UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

GUSTAVO VRECH RIGO

**Modelando a atenção seletiva e a saliência visual através de redes
complexas**

SÃO CARLOS
2010

GUSTAVO VRECH RIGO

Modelando a atenção seletiva e a saliência visual através de redes complexas

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de mestre em Ciência.

Área de concentração: Física Computacional
Orientador: Prof. Dr Luciano da Fontoura Costa

SÃO CARLOS
2010

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES
TRABALHOS, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA
FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC

Rigo, Gustavo Vrech.

Modelando a atenção seletiva e a saliência visual através de redes
complexas / Gustavo Vrech Rigo; orientador Luciano da Fontoura
Costa.-- São Carlos, 2010.

130 p.

Dissertação (Mestrado – Programa de Pós-Graduação em Física -
Área de concentração: Física Aplicada) – Instituto de Física de São
Carlos da Universidade de São Paulo

1. Redes complexas. 2. Atenção seletiva. 3. Saliência visual. I.
Título.

FOLHA DE APROVAÇÃO

Gustavo Vrech Rigo

Dissertação apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências.
Área de Concentração: Física Aplicada -
Opção: Física Computacional

Aprovado(a) em: 22.07.2010

Comissão Julgadora

Prof(a). Dr(a). Francisco Aparecido Rodrigues

Instituição: ICMC/USP

Assinatura

Prof(a). Dr(a). Alneu de Andrade Lopes

Instituição: ICMC/USP

Assinatura

Prof(a). Dr(a). Gonzalo Travieso

Instituição: IFSC/USP

Assinatura

AGRADECIMENTOS

Agradeço à minha família.

Aos meus pais, pelo amor, conforto, presença e amizade. Principalmente pelo conhecimento, pedidos e conselhos, mesmo quando estou relutante em aceitar. Agradeço à minha irmã, pela sua companhia e apoio, e por ter me ensinado que eu não sou o único.

Meus tios e tias, e aos meus primos, meus amigos de mais longa data.

Aos meus amigos, tão presentes e fundamentais.

Ao pessoal de São Paulo, que por mais que tenhamos perdido o contato ainda me fazem lembrar da primeira parte da minha vida.

Ao Eduardo, e todo o pessoal da minha escola de ensino médio, que representam para mim um dos períodos mais intensos e festivos.

À Anna e as meninas, pelos incansáveis passeios, filmes e programas de índio que me ensinaram muito sobre o mundo.

Ao Waldomiro e o pessoal do RPG, que além da amizade, ensinaram para mim que as simulações podem ser muito divertidas.

Ao Tiago, meu amigo que conheço a mais tempo em São Carlos, pelas suas opiniões ácidas que apesar de tudo mudaram muito do jeito que eu enxergo a vida.

À Samarina pela companhia nem sempre pacífica, porém especial.

Ao Magali, o Foca, e o pessoal da Física, que foram muito além da amizade, companhia, bom humor e listas de mecânica quântica.

A todos que contribuíram para o meu desenvolvimento profissional.

Aos professores do Instituto de Física de São Carlos, pela minha excelente formação, e especialmente aos professores que a muito contragosto meu foram demasiadamente exigentes, me mostraram que muita da minha competência advém da superação de limites.

Ao professor Luciano, que me ajudou e orientou em praticamente toda a minha vida dentro da universidade.

Aos funcionários do Instituto de Física de São Carlos, pelo apoio fundamental ao meu crescimento acadêmico, e em especial ao pessoal da biblioteca, que sempre me atenderam prontamente em tudo que precisei, e ao pessoal dos setores de graduação e pós-graduação, que sempre me ajudaram a resolver meus problemas burocráticos (estes não foram poucos).

Ao pessoal do laboratório, que naturalmente permearam o limite entre o profissional e o pessoal.

“Não há fatos eternos, como não há verdades absolutas.”

Friedrich Nietzsche

*“Escolhe um trabalho de que gostes, e não terás que trabalhar
nem um dia na tua vida.”*

Confúcio

*“Desamparado, no meio da rua, sentindo que o chão lhe fugia
debaixo dos pés, tentou conter a aflição que lhe subia pela
garganta. Agitava as mãos à frente da cara, nervosamente, como
se nadasse naquilo a que chamara um mar de leite, mas a boca já
se lhe abria para lançar um grito de socorro.”*

Saramago

RESUMO

RIGO, Gustavo Vrech. **Modelando a atenção seletiva e a saliência visual através de redes complexas**. 2010. p. 130, Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

A atenção seletiva é uma característica central do sistema visual humano, uma vez que todo o cérebro é otimizado de modo a perceber as informações ao seu redor da forma mais rápida possível. Porém, em geral os trabalhos nesta área apenas verificam quais são as regiões de maior frequência da atenção seletiva, dando pouca importância para a sua mecânica. A presente dissertação propõe um modelo que represente a atenção seletiva como uma rede complexa, combinando naturalmente as áreas de redes complexas, cadeias de Markov, análise de imagens, atenção seletiva e saliência visual num modelo biologicamente plausível para simular a atenção seletiva. O modelo propõe que pontos importantes da imagem, pontos salientes, sejam caracterizados como vértices da rede complexa, e que as arestas sejam distribuídas de acordo com a probabilidade da mudança de atenção entre dois vértices. Desta forma, a mecânica da atenção seletiva seria simulada pela mecânica da rede complexa correspondente. Foram estudadas imagens em níveis de cinza, sendo estas correspondentes à cena observada. A probabilidade de mudança entre duas regiões, as arestas da rede, foram definidas através de diversos métodos de composição da saliência visual, e as redes resultantes comparadas com redes complexas provenientes de um experimento protótipo realizado. A partir deste experimento foram propostos refinamentos no modelo original, tornando assim a mecânica do modelo o mais próximo possível da mecânica humana da atenção seletiva.

Palavras-chave: Redes complexas. Atenção seletiva. Saliência visual.

ABSTRACT

RIGO, Gustavo Vrech. **Modeling the selective attention and visual saliency using complex networks**. 2010. p.130, Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

Selective attention is a central feature of the human visual system, since the entire brain is optimized in order to understand the information around as quickly as possible. In general works in this area only search which regions has a higher frequency of selective attention, with little consideration for their mechanics. This study proposes a model that represents the selective attention as a complex network, combining naturally areas of complex networks, Markov chains, image analysis, selective attention and visual salience in a biologically plausible model to simulate the selective attention. The model proposes that the important points of the image, salient points, are identified as vertices of the complex network, and the edges are distributed according to the probability of shift of attention between two vertices. Thus, the mechanics of selective attention would be simulated by the mechanics of correspondent complex network. We studied images in gray levels, which are corresponding to the scene observed. The probability of switching between two regions, the edges of the network were identified through various methods of visual saliency composition, and the resulting networks compared with complex networks from a prototype experiment performed. From this experiment were proposed refinements to the original model, thereby making the mechanical design as close as possible to the mechanics of human selective attention.

Key words: Complex networks. Selective attention. Visual saliency.

LISTA DE FIGURAS

Figura 1.1 - Ilustração básica do modelo. Em (a) é apresentada uma imagem simples contendo três objetos. Em (b), é a reação básica esperada de um observador, segmentar cada um dos objetos, e voltar a sua atenção singularmente para cada um deles.....	24
Figura 2.1 - Comparação entre o esquema ótico de uma câmera <i>pin-hole</i> (a), uma câmera fotográfica com lente (b) e o olho humano (c)..	29
Figura 2.2 - Perfil anatômico do olho humano.....	30
Figura 2.3 - Estrutura das células presentes na retina. À direita estão presentes um cone e oito bastonetes, também estão presentes células bipolares e ganglionares.....	32
Figura 2.4 - Esboço da acuidade visual em função da posição ocular. A região da fóvea possui a maior acuidade visual, e esta cai ao se distanciar da fóvea. A exceção é o ponto cego, local desprovido de fotorreceptores.....	34
Figura 2.5 - Esquema de ativação das células ganglionares presentes na retina. Em (a) é evidenciada a ativação padrão, (b) evidencia uma taxa de ativação positiva, enquanto que (c) uma taxa de ativação negativa.....	37
Figura 2.6 - Exemplo simples de saliência visual. A roupa vermelha naturalmente chama mais atenção de um observador, fazendo com que este dê preferência da sua atenção àquela roupa. Neste caso é definido que a saliência da roupa vermelha é maior do que a saliência das outras roupas.....	39
Figura 2.7 - Exemplo de experimento psicofísico para detecção de saliência. A região composta por caracteres “T”s rotacionados se destaca mais da região composta por caracteres “T” do que	

a região composta por caracteres “L”s.....	40
Figura 2.8 - Conjuntos apresentados para observadores para procura. Em (a) existem 3 caracteres “O” e um “Q”, enquanto que (b) possui um “O” e 3 “Q”s. Em (c) encontram-se 30 caracteres “O” e um “Q”, enquanto que em (b) existem 30 “Q”s e um “O”. Comparando-se (a) e (c), evidencia-se que a eficiência para determinadas procuras independe do número de distratores. Comparando (c) e (d) pode-se notar que é mais fácil encontrar um “Q” dentre um conjunto de “O”s do que um “O” dentro de um conjunto de “Q”s.....	
	42
Figura 2.9 - Exemplo de matriz tratada (a), sua representação como imagem (b), a aproximação do gradiente das posições centrais (c), a imagem deste gradiente (d). Em (e) é possível ver o gradiente obtido com o operador de Sobel da imagem (f).....	
	45
Figura 2.10 - Exemplo de um sistema representado por um grafo. Os estados são representados pelas seguintes figuras geométricas: círculo, quadrado e triângulo. As probabilidades de transição são evidenciadas pelas setas.....	
	47
Figura 2.11 - Matriz M escrita de uma forma didática, evidenciando os índices de cada estado possível.....	
	48
Figura 2.12 - Rede complexa baseada numa rede de amizades. Os vértices são caracterizados por pessoas, enquanto as arestas por ligações de amizade entre estas pessoas.....	
	51
Figura 2.13 - Rede complexa geográfica baseada na distribuição dos estados do Brasil. Os vértices foram distribuídos de forma a representar cada um estado, enquanto que as arestas foram distribuídas de acordo com a presença ou não de uma fronteira entre os dois estados. A rede foi sobreposta com o mapa Brasileiro de modo a facilitar o seu entendimento.....	
	56

Figura 2.14 - Redes regulares. Os vértices são representados pelos círculos, sendo os azuis vértices genéricos, os verde claros são os vizinhos do vértice verde escuro. Em (a) é observada a condição teórica, onde todo o espaço é preenchido pela rede, enquanto que em (b) é adotada a condição periódica de contorno, necessária para representar tal rede num ambiente computacional.....	57
Figura 3.1 - Em (a) imagina-se que o observador está num lugar desconhecido, e começa a reconhecer a cena olhando inicialmente para a região delimitada pelo círculo vermelho. Em (b) são evidenciadas por círculos azuis as regiões as quais o observador poderá voltar sua atenção depois de terminar de observar a região inicial.....	63
Figura 3.2 - Imagem segmentada em n regiões arbitrárias.....	66
Figura 3.3 - Representação gráfica das probabilidades de transição entre as regiões 1, 2 e 3.....	67
Figura 3.4 - Imagem para uma alternativa à região dois – nesse caso a saliência desta região seria maior.....	70
Figura 3.5 - Imagem simples de 12x12 pixels utilizada para ilustrar o modelo (a) e a mesma imagem sobreposta às regiões que delimitam as a área na qual as propriedades que serão extraídas.....	71
Figura 3.6 - Ilustração da matriz computadorizada da figura 3.5 Nesta matriz, os elementos em azul escuro correspondem ao centro de cada região analisada, e em azul claro os elementos pertencentes a esta região.....	71
Figura 4.1 - Exemplo de um <i>frame</i> obtido a partir do vídeo experimental. Tal filme foi resultado da utilização de um eye tracker, equipamento que filma o campo visual de um observador e adiciona uma	

marcação (círculo vermelho) na posição do foco de atenção.....	78
Figura 4.2 - Conjunto das oito imagens utilizadas no experimento.....	79
Figura 4.3 - Segmentação almejada a partir de um <i>frame</i> obtido do filme experimental. A parte amarela é a região que será desprezada por não fazer parte do campo visual simulado, a região em azul compreende o campo visual trabalhado, e a região vermelha a região do foco de atenção do observador.....	82
Figura 4.4 - Um exemplo de uma figura obtida a partir de um <i>frame</i> do filme (a), a correspondente verde (RBG) desta figura (b), e a figura limiarizada a fim de ressaltar os componentes importantes (c). (e) é o resultado da aplicação de transformadas Hough retas e circulares sobre a figura, e (f) é a correspondente imagem com um círculo evidenciando o local onde um observador estaria olhando, com base na figura resposta da transformada Hough.....	83
Figura 4.5 - Seqüência de cinco <i>frames</i> utilizando a amostragem utilizada (a) e a correspondente posição do foco de atenção tomando como base apenas a imagem projetada (b). Em (c) é evidenciada de forma visual a o caminho interpretado pelas imagens.....	85
Figura 4.6 - Redes complexas sobrepostas às imagens. As redes foram geradas a partir de diversos resultados da aplicação do método aglomerativo sobre o caminho gerado pelo foco visual. A distância de aglomeração de (a), (b), (c) e (d) respectivamente foi de 24, 32, 40 e 48 pixels. As imagens (e-l) também obedecem as mesmas distâncias, representam cada uma das outras sete imagens projetadas para o observador.....	88
Figura 6.1 - Saliências computacionais obtidas utilizando diversas combinações de parâmetros. Em (a) evidencia-se a figura	

utilizada para o calculo das saliências, (b) é a saliência computacional calculada utilizando-se como parâmetro os níveis de cinza, (c) esta foi calculada utilizando o gradiente da imagem, (d) a saliência foi baseada no produto entre o gradiente e o nível de cinza, em (e) o calculo foi feito utilizando o desvio padrão de cada região, e (f) a saliência foi dada a partir do produto desvio padrão com o gradiente..... 103

Figura 6.2 - Imagem 5 sobreposta aos círculos vermelhos que evidenciam cada uma das regiões onde serão calculadas as médias da saliência computacional. No centro de cada região foi impresso um círculo cujo nível de cinza é proporcional à média da saliência computacional naquela região..... 105

Figura 6.3 - Gráfico da f.d.p. utilizada para modelar os histogramas utilizados no experimento (a), sendo o eixo x a distancia da sacada, enquanto que no eixo y é graficada a probabilidade de mudança. A função correspondente (b) circular calculada sobre uma região mapeada como um vértice da rede, que será utilizada como máscara para compor a saliência computacional. Em (c) pode-se observar a saliência computacional gerada a partir da Figura 6.1 (c), o módulo do gradiente, e também utilizando tal máscara de distância..... 110

Figura 6.4 - Figura que mapeia a matriz W resultante da Figura 2.4, utilizando uma saliência computacional dada pelo produto dos níveis de cinza com o módulo do gradiente da imagem. Tal matriz foi renormalizada no valor máximo possível para impressão de modo a facilitar a visualização. Pode-se observar um alto valor de auto conexão, assim como a presença de vértices preferenciais..... 111

LISTA DE GRÁFICOS

Gráfico 1 - Histogramas baseados nas distancias das sacadas realizadas pelos olhos do observador no experimento realizado O eixo x representa a distancia da sacada, enquanto que o y a freqüência relativa destas distâncias. Em (a) observa-se as sacadas com base no método aglomerativo de 32 pixels, enquanto que (b) são os resultados a partir da aglomeração de 40 pixels.....	106
--	-----

SUMÁRIO

1	INTRODUÇÃO	19
2	REVISÃO DE LITERATURA	27
2.1	ANATOMIA BÁSICA DO OLHO HUMANO	29
2.2	RETINA E FÓVEA	32
2.3	CÓRTEX	35
2.4	SALIÊNCIA VISUAL	38
2.5	DEFININDO UMA IMAGEM	43
2.5.2	GRADIENTE	44
2.6	CADEIAS DE MARKOV	45
2.7	REDES COMPLEXAS	50
2.7.1	REDES COMPLEXAS: DEFINIÇÕES	52
2.7.2	REDES COMPLEXAS: MEDIDAS ESTRUTURAIS	53
2.7.3	REDES COMPLEXAS: MODELOS DE REDES COMPLEXAS	55
2.7.3.1	REDE REGULAR	55
2.7.3.2	REDE DE ERDÖS-RÉNYI	58
2.7.3.3	REDE DE BARABÁSI-ALBERT	59
3	MODELO: DESCREVENDO UMA IMAGEM EM TERMOS DE UMA REDE COMPLEXA	61
3.1	CONSIDERAÇÕES TEÓRICAS SOBRE O MODELO	6
3.2	EXEMPLO SIMPLES DO MODELO	70
4	EXPERIMENTO	75

4.1	OBTENDO OS DADOS EXPERIMENTAIS	81
4.2	RESULTADOS DO EXPERIMENTO	87
5	AMBIENTE COMPUTACIONAL	93
6	DESENVOLVIMENTO E RESULTADOS: SIMULAÇÕES E O EXPERIMENTO	97
7	CONCLUSÃO	117
	REFERÊNCIAS	125

1 INTRODUÇÃO

Dentro dos sentidos humanos, a visão é a nossa mais forte conexão com o mundo em que vivemos (1). Grande parte da nossa vida é fundamentada em nossa capacidade de ver, e interpretar a informação obtida ao olharmos ao nosso redor. Assim, aterroriza-nos a possibilidade de perdermos a visão repentinamente, a escuridão pode nos provocar medo se estivermos em ambiente desconhecido, como no romance ensaio sobre a cegueira, de Saramago (2).

Tão fundamental para nossa existência, a visão já foi vastamente estudada, porém seu entendimento aprofundado, sob o ponto de vista da captação da informação, só foi possível com o desenvolvimento de técnicas científicas avançadas. Os gregos definiam a visão como uma forma de tato à distância (3), percebendo-se assim objetos distantes. Por mais estranho que essa definição pareça, ela pode ser considerada correta com relação à prática da visão, pois nós a utilizamos para detectar informações remotas como se estivessem próximas, como se estivéssemos tocando no objeto. Mais formalmente, a visão é entendida como a habilidade de detectar a luz e interpretar as consequências deste estímulo.

O sentido visão, porém, é mais complexo do que apenas o aparato para capturar a informação (cristalino, retina, nervo ótico), engloba também a capacidade de interpretar a cena, que biologicamente corresponde ao processamento feito pelo cérebro. De fato, a informação visual transmitida a partir dos olhos é diferente da percepção do mundo que experimentamos com a visão. O primeiro elemento da nossa visão naturalmente são os olhos, órgão especialmente adaptado para captar a luz presente na cena e transformá-la em impulsos elétricos. A luz proveniente de um objeto é projetada pelo cristalino sobre a retina, formando a imagem observada. Esta é captada por células fotossensíveis denominadas cones e bastonetes, células estas responsáveis por converter informação proveniente dos fótons de luz em impulsos elétricos utilizados pelo cérebro para o processamento da informação.

Porém a informação transmitida pelo nervo ótico não é a mesma recebida pelos fotorreceptores (1), pois existe um pré-processamento da informação na retina, e ainda assim a informação transmitida pelo nervo ótico não é o resultado da visão propriamente dita. Na verdade o cérebro recebe uma série de informações que poderíamos classificar como incompletas e as interpreta de forma que

experimentamos a sensação da visão. Um exemplo disso é o que ocorre na região da fóvea.

