

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

CAYO MAGNO DA CRUZ FONTANA

**MOVIMENTOS SACÁDICOS VIRTUAIS BASEADOS EM VG-RAM NA
DETECÇÃO AUTOMÁTICA DE PLACAS DE TRÂNSITO**

VITÓRIA

2013

CAYO MAGNO DA CRUZ FONTANA

**MOVIMENTOS SACÁDICOS VIRTUAIS BASEADOS EM VG-RAM NA
DETECÇÃO AUTOMÁTICA DE PLACAS DE TRÂNSITO**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Informática.

VITÓRIA

2013

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

F679d Fontana, Cayo Magno da Cruz, 1983-
Movimentos sacádicos virtuais baseados em VG-RAM na
detecção automática de placas de trânsito / Cayo Magno da
Cruz Fontana. – 2013.
103 f. : il.

Orientador: Alberto Ferreira de Souza.

Coorientador: Edilson de Aguiar.

Dissertação (Mestrado em Informática) – Universidade
Federal do Espírito Santo, Centro Tecnológico.

1. Redes neurais (Computação). 2. Trânsito - Sinais e
sinalização. 3. Olhos - Movimentos. 4. Reconhecimento de
padrões. 5. Memória de acesso aleatório. 6. Simetria. I. Souza,
Alberto Ferreira de. III. Aguiar, Edilson de. III. Universidade
Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 004

Movimentos Sacádicos Virtuais Baseados em VG-RAM na Detecção Automática de Placas de Trânsito

Cayo Magno da Cruz Fontana

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

Aprovada em 29 de agosto de 2013 por:



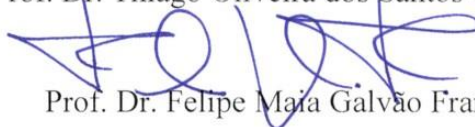
Prof. Dr. Alberto Ferreira De Souza - DI/UFES



Prof. Dr. Edilson de Aguiar - DI/UFES



Prof. Dr. Thiago Oliveira dos Santos - DI/UFES



Prof. Dr. Felipe Maia Galvão França - UFRJ

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória-ES, 29 de agosto de 2013

Dedico este trabalho a tudo e a todos
que me motivaram a realizá-lo.

AGRADECIMENTOS

Primeiramente agradeço a Deus, pela presença constante, força e apoio durante toda esta etapa trilhada com bastante esforço. Ao meu orientador, professor Dr. Alberto Ferreira de Souza, pela oportunidade concedida, por todo o ensinamento durante esta fase e pela ajuda que tornaram este trabalho possível. Ao meu coorientador, professor Dr. Edilson de Aguiar, pela orientação, auxílio e revisão deste trabalho. Aos professores Dr. Felipe Maia Galvão França e Thiago Oliveira dos Santos pelo convite aceito, pela disponibilidade de participar e pelas contribuições pessoais acerca da monografia. À minha mãe Luzinete por toda a educação e orientação familiar que me deu, desde sempre. À minha noiva Pollyanna pelo incentivo para a conclusão deste e também pela compreensão da minha ausência, em vários momentos, ao longo desta jornada. A toda equipe do LCAD pelo suporte, apoio técnico e prestatividade de sempre perante todos os problemas enfrentados durante todo o curso. Aos meus colegas de trabalho pelo incentivo e compreensão nos necessários momentos de ausência. E por fim, agradeço a todos que estiveram envolvidos, diretamente ou indiretamente, na conclusão deste.

RESUMO

A tarefa de detectar e reconhecer placas de trânsito, em ambientes reais, tem sido amplamente pesquisada nos últimos anos.

Recentemente, a quantidade de veículos nas vias urbanas tem crescido exponencialmente. Grandes problemas nestas vias surgiram como consequência deste crescimento. Dados estatísticos da Organização das Nações Unidas (ONU) apontam acidentes de trânsito como uma das principais causas de mortes no planeta.

Com o objetivo de auxiliar os condutores na tarefa de detectar e reconhecer placas de trânsito para alertá-los sobre possíveis alterações nas vias, ou até mesmo atuar no controle do carro, apresentaremos nesta dissertação uma abordagem biologicamente inspirada para a detecção de placas de trânsito com base em uma Rede Neural Sem Peso com Generalização Virtual da Memória de Acesso Randômico (Virtual Generalizing Random Access Memory Weightless Neural Network – VG-RAM WNN).

Redes Neurais sem Peso (RNSP) baseadas em neurônios VG-RAM apresentam um conjunto de eficientes funcionalidades para o aprendizado de máquina, oferecendo uma implementação simples, com treinos e testes rápidos.

A arquitetura da nossa RNSP VG-RAM modela o movimento sacádicos dos olhos, e as transformações sofridas pelas imagens capturadas pela retina dos olhos para o colículo superior no cérebro dos mamíferos. Nós avaliamos o desempenho do nosso sistema na detecção de placas de trânsito utilizando como referência um conjunto de placas de trânsito de origem alemã (GTSDb). Utilizando apenas 12 imagens de placas de trânsito para a etapa de treinamento, nosso sistema foi classificado na 16ª posição, em um total de 53 métodos submetidos entre as 18 equipes participantes, para a detecção de placas de trânsito alemão na categoria proibitiva, na Competição Alemã de Placas de Trânsito, etapa do IJCNN'2013. Os resultados experimentais evidenciaram que nossa abordagem é capaz de detectar uma grande variedade de placas de trânsito, de forma confiável e eficiente, utilizando apenas algumas amostras para o treinamento.

ABSTRACT

The task of detecting and recognizing road signs in real environments have been widely researched in recent years.

Recently, the number of vehicles on urban roads has grown exponentially. Big problems in these pathways have emerged as a result of this growth. Statistics of the United Nations (UN), points traffic accidents as a leading cause of death in the world.

With the aim of assisting drivers in the task of detecting and recognizing road signs to alert them about possible changes in the way, or even act to control the car, we present in this dissertation a biologically inspired approach to detect traffic signs based on a Virtual Generalizing Random Access Memory Weightless Neural Networks - VG-RAM WNN.

VG-RAM WNN are effective machine learning tools that offer simple implementation and fast training and test.

Our VG-RAM WNN architecture models the saccadic eye movement system and the transformations suffered by the images captured by the eyes from the retina to the superior colliculus in the mammalian brain. We evaluated the performance of our VG-RAM WNN system on traffic sign detection using the German Traffic Sign Detection Benchmark (GTSDb). Using only 12 traffic sign images for training, our system was ranked in the 16th position, in the total 53 methods submitted among 18 teams, for the prohibitory category in the German Traffic Sign Detection Competition, part of the IJCNN'2013. Our experimental results showed that our approach is capable of reliably and efficiently detect a large variety of traffic sign categories using a few training samples.

LISTA DE SIGLAS E ABREVIATURAS

WNN	<i>Weightless Neural Network</i>
RNSP	Redes Neurais Sem Peso
RAM	<i>Random Access Memory</i>
VG-RAM	<i>Virtual Generalizing of RAM</i>
OMS	Organização Mundial da Saúde
IJCNN	International Joint Conference on Neural Networks
GTSDb	<i>German Traffic Sign Recognition Benchmark</i>
MAE	Máquina Associadora de Eventos
LCAD	Laboratório de Computação de Alto Desempenho
UFES	Universidade Federal do Espírito Santo
UV	Ultravioleta
IV	Infravermelho
AV	Acuidade Visual
SC	Colículo Superior
TSD	<i>Traffic Sign Detection</i>
TSC	<i>Traffic Sign Centralization</i>
SNC	Sistema nervoso central
PA	Potencial de Ação
QV	<i>Quasi-Visual</i>
cMRF	Formação Reticular Mesencefálica Central
ISI	Intervalo Inter-Sacádico

GPU	<i>Graphics Processing Unit</i>
CUDA	<i>Compute Unified Device Architecture</i>
GB	<i>Gigabyte</i>

LISTA DE FIGURAS

Figura 1 - Espectro visível: O aumento de intensidade pode tornar perceptíveis ondas até então invisíveis, tornando os limites do espectro visível algo elástico	21
Figura 2 - Anatomia interna do olho humano (direito)	22
Figura 3 - Trajetória da luz até as células fotorreceptoras localizadas no nervo óptico.....	23
Figura 4 - A acuidade visual (AV) cai acentuadamente nas regiões periféricas da fóvea	24
Figura 5 - Representação das vias visuais e do fluxo das informações da retina para o interior do cérebro	26
Figura 6 - Estruturas de camadas das células do NGL	27
Figura 7 - Representação do mapeamento retinotópico no V1	28
Figura 8 - Representação do mapa retinotópico e o campo visual no córtex V1	30
Figura 9 - Localização do Colículo Superior no cérebro humano.....	31
Figura 10 – Comparação entre as duas hipóteses das camadas intermediárias do SC. Os números representam as faixas de mesma amplitude sacádica no mapa motor.	33
Figura 11 - Comparação entre a) e b) cujos distratores são muito semelhantes ou muito distintos, respectivamente.....	36
Figura 12 - Padrões de sacadas quando uma imagem de um rosto é observada por 3 minutos	38
Figura 13 - Ilusão de ótica ocasionada pelas microssacadas	41
Figura 14 - Representação do mapeamento retinotópico no córtex visual de um primata, em b), à partir da imagem a).....	44
Figura 15 - Transformação log-polar (mapeamento retinotópico)	45
Figura 16 – Transformação da imagem cartesiana (a) para o plano log-polar (b) e em seguida, em (c), a representação inversa (retinal).....	45
Figura 17 - Nossa variante da transformação log-polar	46
Figura 18 – Estrutura física do neurônio biológico.	48
Figura 19 - Modelo físico de um neurônio artificial	49
Figura 20 - Processamento básico de uma RNSP.....	50
Figura 21 - Estrutura básica de um neurônio do tipo RAM	51
Figura 22 - Fluxo de trabalho da solução proposta	54
Figura 23 - Diagrama de arquitetura da RNSP VG-RAM para detecção de placas de trânsito	54
Figura 24 - Padrão de interconexão sináptica do neurônio Nm2, n2 de TSD	55

Figura 25 - Representação de profundidade (zoom) definida por α em uma imagem.....	57
Figura 26 – Valor obtido por uma sinapse à partir de um elemento da entrada.....	58
Figura 27 - Comparação entre uma imagem original e sua cópia com aplicação do Filtro de Gauss	59
Figura 28 - Ilustração de uma instância de treino do módulo TSD para detecção de placas de trânsito	61
Figura 29 - Ilustração de uma instância de teste do módulo TSD para detecção de placa de trânsito	63
Figura 30 - Descrição das informações que compõem uma tupla gerada por TSD	65
Figura 31 - Representação das delimitações de uma placa de trânsito e as distâncias entre o centro de massa e o centro sacádico	66
Figura 32 - Comparação entre a região da imagem utilizada por TSD e TSC para treinar a RNSP	69
Figura 33 - Busca pela simetria presente na placa através do deslocamento horizontal	70
Figura 34 - Pseudocódigo representando o modelo do algoritmo de TSC.....	72
Figura 35 - Fluxograma da fase de validação do nosso sistema.....	75
Figura 36 - Fluxograma da fase de teste do nosso sistema.....	77
Figura 37 - Desempenho do sistema para os cenários 1, 2 e 3.....	81
Figura 38 - Desempenho do sistema considerando TSD + BI + TSC (cenário 3) nas bases de dados de validação (tendência de cor azul) e de avaliação - teste (tendência de cor vermelha)	82
Figura 39 - Gráfico gerado após a calibração utilizando todo o conjunto de parâmetros	87
Figura 40 - Gráfico de tendências de desempenho da variação das sinapses em função da dimensão da RNSP	88
Figura 41 - Gráfico de tendências de desempenho da variação do desvio padrão da distribuição normal das sinapses em função da dimensão da RNSP.....	90
Figura 42 - Gráfico de tendências de desempenho da variação do fator de escala (fator de ampliação) em função da dimensão da RNSP.....	91
Figura 43 – Comparativo entre as tendências de desempenho com a melhor configuração paramétrica (em vermelho) e a configuração utilizada na competição (em azul)	94

LISTA DE TABELAS

Tabela 1 - Características dos movimentos oculares durante a leitura, percepção de cenas e busca visual	35
Tabela 2 - tabela de pesquisa do neurônio do tipo VG-RAM	52
Tabela 3 - Tabela de Contingência com seus quatro possíveis estados.....	79
Tabela 4 - Descrição dos valores retornados pelas tuplas durante as sacadas na etapa de calibração.....	84
Tabela 5 - Quantidade de execuções do programa para cada configuração	86
Tabela 6 – Matriz de desempenho do número de sinapses utilizado na calibração em função da dimensão da RNSP	89
Tabela 7 - Matriz de desempenho dos valores utilizados para o desvio padrão da distribuição normal na calibração em função da dimensão da RNSP	91
Tabela 8 - Matriz de desempenho dos valores utilizados para o fator de escala (fator de ampliação) na calibração em função da dimensão da RNSP.....	92

SUMÁRIO

RESUMO.....	7
ABSTRACT	8
1. INTRODUÇÃO.....	16
1.1. Motivação	17
1.2. Objetivos	18
1.3. Contribuições	18
1.4. Estrutura do Trabalho	19
2. A VISÃO E OS MOVIMENTOS OCULARES	20
2.1. Percepção Visual	21
2.2. O Olho	22
2.3. Vias Visuais.....	25
2.4. Córtex Visual Primário (V1).....	27
2.5. Colículo Superior	30
2.6. Movimentos Oculares.....	33
2.6.1. Características Básicas dos Movimentos Oculares na Cognição visual	34
2.6.2. Movimentos Oculares Durante a Busca Visual	35
2.7. Movimentos Oculares Sacádicos	37
2.7.1. Classificação dos Movimentos Oculares Sacádicos	39
3. MODELO MATEMÁTICO-COMPUTACIONAL SACÁDICO	42
3.1. Mapeamento Retinotópico (Log-Polar)	43
3.2. Redes Neurais Sem Peso e VG-RAM	47
4. DETECÇÃO DE PLACAS DE TRÂNSITO COM RNSP VG-RAM	53
4.1. Módulo de Detecção de Placas de Trânsito (TSD)	54

4.2.	Módulo de Inferência Bayesiana (BI).....	64
4.3.	Módulo de Centralização de Placas de Trânsito (TSC)	68
5.	AVALIAÇÃO EXPERIMENTAL	73
5.1.	Metodologia.....	73
5.2.	Resultados.....	77
5.3.	Calibração	83
5.3.1.	Variação do Número de Sinapses	87
5.3.2.	Variação do Desvio Padrão da Distribuição Normal	89
5.3.3.	Variação do Fator de Escala	91
5.4.	Resultados Após Calibração	93
6.	CONCLUSÃO	95
6.1.	Sumário	95
6.2.	Análise Crítica.....	96
6.3.	Conclusões	96
6.4.	Trabalhos Futuros	97

1. INTRODUÇÃO

Segurança durante a condução veicular é um tópico de pesquisa de grande relevância para a indústria automotiva. Uma forma de tornar os carros mais seguros durante a atividade de condução é a detecção e o reconhecimento de placas de trânsito, em tempo real. Esta tecnologia tem como objetivo alertar o condutor de ações inadequadas, tais como: excesso de velocidade, tomar um caminho errado (contramão), além de auxiliá-lo em situações difíceis como mau tempo, cansaço, insônia, entre outros. Embora o processo de detecção de placas de trânsito possa ser simplificado, pelo fato da maioria das placas de trânsito ser fixas, às vezes até mesmo descritas por lei, em situações do mundo real onde ocorrem variações estruturais, como por exemplo, diferentes condições de luz, clima, mudanças de ponto de vista, o envelhecimento da placa de trânsito e até deformações, abordagens simples não são confiáveis. Desta forma, métodos robustos são necessários.

Enquanto os humanos são capazes de detectar a existência de uma grande variedade de placas de trânsito eficientemente, sistemas automáticos ainda é um desafio. Dada à alta relevância industrial automotiva, detecção e reconhecimento de placas de trânsito vêm atraindo as atenções de muitos pesquisadores nos últimos anos [1]. Apesar dos recentes grandes avanços, detecção de placas de trânsito ainda é um problema complexo para o mundo real, o que a torna uma importante aplicação para sistemas de condução avançada e autônoma. Um sistema de detecção útil precisa tratar situações onde as placas podem estar rotacionadas, diferentes condições de iluminação, mudanças de perspectiva, oclusão e todos os tipos de condições climáticas. Dada a imagem de uma cena, o problema geral de identificação de placas de trânsito é conseguir identificar uma ou mais placas nesta imagem, usando uma informação a priori sobre a forma, a cor ou as características presentes nas possíveis placas de trânsito.

As atuais soluções na literatura geralmente envolvem segmentação das placas de trânsito nas imagens para solucionar a etapa de detecção da placa e extração de características das regiões das placas de trânsito para a tarefa de reconhecimento [1]. Nesta dissertação, estamos interessados apenas na etapa da detecção de placas de trânsito.

Para detectar placas de trânsito, uma variedade de técnicas tem sido proposta na literatura (ver [1], [2] e [3]). Podemos agrupá-las em 3 principais categorias: métodos baseados em cores, métodos baseados em formas e métodos baseados em características. Na metodologia baseada

em cores [4], geralmente aplica-se a segmentação das cores combinado com técnicas de detecção de bordas para encontrar formas específicas correspondentes às placas de trânsito, nas imagens. Métodos baseados em formas são essencialmente dependentes de informações sobre variações de bordas na imagem, para extrair restrições geométricas que possivelmente correspondem a placas de trânsito, como elipses nas imagens [5]. Além disso, a simetria radial [6] pode ser utilizada para detectar formas regulares, tais como: triângulos, quadrados, octógonos, entre outros. Métodos baseados em características (*Feature-based methods*) utilizam técnicas de aprendizado de máquina para detectar características especiais nas imagens. Entre as técnicas mais utilizadas, estão as Redes Neurais [7], *Support Vector Machine* [8] e métodos *AdaBoost* [9].

Nesta dissertação, apresentamos uma abordagem biologicamente inspirada para a detecção de placas de trânsito baseada na Generalização Virtual de Acesso a Memória Randômica em Redes Neurais Sem Peso (*Virtual Generalizing Random Access Memory Weightless Neural Network* – VG-RAM WNN [10]). RNSP's com neurônios do tipo VG-RAM são eficientes ferramentas de aprendizagem de máquina, dotadas de uma arquitetura simples de implementar como também proporcionam alto desempenho nas etapas de treinamento e testes. Nossa RNSP VG-RAM modela o sistema biológico de movimentos oculares sacádicos e a transformação sofrida pelas imagens captadas pelos olhos da retina para o colículo superior do cérebro de mamíferos.

Nós avaliamos o desempenho do nosso sistema utilizando o conjunto de placas de trânsito da Conferência de Detecção de Placas de Trânsito Alemãs (GTSDb) (<http://benchmark.ini.rub.de>) [11]. Os resultados experimentais mostraram que a nossa abordagem é capaz de detectar de maneira confiável e eficiente uma grande variedade de categorias de placas de trânsito, utilizando apenas doze imagens (contendo uma placa de trânsito em cada uma destas imagens) para o treinamento.

1.1. Motivação

A principal motivação deste trabalho é o desenvolvimento de um modelo matemático-computacional capaz de detectar placas de trânsito em imagens bidimensionais, utilizando

uma metodologia diferente das principais encontradas na literatura atual para a atividade proposta.

Outra grande motivação para o desenvolvimento deste trabalho são os alarmantes dados estatísticos da OMS – Organização Mundial da Saúde – que apontam os acidentes de trânsito como uma das principais causas de mortes em todo o mundo. Por ano, cerca de 1,24 milhão de pessoas são mortas no trânsito em todo o mundo [19]. Sistemas de detecção de placas de trânsito em tempo real podem auxiliar os condutores a respeitarem a legislação de trânsito evitando com que se envolvam em possíveis acidentes, ou ainda, acionar sistemas mecânicos do automóvel – como freio, aceleração, entre outros – atuando de forma autônoma, mantendo a segurança durante a condução veicular.

1.2. Objetivos

O principal objetivo deste trabalho é apresentar um modelo computacionalmente viável para a tarefa de detecção de placas de trânsito em imagens bidimensionais, baseado na neurofisiologia central da visão dos mamíferos, localizada em uma região cerebral conhecida como córtex visual primário [20], responsável pela realização da busca visual do objeto de interesse em uma imagem.

1.3. Contribuições

O Laboratório de Computação de Alto Desempenho (LCAD), da Universidade Federal do Espírito Santo (UFES), possui um projeto em desenvolvimento que tem como objetivo permitir que um veículo de passeio seja capaz de trafegar de forma autônoma em ambientes urbanos, considerando as leis de trânsito e evitando colisões. Alguns módulos deste projeto serão responsáveis por detectar objetos de interesse nas imagens bidimensionais obtidas a partir de câmeras instaladas no exterior do veículo. O modelo proposto neste trabalho permite que placas de trânsito, espalhadas ao longo das vias urbanas, sejam detectadas durante a condução autônoma do veículo.

