right] \tilde{I}(x) \quad \text{(II.5)}$$

onde

$$\tilde{I}(x) = \int_0^\infty \exp \left[i(k - \omega) \xi \right] \exp \left[\frac{-(a - \alpha)}{\sigma^2} \xi \right] I(x - \xi) d\xi \quad \text{(II.6)}$$

Agora, vamos definir $\gamma = k - \omega$ e $\beta = (a - \alpha) / \sigma^2$, de modo que $\tilde{I}(x)$ seja reescrito como:

$$\tilde{I}(x) = \int_0^\infty \exp(i\gamma\xi) \exp(-\beta\xi) I(x - \xi) d\xi \quad \text{(II.7)}$$

ou

$$\tilde{I}(x) = \int_0^\infty \exp[-(1 - i)\beta\xi] \exp[i(\gamma - \beta)\xi] I(x - \xi) d\xi \quad \text{(II.8)}$$

e, finalmente,

$$\tilde{I}(x) = \exp[-i(\gamma - \beta)x] \int_0^\infty \exp[-(1 - i)\beta\xi] \exp[i(\gamma - \beta)(x - \xi)] I(x - \xi) d\xi \quad \text{(II.9)}$$

Comparando esta com (4.6) e (4.7), observamos que a integral acima é uma convolução com o filtro $K_u(x)$, para $u = 1/\beta = \sigma^2 / (a - \alpha)$. A equação (II.9) pode então ser rescrita como:

$$\tilde{I}(x) = \frac{1}{2\beta} \exp[-i(\gamma - \beta)x] \exp[i(\gamma - \beta)(x - u)] I(x - u) \quad \text{(II.10)}$$

Voltando à equação (II.4), finalmente obtemos:

$$I_2(x) = ku \exp[i(\omega x + \psi)] \exp\left[-\frac{(x + \alpha)^2}{2\sigma^2}\right] I(x - u) \quad \text{(II.11)}$$

onde $\psi = \phi + \Delta\phi$, para

$$\Delta\phi = 1 - \left(\frac{k - \omega}{a - \alpha}\right) \sigma^2 \equiv 1 - (k - \omega) u \quad \text{(II.12)}$$

A equação (4.22), relacionado com o deslocamento para a esquerda, pode ser deduzida de modo similar.

Referências Bibliográficas

1. J.R.A. Torreão, J.L. Fernandes. Matching photometric stereo images, *Journal Optical Society of America*. A15 (12) (1998) 2996-2975.
2. P.C.P Carvalho, J.R.A. Torreão, P.E. Ferreira Jr. Data-based Motion Simulation Through a Green's Function Approach, *sibgrapi*, pp. 193-199, *Computer Graphics and Image Processing, XVII Brazilian Symposium on (SIBGRAP'04)*, 2004.
3. B.K.P Horn, *Robot Vision*, Cambridge Mass.: MIT Press, 1986
4. Marr, D. and Poggio, T. A computational theory of human stereo vision. *Proceedings Royal Society London*, B(204):301—328,1979.
5. Wei, G.Q., Brauer, W., Hirzinger, G., 1998. Intensity- and gradient-based stereo matching using hierarchical Gaussian basis functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1143-1160.
6. N.Slesareva, A.Bruhn, and J.Weickert, Optic Flow Goes Stereo: A Variational Method for Estimating Discontinuity-Preserving Dense Disparity Maps., In *Pattern Recognition, Proc. 27th DAGM Symposium DAGM 2005, Vienna, Austria, September 2005* - W.Kropatsch, R.Sablatnig, A.Hanbury (Eds.) *Lecture Notes in Computer Science*, Vol.3663, Springer, Berlin, 33-40.
7. Qian, N., Zhu, Y, Binocular Receptive Field Models, Disparity Tuning, and Characteristic Disparity, *Neural Computation* 8, 1611-1641, 1996, MIT.
8. Wagner, H. , Frost. B, Disparity-sensitive cells in the owl have a characteristic disparity. *Nature*, 364:796-798.1993
9. Freeman, R. D., Ohzawa, I., 1990. On the neurofisiological organization of binocular vision. *Vision Research*, 30:1661-1676.
10. Julesz, B., Foundations of cyclopean perception, *Journal of the Optical Society of America* 59, p. 1544 (1969).
11. D. H. Hubel. *Eye, Brain, and Vision*. Scientific American Library, 1988-1995. <http://neuro.med.harvard.edu/site/dh/bcontext.htm>
12. Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology of London*, 160:106—154