A fóvea é uma depressão na retina, local onde está presente o maior número de fotorreceptores, de forma a ser a região de maior resolução da visão (8-9). A região mais periférica possui apenas um número limitado destes, fazendo que a imagem seja clara no centro de observação e perca qualidade à medida que nos afastamos. O mesmo ocorre no cérebro, onde existe uma grande região dedicada a interpretar a informação proveniente da fóvea e uma região proporcionalmente menor para interpretar a visão periférica. Esta característica provavelmente teve uma razão evolutiva, pois seria impossível para o cérebro interpretar totalmente toda a informação recebida. Mesmo o nervo ótico seria incapaz de enviar o total de informações caso toda a retina tivesse a mesma resolução da fóvea. O cérebro, então, recebe essas informações peculiares e determina mudanças no ponto de observação, que se dá através dos movimentos sacádicos. Assim os olhos movimentam-se, e seqüencialmente focalizam sobre a fóvea outros pontos da cena. O cérebro por sua vez interpreta todas essas informações parciais e compõe a cena global, fazendo com que tenhamos a sensação de estarmos sempre observando toda a informação, e não apenas parte dela.

Naturalmente, para compor a cena com mais eficiência é necessário que os movimentos sacádicos sejam dirigidos aos pontos mais importantes da cena, de modo que estes pontos recebam maior atenção do cérebro (11). Ainda são discutidos quais são os pontos de uma cena que chamam mais atenção, porém algumas características são consensuais. Uma simples de ser abordada, porém complexa de ser trabalhada por completo é a saliência que é definida pelas propriedades ímpares das regiões da cena. Ao observar uma revoada de pássaros, todos devem chamar a mesma atenção. Entretanto, se um deles mudar de direção e começar a voar na direção contrária aos seus companheiros será uma atitude ímpar com relação aos outros pássaros, provavelmente fará com que um observador comece a observá-lo: aquele pássaro chamará mais atenção.

Dessa forma, definimos a saliência de um objeto com relação a um observador, e essa saliência deverá dirigir cada um dos seus movimentos sacádicos. Outro exemplo é o armário contendo apenas roupas pretas, com apenas

uma peça vermelha. A roupa vermelha naturalmente chamará mais atenção de um observador, ele provavelmente irá observar inicialmente a roupa vermelha, e depois voltará sua atenção às outras roupas. Neste caso, dizemos que a saliência da roupa vermelha é maior que a saliência das outras roupas, assim a atenção recebida pela roupa vermelha é maior do que a recebida pelas outras roupas.

Esta dissertação propõe um modelo computacional estocástico para a saliência das cenas e a atenção seletiva. Para tanto, o modelo propõe que os relacionamentos de atenção sejam interpretados como uma rede complexa, onde os pontos de interesse da cena definem os nós da rede, e as possíveis mudanças sacádicas (mudanças de observação) definem as ligações (formalmente arestas). Uma caminhada aleatória para a rede do modelo corresponderia ao caminho que o olho de um observador iria fazer ao percorrer essa cena. Deste modo, a dinâmica da atenção seletiva seria equivalente à dinâmica na rede complexa criada.

A idéia básica do modelo pode ser ilustrada com a figura 1.1. Para observar os três objetos o observador deverá olhar para cada um deles, fazendo com que cada um deles seja um nó na rede criada. As arestas serão a probabilidade de estar observando um desses objetos e começar a observar outro objeto. Pode-se visualizar a dinâmica gerada sobrepondo a imagem original com a sua rede de atenção, conforme a figura 1.1 (b).

Figura 1.1 - Ilustração básica do modelo. Em (a) é apresentada uma imagem simples contendo três objetos. Em (b), é a reação básica esperada de um observador, segmentar cada um dos objetos, e voltar a sua atenção singularmente para cada um deles.

Para tal estudo tomamos um conjunto de imagens, onde cada uma delas representa uma cena a ser observada. As imagens foram escolhidas de forma a simplificar seu entendimento pelo observador, fazendo com que seja necessária menor potência computacional e algorítmica para a aplicação desse modelo. Foram utilizadas imagens em níveis de cinza, assim como imagens de texturas cotidianas, como pedras, gramados e madeira. Essas imagens não possuem nenhuma interpretação ou complexidade particular (complexidade em sua interpretação porque nosso interesse está apenas em estudar o caminho observado pelos olhos), de modo que nosso interesse meramente por simplicidade está em estudar o caminho observado devido apenas à imagem e a textura, não devido a uma interpretação mais complexa da cena, tal como seu conteúdo semântico.

Destas imagens retiramos uma série de propriedades locais para o estudo da atenção seletiva. Essas foram às características fundamentais para definir a dinâmica simulada, definindo os pontos principais da imagem, numa rede complexa, os seus vértices, assim como as arestas da rede (pontos mais salientes devem ter

maior probabilidade de serem visitado) devem ditar a forma com que elas se conectam.

Já as redes complexas têm como sua principal característica a sua grande adaptabilidade a uma grande variedade de situações. Essa vantagem advém de um conceito inicialmente simples, porém robusto. Uma rede complexa consiste num conjunto de elementos com características comuns (denominados nós ou vértices) e uma regra de interação entre esses elementos (arestas ou ligações). Praticamente qualquer coisa poderia ser definida como um nó assim como qualquer tipo de interação poderia ser uma aresta, fazendo com que existam aplicações nas mais diversas áreas da ciência (4).

Considera-se que o matemático Leonard Euler foi quem primeiramente utilizou a teoria das redes complexas (nesse caso mais especificamente teoria dos grafos) ao resolver o problema das pontes de Königsberg utilizando o conceito primitivo de nós e arestas provando que não era possível viajar pela cidade atravessando cada uma das sete pontes existentes uma única vez (5). Porém muitos outros trabalhos evidenciaram a versatilidade das redes complexas, como (6).

O estudo das redes complexas baseadas nas imagens delimita o modelo utilizado. Porém tornou-se necessária a comparação dos resultados obtidos pelo modelo com os dados de algum experimento, norteando as escolhas dos parâmetros utilizados respectivamente a dados obtidos de um observador, e não de suposições já conhecidas.

Utilizando uma colaboração com o laboratório PAMELA (*Pedestrian Accessibility and Movement Monitoring Environment Laboratory*), de Londres, representado pelo professor Sérgio Velastin, foi possível utilizando um *eye tracker* obter dados experimentais a respeito da posição do foco de atenção de um observador. Estes dados consistem num filme com uma marcação específica – um filme com as imagens utilizadas e uma marca móvel, esta representando o ponto onde o observador estava olhando. As coordenadas desta marca em função do tempo representam o caminho que o olho daquele observador fez ao observar cada uma da imagem.

A partir do estudo desse caminho, da localização de cada uma das marcas, do seu agrupamento e posterior estatística, construímos uma nova rede complexa, totalmente baseada no experimento. A partir da comparação das propriedades desta nova rede com redes obtidas puramente a partir do modelo foi possível determinar o quão distante os parâmetros do modelo estavam dos parâmetros utilizados por um observador real, de modo que modificamos os parâmetros de entrada do modelo para que a distância entre as duas redes fosse mínima.

O capítulo 2 fornecerá as informações básicas para o entendimento do modelo, os mecanismos fisiológicos básicos utilizados como argumentos ao se construir a rede complexa utilizada. Também foram abordadas de maneira sucinta propriedades simples de imagens transcritas para um computador, assim como uma abordagem básica dos temas estatísticos utilizados. O capítulo 3 fornece uma visão teórica básica sobre o modelo, definindo suas características, limitações e potencialidades. O capítulo 4 descreve o experimento utilizado para confrontar as redes teóricas utilizadas, assim como quaisquer peculiaridades a respeito da análise dos seus resultados, em especial como interpretá-los como uma rede complexa. O capítulo 5 fornecerá detalhes sobre o ambiente computacional utilizado, enquanto que o capítulo 6 finalmente trata da produção de redes complexas teóricas baseadas no modelo, assim como seu confronto com as redes experimentais. As conclusões e perspectivas para trabalhos futuros encontram-se no capítulo 7.

2 REVISÃO DE LITERATURA

2.1 ANATOMIA BÁSICA DO OLHO HUMANO

O olho tem como principal função captar a luz proveniente das cenas e transformar essa informação luminosa em impulsos nervosos. Para tal, ele é composto de uma série de componentes óticos que viabilizam esta detecção.

O esquema visual do olho pode ser comparado ao de uma câmera do tipo *pin-hole* (7), como nas figuras abaixo:

Figura 2.1 - Comparação entre o esquema ótico de uma câmera *pin-hole* (a), uma câmera fotográfica com lente (b) e o olho humano (c).

Na figura 2.1 (a) é apresentado o esquema básico de uma câmera do tipo *pin-hole*. A luz proveniente de um objeto entra na câmera por um pequeno furo, fazendo com que apenas um raio de luz de cada ponto do objeto atinja exatamente um ponto do anteparo na câmera, fazendo assim uma imagem do objeto no interior da câmera. Porém esse tipo de instrumento tem uma grande limitação: para existir uma boa focalização o furo tem de ser pequeno o suficiente para que apenas um feixe de luz de cada ponto do objeto atinja um ponto do anteparo. Mas um pequeno furo causa aberrações provenientes da difração da luz, assim como também faz com

que apenas uma pequena quantidade de energia luminosa penetre na câmara, fazendo assim com que a imagem formada seja de pouca intensidade.

Essa limitação foi resolvida em câmeras fotográficas modernas ao introduzir as lentes. Estas exercem o mesmo papel que o furo numa câmara *pin-hole*, focaliza os feixes provenientes do objeto no seu interior, porém permitem que uma quantidade maior de luz entre na câmara, eliminando assim as desvantagens do furo na câmara *pin-hole*, conforme a figura 2.1 (b). É assim que o olho humano funciona (8), uma lente (cristalino) capta as imagens das cenas e as projeta sobre um anteparo (retina) onde essa imagem é convertida em impulsos elétricos, e esses enviados ao cérebro. Essa simplificação é denominada como Olho reduzido (8).

Figura 2.2 - Perfil anatômico do olho humano. Fonte: Rhcastilhos. Disponível em: http://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_pt.svg Acesso em: 20 jun 2010

Um esquema sobre os componentes anatômicos do olho é a figura 2.2, e uma descrição mais detalhada dos componentes básicos do olho será apresentada a seguir:

O cristalino faz o papel de lente para os olhos. É uma estrutura biconvexa, que ajuda a focalizar a luz proveniente do ambiente na retina. Sua principal função no sistema visual é a de uma lente de distância focal variável, de modo que objetos a distâncias variáveis possam ser focalizados corretamente sobre a retina. Para tal, o cristalino é formado por fibras transparentes viscosas de natureza protéica. Quando este está relaxado, ele assume uma forma quase esférica. No entanto, este pode ser tracionado radialmente por um conjunto de ligamentos, achatando o cristalino num elipsóide, modificando consideravelmente o foco do sistema ótico.

A pupila tem como função controlar a profundidade de campo ao se contrair ou relaxar. Nas máquinas fotográficas tal função é exercida por um diafragma. A íris é o disco colorido do olho, e seu orifício central é que recebe o nome de pupila (9).

Retina é a túnica nervosa do olho, local onde se dá a percepção das imagens, segundo Castro (9). Com base na analogia com uma máquina fotográfica, a retina seria o filme fotográfico, o local que serve de anteparo para as imagens projetadas. Para captar a imagem a retina conta com duas células especializadas em transformar informações visuais em impulsos nervosos, os cones e os bastonetes. Duas estruturas notáveis presentes na retina são o disco ótico, local de saída do nervo ótico para o cérebro, e a fóvea ou mácula, uma depressão na retina caracterizada pelo aumento do número de células fotossensíveis e por consequência da resolução da visão. A fóvea é um tema fundamental para o entendimento desta dissertação, sendo tratada com mais detalhes nas seções posteriores.

É importante afirmar que aqui são apenas descritos os componentes básicos da visão, necessários para a compreensão da atual dissertação. Uma descrição mais detalhada do sistema visual pode ser encontrada em (8-9).

2.2 RETINA E FÓVEA

Depois de atravessar o sistema ótico do olho, a luz é projetada pelo cristalino sobre a retina. A retina cobre praticamente todo o interior do olho, com exceção da região da íris, responsável pela entrada da luz, porém, a curta distância desta, pára de perceber imagens (9).

A retina é formada por um conjunto de células divididas em camadas, esquematizadas na figura 2.3 a seguir:

Figura 2.3 - Estrutura das células presentes na retina. À direita estão presentes um cone e oito bastonetes, também estão presentes células bipolares e ganglionares. Fonte: CAJAL, Histologie Du Système Nerveux de l'Homme et Des Vertébrés. Adaptado. Disponível em: <http://pt.wikipedia.org/wiki/Ficheiro:Fig_retine.png> Acesso em: 14 jun 2010.

Depois que a luz atravessa o sistema do cristalino do olho e o humor vítreo (região interna do olho), ela atinge a retina, conforme a figura 2.3, passando primeiro pelas células ganglionares, depois pelas camadas plexiformes, camada nuclear e membranas limitantes antes de, finalmente, atingir os cones e bastonetes, células estas que conseguem perceber a luz. Toda a estrutura anterior aos cones e bastonetes foge do escopo desta dissertação, aqui basta saber que essas estruturas são responsáveis por um pré-processamento da imagem percebida (1), além de responsáveis pela transmissão da informação recebida ao cérebro - os axônios das células ganglionares formam o nervo ótico, conectando o olho ao cérebro (8).

Os cones e bastonetes são células especialmente modificadas para serem sensíveis à presença de radiação eletromagnética de comprimento de onda restrito, a luz visível. Para tal, contam com um sistema bioquímico complexo, onde substâncias químicas se decompõem quando expostas à luz e, nesse processo,

excitam fibras nervosas. Os bastonetes, mais sensíveis à radiação, são responsáveis pela percepção da intensidade da luz das imagens, enquanto que os cones fazem com que o olho humano perceba as cores, utilizando um sistema de três tipos diferentes de percepção. Os cones e os bastonetes são distribuídos sobre a retina, porém existe uma região denominada fóvea onde a densidade de bastonetes diminui ao aproximar-se dela, enquanto que a densidade de cones aumenta gradualmente, até que na fóvea central praticamente não existem bastonetes e a densidade de cones é extremamente alta.

A mácula representa o centro do fundo do bulbo ocular, formando uma região amarelada com uma pequena depressão denominada fóvea central (9). A fóvea pode ser considerada ponto central da retina, onde é encontrado a maior acuidade visual, conforme a figura 2.4. Isso se deve a duas características, primeiro a densidade de cones ali presentes, e segundo, que no restante da retina, antes da luz chegar às células fotorreceptivas ela tem de ultrapassar todas as outras camadas da retina, diminuindo assim a acuidade visual. Na região da fóvea, essas camadas se deslocam lateralmente, expondo assim os cones e bastonetes a luz direta. Esse deslocamento faz com que a fóvea seja uma depressão na retina.

Figura 2.4 - Esboço da acuidade visual em função da posição ocular. A região da fóvea possui a maior acuidade visual, e esta cai ao se distanciar da fóvea. A exceção é o ponto cego, local desprovido de fotorreceptores. Fonte: Desenho a mão baseado em arquivo por Hans-Werner Hunziker. Adaptado. Disponível em: <http://en.wikipedia.org/wiki/File:AcuityHumanEye.svg> Acesso em: 14 jun 2010.

Essa configuração peculiar na acuidade visual também tem reflexos na interpretação do estímulo pelo cérebro. Uma grande parte do córtex se empenha em interpretar as informações provenientes da fóvea, enquanto que uma porção menor é gasta na visão periférica. O que ocorre na interpretação de uma cena é que existe uma movimentação dos olhos, movimentando a pequena janela correspondente a fóvea sobre os pontos mais importantes da cena. De fato, a visão da fóvea central nos qualifica a ver detalhes, enquanto que a visão periférica é mais sensível a mudanças

Quando observando uma nova cena, os olhos se fixam em um ponto importante após o outro no campo visual, saltando de um para o outro duas a três vezes por segundo. Esses saltos extremamente rápidos são chamados sacadas. Estas ocorrem tão rapidamente que menos de um décimo do tempo total de observação de uma cena é gasto fazendo a movimentação dos olhos. Este é um movimento balístico, onde apenas são considerados os pontos iniciais e finais, assim como o cérebro suprime a imagem visual durante a sacada, de modo que a

pessoa não tem consciência dos movimentos de um ponto a outro. Para dirigir a sacada o sistema visual conta com músculos especiais que movimentam o olho, assim como critérios de decisão, para que as cenas mais importantes sejam focalizadas. Tal mecânica é denominada atenção seletiva, um observador não observa de maneira integral uma cena, porém de maneira seletiva, prioritariamente às regiões de maior importância. Mais detalhes sobre isso serão abordados na seção 2.4.

2.3 CÓRTEX

A presente seção foi baseada nas notas de aula do autor desta dissertação sobre a disciplina ministrada pelo professor Luciano da Fontoura Costa. Mais informações poderão ser encontradas em (1, 8-10).

A luz proveniente da informação das cenas é captada pelos olhos, focalizada pelo cristalino sobre a retina, sendo convertida em um sinal nervoso por células fotossensíveis. A informação é então enviada ao cérebro pelo nervo ótico, que é formado pelos axônios das células ganglionares presentes na retina. Porém, a informação enviada ao cérebro não é a mesma captada pela retina.

Inicialmente, é feito na retina um pré-processamento a fim de eliminar redundâncias, enviando assim apenas informação considerada útil. Uma estrutura notável entre os olhos e o córtex é o quiasma ótico, um cruzamento entre os nervos óticos de cada olho. Sua importância é que as informações provenientes da parte direita de uma cena são captadas pelos dois olhos, porém são processadas na mesma região do cérebro. Desta forma, as informações provenientes do lado esquerdo do olho esquerdo vão para o lado esquerdo do cérebro, assim como as informações provenientes do lado esquerdo do olho direito também vão para a mesma região do cérebro.

A sinapse do nervo ótico acontece no núcleo geniculado lateral, localizado na parte posterior do cérebro, fazendo a informação atingir o córtex visual primário. Ali a informação começa a ser processada, e espalhada resto do cérebro, fazendo assim

o sentido da visão – a percepção do espaço baseada na interpretação da luz. Aproximadamente 50% do córtex humano é dedicado a algum tipo de processamento da visão, porém ainda não é claro qual é a função de cada parte do cérebro. Toda a estrutura do cérebro ainda é alvo de pesquisas e discussões, sendo que aqui serão apenas citados resultados simples e bem entendidos, necessários para a compreensão básica do modelo que será apresentado.

De forma geral, o processamento do cérebro é feito de modo pouco específico, cada neurônio atua processando diversos tipos de informações. Porém existem regiões de processamento restrito, como fala, memória, e visão. Existe apenas um tipo de neurônio especializado, a célula-avó (*grandmother cell*) cuja função é o reconhecimento de faces e mãos.

A organização é feita de modo bidimensional, “embrulhando” o córtex sobre a região denominada massa branca. O córtex é composto por seis camadas de neurônios, onde cada um tem da ordem de mil conexões. O isolamento elétrico é feito pela massa branca, uma região gordurosa abaixo do córtex.

Neste ponto, torna-se importante definir o conceito de campo receptivo: “É a região do espaço visual capaz de modificar a atividade do neurônio em questão”. O campo receptivo é a resposta que um neurônio apresenta de acordo com um determinado estímulo. Note que existe um campo receptivo específico para cada neurônio, porém ele não será o único responsável por aquela informação, pois este é conectado a uma série de outros neurônios. O campo receptivo caracteriza funcionalmente um neurônio.