As principais contribuições deste trabalho são:

- Implementação de um modelo matemático-computacional para a solução da tarefa de detecção de placas de trânsito em imagens bidimensionais, baseado no sistema visual biológico dos mamíferos.
- Integração de novas funcionalidades ao *framework* MAE (Máquina Associadora de Eventos), acrescentando um módulo de refinamento de detecção de objetos pela simetria geométrica dos mesmos e inclusão de funções genéricas que permitam a visualização computacional das placas de trânsito, na representação log-polar, com as mesmas dimensões da imagem bidimensional de origem.
- Ajuste de parâmetros utilizando técnicas de calibração, com o objetivo de encontrar a configuração com o melhor desempenho na tarefa de detecção de placas de trânsito, através do modelo matemático-computacional de movimentos oculares sacádicos apresentado.

1.4. Estrutura do Trabalho

Esta dissertação está estruturada da seguinte maneira:

- **A Visão e os Movimentos Oculares:** este capítulo apresenta um estudo desenvolvido sobre a fisiologia dos movimentos oculares da visão humana, abordando desde a reflexão da luz na retina dos olhos até o mapeamento destas informações no cérebro humano, enfatizando os movimentos oculares sacádicos, de forma detalhada e explicativa, para que seja passado o entendimento do modelo matemático-computacional baseado neste estudo.
- **Modelo Matemático-Computacional Sacádico:** este capítulo revela a arquitetura e o modelo matemático-computacional biologicamente plausível desenvolvido, com o objetivo de descrever a transformação das informações visuais obtidas pela retina, ocorridas no interior do cérebro, em uma região conhecida como Colículo Superior (SC), utilizando o sistema de coordenadas log-polar.

- **Detecção de Placas de Trânsito com RNSP VG-RAM:** este capítulo apresenta um sistema computacional baseado na arquitetura de uma rede neural sem peso, constituída de neurônios do tipo VG-RAM, que se mostrou bastante eficaz na tarefa de detecção placas de trânsito, em imagens reais de vias urbanas, com fundamentos contextualizados no modelo computacional apresentado no capítulo 3.
- **Avaliação Experimental:** este capítulo apresenta nossa avaliação experimental do sistema proposto. Para tanto, contamos com o apoio de um conjunto de imagens de placas de trânsito alemão fornecido por um instituto interessado em realizar uma competição de sistemas computacionais capazes de detectar placas de trânsito em ambientes reais.
- **Conclusão:** este capítulo apresenta nossa conclusão sobre o trabalho feito, as contribuições dentro do contexto apresentado assim como a direção proposta para o mesmo e o que é esperado nos trabalhos futuros.

2. A VISÃO E OS MOVIMENTOS OCULARES

A essência deste trabalho está baseada nos movimentos oculares compreendidos na fisiologia da visão humana. Nesta seção será apresentado o estudo desenvolvido sobre a visão biológica e os movimentos dos olhos, com foco nos movimentos oculares sacádicos, para que seja possível a compreensão e o entendimento do modelo computacional proposto, explicado posteriormente.

Os olhos humanos são estruturas complexas embora tenham sido muitas vezes comparados com uma câmera [21] pelo fato de que, assim como a lente de uma câmera, a lente do olho foca em uma imagem invertida na retina. Esta analogia passa bem distante do que realmente representa o sistema visual humano. Além de criar uma perspectiva tridimensional do mundo externo, diferente das imagens bidimensionais, o sistema visual possui uma complexa e poderosa função cognitiva que nos permite perceber um objeto de interesse mesmo que este seja apresentado em diferentes condições visuais que consequentemente ocasionam amplas variações na retina, de forma bastante nítida, direcionando os olhos para este [22].

2.1. Percepção Visual

A percepção visual inicia-se na retina no momento que a luz incide na córnea e então é projetada na parte traseira do globo ocular onde é convertida em sinais elétricos. O sistema visual humano é sensível a ondas de comprimento entre aproximadamente 400 e 750 nanômetros, conforme apresentado na figura 1. A figura 1 também ilustra o espectro da radiação eletromagnética capaz de sensibilizar os olhos humanos para possibilitar a visibilidade da luz por radiação.

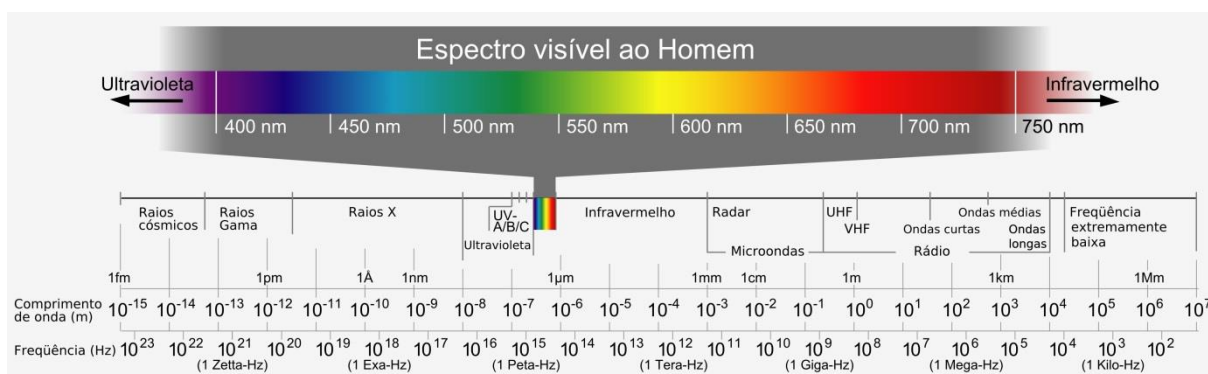


Figura 1 - Espectro visível: O aumento de intensidade pode tornar perceptíveis ondas até então invisíveis, tornando os limites do espectro visível algo elástico

Fonte: retirado de

http://meteoropole.com.br/site/wp-content/uploads/2012/12/espectro_eletromagnetico.jpg

É possível perceber que o comprimento de onda é inversamente proporcional à frequência. A menor frequência estimulante presente no espectro visível é observada pela cor vermelha. À medida que as frequências tornam-se ainda mais baixas que as percebidas pela cor vermelha, o espectro visível entra numa região conhecida como infravermelha (IV), invisível ao olho humano. A mesma analogia ocorre no limite das mais altas frequências, observadas pela cor violeta. Acima destas altas frequências o espectro de radiação entra em uma região chamada ultravioleta (UV). Também invisíveis ao olho humano.

Ao incidirem nos olhos, estas informações visuais – luzes – são então enviadas através do nervo óptico para centros superiores do cérebro para processamento adicional necessário para a percepção.

2.2. O Olho

O olho humano é um órgão altamente complexo e delicado, dotado de diferentes processos e componentes. Possui uma forma física semelhante a uma esfera com aproximadamente 24.4 milímetros de diâmetro [22].

A luz penetra nos olhos através da Córnea: uma membrana de aproximadamente 0.5 mm de espessura, apresentando um índice de refração – número que descreve como a luz, ou qualquer outro tipo de radiação, se propaga através do meio – de $n = 1.377$. A córnea é uma membrana extremamente sensível, fechando a pálpebra dos olhos a qualquer toque. Em seguida esta luz atravessa outra região chamada Câmara Anterior, uma membrana aquosa com cerca de 3.04 mm de espessura, possuindo um índice de refração de $n = 1.336$. A figura 2 apresenta a anatomia do olho direito, por um corte médio, na perspectiva de cima para baixo.

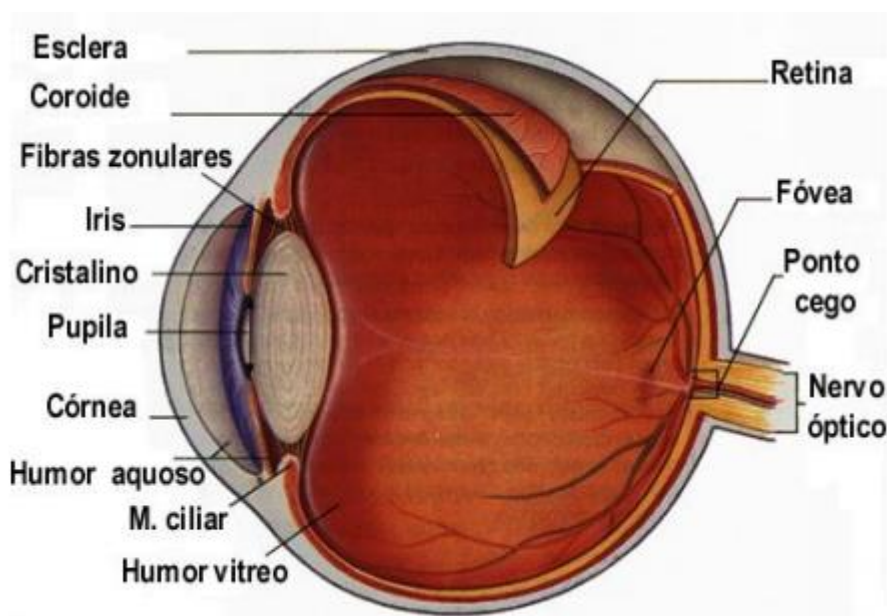


Figura 2 - Anatomia interna do olho humano (direito)
Fonte: adaptado de [22]

Após atravessar a Câmara Anterior, a radiação da luz chega ao Cristalino (ou Lente). A Lente é responsável pelo ajuste de foco quando olhamos para um objeto de interesse longe ou perto. Tornando-se mais ou menos convergente, esta membrana permite que a visão seja limpa e nítida. Por trás do Cristalino, a luz passa pelo Humor Vítreo: uma massa transparente,

semelhante a um gel, responsável por preencher todo o espaço entre a Lente, que se encontra desde a região frontal dos olhos até a Retina, na região posterior do globo ocular. Após atravessar o Humor Vítreo, a luz encontra a Retina, onde efetivamente ocorre a detecção da luz pelos olhos.

A retina é um fino tecido neural, de aproximadamente 0.25 mm de espessura, que reveste a parte posterior do globo ocular [23]. Além das células fotorreceptoras sensíveis a luz – os cones e os bastonetes – a retina possui quatro classes básicas de neurônios: horizontais, bipolares, amácrinas e células ganglionares. Estes neurônios presentes na retina são segregados em três camadas de células e duas camadas sinápticas, denominadas plexiformes.

Todas as sinapses presentes na retina são encontradas nas camadas plexiformes cujas informações visuais externas penetram em pelo menos duas delas antes de sair da região do globo ocular: camada plexiforme exterior e a camada plexiforme interior. Os axônios das células ganglionares que ocupam a superfície da retina formam o nervo óptico. O nervo óptico é responsável por enviar as informações visuais para as demais regiões do cérebro. Após atravessar a retina, a luz encontra as células fotorreceptoras: cones e bastonetes. Os cones e os bastonetes são responsáveis pelo reconhecimento de cores e luminosidade, respectivamente, presentes na reflexão da radiação. Os cones são densamente presentes na região da visão. A figura 3 ilustra a trajetória da luz desde a córnea até os fotorreceptores.

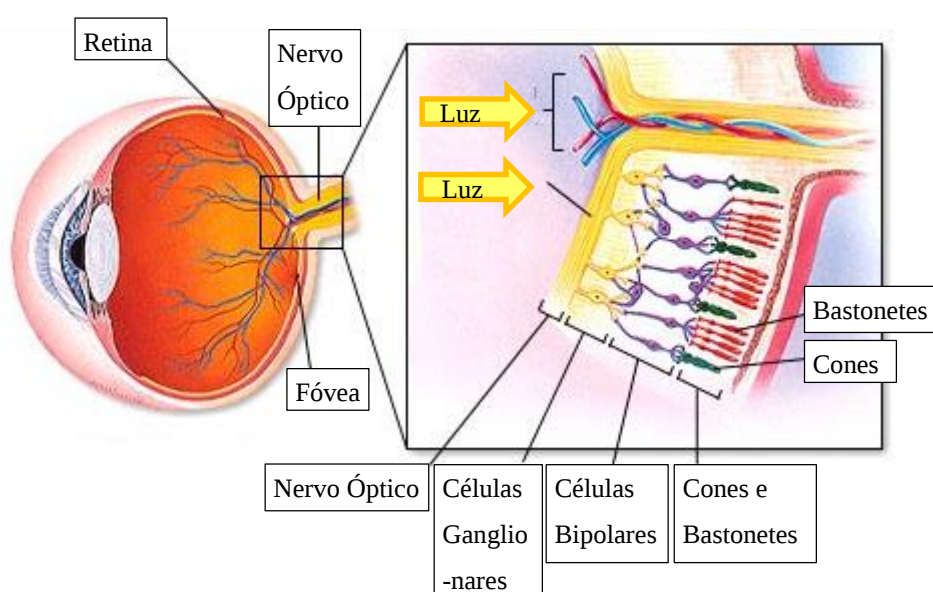


Figura 3 - Trajetória da luz até as células fotorreceptoras localizadas no nervo óptico.

Fonte: adaptado de

http://www.mhhe.com/biosci/esp/2001_gbio/folder_structure/an/m3/s3/assets/images/anm3s3_4.jpg

A alta sensibilidade a luz pelos fotorreceptores é resultado da presença de uma molécula de pigmentos visuais, contidas em suas células, chamada segmento externo. Os cones possuem 3 pigmentos que absorvem o máximo da luz nas cores vermelho, verde e azul, do espectro. Estes 3 tipos de cones responsável pela absorção das cores do espectro chamam-se Opsina Verde, Opsina Vermelha e Opina Azul. A grande maioria dos seres humanos misturam as intensidades destas cores para enxergar todas as demais cores do espectro. Este modelo, chamado de triclomatismo – baseada na representação destas três cores primárias – influenciou o padrão internacional de cores RGB (*Red*, *Green* e *Blue*). Existem aproximadamente 4.5 milhões de cones e 91 milhões de bastonetes presentes na retina do olho humano. Apesar da maioridade, não há presença de bastonetes em uma região onde temos uma grande percepção e detalhamento das informações visual, a Fóvea.

Caracterizada por uma depressão na superfície central da retina, a fóvea é responsável pela visão em alta definição do sistema visual humano. Todos os detalhes visuais são obtidos na região da fóvea, nos permitindo realizar diversos tipos de atividades como dirigir, ler, observar cenas detalhadamente, e outras atividades visuais que requerem alta definição. Nesta região, apenas os cones são presentes e em uma concentração 200 vezes maior que nas demais regiões da retina. Por ser uma região completamente avascular a luz é detectada na fóvea sem nenhuma perda ou dispersão de radiação, tornando a fóvea como uma região de máxima acuidade visual [25]. Acuidade visual (AV) representa a clareza da visão em relação ao foco da retina e a sensibilidade cerebral.

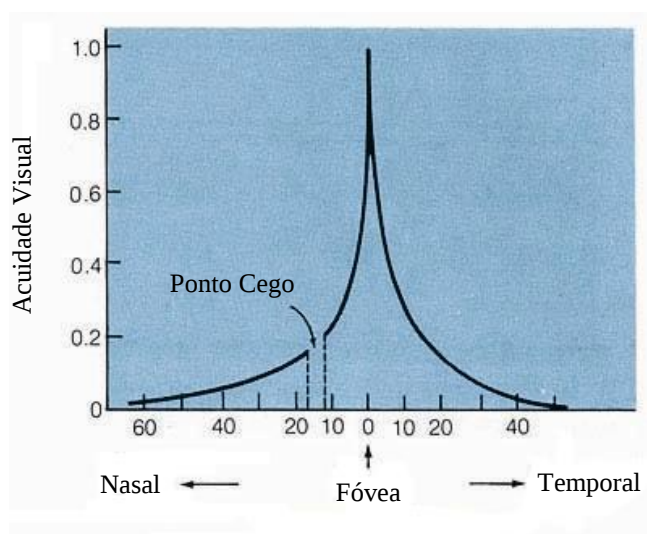


Figura 4 - A acuidade visual (AV) cai acentuadamente nas regiões periféricas da fóvea
 Fonte: adaptado de <http://www.scholarpedia.org/w/images/a/a8/Resolution.jpg>

A figura 4 mostra o gráfico da acuidade visual (visão em alta definição) no contexto do olho humano. A fóvea está situada no ponto 0, onde ocorre a máxima acuidade visual. À medida que as informações visuais se dispersam desta região, ou seja, nas regiões periféricas da fóvea, a acuidade visual cai acentuadamente. É possível notar que temos uma região, na retina, onde a detecção da luz não ocorre. Localizada na região nasal do olho humano, esta região é conhecida como ponto cego e é totalmente desprovida de células fotorreceptoras, justificando assim o seu nome.

Vimos que as informações visuais que chegam à nossa retina são transformadas em pulsos nervosos pelos fotorreceptores e pelas células ganglionares. Após esta transformação, estas informações são enviadas para as demais áreas do cérebro, conhecidas como as vias visuais.

2.3. Vias Visuais

Os axônios das células ganglionares presentes na retina do olho humano, através do nervo óptico, projetam informações visuais para múltiplas áreas do cérebro, dentre elas o Córtex Visual Primário (V1) que possui a projeção visual mais importante do cérebro [37]. Essas informações atingem uma estrutura chamada quiasma óptico que inverte o fluxo destas informações visuais, oriundas da retina, para o interior do cérebro. Após esta inversão, as informações visuais chegam a uma região chamada de Núcleo Geniculado Lateral (NGL).

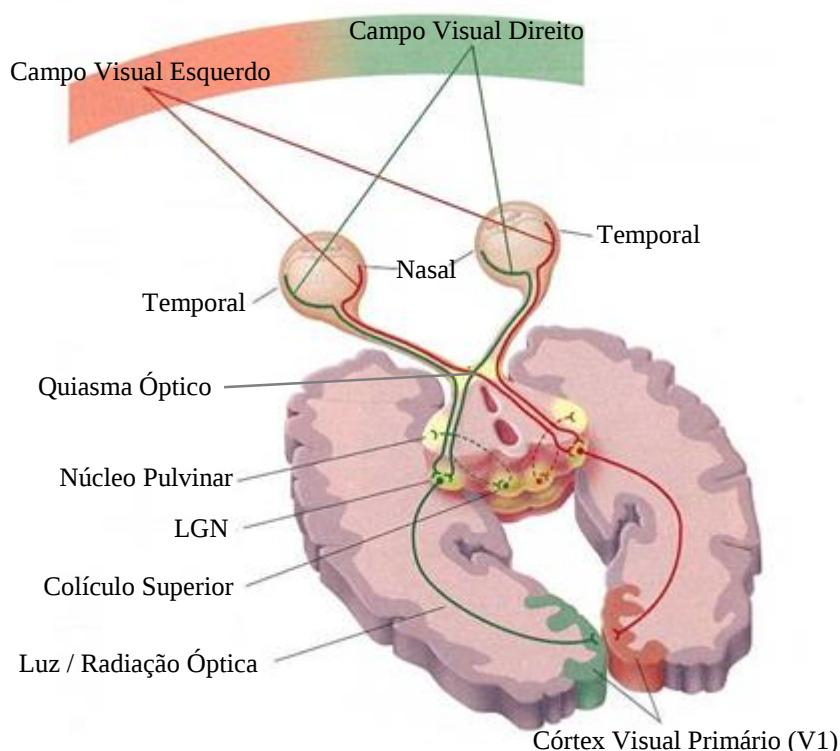


Figura 5 - Representação das vias visuais e do fluxo das informações da retina para o interior do cérebro
 Fonte: adaptado de http://www.edoctoronline.com/media/19/photos_82DF0D2A-BFD1-4043-B3B8-644A45DD0F38.jpg

O cérebro possui duas estruturas NGL sendo uma no hemisfério direito e outra no esquerdo, específicos para cada olho. Cada NGL possui 6 camadas de células empilhadas, uma sobre a outra [22], segregadas em dois subgrupos: células do tipo Magno e do tipo Parvo. As células do tipo Magno são maiores que as células Parvo e são responsáveis pelas informações sobre os movimentos, não possuindo grande nível de detalhamento das informações. Trabalham em altas taxas de transmissão de informações enquanto as células do tipo Parvo possuem detalhes mais ricos, principalmente em relação às cores e formas. As células Parvo possuem uma alta definição espacial, porém operam em taxas mais baixas que as células Magno [38]. A combinação destas informações por estas 6 camadas são chamadas de radiação óptica e são enviadas ao Córtex Visual Primário (ver figura 5). As estruturas das camadas do NGL são similares as estruturas ganglionares da retina, possuindo organizações similares. Este modelo é chamado de retinotopia ou mapa retinotópico.

A representação retinotópica não é linear por haver uma concentração muito maior de neurônios na região da fóvea do que nas demais regiões da retina (ver figura 4), proporcionando uma alta resolução na parte central.

