HALLYSSON OLIVEIRA

Uma Modelagem Computacional de Áreas Corticais do Sistema Visual Humano Associadas à Percepção de Profundidade

Dissertação apresentada ao Mestrado de Informática do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Informática.
Orientador: Prof. Dr. Alberto Ferreira De Souza.
Co-orientador: Prof. Dr. Neyval Costa Reis Jr.

Vitória
2005

HALLYSSON OLIVEIRA

Uma Modelagem Computacional de Áreas Corticais do Sistema Visual Humano Associadas à Percepção de Profundidade

Dissertação apresentada ao Mestrado de Informática do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Informática.

Aprovada em 15 de julho de 2005.

COMISSÃO EXAMINADORA

Prof. Dr. Alberto Ferreira de Souza
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Neyval Costa Reis Jr.
Universidade Federal do Espírito Santo
Co-orientador

Prof. Dr. Elias de Oliveira
Universidade Federal do Espírito Santo

Prof. Dr. Laércio Evandro Ferracioli da Silva
Universidade Federal do Espírito Santo

Prof[a]. Dr[a]. Priscila Machado Vieira Lima
Universidade Federal do Rio de Janeiro

Vitória
2005

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

Oliveira, Hallysson, 1979-

O48m Uma modelagem computacional de áreas corticais do sistema visual humano associadas à percepção de profundidade / Hallysson Oliveira. – 2005.

127 f. : il.

Orientador: Alberto Ferreira de Souza.

Co-Orientador: Neyval Costa Reis Júnior.

Dissertação (mestrado) – Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Profundidade - Percepção. 2. Visão por computador. 3. Modelagem computacional. I. Souza, Alberto Ferreira de. II. Reis Júnior, Neyval Costa. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 004

Aos meus pais, meu irmão e minha irmã.

AGRADECIMENTOS

A Deus em primeiro lugar por sempre ter me dado forças para fazer tudo.

Ao professor Alberto Ferreira de Souza pela orientação deste trabalho.

Aos meus amigos Dijalma, Sotério, João Paulo, Felipe, Sérgio, Leonardo e Fernando do LCAD, e em especial para Stiven que me ajudou em vários momentos.

Aos meus pais Getulio de Oliveira e Olga Rosa Vieira pelo apoio e incentivo.

A minha namorada Mileani pela paciência e compreensão, além de ter me ajudado quando não pude estar presente.

E a todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

EPÍGRAFE

"Ninguém é tão grande que não possa aprender, nem tão pequeno que não possa ensinar."

Píndaro, poeta romano.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

As imagens projetadas nas retinas humanas são bidimensionais, entretanto o cérebro é capaz de sintetizar uma representação tridimensional com informações sobre cor, forma e profundidade a respeito os objetos no ambiente ao seu redor. Para tanto, após escolher um ponto no espaço 3-D, os olhos vergem para este ponto de fixação e, em paralelo, o sistema visual é realimentado com informações sobre o posicionamento dos olhos, fornecidas pelo sistema óculo-motor, interpretando-as como a distância deste ponto ao observador. A percepção de profundidade ao redor do ponto de vergência pode ser obtida estimando a disparidade, isto é, a diferença de posição entre pontos correspondentes nas retinas, causada pela separação horizontal dos olhos na cabeça. A maior parte do processamento necessário à percepção da profundidade é realizada pelo córtex visual, principalmente nas áreas V1 e MT.

Neste trabalho é apresentada uma modelagem e implementação matemático-computacional de uma arquitetura neural, representando as áreas corticais V1 e MT, baseada em estudos da neurofisiologia do córtex visual com o intuito de criar num computador uma representação tridimensional do mundo externo, a partir de imagens obtidas de duas câmeras acopladas a este computador. A implementação deste modelo foi baseada em filtros de Gabor e mapeamento log-polar da retina para o córtex. Com este modelo é possível construir um mapa das disparidades entre as imagens e, a partir deste mapa, calcular a posição no espaço tridimensional de cada ponto da imagem representado no mapa. Desta forma, é possível criar uma representação tridimensional interna ao computador do ambiente ao seu redor.

ABSTRACT

The images formed on the human retinas has only two dimensions, however the brain is capable of synthesizing a three-dimensional representation with color, shape and depth information about the environment surroundings. For in such a way, after choosing a point in the 3-D space, the eyes verge to this fixation point and, at the same time, the visual system is fed back with eyes placement information, supplied by ocular motor system, considering them as the distance of this point to the observer. The depth perception around the vergence point may be got by estimating the disparity, the positional difference between the two retinal projections of a given point in space results from the horizontal separation of the eyes. Most of the depth perception processing is done by visual cortex, mainly in V1 and MT areas.

This work presents a modeling and implementation of a mathematical computational neural architecture that represents the V1 and MT cortical areas, based in visual cortex neurophysiology, used to create in a computer a 3-D representation of the surrounds, using images from two cameras connected to this computer. This model implementation was based upon Gabor filters and log-polar mapping. With this model it is possible to make a disparity map of the images and thus calculate the 3-D spatial position of each image point represented in the disparity map. This way it is possible to create a 3-D representation of the surrounds in a computer.

1. INTRODUÇÃO

Do mesmo modo que para seres biológicos, a visão constitui ferramenta importante no interfaceamento de robôs com o mundo real em várias aplicações. Dar a capacidade de fazer representações visuais a sistemas robóticos facilitaria a busca de objetos no meio físico, a visualização de seus detalhes, a transformação de instruções verbais de manipulação destes objetos em pictóricas, a memorização de conceitos a respeito deles, a sua manipulação espacial, entre outras tarefas. No entanto, os mecanismos subjacentes à visão e que possibilitam o cumprimento de tais tarefas não são tão óbvios e merecem especial atenção pelos estudiosos da ciência da cognição.

Estudos realizados nas áreas de neurofisiologia [16], reconhecimento de padrões e inteligência artificial mostram que o cérebro reconhece de forma paralela e distribuída várias características das imagens capturadas pelos olhos, tais como cor, forma, profundidade e movimento, utilizando, para isto, estratégias ainda não reproduzidas fielmente por computador. Desta forma, a compreensão da maneira pelo qual o cérebro humano efetua o processamento visual é importante para dotar o computador de capacidade de criar representações do mundo ao seu redor utilizando a visão.

Os neurofisiologistas da visão tentam identificar e compreender como funcionam os mecanismos e sistemas celulares que implementam a percepção visual, reconhecida pela maioria dos neurofisiologistas como sendo um processo construtivo, no qual o mundo físico externo é modelado e organizado no sistema perceptual, culminando em representações internas de objetos. Desta forma, o sistema visual integraria os sinais provenientes da retina em superfícies, e estas em objetos perceptuais. As imagens captadas pelos olhos seriam trabalhadas e modificadas, de várias formas, pelo conjunto das áreas visuais do cérebro para possibilitar esta integração.

Segundo Kandel [16] a percepção visual é um processo criativo. Os seres humanos possuem um “gênio” visual encarregado de resolver as diversas dificuldades inerentes ao processo de ver e a grande maioria nem se dá conta do processo. Ver é uma tarefa tão simples que a maioria não se atenta aos problemas a ela associados. Um exemplo da complexidade inerente à percepção visual é capacidade de criar uma percepção tri-dimensional do mundo apesar do fato das imagens projetadas nas retinas possuírem apenas duas dimensões.

Não se sabe ao certo de que forma o cérebro combina e transforma as imagens bidimensionais projetadas nas retinas do olho direito e do olho esquerdo numa imagem tridimensional, ou como o cérebro, a partir da percepção visual, distingue que um objeto está mais próximo ou distante de que outro, ou, ainda, como é estimada a profundidade relativa de um objeto dentro do campo visual. Estudos psicofísicos indicam que a transformação da imagem de bidimensional para tridimensional se baseia em dois tipos de indícios: pistas monoculares para profundidade e pistas estereoscópicas para disparidade binocular. As pistas monoculares são, por exemplo, o tamanho familiar dos objetos, oclusão, perspectiva, entre outros. As pistas estereoscópicas são oriundas da ligeira diferença entre as imagens projetadas na retina do olho direito e do olho esquerdo, chamada de disparidade binocular, devido ao afastamento horizontal entre os olhos.

O entendimento de como o cérebro combina todas estas informações permitiria realizar uma modelagem computacional de partes do cérebro capaz de criar uma representação tridimensional interna ao computador do mundo externo. Entretanto, para criar esta representação, ainda seria necessário resolver um outro problema: Como representar o mundo externo no sistema computacional? Que estrutura de dados utilizar para armazenar as informações extraídas de forma que possibilite a navegação de um sistema robótico, por exemplo? Para que o sistema robótico consiga navegar no meio externo é preciso que esta estrutura seja capaz de representar o ambiente externo de uma forma concisa e viável para navegação.

1.1. MOTIVAÇÃO

A visão é o sentido mais poderoso e complexo que se possui. Atualmente, o conhecimento acerca da visão biológica, em particular a visão humana, não é completo nem bastante detalhado. Isto tem motivado muitos pesquisadores a propor teorias e arquiteturas computacionais sobre como é feito o processamento da visão biológica para posterior comparação com o processo biológico, a fim de tentar compreendê-la cada vez mais. Alguns exemplos são os trabalhos de Adelson [1], Qian e Zhu [24] e [25], Fardin [12], Sturzl [26], Komati [17] e Ohzawa [21] e [22]. Essas teorias e arquiteturas têm evoluído na medida em que aumenta a compreensão sobre este processo.

Este trabalho está inserido numa linha de pesquisa do Departamento de Informática da UFES cujo objetivo de longo prazo é implementar uma arquitetura neural artificial com capacidades semelhantes, ou até mesmo superiores, que a capacidade humana. O grupo de

pesquisa envolvido iniciou seus trabalhos pelo estudo do sentido da visão, por ser o sentido que fornece mais informações sobre o mundo externo. Dentro deste contexto, neste trabalho é realizado um estudo de como modelar e implementar uma parte da percepção visual envolvida na construção de uma representação interna do mundo externo baseada em visão estereoscópica.

A introdução de uma representação fiel e consistente do mundo externo num sistema robótico forneceria informações suficientes para possibilitar que este sistema robótico possa navegar de forma autônoma no ambiente ao seu redor. Além disso, realizar esta tarefa baseando-se na neurofisiologia do cérebro humano implica em criar uma modelagem matemático-computacional de algumas partes de cérebro responsáveis por este processamento.

1.2. OBJETIVOS

Os principais objetivos deste trabalho foram: estudar como o cérebro humano, a partir de um par de imagens bidimensionais projetadas nas retinas, cria uma representação interna tridimensional do mundo externo e, utilizando este conhecimento, a partir de aspectos conhecidos da arquitetura visual humana, modelar e implementar computacionalmente parte do processamento realizado no cérebro humano, mais precisamente, o processo de percepção de profundidade que ocorre no córtex visual, devido à disparidade binocular dos olhos.

Este modelo poderá ser usado para capacitar um sistema robótico a criar uma representação interna do ambiente ao seu redor, utilizando visão estereoscópica, de forma a poder navegar e interagir com este ambiente.

1.3. CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

- A modelagem matemático-computacional de importantes áreas cerebrais envolvidas com a percepção de profundidade;
- Proposta de um modelo computacional para a uma área do córtex visual denominada MT (esta área será detalhada no próximo capítulo) que é de suma importância para a percepção de profundidade e que ainda não possui um modelo computacional consistente;

- Reconstrução tridimensional interna ao computador de superfícies do mundo externo a partir da imagem capturada por um par de câmeras, fornecendo assim dados suficientes para um sistema robótico navegar e interagir com o ambiente ao seu redor.

Neste trabalho de pesquisa foi modelado matematicamente o comportamento de células individuais agrupando estas células de modo a formar uma arquitetura, sempre buscando, com esta arquitetura, emular partes do córtex visual humano cuja resposta eletroquímica observada é conhecida.

Estimulada por imagens estéreo do mundo real, esta arquitetura responde com um padrão global de comportamento dos seus neurônios que imita parte do comportamento neural de uma importante região do córtex envolvida com a detecção da disparidade das imagens capturadas pelos olhos (área MT). Foi desenvolvido todo o código computacional que, a partir da disparidade disponibilizada pela arquitetura e da geometria do sistema composto pelas câmeras (seu afastamento físico e suas características óticas), é capaz de gerar uma representação tridimensional das superfícies na frente das câmeras. Ou seja, a arquitetura desenvolvida mais o código que gera a representação tridimensional permitem ao computador a percepção de profundidade de um modo similar ao que o sistema visual biológico nos permite a percepção de profundidade.

2. SISTEMA VISUAL HUMANO

Os mecanismos subjacentes ao sistema visual humano não são óbvios. Não se sabe ao certo como o cérebro nos permite perceber características do mundo visual, tais como forma, profundidade, cor e movimento. As estratégias adotadas pelo cérebro para reconhecer estas características envolvem processos complexos e ainda não totalmente compreendidos.

As várias características do mundo visual reconhecidas pela percepção visual (forma, profundidade, cor e movimento), são processadas por diferentes regiões do cérebro, que aparentemente não estão correlacionadas, mas são, de alguma forma, integradas numa única percepção. Neste capítulo é apresentada uma visão geral sobre estas regiões do cérebro humano, que fazem parte do sistema visual humano, analisando como os sinais luminosos capturados pelos olhos são processados em cada uma delas. Não serão abordados os processamentos de cor e movimento, uma vez que estes processamentos não são essenciais para a percepção de profundidade por meio de disparidade binocular.

2.1. O OLHO

O globo ocular, com cerca de 25 milímetros de diâmetro, é o responsável pela captação da luz refletida pelos objetos. Anatomicamente, o globo ocular fica alojado em uma cavidade formada por vários ossos chamada órbita e é constituído por três túnicas (camadas): túnica fibrosa externa, túnica intermédia vascular pigmentada e túnica interna nervosa.

Na Figura 2-1, estão representadas todas as partes que formam as túnicas fibrosa externa, intermédia vascular pigmentada e a túnica interna nervosa. A túnica fibrosa externa ou esclerótica, também chamada de “branco do olho”, tem uma função protetora. É resistente, de tecido fibroso e elástico, e envolve externamente o olho (globo ocular). A maior parte da esclerótica é opaca e chama-se esclera. A ela estão conectados os músculos extra-oculares que movem os globos oculares, dirigindo-os ao seu objetivo visual. A parte anterior da esclerótica chama-se córnea, que é transparente e atua como uma lente convergente.

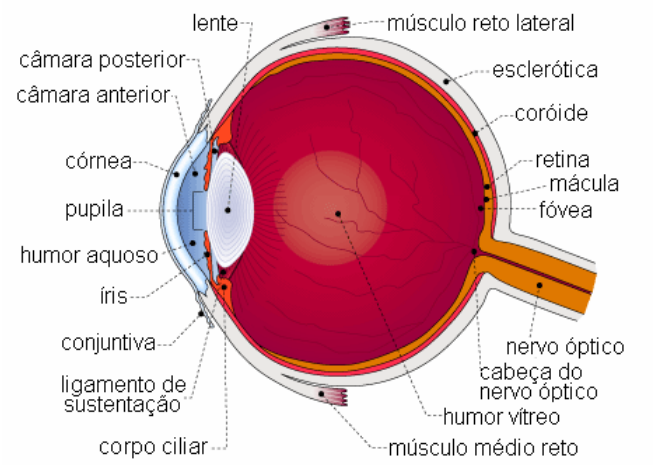

Figura 2-1 - Anatomia do olho humano. Corte medial horizontal do olho direito visto de cima. Fonte: http://www.escolavesper.com.br/olho_humano.htm. Acesso em: 22 jan. 2005.

A túnica intermédia vascular pigmentada ou úvea compreende a coróide, o corpo ciliar e a íris. A coróide está situada abaixo da esclerótica e é bastante pigmentada para absorver a luz que chega a retina, evitando sua reflexão dentro do olho. Ela é intensamente vascularizada e tem também como função nutrir a retina. A íris é uma estrutura muscular de cor variável (parte circular que dá cor aos olhos), é opaca e tem uma abertura central, chamada pupila, por onde a luz passa. O diâmetro da pupila varia, aproximadamente de 2mm a 8mm, de acordo com a intensidade luminosa do ambiente. Em ambientes claros a pupila se estreita, diminuindo a passagem de luz, evitando a saturação das células detectoras de luz da retina. No escuro a pupila se dilata, aumentando a passagem de luz e sua captação pela retina. A luz que passa pela pupila atinge imediatamente o cristalino, uma lente gelatinosa que focaliza os raios luminosos sobre a retina.

O corpo ciliar é uma estrutura formada por musculatura lisa e que envolve o cristalino (a lente do olho). Ele á capaz de mudar a forma do cristalino permitindo assim ajustar a visão para objetos próximos ou distantes. Este processo é conhecido como acomodação visual. A convergência correta do cristalino faz com que a imagem seja projetada nitidamente na retina. Se a imagem for maior ou menor que a necessária, fica fora de foco. Se o cristalino está ajustado para uma certa distância de um objeto, e este objeto se aproxima, a imagem perde a nitidez. Para recuperá-la, o corpo ciliar aumenta a convergência do cristalino, acomodando-o, diminuindo a distância focal. Caso o objeto se afaste, ocorre o processo inverso. Este processo está ilustrado na Figura 2-2.

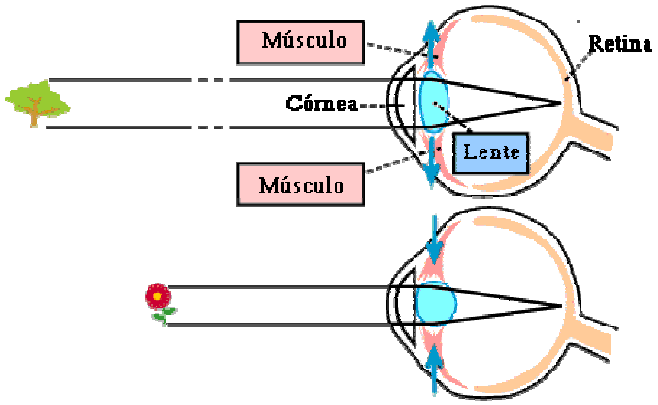

Figura 2-2 - Figura mostrando a acomodação visual do cristalino.

A túnica interna nervosa é a retina. É na retina que se formam as imagens visualizadas. A imagem projetada na retina é invertida, mas isto não causa nenhum problema já que o cérebro se adapta a isto desde o nascimento. Para que a imagem seja projetada na retina, a luz percorre o seguinte caminho: primeiramente a luz atinge a córnea, que conforme visto anteriormente é um tecido transparente, passa pela pupila, que é a abertura situada na íris que regula a intensidade de luz que entra no olho, atravessa o cristalino, que é a lente gelatinosa e que tem a função de focalizar a imagem na retina, atravessa um fluido viscoso chamado humor vítreo, que preenche a região entre o cristalino e a retina, e finalmente atinge a retina.

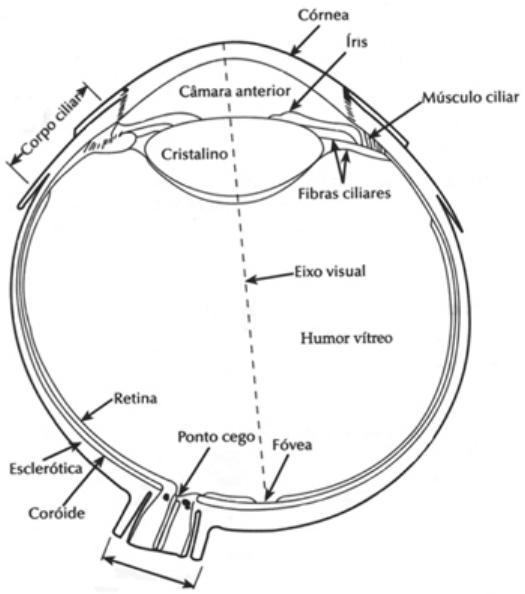

Figura 2-3 - Eixo visual. Fonte: http://www.on.br/glossario/alfabeto/o/olho_humano.html. Acesso em: 22 jan. 2005.

A retina é composta por mais de 100 milhões de células fotossensíveis: cerca de 7 milhões de cones e entre 75 milhões e 150 milhões de bastonetes. Estas células, quando excitadas pela energia luminosa, estimulam as células nervosas adjacentes, gerando um impulso nervoso que é propagado pelo nervo óptico. A imagem fornecida pelos cones é mais nítida e rica em detalhes. Os cones são sensíveis às cores. Há três tipos de cones: um que se excita com luz vermelha, outro com luz verde e outro com luz azul. Os bastonetes não têm poder de resolução visual tão bom nem conseguem detectar cores, mas são mais sensíveis à luz. Em situações de pouca luminosidade a visão passa a depender exclusivamente dos bastonetes.

As imagens dos objetos visualizados diretamente são projetadas normalmente numa região da retina chamada *fovea centralis* ou simplesmente fóvea com cerca de 1,5 mm de diâmetro e que fica na direção da linha (eixo visual) que passa pela córnea, pupila e pelo centro do cristalino (Figura 2-3). Na Figura 2-4, é mostrada a distribuição de cones e bastonetes na retina. Os cones são encontrados na retina central, em um raio de aproximadamente 10 graus a partir da fóvea. Os bastonetes estão localizados principalmente na retina periférica. O local da retina de onde sai o nervo óptico não possui cones nem bastonetes. Este local é chamado de ponto cego porque uma imagem que forme sobre este ponto, não é vista.

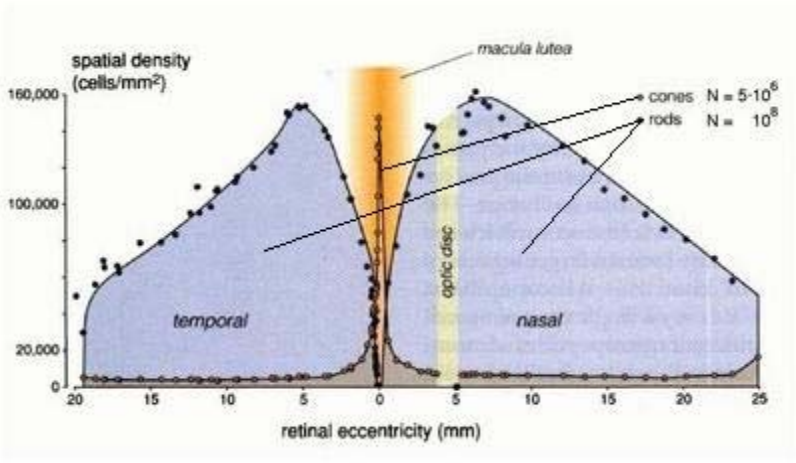

Figura 2-4 - Distribuição de cones e bastonetes (*rods*) na retina. A *macula lutea* (região da fóvea e vizinhanças) possui alta densidade de cones, enquanto que os bastonetes se concentram na periferia. O ponto cego fica na região do disco óptico (*optic disc*). Fonte: http://www.brainworks.uni-freiburg.de/group/wac. Acesso em: 22 jan. 2005.

Os neurônios de saída da retina são as células ganglionares que projetam seus axônios através do nervo óptico, levando a informação visual para o cérebro. Entre as células ganglionares e as células fotorreceptoras (cones e bastonetes), existem três classes de interneurônios: células bipolares, horizontais e amácrinas. Estas células transmitem os sinais dos fotorreceptores para as células ganglionares combinando sinais de vários fotorreceptores e fazendo com que a saída de cada célula ganglionar da retina dependa de um padrão preciso de luz e de como este padrão muda com o tempo [16].

Cada célula ganglionar recebe informações de um conjunto de células fotorreceptoras vizinhas em uma área circunscrita na retina que é o seu **campo receptivo**. O campo receptivo de uma célula ganglionar é área da retina na qual esta célula controla. Duas características importantes podem ser percebidas nos campos receptivos das células ganglionares. Primeiro, são aproximadamente circulares. Segundo, são divididos em duas partes: um círculo central e um anel periférico, conforme pode ser visto na Figura 2-5.

As células ganglionares respondem bem a uma iluminação diferencial entre o centro e a periferia dos seus campos receptivos. Assim, é possível identificar dois tipos de células ganglionares: *on center* e *off center*. Células ganglionares com campos receptivos *on center* ficam excitadas quando a luz estimula o centro e ficam inibidas quando a luz estimula o contorno do campo receptivo. Células *off center*, funcionam ao contrário, ficam excitadas quando a luz estimula a periferia e ficam inibidas quando a luz estimula o centro do campo receptivo. Os sinais visuais de intensidade luminosa (na verdade, de contraste), depois das transformações feitas na retina, são levados até o cérebro pelo nervo ótico.

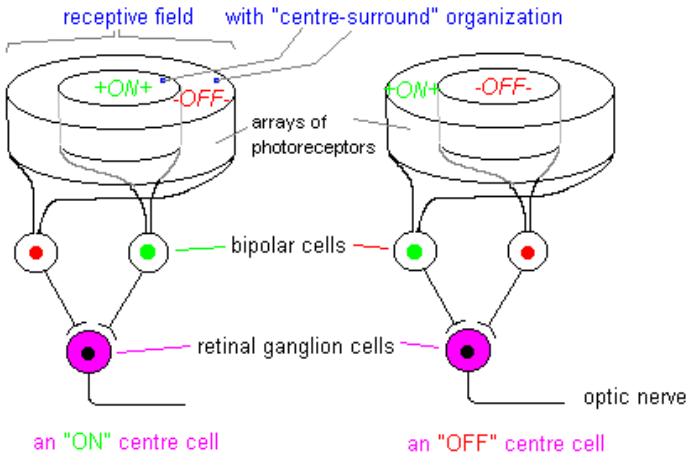

Figura 2-5 - Campo receptivo das células ganglionares. Fonte: http://www.cf.ac.uk/biosi/staff/jacob/teaching/sensory/vision.html. Acesso em: 25 jan. 2005.

2.2. FLUXO DE INFORMAÇÕES VISUAIS

Nesta seção será descrito o fluxo de informações visuais em dois estágios: primeiro, a informação visual saindo da retina e indo para o mesencéfalo e tálamo (Primeiro Estágio na Figura 2-6), e depois, a informação saindo do tálamo para o córtex visual primário (Segundo Estágio na Figura 2-6). Para tanto, serão definidos alguns conceitos a seguir.

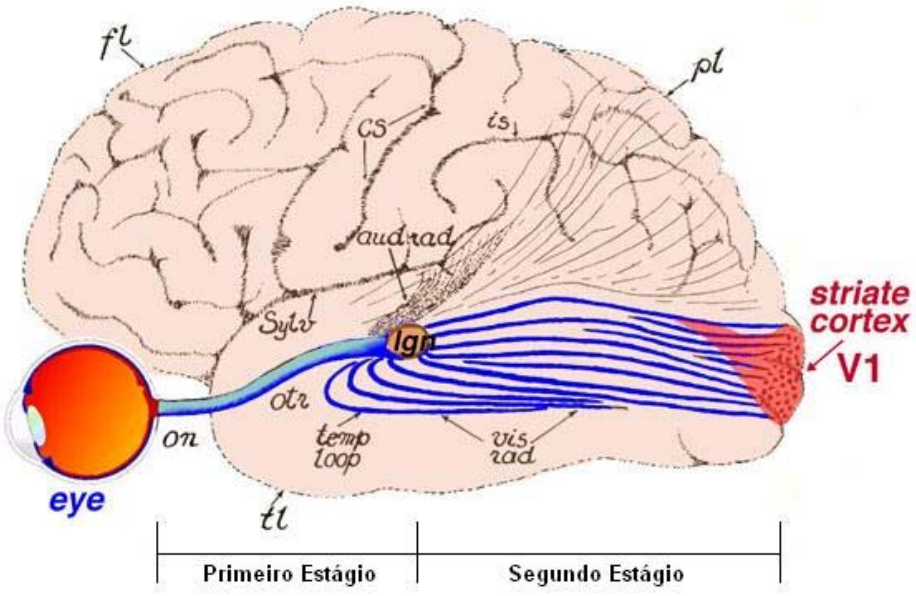

Figura 2-6 - Fluxo das informações visuais. Fonte: http://webvision.med.utah.edu/VisualCortex.html com alterações. Acesso em: 17 nov. 2004.

A retina pode ser dividida em duas partes: hemirretina nasal e hemirretina temporal, cuja separação é uma linha imaginária que corta o olho de cima a baixo passando pela fóvea. Numa situação em que as fóveas de ambos os olhos estão fixas num ponto do espaço situado em linha reta com o nariz, é possível dividir o campo visual em *left hemifield* (campo visual esquerdo) à esquerda do ponto fixo no espaço e *right hemifield* (campo visual direito) à direita do ponto fixo no espaço. O *left hemifield* é projetado na hemirretina nasal do olho esquerdo e na hemirretina temporal do olho direito. O *right hemifield* é projetado na hemirretina nasal do olho direito e na hemirretina temporal do olho esquerdo, conforme mostrado na Figura 2-7.

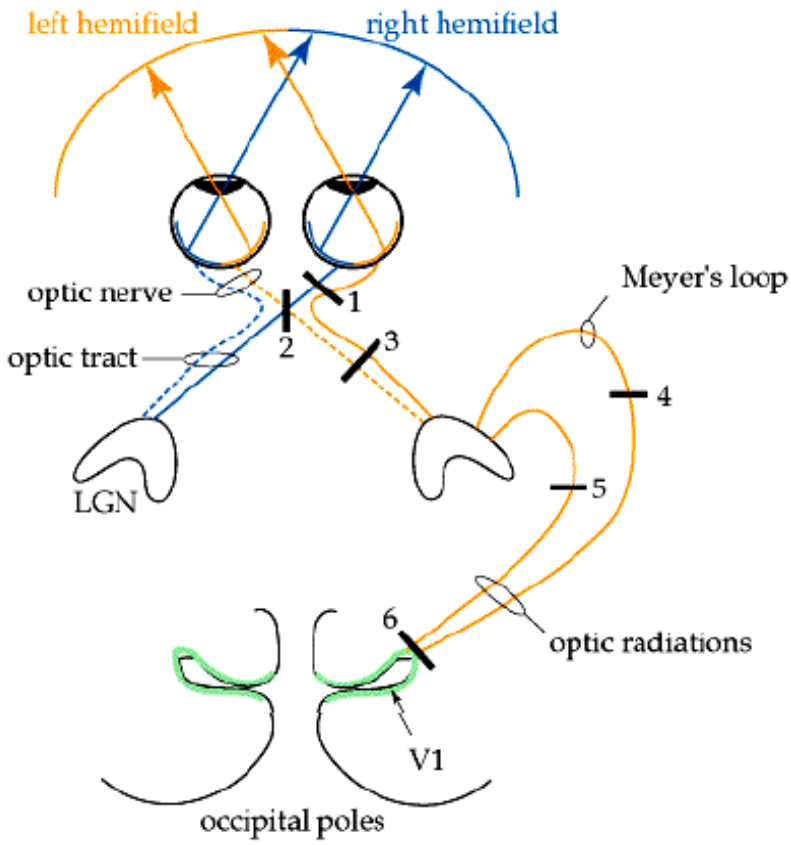

Figura 2-7 - Campo visual. (1) Nervo óptico, (2) Quiasma óptico, (3) Trato óptico. Fonte: http://thalamus.wustl.edu/course/basvis.html. Acesso em: 28 nov. 2004.

O nervo óptico de cada olho projeta-se para o quiasma óptico que é a região onde é feita a separação das fibras de cada olho, em tratos ópticos, destinadas para um mesmo lado do cérebro. Os tratos ópticos se projetam cada uma para três áreas subcorticais simétricas (que existem nos dois lados de cérebro): região pretectal ou *pretectum*, o *superior culliculus* do mesencéfalo e o *lateral geniculate nucleus* (LGN) do tálamo conforme mostra a Figura 2-8.

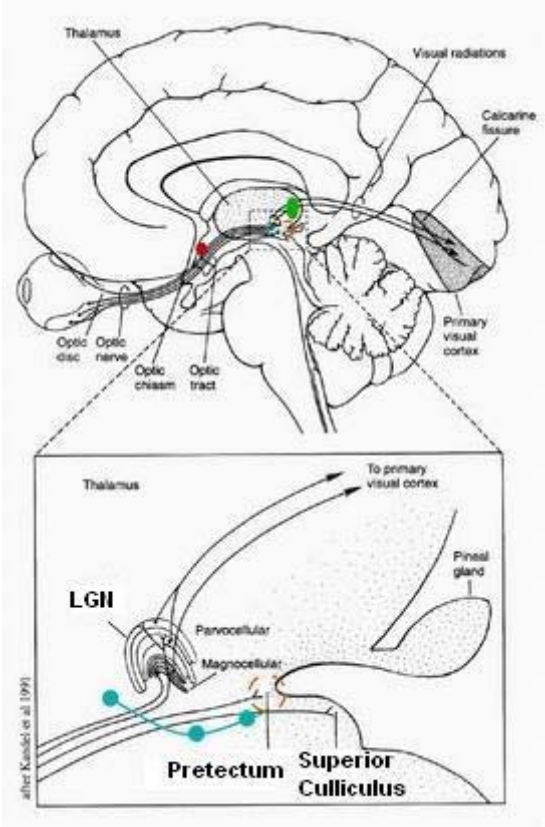

Figura 2-8 - Projeções da retina no mesencéfalo. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Acesso em: 22 jan. 2005.

A região pretectal do mesencéfalo possui células que se projetam bilateralmente para os neurônios do sistema simpático/parassimpático que controla os reflexos pupilares, contraindo e dilatando a pupila de acordo com a quantidade de luz que incide nos olhos. O *superior culliculus* é uma estrutura de camadas alternantes cinzentas e brancas localizada no teto do mesencéfalo. As células das camadas superficiais projetam-se para uma vasta área do córtex cerebral, formando uma via indireta da retina para o córtex cerebral. As camadas superficiais também recebem sinais provenientes do córtex visual enquanto que as camadas mais profundas recebem projeções de várias outras áreas do córtex ligadas a outros sentidos. O *superior culliculus* possui um mapeamento visual além de responder a estímulos auditivos e somatossensórios.

Células das camadas mais profundas do *superior culliculus* respondem positivamente antes dos movimentos sacádicos dos olhos, no qual os olhos trocam rapidamente de um ponto de fixação para outro numa cena. Estas células formam um mapa de movimento sacádico ordenado com o mapa visual. Para controlar os movimentos sacádicos, o *superior culliculus* recebe informações não só da retina, mas também do córtex cerebral.

O *lateral geniculate nucleus* (LGN) é o ponto de retransmissão das informações visuais provenientes da retina para o córtex visual. Cerca de 90% dos axônios da retina chegam até o LGN, que possui uma representação retinotópica da metade contralateral (lado oposto) do campo visual.

A razão entre uma área do LGN e uma área correspondente da retina que representa um grau do campo visual é chamada de fator de magnificação daquela área do LGN. A fóvea possui uma representação relativamente maior que a retina periférica no LGN, ou seja, as regiões do LGN que monitoram a fóvea possuem um maior fator de magnificação.

Nos primatas, incluindo os humanos, o LGN é formado por seis camadas numeradas de 1 (ventral) a 6 (dorsal). As duas camadas mais ventrais (camadas 1 e 2) contém células relativamente grande que recebem conexões das células ganglionares M da retina e são conhecidas como camadas magnocelulares enquanto que as outras 4 camadas dorsais (camadas 3, 4, 5 e 6) contém células que recebem conexões das células ganglionares P da retina e são conhecidas como camadas parvocelulares. A Figura 2-9 mostra de forma esquemática o mapeamento da retina nas diversas camadas do LGN. Nela é possível ver como o fator de magnificação varia da fóvea para a periferia no LGN.

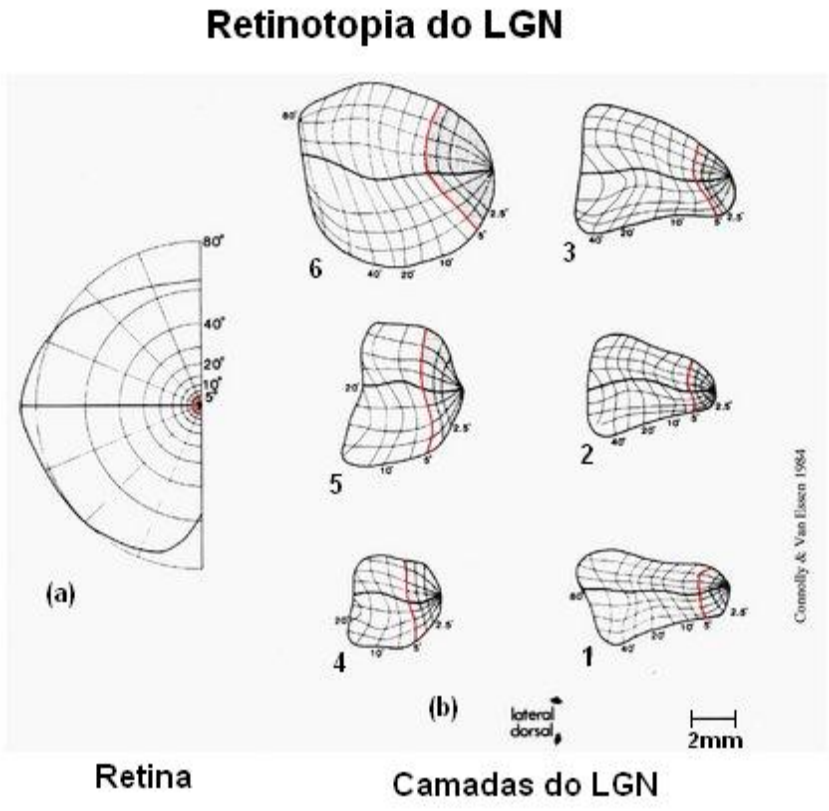

Figura 2-9 - Retinotopia do LGN. (a) Mapeamento da retina. (b) Mapeamento da retina nas camadas 1 à 6 do LGN. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Aceso em: 22 jan. 2005.

Todas as camadas do LGN possuem células com campo receptivo *on center* e *off center*, sendo que cada camada recebe sinais somente de um olho. As fibras da hemirretina nasal contralateral (outro lado) são projetadas nas camadas 1, 4 e 6, enquanto que as fibras da hemirretina temporal ipsilateral (mesmo lado) são projetadas nas camadas 2, 3 e 5 conforme mostrado na Figura 2-10.

