

**UMA MODELAGEM MATEMÁTICO-
COMPUTACIONAL DO SISTEMA BIOLÓGICO DE
PERCEPÇÃO DE MOVIMENTO E VELOCIDADE**

Raphael Vivacqua Carneiro

Dissertação de Mestrado em Informática

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Vitória, Dezembro de 2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

6 cm

13 cm

7 cm

4 cm

RAPHAEL VIVACQUA CARNEIRO

Diss. MESTRADO

- 2006

UMA MODELAGEM MATEMÁTICO- COMPUTACIONAL DO SISTEMA BIOLÓGICO DE PERCEPÇÃO DE MOVIMENTO E VELOCIDADE

Raphael Vivacqua Carneiro

Dissertação submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para obtenção do grau de Mestre em Informática.

Aprovada em 16/12/06 por:

Prof. Dr. Aberto Ferreira de Souza - PPGI/UFES

Prof. Dr. Elias Silva de Oliveira - PPGI/UFES

Prof. Dr. Felipe Maia Galvão França - COPPE/UFRJ

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Vitória, Dezembro de 2006

Carneiro, Raphael Vivacqua, 1965

Uma Modelagem Matemático-Computacional do Sistema Biológico de Percepção de Movimento e Velocidade. [Vitória] 2006

xiii, 108 p., 29,7 cm (UFES, M. Sc., Informática, 2006)

Dissertação, Universidade Federal do Espírito Santo, PPGI.

1. Ciência da Computação. 2. Neurocomputação. I. PPGI/UFES. II. Título.

A Cris, Leo e Lulu.

“A imaginação é mais importante do que o conhecimento.”
Albert Einstein

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	OBJETIVOS	14
1.2	RELEVÂNCIA DO ESTUDO	15
1.3	CONTRIBUIÇÕES DESTE TRABALHO	16
2	REFERENCIAL TEÓRICO.....	17
2.1	O SISTEMA VISUAL HUMANO	17
2.2	MODELOS TEÓRICOS	21
2.2.1	<i>Modelagem do mapeamento da retina em V1.....</i>	<i>21</i>
2.2.2	<i>Modelagem das células de V1</i>	<i>23</i>
2.2.3	<i>Modelagem das células de MT.....</i>	<i>27</i>
2.3	MODELO PROPOSTO PARA A PERCEPÇÃO DE MOVIMENTO	28
3	METODOLOGIA	33
3.1	PLATAFORMA.....	33
3.2	PROCEDIMENTOS.....	36
3.2.1	<i>Captação das imagens.....</i>	<i>36</i>
3.2.2	<i>Identificação da região de movimento.....</i>	<i>37</i>
3.2.3	<i>Emulação do movimento ocular de sacada</i>	<i>38</i>
3.2.4	<i>Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida ..</i>	<i>39</i>
3.2.5	<i>Criação do mapa de velocidades.....</i>	<i>40</i>
3.2.6	<i>Representação visual da velocidade de cada ponto na imagem.....</i>	<i>41</i>
3.2.7	<i>Comparação dos resultados obtidos com os esperados</i>	<i>42</i>
4	EXPERIMENTOS	43
4.1	PERCEPÇÃO DE MOVIMENTO HORIZONTAL DE UM OBJETO.....	44
4.2	PERCEPÇÃO DE MOVIMENTO COM COMPONENTES DE VELOCIDADE HORIZONTAL E VERTICAL E PERSEGUIÇÃO SUAVE A UM OBJETO	58
4.3	PERCEPÇÃO DE MOVIMENTOS DE TRANSLAÇÃO SIMULTÂNEOS E DIVERSOS DE MÚLTIPLOS OBJETOS ..	73
4.4	PERCEPÇÃO DE MOVIMENTO DE ROTAÇÃO DE UM OBJETO.....	86
5	DISCUSSÃO	98
5.1	TRABALHOS ANTERIORES.....	98
5.2	ANÁLISE CRÍTICA	99
6	CONCLUSÃO.....	102
6.1	TRABALHOS FUTUROS	103
	REFERÊNCIAS	105

LISTA DE FIGURAS

Figura 1 – Sistema visual humano e suas principais áreas corticais.....	17
Figura 2 – Mapeamento do campo visual em V1.....	18
Figura 3 – Respostas de uma célula simples em V1 a diversos estímulos.....	19
Figura 4 – Respostas de uma célula complexa em V1 a diversos estímulos.....	19
Figura 5 – Modelo de células de MT sensíveis a diversos movimentos e profundidades.....	20
Figura 6 – Representação do mapeamento log-polar da imagem projetada na retina em V1.....	22
Figura 7 – Exemplo de transformação log-polar, usando o sistema de coordenadas u-v acima. O eixo vertical no centro representa $u = 0$, sendo que u cresce positivamente tanto à direita quanto à esquerda. À direita, o eixo da coordenada angular v varia de $-\pi/2$ a $\pi/2$ no sentido para cima, e à esquerda v varia de $\pi/2$ a $3\pi/2$ no sentido para baixo.	23
Figura 8 – Gráficos da função de Gabor para diversos valores de fase da portadora.....	25
Figura 9 – Modelo de Hubel e Wiesel para o campo receptivo das células simples binoculares.....	25
Figura 10 – Modelo de célula simples binocular com diferença de fase igual a π	26
Figura 11 – Modelo de célula de MT.....	27
Figura 12 – Modelo proposto para percepção de movimento e profundidade. As estruturas neurais sintonizadas na detecção de disparidades binoculares (d_1, d_2, \dots, d_n) possibilitam a percepção de profundidade, enquanto as estruturas sintonizadas na detecção de disparidades espaço-temporais (v_1, v_2, \dots, v_n) possibilitam a percepção de velocidade.....	29
Figura 13 – Modelo representando uma estrutura neural de V1 e MT sintonizada para detecção de uma disparidade binocular específica.....	30
Figura 14 – Modelo representando uma estrutura neural de V1 e MT sintonizada para detecção de uma disparidade espaço-temporal específica.....	32
Figura 15 – Utilização do <i>framework</i> MAE.....	34
Figura 16 – Exemplo de gráfico que representa a saída das células de MT para cada ponto (x, y) da imagem. Uma vez que esta camada de células esteja sintonizada para detecção de disparidade nula, pode-se identificar as regiões de movimento e de repouso na imagem.	37
Figura 17 – Experimento de Johansson com as luzes andantes. À esquerda há um conjunto de pontos luminosos; à direita a representação da conexão desses pontos, construída mentalmente. ...	41
Figura 18 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (600, 240).	45
Figura 19 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada. O objeto se movimentou cerca de 15 pixels à direita.	45
Figura 20 – Círculo que representa o campo visual em torno do centro de atenção em (600, 240). ...	46
Figura 21 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.	46
Figura 22 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.	46
Figura 23 – Representação da região de movimento ou de incerteza, centro de atenção (600, 240).	47
Figura 24 – (a) Cena no instante 1. O centro de atenção foi deslocado para a coordenada (308, 244), por meio da emulação do movimento ocular de sacada. Em (b) e (c) vemos o mapeamento log-polar das imagens, em torno do novo centro de atenção, nos instantes 1 e 2, após a sacada.	48
Figura 25 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (246, 239), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.	49

Figura 26 – Mapeamento log-polar das imagens, em torno do centro de atenção (246, 239), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.....	50
Figura 27 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase. ...	50
Figura 28 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.....	51
Figura 29 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.	51
Figura 30 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com <i>spatial pooling</i> . 52	
Figura 31 – Representação da região de movimento ou de incerteza, centro de atenção (246, 239). 52	
Figura 32 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, no movimento horizontal, sentido para a direita.	53
Figura 33 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical com módulo igual a 16.....	54
Figura 34 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.	55
Figura 35 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$	56
Figura 36 – Representação visual das velocidades resumidas num grid de largura 16 x 16 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT. 57	
Figura 37 – Cena nos instantes 1 (a), 2 (b), 3 (c) e 4 (d). O centro de atenção está localizado na coordenada (320, 240). O objeto se movimentou cerca de 14 pixels à esquerda e 11 abaixo.....	58
Figura 38 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240). ..	59
Figura 39 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.	59
Figura 40 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.	60
Figura 41 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240). 60	
Figura 42 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (399, 346), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência. 61	
Figura 43 – Mapeamento log-polar das imagens, em torno do centro de atenção (399, 346), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.....	62
Figura 44 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase. ...	62
Figura 45 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.....	63
Figura 46 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma	

retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.	63
Figura 47 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com <i>spatial pooling</i> .64	64
Figura 48 – Representação da região de movimento ou de incerteza, centro de atenção (399, 346). 64	64
Figura 49 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, no movimento diagonal, sentido para baixo à esquerda.....	65
Figura 50 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical com módulo igual a 12.....	66
Figura 51 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.	67
Figura 52 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$	68
Figura 53 – Representação visual das velocidades resumidas num grid de largura 12 x 12 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT. 69	69
Figura 54 – Mapeamento log-polar das imagens, em torno do centro de atenção (387, 334), nos instantes 2 e 3, após o movimento ocular de perseguição suave.	70
Figura 55 – Representação da região de movimento ou de incerteza, centro de atenção (387, 334), após o movimento ocular de perseguição suave.....	70
Figura 56 – Mapeamento log-polar das imagens, em torno do centro de atenção (375, 322), nos instantes 3 e 4, após o movimento ocular de perseguição suave.	71
Figura 57 – Representação da região de movimento ou de incerteza, centro de atenção (375, 322), após o movimento ocular de perseguição suave.....	71
Figura 58 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (320, 240).	74
Figura 59 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada.	74
Figura 60 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240). ..	75
Figura 61 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.	75
Figura 62 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; as áreas claras correspondem às regiões de movimento. 75	75
Figura 63 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240). 76	76
Figura 64 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (358, 271), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência. 77	77
Figura 65 – Mapeamento log-polar das imagens, em torno do centro de atenção (358, 271), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.....	78
Figura 66 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase. ...	78
Figura 67 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.....	79
Figura 68 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.	79

Figura 69 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com <i>spatial pooling</i> .	80
Figura 70 – Representação da região de movimento ou de incerteza, centro de atenção (358, 271).	80
Figura 71 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, em movimentos horizontais e verticais.	81
Figura 72 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 8.	82
Figura 73 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.	83
Figura 74 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$.	84
Figura 75 – Representação visual das velocidades resumidas num grid de largura 10 x 10 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.	85
Figura 76 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (320, 240).	86
Figura 77 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada. O objeto se movimentou cerca de 15 pixels à direita.	87
Figura 78 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240).	87
Figura 79 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.	88
Figura 80 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.	88
Figura 81 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240).	88
Figura 82 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (299, 268), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.	89
Figura 83 – Mapeamento log-polar das imagens, em torno do centro de atenção (299, 268), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.	90
Figura 84 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase.	90
Figura 85 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.	91
Figura 86 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.	91
Figura 87 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com <i>spatial pooling</i> .	92
Figura 88 – Representação da região de movimento ou de incerteza, centro de atenção (299, 268).	92
Figura 89 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 24.	93
Figura 90 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 12.	94

Figura 91 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.	95
Figura 92 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,7$	96
Figura 93 – Representação visual das velocidades resumidas num grid de largura 16×16 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.	97

LISTA DE TABELAS

Tabela 1 – Velocidades (v_x , v_y) predefinidas, expressas em pixels por Δt	40
Tabela 2 – Parâmetros dos filtros utilizados em todos os experimentos.....	43
Tabela 3 – Experimento 1: Seqüência de convergência das sacadas	49
Tabela 4 – Experimento 2: Seqüência de convergência das sacadas	62
Tabela 5 – Experimento 3: Seqüência de convergência das sacadas	77
Tabela 6 – Experimento 4: Seqüência de convergência das sacadas	89

RESUMO

Tudo o que vemos é construído mentalmente pelo sistema visual do cérebro humano, a partir dos estímulos recebidos nas retinas: a percepção de cor, forma, profundidade e movimento, os objetos e as cenas visuais completas. Pesquisas na área de inteligência artificial, especialmente aplicadas à robótica, buscam emular esta função biológica, a fim de capacitar sistemas artificiais na interação com o mundo real. Entretanto, tarefas simples para um ser humano são muito complexas para um robô.

Este trabalho propõe uma modelagem matemático-computacional do sistema visual humano, que seja biologicamente plausível e que seja capaz de emular a função de percepção de movimento e de velocidade dos objetos presentes no campo visual. A arquitetura neural utilizada representa células das camadas V1 e MT do córtex visual, com campos receptivos que utilizam convolução dos estímulos visuais com uma função de Gabor, e um mapeamento log-polar das células da retina para as equivalentes no córtex visual. Essa modelagem foi implementada em um sistema computacional que calcula um mapa de velocidades, abrangendo todo o campo visual, a partir do processamento de uma seqüência de imagens em movimento, captadas por uma câmera, e do subsequente processamento das camadas neurais envolvidas no modelo. A fim de que a sua eficácia fosse verificada, diversos experimentos foram realizados, obtendo-se resultados bastante satisfatórios.

Palavras-chave: Percepção de Movimento, Sistema Visual Humano, Visão Artificial

ABSTRACT

Everything that we see is mentally built by the human brain's visual system, from the stimuli received by the retinas: the perception of color, shape, depth and motion, the objects and the complete visual scenes. Researches on artificial intelligence, especially applied to robotics, look for emulating this biological function, in order to enable artificial systems to interact with the real world. However, simple tasks for a human being might be very complex for a robot.

This work proposes a mathematical-computational modeling of the human visual system that is biologically plausible and capable to emulate motion perception and velocity perception of the objects present in the visual field. The proposed neural architecture represents V1 and MT cells of the visual cortex, with receptive fields using a convolution of the visual stimuli with a Gabor function, and a log-polar mapping of the retina cells to the equivalent ones in the visual cortex. That modeling was implemented in a computational system that calculates a map of velocities, comprehending the whole visual field, from the processing of a sequence of motion images, captured by a camera, and the subsequent neural layers processing. In order to evaluate the effectiveness of the system, several experiments were accomplished, obtaining quite satisfactory results.

Keywords: Motion Perception, Human Visual System, Artificial Vision

1 INTRODUÇÃO

A visão é o sentido humano mais complexo e importante para os processos de cognição e de interação do indivíduo com o Universo. No cérebro, os sistemas visuais influenciam os sistemas motores e fornecem informações para a percepção humana. Neurofisiologistas buscam identificar e compreender como funcionam os mecanismos e sistemas celulares envolvidos no processo de percepção visual. Trata-se de um processo construtivo, em que o mundo exterior é modelado e organizado em representações internas de objetos. O sistema visual recebe os sinais das imagens captadas pelos olhos e, por meio de transformações e processamentos diversos, integra estes sinais em superfícies e estas, em objetos perceptuais. Segundo Hoffman, tudo o que vemos é construído por nosso sistema visual: cor, sombra, textura, movimento, forma, objetos visuais e cenas visuais completas [HOF00].

Pesquisas na área de inteligência artificial, especialmente aplicadas à robótica, buscam emular esta função biológica, a fim de capacitar sistemas artificiais na interação com o mundo real em várias aplicações. Entretanto, tarefas simples para um ser humano são muito complexas para um robô, como por exemplo: reconhecer e manipular objetos no mundo real, locomover-se espacialmente num ambiente desconhecido, calcular a distância e a velocidade de objetos, entre outras. Os mecanismos subjacentes à visão que possibilitam o cumprimento de tais tarefas não são totalmente conhecidos.

1.1 Objetivos

Como representar o mundo externo num sistema computacional? Como perceber o movimento dos objetos exteriores? Faz-se necessária a criação de representações visuais internas do ambiente externo, com precisão suficiente para viabilizar a navegação de um sistema robótico.

Este trabalho visa investigar uma modelagem matemático-computacional que possibilite a implementação de um sistema de visão artificial capaz de emular uma função importante do sistema óculo-motor biológico: a percepção de movimento e de velocidade. Para tal implementação será realizada uma modelagem de partes do

córtex visual humano, a partir de aspectos conhecidos da arquitetura visual humana. Utilizando imagens bidimensionais, capturadas por câmeras comuns, será construída uma representação interna ao nosso sistema de visão, dos objetos presentes nessas imagens. Com esta representação interna, espera-se dar ao sistema uma capacidade tal que possibilitaria navegar no ambiente externo e manipular objetos nele presentes.

Este trabalho está inserido numa linha de pesquisa do Departamento de Informática da Universidade Federal do Espírito Santo (UFES), cujo objetivo de longo prazo é implementar uma arquitetura neural artificial com algumas capacidades análogas às do sistema visual humano. Trabalhos anteriores nessa linha de pesquisa exploraram a função de controle de vergência do sistema óculo-motor [KOM02] e a percepção de profundidade a partir de imagens estereoscópicas bidimensionais [OLI05].

1.2 Relevância do estudo

O conhecimento atual acerca da visão biológica não é completo nem bastante detalhado. Isto tem motivado muitos pesquisadores a propor teorias e arquiteturas computacionais relativas ao processamento da visão, para posterior comparação com o processo biológico. Essas teorias e arquiteturas têm evoluído na medida em que aumenta a compreensão sobre este processo.

Estudos realizados nas áreas de neurofisiologia, 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 pela qual o nosso cérebro efetua o processamento visual é importante para podermos dotar o computador de capacidade de criar representações do mundo ao seu redor utilizando a visão.

1.3 Contribuições deste trabalho

A principal contribuição deste trabalho foi a implementação de um modelo matemático-computacional capaz de:

- perceber objetos em movimento, possibilitando a emulação do movimento ocular de sacada;
- identificar as velocidades dos objetos, possibilitando a emulação do movimento ocular de perseguição suave;
- viabilizar a representação integrada das percepções de movimento e de profundidade dos objetos.

2 REFERENCIAL TEÓRICO

Há muitas pesquisas que servem de subsídio a este trabalho, em diferentes áreas do conhecimento, tais como: psicologia cognitiva, neurociência, informática. É especialmente relevante o referencial teórico relativo ao sistema visual humano e aos modelos teóricos propostos em trabalhos anteriores.

2.1 O sistema visual humano

As informações visuais recebidas pelas retinas chegam a duas áreas subcorticais simétricas denominadas *lateral geniculate nucleus* (LGN), sendo daí retransmitidas para o córtex visual. Diversas áreas corticais processam essas informações, cada uma contribuindo de forma diferente para a percepção de movimento, profundidade, forma e cor. Kandel (2000) destaca o papel das áreas V1, V2, V3, V4 e V5, sendo esta última também conhecida como MT. Essas áreas corticais visuais ocupam aproximadamente metade do córtex de um macaco.

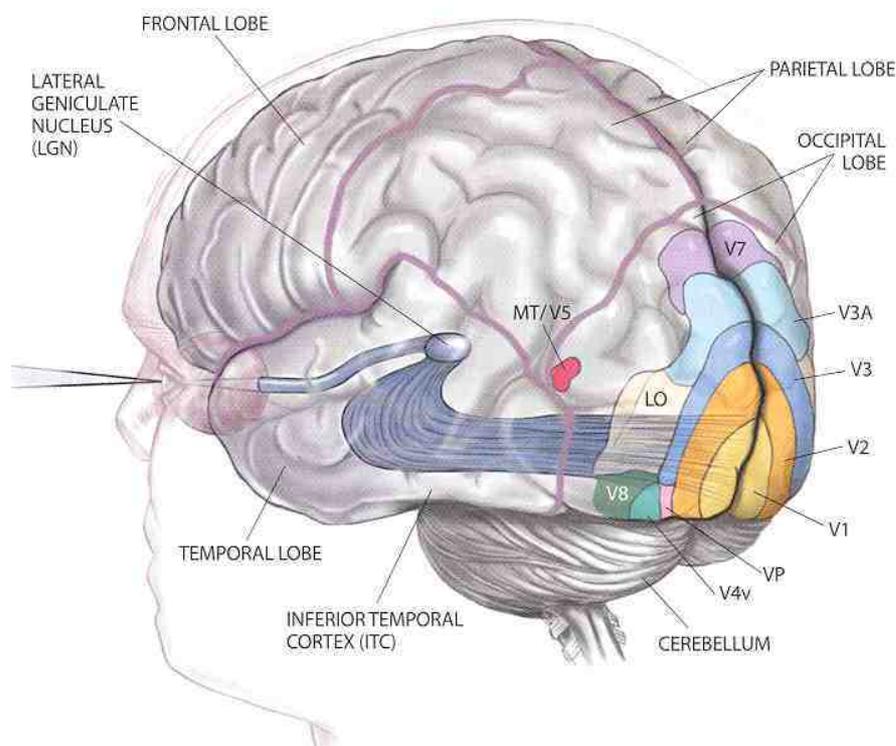


Figura 1 – Sistema visual humano e suas principais áreas corticais
Fonte: Pomeranz e Portillo, 2006.

Cada neurônio de saída da retina recebe informações de um conjunto de células de uma área circunvizinha na retina, que constitui o seu campo receptivo.

Tootell *et al.* (1982) constataram que metade das projeções da retina sobre o córtex visual primário (V1) são provenientes de uma região central da retina, com cerca de 1,5 mm de diâmetro, chamada *fovea centralis*, ou simplesmente fóvea. Isso caracteriza a maior acuidade visual desta área. O ângulo total de visão humana é de cerca de 200°, sendo que a fóvea representa cerca de 2° no centro do campo de visão. Isso corresponde ao tamanho da unha do polegar vista à distância de um braço estendido. Áreas do campo visual vizinhas na retina são também vizinhas em V1, conforme Figura 2.

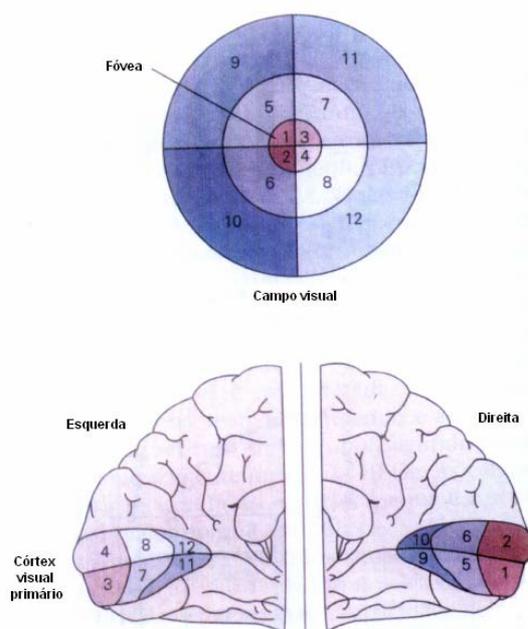


Figura 2 – Mapeamento do campo visual em V1
Fonte: Kandel, 2000.

Hubel e Wiesel (1962) classificaram as células de V1 de acordo com a complexidade de sua resposta. As células *simples* possuem um campo receptivo com regiões excitatórias e inibitórias e respondem de forma mais vigorosa a estímulos na forma de barras com uma orientação específica. A Figura 3-a representa uma célula simples cujo campo receptivo responde de forma vigorosa a um estímulo na forma de uma barra na vertical, localizada no centro do campo e com largura aproximadamente de um terço da largura do campo. As Figuras 3-b, 3-c e 3-d

representam essa mesma célula com resposta nula para estímulos em posição ou fase diferente (b), em orientação diferente (c) e com largura ou frequência diferente (d).

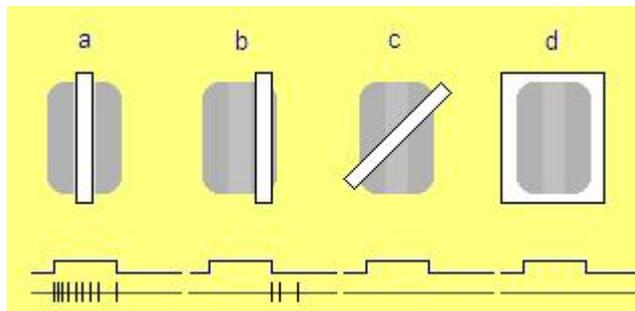


Figura 3 – Respostas de uma célula simples em V1 a diversos estímulos
Fonte: Mather, 2006.

As células *complexas* diferem das simples por responder de forma invariável em relação à 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. A Figura 4-a e 4-b representam uma célula complexa cujo campo receptivo responde de forma vigorosa a um estímulo na forma de uma barra vertical, localizada em qualquer posição do campo. Neste exemplo, a célula tem resposta vigorosa quando o estímulo está se movimentando no sentido para a esquerda (Figura 4-c), e resposta nula quando o estímulo está se movimentando no sentido para a direita (Figura 4-d).

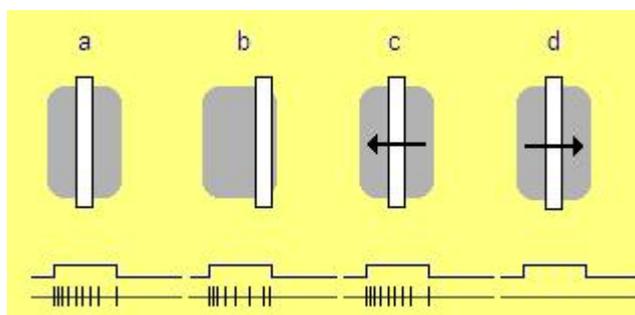


Figura 4 – Respostas de uma célula complexa em V1 a diversos estímulos
Fonte: Mather, 2006.

As células de V2, assim como as células de V1, são sensíveis a orientação, cor e profundidade dos estímulos, ou seja, estas células continuam a análise iniciada em

V1. Von der Heydt, Petherhans e Baumgartner (1984) observaram que em V2 é feito um processamento de contornos num nível acima do que ocorre em V1, evidenciando a análise progressiva que ocorre no córtex visual.

Segundo Gegenfurtner (1997), grande parte das células de V3 é seletiva 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.

Zeki (1973) e estudos posteriores mostraram que as células de V4 são sensíveis a combinações de cores e formas.

Movshon *et al.* (1985) testaram a percepção de movimento que começa de forma rudimentar em V1 atingindo formas bem mais abstratas na área temporal média (MT), numa abstração sucessiva, ou seja, em etapas. A percepção de profundidade também é processada em MT. A Figura 5 apresenta um modelo esquemático das células de MT. As setas indicam as direções de movimento a que cada grupo de células é sensível. As faixas de cores representam as células sensíveis a uma determinada profundidade de afastamento do estímulo em relação ao ponto de fixação do observador. As células azuis são sensíveis somente a movimento, enquanto as demais são sensíveis ao mesmo tempo a movimento e profundidade.

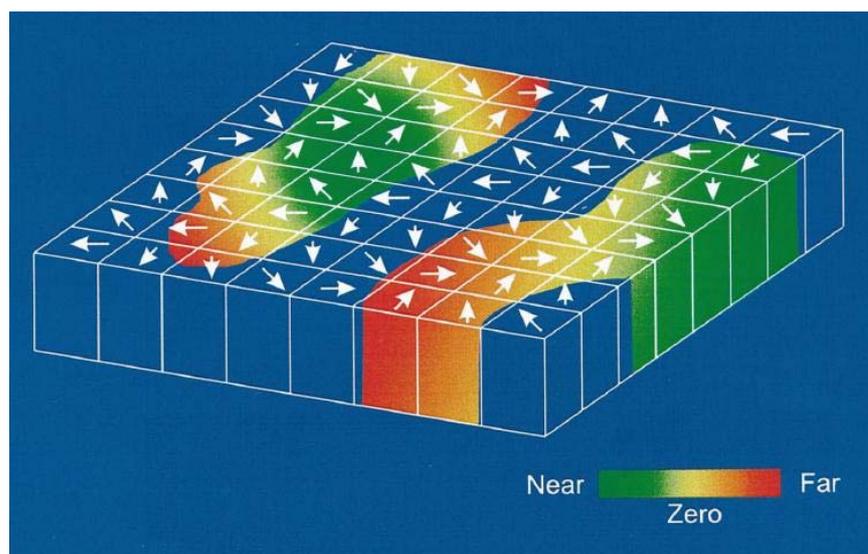


Figura 5 – Modelo de células de MT sensíveis a diversos movimentos e profundidades
Fonte: DeAngelis, 1999.

Outro aspecto de interesse no sistema visual humano são os movimentos oculares. Os movimentos dos olhos ocorrem em três eixos de rotação: vertical, horizontal e torsional. O sistema é capaz de mover os olhos rapidamente em busca de um ponto de interesse. Uma vez localizado o ponto de interesse, o sistema estabiliza essa imagem na retina, na região da fóvea, mesmo que a cabeça se movimente numa direção qualquer.

Dodge (1903) identificou cinco movimentos oculares distintos:

- Vestíbulo-ocular: mantém as imagens estáveis na retina durante rápidas rotações da cabeça;
- Optocinético: mantém as imagens estáveis na retina durante rotações lentas e contínuas da cabeça;
- Sacada: movimenta bruscamente o olho a fim de trazer novos pontos de interesse para a fóvea;
- Perseguição suave: mantém na fóvea a imagem de um alvo em movimento;
- Vergência: movimenta distintamente os olhos esquerdo e direito, de modo que o mesmo ponto de interesse seja trazido para ambas fóveas.

2.2 Modelos teóricos

Os modelos teóricos são implementados por meio de estruturas de dados que armazenam as imagens projetadas nas retinas direita e esquerda, e funções de mapeamento, transformação e extração de características dessas imagens. Essas funções modelam o processamento feito pelas células das áreas corticais, numa arquitetura com níveis crescentes de abstração.

2.2.1 Modelagem do mapeamento da retina em V1

Tootell *et al.* (1982), por meio de experimentos realizados com aplicação de substância radioativa em primatas, verificaram como é feito o mapeamento da retina em V1. Este mapeamento pode ser modelado matematicamente utilizando uma transformação log-polar da imagem projetada na retina para o córtex visual primário.

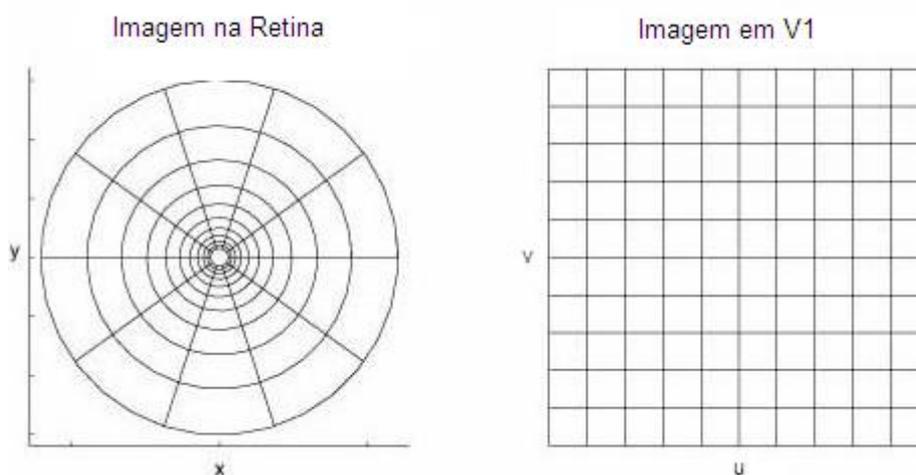


Figura 6 – Representação do mapeamento log-polar da imagem projetada na retina em V1

A Figura 6 mostra no lado esquerdo um plano cartesiano x-y com uma série de círculos concêntricos cujos raios guardam uma proporção logarítmica entre si e no lado direito um plano de coordenadas u-v em que cada retângulo corresponde a um segmento de setor circular no plano x-y. As equações (1) e (2) estabelecem matematicamente essa transformação. Assim como ocorre no sistema visual humano, as regiões mais próximas do centro dos círculos no plano x-y (fóvea da retina) são representadas por uma área relativamente maior no plano u-v (córtex V1).

$$u = \log\left(\sqrt{x^2 + y^2}\right) \quad (1)$$

$$v = \arctan\left(\frac{y}{x}\right) \quad (2)$$

A Figura 7 mostra um exemplo de transformação log-polar utilizada neste trabalho, considerando que a fóvea está centrada no ponto com a cruz vermelha, na parte inferior da bola de futebol. A fim de preservar a relação de vizinhança entre os pontos da imagem no centro de transformação, neste trabalho foi adotada uma convenção de coordenadas do plano u-v conforme esquematizado na Figura 7.

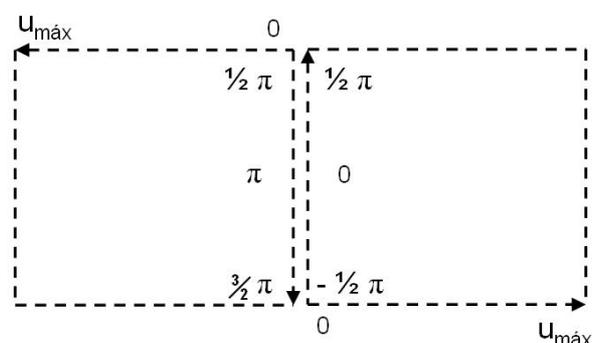
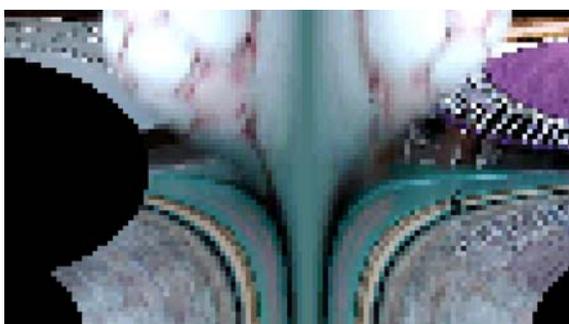
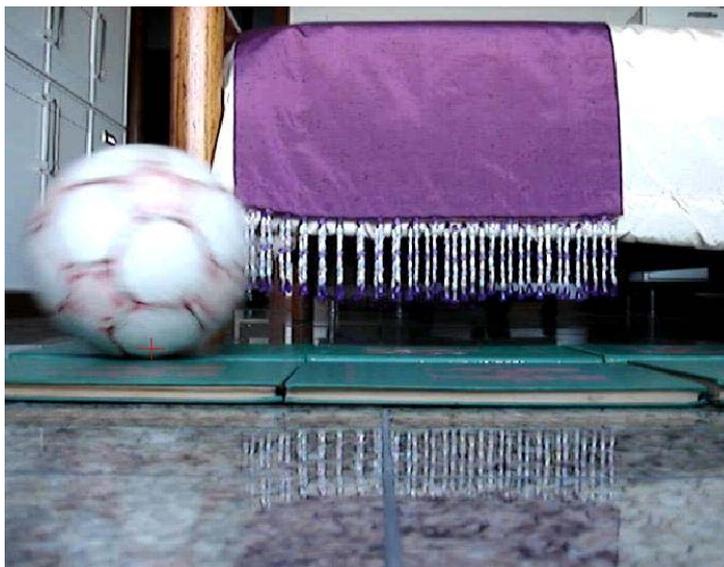


Figura 7 – Exemplo de transformação log-polar, usando o sistema de coordenadas u - v acima. O eixo vertical no centro representa $u = 0$, sendo que u cresce positivamente tanto à direita quanto à esquerda. À direita, o eixo da coordenada angular v varia de $-\pi/2$ a $\pi/2$ no sentido para cima, e à esquerda v varia de $\pi/2$ a $3\pi/2$ no sentido para baixo.

2.2.2 Modelagem das células de V1

Jones e Palmer (1987) descreveram uma função de Gabor, que é o produto de uma cossenóide complexa denominada portadora, por uma função gaussiana bidimensional denominada envoltória. A parte real dessa função complexa se ajusta muito bem com os campos receptivos encontrados nas células simples no córtex estriado em gatos e também em macacos.