Experimentalmente, um campo receptivo pode ser determinado através de métodos eletrofisiológicos. Um neurônio pode ser entendido como uma pequena bateria. Ao estimulá-lo de alguma forma pode ocorrer um disparo fisiológico mais constantemente, ou pode-se inibir a taxa de disparo. Esse tipo de medida pode ser feita fixando-se a cabeça de um animal e exibindo um estímulo visual, como por exemplo, a imagem de uma reta. Com um eletrodo é feita a medição da resposta a essa imagem de um conjunto de neurônios.

No caso das células ganglionares da retina, observa-se o campo receptivo de acordo com a figura 2.5 (a).

Figura 2.5 - Esquema de ativação das células ganglionares presentes na retina. Em (a) é evidenciada a ativação padrão, (b) evidencia uma taxa de ativação positiva, enquanto que (c) uma taxa de ativação negativa.

A região com sinais positivos indicam aumento na faixa de disparo, enquanto a região com os sinais negativos representam diminuição. Com um estímulo como representado pelos retângulos tracejados, é possível inferir o tipo de resposta desse tipo de célula. Na figura 2.5 (a), é observada a taxa de disparo padrão. No caso da figura 2.5 (b), toda a parte positiva é encoberta, resultando no aumento na taxa de disparo. O mesmo ocorre na figura 2.5 (c), mas de modo inverso: apenas uma região negativa é encoberta, resultando na diminuição da taxa de disparo. Assim, este campo receptivo atua como um detector de bordas.

No caso do córtex visual primário o campo receptivo é semelhante ao do campo da retina, porém alongado em uma determinada direção, fazendo com que este seja um detector de segmentos de reta. Note que neste caso cada neurônio é sensível tanto quanto à posição tanto quanto à orientação. Esse campo receptivo é denominado campo receptivo simples. É interessante salientar que não existe um campo receptivo sensível à curvatura, apenas para detectar a falta dela, as retas.

O campo receptivo complexo é semelhante ao simples, porém este deixa de ser específico com relação à posição, sendo sensível apenas à orientação. O campo receptivo hipercomplexo tem suas características discutidas até hoje, sendo que ele terá as características de um campo receptivo complexo, porém o comprimento relativo do estímulo torna-se importante. Uma característica desse sistema é a aproximação da linearidade, estímulos combinados têm como resultado aproximado o resultado singular de cada estímulo, combinados ao final. Matematicamente,

dados duas estímulos quaisquer, representados matematicamente por a e b , submetidas a um campo receptivo, representado por S , assim $S(a+b) = S(a) + S(b)$.

A informação visual chega ao córtex na forma de um mapa conforme, pontos adjacentes são processados próximos na retina também são processados próximos no córtex. Uma vez que cada região do córtex recebe informações dos dois olhos, essas informações são processadas em regiões próximas no córtex, formando estruturas denominadas colunas de dominância ocular. Os campos receptivos de cada coluna comportam-se da seguinte forma: A medição ao longo de uma linha tangente ao córtex resulta em campos receptivos semelhantes. Porém existem pontos de descontinuidade, denominados fraturas. Isso ocorre devido à topologia na região, onde os campos receptivos são distribuídos na forma de estruturas circulares denominadas cata-ventos.

2.4 SALIÊNCIA VISUAL

A presente seção foi baseada nos trabalhos realizados pela psicóloga Anne Treisman (11).

A saliência visual de um objeto é definida como a capacidade que este objeto tem de chamar a atenção numa determinada cena. Um bom exemplo disso é o que ocorre na figura 2.6.

Figura 2.6 - Exemplo simples de saliência visual. A roupa vermelha naturalmente chama mais atenção de um observador, fazendo com que este dê preferência da sua atenção àquela roupa. Neste caso é definido que a saliência da roupa vermelha é maior do que a saliência das outras roupas.

Neste caso, dentro de um conjunto de roupas escuras, uma roupa vermelha naturalmente chama mais atenção, ela se destaca, fazendo com que um observador olhe primeiro para ela, e só depois para as outras roupas presentes. Nesse caso, a saliência visual da roupa vermelha é maior que a saliência visual dos outros objetos em questão.

Esse efeito de saliência parece ser algo automático para o cérebro, não é necessário nenhum esforço para o observador para perceber o que é ou não saliente. Existem indícios de que antes do reconhecimento de objetos dentro de imagens, o cérebro faz o reconhecimento de componentes básicos, como retas, superfície e volumes, e só então reconhece o objeto. Este se relaciona com os campos receptivos do cérebro, evidenciando como a informação é transmitida a partir dos olhos até atingir todas as partes do córtex. De fato, sabe-se que o cérebro tem regiões de dominância, uma região reconhece cor, outra bordas, outra orientação, e só depois da informação passar por essas áreas é que ela irá atingir as regiões que de fato reconhecerão o objeto.

Porém todo esse processamento é feito independentemente da vontade do observador. Um exemplo utilizado por Treisman seria o caso uma pessoa fosse magicamente transportada para uma cidade desconhecida. Esse observador enxergaria construções, pessoas e carros, mas não daria importância para cores, bordas, movimentos e distâncias, todas essas características seriam automaticamente integradas de modo ao observador dar importância apenas ao que parece ser interessante – o reconhecimento básico da nova cidade em questão.

Ao tentar procurar quais características são responsáveis por essa saliência, algumas conclusões são bem fundamentadas e aceitas. Uma estratégia sugerida é que um objeto do mundo real será constituído de partes que tentem a ter características comuns, como a mesma cor e textura, sua borda tente a ser contínua de linhas e superfícies, eles movem-se juntos. Um exemplo desse tipo de característica pode ser visualizado na figura 2.7.

Figura 2.7 - Exemplo de experimento psicofísico para detecção de saliência. A região composta por caracteres “T”s rotacionados se destaca mais da região composta por caracteres “T” do que a região composta por caracteres “L”s. Fonte: Treisman, 1986. Adaptado

Na pode-se observar três regiões distintas, porém é notável que a região formada por caracteres “T” se segrega muito bem da região formada pelos caracteres “T” rotacionados, mas não é possível observar o mesmo fenômeno entre a região formada pelos caracteres “T” e os caracteres “L”. Observe que os caracteres utilizados têm basicamente as mesmas características, figuras geométricas retas, de tamanho, perímetro e curvatura semelhantes. Porém a região dos “T”s rotacionados tem orientação radicalmente distinta das outras duas regiões.

O fato de essa fronteira ter uma saliência muito maior poderia ser explicado pelo fato do córtex possuir campos receptivos específicos para orientação, enquanto que a resposta dos neurônios para as outras características dos caracteres seriam menos marcantes. De fato, acredita-se que as propriedades que fazem objetos salientes são as mesmas processadas inicialmente pelo cérebro, de forma que objetos com características congruentes com campos receptivos tendem a ser salientes.

Outro modo de encontrar saliências em objetos será a análise de experimentos de busca, onde observadores humanos têm de procurar objetos específicos dentro de grupos com vários outros objetos, chamados distratores. Experimento esse semelhante ao provérbio “procurar agulha no palheiro”, supõe que, caso a agulha tivesse propriedades distintas às da palha, como uma cor distonante, ou um tamanho ou forma distintos, tal tarefa seria significativamente mais fácil. Quando tais experimentos eram realizados perguntando para os observadores encontrarem objetos com características diferentes dos distratores, como orientação, cor ou curvatura, os objetos eram encontrados em tempo praticamente constante em conjuntos tanto com três, tanto com trinta objetos. Essa afirmação continuava praticamente verdadeira quando a propriedade saliente do objeto não era informada ao observador, este levava um pouco mais de tempo para encontrar o objeto, porém a quantidade de distratores no grupo não afetava significativamente os resultados.

Também foi investigada a saliência de objetos como alvo de uma busca e como agente distrator. Neste caso, observadores observavam conjuntos de letras “Q”s, onde havia uma letra “O” como alvo da busca, tal experimento pode ser visto na figura 2.8. O inverso também foi investigado, observadores procurando por uma letra “Q” num conjunto de distratores “O”s. Note que a diferença entre as duas letras está apenas na presença ou não do traço na parte inferior do caractere. Verificou-se como resultado que a letra “Q” num conjunto de letras “O”s era mais saliente do que o inverso. Esse resultado evidencia que a saliência não é dependente apenas das propriedades relativas entre os pontos da imagem, mas também do objeto em si que esta sendo procurado. Sugere-se que esse tipo de resultado é devido à presença de campos receptivos pertinentes para detectar a presença positiva de elementos (isso é, o campo responde caso exista a presença do elemento), porém não existe nenhum tipo de campo que indica a presença negativa (caso o objeto não esteja lá).

Resultados semelhantes são obtidos ao realizar buscas sobre objetos com tamanhos diferentes, caso o alvo da busca seja maior que os objetos distratores a busca é mais rápida do que caso o alvo seja menor do que os outros objetos. Aparentemente o cérebro parece prestar mais atenção à presença de objetos do que quanto a sua falta.

Figura 2.8 - Conjuntos apresentados para observadores para procura. Em (a) existem 3 caracteres “O” e um “Q”, enquanto que (b) possui um “O” e 3 “Q”s. Em (c) encontram-se 30 caracteres “O” e um “Q”, enquanto que em (d) existem 30 “Q”s e um “O”. Comparando-se (a) e (c), evidencia-se que a eficiência para determinadas procuras independe do número de distratores. Comparando (c) e (d) pode-se notar que é mais fácil encontrar um “Q” dentro um conjunto de “O”s do que um “O” dentro de um conjunto de “Q”s. Fonte: Treisman, 1986. Adaptado

Outros experimentos desse tipo revelaram que as propriedades que tornavam alvos salientes em relação aos seus agentes distratores foram cor, tamanho, contraste, rotação, curvatura e finais de retas.

2.5 DEFININDO UMA IMAGEM

Como usado nesta dissertação, uma imagem (12) refere-se a uma função bidimensional de intensidade de luz $f(x,y)$, onde x e y denotam a coordenada espacial, e o valor de f num (x,y) ponto é proporcional a intensidade luminosa naquele ponto. A função $f(x,y)$ em geral é definida como o nível de cinza no ponto (x,y) .

Uma imagem digital é uma imagem onde foram digitalizados tanto os valores dos níveis de cinza, tanto os valores válidos das coordenadas (x,y) . Aqui as imagens são representadas como matrizes, onde a coordenada correspondente é indexada pelas linhas e colunas desta matriz, e o valor do nível de cinza é dado pelo valor correspondente na matriz. O elemento básico desta matriz é chamado pixel. Por simplicidade, essa matriz será chamada na presente dissertação de imagem.

Uma imagem simples, dita imagem em níveis de cinza, é a imagem apresentada até então. Uma imagem colorida é uma imagem representada por três matrizes de níveis de cinza, onde cada uma delas tem o valor da componente RGB daquele ponto. Uma imagem binária, ou imagem em preto e branco, é uma imagem que consiste apenas em dois tipos de pixels, brancos e pretos. Na presente dissertação, uma imagem sempre será em níveis cinza, a não ser quando especificado o contrário.

O valor do nível de cinza usualmente pode variar entre 0 e 255, ou entre 0 e 1, dependendo da definição inicial. A quantidade de divisões disponíveis varia de acordo com a *palette* utilizada, em geral variando entre 256 e alguns milhões de tonalidades. A matriz imagem poderá ter qualquer tamanho escolhido, este tamanho definirá a resolução da imagem, quanto maior o número de elementos dessa matriz,

maior será a qualidade gráfica da imagem. Observe que pelo fato da matriz ser representada graficamente como a linha 1 na parte de cima, e crescendo de cima para baixo, fazendo com que a definição geométrica de $f(x,y)$ torne-se invertida.

2.5.1 GRADIENTE

O gradiente, ou contraste nada mais é do que a variação dos níveis de cinza na região. Existem diversos algoritmos para fazer este cálculo, e utilizaremos neste trabalho o operador de Sobel (13).

Este operador representa uma aproximação para a gradiente matemático daquele ponto, com alguma suavização. Utilizaremos o operador de Sobel neste trabalho devido a sua menor complexibilidade algorítmica. A aproximação do gradiente em uma determinada direção nada mais é do que a convolução da imagem por uma máscara especial. Estas máscaras, nas direções coordenadas de uma imagem serão:

$$\Delta x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad \Delta y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad (2.1)$$

De uma maneira operacional, o operador de Sobel terá como resultado parcial a convolução da imagem com matrizes específicas relacionadas às variações ao longo de x e de y:

$$g_x = \Delta x * A_{im} \quad g_y = \Delta y * A_{im} \quad (2.2)$$

Onde g_x e g_y denotam a variação em cada coordenada, $\otimes$ denota convolução e A_{im} é a função a qual se deseja o gradiente, neste caso, uma imagem. O módulo do gradiente será:

$$|\vec{\nabla} A_{im}| \cong \sqrt{g_x^2 + g_y^2} \quad (2.3)$$

A direção do gradiente:

$$\alpha \cong \text{tg}^{-1} \left(\frac{g_x}{g_y} \right) \quad (2.4)$$

A figura 2.9 mostra a ilustração de uma imagem simples (b) convertida numa matriz (a), o respectivo módulo gradiente aproximado do centro da imagem (c), e a correspondente imagem do gradiente (d). Um exemplo aplicado numa figura extensa pode ser visto em figura 2.9 (e), e seu respectivo módulo do gradiente figura 2.9 (f).

Figura 2.9 - Exemplo de matriz tratada (a), sua representação como imagem (b), a aproximação do gradiente das posições centrais (c), a imagem deste gradiente (d). Em (e) é possível ver o gradiente obtido com o operador de Sobel da imagem (f). Fonte: Simpsons, 2008. <http://en.wikipedia.org/wiki/File:Valve_original_%281%29.PNG> Acesso: 20 jun 2010.

2.6 CADEIAS DE MARKOV

Cadeias de Markov estruturas estatísticas formadas por processos discretos aleatórios, com a propriedade de que o próximo passo dependerá unicamente do resultado do passo anterior.

Tais cadeias têm aplicações em diversas áreas da ciência, como química, física, economia e ciências sociais. Deste modo, existem diversas fontes matematicamente abrangentes, como (14-16). Aqui foram abordados conceitos básicos dessa extensa, necessárias apenas para o entendimento da presente dissertação.

Uma cadeia de Markov trata um sistema composto de estados discretos, e estes estados evoluem temporalmente. Cada estado deverá ter probabilidades de transição para todos os outros, desta forma, a probabilidade para ir de um estado para outro depende unicamente do estado atual deste sistema. Dado um estado i do sistema, existirá um conjunto de probabilidades P_i que governará a probabilidade do sistema evoluir do estado i para qualquer outro estado do sistema. Sendo o sistema discreto, contendo n diferentes estados:

$$P_i = \{p_{i,1}, p_{i,2}, p_{i,2}, p_{i,2}, \dots, p_{i,n}\} \quad (2.5)$$

Onde:

$p_{i,k}$: É a probabilidade de mudança a partir do estado i para o estado k .

Note que necessariamente o sistema deverá evoluir para um estado válido para o sistema, de forma que:

$$\sum_{k=1}^n p_{i,k} = 1 \quad (2.6)$$

Tal conjunto poderia ser organizado na forma de uma matriz, de forma a contabilizar quais serão as probabilidades de mudança de sejam compiladas nas colunas da matriz. Deste modo, a coluna i desta matriz conterá em duas linhas os componentes do conjunto P_i . Esta matriz é denominada matriz de transição.

$$M = \begin{pmatrix} p_{1,1} & p_{2,1} & \cdots & p_{n,1} \\ p_{1,2} & p_{2,2} & & \\ \vdots & & \ddots & \\ p_{1,n} & & & p_{n,n} \end{pmatrix} \quad (2.7)$$

Existe na literatura uma discordância quanto à definição ao organizar o conjunto P_i na forma de linhas de uma matriz, ou na forma das colunas. Aqui utilizamos como as colunas da matriz, forma esta que é denominada matriz esquerda.

Tome como exemplo de uma matriz de transição o sistema proposto pela figura 2.10. Neste exemplo, as probabilidades de transição são denotadas por um grafo.

Figura 2.10 - Exemplo de um sistema representado por um grafo. Os estados são representados pelas seguintes figuras geométricas: círculo, quadrado e triângulo. As probabilidades de transição são evidenciadas pelas setas.

Tal exemplo pode ser denotado como uma simplificação do modelo proposto por esta dissertação. Imagine que um observador está olhando para o círculo. Ele terá uma probabilidade zero de continuar olhando para o mesmo, uma probabilidade 0,7 de começar a olhar para o triângulo, e uma probabilidade 0,3 de olhar para o quadrado.

A matriz de transição transcrita a partir da figura 2.10:

$$M = \begin{pmatrix} 0,0 & 0,5 & 0,8 \\ 0,3 & 0,0 & 0,1 \\ 0,7 & 0,5 & 0,1 \end{pmatrix} \quad (2.8)$$

Indexando o círculo como objeto (1), o quadrado como (2), e o triângulo como (3) pode-se visualizar a matriz M de uma maneira mais didática de acordo com a figura 2.11:

$$\begin{array}{c} \begin{array}{ccc} & (1) & (2) & (3) \\ (1) & 0,0 & 0,5 & 0,8 \\ (2) & 0,3 & 0,0 & 0,1 \\ (3) & 0,7 & 0,5 & 0,1 \end{array} \end{array}$$

Figura 2.11 - Matriz M escrita de uma forma didática, evidenciando os índices de cada estado possível.

Observando apenas a primeira coluna, temos o índice (1), que representa o objeto (1). Esta coluna seria utilizada para o caso de um observador estar olhando para o círculo, e relaciona a probabilidade deste observador continuar olhando esse mesmo objeto, ou começar a olhar algum outro. No caso, temos uma probabilidade zero dele manter o foco no mesmo lugar, uma probabilidade 0,3 do foco ir para o objeto (2), e 0,7 de observar o objeto (3). O raciocínio é completamente análogo para as outras colunas, sendo a segunda relacionada com o objeto (2), e a terceira coluna com o objeto (3).

As matrizes de transição têm diversas propriedades, e aqui será relacionada apenas uma delas que será utilizada posteriormente neste trabalho.

Imagine um conjunto de mil observadores olhando para os objetos da figura 2.10, distribuindo sua atenção da seguinte forma: 400 deles olhando para o objeto (1), 300 para o (2), e 300 para (3). Depois de um passo, pode-se perguntar qual será o número de observadores, em média, que estará observando cada objeto. O método usual de responder tal exercício seria utilizando as probabilidades de transição de cada estado, descobrir qual seria a média de transições para cada estado, e fazer sua contabilidade final. Tal método, baseado na estrutura de M poderia ser descrito como uma multiplicação matricial. Escrevendo um vetor $\vec{S}$ contendo o número de observadores sobre cada estado, o vetor resultado $\vec{R}$ poderia ser obtido a partir da seguinte multiplicação matricial:

$$\vec{S} = \begin{pmatrix} 400 \\ 300 \\ 300 \end{pmatrix} \rightarrow M\vec{S} = \vec{R} \rightarrow \begin{pmatrix} 0,0 & 0,5 & 0,8 \\ 0,3 & 0,0 & 0,1 \\ 0,7 & 0,5 & 0,1 \end{pmatrix} \begin{pmatrix} 400 \\ 300 \\ 300 \end{pmatrix} = \begin{pmatrix} 390 \\ 150 \\ 460 \end{pmatrix} \rightarrow \vec{R} = \begin{pmatrix} 390 \\ 150 \\ 460 \end{pmatrix} \quad (2.9)$$

Tal método pode ser usado indefinidamente, determinando assim a quantidade de observadores depois de L passos temporais:

$$\vec{R}_L = M^L \vec{S} \quad (2.10)$$

A partir de tal conceito torna-se interessante saber qual será o ponto estacionário do sistema. Tomando o mesmo exemplo, à medida que um objeto tem muitos observadores, a média de observadores que deixam aquele objeto, mesmo que este tenha uma probabilidade pequena de fazê-lo, tende a crescer com o número de observadores. Dessa forma, existe um L no qual o número de observadores que deixam de olhar um objeto se iguala ao número de observadores que saem deste objeto, fazendo com que o vetor $\vec{R}_L$ deixe de ter o seu valor modificado. $\vec{R}_L$ estacionário será denotado como $\vec{R}_L^{final}$. Este poderia ter seu valor calculado a partir da relação anterior, fazendo com que L seja suficientemente grande.