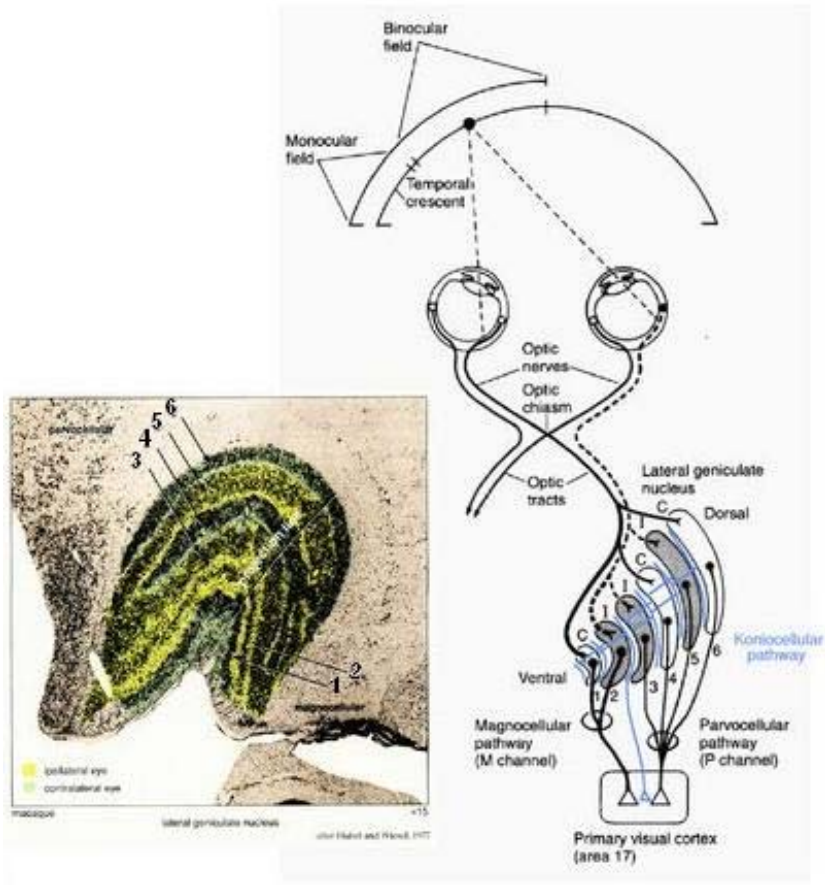

Figura 2-10 - LGN. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Acesso em 22 jan. 2005.

As células das camadas magnocelulares e parvocelulares do LGN projetam-se para o córtex visual formando duas vias independentes (vias M e P) que se estendem desde a retina até o córtex visual primário. A via P é essencial para a visão de cores e sensível a estímulos de alta freqüência espacial e baixa freqüência temporal da imagem na retina. A via M é mais sensível a estímulos de baixa freqüência espacial e alta freqüência temporal.

Na Figura 2-11 é ilustrado o fluxo visual desde a retina até o córtex estriado (V1), mostrando como a imagem é subdividida ao chegar no córtex estriado. O campo visual do olho esquerdo corresponde do ponto 1 ao ponto 8 e o campo visual do olho direito

corresponde do ponto 2 ao ponto 9. É possível notar que, que conforme esquematizado na Figura 2-7, a parte esquerda da imagem (campo visual esquerdo ou *left hemifield*) compreendida dos pontos 1 à 5, é projetada na hemirretina nasal do olho esquerdo e na hemirretina temporal do olho direito enquanto que a parte direita da imagem, compreendida do pontos 5 à 9 é projetada na hemirretina nasal do olho direito e na hemirretina temporal do olho esquerdo. A informação visual então é conduzida pelos nervos ópticos e separadas no quiasma óptico para que as informações de um mesmo lado da imagem seja destinada a um mesmo lado do cérebro.

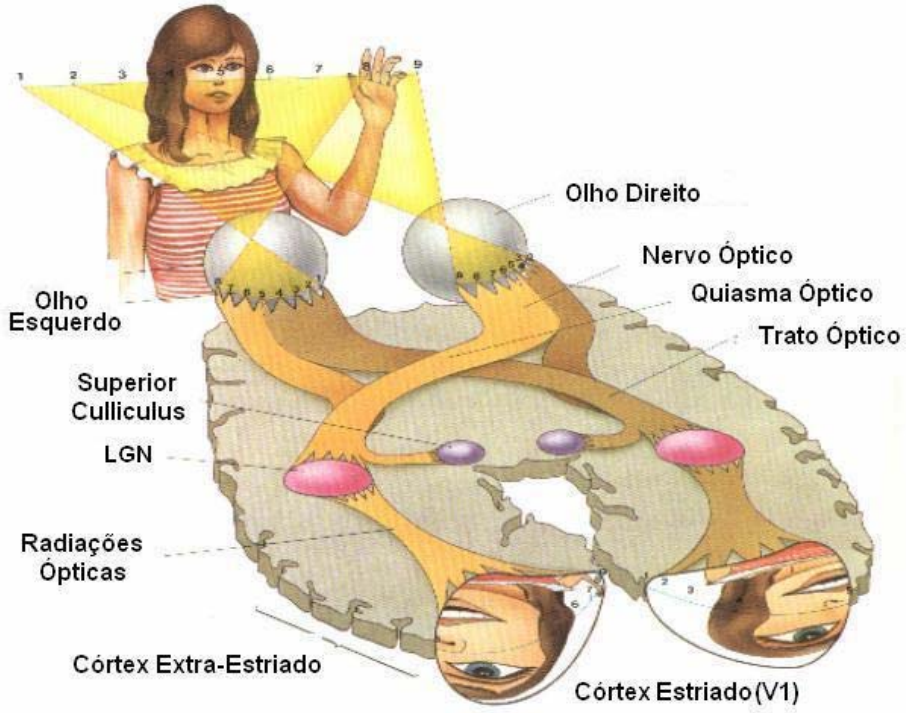

Figura 2-11 - Exemplo ilustrando o fluxo de informações visuais da retina ao córtex estriado. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Acesso em: 22 jan. 2005.

2.3. ORGANIZAÇÃO DO CÓRTEX VISUAL

A informação visual é processada em diversas áreas corticais, sendo que cada uma delas contribui diferencialmente para o processamento da percepção de movimento, profundidade, forma e cor. Serão descritas brevemente cinco áreas corticais visuais mais diretamente ligadas a este trabalho: V1, V2, V3, V4 e MT (também conhecida como V5). Na Figura 2-12-A é apresentada uma vista lateral de um hemisfério do cérebro de um macaco e na Figura 2-12-B é mostrado este hemisfério estendido de modo a formar um plano, onde são indicadas com uma tonalidade mais escura as áreas corticais visuais e são apontadas as áreas

V1, V2, V3, V4 e MT. Como é possível observar na Figura 2-12-A, as áreas corticais visuais ocupam aproximadamente metade do córtex de um macaco.

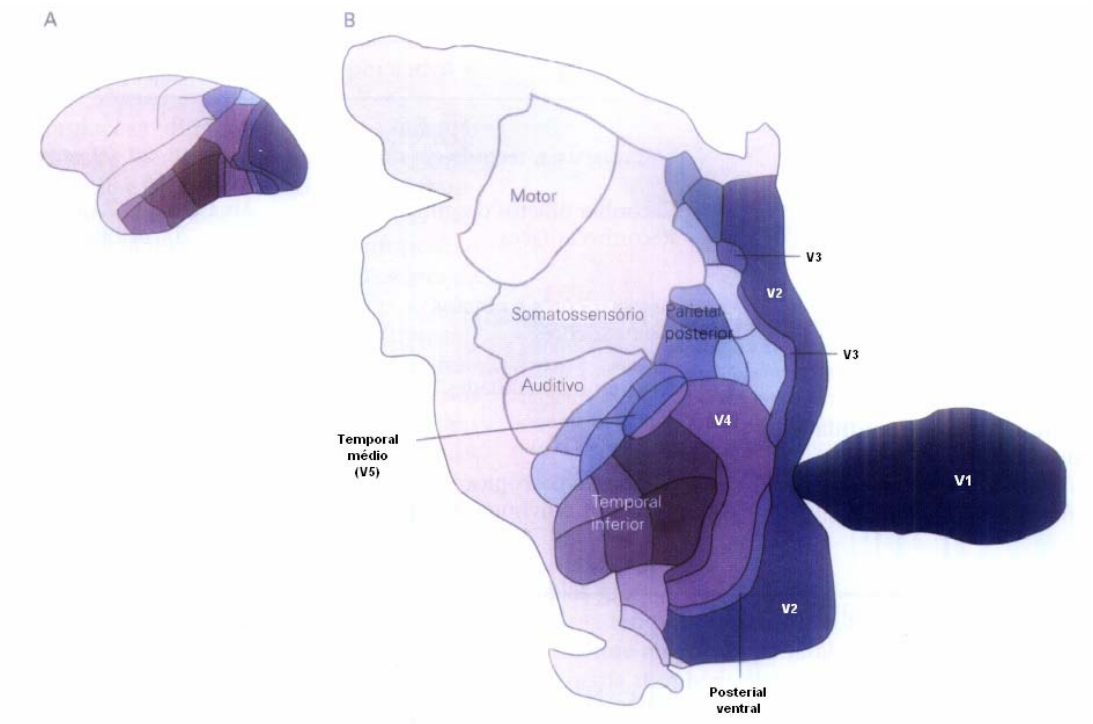

Figura 2-12 - Áreas corticais. Fonte: [16].

2.3.1. V1

Quase toda informação visual vinda da retina entra no córtex via a área V1 que, devido a sua aparência estriada, também é conhecida como córtex estriado. As outras áreas são conhecidas como córtex extra-estriado. V1 também é chamada de córtex visual primário e de área 17 de Brodmann [16]. Nos humanos o córtex visual primário possui cerca de 2mm de espessura e é dividido em 6 camadas numeradas de 1 à 6 (Figura 2-13). A camada 4 é a que recebe a maioria das projeções dos axônios do LGN e pode ser dividida em 4 sub camadas: 4A, 4B, 4Cα e 4Cβ. Os axônios de células das camadas parvocelulares (P) do LGN terminam principalmente na camada 4Cβ com algumas poucas projeções para 4A e 1, enquanto que os axônios de células das camadas magnocelulares (M) terminam principalmente na camada 4Cα.

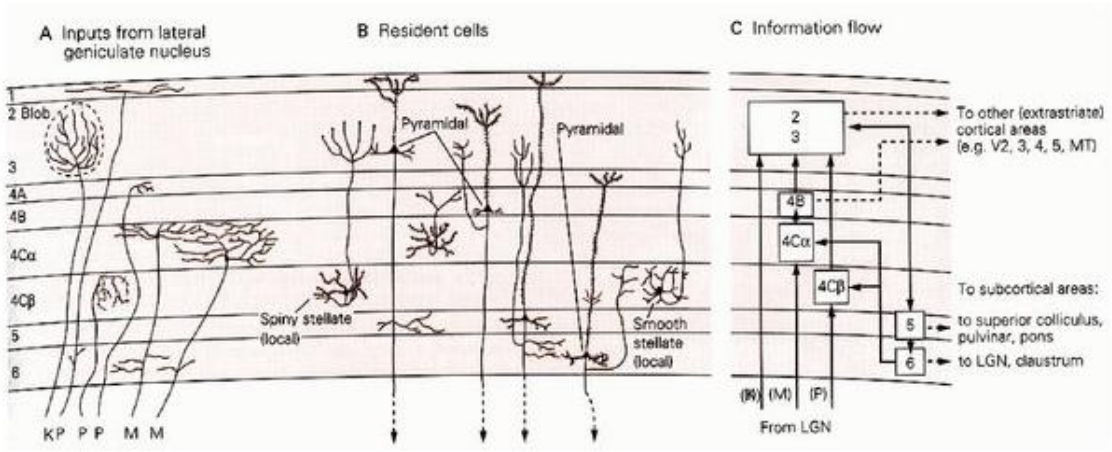

Figura 2-13 - Organização de V1. A) Os axônios dos neurônios P e M do LNG terminam na camada 4; B) Células de V1; C) Concepção do fluxo de informação em V1. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Acesso em 22 jan. 2005.

Através de estudos feitos em macacos verificou-se que o córtex estriado, assim como LGN, possui um mapa retinotópico, isto é, áreas do campo visual vizinhas da retina são também vizinhas em V1, do campo visual contralateral [33]. O aspecto mais importante deste mapa é que cerca da metade das projeções da retina sobre o córtex visual primário são provenientes da fóvea e regiões circunvizinhas. Este aspecto pode ser visto na Figura 2-14. Esta área apresenta a maior acuidade visual.

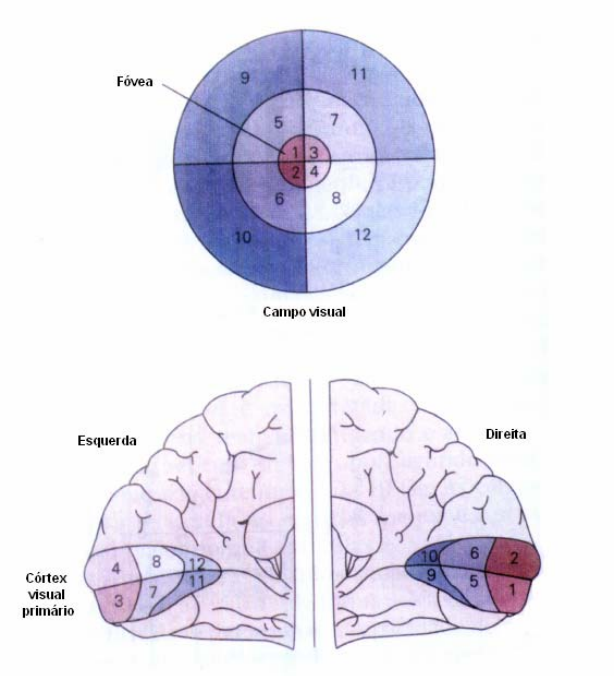

Figura 2-14 - Campo visual representado no córtex visual primário humano. Fonte: [16].

O formato do campo receptivo das células de V1 é diferente do formato dos campos receptivos das células da retina e do LGN que são circulares. Em V1, os campos receptivos das células são alongados e, conseqüentemente, respondem melhor à estímulos alongados do que à estímulos pontuais. Hubel e Wiesel [14] classificaram as células de V1 de acordo com a complexidade de sua resposta, dividindo-as em dois grupos chamados simples e complexos.

As células simples também possuem campo receptivo com regiões excitatórias e inibitórias, contudo estas regiões têm seu formato alongado. Estas células respondem melhor à estímulos na forma de barras com uma orientação específica. Uma célula simples que responde melhor a um estímulo vertical não responderá bem à um estímulo horizontal ou oblíquo, e vice-versa.

Na Figura 2-15 é apresentado de forma esquemática o comportamento de uma célula simples quando um estímulo em forma de barra é projetado em seu campo receptivo. Para produzir os resultados mostrados na Figura 2-15 o pesquisador, tipicamente, monitora a tensão no interior da célula através de um micro eletrodo (na verdade, uma micropipeta que perfura a parede celular), ao mesmo tempo em que o animal cuja célula está sendo monitorada observa um estímulo. Na Figura 2-15, quatro estímulos na forma de barra são mostrados posicionados sobre uma representação do campo receptivo da célula; os três primeiros, da esquerda para a direita, são barras brancas estreitas e de mesma largura, e o quarto é uma barra branca larga que cobre todo o campo receptivo. O campo receptivo em questão possui orientação vertical, sendo que a parte do mesmo que excita a célula é central e as que inibem ficam nas laterais esquerda e direita. Abaixo de cada conjunto estímulo-campo receptivo são mostrados dois gráficos. O primeiro, imediatamente abaixo de cada conjunto estímulo-campo receptivo, mostra o momento em que o estímulo está desligado ou ligado (trata-se do comportamento de um sinal elétrico ao longo do tempo que, no nível baixo, indica que o estímulo está inativo e, no nível alto, indica que o estímulo está ativo). O segundo representa o sinal capturado pelo micro eletrodo conectado à célula.

Como o primeiro par de gráficos Figura 2-15 (o mais a esquerda) mostra, a célula simples responde fortemente (emitindo vários pulsos pouco afastados no tempo, que é o modo como as células do córtex sinalizam sua ativação) quando o estímulo é ligado estando corretamente orientado e posicionado sobre a parte central do campo receptivo. A resposta da célula é mais vigorosa imediatamente após o acionamento do estimulo, o que mostra um aspecto temporal da resposta da célula. Na verdade, permanecendo o estímulo por muito tempo (de dezenas de segundos a alguns minutos), a resposta da célula desapareceria

totalmente, por um processo conhecido como acomodação. Mas um estímulo constante por muito tempo não ocorre naturalmente, uma vez que os olhos são movimentados continuamente.

Como o segundo par de gráficos da Figura 2-15 mostra, quando o estímulo é posicionado sobre a parte do campo receptivo que inibe a célula ela não responde no momento em que o estímulo é ligado, embora responda, fracamente, imediatamente após o estímulo ser desligado (também uma evidência do aspecto temporal da resposta da célula). Nos outros casos a célula não responde.

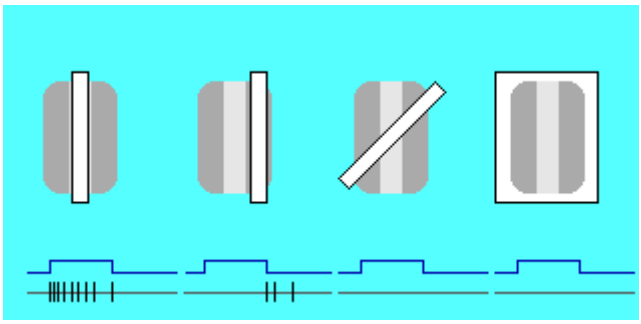

Figura 2-15 – Resposta de uma célula simples em função da projeção de um estímulo em forma de barra. Fonte: [18]

As células complexas são mais numerosas em V1 do que as células simples e, assim como as células simples, respondem bem apenas para um estímulo com uma orientação específica. Porém, diferentemente das células simples, a resposta das células complexas não é seletiva à posição espacial do estímulo, ou seja, não varia com a posição do estímulo dentro do seu campo receptivo. Muitas células complexas são sensíveis ao sentido e direção do movimento do estímulo dentro do seu campo receptivo, respondendo somente quando este estímulo se move numa determinada direção e sentido.

Na Figura 2-16 é apresentado o comportamento de uma célula complexa quando um estímulo em forma de barra é projetado no seu campo receptivo. A célula complexa responde independente da posição do estímulo no campo receptivo, diferentemente da célula simples. As células complexas também são sensíveis à movimentação do estímulo dentro de seu campo receptivo. Caso o estímulo se mova no mesmo sentido em que o campo receptivo esteja sintonizado, a célula continua respondendo, caso contrário para de responder ao estímulo.

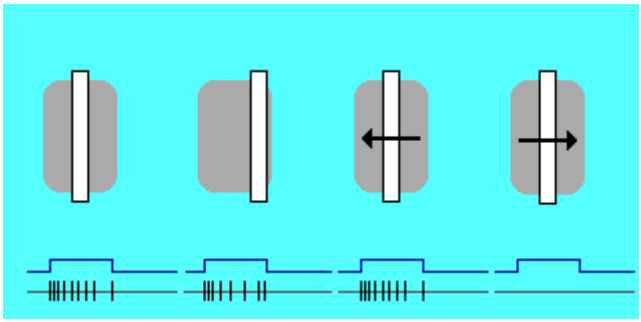

Figura 2-16 - Resposta de uma célula complexa em função da projeção de um estímulo em forma de barra. Fonte: [18]

Hubel e Wiesel foram os primeiros a descobrir que as células de V1 são arranjadas e organizadas de uma forma precisa em relação à sensibilidade à orientação. Ao longo da superfície de V1, a sensibilidade à orientação varia gradualmente, mas permanece constante ao longo dos 2mm de córtex (de uma coluna do córtex), conforme visto na Figura 2-17A na qual cada cor representa que as células naquela região possui sensibilidade à direção das barras da mesma cor. Hubel e Wiesel também descobriram que a resposta das células varia de acordo com o olho estimulado. Muitas células de V1 respondem de forma aproximadamente equivalente a estímulos provenientes de ambos os olhos, mas a maioria das células de V1 respondem preferencialmente à estímulos provenientes de um determinado olho. Esta característica, chamada de dominância ocular, é organizada no córtex visual primário de uma forma que não varia verticalmente (em colunas), mas alterna ao longo da superfície de V1 como pode ser visto na Figura 2-17B.

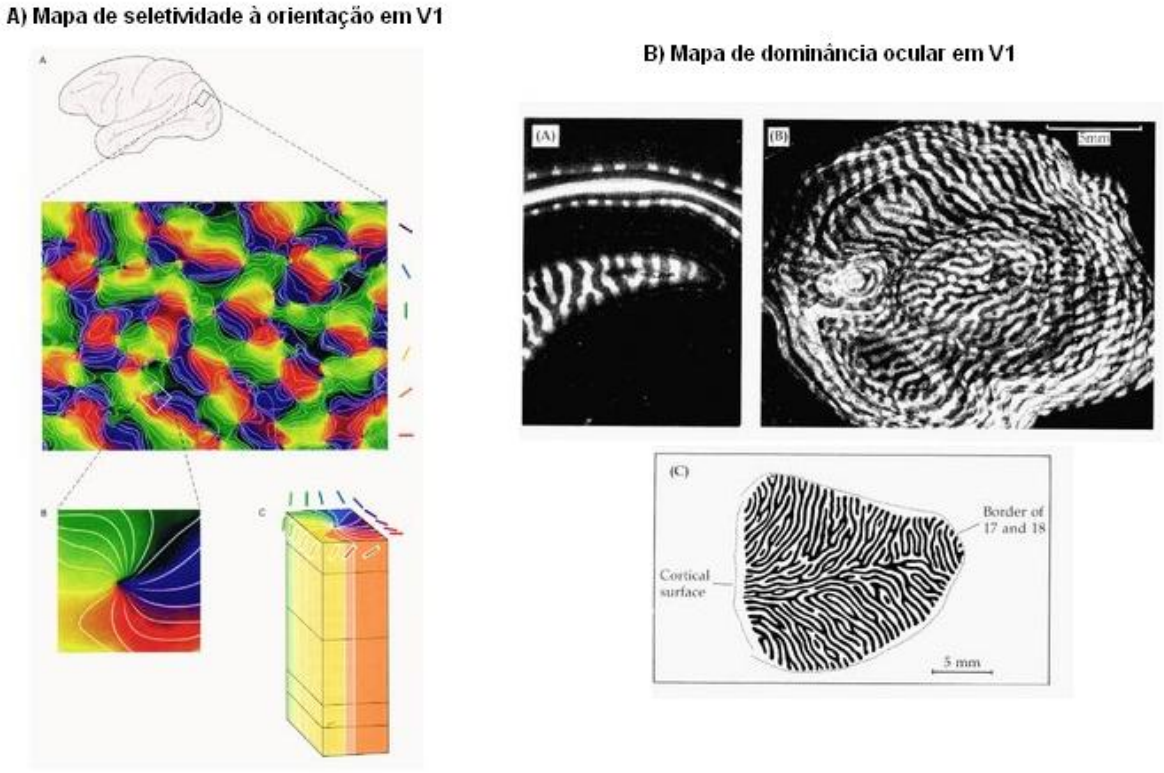

Figura 2-17 – (A) Seletividade à orientação e a (B) dominância ocular, variam ao longo da superfície de V1, porém não se alteram numa mesma coluna. Fonte: http://www.brainworks.uni-freiburg.de/group/wachtler/VisualSystem/. Acesso em: 22 jan. 2005.

Um grupo de colunas que respondem à linhas (estímulos) com todas as orientações numa região particular do campo visual foi denominado de hipercoluna por Hubel e Wiesel. Essas hipercolunas aparecem repetidas regularmente e precisamente sobre a superfície de V1, ocupando cada uma cerca de 1mm2. Essa organização sugere uma modularização do córtex cerebral, na qual cada módulo de uma área do córtex processaria todas as variantes locais da informação visual tratada naquela área cortical. Assim, se uma determinada área processa orientações do estímulo visual, uma hipercoluna dela codifica todas as orientações da região do campo visual monitorada pela hipercoluna. O mesmo ocorrendo para áreas corticais responsáveis por processar. profundidade, movimento, etc.

2.3.2. V2

A área V2 possui uma extensa fronteira com a área V1. Esta área apresenta regiões chamadas de faixas grossas e faixas finas separadas por regiões chamadas de interfaixas (vide Figura 2-18). As faixas finas (em coloração vermelha na Figura 2-18) e interfaixas (em coloração cinza na Figura 2-18) recebem projeções da via parvocelular que vêm das camadas

2 e 3 da área V1 enquanto que as faixas grossas (em coloração amarela na Figura 2-18) recebem projeções da via magnocelular que vêm das camadas 4B, 4Cα e 4Cβ. As faixas finas e as interfaixas se projetam para a área V4, enquanto que as faixas grossas se projetam para área temporal média (MT ou V5). Estes caminhos não são totalmente separados conforme descrito, pois existem conexões entre as faixas finas e grossas e também projeções da área V4 de volta para as faixas finas de V2. Também existem conexões das faixas grossas com a área V3.

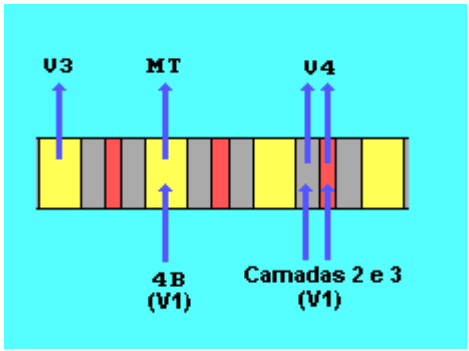

Figura 2-18 - Esquemático de V2. Fonte: [18]

As células de V2, assim como as células de V1, são sensíveis à orientação, cor e à profundidade dos estímulos, ou seja, estas células continuam a análise iniciada em V1. A resposta das células de V2 para contornos reais e ilusórios foram testados juntamente com as células de V1 em alguns experimentos. Um exemplo de percepção de contornos ilusórios pode ser visto na Figura 2-19. No desenho da esquerda, existe realmente um quadrado desenhado. No desenho do centro, apesar de não existir um quadrado desenhado, é possível facilmente "enxergar" um contorno ilusório de um quadrado, enquanto que no desenho da direita, apesar dos objetos serem os mesmos que no desenho do centro, não é possível "enxergar" o mesmo quadrado.

Muitas células de V2 responderam aos contornos ilusórios exatamente como responderam às bordas, enquanto que poucas células de V1 responderam aos mesmos contornos ilusórios da mesma forma como responderam às bordas [13]. Essas observações sugerem que na área V2 é feito um processamento de contornos num nível acima do processamento que ocorre em V1, constituindo assim uma evidência da análise progressiva que ocorre no córtex visual.

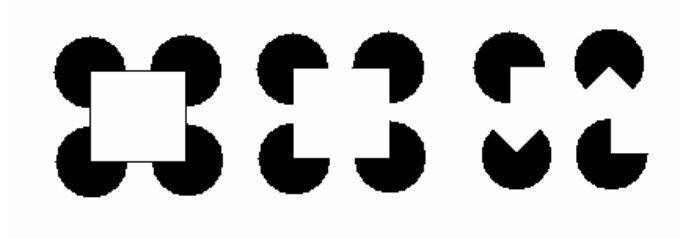

Figura 2-19 - Contorno ilusório. É possível "visualizar" um quadrado branco na figura do meio, apesar de não existir um quadrado desenhado explicitamente.

2.3.3. V3

Pouco se sabe sobre as propriedades funcionais dos neurônios da área extra-estriada V3. Esta área recebe informações das faixas grossas da área V2 e da camada 4B da área V1, e faz projeções tanto para a área temporal média (MT ou V5) quanto para a área V4. Grande parte das células de V3 são seletivas em relação à orientação e direção do estímulo visual, sendo que algumas células são seletivas a cores, o que sugere que em V3 ocorre uma interação entre o processamento de cor e movimento [10].

2.3.4. V4

A área extraestriada V4 foi estudada em profundidade inicialmente por Semir Zeki [29]. Esta área recebe projeções principalmente das faixas finas e interfaixas de V2, provenientes do caminho parvocelular que vêm do LGN e retina, mas também recebe projeções das áreas V1 e V3. Inicialmente pensava-se que as células de V4 fossem exclusivamente dedicadas ao processamento de cores, porém estudos posteriores mostraram que as células de V4 são sensíveis a combinações de cores e formas. As células de V4 se projetam principalmente para o córtex temporal inferior, onde é feito o reconhecimento de faces e de outras formas complexas.

2.3.5. V5 ou MT

A área V5, também conhecida como área temporal média ou MT, recebe projeções da camada 4B do córtex visual primário (V1) e das faixas grossas de V2. Estas conexões são provenientes do caminho magnocelular que parte das células M da retina, passa pelas camadas magnocelulares do LGN e chega no córtex MT. Assim como a área V1, MT possui

um mapa retinotópico do campo visual contralateral, porém os campos receptivos das células de MT são bem maiores que os campos receptivos das células de V1.

O processamento de movimento começa de forma rudimentar em V1 atingindo formas bem mais abstratas em MT numa abstração sucessiva, ou seja em etapas. Anthony Movshon *et al*. [20] testou a hipótese de que o movimento é processado em 2 etapas, registrando a resposta de células em V1 e MT para um padrão de linhas cruzadas em forma de xadrez em movimento. As células de V1 responderam ao movimento dos elementos isolados do padrão, que são as linhas, enquanto que as células em MT responderam ao movimento do padrão em forma de xadrez por completo.

A percepção de profundidade também é processada em MT. Embora células sensíveis à disparidade binocular sejam encontradas em várias áreas corticais, como V1, V2 e V3, as células de MT respondem melhor a estímulos em distâncias específicas do plano de fixação (mais próximo, ou mais distante do ponto de fixação). Esse processamento da disparidade binocular pode ser utilizado tanto para a percepção de profundidade quanto para o controle do movimento de vergência dos olhos.

A Figura 2-20 apresenta um modelo esquemático da arquitetura funcional de MT mostrando que esta área processa tanto informações de movimento (direção) quanto de profundidade. As setas indicam a direção preferencial dos neurônios numa coluna. Estas direções preferenciais variam suavemente através da superfície de MT. A percepção de disparidade é representada pela faixa colorida que varia de *near* (estímulo afastado do ponto de fixação, mas perto do observador), em verde, até *far* (estímulo afastado do ponto de fixação, mas longe do observador), em vermelho, passando pela disparidade zero (estímulo à mesma distância do observador do que o ponto de fixação), codificada em amarelo. As regiões do modelo que estão em azul são regiões da área MT que aparentemente têm pouca seletividade para disparidade.

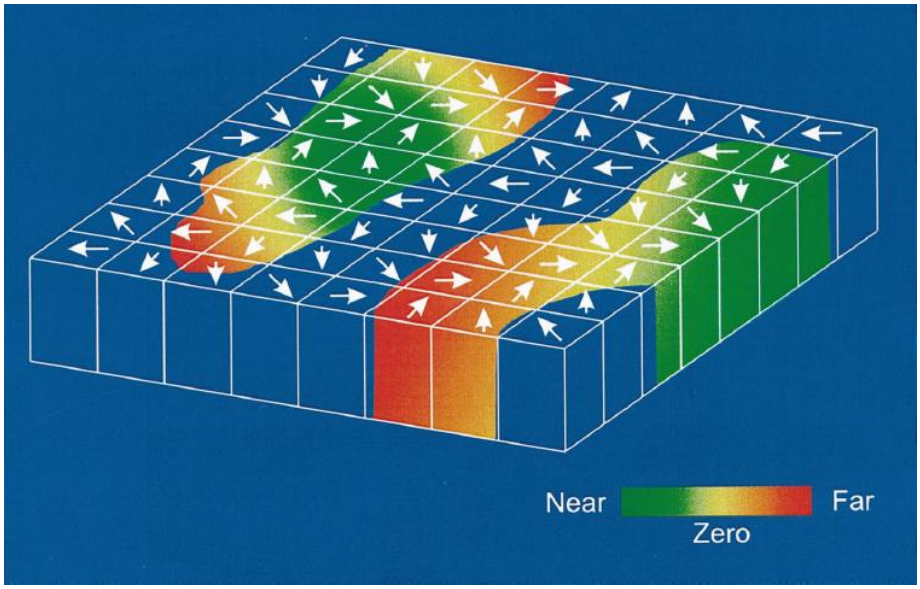

Figura 2-20 - Modelo esquemático da arquitetura funcional de MT. Fonte: [7].

2.4. VIAS PARALELAS

As informações visuais vindas da retina são conduzidas através de vias paralelas que se iniciam na retina, passam pelo LGN, chegam em V1, e depois continuam até os córtices parietal posterior (via dorsal) e temporal inferior (via ventral), como mostra a Figura 2-21. As células P da retina se projetam para as camadas parvocelulares (camadas 3, 4, 5 e 6) do LGN e seguem para o córtex visual primário, recebendo o nome de via P ou via parvocelular. A partir de V1, esta via se projeta para as faixas finas e interfaixas de V2, que depois seguem para a área V4, formando assim a via ventral que alcança o córtex temporal inferior (Figura 2-21). Os neurônios que fazem parte da via ventral são mais sensíveis em relação ao contorno das imagens, sua orientação e bordas. Um outro aspecto importante que é processado nesta via é a percepção de cores. Estas células possuem alta resolução espacial, baixa resolução temporal e alta sensibilidade a cores e bordas, o que proporciona a este sistema a capacidade de analisar “o quê” é visto. Lesões no lobo temporal inferior causam deficiências relacionadas ao reconhecimento de objetos complexos, inclusive o reconhecimento de faces.

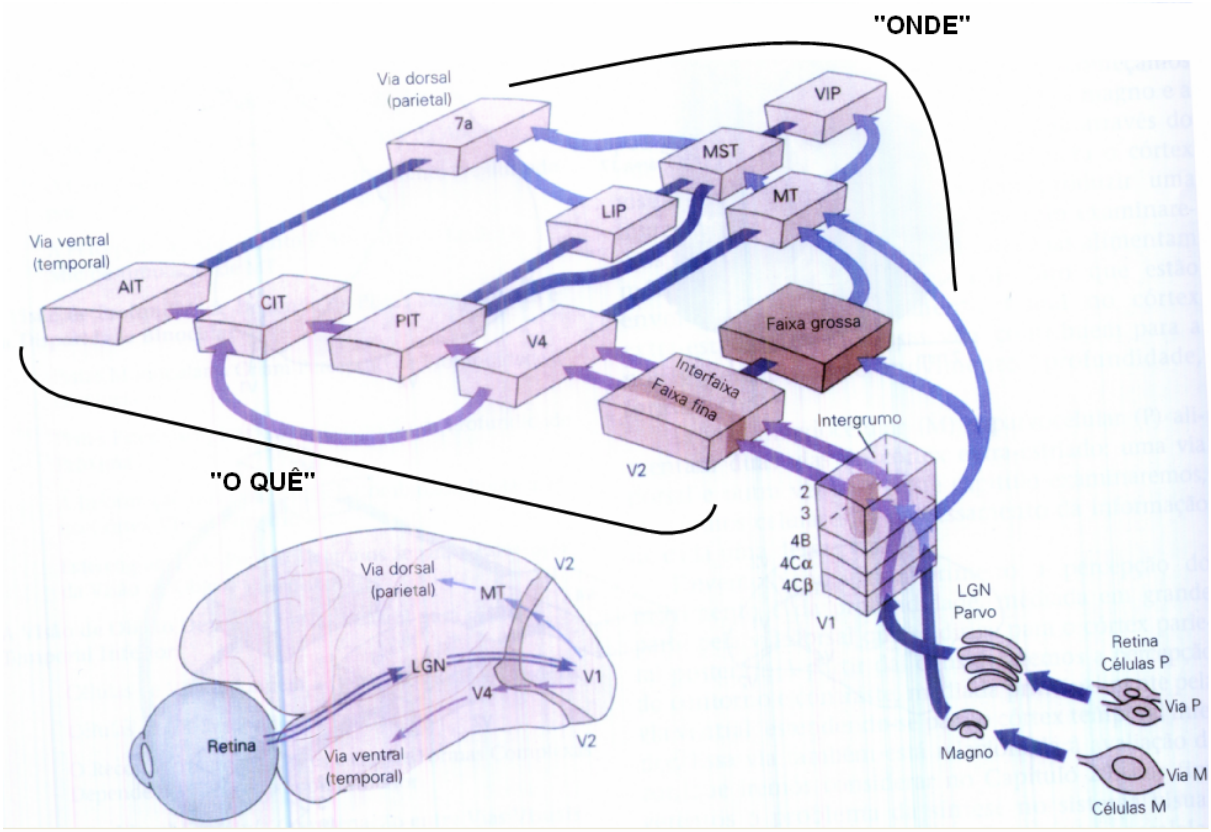

Figura 2-21 - As vias paralelas M e P projetam-se para o córtex visual passando pelo LGN. Fonte: [16].

As células M da retina se projetam para as camadas magnocelulares (camadas 1 e 2) do LGN e também seguem para o córtex visual primário, recebendo o nome de via M ou via magnocelular. A via M se estende de V1 até as faixas grossas de V2, que depois se projetam para a área temporal média (MT) formando a via dorsal que se estende até o córtex parietal posterior (Figura 2-21). Conforme visto anteriormente, o MT (também chamado de V5) está relacionado ao processamento do movimento e profundidade. Os neurônios que formam este sistema são poucos sensíveis a cor e objetos parados, diferentemente dos neurônios associados à via ventral, mas possuem alta resolução temporal e sensibilidade a disparidade binocular, o que faz com que este sistema tenha capacidade de analisar “onde” estão os objetos vistos. Lesões na via dorsal causam deficiência na percepção de movimentos e nos movimentos dos olhos dirigidos a alvos em movimento (movimento de perseguição suave).

A via dorsal, responsável pela análise de “o quê” é visualizado, continua até terminar numa região do córtex pré-frontal especializada na memória de trabalho visual espacial enquanto que a via ventral, responsável pela análise de “onde” estão os objetos visualizados, continua até terminar numa outra região também do córtex pré-frontal especializada na memória de trabalho de cognição. Esta análise mostra que o sistema visual está organizado

em vias paralelas bem definidas, com uma organização seqüencial e hierárquica em cada uma delas

2.5. SISTEMA ÓCULO-MOTOR

O ângulo total de visão humana possui arco de cerca de 200 graus. A melhor definição fica na fóvea, que tem pouco menos de 1 mm de diâmetro e representa cerca de 2 graus de arco no centro do campo de visão (aproximadamente o tamanho da unha do polegar à distância do braço estendido). Cerca de metade de V1 é devotada inteiramente às fóveas e esta concentração de recursos neurais, que vem da retina e persiste nas outras áreas corticais visuais além de V1, resulta em uma percepção visual muito melhor das imagens que estão sobre as fóveas. Por essa razão, a visão um sistema bastante elaborado para a movimentação rápida e precisa dos olhos.

Os movimentos dos olhos se dão em 3 eixos de rotação (Figura 2-22): vertical (eixo X, movimento para baixo e para cima - *depression* e *elevation*), horizontal (eixo Y, movimento de lado para outro, *abduction* e *adduction*) e torsional (eixo Z, *extorsions* e *intorsions*)

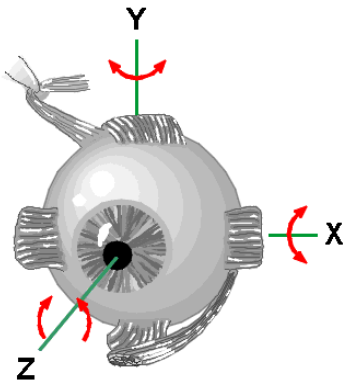

Figura 2-22 - Movimentos oculares. Fonte: http://www.auto.ucl.ac.be/EYELAB/Welcome.html. Acesso em: 25 out. 2004.

Cada um dos olhos possui 3 pares de músculos extra-oculares, que operam antagonicamente (Figura 2-23): *Medial Rectus (adduction)* e *Lateral Rectus (abduction)*; *Superior Rectus (elevation)* e *Inferior Rectus (depression)*; *Superior Oblique (extorsion)* e *Inferior Oblique (introsion)* [16].