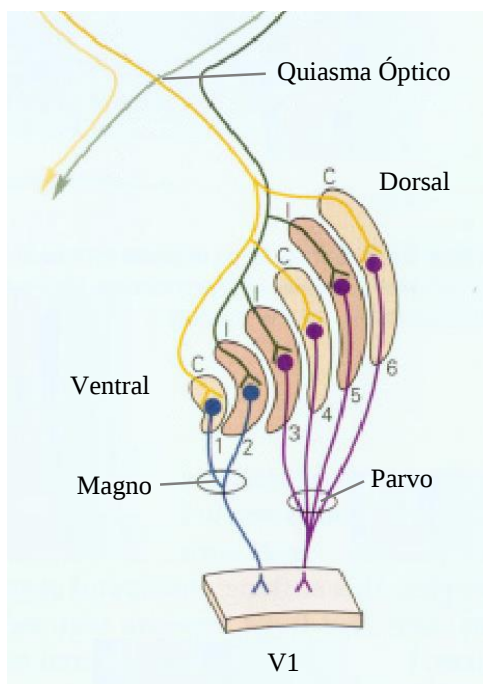


Figura 6 - Estruturas de camadas das células do NGL
Fonte: adaptado de [21]

A figura 6 ilustra o NGL do hemisfério direito do cérebro. As camadas 1 e 2 representam as células Magno enquanto que as camadas 3, 4, 5 e 6, as células Parvo. As conexões ganglionares oriundas do lado temporal da retina chegam às camadas 2 (magno), 3 e 5 (parvo). Esta organização é chamada de ipsilateral (mesmo lado), ou seja, estas informações são obtidas pelo olho correspondente ao mesmo hemisfério do cérebro. Não ocorre a inversão das informações retinotópicas nesta situação. Por outro lado, as conexões do lado nasal nas camadas 1 (magno), 4 e 6 (parvo), conhecidas como contralateral. Esta camada recebe informações retinotópicas do olho (lado nasal) de lado oposto. Após as informações retinotópicas serem organizadas nesta estrutura, elas continuam segregadas pelos hemisférios e são levadas ao Córtex Visual Primário (V1).

2.4. Córtex Visual Primário (V1)

O Córtex Visual Primário – também chamado de Córtex Estriado ou V1 – está localizado no polo posterior do Córtex Occipital, sendo responsável pelo processamento das informações

visuais e umas das áreas mais estudadas do cérebro [39]. É uma das regiões de maior densidade de neurônios do cérebro, possuindo cerca de 140 milhões em cada hemisfério. O Córtex Visual Primário é dividido em 6 camadas distintas enumeradas de 1 a 6. Assim como as demais estruturas das vias visuais o V1 possui dois hemisférios: o direito recebendo informações do olho esquerdo e vice-versa. Cada hemisfério do V1 recebe informações de dois caminhos chamados de fluxo dorsal e fluxo ventral. O fluxo dorsal inicia-se no V1 passando por demais como V2, V5 e Parietal Superior. Está associado com o movimento e representação das posições de objetos na cena mapeada. O fluxo ventral inicia-se no V1 passando pelo V2, V4 e pelo Córtex Temporal Inferior. Está associado ao reconhecimento de formas primitivas e complexas e também com o armazenamento da memória de longo prazo.

A principal função do V1 é o processamento visual das informações do LGN e enviar o resultado deste processamento para demais áreas visuais corticais e outras regiões do cérebro, como o Colículo Superior. Além de especializada em processamento de informações sobre objetos, tanto estáticos quanto dinâmicos, é um excelente reconhecedor de padrões [40]. A figura 7 ilustra o mapeamento de uma cena observada pelos olhos e mapeada no V1.

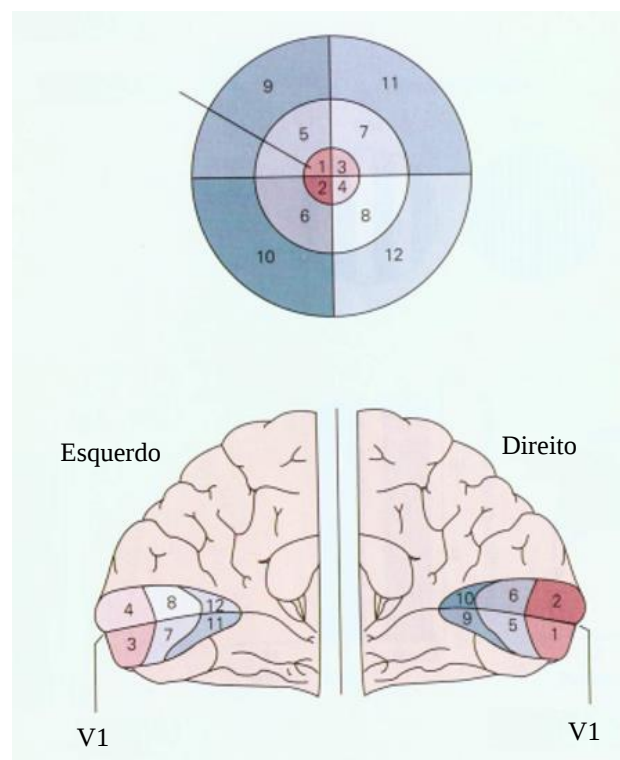


Figura 7 - Representação do mapeamento retinotópico no V1
Fonte: adaptado de [21]

Pela alta sensibilidade e densidades de fotorreceptores, aproximadamente um quarto das informações retinotópicas oriundas da região da fóvea são representadas no córtex visual primário. Estas informações possuem um alto nível de detalhamento no V1.

Retinotopia

O termo retinotopia, ou mapeamento retinotópico, refere-se a uma relação não aleatória entre o padrão de distribuição dos neurônios presente no V1 e os valores de seus campos receptivos [41]. Mapas retinotópicos podem ser caracterizados quantitativamente por uma taxa de mudança de posição no córtex v1 como uma função da posição no espaço visual em mm / grau. Esta taxa quantitativa é chamada de Fator de Ampliação (M). Um alto valor de M significa que uma grande região do V1 é alocada para analisar uma pequena região no espaço visual. Nos primatas, M é maior na região da Fóvea e diminui nas regiões periféricas (ver figura 7).

A relação entre as informações da retina e a sua representação no córtex v1, pode ser descrita por uma função logarítmica, $w = a \log(b + z)$, onde $w = (x + iy)$ obtém a posição no córtex em mm, $z = re^{i\theta}$, onde r é o campo visual medido em graus, θ é o ângulo meridional medido em radianos e a e b são constantes experimentais. Esta função foi inicialmente proposta por Schwartz (1980), e descreve o mapeamento observado em primatas, incluindo humanos, evidenciando a grande representação visual da fóvea nos hemisférios do V1 e a tendência das linhas de ângulos meridionais sendo mapeadas ortogonalmente na superfície do V1 (ver figura 8).

Na imagem à direita da figura 8, temos uma representação do campo visual nas mesmas coordenadas em que aparecem no córtex V1 aberto e aplainado. O ponto HM é o meridiano horizontal – representação da origem do plano apresentado – e o ponto escuro é o ponto cego presente na retina do olho. Os números informam o ângulo das meridionais com as linhas ortogonais (os possíveis valores de θ na proposta de Schwartz).

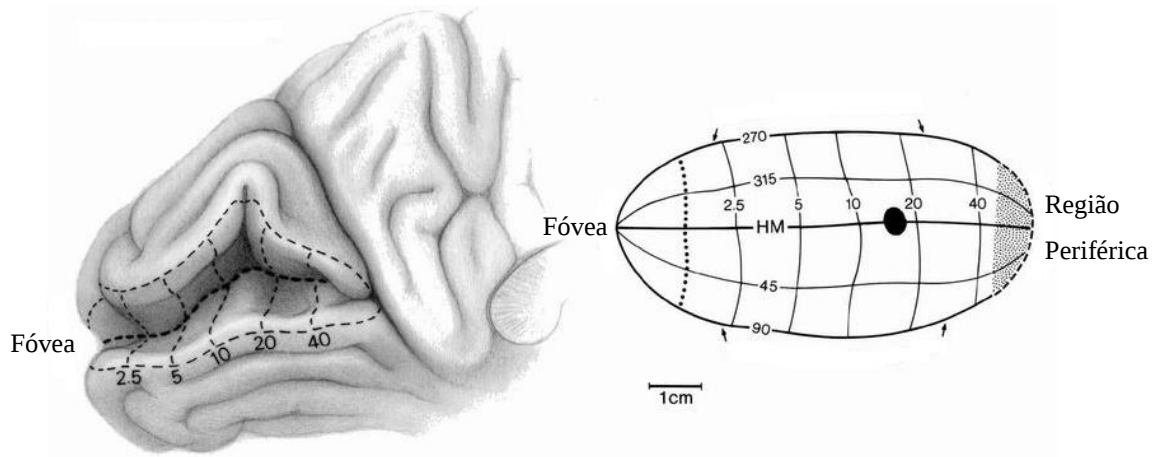


Figura 8 - Representação do mapa retinotópico e o campo visual no córtex V1

Fonte: adaptado de

http://www.scholarpedia.org/w/images/thumb/0/0d/Visual_map_Swindale_Human_visual_fields_Horton_2006.jpg/500px-Visual_map_Swindale_Human_visual_fields_Horton_2006.jpg

Apesar desta representação uniformemente logarítmica pesquisas recentes sugerem que a taxa do fator de ampliação, M , pode ser localmente isotrópica, sofrendo bastante influência de direção de acordo com os índices de refração da luz nos olhos.

A precisão dos mapas locais coexiste com possíveis distorções globais, particulares em primatas [40]. Um exemplo de distorção é a ampliação das informações retinotópicas que favorece o campo visual central em relação à visão periférica onde 50% da área do V1 são responsáveis por processar 2% do campo visual. Esta distorção é de suma importância e relevância para a atuação de outra estrutura do cérebro com papel fundamental no sistema visual humano: o Colículo Superior (SC), responsável pelos movimentos oculares dos olhos para regiões previamente mapeadas em alta definição pelo córtex V1.

2.5. Colículo Superior

O sistema motor visual dos mamíferos é um dos melhores modelos estudados para elucidar os princípios computacionais no controle dos movimentos oculares e os mecanismos de aplicação neural. Neste modelo encontra-se o Colículo Superior (SC), desempenhando um papel fundamental no direcionamento do olhar, a partir dos olhos e dos movimentos da cabeça, integrando múltiplas entradas sensoriais e cognitivas. Localizado no mesencéfalo dorsal [23], sua estrutura é composta de 3 camadas conhecidas como Superficial,

Intermediária e Profunda. A camada superficial do SC recebe informações da retina e de outras áreas visuais corticais. A camada intermediária do SC recebe as informações visuais diretamente da camada superficial através de um grupo de células nervosas localizadas na posição lateral do mesencéfalo, adjacente ao colículo inferior ao qual possui interligações, chamado de *Parabigeminalis*. Esta estrutura intermediária é responsável por avaliar a saliência de curto prazo das informações sensoriais visuais. A camada intermediária também é responsável por representar, de forma retinotópica, as informações visuais do alvo sacádico. Neste momento, seus neurônios de saída sofrem ativações que são reenviadas para a Ponte de Varólio – estrutura pertencente ao tronco cerebral de grande importância na condução de informações para as vias ascendentes e descendentes do cérebro – e para regiões do mesencéfalo, para então, finalmente, permitir a execução motora do sistema visual, ou seja, permitir os movimentos verticais e horizontais dos olhos e a ativação da musculatura da cabeça, respectivamente.

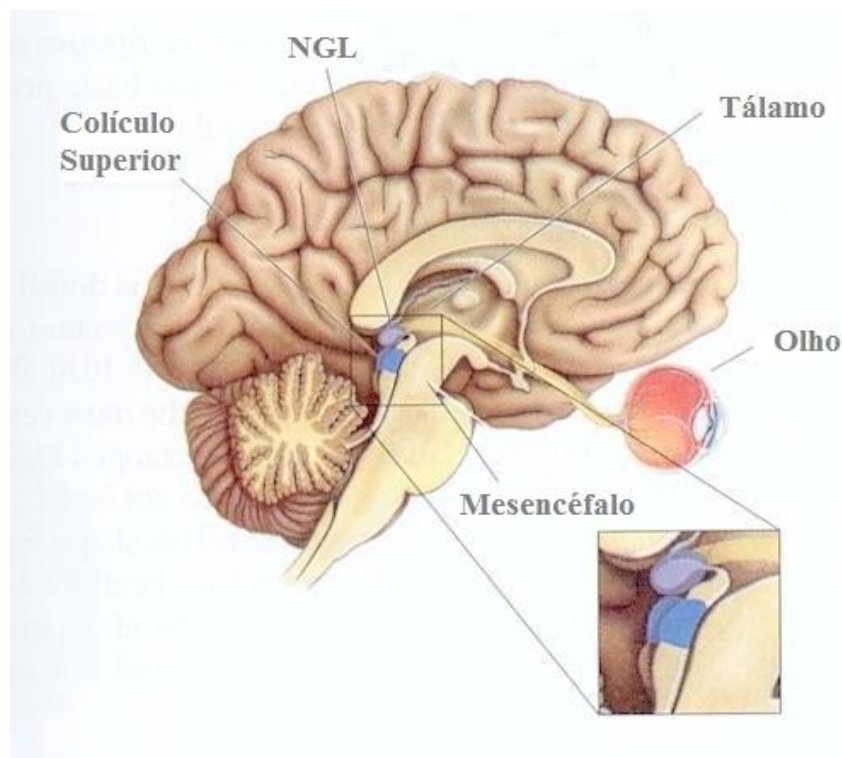


Figura 9 - Localização do Colículo Superior no cérebro humano
Fonte: adaptado de http://editthis.info/images/psy3242/b/bc/Superior_col.jpg

Fisiologia

A camada superficial do SC é uma estrutura interconectada por outras duas camadas que possuem dois principais tipos de células – um tipo com seu campo receptivo largo (*wide-field*) e outro tipo com seu campo receptivo estreito (*narrow-field*).

As camadas intermediária e profunda são formadas de células com variados tipos de ativação, embora sejam morfologicamente similares. Neurônios de excitação encontrados na camada intermediária são classificados como *Quasi-Visual* (QV), *Burst* e *Buildup*. Neurônios de inibição são encontrados em ambas as camadas (intermediária e profunda).

Os neurônios do tipo *Burst* enviam sua saída para o núcleo de controle dos movimentos oculares, no tronco cerebral. Eles atuam como neurônios reguladores quando diretamente estimulados, mas exibem um comportamento explosivo quando recebem informações da camada superficial, enviadas pelos neurônios do tipo *Quase-Visual* e *Buildup*. Neurônios do tipo *Burst* possuem receptores glutamatérgicos – ativados por um grupo de aminoácidos do tipo glutamato, sendo o principal neurotransmissor do SNC.

Os neurônios do tipo *Buildup* possuem interligações de curto alcance, e propagam sua ativação durante os movimentos oculares sacádicos, embora este propósito ainda seja incerto.

Além das informações inibitórias que chegam pela camada intermediária do SC, os neurônios do tipo *Burst* também recebem informações nervosas de áreas cerebrais que estão diretamente relacionadas com os movimentos oculares sacádicos, como a Formação Reticular Mesencefálica Central (cMRF), agindo como reguladora da energia emitida pelas sacadas e do tipo do movimento ocular, através de mecanismos inibitórios presentes nas camadas intermediária e profunda do SC. Sendo assim, podemos concluir que o SC é parte do ciclo da geração dos movimentos oculares sacádicos.

Pesquisas recentes mostraram que os neurônios do tipo *Burst*, na região rostral (ou zona de fixação) do SC geram microssacadas durante a fixação ocular, de forma similar com que as sacadas são geradas na região caudal do SC (Figura 10) [28].

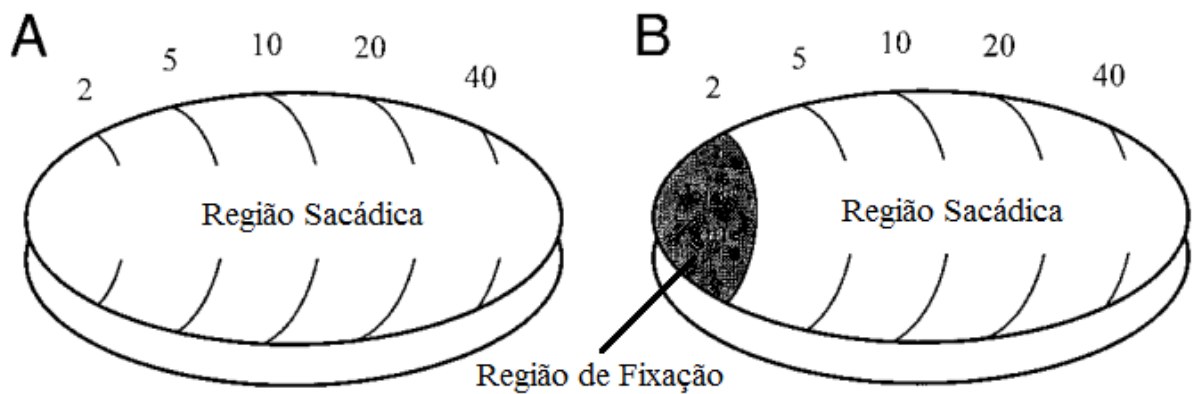


Figura 10 – Comparação entre as duas hipóteses das camadas intermediárias do SC. Os números representam as faixas de mesma amplitude sacádica no mapa motor.
 Fonte: adaptado de [31]

Esta é uma forte evidência de que ocorrem ativações assimétricas, em grande escala, propagando-se na camada intermediária do SC. Enquanto as conexões que ocorrem na camada superior do SC possuem longo alcance sináptico, as interconexões de excitação e de inibição ocorridas na camada intermediária apresentam um limite de 500 mm ou menos. Estas ativações assimétricas ocorrem a partir do mapeamento retinotópico das informações visuais que chegam ao SC através da retina.

2.6. Movimentos Oculares

A análise dos movimentos oculares é uma fonte importante de informação em estudos de aspectos cognitivo-comportamentais. Através do movimento conjunto da cabeça e do globo ocular, sob comandos do cérebro, a visão percebe características de movimento de um objeto sendo acompanhado pela fóvea. Ou seja, sua imagem, projetada na retina, permanece estacionária, excetuando-se os movimentos sacádicos superpostos ao movimento médio do objeto. Nestas condições, as características de resolução espacial da visão são mantidas, apesar do movimento.

Os movimentos oculares são comportamentos que podem ser medidos, e sua medição provê uma forma sensível de aprender sobre o processamento cognitivo e visual. Apesar do considerável tempo de pesquisas sobre os movimentos oculares, foi apenas nas últimas

décadas que sua medição levou a importantes descobertas sobre os processos psicológicos que ocorrem durante tarefas como leitura, pesquisa visual e percepção da cena.

2.6.1. Características Básicas dos Movimentos Oculares na Cognição visual

Embora tenhamos a sensação de que é possível processar todo o campo visual em uma única fixação, na realidade, o cérebro seria incapaz de processar informações do ambiente fora da região central da fóvea, se fôssemos incapazes de mover nossos olhos [32].

Devido as limitações na acuidade visual da retina, os movimentos oculares são necessários para o processamento do campo visual. A nossa capacidade de discriminar detalhes no campo visual, cai acentuadamente fora da região central da fóvea (estendendo-se até 5° em cada lado do ponto de fixação) e na região periférica da fóvea.

Enquanto estamos lendo ou procurando um ponto de fixação em um alvo ou simplesmente olhando para uma nova cena, nossos olhos se movem a cada 200-350 ms. Estes movimentos oculares servem para mover a fóvea (a parte de alta resolução da retina compreendida em dois graus no centro do campo visual) para uma área de interesse, com o objetivo de processar essa região de forma bem detalhada.

Durante estes movimentos oculares – conhecidos como sacadas – a visão é eliminada e novas informações são adquiridas somente durante a fixação, ou seja, durante o período de tempo em que os olhos permanecem relativamente parados [33].

Embora seja verdade que podemos mover nosso centro de atenção, independentemente de onde os olhos estão fixados, a sensação que temos é de que isso não ocorre. A diferença entre a atenção e a fixação acontece frequentemente em simples tarefas [34], no entanto, em tarefas como leitura, busca visual e percepção de cena, a atenção escondida e a atenção evidente – localização exata do olho – estão fortemente ligadas.

Pelo fato dos movimentos oculares serem essencialmente motores, é necessário um tempo para planejar e executar uma sacada. Além disso, o ponto focal final é pré-selecionado antes do início do movimento.

Embora seja assumida a teoria de que os dois olhos se movem em sincronismo e que se fixam no mesmo ponto no espaço, movimento ocular conhecido como vergência, uma pesquisa recente demonstra claramente que este não é o caso, e os dois olhos são frequentemente desviados um do outro [35] [36].

Existem evidências consideráveis de que a natureza da tarefa influencia nos movimentos oculares. Um resumo da quantidade média de tempo gasto em cada fixação e a distância média que os olhos se movem durante as atividades da leitura, busca visual, percepção da cena são apresentados na tabela 1.

Tabela 1 - Características dos movimentos oculares durante a leitura, percepção de cenas e busca visual

Atividade	Período Médio de Fixação (ms)	Tamanho Médio das Sacadas (graus)
Leitura Silenciosa	225 - 250	2 (8 - 9 caracteres)
Leitura Oral	275 - 325	1.5 (6 - 7 caracteres)
Percepção de Cena	260 - 330	4
Busca Visual	180 - 275	3

A partir das informações presentes na tabela 1 é possível perceber que, apesar dos valores apresentados serem bastante distintos nas diferentes atividades, há uma faixa intermediária de duração da fixação dos olhos e, para cada tarefa, há uma variação considerável na combinação entre o período de fixação e o comprimento das sacadas.