Algebricamente, procura-se um vetor $\vec{R}_L^{final}$ que obedeça a seguinte propriedade:

$$W\vec{R}_L^{final} = \vec{R}_L^{final} \quad (2.11)$$

A partir de uma equação de autovalores:

$$A\vec{X}_k = \lambda_k \vec{X}_k \rightarrow \lambda_k = 1 \rightarrow A\vec{X}_k^{\lambda=1} = \vec{X}_k^{\lambda=1} \quad (2.12)$$

Onde:

A : é uma matriz quadrada, neste caso, uma matriz de transição.

$\vec{X}_k$: é o k -ésimo autovetor associado à matriz A .

$\lambda_k = 1$: é o k -ésimo autovalor associado à matriz A .

$\vec{X}_k^{\lambda=1}$: é o auto-vetor associado à matriz A cujo autovalor é 1.

Deste modo, para determinar $\vec{R}_L^{final}$ basta procurar pelo auto-vetor da matriz M cujo autovalor associado seja 1. Este auto-vetor será o $\vec{R}_L^{final}$ procurado.

2.7 REDES COMPLEXAS

O experimento de Milgram (17) é provavelmente o exemplo mais conhecido de uma rede de relações pessoais. Esse experimento deu início à expressão “Seis graus de separação” (*Six degrees of separation*), que afirma que cada indivíduo pertencente a um grande grupo (como por exemplo, a população de um país) conhecerá qualquer outro membro arbitrário do mesmo grupo a partir de uma cadeia de amizades de tamanho médio seis.

Milgram, em 1967, enviou cartas para uma série de pessoas nos Estados Unidos perguntando se o destinatário conheceria uma determinada pessoa de outra região daquele país. Caso a pessoa o conhecesse, aquela cadeia acabaria ali. Se não fosse esse o caso, a pessoa escolhida deveria enviar a mesma carta para um amigo, na esperança que este amigo conheça o tal destinatário, e se não, repetir o procedimento. Generalizando, uma pessoa “a” poderia não conhecer uma pessoa “z”, porém a pessoa “a” poderia conhecer “b”, que por sua vez conhece “c”, formando a cadeia “a”, “b”, “c”, ..., “y”, “z”, conectando assim a pessoa “a” a pessoa “z”. O tamanho aproximado médio dessas cadeias foi de seis, ou seja, concluiu-se que naquele ano, nos Estados Unidos, quaisquer pessoas eram conectadas entre si por seis intermediários em média. Algo curioso é que o termo popular “Seis graus de separação” não foi criado por Milgram, mas pelo escritor John Guare (18).

O conceito utilizado por Milgram para descrever sua rede social é muito parecido com o conceito moderno de rede complexa. Cada elemento desta rede (cada pessoa) se conecta com outro elemento de acordo com uma regra pré-estabelecida (nesse caso a amizade), formando uma rede social, e na presente dissertação, rede complexa. É comum denominar um elemento da rede complexa (no experimento de Milgram, uma pessoa) como um nó ou vértice, e cada interação (amizade) como aresta ou ligação. A figura 2.12 é um exemplo de uma rede

complexa formada a partir de uma suposta rede de amizades. Nesta rede, Vitor não conhece Edi, porém, Vitor conhece Poliana, que por sua vez conhece Isabella, que conhece Edi. Observe que é perfeitamente possível que um elemento da rede não esteja conectado, como evidenciado na figura pelo vértice Stanley.

Figura 2.12 - Rede complexa baseada numa rede de amizades. Os vértices são caracterizados por pessoas, enquanto as arestas por ligações de amizade entre estas pessoas.

Uma rede complexa baseia-se num conjunto de elementos com características comuns (vértice) e uma regra de interação entre esses elementos (aresta). Assim, praticamente qualquer sistema poderia ser descrito através de uma rede complexa, justificando a grande variedade de aplicações em inúmeros campos da ciência.

Conforme discutido nos tópicos iniciais de (4), pode-se utilizar as redes complexas para modelar redes sociais, mas não apenas de amizades como no experimento de Milgram, mas redes de relações comerciais, casamentos entre famílias, e até mesmo redes de contato sexual. Redes de informações também podem ser estudadas, como por exemplo, a rede formada pelas citações entre artigos científicos, ou a *World Wide Web*, a rede de páginas de internet conectadas entre si a partir de hiperligações. A estrutura física da rede de computadores, seu

cabeamento e roteadores, também pode ser interpretados como uma rede complexa, assim como a estrutura da rede de distribuição elétrica de uma região, ou mesmo a conexão dos vôos entre aeroportos.

Sistemas biológicos também podem ser estudados a partir de redes complexas, como redes de predador-presa, onde cada espécie é interpretada como um vértice da rede, e a aresta seria a interação predatória entre as duas espécies. Outros exemplos são as redes formadas por interações mecânicas físicas entre proteínas, redes formadas por neurônios, ou mesmo redes formadas a partir do sistema vascular de plantas. Mais detalhes, e bibliografia adequada para cada um desses exemplos podem ser encontrados em (4).

2.7.1 REDES COMPLEXAS: DEFINIÇÕES

Uma rede complexa Γ é um objeto abstrato (19), (20) formado por um conjunto V de vértices ou nós, e por um conjunto E de arestas ou ligações, sendo que estas relacionam dois pares de vértices. O conjunto de vértices de uma rede Γ é denotado como $V(\Gamma)$, enquanto que o conjunto de arestas $E(\Gamma)$. Caso dois vértices $v_i, v_j \in V(\Gamma)$ sejam conectados por uma aresta E , estes vértices v_i e v_j são chamados vizinhos, e a aresta de conexão é denotada como o par ordenado $E = (v_i, v_j)$. Existem redes que possuem arestas denominadas hiper-arestas, as quais podem ligar um vértice com todo um conjunto de outros vértices, porém neste trabalho serão abordadas apenas arestas simples, de modo que a igualdade $(v_i, v_j) = (v_i, v_k)$ será válida se, e somente se, $v_j = v_k$. Note que desta forma as arestas terão uma direção preferencial, sendo que a rede será denominada rede direcionada, $(v_i, v_j) \neq (v_j, v_i)$. Seu contraponto será uma rede não-direcionada, onde a igualdade $(v_i, v_j) = (v_j, v_i)$ é válida.

Um método usual de representar uma rede complexa é utilizando uma matriz A , denominada matriz de adjacência. Esta matriz será $N \times N$, onde N é o número de vértices da rede. Cada elemento a_{ij} da matriz A terá valor baseado nas

conexões entre os vértices v_i e v_j . Caso $(v_i, v_j) \in E(\Gamma)$ o valor de a_{ij} será 1, caso não exista aresta conectando os dois vértices o valor de a_{ij} será zero.

Geralmente é necessário associar um valor numérico a cada uma das arestas (pesos), sendo que estes têm algum significado para o sistema apresentado. Exemplos seriam tempo e/ou distância, no caso de redes geográficas, força de reação no caso de redes protéicas, ou proximidade social no caso de redes sociais. Dessa forma, mapeia-se cada elemento (v_i, v_j) num peso correspondente $w_{i,j}$ tal que $(v_i, v_j) \rightarrow w_{i,j}$. Num conjunto $V(\Gamma)$ com elementos N , defini-se uma matriz de pesos $W_{N \times N}$ onde cada elemento $w_{i,j}$ representa um peso $w_{i,j}$ e por conseqüência (v_i, v_j) .

No presente trabalho, cada peso $w_{i,j}$ representa a probabilidade de mudança entre os vértices v_i e v_j , $P(i \rightarrow j)$, e a matriz W é uma matriz estocástica. Uma diferença quanto à estrutura de W é que no modelo proposto o elemento associado à aresta entre v_i e v_j é dado por $w_{j,i}$ (seus índices são invertidos em relação à convenção apresentada, devido à matriz estocástica ser construída de forma esquerda ou direita).

2.7.2 REDES COMPLEXAS: MEDIDAS ESTRUTURAIS

As redes complexas têm uma série de propriedades locais e globais as quais já foram alvo de sucessivos estudos, conforme (4).

Uma das propriedades mais utilizadas é a conectividade local de um vértice v_i a partir do seu grau (21). O grau k_i de um vértice v_i será o número de vizinhos que este vértice possui dentro da rede Γ . A vizinhança de v_i é definida como a sub-rede Γ' onde todos os seus $v_k \in \Gamma'$ são vizinhos de v_i , assim como $V(\Gamma') \subseteq V(\Gamma)$ e $E(\Gamma') \subseteq E(\Gamma)$. A partir dos elementos da matriz de adjacência da rede, o grau k_i será:

$$k_i = \sum_{j=1}^N a_{j,i} \quad (2.13)$$

Numa rede onde as arestas têm pesos o conceito de grau naturalmente evolui para a propriedade denominada *strenght* s_i . Essa é dada pela soma de todos os pesos $w_{j,i}$ que representam arestas (v_j, v_i) , que conectam o conjunto de vizinhos v_i de com v_i . Assim:

$$s_i = \sum_{j=1}^N w_{j,i} \quad (2.14)$$

Outra propriedade é o coeficiente de aglomeração (*clustering coefficient*) que quantifica a conectividade da vizinhança do vértice em questão v_i . Sendo n_i o número de vizinhos de v_i , e e_i o número de conexões entre esses vizinhos, o coeficiente de aglomeração ca_i é definida como a probabilidade de dois vizinhos de v_i estarem conectados:

$$ca_i = \frac{2e_i}{n_i(n_i - 1)} \quad (2.15)$$

Num sistema onde as arestas têm pesos, é necessário ponderar o valor da conexão de cada uma delas, sendo o coeficiente dado por:

$$ca_i = \frac{1}{s_i(k_i - 1)} \sum_{j,k=1}^{N,N} \frac{w_{i,j} + w_{i,k}}{2} a_{i,j} a_{i,k} w_{j,k} \quad (2.16)$$

Tais propriedades são locais, todas calculadas sobre um nó em questão. Algumas redes podem ser caracterizadas pela média das propriedades locais de seus vértices, tornando úteis os conceitos de conectividade média, *strenght* médio e coeficientes de aglomeração médios. Aqui novamente o índice i refere-se às propriedades do vértice v_i .

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i \quad (2.17)$$

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^N s_i \quad (2.18)$$

$$\langle ca \rangle = \frac{1}{N} \sum_{i=1}^N ca_i \quad (2.19)$$

2.7.3 REDES COMPLEXAS: MODELOS DE REDES COMPLEXAS

Os métodos de construção de redes complexas são resultado do vasto estudo dentro desta área, onde basicamente foram propostos modelos teóricos para tentar explicar uma construção de uma rede experimental. Aqui serão expostos três modelos usuais de redes complexas, mais detalhes poderão ser encontrados em (4), (21), (22) e (18).

2.7.3.1 REDE REGULAR

A rede geográfica é um tipo de rede conceitualmente simples, utilizada em diversos ramos da ciência. Esta advém da discretização de um espaço contínuo em subdivisões. Dada uma região R definida sobre um espaço qualquer, defini-se $r_i \subset R$ tal que $\sum_{i=1}^N r_i = R$, onde N é o número de subdivisões r_i . Definindo-se um v_i correspondente a cada r_i , a rede Γ onde $v_i \in V(\Gamma)$ para todo e qualquer que seja $i \in [1, N]$, será uma rede regular baseada na região R .

As arestas são definidas de maneira arbitrária, podendo conter ou não correspondência geográfica. A figura 2.13 (a) e (b) mostra uma rede geográfica baseada na divisão dos estados do Brasil. Neste caso cada estado é representado como um vértice, e as arestas conectam estados fronteiriços.

Figura 2.13 - Rede complexa geográfica baseada na distribuição dos estados do Brasil. Os vértices foram distribuídos de forma a representar cada um um estado, enquanto que as arestas foram distribuídas de acordo com a presença ou não de uma fronteira entre os dois estados. A rede foi sobreposta com o mapa Brasileiro de modo a facilitar o seu entendimento. Fonte: Felipe Menegaz, 2007. Adaptado. Disponível em: http://pt.wikipedia.org/wiki/Ficheiro:Brazil_Labelled_Map.svg Acesso: 20 jun 2010

Um caso particular de rede geográfica é a rede regular, rede esta onde cada r_i é dividido de maneira semelhante. Tal arranjo faz com que a rede apresente uma forte simetria topológica. Um exemplo desse tipo de arranjo é a figura 2.14, uma rede onde r_i são quadrados de um espaço R bidimensional.

Figura 2.14 - Redes regulares. Os vértices são representados pelos círculos, sendo os azuis vértices genéricos, os verde claros são os vizinhos do vértice verde escuro. Em (a) é observada a condição teórica, onde todo o espaço é preenchido pela rede, enquanto que em (b) é adotada a condição periódica de contorno, necessária para representar tal rede num ambiente computacional.

As redes regulares têm propriedades notáveis. A partir da rede formada a partir de quadrados (figura 2.14), é possível calcular de maneira trivial a conectividade média, uma vez que todo e qualquer vértice dessa rede tem quatro vizinhos como mostrado na figura 2.14 (a) pelas cores verde e verde escuro, a conectividade média será $\langle k \rangle = 4$. De modo semelhante, o coeficiente de aglomeração de todos os vértices será o mesmo devido à topologia da rede, dois vértices conectados ao mesmo vértice nunca serão conectados, de modo que $\langle ca \rangle = 0$. Caso a rede fosse hexagonal, resultante da divisão do espaço em hexágonos, e não em quadrados, os valores dessas duas propriedades iriam se alterar, porém o raciocínio utilizado para o cálculo do valor não.

A figura 2.14 evidencia outra característica das redes regulares. Em geral esta é definida sobre um espaço infinito, gerando uma rede de infinitos vértices. Quando se opera com uma rede deste tipo num ambiente computacional é necessário criar uma condição periódica de contorno, uma vez que não é possível alocar memória infinita num computador. Um exemplo de tal condição é a figura 2.14.

2.7.3.2 REDE DE ERDÖS-RÉNYI

A rede de Erdős-Rényi foi inicialmente proposta por Solomonoff e Rapoport (23), porém é mais conhecida pelo estudo de Erdős e Rényi (24), que publicaram um conjunto de artigos no final da década de 50 sobre este assunto. Tal rede pode ser interpretada como a oposição às redes regulares, esta é formada aleatoriamente a partir de um conjunto de premissas. Este tipo de rede também é chamada de rede aleatória.

O primeiro caso de estudos supunha a distribuição aleatória de m arestas entre vértices n . Porém tal definição mostrou-se complexa para o cálculo de propriedades básicas, sendo substituída por um modelo que lida com a probabilidade de conexão entre dois vértices. Matematicamente (18) este modelo é denominado $G(n, p)$, este dita que dois vértices de um conjunto de n vértices terão probabilidade p de serem conectados. Tecnicamente $G(n, p)$ é definido como um ensemble, a probabilidade de uma conformação específica G é dada por:

$$P(G) = p^m (1 - p)^{\binom{n}{2} - m} \quad (2.20)$$

Onde:

p : é a probabilidade de dois vértices serem conectados.

m : é o número de arestas formadas

n : é o número total de vértices

As propriedades desta rede são:

$$\langle k \rangle = p(N - 1) \quad (2.21)$$

$$\langle ca \rangle = \frac{\langle k \rangle}{N} \quad (2.22)$$

2.7.3.3 REDE DE BARABÁSI-ALBERT

Um dos tipos de rede mais interessantes na opinião deste autor são as redes que possuem um elemento chamado *hub*, que é um vértice centralizador, com muitas conexões. Estes aparecem em uma série de redes, como por exemplo, as redes sociais, um *hub* seria uma pessoa com um número de amizades muito elevado. Numa rede de aeroportos os *hubs* são aqueles pontos concentradores de vôos, preferidos para fazer conexões de longa distância.

Tais redes também são conhecidas como redes que possuem uma lei de potencias, neste caso, a probabilidade de um vértice ter uma conectividade k é dada por:

$$P(k) = \beta k^{-\alpha} \quad (2.23)$$

Onde β e α são constantes reais.

Tal lei dita que existe uma probabilidade não nula de existirem vértices com altos valores de conectividade, sendo estes os *hubs*. Redes que obedecem a leis de potência também são conhecidas como redes livre de escala.

Barabási e Albert em 1999 (22) propuseram um modelo promissor para modelar tais redes livre de escala baseado no estudo de ponteiros dentro da *World Wide Web*. Tal modelo prevê o crescimento de uma rede a partir de algumas características. Um novo vértice é introduzido na rede, e este automaticamente terá um número de conexões c com os outros vértices que até então fazem parte da rede. Estas novas conexões deverão prioritariamente se dar com os vértices que possuem mais vizinhos, de forma que a probabilidade $P_{novo \rightarrow i}$ de um novo vértice v_{novo} se conectar ao vértice v_i cujo grau é k_i será γ , sendo uma constante de normalização positiva:

$$P_{novo \rightarrow i} = \gamma k_i \quad (2.24)$$

O fato dos novos vértices se conectarem aos vértices mais conectados (fenômeno parafraseado como o rico fica mais rico, também conhecido como vantagem acumulada (25)) faz com que a maior parte das conexões se dêem com os vértices mais antigos da rede, fazendo com que os mais novos fiquem marginalizados. Tais redes possuem $\langle k \rangle = 2c$ e $\langle ca \rangle = N^{-75}$ (26).

3 MODELO: DESCRREVENDO
UMA IMAGEM EM TERMOS
DE UMA REDE COMPLEXA

O modelo aqui proposto reúne características biológicas, assim como conceitos das áreas de caminhadas aleatórias, cadeias de Markov, redes complexas e análises de imagens a fim de simular a capacidade visual de um observador.

Imagine inicialmente um observador com os olhos fechados e este é levado a um lugar o qual ele não conhece. Ao abrir os olhos, ele se depara com uma paisagem desconhecida, e irá então observá-la. Tome como exemplo de cena a figura 3.1 (a) a seguir. Inicialmente o observador irá voltar sua atenção para um ponto arbitrário da cena, como, por exemplo, a região delimitada pelo círculo vermelho. Conforme discutido em seção 2.2 é impossível o reconhecimento completo da cena com apenas uma região sendo projetada sobre a fóvea. É necessário um movimento sacádico para outra região de interesse da mesma imagem, a fim de obter o máximo possível de informação, resultando na interpretação completa da cena. Dessa forma, o observador muda o foco da sua atenção e começa a observar outra parte da cena.

Figura 3.1 - Em (a) imagina-se que o observador está num lugar desconhecido, e começa a reconhecer a cena olhando inicialmente para a região delimitada pelo círculo vermelho. Em (b) são evidenciadas por círculos azuis as regiões as quais o observador poderá voltar sua atenção depois de terminar de observar a região inicial.