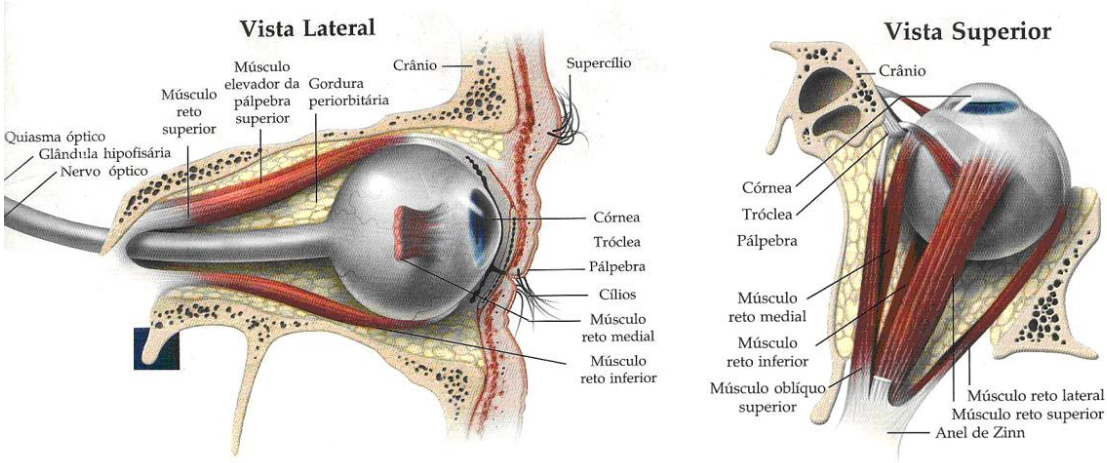

Figura 2-23 - Músculos oculares. A) Vista Lateral; B) Vista Superior. Fonte: http://www20.brinkster.com/tonho/olho/olhohumano.html. Acesso em: 15 maio 2004.

Olhar de forma exploratória em busca de um ponto de interesse requer mover os olhos rapidamente de modo que a imagem dos objetos seja projetada sobre nossas fóveas. Uma vez localizado o ponto de interesse, contudo, é preciso estabilizar sua imagem na retina, mesmo que a cabeça se movimente. O sistema óculo-motor tem, então, duas grandes funções:

1. Posicionar a imagem do ponto de interesse – o alvo – na parte da retina com maior acuidade, a fóvea;
2. Manter a imagem estacionária na fóvea, independente de movimentos do alvo ou da cabeça.

Por volta de 1902, Raymond Dodge descreveu 5 sistemas separados de controle da posição dos olhos [9]. Estes 5 sistemas podem ser divididos em dois grupos, segundo as duas grandes funções descritas do sistema óculo-motor, como mostrado na Tabela 2-1.

Os primeiros 4 movimentos são conjugados, cada olho se move na mesma direção e na mesma quantidade. O último é desconjugado: os olhos se movem em direções diferentes e, muitas vezes, de diferentes quantidades.

Movimento	Função
Movimentos que estabilizam o olho quando a cabeça se move	
Vestíbulo-ocular	Mantém as imagens estáveis na retina durante rápidas rotações da cabeça
Optokinético	Mantém as imagens estáveis na retina durante rotação lenta e contínua da cabeça
Movimentos que mantém a fóvea no alvo	
Sacada	Trás novos pontos de interesse para a fóvea
Perseguição suave	Mantém a imagem de um alvo em movimento na fóvea
Vergência	Ajusta os olhos para que o mesmo ponto seja levado a ambas as fóveas

Tabela 2-1 - Movimentos Oculares

Existem outros tipos de movimentos que têm com principal característica amplitudes muito pequenas. Estes movimentos são involuntários e ocorrem quando se está observando um objeto fixo. Estes movimentos são chamados de movimentos de *sustaining* e possuem como principal função manter o foco sobre o objeto que está sendo observado, e produzir variações constantes, mesmo que pequenas, da imagem na retina. Sem estas variações a imagem "desapareceria" devido à acomodação dos neurônios do sistema visual.

2.6. VISÃO BINOCULAR

Uma das principais funções do sistema visual é transformar as imagens bidimensionais projetadas na retina numa imagem tridimensional. Estudos indicam que esta transformação é baseada tanto em pistas monoculares para a profundidade como o tamanho familiar dos objetos (sabendo previamente o tamanho aproximado de um objeto, é possível julgar a distância deste objeto), oclusão (caso um objeto A esconda parcialmente um objeto B, assume-se que o objeto A está mais próximo que o objeto B), entre outras pistas, quanto em pistas estereoscópicas oriundas da disparidade binocular causada pela leve diferença das imagens projetadas nas retinas.

Quando os objetos observados estão à distâncias maiores do que cerca de 30 metros, as imagens projetadas nas retinas são praticamente idênticas, eliminando quase que totalmente a disparidade binocular, fazendo com que a percepção de profundidade seja baseada principalmente nas pistas monoculares. Entretanto, a percepção de profundidade a distâncias substancialmente menores do que 30 metros é mediada preferencialmente pela *stereopsis* (originário do grego '*stereo*', que significa sólido, referenciando a imagem em 3D), que é a percepção de profundidade visual baseada na disparidade binocular. A visão estereoscópica é possível porque os dois olhos estão separados horizontalmente, a uma distância média de

65mm em adultos, sendo que cada olho observa o mundo através de uma posição ligeiramente diferente. Desta forma, objetos a distâncias diferentes produzem imagens com afastamento relativo (disparidade) diferente nas retinas como mostra a Figura 2-24.

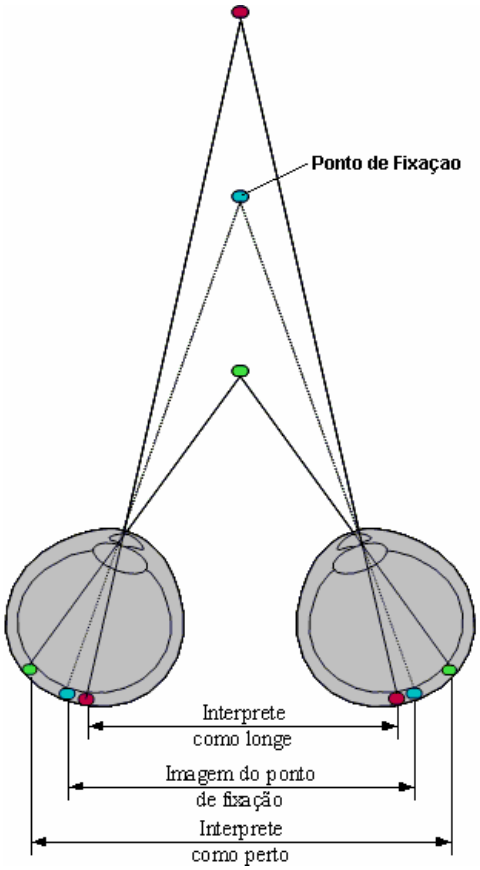

Figura 2-24 - Visão Binocular. Fonte: [16].

Ao olhar um ponto no espaço, a imagem deste ponto é projetada em pontos correspondentes na retina de cada olho, mais precisamente na fóvea de cada olho. Este ponto no espaço é chamado de ponto de fixação e o plano vertical de pontos no espaço onde se localiza o ponto de fixação é chamado de plano de fixação. Pontos no espaço situados antes do plano de fixação, ou seja, pontos que estão mais próximos do que o ponto de fixação, são projetados em pontos na retina que possuem disparidade positiva (um ponto no olho esquerdo tem seu correspondente no olho direito mais a direita com relação à fóvea), enquanto que os pontos mais distantes do que o ponto de fixação possuem uma disparidade negativa (um ponto no olho esquerdo tem seu correspondente no olho direito mais a esquerda). Pontos que estão a mesma distância, possuem disparidade zero (um ponto no olho esquerdo tem seu correspondente no olho direito na mesma posição com relação à fóvea).

3. MODELAGEM DO SISTEMA VISUAL HUMANO

A construção de uma imagem tridimensional interna ao computador do mundo externo a partir de duas imagens bidimensionais, que busque similaridade com o sistema biológico, requer uma modelagem das transformações na informação visual que ocorrem durante o caminho que ela faz, saindo da retina, passando pelo LGN até o córtex, e do processamento ocorrido no próprio córtex, em especial nas partes do córtex visual que estão envolvidas no processamento e percepção de profundidade, que são as áreas V1 e MT.

Esta modelagem foi feita iniciando com uma transformação nas imagens direita e esquerda (imagens projetadas nas retinas direita e esquerda respectivamente), que tem como objetivo de reduzir a quantidade de informação a ser processada e fornecer mais atenção à informação visual situada no ponto focal, semelhante à transformação ocorrida na imagem ao chegar no LGN, e depois propagada para V1. Após esta transformação é feita uma extração de características das imagens através de filtros que modelam o processamento feito pelas células de V1 e MT. Esta extração é feita numa arquitetura com níveis crescentes de abstração. Combinando as informações obtidas através destas características de cada imagem, são obtidas as informações necessárias para a reconstrução 3D. Foi implementada também uma inversa desta modelagem para permitir a visualização da representação tridimensional interna ao computador.

Esta seção está divida em três partes nas quais serão explicadas detalhadamente como foi modelada cada parte do sistema visual humano, na mesma seqüência no qual é feito o processamento da disparidade binocular. Na primeira parte será abordada a transformação na imagem ocorrida entre a retina e o córtex V1, passando pelo LGN. Na segunda parte será aborda a modelagem das células do córtex V1 que realizam o processamento de percepção de profundidade através da disparidade binocular, e como se interagem. Na terceira parte será apresentada a modelagem da área MT, mostrando como esta área recebe as informações de V1 e as organiza de forma a permitir a extração das informações necessárias para a reconstrução tridimensional.

3.1. MODELAGEM RETINA-V1 (MAPEAMENTO LOG-POLAR)

Num dado momento, apenas uma fração da informação disponível da cena visual que recai nas duas retinas pode ser processada. Esta fração é representada principalmente pelas fóveas que, por serem as regiões das retinas que possuem maior quantidade de fotorreceptores, possuem uma maior representação no córtex visual primário.

Através de estudos feitos em primatas, verificou-se como é feito este mapeamento da retina em V1. A Figura 3-1 mostra a representação de uma imagem contendo círculos concêntricos no córtex V1. Esta representação foi obtida por Tootell [27] através da aplicação de 2-deoxyglucose radiativa em um olho e do controle da fixação da fóvea deste olho no centro da imagem. A imagem não era estática; os pontos brancos que a compõem se moviam ao longo do círculo para garantir que as respostas das células do sistema visual não se acomodassem. Após 45 minutos, o macaco foi sacrificado e a parte correspondente a V1 do hemisfério esquerdo do cérebro do macaco foi recortada, aplainada e posicionada sobre um filme fotográfico sensível à radiação que, quando revelado, mostrou a imagem representada na Figura 3-1.

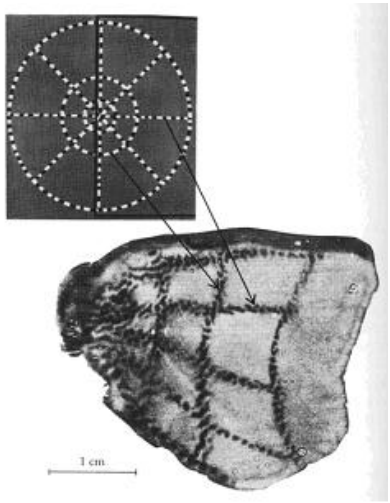

Figura 3-1 - Representação da imagem projetada na retina em V1. Fonte: [27]

A 2-deoxyglucose radiativa é confundida pelas células do sistema visual como sendo glicose, e é passada pelas sinapses das células mais ativadas de um estágio para outro do sistema visual, segundo o mapeamento da imagem na retina e seus correspondentes em cada estágio. Este mapeamento, que ocorre na passagem das informações visuais no caminho entre a retina e V1, pode ser modelado matematicamente utilizando uma transformação log-polar da

imagem projetada na retina para o córtex visual primário. Esta transformação realiza o mapeamento dos pontos do plano cartesiano (coordenadas x e y) para pontos no plano log-polar (coordenadas η e ξ) de acordo como mostrado na Figura 3-2.

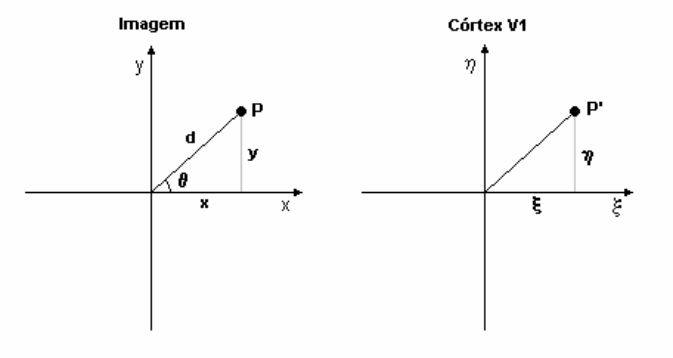

Figura 3-2 - Transformação log-polar

O mapeamento log-polar é descrito como uma transformação **F:P→P'** na qual:

$$\xi \propto \log(d) = \log\left(\sqrt{x^2 + y^2}\right) \tag{1}$$

$$\eta \propto \theta = \arctan\left(\frac{y}{x}\right) \tag{2}$$

Uma característica importante desta transformação é que, devido ao comportamento logarítmico da variável ξ, a qual é diretamente proporcional ao logaritmo da distância *d* da origem ao ponto **P,** as regiões mais próximas ao centro das coordenadas cartesianas possuem uma maior representação, e portanto uma melhor definição, no plano log-polar, enquanto que as regiões mais afastadas do centro possuem uma menor representação, e portanto uma pior definição no plano log-polar, semelhante ao que ocorre no sistema visual humano. Este comportamento pode ser visto na Figura 3-3. A Figura 3-3-a mostra o plano x-y repartido em fatias e estas recortadas por círculos concêntricos cujos afastamentos (entre os círculos) segue uma função logarítmica da distância ao centro. A Figura 3-3-b mostra o plano η–ξ, no qual cada retângulo representa um dos segmentos de fatia da Figura 3-3a, e deixa claro, nesta representação, como as regiões no centro do plano x-y são mais bem representadas no plano η–ξ.

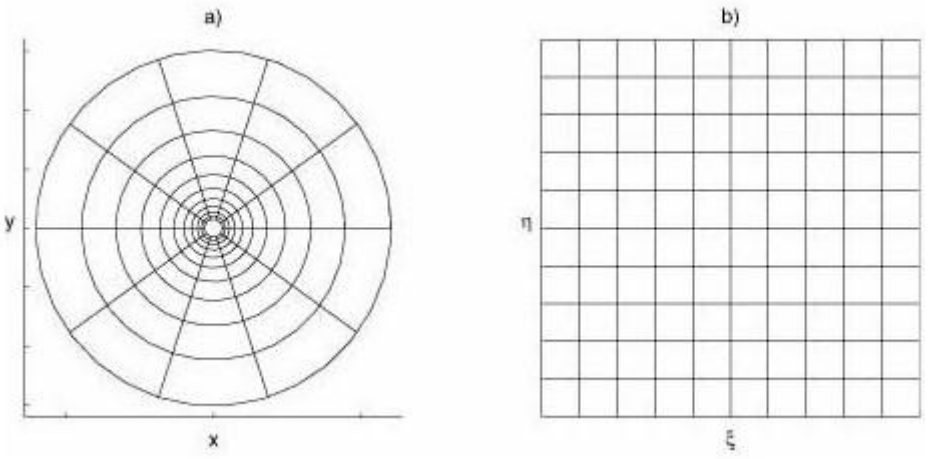

Figura 3-3 - Representação do comportamento logarítmico da transformação log-polar. (a) plano x-y, (b) plano η–ξ, Fonte: http://omni.isr.ist.utl.pt/~alex/Projects/TemplateTracking/logpolar.htm. Acesso em: 22 dez. 2004.

Na Figura 3-4 é apresentado um exemplo da aplicação da transformação log-polar numa imagem. Uma imagem (Figura 3-4a) de 128x128 pixels sofre uma transformação log-polar com uma amostragem de 64x32 (η e ξ respectivamente) representada na Figura 3-4b. Realizando a transformação inversa (Figura 3-4c) é possível verificar facilmente a perda de definição a medida em que se afasta do centro.

Figura 3-4 - Exemplo da aplicação da transformação log-polar. Fonte: http://omni.isr.ist.utl.pt/~alex/Projects/TemplateTracking/logpolar.htm. Acesso em: 22 dez. 2004.

Neste trabalho foi feita uma modificação na forma de visualização da imagem no plano log-polar com a finalidade de preservar a relação de vizinhança entre os pontos da imagem no centro da transformação log-polar de acordo com as vizinhanças observadas no córtex relativas às imagens projetadas na fóvea. A Figura 3-5 mostra um exemplo da transformação log-polar utilizada neste trabalho. A cruz vermelha na Figura 3-5a representa o centro da transformação. Na Figura 3-5b, a metade esquerda representa a metade esquerda da

imagem original, conforme seria projetada na retina e propagada para V1 se ponto de atenção fosse a cruz vermelha, e a metade direita representa a metade direita da imagem original na mesma situação.

Figura 3-5 – Exemplo da transformação log-polar utilizada neste trabalho.

3.2. MODELAGEM DA ÁREA V1

Conforme descrito na Seção 2.3.1, quase toda informação visual vinda dos olhos em direção ao córtex visual passa inicialmente pelo córtex visual primário (V1), fazendo com que sua modelagem seja necessária para praticamente qualquer sistema que queira reproduzir o comportamento da visão humana. Neste trabalho, além de ter sido modelado o mapeamento retina-V1, foram modeladas também as células simples e células complexas de V1 (Seção 2.3.1). Para esta modelagem foram utilizadas funções Gabor [19], conforme descrito a seguir.

3.2.1. Células Simples

A informação visual proveniente das células ganglionares da retina passa primeiramente pelo LGN antes de chegar a V1. Antes de chegar em V1 a informação vinda dos olhos ainda está segregada; entretanto, a partir de V1 surgem células que recebem informação de ambos os olhos e possuem seletividade para a disparidade binocular [6],[23]. As projeções do LGN chegam em V1 principalmente nas células simples que, conforme visto na Seção 2.3.1, são seletivas à orientação do estímulo visual e possuem campo receptivo

alongado com regiões excitatórias e inibitórias, sendo também seletivas a uma determinada freqüência espacial do estímulo.

Hubel e Wielsel [14] sugeriram um padrão de interconexão entre o LGN e V1 que mostra como o campo receptivo das células simples binoculares no córtex visual primário pode ser originado a partir do das células do LGN. Através deste padrão de interconexão, mostrado na Figura 3-6, os campos receptivos de várias células do LGN são sobrepostos para formar o campo receptivo das células simples de V1, mostrado na Figura 2-15. Esta sobreposição pode ser tanto de células *on center* quanto *off center*, sendo que na Figura 3-6 é apresentado um esquema para células *on center*.

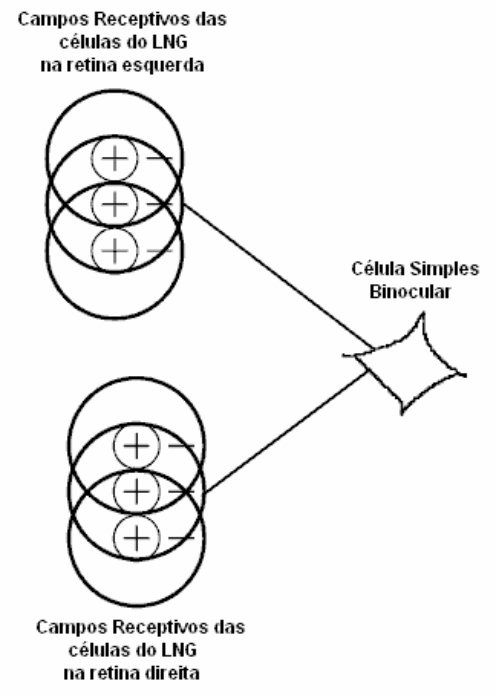

Figura 3-6 - Modelo de Hubel e Wiesel para o campo receptivo das células simples binoculares. Fonte: [25].

Jones e Palmer [15] analisaram o perfil do campo receptivo de células simples no córtex estriado em gatos e o descreveram como sendo uma função de Gabor que é o produto de uma cossenóide complexa denominada **portadora**, por uma função gaussiana bidimensional denominada **envoltória** [19]. Jones e Palmer mostraram que a parte real da função de Gabor complexa se ajusta muito bem com os campos receptivos encontrados nas células simples no córtex estriado em gatos. Estudos posteriores mostraram que, semelhante ao campo receptivo das células simples do córtex visual de gatos, o perfil do campo receptivo das células simples dos macacos também podem ser descrito por uma função de Gabor.

A fórmula de uma função de Gabor complexa bidimensional *g(x,y)*, no domínio do espaço é:

$$g(x,y) = s(x,y) * w(x,y) \tag{3}$$

onde *x e y* são as coordenadas espaciais de um ponto na retina , *s(x,y)* é a cossenóide complexa (portadora) e *w(x,y)* é a função gaussiana bidimensional (envoltória). Considerando somente a parte real da função de Gabor, a portadora é:

$$s(x,y) = \cos\left(2\pi\left(u_0 x + v_0 y\right) + \phi\right) \tag{4}$$

onde o parâmetro Φ define a fase da portadora enquanto que os parâmetros u_0 e v_0 definem a freqüência espacial da portadora em coordenadas cartesianas. Esta freqüência espacial também pode ser expressa em coordenadas polares com uma magnitude F_0 e uma orientação θ_0:

$$\begin{aligned} F_0 &= \sqrt{{u_0}^2 + {v_0}^2} \\ \theta_0 &= \arctan\left(\frac{v_0}{u_0}\right) \end{aligned} \tag{5}$$

ou seja,

$$\begin{aligned} u_0 &= F_0 \cos(\theta_0) \\ v_0 &= F_0 \sin(\theta_0) \end{aligned} \tag{6}$$

A envoltória da função de Gabor é modelada da seguinte forma:

$$w(x,y) = K \exp\left(-\pi\left(a^2 {(x - x_0)_r}^2 + b^2 {(y - y_0)_r}^2\right)\right) \tag{7}$$

no qual o parâmetro *K* é um fator de escala para a amplitude da envoltória, x_0 e y_0 representam a coordenada do pico da gaussiana e, a e b são parâmetros de escala da envoltória que controlam a largura de banda ω, em oitavas, da portadora:

$$a = F_0 \omega$$
$$b = \frac{a}{aspect\_ratio} \tag{8}$$

no qual *aspect_ratio* é a relação entre a largura e a altura do campo receptivo. A Figura 3-7 mostra o gráfico de uma gaussiana bidimensional simétrica com *aspect_ratio* igual a 1.

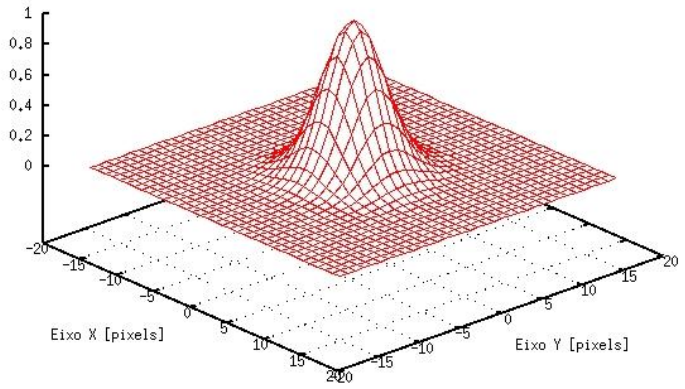

Figura 3-7 – Ilustração de uma gaussiana bidimensional simétrica, ou seja, com *aspect_ratio* = 1.

O subscrito r na equação (7) indica a possibilidade de uma operação de rotação bidimensional do campo receptivo da célula simples. No caso desta rotação, os termos entre parêntesis com subscrito r na equação (7) devem ser substituídos por:

$$\begin{cases} (x - x_0)_r = (x - x_0)\cos(\theta_0) + (y - y_0)\sin(\theta_0) \\ (y - y_0)_r = -(x - x_0)\sin(\theta_0) + (y - y_0)\sin(\theta_0) \end{cases} \tag{9}$$

Dessa forma, combinando as equações (3), (4) e (7) a parte real da função de Gabor fica:

$$g(x, y) = K \exp\left(-\pi\left(a^2 (x - x_0)_r^{\,2} + b^2 (y - y_0)_r^{\,2}\right)\right) * \cos\left(2\pi\left(u_0 x + v_0 y\right) + \phi\right) \tag{10}$$

Na verdade, a equação (10) possui uma componente DC indesejável [19], uma vez que a célula não deve responder para estímulos constantes, independente de sua intensidade. Para resolver este problema, é incluída uma compensação da componente DC [19], na portadora descrita anteriormente na equação (4):

$$s(x,y) = \cos\left(2\pi\left(u_0 x + v_0 y\right) + \phi\right) - \exp\left(-\pi\left(\frac{{u_0}^2}{{a_0}^2} + \frac{{v_0}^2}{{b_0}^2}\right)\right)\cos(\phi) \tag{11}$$

Assim, a forma final da função de Gabor que modela o campo receptivo de uma célula simples é:

$$g(x,y) = K\exp\left(-\pi\left(a^2\left(x - x_0\right)_r^{\,2} + b^2\left(y - y_0\right)_r^{\,2}\right)\right)$$
$$\left[\cos\left(2\pi\left(u_0 x + v_0 y\right) + \phi\right) - \exp\left(-\pi\left(\frac{{u_0}^2}{{a_0}^2} + \frac{{v_0}^2}{{b_0}^2}\right)\right)\cos(\phi)\right] \tag{12}$$

A Figura 3-8 mostra o gráfico de funções de Gabor para diferentes valores de ϕ.

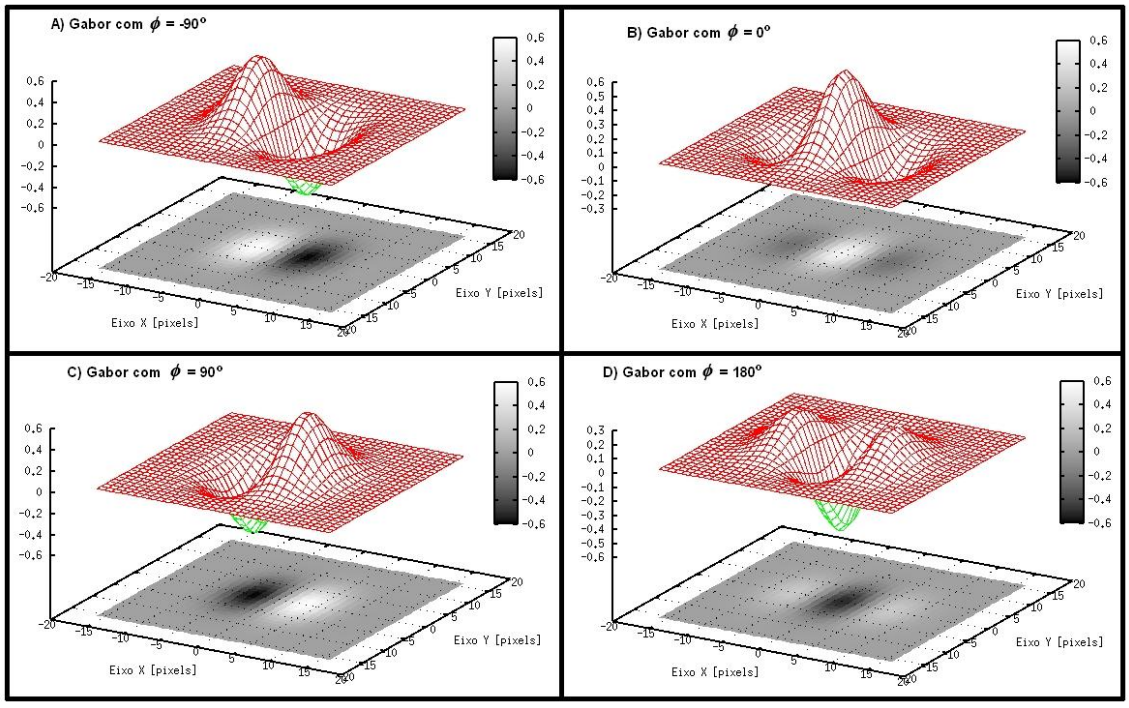

Figura 3-8 - Figura com funções de Gabor para diferentes valores de fase. A) Φ = -90°. B) Φ = 0°. C) Φ = 90°. D) Φ = 180°. Todos os gráficos estão na mesma escala e a área coberta pela função de Gabor é de 33x33 unidades (pixels).

A resposta de uma célula simples, representada por $R_s(x,y)$ na equação (13), é aproximada pela convolução espacial de uma função de Gabor com o estímulo visual, ou seja, a convolução da equação (12) com o estímulo, representado por *Estimulo*(x,y). Realizar esta

convolução é o equivalente a extrair a soma das amplitudes das componentes de freqüência espacial que compõem o estimulo de acordo com o filtro de Gabor bidimensional, de freqüências u_0 e v_0 e largura de banda ω.

$$R_S = g(x,y) * Estimulo(x,y) \tag{13}$$

Considerando o campo receptivo das células simples binoculares, como sendo a composição de dois campos receptivos conforme descrito na equação (12), um para o olho direito e outro para o olho esquerdo, existem basicamente três maneiras de se detectar disparidades binoculares com células simples binoculares: através da diferença da posição dos campos receptivos nos dois olhos, ou da diferença de fase dos campos receptivos nos dois olhos, ou de ambas. No modelo de diferença de posição (Figura 3-9) os campos receptivos em cada olho possuem a mesma forma, porém não estão centrados na mesma posição relativa à fóvea nas duas retinas. No modelo de diferença de fase (Figura 3-9) os campos receptivos estão centrados em posições correspondentes nas duas retinas porém, devido à diferença de fase, possuem perfil diferente. Ambos os modelos podem ser combinados em um modelo híbrido, de posição e fase.

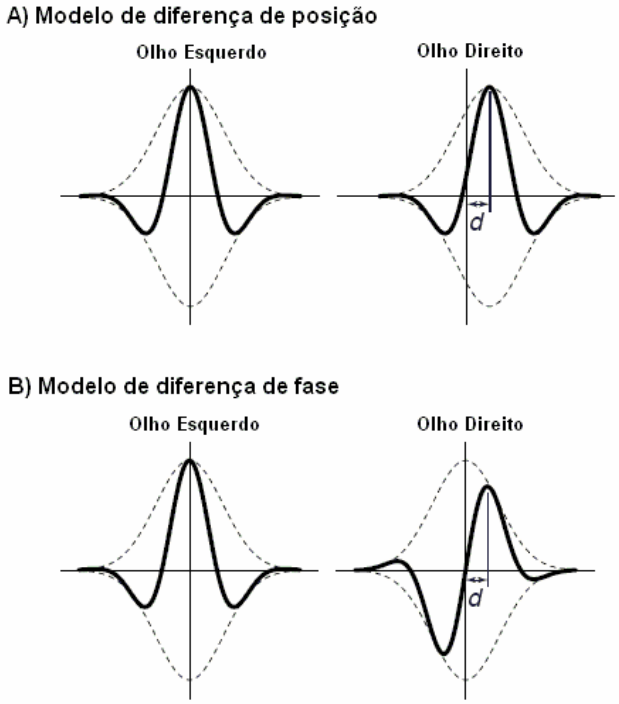

Figura 3-9 - Modelos de codificação de disparidade. A) No modelo de diferença de posição os campos receptivos são idênticos porém centrados em posições não correspondentes nas duas retinas. B) No modelo de diferença de fase os campos receptivos possuem formas diferentes porém estão centrados em posições correspondentes nas duas retinas. Fonte: [8] com alterações.

Existe uma diferença no valor máximo de disparidade *d* que cada modelo pode codificar. No modelo de diferença de fases, a faixa de disparidades binoculares que é possível codificar é inversamente proporcional à freqüência espacial central da célula, já que a reposta da célula segue uma função periódica – a portadora da função Gabor. Assim, quanto maior a freqüência, menor a região coberta por um ciclo da função portadora da Gabor, e menor a disparidade que se consegue detectar. O modelo de diferença de posição não possui esta restrição.

Mapeando simultaneamente o campo receptivo do olho direito e do olho esquerdo em pares de células simples de V1, Anzai *et al.* [4] mostraram que existe tanto disparidade por diferença de posição quanto por diferença de fase dando suporte para o modelo híbrido (fase e posição) de codificação da disparidade binocular.

No modelo proposto foram implementados dois tipos de células simples: células simples monoculares e células simples binoculares. As células simples monoculares foram implementadas utilizando um campo receptivo modelado por uma função de Gabor, de acordo com a equação (12), centrado num ponto da retina de um dos dois olhos (na verdade, câmeras), conforme mostrado na Figura 3-10. A saída desta célula simples monocular é a convolução da imagem centrada neste ponto com o campo receptivo, conforme equação (13).

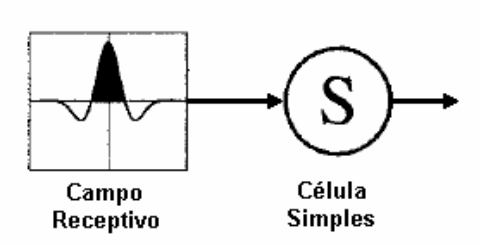

Figura 3-10 - Modelo de célula simples monocular implementada com o campo receptivo ajustado com fase $\Phi = 0$.

As células simples binoculares foram implementadas utilizando dois campos receptivos, cada um recebendo informações de um olho com diferença de fase de 180º entre eles. A resposta desta célula simples binocular é computada somando a contribuição de cada campo receptivo. Ou seja, a saída da célula simples binocular é computada como sendo a soma de duas células simples monoculares, uma de cada olho, com diferença de fase de 180º. O diagrama da Figura 3-11 mostra como as células simples binoculares foram implementadas.

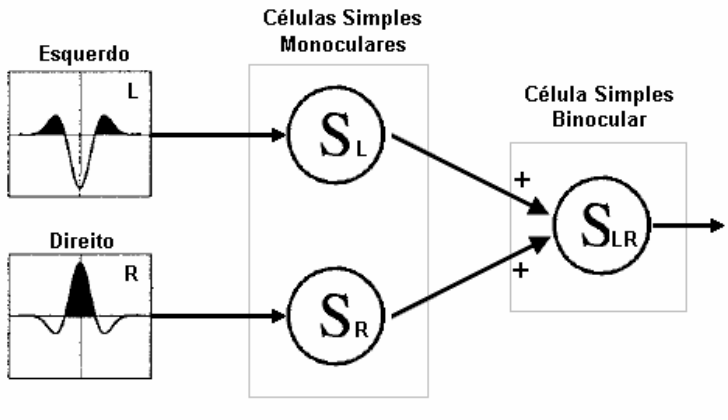

Figura 3-11 - Modelo de célula simples binocular implementada cuja resposta é a soma das respostas de duas células simples monoculares.

Na célula simples binocular implementada, quando um estímulo é projetado sobre seus campos receptivos em ambas as câmeras simultaneamente, a resposta da célula simples binocular é nula, pois a contribuição de um campo receptivo anula a do outro devido à defasagem de 180º na portadora dos dois campos receptivos. Esta situação é mostrada na Figura 3-12. Caso o estímulo não seja projetado em pontos correspondentes, as contribuições dos campos receptivos terão amplitudes diferentes, fazendo com que a célula simples binocular tenha uma resposta não nula.

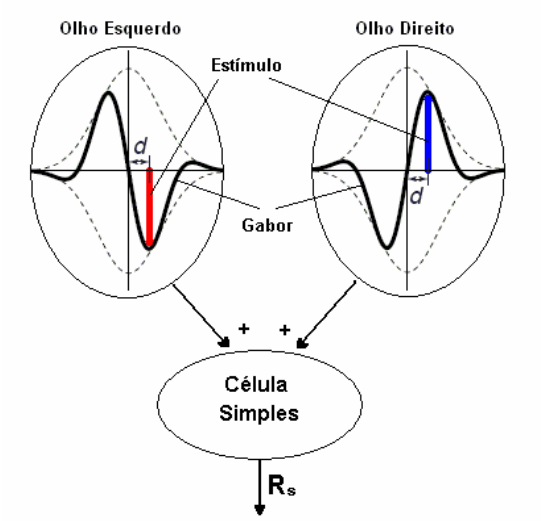

Figura 3-12 - Resposta da célula simples binocular implementada. O mesmo estímulo é projetado em posições correspondentes nos campos receptivos esquerdo e direito. Desta forma, a contribuição de um campo receptivo é anulada pela do outro, fazendo com que a reposta R_S da célula simples seja nula.

O primeiro estágio de codificação da profundidade através da disparidade binocular é feito pelas células simples binoculares. Entretanto, devido à característica de seu campo receptivo, que possui claramente regiões excitatórias alternadas com regiões inibitórias, a

resposta das células simples binoculares é influenciada pela posição do estímulo dentro de seu campo receptivo, ou seja, pela fase do estímulo. As células complexas, descritas a seguir, não são influenciadas pela fase do estímulo.

3.2.2. Células Complexas

Para codificar a profundidade dos objetos no campo visual, o sistema visual humano precisa resolver o problema da correspondência, ou seja, descobrir, para cada ponto da imagem projetada na retina direita, qual é o ponto correspondente na imagem projetada na retina esquerda, e vice-versa. As células complexas computam uma aproximação da solução para o problema da correspondência [5].

Analisando a resposta das células complexas, Hubel e Wiesel [14] observaram que estas células não possuem regiões excitatórias e inibitórias bem definidas e que, de fato, as células complexas respondem a estímulos, independente da sua posição dentro do campo receptivo (vide Figura 2-16). Como uma possível explicação para o campo receptivo das células complexas, Hubel e Wiesel propuseram um modelo hierárquico no qual o campo receptivo de uma célula complexa seria formado por uma composição dos campos receptivos de células simples [14], conforme mostrado na Figura 3-13.

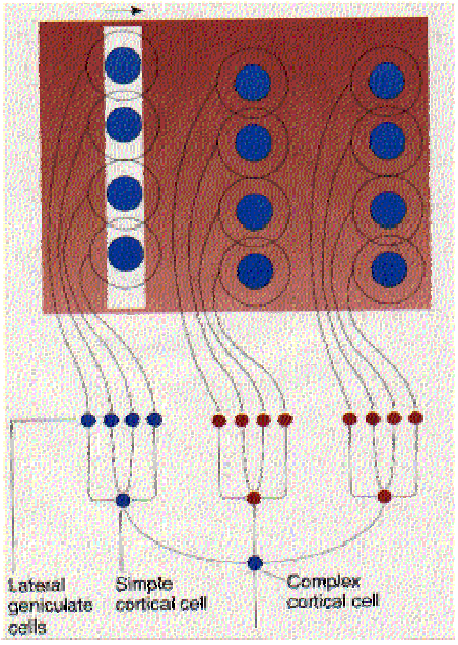

Figura 3-13 – Modelo Hubel e Wiesel de células complexas. O campo receptivo é composto de campos receptivos de células simples que por sua vez são compostos por campos receptivos de células do LGN. Fonte: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/V/VisualProcessing.html. Acesso em: 28 jun. 2004.