2.6.2. Movimentos Oculares Durante a Busca Visual

O período de tempo das fixações pode ser imensamente variável. Sabe-se que o período médio das fixações de curta duração é em torno de 180 ms, enquanto a mesma medida, das fixações de longas durações é de 275 ms. Sem dúvida alguma, isto ocorre devido ao fato de que o nível de dificuldade da pesquisa do objeto de interesse (dependendo da densidade e desordenação da imagem como um todo), e o tipo / objetivo da pesquisa, influenciarão fortemente na duração das fixações em cada item.

Normalmente o tamanho das sacadas são pouco maiores que na leitura, embora as sacadas podem ser bem pequenas em imagens densamente complexas e confusas. Quando a cena é muito complexa (com muitos objetos e confusa) a pesquisa torna-se mais refinada do que em cenas simples. À medida que a cena torna-se mais complexa e confusa, ocorre um aumento na duração e no tempo das fixações e uma redução no tamanho médio das sacadas.

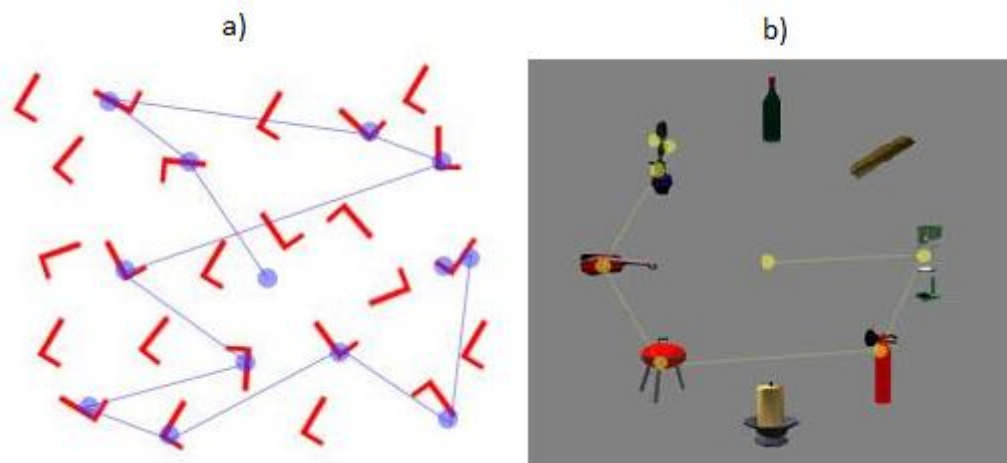


Figura 11 - Comparação entre a) e b) cujos distratores são muito semelhantes ou muito distintos, respectivamente

Fonte: adaptado de <http://www.scholarpedia.org/w/images/0/01/Search.jpg> e <http://www.scholarpedia.org/w/images/1/12/Visualesearch.jpg>

Pesquisas também mostraram que, na busca visual, o espaço perceptivo varia em relação aos objetos distratores, isto é, o espaço perceptivo é menor quando os objetos distratores são bem semelhantes ao objeto de interesse, e maior caso contrário. A figura 11 ilustra esta propriedade da busca visual: na figura a), foi questionado ao observador que localizasse as letras “L” que estivessem invertidas; na letra b), o observador foi alertado de que deveria encontrar um jarro contendo a planta cacto.

Isto sugeriu que haviam duas regiões qualitativamente diferentes dentro do intervalo:

- Região de Decisão: quando, de fato, a informação sobre a presença ou a ausência do objeto de interesse é alcançada;
- Região de Pré-Visualização: onde algum item informativo está disponível, mas a informação sobre a ausência do objeto de interesse ainda não foi alcançada.

Predizer a região com maior probabilidade de onde o observador fixará seus olhos em uma busca visual, é frequentemente descrito utilizando o mapa de saliência [45]. O Modelo de Pesquisa Guiada é um dos mais conhecidos na literatura e tem despertado um grande interesse nos mecanismos de busca visual. De acordo com este modelo, a orientação da pesquisa de uma cena, basicamente é feita de duas maneiras:

- *bottom-up*: prioriza a pesquisa pelos objetos que mais se diferenciam dos seus vizinhos;
- *top-down*: prioriza a busca de acordo com o quanto os objetos pesquisados compartilham características comuns com o objeto de interesse.

2.7. Movimentos Oculares Sacádicos

Os movimentos sacádicos são os mais rápidos movimentos oculares. Sua principal função é redirecionar o eixo visual para um novo local de interesse, mantendo-o alinhado com a fóvea. Os comandos cerebrais enviados para os músculos oculares resultam em uma rápida rotação dos olhos até que permaneçam parados em sua nova posição [46].

Estes movimentos são conhecidos como sacadas ou movimentos oculares sacádicos. As sacadas direcionam a fóvea a um objeto ou uma região de interesse, possibilitando uma alta acuidade visual e então permitindo um grande nível de detalhamento da região alvo. Normalmente, o sistema visual executa várias sacadas por segundo e suas direções são selecionadas por um processo cognitivo no cérebro, sem envolvimento algum da consciência. A visão depende da informação que é adquirida durante as fixações entre as sacadas. Nenhuma informação visual é útil durante os movimentos sacádicos dos olhos.

Têm-se aprendido bastante sobre o controle das sacadas através de estudos detalhados em simples situações onde um observador é orientado a mover seus olhos para um objeto novo em uma cena. Nesta situação, é possível medir o tempo entre o surgimento do alvo sacádico e o início da sacada. Também é possível medir com precisão as métricas sacádicas, isto é, a dimensão e a orientação usando a medida angular para descrever a rotação ocular (ângulo visual em graus). Desta forma, torna-se factível medir detalhadamente a progressão de um único movimento sacádico e toda a trajetória feita entre as sacadas.

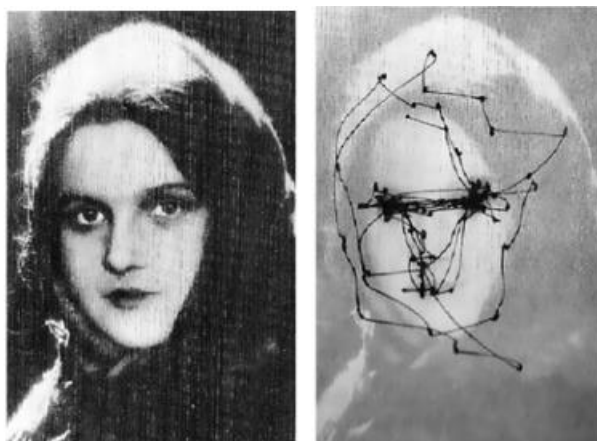


Figura 12 - Padrões de sacadas quando uma imagem de um rosto é observada por 3 minutos
Fonte: adaptado de http://www.scholarpedia.org/w/images/7/72/Fig1_Yarbus115.jpg

Surpreendentemente as sacadas mostram um padrão similar. A duração das sacadas aumentam, de forma não linear, assim como sua amplitude (tamanho), em cerca de 20 ms para os movimentos menores e aumentam em torno de 100 ms para os maiores movimentos.

A velocidade com que os olhos se movem mostra um aumento padrão, com a velocidade máxima em torno de 10 graus / s para sacadas muito pequenas (movimentos microscádicos) e mais de 300 graus / s para grandes movimentos scádicos. A trajetória percorrida pela rotação dos olhos é semelhantemente estereotipada, embora esteja sujeita a variações de origem desconhecidas.

Este comportamento estereotipado não se aplica apenas para direcionar sacadas dentro de determinados contextos visuais, mas sim sobre toda e qualquer situação onde as sacadas são executadas. Pequenos desvios sistemáticos são percebidos, por exemplo, sacadas executadas em ambientes com pouca iluminação (ou até mesmo com ausência total de luz) são mais lentas do que as feitas em ambientes com iluminação.

Outra grande característica encontrada nas sacadas é o comportamento balístico. Em teoria, o movimento balístico é aquele onde o seu destino está previamente determinado desde a sua origem. Os Movimentos Scádicos foram biologicamente modelados para possuírem esta característica. Uma nova informação visual apenas influenciará na sacada se esta for processada pelo cérebro com um mínimo de 70 ms antes do início do movimento. Entretanto, durante sua trajetória, as sacadas mostram uma pequena estabilização para o alvo scádico.

Isso demonstra que há um sinal de referência interno usado para fazer pequenas correções durante o processo de geração da sacada [46].

Sacadas Corretivas e Paralelismo Sacádico

Durante as sacadas percebeu-se que frequentemente, uma pequena segunda sacada corretiva é efetuada, objetivando aproximar o olhar para o alvo sacádico. Estas sacadas corretivas ocorrem tipicamente após a sacada primária, com um atraso extremamente pequeno. Este curto intervalo de atraso é chamado de ISI (*Inter-Saccadic Interval* – Intervalo Inter-Sacádico). O ISI entre as sacadas evidenciou a existência da programação paralela dos movimentos sacádicos. Sacadas corretivas podem acontecer antes mesmo que o alvo sacádico desapareça e os olhos não o alcancem. Todavia, a probabilidade de que sua execução aconteça é maior quando o alvo está visível. Durante a leitura e outras tarefas visuais, ocorrem pequenas sacadas similares de correção. As sacadas corretivas são frequentes pelo fato de que, em média, a precisão sacádica está entre 5% a 10% da amplitude visual. Entretanto, mecanismos adaptativos operam sobre os múltiplos movimentos para manter o sistema sacádico calibrado por um longo período [47].

2.7.1. Classificação dos Movimentos Oculares Sacádicos

Os movimentos oculares sacádicos são classificados em 4 categorias: reflexivos e voluntários, atrasados e direcionados pela memória, anti-sacadas e microsacadas.

- **Reflexivos e Voluntários:** orientação e condução visual para um alvo recentemente novo em uma cena é completamente natural pelo movimento ocular sacádico. Uma sacada feita para um alvo que surge repentinamente em uma cena é chamada de Reflexiva (ou descoberta de estímulo) em comparação com as sacadas feitas em situações que dependem do controle de processos cognitivos (por exemplo, quando efetuamos um movimento ocular sacádico do tipo “olhar para a esquerda”) chamada de Voluntária. A forte distinção de ambos os contextos é percebida pelas suas características comportamentais básicas [48]. No entanto, a rigor, todos os movimentos sacádicos são considerados de natureza voluntária, assim como o observador pode decidir por não fazer quaisquer movimentos com seus olhos. Além

disso, se o local e o tempo do aparecimento de um alvo sacádico podem ser previsto, uma sacada antecipada pode frequentemente aparecer antes mesmo que o destino seja exibido ou tenha iniciado com muita rapidez logo após a ocorrência da orientação visual. No dia-a-dia, os olhos estão frequentemente muito ativos, efetuando inúmeras sacadas por segundo enquanto escaneia uma cena.

- **Atrasados e Direcionados pela Memória:** neste contexto, os observadores são orientados a efetuarem uma sacada para um estímulo presente na região periférica, logo após um curto período de atraso (alguns segundos), e frequentemente a origem deste deslocamento dá-se a partir do ponto central da fixação. Desta forma, o paradigma das sacadas atrasadas tem sido utilizado para investigar profundas perdas das funções motoras e comportamentais humanas, tais como o mal de Parkinson e a esquizofrenia [49]. A sacada direcionada (ou guiada) pela memória é similar a Atrasada, mas neste caso, o alvo sacádico é visualizado muito brevemente de modo que as sacadas seguintes são direcionadas para a região do alvo. Estas sacadas tendem a apresentarem um decréscimo na precisão, em indivíduos normais e principalmente em pacientes com problemas no ganglio basal e regiões do lobo frontal, envolvidas no processo da memória de trabalho.
- **Antisacadas:** a tarefa da antisacada foi introduzida por Hallett (1978) e desde então tornou-se um dos principais conceitos de sacadas endógenas [50]. A antisacada é um movimento ocular gerado deliberadamente em direção oposta a um alvo visual subitamente apresentado. É na verdade a possibilidade de desviar a atenção, sem executar um desvio ocular a um estímulo visual inesperado. Inicialmente, o observador necessita localizar o alvo e, em seguida, determinar a direção da antisacada sendo que, para tal, deverá eliminar a sacada reflexiva na direção do alvo. Possuem mais tempo de latência (em torno de 250-350 ms) do que as sacadas feitas para o alvo sacádico (média entre 150-250 ms) e também possuem uma quantidade razoável de erros que são observados durante os movimentos antisacádicos, nas situações em que o observador não consegue ter uma tomada de decisão ao objetivo. Nós a utilizamos com muita frequência quando, por exemplo, dirigimos um veículo focando os olhos na estrada e não desviando a atenção na direção de muitos estímulos visuais que surgem incessantemente, ao longo do caminho, estimulando nosso campo visual.

- **Microssacadas:** Embora se assuma que nossos olhos não se movimentam quando estão fixos em um alvo sacádico, na verdade eles estão em constantes movimentos em pequena escala, apresentando tremores e direções irregulares, intercalados por minúsculos movimentos sacádicos (menores que 0.5 graus). Estes movimentos oculares estabilizadores são imperceptíveis, do ponto de vista do observador, e essenciais para que nossa percepção visual não diminua [51], sendo necessários para o preenchimento da visão periférica quando os olhos permanecem fixados no alvo. As microssacadas ocorrem entre uma e duas vezes a cada segundo. Nos últimos anos, pesquisadores relacionaram os movimentos oculares microssacádicos com o problema da ilusão visual em determinadas cenas, como as imagens enigmáticas (figura 13).

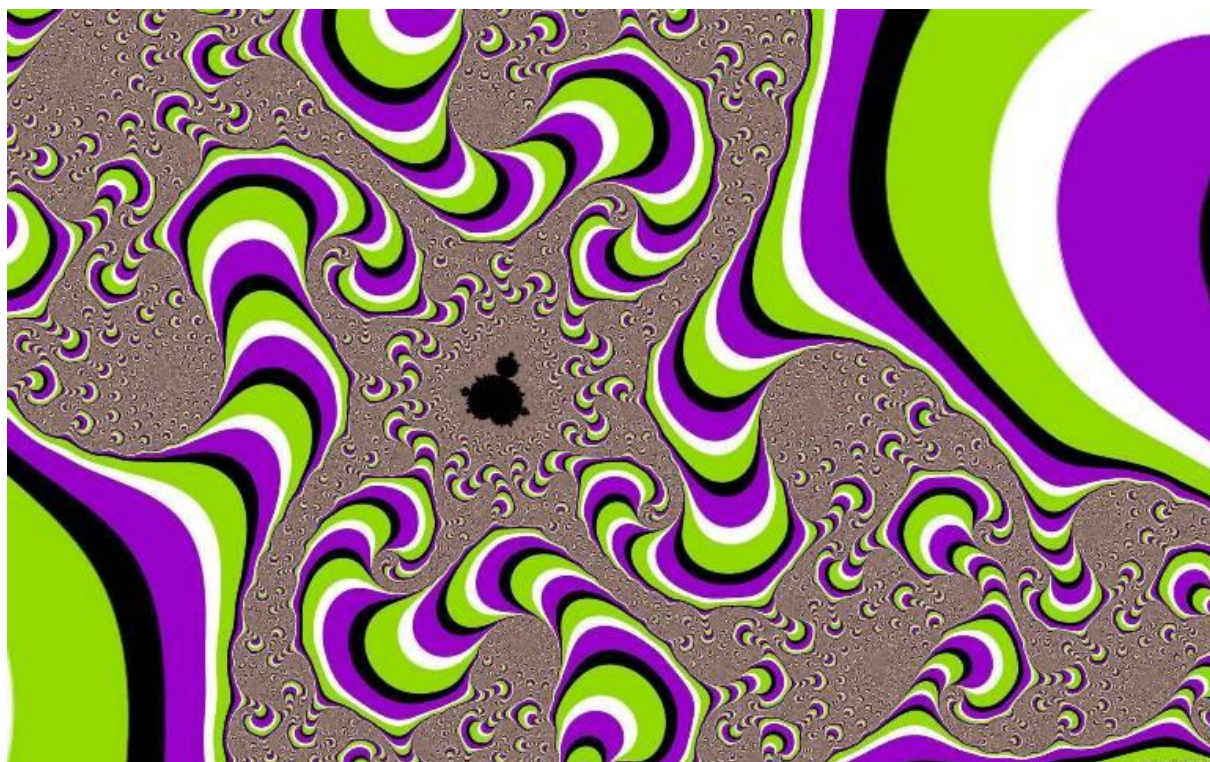


Figura 13 - Ilusão de ótica ocasionada pelas microssacadas
Fonte: retirado de <http://upix.me/files/B8gqbe/mOsLH3x-05212013135927000000.jpg>

Recentes estudos desenvolvidos por pesquisadores do *Barrow Neurological Institute*, sugerem que as microssacadas são responsáveis pelo fenômeno da ilusão visual, causados quando visualizamos imagens que contém repetitivos arranjos de gradientes luminosos, assim como na figura 13, provocando o movimento ilusório conhecido como Ilusão das Serpentes Giratórias (*Rotating Snakes Illusion*). De acordo com relatórios recentes, movimentos

oculares lentos – um tipo de movimento ocular não sacádico – leva a ilusão visual quando as microscadadas, responsáveis pela definição da imagem periférica durante a fixação visual, não conseguem fazê-lo [25].

3. MODELO MATEMÁTICO-COMPUTACIONAL SACÁDICO

Como mencionado no capítulo anterior, o subsistema de movimentos oculares sacádicos do sistema visual dos mamíferos é o principal responsável por apontar a fóvea para objetos de interesse [12]. A fóvea é a região central da retina, que tem a mais elevada densidade de receptores e, portanto, proporciona a maior acuidade visual. Estas informações são enviadas a uma região do cérebro chamada Córtex Visual Primário (V1) onde ocorre a transformação destas informações. O sistema sacádico produz movimentos rápidos dos olhos (sacadas) que mudam a fóvea rapidamente para um alvo visual (alvo sacádico) no campo visual. O objetivo das sacadas é mover os olhos tão rapidamente quanto possível. Como não há tempo de uma resposta visual detalhada de todo o ambiente, afim de modificar significativamente o trajeto de uma sacada e as correções para a direção do alvo sacádico, normalmente são executadas sucessivas sacadas [20].

Um modelo computacional capaz de realizar as atividades mencionadas deverá simular as principais características encontradas na fisiologia da visão humana, desde a obtenção das informações visuais captadas pela retina, através da reflexão da luz, transportando estas informações nervosas através do Núcleo Lateral Geniculado até chegar ao V1, onde ocorre o processamento retinotópico destas informações, mapeando a cena externa para que estas informações sejam enviadas para o Colículo Superior (SC) onde será estimado o próximo centro de atenção (alvo sacádicos).

Baseado nestas premissas, o modelo proposto obtém as informações visuais através de uma câmera de vídeo, de onde são retirados *frames* (quadros de cenas) obtidos de uma gravação prévia feita por uma câmera, representando imagens monoculares (2 dimensões) de maneira análoga a representação visual da retina de um olho biológico. Estas imagens são computacionalmente escaladas e filtradas representando, respectivamente, a aproximação ou afastamento e ajuste de foco no contexto biológico. Após estes tratamentos nas imagens, ou

seja, nas informações retinotópicas, é realizada a transformação da imagem, que atualmente encontra-se representada em um plano cartesiano oriunda da câmera de vídeo, para o plano de coordenadas polar através da função de transformação log-polar, simulando a transformação retinotópica das informações oriundas da retina ao Córtex Visual Primário (V1). Após esta transformação, o modelo apresentado é capaz de aprender um conjunto de características de interesse presentes na imagem, utilizando uma metodologia computacional baseada no comportamento neurológico chamada Redes Neurais Sem Peso. Nesta seção serão apresentadas estas habilidades computacionalmente modeladas fundamentadas no comportamento biológico da visão.

3.1. Mapeamento Retinotópico (Log-Polar)

O mapeamento retinotópico está presente em muitas, possivelmente todas, as áreas visuais do córtex cerebral além do V1 [41]. Estes mapeamentos também ocorrem em estruturas não corticais como o SC, sendo este responsável por controlar os movimentos motores dos olhos durante as sacadas, que são direcionados para uma região de interesse. Estes movimentos requerem grandes detalhes das informações visuais, captadas pelos olhos, para que sejam completamente precisos. Estas informações visuais são transformadas em impulsos elétricos pela retina que, por meio do nervo óptico, são projetados para o Córtex Visual Primário (V1) e, após sofrer algumas transformações, são enviados ao SC e outras áreas cerebrais [12].

Em 1982 o psicólogo e pesquisador Roger Tootell, juntamente com uma equipe de outros pesquisadores, fez experiências com alguns primatas com o objetivo de entender e analisar a organização retinotópica presente na fisiologia da visão desta espécie. Utilizando uma substância química composta de moléculas D-2-deoxiglicose – método que, na ocasião, o próprio pesquisador classificou-o como superior aos demais métodos por fornecer uma visão bastante detalhada e completa da organização retinotópica – em um dos olhos do primata, mantendo-o anestesiado e paralisado, e utilizando padrões de procedimento ópticos, levaram o centro da imagem da tela apresentada na imagem (a) da figura 14 – dotada de 8 raios e 3 círculos concêntricos – até a fóvea de um dos olhos do primata. Após um período aproximado entre 25 à 30 minutos, o animal foi sacrificado e córtex foi cortado de forma histológica, aplainado para que então sua superfície fosse analisada. Os pesquisadores concluíram que a

transformação espacial do campo visual para o córtex V1 foi representada analiticamente como um complexo mapeamento logarítmico.