As equações (3), (4) e (5) descrevem respectivamente a função de Gabor complexa bidimensional $g(x,y)$, a parte real da função portadora $s(x,y)$ e da função envoltória $w(x,y)$ [MOV06]. As coordenadas espaciais x e y representam um ponto na retina; o parâmetro Φ corresponde à fase da portadora; os parâmetros u_0 e v_0 correspondem

à frequência espacial da portadora em coordenadas cartesianas; o parâmetro K é um fator de escala da amplitude da envoltória; os parâmetros x_0 e y_0 representam as coordenadas de pico da gaussiana e os parâmetros a e b são parâmetros de escala da envoltória que controlam a largura de banda. O subscrito r indica a possibilidade de uma operação de rotação no campo receptivo.

$$g(x, y) = s(x, y) * w(x, y) \quad (3)$$

$$s(x, y) = \cos(2\pi(u_0x + v_0y) + \Phi) \quad (4)$$

$$w(x, y) = K \exp\left(-\pi\left(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2\right)\right) \quad (5)$$

A equação final (6) da função de Gabor que modela o campo receptivo de uma célula simples contém uma compensação da componente DC na portadora, uma vez que a célula não deve responder a estímulos uniformes em seu campo, independentemente da intensidade do estímulo. A Figura 8 representa graficamente a função de Gabor para diferentes valores de Φ .

$$g(x, y) = K \exp\left(-\pi\left(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2\right)\right) \left[\cos(2\pi(u_0x + v_0y) + \Phi) - \exp\left(-\pi\left(\frac{u_0^2}{a^2} + \frac{v_0^2}{b^2}\right)\right) \right] \cos(\Phi) \quad (6)$$

Segundo esse modelo, a resposta de uma célula simples no ponto (x, y) equivale à convolução espacial da função de Gabor com o estímulo visual presente no campo receptivo da célula, conforme equação (7).

$$R(x, y) = g(x, y) * \text{Estímulo}(x, y) \quad (7)$$

Hubel e Wiesel (1962) 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 das células do LGN. Nesse modelo, os campos receptivos de várias células do LGN são superpostos para formar o campo receptivo de uma célula simples binocular em V1, conforme Figura 9.

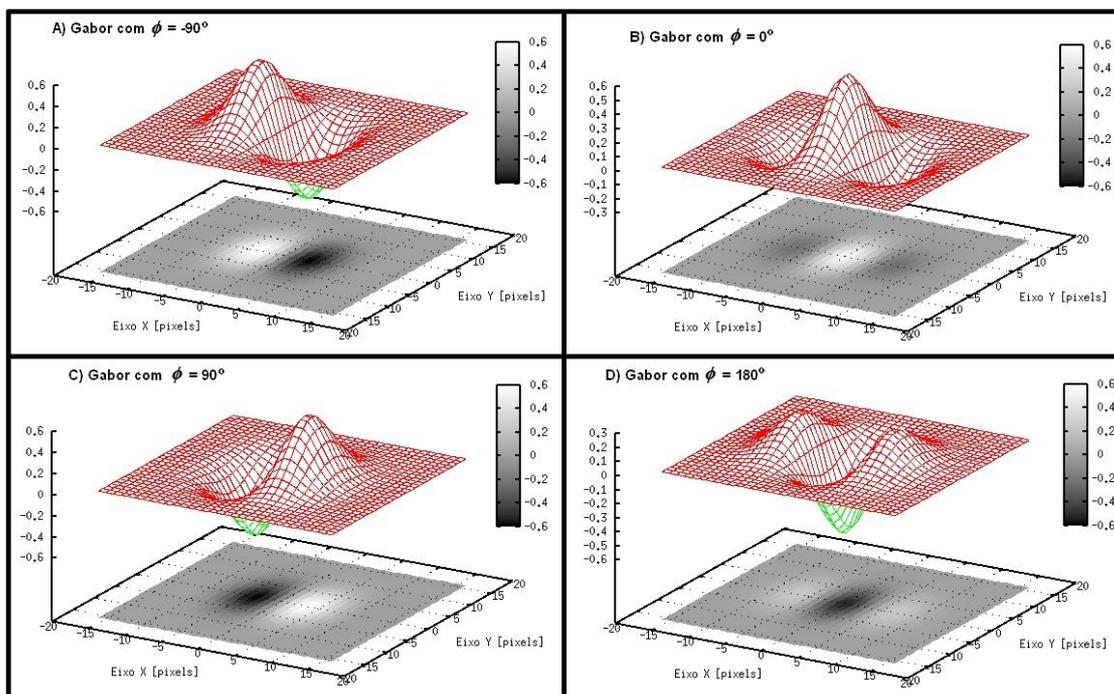


Figura 8 – Gráficos da função de Gabor para diversos valores de fase da portadora
 Fonte: Oliveira e DeSouza, 2005.

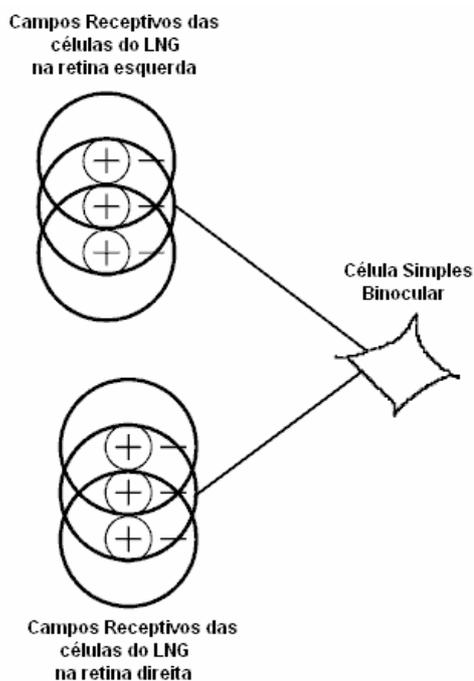


Figura 9 – Modelo de Hubel e Wiesel para o campo receptivo das células simples binoculares
 Fonte: Qian, 1997b.

Para detecção de disparidades binoculares com células simples binoculares, há três modelos: através da diferença de posição dos campos receptivos nos olhos esquerdo e direito; através da diferença de fase dos campos receptivos; ou um modelo híbrido, de posição e fase [DEA00].

Oliveira e DeSouza (2005) implementaram um modelo com dois tipos de células simples: monoculares e binoculares. As células simples monoculares utilizam um campo receptivo modelado por uma função de Gabor, centrado numa posição da retina de um dos olhos. As células simples binoculares somam a contribuição de dois campos receptivos, cada um proveniente de um olho, com diferença de fase igual a π , conforme Figura 10. Este modelo é do tipo *tuned inhibitory*, isto é, quando o mesmo estímulo é projetado simultaneamente na mesma posição em ambos olhos, a resposta da célula simples binocular é nula, caso contrário a resposta é não nula.

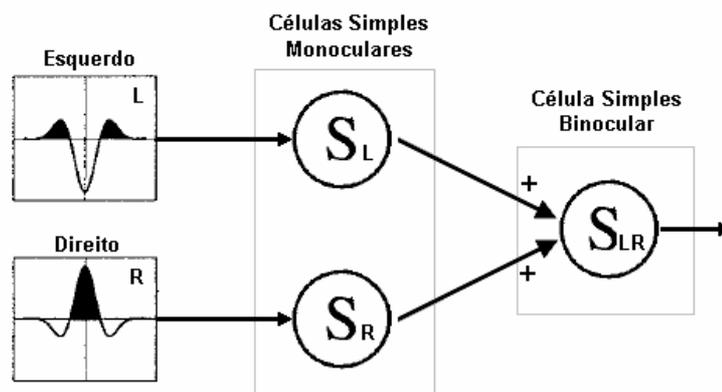


Figura 10 – Modelo de célula simples binocular com diferença de fase igual a π
 Fonte: Oliveira e DeSouza, 2005.

Para as células complexas, uma modelagem proposta é por meio de filtros em quadratura de fase, isto é, com diferença de fase igual a $\pi/2$, cujas saídas são elevadas ao quadrado e somadas [ADE85].

Oliveira e DeSouza (2005) implementaram um modelo com dois tipos de células complexas: monoculares e binoculares. As células complexas monoculares são modeladas recebendo projeções de duas células simples monoculares, com campos receptivos centrados num mesmo ponto da retina, de um mesmo olho, com diferença de fase igual a $\pi/2$. As células complexas binoculares foram

implementadas a partir de uma variação do modelo proposto por Ohzawa (1987), tornando-as do tipo *tuned inhibitory*, em vez de *tuned excitatory*. Em ambos os casos, as células complexas computam a soma dos quadrados dos estímulos provenientes das células simples.

2.2.3 Modelagem das células de MT

O modelo proposto por Oliveira e De Souza (2005) para as células de MT normaliza o estímulo proveniente de uma célula complexa binocular pela soma dos estímulos provenientes de duas células complexas monoculares, com campos receptivos centrados nos mesmos pontos da célula complexa binocular, conforme Figura 11.

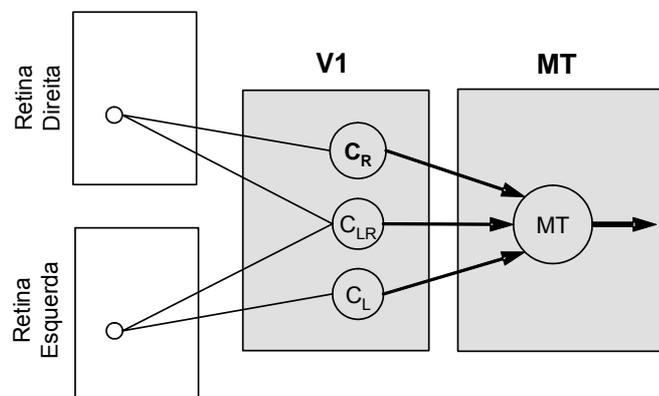


Figura 11 – Modelo de célula de MT
Fonte: Oliveira e DeSouza, 2005.

Na equação (8), MT representa a saída de uma célula de MT; C_{LR} representa a saída da célula complexa binocular; C_L e C_R representam as saídas das células complexas monoculares dos olhos esquerdo e direito. Uma constante k foi adicionada a esta normalização para controlar a seletividade da célula.

$$MT = \frac{C_{LR}}{C_L + C_R + k} \quad (8)$$

Seguindo o esquema apresentado na Figura 5 (seção 2.1), a modelagem desta arquitetura foi feita utilizando várias camadas de células complexas, onde cada camada está ajustada para detectar uma determinada disparidade binocular.

Camadas sucessivas detectam disparidades crescentes sucessivamente, com diferença de distância de um pixel entre uma camada e a imediatamente anterior.

Qian e Zhu (1997) propuseram um modelo que considera uma cooperação e influência entre as células vizinhas de uma mesma camada. Isto foi modelado matematicamente no modelo de Oliveira e DeSouza (2005) aplicando uma convolução com uma função peso espacial, processo no qual é chamado de *spatial pooling*. A utilização do *spatial pooling* aprimora muito os resultados, porque suaviza a superfície formada pelas saídas das células vizinhas de uma mesma camada, eliminando ruídos.

2.3 Modelo proposto para a percepção de movimento

Experimentos psicofísicos indicam que há uma forte interação entre percepção estereoscópica de profundidade e percepção de movimento. Qian e Andersen (1997) propuseram um modelo integrado de célula que detecta simultaneamente a disparidade binocular e velocidade, usando filtros de Gabor espaço-temporais, com dimensões x , y e t .

Para efeito de simplificação, neste trabalho utilizamos um modelo em que o movimento e a profundidade são processados em estruturas neurais similares, porém cada uma especificamente sintonizada em um ou em outro tipo de percepção. A Figura 5 ilustra ambos os modelos. Ambos os processamentos utilizam estruturas análogas, com células simples e complexas em V1 e células de MT, o que facilita a implementação e a integração de todo o modelo. A Figura 12 representa esquematicamente as dimensões e as estruturas presentes no modelo.

Na parte inferior da Figura 12, vemos representadas as estruturas neurais para detecção de disparidades binoculares. Assumindo-se que o movimento ocular de vergência ocorre apenas em torno do eixo vertical, essas estruturas neurais devem estar sintonizadas para detectar distâncias específicas d_1 , d_2 , ..., d_n somente no eixo horizontal (x). A unidade de medida utilizada nessas estruturas é um pixel, que representa uma unidade de deslocamento entre as imagens dos olhos esquerdo e direito. Uma vez que a distância entre os olhos é uma constante conhecida pelo cérebro humano, conhecendo-se o deslocamento de um determinado ponto

presente nas imagens projetadas nas retinas esquerda e direita, tem-se condições para se calcular a profundidade Z real desse ponto, utilizando-se uma relação trigonométrica trivial.

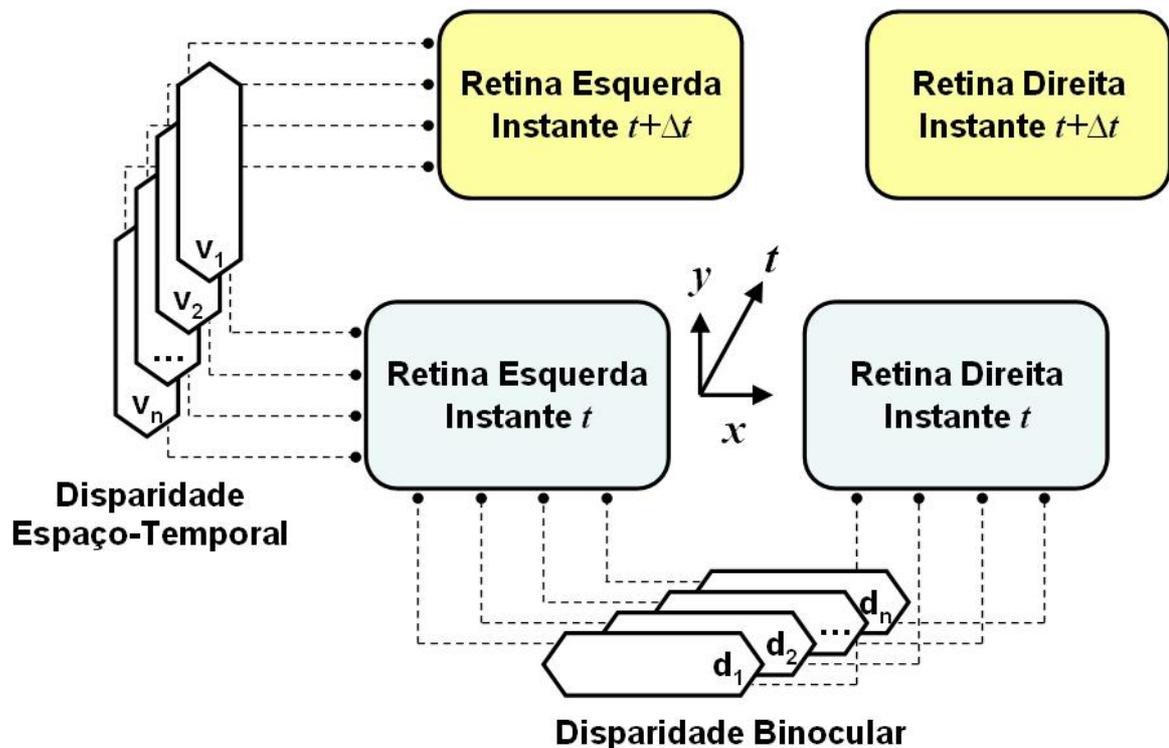


Figura 12 – Modelo proposto para percepção de movimento e profundidade. As estruturas neurais sintonizadas na detecção de disparidades binoculares (d_1, d_2, \dots, d_n) possibilitam a percepção de profundidade, enquanto as estruturas sintonizadas na detecção de disparidades espaço-temporais (v_1, v_2, \dots, v_n) possibilitam a percepção de velocidade.

As estruturas de disparidade binocular processam as imagens captadas pelas retinas esquerda e direita num mesmo instante de tempo. Consideramos o mesmo modelo utilizado por Oliveira e DeSouza (2005), conforme Figura 13. Neste modelo, as células S_L e S_R são células simples monoculares de V1, sintonizadas respectivamente no mapeamento log-polar das retinas esquerda e direita, guardando entre si uma disparidade binocular específica d , e com aplicação de filtros de Gabor com iguais orientação e frequência, porém com diferença de fases igual a π . As células S_L^Q e S_R^Q são células análogas a S_L e S_R , porém em quadratura de fase, ou seja, com diferença de fases igual a $\pi/2$. S_{LR} é uma célula simples binocular de V1, com característica *tuned inhibitory*, que computa as saídas das células simples monoculares S_L e S_R . S_{LR}^Q é a sua correspondente em quadratura

de fase. C_L é uma célula complexa monocular de V1, que computa os quadrados das saídas das células simples monoculares S_L e S_L^Q . C_R é análoga a C_L , para o olho direito. C_{LR} é uma célula complexa binocular de V1, com característica *tuned inhibitory*, que computa os quadrados das saídas das células simples binoculares S_{LR} e S_{LR}^Q . MT é uma célula de MT, que normaliza a saída de C_{LR} pela soma das saídas de C_L e C_R , conforme apresentado na seção anterior.

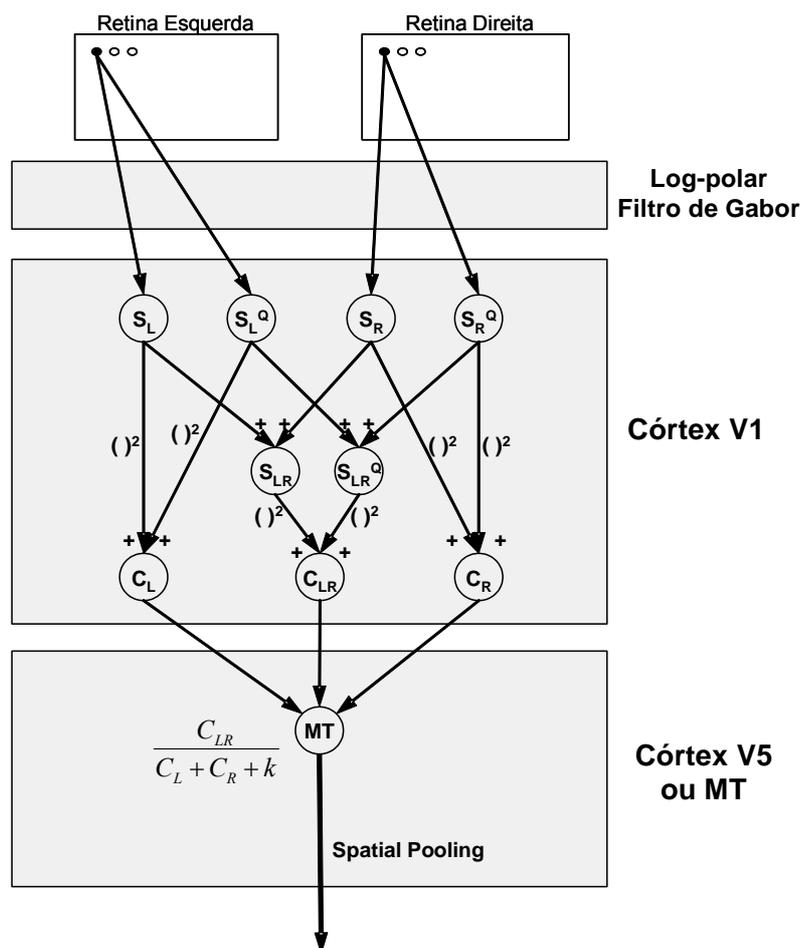


Figura 13 – Modelo representando uma estrutura neural de V1 e MT sintonizada para detecção de uma disparidade binocular específica
Fonte: Oliveira e DeSouza, 2005.

Para cada ponto da retina, o sistema neural processa os valores de saída das células de MT sintonizadas para detecção de cada disparidade binocular específica. Como o modelo é *tuned inhibitory*, a célula de MT que apresentar o menor valor de saída é aquela que representa a disparidade binocular mais ajustada àquele ponto da retina. Dessa forma, aplicando regras trigonométricas simples, o sistema consegue obter o valor da profundidade Z de cada ponto da imagem.

No lado esquerdo da Figura 12, vemos representadas as estruturas neurais para detecção de disparidades espaço-temporais. A imagem captada por cada retina é bidimensional, portanto essas estruturas devem estar sintonizadas para detectar velocidades específicas na forma de vetores no plano x-y, tais como: $v_1 = (v_{1x}, v_{1y})$, $v_2 = (v_{2x}, v_{2y})$ etc. A unidade de medida utilizada para representar a distância entre os pontos da retina é igualmente um pixel. A unidade de tempo Δt é o intervalo entre a captação de duas imagens sucessivas na retina. Para efeito de simplificação, assumiremos que esse intervalo é constante no sistema. Portanto, uma velocidade igual a $(-5, 10)$ significa um deslocamento de cinco pixels à esquerda e dez pixels para cima, no intervalo Δt . Um pixel representa também uma unidade de deslocamento angular $\Delta\theta$ no campo visual da retina. Dessa forma, uma vez conhecida a profundidade Z real de um ponto em relação ao observador, o cérebro humano tem condições de calcular a velocidade real desse ponto no espaço, utilizando uma relação trigonométrica trivial.

As estruturas de disparidade espaço-temporal processam as imagens captadas por uma retina em dois instantes de tempo consecutivos. Utilizamos o modelo mostrado na Figura 14. Neste modelo, as células S_1 e S_2 são células simples monoculares de V1, sintonizadas respectivamente no mapeamento log-polar de uma retina em dois instantes de tempo (t) e $(t+\Delta t)$, guardando entre si uma disparidade espaço-temporal específica $v = (v_x, v_y)$, e com aplicação de filtros de Gabor com iguais orientação e frequência, porém com diferença de fases igual a π . As células S_1^Q e S_2^Q são células análogas a S_1 e S_2 , porém em quadratura de fase, ou seja, com diferença de fases igual a $\pi/2$. S_T é uma célula simples temporal de V1, com característica *tuned inhibitory*, que computa as saídas das células simples monoculares S_1 e S_2 . S_T^Q é a sua correspondente em quadratura de fase. C_1 é uma célula complexa monocular de V1, que computa os quadrados das saídas das células simples monoculares S_1 e S_1^Q . C_2 é análoga a C_1 , para o instante de tempo $(t+\Delta t)$. C_T é uma célula complexa temporal de V1, com característica *tuned inhibitory*, que computa os quadrados das saídas das células simples temporais S_T e S_T^Q . MT é uma célula de MT, que normaliza a saída de C_T pela soma das saídas de C_1 e C_2 , de forma análoga às células de MT das estruturas de disparidade binocular.

Para cada ponto da retina, o sistema neural processa os valores de saída das células de MT sintonizadas para detecção de cada disparidade espaço-temporal

específica. Como o modelo é *tuned inhibitory*, a célula de MT que apresentar o menor valor de saída é aquela que representa a disparidade espaço-temporal mais ajustada àquele ponto da retina. Dessa forma, aplicando regras trigonométricas simples, o sistema consegue obter o valor da velocidade real de cada ponto da imagem.

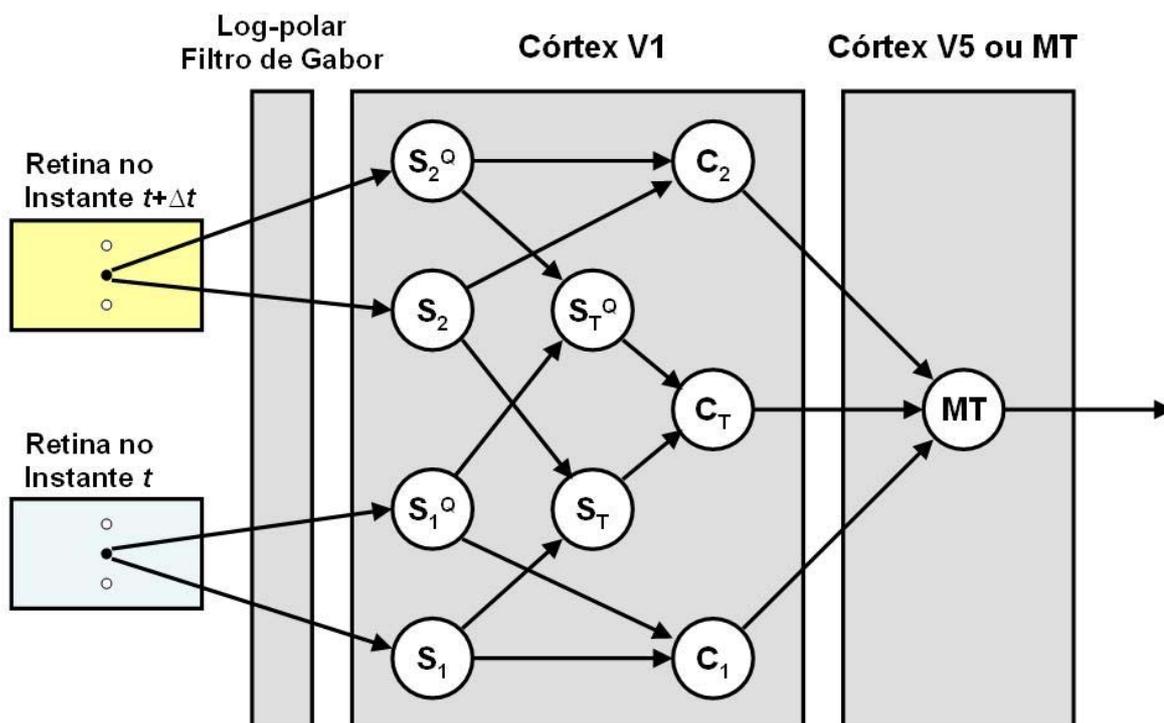


Figura 14 – Modelo representando uma estrutura neural de V1 e MT sintonizada para detecção de uma disparidade espaço-temporal específica

Outro aspecto a destacar na Figura 12 é que o modelo pode ser expandido de modo que também haja estruturas neurais sintonizadas na detecção de disparidades binoculares (d_1, d_2, \dots, d_n) nas imagens captadas pelas retinas esquerda e direita no instante de tempo ($t+\Delta t$). Dessa forma o modelo poderia calcular a profundidade Z real de cada ponto no instante ($t+\Delta t$) e, por meio de comparação com as profundidades calculadas no instante t , poderia obter a componente de velocidade no eixo z (v_z).

As estruturas neurais sintonizadas na detecção de disparidades espaço-temporais (v_1, v_2, \dots, v_n) podem estar acopladas tanto na retina esquerda quanto na direita.

3 METODOLOGIA

Nesta seção descrevemos a plataforma utilizada neste trabalho e os procedimentos adotados para a realização dos experimentos.

3.1 Plataforma

Para a implementação do modelo computacional de visão artificial capaz de emular a percepção de movimento, será utilizada a MAE (Máquina Associadora de Eventos), que é um *framework* aberto que permite projetar estruturas modulares usando redes neurais sem peso e estudar suas propriedades. Este *framework* foi desenvolvido na Universidade Federal do Espírito Santo e primeiramente utilizado por Komati e DeSouza (2002). A MAE é similar em alguns aspectos ao NRM (*Neural Representation Modeler*), desenvolvido pelo *Neural Systems Engineering Group*, do *Imperial College London*, e comercializado por *Novel Technical Solutions* [ALE89]. MAE difere do NRM em três principais aspectos: é *open source*, utiliza plataforma UNIX (Solaris ou Linux), e utiliza uma linguagem textual.

A MAE permite a descrição de estruturas neurais por meio de uma linguagem de alto nível e geração automática dos programas fonte e executável. Os módulos básicos de definição são camadas neurais bidimensionais e filtros de processamento associados a essas camadas. Os neurônios utilizados nas camadas neurais podem ser definidos com vários atributos que estabelecem comportamentos especiais (tipo, sensibilidade a cor, tamanho da memória etc.) e pode-se livremente criar as lógicas dos filtros de processamento utilizando a linguagem C.

O usuário do *framework* MAE especifica o modelo usando a linguagem NADL (*Neural Architecture Description Language*) e gravando um arquivo fonte com a extensão “.con”. O tradutor *netcomp* transforma esse arquivo “.con” num programa fonte em linguagem C. O script *netcompiler* executa a compilação desse programa fonte e de todos os programa C criados pelo usuário para tratar especificamente os filtros e a manipulação da rede neural. Em seguida ocorre a *linkagem* desses programas com a biblioteca MAE, gerando uma aplicação MAE executável, conforme mostrado na Figura 15.

A biblioteca MAE oferece uma representação visual semelhante ao OpenGL, amigável e flexível. Os pacotes adicionais utilizados pela MAE são:

- MESA 3-D: biblioteca gráfica 3D;
- GLUT: ferramentas para escrever programas em OpenGL, que também pode ser utilizado com o MESA 3D;
- XForms: ferramentas de interface gráfica com o usuário baseada no sistema de janelas X Window;
- Flex: gerador de código de analisadores léxicos;
- Bison: gerador de código de analisador sintático.

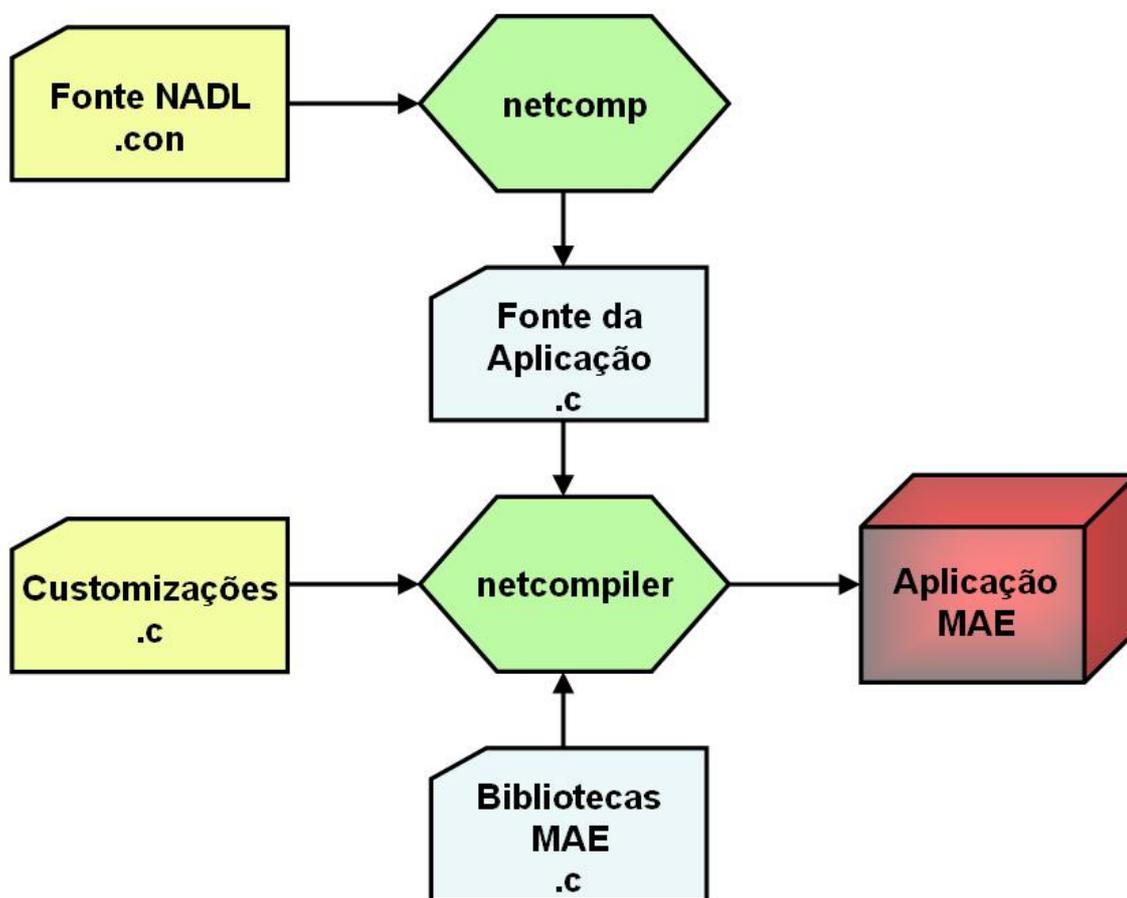


Figura 15 – Utilização do *framework* MAE

Neste trabalho os seguintes componentes foram utilizados na aplicação MAE:

- **neuronlayer** – implementa uma matriz de neurônios que formam uma camada neural. Foram utilizadas camadas neurais para representar cada grupo de células em V1 e MT: células simples monoculares e temporais, células complexas monoculares e temporais, e células de MT. A fim de limitar a ocupação de memória e o tempo de processamento, cada camada foi dimensionada em uma matriz de tamanho fixo de 128 x 64 neurônios, ocupando menos de 3% do tamanho da imagem originalmente captada pela retina. A saída dos neurônios foi configurada do tipo *greyscale_float*, ou seja, a saída pode ser um valor real e é representada em 256 tons de cinza. Quanto mais escuro, mais próximo de zero é o valor; quanto mais claro, maior é o valor. Se houver valores de saída positivos e negativos numa mesma camada, então serão representados respectivamente por 256 tons de verde e vermelho.
- **input** – implementa uma camada especial, que contém uma entrada para a rede neural. Associada a essa camada é criada automaticamente uma *neuronlayer* para armazenar os valores dos estímulos de entrada e uma janela para o usuário poder visualizar o estado desses estímulos. Neste trabalho foram utilizadas duas *inputs* com tamanho fixo de 640 x 480 pixels cada: uma para representar a imagem captada pela retina no instante t e outra para representar a imagem nessa mesma retina no instante $(t+\Delta t)$. Essas *inputs* admitem pixels coloridos, mas, para efeito de simplificação, neste trabalho os estímulos coloridos foram convertidos em estímulos em tons de cinza, calculando-se a média aritmética dos valores dos três componentes RGB (*red*, *green*, *blue*) de cada pixel.
- **output** – implementa uma janela de visualização das saídas de uma *neuronlayer*. Neste trabalho, cada *neuronlayer* possui uma *output* associada.
- **outputconnect** – estabelece a conexão de uma *output* a uma *neuronlayer*.
- **filter** – define que uma ou mais *inputs* ou *neuronlayers* terão seus estímulos ou suas saídas processadas por um filtro, e o resultado desse processamento será armazenado em outra *neuronlayer*. Neste trabalho foram implementados *filters* para o processamento do mapeamento log-polar, da função de Gabor, do cômputo das células simples, complexas e de MT.

3.2 Procedimentos

O escopo deste trabalho restringiu-se à implementação e teste do modelo de percepção de movimento, representado na Figura 14.

Os seguintes passos foram utilizados no processo de detecção de movimento e identificação das velocidades de cada ponto das imagens:

- Apresentar duas imagens ao sistema, captadas em dois instantes distintos;
- Identificar a região de movimento entre as duas imagens;
- Emular um movimento ocular de sacada para a região de movimento;
- Computar todas as camadas MT sintonizadas com cada uma das disparidades espaço-temporais;
- Criar um mapa de velocidades, indicando a velocidade mais ajustada para cada ponto da imagem;
- Representar visualmente a velocidade de cada ponto na imagem;
- Comparar os resultados obtidos com os esperados.

3.2.1 Captação das imagens

A captação das imagens foi feita por meio de uma câmera digital Sony DSC-P150, com gravação de arquivos em formato padrão MPEG-1 (*Moving Picture Experts Group*), com resolução de 640 x 480 pixels, 24 bits de cores e taxa de 25 ou 30 quadros por segundo. A fim de desconsiderar os movimentos oculares compensatórios de rotações da cabeça (vestíbulo-ocular e optocinético), a câmera foi mantida numa posição fixa em relação ao solo, durante a captação das imagens.

A fim de possibilitar ao sistema trabalhar as imagens como matrizes de pixels, cada quadro do filme foi primeiramente convertido em um *bitmap* com 24 bits de cores por pixel, utilizando uma função de decodificação de imagens MPEG-1. Antes de serem processadas pelo filtro de Gabor, cada quadro foi então convertido para um *bitmap* em que cada pixel possui um valor de intensidade com 256 níveis de variação. Essa conversão foi necessária, para que o filtro de Gabor pudesse operar sobre valores escalares dos estímulos apresentados à retina, em vez de valores tridimensionais RGB (*red, green, blue*).

O módulo de captação de imagens possibilita também que sejam selecionados quadros não consecutivos do filme, permitindo assim alterar a constante Δt utilizada para o cômputo das velocidades. Permite também utilizar seqüências de imagens em formato JPEG (*Joint Photographic Experts Group*).