Vários estudos demonstraram que as respostas de células complexas não poderiam ser resultado de uma combinação linear das respostas de células simples [5]. Contudo, as células complexas aparentam prover um nível seguinte de abstração para o processamento da stereopsis baseado na atividade de células simples, mesmo que não uma combinação linear da resposta destas.

Vários modelos de células complexas têm sido propostos utilizando uma combinação entre subunidades de campos receptivos de células simples. Uma modelagem bastante popular de células complexas é a denominada de **modelo de energia**, que é construída com dois filtros passa banda em quadratura de fase (fases com diferença de 90º) cujas saídas são elevadas ao quadrado e então somadas [1]. Ohzawa *et al.* [22] propôs um modelo de energia de segunda ordem para células complexas binoculares, o qual provê uma boa aproximação do comportamento exibido por células complexas do córtex estriado de gatos. Este modelo computa a soma do quadrado da saída de um par de células simples em quadratura. Cada célula simples soma a contribuição de seus dois campos receptivos, um na retina direita e outro na retina esquerda (Figura 3-14).

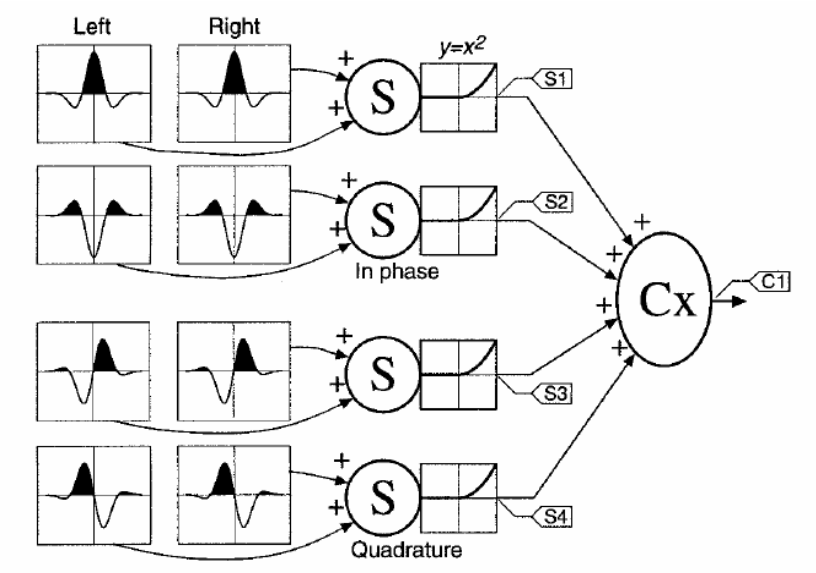

Figura 3-14 - Modelo de energia para célula complexa proposto por Ohzawa para detecção de disparidade binocular. Fonte: [22].

Neste modelo foram propostos dois tipos de células complexas: células complexas monoculares e células complexas binoculares. Ambos os tipos recebem informações de duas células simples e computam a soma do quadrado da saída destas células. As células complexas monoculares são modeladas recebendo projeções de duas células simples também

monoculares, com campos receptivos centrados num mesmo ponto da retina de um mesmo olho, porém em quadratura de fase, ou seja, com diferença de fase $\Phi = 90º$. A saída de cada célula simples é elevada ao quadrado e somadas, formando a resposta da célula complexa monocular. A Figura 3-15 mostra de forma esquemática a arquitetura da célula complexa monocular implementada.

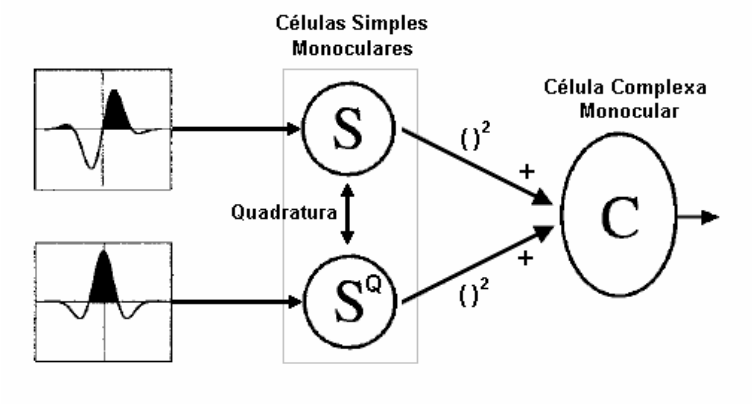

Figura 3-15 - Célula complexa monocular formada pela composição de duas células simples em quadratura de fase.

A célula complexa binocular deste modelo, tem como base o modelo proposto por Ohzawa [22] (Figura 3-14), porém no modelo de célula complexa proposto pelo Ohzawa, os campos receptivos direito e esquerdo de cada célula simples estão em fase (diferença de fase nula), enquanto que as células simples entre si (S1, S2, S3 e S4) possuem diferença de fase (Figura 3-14). No modelo proposto, os campos receptivos de uma mesma célula simples possuem uma diferença de fase de 180º como mostrado na Figura 3-11. Contudo, assim como no modelo de Ohzawa, neste modelo há uma diferença de fase de 90º entre as células simples (elas estão em quadratura). A Figura 3-16 mostra de forma esquemática o modelo implementado de célula complexa binocular.

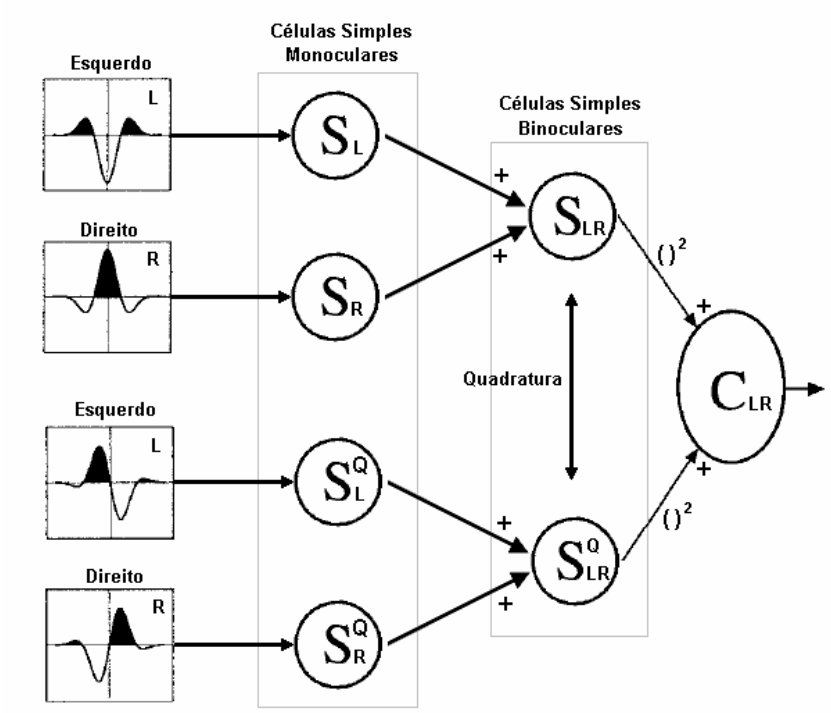

Figura 3-16 - Modelo implementado de célula complexa binocular. Este modelo consiste em subunidades de células simples binoculares em quadratura, que por sua vez são compostas de células simples monoculares.

Esta diferença de fase de 180º graus entre os campos receptivos das células simples binoculares que compõem a célula complexa binocular, faz com que o comportamento desta célula complexa seja do tipo *tuned inhibitory*. Células *tuned inhibitory* têm resposta mínima quando os estímulos nos seus campos receptivos no olho esquerdo e direito são iguais. O modelo proposto por Ohzawa (Figura 3-14) é do tipo *tuned excitatory*, no qual a célula complexa responde de forma máxima quando os estímulos nos seus campos receptivos no olho esquerdo e direito são iguais. A Figura 3-17 mostra o perfil da resposta de células *tuned inhibitory* e *tuned excitatory*.

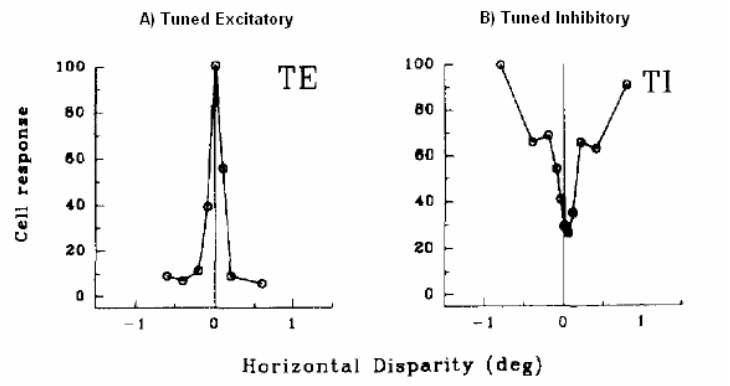

Figura 3-17 - Resposta de células do tipo *tuned excitatory* e *tuned inhibitory*. Fonte: [11].

3.3. MODELAGEM DA ÁREA MT

A área do córtex conhecida como MT ou V5 recebe projeções principalmente da camada 4B de V1 e das faixas grossas de V2 (seção 2.3.5), sendo que a área MT, além de processar informações sobre profundidade, também processa informações sobre movimento. Neste trabalho é abordado somente o processamento de profundidade e, por essa razão, a modelagem de MT foi especificamente baseada na continuidade do processamento da percepção da profundidade partindo do processamento já realizado em V1, pelas células simples e complexas. As respostas das células complexas de V1 são propagadas para a área MT, onde continuará o processamento da percepção de profundidade. Basicamente, em MT, esta informação de profundidade vinda de V1 sofrerá um refinamento e, em seguida, será agrupada por níveis superiores do córtex de modo a selecionar a melhor disparidade para cada ponto processado.

Na modelagem da área MT feita neste trabalho, cada célula de MT recebe projeções de uma célula complexa binocular (C_{LR}) e de duas células complexas monoculares (C_L e C_R), com campos receptivos centrados nos mesmos pontos da célula complexa binocular. Desta forma, cada célula de MT é binocular e possui um campo receptivo formado pela combinação de uma célula complexa binocular e duas células simples monoculares, conforme mostrado na Figura 3-18.

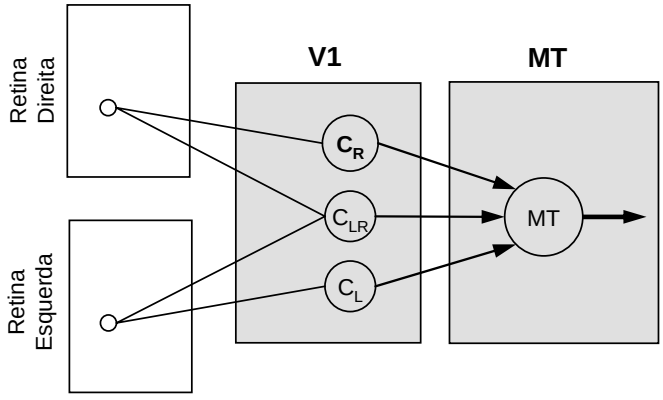

Figura 3-18 - Uma célula de MT implementada recebe projeções de células complexas de V1 com campos receptivos centrados nos mesmos pontos.

Neste modelo, uma célula de MT normaliza a saída de uma célula complexa binocular pela soma das saídas das células complexas monoculares. Uma constante *k* foi adicionada a esta normalização para controlar a seletividade da célula. Desta forma, a saída de uma célula de MT é dada pela equação (14).

$$MT = \frac{C_{LR}}{C_L + C_R + k} \tag{14}$$

De acordo com o modelo esquemático da arquitetura funcional de MT mostrado na Figura 2-20 (seção 2.3.5), a percepção de profundidade é codificada em faixas que variam suave e continuamente entre disparidades *near* (perto) e disparidades *far* (longe) passando pela disparidade zero. A modelagem desta arquitetura foi feita utilizando várias camadas de células complexas, onde cada camada esta ajustada para detectar uma determinada disparidade. Camadas sucessivas detectam disparidades sucessivas.

Neste modelo, uma camada ajustada para disparidade *d* em MT é modelada como um conjunto de células que recebe a projeção das repostas das células complexas de V1 que possuem seus campos receptivos ajustados para detectar esta disparidade. Para modelar uma camada ajustada para uma disparidade *d+1*, os campos receptivos que recebem informações sobre a imagem projetada na retina direita são centrados na mesma posição que os mesmos campos receptivos da camada *d*, enquanto que os campos receptivos que recebem informações sobre a imagem projetada na retina esquerda são centrados numa posição deslocada de uma unidade (um *pixel*) para a direita, aumentando a disparidade binocular em uma unidade, conforme mostrado na Figura 3-19. Logo, neste modelo a disparidade é computada segundo a diferença de posição dos campos receptivos esquerdo e direito das células e não segundo a fase (vide Figura 3-9).

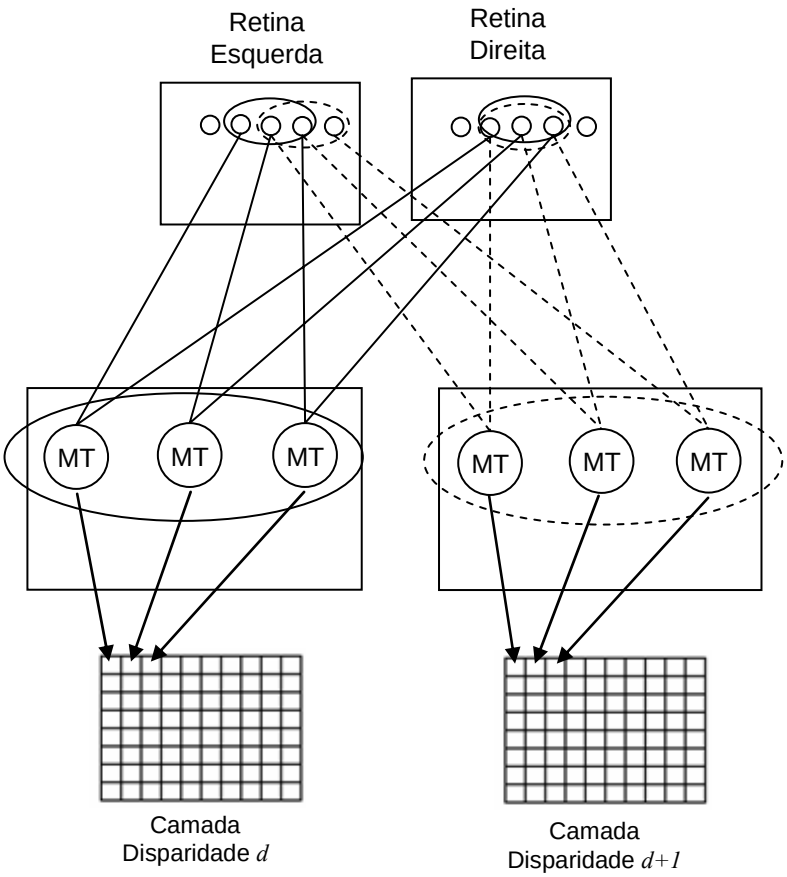

Figura 3-19 - Camadas com disparidade *d* e *d+1* em MT .

Várias camadas com disparidades sucessivas são agrupadas em ordem, conforme mostrado na Figura 3-20, de modo a modelar a arquitetura de MT de acordo com o modelo funcional biológico mostrado na Figura 2-20 (seção 2.3.5). No cérebro, as células que codificam diferentes disparidades associadas à um mesmo ponto da imagem são vizinhas na superfície do córtex MT; neste modelo, apenas por conveniência, esta vizinhança ocorre na dimensão da profundidade da matriz tridimensional da Figura 3-20. No cérebro este arranjo não seria possível, devido à organização planar bidimensional do córtex

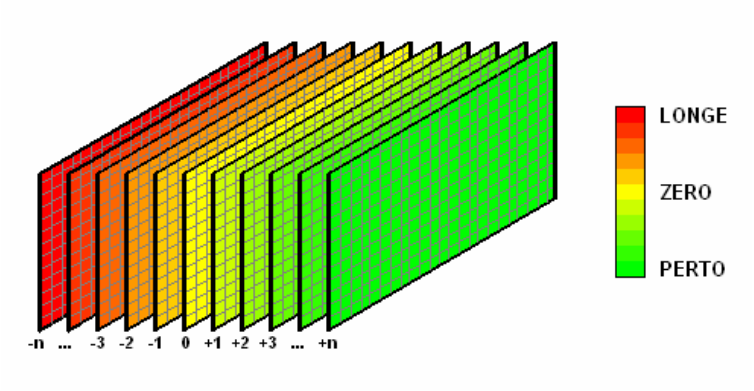

Figura 3-20 - Modelagem em camadas da arquitetura funcional de MT.

Em cada camada de MT, foi incluída uma forma de cooperação entre as células vizinhas de uma mesma camada, o que melhora significantemente a percepção de profundidade. Uma célula de MT influência no resultado de outras células em sua vizinhança, numa mesma camada, e vice versa. Isto é baseado no fato que o tamanho do campo receptivo das células complexas é, em média, maior que os campos receptivos das células simples de mesma excentricidade [14], e neste modelo, as células de MT recebem projeções exclusivamente de células complexas de V1.

Foi proposto por Qian & Zhu [24] que esta influência entre as células vizinhas pode ser incorporada no modelo de uma célula complexa calculando a média dos pares de células simples em quadratura próximas umas das outras, sobrepondo seus campos receptivos. Isto pode ser modelado matematicamente aplicando uma convolução com uma função peso espacial, processo no qual é chamado de ***spatial pooling***.

Com esta cooperação entre as células vizinhas, a resposta final de uma célula complexa (R_c) é dada pela convolução espacial das respostas das células complexas ao seu redor (R_q) com uma função que dá pesos diferentes para esta cooperação em função da distância da célula vizinha à célula complexa em questão. A função peso (w) utilizada foi uma Gaussiana bidimensional assim como descrito na equação (7).

$$R_c = R_q * w \tag{15}$$

3.4. ARQUITETURA PROPOSTA

A arquitetura de MT proposta é apresentada na Figura 3-21. As células da área MT recebem projeções das células complexas de V1 (tanto células complexas binoculares quanto monoculares) que por sua vez recebem informações das células simples binoculares e monoculares respectivamente.

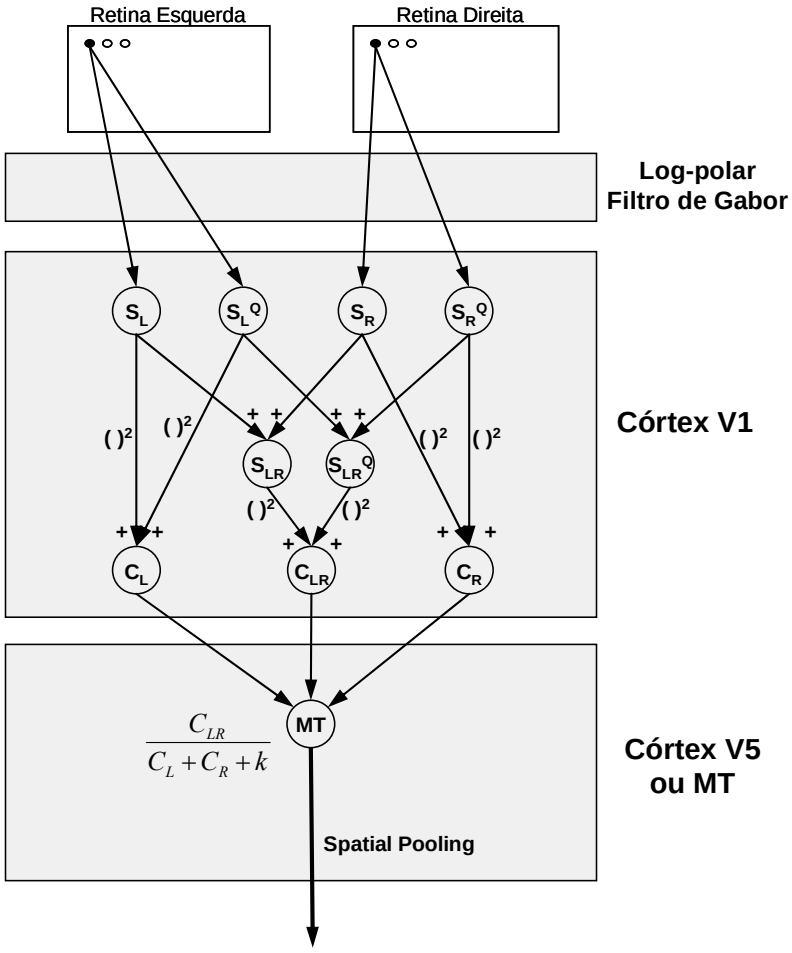

Figura 3-21 - Modelo representando a arquitetura de V1 e MT. Cada célula de MT recebe projeções das células complexas de V1.

Neste modelo, as imagens projetadas nas retinas são conduzidas até ao córtex V1 passando pelo LGN. Neste caminho, devido aos fatores de magnificação e ao tipo de mapeamento retinotópico existente no LGN (Figura 2-9 – seção 2.2) e no próprio córtex (Figura 2-14 – seção 2.3.1), as imagens sofrem uma transformação log-polar. Os filtros de Gabor modelam o campo receptivo das células simples em V1. As células simples monoculares S_L e S_L^Q, com campo receptivo na retina esquerda, estão em quadratura de fase,

assim como S_R e S_R^Q, na retina direita. As células simples binoculares S_{LR} e S_{LR}^Q recebem projeções das células simples monoculares em fase e em quadratura respectivamente e, portanto, também estão em quadratura de fase.

A célula complexa monocular C_L tem como entrada a saída das células simples monoculares S_L e S_L^Q, enquanto que C_R tem como entrada a saída das células simples monoculares S_R e S_R^Q. A célula complexa binocular C_{LR} recebe projeções das células simples monoculares S_{LR} e S_{LR}^Q. Uma célula em MT recebe projeções de três células complexas, sendo duas delas monoculares e uma binocular. Desta forma, cada célula em MT recebe informações provenientes de um processamento de um conjunto de células simples e complexas em V1 composto de:

1. Duas células simples monoculares em quadratura de fase com campo receptivo no olho direito;
2. Duas células simples monoculares em quadratura de fase com campo receptivo no olho esquerdo;
3. Duas células simples binoculares em quadratura compostas pelas células dos itens 1 e 2;
4. Uma célula complexa monocular, com campo receptivo no olho direito, formada pelas células do item 1;
5. Uma célula complexa monocular, com campo receptivo no olho esquerdo, formada pelas células do item 2;
6. Uma célula complexa binocular formada pelas células do item 3.

Considerando os itens citados como um bloco de processamento em V1, a Figura 3-22 mostra um esquema da propagação da informação visual da retina até MT para detecção de disparidade. Em camadas de MT com células sintonizadas em disparidades mais elevadas (perto), os campos receptivos provenientes do olho esquerdo estão centrados em posições deslocadas para a direita, já em camadas de MT com células sintonizadas em disparidades mais baixas (longe), estes campos receptivos estão centrados em posições deslocadas para a esquerda. Os campos receptivos provenientes do olho direito não são deslocados. A Figura 3-22 mostra este deslocamento dos campos receptivos para camadas com disparidade d e $d+1$, juntamente com os blocos de processamento em V1.

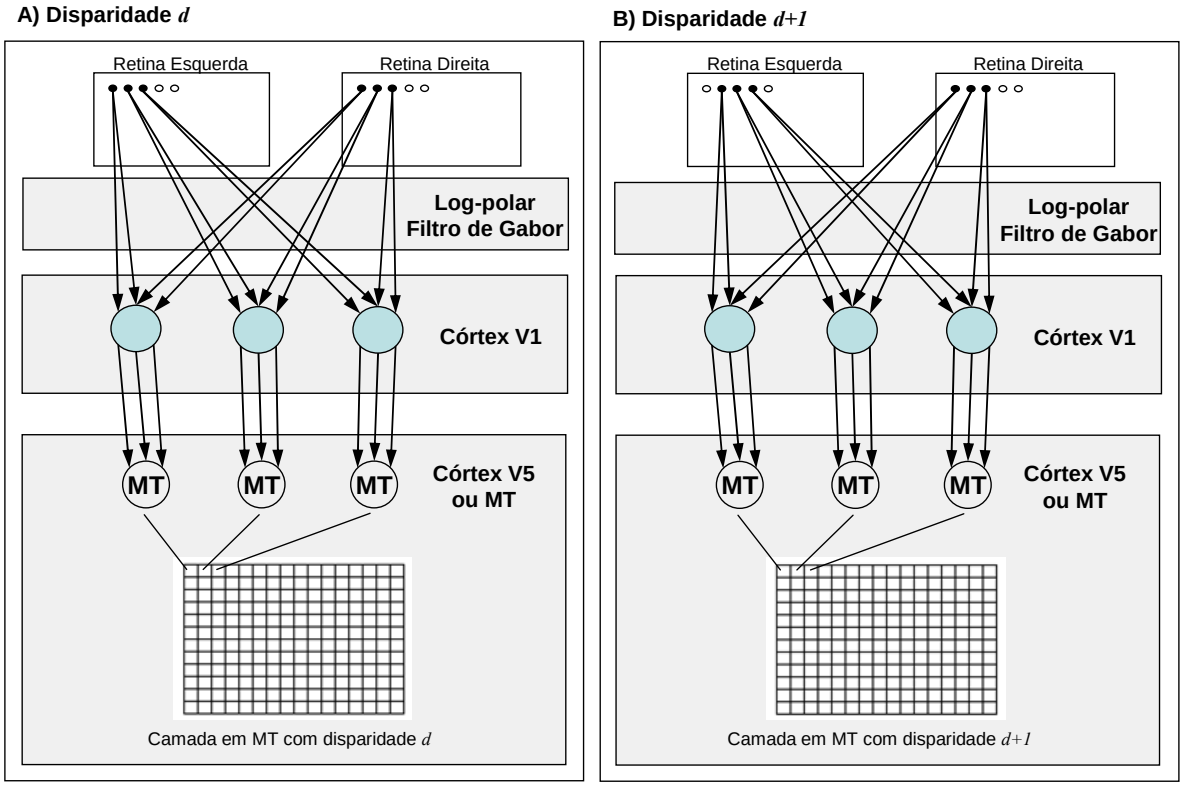

Figura 3-22 – Esquema da arquitetura proposta mostrando que cada célula de MT recebe projeções de um bloco de processamento em V1. A camada com disparidade *d+1* possui os campos receptivos provenientes da retina esquerda centrados em posições deslocadas de uma unidade (pixel) em relação à camada com disparidade *d*.

4. METODOLOGIA

Neste capítulo será apresentada toda a metodologia utilizada para implementar a arquitetura proposta neste trabalho de pesquisa e calcular a disparidade binocular entre duas imagens estéreo bidimensionais utilizando esta arquitetura. Também serão apresentadas as ferramentas utilizadas para implementar a arquitetura e a plataforma robótica que foi construída para este trabalho.

4.1. *FRAMEWORK* MAE

Nesta seção será apresentado o *framework* MAE – Máquina Associadora de Eventos - que foi a principal ferramenta utilizada para o desenvolvimento deste trabalho. A MAE é um *framework* construído inicialmente para ser utilizado em plataformas UNIX, que permite projetar tanto estruturas modulares usando RNSP (Redes Neurais Sem Peso) com neurônios do tipo VGRAM (*Virtual Generalising Random Access Memory*) quanto estruturas com arquitetura em camadas, definindo um processamento específico para cada camada. Neste trabalho não foi utilizado RNSP, portanto não será abordada a utilização da MAE para projetar este tipo de estrutura. Maiores detalhes da utilização deste *framework* em projetos de arquiteturas RNSPs, pode ser encontrado em [17].

A utilização do *framework* MAE se deve às várias facilidades fornecidas tais como:

- Simplicidade na descrição de arquiteturas neurais e/ou baseada em camadas: A MAE permite a descrição de arquiteturas neurais e/ou camadas através de uma linguagem de alto nível, com geração automática do código que as implementa.
- Flexibilidade: Na linguagem de descrição de arquiteturas, vários parâmetros podem ser alterados; além disso, a ferramenta permite a incorporação de código do usuário.
- Facilidade visual. Toda camada descrita pela linguagem utilizada pela MAE pode ter uma representação visual, o que torna possível o acompanhamento da resposta temporal de cada camada.
- Portabilidade. A MAE é implementada utilizando uma linguagem padrão, o C ANSI, possibilitando que seja portada para qualquer sistema que possua um compilador C padrão. Os pacotes e bibliotecas importadas pelo *framework* são

pacotes portáteis como Mesa3D, Flex, Bison, entre outros. A MAE já foi portada para vários sistemas UNIX, como Conectiva Linux, Red Hat, Solaris, Fedora e Debian, e atualmente existe uma versão em fase de testes para a plataforma Windows.

O sistema MAE é composto, basicamente, de duas partes. A primeira é um programa chamado netcomp, que recebe como entrada um arquivo com extensão .con contendo a descrição da arquitetura (neural ou de camadas), e gera como saída um arquivo com extensão .c, contendo código C correspondente à tradução do arquivo .con para C. A segunda é a biblioteca de rotinas MAE, que implementa os neurônios, seu treinamento, processamento específico para cada camada, interface com o usuário, etc.

O netcomp é um tradutor da linguagem de descrição de arquitetura descrita no arquivo .con para C. O arquivo .c gerado e os arquivos de rotinas do usuário são compilados e linkados com outras bibliotecas necessárias por um script chamado netcompiler. É este script que gera o arquivo executável final, ou aplicação MAE, que implementa a arquitetura descrita no arquivo .con, além de uma interface que permite ao usuário manipular a arquitetura. Na Figura 4-1 é mostrado um esquema que ilustra o processo para gerar uma aplicação MAE. Na verdade, este processo é transparente ao usuário. O netcompiler se encarrega de realizar todo este processo automaticamente, gerando como saída uma aplicação MAE.

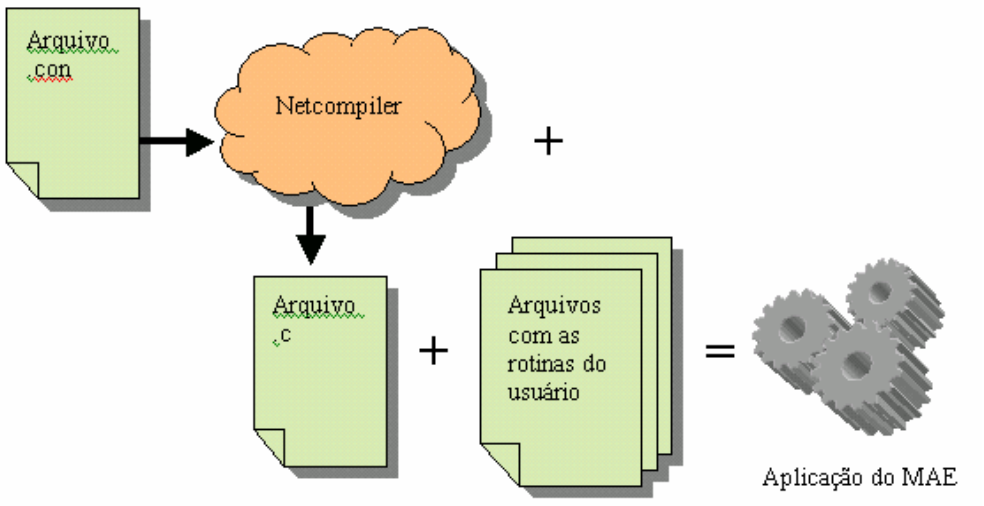

Figura 4-1 - Esquema de criação de uma aplicação utilizando o *framework* MAE.

A MAE utiliza uma biblioteca própria implementada em ANSI C e um conjunto de pacotes e bibliotecas externas que são utilizados para fornecer algumas facilidades sem necessidade de implementá-las. Os pacotes e bibliotecas utilizadas pela MAE são:

- Biblioteca MAE: Biblioteca implementada em C, com código fonte escrito em ANSI C, com as rotinas necessárias para efetuar todo o controle da aplicação. Contém também todas as rotinas utilizadas para criação e manipulação de cada uma das partes que compõe a arquitetura (neural e/ou camadas), inclusive rotinas para treinamento e execução da rede neural, quando a aplicação for uma RNSP.
- MESA 3-D e GLUT: MESA 3-D é uma biblioteca gráfica 3D com uma API bastante semelhante ao OpenGL e que utiliza o mesmo tipo de máquina de estados e sintaxe de comandos do OpenGL, porém de distribuição livre. O GLUT (OpenGL Utility Toolkit) é um conjunto de ferramentas independente do sistema de janelas utilizado pelo sistema operacional para escrever programas em OpenGL, que também pode ser utilizado com o MESA 3D.
- XForms: Conjunto de ferramentas para desenvolver interface gráfica com o usuário baseada no sistema de janelas X Window. Utilizada para a criação do painel de controle de uma aplicação MAE como controle dos eventos do mouse, teclado, menus e botões.
- Flex & Bison: Flex é um gerador de código de analisadores léxicos. É uma implementação de distribuição livre da ferramenta Lex. Bison é um gerador de código de analisador sintático (*paser*) baseado numa gramática descrita num arquivo texto entrado como parâmetro. A MAE utiliza o Flex e o Bison em conjunto para descrever e interpretar as duas linguagens (linguagem de descrição da arquitetura e linguagem de execução de scripts para RNSP) utilizadas.

4.1.1. Linguagem de descrição das camadas

O conteúdo dos arquivos com a definição léxica e sintática da linguagem de descrição das camadas se encontra nos arquivos `netcomp_lex.l` e `netcomp_yacc.y` no diretório `<diretório da MAE>/src`. Nesta seção será descrita resumidamente os comandos da linguagem utilizados. Esta linguagem de descrição da arquitetura de camadas da aplicação também é utilizada para descrever as camadas de redes neurais para o caso de uma aplicação de RNSP, com comandos que criam as conexões (sinapses) entre as camadas de redes neurais e, portanto, permitindo o treinamento das mesmas, entretanto, como não foi utilizado RNSP neste trabalho, serão descritos somente os comandos utilizados na arquitetura

de camadas que foi utilizada neste trabalho. Os comandos para descrição das camadas são: `neuron_layer`, `input`, `output`, `output_connect`, `filter`, e `set`.

- **`neuron_layer`**

Uma `neuron_layer` é, basicamente, uma matriz de neurônios, que neste trabalho são utilizadas para implementar as camadas de neurônios (células simples, células complexas, células de MT, etc.). A sintaxe para especificar uma `neuron_layer` é mostrada abaixo:

```
neuron_layer <nome><dimensão> with <tipo de saída> outputs;
```

- `<nome>` : Nome pelo qual a `neuron_layer` poderá ser referenciada.
- `<dimensão>` : Dimensão da `neuron_layer`. Definida entre colchetes. Exemplo da dimensão de uma `neuron_layer` bidimensional 10x10 com um total de 100 neurônios: [10][10].
- `<tipo de saída>` : Define o tipo de saída dos neurônios de uma `neuron_layer`. Pode ser: `b&w`, `grayscale`, `grayscale_float` e `color`. Quando as saídas são definidas como `b&w` elas podem assumir os valores 0 ou 1 apenas, representando, o preto e o branco. Quando as saídas são definidas como `grayscale` elas podem assumir valores inteiros de 0 à 255, representando uma escala de cinza com 256 valores. Quando as saídas são definidas como `grayscale_float` elas podem assumir qualquer valor de ponto flutuante, entretanto para a visualização, é selecionado o maior valor em módulo (|s_max|) da saída dos neurônios desta `neuron_layer`, e este valor é utilizado para realizar uma discretização na visualização: a parte positiva, definida pelo intervalo **[0, |s_max|]**, é discretizada e visualizada como 256 níveis de verde, enquanto que a parte negativa, definida pelo intervalo **[-|s_max|, 0[**, é discretizada e visualizada como 256 níveis de vermelho. Caso só possua valores positivos, a visualização é feita em escala de cinza. Quando as saídas são definidas como `color`, elas podem assumir um valor RGB de 24 bits, com 8 bits por cor.

- **input**

Através do comando input , é criada uma neuron_layer com nome, dimensão e saída especificadas, entretanto, um tratamento especial é dada nestas neuron_layer's. É criada uma janela para cada input gerada que mostra o estado dos neurônios desta input, e podem ser associadas duas funções, uma para geração e outra para manipulação da input. A neron_layer criada pela input pode ser manipulada por eventos de mouse e teclado. A sintaxe para especificar uma input é mostrada abaixo:

```
input <nome><dimensão> with <tipo de saída> outputs [produced by <função
geradora>] [controled by <função de controle>];
```

- <nome> : Nome pelo qual a input poderá ser referenciada.
- <dimensão> : Dimensão da input. Possui a mesma descrição do parâmetro <dimensão> da neuron_layer, citado anteriormente.
- <tipo de saída> : Define o tipo de saída dos neurônios de uma input. Podem ser: b&w, grayscale, grayscale_float e color. Possui a mesma descrição do parâmetro <tipo de saída> da neuron_layer, citado anteriormente.
- <função geradora> : O objetivo da função de geração é inicializar parâmetros da input, criar sua janela, associá-la às funções de controle de mouse, teclado, além do controle de atualização do conteúdo da janela. Podem-se passar quantos parâmetros forem necessários para a função de geração, desde que sejam constantes, expressões aritméticas ou nomes de objetos já definidos no arquivo de definição. Nesta função, normalmente é definida como a input receberá as informações de entrada, por exemplo, de um arquivo de entrada. A declaração de uma função de geração é opcional. Se não for definida, a função pré-estabelecida na biblioteca MAE será utilizada, que é a que trata de imagens estáticas. Esta função procura no diretório em que está sendo executada a aplicação MAE, um arquivo com o mesmo nome da input com extensão .in. Uma string na primeira linha deste arquivo deve informar o nome de um arquivo .bmp contendo a imagem estática a ser usada como entrada. Os valores de cada pixel deste arquivo são passados para os neurônios da input.

- o <função de controle> : Também é opcional. O objetivo da função de controle é fornecer o código necessário para tratar a interação da input com o usuário, por exemplo, eventos de clique de mouse sobre a janela de visualização da input. Podem-se passar quantos parâmetros quiser, desde que sejam constantes, expressões aritméticas ou nomes de objetos já definidos no arquivo de definição.

- **output**

Uma output é apenas uma visualização das saídas de uma neuron_layer. Através de uma output, pode-se fornecer uma interação do usuário com uma determinada neuron_layer, semelhante a input. A sintaxe para especificar uma output é mostrada abaixo:

```
output <nome><dimensão> [handled by <função manipuladora>];
```

- o <nome> : Nome pelo qual a output poderá ser referenciada.
- o <dimensão> : Dimensão da output. Possui a mesma descrição do parâmetro <dimensão> da neuron_layer, citado anteriormente.
- o <função manipuladora> : A declaração desta função é opcional. O objetivo da função manipuladora é fornecer o código necessário para tratar a interação da output, e portanto com a neuron_layer na qual está associada (vide comando output_connect), com o usuário. Podem-se passar quantos parâmetros quiser, desde que sejam constantes, expressões aritméticas ou nomes de objetos já definidos no arquivo de definição.