A região escolhida provavelmente será aquela que tiver a maior saliência, conforme debatido na seção 2.4. Aqui adotamos uma teoria probabilística para prever qual será o ponto do próximo salto. A abordagem estatística é razoável neste caso porque em experimentos foram observados no mínimo uma grande variabilidade de resultados (27). Considere a divisão da imagem conforme a figura 3.1 (b) em outras seis outras regiões. Definimos então uma probabilidade de mudança entre o ponto inicial (círculo vermelho) e quaisquer uma das outras regiões

(círculos azuis). Note que a probabilidade de mudança deve levar em consideração a saliência local da região, locais que chamam mais atenção deverão ter maior probabilidade de serem visitados.

Biologicamente estamos propondo que a saliência visual irá dirigir a atenção seletiva, enquanto que a atenção seletiva se torna eficiente devido à saliência visual.

Com base na figura 3.1 (b) o observador então voltará sua atenção para a próxima região. Nesta nova região ele terá novamente probabilidade de voltar sua atenção para cada uma das outras regiões. É importante ressaltar que em qualquer um dos passos é possível que o observador continue a observar a região em que está – talvez ele esteja procurando por mais detalhes de uma região rica, ou mesmo observar uma região não demarcada na figura 3.1 (b), todas as regiões da cena podem ser visitadas, e todas elas têm necessariamente uma probabilidade de mudança associada.

O modelo propõe que essa mecânica ocorra repetidamente, o observador irá observar cada uma das regiões que achar necessária para interpretar integralmente a cena, e cada vez que ele visitar uma região haverá probabilidade da sua atenção se voltar para cada uma das outras regiões (ou continuar a observar a mesma região).

Note que o modelo propõe a divisão discreta da imagem, e de forma semelhante, saltos discretos entre cada uma das regiões. Em outras palavras, o observador muda o seu foco de atenção de um ponto a outro sem observar nenhuma das regiões que por ventura estão entre esses dois pontos, tal qual o movimento balístico utilizado pela sacada. De fato, as sacadas que os olhos executam, “saltos” de longa distância da atenção permitem que a mecânica desta seja modelada pela mecânica uma rede complexa, utilizando regiões distintas que concentram a atenção como um todo – os vértices – e ligações discretas entre essas regiões, saltos de atenção – as arestas da rede.

Tais características do sistema visual fazem com que a rede complexa resultante tenha uma topologia peculiar. O processo de formação dos vértices é geográfico, cada região geográfica do campo visual do observador é mapeada por vértices. Porém suas conexões, as arestas podem ou não serem distribuídas de

maneira geográfica. É possível inclusive em alguns casos a identificação de *hubs*, que no contexto da atenção representam objetos que monopolizam a atenção do observador. Mais detalhes a respeito da formação prática das arestas da rede podem ser encontrados no capítulo 6.

Em resumo o modelo interpreta a atenção seletiva de um observador como uma rede complexa. Os vértices desta rede são dados pelos elementos salientes do espaço visual do observador – aqueles elementos que o observador focalizará sua atenção – e as arestas são distribuídas com base na existência ou não de uma probabilidade de mudança da atenção a partir de um vértice para outro. Deste modo a mecânica da atenção seletiva do observador será simulada a partir da mecânica na rede construída.

Observe que o modelo pode lidar com redes dinâmicas, no caso de um campo visual dinâmico, objetos poderiam modificar-se, aparecer ou sumir de cena, fazendo com que existam modificações na configuração de arestas e vértices, podendo até incluir ou excluir um novo vértice. Porém a presente dissertação está restrita à análise de imagens, fazendo com que as redes estudadas sejam estáticas.

3.1 CONSIDERAÇÕES TEÓRICAS SOBRE O MODELO

Define-se uma imagem como uma função tridimensional $Im(x, y)$, onde o valor de cada pixel corresponde à altura dessa função, conforme seção 2.5. $Im(x, y)$ é subdividida em n regiões, conforme ilustrado na figura 3.2 a seguir.

Figura 3.2 - Imagem segmentada em n regiões arbitrárias. Fonte: xkcd. Adaptado. Disponível em: <http://xkcd.com/695/> Acesso em: 20 jun 2010.

De acordo com o modelo proposto, para cada região i existe uma probabilidade de mudança para uma região j :

$$P(i \rightarrow j) = P(X_{t+1} = j \mid X_t = i, C = C_t') \quad (3.1)$$

Onde:

$P(i \rightarrow j)$: é a probabilidade de mudança da região i para a região j .

t : Variável discreta temporal, representa o número iterações ou passos.

X_t : Variável aleatória que representa a posição de observação num tempo t .

C : é o conjunto das variáveis aleatórias $\{X_1, X_2, \dots, X_t\}$, sendo que estas representam a região observada em cada tempo t .

C_t' : é o caminho percorrido na caminhada em questão.

Imaginando esse sistema como uma rede complexa, cada região da imagem é interpretada como um vértice, e as probabilidades de mudança serão arestas, conforme evidenciado na figura 3.3. Observe que a rede complexa resultante tem uma distribuição geográfica de vértices, cada região r_i é descrita por um vértice v_i , e cada aresta (v_i, v_j) terá como peso o valor da probabilidade de transição, $P(i \rightarrow j)$.

Figura 3.3 - Representação gráfica das probabilidades de transição entre as regiões 1, 2 e 3.
 Fonte: xkcd. Adaptado. Disponível em: <<http://xkcd.com/695/>> Acesso em: 20 jun 2010.

De posse do conjunto de todos os $P(i \rightarrow j)$ possíveis para todos os valores de i e j presentes é possível construir a matriz estocástica W_t que representa os valores das probabilidades de mudanças para um certo tempo t .

$$W_{j,i}^t \equiv P(i \rightarrow j) = P(X_t = j | X_{t-1} = i, C = C_t') \quad (3.2)$$

De forma que:

$$W_t = \begin{bmatrix} W_{1,1}^t & W_{1,2}^t & \dots & W_{1,n}^t \\ W_{2,1}^t & W_{2,2}^t & & \\ \dots & & \dots & \\ W_{n,1}^t & & & W_{n,n}^t \end{bmatrix} \quad (3.3)$$

É importante ressaltar que as colunas da matriz W_t têm de ser normalizadas, de modo que para um i fixo:

$$\sum_{j=1}^n W_{j,i}^t = 1 \quad (3.4)$$

Para todo e qualquer que seja i natural $\in [1, n]$.

É necessário para estudar o valor de cada $P(i \rightarrow j)$ definir inicialmente o escopo exato deste trabalho. Para tal definimos o um vetor $\vec{f}_i$ (*features*) contendo propriedades locais de cada região r_i de $\text{Im}(x, y)$, como por exemplo, média, variância, entropia, etc..

$$\vec{f}_i = \begin{bmatrix} media_i \\ variancia_i \\ entropia_i \\ \dots \end{bmatrix} \quad (3.5)$$

Também sendo necessário definir o conjunto F , o conjunto de todos os vetores $\vec{f}_i$ sobre suas regiões:

$$F = \{f_1, f_2, \dots, f_i, \dots, f_n\} \quad (3.6)$$

Assim, a probabilidade pode ter todas suas dependências explicitadas na forma:

$$P(i \rightarrow j) = h_{(i \rightarrow j)}(F, C_t') \quad (3.7)$$

A função $h_{(i \rightarrow j)}$ será a probabilidade de mudança da região i para a j , dado o conjunto de propriedades F , e o caminho percorrido até então C_t' . A determinação explícita de $h_{(i \rightarrow j)}$ não é conhecida, sendo este é um dos escopos deste trabalho. No capítulo 6 serão apresentados mais detalhes a respeito das tentativas em determinar $h_{(i \rightarrow j)}$. Porém aqui se torna necessário fundamentar de acordo com o objeto de estudo (saliência visual e atenção seletiva) quais os fundamentos para a dependência de $h_{(i \rightarrow j)}$.

Com relação ao caminho C_t' , a dependência pode ser interpretada como a presença da memória humana no modelo. À medida que o observador permanece observado uma mesma região é razoável supor-se que seu interesse sobre aquela região irá diminuir (28), uma vez que ele já absorveu boa parte da informação presente, de modo que:

$$h_{(i \rightarrow i)}(F, C_{t+1}') \leq h_{(i \rightarrow i)}(F, C_t') \quad (3.8)$$

Assim a probabilidade de um observador continuar a observar a mesma região i sempre cairá em relação à probabilidade anterior de observação. O mesmo ocorre de maneira geral, quando uma região recebe uma visita, a probabilidade de ela ser visitada novamente a partir de qualquer região presente também diminui.

Note que esse efeito é válido apenas para análises em passos necessariamente subseqüentes. Devido à $h_{(i \rightarrow j)}$ representar uma probabilidade, quando o seu valor numa região diminui, o seu valor nas outras regiões aumenta, conforme a equação (3.4). Dessa forma o valor de $h_{(i \rightarrow j)}$ tende a ser bastante dinâmico, fazendo com que cada passo influencie a probabilidade de mudança para cada uma das regiões.

Essa forma de $h_{(i \rightarrow j)}$ torna necessário recalcular W_t em cada passo, requerendo maior poder computacional, saindo do escopo deste trabalho. Assim o efeito da memória neste trabalho será desprezado, ficando a cargo de investigações futuras.

Desta forma:

$$P(i \rightarrow j) = P(X_{t+1} = j \mid X_t = i) \quad (3.9)$$

$$W_{j,i} = P(i \rightarrow j) = P(X_t = j \mid X_{t-1} = i) \quad (3.10)$$

$$W = \begin{bmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,n} \\ W_{2,1} & W_{2,2} & & \\ \vdots & & \ddots & \\ W_{n,1} & & & W_{n,n} \end{bmatrix} \quad (3.11)$$

$$\sum_{j=1}^n W_{j,i} = 1 \quad (3.12)$$

Por fim

$$P(i \rightarrow j) = h_{(i \rightarrow j)}(F) \quad (3.13)$$

Note que apesar de ser esperada uma diferença entre os resultados simplificados e os que apresentam dependência temporal, os resultados obtidos podem ser considerados boas aproximações para observações iniciais, isso é, quando o observador começa a observar uma cena a memória não promoverá grande diferença entre os resultados, de forma que esta começa a ser realmente importante depois de algum tempo de observação da cena. Também se espera minimizar o efeito da memória ao trabalharmos com imagens sem conteúdo

semântico, tais imagens não fornecem informação suficiente para a memória tornar uma região desinteressante.

Utilizando um raciocínio análogo, poderia ser proposto por simplicidade um $h_{(i \rightarrow j)}$ dependente apenas de $\vec{f}_i$ e $\vec{f}_j$, os dois vetores propriedades das duas regiões em questão, e não uma dependência global de todo o conjunto F , porém este apresenta claros desvios de resultados, conforme o contra-exemplo a seguir.

Neste caso, imagina-se que estamos procurando pela probabilidade $P(1 \rightarrow j)$, a probabilidade de mudança a partir da região 1, para qualquer outra região da imagem. Tome para este exemplo a rede formada a partir da imagem, a figura 3.2. É razoável supor-se que a maior probabilidade seria para a região 3, pois ali está o personagem da cena (o personagem é a sonda *spirit*, enviada à Marte). Também é razoável que a probabilidade $P(1 \rightarrow 2)$ seja pequena, o fundo da imagem parece desinteressante. Porém, caso a região fosse substituída pela figura 3.4, muito provavelmente a probabilidade $P(1 \rightarrow 2)$ seria muito maior, de modo que a probabilidade $P(1 \rightarrow 3)$ seria menor, mesmo sem nenhuma alteração nas $\vec{f}_1$ e $\vec{f}_3$. Assim a expressão da equação 3.13 continua válida.

Figura 3.4 - Imagem para uma alternativa à região dois – nesse caso a saliência desta região seria maior.

3.2 EXEMPLO SIMPLES DO MODELO

Considere a imagem simples exemplificada na figura 3.5 (a) como exemplo:

Figura 3.5 - Imagem simples de 12x12 pixels utilizada para ilustrar o modelo (a) e a mesma imagem sobreposta às regiões que delimitam as a área na qual as propriedades que serão extraídas.

Sua matriz computadorizada será ilustrada na figura 3.6.

0,96	1,00	0,92	1,00	1,00	0,98	1,00	0,99	0,99	1,00	1,00	0,94
1,00	0,96	1,00	0,95	0,95	1,00	0,98	1,00	1,00	1,00	0,96	1,00
1,00	0,96	0,96	1,00	1,00	0,97	1,00	1,00	1,00	0,93	1,00	0,96
0,94	1,00	0,98	1,00	0,96	1,00	0,00	0,98	0,99	1,00	0,94	0,03
1,00	0,94	1,00	0,96	1,00	1,00	0,04	1,00	0,03	0,02	0,98	0,00
0,98	0,97	1,00	1,00	0,89	0,97	1,00	0,00	0,00	0,00	0,00	1,00
1,00	1,00	0,02	0,00	0,02	0,07	0,00	0,00	0,00	0,01	0,02	1,00
0,01	0,00	0,03	0,00	0,03	0,00	0,00	0,04	0,01	0,00	0,97	1,00
0,96	0,07	0,00	0,00	0,04	0,00	0,00	0,00	1,00	0,97	1,00	1,00
1,00	0,9	0,07	0,00	0,00	0,04	0,02	0,03	0,91	1,00	0,96	0,93
1,00	0,01	0,01	1,00	1,00	1,00	0,00	0,95	0,09	0,98	1,00	1,00
1,00	0,00	0,99	0,01	0,96	0,00	1,00	1,00	0,00	1,00	0,99	0,99

Figura 3.6 - Ilustração da matriz computadorizada da figura 3.5 Nesta matriz, os elementos em azul escuro correspondem ao centro de cada região analisada, e em azul claro os elementos pertencentes a esta região.

Neste exemplo assumiremos três regiões distintas de observação, círculos de raio 3 pixels cujos centros são dados pelas coordenadas (4,4), (5,9) e (9,4), conforme a figura 3.5 (b). Note que os valores apresentados estão na forma (linha,coluna).

Utilizaremos duas propriedades locais de cada região: A média dos seus níveis de cinza, e a média sobre a região do módulo do gradiente. Também serão utilizadas as coordenadas do centro de cada círculo, de modo que o vetor de características para cada região i será:

$$\vec{f}_i = \begin{bmatrix} \text{linha}_i \\ \text{coluna}_i \\ \text{Cinza}_i \\ |\text{gradiente}_i| \end{bmatrix} \quad (3.14)$$

O gradiente pode ser aproximado de acordo com a seção 2.5.1 como a aplicação do operador de Sobel sobre a figura 3.5. Dessa forma, o conjunto F deste exemplo será:

$$F = \left\{ \vec{f}_1 = \begin{bmatrix} 4,00 \\ 4,00 \\ 0,98 \\ 0,27 \end{bmatrix}, \vec{f}_2 = \begin{bmatrix} 5,00 \\ 9,00 \\ 0,46 \\ 0,98 \end{bmatrix}, \vec{f}_3 = \begin{bmatrix} 9,00 \\ 4,00 \\ 0,09 \\ 0,69 \end{bmatrix} \right\} \quad (3.15)$$

Torna-se necessário definir a função $h_{(i \rightarrow j)}$ (eq. 3.13) para este exemplo:

$$h_{(i \rightarrow j)} = \alpha_i \frac{1}{d_{i,j}} \left(\frac{\vec{f}_j(4)}{\vec{f}_i(4)} \right) \quad (3.16)$$

Onde:

α_i : é uma constante de proporcionalidade, será debatida adiante.

$d_{i,j}$: é a distância euclidiana entre as duas regiões em questão, i e j , calculadas de acordo com os elementos $\vec{f}_i(1)$, $\vec{f}_i(2)$, $\vec{f}_j(1)$, e $\vec{f}_j(2)$.

$\vec{f}_i(4)$: é o quarto elemento do vetor $\vec{f}_i$, no caso deste exemplo a média do módulo do gradiente dos níveis de cinza na região i .

Devido ao termo inversamente proporcional a distância é necessário definir a probabilidade de mudança de um nó para ele mesmo. Por simplificação, será zero nesse exemplo, o observador nunca permanecerá no mesmo local. Dessa forma:

$$h_{(i \rightarrow i)} = 0 \quad (3.17)$$

Note que a forma da eq. 3.16 é arbitrária, foi utilizada aqui uma expressão simples para melhor entendimento do exemplo.

O termo α_i refere-se a uma constante comum a todas as probabilidades de mudança a partir do nó i , e este será apenas o valor de correção para que a soma de cada uma das colunas seja unitária (eq. 3.12)

.Uma vez conhecidos os termos de $h_{(i \rightarrow j)}$ é possível construir a matriz estocástica desnormalizada. De acordo com a eq. 3.10:

$$W_{j,i}' = h_{(i \rightarrow j)}' = \frac{1}{d_{i,j}} \left(\frac{\vec{f}_j(4)}{\vec{f}_i(4)} \right)$$

$$W_{1,1}' = h_{(1 \rightarrow 1)}' = 0$$

$$W_{1,2}' = h_{(2 \rightarrow 1)}' = \frac{1}{d_{2,1}} \left(\frac{\vec{f}_1(4)}{\vec{f}_2(4)} \right) = \frac{1}{5,09} \times \left(\frac{0,27}{0,98} \right) = 0,05$$

$$\dots$$

$$W' = \begin{bmatrix} 0,00 & 0,05 & 0,07 \\ 0,71 & 0,00 & 0,22 \\ 0,51 & 0,10 & 0,00 \end{bmatrix} \quad (3.18)$$

Normalizando cada uma das colunas da eq. 3.12, a matriz estocástica da eq. 3.11:

$$W = \begin{bmatrix} 0,00 & 0,33 & 0,26 \\ 0,58 & 0,00 & 0,74 \\ 0,42 & 0,67 & 0,00 \end{bmatrix} \quad (3.19)$$

A análise da matriz W obtida depende diretamente da função h escolhida.

Neste exemplo h era composta por dois fatores multiplicativos: $\frac{1}{d_{i,j}}$ e $\frac{\vec{f}_j(4)}{\vec{f}_i(4)}$. O primeiro termo é proporcional ao inverso da distância pode ser interpretado como a falta de resolução à medida que a imagem é projetada distante da fóvea, dentro do

olho. Uma vez que a imagem não é nítida longe do centro de observação, pode-se esperar que não exista grande probabilidade da atenção se voltar para uma parte distante da imagem. Porém essa abordagem adiciona um novo problema, com relação à atenção permanecer no mesmo ponto. Em grande parte do tempo nossa atenção se volta para um único objeto ou região, e apenas depois de algum tempo de observação a atenção se volta para outra região.

Já o termo $\frac{\vec{f}_j(4)}{\vec{f}_i(4)}$ é interpretado como uma tendência a favorecer regiões de gradientes mais altos. Caso a região destino ($\vec{f}_j(4)$) tenha a média do módulo do gradiente mais alto do que a região inicial ($\vec{f}_i(4)$) o termo em questão será maior que 1, favorecendo a mudança. De maneira análoga, caso $\vec{f}_j(4)$ seja menor que $\vec{f}_i(4)$, a razão será menor do que 1, fazendo com que a mudança seja menos provável.

Com base na análise dos termos de h pode-se interpretar a matriz W obtida. Uma vez que as distâncias entre cada uma das regiões têm valores semelhantes, a média sobre o gradiente torna-se fundamental para determinar as probabilidades de mudança. Neste exemplo a região 2 - a região de maior gradiente - tem maiores probabilidades de mudanças que as duas outras regiões, o que significa que de acordo com a função h escolhida, um observador iria observar essa região mais atentamente do que as outras.

4 EXPERIMENTO

O experimento para validar o modelo tornou-se viável a partir de uma colaboração com o laboratório PAMELA - *Pedestrian Accessibility and Movement Monitoring Environment Laboratory* - da *University College*, Londres, através do professor Sérgio Velastin. Este laboratório possui o equipamento necessário para realizar os experimentos, um *eye tracker*. Vale ressaltar que tal experimento foi não invasivo, tendo envolvido um único indivíduo.