3.2.2 Identificação da região de movimento

Conforme o modelo proposto na Figura 12, o sistema trabalha com várias estruturas em camadas paralelas para detecção de disparidades espaço-temporais, cada uma sintonizada numa velocidade específica $(v_{1x}, v_{1y}), (v_{2x}, v_{2y}), \dots, (v_{nx}, v_{ny})$. Dessa forma, usando-se a camada que está sintonizada para detecção da disparidade nula ($v_x = 0, v_y = 0$), podemos identificar na imagem quais são as regiões de repouso e as regiões de movimento.

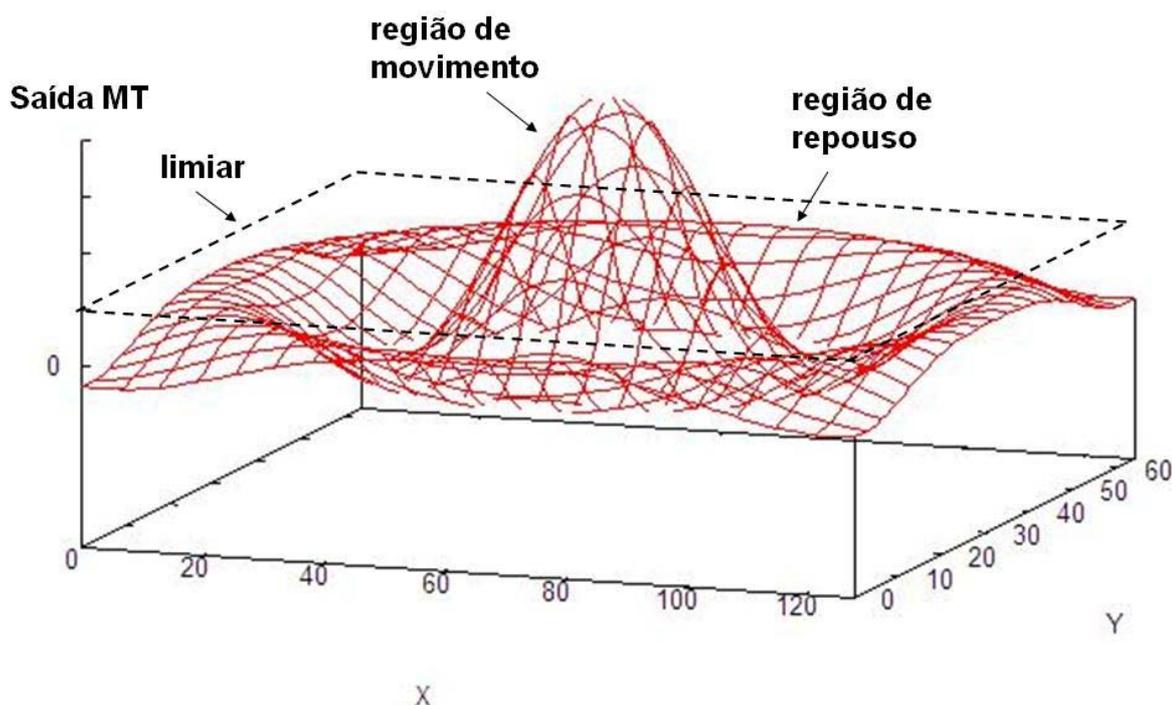


Figura 16 – Exemplo de gráfico que representa a saída das células de MT para cada ponto (x, y) da imagem. Uma vez que esta camada de células esteja sintonizada para detecção de disparidade nula, pode-se identificar as regiões de movimento e de repouso na imagem.

Uma vez que o modelo é *tuned inhibitory*, as células de MT da camada sintonizada para detecção da disparidade nula que apresentarem valores mais próximos de zero (abaixo do limiar L) serão aquelas que indicarão pontos com velocidade nula na

imagem. Valores acima do limiar L indicarão pontos com velocidade não nula, conforme exemplo mostrado na Figura 16.

O valor de limiar foi determinado empiricamente durante os experimentos como sendo $L = 0,1$. Este foi o valor que representou com maior confiabilidade a separação entre as regiões de repouso e de movimento. Um método alternativo para a determinação do valor de limiar seria a utilização de um algoritmo que testasse diversos valores e calculasse as médias das saídas das células em cada região. O valor de limiar que proporcionasse a maior distância entre as médias, seria o escolhido. É importante observar que o mecanismo biológico de ativação das células neurais é igualmente baseado em valores de limiar. A utilização de drogas psicotrópicas, como o álcool, altera a química cerebral e conseqüentemente altera os valores de limiar das células responsáveis pela percepção de movimento e profundidade, prejudicando a acurácia dessas funções cerebrais.

A identificação da região de movimento na imagem é importante por dois motivos. Primeiramente porque estabelece qual é a região que merece maior atenção do cérebro e viabiliza a realização do movimento ocular de sacada, conforme abordado na próxima etapa. Em segundo lugar porque é um passo relevante para um processamento de mais alto nível realizado pelo cérebro: a identificação de objetos e a construção de um modelo simbólico do mundo real. Segundo Hoffman (2000), “construir movimentos e construir objetos caminham lado a lado”. Ao definir que regiões apresentam movimento numa imagem, o cérebro identifica elementos que se destacam em relação à paisagem de fundo e que, portanto, devem ser elaborados separadamente na mente humana.

3.2.3 Emulação do movimento ocular de sacada

Na visão humana, o cérebro escolhe um ponto de atenção visual e direciona os olhos de modo que a imagem deste ponto de atenção fique projetada na retina exatamente sobre a região da fóvea. Este movimento ocular é denominado *sacada* e é realizado com extrema rapidez, numa velocidade angular de até 900° por segundo. O efeito de situar o ponto de atenção visual sobre a fóvea é fazer uma magnificação dos detalhes da imagem ao redor daquele ponto, permitindo uma exploração visual muito mais acurada. Conforme mencionado na seção anterior, cerca de metade das células de V1 é dedicada aos estímulos provenientes da fóvea.

No caso da percepção de movimento, obviamente a escolha do ponto de atenção visual recairá justamente na região do campo visual em que o movimento tenha sido detectado com maior intensidade. Esta região foi determinada na etapa descrita no tópico anterior.

O movimento de sacada foi implementado no sistema por meio de uma função que calcula o *centro de massa* da região de movimento, sendo o peso de cada ponto substituído pelo valor de saída da célula correspondente em MT. Em outras palavras, para o conjunto de células de MT que possuem valor de saída superior ao limiar L , calcula-se a média ponderada das coordenadas (x,y) dos pontos correspondentes localizados na retina, usando como fator de ponderação o valor de saída de cada célula de MT. As equações (9) mostram esse cálculo das coordenadas de sacada.

$$\forall c \in MT : saída_c > L \therefore x_{SACADA} = \frac{\sum_c x_c * saída_c}{\sum_c saída_c}, y_{SACADA} = \frac{\sum_c y_c * saída_c}{\sum_c saída_c} \quad (9)$$

3.2.4 Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida

O modelo proposto trabalha com estruturas modulares, sendo cada uma delas dedicada a captar uma determinada disparidade espaço-temporal, ou seja, uma determinada velocidade (v_x, v_y) . Na Figura 14 está representado o esquema de um módulo. Em princípio não há limite fixado para a quantidade de módulos que o sistema terá – o limite é a capacidade de memória e de processamento existente no sistema. Neste trabalho o processamento dos módulos foi feito de forma seqüencial, a fim de tornar mais simples a sua implementação. Idealmente o processamento dos módulos poderia ser totalmente paralelizado.

Além da velocidade nula $(0,0)$, neste trabalho foram criadas estruturas para detecção de 64 velocidades predefinidas, conforme apresentado na Tabela 1. Esse conjunto de velocidades foi escolhido de modo a abranger uma gama variada de direções e valores, suficientemente capaz de testar a validade do modelo. Caso isso não fosse verificado, estruturas adicionais poderiam ser acrescentadas.

Uma vez processados os valores de saída da camada MT, para cada estrutura espaço-temporal predefinida, em cada ponto da imagem na região de movimento, esses valores são guardados em memória, para serem utilizados na etapa seguinte.

Tabela 1 – Velocidades (v_x, v_y) predefinidas, expressas em pixels por Δt

							
(4, 0)	(4, 4)	(0, 4)	(-4, 4)	(-4, 0)	(-4, -4)	(0, -4)	(4, -4)
(8, 0)	(8, 8)	(0, 8)	(-8, 8)	(-8, 0)	(-8, -8)	(0, -8)	(8, -8)
(12, 0)	(12, 12)	(0, 12)	(-12, 12)	(-12, 0)	(-12, -12)	(0, -12)	(12, -12)
(16, 0)	(16, 16)	(0, 16)	(-16, 16)	(-16, 0)	(-16, -16)	(0, -16)	(16, -16)
(20, 0)	(20, 20)	(0, 20)	(-20, 20)	(-20, 0)	(-20, -20)	(0, -20)	(20, -20)
(24, 0)	(24, 24)	(0, 24)	(-24, 24)	(-24, 0)	(-24, -24)	(0, -24)	(24, -24)
(28, 0)	(28, 28)	(0, 28)	(-28, 28)	(-28, 0)	(-28, -28)	(0, -28)	(28, -28)
(32, 0)	(32, 32)	(0, 32)	(-32, 32)	(-32, 0)	(-32, -32)	(0, -32)	(32, -32)

3.2.5 Criação do mapa de velocidades

As células de MT modeladas neste trabalho são *tuned inhibitory*, ou seja, as respostas são mínimas quando um estímulo esteja sendo projetado em posições equivalentes nos campos receptivos nos instantes t e $(t+\Delta t)$. A escolha da disparidade espaço-temporal mais ajustada a cada ponto da imagem baseia-se nesse princípio de competição entre as saídas das estruturas de MT. Aquela estrutura que apresentar o menor valor de saída será a vencedora. Valores de saída acima do limiar L serão desprezados, porque não apresentam um grau de confiabilidade minimamente aceitável.

A criação do mapa de velocidades parte da premissa de que os pontos presentes na imagem podem estar em movimentos diferentes, com velocidades distintas. Por isso faz-se necessário que o processo de seleção da disparidade vencedora seja feito separadamente para cada ponto da imagem. A reunião de todas as disparidades

vencedoras na imagem formará um mapa de velocidades, no plano de coordenadas $u-v$ (log-polar), que será utilizado na etapa seguinte.

3.2.6 Representação visual da velocidade de cada ponto na imagem

O mapa de velocidades foi utilizado para representar visualmente a velocidade de cada ponto. Para isso, a representação no plano de coordenadas log-polares de MT foi convertida novamente para coordenadas cartesianas da retina. A região de movimento foi destacada na imagem e a região de repouso foi apagada.

Dentro da região de movimento, foram representados somente os pontos cuja determinação de velocidade atingiu a confiabilidade mínima. O critério de confiabilidade utilizado foi o valor da saída vencedora de MT ser inferior ao valor de limiar L .

Usando o *princípio da rigidez*, o sistema visual humano não necessita de conhecer a velocidade de todos os pontos de um objeto rígido, para construir mentalmente o seu movimento [HOF00]. Johansson (1973) demonstrou esse princípio com um experimento simples, porém muito convincente, representado na Figura 17.



Figura 17 – Experimento de Johansson com as luzes andantes. À esquerda há um conjunto de pontos luminosos; à direita a representação da conexão desses pontos, construída mentalmente. Fonte: McIntyre, 2003.

Foram colocadas pequenas luzes nas articulações de dançarinos e estes foram filmados no escuro. Quando os dançarinos estavam parados, via-se somente um conjunto aleatório de doze pontos brilhantes. Mas quando os dançarinos se mexiam, suas formas e ações eram facilmente vistas, a partir dos movimentos dos pontos. O

que é notável em relação a esse experimento é o quanto o cérebro humano cria, a partir de tão pouca informação.

Partindo-se desse princípio, é razoável assumir que basta ao sistema conhecer a velocidade de um pequeno conjunto de pontos de um objeto em movimento. A partir desses pontos de velocidade conhecida, o sistema poderá derivar a velocidade dos demais pontos de interligação.

A velocidade dos pontos conhecidos é representada no sistema por uma flecha na direção do movimento, com tamanho proporcional ao valor da velocidade.

3.2.7 Comparação dos resultados obtidos com os esperados

Os experimentos foram concebidos de forma a permitir a comparação dos resultados obtidos com os esperados.

Uma vez que o sistema esteja capacitado a reconhecer a velocidade de cada objeto, será viabilizada a emulação do movimento ocular de *perseguição suave*, descrito na seção anterior.

4 EXPERIMENTOS

Usando a metodologia descrita na seção anterior, foi realizada uma série de experimentos. Pela análise dos resultados, foram validadas a arquitetura e a metodologia adotadas, conforme relatado a seguir.

O primeiro experimento teve por objetivo perceber o movimento de um único objeto se deslocando apenas na dimensão horizontal. Esse objeto possui uma sombra projetada na parede e um reflexo sobre a superfície em que se encontra. A sombra e o reflexo acompanham o objeto na mesma velocidade. O segundo experimento objetivou perceber o movimento diagonal de um objeto, ou seja, com componentes de velocidade em duas dimensões: horizontal e vertical. Além disso, esse experimento apresentou uma seqüência de quatro imagens, com as quais o sistema devia ser capaz de emular movimentos de sacada e perseguição suave. O terceiro experimento objetivou perceber movimentos distintos e simultâneos de quatro objetos numa mesma cena. Esses objetos se deslocavam em movimentos de translação com direções e valores de velocidade variados. O último experimento teve por objetivo perceber o movimento de rotação de um objeto.

A Tabela 2 apresenta os valores utilizados em todos os experimentos como parâmetros dos filtros.

Tabela 2 – Parâmetros dos filtros utilizados em todos os experimentos

Filtro	Parâmetro	Valor
biological_gabor_filter	camera_opening_degree	47.05
"	cycles_per_degree	0.85
"	band_width	2.0
"	aspect_ratio	1.0
"	orientation	0.0
"	phase	$t_1: \pi/4.0, t_1^Q: 3.0*\pi/4.0,$ $t_2: -3.0*\pi/4.0, t_2^Q: -\pi/4.0$
"	shift	0
"	mapping	MAP_LOG_POLAR
"	log_factor	2.0
"	kernel_type	KERNEL_2D
robot_mt_cell	mt_sensibility	0.1
robot_gaussian_filter	kernel_size	9
"	sigma	2.0
robot_motion_area_filter	log_factor	2.0
"	threshold_0	0.1

Em todos os experimentos, foi utilizado um microcomputador Toshiba com processador Intel Pentium M 1.6 GHz, com memória RAM de 512 megabytes e sistema operacional Linux Fedora Core 2.6.5-1.358. O tempo total de execução foi de aproximadamente dois minutos para cada experimento.

4.1 Percepção de movimento horizontal de um objeto

Neste experimento, uma estatueta de cerâmica foi colocada sobre uma superfície reflexiva e um fundo opaco. No instante seguinte, esse objeto se movimentou 5 mm à direita, correspondendo a cerca de 15 pixels na imagem. As Figuras 18 e 19 apresentam as cenas capturadas nos dois instantes.

A Figura 20 mostra que, devido ao afastamento em relação ao centro de atenção, somente uma pequena parte do objeto se encontra dentro do campo visual.

A Figura 21 representa o mapeamento log-polar do campo visual, em torno do centro de atenção, nos dois instantes. Esse mapeamento, que é característico da camada V1 do sistema visual, faz com que as áreas distantes do centro de atenção tenham uma representação pequena no conjunto total de células. Neste exemplo, a parte do objeto presente no campo visual corresponde apenas a cerca de 1% da área total de V1.

Identificação da região de movimento

Usando a estrutura de detecção de disparidade espaço-temporal nula, podemos identificar na imagem quais são as regiões de repouso e as regiões de movimento. A Figura 22 mostra a saída da camada MT para essa estrutura. A área mais clara corresponde às células com valor de saída maior; a área escura corresponde às células com valor de saída nulo ou bem pequeno.

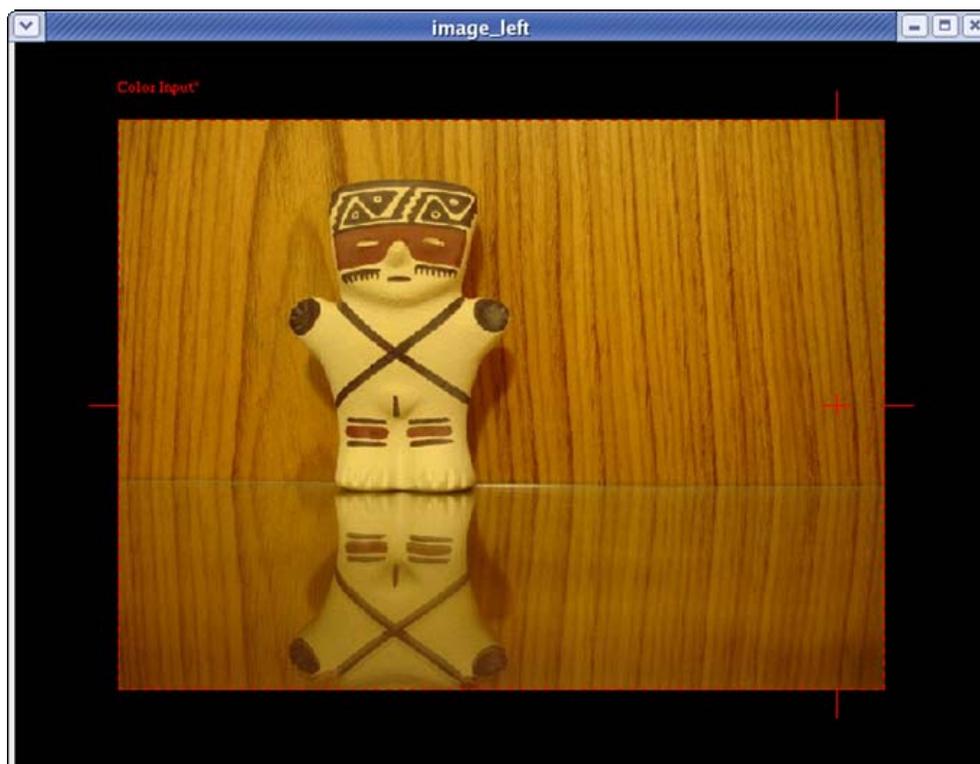


Figura 18 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (600, 240).

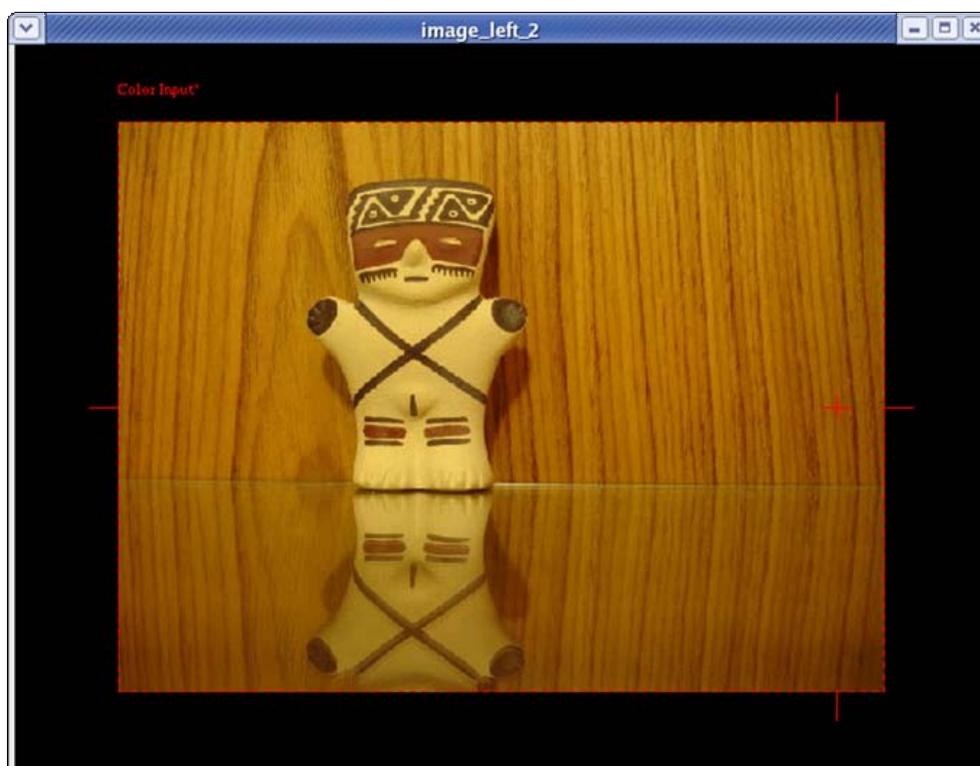


Figura 19 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada. O objeto se movimentou cerca de 15 pixels à direita.

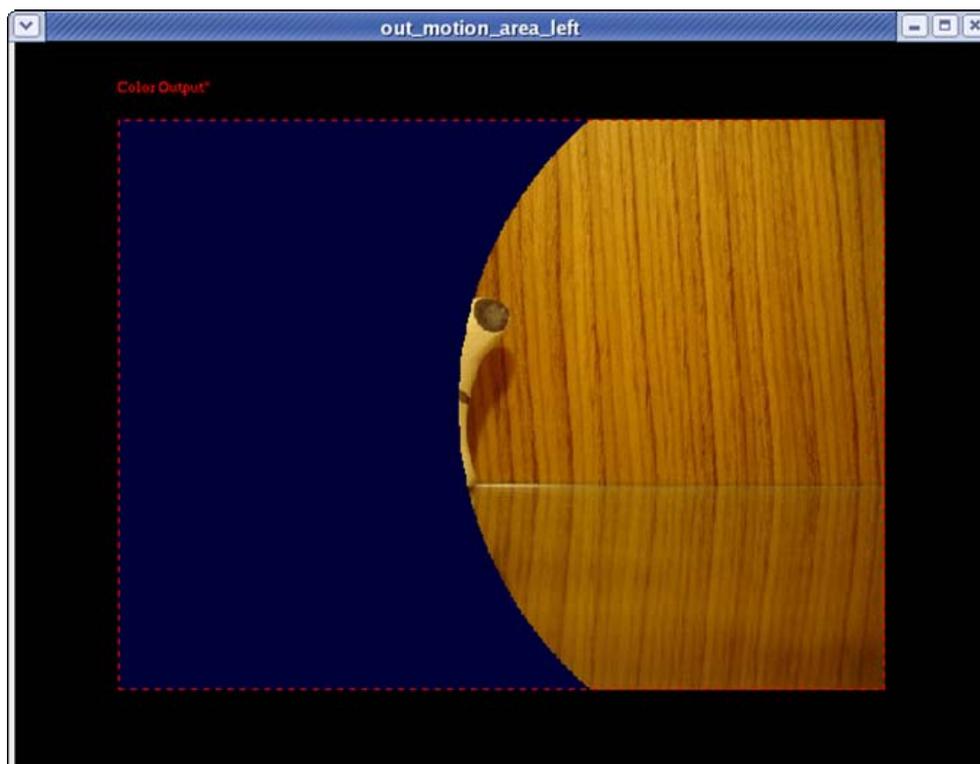


Figura 20 – Círculo que representa o campo visual em torno do centro de atenção em (600, 240).



Figura 21 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.

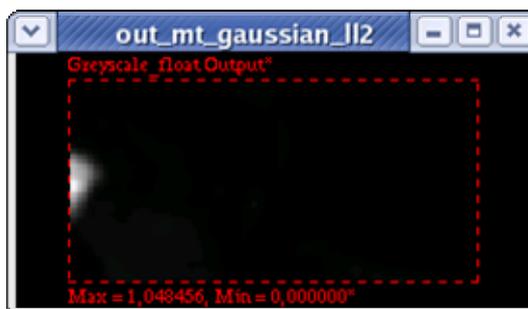


Figura 22 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.

Mesmo tendo no campo visual apenas uma pequena fração do objeto, o sistema foi capaz de separar com boa precisão as regiões de repouso e de movimento entre as duas cenas apresentadas. A região detectada com velocidade nula foi apagada na Figura 23, ficando representado apenas o objeto que se movimentou e uma região de incerteza, que corresponde ao contorno do objeto.

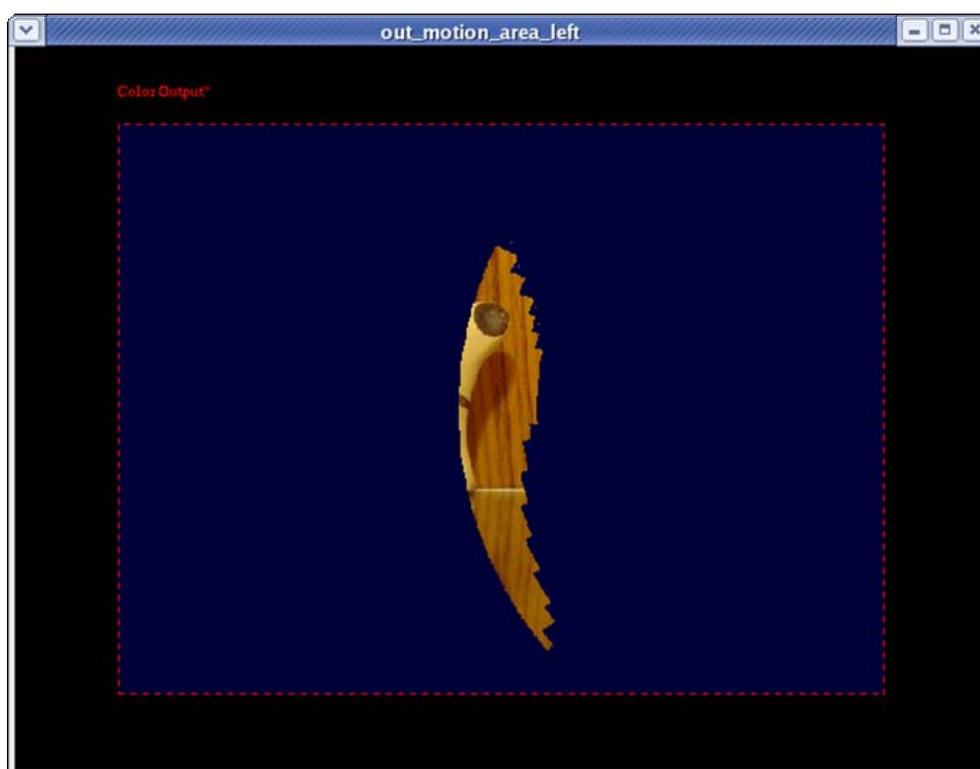
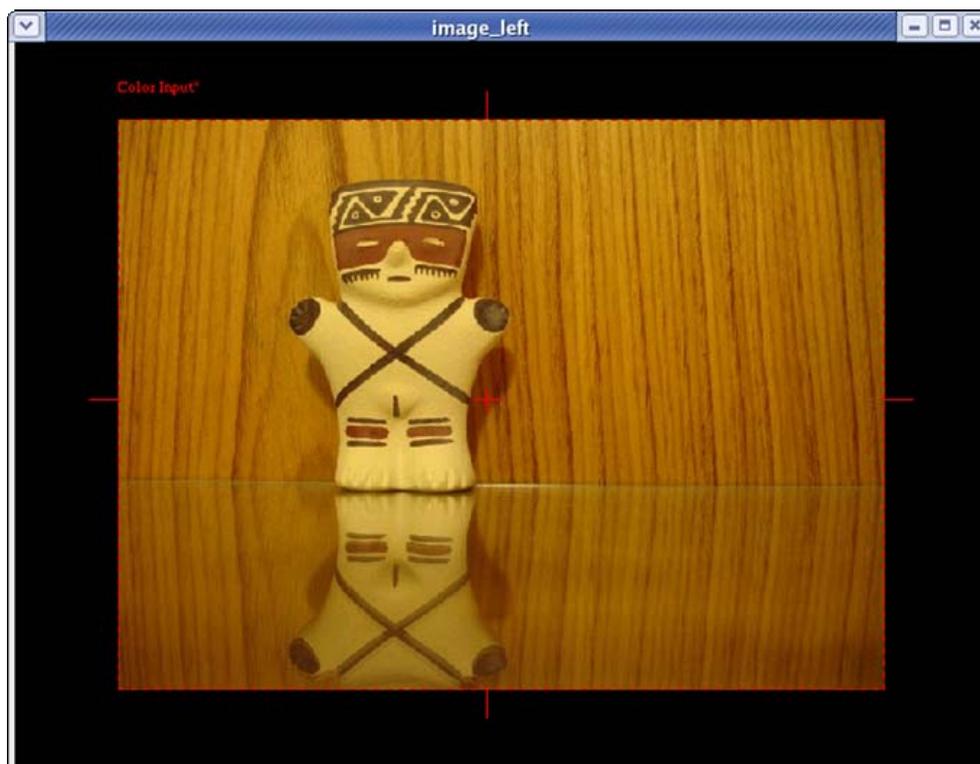


Figura 23 – Representação da região de movimento ou de incerteza, centro de atenção (600, 240).

Emulação do movimento ocular de sacada

A fim de magnificar a representação do objeto em V1, emulamos um movimento ocular de sacada, colocando a região de movimento na fóvea. Neste exemplo, a operação de sacada deslocou o centro de atenção para a coordenada (308, 244).

O resultado dessa operação no sistema é apresentado na Figura 24. Observa-se que o objeto presente no campo visual passou a ocupar cerca de 30% da área total de V1, permitindo com isso uma representação muito maior de seus detalhes.



(a)



(b)



(c)

Figura 24 – (a) Cena no instante 1. O centro de atenção foi deslocado para a coordenada (308, 244), por meio da emulação do movimento ocular de sacada. Em (b) e (c) vemos o mapeamento log-polar das imagens, em torno do novo centro de atenção, nos instantes 1 e 2, após a sacada.

Este nível de detalhe já seria suficiente para permitir ao sistema trabalhar a percepção de velocidade com um razoável grau de precisão. Contudo, a fim de aprimorar, realizamos mais alguns movimentos de sacada, até convergir para o centro de atenção final, localizado nas coordenadas (246, 239).

A Tabela 3 mostra a seqüência de convergência das sacadas. As Figuras 25 e 26 mostram o resultado para o centro de atenção final.

Tabela 3 – Experimento 1: Seqüência de convergência das sacadas

Sacada	Deslocamento	Coordenadas
1	(-292, +4)	(308, 244)
2	(-34, -4)	(274, 240)
3	(-15, -1)	(259, 239)
4	(-7, 0)	(252, 239)
5	(-3, 0)	(249, 239)
6	(-2, 0)	(247, 239)
7	(-1, 0)	(246, 239)
8	(0, 0)	(246, 239)

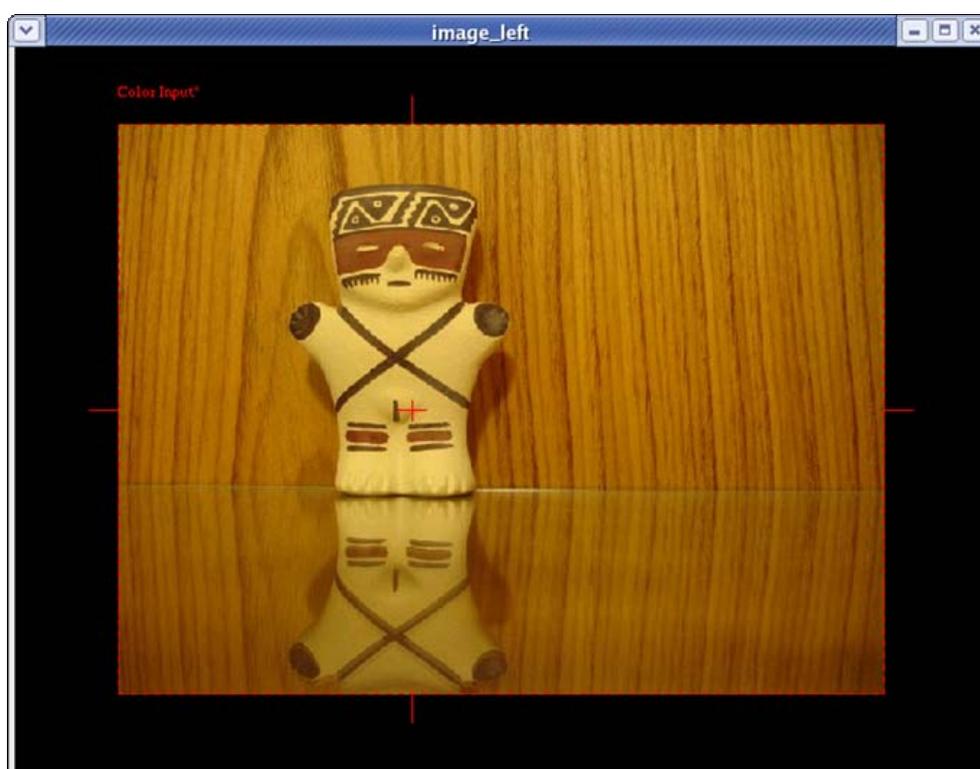


Figura 25 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (246, 239), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.

A Figura 26 mostra que, no centro de atenção final, o objeto de interesse ocupa cerca de 50% da área total de V1.

Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida

O processamento de cada camada MT foi feito por meio de uma estrutura em camadas, conforme a Figura 14.

Foram processadas 65 estruturas separadamente, sendo uma estrutura sintonizada com a velocidade nula e as demais sintonizadas em cada uma das velocidades relacionadas na Tabela 1. A título de exemplificação, mostramos nas Figuras 27 a 29 os resultados das camadas intermediárias somente para a estrutura sintonizada na velocidade nula.

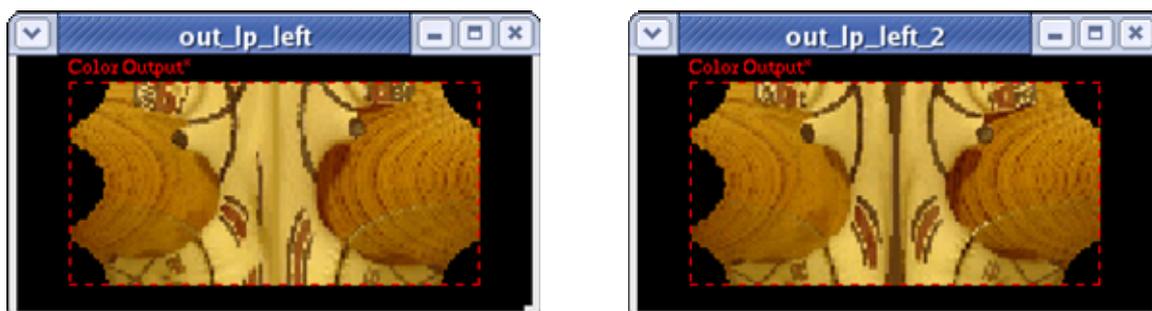


Figura 26 – Mapeamento log-polar das imagens, em torno do centro de atenção (246, 239), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.

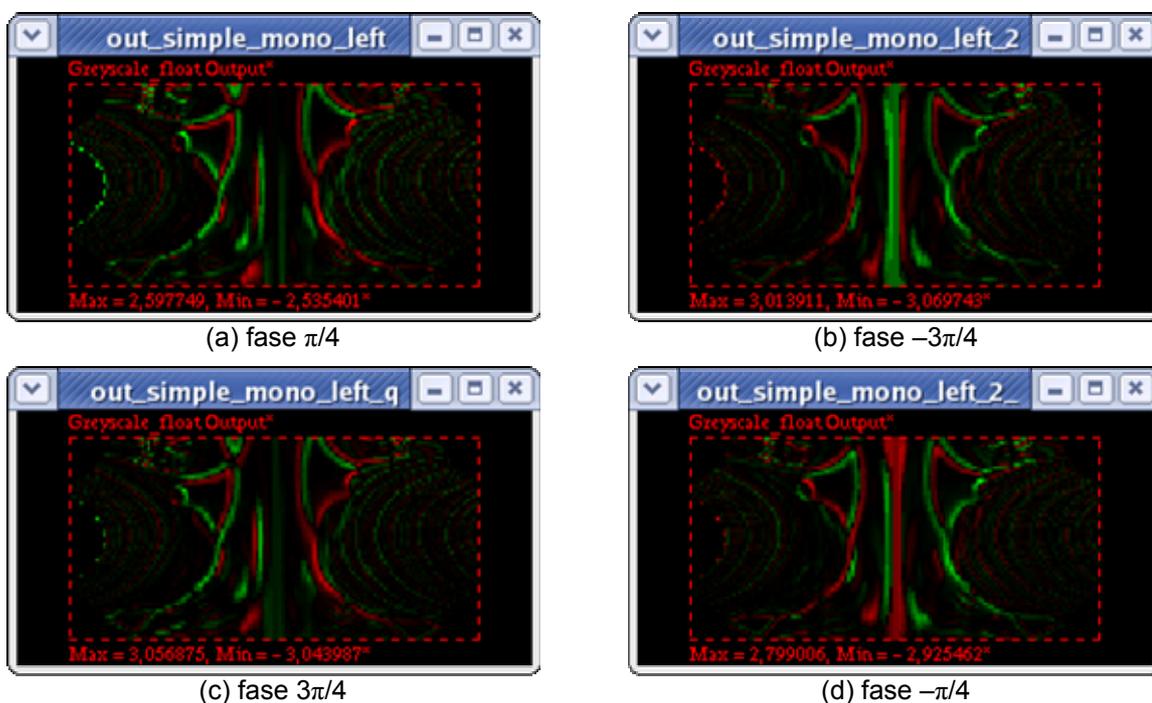


Figura 27 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase.