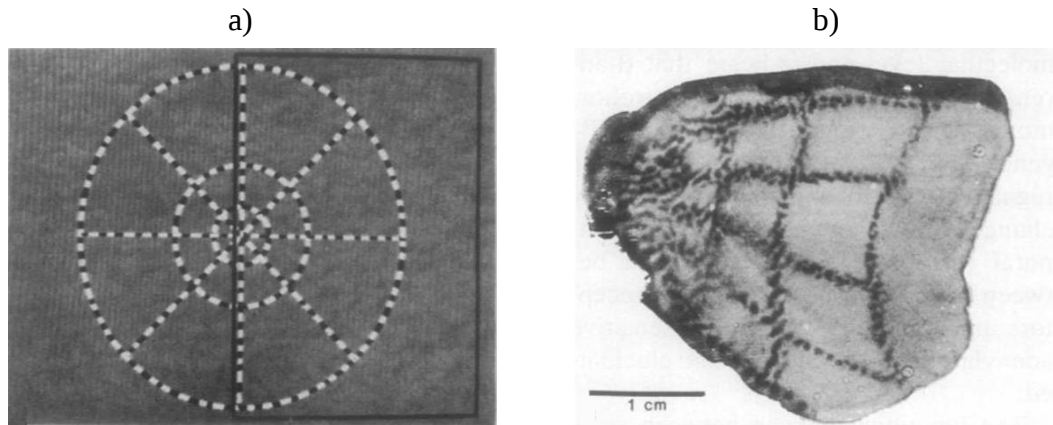


Figura 14 - Representação do mapeamento retinotópico no córtex visual de um primata, em b), à partir da imagem a)

O resultado da análise pode ser visualizado na figura 14. Nota-se que as linhas concêntricas que estavam apresentadas na tela do monitor (imagem (a)) transformaram-se em traços quase retos no córtex V1 (b). É possível perceber também que a região central de (a), onde estão os círculos mais internos, possuem uma representação bem maior no V1 (b), possibilitando uma alta acuidade visual nesta região, que é exatamente para onde a fóvea do primata estava focada.

Baseado neste comportamento biológico é possível representar matematicamente este modelo utilizando o plano polar de coordenadas (ρ , θ), conforme mostra a figura 15, no qual simulamos o mapeamento das informações visuais que chegam da retina para o córtex visual primário, onde ocorre o mapeamento retinotópico das mesmas, para que então sejam levadas ao SC e para outras regiões do cérebro.

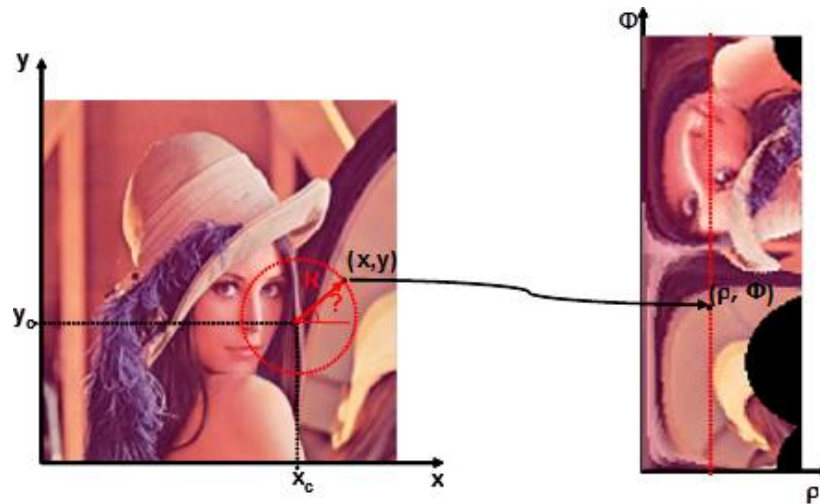


Figura 15 - Transformação log-polar (mapeamento retinotópico)

O mapeamento das informações visuais da retina, passando pelo V1 até chegarem ao colículo superior, segue a função log-polar [13]. A figura 15 mostra a transformação log-polar na imagem, centrada no ponto (x_c, y_c) – este ponto corresponde ao centro de atenção, no campo visual. É possível perceber que o círculo (em vermelho) na imagem à esquerda da figura 15 torna-se uma linha reta na imagem à direita e as regiões em torno do centro do círculo (a fóvea do modelo), na imagem da esquerda, ocupam uma área muito maior na imagem à direita. Esta simulação é análoga a descoberta feita por Tootell, detalhada no começo desta seção.

Um grande benefício apresentado neste modelo matemático é a qualidade da compressão de dados que, quando comparada a imagens cartesianas (como em imagens bidimensionais figura 16 (a)) as imagens log-polares permitem rápidas taxas de amostragem sem perder a alta resolução na área da fóvea, conforme podemos visualizar na figura 16 c).



Figura 16 – Transformação da imagem cartesiana (a) para o plano log-polar (b) e em seguida, em (c), a representação inversa (retinal)

A figura 16 mostra as etapas da transformação log-polar partindo de uma imagem cartesiana de coordenadas x e y . A imagem cartesiana bidimensional (a) possui uma resolução de 128×128 pixels. Após sofrer a transformação log-polar (b), a mesma passou a ter uma resolução de 64×32 pixels, o que demonstra uma redução na resolução inicial, atingindo uma desempenho 8 vezes maior que a projeção cartesiana. A imagem (c) mostra o mapeamento inverso da projeção log-polar em um plano cartesiano. É possível notar, em (c), que a região da fóvea (centro de atenção) possui uma alta resolução em relação às regiões periféricas [42].

A modelagem matemática da transformação log-polar, comumente utilizada na literatura, é dada por:

$$R = \sqrt{(x - xc)^2 + (y - yc)^2} \rightarrow \rho \propto \log(R) \quad (1)$$

$$\theta = \arctan\left(\frac{(y - yc)}{(x - xc)}\right) \rightarrow \phi \propto \theta \quad (2)$$

Neste trabalho, a transformação log-polar não está sendo utilizada exatamente como mostrada acima. Uma variante foi criada para gerar uma melhor precisão no mapeamento da retina para o SC [20]. A figura 17 mostra esta variante da transformação log-polar, onde regiões vizinhas em torno do centro do círculo (a fóvea do modelo) também são vizinhas na transformação log-polar (retinotópica), como ocorre no SC. Isto não ocorre na transformação padrão da literatura, ilustrada na figura 15.

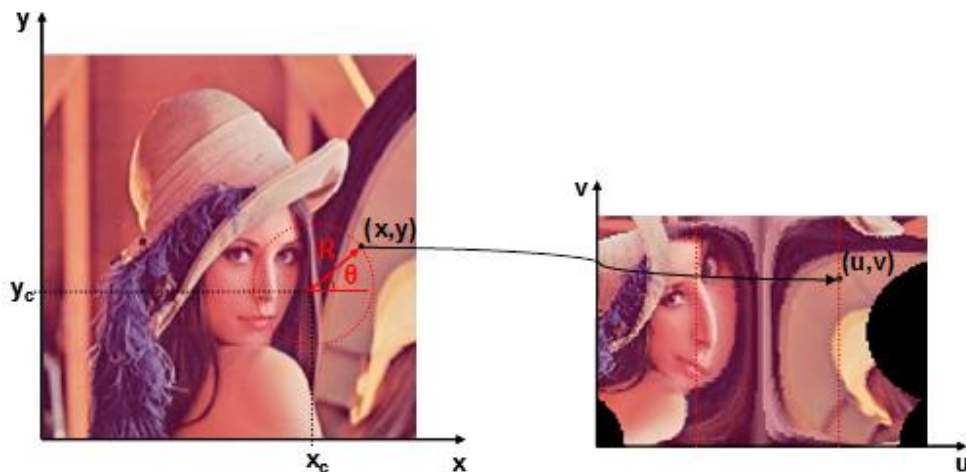


Figura 17 - Nossa variante da transformação log-polar

Na seção anterior vimos que nosso córtex visual é segregado em dois hemisférios, direito e esquerdo. A figura 17 representa nossa variante da transformação log-polar com o objetivo de representar os hemisférios do córtex V1, evitando perdas de informações nas regiões vizinhas da fóvea.

3.2. Redes Neurais Sem Peso e VG-RAM

Estudar o processo cognitivo do cérebro humano tem despertado grande interesse pelos pesquisadores nos últimos anos. O cérebro humano é capaz de executar tarefas como reconhecimento de padrões, percepção e controle motor, com muito mais eficiência e desempenho em relação aos mais rápidos dos computadores atuais [43]. Para que seja possível aplicar o conceito biológico dos movimentos sacádicos, planejados e executados pelo Colículo Superior (SC), utilizamos uma metodologia computacional baseada no comportamento neurológico, chamada Redes Neurais Artificiais.

Redes Neurais Artificiais (RNAs) consistem em modelar computacionalmente o comportamento cognitivo do cérebro humano, sendo este, composto por bilhões de neurônios – estima-se que um humano adulto possui 10^{11} neurônios – que reagem a estímulos oriundos do próprio cérebro (como na tarefa de pensar) ou pelos sensores biológicos (visão, audição, tato, entre outros). A comunicação entre estes neurônios é essencialmente fundamental para as funções cerebrais. Conforme mostra a figura 18, basicamente os neurônios biológicos são compostos por 3 estruturas: corpo celular ou soma, axônio e dendritos, e são interligados através das sinapses.

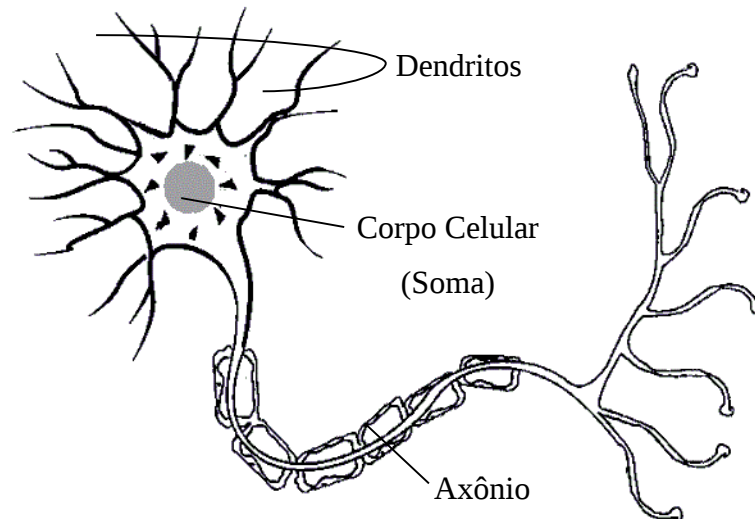


Figura 18 – Estrutura física do neurônio biológico.
 Fonte: adaptado de <http://faculty.washington.edu/chudler/gif/colorb1.gif>

Os neurônios recebem informações (estímulos nervosos) de outros neurônios através dos dendritos. Cada neurônio possui um grande número de dendritos (possivelmente dezenas de milhares) e enviam estas informações nervosas para o corpo celular (ou soma) onde estas informações sofrem reações eletroquímicas, resultando em uma série de pulsos elétricos (chamado de potencial de ação), que são propagados para outros neurônios através do axônio [43]. A interconexão sináptica entre os neurônios ocorre na ligação do axônio de um neurônio com o dendrito de outro neurônio, causando excitação ou inibição.

Baseado neste comportamento biológico, durante a década de 40 os pesquisadores McCulloch e Pits desenvolveram, de forma bastante simplificada, um modelo matemático do neurônio biológico. Este modelo é utilizado em diversas aplicações que envolvem esta técnica computacional, podendo ser descrito pelos termos das equações (1) e (2). Basicamente estas operações efetuam a soma ponderada de todas as possíveis entradas de um neurônio e em seguida submetem este resultado, chamado de ativação, a uma função de ativação. Esta função gera uma saída chamada de Potencial de Ação.

$$u = \sum_{j=1}^n w_j x_j \quad (1)$$

$$y = \varphi(u + b) \quad (2)$$

Onde x são os possíveis valores de entradas (sinapses); w são os pesos atribuídos a cada entrada (pesos sinápticos); u , valor de ativação do neurônio, é o resultado da combinação linear entre os pares de pesos e entradas; b (bias) é um parâmetro externo que funciona como um limiar do potencial de ativação do neurônio; $\varphi(.)$ é a função de ativação e y é o valor da saída [43]. A figura 19 mostra a estrutura física de um neurônio artificial.

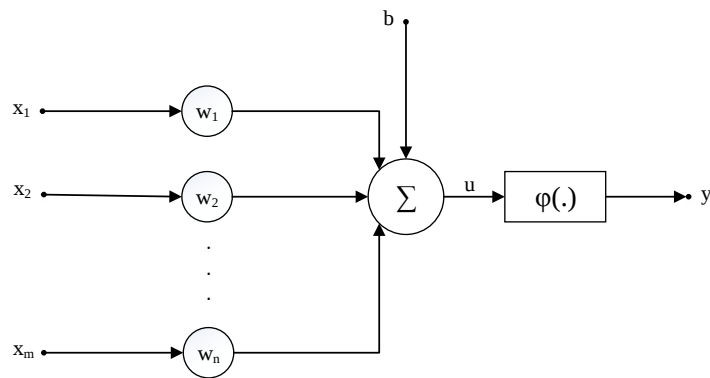


Figura 19 - Modelo físico de um neurônio artificial

Na literatura podemos encontrar variadas arquiteturas baseadas nas redes neurais artificiais para solucionar problemas de reconhecimento de padrões. Neste trabalho, utilizaremos as Redes Neurais Sem Peso baseadas em neurônios do tipo VG-RAM.

Utilizando uma abordagem diferente das demais metodologias, as Redes Neurais Sem Peso (RNSP) – comumente encontrada na literatura como Método N-Tuplas ou ainda Rede Neural Baseada em Memória de Acesso Randômico (RAM - *Random Access Memory*) – como o próprio nome sugere, não utilizam pesos em suas entradas (sinapses). Basicamente, estas arquiteturas aprendem a reconhecer um padrão utilizando funções e operações lógicas. Um determinado padrão apresentado a uma RNSP é descrito de forma lógica. Cada função lógica avalia a entrada apresentada e retorna um valor booleano (verdadeiro ou falso) [44].

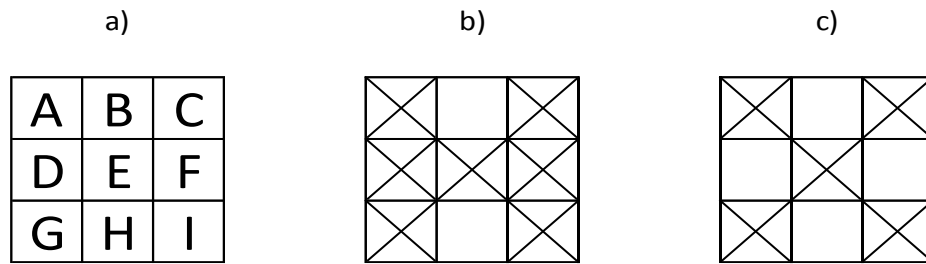


Figura 20 - Processamento básico de uma RNSP

A figura 20 mostra um exemplo básico de classificação de uma RNSP do tipo RAM. A figura 20 (a) representa uma matriz bidimensional 3×3. Cada célula possui uma função lógica que avalia seu valor apresentado e retorna um valor booleano. Todos os valores booleanos retornados formam um padrão. Por exemplo, a figura 20 (b) é representada pela função lógica abaixo:

$$R_{3b)} = A \cdot \neg B \cdot C \cdot D \cdot E \cdot F \cdot H \cdot \neg H \cdot I$$

Da mesma maneira apresentada para 20 b), a figura 20 c) é representada pela função abaixo:

$$R_{3c)} = A \cdot \neg B \cdot C \cdot \neg D \cdot E \cdot \neg F \cdot H \cdot \neg H \cdot I$$

Para aumentar a capacidade de generalização da rede, ao invés de juntar todos os mínimos termos lógicos mencionados acima as entradas são agrupadas em tuplas e representadas pela soma dos termos, quando estes retornarem verdadeiro na função lógica. Por exemplo: na figura 20, existem 3 tuplas contendo 3 informações lógicas em cada uma. Neste caso, teríamos:

- Tupla 1: $A \vee B \vee C$
- Tupla 2: $D \vee E \vee F$
- Tupla 3: $G \vee H \vee I$

Esta abordagem foi apresentada por Aleksander e Stonham, em 1966, onde uma determinada classe é representada por funções lógicas. O modelo utiliza essencialmente um decodificador binário 1-para-N para que sejam armazenadas todas estas combinações lógicas de cada tupla e, em seguida, define uma unidade para o resultado encontrado [44]. Em outras palavras, o decodificador lógico retorna um valor binário de ordem 2^n (onde n é o número de tuplas) e atribui um identificador para este resultado. Quando um novo grupo de dados (tupla) for apresentado a esta arquitetura a mesma operação é executada, mas neste caso os dados antes decodificados são acessados e somados a esta nova entrada para que seja feito o cálculo de aproximação a uma determinada classe.

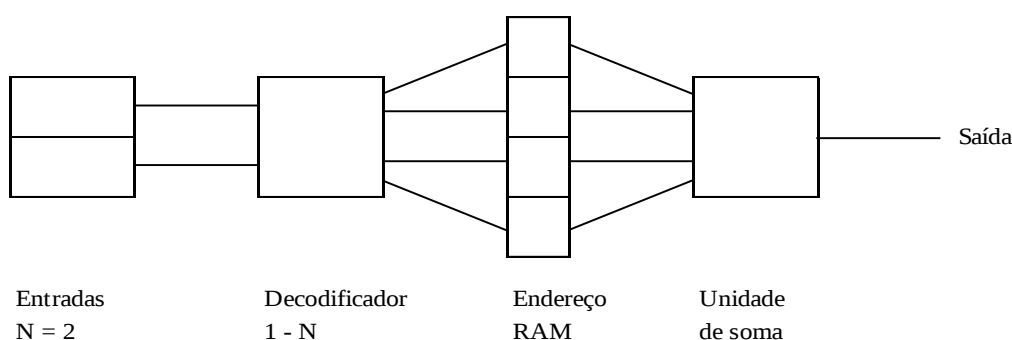


Figura 21 - Estrutura básica de um neurônio do tipo RAM

A figura 21 define um nó (ou neurônio) do tipo RAM. Neste exemplo é apresentada à rede, uma classe com duas tuplas ($N = 2$). Em seguida, as funções booleanas que compõem o decodificador binário calcularão as entradas apresentadas e armazenarão o resultado em um endereço do tipo RAM (o endereço armazenado terá quatro bits, pois conforme explicado anteriormente, o tamanho do endereço é de ordem 2^n). Este procedimento pode ser executado apenas uma vez durante a fase de aprendizado de cada classe. Para classificar uma nova entrada apresentada à rede, a unidade de soma encontra a classe previamente aprendida que apresenta um resultado mais próximo com a nova entrada. Tanto os decodificadores quanto os somadores (unidades de soma) podem assumir diferentes modelos lógicos e matemáticos. Utilizando este modelo de memória compartilhada (RAM), quando um valor é apresentado à rede, cada neurônio procura na memória de todos os neurônios da rede. A saída de cada neurônio VG-RAM da RNSP é feita a partir do par cuja entrada está mais próxima da entrada apresentada. A função de distância empregada pelos neurônios das VG-RAM é a distância de

Hamming. Se houver mais do que um par com a mesma distância mínima a partir da entrada apresentada, a saída do neurônio é escolhida aleatoriamente entre estes pares [20].

Distância de Hamming (Unidade de Soma da RNSP)

Considerando o modelo de neurônio apresentado onde sua memória é compartilhada entre os demais neurônios que constituem a rede, a Tabela 2 mostra a tabela de consulta de um neurônio VG-RAM com três sinapses (X_1 , X_2 e X_3). Esta tabela de pesquisa contém três pares de entrada, que foram armazenados durante a fase de treinamento (entrada #1, entrada #2 e entrada #3). Durante a fase de teste, quando um vetor de entrada (*input*) é apresentado à rede, o algoritmo de teste da rede calcula a distância entre este vetor de entrada com cada uma das entradas dos pares de entrada-saída armazenados na tabela de consulta do neurônio. No exemplo da Tabela 2, a distância de Hamming a partir da entrada para a entrada #1 é dois, porque ambos os bits X_2 e X_3 bits não coincidem com o vetor de entrada. A distância da entrada #2 é um, porque X_1 é o único bit não coincidente. A distância da entrada #3 é três. Assim, para este vetor de entrada, o algoritmo avalia a saída do neurônio Y como o rótulo 2, uma vez que é o valor armazenado na saída da entrada #2 [20].