Um *eye tracker* é um aparelho capaz de, a partir de uma calibração prévia mostrar as coordenadas do foco de atenção da pessoa que o está usando. O foco de atenção nesse contexto seria a região da imagem projetada sobre a fóvea no olho humano, conforme debatido nas seções relativas à base biológica desta dissertação. Nesse caso em específico, foi projetada sobre um anteparo fixo uma seqüência de oito imagens. Com o observador sentado sobre uma cadeira, e o equipamento pré-calibrado, foi possível filmar o que seria o campo visual do observador, o equipamento adicionava ao filme um círculo vermelho, sendo esta a posição do foco de atenção do observador. Tal filme tinha resolução de 720x576 pixels e continha 25 *frames* por segundo. Um exemplo de um *frame* encontrado no filme é a figura 4.1, e o vídeo completo para download pode ser encontrado no seguinte link da web: http://cyvision.ifsc.usp.br/~gvrech/filme_eye_tracker.mpg.

Figura 4.1 - Exemplo de um *frame* obtido a partir do vídeo experimental. Tal filme foi resultado da utilização de um eye tracker, equipamento que filma o campo visual de um observador e adiciona uma marcação (círculo vermelho) na posição do foco de atenção.

Inicialmente esperava-se exibir todas as oito imagens para 30 observadores distintos. Porém o laboratório em questão teve um problema com a calibragem original do aparelho, não sendo possível no tempo hábil realizar o experimento com todos os observadores. Com essa limitação, foi possível verificar o campo visual de um observador, sendo que este servirá de protótipo para investigações futuras.

O experimento consistiu em exibir para um observador um conjunto de oito imagens escolhidas previamente, as quais foram exibidas sem interrupção durante aproximadamente 50 segundos. As imagens encontram-se compiladas na figura 4.2, na ordem correspondente a que foram exibidas. Foram escolhidas imagens em níveis de cinza, assim como imagens sem nenhuma interpretação prévia, apenas imagens compostas de texturas.

continua

continuação

Figura 4.2 - Conjunto das oito imagens utilizadas no experimento.

Embora o modelo proposto seja capaz de lidar com figuras complexas coloridas, para o experimento adotamos apenas figuras de interpretação simples. Essa decisão foi tomada para manter simples os algoritmos utilizados na obtenção da saliência visual de cada imagem. A presença da interpretação da cena, como por exemplo, o reconhecimento de uma face, modifica radicalmente a saliência de uma imagem, o cérebro humano possui até mesmo um neurônio especializado no reconhecimento de faces. Porém a complexibilidade computacional necessária para segmentar uma face numa imagem é enorme (29) e os algoritmos não necessariamente tem ótimo rendimento. Desse modo, figuras contendo algum componente interpretativo fogem do escopo deste trabalho de mestrado.

O experimento teve como resultado um vídeo de 7 minutos e 50 segundos, contendo a posição do olho do observador nesse período.

4.1 OBTENDO OS DADOS EXPERIMENTAIS

Uma vez que o resultado do experimento não foi diretamente as coordenadas do foco visual do observador, mas um filme contendo suas coordenadas, foi necessário desenvolver um algoritmo que automaticamente analisasse o filme e retornasse as coordenadas.

Inicialmente definimos quais seriam as informações a serem retiradas do filme. As coordenadas do foco de atenção do observador são as primeiras informações úteis, porém não são as únicas. O modelo foi desenvolvido para ser aplicado sobre uma imagem, porém tal imagem foi projetada sobre um anteparo, e este não era o único elemento no campo visual do observador, como pode ser visto na figura 4.1. Como foi requisitado ao observador que este prestasse atenção somente na imagem projetada, é uma aproximação eficaz desprezar os outros elementos do campo visual do observador, levando em consideração apenas a imagem. Assim torna-se necessário segmentar as bordas da imagem dos outros componentes do campo visual do observador, assim como segmentar a localização do círculo que revela as coordenadas do foco de atenção. A figura 4.3 mostra a segmentação almejada a partir do mesmo *frame* do filme mostrado na figura 4.1. A região em amarelo pertence à parte do campo visual do observador que será desprezada, o polígono azul contorna a imagem projetada, que de acordo com o modelo é todo o campo visual simulado, e o círculo vermelho evidencia qual região é o alvo da atenção do observador.

Figura 4.3 - Segmentação almejada a partir de um *frame* obtido do filme experimental. A parte amarela é a região que será desprezada por não fazer parte do campo visual simulado, a região em azul compreende o campo visual trabalhado, e a região vermelha a região do foco de atenção do observador.

Assim, o objetivo do algoritmo criado era receber como entrada uma imagem de um *frame* do filme e identificar sem supervisão humana 5 pontos: O ponto central do círculo vermelho, assim como cada um dos 4 vértices do polígono evidenciado pela cor azul na figura 4.3.

Inicialmente foram aplicados diversos limiares (*thresholds*), (12), procurando segmentar ao máximo a imagem, a fim de maximizar a eficiência dos métodos posteriormente utilizados. O melhor resultado encontrado foi utilizando um limiar médio igual a 0,5 apenas na componente verde (RGB) das imagens. A figura 4.4 mostra a evolução das imagens a partir da imagem original figura 4.4 (a), a componente verde RGB desta figura 4.4 (b), e por fim a componente verde limiarizada figura 4.4 (c):

Figura 4.4 - Um exemplo de uma figura obtida a partir de um *frame* do filme (a), a correspondente verde (RGB) desta figura (b), e a figura limiarizada a fim de ressaltar os componentes importantes (c). (e) é o resultado da aplicação de transformadas Hough retas e circulares sobre a figura, e (f) é a correspondente imagem com um círculo evidenciando o local onde um observador estaria olhando, com base na figura resposta da transformada Hough.

Sobre uma imagem do tipo da figura 4.4 (c), foram aplicadas dois tipos de transformada Hough, sobre retas e a circular. A transformada Hough sobre retas utiliza o fato de serem necessárias duas variáveis para se definir uma reta, seu coeficiente angular e linear. Cada par de pontos da imagem são mapeados num espaço de fases de forma semelhante a um histograma, onde as coordenadas dadas são os coeficientes linear e angular da reta que une estes dois pontos. Ao repetir esse procedimento para todos os conjuntos de pontos da imagem, regiões bastante populosas mapeiam prováveis retas na imagem original. O mesmo ocorre na transformada Hough circular, porém de uma forma mais complexa, uma vez que é necessário um espaço de fase tridimensional para mapear as três coordenadas necessárias para se especificar um círculo. Foram utilizadas as retas formadas a partir das 4 regiões cujo espaço de fase era mais denso no caso da transformada reta, e foi utilizada a região mais densa do espaço de fase para a coordenada do

foco de atenção, no caso da transformada circular. A imagem sobreposta aos elementos segmentados pode ser vista na figura 4.4 (d).

Uma vez que a transformada Hough reta fornece apenas retas, e não segmentos de reta, foi necessário procurar pela intersecção de das 4 retas resultantes para um resultado esperado, isso é, os pontos das bordas do campo visual simulado pelo modelo. Já a coordenada do círculo foi dada diretamente pelo resultado da transformada, não sendo necessárias quaisquer transformações posteriores.

De posse das 4 coordenadas que forneciam a borda da imagem foi possível construir um mapa conforme que convertia a coordenada já identificada no vídeo do foco de atenção para a coordenada correspondente utilizada para as simulações. A posição do foco de atenção convertido resultante da figura 4.4 (a) pode ser visualizada na figura 4.4 (e). Um exemplo do caminho que o foco de atenção faz pela imagem pode ser visualizado na figura 4.5.

Embora a mudança no foco de atenção possa ocorrer até três vezes por segundo (10), devido às características do filme foram observadas mudanças menos constantes, em média uma por segundo, de forma que esse foi o intervalo de amostragem para os *frames*, a cada segundo do filme foi extraída a imagem de um *frame* para ser submetida ao algoritmo de segmentação descrito anteriormente. Uma vez que cada uma das 8 imagens (figura 4.2) submetidas ao observador ao se realizar o experimento foi exibida durante 50 segundos, obteve-se como resultado 50 imagens de *frames*, e destas foram extraídas 50 coordenadas de passos sucessivos do foco de atenção do observador.

(a)

(b)

(c)

Figura 4.5 - Seqüência de cinco *frames* utilizando a amostragem utilizada (a) e a correspondente posição do foco de atenção tomando como base apenas a imagem projetada (b). Em (c) é evidenciada de forma visual a o caminho interpretado pelas imagens.

A figura 4.5 fornecesse como exemplo imagens entrada, e seus respectivos resultados. A figura 4.5 (a) mostra uma seqüência de cinco *frames* consecutivos

amostrados, da maneira como foi fornecida a entrada para o algoritmo. A figura 4.5 (b) evidencia como a simulação do campo visual irá tratar essas imagens, como se o foco de atenção estivesse restrito apenas a imagem projetada sobre o anteparo, já evidenciando qual seria a posição do foco de atenção nesta imagem. Figura 4.5 (c) evidencia o caminho interpretado, num arranjo já semelhante a da rede implementada.

Existe como limitação técnica do *eye tracker* utilizado algum ruído eletrônico, caracterizado como um salto para uma região periférica do campo visual (figura 4.1) do filme. Algumas vezes a imagem selecionada como amostragem conhecia com uma imagem do ruído, resultando numa posição inválida do foco de atenção (verificada através da comparação simples entre a coordenada do foco visual e as coordenadas da borda da imagem). Neste caso o algoritmo automaticamente excluía aquele resultado e procurava por um novo *frame*, a dois *frames* da posição original do *frame* amostrado (menos do que um dose avos de segundo). Tal método resolveu os problemas relacionados com o ruído do filme adiantando não mais do que quatro *frames* em aproximadamente 7% das imagens utilizadas para caracterizar o caminho do foco de atenção.

De maneira sucinta, um aparelho filma o campo visual de um observador e adiciona ao filme uma marcação evidenciando o foco de atenção deste. Um algoritmo desenvolvido especialmente para esta finalidade lê tal filme, e obtém automaticamente a coordenada deste foco de atenção. Com base nessa coordenada foi composto o caminho que foco de atenção do observador realizou sobre cada uma de oito imagens distintas pré-selecionadas, resultando assim oito caminhos distintos.

4.2 RESULTADOS DO EXPERIMENTO

A partir do caminho efetuado pelo foco de atenção do observador foi possível construir a estrutura matemática utilizada pelo modelo, uma rede complexa. De início procurou-se pelos vértices da rede, de acordo com o modelo proposto nas seções anteriores, regiões de interesse da imagem. Uma vez que foram utilizadas

imagens sem nenhuma interpretação inicial não é coerente segmentar a imagem de maneira semântica, de modo que as regiões foram separadas por similaridade geográfica, pela distância simples entre elas. Com base nas características biológicas apresentadas nas seções anteriores, sabe-se que as sacadas realizadas pelo olho em geral atingem regiões a grandes distâncias, justificando assim que pontos próximos pertençam à mesma região.

Para definir qual coordenada faria parte de cada região foi implementado um método aglomerativo hierárquico, onde caso dois pontos estiverem a uma distância euclidiana menor do que um valor pré determinado, estes fariam parte da mesma região. Caso um novo ponto também estivesse a uma distância menor que o valor pré-determinado de quaisquer uns dos pontos da região, aquele novo ponto também faria parte de tal região. Tal procedimento foi repetido para todos os pontos do caminho obtido anteriormente, formando assim as regiões que foram definidas como vértices da rede complexa.

Foram construídas redes complexas utilizando como distância de aglomeramento de pontos os valores de 8, 16, 24, 32, 40 e 48 pixels. As redes resultantes podem ser visualizadas na figura 4.6. Os resultados gerados a partir das distâncias 8 e 16 foram desconsiderados, pois nenhuma delas gerava regiões visitadas constantemente pelo observador.

As arestas são definidas pela conexão entre uma região e outra, nesse contexto, uma aresta é definida pela mudança do foco de atenção do observador de uma região para outra. Uma vez que as regiões foram definidas, não foi necessário nenhum tratamento especial para assinalar as arestas.

(a)

(b)

(c)

(d)

(e)

continua

continuação

(f)

(g)

(h)

(i)

(j)

continua

continuação

(l)

Figura 4.6 - Redes complexas sobrepostas às imagens. As redes foram geradas a partir de diversos resultados da aplicação do método aglomerativo sobre o caminho gerado pelo foco visual. A distância de aglomeração de (a), (b), (c) e (d) respectivamente foi de 24, 32, 40 e 48 pixels. As imagens (e-l) também obedecem às mesmas distâncias, representam cada uma das outras sete imagens projetadas para o observador.

A figura 4.6 evidencia claramente qual o resultado topológico sobre a rede da modificação da distância de agrupamento. Ao aumentar a distância de agrupamento verificasse que regiões inicialmente distintas unam-se para formar uma nova região, localizada entre as duas anteriores. Pequenas distâncias fazem com que as regiões de agrupamento consistam de uma ou duas coordenadas apenas. De acordo com as propriedades estatísticas do modelo, é necessário que cada região seja visitada mais de uma vez, resultado este que corrobora com resultados biológicos já discutidos, apenas uma pequena região fazia parte do foco de atenção, porém essa região era composta de mais do que apenas alguns pontos da imagem. Deste modo, da mesma forma que as redes geradas por distâncias 8 e 16, as redes geradas por distância 24 ainda não conseguiam popular de maneira satisfatória a distribuição das regiões.

Macro regiões, regiões de grande tamanho, também apresentam problemas quanto à estrutura geral da rede. Biologicamente, apenas uma pequena região é focalizada sobre a fóvea, de modo que regiões de grande tamanho também devem ser desconsideradas. Tal problema é evidenciado analisando qualitativamente as propriedades topográficas das redes resultantes. Em dois exemplos claros, figura 4.6 (g) e (j), a topologia da rede foi drasticamente alterada quando a rede foi compilada utilizando distância 48, em comparação a rede com distância 40. Desse modo, as redes resultantes da distância 48 pixels também foram desconsideradas.

Por fim, foram utilizadas as redes formadas pelo método aglomerativo utilizando distâncias de 32 e 40 pixels.

Uma vez definidas quais coordenadas pertenciam a cada região, foram assinalados índices a cada uma destas, de modo que se tornasse possível indexar cada uma das regiões. O caminho realizado pelo observador foi então convertido de acordo com esses índices, de modo que agora o caminho era composto pelas regiões - vértices da rede - que o observador estaria visitando ao observar cada uma das imagens, formando assim um caminho dentro da rede complexa modelada. O próximo passo para a caracterização da rede obtida baseada no experimento foi a construção da matriz estocástica W (eq. 3.11). Tal matriz foi construída observando-se a ocorrência de transições a partir de um índice específico de um vértice no caminho percorrido pelo observador, isso é, para um vértice i do caminho, verificou-se cada passo subsequente, e a partir da distribuição de vértices subsequentes de i foi assinalada a probabilidade de transição.

Num experimento contendo mais observadores, este seria o ponto de convergência, a rede complexa resultante de uma imagem seria confeccionada a partir da distribuição de vértices subsequentes de todos os caminhos.

O experimento for fim resultou em 16 redes complexas, uma rede para cada uma das oito imagens projetadas para o observador agrupadas utilizando distancias de aglomeração 32 pixels, e de modo semelhante, uma rede para cada imagem ao utilizar aglomeração de distância 40 pixels.

5 AMBIENTE COMPUTACIONAL

Pode-se definir uma simulação de diversas formas distintas, dentre elas, Pegden (30) a define como: “a simulação é um processo de projetar um modelo computacional de um sistema real e conduzir experimentos com este modelo com o propósito de entender seu comportamento e/ou avaliar estratégias para sua operação”. Para tal é necessário converter um sistema real, mensurado a partir de características reais, para um sistema computacional, utilizando uma linguagem apropriada. Em geral é necessário discretizar grandezas reais em variáveis computacionais.

No modelo aqui proposto foi simulado a atenção seletiva de um observador a partir de uma rede complexa (capítulo 3). Devido à complexibilidade intrínseca do modelo, foi necessária uma simulação computacional, tornando possível o cálculo de todas as dependências necessárias do modelo empregado. O uso do poder computacional também tornou possível a variação das características do modelo, assim como sua eventual prova, tornando possível a evolução do modelo a partir de experimentação. Os detalhes da simulação computacional serão descritos no capítulo 6, aqui apenas serão abordadas características do ambiente computacional utilizado.

Durante as simulações foi utilizado o *software* livre *scilab* (<http://www.scilab.org>). O *scilab* fornece um ambiente computacional onde estão disponíveis diversas subrotinas e funções utilizadas para tratamento matemático, estatístico e gráfico. Existe uma perda considerável ao utilizar o *scilab* em comparação com linguagens usualmente utilizadas para programação científica, como *C/C++* ou *fortran*, porém devido aos protótipos em *scilab* demorarem aproximadamente uma hora para o tempo de execução no pior caso (utilizando uma ordem aproximada de mil vértices), e em melhor caso (redes geradas pelo experimento, tipicamente 25 vértices) aproximadamente um minuto optou-se por não fazer a conversão.

Durante as simulações foram utilizados os computadores do Grupo de Visão Cibernética, e quando necessário foi utilizado o *cluster* do grupo de pesquisa. Tais computadores continham processador Intel®, modelo Pentium® IV de 3,2 GHz, e 2 GB de memória RAM.

6 DESENVOLVIMENTO E
RESULTADOS: SIMULAÇÕES
E O EXPERIMENTO

Com base no modelo apresentado no capítulo 3, foi construído um conjunto de redes complexas teóricas com a finalidade de compará-las às redes complexas experimentais obtidas pelo processo já descrito no capítulo 4.

O modelo propõe que tal rede complexa terá seus vértices relacionados com a saliência local de cada posição da imagem, e suas arestas relacionadas com a probabilidade da mudança da atenção do observador de um vértice para outro. O primeiro passo para o desenvolvimento da rede será a escolha dos seus vértices.

Uma vez que trabalhamos com imagens, a rede criada será estática, não sendo possível criar ou excluir vértices ou arestas. Para dividir a imagem em regiões de interesse, foram inicialmente consideradas três opções. A primeira é a divisão aleatória, distribuindo os vértices sobre as imagens aleatoriamente. Tal proposição foi desconsiderada uma vez que, claramente, ao observar uma imagem existe uma segmentação em função dos elementos presentes, fazendo com que a proposição de que a observação se dê de maneira aleatória torne-se incoerente. Uma segunda opção seria distribuir os vértices uniformemente sobre a região, delimitando assim uma rede regular. O modelo propõe que é possível para um observador focar sua atenção sobre praticamente qualquer ponto da cena, mesmo que estes não sejam necessariamente salientes (neste caso estes vértices seriam muito pouco visitados). Porém tal método envolve um número elevado de vértices, fazendo com que as simulações tornem-se demasiadamente custosas sem acrescentar grande desvio de resultados, uma vez que a maior parte do tempo a atenção estará focalizada nos pontos mais salientes. Desse argumento emerge a terceira opção, de considerar como vértices da rede apenas as regiões da figura de maior saliência.

Uma vez que procurávamos a confirmação do modelo teórico a partir do experimento proposto, utilizaram-se como vértices as regiões salientes definidas pelo experimento, a fim de preservar a topologia geral das redes geradas, e de simplificar a comparação entre elas. Neste caso, pode-se associar vértices às regiões não visitadas no experimento de modo que toda a imagem é mapeada sobre vértices da rede complexa. Porém tais vértices não terão probabilidade de mudança, sendo estes desconectados da rede.