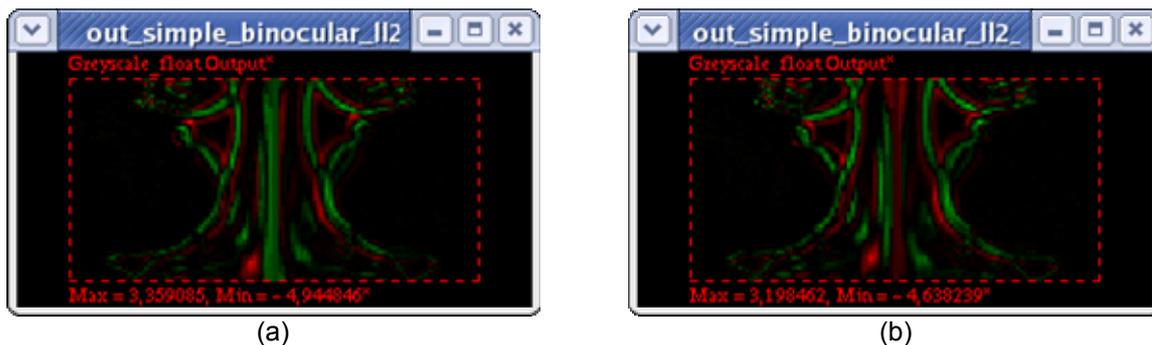


Figura 28 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.

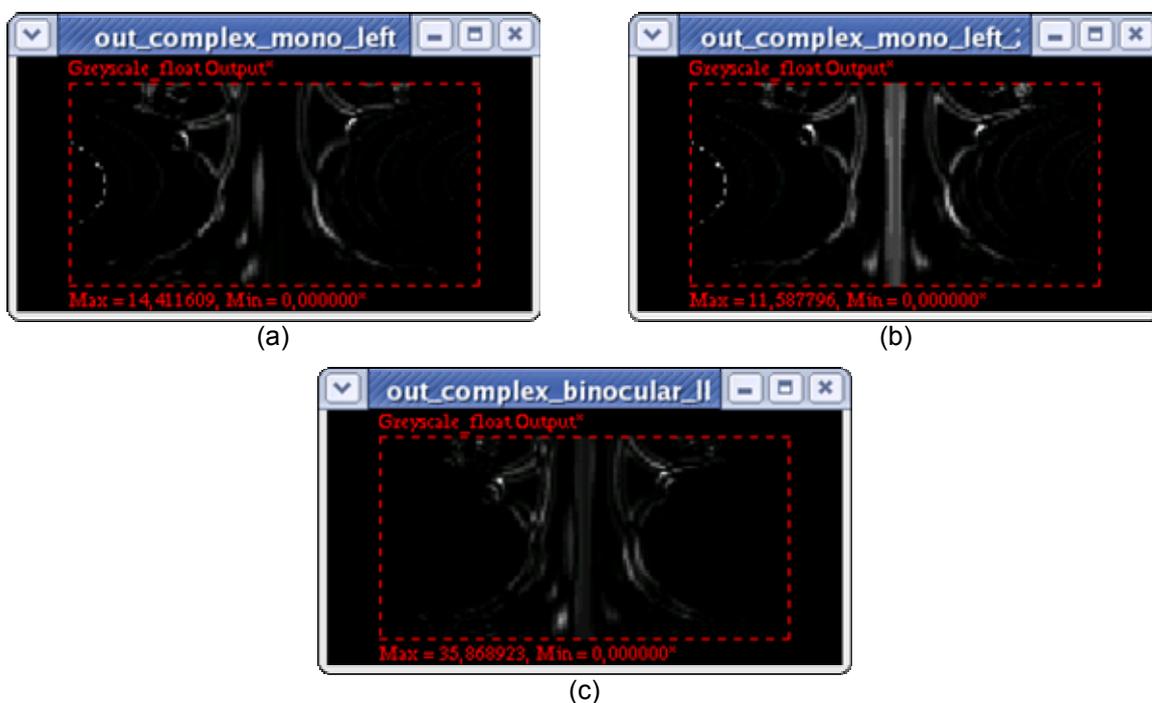


Figura 29 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.

Na Figura 30, apresentamos o resultado da camada MT sintonizada na velocidade nula, em duas versões: (a) antes de aplicar o algoritmo de *spatial pooling*; (b) após aplicar o algoritmo de *spatial pooling*. O efeito visivelmente notável é a suavização da função de saída de MT, por meio da cooperação entre as células vizinhas nesta camada.

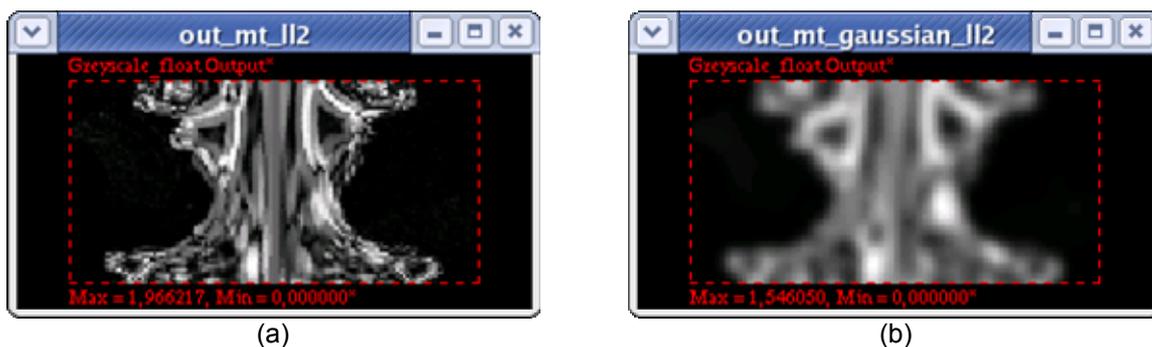


Figura 30 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com *spatial pooling*.

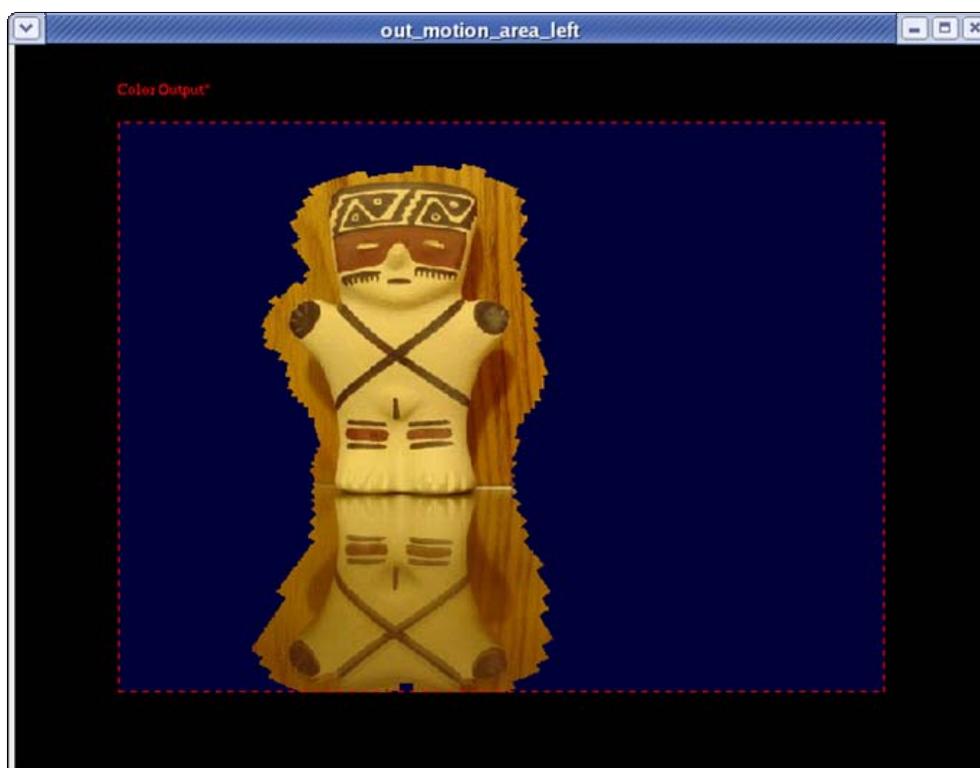


Figura 31 – Representação da região de movimento ou de incerteza, centro de atenção (246, 239).

Utilizando uma função de limiar sobre a saída de MT, com valor de limiar $L = 0,1$, chegamos à separação das regiões de repouso e de movimento, conforme mostrado na Figura 31. Observamos que houve uma melhora significativa em relação ao resultado inicial obtido na Figura 23. Outro aspecto importante é que a região do reflexo da estatueta sobre a superfície também é uma imagem que se move, e isso foi corretamente detectado pelo sistema.

Criação do mapa de velocidades

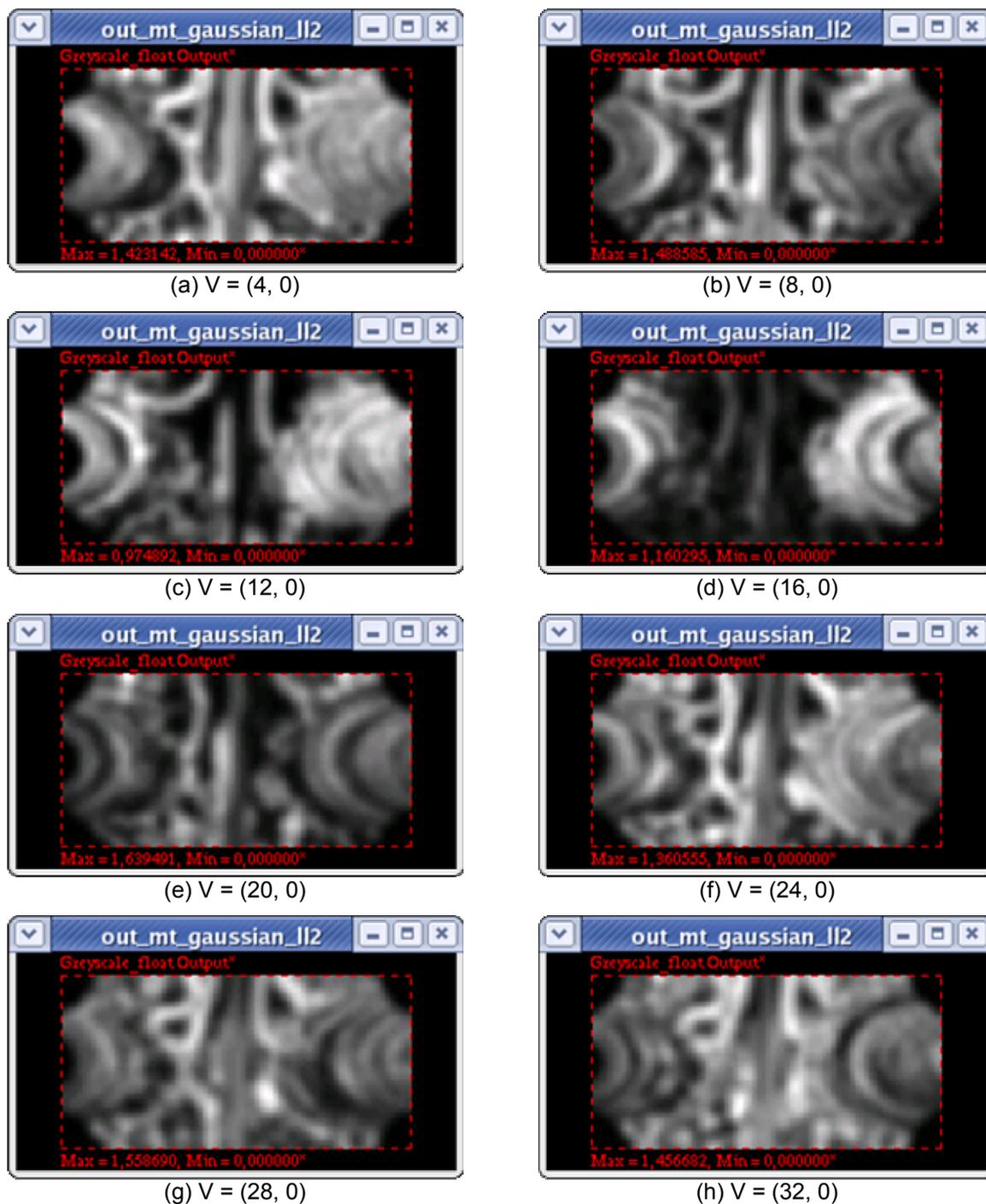


Figura 32 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, no movimento horizontal, sentido para a direita.

As Figuras 30-b, 32 e 33 mostram as camadas MT sintonizadas em 16 velocidades distintas, para ilustração do processo de criação do mapa de velocidades. O processo completo neste experimento comparou 65 camadas MT.

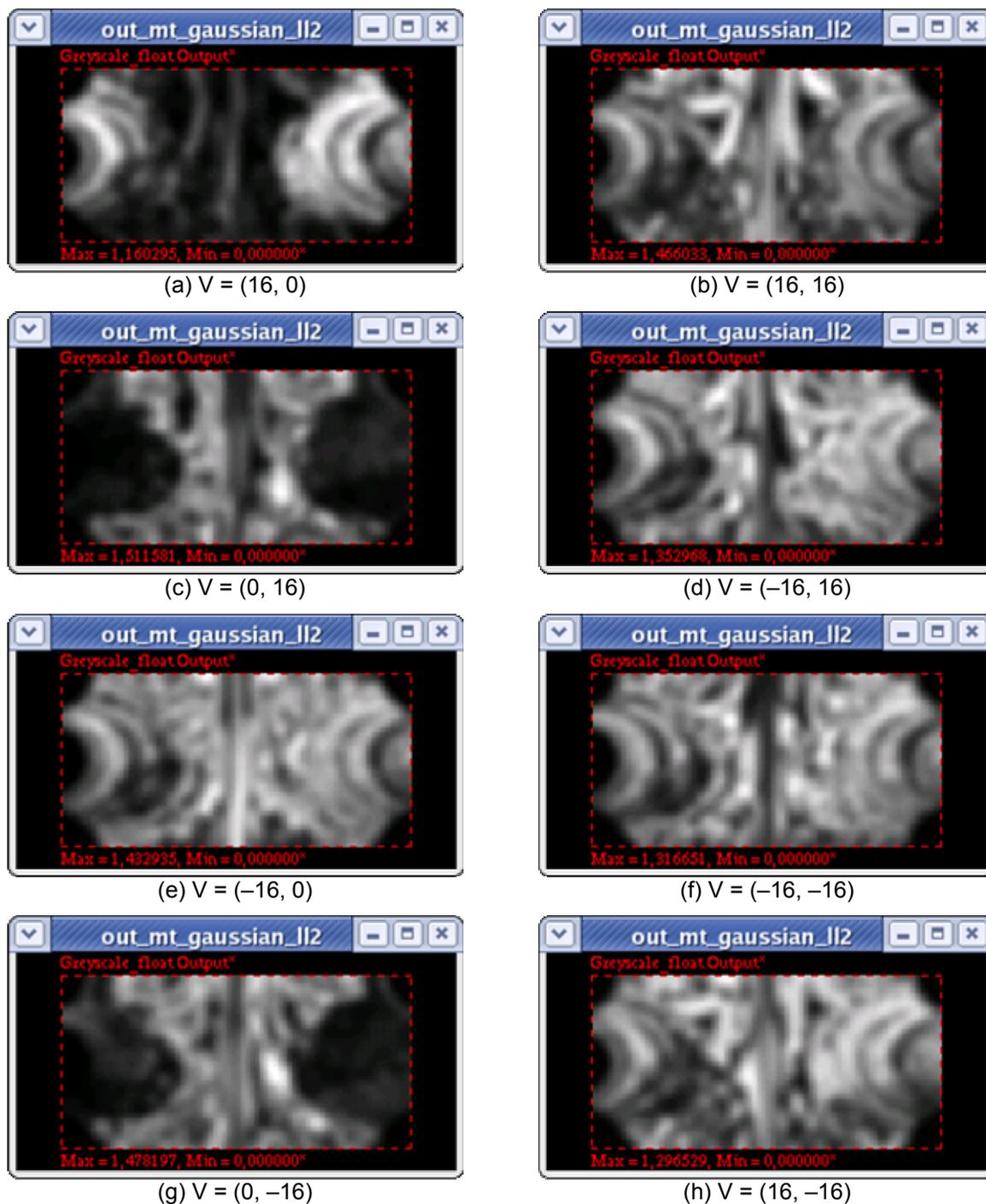


Figura 33 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical com módulo igual a 16.

Nas Figura 30-b, 32 e 33 podemos concluir, visualmente, que, dentre as 16 camadas MT apresentadas, aquela que possui a área central mais escura e, portanto, com valores mais próximos de zero, é a camada sintonizada na velocidade $(16, 0)$.

Essa constatação visual também foi confirmada pelo processamento feito com as 65 camadas MT, produzindo o resultado mais próximo possível do esperado, uma vez que o resultado exato seria (15, 0).

O método proposto para este experimento estabelece que esta comparação seja feita para cada célula de V1. A reunião dos valores mínimos de MT obtidos para cada célula constitui o mapa de velocidades, conforme mostrado na Figura 34.

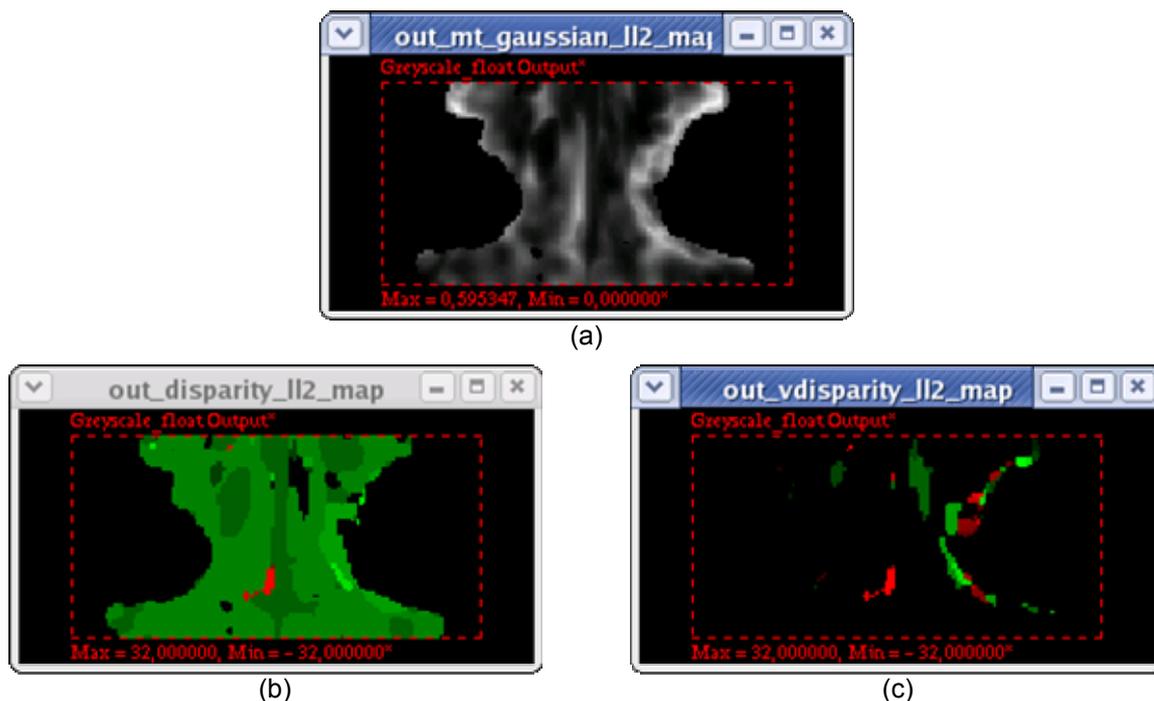


Figura 34 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.

Na Figura 34-a vemos o mapa de valores mínimos de MT obtidos para cada célula. As áreas mais claras indicam valores maiores e as áreas escuras, valores mais próximos de zero. Utilizando uma função de limiar, identificamos áreas em que a disparidade espaço-temporal possui grau elevado de incerteza. Essa incerteza trouxe reflexos visíveis na Figura 34-c, onde vemos algumas áreas em que a disparidade obtida é não nula. Na Figura 34-b, vemos que a maior área foi calculada corretamente, com disparidade horizontal igual a 16. Os valores mais imprecisos coincidem com as áreas em que o valor mínimo de MT supera o limiar.

Representação visual da velocidade de cada ponto na imagem

A representação visual das velocidades foi feita a partir dos resultados obtidos nos mapas de velocidades horizontal e vertical. A representação do plano de coordenadas log-polares de MT foi convertida em coordenadas cartesianas da imagem da retina. Dentro da região de movimento obtida na Figura 31, foram representados os pontos cujo valor mínimo de MT respeitou o limiar estabelecido. Cada velocidade foi representada por uma flecha na direção do movimento, com tamanho proporcional ao seu valor, conforme apresentado nas Figuras 35 e 36.

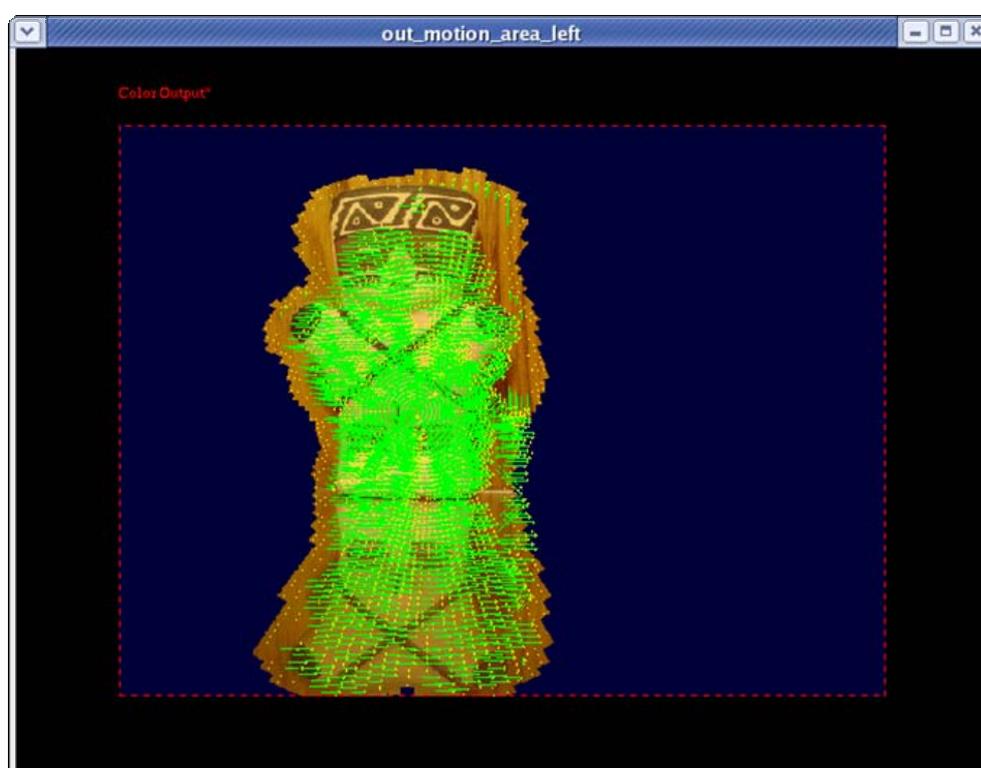


Figura 35 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$.

Comparação dos resultados obtidos com os esperados

Observando os resultados obtidos na Figura 35, vemos que na área central da estatueta as velocidades foram percebidas corretamente pelo sistema, bem como na área do reflexo da estatueta sobre a superfície. Isso pode ser mais facilmente notado na Figura 36, que representa as velocidades resumidas num grid. A figura foi subdividida em quadrados de 16 x 16 pixels, e em cada quadrado foi representada

apenas uma flecha de velocidade, correspondente à célula que apresentou o menor valor de saída MT.

Ainda na Figura 35, observamos que na região no alto da cabeça da estatueta, alguns pontos obtiveram valor mínimo de MT acima do limiar e, por isso, não foram representados. Notamos também que alguns pontos na área de contorno tiveram suas velocidades bem próximas de zero, como esperado. Na área de contorno à direita obtivemos alguns resultados incorretos. A área de resultados menos confiáveis é justamente aquela que fica encoberta pelo objeto após o seu movimento.

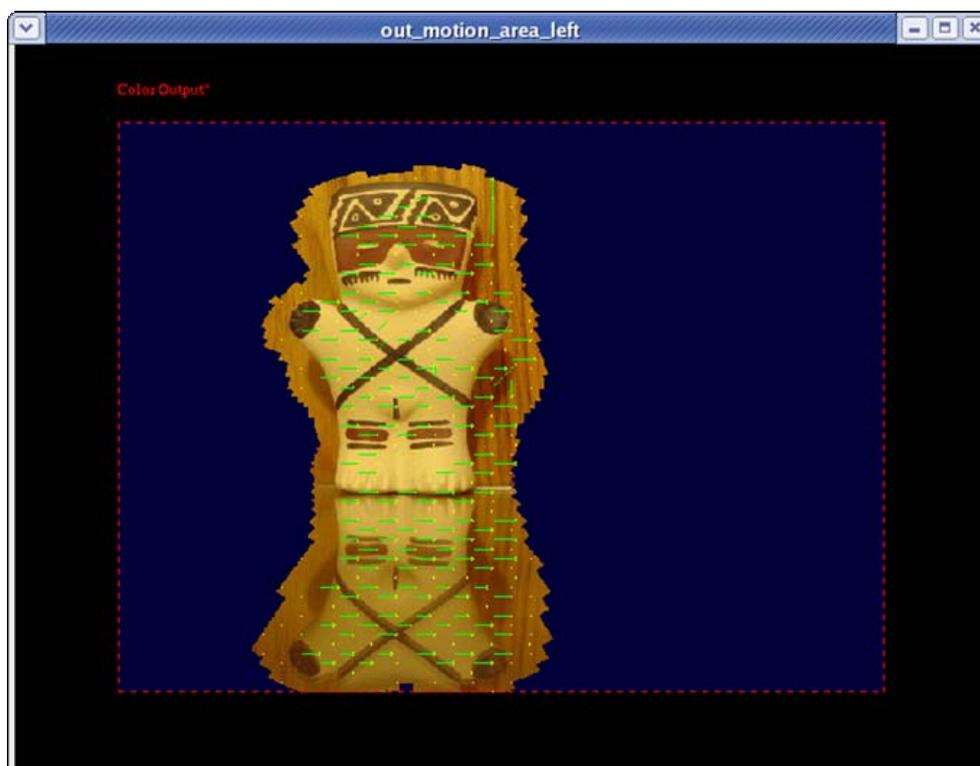


Figura 36 – Representação visual das velocidades resumidas num grid de largura 16 x 16 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.

4.2 Percepção de movimento com componentes de velocidade horizontal e vertical e perseguição suave a um objeto

Neste experimento, três objetos imantados foram colocados sobre uma superfície metálica branca. As imagens foram capturadas em quatro instantes consecutivos, sendo que apenas o objeto com a figura de Papai Noel efetuou um movimento de 11,5 mm na direção diagonal abaixo à esquerda, correspondendo a cerca de 14 pixels à esquerda e 11 pixels abaixo na imagem $(-14, -11)$. A Figura 37 apresenta as cenas capturadas nos quatro instantes. O objetivo deste experimento foi avaliar o modelo quanto à sua capacidade de percepção de um objeto em movimento que possua velocidades tanto no eixo horizontal quanto no vertical, e também avaliar a capacidade de emular o movimento ocular de perseguição suave.

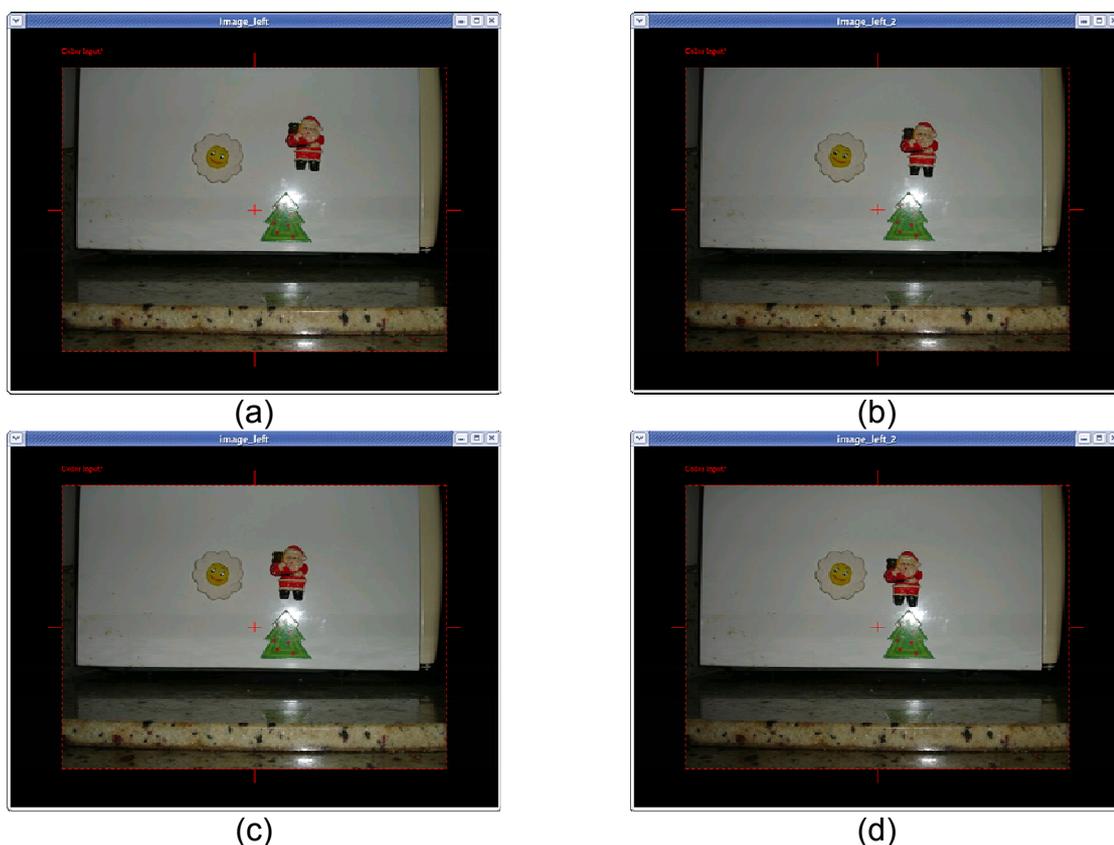


Figura 37 – Cena nos instantes 1 (a), 2 (b), 3 (c) e 4 (d). O centro de atenção está localizado na coordenada $(320, 240)$. O objeto se movimentou cerca de 14 pixels à esquerda e 11 abaixo.

A Figura 38 apresenta o campo visual em torno do centro de atenção. A Figura 39 representa o mapeamento log-polar do campo visual, em torno do centro de atenção, nos instantes 1 e 2. Nesse mapeamento, que é característico da camada

V1 do sistema visual, a parte do objeto que se move no campo visual corresponde apenas a cerca de 1,5% da área total de V1.



Figura 38 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240).

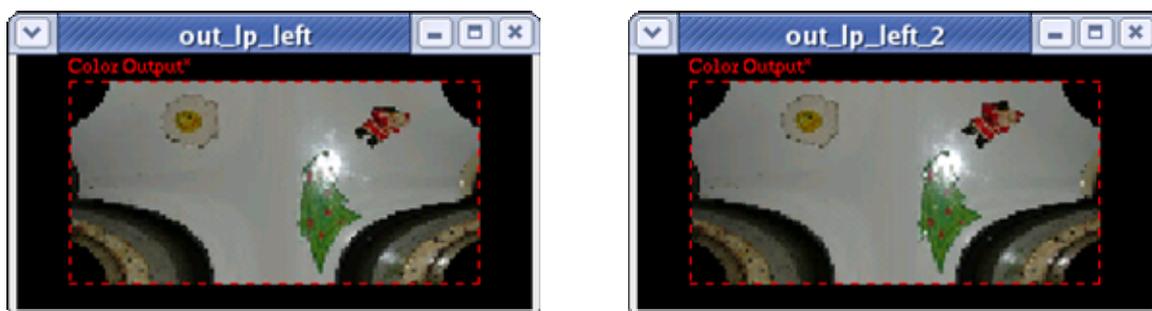


Figura 39 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.

Identificação da região de movimento

Usando a estrutura de detecção de disparidade espaço-temporal nula, podemos identificar na imagem quais são as regiões de repouso e as regiões de movimento. A Figura 40 mostra a saída da camada MT para essa estrutura. A área mais clara corresponde às células com valor de saída maior; a área escura corresponde às células com valor de saída nulo ou bem pequeno.

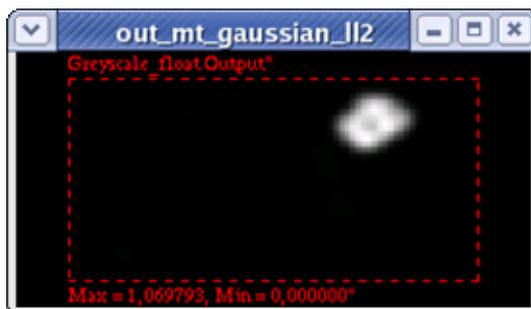


Figura 40 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.