- **output_connect**

Comando utilizado para associar uma output com uma neuron_layer que, para isto, devem possuir as mesmas dimensões. A sintaxe do comando output_connect é mostrada abaixo:

```
output_connect <nome da neuro_layer> to <nome da output>;
```

- `<nome da neuron_layer>` : Nome da `neuron_layer` que terá a saída de seus neurônios mostrada na `output`.
- `<nome da output>` : Nome da `output` que mostrará a saída dos neurônios da `neuron_layer`.

- **`filter`**

Este comando define que uma ou mais `neuron_layers`, terão suas saídas processadas e, o resultado deste processamento será a entrada de uma outra `neuron_layer`. Esta é a definição de filtro para a MAE. Alguns filtros mais utilizados, por exemplo, um filtro que computa a soma da saída de um conjunto de `neuron_layer`'s colocando o resultado numa outra `neuron_layer`, estão implementados no arquivo `<diretório da MAE>/src/filter.c`, entretanto podem ser implementados filtros específicos para uma determinada aplicação, bastando que seja codificada no arquivo de funções do usuário que será descrito posteriormente. A sintaxe do comando `filter` é mostrada abaixo:

```
filter <neuron_layer's de entrada> with <nome do filtro><parâmetros> producing <neuron_layer de saída>;
```

- `<neuron_layer's de entrada>` : Nome das `neuron_layer`'s que são entrada para o filtro. Caso tenha mais de uma `neuron_layer`, deverão ser separadas por vírgula.
- `<nome do filtro>` : Nome do filtro, que deve ser o mesmo nome da função que o implementa.
- `<parâmetros>` : Conjunto de parâmetros passado para a função que implementa o filtro, cuja sintaxe é semelhante à sintaxe do comando `printf` da linguagem C.
- `<neuron_layer de saída>` : Nome da `neuron_layer` que receberá a saída do processamento do filtro.

- **`set`**

O comando `set` atribui um valor a uma variável global qualquer da aplicação MAE, definida em seu código fonte. A sintaxe do comando `set` é mostrada abaixo:

```
set <variável> = <valor>;
```

  - `<variável>` : Nome da variável global.
  - `<valor>` : Valor a ser atribuído na variável global.

4.1.2. Rotinas implementadas pelo usuário

Na construção de uma aplicação utilizando o *framework* MAE, é permitido incluir funções implementadas pelo usuário. Nesta seção serão descritas as funções de usuário predefinidas pelo *framework*, juntamente com a sintaxe, objetivos e quando são chamadas. Estas funções devem estar dentro de um arquivo cujo nome deve estar de acordo com o nome do arquivo .con que define a arquitetura da aplicação. Se o arquivo se chama <arq>.con, este arquivo deve se chamar <arq>_user_functions.c e deve estar localizado dentro de um diretório localizado no mesmo diretório do arquivo <arq>.con de nome <arq>_user_functions.

- Arquivo de descrição da arquitetura
  - `<arq>.con`

- Arquivo do rotinas do usuário:
  - `<arq>_user_functions/<arq>_user_functions.c`

- `int init_user_functions (INPUT_DESC *input, int status);`

Nesta função o usuário pode implementar qualquer inicialização necessária para a aplicação desenvolvida. Recebe como parâmetro um ponteiro para uma `input` e o código de status da execução da aplicação, isto é, se está em movimentação (MOVE), processando (RUN) ou apenas posicionando-o (SET_POSITION). É chamada uma única vez depois que as `neuron_layer`'s já foram instanciadas e antes de entrar no ciclo de espera por algum evento do usuário. Função de implementação obrigatória.

- `void input_generator (INPUT_DESC *input, int status);`

Função de geração da `input`. Pode ter qualquer nome, contudo igual ao nome do parâmetro `<função de geração>` do comando de descrição da `input.` É chamada uma vez por ciclo de execução da aplicação. Recebe como parâmetro um ponteiro para a `input` na qual foi definida e o código de status da execução da aplicação. Esta função só é obrigatória caso seja definida no comando de criação da `input`.

- `void input_controler (INPUT_DESC *input, int status);`

Função de controle da `input`. Pode ter qualquer nome, contudo igual ao nome do parâmetro `<função de controle>` do comando de descrição da `input`. É chamada toda vez que ocorre um evento de clique de mouse dentro da janela da `input` na qual foi definida. Recebe como parâmetro um ponteiro para a `input` na qual foi definida e o código de status da execução da aplicação. Esta função só é obrigatória caso seja definida no comando de criação da `input`.

- `void output_handler (OUTPUT_DESC *output, int status);`

Função de manipulação da `neuron_layer` na qual a `output` foi associada. Pode ter qualquer nome, contudo igual ao nome do parâmetro `<função de manipulação>` do comando de descrição da `output`. É chamada toda vez que ocorre um evento de clique de mouse dentro da janela da `output` na qual foi definida. Recebe como parâmetro um ponteiro para a `output` na qual foi definida e o código de status da execução da aplicação. Esta função só é obrigatória caso seja definida no comando de criação da `output`.

- `void <nome do filtro> (FILTER_DESC *filter_desc);`

Função que implementa o processamento de um filtro definido no arquivo de descrição da arquitetura da aplicação (arquivo .con). Pode ter qualquer nome, contudo igual ao nome do parâmetro `<nome do filtro>` de algum comando `filter`. É chamada uma vez por ciclo de execução da aplicação. Recebe como parâmetro um ponteiro para uma estrutura do tipo `FILTER_DESC` no qual possui todas as informações sobre este filtro, como as `neuron_layer`'s e os parâmetros de entrada e a `neuron_layer` de saída. A consistência quanto à quantidade de camadas de entrada e saída deve ser implementada pelo usuário dentro

do código do filtro. Só é obrigatória se a mesma tiver sido usada no comando `filter` e não existir na biblioteca MAE.

4.2. IMPLEMENTAÇÃO DA PLATAFORMA ROBÓTICA

O desenvolvimento deste trabalho demandou a necessidade de construir uma plataforma (hardware e software) na qual é possível capturar 2 imagens estereoscópicas do mundo externo (separadas a uma distância equivalente a distância entre os olhos humanos) e a partir destas imagens, realizar todo o processamento necessário para a construção de uma representação tridimensional que possibilite um sistema robótico navegar neste mundo.

A plataforma robótica utilizada nesse estudo foi construída sobre um robô de uso doméstico, o aspirador de pó Roomba Pro Elite (www.irobot.com). Uma vez que o robô original não possui uma interface pronta para a comunicação com o PC, mas pode ser acionado por um controle remoto infravermelho, optou-se por criar um módulo de acionamento infravermelho para controlar a plataforma robótica a partir do PC.

Tal módulo é constituído por um hardware capaz de receber e emitir sinais na faixa do espectro infravermelho, uma interface de comunicação serial com o PC e uma biblioteca na linguagem C que permite identificar, decodificar e armazenar sinais para retransmiti-los posteriormente.Assim, é possível ensinar ao computador a controlar os movimentos da plataforma robótica no plano. Os movimentos possíveis para o robô são:

1. Frente
2. Rotação para a esquerda
3. Rotação para a direita
4. Composição: Frente + Rotação para a esquerda
5. Composição: Frente + Rotação para a direita

Para simplificar a solução utilizada, os movimentos 4 e 5 não estão sendo utilizados. Desta forma, com os movimentos 1, 2 e 3 o robô pode andar em linha reta para qualquer direção, bastando que, após escolhida uma direção e sentido, utilize ou movimento 2 ou o movimento 3 para se orientar e depois o movimento 1 para se movimentar na direção e sentido escolhido. Para controlar os movimentos do robô, foi implementada uma interface entre o robô e um computador (denominado computador central) através da porta paralela.

Duas câmeras foram colocadas em cima do robô paralelamente, separadas uma da outra por uma distância de 6,9 cm. Esta distância foi ajustada para que ficasse entre 6,5 cm e 7 cm que é a distância média entre os olhos de um adulto.

Para capturar a imagem das câmeras foram utilizados 2 computadores, um para cada câmera, cada um com uma placa capturadora de vídeo. Cada câmera foi conectada ao computador através de um cabo S-Vídeo. Estes dois computadores (computador 1 e computador 2) e o computador central estão conectados por uma rede *Ethernet* através de um *switch* de 10/100Mbits. O robô, já com as câmeras acopladas na parte superior, é mostrado Figura 4-2A e a arquitetura do sistema é apresentada na Figura 4-2B.

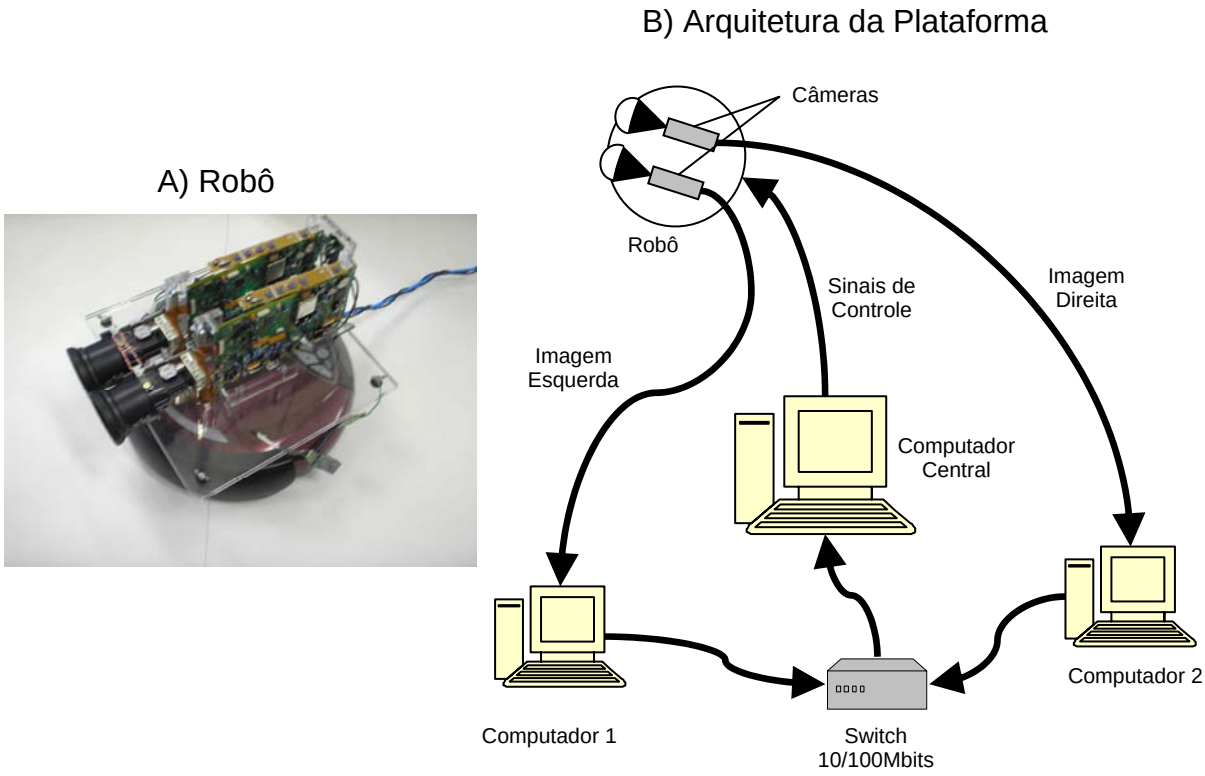

Figura 4-2 – A) Imagem do robô com 2 câmeras acopladas na parte superior. B) Esquema da arquitetura de hardware do sistema para captura de imagens e controle do robô.

Com esta arquitetura o computador central pode receber as imagens das duas câmeras, processar estas informações simulando o sistema visual humano, gerando uma imagem tridimensional da onde o robô está "olhando". Com base nesta representação o computador central pode tomar uma decisão para movimentar o robô recebendo novas imagens, repetindo o ciclo.

O software de captura das imagens das câmeras foi desenvolvido na plataforma Windows, utilizando a linguagem Visual C++ 6.0 com a biblioteca DirectX 9.0. No computador central foi utilizado o *framework* MAE (Máquina Associadora de Eventos) para desenvolver uma aplicação que recebe as imagens das câmeras esquerda e direita, realiza o

processamento desta informação visual reconstruindo uma representação tridimensional e, a partir desta representação, tomar as decisões necessárias para controlar a negação do robô. Esta aplicação foi desenvolvida numa plataforma Linux (Fedora Core 2), utilizando a linguagem C, com algumas bibliotecas gráficas, como Mesa3-D e Xforms, conforme visto na seção 4.1.

4.3. MÉTODO UTILIZADO

Nesta seção será descrita toda a seqüência do processo realizado para computar a disparidade binocular das imagens direita e esquerda, e reconstruir uma imagem tridimensional a partir destas informações. Será discutido o processo desde a captura das imagens pelas câmeras instaladas no robô, até o processamento das imagens, incluindo a detecção da disparidade binocular entre elas e a reconstrução tridimensional interna ao computador. Este processo consiste nos seguintes passos:

1. Capturar as imagens direita e esquerda e enviar para a MAE
2. Escolher um ponto de atenção na imagem direita
3. Vergência
4. Calcular a disparidade entre as imagens direita e esquerda
5. Criar um mapa de disparidade binocular com a disparidade mais adequada para cada ponto no mapa
6. Reconstruir a imagem tridimensional a partir do item 5.

De todos os passos citados existe um, o passo 2, que é feito manualmente; os demais são realizados pelos computadores pertencentes à arquitetura do sistema desenvolvido. Os passos 3, 4 e 5 são realizados num mesmo laço (*loop*) de iterações. Como resultado do processamento deste laço é obtido um mapa final de disparidades entre as imagens direita e esquerda. Com este mapa de disparidades é possível realizar um mapeamento inverso, reconstruindo uma imagem tridimensional a partir das imagens bidimensionais capturadas.

Contudo, antes de prosseguir é necessário descrever como é calculada a coordenada de um ponto no espaço a partir da projeção deste ponto na retina direita e esquerda. Após esta descrição serão discutidos os 6 passos do processo de computo da disparidade binocular mencionados acima.

4.3.1. Cálculo da coordenada espacial de um ponto

A vergência constitui um dos movimentos oculares mais importantes e é imprescindível para a percepção de distância em relação aos objetos do mundo no campo de visão. Ela permite direcionar os olhos para um determinado ponto, ou seja, fixar as projeções do ponto em questão sobre as fóveas. Em paralelo, o sistema visual é realimentado com informações sobre o posicionamento dos olhos, fornecidas pelo sistema óculo-motor e as interpreta como distância. Desta forma, o observador é capaz de estimar a distância até um ponto em particular do mundo. O processo através do qual se realiza a vergência será abordado na seção 4.3.4. Nesta seção será discutida a formulação utilizada para calcular a posição de um ponto no espaço a partir das coordenadas deste ponto projetada nas retinas, que neste caso são as câmeras.

Dada a vergência num ponto qualquer, tem-se a projeção deste ponto nas imagens e, partir daí, localizá-lo no espaço é relativamente simples, a principal dificuldade para viabilizar tal processo é conseguir realizar uma vergência razoavelmente precisa no ponto em questão. A Figura 4-3 ilustra de forma sucinta a formulação desenvolvida para determinar as coordenadas de um ponto no mundo (espaço 3D).

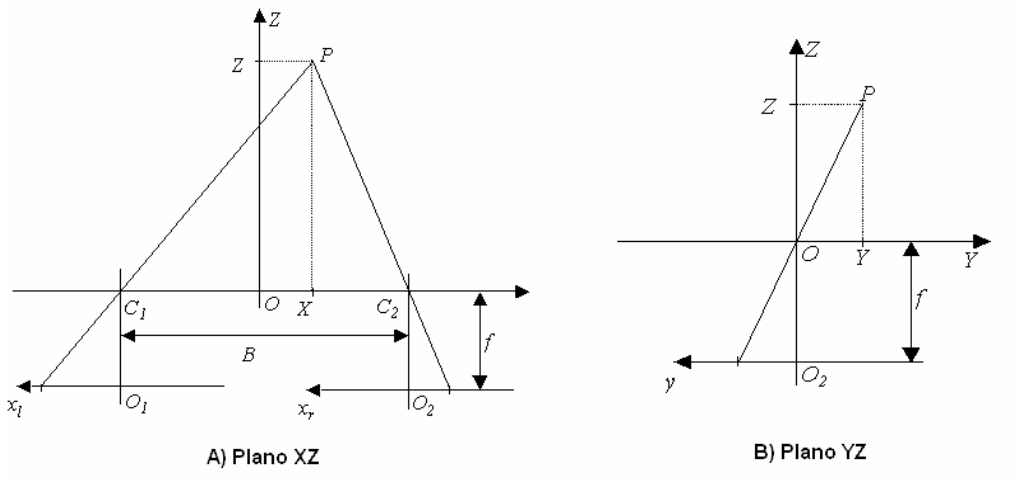

Figura 4-3 - Geometria do cálculo de um ponto no espaço a partir da projeção nas câmeras.

Seja *f* a distância focal das câmeras medida em pixels e *B* a separação entre elas em metros (Figura 4-3). Por semelhança de triângulos, a coordenada *Z* do ponto no mundo em relação ao referencial *O* (Figura 4-3-A) é dada por:

$$Z = \frac{f \cdot B}{x_l - x_r} \tag{16}$$

onde x_l e x_r são as abscissas das projeções do ponto nas imagens direita e esquerda, respectivamente.

De forma análoga, as coordenadas X e Y do ponto P do mundo são dadas por:

$$X = \frac{B}{2} \cdot \frac{(x_l + x_r)}{(x_l - x_r)} \tag{17}$$

$$Y = \frac{y \cdot B}{(x_l - x_r)} \tag{18}$$

no qual $y = y_l = y_r$.

4.3.2. Captura das imagens

Todo o processo de computar a disparidade binocular das imagens direita e esquerda, e reconstruir uma imagem tridimensional a partir destas informações é iniciado pela captura das imagens pelas câmeras que, conforme mencionado na seção 4.2, estão instaladas paralelamente em cima do robô, separadas a uma distância de 6,9 cm. Os programas de captura de imagens rodando nos computadores 1 e 2 da Figura 4-2 (seção 4.2), recebem as imagens das câmeras esquerda e direita, respectivamente, a uma taxa de até 30 quadros por segundo. A aplicação MAE rodando no computador central se conecta simultaneamente com os programas de captura dos computadores 1 e 2, via *socket*, e envia uma mensagem solicitando uma imagem de cada câmera. Os programas de captura de imagens rodando nos computadores 1 e 2 foram desenvolvidos utilizando duas *threads*, uma responsável pela captura, e outra responsável pela conexão e transmissão de dados, de forma a minimizar a influência da conexão e transmissão de dados na captura das imagens, diminuindo a interdependência entre ambas as *threads*.

A aplicação MAE, no computador central, recebe as imagens direita e esquerda colocando-as nas `input`'s direita e esquerda respectivamente. Para implementar o modelo proposto do sistema visual humano, cada conjunto de células de um mesmo tipo foi agrupado numa mesma camada, implementada como uma `neuron_layer`. Filtros foram utilizados para implementar o processamento que é feito na passagem da informação visual de uma camada para outra. Na Figura 4-4 é mostrado em esquemático da utilização de filtros e `neuron_layer`'s para implementação de um modelo de uma célula simples binocular

numa aplicação MAE. As imagens direita e esquerda são filtradas pelo filtro F1, cada uma gerando uma `neuron_layer` de células simples monoculares, que por sua vez são as entradas do filtro F2 que as soma gerando assim uma nova `neuron_layer` de células simples binoculares.

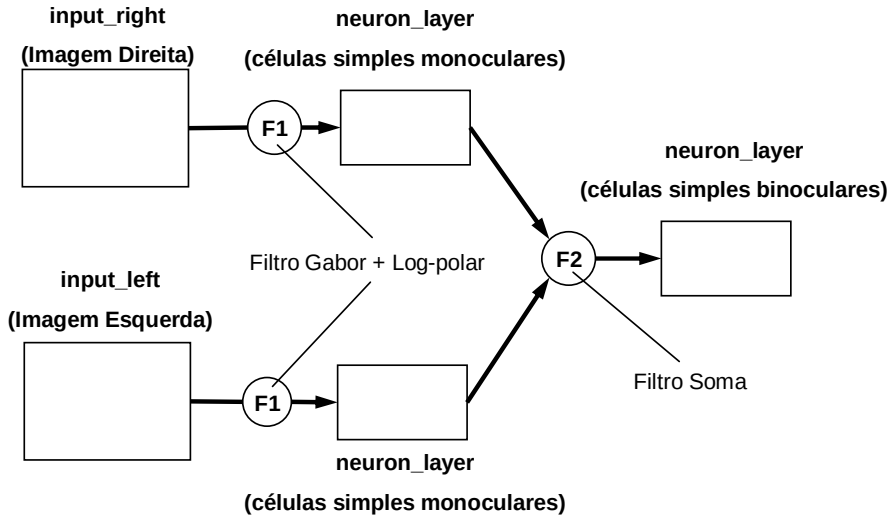

Figura 4-4 - Ilustração da implementação de uma camada de células simples binoculares, utilizando `input`, `neuron_layer` e filtros numa aplicação utilizando a MAE.

Cada imagem capturada possui resolução de 320x240 pixels e 24 bits de cores. Antes de serem transmitidas para a aplicação MAE, as imagens são convertidas de colorida com 24 bits/pixel para monocromática com escala de cinza de 8 bits/pixel. Esta conversão é feita para diminuir o volume de informação a ser transmitido e processado. Esta conversão não gera problemas pois a percepção de cores não é essencial para o processamento de já que cada um destes processamentos ocorre numa região diferente do córtex visual. Enquanto o processamento de percepção de profundidade ocorre principalmente em MT (seção 2.3.5), o processamento de cores ocorre principalmente em V4 (seção 2.3.4).

4.3.3. Escolha do ponto de atenção

É necessário escolher um ponto para qual será destinada a atenção visual, assim como ocorre na visão humana. Para escolher um ponto de atenção (ponto para onde o cérebro humano direcionará os olhos) na imagem projetada nas retinas, o sistema visual humano realiza o movimento de sacada, que é o movimento rápido dos olhos, com cerca de 900º/s, na direção do ponto de atenção (Tabela 2-1).

Conforme discutido na seção 2.2, o *superior culliculus* é a região do cérebro responsável pelo controle dos movimentos de sacadas dos olhos. Entretanto, este não é o foco deste trabalho e, portanto, a escolha dos pontos de atenção é feita manualmente pelo usuário, utilizando o mouse para clicar no ponto desejado da imagem de entrada.

4.3.4. Vergência, disparidade binocular e mapa de disparidades

Realizar a vergência significa encontrar nas duas imagens o mesmo ponto para o qual está voltada a atenção. Neste trabalho, quando é escolhido um ponto na imagem direita, realizar a vergência significa achar o ponto correspondente a este na imagem esquerda. Calcular a disparidade binocular implica em armazenar em cada camada de MT, a medida das diferenças entre as coordenadas dos pontos correspondentes no olho direito e no esquerdo. Construir um mapa de disparidades, significa encontrar, para cada ponto ao redor do ponto de vergência numa imagem, o ponto correspondente na outra imagem, ou seja, para cada ponto na imagem direita, achar o ponto correspondente na imagem esquerda. Tanto para resolver vergência, quanto para construir o mapa de disparidades, é preciso resolver o problema da correspondência (seção 3.2.2).

Após a escolha do ponto de atenção é necessário vergir os olhos para este ponto, ou seja, movimentar os olhos de forma que o ponto de atenção selecionado seja projetado na fóvea de cada retina (seção 2.5). A partir do mesmo processo no qual é feita a vergência, obtém-se as informações necessárias para calcular um mapa de disparidade binocular, que é a representação interna e será utilizado posteriormente como base para a reconstrução tridimensional. Portanto, como estes dois processos são realizados em conjunto, serão discutidos nesta mesma seção.

Imediatamente após a escolha manual de um ponto da imagem direita, automaticamente o foco do olho esquerdo é movimento para a posição de mesmas coordenadas x e y na imagem esquerda. Desta forma, os olhos estão “olhando para o infinito”, na direção do ponto escolhido na imagem direita. A partir deste ponto é feita uma varredura no olho esquerdo, a partir deste ponto inicial, movendo para a direita o foco do olho esquerdo, como mostrado na Figura 4-5.

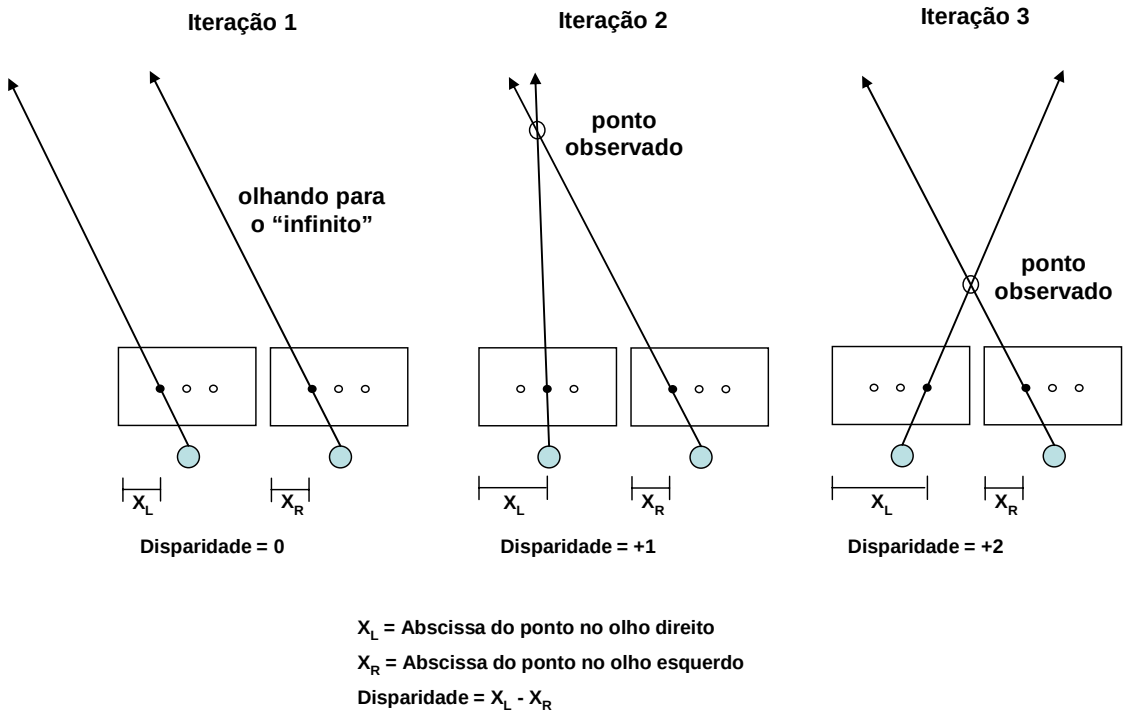

Figura 4-5 - A cada iteração o centro da focalização do olho esquerdo é deslocado um pixel para a direita, enquanto o olho direito fica parado. Desta forma a cada iteração são calculadas informações sobre uma disparidade diferente, ou seja, cada vez focalizando (vergindo os olhos) mais perto.

Em cada iteração é construída uma camada em MT calculando a resposta de cada célula desta camada utilizando a arquitetura proposta na seção 3.4. Se na iteração *n* foram calculadas as respostas das células da camada *d* em MT, na iteração *n+1*, serão calculadas as respostas das células da camada *d+1*, e assim sucessivamente (Figura 3-22). Após o cálculo de todas as iterações, e consequentemente, todas as camadas de MT, é obtida a estrutura de MT de acordo com a Figura 3-20. Note que esta forma seqüencial de computar as camadas de MT é resultado de sua implementação computacional, mas o modelo é totalmente paralelo.

A partir da informação contida na área cortical MT implementada é feita a vergência e construído um mapa de disparidade binocular que será a base para a reconstrução tridimensional. Cada camada da MT implementada é uma matriz com o resultado de um conjunto de células com mesma disparidade, isto é, para uma célula numa camada com uma disparidade *d*, o centro do campo receptivo na imagem direita é definido pelas coordenadas (x_R, y_R), e o centro do campo receptivo na imagem esquerda é dado pelas coordenadas (x_L, y_L), para uma célula nas mesmas condições, porém na camada *d+1*, as coordenadas dos centros dos campos receptivos nas imagens direita e esquerda são dadas pelas coordenadas (x_R, y_R) e (x_L+1, y_L) respectivamente. Na Figura 3-22 é mostrado este deslocamento, com 3 células para cada camada.

Na condição de vergência o ponto visualizado no espaço é projetado na fóvea de cada retina. Como este ponto está projetado em posições correspondentes em cada retina, a disparidade é nula. As células de MT modeladas neste trabalho são *tuned inhibitory*, cujas respostas são mínimas quando um estímulo está sendo projetado em posições equivalentes nos seus campos receptivos direito e esquerdo e, portanto, não há disparidade entre os campos receptivos direito e esquerdo. Dessa forma, para calcular a coordenada de vergência, é calculado o somatório das respostas das células de cada camada em MT, e é selecionada a coordenada (x,y) da camada em MT que possui o menor somatório, ou seja, a camada que possui a melhor resposta para disparidade nula (Figura 4-6 e Figura 4-7).

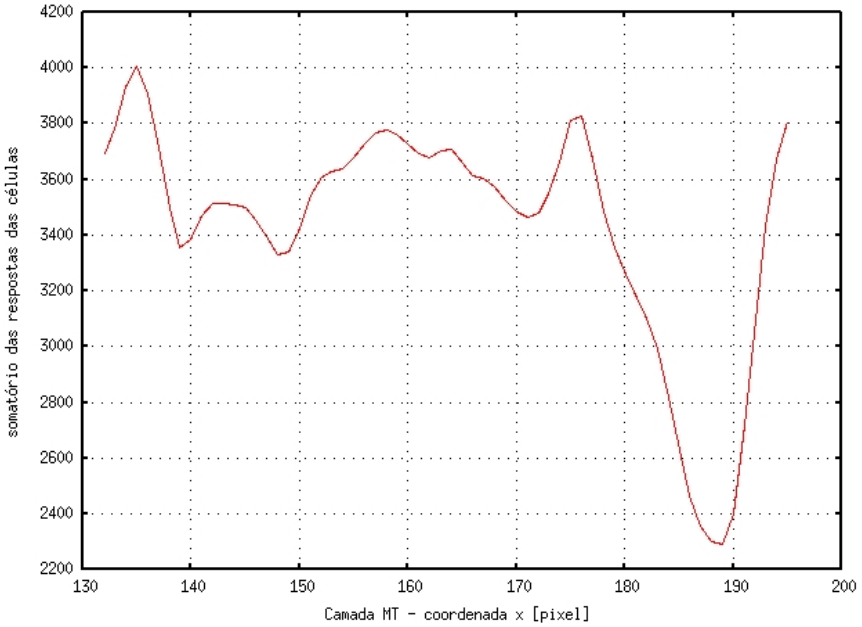

Figura 4-6 - Gráfico mostrando a resposta total de cada camada de MT versus a coordenada (na verdade, a disparidade) associada a cada camada para as imagens de entrada da Figura 4-7 (a disparidade zero é a da coordenada x = 132) . Neste exemplo o ponto de vergência está na coordenada x = 189, ou seja, das 64 camadas de MT modeladas (132 à 195), a camada relativa a coordenada 189 teve o menor somatório da resposta total das células, indicando o ponto de vergência.

Figura 4-7 – Imagens (320x240) direita e esquerda de uma mesma cena. A cruz vermelha indica a posição para onde o olho está olhando, ou seja, o ponto da imagem projetado diretamente na fóvea. (A) Ao escolher um ponto na imagem direita, coordenada [132,135], o olho esquerdo se move para a posição correspondente, coordenada [132,135], fazendo com que o observador esteja "olhando para o infinito". (B) Após o processo de vergência os dois olhos estão orientados de forma que o ponto escolhido esteja sendo projetado diretamente na fóvea de cada retina. O olho direito continua na coordenada [132,135], enquanto que o olho esquerdo moveu para a coordenada [189,135] que é o ponto correspondente ao ponto focalizado pelo olho direito [132,135].

Para construir o mapa de disparidade binocular entre as imagens, o procedimento é semelhante ao da vergência e, além disto, são utilizadas as mesmas informações que foram utilizadas para calcular o ponto de vergência. Na vergência é preciso encontrar a coordenada (o plano de disparidade *d* no modelo MT) no qual toda a imagem possui a menor disparidade (com destaque para a fóvea, devido ao mapeamento log-polar), portanto é selecionada a camada com o menor valor total (menor somatório das respostas das células), enquanto que para construir o mapa de disparidades é preciso encontrar, ponto a ponto, a menor disparidade (Figura 4-8).

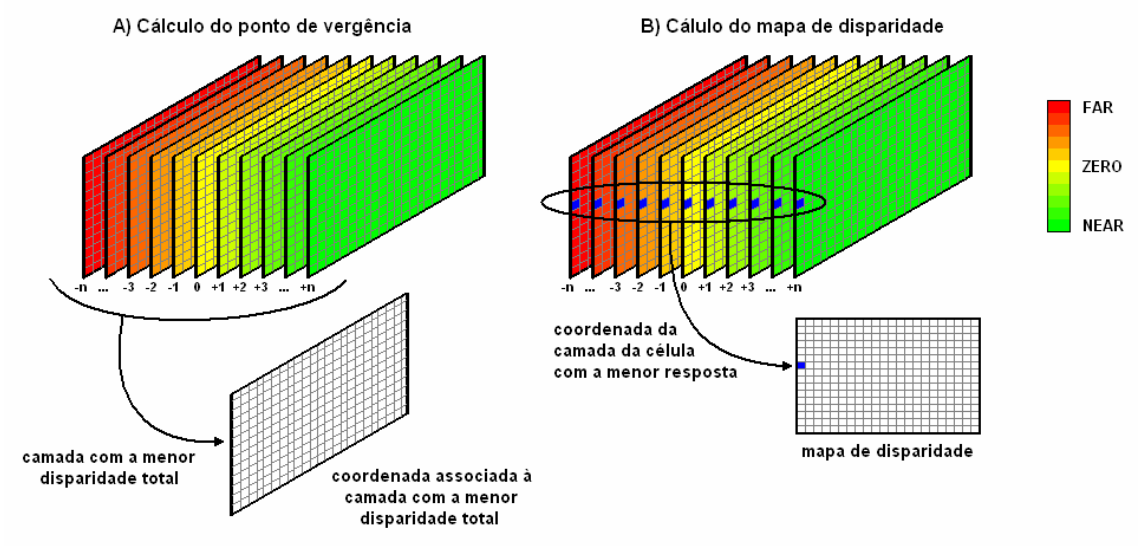

Figura 4-8 - No cálculo da vergência é selecionada a coordenada da camada com a menor disparidade total (a). Na construção do mapa de disparidades, para cada conjunto de células correspondentes entra as camadas, é selecionada a disparidade da célula com a menor resposta (b).

Em cada camada ocorre uma cooperação entre as células, ou seja, a resposta de uma célula é influenciada positivamente pelas respostas das células vizinhas (ver *spatial pooling*, Seção3.3, página 63), enquanto que entre as camadas ocorre uma competição entre as células. Uma ilustração de como ocorre esta interação entre as células, tanto de camadas iguais quanto de camadas diferentes pode ser vista na Figura 4-9.

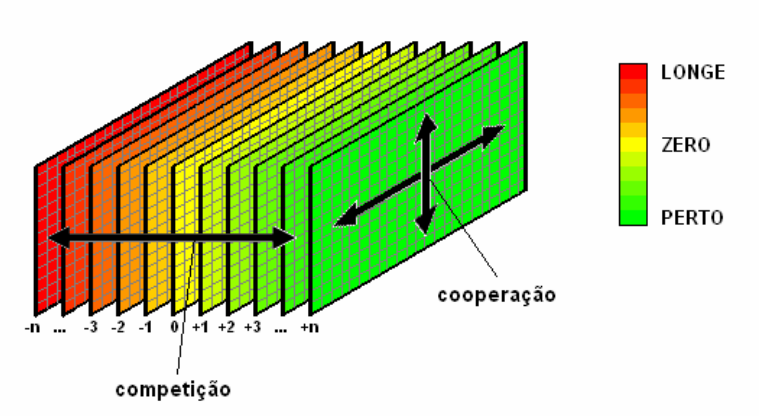

Figura 4-9 - Figura mostrando com interagem as células de MT. Em cada camada, as respostas das células influenciam nas repostas de suas vizinhas (cooperação entre as células). Entre as camadas há uma competição para selecionar tanto cada célula individualmente com a menor disparidade (mapa de disparidade), quanto uma camada inteira com a menor disparidade (vergência).

Após cada iteração do loop que calcula cada camada de MT, o mapa de disparidades (Figura 4-4) é atualizado. Os valores colocados no mapa a cada iteração são as coordenadas x (abscissas), na imagem, da célula em MT com menor valor de saída, ou seja, há uma

competição entre as células. Assim, após o computo de todas as camadas de MT, o mapa de disparidades fica somente com valores positivos (abscissas das coordenadas de cada camada na qual estava a célula com a menor resposta). Esta é a disparidade absoluta entre as imagens. Contudo, a reconstrução tridimensional através da disparidade binocular se baseia na disparidade com relação ao ponto de vergência. Para se obter a disparidade com relação ao ponto de vergência é subtraído de cada ponto no mapa de disparidade o valor da abscissa da coordenada do ponto de vergência, obtendo-se, assim, um mapa de disparidades relativo ao ponto de vergência. Na Figura 4-10 é mostrado o mapa de disparidades antes e depois da compensação da vergência.

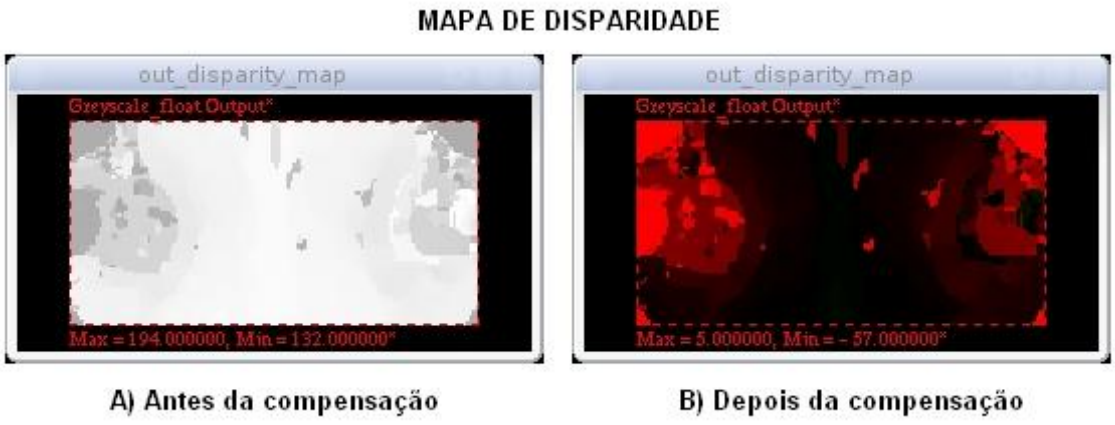

Figura 4-10 - Figura mostrando a compensação feita no mapa de disparidade após a vergência. (A) Antes da vergência, o mapa possui as disparidades associadas as coordenadas de cada uma das 64 camadas de MT, que neste caso variam de 132 à 194. (B) A vergência (neste caso) foi escolhida pela camada de coordenada 189, portanto, todo o mapa foi compensado, subtraindo a coordenada de vergência, onde a disparidade deve ser igual a zero, tendo então disparidades variando de -57 à +5.