Também foram utilizadas para confeccionar as redes complexas teóricas as mesmas imagens utilizadas no experimento, conforme a figura 4.2. Quando se tornar conveniente tratar da imagem computadorizada, esta será chamada de $\text{Im}(x, y)$, conforme seção 2.5.

Se definido os vértices é necessário procurar pelas arestas, e pelos seus respectivos pesos. Conforme debatido na seção 3.1, as arestas são dadas pela função $h_{(i \rightarrow j)}$ (equação 3.13), sendo que esta é dependente do conjunto F das propriedades locais $\vec{f}_i$, conforme as equações 3.5 e 3.6.

As propriedades locais $\vec{f}_i$ foram escolhidas de modo a corroborar com os resultados apresentados na seção 2.4, tais propriedades devem ter a característica de chamar a atenção de um observador. Utilizamos aqui três propriedades básicas: Níveis de Cinza, o módulo do gradiente e o desvio padrão dos níveis de cinza. Os níveis de cinza podem ser caracterizados importantes quando o observador segmenta um objeto do seu fundo, enquanto que o módulo do gradiente e o desvio padrão dos níveis de cinza possuem campos receptivos seção 2.3 sensíveis a estas características.

Porém tais propriedades não podem ser calculadas na posição média da região que representa um vértice, tais propriedades tendem a mudar drasticamente avançando poucos pixels. Tampouco é coerente fazê-lo, uma vez que cada vértice representa a capacidade de uma região chamar a atenção, não apenas a de um ponto. Deste modo é necessário fazer o cálculo de $\vec{f}_i$ numa região de interesse que caracterize a região correspondente da imagem. Desta forma define-se $r_{propriedads}$, como o raio de alcance das propriedades, cada uma das propriedades é calculada sobre a região delimitada por um círculo de raio $r_{propriedads}$, e então $\vec{f}_i$ será caracterizado pela média de cada uma das propriedades sobre esta região.

A partir do tamanho médio dos grupos formados no experimento, adotou-se como valor do raio de alcance das propriedades:

$$r_{propriedads} = 50 \text{ pixels}$$

A média dos níveis de cinza na região delimitada por $r_{propriedads}$ será definida como $\vec{f}_i(1)$, a média do módulo do gradiente nesta mesma região será $\vec{f}_i(2)$, e o desvio padrão do nível de cinza nesta região será denominado $\vec{f}_i(3)$.

Observe que estaremos utilizando por simplificação regiões circulares definidas por $r_{propriedads}$. Devido ao fato de que os vértices da rede teórica serem os mesmos vértices do experimento, é provável que exista sobreposição das regiões utilizadas, como pode ser observado na figura 6.2.

Uma vez determinado o conjunto F é possível estudar o comportamento de $h_{(i \rightarrow j)}$ a partir de diversas combinações dos elementos de F . Um caso de estudo seria a seguinte combinação:

$$h'_{(i \rightarrow j)} = \frac{\vec{f}_j(1)}{\vec{f}_i(1)} \quad (6.1)$$

Observe que neste caso estamos utilizando $h'_{(i \rightarrow j)}$ para denotar a função desnormalizada, a função que relaciona a probabilidade de mudança $h_{(i \rightarrow j)}$ é dada pelo fator normalização $h_{(i \rightarrow j)} = \alpha h'_{(i \rightarrow j)}$ tal que $\sum_i^N \alpha h'_{(i \rightarrow j)} = 1$, sendo N o total de vértices de rede e α uma constante real de proporcionalidade.

Tal configuração da equação 6.1 permite a valorização de regiões com níveis de cinza mais altos, uma vez que $h'_{(i \rightarrow j)}$ terá um maior que 1 apenas se o valor de $\vec{f}_j(1)$ seja maior que $\vec{f}_i(1)$. A mesma forma da eq. 6.1 poderia ser estendida para quaisquer propriedades de F . Porém tal forma de $h'_{(i \rightarrow j)}$ introduz um novo problema, um sistema sob condições específicas poderia gerar uma função $h'_{(i \rightarrow j)}$ indeterminada no caso da propriedade no denominador for nula (o que resultaria numa divisão por zero). Tal configuração não é rara, uma vez que comumente existem regiões constantes na imagem, fazendo com que $\vec{f}_i(2)$ nesta região seja zero.

Ao reformular $h'_{(i \rightarrow j)}$ é necessário que esta obedeça algumas propriedades, como ser livres de valores indeterminados independente das configurações de F , valorizar regiões ditas salientes, enquanto que marginaliza regiões de baixo interesse. Uma solução simples para tal problema seria:

$$h'_{(i \rightarrow j)} = \vec{f}_j(1) \quad (6.2)$$

Tal forma é trivialmente livre de indeterminações caso o conjunto F também o seja, assim como naturalmente prioriza regiões salientes. Esta forma também marginaliza regiões de baixo interesse, porém de uma maneira indireta: uma região com valores altos de $\vec{f}_i$ terá uma alta probabilidade de manter ali a atenção de um observador, enquanto que uma região de baixos valores não chamará atenção, de acordo com a normalização $\sum_i^N \alpha h'_{(i \rightarrow j)} = 1$, sendo $h_{(i \rightarrow j)} = \alpha h'_{(i \rightarrow j)}$. Com base nesta definição da eq. 6.2 é possível introduzir uma evolução no método utilizado para combinar F formando $h_{(i \rightarrow j)}$, a construção intermediária de uma nova função, a $S(x, y)$.

Define-se $S(x, y)$ como o produto de algumas das funções que formam $\text{Im}(x, y)$, porém estas não estão inicialmente restritas às regiões de interesse, mas sim sobre toda a imagem tratada de $\text{Im}(x, y)$. A função $S(x, y)$ será denominada como saliência computacional, sua importância será discutida adiante. Foram utilizadas cinco formas distintas de $S(x, y)$.

$$S(x, y) = \text{Im}(x, y) \quad (6.3)$$

$$S(x, y) = \text{Grad}(x, y) \quad (6.4)$$

$$S(x, y) = \text{Grad}(x, y) \text{Im}(x, y) \quad (6.5)$$

$$S(x, y) = \text{Des}(x, y) \quad (6.6)$$

$$S(x, y) = \text{Grad}(x, y) \text{Des}(x, y) \quad (6.7)$$

Onde:

$Grad(x, y)$: É o gradiente de $Im(x, y)$, calculado a partir do operador de Sobel.

$Des(x, y)$: É o desvio padrão de uma região circular de $Im(x, y)$, cujo raio é

$r_{propriedades}$.

A saliência computacional pode ser interpretada como a versão computacional da saliência básica de objetos, conforme discutido na seção 2.4. Assim, tal função seria a conexão entre cada uma das propriedades que formam a saliência de uma imagem, e da forma aqui definida, $S(x, y)$ terá um alto valor nas regiões de $Im(x, y)$ que chamam naturalmente a atenção do observador. A principal vantagem da definição da saliência computacional é que se torna possível visualizar a saliência trabalhada, viabilizando a inspeção supervisionada do modo que as propriedades de F combinam-se.

A figura 6.1 exemplifica utilizando a quinta imagem (a quinta imagem da figura 4.2) trabalhada no experimento, as saliências computacionais obtidas pelas cinco combinações. A figura 6.1 (a) mostra a imagem original, enquanto (b) evidencia a saliência computacional calculada com base na eq. 6.3, (c) calculada com base na eq. 6.4, (d) de acordo com a eq. 6.5, assim como (e) com a eq. 6.6 e (f) com a eq. 6.7.

(a)

(b)

continua

continuação

(c)

(d)

(e)

(f)

Figura 6.1 - Saliências computacionais obtidas utilizando diversas combinações de parâmetros. Em (a) evidencia-se a figura utilizada para o cálculo das saliências, (b) é a saliência computacional calculada utilizando-se como parâmetro os níveis de cinza, (c) esta foi calculada utilizando o gradiente da imagem, (d) a saliência foi baseada no produto entre o gradiente e o nível de cinza, em (e) o cálculo foi feito utilizando o desvio padrão de cada região, e (f) a saliência foi dada a partir do produto desvio padrão com o gradiente.

Desta vez $h'_{(i \rightarrow j)}$ será distribuída de maneira proporcional à média de $S(x, y)$ sobre a região delimitada pelo $r_{propriedades}$. A figura 6.2 evidencia as regiões onde as médias foram feitas a partir da saliência computacional da eq. 6.5, a figura 6.1 (d). A figura 6.2 mostra cada uma das regiões formadas utilizando círculos vermelhos. No centro de cada região foi impresso um novo círculo com um nível variável de

intensidade, tal círculo informa o nível de cinza proporcional a média da saliência computacional da região em questão.

Figura 6.2 - Imagem 5 sobreposta aos círculos vermelhos que evidenciam cada uma das regiões onde serão calculadas as médias da saliência computacional. No centro de cada região foi impresso um círculo cujo nível de cinza é proporcional à média da saliência computacional naquela região.

A função $h_{(i \rightarrow j)}$ então será definida como:

$$h_{(i \rightarrow j)} = \alpha \iint_{\text{região } i} S(x, y) dx dy \quad (6.8)$$

$$\sum_{i=1}^N \alpha \iint_{\text{região } i} S(x, y) dx dy = 1 \rightarrow \alpha = \frac{1}{\sum_{i=1}^N \iint_{\text{região } i} S(x, y) dx dy} \quad (6.9)$$

A integral da eq. 6.8 é facilmente calculada computacionalmente a partir de qualquer $S(x, y)$ dada, enquanto que a constante α é obtida pela normalização da eq. 6.9.

Tal versão de $h_{(i \rightarrow j)}$ ainda pode ser melhorada, uma vez que outra característica biológica até então não foi trabalhada. O cérebro não interpreta a

informação visual de modo uniforme, pois existe uma maior acuidade na região da fóvea, conforme explorado na seção 2.2. Esta característica torna pouco coerente utilizar a função $S(x, y)$ sem qualquer dependência de qual ponto o observador foca sua atenção. Foi proposta então uma nova versão para $S(x, y)$ levando em consideração a posição geográfica que o observador foca sua atenção quando se procura um novo vértice para a mudança de atenção.

A partir do experimento realizado, procurou-se a distribuição das mudanças de atenção a partir da distância entre o vértice atual, e o próximo vértice o qual o observador irá focalizar, caracterizando assim a distribuição das distancias médias das sacadas efetuadas pelo observador. Os histogramas resultantes encontram-se compilados no gráfico 1.

continua

continuação

(a)

continua

continuação

(b)

Gráfico 1 - Histogramas baseados nas distancias das sacadas realizadas pelos olhos do observador no experimento realizado. O eixo x representa a distancia da sacada, enquanto que o y a freqüência relativa destas distâncias. Em (a) observa-se as sacadas com base no método aglomerativo de 32 pixels, enquanto que (b) são os resultados a partir da aglomeração de 40 pixels.

A análise dos histogramas dita à existência de uma alta probabilidade da observação permanecer no mesmo vértice, conforme a alta incidência de sacadas de tamanho médio zero. Outros comportamentos distintos são picos menores a uma

distância média do ponto de observação, assim como uma distribuição relativamente uniforme ao longo de grandes distâncias do vértice atual.

Foi proposta a presença de três comportamentos distintos sobrepostos nos histogramas, onde aqui serão modelados a partir da composição de três funções gaussianas distintas. Deste modo, a função densidade de probabilidade (f.d.p.) que descreve o comportamento dos histogramas presentes em gráfico 1 será dada por $f(d)$, onde d é a distancia da sacada, sendo $f(d)$ composta pela soma de três funções gaussianas distintas $g_{zero}(d)$, $g_{medio}(d)$ e $g_{longo}(d)$, de modo que $f(d) = g_{zero}(d) + g_{medio}(d) + g_{longo}(d)$. A primeira delas propõe uma alta probabilidade da sacada não existir, fazendo com que o observador permaneça no vértice atual, de modo que se utilizou como parâmetro para $g_{zero}(d)$ uma média $\mu = 0$ pixels e um desvio padrão $\sigma = 10$ pixels. A segunda delas, $g_{medio}(d)$, compreende a situação de uma sacada padrão, o observador muda o seu foco de atenção para um ponto próximo a fim de continuar explorando a imagem. Para esta utilizou-se como parâmetros os valores de $\mu = 80$ pixels e $\sigma = 30$ pixels, uma vez que tal distribuição deve estar a uma distancia média de zero, e conforme evidenciado em gráfico 1, possui um desvio padrão considerável. Já $g_{longo}(d)$ é interpretado como uma probabilidade do observador se desinteressar completamente pela região visitada, fazendo assim uma sacada de longa distância. Os parâmetros utilizados foram $\mu = 200$ pixels e $\sigma = 300$ pixels.

Definindo-se a função $G_i(x,y)$ como a versão bivariada da função $f(d)$, pode-se determinar a nova saliência computacional, esta levando em consideração a posição geográfica entre os vértices origem e destino da sacada. Note que $G_i(x,y)$ possui uma nova dependência, a região i que o observador se encontra. Tal função fará com que a saliência computacional também se torne dependente da região i , formando assim não uma única saliência computacional para uma imagem, mas uma saliência distinta de acordo com a região observada. A figura 6.3 mostra de forma gráfica $f(d)$ e $G_i(x,y)$ calculada sobre um determinado vértice, assim como a saliência computacional resultante para aquela região. A saliência computacional final será o produto de $S(x,y)$ por $G_i(x,y)$.

(a)

(b)

(c)

Figura 6.3 - Gráfico da f.d.p. utilizada para modelar os histogramas utilizados no experimento (a), sendo o eixo x a distancia da sacada, enquanto que no eixo y é graficada a probabilidade de mudança. A função correspondente (b) circular calculada sobre uma região mapeada como um vértice da rede, que será utilizada como máscara para compor a saliência computacional. Em (c) pode-se observar a saliência computacional gerada a partir da Figura 6.1 (c), o módulo do gradiente, e também utilizando tal máscara de distância.

Observe que a figura 6.3 (c) ainda mantém suas características básicas observadas na figura 6.1 (c), porém estas foram modularizadas pela função apresentada em figura 6.3 (b).

Por fim, a função $h_{(i \rightarrow j)}$ poderá ser calculada:

$$h_{(i \rightarrow j)} = \alpha \iint_{\text{região } i} S(x, y) G_i(x, y) dx dy \quad (6.10)$$

$$\sum_{i=1}^N \alpha \iint_{\text{região } i} S(x, y) G_i(x, y) dx dy = 1 \rightarrow \alpha = \frac{1}{\sum_{i=1}^N \iint_{\text{região } i} S(x, y) G_i(x, y) dx dy} \quad (6.11)$$

De posse de $h_{(i \rightarrow j)}$ pode-se construir trivialmente a matriz W , de acordo com a equação 3.10. A matriz c do experimento 5, utilizando a saliência computacional da eq. 6.5, pode ser visualizada como figura na figura 6.4, porém para tal figura teve seu nível de cinza renormalizado para o valor máximo permitido, melhorando a distinção entre os níveis de cinza.

Figura 6.4 - Figura que mapeia a matriz W resultante da Figura 2.4, utilizando uma saliência computacional dada pelo produto dos níveis de cinza com o módulo do gradiente da imagem. Tal matriz foi renormalizada no valor máximo possível para impressão de modo a facilitar a visualização. Pode-se observar um alto valor de auto conexão, assim como a presença de vértices preferenciais.

Observe que a diagonal da matriz representada na figura 6.4 é clara, evidenciando alta probabilidade da auto conexão do vértice observado, resultado consistente com o esperado, de acordo com os argumentos já discutidos. Também se pode observar a probabilidade maior de conexão em nós esparsos, evidenciando

que cada vértice de fato exibe mudanças preferenciais para determinados vértices, de acordo com a distância preferencial de sacadas, além da saliência evidente de cada região. Outra característica da matriz W é o seu auto-vetor associado ao seu auto valor 1, conforme a eq. 2.12, este auto-vetor corresponderá à frequência de visitação de cada um dos vértices, caso o observador fique muito tempo observando a cena – neste caso percorrendo a imagem. Embora tal medida não tenha sido utilizada na presente dissertação, ela poderia ser utilizada para comparar o resultado obtido pelo modelo com experimentos que verificam apenas a frequência de visitas de uma determinada região.

Foi formado então um conjunto de 80 matrizes consistindo nas combinações das seguintes características: forma de saliência computacional (cinco formas), imagens utilizadas no experimento (8 imagens ao todo) e distância de agrupamento (foram consideradas duas distâncias, 32 e 40 pixels).

Uma vez prontas as redes complexas teóricas foi possível comparar a distância de cada um dos modelos de saliência computacional com relação às redes complexas geradas pelo experimento. Para tal foi calculada a distância euclidiana total entre cada uma das redes em relação à rede experimental, utilizando-se a seguinte fórmula:

$$d_{i,j} = \sum_{k=1}^N \sum_{l=1}^N \sqrt{(w_{k,l}^{i,j} - w_{\text{exp},k,l}^j)^2} \quad (6.12)$$

Onde:

$d_{i,j}$: É a distância entre as matrizes estocásticas $W^{i,j}$ e W_{exp}^j , sendo i o índice do método de obtenção da saliência computacional utilizado, e a j o índice da imagem utilizada.

$w_{k,l}^{i,j}$: Elemento indexado pela linha k e coluna l da matriz $W^{i,j}$.

$w_{\text{exp},k,l}^j$: Elemento indexado pela linha k e coluna l da matriz W_{exp}^j .

As matrizes $d_{i,j}$ compiladas a partir de $d_{i,j}$ encontram-se a seguir. Estas foram normalizadas em função do seu menor valor a fim de salientar o melhor valor

obtido, assim como facilitar comparações entre os métodos obtidos. O índice superior indica qual método de agrupamento foi utilizado ao fazer a comparação:

$$d^{32} = \begin{pmatrix} 1,0010 & 1,0000 & 1,0096 & 1,0127 & 1,0000 & 1,0704 & 1,0012 & 1,0834 \\ 1,0150 & 1,0179 & 1,0009 & 1,0262 & 1,0583 & 1,0398 & 1,0096 & 1,0000 \\ 1,0132 & 1,0002 & 1,0011 & 1,0000 & 1,0548 & 1,0000 & 1,0089 & 1,0405 \\ 1,0000 & 1,0089 & 1,0081 & 1,0194 & 1,0097 & 1,0692 & 1,0000 & 1,0604 \\ 1,0139 & 1,0064 & 1,0000 & 1,0033 & 1,0327 & 1,0116 & 1,0024 & 1,0239 \end{pmatrix}$$

$$d^{40} = \begin{pmatrix} 1,0032 & 1,0165 & 1,0032 & 1,0209 & 1,0000 & 1,0480 & 1,0055 & 1,0576 \\ 1,0237 & 1,0354 & 1,0000 & 1,0401 & 1,1063 & 1,0397 & 1,0136 & 1,0000 \\ 1,0210 & 1,0000 & 1,0024 & 1,0000 & 1,1070 & 1,0000 & 1,0162 & 1,0313 \\ 1,0000 & 1,0348 & 1,0029 & 1,0189 & 1,0180 & 1,0502 & 1,0000 & 1,0423 \\ 1,0213 & 1,0154 & 1,0023 & 1,0015 & 1,0812 & 1,0129 & 1,0050 & 1,0166 \end{pmatrix}$$

Um pequeno valor de um elemento de d^{32} ou d^{40} mostraria que aquela saliência computacional fornece uma rede complexa mais próxima à rede complexa gerada a partir do experimento, e neste caso, os menores valores foram normalizados em 1.