13. Pettigrew, J., Nikara, T., and Bishop, P. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research*, 6:391--410.
14. Poggio, G., Motter, B., Squatrito, S., and Trotter, Y. (1985). Responses of neurons in visual cortex (v1 and v2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, 25:397--406.
15. Wagner, H. and Frost, B. (1993). Disparity-sensitive cells in the owl have a characteristic disparity. *Nature*, 364:796--798.
16. Maske, R., Yamane, S., Bishop, P. O. 1984. Binocular simple cells for local stereopsis: Comparison of receptive field organization for two eyes. *Vision Res.* 24, 1921-1929.
17. Bishop, P.O., Henry, G. H., Smith, C.J. 1971. Binocular interaction fields of single units in the cat striate cortex. *J. Physiol.* 216, 39-68.
18. Ohzawa, I., DeAngelis, G. C., Freeman, R.D. 1990. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science* 249, 1037-1041.
19. DeAngelis, G.C., Ohzawa, I., Freeman, R. D. 1991. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352, 156-159.
20. D. J. Fleet, D. J. Heeger, H. Wagner. Modelling Binocular Neurons in the Primary Visual Cortex, in "Computational and Biological Mechanisms of Visual Coding", M. Jenkin, K. Harris (Eds.), Cambridge University Press, 1997, 103-130
21. Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, 18, 359-368.
22. Qian, N. and Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research*, 37:1811-1827
23. S. T. Barnard. A stochastic approach to stereo vision, Proc. Fifth Nat. Conf. Artificial Intell. (Philadelphia, PA), 1986, pp. 676-680
24. B.K.P. Horn and B.G. Schunck. Determining optical flow. AI Memo 572. Massachusetts Institute of Technology, 1980.
25. Maki, A.; Bretzner, L.; Eklundh, J. Local Fourier Phase and Disparity Estimates: An Analytical Study. In Hlavac, Sara (eds.): Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns, Springer-Verlag, Berlin, Heidelberg, New York, pp. 868-873, September 1995.
26. Udo Ahlvers, Udo Zölzer. Improvement of phase-based algorithms for disparity estimation by means of magnitude information. *ICIP 2004*: 3025-3028
27. T. Sanger. Stereo Disparity Computation using Gabor Filters, *Biological Cybernetics*, vol. 59, 1988.

28. D.J. Fleet, A.D. Jepson, A.D. and M.R.M. Jenkin. Phase-based disparity measurement, *CVGIP: Image Understanding*, 53(2) (1991), pp.198-210
29. A. Fusiello, E. Trucco, A. Verri. A compact algorithm for rectification of stereo pairs, in *Machine Vision and Applications* (2000) 12: 16–22
30. Fusiello, A. Image Gallery. <http://profs.sci.univr.it/~fusiello/demo/smw/gallery/>. Consultada em 25/03/2007
31. CMU-VASC (1997). Carnegie Mellon University Image Database. <http://vasc.ri.cmu.edu/idb/>. Consultada em 25/03/2007
32. Ahlvers U (2005). Stereobilder Datenbank. http://www.hsu-hh.de/ant/index_pfbUvcu5IULdjkEN.html. Consultada em 25/03/2007
33. Watson, K. et al. (1998). Development of Stereoscopic-Haptic Virtual Environments. *Proc. of the 12th IEEE Symposium on Computer-Based Medical Systems*.
34. StereoGraphics Corporation (1997). *Stereographics Developers' Handbook: Background on Creating Imagens for CrystalEyes and SimulEyes*. http://www.stereographics.com/support/downloads_support/handbook.pdf. Acessado em 25/03/2007.
35. Fontoura, F.N.F. (2001) *Estereoscopia*. Curso de Especialização em Informática com Ênfase em Internet e aplicações de ensino. <http://atlas.ucpel.tche.br/~magic/compgraf/estereoscopia.html>. Acessado em 18/04/2002.
36. Chen, Y. and Qian, N. A Coarse-to-fine Disparity Energy Model with both Phase-shift and Position-shift Receptive Field Mechanisms, *Neural Computation*, 2004, 16:1545-1577
37. W.K. Pratt. *Digital Image Processing*, Wiley, New York, 1978.
38. Fleet DJ, Wagner H and Heeger DJ (1996). Neural Encoding of Binocular Disparity: Energy Models, Position Shifts and Phase Shifts. *Vision Res.* Vol 36: 1839-1857
39. D. G. Stork and H. R. Wilson. Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *J. Opt. Soc. Am. A* 7, 1362-1373 (1990).
40. Torreão, J. R. A. (2001). A Green's Function Approach to Shape from Shading, *Pattern Recognition* 34, pp. 2367-2382.
41. Guo-Qing Wei, Wilfried Brauer, Gerd Hirzinger. Intensity- and Gradient-Based Stereo Matching Using Hierarchical Gaussian Basis Functions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1143-1160, Nov., 1998.

42. Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2), 284–299.
43. Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *J. Neurophysiol.*, 77, 2879–2909.
44. Hassani, S. *Mathematical Physics*, Springer-Verlag, New York, 2002.
45. Torreão, J.R.A. & Amaral, M.S. (2006). Efficient, Recursively Implemented Differential Operator, with Application to Edge Detection, *Pattern Recognition Letters* 27(9), pp. 987-995.
46. Fetter AL, Walecka JD (1980). *Theoretical Mechanics of Particles and Continua*. McGraw-Hill, New York
47. Torreão JRA (2003). Geometric-photometric approach to monocular shape estimation. *Image and Vision Computing* 21: 1045-1061
48. Qian N (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Comput.* 6: 390-404
49. Qian N, Mikaelian S (2000). Relationship between phase and energy methods for disparity computation. *Neural Comput.* 12: 303-316
50. T. P. Weldon, W. E. Higgins, D. F. Dunn. Efficient Gabor filterdesign for texture segmentation, *Pattern Recognition*, vol. 29, no. 12,pp. 2005-2016, Dec. 1996