Tabela 2 - tabela de pesquisa do neurônio do tipo VG-RAM

	X_1	X_2	X_3	Y
<i>entrada #1</i>	1	1	0	<i>rótulo 1</i>
<i>entrada #2</i>	0	0	1	<i>rótulo 2</i>
<i>entrada #3</i>	0	1	0	<i>rótulo 3</i>
	↑	↑	↑	↓
<i>nova entrada</i>	1	0	1	<i>rótulo 2</i>

Uma característica importante encontrada nas redes do tipo RAM é que os neurônios são preenchidos por um subconjunto de amostras (tuplas) oriundas das classes apresentadas de maneira que a precisão de classificação da rede seja diretamente proporcional a este tamanho. Quanto maior o tamanho da tupla, maior a habilidade de classificação da rede. Outra

importante vantagem das RNSP's é o armazenamento de valores binários em seus nós, o que permite um rápido processamento na classificação dos possíveis padrões submetidos à rede.

Apesar das RNSP's baseadas em RAM serem conceitualmente dotadas de uma arquitetura simples, capaz de apresentar bom desempenho em tarefas de reconhecimento de padrões mostrando alto desempenho computacional no aprendizado da rede e de simples implementação se comparada às demais arquiteturas neurais, a memória física presente nas atuais arquiteturas computacionais pode tornar-se um problema dependendo do tamanho da tupla (tamanho de N) pelo fato de que sua representação é de ordem 2^n .

4. DETECÇÃO DE PLACAS DE TRÂNSITO COM RNSP VG-RAM

Utilizando os conceitos apresentados nas seções anteriores desenvolvemos um sistema para detecção de placas de trânsito, em imagens bidimensionais extraídas de ambientes reais, com o objetivo de solucionar a tarefa em questão. Sumariamente, este sistema possui três módulos principais:

- Módulo de Detecção de Placa de Trânsito (*Traffic Sign Detection – TSD*)
- Módulo de Inferência Bayesiana (*Bayesian Inference – BI*)
- Módulo de Centralização de Placa de Trânsito (*Traffic Sign Centralization – TSC*)

O módulo TSD foi desenvolvido utilizando os conceitos de RNSP e VG-RAM com o objetivo de detectar placas de trânsito em imagens bidimensionais. TSD é treinado com mínimos exemplares de placas e, durante a busca por placas nas imagens, geram um conjunto de informações. O módulo BI é responsável por tratar estas informações oriundas de TSD inferindo, através do Teorema de Bayes aplicado a algumas variáveis contidas no conjunto de dados de TSD, detecções realizadas que de fato possuem grande probabilidade de ser uma placa de trânsito na imagem. Por fim, o módulo TSC recebe estas informações filtradas por BI e explora a simetria presente na forma geométrica de uma placa de trânsito, possivelmente detectada. TSC garante que uma placa foi detectada pelo fato de que toda placa de trânsito possui a forma física de uma figura geométrica primitiva, sendo simétrica em vários pontos. Sucintamente, o diagrama da figura 22 apresenta o fluxo de trabalho do sistema desenvolvido.

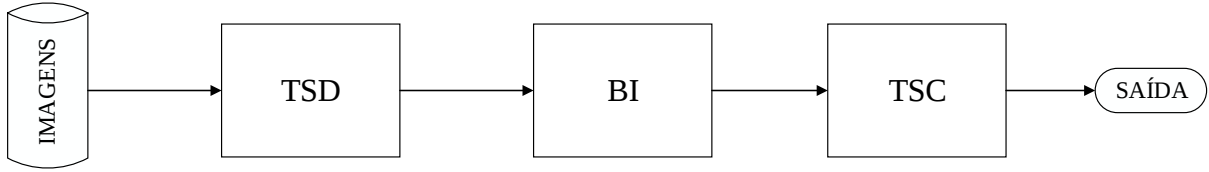


Figura 22 - Fluxo de trabalho da solução proposta

Nesta seção aprofundaremos o conhecimento referente aos módulos mencionados, assim como às demais estruturas que compõem a arquitetura do sistema.

4.1. Módulo de Detecção de Placas de Trânsito (TSD)

TSD (*Traffic Sign Detection*) é o principal módulo do sistema proposto neste trabalho. Este módulo trabalha efetivamente na busca pelas placas de trânsito nas imagens bidimensionais extraídas de ambientes reais. A arquitetura de TSD é baseada em RNSP com neurônios do tipo VG-RAM, composta por única matriz bidimensional de neurônios $m \times n$, denominada N . A figura 23 contempla toda a arquitetura utilizada no desenvolvimento de TSD.

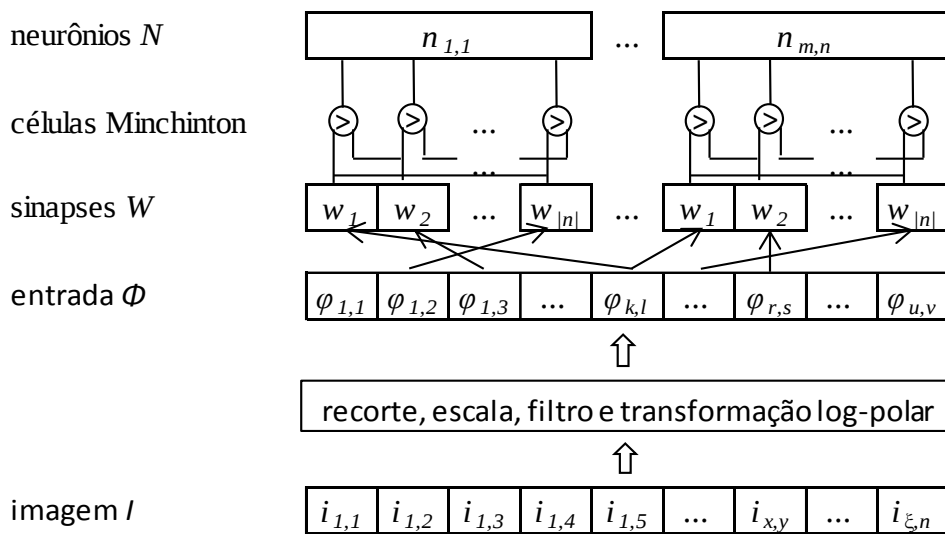


Figura 23 - Diagrama de arquitetura da RNSP VG-RAM para detecção de placas de trânsito

Cada neurônio n_{ij} apresentado no diagrama acima possui um conjunto de estruturas denominadas sinapses, representado por $W = (w_1, w_2, \dots, w_{|W|})$. Estas sinapses são responsáveis por obter informações do elemento de entrada desta arquitetura, denominado Φ . Φ é a representação bidimensional de pixels da imagem de entrada, I , no qual as sinapses de n_{ij} estão conectadas. Esta representação é modelada através do mapeamento dos elementos de Φ para o centro do campo de recepção de cada neurônio de N , seguindo a função log-polar, assim como é realizado no contexto biológico onde é feito o mapeamento das informações que chegam à retina e seguem para o SC (Seção 3).



Figura 24 - Padrão de interconexão sináptica do neurônio $N_{\frac{m}{2}, \frac{n}{2}}$ de TSD

A figura 24 apresenta o padrão de interconexão sináptica de cada neurônio n_{ij} que concretiza o seu campo receptivo, conotativamente representada por $\Omega_{i,j,\sigma}(W)$. Dada uma imagem arbitrária a ser explorada por TSD, a distribuição das sinapses de um neurônio que monitoram uma determinada região desta imagem (à esquerda) é representada pela Função Densidade de Probabilidade (FDP) da Distribuição Normal (ou Curva de Gauss) Bidimensional, com variância σ^2 , centrada no ponto φ_{μ_k, μ_l} . As coordenadas μ_k e μ_l (plano polar) de Φ são obtidas através do resultado da função log-polar inversa aplicada nas coordenadas i e j (plano cartesiano) de N . Assim, a distribuição de coordenadas k e l dos pixels de Φ , conectados a n_{ij} via W , dá-se através das seguintes FDP's:

$$\omega_{\mu_k, \sigma_2}(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(k-\mu_k)^2}{2\sigma^2}} \quad (3)$$

$$\omega_{\mu_l, \sigma_2}(l) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(l-\mu_l)^2}{2\sigma^2}} \quad (4)$$

Onde o desvio padrão, σ , é um parâmetro de entrada da arquitetura. Este parâmetro é responsável pela dispersão da FDP que utilizamos no modelo, atuando no espalhamento da distribuição sináptica dos neurônios. As coordenadas μ_k e μ_l do pixel de Φ , que representam o máximo global da FDP bidimensional, são obtidas pelos termos das equações (5) e (6).

$$\mu_k = \frac{u}{2} + d \cdot \cos\theta \quad (5)$$

$$\mu_l = \frac{v}{2} + d \cdot \sin\theta \quad (6)$$

Onde

$$d = \frac{u}{2} \cdot \left(\frac{\alpha^{\left| \frac{i-m/2}{m/2} \right|} - 1}{\alpha - 1} \right) \quad (7)$$

E

$$\theta = \begin{cases} \pi \cdot \left(\frac{3n}{2} - \frac{j}{n} \right) + \frac{\pi}{2n}; & \text{if } k < \frac{m}{2} \\ \pi \cdot \left(\frac{3n}{2} + \frac{j}{n} \right) + \frac{\pi}{2n}; & \text{if } k > \frac{m}{2} \end{cases} \quad (8)$$

E α é o fator de escala da função log-polar sendo também um parâmetro da arquitetura. O fator de escala da log-polar define a distribuição de neurônios na imagem de entrada, no plano polar, responsáveis pelo redimensionamento da região monitorada, ou seja, este parâmetro é incubido por definir a profundidade da acuidade visual da RNSP.

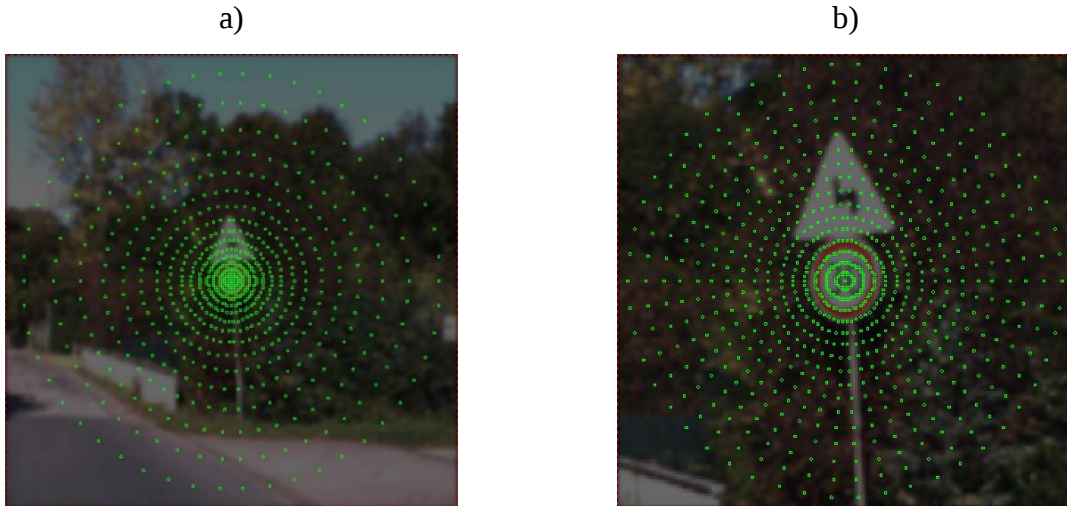


Figura 25 - Representação de profundidade (zoom) definida por α em uma imagem

A figura 25 representa o redimensionamento da imagem de entrada da rede, definida pelo resultado da função log-polar inversa em função do parâmetro α . Na imagem a) o valor do parâmetro α é maior que na imagem b). É possível perceber que quanto maior o zoom, maior a sensibilidade da rede em áreas mais distantes da coordenada de origem da log-polar, no plano polar. Em outras palavras, análogo à visão biológica, quanto mais distante é o ponto focal da retina, maior será a acuidade visual em áreas mais distantes do ponto central da fóvea.

O $\Omega_{i,j,\sigma}(W)$ simula um padrão de interconexões sinápticas, observado em muitas classes de neurônios biológicos [12]. Este padrão é criado aleatoriamente quando a rede é construída, permanecendo estático posteriormente. Embora criado de maneira aleatória, o padrão de interconexões sinápticas é o mesmo para todos os neurônios. Além disso, a memória de todos os neurônios é compartilhada (seção 3.1). Assim, quando a rede é treinada, cada neurônio aprende a associação entre as informações coletadas por suas sinapses e um valor de saída (metodologia padrão de redes baseadas em neurônios do tipo VG-RAM), mas todos os neurônios compartilham os demais conhecimentos aprendidos previamente [20].

Decodificador Lógico – Células *Minchinton Type 0 Cell*

Cada sinapse da RNSP VG-RAM só pode obter um único bit de entrada (função do decodificador lógico explanado na seção anterior). Assim, com o objetivo de que nosso modelo seja ideal para tratar imagens de ambientes reais, nas quais um pixel pode assumir uma variedade de diferentes valores, utilizamos células *Minchinton* [17].

Desenvolvida na Universidade de *Reading*, as células *Minchinton* representam uma das técnicas utilizadas para formar as tuplas (vetores de bits) que irão compor a memória (RAM) dos neurônios da RNSP [44]. Neste método cada sinapse de um neurônio, w_t , forma uma célula *Minchinton* com a próxima, w_{t+1} ($w_{|W|}$ forma uma célula *Minchinton* com w_1). O tipo de célula *Minchinton* que utilizamos é comumente chamado na literatura de *Type 0 Cell*. Estas células comparam os valores coletados entre duas sinapses, retornando 1 (verdadeiro) se a sinapse w_t , que está conectada a um elemento da entrada, $\phi_{k,l}$, possuir um valor maior que o valor do elemento $\phi_{r,s}$ para qual a sinapse w_{t+1} está conectada, ou seja, $\phi_{k,l} > \phi_{r,s}$. Caso contrário, retorna zero (ver as sinapses w_1 e w_2 do neurônio $n_{1,1}$ da Figura 30).

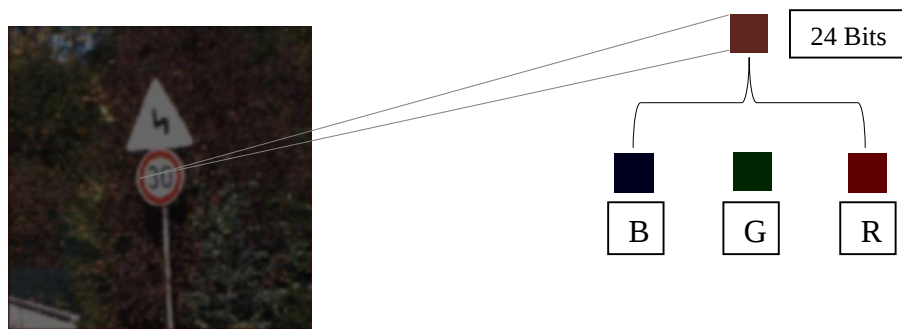


Figura 26 – Valor obtido por uma sinapse à partir de um elemento da entrada

A figura 26 mostra o processo de extração do valor de um elemento da entrada, $\phi_{k,l}$, produzido por uma sinapse w_t de um neurônio aleatório n_{ij} de TSD. Como já explicado, o sistema apresentado trabalha com imagens bidimensionais extraídas de ambientes reais. Estas imagens são formadas por uma matriz de pixels de 3 canais de cores, RGB, ordenados pelo sistema em BGR, onde cada elemento possui um valor de 24 bits. Estes valores são coletados pelas sinapses para que, através das células *Minchinton*, sejam somados e comparados entre as sinapses vizinhas retornando um valor booleano.

TSD em Atividade

Durante o treinamento, a imagem de treino é recortada por um quadrado centrado no centro da imagem de uma placa de trânsito. As dimensões deste quadrado são estimadas para conter a imagem da placa de trânsito e parte do fundo da imagem. Utilizamos diferentes fatores de escala, para que o tamanho do quadrado recortado seja correspondente entre os variados tamanhos de imagens de placas de trânsito.

Imagens reais possuem um nível bastante elevado de detalhamento, além de serem complexas e desordenadas. Para que as RNSPs tenham uma boa eficiência na generalização das suas entradas, a redução do excesso de detalhamento em imagens deste tipo torna-se factível. Desta forma, padrões de entrada mais próximos aos padrões previamente treinados tendem a serem categorizados em uma mesma classe pela RNSP. Para utilizarmos este conceito, aplicamos o Filtro de Gauss na imagem recortada, mencionada anteriormente.



Figura 27 - Comparação entre uma imagem original e sua cópia com aplicação do Filtro de Gauss

A figura 27 apresenta duas imagens: à esquerda está a imagem original; à direita mostra o resultado da aplicação do filtro Gaussiano na imagem original. É possível perceber que a aplicação do filtro suaviza os traços presentes na imagem, removendo os detalhes de alta resolução que poderiam impactar negativamente na generalização dos padrões das entradas da RNSP.

O centro da imagem da placa é utilizado como ponto de origem da função log-polar (centro de atenção, sacada, no campo visual) que mapeia Φ em N . Assim, cada neurônio é treinado para produzir um valor diferente de zero se o centro do campo receptivo estiver dentro de um círculo de raio r , referenciado no centro da imagem da placa de trânsito, ou zero caso contrário, onde r é um parâmetro da arquitetura. O valor treinado diminui à medida que o centro do campo receptivo dos neurônios se move para longe do centro do círculo, até atingir zero na sua borda. Este procedimento é repetido para todas as imagens do conjunto de imagens de treino. Ainda durante esta fase, o fator de escala mencionado acima está definido para um valor que coloca a imagem da placa de trânsito exatamente dentro do círculo de raio r .

A figura 28 mostra um exemplo de uma instância de treino da rede neural do módulo TSD para detectar uma placa de trânsito em uma imagem do conjunto de treino. A figura 28 (a) mostra a imagem de treinamento com a borda e o centro da placa de trânsito delimitados por um quadrado vermelho e uma cruz vermelha, respectivamente; a figura 28 (b) mostra a transformação (recorte, escala e filtro) da imagem da placa extraída, a figura 28 (c) mostra o mapeamento da função log-polar de Φ para N , utilizados para visualização do modelo e a Figura 28 (d) mostra a saída de N após do treino. Conforme mostrado na figura 28 (d), os neurônios com o centro dos seus campos receptivos dentro do círculo centrado no centro da imagem da placa de trânsito (comparar a figura 28 (c), com a Figura 28 (d)) são formados para produzir saídas com valores superiores a zero (branco ou cinzento), enquanto que aqueles que possuem o centro do seu campo receptivo longe do centro do círculo, são treinados para gerar um valor de saída igual a zero (preto) [20]. É possível perceber que os círculos tornam-se retângulos após a transformação log-polar, e as regiões em torno do centro da transformação ocupam uma área muito maior (comparar as representações da placa de trânsito circular com a representação da placa de trânsito triangular na figura 28 (b) e (c)).

Durante os testes, as imagens são sondadas pelo nosso sistema em vários pontos regularmente espaçados. São utilizados doze diferentes valores de escala para estas sondagens, resultando em vários tamanhos de imagens da placa de trânsito objetivando simular o processo de olhar para uma imagem de diferentes distâncias. A sondagem é feita através de deslocamentos horizontais e verticais do centro da função log-polar (centro de atenção), da imagem de teste proporcional ao tamanho do retângulo recortado, referente à placa de trânsito (janela de atenção), mencionado acima. O retângulo é então deslocado produzindo novas imagens sobrepostas, permitindo a detecção de imagens de placas de trânsito nos limites da janela de atenção.

a)



b)



c)



d)

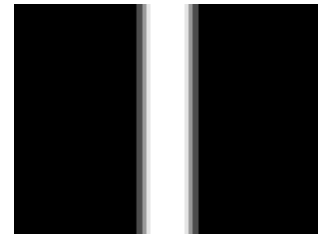


Figura 28 - Ilustração de uma instância de treino do módulo TSD para detecção de placas de trânsito

Cada janela de atenção extraída da imagem de teste é redimensionada (para caber em Φ), filtrada por um filtro gaussiano para suavizar os possíveis ruídos, e seus pixels são copiados para Φ . Os neurônios, em seguida, geram suas saídas de acordo com seus campos receptivos. Como já foi mencionado, neurônios com campo receptivo para as regiões do recorte da imagem, semelhante a regiões de imagens de placas de trânsito previamente treinados, geram saídas com valores elevados. Após os neurônios ativos serem submetidos a um procedimento equivalente ao *winner-takes-all* (vencedor-leva-tudo) – comportamento observado no SC [14] – o neurônio com a maior saída é selecionado e então é utilizada a função log-polar inversa, nas coordenadas deste neurônio, para calcular as coordenadas de um ponto (alvo sacádico) dentro do retângulo da janela de atenção, que possivelmente pertence a imagem de uma placa de trânsito. O centro da função log-polar (centro de atenção / sacada) é movido (sacado) para

este ponto na imagem de teste, e as saídas dos neurônios são recalculadas. Este procedimento é repetido uma ou mais vezes, ou até que o centro da função log-polar não mude. Em seguida, utilizamos um cálculo de correspondência entre valores, que quantifica a semelhança entre o centro da janela de atenção da imagem e uma placa de trânsito, comparando a imagem da saída de N depois de testar com o que foi previamente treinado (Figura 28 (d)). Se a saída de N depois do teste é semelhante à saída de N , após o treino, a imagem de uma placa de trânsito pode ter sido detectado na janela de atenção recortada da imagem.