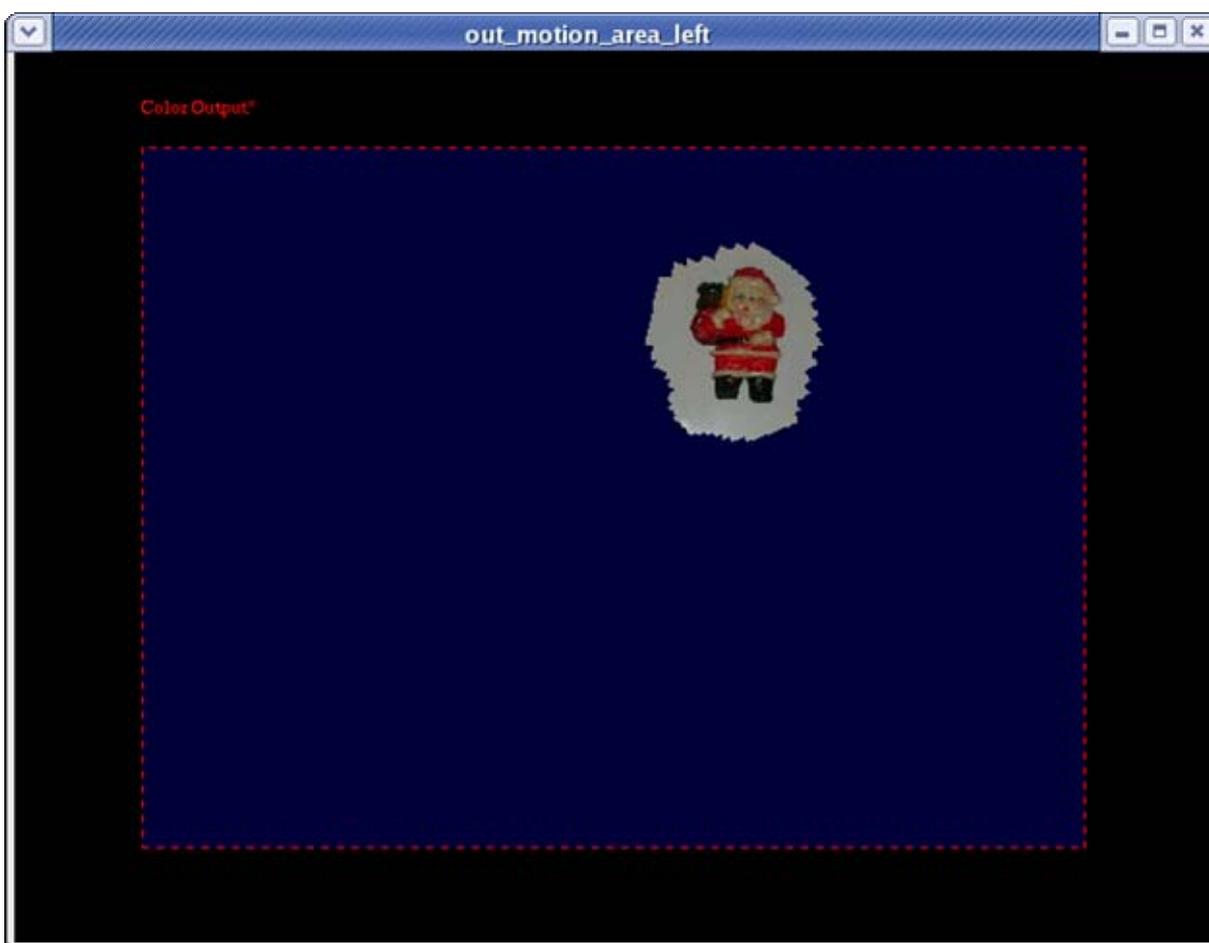


Figura 41 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240).

Mesmo tendo no campo visual apenas uma pequena fração do objeto, o sistema foi capaz de separar com boa precisão as regiões de repouso e de movimento entre as duas cenas apresentadas. A região detectada com velocidade nula foi apagada na

Figura 41, ficando representado apenas o objeto que se movimentou e uma região de incerteza, que corresponde ao contorno do objeto.

Emulação do movimento ocular de sacada

A fim de magnificar a representação do objeto em V1, emulamos uma seqüência de movimentos oculares de sacada, colocando a região de movimento na fóvea. Neste exemplo, a operação final de sacada deslocou o centro de atenção para a coordenada (399, 346). A Tabela 4 mostra a seqüência de convergência das sacadas.



Figura 42 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (399, 346), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.

O resultado dessa operação no sistema é apresentado nas Figuras 42 e 43. Observa-se que o objeto presente no campo visual passou a ocupar cerca de 20% da área total de V1, permitindo com isso uma representação muito maior de seus detalhes.

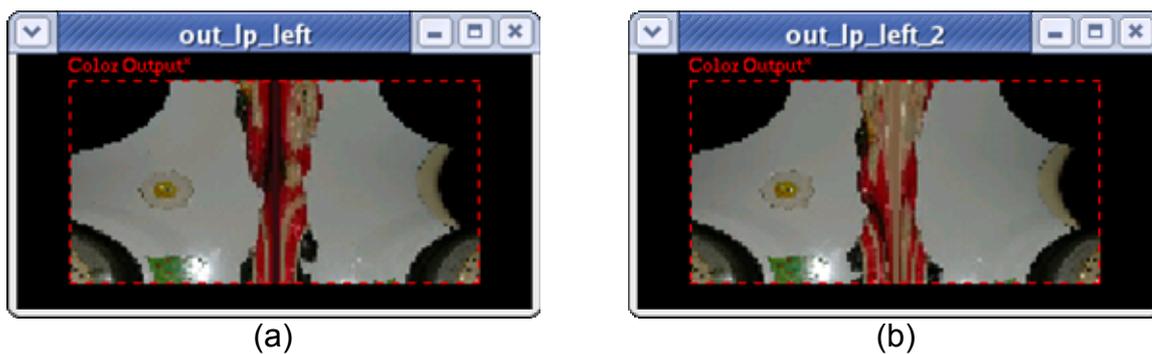


Figura 43 – Mapeamento log-polar das imagens, em torno do centro de atenção (399, 346), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.

Tabela 4 – Experimento 2: Seqüência de convergência das sacadas

Sacada	Deslocamento	Coordenadas
1	(+76, +98)	(396, 338)
2	(+2, +4)	(398, 342)
3	(+1, +2)	(399, 344)
4	(0, +1)	(399, 345)
5	(0, +1)	(399, 346)
6	(0, 0)	(399, 346)

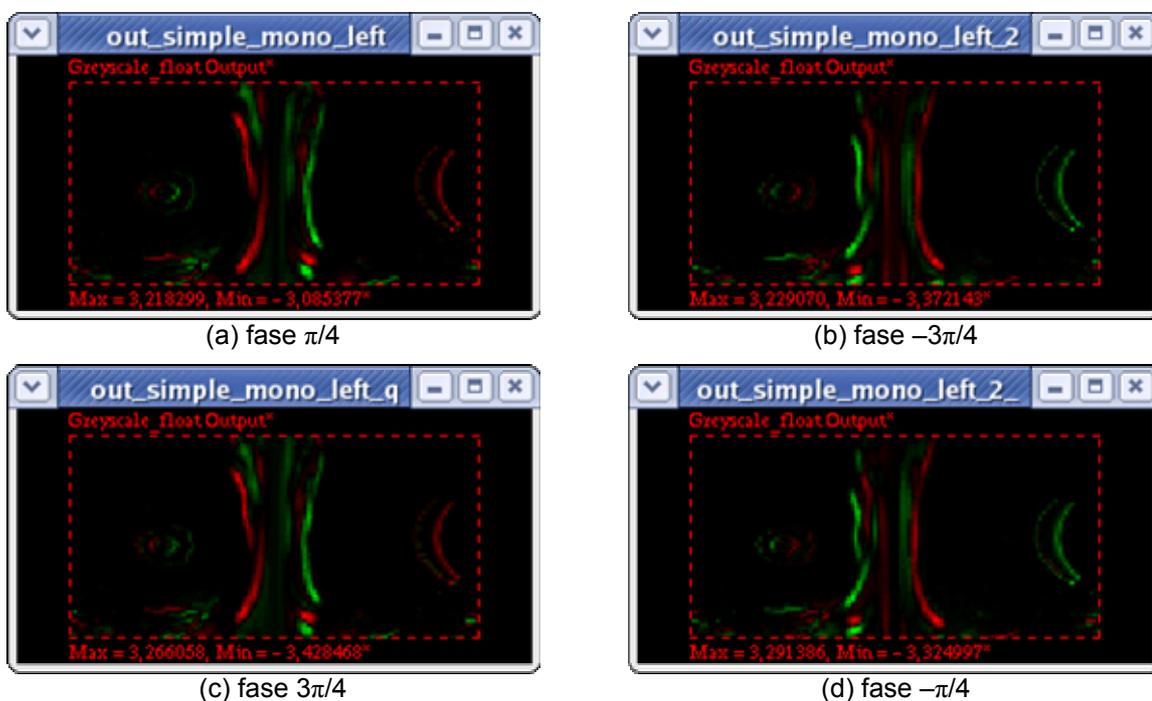


Figura 44 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase.

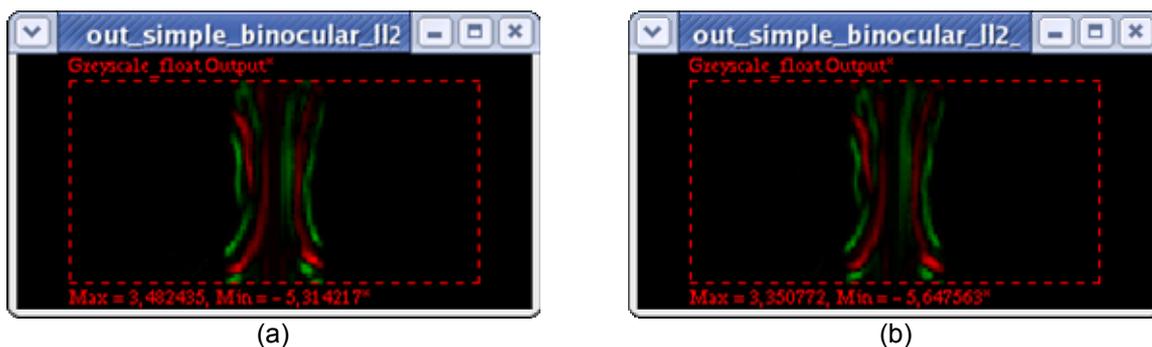


Figura 45 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.

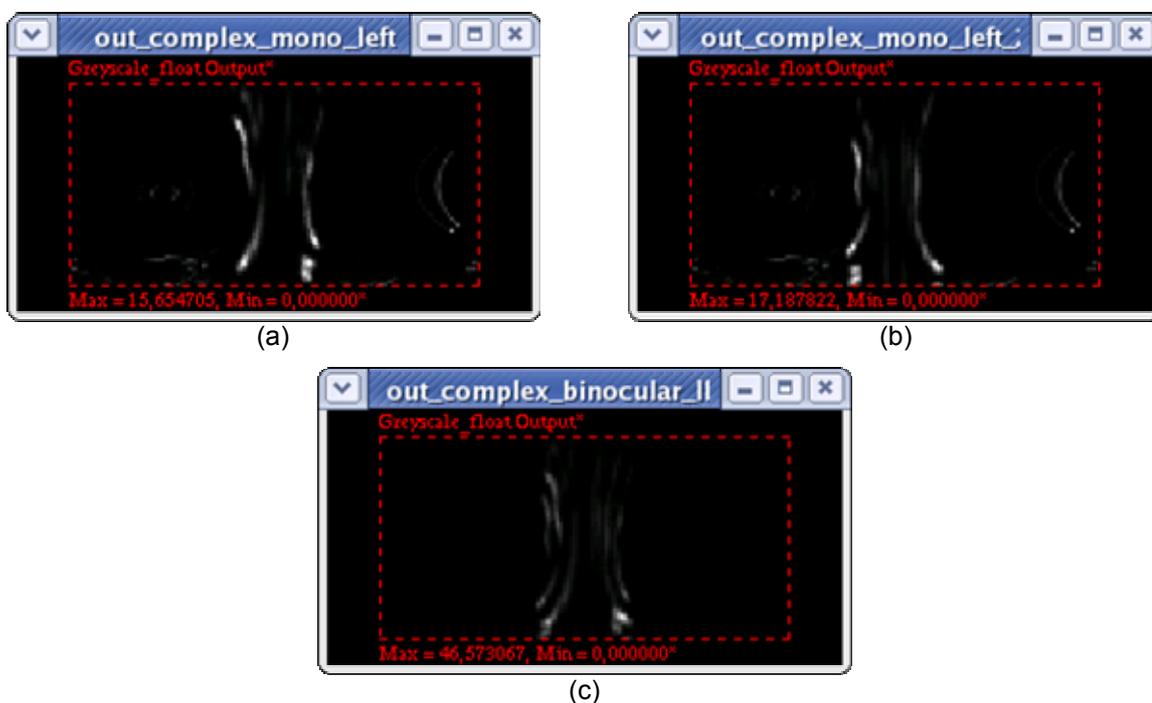


Figura 46 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.

Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida

O processamento de cada camada MT foi feito por meio de uma estrutura em camadas, conforme a Figura 14.

Foram processadas 65 estruturas separadamente, sendo uma estrutura sintonizada com a velocidade nula e as demais sintonizadas em cada uma das velocidades relacionadas na Tabela 1. A título de exemplificação, mostramos nas Figuras 44 a 46 os resultados das camadas intermediárias somente para a estrutura sintonizada na velocidade nula.

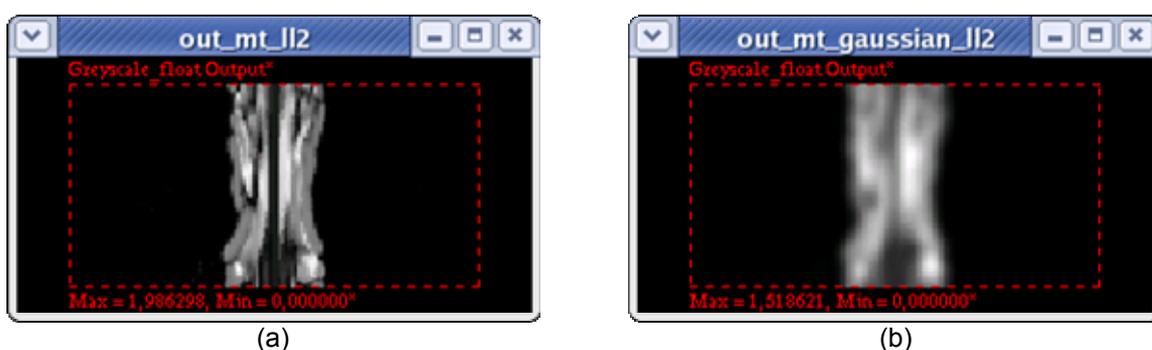


Figura 47 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com *spatial pooling*.

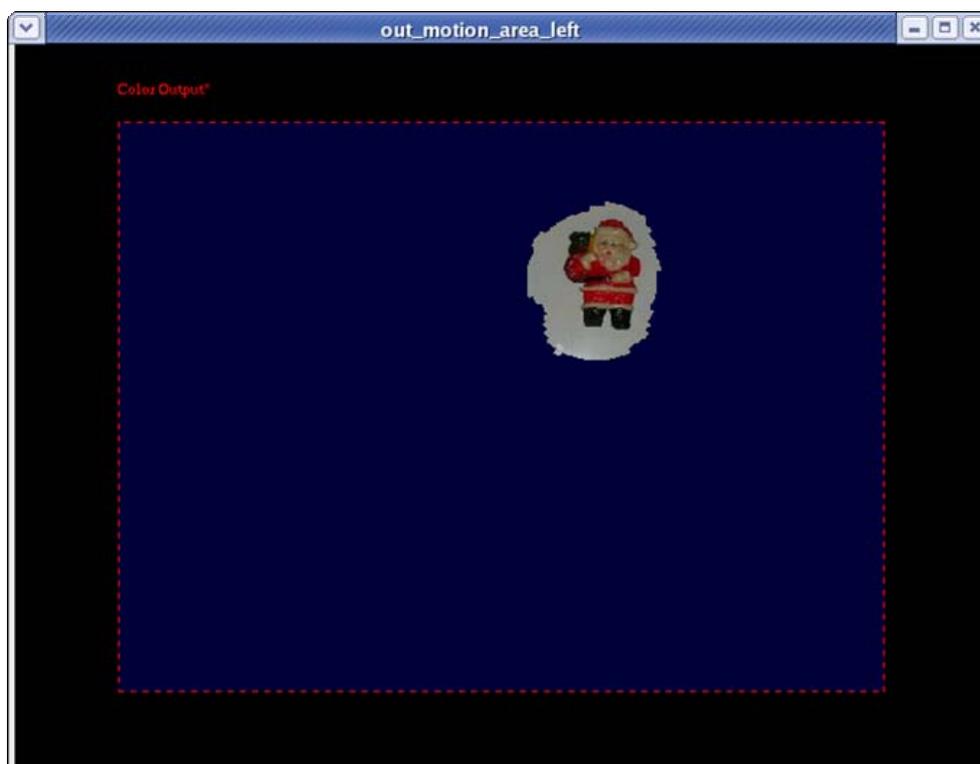


Figura 48 – Representação da região de movimento ou de incerteza, centro de atenção (399, 346).

Na Figura 47, apresentamos o resultado da camada MT sintonizada na velocidade nula, em duas versões: (a) antes de aplicar o algoritmo de *spatial pooling*; (b) após aplicar o algoritmo de *spatial pooling*. O efeito visivelmente notável é a suavização da função de saída de MT, por meio da cooperação entre as células vizinhas nesta camada.

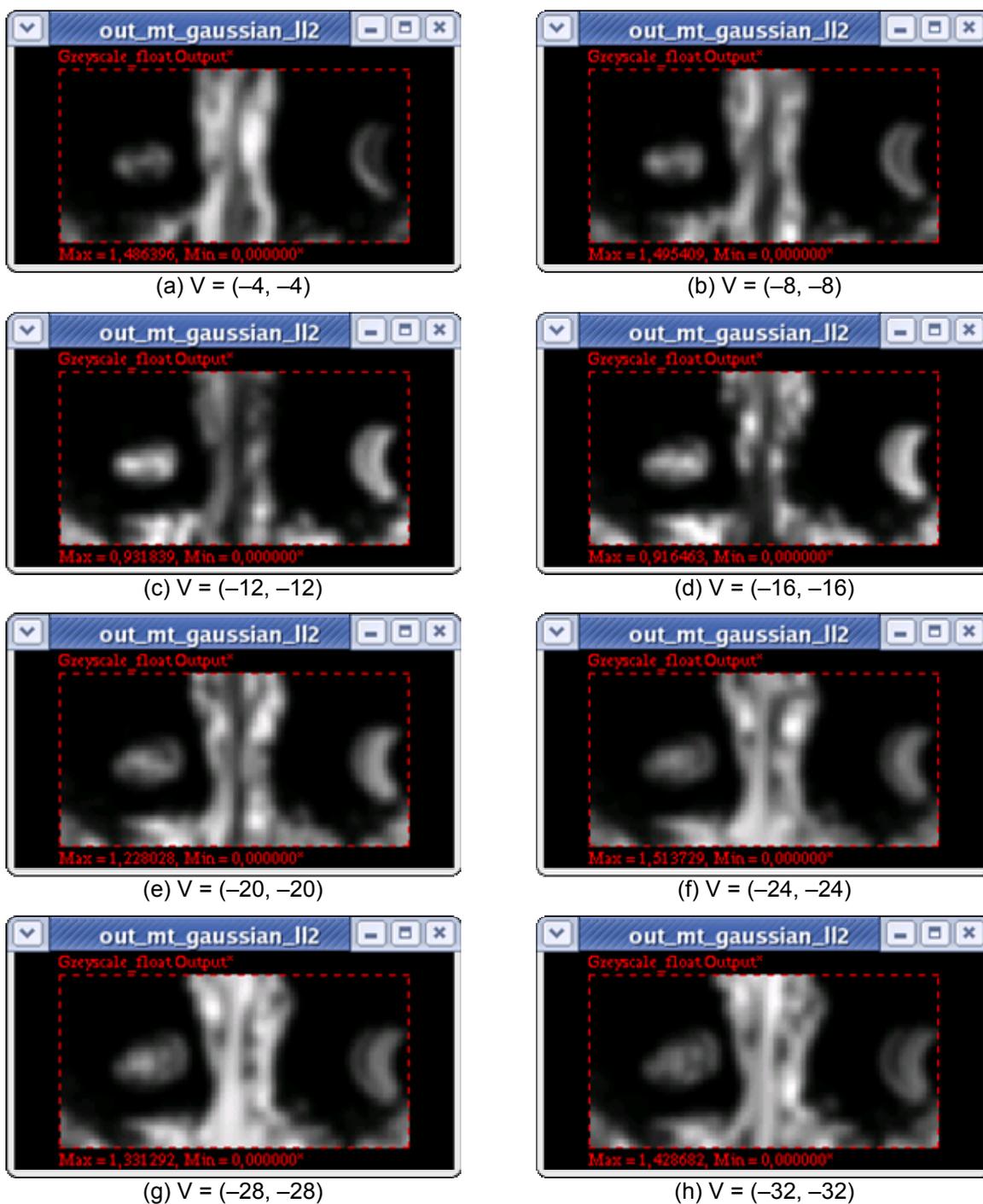


Figura 49 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, no movimento diagonal, sentido para baixo à esquerda.

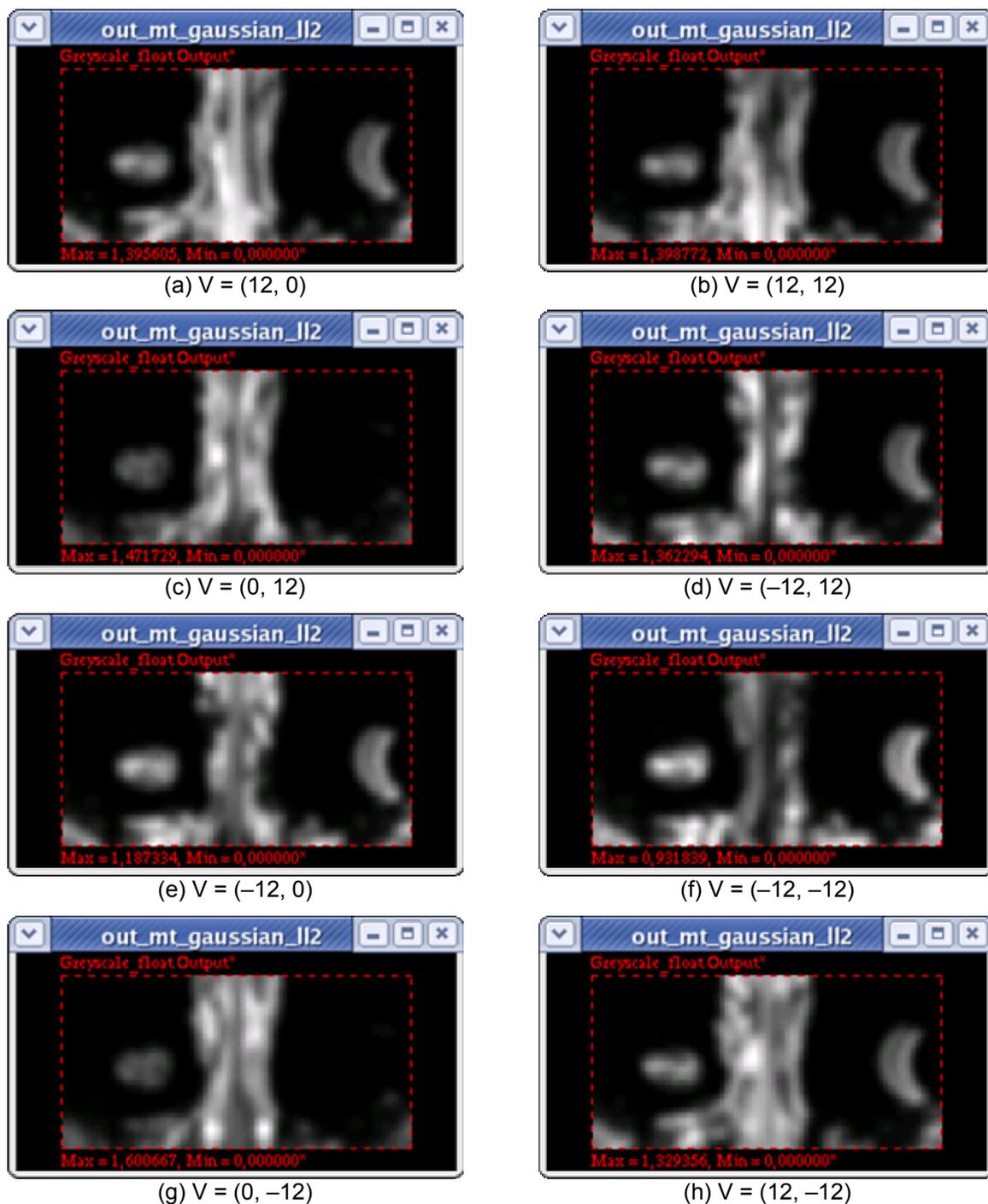


Figura 50 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical com módulo igual a 12.

Utilizando uma função de limiar sobre a saída de MT, com valor de limiar $L = 0,1$, chegamos à separação das regiões de repouso e de movimento, conforme mostrado na Figura 48. Observamos que não houve variação significativa em relação ao resultado inicial obtido na Figura 41, quanto à determinação da região de movimento.

Criação do mapa de velocidades

As Figuras 47-b, 49 e 50 mostram as camadas MT sintonizadas em 16 velocidades distintas, para ilustração do processo de criação do mapa de velocidades. O processo completo neste experimento comparou 65 camadas MT. Podemos concluir, visualmente, que, dentre as 16 camadas MT apresentadas, aquela que possui a área central mais escura e, portanto, com valores mais próximos de zero, é a camada sintonizada na velocidade $(-12, -12)$.

Essa constatação visual também foi confirmada pelo processamento feito com as 65 camadas MT, produzindo o resultado mais próximo possível do esperado, uma vez que a velocidade exata seria $(-14, -11)$.

O método proposto para este experimento estabelece que esta comparação seja feita para cada célula de V1. A reunião dos valores mínimos de MT obtidos para cada célula constitui o mapa de velocidades, conforme mostrado na Figura 51.

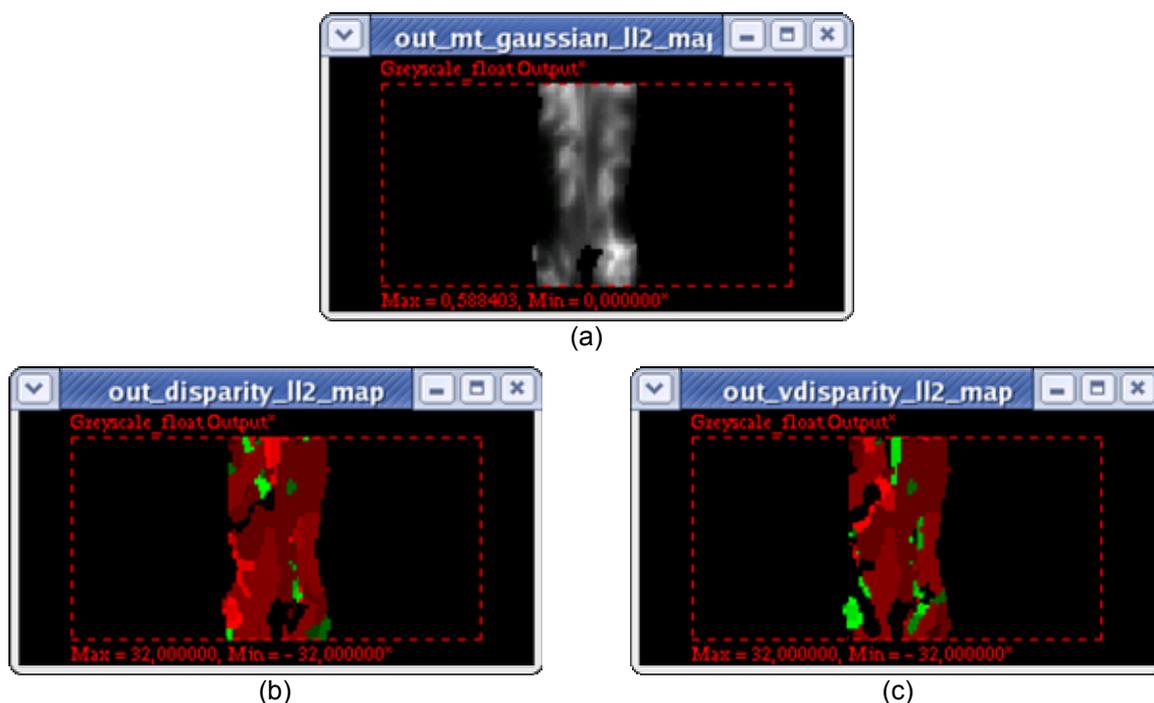


Figura 51 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.

Na Figura 51-a vemos o mapa de valores mínimos de MT obtidos para cada célula. As áreas mais claras indicam valores maiores e as áreas escuras, valores mais próximos de zero. Utilizando uma função de limiar, identificamos áreas em que a

disparidade espaço-temporal possui grau elevado de incerteza. Essa incerteza trouxe reflexos nas Figuras 51-b e 51-c, onde vemos que a maior área foi calculada corretamente, com velocidade igual a -12 , mas algumas áreas apresentam ruído. Os valores mais imprecisos coincidem com as áreas em que o valor mínimo de MT supera o limiar.

Representação visual da velocidade de cada ponto na imagem

A representação visual das velocidades foi feita a partir dos resultados obtidos nos mapas de velocidades. A representação do plano de coordenadas log-polares de MT foi convertida em coordenadas cartesianas da imagem da retina. Dentro da região de movimento obtida na Figura 48, foram representados os pontos cujo valor mínimo de MT respeitou o limiar estabelecido. Cada velocidade foi representada por uma flecha na direção do movimento, com tamanho proporcional ao seu valor, conforme apresentado nas Figuras 52 e 53.

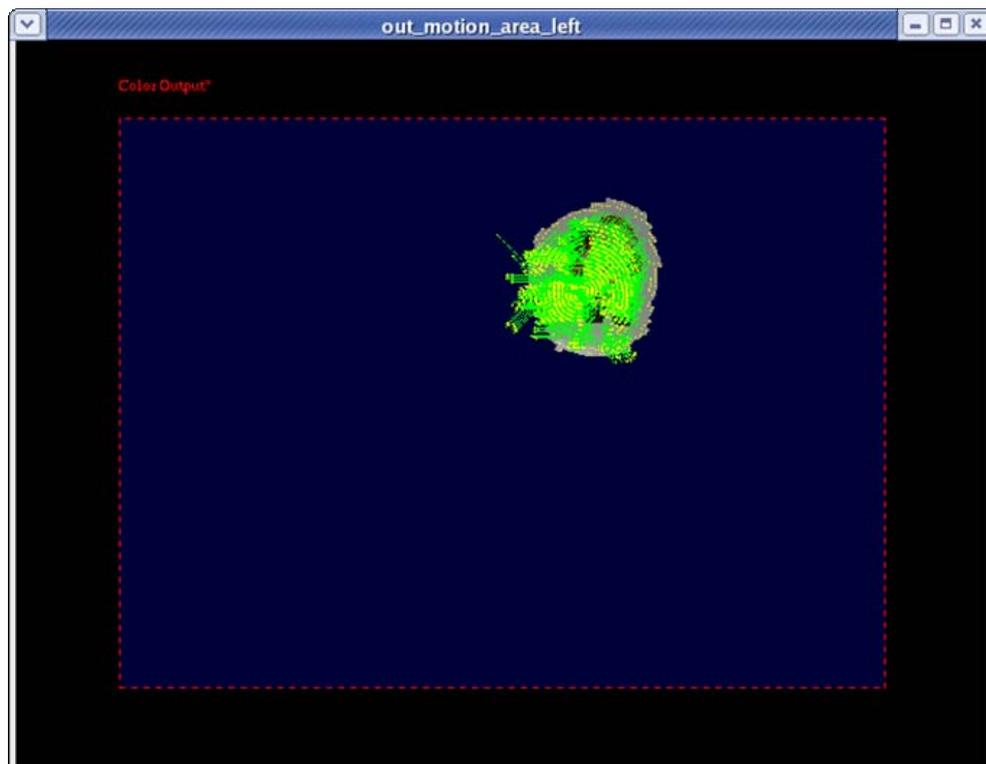


Figura 52 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$.

Comparação dos resultados obtidos com os esperados

Observando os resultados obtidos na Figura 52, vemos que na área central do objeto que se move as velocidades foram percebidas corretamente pelo sistema. Isso pode ser mais facilmente notado na Figura 53, que representa as velocidades resumidas num grid. A figura foi subdividida em quadrados de 12 x 12 pixels, e em cada quadrado foi representada apenas uma flecha de velocidade, correspondente à célula que apresentou o menor valor de saída MT.

Na Figura 53, observamos que alguns pontos na área de contorno tiveram suas velocidades bem próximas de zero, como esperado. Na área de contorno inferior e esquerdo obtivemos resultados incorretos. A área de resultados menos confiáveis é justamente aquela que fica encoberta pelo objeto após o seu movimento.

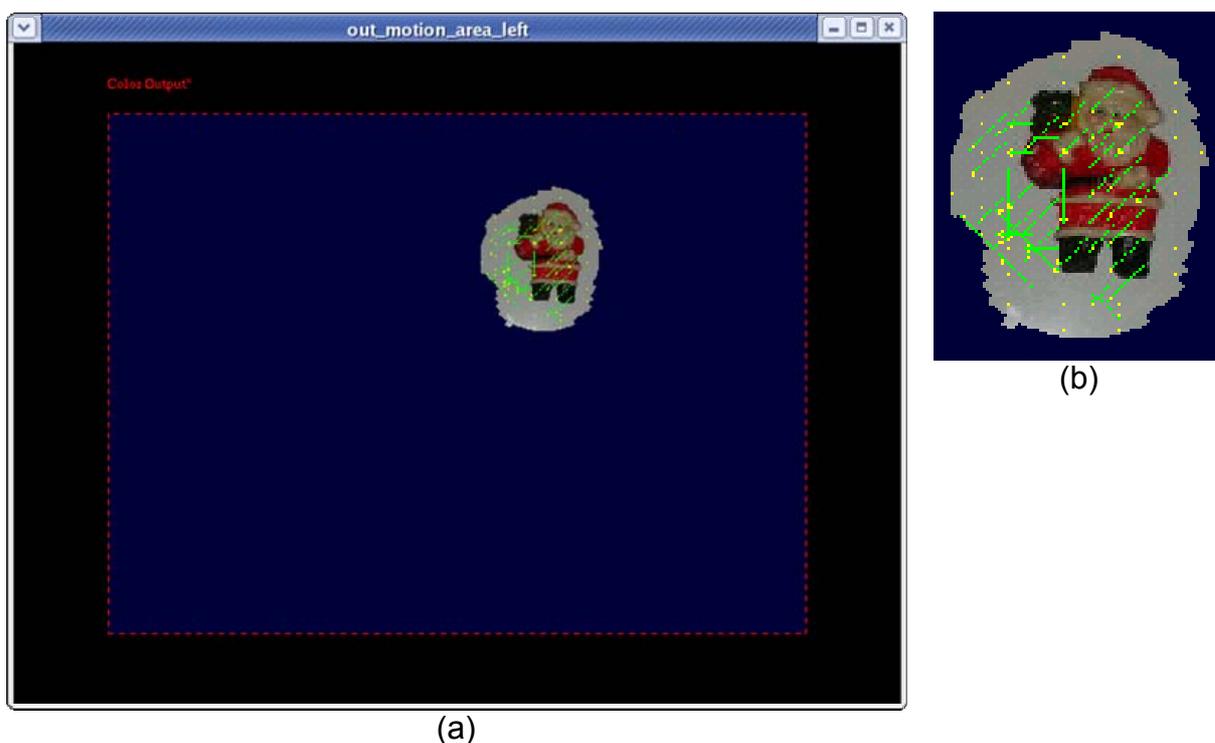


Figura 53 – Representação visual das velocidades resumidas num grid de largura 12 x 12 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.

Emulação do movimento ocular de perseguição suave

Uma vez conhecidos o centro de atenção num determinado instante e a velocidade média dos pontos da região de movimento, o sistema é capaz de prever o novo centro de atenção antes que a próxima imagem na seqüência seja apresentada à

retina. Isso equivale a emular o movimento ocular de perseguição suave, em que o objeto em movimento fica mantido na fóvea.

Na seqüência do movimento analisado, a Figura 37-c, correspondente ao instante 3, foi trazida à retina e comparada com a Figura 37-b, que corresponde ao instante 2. O centro de atenção anterior era (399, 346) e a velocidade média dos pontos em movimento era (-12, -12). O sistema então deslocou o centro de atenção automaticamente para o ponto (387, 334).