4.3.5. Reconstrução da imagem tridimensional

A informação tridimensional da imagem do mundo externo, projetada nas retinas direita e esquerda, está codificada no mapa de disparidade, após todo o processamento descrito na seção anterior. O mapa de disparidades contém a disparidade entre os campos receptivos nas câmeras esquerda e direita de cada célula vencedora de MT.

Na Figura 4-11-A é mostrado como são projetados no espaço 3 pontos consecutivos com a mesma disparidade e, na Figura 4-11-B, é mostrado como um mesmo ponto na imagem direita pode ser projetado em diferentes pontos do espaço dependendo de sua disparidade.

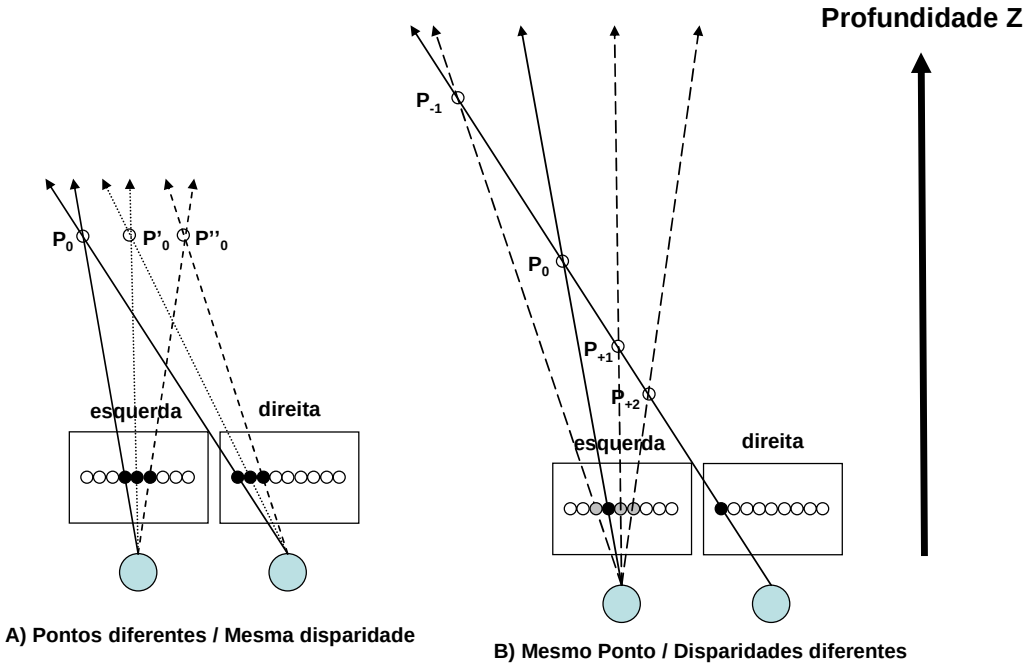

Figura 4-11 – Projeções de pontos diferentes no espaço mas com a mesma disparidade (A). Projeção do mesmo ponto no espaço considerando diferentes disparidades (B).

A reconstrução tridimensional interna ao computador é baseada no mapa de disparidades, no qual toda a informação visual vinda da retina, sofreu uma transformação log-polar. Portanto, para calcular os pontos no espaço tridimensional externo a partir do mapa de disparidades é preciso seguir os seguintes passos para cada ponto do mapa:

1. Realizar o mapeamento log-polar inverso do córtex (mapa) para as retinas esquerda e direita.
2. Adicionar a disparidade associada a este ponto, deslocando horizontalmente o ponto na retina esquerda.
3. Calcular a coordenada espacial deste ponto.

Na Figura 4-12-A, mostra o exemplo de mapeamento log-polar inverso (Córtex → Retina) de um ponto qualquer no mapa de disparidade para as retinas (na verdade, câmeras) sem levar em consideração a disparidade, mas apenas a vergência, isto é, tratando o ponto como tendo disparidade zero, inicialmente. Com a projeção nas retinas, é possível calcular a posição deste ponto no espaço usando a formulação discutida na Seção 4.3.1, obtendo-se assim o ponto P (Figura 4-12-B). Entretanto, de acordo com o mapa de disparidades, a disparidade deste ponto é -1 (Figura 4-12-A). Isto quer dizer que, na verdade, a projeção deste ponto na retina esquerda está deslocada -1 pixel na horizontal (1 pixel para a esquerda). A

projeção na retina direita não é alterada. Desta forma, recalculando a posição do ponto no espaço, é obtido o ponto P' (Figura 4-12-B).

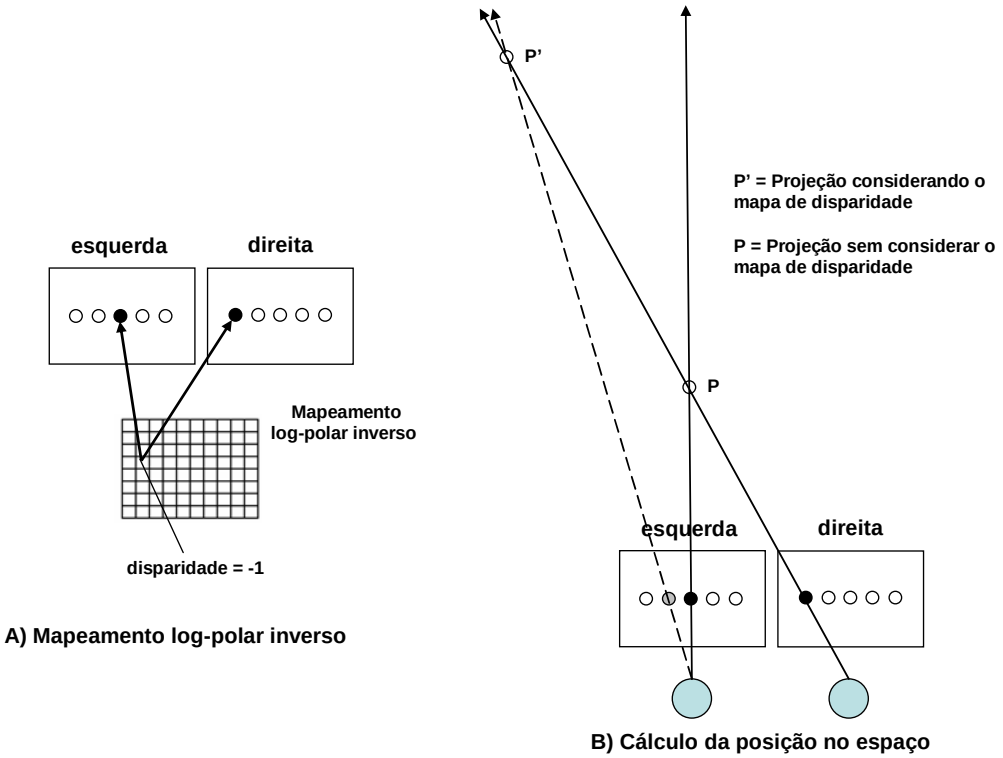

Figura 4-12 – Mapeamento inverso de um ponto no córtex para as retinas direita e esquerda (A). Calculando a posição deste ponto no espaço, é obtido P. O ponto P' é a projeção considerando uma disparidade = -1, no qual determina que a projeção deste ponto na retina esquerda deve ser deslocada um pixel para a esquerda.

4.4. CINCO AMOSTRAS

Numa situação ideal, a disparidade de um ponto no espaço tridimensional corresponde ao ponto de mínimo global desta função. Contudo, num ambiente real, escolher a disparidade associada à célula de MT menos responsiva para um ponto no mapa nem sempre vai corresponder à determinar a disparidade correta para este ponto. Isto pode ser explicado pela diferença de brilho, contraste e foco entre as imagens esquerda e direita, que vão afetar as respostas das células da arquitetura de camadas neurais que culmina em MT, afetando as respostas das células em MT. Vários fatores podem interferir nas relações entre o brilho, contraste e foco de regiões correspondentes na imagem esquerda e direita, como, por exemplo:

- As câmeras utilizadas para captura das imagens, apesar de serem exatamente do mesmo modelo e estarem configuradas da mesma forma podem ter uma sensibilidade diferente à luz.
- Variação na iluminação ambiente
- Diferença na quantidade luz que incide em cada câmera

Na maioria dos casos observados com este problema, a disparidade correta, apesar de não estar associada ao mínimo global, está associada a um mínimo local, com amplitude similar ao mínimo global. Assim, foi utilizada uma heurística no qual a resposta de um ponto no mapa de disparidades tem que estar de acordo com as respostas dos pontos vizinhos.

Esta heurística consiste em, para cada ponto do mapa de disparidade, armazenar um conjunto (*CONJ*) de pares com as informações de intensidade da resposta para uma disparidade (*I(d)*) e esta disparidade associada (*d*) dos cinco menores mínimos da função, podendo assumir como disparidade correta para este ponto, qualquer uma das disparidades de um destes cinco mínimos.

$$CONJ = \{(d_i, I(d_i))\}_{1 \le i \le 5} \tag{19}$$

Inicialmente, a disparidade de um ponto no mapa de disparidades (d_p) é selecionada como sendo igual à disparidade do menor dentre os cinco mínimos armazenados, ou seja, o mínimo global.

$$d_p = d_i \Big| I(d_i) = \min_{1 \le k \le 5} \{I(d_k)\} \tag{20}$$

Em seguida são feitas algumas iterações de forma a fazer com que cada ponto P do mapa de disparidades fique em conformidade com a média dos 8 pontos vizinhos (Figura 4-13).

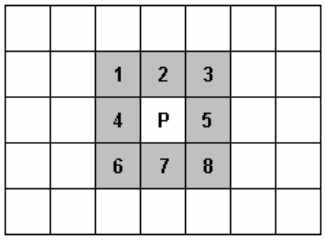

Figura 4-13 - Figura representando os oito pontos vizinhos de um ponto P específico.

Em cada iteração, para cada ponto no mapa de disparidades, é calculada a média (I_M) da intensidade da resposta das oito células vizinhas ao ponto, e escolhida uma nova disparidade, dentre as cinco previamente armazenadas que minimize a distância para esta média.

$$d_p = d_i \left| I(d_i) = \min_{1 \le k \le 5} \left\{ I(d_k) - I_M \right| \right\} \tag{21}$$

Estas iterações são executadas para todo o mapa até atingir a convergência, no qual nenhum ponto muda o valor de uma iteração para outra, ou atingir um número máximo de iterações predefinidas. Na Figura 4-14 é mostrada um mapa de disparidades antes deste processamento e após todas as iterações. É possível verificar visualmente a melhora no mapa de disparidades, após as iterações, no qual várias regiões ruidosas foram eliminadas. Esta heurística equivale a iterações possivelmente existentes na área MT biológica.

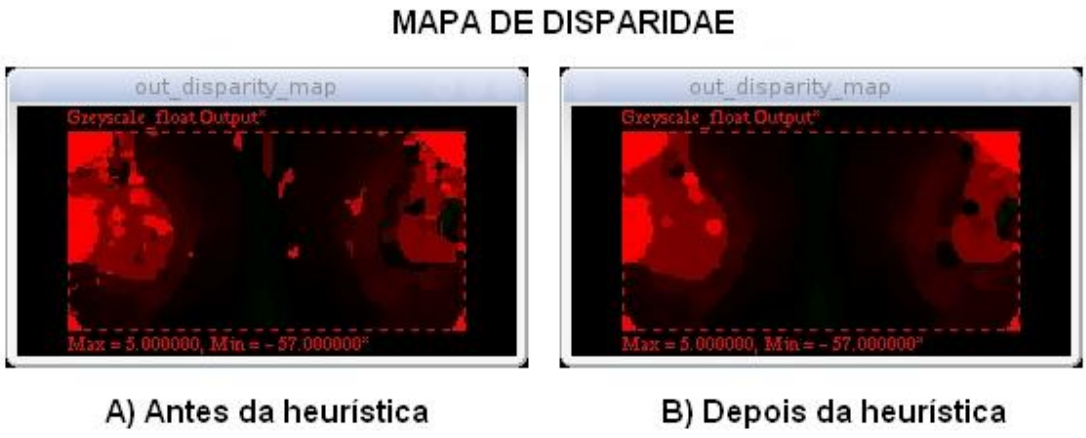

Figura 4-14 - Mapa de disparidade antes e depois de utilizar uma heurística para melhorar a conformidade do mapa de disparidades. É nítida a eliminação de algumas inconformidades do mapa após a utilização da heurística.

4.5. ESTRUTURA DE ARMAZENAGEM DA RECONSTRUÇAO TRIDIMENSIONAL

É preciso armazenar toda a reconstrução tridimensional numa estrutura de dados na qual seja possível utilizar todas as informações para projetar os pontos no espaço e permitir que essa possa ser utilizada, em trabalhos futuros, como base para a navegação do robô

mostrado na Figura 4-2. Quanto mais simples esta estrutura, menor será o esforço computacional para percorrê-la e utilizá-la para a navegação do robô.

Apesar de a reconstrução ser tridimensional, foi utilizada uma estrutura (**estrutura TMap**) bidimensional para armazená-la. Isto porque, quando se olha para um objeto, obtém-se uma informação tridimensional sobre a superfície visível deste objeto, não se podendo afirmar, a princípio, nada acerca do que está por trás desta superfície. Desta forma, a estrutura TMap foi modelada como uma matriz bidimensional de pontos que representa a superfície visível da sua janela de visão, conforme visto na Figura 4-15, no qual cada ponto possui os seguintes atributos:

- Intensidade luminosa
- Distância
- Ângulo horizontal (α)
- Ângulo vertical (β)

Cada ponto na estrutura bidimensional TMap representa um ponto no espaço a uma direção fixa a partir de um ponto central imaginário entre as duas câmeras do robô. A posição do ponto no espaço associado a cada ponto da estrutura é codificada em coordenadas polares, e representada por α e β e distância. A distância é obtida a partir da disparidade e geometria das câmeras, e os ângulos α e β são dados pelas Equações (22), nas quais n_H e n_V são os índices horizontal e vertical em TMap, H e V são as dimensões horizontal e vertical da estrutura TMap, e FOV_H e FOV_V são os ângulos que representam a abertura dos campos de visão horizontal e vertical das câmeras respectivamente (Figura 4-5)

$$\begin{cases} \alpha = -\dfrac{FOV_H}{2} + \dfrac{n_H . FOV_H}{H} \\ \beta = -\dfrac{FOV_V}{2} + \dfrac{n_V . FOV_V}{V} \end{cases} \tag{22}$$

Nesta estrutura é possível armazenar qualquer superfície visível que esteja no campo de visão, bastando que a distância de cada ponto seja ajustada de forma adequada para coincidir com a superfície. No caso em que todos os pontos estão à mesma distância do centro das câmeras, é obtida uma superfície que representa uma casca esférica. Quanto maior a distância de um ponto representado no TMap, menor a sua resolução, assim como ocorre na visão humana.

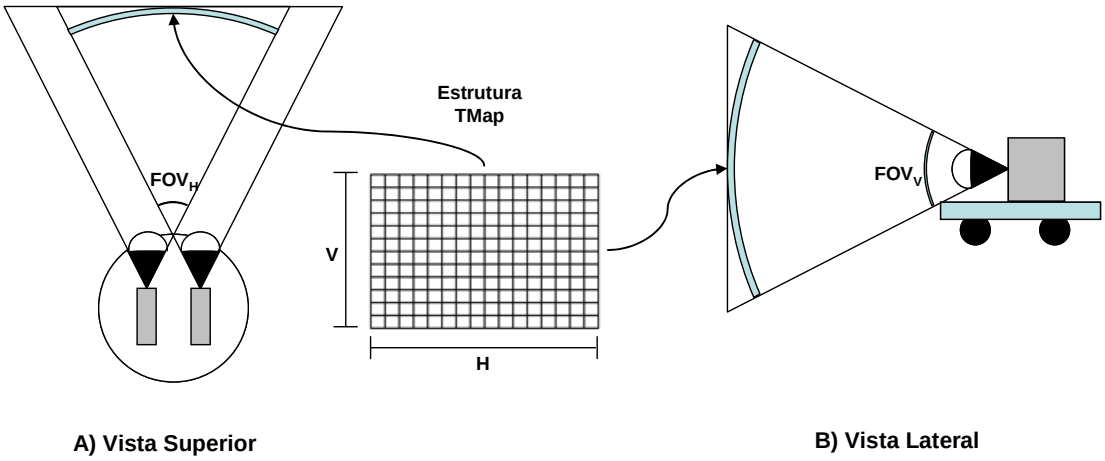

Figura 4-15 - Representação de como a estrutura TMap é utilizada.

A Figura 4-16 mostra um par de imagens estéreo capturadas pelo robô, o mapa de disparidades computado e a reconstrução tridimensional vista de três ângulos.

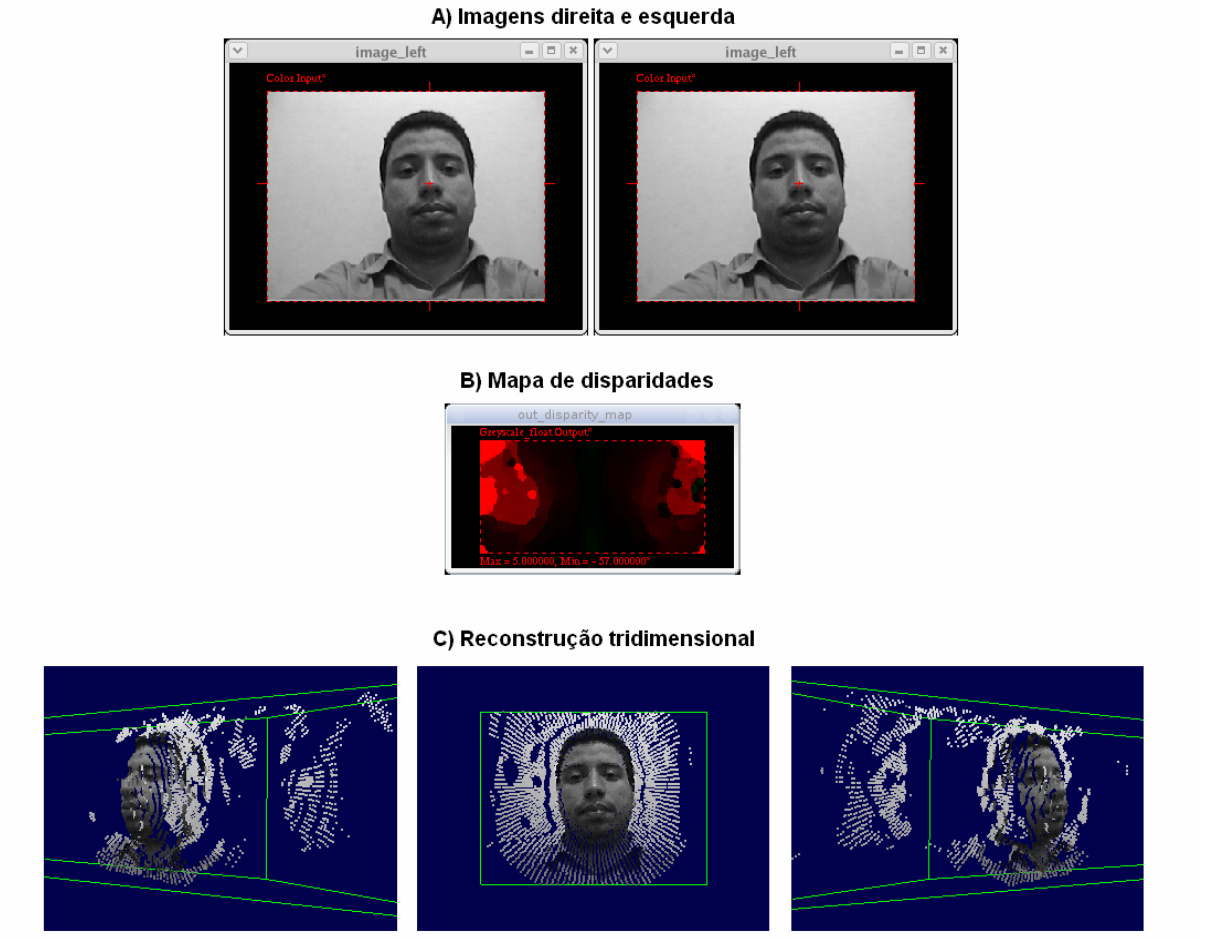

Figura 4-16 – Reconstrução tridimensional a partir de um par de imagens estereoscópicas.

4.6. COMPENSAÇÃO DO EFEITO DA DISCRETIZAÇÃO

A utilização de imagens matriciais em todo este processo leva a um problema de discretização na representação da imagem tridimensional, que afeta fortemente o eixo Z (profundidade) devido à própria discretização da imagem. Este efeito pode ser visto na Figura 4-11 e na Figura 4-12, no qual apenas um pixel de diferença afasta ou aproxima a projeção do ponto no espaço consideravelmente. Isto é agravado pelo fato de que o mapa de disparidades obtido possui somente valores inteiros, uma vez que a disparidade está expressa nele em pixels. O impacto da discretização é tão mais expressivo quanto maior for a distância do ponto no mundo real.

Para reduzir este efeito e produzir representações tridimensionais mais realistas foi utilizada novamente a técnica de *spatial pooling* mostrado na seção 3.3. A Figura 4-17 mostra o resultado da suavização no mapa de disparidade obtida aplicando o *spatial pooling* após sua construção.

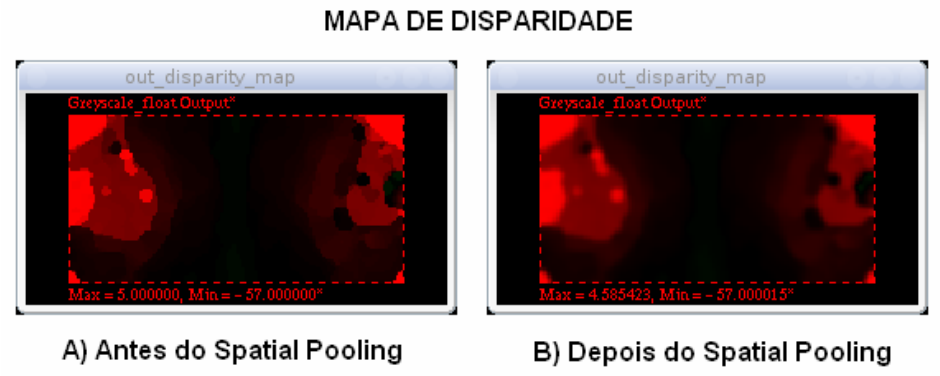

Figura 4-17 - Figura mostrando o resultado da utilização do *spatial pooling* após a construção do mapa de disparidades. As transições entre as disparidades que eram abrutas (apenas disparidades inteiras), são suavizadas pelo *spatial pooling*, semelhante ao efeito de um filtro passa baixa.

A Figura 4-18 mostra um par de imagens estéreo capturadas pelo robô, o mapa de disparidades computado utilizando a técnica de *spatial pooling* e a reconstrução tridimensional vista de três ângulos. É possível notar visivelmente a melhora na qualidade da reconstrução gerada utilizando *spatial pooling*, comparando as figuras Figura 4-16 (sem *spatial pooling*) e Figura 4-18 (com *spatial pooling*).

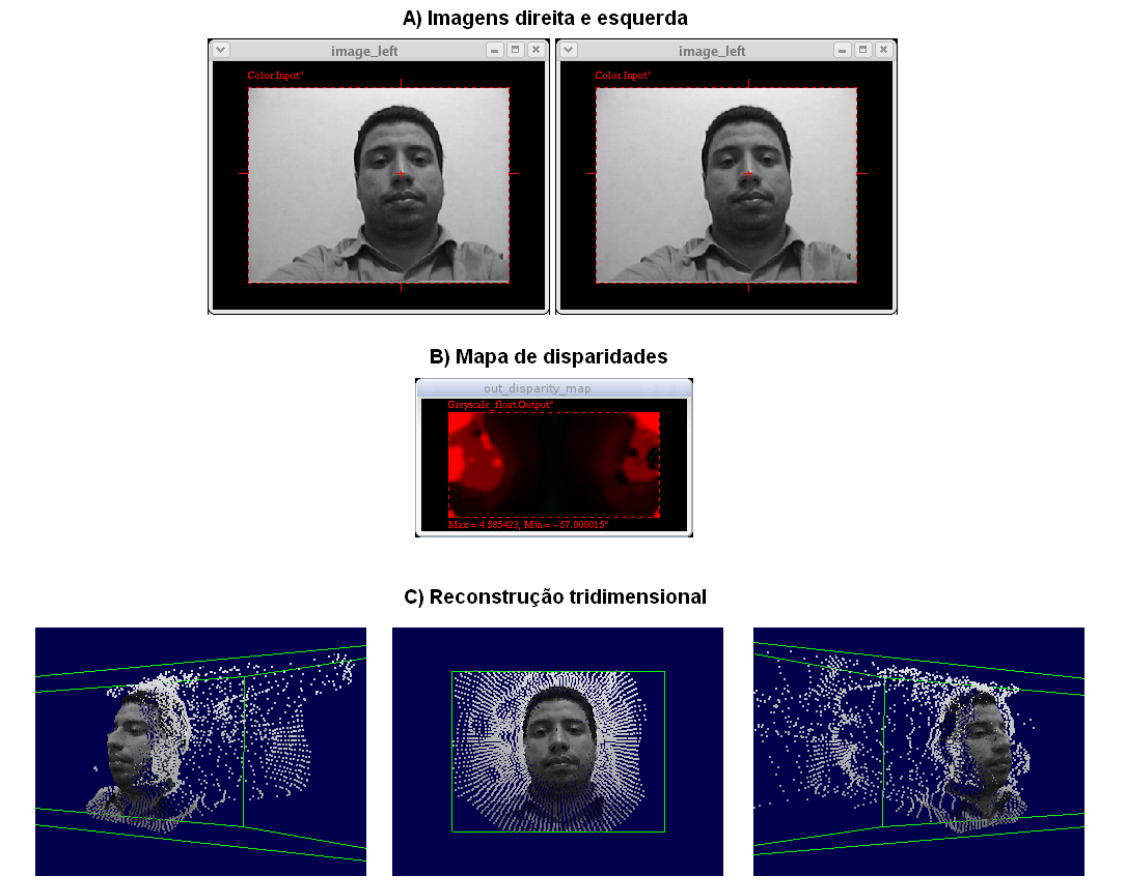

Figura 4-18 - Reconstrução tridimensional a partir de um par de imagens estereoscópicas. As imagens são as mesmas da Figura 4-16. A técnica *spatial pooling* suaviza o mapa de disparidades de melhora significativamente a reconstrução tridimensional.

5. EXPERIMENTOS

Neste capítulo são apresentados os experimentos realizados para analisar tanto as resposta de cada um tipo das células modeladas quanto a resposta da arquitetura proposta como um todo. Também foram realizados experimentos avaliando a influência da freqüência espacial, a variação do parâmetro de seletividade da célula modelada de MT e algumas reconstruções utilizando a arquitetura proposta neste trabalho.

5.1. RESPOSTA DAS CÉLULAS SIMPLES, COMPLEXAS E DE MT

Os tipos de células modeladas (vide Figura 3-21) são:

- S_L - Célula simples monocular com dominância ocular esquerda
- S_L^Q - Célula simples monocular com dominância ocular esquerda, em quadratura de fase com S_L
- S_R - Célula simples monocular com dominância ocular direita
- S_R^Q - Célula simples monocular com dominância ocular direita, em quadratura de fase com S_R
- S_{LR} - Célula simples binocular
- S_{LR}^Q - Célula simples binocular, em quadratura de fase com S_{LR}
- C_L – Célula complexa monocular com dominância ocular esquerda
- C_R – Célula complexa monocular com dominância ocular direita
- C_{LR} - Célula complexa binocular
- MT – Célula da área MT

Para analisar a resposta de cada tipo de célula modelada foram realizados tanto experimentos com as células monoculares quanto com as células binoculares. O experimento com as células monoculares foi feito da seguinte forma: um estímulo com a mesma orientação preferencial das células é movimentado dentro do campo receptivo da célula e, para cada posição do estímulo, é medida a resposta da célula. Foi medida a resposta da célula em função da posição do estímulo tanto para um estímulo claro quanto para um estímulo escuro num fundo cinza. Estes estímulos são mostrados na Figura 5-1.

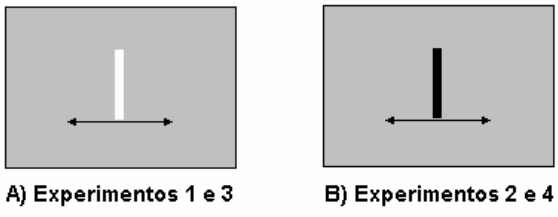

Figura 5-1 - O estímulo é claro nos experimentos 1 e 3 e escuro nos experimentos 2 e 4.

Nos quatro experimentos citados na Figura 5-1, a largura do estímulo é de 0,5º com orientação vertical, a freqüência preferencial das células é de 0,31 ciclos por grau, e o campo receptivo delas têm, também, orientação vertical. O centro dos campos receptivos das células foi posicionado na posição 2,5º (é usual medir-se posições de campos receptivos em graus, que é uma unidade de medida mais apropriada para coordenadas na retina). Nos experimentos 1 e 2, foram avaliadas as células simples monoculares S_L, S_L^Q, S_R, S_R^Q. Seus resultados são mostrados na Figura 5-2 e na Figura 5-3.

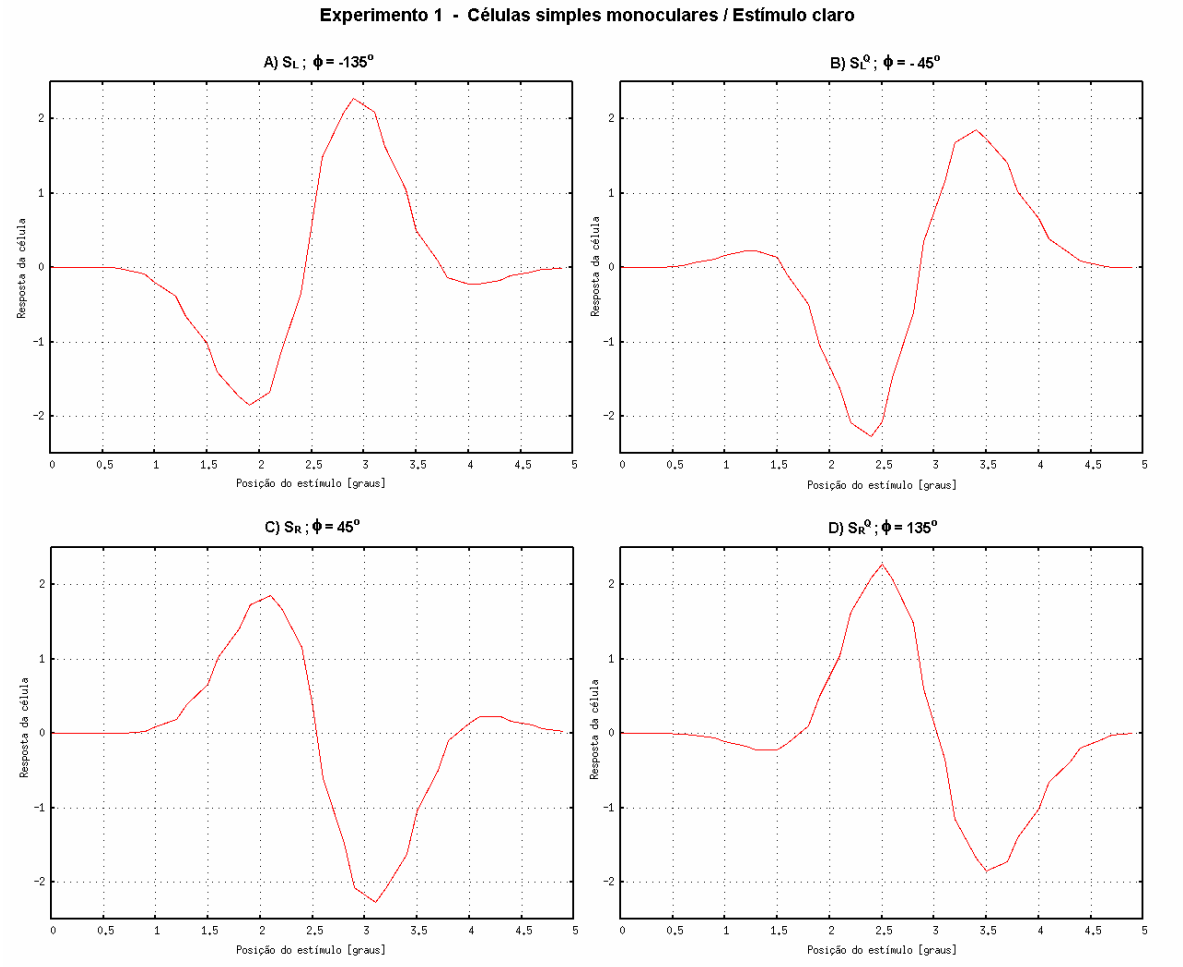

Figura 5-2 - Experimento 1. Gráfico mostrando a amplitude da resposta de células simples monoculares com freqüência preferencial de 0.31 ciclos/grau e diferentes fases em função da posição de um estímulo claro com 0,5º de largura projetado no campo receptivo.

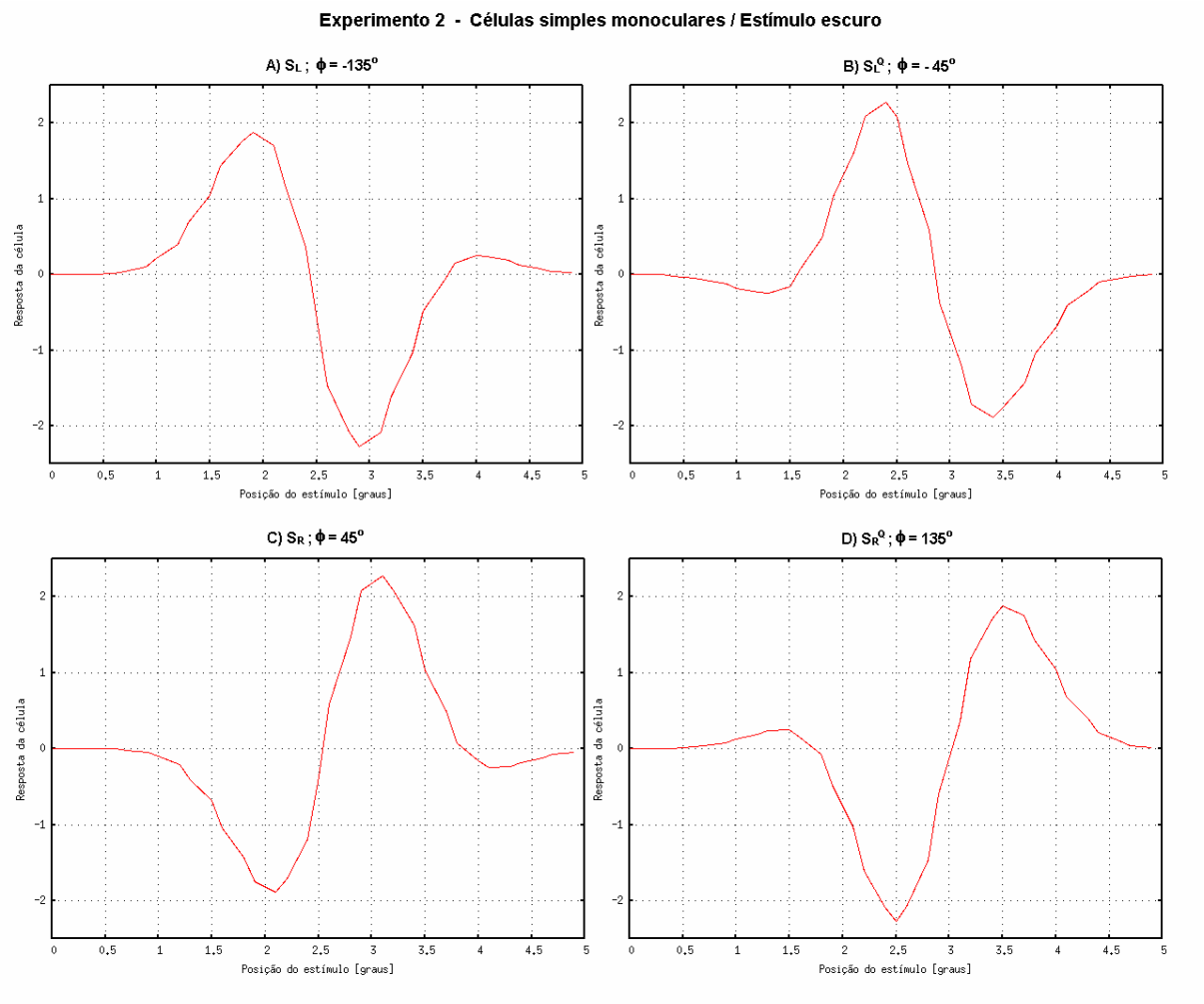

Figura 5-3 - Experimento 2. Gráfico mostrando a amplitude da resposta de células simples monoculares com freqüência preferencial de 0.31 ciclos/grau e diferentes fases em função da posição de um estímulo escuro com 0,5° de largura projetado no campo receptivo.

Como a Figura 5-2 e a Figura 5-3 mostram, a resposta das células simples é influenciada pela fase do estímulo, ou seja, pela posição do estímulo em relação ao seu campo receptivo. Tomando como exemplo o experimento 1, na Figura 5-2-A, quando o estímulo está à esquerda do centro do campo receptivo a resposta da célula é negativa, enquanto que se estiver à direita, a resposta é positiva. Uma outra característica das células simples é que a resposta desta célula em função da posição do estímulo claro é oposta em relação à resposta quando o estímulo é escuro.

Nos experimentos 3 e 4, Figura 5-4 e Figura 5-5 respectivamente, foram avaliadas as respostas de células complexas monoculares C_L e C_R em função da posição do estímulo. A resposta das células complexas monoculares é computada a partir da soma do quadrado de repostas de células simples monoculares em quadratura de fase (vide Figura 3-15 que mostra

de forma esquemática a arquitetura de nossa célula complexa monocular), como é o caso das células A e B, ou C e D na Figura 5-2 e na Figura 5-3. Os experimentos 3 e 4 possuem o mesmo resultado porque a soma do quadrado das respostas das células A e B, ou das células C e D na Figura 5-2 e na Figura 5-3 produzem o mesmo resultado. Também é possível observar que as células complexas monoculares não são muito sensíveis à fase do estímulo.