A figura 29 mostra um exemplo de testes onde os neurônios da rede, treinados para detectar placas de trânsito, geram os seus resultados de acordo com a região da imagem monitorada por seus campos receptivos. A figura 29 (a) mostra a imagem de teste com o centro de atenção marcado com uma cruz vermelha; a figura 29 (b) mostra o recorte da janela de atenção (escalada e filtrada) extraída a partir da imagem de teste da figura 29 (a); a figura 29 (c) mostra o mapeamento log-polar de Φ para N (apenas para informação visual), e a figura 29 (d) mostra a saída de N antes da sacada. Ainda na figura 29 (d), é possível visualizar que os neurônios com o centro dos seus campos receptivos em uma imagem de uma placa de trânsito, geram saídas com altos valores (comparar a Figura 29 (c) com a figura 29 (d)). Entre a sequência de figuras que se inicia à partir da Figura 29 (e) até a figura 29 (h), há uma equivalência em relação a sequência que vai da Figura 29 (a) até a figura 29 (d). Entretanto, na primeira sequência citada, o ponto de atenção (centro da janela) não está focado em uma placa de trânsito. Na segunda sequência, após a execução do movimento sacádico, é possível perceber que o centro de atenção é deslocado para a região central da placa de trânsito. Conforme ilustrado na figura 29 (h), a saída de N após a sacada é muito semelhante à saída de N após o treino (compare a figura 28 (d) com a figura 29 (h)), o que indica que a imagem de uma placa de trânsito pode ter sido detectada no recorte do retângulo da janela de atenção da imagem de teste. Um vídeo de um único treinamento e várias sacadas subsequentes está disponível em http://youtu.be/H_LdE8fcbF4.

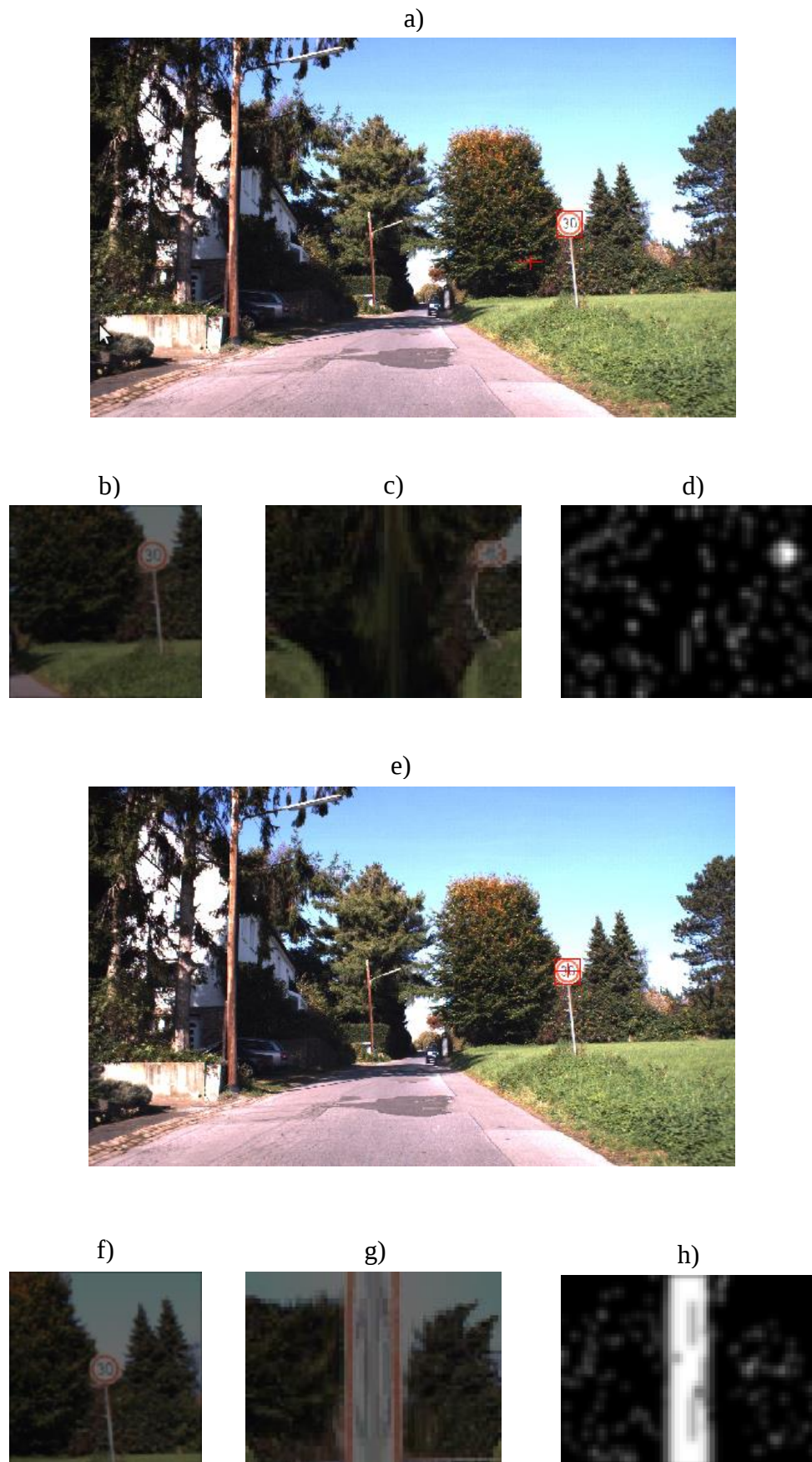


Figura 29 - Ilustração de uma instância de teste do módulo TSD para detecção de placa de trânsito

O valor probabilístico indicando que TSD realmente detectou a imagem de uma placa de trânsito é estimado utilizando inferência Bayesiana conforme descrito a seguir na seção 4.2. Toda a decisão final do sistema é regulada por um limiar: se o grau de similaridade é maior do que este limiar, então o modelo assume que uma placa de trânsito foi detectada. No entanto, a imagem de uma placa de trânsito pode ser detectada fora do centro de massa da imagem extraída a partir da imagem de teste (como mostra a Figura 29 (e)). Para encontrar exatamente o centro da imagem que representa uma placa de trânsito possivelmente detectada, empregamos uma segunda RNSP VG-RAM que explora a simetria presente na imagem de uma placa de trânsito, conforme descrito a seguir na seção 4.3.

4.2. Módulo de Inferência Bayesiana (BI)

Na solução que propomos neste trabalho, empregamos o módulo BI (*Bayesian Inference*) para estimarmos um valor probabilístico indicando que TSD realmente detectou a imagem de uma placa de trânsito. Esta probabilidade gerada é submetida a uma constante (limiar) que definirá um valor mínimo para determinar se uma placa de trânsito foi, ou não, detectada em uma sacada de TSD. Nesta seção apresentaremos os detalhes de BI e sua importância para o resultado final do sistema.

Nosso sistema faz o mapeamento da pontuação correspondente, que quantifica a similaridade entre o centro de atenção de uma imagem de uma placa de trânsito (seção 4.1), em uma medida de probabilidade. Esta medida é expressa com a probabilidade de $p(D \mid M, S, X \text{ e } Y)$, onde D é uma variável binária aleatória. $D = True$ (verdadeiro) se uma placa de trânsito foi detectada, e $D = False$ (falso) caso contrário. As variáveis aleatórias discretas $M = \{m_1, m_2, \dots, m_{|m|}\}$, $S = \{s_1, s_2, \dots, s_{|s|}\}$, $X = \{x_1, x_2, \dots, x_{|x|}\}$ e $Y = \{y_1, y_2, \dots, y_{|y|}\}$ representam a discretização dos possíveis valores de similaridade, fatores de escala e coordenadas x e y do alvo das sacadas na imagem de teste, respectivamente [20]. Utilizando o Teorema de Bayes, a probabilidade de que a imagem de uma placa de trânsito foi detectada no recorte da imagem (janela de atenção) ($p(D = TRUE)$), dado o valor retornado pela rede neural no intervalo m_i , com o recorte da janela de atenção extraída da imagem de teste de acordo com o fator de escala s_i , e o alvo sacádico dentro da faixa de coordenadas horizontais e verticais x_i, y_i da imagem de teste, pode ser formulada como:

$$p(D | M, S, X, Y) = \frac{p(D).p(M, S, X, Y | D)}{p(M, S, X, Y)} \quad (9)$$

Onde

$$p(M, S, X, Y | D) = P(M | D).P(S | D).P(X | D).P(Y | D) \quad (10)$$

E

$$p(M, S, X, Y) = p(D).P(M | D).P(S | D).P(X | D).P(Y | D) + P(\neg D).P(M | \neg D).P(S | \neg D).P(X | \neg D).P(Y | \neg D) \quad (11)$$

Para estimar os valores de $p(D)$, $p(\neg D)$, $p(M | D)$, $p(M | \neg D)$, $p(S | D)$, $p(S | \neg D)$, $p(X | D)$, $p(X | \neg D)$, $p(Y | D)$ e $p(Y | \neg D)$, foi utilizado um subconjunto de treino e um subconjunto de validação. A RNSP do módulo TSD foi treinada com imagens do conjunto de treino e seu resultado foi avaliado com a utilização do conjunto de validação, sendo possível obter a estimativa dos valores dos termos das equações (10) e (11) [20].

A cada sacada, TSD gera uma tupla com as informações de m_i , s_i , x_i , e y_i . A figura 30 descreve uma das saídas geradas por TSD, utilizada pelo módulo BI para estimar qual a probabilidade de que uma placa de trânsito seja detectada.

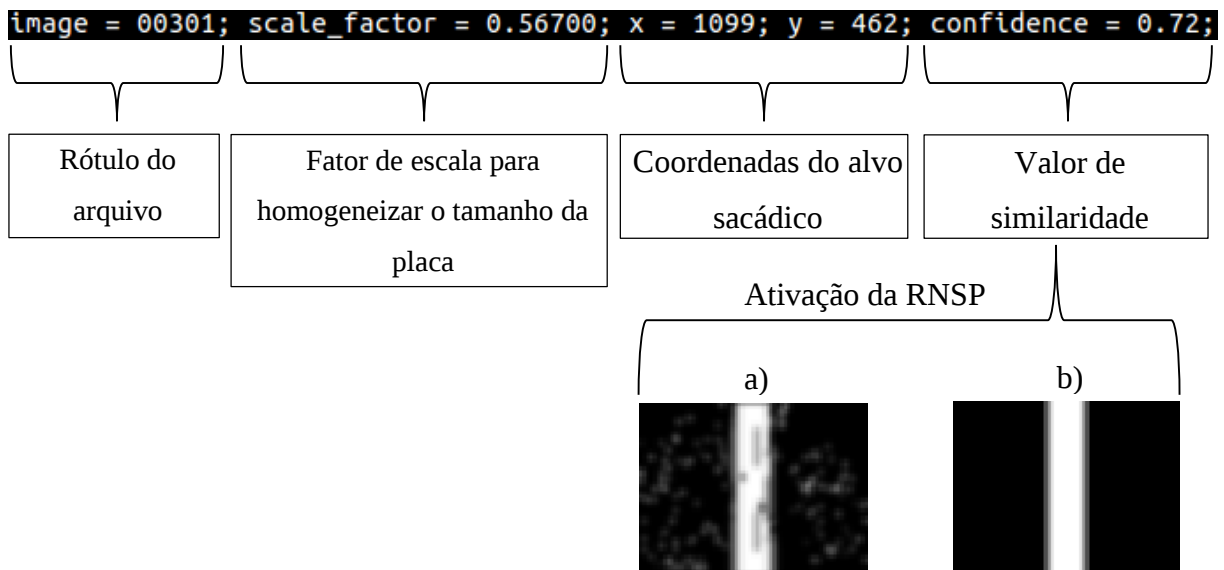


Figura 30 - Descrição das informações que compõem uma tupla gerada por TSD

O valor de similaridade, m_i , corresponde a comparação entre o mapa de ativação dos neurônios após uma sacada de TSD com o mapa de ativação dos neurônios no estágio de aprendizado para um determinado padrão, respectivamente representados pelas imagens a e b na figura 30.

A probabilidade de que uma placa de trânsito foi detectada, $p(D = True)$, ou não foi detectada, $p(D = False)$, pode ser estimada como o percentual de sacadas que atingiu, ou não atingiu, as coordenadas que formam o retângulo que delimita a área da placa de trânsito, presente em uma imagem, do conjunto de validação de acordo com as quatro limitações descritas a seguir nesta seção. As probabilidades $p(M = m_i | D = TRUE)$ (ou $p(M = m_i | D = FALSE)$) podem ser calculadas como o percentual de correspondência em cada valor de m_i , uma vez que uma placa de trânsito foi (ou não) detectada, ou seja, $D = True$ (ou $D = False$). As probabilidades $p(S = s_i | D = TRUE)$ (ou $p(S = s_i | D = FALSE)$) podem ser calculadas como o percentual de imagens extraídas de acordo com um fator de escala s_i , dado que $D = True$ (ou $D = False$). As probabilidades $p(X = x_i | D = TRUE)$ (ou $p(X = x_i | D = FALSE)$) podem ser calculadas como o percentual de alvos das sacadas dentro da faixa de coordenadas horizontais x_i , dado $D = True$ (ou $D = False$). E finalmente, as probabilidades $p(Y = y_i | D = TRUE)$ (ou $p(Y = y_i | D = FALSE)$) podem ser estimadas como o percentual de alvos sacádicos dentro da faixa de coordenadas verticais y_i , dado que a $D = True$ (ou $D = False$).

No módulo de Inferência Bayesiana utilizamos quatro restrições para determinar que uma placa de trânsito tenha sido realmente detectada, ou seja, a sacada atinge as delimitações de uma placa de trânsito no arquivo de *ground truth*, se todas as quatro restrições à seguir forem alcançadas.

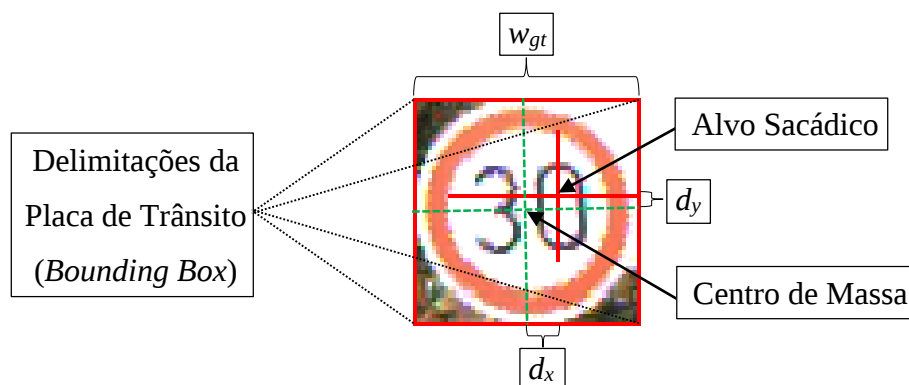


Figura 31 - Representação das delimitações de uma placa de trânsito e as distâncias entre o centro de massa e o centro sacádico

Utilizamos duas variáveis, d_x e d_y , representando as distâncias entre os eixos x e os eixos y, respectivamente, referentes às coordenadas do alvo sacádico e as coordenadas do centro de

massa do retângulo delimitador (*bounding box*) da placa de trânsito do arquivo de *ground truth* (ver figura 31). Outras duas variáveis também foram utilizadas, sendo w_{gt} a largura da placa de trânsito no arquivo de *ground truth* e w_{ei} a largura da placa de trânsito detectada, estimada em função de s_i ($w_{ei} = c / s_i$, onde s_i é o fator de escala utilizado na sacada e c é um parâmetro do nosso sistema) [20]. A primeira restrição especifica que o alvo sacádico deve estar próximo ao centro de massa do retângulo delimitador (*bounding box*) do arquivo de *ground truth*, ou seja:

$$d_x \leq 0.8 \times \frac{w_{gt}}{2} \quad (12)$$

E

$$d_y \leq 0.8 \times \frac{w_{gt}}{2} \quad (13)$$

A segunda restringe que o tamanho do retângulo que delimita a altura e a largura da imagem da placa de trânsito detectada, deve estar próximo ao tamanho do retângulo delimitador (*bounding box*) do arquivo de *ground truth*, ou seja:

$$w_{gt} \times 0.7 \leq w_{ei} \leq w_{gt} \times 1.2 \quad (14)$$

A terceira restringe que a probabilidade m_i , correspondente ao percentual de similaridade da ativação da rede na sacada, deve ser maior ou igual a 0.3. E finalmente, a quarta e última restrição implica que a imagem da placa de trânsito deve pertencer a uma categoria pré-definida de placas de trânsito.

A decisão final do sistema é definida por um limiar: Se $p(D = TRUE \mid M = m_i, S = s_i, X = x_i, Y = y_i)$ é maior do que o limiar, então uma placa de trânsito foi detectada. O valor deste limiar pode ser especificado pelo utilizador do sistema ou ajustado automaticamente utilizando um subconjunto de treino e um subconjunto de validação, através da variação do valor do limiar,

até que o valor ótimo do desempenho de interesse da precisão e *recall* seja alcançado (o subconjunto de validação é parte do conjunto de dados de treino, mas não do conjunto de dados de teste [18]).

4.3. Módulo de Centralização de Placas de Trânsito (TSC)

Mesmo que uma sacada gerada pela RNSP do módulo TSD reconheça uma placa de trânsito, alcançando as delimitações (em pixels) da mesma na imagem corrente, suas coordenadas podem estar afastadas do centro de massa desta placa. Este é um efeito indesejado uma vez que, para submetermos o resultado do nosso modelo para a Competição Alemã de Placas de Trânsito (GTSDb) – uma das motivações para o desenvolvimento do trabalho – deveríamos registrar as coordenadas dos centros das placas detectadas. Nesta seção apresentaremos o emprego do módulo TSC (*Traffic Sign Centralization*), essencial para eliminar estes problemas do nosso sistema.

O módulo TSC explora a simetria presente nas imagens das placas de trânsito detectadas, com o objetivo de encontrar seu centro. Sua arquitetura é exatamente a mesma apresentada no módulo TSD, apresentada previamente no capítulo 4.1, e opera em três etapas:

1. Aprendizagem da Placa de Trânsito Detectada;
2. Busca pela simetria horizontal;
3. Busca pela simetria vertical.

Aprendizagem da Placa de Trânsito Detectada

Nesta primeira etapa, o centro de atenção de TSC (o centro da log-polar, ou ainda, o alvo sacádico) é definido para as coordenadas cartesianas geradas pelo módulo TSD, descritas nas tuplas geradas pelas sacadas (ver figura 30). Para isso, uma janela de atenção da imagem, centrado no alvo sacádico de TSD, é extraída da imagem de teste por um recorte retangular com o mesmo fator de escala usado pelo TSD na detecção. Este recorte é redimensionado e filtrado, e seus pixels são copiados para as entradas de Φ do módulo TSC. Em seguida, a rede neural TSC é treinada para aprender a aparência da imagem em torno do alvo sacádico do TSD. O processo de treino é idêntico ao executado em TSD e previamente descrito no

capítulo 4.1, exceto pelo parâmetro r que no caso do TSC é definida de tal forma que os neurônios da rede aprendem valores de saída diferentes de zero em uma região menor da imagem da placa de trânsito, possivelmente detectada por TSD, invés de toda a imagem da placa [20]. A figura 32 mostra a diferença entre a região da image da placa de trânsito, utilizada por TSD, para treinar os neurônios da sua arquitetura e a região utilizada por TSC para o mesmo propósito. O parâmetro r mencionado acima indica o valor do raio da área circular que será enviada para treino em ambas as redes. O valor do raio da área enviada por TSC (envolvida por um círculo em vermelho) é aproximadamente à metade deste mesmo valor em TSD.



Figura 32 - Comparação entre a região da imagem utilizada por TSD e TSC para treinar a RNSP

Busca pela simetria horizontal

Neste estágio, a janela de atenção extraída da imagem de validação, que contém a placa de trânsito, é invertida horizontalmente e os neurônios de TSC geram uma saída de acordo com os seus campos receptivos (treinados na etapa anterior). Apenas a imagem é invertida, as coordenadas do alvo sacádicos permanecem estáticas. Se o alvo sacádicos de TSD atingir exatamente o centro de massa da imagem da placa de trânsito, a sua imagem invertida possuirá uma representação bastante semelhante devida a simetria geométrica da placa de trânsito. Portanto, neste caso, os neurônios da rede TSC mostrarão uma ativação muito semelhante ao de um aprendizado e a execução de uma sacada por TSC não moverá o centro de atenção. Por outro lado, se a sacada de TSD atingir um ponto diferente do centro da imagem da placa de trânsito, a região da imagem que se mostrará semelhante será uma região simétrica ao aprendido por TSC. Portanto, após uma sacada de TSC, os neurônios que mostrarão ativação serão aqueles que têm campo receptivo em regiões onde a imagem trocada é simétrica a imagem aprendida. Em seguida, a sacada de TSC moverá o seu centro de atenção para esta região. As distâncias do eixo x , d_x^h , e do eixo y , d_y^h , entre o centro de

atenção atual (alvo sacádico de TSC), (x_c, y_c) , e o alvo sacádico de TSD, (x_d, y_d) , são calculadas e armazenadas ($d_x^h = x_d - x_c$ e $d_y^h = y_d - y_c$). Considerando a aparência simétrica das imagens de placas de trânsito, é esperado que a maioria dos movimentos sacádicos de TSC, neste passo, sejam essencialmente horizontais (ou seja, $d_y^h \approx 0$).