Porém pela análise de d^{32} e d^{40} não foi possível observar um método absoluto como menor valor de diferença entre as redes. O método da eq. 6.3 teve melhor acerto em duas imagens no agrupamento 32, e melhor desempenho em uma imagem quando o agrupamento era 40; O método da eq. 6.4 teve um rendimento mínimo, e um muito próximo do mínimo em d^{32} , obtendo também dois valores mínimos em d^{40} . O terceiro método da eq. 6.5, conseguiu melhor desempenho em dois casos em d^{32} , e em três em d^{40} . Já o método da eq. 6.6 conseguiu o melhor rendimento em dois casos nos dois agrupamentos. O método da eq. 6.7 conseguiu melhor resultado em um método no agrupamento 32, e não conseguiu melhor resultado no caso do agrupamento 40.

Tais resultados poderiam apontar o método de formação da saliência computacional de acordo com a equação 6.5 como o método que produz os melhores resultados, enquanto a eq. 6.7 apontaria o método como o que produz os piores. Porém o método resultante da eq. 6.7 produziu resultados próximos de 1,

sendo este inclusive o método que tem a menor soma dos erros ao comparar todos os métodos utilizados a todas as imagens.

Estes resultados poderiam apontar uma falha nos métodos utilizados para construir as saliências computacionais utilizadas. Antes é conveniente comparar diretamente d^{32} e d^{40} procurando pelos motivos das divergências encontradas. Embora não existam muitas diferenças quanto aos resultados individuais de cada modelo de saliência computacional, pode-se comparar o resultado individual de cada modelo de para o caso da aglomeração com distâncias 32 e 40. De uma forma geral, ao comparar o desempenho dos modelos na mesma imagem, pode-se notar diferenças percentuais, porém essas diferenças são sempre maiores em d^{40} . A diferença entre d^{32} e d^{40} conceitualmente é a distancia a qual os pontos que foram resultados do experimento foram agrupados. A distância 40 agrupou pontos que a distância 32 não conseguiu, produzindo assim vértices que representavam regiões maiores, fazendo com que cada vértice na rede complexa resultante fosse mais visitado. De maneira oposta, o espaço amostral do agrupamento 32 se mostrou menos denso do que o de 40, tornando assim qualquer método estatístico (como o aqui utilizado) menos confiável.

Essa limitação evidencia que é necessário, antes de propor novos métodos para saliência computacional, deve-se procurar adequar melhor à rede complexa teórica à verificada experimentalmente. Para isto procurou-se limitar mais a rede teórica, baseando-se à experimental, tornando-a mais próxima possível a topologia das duas redes.

Limitou-se então a conectividade dos vértices da rede teórica, baseando-se na rede experimental. Dois vértices da rede complexa teórica só poderiam se conectar caso exista tal conexão na rede experimental. Utilizando tal técnica, fez-se com que a distribuição de probabilidades não se desse sobre todos os vértices da rede como até então, porém dividida entre poucos vértices, fazendo assim com que a disputa por altos valores de saliência computacional naquelas regiões se tornasse mais acirrada.

Foram geradas novamente novas funções $h_{(i \rightarrow j)}$, porém estas possuem um novo termo $\delta_{i,j}$, sendo que este será 1 caso $W_{\text{exp}} > 0$ para aquela imagem, ou seja, caso exista probabilidade de mudança de atenção entre estes dois nós na rede complexa experimental, e será zero caso contrário:

$$h_{(i \rightarrow j)} = \delta_{i,j} \left(\alpha \iint_{\text{região } i} S(x, y) G_i(x, y) dx dy \right) \quad (6.13)$$

A normalização ocorre da mesma forma da eq. 6.9.

A partir de $h_{(i \rightarrow j)}$ gerou-se novamente o conjunto de matrizes W , assim como as distâncias para os dois métodos aglomerativos, d^{32} e d^{40} :

$$d^{32} = \begin{pmatrix} 1,0371 & 1,0248 & 1,0001 & 1,0000 & 1,0322 & 1,0278 & 1,0191 & 1,0080 \\ 1,0000 & 1,0359 & 1,0428 & 1,0828 & 1,0000 & 1,0086 & 1,0123 & 1,0035 \\ 1,0064 & 1,0484 & 1,0000 & 1,0429 & 1,0202 & 1,0000 & 1,0000 & 1,0091 \\ 1,0367 & 1,0000 & 1,0014 & 1,0265 & 1,0247 & 1,0292 & 1,0247 & 1,0000 \\ 1,0023 & 1,0262 & 1,0004 & 1,0568 & 1,0127 & 1,0038 & 1,0004 & 1,0026 \end{pmatrix}$$

$$d^{32} = \begin{pmatrix} 1,0096 & 1,0153 & 1,0253 & 1,0000 & 1,0204 & 1,0234 & 1,0209 & 1,0095 \\ 1,0000 & 1,0348 & 1,0110 & 1,0908 & 1,0304 & 1,0091 & 1,0172 & 1,0096 \\ 1,0021 & 1,0319 & 1,0030 & 1,0457 & 1,0000 & 1,0000 & 1,0000 & 1,0145 \\ 1,0086 & 1,0000 & 1,0235 & 1,0284 & 1,0178 & 1,0259 & 1,0267 & 1,0000 \\ 1,0006 & 1,0176 & 1,0000 & 1,0606 & 1,0045 & 1,0045 & 1,0014 & 1,0068 \end{pmatrix}$$

Este procedimento novamente não detectou um método mais eficiente em todos os casos, porém permitiu que o método 3 do cálculo da saliência computacional baseado na eq. 6.5, que descreve tal propriedade como o produto entre o módulo do gradiente e o nível de cinza obteve melhor resultado em 3 imagens nos dois métodos de aglomeração, e conseguiu desvio relativo menor do que 1% em outros dois casos, ficando próximo do menor valor possível da distancia entre a rede teórica e a experimental em 5 das 8 imagens utilizadas.

Outro método de destaque foi o baseado na eq. 6.7, o produto do módulo do gradiente com o desvio padrão do nível de cinza, embora este método não tenha obtido o melhor caso na maior parte das figuras, seu desvio acumulado foi o menor obtido, seu valor da distância em geral também foi menor do que 1% em 5 das 8

imagens utilizadas. Este resultado expõe o módulo do gradiente como característica chave na determinação da saliência computacional, porém está não é suficiente para descrevê-la completamente, conforme os resultados obtidos pela utilização pura do módulo do gradiente como parâmetro (eq. 6.4).

7 CONCLUSÃO

A interpretação humana compreende enorme peso ao sentimento de humanidade – sentir-se humano. Interpretar fatos, retirar conclusões e prever resultados fazem parte da existência do indivíduo como ser humano. A interpretação visual, por sua vez, embora presente em diversos seres vivos, atingiu seu ápice na espécie humana, recebendo esta enorme peso na interpretação geral do mundo em que vivemos.

Diversos trabalhos se concentraram em descrever características do sistema visual humano, e em especial à atenção seletiva – uma estratégia utilizada pelo sistema visual com o fim de minimizar o tempo de interpretação de uma cena. Porém, até então todos os trabalhos preocupavam-se em descrever qual era a frequência de visita de qualquer região interessante da cena, sem preocupar-se qual seria a dinâmica da atenção seletiva durante a sua interpretação. Na presente dissertação desenvolveu-se um modelo biologicamente plausível que interpretava a atenção seletiva como uma rede complexa.

As redes complexas tornaram-se evidentes ao modelar diversos sistemas reais em áreas do conhecimento distinto, utilizando conceitos matemáticos simples, porém robustos. Nesta dissertação a atenção seletiva foi modelada imaginando que os pontos da cena para onde um observador pudesse voltar sua atenção sejam vértices de uma rede complexa, e a possibilidade da mudança de atenção entre dois vértices – dois pontos importantes da cena – sejam arestas.

A distribuição dos vértices e arestas, seus detalhes implementativos e características delimitaram o escopo deste trabalho. Como protótipo para tal modelo foi realizado um experimento utilizando-se um *eye tracker*, aparelho este capaz de identificar o foco da atenção de um observador que utiliza tal equipamento. Comparando-se as características do experimento com as do modelo foi possível refinar os elementos utilizados, na tentativa de simular-se de maneira mais próxima a atenção seletiva do observador.

Para a concepção do modelo foi necessário o estudo de características biológicas do sistema visual humano. O conhecimento da estrutura básica dos fotorreceptores da retina, formando a fóvea é fundamental para a atenção seletiva, sendo esta a justificativa para a estrutura do modelo – o olho de fato faz saltos

discretos através de regiões de da imagem, tal qual a mecânica de um caminhar aleatório numa rede complexa.

Também foram estudados alguns campos receptivos simples de neurônios, tornando possível aproximar o modelo matemático apresentado à sua inspiração biológica. Foi estudada também a saliência visual, característica essa que faz com que objetos distintos das cenas tornem-se salientes para um observador, fazendo com que sua atenção seja concentrada em tais objetos.

Fez parte do escopo o estudo de características estatísticas e computacionais para desenvolver o modelo, como propriedades das cadeias de Markov, estruturas das redes complexas, assim como o estudo de visão computacional, tais quais suas propriedades e medidas.

Através da intersecção das áreas estudadas, campos receptivos, atenção seletiva, saliência visual, visão computacional, cadeias de Markov e redes complexas, construiu-se o modelo, integrando naturalmente as áreas – a atenção seletiva foi modelada por uma rede complexa, biologicamente inspirada. O modelo, a partir de uma imagem em níveis de cinza, simulava a atenção seletiva de um observador.

Os vértices da rede proposta foram dados de acordo com a saliência visual das imagens utilizadas. Uma alta saliência significava que a atenção do observador iria ser atraída por aquele vértice. Foram utilizadas técnicas de visão computacional a fim de retirar das imagens selecionadas a saliência visual.

As arestas, tal qual numa cadeia de Markov, eram dadas pelas probabilidades de mudança entre dois vértices da rede. Tal probabilidade também foi inspirada nas características biológicas da atenção seletiva, como os campos receptivos, e a saliência visual – quanto maior a saliência de uma região maior seria a probabilidade do observador que está em outra região dirigir sua atenção para a região original.

A partir da integração natural de toda a literatura utilizada, concluiu-se que é possível simular a saliência visual a partir de uma rede complexa, levando em consideração grande parte das características biológicas envolvidas, resultando num modelo cuja mecânica é muito próxima da mecânica real da atenção seletiva. Deste

modo, foi feito um experimento procurando refinar ao máximo o modelo utilizado, procurando diminuir a distância entre a rede complexa teórica proposta, e uma eventual rede real.

A partir dos resultados do experimento foi possível construir uma rede complexa tal qual a rede construída pelo modelo, sendo possível assim uma comparação direta.

As simulações foram feitas de maneira progressiva, sendo adicionadas características biológicas ao longo da sua construção. De acordo com o modelo foi definida uma função $h_{(i \rightarrow j)}$, sendo esta relacionada com a probabilidade de transição entre dois vértices. A partir da capacidade de alguns objetos naturalmente chamar a atenção a partir de suas características básicas, concluiu-se que $h_{(i \rightarrow j)}$ deveria ser proporcional a esta característica, sendo esta mensurada no como a saliência computacional. Assim $h_{(i \rightarrow j)}$ foi descrito como algo proporcional ao valor da saliência computacional de uma determinada região.

Foram propostos cinco tipos de saliência computacional, baseadas nas características comumente conhecidas de saliência visual, sendo os métodos para determinação da saliência computacional: proporcional ao nível de cinza de uma região, proporcional ao módulo do gradiente, ao produto nível de cinza pelo gradiente, pela média do desvio padrão da região, assim como pelo produto gradiente pela média do desvio padrão.

Com base na distribuição de fotorreceptores da retina, foi necessário adicionar a saliência computacional um novo termo, sendo que este levava em consideração a posição relativa dos vértices da rede. Para tal foi feito o histograma das distancias relativas das sacadas sobre os resultados obtidos do experimento. Tal histograma revelou três regimes distintos, levando à conclusão de que a atenção seletiva tem três comportamentos: a atenção continua focada no mesmo vértice, provavelmente ainda recolhendo informações importantes, ou ela faz uma sacada de média distância, procurando por um novo vértice, porém este está a uma distância média do original, provavelmente o observador procurou por vértices próximos ao original procurando por informação nova, porém correlacionada com a anterior, ou então uma sacada de longo alcance, procurando por informações completamente

novas da cena. Tal comportamento foi introduzido no modelo na forma de uma máscara que modela a saliência determinada anteriormente.

Uma vez determinadas as redes complexas teóricas resultantes de cada uma das formas de saliência computacional, comparou-se cada uma destas com as formadas pelo observador no experimento. Apesar de alguns métodos se destacarem, não foi possível determinar um método ótimo. Comparando-se duas topologias distintas de formação da rede experimental (utilizando um agrupamento de distâncias 32 e 40 pixels para se definir as regiões no experimento) observou-se uma melhora no rendimento quando eram formadas regiões maiores. Pode-se concluir que a mudança na densidade estatística utilizada tem influência direta sobre a comparação entre as redes.

Procurou-se então restringir a topologia da rede teórica de modo que esta ficasse ainda mais próxima da rede experimental. Tal ação promoveu o destaque de dois métodos de determinação da saliência computacional: no primeiro, utilizando o produto entre o módulo do gradiente e o nível de cinza, e, no segundo o produto entre o módulo do gradiente e a média do desvio padrão. Este primeiro método ficou com desvio menor do que 1% em relação ao melhor método em 5 das 8 imagens utilizadas. O segundo conseguiu o menor desvio entre a rede teórica e a experimental ao contabilizar todas as 8 imagens.

Conclui-se, com esse resultado, que o módulo do gradiente é de fato um fator chave para o cálculo da saliência computacional, porém tal característica sozinha não é de grande interesse, uma vez que ao utilizar apenas o módulo do gradiente para o cálculo da saliência computacional não resultou em bons resultados. A importância do módulo do gradiente pode ser associada à presença de diversos tipos de campos receptivos que são sensíveis a alguma forma de mudança (gradiente).

Finalmente, conclui-se que é possível simular a atenção seletiva a partir de uma rede complexa, assim como é possível a comparação do modelo proposto com um experimento. Porém a consistência do resultado obtido é abalada pelo fato do espaço amostral não ser corretamente populado – é necessário maior número de

observadores ao realizar o experimento, de modo que os algoritmos aqui apresentados servem de protótipos para futuras investigações.

Esta conclusão prevê claramente a possibilidade de trabalhar futuramente com uma maior densidade amostral, evidenciando a relevância deste trabalho para futuras investigações. Outras modificações que fugiram do escopo desta dissertação de mestrado, porém evidenciam claro clamor por investigação é a presença da memória, assim como utilizar imagens com características semânticas.

Introduzir a memória no modelo significaria que durante um passeio aleatório pela rede complexa, a saliência computacional de um vértice visitado naturalmente diminuiria, fazendo assim com que vértices salientes tornem-se desinteressantes com o tempo, à medida que o observador já adquiriu informações suficientes sobre ele. Utilizar imagens que contenham significado, como faces e pessoas, poderia ser naturalmente introduzindo um valor modificado da saliência para cada componente semântico da cena. Porém, para viabilizar tal abordagem é necessário desenvolver segmentadores eficientes de imagens, de forma a tornar possível o cálculo da saliência computacional.

REFERENCIAS

1 MARR, D. **Vision a computational investigation into the human representation and processing of visual information**. San Francisco: W.H. Freeman, 1982

2 SARAMAGO, J. **Ensaio sobre a cegueira**. São Paulo: Companhia das Letras, 2008.

3 REALE, G.; ANTISERE, D.; STORNILOLO, I. (trad). **História da filosofia**. 4a. ed. São Paulo: Paulus, 2009

4 NEWMAN, M. E. J. The structure and function of complex networks. **SIAM Reviews**, v. 45, n. 2, p. 167 – 256, 2003.

5 EULER, L. Solutio problematis ad geometriam situs pertinentis. **Commentarii Academiae Scientiarum Imperialis Petropolitanae**, v. 8, p. 128 – 140, 1741.

6 JEONG, H.; MASON, S.; BARABÁSI, A.-L.; OLTVAI, Z. N. Lethality and centrality in protein networks, **Nature**, v. 411, n. 6833, p. 41-42, 2001.

7: NUSSENZVEIG, H. M. **Curso de física básica** ótica, relatividade, física quântica. São Paulo: Edgard Blucher, 2000.

8 GUYTON, A. C.; HALL, J. E. **Tratado de fisiologia médica**. 9a. ed. Rio de Janeiro: Guanabara Koogan, 1997.

9 CASTRO, S. V. **Anatomia fundamental**. 3a. ed. São Paulo: Mcgraw-Hill, 1985.

10 LEIBOVIC, K N (ed). **Science of vision**. New York: Springer-Verlag, c1990

11 TREISMAN, A. Features and objects in visual processing. **Scientific American**, Vol. 255, n. 5, p. 106-118, 1986.

12 GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. 3nd. ed. Upper Saddle River: Pearson/Prentice Hall, 2008.

13 COSTA, L. da F.; CESAR, R. M. **Shape classification and analysis theory and practice**. 2nd ed. Boca Raton: CRC Press, 2009.

14 ROSS, S. M. **A first course in probability**. 8nd. ed. Upper Saddle River: Pearson Prentice Hall, 2010.

15 RESNICK, S. I. **Adventures in stochastic processes**. Boston: Birkhäuser, 2002.

16 FREEDMAN, D. **Markov chains**. San Francisco: Holden-Day, 1971.

17 MILGRAM, S. The small world problem. **Psychology Today**, v. 2, p. 60 – 67, 1967.

18 NEWMAN, M. **Networks**: an introduction. Oxford: Oxford University Press, 2010

19 BRANDES, U.; ERLEBACH, T.(Eds.) **Network analysis**: methodological foundations. Berlin: Springer Verlag, 2005. Cap.2, p. 7-15. ISBN 3-540-24979-6 (Lecture Notes in Computer Science, v.3418)

20 ROCHA, L. E. C. da **Redes acopladas: estrutura e dinâmica**. 2007. 123 f. Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2007.

21 COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G.; BOAS, P. R. V. Characterization of complex networks: a survey of measurements. **Advances in Physics**, v.56, n.1, p.167-242, 2007.

22 BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n.5439, p. 509–512, 1999.

23 SOLOMONOFF, R.; RAPOPORT, A. Connectivity of random nets. **Bulletin of Mathematical Biology**., v. 13, n.2 p. 107 – 117, 1951.

24 ERDÖS, P.; RÉNYI, A. On random graphs. **Publicationes Mathematicae**, v. 6, p. 290–297, 1959

25 PRICE, D. J. d. S. A general theory of bibliometric and other cumulative advantage processes. **Journal of the American Society for Information Science**, v. 27, n. 5, p. 292 – 306, 1976.

26 ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. **Reviews of Modern Physics**, v. 74, n.1, p. 47–97, 2002.

27 KRIEGER G.; RENTSCHLER I.; HAUSKE G.; SCHILL K.; ZETZSCHE C. Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. **Spatial Vision**. v. 13, n. 2-3, p. 201-14, 2000.

28 ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE transactions on pattern analysis and machine intelligence**, v. 20, n.11, p. 1254-1259, 1998.

29 SINHA, P.; BALAS, B.; OSTROVSKY, Y.; RUSSEL, R. Face recognition by humans: nineteen results all computer vision researchers should know about. **Proceedings of the IEEE**, v.94, n.11, p. 1948-1962, 2006.

30 PEGDEN, C. D.; SHANON, R. E.; SADOWSY, R. **Introduction to simulation using SIMAN**. New Jersey: McGraw-Hill, 1990.