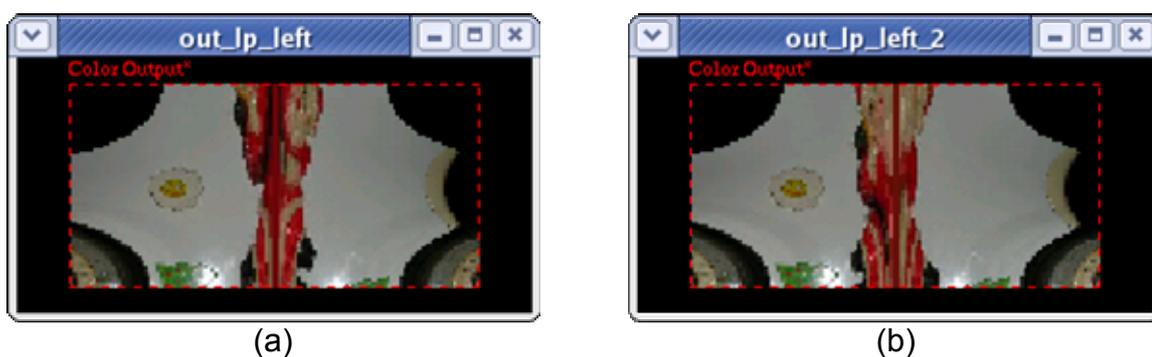


Figura 54 – Mapeamento log-polar das imagens, em torno do centro de atenção (387, 334), nos instantes 2 e 3, após o movimento ocular de perseguição suave.

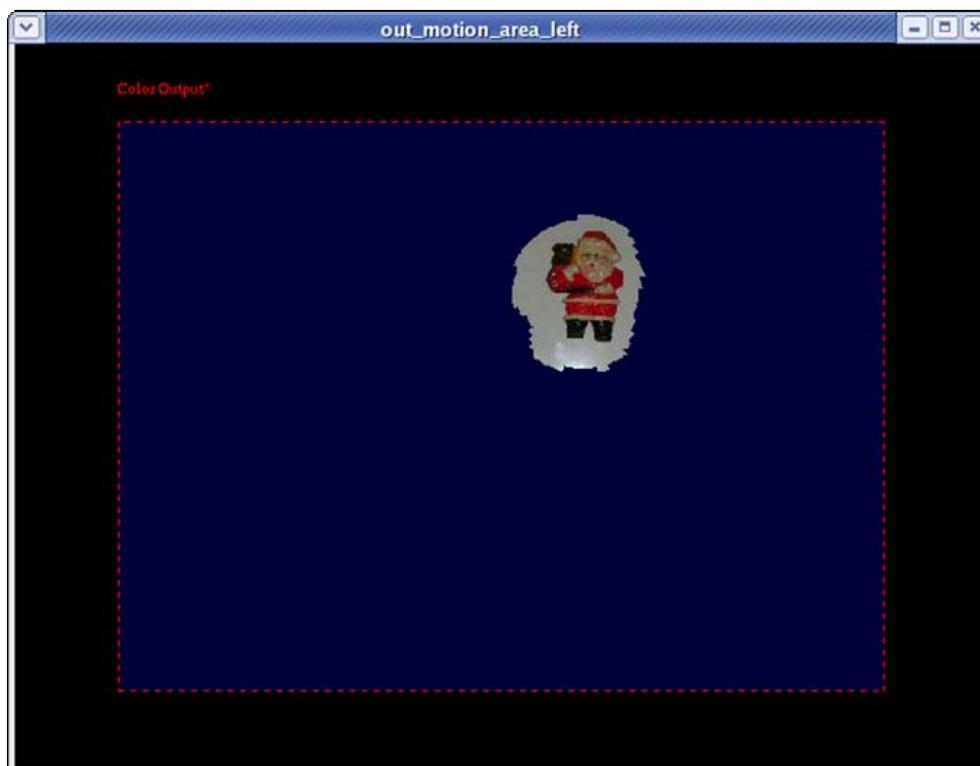


Figura 55 – Representação da região de movimento ou de incerteza, centro de atenção (387, 334), após o movimento ocular de perseguição suave.

O resultado obtido pode ser visualizado nas Figuras 54 e 55. Sem que fosse necessário efetuar um novo processamento de sacada, a região de movimento manteve-se localizada sobre a fóvea, ocupando cerca de 20% da área de V1.

Após efetuar um novo processamento das camadas MT e cálculo do mapa de velocidades, o resultado obtido para a velocidade média dos pontos em movimento foi novamente $(-12, -12)$.

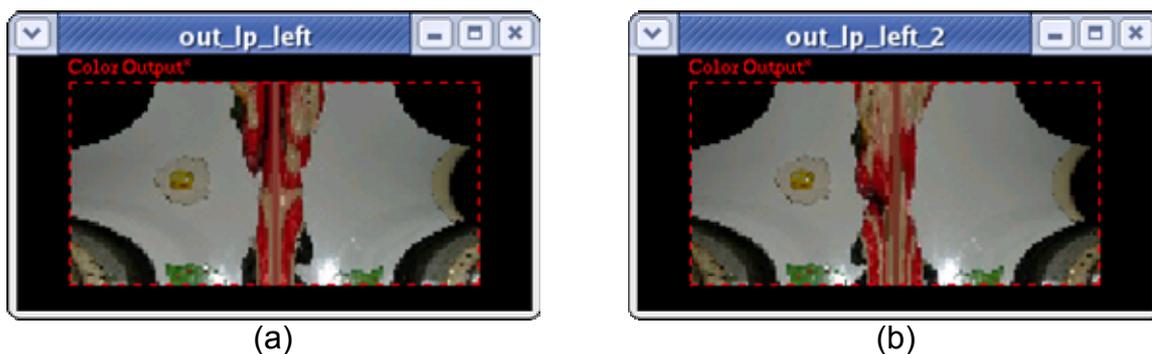


Figura 56 – Mapeamento log-polar das imagens, em torno do centro de atenção (375, 322), nos instantes 3 e 4, após o movimento ocular de perseguição suave.

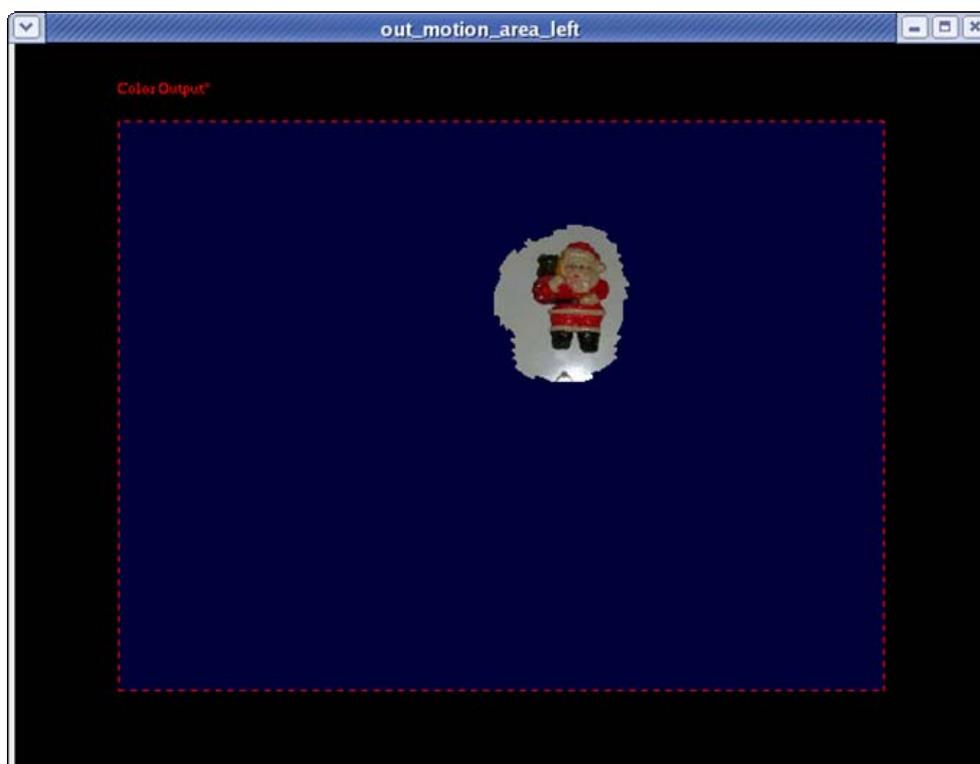


Figura 57 – Representação da região de movimento ou de incerteza, centro de atenção (375, 322), após o movimento ocular de perseguição suave.

Na seqüência do movimento analisado, a Figura 37-d, correspondente ao instante 4, foi trazida à retina e comparada com a Figura 37-c, que corresponde ao instante 3. O centro de atenção anterior era (387, 334), e o novo centro de atenção foi automaticamente para o ponto (375, 322). O resultado está na Figuras 56 e 57.

Novamente, a emulação do movimento de perseguição suave manteve o alvo na fóvea, sem que fosse necessário efetuar um novo processamento de sacada.

4.3 Percepção de movimentos de translação simultâneos e diversos de múltiplos objetos

Neste experimento, quatro objetos realizam movimentos de translação simultâneos, independentes e distintos. As Figuras 58 e 59 apresentam as cenas capturadas nos dois instantes. O objeto que está mais no alto movimentou-se 9,5 mm à direita, correspondendo a cerca de (+9, -2) pixels na imagem; o que está mais à direita movimentou-se 9 mm para baixo, correspondendo a cerca de (0, -9) pixels; o que está mais embaixo movimentou-se 10,5 mm à esquerda, correspondendo a cerca de (-10, -2) pixels; o que está mais à esquerda movimentou-se 6 mm para cima, correspondendo a cerca de (+1, +6) pixels. O objetivo deste experimento é avaliar o modelo quanto à sua capacidade de percepção de velocidades distintas na mesma cena.

A Figura 60 apresenta o campo visual em torno do centro de atenção. A Figura 61 representa o mapeamento log-polar do campo visual, em torno do centro de atenção, nos instantes 1 e 2. Nesse mapeamento, que é característico da camada V1 do sistema visual, os objetos que se movem no campo visual correspondem a cerca de 6,5% da área total de V1.

Identificação da região de movimento

Usando a estrutura de detecção de disparidade espaço-temporal nula, podemos identificar na imagem quais são as regiões de repouso e as regiões de movimento. A Figura 62 mostra a saída da camada MT para essa estrutura. As áreas mais claras correspondem às células com valor de saída maior; a área escura corresponde às células com valor de saída nulo ou bem pequeno.

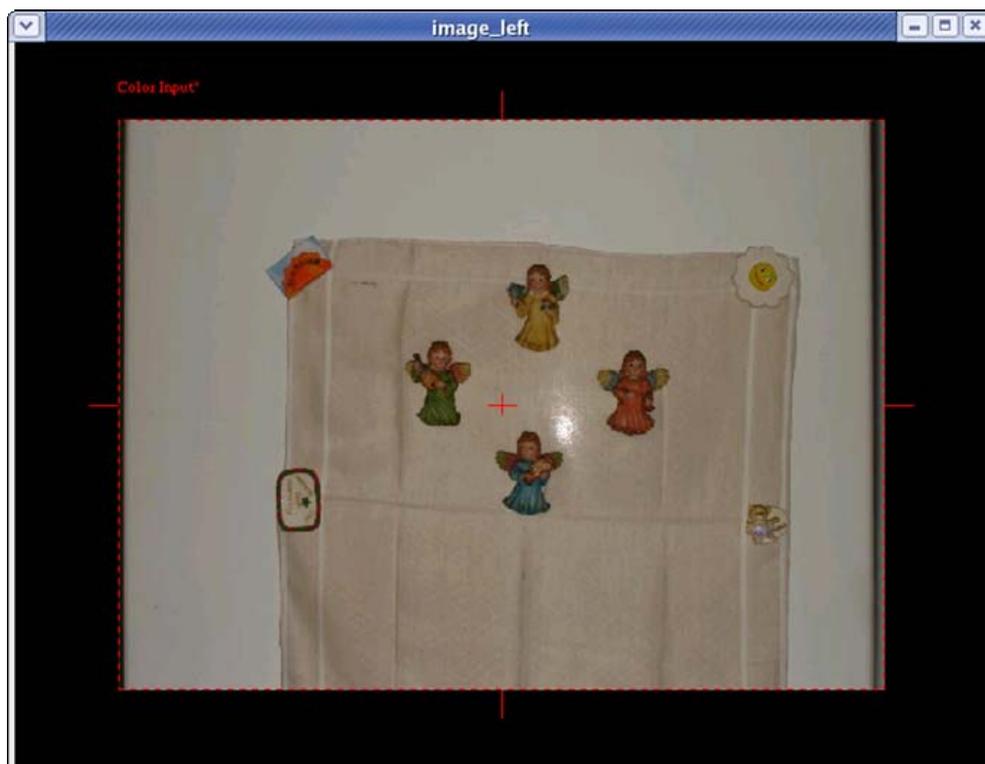


Figura 58 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (320, 240).



Figura 59 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada.



Figura 60 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240).



Figura 61 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.

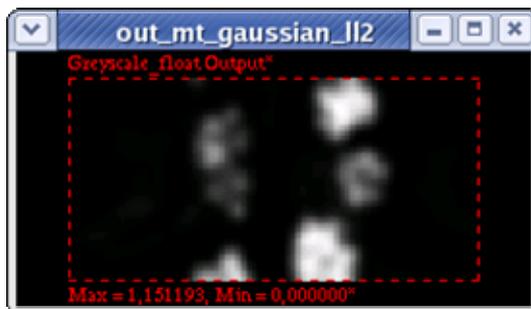


Figura 62 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; as áreas claras correspondem às regiões de movimento.

Mesmo tendo no campo visual apenas uma pequena fração do objeto, o sistema foi capaz de separar com boa precisão as regiões de repouso e de movimento entre as duas cenas apresentadas. A região detectada com velocidade nula foi apagada na Figura 63, ficando representados apenas os objetos que se movimentaram e uma região de incerteza, que corresponde aos contornos dos objetos.

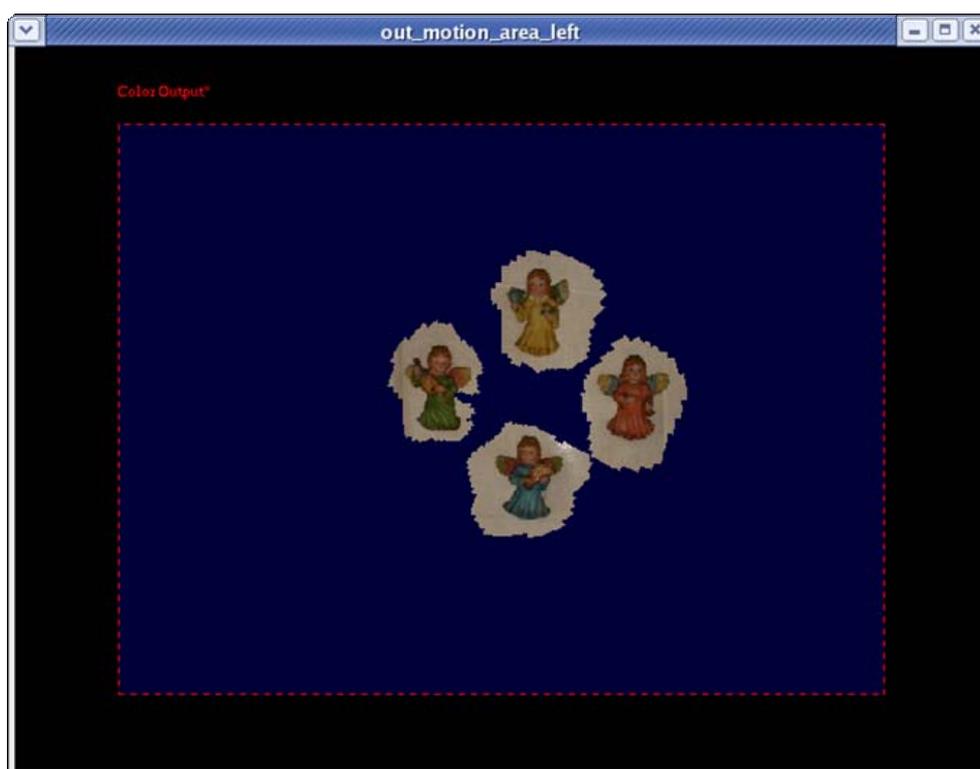


Figura 63 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240).

Emulação do movimento ocular de sacada

A fim de magnificar a representação do objeto em V1, emulamos um movimento ocular de sacada, colocando a região de movimento na fóvea. Neste exemplo, a seqüência de operações de sacada deslocou o centro de atenção para a coordenada (358, 271).

A Tabela 5 mostra a seqüência de convergência das sacadas. O resultado dessa operação no sistema é apresentado nas Figuras 64 e 65. Os objetos que se movem no campo visual passaram a ocupar cerca de 7% da área total de V1. Não houve um acréscimo significativo em relação à situação anterior (Figura 61). O centro geométrico dos objetos em movimento situa-se numa região sem movimento.

Tabela 5 – Experimento 3: Seqüência de convergência das sacadas

Sacada	Deslocamento	Coordenadas
1	(+24, +1)	(344, 241)
2	(+8, +3)	(352, 244)
3	(+3, +4)	(355, 248)
4	(+1, +4)	(356, 252)
5	(+1, +4)	(357, 256)
6	(0, +3)	(357, 259)
7	(+1, +2)	(358, 261)
8	(0, +2)	(358, 263)
9	(0, +1)	(358, 264)
10	(0, +1)	(358, 265)
11	(0, +1)	(358, 266)
12	(0, +1)	(358, 267)
13	(0, +1)	(358, 268)
14	(0, +1)	(358, 269)
15	(0, +1)	(358, 270)
16	(0, +1)	(358, 271)
17	(0, 0)	(358, 271)



Figura 64 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (358, 271), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.



Figura 65 – Mapeamento log-polar das imagens, em torno do centro de atenção (358, 271), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.

Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida

O processamento de cada camada MT foi feito por meio de uma estrutura em camadas, conforme a Figura 14.

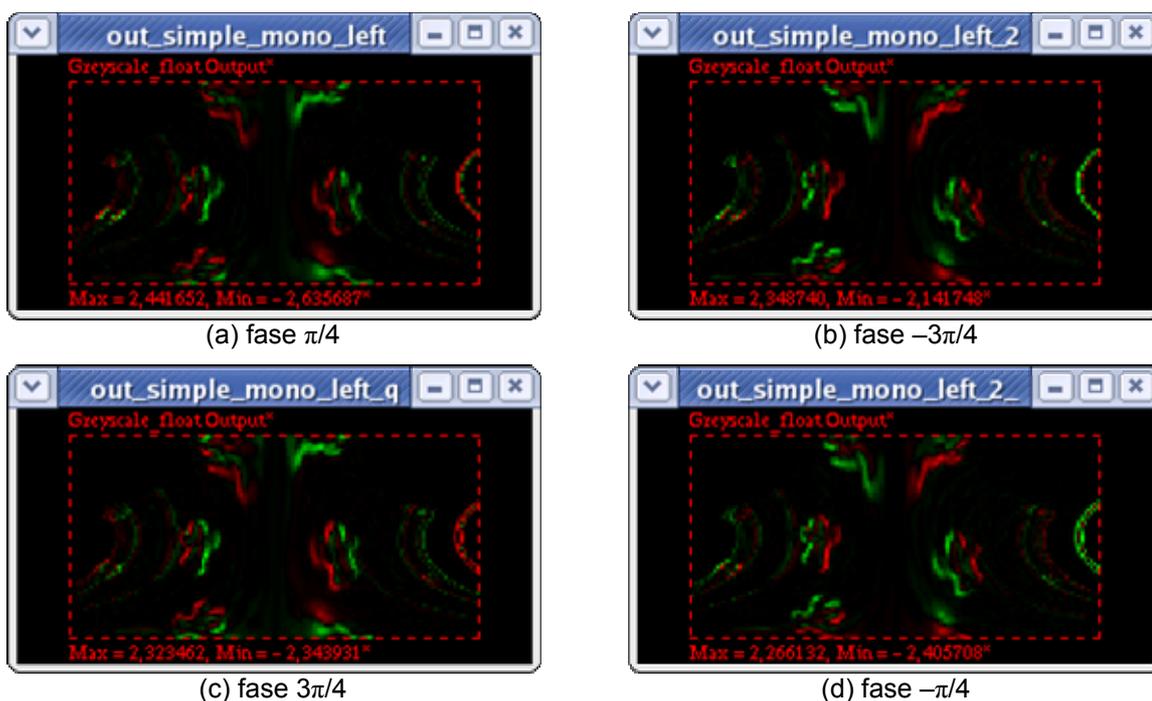


Figura 66 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase.

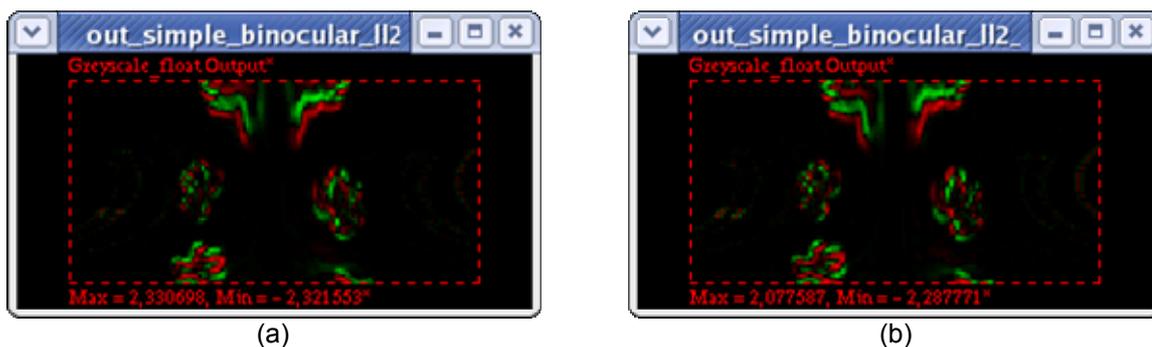


Figura 67 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.

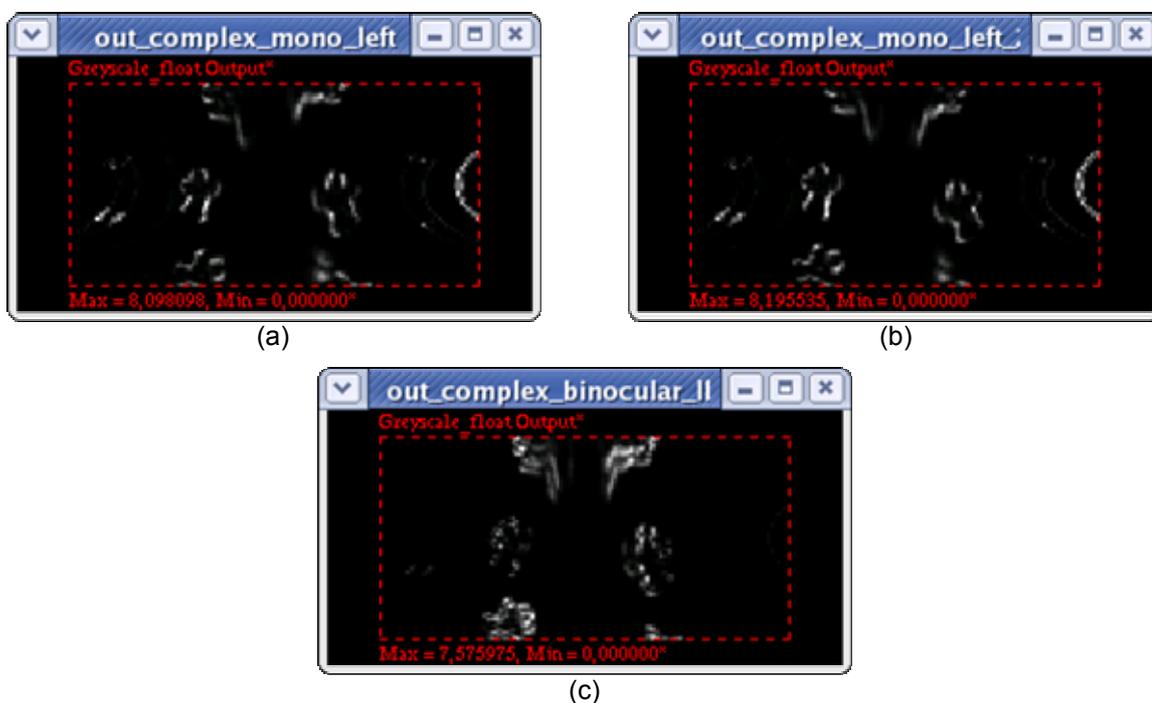


Figura 68 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.

Foram processadas 65 estruturas separadamente, sendo uma estrutura sintonizada com a velocidade nula e as demais sintonizadas em cada uma das velocidades relacionadas na Tabela 1. A título de exemplificação, mostramos nas Figuras 66 a 68 os resultados das camadas intermediárias somente para a estrutura sintonizada na velocidade nula.

Na Figura 69, apresentamos o resultado da camada MT sintonizada na velocidade nula, em duas versões: (a) antes de aplicar o algoritmo de *spatial pooling*; (b) após aplicar o algoritmo de *spatial pooling*. O efeito visivelmente notável é a suavização da função de saída de MT, por meio da cooperação entre as células vizinhas nesta camada.

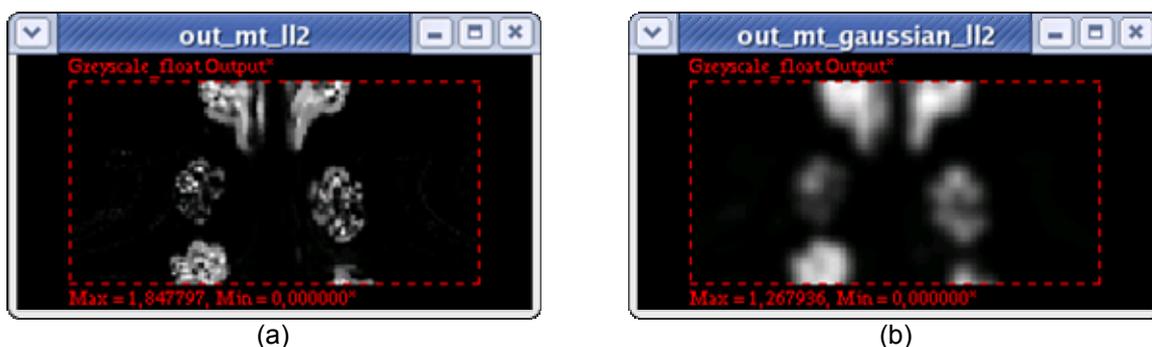


Figura 69 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com *spatial pooling*.

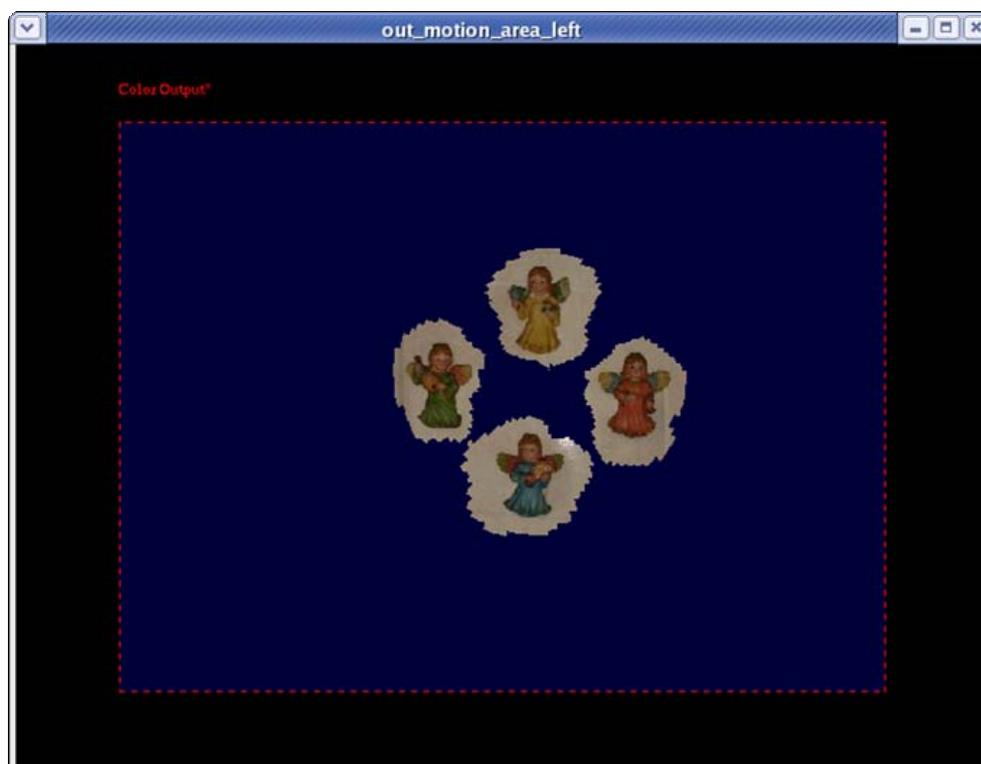


Figura 70 – Representação da região de movimento ou de incerteza, centro de atenção (358, 271).

Utilizando uma função de limiar sobre a saída de MT, com valor de limiar $L = 0,1$, chegamos à separação das regiões de repouso e de movimento, conforme mostrado na Figura 70. Observamos que houve uma melhora em relação ao resultado inicial obtido na Figura 63.

Criação do mapa de velocidades

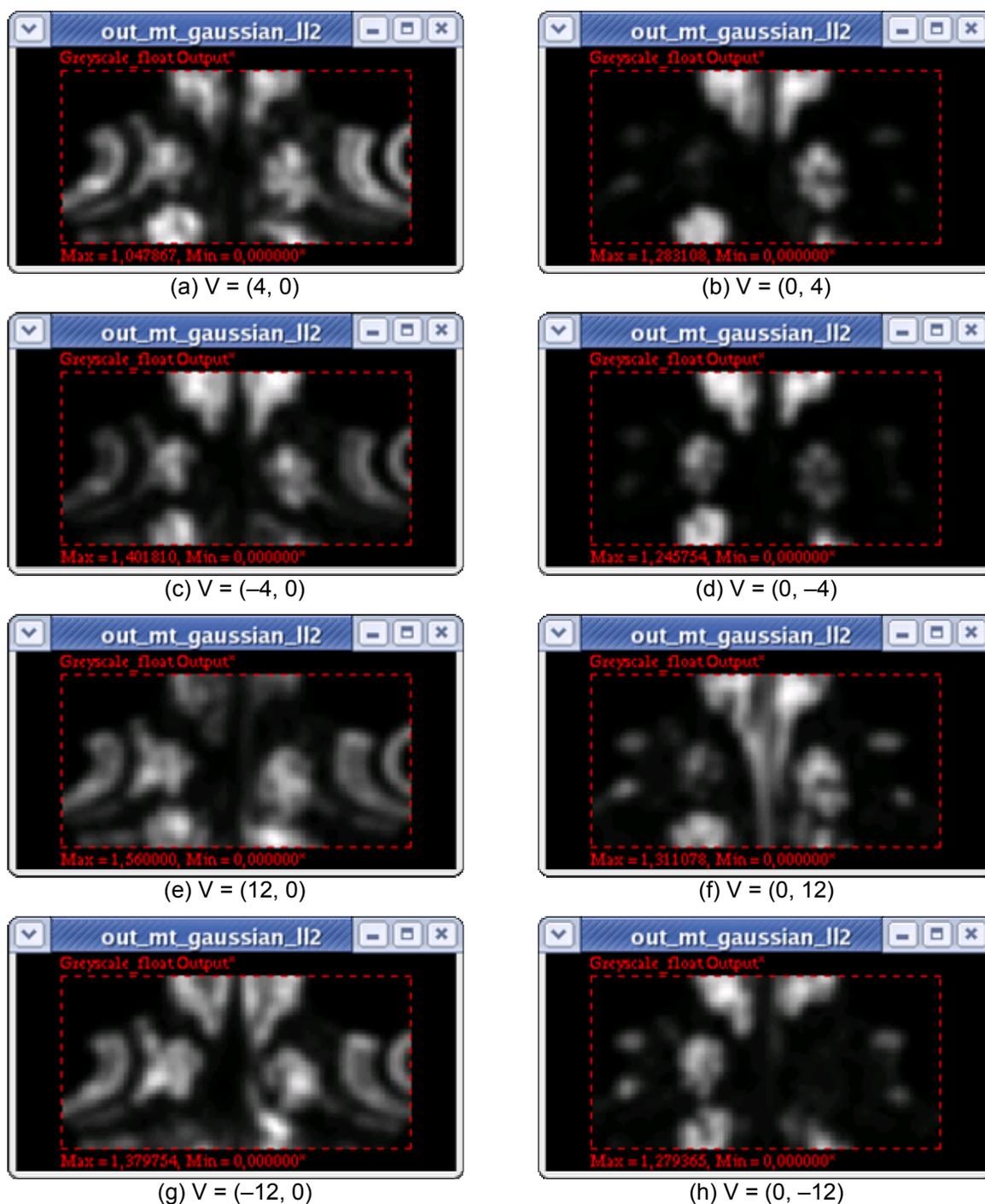


Figura 71 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, em movimentos horizontais e verticais.

As Figuras 69-b, 71 e 72 mostram as camadas MT sintonizadas em 17 velocidades distintas, para ilustração do processo de criação do mapa de velocidades. O processo completo neste experimento comparou 65 camadas MT.

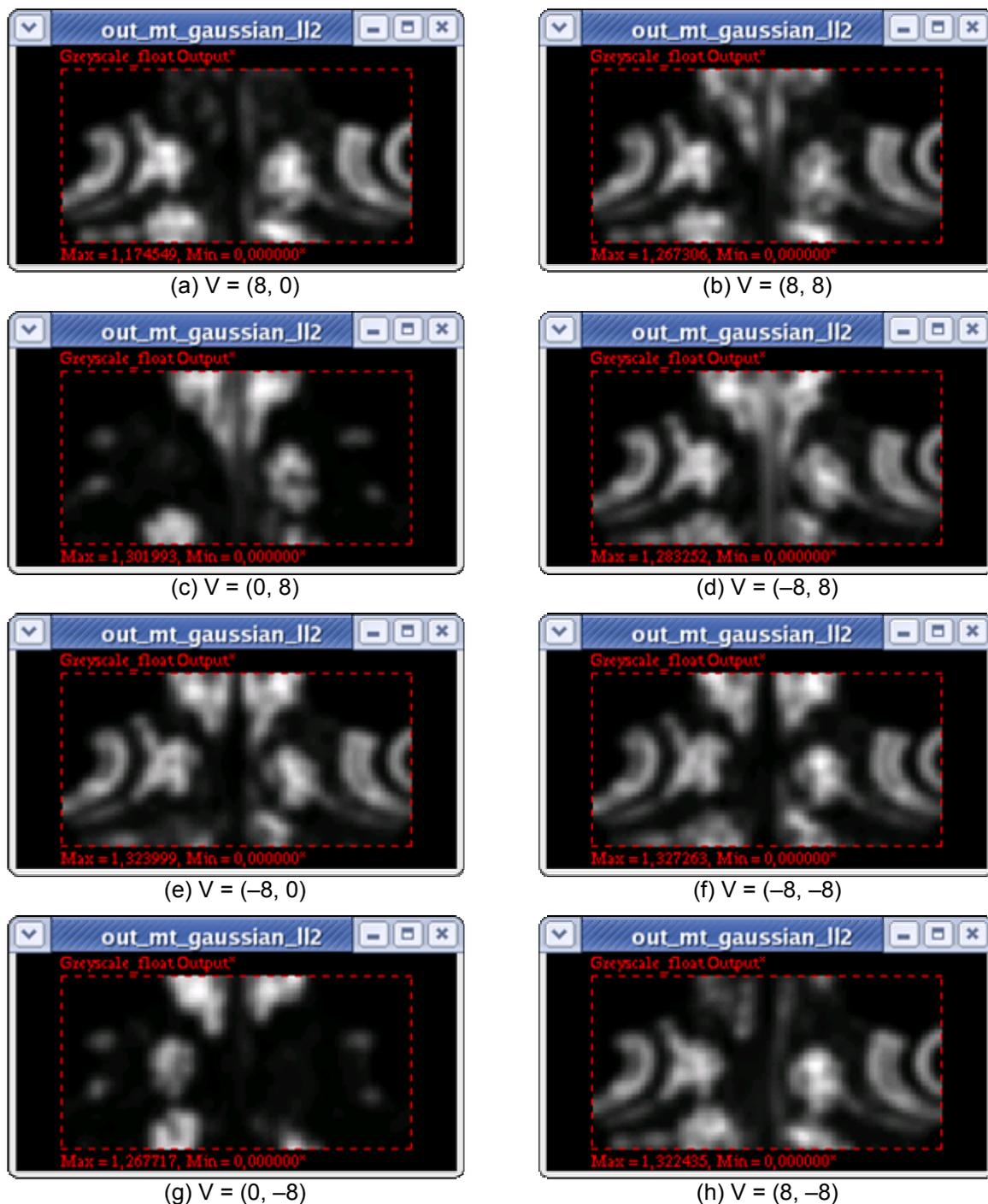


Figura 72 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 8.