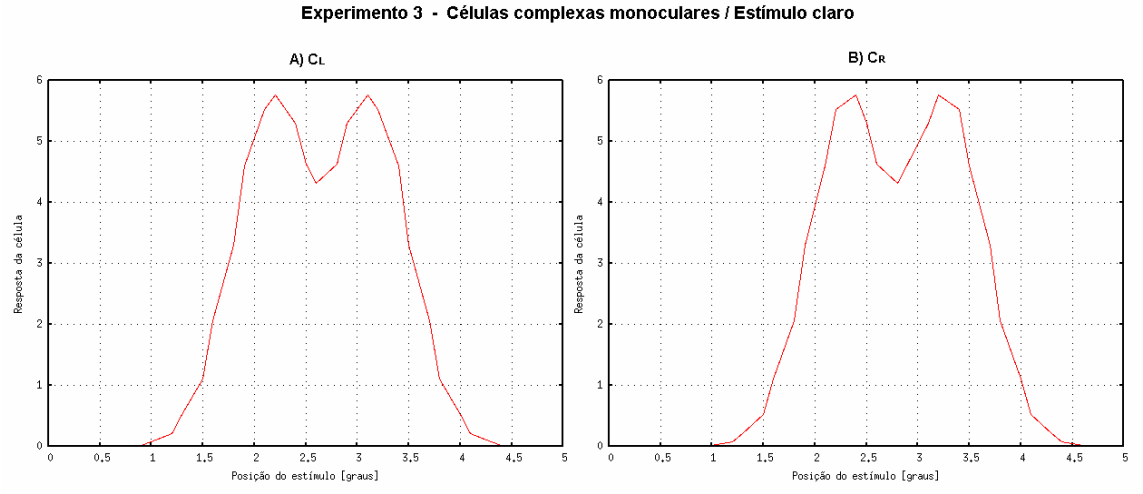

Figura 5-4 - Experimento 3. Resposta de células complexas monoculares com freqüência preferencial de 0.31 ciclos/grau em função da posição de um estímulo claro com 0.5º de largura.

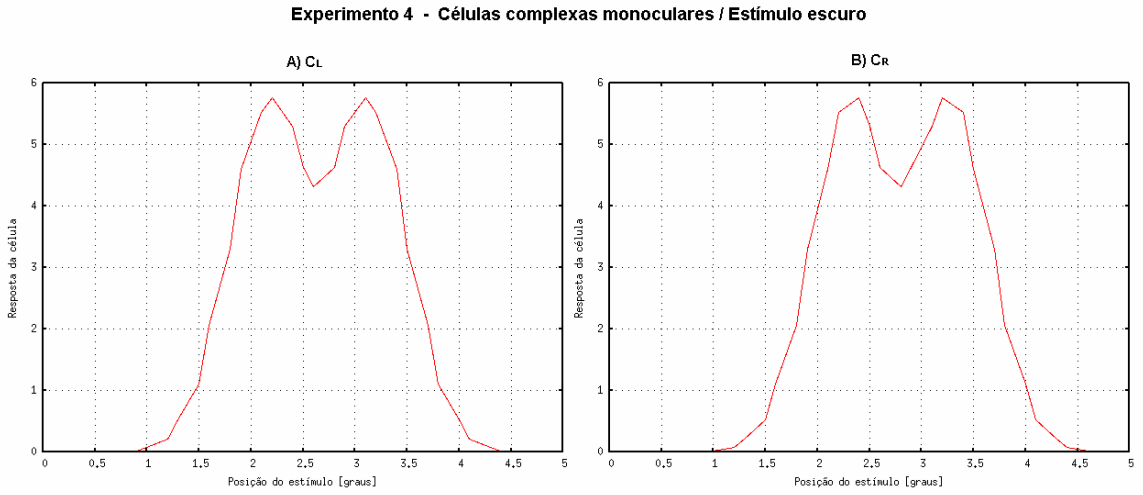

Figura 5-5 - Experimento 4. Resposta de células complexas monoculares com freqüência preferencial de 0.31 ciclos/grau em função da posição de um estímulo escuro com 0.5º de largura.

Para analisar a resposta das células binoculares foram utilizados dois estímulos simultaneamente, um projetado no olho esquerdo e outro no olho direito. Assim como nos experimentos das células monoculares, a largura do estímulo é de 0.5º com orientação

vertical, a freqüência preferencial das células é de 0.31 ciclos por grau, e o campo receptivo está centrado na posição 2,5º.

Foram feitos experimentos com estímulos escuro no campo receptivo do olho esquerdo e escuro no campo receptivo do olho direito, ou estímulo escuro-escuro, assim como estímulos claro-claro, escuro-claro e claro-escuro nos olhos esquerdo e direito, respectivamente. A Figura 5-6 mostra o resultado destes experimentos. Nela, a linha de gráficos superior mostra a resposta da célula S_{LR} aos vários estímulos usados e a linha de gráficos inferior, a resposta da célula S_{LR}^{Q}. Cada coluna de dois gráficos mostra a reposta das células para cada um dos quatro estímulos descritos (escuro-escuro, claro-claro, escuro-claro e claro-escuro) – a legenda acima de cada coluna permite distinguir o estímulo usado em cada coluna. O eixo das abscissas dos gráficos indica a posição do primeiro estimulo do par com relação ao centro do campo receptivo das células no olho esquerdo, enquanto que o eixo das ordenadas indica a posição do segundo estimulo do par com relação ao centro do campo receptivo das células no olho direito. A cor do gráfico indica a intensidade da resposta das células para o par de estímulos em uma dada posição específica; o valor numérico da resposta das células para cada cor aparece na legenda à direita da Figura 5-6.

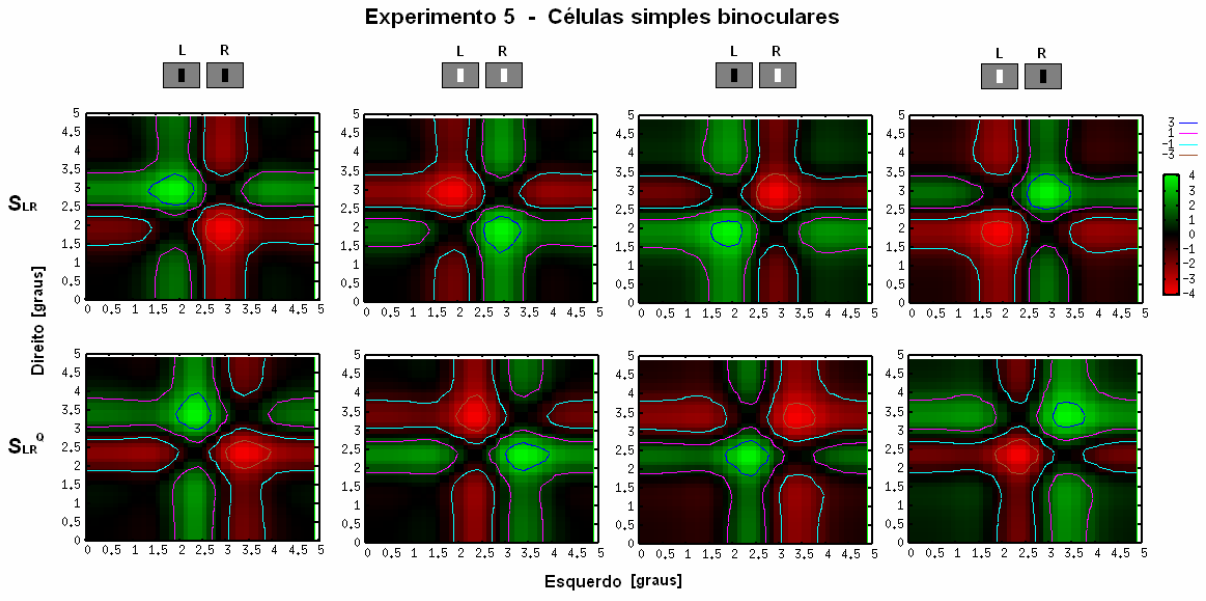

Figura 5-6 - Experimento 5. Gráfico mostrando a resposta de células simples binoculares com freqüência preferencial de 0.31 ciclos/grau, em função da posição de dois estímulos (todas as combinações de claro e escuro), um em cada olho. Os estímulos possuem 0.5º de largura. O eixo horizontal indica a posição do estímulo em graus no campo receptivo esquerdo, enquanto que o eixo vertical indica a posição do estímulo em graus no campo receptivo direito.

Da mesma forma que para as células simples monoculares (Figura 5-2 e Figura 5-3), as células simples binoculares também são influenciadas pela fase do estímulo. Este fato pode ser observado na Figura 5-6, a qual mostra que a resposta da célula pode ser positiva ou negativa, dependendo da posição (fase) do estímulo dentro do seu campo receptivo, tanto esquerdo quando direito. Também é possível observar que, assim como as células simples monoculares, a resposta das células simples binoculares recebendo estímulos escuro-escuro são exatamente o oposto de quando estas mesmas células recebem estímulos claro-claro. Esta oposição também ocorreu no teste com estímulos escuro-claro e claro-escuro. As características das células simples monoculares e binoculares são equivalentes, haja vista que as repostas das células simples binoculares são, na verdade, o resultado da soma das respostas de duas células simples monoculares.

A equivalência de características também ocorre entre as células complexas monoculares e binoculares que, devido ao comportamento de segunda ordem e por possuírem subunidades lineares (células simples) em quadratura de fase (modelo de energia – Seção 3.2.2), a informação sobre a fase (posição) do estímulo dentro do campo receptivo é perdida. Isto pode ser observado no experimento 6 mostrado na Figura 5-7 , que segue o mesmo padrão da figura anteriormente descrita (Figura 5-6).

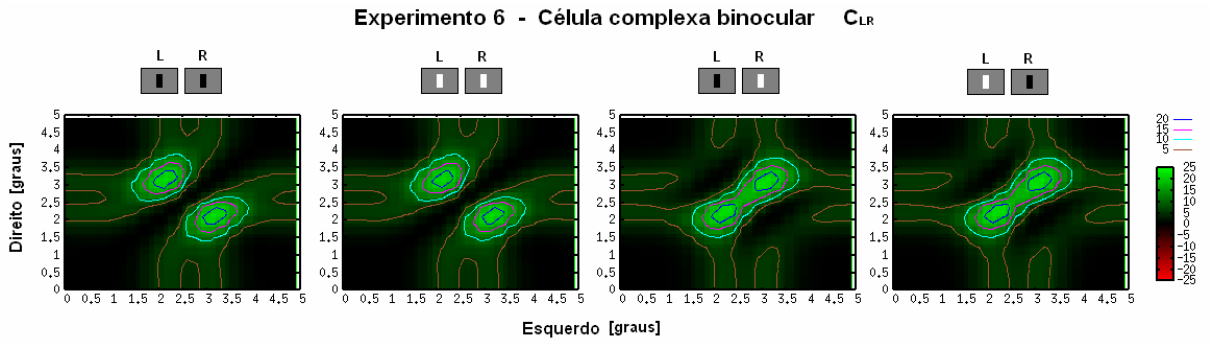

Figura 5-7 - Experimento 6. Resposta de uma célula complexa binocular com freqüência preferencial de 0.31 ciclos/grau em função da posição de estímulos com 0.5º de largura. O eixo horizontal indica a posição do estímulo em graus no campo receptivo esquerdo, enquanto que o eixo vertical indica a posição do estímulo em graus no campo receptivo direito.

No experimento 6 (Figura 5-7) é possível notar a capacidade da célula complexa de detectar disparidade binocular. Esta célula complexa binocular tende a apresentar uma resposta próxima do seu mínimo para estímulos iguais (escuro-escuro e claro-claro) e próxima do seu máximo para estímulos invertidos (escuro-claro e claro-escuro) quando os estímulos direito e esquerdo estão em posição correspondentes nos campos receptivos direito e

esquerdo. Estas posições de correspondência estão localizadas na diagonal dos gráficos da Figura 5-6 e da Figura 5-7 em que os valores das ordenadas e das abscissas são iguais

Na Figura 5-7 a resposta da célula complexa binocular a estímulos iguais (escuro-escuro e claro-claro) é nula na diagonal. Para estímulos invertidos (escuro-claro e claro-escuro), apesar da resposta na diagonal não ser homogênea, existe uma região na diagonal onde a resposta da célula é máxima, indicando disparidade zero. Este comportamento da célula complexa é desejável, já que se assemelha ao comportamento de um detector de disparidade perfeito, conforme proposto por Ohzawa [21]. No gráfico da Figura 5-8, que mostra como seria um detector de disparidade perfeito segundo Ohzawa, a resposta segue um mesmo padrão em toda a diagonal.

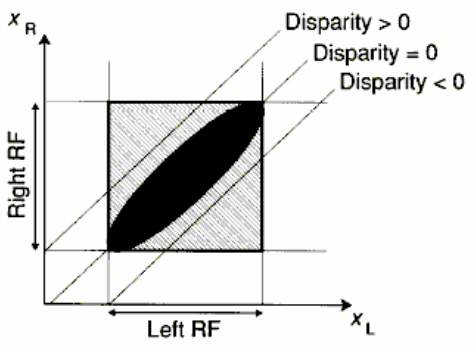

Figura 5-8 – Resposta desejada para um detector de disparidade binocular. Fonte: [21]

Os experimentos citados (experimentos 1 a 6), foram feitos utilizando as células modeladas neste trabalho que são do tipo *tuned inhibitory* (seção 3.2.2). Ohzawa *et al.* [22] analisaram a resposta de células binoculares simples e complexas, porém *tuned excitatory*, do córtex estriado de gatos contra estímulos escuro-escuro, claro-claro, escuro-claro e claro-escuro. Os estímulos possuíam, assim como em nos experimentos mostrados, largura de 0,5º e a freqüência preferencial das células de 0,31 ciclos/grau. Na verdade, os parâmetros utilizados nos experimentos deste trabalho foram os mesmo dos experimentos feitos por Ohzawa *et al.* [22] para facilitar a análise e comparação de resultados.

Na Figura 5-9-A é mostrada a resposta de células complexas binoculares do córtex estriado (camadas 2 e 3) a estímulos projetados em seus campos receptivos direito e esquerdo, enquanto que na Figura 5-9-B, a resposta de células simples binoculares do córtex estriado (camada 4). Quanto mais escuro o gráfico, mais alta é a resposta da célula.

Como visto na Seção 3.2, os modelos propostos de células simples e complexas são *tuned inhibitory*, contudo os dados obtidos por Ohzawa *et al.* [22] (Figura 5-9), são de células

tuned excitatory. Com isto, apenas com o intuito de comparar os modelos propostos com os dados obtidos por Ohzawa *et al.*, foi feito um experimento (experimento 10 na Figura 5-10) eliminando a diferença de fase de 180º entre os campos receptivos direito e esquerdo das células simples binoculares. Desta forma, ao invés de um campo receptivo ser o oposto do outro, se anulando quando o estímulo está na mesma posição nos dois campos receptivos (*tuned inhibitory*), os campos são idênticos, se somando neste caso, fazendo com que a célula fique *tuned excitatory* (Figura 3-17).

Para facilitar a visualização e comparação do modelo apresentado com os dados de células reais apresentados por Ohzawa *et al.* [22], os gráficos do experimento 10, apresentados na Figura 5-10, estão em escala de cinza, sendo que quanto mais escuro o gráfico, maior é a resposta da célula, da mesma forma que os gráficos apresentados por Ohzawa *et al.* [22] na Figura 5-9.

A resposta das células complexas modeladas neste trabalho é bastante semelhante com a resposta das células complexas do córtex V1 de gatos medidas por Ohzawa. Esta semelhança fica evidente comparando os gráficos da Figura 5-10 onde estão as respostas apresentadas pelo modelo proposto, com os gráficos da Figura 5-9 onde estão as respostas medidas por Ohzawa.

É possível notar que a reposta de uma célula complexa ainda não é ideal para detectar disparidades binoculares. Isto pode ser visto comparando tanto as respostas das células complexas modeladas (Figura 5-10) quanto das células complexas biológicas (Figura 5-9) com a resposta de um detector de disparidades perfeito mostrado no gráfico da Figura 5-8. Contudo, será visto adiante que a resposta da célula de MT implementada se assemelha bastante com o detector de disparidades ideal proposto por Ohzawa e mostrado na Figura 5-8.

O mesmo teste feito com as células simples e complexas (experimentos 1 a 6) foi feito com o modelo de células do MT proposto neste trabalho (Seção 3.3). O resultado deste experimento (experimento 8) é apresentado na Figura 5-11. Para estímulos iguais (escuro-escuro e claro-claro), este modelo de células apresenta uma resposta praticamente homogenia na região de disparidade zero (diagonal principal), bem semelhante à resposta desejada para um detector de disparidades (Figura 5-8). Isto faz com que este modelo de célula seja bastante eficiente para detecção de disparidade binocular.

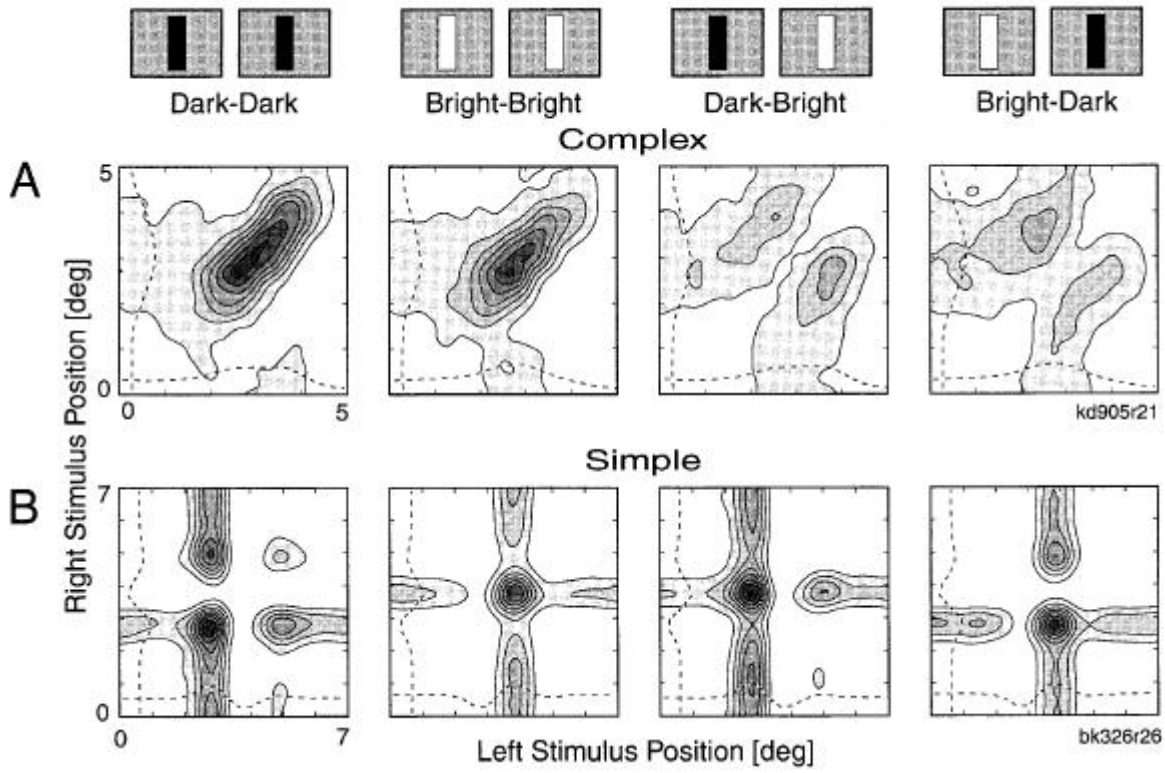

Figura 5-9 - Resposta de células binoculares simples e complexas do córtex estriado (córtex V1) de gatos em função da projeção de estímulos no olho direito e esquerdo. (A) Célula complexa binocular (córtex V1 – camadas 2 e 3). (B) Célula simples binocular (córtex V1 – camada 4). O eixo horizontal indica a posição do estímulo em graus no campo receptivo esquerdo, enquanto que o eixo vertical indica a posição do estímulo em graus no campo receptivo direito. Fonte: [22].

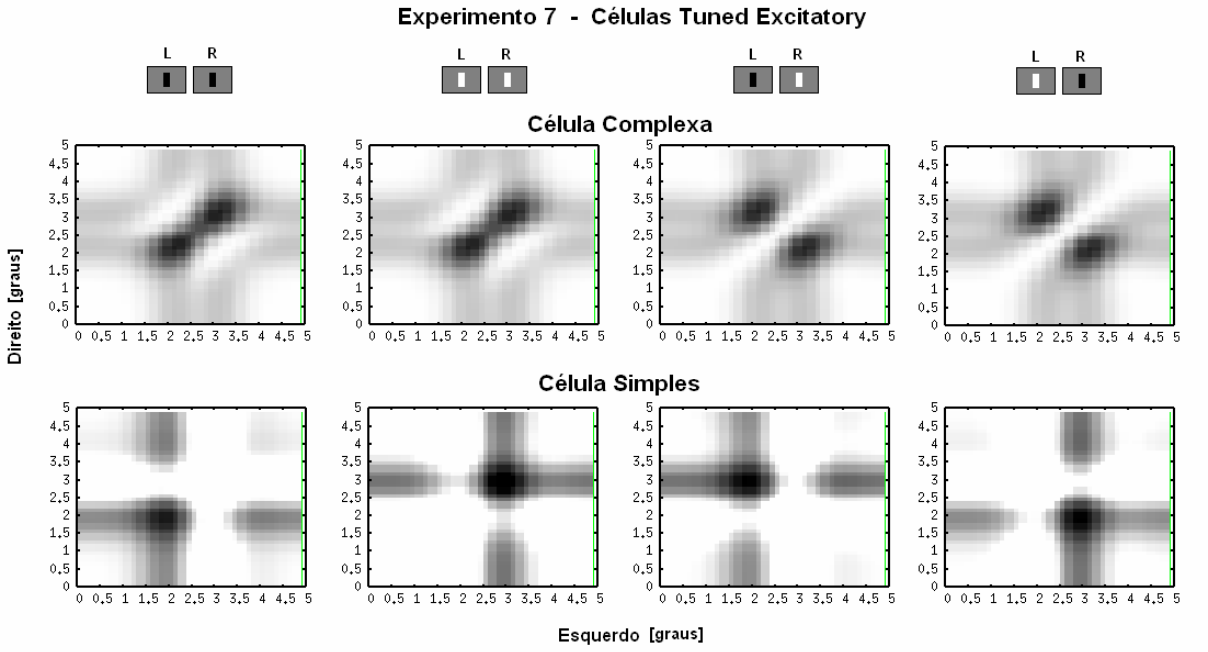

Figura 5-10 – Experimento 7. Resposta do modelo proposto para células binoculares simples e complexas *tuned excitatory*. A resposta do modelo se assemelha bastante com a resposta das células do córtex estriado de gatos, medidas por Ohzawa *et al.* [22] e apresentada na Figura 5-9.

Nas Figura 5-11-A e Figura 5-11-B são apresentadas as respostas do modelo de células de MT sem e com a utilização do *spatial pooling*, respectivamente. A diferença nos gráficos apresentados é sutil, entretanto esta diferença é bastante visível na reconstrução tridimensional da imagem. Analisando somente as respostas para estímulos iguais, pelo fato deste modelo ter sido feito com células complexas *tuned inhibitory*, a região da diagonal onde a disparidade é zero, é a região com a menor reposta. No modelo de célula de MT sem *spatial pooling* (Figura 5-11-A), as bordas da diagonal oposta também possuem valores baixos, podendo inclusive ser mais baixo que a resposta correta na diagonal onde a disparidade é zero. Já no modelo com *spatial pooling* (Figura 5-11-B), esta região com valores baixos na diagonal oposta é bastante reduzido, melhorando substancialmente a reconstrução tridimensional.

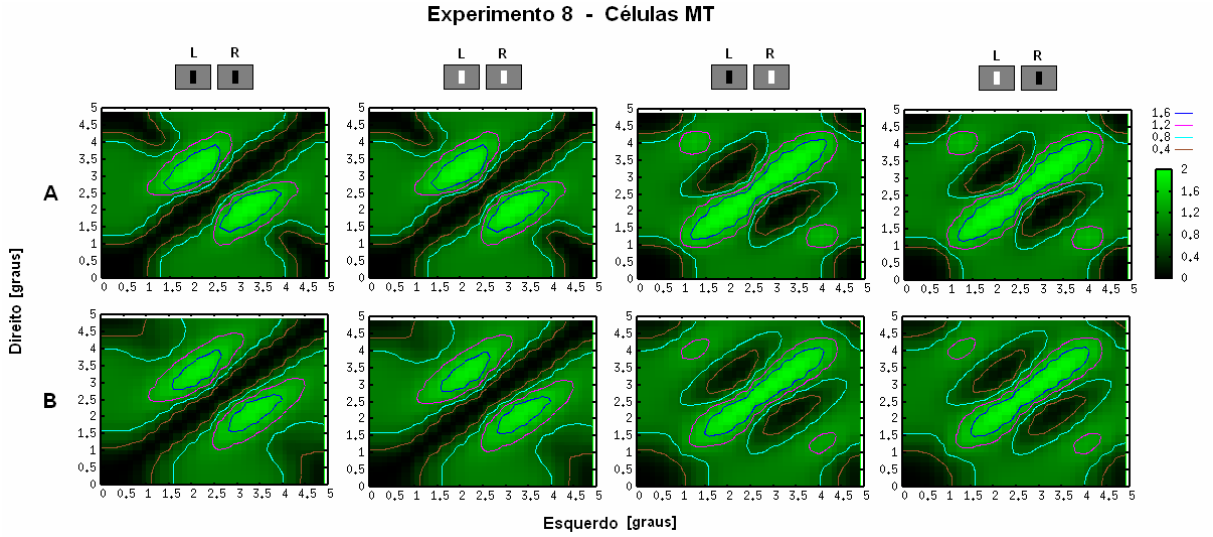

Figura 5-11 - Experimento 8. Resposta apresentada pelo modelo proposto de células de MT em função da posição de estímulos com 0.5º de largura. (A) Resposta do modelo de célula MT sem *spatial pooling*. (B) Resposta do modelo de célula MT utilizando *spatial pooling*. O eixo horizontal indica a posição do estímulo em graus no campo receptivo esquerdo, enquanto que o eixo vertical indica a posição do estímulo em graus no campo receptivo direito.

Pode ser observado que a reposta destas células vai se aproximando o cada estágio (simples - Figura 5-6 → complexa - Figura 5-7→ MT - Figura 5-10) da resposta ideal de um detector de disparidades (Figura 5-8). Um outro ponto a ser observado é que as células do tipo *tuned inhibitory* aparentam ser mais eficientes na detecção de disparidade binocular do que as células *tuned excitatory* devido a uma maior semelhança à resposta ideal de um detector de disparidades do que a resposta das *tuned excitatory* (vide Figura 5-8).

5.2. RECONSTRUÇÃO

A seguir são mostrados dois exemplos de reconstrução tridimensional utilizando as informações dos mapas de disparidades produzidos pelo processamento da arquitetura proposta neste trabalho com freqüência preferencial igual a 0,85 ciclos/grau, largura de banda igual a 2 (duas oitavas) e seletividade MT igual a 0,1.

A Figura 5-13 possui uma reconstrução tridimensional, vista de diversos ângulos, das imagens bidimensionais da Figura 5-12. As cruzes vermelhas nas imagens da Figura 5-12 representam os locais onde foram realizadas as vergências.

Figura 5-12 - Imagens direita e esquerda de um rosto.

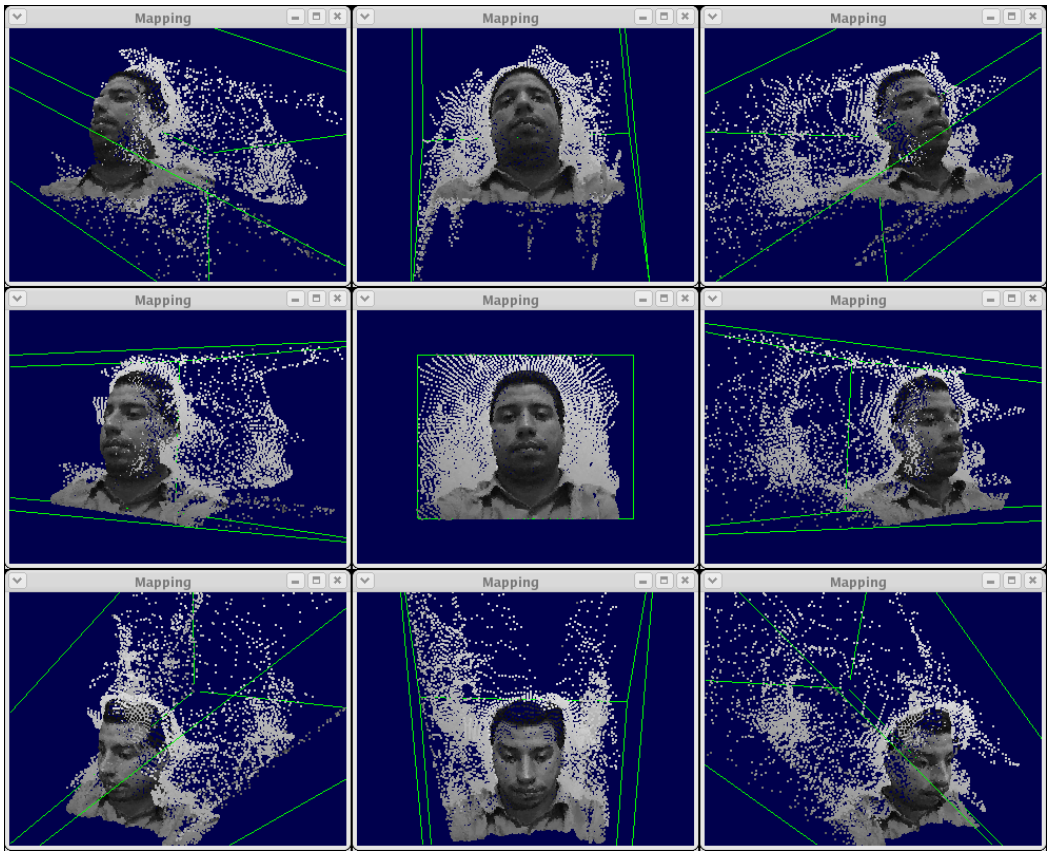

Figura 5-13 - Reconstrução tridimensional das imagens da Figura 5-12.

A Figura 5-15 possui uma reconstrução tridimensional, vista de diversos ângulos, das imagens bidimensionais da Figura 5-14. As cruzes vermelhas nas imagens da Figura 5-14 representam os locais onde foram realizadas as vergências.

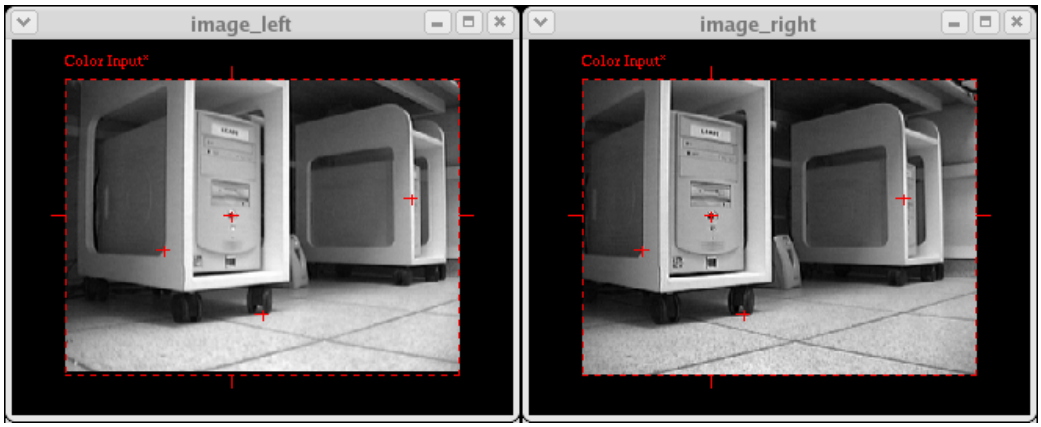

Figura 5-14 - Imagens direita e esquerda de dois gabinetes de computador.

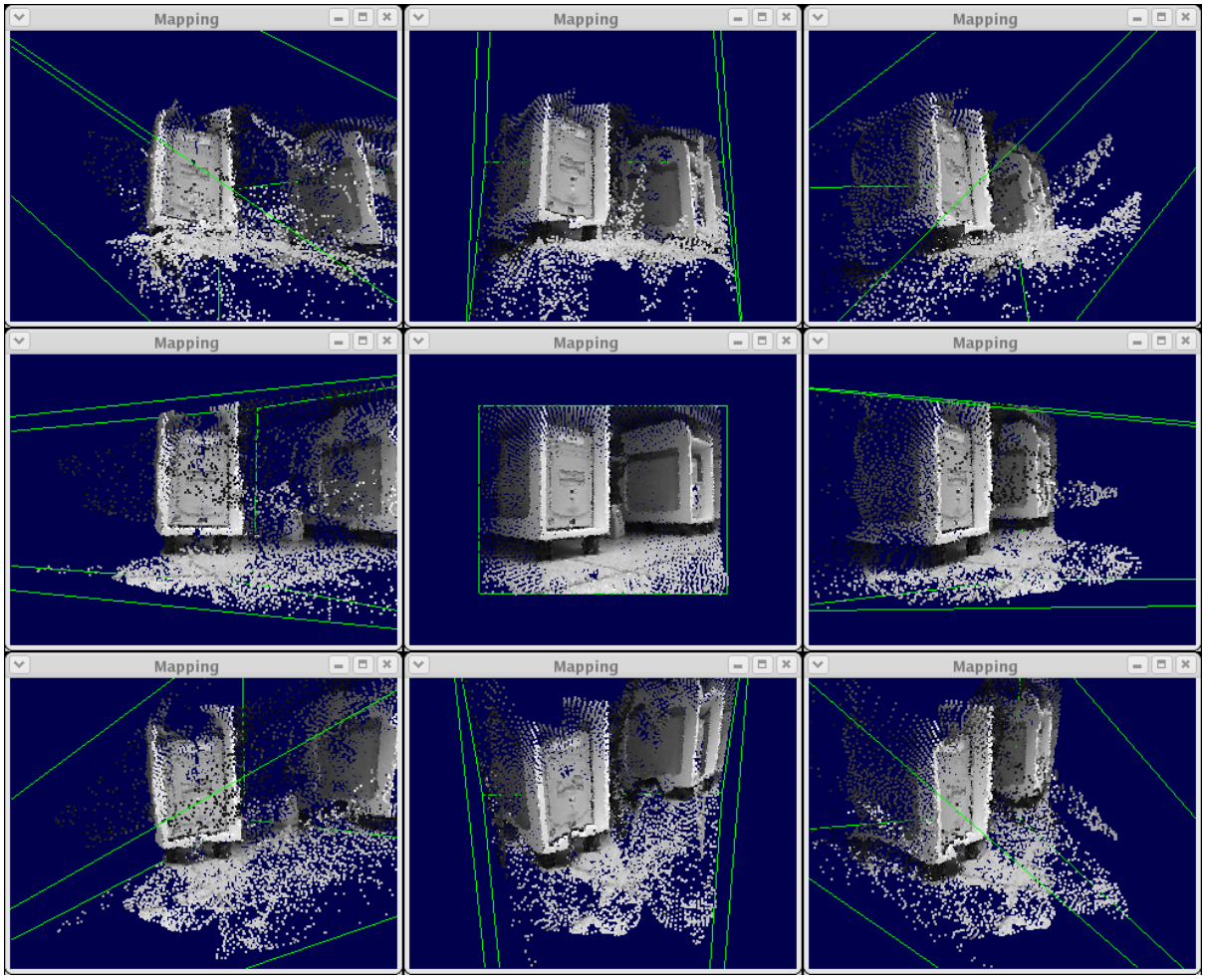

Figura 5-15 - Reconstrução tridimensional das imagens da Figura 5-14.

5.3. FREQÜÊNCIA PREFERENCIAL

A freqüência preferencial dos campos receptivos das células influencia diretamente na qualidade da representação interna criada, e consequentemente na qualidade da reconstrução tridimensional. Foram feitos alguns testes de reconstrução tridimensional, variando esta freqüência preferencial dos campos receptivos das células que compõe a arquitetura proposta. Pelo fato da reconstrução tridimensional ser apenas uma forma de visualizar a representação interna, quanto melhor a qualidade da reconstrução, melhor a qualidade da representação.

As mesmas imagens utilizadas para realizar os experimentos citados na Seção 5.2 (Figura 5-12 e Figura 5-14), foram também utilizadas para realizar os experimentos de avaliação da influência da freqüência preferencial das células da arquitetura proposta na qualidade da representação interna criada.

Nos dois experimentos realizados os parâmetros fixados foram a largura de banda e a seletividade das células de MT. Neste caso a largura de banda é de 2 oitavas e a seletividade das células de MT é 0,1. Para cada par de imagem foram feitos testes com três canais de freqüência diferentes. Os canais de freqüência testados foram 0,42 ciclos/grau, 0,85 ciclos/grau e 1,70 ciclos/grau.

É possível notar em ambos os testes (Figura 5-16 e Figura 5-17) que ao utilizar a freqüência de 1,70 ciclos/grau como freqüência preferencial a reconstrução ficou com muito ruído de alta freqüência. Para o caso no qual a freqüência preferencial é de 0,85 ciclos/grau este ruído de alta freqüência diminuiu. Já para o caso no qual a freqüência preferencial é de 0,42 ciclos/grau este ruído de alta freqüência diminuiu mais ainda, entretanto a superfície reconstruída começou a ter leves deformações.

Quanto menor a freqüência preferencial da célula, maior é o seu campo receptivo cobrindo uma área maior da imagem. Isto diminui o erro relativo introduzido pela discretização da imagem, diminuindo os ruídos de alta freqüência. Contudo, como o campo receptivo fica maior, estímulos com uma freqüência espacial um pouco mais alta (quinas e bordas) não são muito bem detectados por estes campos receptivos. Devido a isto, para tentar obter uma reposta razoável sem comprometer a detecção de bordas e quinas, a freqüência preferencial adotada foi a de 0,85 ciclos/grau.

A Figura 5-16 mostra o resultado do teste de variação da freqüência preferencial dos campos receptivos das células da arquitetura proposta para imagens estereoscópicas de um rosto. As cruzes vermelhas representam os locais onde foram realizadas as vergências.