Nas Figura 69-b, 71 e 72 podemos concluir, visualmente, que, dentre as 17 camadas MT apresentadas, aquela que possui a área superior mais escura e, portanto, com valores mais próximos de zero, é a camada sintonizada na velocidade (8, 0); a camada que possui a área direita com valores mais próximos de zero, é a camada sintonizada na velocidade (0, -8); a camada que possui a área inferior com valores mais próximos de zero, é a camada sintonizada na velocidade (-8, 0); e a camada que possui a área esquerda com valores mais próximos de zero, é a camada sintonizada na velocidade (0, 8).

Essa constatação visual também foi confirmada pelo processamento feito com as 65 camadas MT, produzindo um resultado em conformidade com o esperado.

O método proposto para este experimento estabelece que o esta comparação seja feita para cada célula de V1. A reunião dos valores mínimos de MT obtidos para cada célula constitui o mapa de velocidades, conforme mostrado na Figura 73.

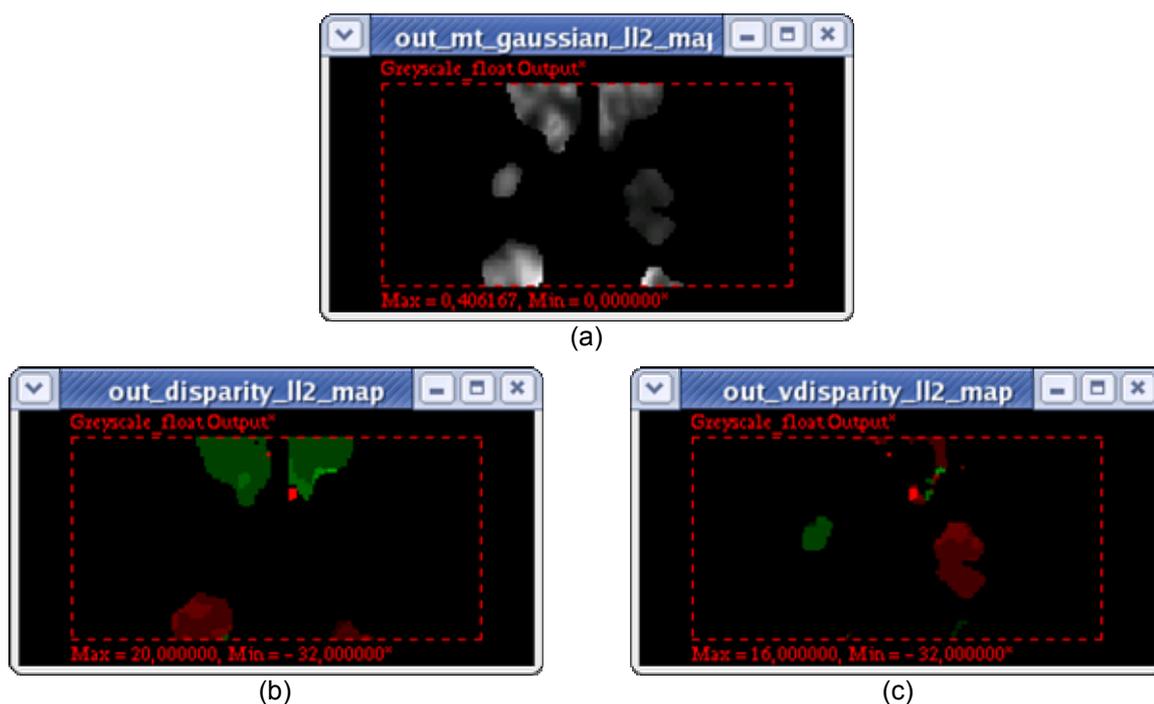


Figura 73 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.

Na Figura 73-a vemos o mapa de valores mínimos de MT obtidos para cada célula. As áreas mais claras indicam valores maiores e as áreas escuras, valores mais próximos de zero. Utilizando uma função de limiar, identificamos áreas em que a disparidade espaço-temporal possui grau elevado de incerteza. Na Figura 73-b,

vemos que a maior área foi calculada corretamente, com disparidade horizontal igual a 8 para a área superior, -8 para a área inferior e nula para as áreas direita e esquerda. Na Figura 73-c, também a maior área foi calculada corretamente, com disparidade vertical igual a 8 para a área esquerda, -8 para a área direita e nula para as áreas superior e inferior. Os valores mais imprecisos coincidem com as áreas em que o valor mínimo de MT supera o limiar.

Representação visual da velocidade de cada ponto na imagem

A representação visual das velocidades foi feita a partir dos resultados obtidos nos mapas de velocidades. A representação do plano de coordenadas log-polares de MT foi convertida em coordenadas cartesianas da imagem da retina. Dentro da região de movimento obtida na Figura 70, foram representados os pontos cujo valor mínimo de MT respeitou o limiar estabelecido. Cada velocidade foi representada por uma flecha na direção do movimento, com tamanho proporcional ao seu valor, conforme apresentado nas Figuras 74 e 75.

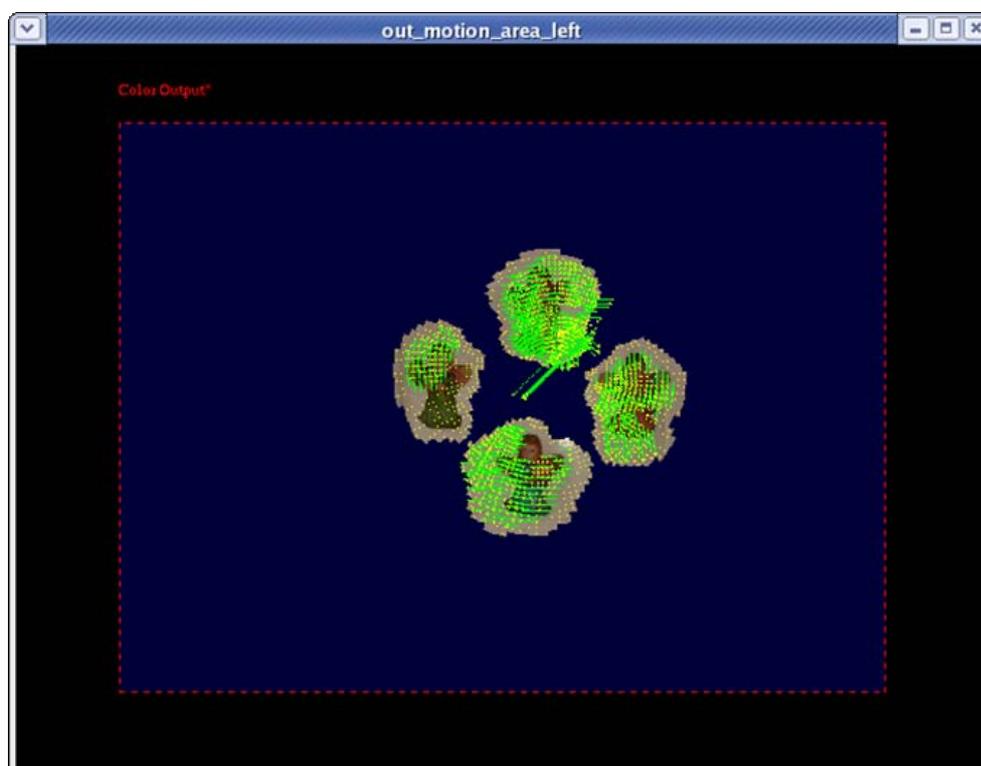


Figura 74 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,3$.

Comparação dos resultados obtidos com os esperados

Observando os resultados obtidos na Figura 74, vemos que na área central de cada objeto que se move as velocidades foram percebidas corretamente pelo sistema. Isso pode ser mais facilmente notado na Figura 75, que representa as velocidades resumidas num grid. A figura foi subdividida em quadrados de 10 x 10 pixels, e em cada quadrado foi representada apenas uma flecha de velocidade, correspondente à célula que apresentou o menor valor de saída MT.

Nas áreas de contorno dos objetos obtivemos alguns resultados incorretos. As áreas de resultados menos confiáveis são aquelas que ficam encobertas pelos objetos após os seus movimentos. O resultado mais distante do esperado foi na parte inferior do objeto localizado à esquerda, que ficou com velocidade nula.



Figura 75 – Representação visual das velocidades resumidas num grid de largura 10 x 10 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.

4.4 Percepção de movimento de rotação de um objeto

Neste experimento, foi utilizado um prato de cerâmica, com diâmetro igual a 260 mm, correspondendo a cerca de 316 pixels na imagem. No instante seguinte, esse objeto sofreu uma rotação de 10,75 graus no sentido horário. Na região da borda do objeto, isso corresponde a um deslocamento tangencial de cerca de 26 pixels na imagem. No centro do objeto, o deslocamento é nulo. As Figuras 76 e 77 apresentam as cenas capturadas nos dois instantes. A Figura 78 mostra o campo visual. A Figura 79 representa o mapeamento log-polar do campo visual, em torno do centro de atenção, nos dois instantes.

Identificação da região de movimento

Usando a estrutura de detecção de disparidade espaço-temporal nula, podemos identificar na imagem quais são as regiões de repouso e as regiões de movimento. A Figura 80 mostra a saída da camada MT para essa estrutura. A área mais clara corresponde às células com valor de saída maior; a área escura corresponde às células com valor de saída nulo ou bem pequeno.

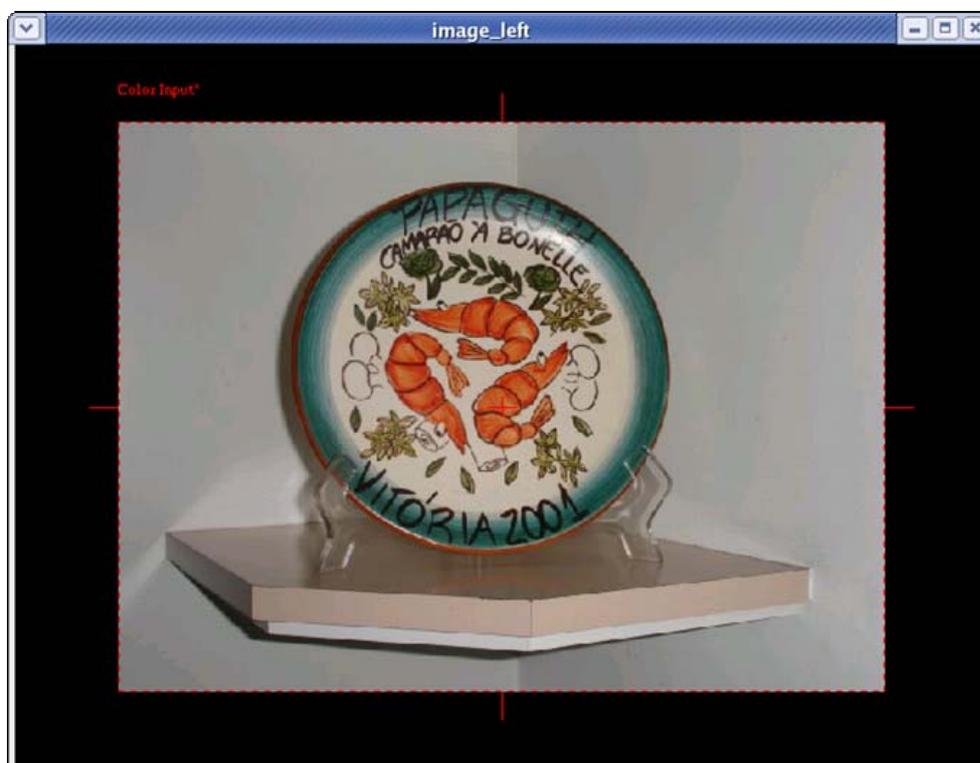


Figura 76 – Cena no instante 1. O centro de atenção está assinalado com uma cruz, localizado na coordenada (320, 240).

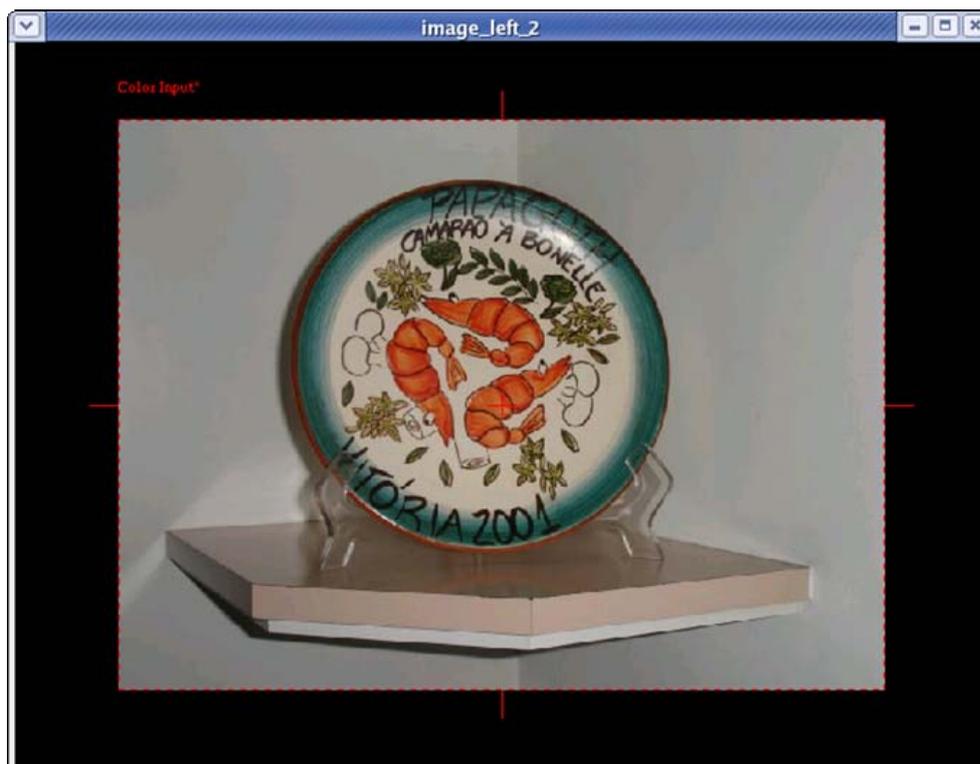


Figura 77 – Cena no instante 2. O centro de atenção está localizado na mesma coordenada. O objeto se movimentou cerca de 15 pixels à direita.



Figura 78 – Círculo que representa o campo visual em torno do centro de atenção em (320, 240).



Figura 79 – Mapeamento log-polar das imagens, em torno do centro de atenção, nos instantes 1 e 2.

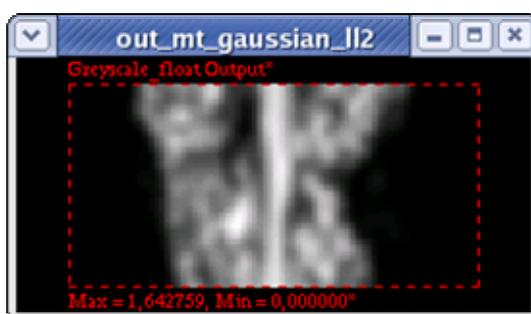


Figura 80 – Saída da camada MT para a estrutura que detecta velocidade nula. A área escura corresponde à região de velocidade nula; a área clara corresponde à região de movimento.

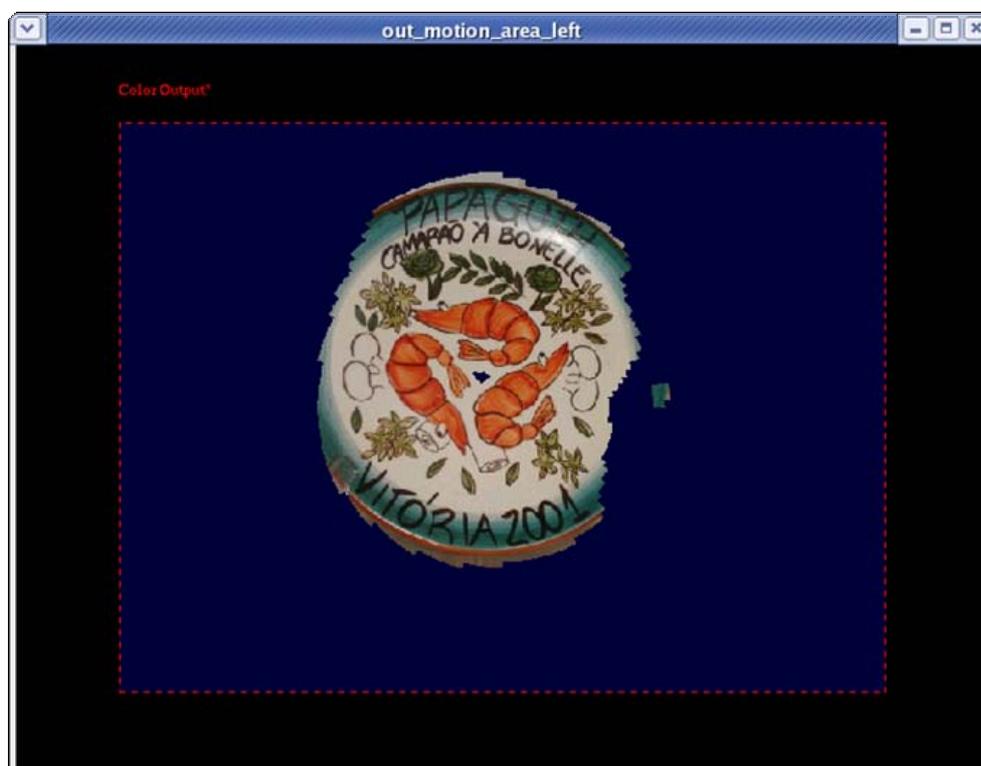


Figura 81 – Representação da região de movimento ou de incerteza, centro de atenção (320, 240).

A região detectada com velocidade nula foi apagada na Figura 81, ficando representada apenas o objeto que se movimentou e uma região de incerteza, que corresponde ao contorno do objeto.

Emulação do movimento ocular de sacada

A fim de melhorar a representação do objeto em V1, emulamos uma seqüência de movimentos oculares de sacada, colocando a região de movimento na fóvea. O resultado dessa operação no sistema é apresentado nas Figuras 82 e 83. A Tabela 6 mostra a seqüência de convergência das sacadas.

Tabela 6 – Experimento 4: Seqüência de convergência das sacadas

Sacada	Deslocamento	Coordenadas
1	(-9, +9)	(311, 249)
2	(-5, +6)	(306, 255)
3	(-3, +5)	(303, 260)
4	(-2, +3)	(301, 263)
5	(-1, +2)	(300, 265)
6	(-1, +2)	(299, 267)
7	(0, +1)	(299, 268)
8	(0, 0)	(299, 268)

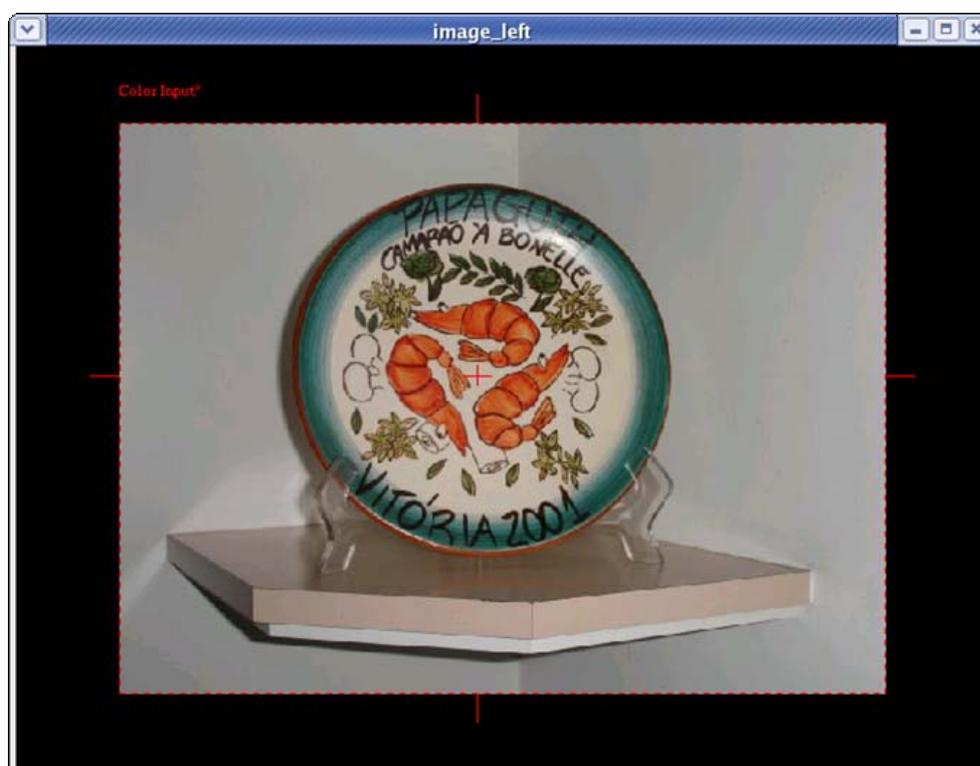


Figura 82 – Cena no instante 1. O centro de atenção foi deslocado para a coordenada (299, 268), por meio da emulação de uma seqüência de movimentos oculares de sacada, até atingir convergência.

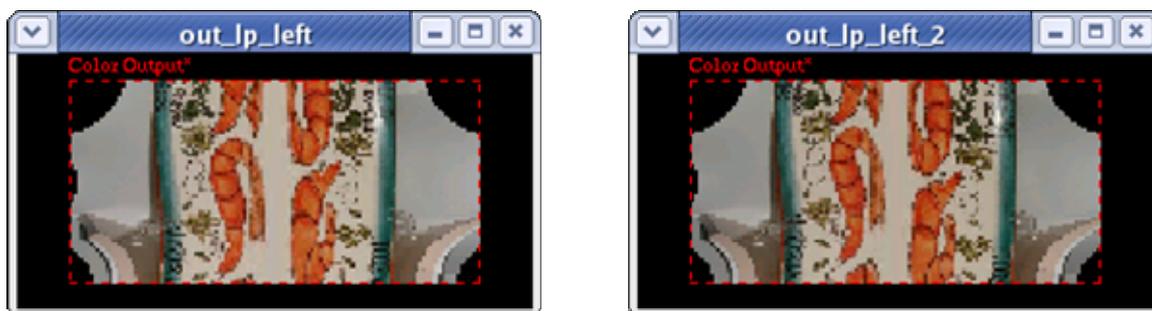


Figura 83 – Mapeamento log-polar das imagens, em torno do centro de atenção (299, 268), nos instantes 1 e 2, após a seqüência de movimentos oculares de sacada.

Computação de cada camada MT sintonizada numa disparidade espaço-temporal predefinida

O processamento de cada camada MT foi feito por meio de uma estrutura em camadas, conforme a Figura 14.

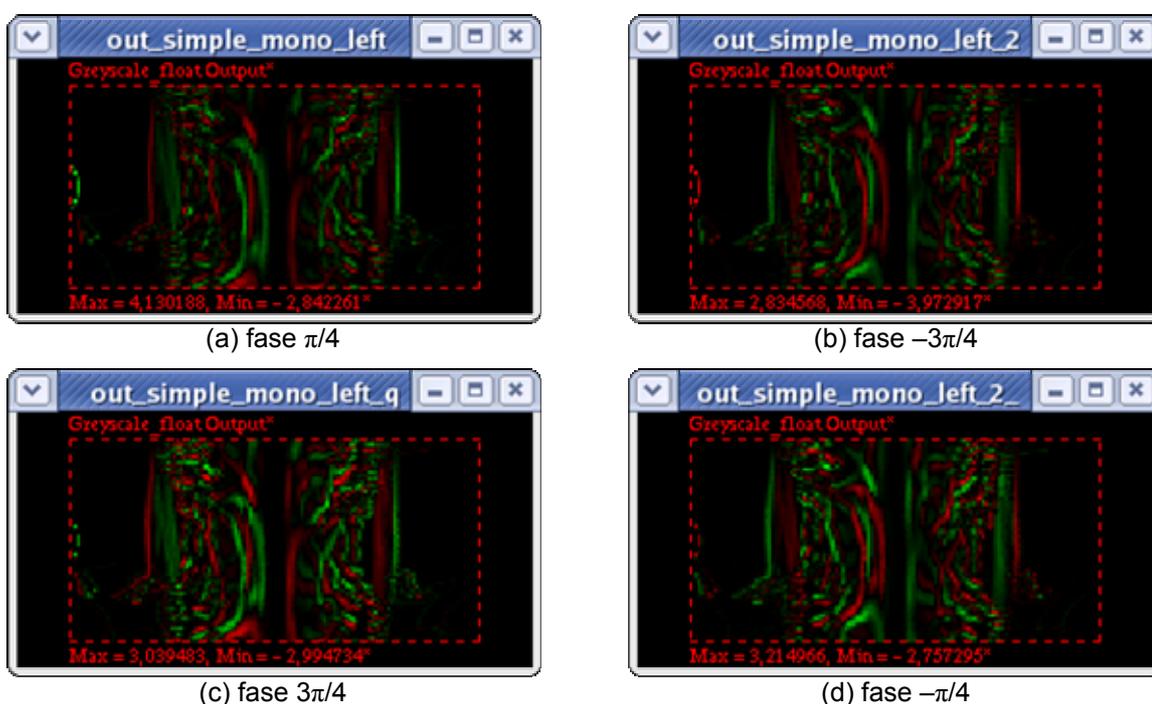


Figura 84 – Representação de camadas de células em V1. (a) células simples monoculares sintonizadas numa retina no instante 1; (b) células simples monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a), e com diferença de fase igual a π ; (c) e (d) células simples análogas a (a) e (b), em quadratura de fase.

Foram processadas 65 estruturas separadamente, sendo uma estrutura sintonizada com a velocidade nula e as demais sintonizadas em cada uma das velocidades relacionadas na Tabela 1. A título de exemplificação, mostramos nas Figuras 84 a 86 os resultados das camadas intermediárias somente para a estrutura sintonizada na velocidade nula.

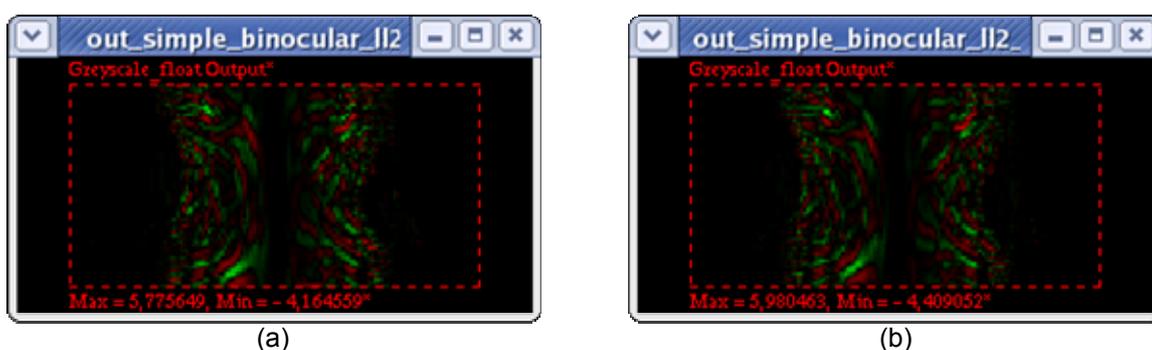


Figura 85 – Representação de camadas de células em V1. (a) células simples temporais que computam as saídas de uma mesma retina nos instantes 1 e 2, com disparidade espaço-temporal nula e com diferença de fase igual a π ; (b) células simples temporais análogas a (a), em quadratura de fase.

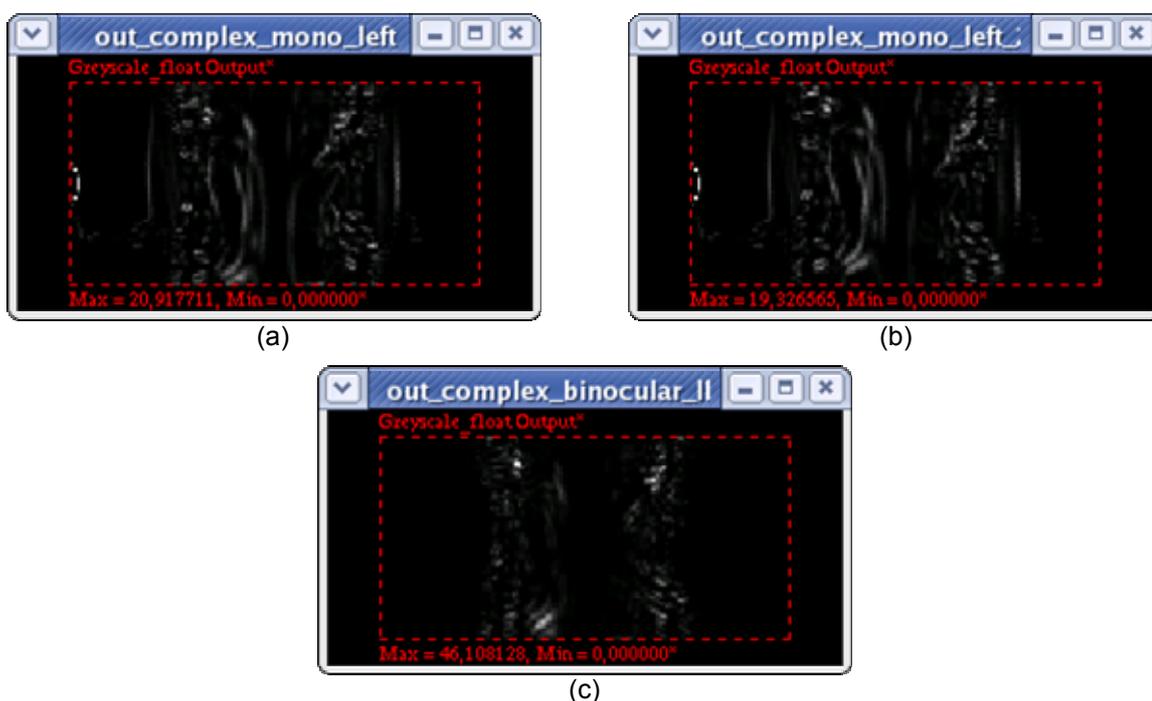


Figura 86 – Representação de camadas de células em V1. (a) células complexas monoculares sintonizadas numa retina no instante 1; (b) células complexas monoculares sintonizadas na mesma retina no instante 2, guardando uma disparidade espaço-temporal nula em relação a (a); (c) células complexas temporais, com disparidade espaço-temporal nula.

Na Figura 87, apresentamos o resultado da camada MT sintonizada na velocidade nula, em duas versões: (a) antes de aplicar o algoritmo de *spatial pooling*; (b) após aplicar o algoritmo de *spatial pooling*. O efeito visivelmente notável é a suavização da função de saída de MT, por meio da cooperação entre as células vizinhas nesta camada.

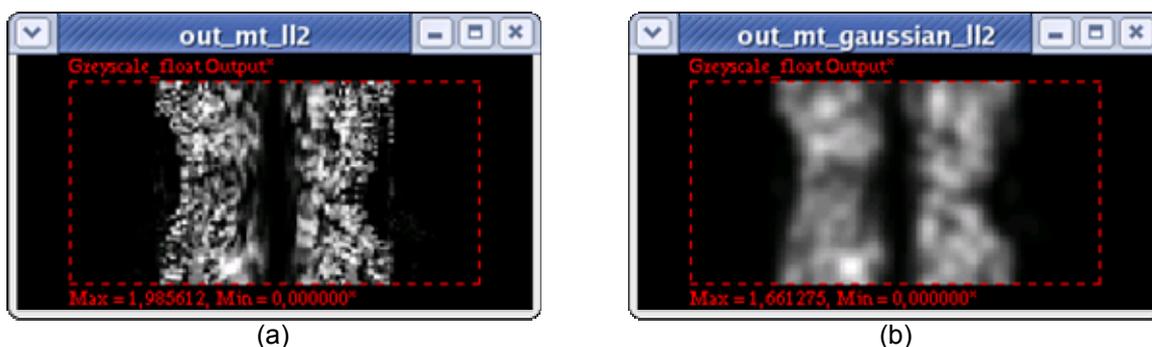


Figura 87 – Representação de camadas de células em MT. (a) células de MT que normalizam a saída das células complexas temporais pela soma das células complexas monoculares, sintonizadas na mesma retina, com disparidade espaço-temporal nula; (b) células de MT de (a) com *spatial pooling*.

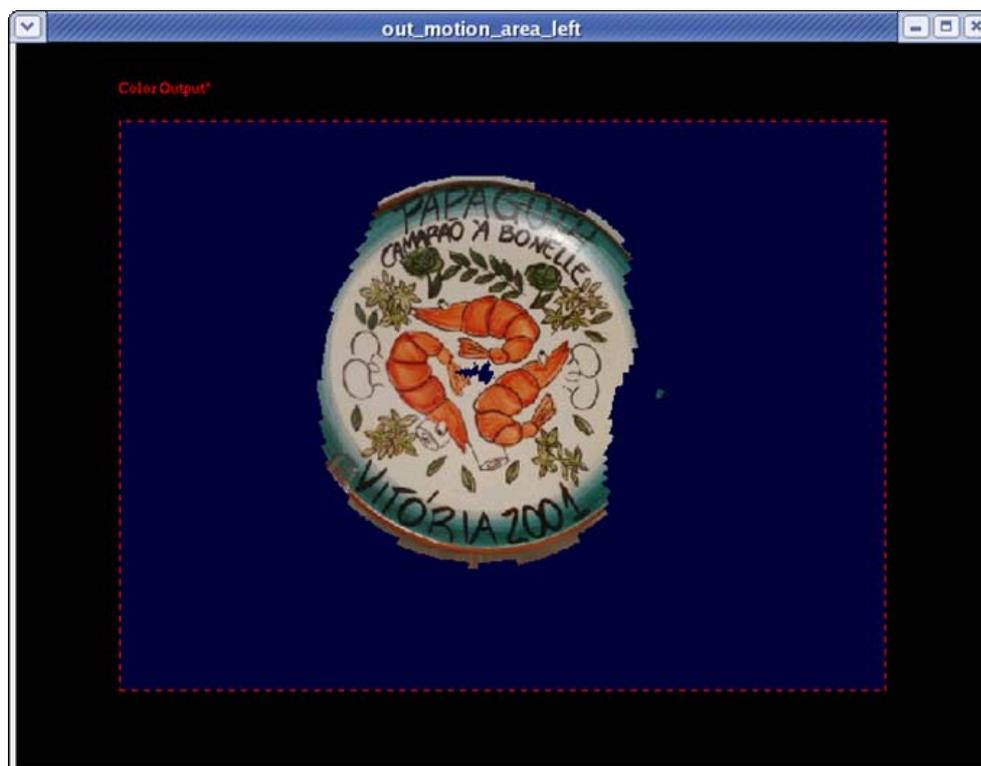


Figura 88 – Representação da região de movimento ou de incerteza, centro de atenção (299, 268).