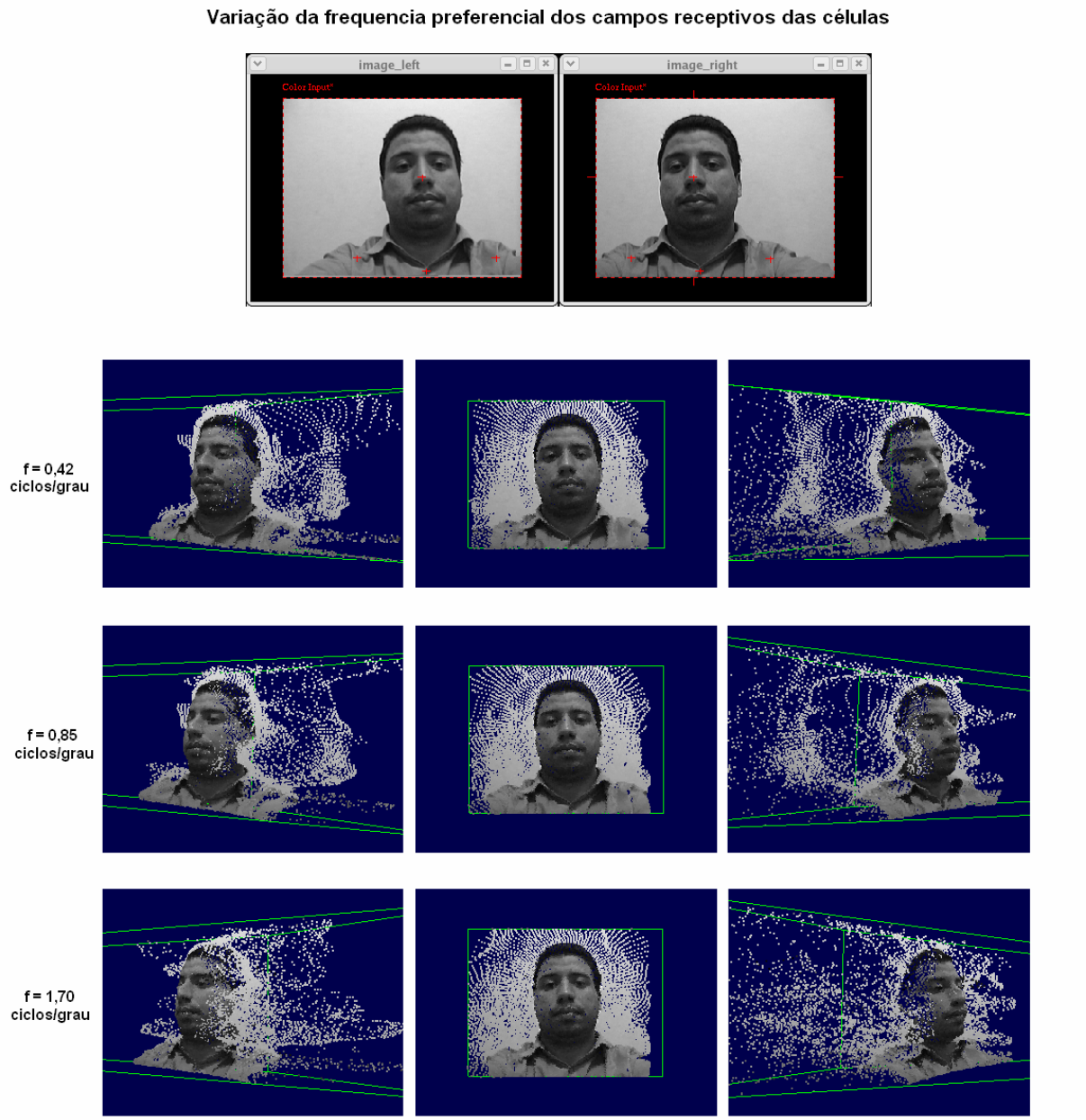

Figura 5-16 – Teste de variação da freqüência preferencial dos campos receptivos das células da arquitetura proposta em imagens estereoscópicas de um rosto. As freqüências preferenciais testadas foram f=0,42 ciclos/grau, f=0,85 ciclos/grau e f=1,70 ciclos/grau.

A Figura 5-17 mostra o resultado do teste de variação da freqüência preferencial dos campos receptivos das células da arquitetura proposta para imagens estereoscópicas de dois gabinetes de computador. As cruzes vermelhas representam os locais onde foram realizadas as vergências.

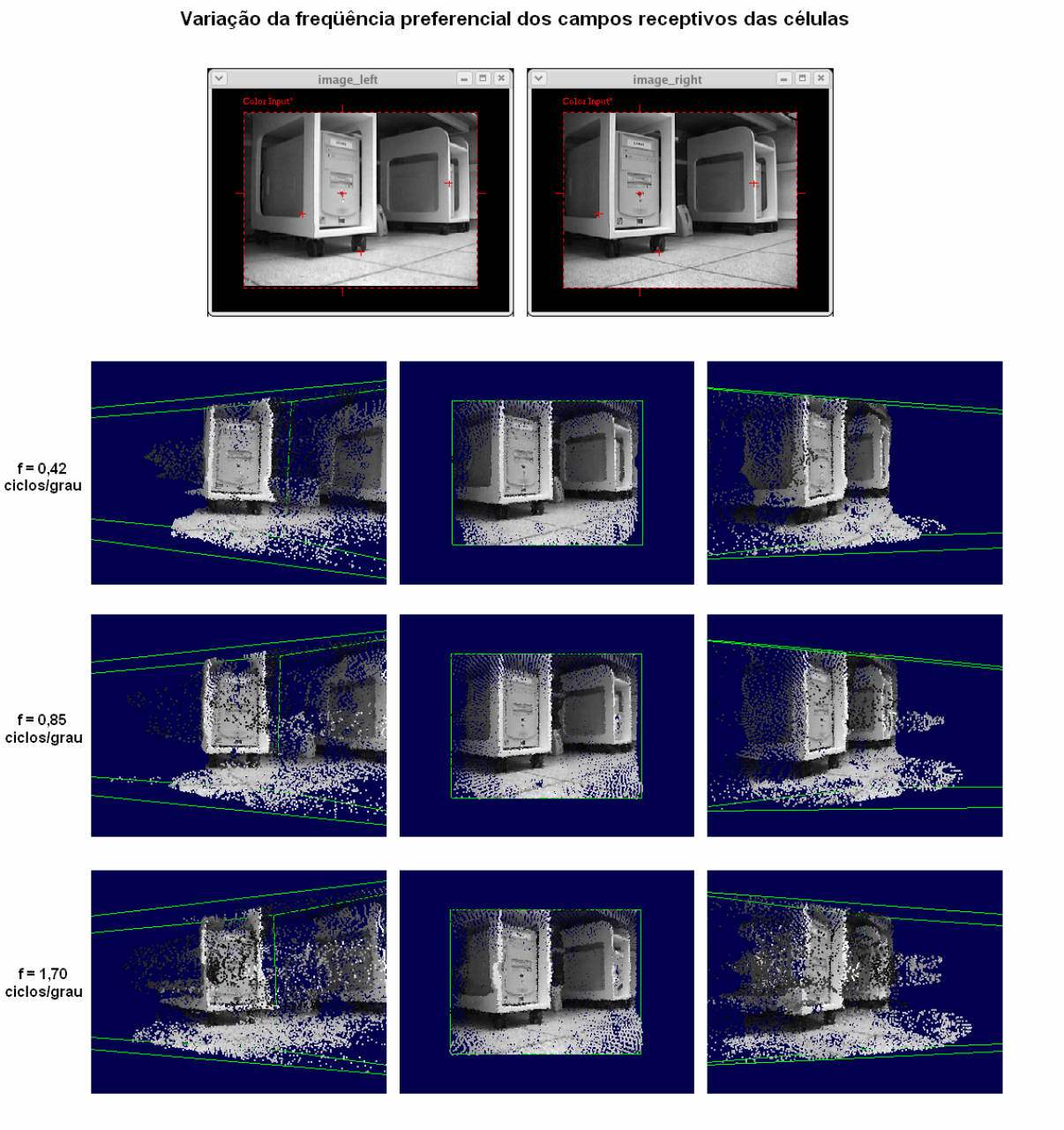

Figura 5-17 - Teste de variação da freqüência preferencial dos campos receptivos das células da arquitetura proposta em imagens estereoscópicas dois gabinetes de computador. As freqüências preferenciais testadas foram f=0,42 ciclos/grau, f=0,85 ciclos/grau e f=1,70 ciclos/grau

5.4. SELETIVIDADE DA CÉLULA MT

No experimento 9 (Figura 5-18), foi analisada a resposta da célula de MT modelada em função da variação do parâmetro *k* de seletividade, conforme descrito na equação (14). Para este experimento foram utilizados somente estímulos escuro-escuro. O tamanho dos estímulos e a freqüência espacial preferencial da célula são as mesmas utilizadas nos experimentos anteriores. O parâmetro foi variado em escala logarítmica de 0,001 a 100.

A seletividade deste modelo proposto de célula de MT é inversamente proporcional ao parâmetro *k*. Para valores de $k >= 1$ a resposta da célula não foi satisfatória, já que foi gradualmente afastando da resposta do modelo ideal para um detector de disparidades. Para valores de $k <= 0{,}1$ a resposta da célula semelhante a resposta ideal, entretanto, foram feitos alguns testes de reconstrução para estes valores de k, e o melhor resultado foi obtido para $k = 0{,}1$. Isto se deve ao fato que, quando a seletividade da célula aumenta demasiadamente, ela pode não responder bem a um estímulo com uma ligeira variação de intensidade luminosa, ou brilho, ou contraste. Neste caso o estímulo deveria ser exatamente igual nos dois olhos, fato que pode não ocorrer devido às variações citadas entre as imagens.

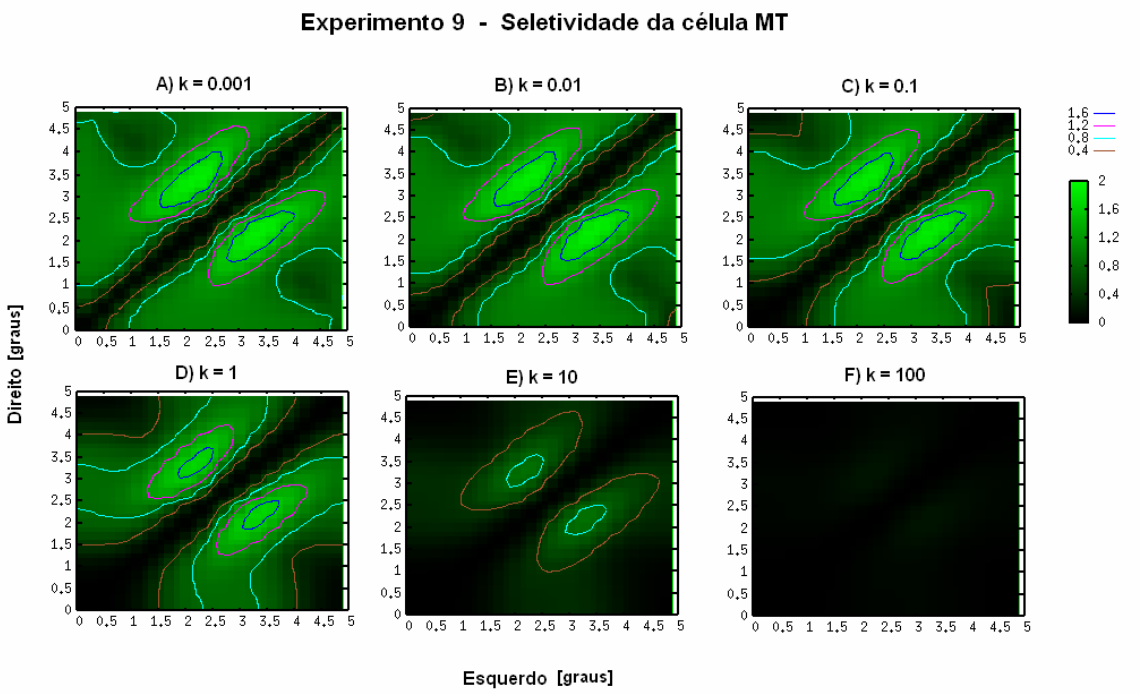

Figura 5-18 - Experimento 9. Resposta apresentada pelo modelo proposto de células de MT proposto neste trabalho para diversos valores do parâmetro de seletividade em função da posição de estímulos escuro-escuro com 0.5º de largura. O eixo horizontal indica a posição do estímulo em graus no campo receptivo esquerdo, enquanto que o eixo vertical indica a posição do estímulo em graus no campo receptivo direito.

6. DISCUSSÃO

Neste capítulo é apresentada uma análise comparativa deste trabalho de pesquisa com dois outros trabalhos correlatos. Um deles busca a plausibilidade biológica, assim como o modelo proposto neste trabalho, utilizando filtros Gabor e células complexas, enquanto o outro segue uma abordagem puramente matemática baseada em transformadas FFT – *Fast Fourier Transform* (Transformada Rápida de Fourier). Em seguida a esta análise comparativa é apresentada uma análise crítica deste trabalho de pesquisa.

6.1. TRABALHOS CORRELATOS

No trabalho realizado por Qian e Zhu [24] e [25] é proposto um modelo de célula complexa para computar disparidade binocular composto por subunidades de células simples em quadratura, entretanto sem a diferença de fase 180º entre os perfis dos campos receptivos direito e esquerdo. Sem esta diferença, as células complexas são *tuned excitatory*, nas quais a resposta é menos eficiente para detecção de disparidade binocular (Seção 5.1). Qian e Zhu utilizaram imagens artificiais para testar seu modelo; o caso, estereogramas constituídos de pontos aleatórios nos quais a parte central é deslocada ligeiramente para um lado e a periferia para outro (Figura 6-1). Por outro lado, neste trabalho de pesquisa foram utilizadas imagens de cenas reais.

O modelo proposto por Qian e Zhu contempla somente células simples e complexas, ou seja apenas o córtex V1, entretanto a percepção de profundidade é iniciada em V1 e continuada na área MT (seções 2.3.5 e 3.3). Na arquitetura proposta por Qian e Zhu, para calcular a disparidade de um ponto é realizado um cálculo com duas células complexas binoculares, com campos receptivos centrados nos mesmos pontos das retinas direita e esquerda, utilizando a diferença de fase entre elas e a resposta de cada uma das células ao estímulo verificado. Já na arquitetura proposta neste trabalho, para calcular a disparidade de um ponto são utilizadas uma célula complexa binocular em conjunto com duas células simples monoculares (Seção 3.3), todas com campos receptivos centrados nos mesmos pontos nas retinas direita e esquerda.

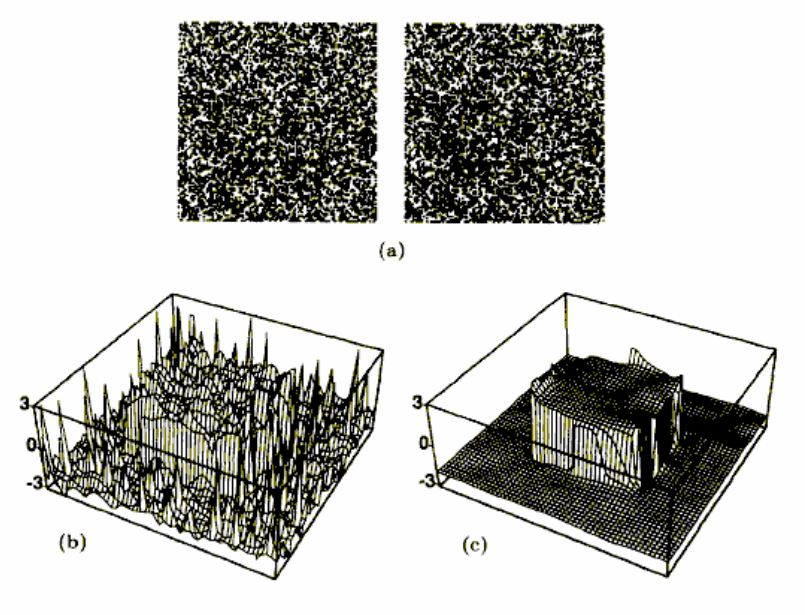

Figura 6-1 - Disparidade binocular computada a partir de estereogramas gerados com pontos aleatórios de 110x110 pixels. A parte central dos estereogramas (50x50 pixels) e a parte externa possuem disparidades de 2 e -2 pixels respectivamente (a). O resultado do processamento (b) é significantemente melhorado utilizando *spatial pooling* (c). Fonte: [24].

Uma abordagem diferente para o problema de detecção de disparidade foi feita por Ahlvers *et al.* nos trabalhos [2] e [3]. Nestes trabalhos utilizou-se um algoritmo baseado em FFT para estimar a disparidade binocular entre imagens estéreo. Uma vantagem deste tipo de abordagem sobre a utilização de filtros Gabor é o fato de que, com a FFT, são utilizados vários canais de freqüência (todos os possíveis dentro da faixa discretizada pela FFT) enquanto que com filtros Gabor é utilizada apenas uma freqüência preferencial. Contudo, isto não se mostrou um problema pois os filtros Gabor possuem uma freqüência preferencial mas também possuem uma largura de banda na qual é possível ajustar uma faixa de freqüências ao redor da freqüência preferencial na qual o filtro responde satisfatoriamente.

Além disto, a envoltória gaussiana dos filtros Gabor amortece as bordas da portadora cossenóide do filtro Gabor que é responsável por extrair as características de freqüência do estímulo, fazendo com que a região central tenha mais importância que a região periférica o que é extremamente desejado já que a célula responderá melhor quando o estímulo estiver projetado na mesma posição nos dois campos receptivos (disparidade zero). Este fato não ocorre quando se utiliza FFT, podendo fazer com que componentes de freqüência sem muita importância para o estímulo que está sendo analisado num determinado momento possam influenciar negativamente no cálculo da disparidade. Ou seja, o filtro Gabor extrai

características locais de freqüência do estímulo enquanto que a FFT extrai todas as características de freqüência ao redor do estímulo, até mesmo as que não são relevantes.

As imagens utilizadas em [2] e [3], também são imagens estéreo artificiais, simplificando drasticamente vários problemas inerentes à percepção de profundidade que o cérebro precisa resolver considerando fatores tais como, diferença de contraste e luminosidade entre as imagens direita e esquerda (devido à iluminação ambiente, reflexos, etc), diferença de foco entre as imagens (devido a defeitos visuais como miopia, hipermetropia e astigmatismo), entre outros. A transformada de Fourier utilizada por Ahlvers *et al.* nos trabalhos [2] e [3] é feita utilizando um *kernel* (estrutura análoga ao campo receptivo) unidimensional, sendo apenas na horizontal já que o objetivo é detectar disparidades horizontais. Isto pode ser uma vantagem computacional mas qualquer diferença vertical existente entre as imagens (mesmo que seja de apenas um pixel) pode gerar resultados significativamente imprecisos.

Um outro fato importante a ser considerado em relação ao trabalho de Ahlvers *et al.* é que já foi demonstrado por Jones e Palmer [15]: as células do córtex visual do cérebro humano extraem as características de freqüência da imagem com células que possuem campos receptivos que podem ser modelados por filtros Gabor. Ou seja, a utilização de FFT para extrair estas características seria abordar o problema de percepção de profundidade por um caminho diferente do utilizado pela natureza. Isso pode não representar um problema em si, mas a saída de um detector de disparidades baseado em FFT pode não ser apropriada para outras camadas em níveis abstratos mais elevados da percepção visual, de reconhecimento de formas ou velocidades de objetos por exemplo.

Por fim, este modelo difere dos dois trabalhos citados anteriormente na forma utilizada para detectar disparidades. Os trabalhos de Qian e Zhu [24] e [25], e Ahlvers *et al.* [2] e [3] são baseados na detecção de disparidade binocular através da diferença de fase (células com os campos receptivos centrados na mesma posição mas com diferença de fase), enquanto que neste modelo foi utilizada a diferença de posição (células com os campos receptivos centrados em posições diferentes mas com a mesma fase), conforme demonstrado na Figura 3-9.

6.2. ANÁLISE CRÍTICA DESTE TRABALHO DE PESQUISA

Atualmente existem vários trabalhos na área de visão cujo objetivo é criar uma imagem ou representação tridimensional do espaço visualizado como os trabalhos de Qian e

Zhu [24] e [25], Ahlvers et al. [2] e [3], Sturzl [26], Anzai [4] e [5] Ohzawa [21] e [22]. Alguns destes trabalhos procuram seguir uma linha biológica (Qian e Zhu [24] e [25], Anzai [4] e [5], Ohzawa [21] e [22]) ou seja, entender como o cérebro realiza este processo para posteriormente simulá-lo, enquanto que outros (Ahlvers et al. [2] e [3]) buscam atingir este objetivo por meio de ferramentas e algoritmos que não são utilizados pelo cérebro humano como, por exemplo, utilizando sensores de ultra-som, sensores a laser, várias câmeras que podem inclusive nem estar paralelas, entre outros. Este trabalho de pesquisa busca a plausibilidade biológica, ou seja, tenta resolver o problema do cálculo da disparidade binocular utilizando soluções baseadas em estudos sobre a neurofisiologia e neuroanatomia do sistema visual.

Embora o modelo proposto tenha obtido resultados qualitativamente bons, alguns itens podem ser questionados. No que diz respeito à arquitetura, a modelagem da área V1 está bem condizente com os estudos sobre a neurofisiologia da visão, nos quais foram caracterizadas células simples com campos receptivos modelados por filtros Gabor e células complexas com comportamento quadrático compostas de subunidade lineares (células simples). O mapeamento da imagem que sai da retina e vai para V1 também está condizente com a literatura, já que foi utilizado um mapeamento do tipo log-polar.

A modelagem da área MT com várias camadas sucessivas de células, na qual cada camada está associada a uma disparidade também está de acordo com a literatura, mas ainda não há dados e estudos suficientes para que se possa afirmar com exatidão que a resposta das células da área cortical MT segue a formulação deste modelo, e descrita na equação (14). Contudo, este trabalho propõe um modelo cujas características, no nível celular, podem ser verificadas em animais através da monitoração de células de MT com micropipetas e da aplicação de estímulos visuais equivalentes aos utilizados aqui.

Um ponto importante a ser observado é que, devido à utilização de filtros Gabor, têm-se apenas um canal preferencial de freqüência espacial. Isto é uma limitação inerente à utilização de filtros Gabor já que sua formulação restringe a freqüência à apenas um canal, porém com uma determinada largura de banda, (vide equação (12)). Apesar de ser um modelo que segue a linha biológica, existe esta restrição em relação à freqüência. O cérebro humano utiliza uma combinação com células sintonizadas em diferentes freqüências, mas neste trabalho foi utilizada somente uma freqüência preferencial. Entretanto este modelo pode ser facilmente estendido para que utilize mais freqüências.

A utilização de um modelo baseado em diferença de posição para codificar a disparidade binocular, ao invés de diferença de fase, tem a vantagem de não limitar a

disparidade máxima que uma célula pode detectar para uma determinada freqüência do estímulo, devido ao fato da diferença de fase ser limitada á ±180º, em uma mesma freqüência, conforme explicado na Seção 3.2.1. Dessa forma, é possível detectar disparidades maiores do que a diferença de fase entre as células permite.

O fato de este modelo utilizar uma abordagem para detecção de disparidade binocular baseada em diferença de posição traz uma discretização na disparidade detectada; discretização esta que aparentemente não existe no sistema visual humano. Esta discretização impossibilita a detecção de disparidades subpixel. Para contornar este efeito foi utilizada novamente a técnica de *spatial pooling* que, a grosso modo, funciona como um filtro passa baixa para suavizar a discretização. Este efeito e sua compensação foram descritos na Seção 4.6.

A utilização de imagens monocromáticas (escala de cinza) pode fazer com que a qualidade da representação interna criada seja inferior, pois utilizar somente a intensidade luminosa pode levar a alguma situação ambígua como, por exemplo, no caso de dois pontos com cores diferentes mas com a mesma intensidade luminosa. Além disto, apesar de alguns mamíferos possuírem visão monocromática, a grande maioria possui pelo menos visão dicromática, sendo que o homem e alguns primatas possuem visão tricromática, ficando em desacordo com o utilizado neste trabalho.

A resolução das imagens utilizada neste trabalho é baixa, inviabilizando a extração de características de alta freqüência espacial. Devido a isto, foi utilizada uma freqüência preferencial baixa em relação às freqüências espaciais que o ser humano consegue detectar. A freqüência preferencial das Gabor's utilizada neste trabalho foi de 0,85 ciclos por grau, enquanto que o sistema visual dos seres humanos é sensível principalmente às freqüências entre 0,5 ciclos por grau a 40 ciclos por grau [28]. A freqüência preferencial utilizada neste trabalho pode ser comparada à faixa de freqüência sensível pelo sistema visual de gatos que varia de cerca de 0,1 ciclos por grau à 2,5 ciclos por grau [28].

Uma série de experimentos com a finalidade de fornecer dados para avaliar melhor este trabalho de pesquisa ainda ficaram pendentes como, por exemplo, experimentos para analisar a precisão da representação interna criada também seriam importantes.

A Tabela 6-1 possui uma comparação entre o modelo proposto e os modelos de Qian e Zhu [24] e [25] e a abordagem de Ahlvers [2] e [3].

	Modelo proposto	**Qian e Zhu**	**Ahlvers**
Plausibilidade Biológica	Sim	Sim	Não
Partes do córtex modelada	LGN, V1 e MT	V1	---
Célula Complexa	Tuned inhibitory	Tuned excitatory	---
Abordagem para extração de características de freqüência e fase do estímulo	Filtros Gabor	Filtros Gabor	FFT
Abordagem para detecção de disparidade	Diferença de posição	Diferença de fase	Diferença de fase
Canal de Freqüência	Faixa de freqüência centrada numa freqüência preferencial	Faixa de freqüência centrada numa freqüência preferencial	Vários canais de freqüência
Formato do Campo Receptivo / *Kernel*	Campo Receptivo Bidimensional	Campo Receptivo Bidimensional	Kernel Unidimensional
Mapeamento da imagem	Log Polar	Plano	Plano
Imagens testadas	Reais	Artificiais	Artificiais

Tabela 6-1 - Tabela comparativa entre o modelo proposto, o modelo de Qian e Zhu [24] e [25] e a abordagem feita por Alhvers [2] e [3]

7. CONCLUSÃO

Este capítulo inicia com um sumário sobre o conteúdo deste trabalho de pesquisa e, em seguida, discute os principais resultados e conclusões alcançados. Por fim, apresenta algumas recomendações para trabalhos futuros.

7.1. SUMÁRIO

Neste trabalho foi proposta uma arquitetura neural para modelar partes do córtex V1 e MT, baseada em estudos neurofisiológicos e neuroanatômicos realizados sobre o sistema visual biológico. Assim como o sistema visual biológico, esta arquitetura é capaz de, a partir de duas imagens bidimensionais (imagens estéreo) de uma cena do mundo real, criar uma representação interna tridimensional do mundo externo com informações sobre as profundidades (distância ao observador) observadas na cena original. Afim de verificar a qualidade da representação interna criada pela arquitetura, foi feito um mapeamento inverso desta representação para um espaço tridimensional interno ao computador. O funcionamento da arquitetura proposta é baseado nas disparidades observadas entre o par de imagens estéreo.

A separação horizontal dos olhos causa pequenas diferenças de posição, chamadas de disparidades binoculares, entre regiões correspondentes entre as duas imagens projetadas na retina. Essas disparidades provêm informações suficientes para a percepção de profundidade. O problema associado à disparidade é determinar que regiões de cada uma das imagens vem do mesmo objeto na cena real. Este problema é chamado de problema da correspondência. Conhecendo a separação entre os olhos e os pontos correspondentes nas duas imagens (a disparidades entre cada ponto das imagens), é possível calcular a posição real de cada ponto da imagem no espaço, conforme visto na Seção 4.3.1.

Para poder identificar os pontos correspondentes entre as imagens e montar um mapa de disparidades, o qual pode ser considerado uma representação interna da informação de profundidade associada ao mundo externo visualizado, foram utilizados modelos de células biológicas com campos receptivos modelados por filtros Gabor. A amostragem dos pontos das imagens de entrada foi feita utilizando um mapeamento log-polar, da mesma forma que é feita pelo sistema visual humano. Este mapeamento enfatiza a região ao redor do ponto de atenção e atenua a representação da região periférica.

Para modelar o córtex V1, foram utilizados modelos de células simples e complexas. A resposta das células simples é basicamente uma associação de filtros Gabor com mapeamento log-polar num ponto da imagem, do qual são extraídas as características de freqüência e fase neste ponto através da Gabor, diminuindo a influência de regiões ao redor deste ponto que não possuem relevância através do mapeamento log-polar. As células complexas combinam a informação de várias células simples obtendo um próximo nível de processamento de profundidade. Neste trabalho foi proposto um modelo de célula do córtex MT que combina a saída de várias células complexas, provendo mais um nível de processamento de disparidade.

As células de MT modeladas neste trabalho são *tuned inhibitory*, respondendo minimamente quando um estímulo está projetado na mesma posição em seus campos receptivos nos olhos direito e esquerdo (disparidade zero). Utilizando o modelo de diferença de posição para detectar disparidade binocular, são disponibilizadas várias células MT com os campos receptivos direito centrados no mesmo ponto na imagem direita, porém com os campos receptivos esquerdo centrados em posições subseqüentes horizontalmente na imagem esquerda. Dessa forma está sendo feita uma comparação entre as características de freqüência espacial entre um ponto na imagem direita com uma série de pontos dispostos horizontalmente na imagem esquerda. Selecionando a célula com a menor resposta (*tuned inhibitory*), obtém-se o ponto na imagem esquerda com a maior probabilidade de corresponder ao ponto na imagem direita.

Agrupando as células modeladas de MT em várias camadas com células de mesma disparidade, obtém-se uma estrutura semelhante à disposição, descrita pela literatura, das células da área cortical MT (vide Seção 3.3). Esta disposição em camadas facilita a seleção das células com a menor resposta e possibilita a cooperação entre células com mesma disparidade conforme mostrado na seção 3.3. Esta disposição em camadas do córtex MT viabiliza tanto o processo de vergência quanto o de *stereopsis* (percepção de profundidade baseada em disparidade).

O processo de realizar a vergência requer que a disparidade global seja minimizada (Figura 4-8). Isto se deve ao fato de que ao olhar para um ponto (ponto de vergência), a imagem deste ponto e ao redor dele é projetada nas fóveas, ou seja, em pontos correspondentes (disparidade zero). Como mapeamento log-polar enfatiza no córtex a informação proveniente da fóvea, minimizar a disparidade global significa encontrar nas imagens direita e esquerda a região projetada na fóvea, e consequentemente o ponto de vergência. Já o processo de *stereopsis* requer que a disparidade para cada ponto seja

minimizada (Figura 4-8). Dessa forma é selecionada a célula de MT com a menor resposta dentre um conjunto de células (que estão disponibilizadas na mesma posição mas em camadas diferentes, vide Figura 4-8). Agrupando cada uma das células de MT com a menor resposta numa nova estrutura chamada mapa de disparidades, é obtida a representação interna sobre disparidades entre as imagens, com informações sobre a profundidade de cada ponto, que pode ser utilizada para recriar a cena original em três dimensões internamente ao computador.

7.2. RESULTADOS E CONCLUSÕES

Neste trabalho foram avaliadas as repostas de cada uma das células da arquitetura proposta dos córtices V1 e MT e, em conjunto, as respostas das células do nível superior, de uma forma bastante satisfatória. Comparando o resultado deste modelo com resultados de dados da literatura sobre células reais e também com modelos puramente matemáticos, idealizados para um detector de disparidades, foi possível observar que este modelo produz resultados semelhantes ou melhores que os existentes, conforme visto na Seção 5.1.

No entanto, avaliar a qualidade da representação interna criada por esta arquitetura é uma tarefa complicada, principalmente pelo fato de que ainda não existem experimentos nem aparatos que consigam coletar este tipo de informação diretamente do cérebro. Dessa forma, este tipo de avaliação ainda fica limitado à comparação dos resultados obtidos com transformações tridimensionais feitas em sistemas gráficos como o Mesa3D (ambiente tipo o OpenGL), que foi utilizado neste trabalho.

Os experimentos feitos avaliando a resposta de células *tuned inhibitory* e *tuned excitatory* (seção 5.1), indicam que as células *tuned inhibitory* são mais eficientes para percepção da disparidade binocular devido à sua resposta mais homogênea, ficando mais próxima do arquétipo do detector ideal de disparidades.

Embora os resultados ainda sejam muito iniciais, sendo necessário ainda bastante estudo nesta área, pode-se considerar que os resultados foram bastante satisfatórios, pois foi conseguido, mesmo ainda sem existir uma validação mais precisa, uma representação interna tridimensional do mundo externo que possui plausibilidade biológica. Este modelo permite, inclusive, fazer predições sobre as respostas de células de MT ainda não modeladas matematicamente.

7.3. TRABALHOS FUTUROS

Os resultados satisfatórios obtidos até o momento motivam continuar as pesquisas para alcançar o objetivo final de dotar um sistema robótico da capacidade de criar representações internas do mundo externo que permitam que esse sistema possa navegar e interagir com o meio externo através desta representação.

Um estudo importante sugerido é alterar o modelo proposto para incorporar mais canais de freqüência. Com esta melhoria, espera-se que muitos problemas de ruído possam ser eliminados ou pelo menos diminuídos, pois atualmente as freqüências altas não estão sendo devidamente contempladas. Para poder aumentar o número de canais de freqüência, principalmente incluindo altas freqüências é imprescindível utilizar imagens com resolução e qualidade mais alta.

Outro trabalho extremamente importante a fazer é alterar o modelo proposto para também detectar disparidades com o modelo de diferença de fase de modo a tentar calcular disparidades com precisão subpixel, evitando a discretização na profundidade que foi observada. Conforme visto na Seção 3.2.1, acredita-se que o sistema visual humano utiliza este tipo de combinação.

Outro ponto a ser estudado é a influência da inclusão da informação de cor no modelo. Esta informação pode ser bastante útil para eliminar alguns erros de ambigüidade causados por pontos com cores diferentes, porém com a mesma intensidade luminosa. Atualmente estes pontos podem ser, erroneamente, considerados como pontos correspondentes.

Um outro trabalho relevante que deve ser feito é examinar formas consistentes para avaliar a qualidade da representação interna criada para que se possa ter uma referência de quão boa é a representação criada e em quais pontos precisa ser melhorada.

8. REFERÊNCIAS BIBLIOGRÁFICAS

[1] ADELSON, Edward H.; BERGEN, James R. **Spatiotemporal energy models for the perception of motion.** J. Opt. Soc. Am. A 2: 284–299, 1985.

[2] AHLVERS, Udo; ZOELZER Udo; RECHMEIER, Stefan. **FFT-Based Disparity Estimation for Stereo Image Coding.** University of the German Federal Armed ForcesHamburg, Germany, 2003.

[3] AHLVERS, Udo; ZOELZER, Udo. **Improvement of phase-based algorithms for disparity estimation by means of magnitude information.** Helmut-Schmidt-University / University of the Federal Armed Forces Hamburg, Germany, 2004.

[4] ANZAI, A.; OHZAWA, I.; FREEMAN, Ralph D. **Neural Mechanisms for Processing Binocular Information I. Simple Cells.** The American Physiological Society. 0022-3077/99, 1999.

[5] ANZAI, A. OHZAWA, I., FREEMAN, Ralph D. **Neural Mechanisms for Processing Binocular Information II. Complex Cells.** The American Physiological Society. 0022-3077/99, 1999.

[6] BARLOW, H.B. et al. **The neural mechanism of binocular depth discrimination.** J. Physiol. 193:327-342, 1967.

[7] DeANGELIS, Gregory C.; NEWSOME, Willian T. **Organization of disparity-selective neurons in macaque area MT.** The Journal of Neuronscience . 19(4):1398-1415, 1999.

[8] DeANGELIS, Gregory C. **Seeing in three dimensions: the neurophysiology of stereopsis.** Trends Cogn Sci 4: 80–90, 2000.

[9] DODGE, R. **Five Types of eye movement in the horizontal meridian plane of the field of regard.** Am. J. Physiol. 8:307-329, 1903.

[10] GEGENFURTNER, Karl R. **Functional Properties of Neurons in Macaque Area V3**. The American Physiological Society, 0022-3077/97, 1997.

[11] GONZALEZ, F.; PEREZ, R. **Neural mechanisms underlying stereoscopic vision**. Progress in Neurobiology. 55:191-224, 1998.

[12] FARDIN JR, Dijalma; KOMATI, Karin S.; DE SOUZA, Alberto F. **Arquitetura do Sistema Visual Humano: Uma Abordagem Computacional** In: III Encontro Regional de Informática RJ/ES, 2003, Vitória. Anais do III Encontro Regional de Informática RJ/ES. , 2003.

[13] VON der Heydt R.; PETHERHANS E.; BAUMGARTNER G. **Illusory contours and cortical neuron responses**. Science, 224:1260-1262, 1984.

[14] HUBEL, D. H.; WEISEL, T. N. **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex**. J. Physiol. 160: 106-154, 1962.

[15] JONES, J. P.; PALMER, L. A. **An evaluation of the two-dimensional Gabor filter model of simple receptive fileds in cat striate cortex.** J. Neurophysiol 58:1233-1258, 1987.

[16] KANDEL, Eric R.; SCHWARTZ, James H.; JESSELL Thomas M. **Principles of Neural Science**. 4th Ed. Prentice-Hall International, Inc., 2000.

[17] KOMATI, Karin Satie; DE SOUZA, Alberto Ferreira. **Vergence Control in a Binocular Vision System using Weightless Neural Networks**. In: Proceedings of the 4th International Symposium on Robotics and Automation. Los Alamitos: IEEE, 2002.

[18] MATHER, George. **The Visual Cortex**. [on line]. Disponível em http://www.biols.susx.ac.uk/home/George_Mather/Linked%20Pages/Physiol/Cortex.html. Acesso em: 30 mar. 2005.

[19] MOVELLAN, Javier R. **Tutorial on Gabor Filters.** [on line]. Disponível em http://mplab.ucsd.edu/tutorials/tutorials.html. Acesso em: 23 jun. 2004.

[20] MOVSHON, J. A.; ADELSON, E. H.; GIZZI, M. S.; NEWSOME, W. T. **The analysis of moving visual patterns**. In C. Chagas, R. Gattass, C. Gross (eds.), Pattern Recognition Mechanisms. New York, Springer, 117-151, 1985.

[21] OHZAWA, I.; DeANGELIS, Gregory C.; FREEMAN, Ralph D. **Stereoscopic Depth Discrimination in the Visual Cortex: Neurons Ideally Suited as Disparity Detectors**. Science, Volume 249, pp. 1037-1041, 1990.

[22] OHZAWA, I.; DeANGELIS, Gregory C.; FREEMAN, Ralph D. **Encoding of Binocular Disparity by Complex Cells in the Cat's Visual Cortex**. The American Physiological Society, 0022-3077, 1997.

[23] PETTIGREW, J.D. et al. **Binocular interaction on single units in cat striate cortex: simultaneous stimulation by songle moving slit with receptive fields in correspondence**. Exp. Brain res. 6:391-410.

[24] QIAN, Ning; ZHU, Yudong. **Physological Computation of Binocular Disparity**. Vision Res. 37: 1811–1827, 1997.

[25] QIAN, Ning. **Binocular Disparity and the Perception of Depth**. Neuron 18: 359-368, 1997

[26] STÜRZL W.; HOFFMANN, U.; MALLOT, Handspeter A. **Vergence Control and Disparity Estimation with Energy Neurons: Theory and Implementation**. ICANN 2002, LNCS, pp. 1255-1260, 2002.

[27] TOOTELL, R. B.; SILVERMAN, M. S.; SWITKES, E.; VALOIS, R. L. **Deoxyglucose analysis of retinotopic organization in primate striate córtex**. Science, 218:902-904, Nov 26, 1982.

[28] VALOIS, Russell L. De; VALOIS, Karen K. De. **Spatial Vision**. New York: Oxford University Press, 1988.

[29] ZEKI, S. M. **Colour coding of the rhesus monkey prestriate cortex**. Brain Research, 422-427, 1973.