Utilizando uma função de limiar sobre a saída de MT, com valor de limiar $L = 0,1$, chegamos à separação das regiões de repouso e de movimento, conforme mostrado

na Figura 88. Observamos que houve uma melhora em relação ao resultado inicial obtido na Figura 81. A área central do objeto possui velocidade nula. As duas regiões de bordas do objeto, para as quais não foi detectado movimento, são regiões com uma textura de pouquíssima variação. A análise de um movimento angular neste objeto fica muito dificultada quando se consideram essas regiões.

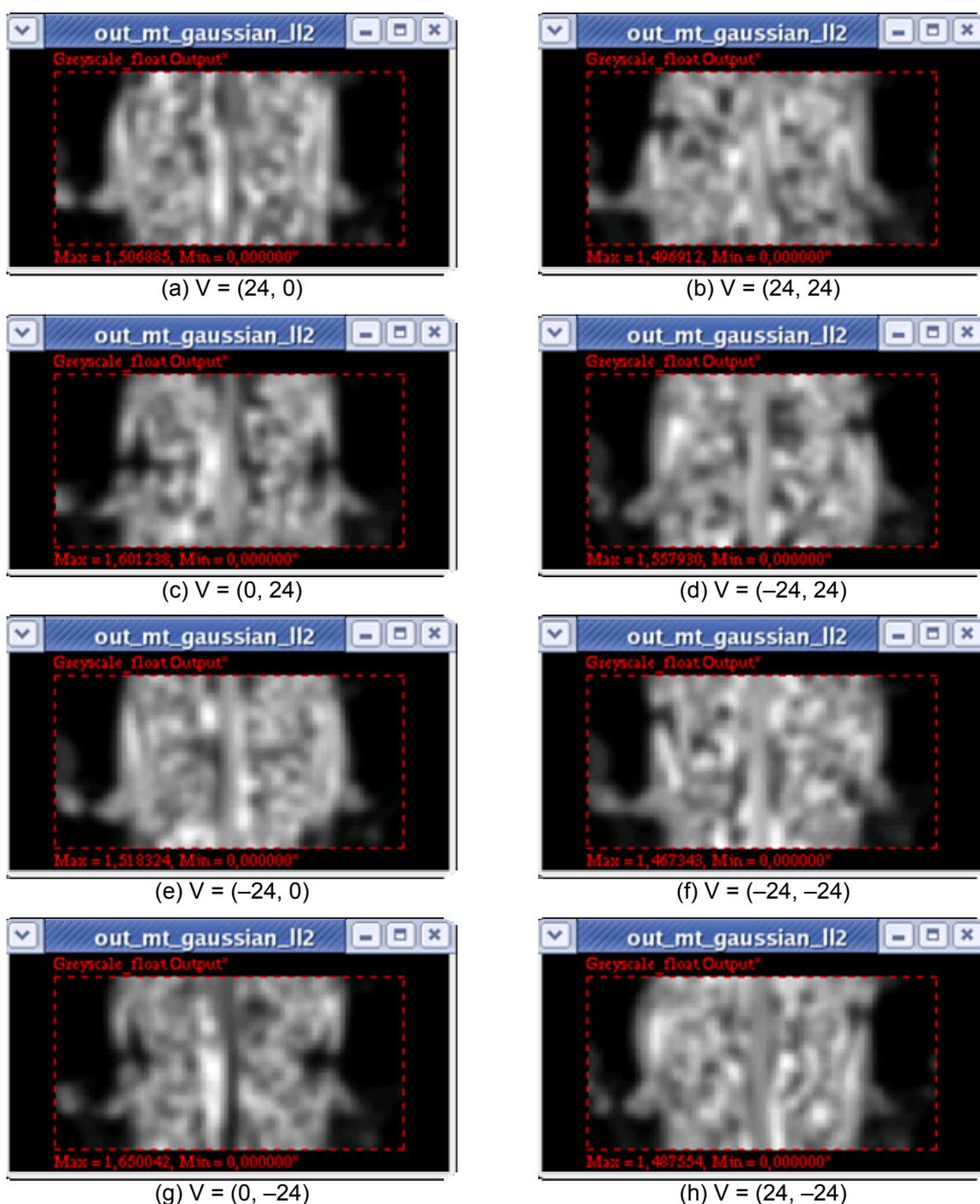


Figura 89 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 24.

Criação do mapa de velocidades

As Figuras 87-b, 89 e 90 mostram as camadas MT sintonizadas em 16 velocidades distintas, para ilustração do processo de criação do mapa de velocidades. O processo completo neste experimento comparou 65 camadas MT.

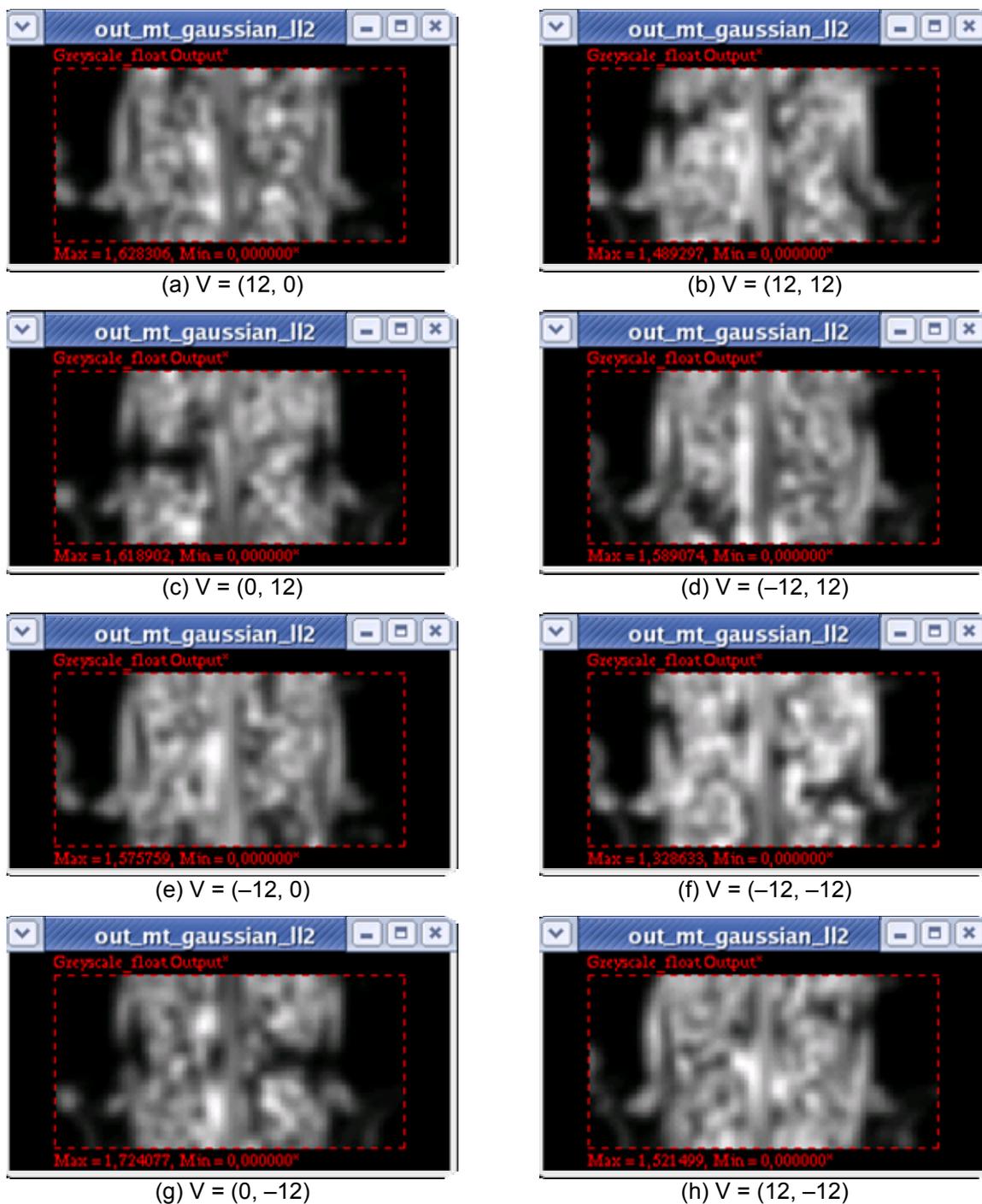


Figura 90 – Representação das camadas de células em MT, sintonizadas em oito velocidades diferentes, cada uma numa direção e sentido distinto, e com pelo menos uma componente horizontal ou vertical de módulo igual a 12.

Nas Figura 87-b, 89 e 90 podemos concluir, visualmente, que, dentre as 16 camadas MT apresentadas, não existe uma que prevaleça sobre todas as demais.

O método proposto para este experimento estabelece que o esta comparação seja feita para cada célula de V1. A reunião dos valores mínimos de MT obtidos para cada célula constitui o mapa de velocidades, conforme mostrado na Figura 91.

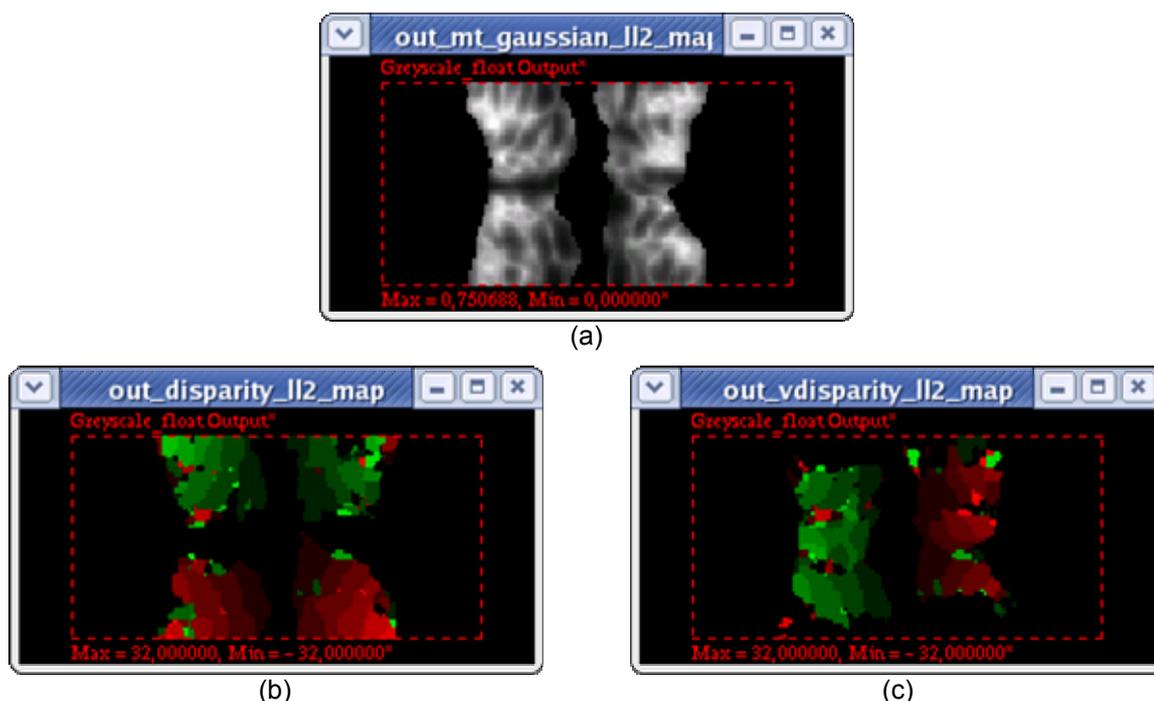


Figura 91 – (a) Mapa de valores mínimos de MT para cada célula; (b) mapa de velocidades no eixo horizontal; (c) mapa de velocidades no eixo vertical.

Na Figura 91-a vemos o mapa de valores mínimos de MT obtidos para cada célula. As áreas mais claras indicam valores maiores e as áreas escuras, valores mais próximos de zero. Utilizando uma função de limiar, identificamos áreas em que a disparidade espaço-temporal possui grau elevado de incerteza. Na Figura 91-b, vemos que a disparidade horizontal é predominantemente positiva na parte superior do mapa de velocidades, e negativa na parte inferior. Além disso, os valores possuem módulos maiores quanto mais distantes do centro. Na Figura 91-c, vemos que a disparidade vertical é predominantemente positiva na parte esquerda do mapa de velocidades, e negativa na parte direita. Além disso, os valores possuem módulos maiores quanto mais distantes do centro. Esses resultados estão de acordo com o esperado para o movimento rotacional.

Representação visual da velocidade de cada ponto na imagem

A representação visual das velocidades foi feita a partir dos resultados obtidos nos mapas de velocidades. A representação do plano de coordenadas log-polares de MT foi convertida em coordenadas cartesianas da imagem da retina. Dentro da região de movimento obtida na Figura 88, foram representados os pontos cujo valor mínimo de MT respeitou o limiar estabelecido. Cada velocidade foi representada por uma flecha na direção do movimento, com tamanho proporcional ao seu valor, conforme apresentado nas Figuras 92 e 93.

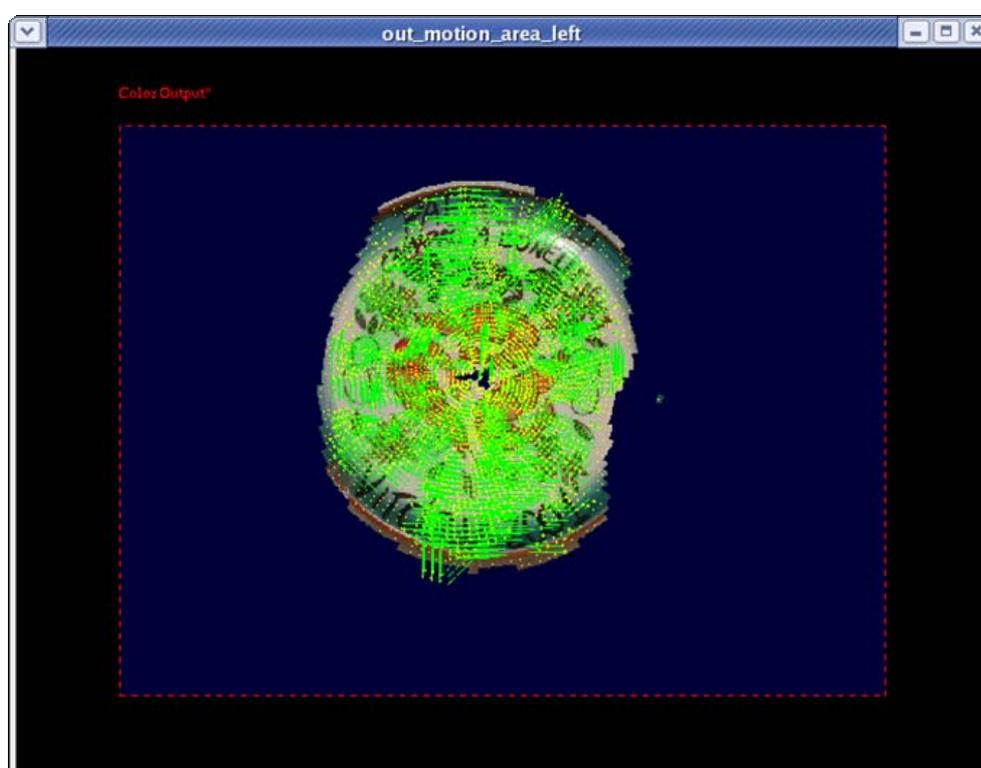


Figura 92 – Representação visual da velocidade em cada ponto na imagem, com limiar $L = 0,7$.

Comparação dos resultados obtidos com os esperados

Observando os resultados obtidos na Figura 92, vemos que o sistema percebeu corretamente a maioria dos pontos na área central do objeto que se move. Isso pode ser mais facilmente notado na Figura 93, que representa as velocidades resumidas num grid. A figura foi subdividida em quadrados de 16×16 pixels, e em cada quadrado foi representada apenas uma flecha de velocidade, correspondente à célula que apresentou o menor valor de saída MT.

Nas áreas de contorno dos objetos obtivemos alguns resultados incorretos. As áreas de resultados menos confiáveis são aquelas que ficam mais distantes da fóvea.

Conforme esperado, os pontos mais próximos do centro do objeto que sofre a rotação possuem velocidades menores, enquanto os pontos mais afastados do centro possuem velocidades maiores.

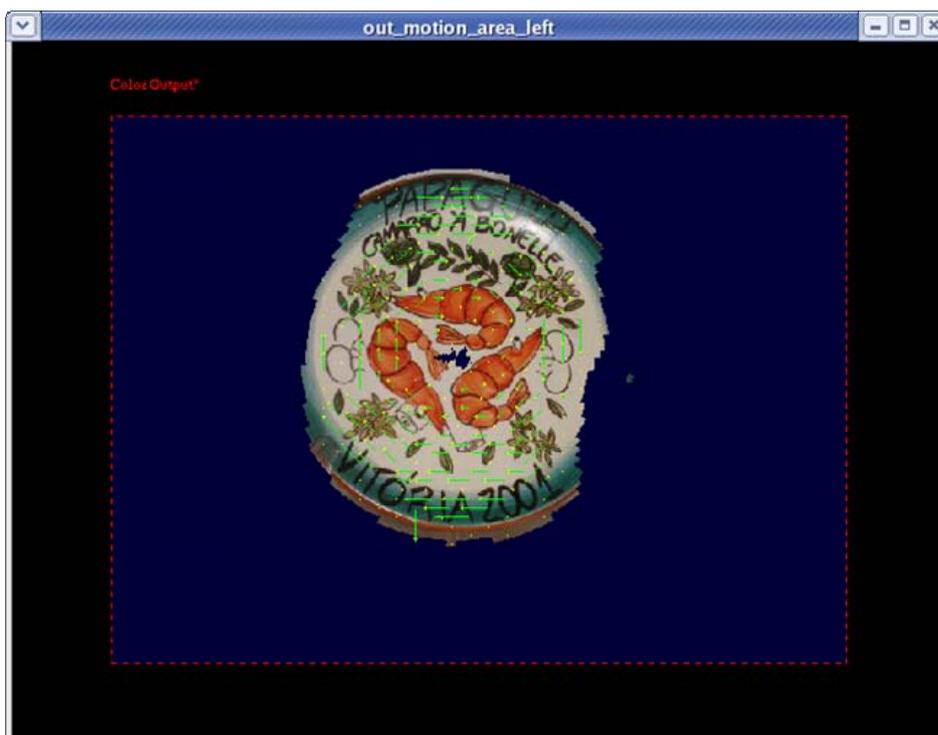


Figura 93 – Representação visual das velocidades resumidas num grid de largura 16 x 16 pixels. Em cada quadrado do grid é representada a velocidade da célula que possui menor valor de saída MT.

5 DISCUSSÃO

A seguir é feita uma comparação da abordagem deste trabalho de pesquisa com a de dois outros trabalhos anteriores relativos ao mesmo tema e uma análise crítica deste trabalho.

5.1 Trabalhos anteriores

Adelson e Bergen [ADE85] propuseram modelos de energia sintonizados em frequências espaço-temporais para percepção de movimento. Assim como o modelo aqui proposto, aqueles são fisiologicamente plausíveis. Os modelos propostos por Adelson e Bergen utilizam células simples e complexas de V1, enquanto o modelo aqui proposto utiliza também células da camada MT, o que possui uma correspondência maior com o sistema biológico. Assim como o modelo aqui proposto, os modelos de Adelson e Bergen não possuem uma abordagem focada no problema abstrato de correspondência de pontos entre duas imagens. Em vez disso, a abordagem utilizada é a de que a orientação espaço-temporal é uma propriedade dos estímulos visuais, resultante da extração de energia por meio de filtros.

Qian e Andersen [QIA97a] evoluíram sobre a proposta de Adelson e Bergen e propuseram um modelo integrado de percepção de movimento e percepção estereoscópica de profundidade, fisiologicamente plausível. No modelo de Qian e Andersen as duas funções de percepção visual são processadas simultaneamente pelo mesmo tipo de neurônio, enquanto no modelo aqui proposto os processamentos das disparidades binoculares e das disparidades espaço-temporais são realizados por células especializadas em uma ou em outra função. Contudo, no modelo proposto neste trabalho (Figuras 12, 13 e 14), existe uma grande similaridade na arquitetura neural proposta para ambos os tipos de processamento, o que facilita a sua integração.

Em ambos os modelos citados, para se determinar a velocidade de um ponto, verifica-se quais são as frequências espaço-temporais preferenciais na célula correspondente no sistema visual. Adicionalmente, no modelo de Qian e Andersen, para se determinar a disparidade binocular, verifica-se quais são a diferença de fase e a frequência espacial horizontal preferenciais. Ou seja, nesses modelos citados, as

células têm os seus campos receptivos centrados na mesma posição, mas com diferentes frequências e diferentes fases. Diferentemente, no modelo aqui proposto, tanto a velocidade quanto a disparidade binocular de um ponto são determinadas pela diferença de posição, ou seja, as células têm os seus campos receptivos centrados em posições diferentes, mas com a mesma diferença de fase e a mesma frequência.

5.2 Análise crítica

O processamento de funções de baixo nível do sistema visual, tais como percepção de movimento e de profundidade, poderia ser resolvido por meio de algoritmos e abordagens puramente abstratas, sem nenhuma correlação com a fisiologia humana. De fato, vários trabalhos têm adotado essa abordagem.

Contudo a abordagem adotada neste trabalho é a de implementar computacionalmente uma modelagem que seja fisiologicamente plausível. O objetivo maior é compreender como o cérebro realiza este processamento e depois emulá-lo no sistema computacional. O benefício esperado para esse tipo de abordagem baseada na fisiologia deverá ser alcançado mais adiante, nos estudos do processamento de funções cerebrais de alto nível. Ademais, se a Natureza levou mais de três bilhões de anos para evoluir as funções fisiológicas dos seres vivos até o nível das existentes no cérebro humano, é porque ali residem formas extremamente eficientes de lidar com tais funções.

Vários trabalhos acadêmicos têm utilizado imagens artificialmente criadas, a fim de eliminar diferenças de luz, sombra, brilho, contraste, cor, foco, profundidade, ângulos e outras imperfeições que surgem quando se trabalha com seqüências de imagens de uma mesma cena. Ao contrário dessa abordagem, neste trabalho optamos por utilizar imagens do mundo real, com imprecisões e diferenças, sem qualquer retoque ou ajuste que tornasse mais fácil o trabalho do sistema computacional. A qualidade imperfeita das imagens é evidente. Mesmo assim, o sistema apresentou bons resultados em todos os experimentos, o que demonstra a robustez do sistema perante as imprecisões dos estímulos visuais recebidos.

Apesar dos bons resultados obtidos nos experimentos realizados neste trabalho, utilizando a modelagem proposta, existem limitações e restrições no método.

A caracterização de células simples com campos receptivos modelados por função de Gabor é bem condizente com o conhecimento existente sobre a neurofisiologia do sistema visual. Contudo o cérebro humano utiliza múltiplas combinações de células com campos receptivos sintonizados em diferentes orientações, frequências e fases. Uma restrição adotada neste trabalho foi a de utilizar células com campos receptivos sempre com a mesma orientação, frequência e fase, sendo que as células correspondentes aos olhos direito e esquerdo guardam uma diferença de fase constante e igual a π .

A abordagem adotada neste trabalho para percepção de velocidade baseando-se em diferença de posição, em vez de diferença de fase, trouxe uma vantagem de não limitar a diferença máxima que uma célula pode detectar a $\pm\pi$ em cada frequência. O sistema é capaz de detectar qualquer velocidade dentro do campo visual. Por outro lado, existe uma limitação quanto à velocidade mínima detectável, que é igual a um pixel. Apesar de ter sido utilizado *spatial pooling*, que suaviza o efeito da discretização em pixels, a limitação existe.

Ao aplicar a convolução da função de Gabor com o estímulo presente no campo receptivo de cada célula simples, foi utilizado um valor escalar de intensidade luminosa de cada pixel, correspondente à média aritmética das intensidades individuais das cores: vermelha, verde e azul. Devido a essa limitação, o sistema fica incapaz de distinguir a diferença entre intensidades luminosas de pontos cujas cores sejam, por exemplo, no sistema RGB: (255, 0, 0) ou (0, 255, 0) ou (0, 0, 255) ou (85, 85, 85). Por outro lado, o cérebro humano distingue muito bem tais diferenças de cor. Isso pode levar a situações em que a velocidade de certos pontos seja detectada erroneamente pelo sistema.

O sistema implementado detecta velocidades no espaço (x, y, t) . Entretanto, no mundo real, o cérebro detecta velocidades no espaço (x, y, z, t) . Um efeito dessa limitação é que um objeto que esteja num movimento no eixo z , aproximando-se do observador, será representado com seus pontos em velocidades centrífugas; enquanto um objeto que se afasta do observador num movimento ao longo do eixo z , será representado com seus pontos em velocidades centrípetas.

Outra premissa utilizada no sistema é que o observador esteja estacionário em relação ao solo. No mundo real, o sistema visual utiliza movimentos oculares especiais para compensar o efeito do deslocamento do observador.

Por fim, uma limitação importante, pensando-se numa futura utilização deste sistema na navegação de um robô no mundo real, é o tempo de processamento do mapa de velocidades. O cérebro humano, utilizando o processamento paralelo de sua rede neuronal, é capaz de perceber as velocidades numa fração de segundo. O sistema implementado neste trabalho não se preocupou com a otimização do tempo de processamento do sistema.

6 CONCLUSÃO

Neste trabalho foi proposta uma modelagem matemático-computacional do sistema biológico da visão humana que é responsável pela percepção do movimento e da velocidade dos objetos. O modelo é biologicamente plausível, porque se baseia em estudos neurofisiológicos do comportamento e do mapeamento de células simples e complexas presentes nas camadas V1 e MT do córtex visual.

Nos experimentos realizados neste trabalho, foi constatado que o sistema implementado foi capaz de processar uma seqüência de imagens do mundo real e identificar quais eram os objetos que se moviam em cada cena e quais permaneciam imóveis.

Uma vez identificados os objetos que se moviam em cada cena, o sistema permitiu a emulação do movimento ocular de sacada, de modo que o alvo de interesse ficasse localizado no centro de atenção da fóvea, magnificando os seus detalhes.

O sistema também foi capaz de identificar, dentre um conjunto de possíveis direções e valores de velocidades, quais eram aquelas que mais se aproximavam da velocidade real de cada objeto em movimento.

A partir da velocidade conhecida dos objetos, o sistema permitiu emular o movimento ocular de perseguição suave, de modo que o objeto em movimento permanecesse no centro de atenção, ao longo de uma seqüência de imagens.

Foram realizados experimentos que avaliaram o comportamento do modelo frente a diversas situações distintas: movimento de um objeto na direção horizontal; movimento de um objeto em diagonal, ou seja, com uma combinação de componentes de velocidade horizontal e vertical; movimento simultâneo e independente de translação de múltiplos objetos; e movimento de rotação de um objeto em torno de um centro estacionário. O resultado apresentado pelo sistema foi bastante satisfatório em todos os experimentos realizados. Ressalte-se que as imperfeições das imagens utilizadas nos experimentos reforçam a constatação da robustez do sistema.

6.1 Trabalhos futuros

Os resultados obtidos neste trabalho, bem como as premissas e restrições existentes, ensejam a realização de trabalhos futuros que certamente trarão contribuições de grande valia.

- O custo dos equipamentos tem sido muito reduzido a cada ano, de modo que a implementação de um processamento paralelo das camadas neurais torna-se economicamente viável e traria um ganho de velocidade de processamento do sistema e uma diminuição significativa do tempo de resposta.
- Neste trabalho o sistema tratou as imagens como se fossem monocromáticas, de modo que um objeto na cor vermelha e outro na cor azul não teriam distinção. O estudo do tratamento das cores e do contraste de luminância das imagens tem um potencial de aprimorar a precisão da percepção das velocidades.
- O sistema implementado neste trabalho uniformizou os campos receptivos das células, mantendo-os com a mesma orientação, frequência e fase. No sistema biológico, as células possuem uma grande variedade. Uma área promissora para trabalhos futuros é a análise do impacto nos resultados do sistema, ao se implementar vários conjuntos de células simples em V1, com diferentes rotações do campo receptivo e diferentes frequências e fases.
- A fim de superar a limitação da velocidade mínima detectável, que é igual a um pixel, pode-se estudar uma forma de aprimoramento no sistema, que o torne capaz de detectar velocidades subpixel.
- Uma evolução natural desta pesquisa é a integração das estruturas de percepção de profundidade e de velocidade, de modo que o sistema adquira capacidade de perceber movimentos no eixo z, ou seja, de afastamento ou de aproximação em relação ao observador, bem como diferenciar tais movimentos de outros movimentos no plano x-y que envolvam diminuição ou aumento do tamanho de objetos.
- A fim de não perder de vista um dos objetivos principais dessa linha de pesquisa, a implementação da emulação dos movimentos oculares vestibulo-ocular e optocinético traria uma importante contribuição no sentido de viabilizar a

utilização deste sistema na navegação de um robô em movimento, no mundo real.

REFERÊNCIAS

- [ADE85] ADELSON, E.H.; BERGEN, J.R. Spatiotemporal energy models for the perception of motion. **Journal Optical Society of America A**, 2: 284–299, 1985.
- [ALE89] ALEKSANDER, I. Towards visual awareness in a neural system. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING ICONIP'97, 1997. **Proceedings...**, 1997.
- [BRA01] BRADDICK, O.; QIAN, N. The organization of global motion and transparency. In: **Motion vision: computational, neural, and ecological constraints** (J. M. Zanker and Z. Jochen, eds.), Springer-Verlag, 2001. p. 86-112.
- [BRA95] BRADLEY, D.C.; QIAN, N.; ANDERSEN, R.A. Integration of motion and stereopsis in cortical area MT of the macaque. **Nature**, 373:609-611, 1995.
- [CHE05] CHEN, Y. *et al.* Effects of attention on motion repulsion. **Vision Research**, 45:1329-1339, 2005.
- [CHE01] CHEN, Y.; MATTHEWS, N.; QIAN, N. Motion rivalry impairs motion repulsion. **Vision Research**, 41:3639-3647, 2001.
- [DEA99] DeANGELIS, G.C.; NEWSOME, W.T. Organization of disparity-selective neurons in macaque area MT. **The Journal of Neuroscience**. 19(4):1398-1415, 1999.
- [DEA00] DeANGELIS, G.C. Seeing in three dimensions: the neurophysiology of stereopsis. **Trends in Cognitive Science**, 4: 80–90, 2000.
- [DOD03] DODGE, R.; Five Types of eye movement in the horizontal meridian plane of the field of regard. **American Journal of Physiology**. 8:307-329, 1903.

- [FER02] FERNANDEZ, J.M.; WATSON, B.; QIAN, N. Computing relief structure from motion with a distributed velocity and disparity representation. **Vision Research**, 42:883-898, 2002.
- [GEE98] GEESAMAN, B.J.; QIAN, N. The effect of complex motion pattern on speed perception. **Vision Research**, 38:1223-1231, 1998.
- [GEG97] GEGENFURTNER, Karl R. Functional properties of neurons in macaque area V3. **The American Physiological Society**, 0022-3077/97, 1997.
- [HOF00] HOFFMAN, Donald D. **Inteligência visual: como criamos o que vemos; tradução de Denise Cabral Carlos de Oliveira**. Rio de Janeiro: Campus , 2000.
- [HUB62] HUBEL, D.H.; WIESEL, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **Journal of Physiology**, 160: 106-154, 1962.
- [JOH73] JOHANSSON, G. Visual perception of biological motion and a model for its analysis. **Perception & Psychophysics**, 14: 201-211, 1973.
- [JON87] JONES, J. P.; PALMER, L. A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. **Journal of Neurophysiology**, 58:1233-1258, 1987.
- [KAN00] KANDEL, Eric R.; SCHWARTZ, James H.; JESSELL Thomas M. **Principles of neural science**. 4th Ed. Prentice-Hall International, Inc.,2000.
- [KOM02] KOMATI, K.S.; DE SOUZA, A.F. Vergence control in a binocular vision system using weightless neural networks. In: INTERNATIONAL SYMPOSIUM ON ROBOTICS AND AUTOMATION, 4th, 2002, Los Alamitos. **Proceedings...** Los Alamitos: IEEE, 2002.
- [LUH04] LU, H.; QIAN, N.; LIU, Z. Learning motion discrimination with suppressed MT. **Vision Research**, 44:1817-1825, 2004.

- [MCI03] MATHER, G. The Visual Cortex. University of Sussex. Disponível em <http://www.lifesci.sussex.ac.uk/home/George_Mather/Linked%20Pages/Physiol/Cortex.html>. Acesso em 13 out. 2006.
- [MAT06] MATHER, G. The Visual Cortex. University of Sussex. Disponível em <http://www.lifesci.sussex.ac.uk/home/George_Mather/Linked%20Pages/Physiol/Cortex.html>. Acesso em 13 out. 2006.
- [MAT00] MATTHEWS, N.; GEESAMAN, B.J.; QIAN, N. Dependence of motion repulsion and rivalry on the distance between moving elements. **Vision Research**, 40:2025-2036, 2000.
- [MAT99] MATTHEWS, N.; QIAN, N. Axis-of-motion affects direction discrimination, not speed discrimination. **Vision Research**, 39:2205-2211, 1999.
- [MCI03] McINTYRE, M.E. Footnote to the 2003 Reith Lecture. University of Cambridge. Disponível em <<http://www.atm.damtp.cam.ac.uk/people/mem/reith.html>>. Acesso em 13 out. 2006.
- [MEN04] MENG, X.; CHEN, Y.; QIAN, N. Both monocular and binocular signals contribute to motion rivalry. **Vision Research**, 44:45-55, 2004.
- [MOV06] MOVELLAN, J.R. Tutorial on Gabor Filters. University of California, San Diego. Disponível em <<http://mplab.ucsd.edu/tutorials/tutorials.html>>. Acesso em 13 out. 2006.
- [MOV85] MOVSHON, J. A.; ADELSON, E. H.; GIZZI, M. S.; NEWSOME, W. The analysis of moving visual patterns. In C. Chagas, R. Gattass, C. Gross (eds.), **Pattern recognition mechanisms**. New York, Springer, 117-151, 1985.
- [OHZ97] 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.
- [OLI05] OLIVEIRA, H.; DE SOUZA, A.F. **Uma modelagem computacional de áreas corticais do sistema visual humano associadas à percepção**

- de profundidade**. 2005. 133 f. Dissertação (Mestrado em Informática) – Universidade Federal do Espírito Santo, Vitória, 2005.
- [POM06] POMERANZ, J.R.; PORTILLO, M. Psychology of Perception. Rice University. Disponível em <<http://www.owl.net.rice.edu/~psyc351/imagelist.htm>>. Acesso em 13 out. 2006.
- [QIA94a] QIAN, N. Computing stereo disparity and motion with known binocular cell properties, **Neural Computation**, 6:390-404, 1994.
- [QIA97a] QIAN, N.; ANDERSEN, R.A. A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. **Vision Research**, 37: 1683-1698(16), 1997.
- [QIA95] QIAN, N.; ANDERSEN, R.A. V1 responses to transparent and non-transparent motions. **Experimental Brain Research**, 103:41-50, 1995.
- [QIA94b] QIAN, N; ANDERSEN, R.A.; ADELSON, E.H. Transparent motion perception as detection of unbalanced motion signals. **Journal of Neuroscience**, 14:7357-7392, 1994.
- [QIA97b] QIAN, Ning; ZHU, Yudong. Physiological computation of binocular disparity. **Vision Research**, 37: 1811–1827, 1997.
- [TAN04] TANAKA, H; TAI, M; QIAN, N. Different predictions by the minimum variance and minimum torque-change models on the skewness of movement velocity profiles. **Neural Computation**, 16:2021-2040, 2004.
- [TOO82] TOOTELL, R. B.; SILVERMAN, M. S.; SWITKES, E.; VALOIS, R. L. Deoxyglucose analysis of retinotopic organization in primate striate cortex. **Science**, 218:902-904, Nov 26, 1982.
- [VON84] VON DER HEYDT R.; PETHERHANS E.; BAUMGARTNER G. Illusory contours and cortical neuron responses. **Science**, 224:1260-1262, 1984.
- [ZEK73] ZEKI, S. M. Colour coding of the rhesus monkey prestriate cortex. **Brain Research**, 422-427, 1973.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)