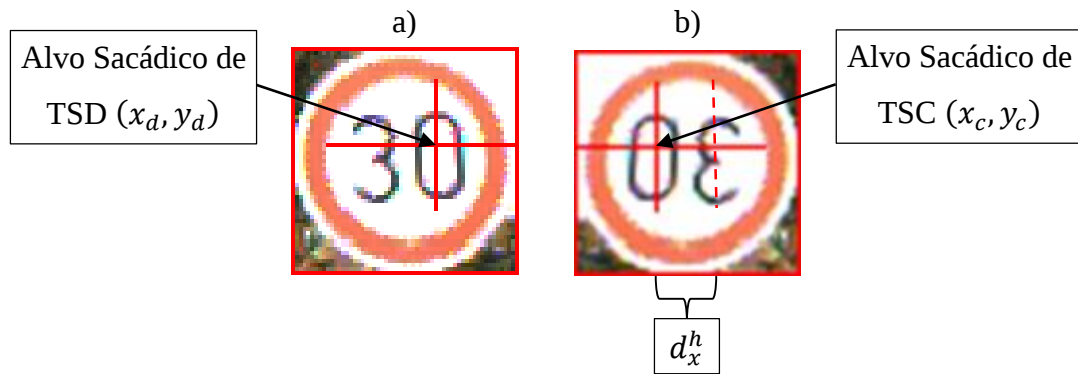


Figura 33 - Busca pela simetria presente na placa através do deslocamento horizontal

A figura 33 ilustra a pesquisa feita por TSC para encontrar a simetria da região treinada ao redor do alvo sacádico, na imagem (a), na imagem (b). É possível notar que houve um deslocamento entre os eixos x, ou seja, $d_x^h > 0$. Por outro lado, não ocorreu deslocamento entre os eixos y ($d_y^h = 0$). Estas duas análises são conclusivas na detecção de regiões simétricas.

Busca pela simetria vertical

A terceira e última etapa é inteiramente análoga a etapa anterior, a única diferença é que a imagem que contém a placa de trânsito é verticalmente invertida. Fora esta particularidade, os mesmos procedimentos do passo anterior são executados. As configurações iniciais são redefinidas, ou seja, a janela de atenção de TSC é novamente preenchida com a imagem da placa de trânsito e seu centro de atenção é posicionado para o alvo sacádico de TSD. Pelo fato da RNSP de TSC já ter sido treinada na etapa anterior, será necessário apenas inverter a imagem na vertical e então executar uma sacada para que as distâncias no eixo x d_x^v , e do eixo y d_y^v , sejam calculadas e armazenadas. Neste caso, espera-se que a maioria dos movimentos sacádicos seja vertical (ou seja, $d_x^v \approx 0$).

O módulo TSC considera que uma placa de trânsito foi detectada na janela de atenção, extraída da imagem original do conjunto de dados de validação, se $|d_y^h| \leq 0.2 \times |d_x^h|$ e $|d_x^v| \leq 0.2 \times |d_y^v|$ e que seu centro foi encontrado no ponto $(x_d - \frac{d_x^h}{2}, y_d - \frac{d_y^v}{2})$. Neste caso, o centro do quadrado da janela é movido para este ponto e o processo de detecção da placa de trânsito é considerado completo. Desta forma, a placa de trânsito delimitada pelo arquivo de *ground truth* pode ser facilmente calculada usando este ponto de coordenadas e o fator de escala atual [20].

```

TSC_ALGORITHM
BEGIN
  FOR I = 1 TO TOTAL_IMAGES
    READ_IMAGE(I)
    COMPUTE_SIMMETRY()
    T = 0
    WHILE COMPUTE_DELTA_XY() EQUALS 0 AND T < 6
      SET_NEW_RANDOM_TARGET()
      IF T < 4 THEN
        SET_NEW_SCALE()
      ENDIF
      COMPUTE_SIMMETRY()
      T = T + 1;
    ENDWHILE
    IF COMPUTE_DELTA_XY() NOT EQUALS 0 THEN
      RETURN TRUE
    ELSE
      RETURN FALSE
    ENDIF
  ENDFOR
END

COMPUTE_SIMMETRY
BEGIN
  TRAIN_RNSP()
  GET_SIMMETRY_HORIZONTAL()
  CLEAR_RNSP_MEMORY()
END

GET_SIMMETRY_HORIZONTAL
BEGIN
  FLIP_HORIZONTAL(I)
  COMPUTE_SACCADE()
  CALCULATE_DELTA(X)
END

GET_SIMMETRY_VERTICAL
BEGIN
  INITIALIZE_POSITION()
  FLIP_VERTICAL(I)
  COMPUTE_SACCADE()
  CALCULATE_DELTA(Y)
END

```

Figura 34 - Pseudocódigo representando o modelo do algoritmo de TSC

O algoritmo de pesquisa do centro de massa de uma placa de trânsito, implementado no módulo TSC, possui a estrutura apresentada no pseudocódigo da figura 34. Os métodos descritos pelos nomes de *GET_SIMMETRY_HORIZONTAL* e *GET_SIMMETRY_VERTICAL* representam, respectivamente, as etapas de busca pela simetria horizontal e busca pela simetria vertical. Durante a busca pelo centro de massa da placa de trânsito, executado pela

função *COMPUTE_SIMMETRY*, um limite de 4 diferentes escalas é definido para que o tamanho da possível placa de trânsito pesquisada não ultrapasse os limites de escala utilizados pelo módulo TSD, na fase de detecção.

Se a primeira sacada de TSC não é mais horizontal ou na segunda não é mais vertical, a memória dos neurônios de TSC é removida para que este procedimento seja repetido a partir de um ponto escolhido aleatoriamente, próximo das coordenadas do ponto atual (x_d, y_d) , utilizando um fator de escala maior. É importante salientar que este ponto escolhido deve ser calculado de forma que o novo alvo sacádicos permaneça dentro das limitações da figura geométrica que caracteriza uma placa de trânsito.

Assim, enquanto $|d_y^h| > 0.2 \times |d_x^h|$ e $|d_x^v| > 0.2 \times |d_y^v|$ e um número máximo de iterações não for alcançado, todo o processo é repetido. Se o número máximo de iterações é alcançado, o nosso sistema considera que a janela de atenção da imagem não contém uma imagem de uma placa de trânsito. O número de iterações pode ser definido pelo manuseador do sistema. Neste trabalho, utilizamos um limite de 6 iterações para que TSC detecte a placa de trânsito através da simetria da figura geométrica primitiva, presente na categoria das placas utilizadas no conjunto de treino e de validação. Uma animação da operação TSC está disponível em <http://youtu.be/SZ9w1XBWJqE>.

5. AVALIAÇÃO EXPERIMENTAL

5.1. Metodologia

Para avaliar o desempenho do sistema desenvolvido, utilizamos o conjunto de dados do modelo referência da Competição Alemã de Detecção de Placas de Trânsito (*The German Traffic Sign Detection Benchmark* - GTSDb) disponibilizado no portal eletrônico do Instituto de Neuroinformática da Universidade de Rurh (<http://benchmark.ini.rub.de>) [11]. A grande motivação e interesse neste conjunto de dados foram dados pela participação da nossa equipe na Competição Alemã de Detecção de Placas de Trânsito, uma das etapas da Conferência Internacional de Redes Neurais 2013 (*International Joint Conference on Neural Networks*

2013 – IJCNN 2013) realizada na cidade de Dallas (estado do Texas, EUA) entre os dias 04 – 09 de Agosto de 2013 [54].

Este conjunto de dados é segregado em dois subconjuntos denominados validação e avaliação (teste), nomenclatura dada pelos seus respectivos propósitos. O conjunto de dados disponibilizado pelo GTSDDB possui:

- 900 imagens das quais:
 - 600 pertencem ao conjunto de validação;
 - 300 pertencem ao conjunto de avaliação (teste);
- 1 arquivo de referência-base (*ground truth*) para o conjunto de validação;
- 1 arquivo contendo as seguintes categorias das placas de trânsito:
 - Proibitivo: 12 classes;
 - Perigo: 15 classes;
 - Obrigatório: 8 classes.

As imagens foram extraídas de uma gravação de vídeo nas proximidades da cidade de Bochum, na Alemanha, durante várias voltas com um veículo de passeio pela cidade. As gravações foram feitas em vias urbanas, rurais e rodovias, durante o dia até o anoitecer e em diferentes condições climáticas [52]. Estas imagens foram convertidas para o espaço de cores RGB, no formato PPM e possuem a dimensão de 1380×800 pixels.

Validação

Durante a competição, a organização do IJCNN 2013 disponibilizou, no mesmo portal mencionado com o conjunto de imagens, uma página onde as equipes competidoras poderiam submeter seus resultados. Mas antes que submetêssemos nossos resultados ao portal, utilizamos o conjunto de imagens de validação para avaliarmos o desempenho do sistema desenvolvido e gerar como saída principal, os resultados probabilísticos do módulo de Inferência Bayesiana (BI), apresentado na seção 4.2. O conjunto de dados de validação foi bastante útil por possuir o arquivo de referência-base (*ground truth*), norteando nossos resultados durante a validação.

Durante esta etapa, dividimos o conjunto de imagens de validação (com 600 imagens ao todo), em dois subconjuntos de 300 imagens cada, de modo que um conjunto foi utilizado para treinar a RNSP do módulo TSD, chamado Conjunto de Treino possuindo as 300

primeiras imagens (0 – 299), e outro para efetivamente validar o seu desempenho, denominado Conjunto de Validação possuindo as 300 últimas imagens (300 – 599).

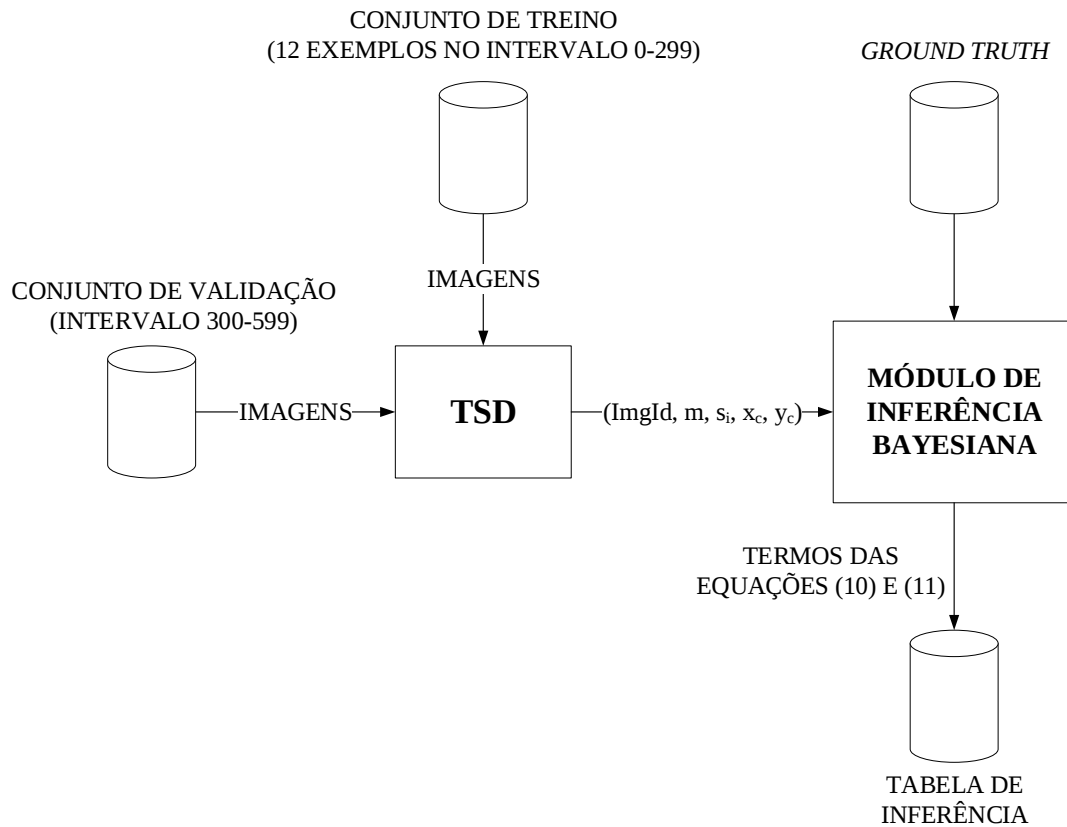


Figura 35 - Fluxograma da fase de validação do nosso sistema

A Figura 35 mostra o diagrama de fluxo da etapa de validação do nosso sistema, que é usado para estimar os valores dos termos das equações (10) e (11) do módulo BI. Inicialmente TSD é treinado com apenas 12 imagens, retiradas aleatoriamente, do total das 300 imagens do conjunto de treino. Estas 12 imagens contêm sinais de trânsito da categoria Proibitivo, sendo cada uma pertencente a uma classe da categoria, conforme mencionado anteriormente (a categoria Proibitivo possui 12 classes).

A rede neural TSD é treinada com as 12 imagens de placas de trânsito do conjunto de treino e testada com as imagens do conjunto de validação. TSD gera uma tupla no padrão $[ImgId, m, s, x_c, y_c]$ para cada execução de teste (sacada), onde $ImgId$ é o número de identificação de imagem, m é a pontuação correspondente, s é o fator de escala, e x_c e y_c são as coordenadas cartesianas do alvo da sacada (usamos uma única sacada para cada instância de teste). Estas

tuplas são enviadas para o módulo de Inferência Bayesiana onde são armazenadas e comparadas com as informações contidas no arquivo de *ground truth* referente ao conjunto de validação do GTSDb, para que o módulo BI processe as informações e defina a estimativa probabilística através dos termos das equações (10) e (11), persistindo o resultado em um arquivo que chamamos Tabela de Inferência (saída de BI).

Avaliação

Com a Tabela de Inferência gerada por BI na etapa de validação, o sistema possui as probabilidades de o módulo TSD detectar uma placa de trânsito a partir das suas variáveis. A Figura 36 ilustra o diagrama de fluxo da fase de avaliação do nosso sistema. Nesta fase, o módulo TSD utiliza como entrada o conjunto de treino (que é referente a primeira metade do conjunto de validação) e o subconjunto de avaliação. O conjunto de treino é o mesmo utilizado na fase de validação onde TSD também utiliza apenas 12 imagens do total de 300 imagens. O conjunto de dados de teste é composto por todas as 300 imagens (faixa de 0-299) do conjunto de dados de teste de GTSDb. O treino de TSD é da mesma forma que na fase de validação e testado com as imagens de todo o conjunto de dados de teste.

A saída de TSD alimenta o módulo de inferência Bayesiana que, nesta fase, através da Tabela de Inferência construída na fase de validação, é capaz de produzir uma tupla que inclui p , que é uma abreviatura de $p(D = TRUE \mid M = m_i, S = s_i, X = x_i, Y = y_i)$, e os mesmos membros da tupla recebida de TSD (Figura 36). A probabilidade gerada é comparada a um limiar definido pelo utilizador do sistema ou automaticamente através variação dos valores do limiar (seção 4.2). Se o valor de p desta tupla estiver acima de um limiar então a tupla é enviada a TSC, caso contrário é descartada.

TSC tenta encontrar uma placa de trânsito, buscando o centro da imagem candidata de uma placa de trânsito com as coordenadas informadas por TSD. Caso não encontre a placa, descarta a tupla. As placas de trânsito devidamente centradas por TSC são armazenadas em um arquivo de saída com os pontos delimitadores (*bounding boxes*), de geometrias retangulares, a partir da imagem de teste, que podem ser comparados com os dados do arquivo de *ground truth* do GTSDb.

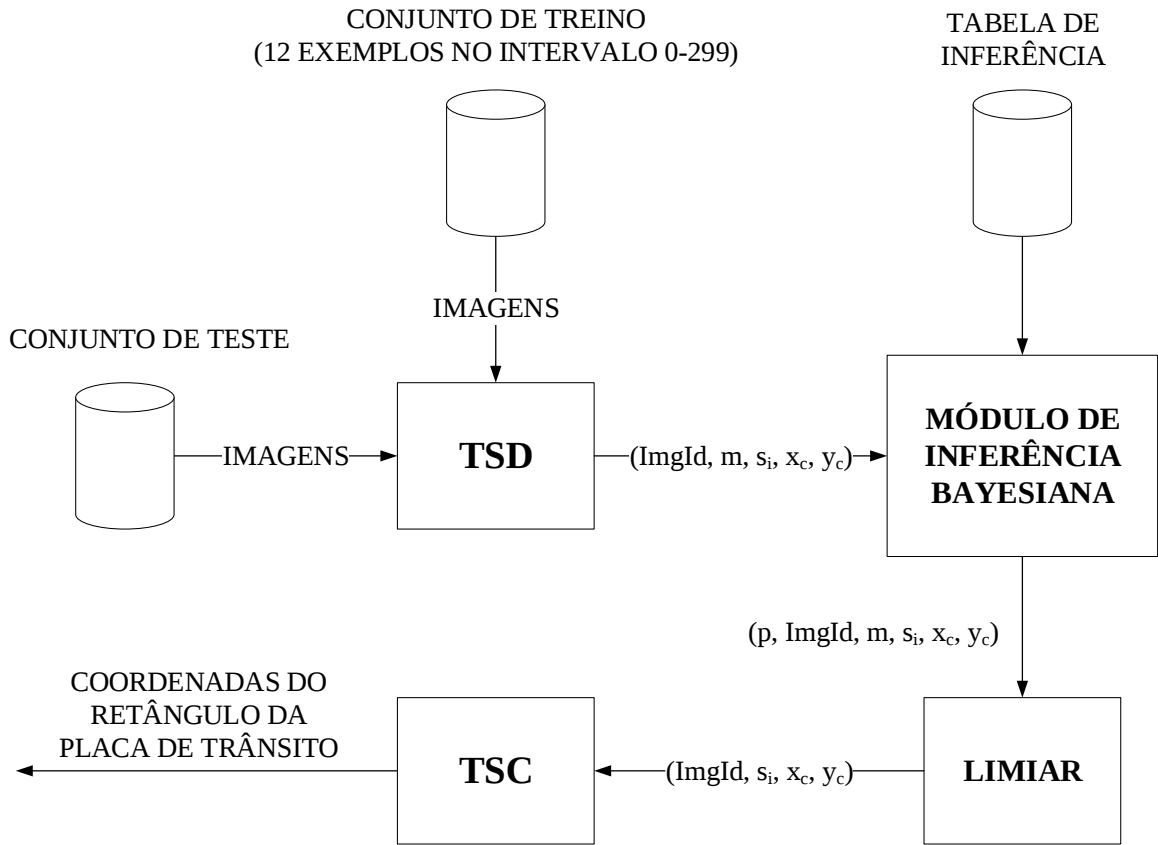


Figura 36 - Fluxograma da fase de teste do nosso sistema

5.2. Resultados

Nesta seção serão apresentados os resultados experimentais obtidos após a execução do modelo proposto. Estes resultados foram submetidos ao website da competição alemã de detecção de placas de trânsito, etapa da IJCNN 2013, no intuito de avaliar o desempenho do modelo na competição.

O ambiente utilizado foi uma máquina *Alienware Aurora R1 Dell*, com um processador *Intel Core i7-930 quad-core* (8M Cache, 2,8 GHz) e 12 GB de RAM. A arquitetura das RNSP de TSD e TSC possuem cinco parâmetros:

1. Número de neurônios, $m \times n$;
2. Número de sinapses por neurônio, $|W|$;
3. Tamanho da entrada da rede, $u \times v$;
4. Desvio padrão, σ , da distribuição normal bidimensional seguido pelo padrão de interligação sináptica dos neurônios, $\mathcal{Q}$;

5. Fator de escala, α , da função log-polar responsável pelo mapeamento de Φ em N .

Utilizamos os seguintes parâmetros em nossos experimentos, os quais foram escolhidos de forma empírica:

1. Número de neurônios da rede igual a 65×49 ;
2. Número de sinapses por neurônio igual a 256;
3. Tamanho da entrada de rede igual a 201×201 ;
4. Desvio padrão igual a 5;
5. Fator de escala igual a 2.

Usamos 12 diferentes fatores de escala (α) escolhidos de tal forma que a maior placa de trânsito contida nas imagens do conjunto de imagens de GTSDb (com o tamanho de 128×128 pixels), se encaixa nos campos receptivos dos neurônios de TSD, treinados com um valor diferente de zero, no menor fator de escala, assim como a menor imagem de uma placa de trânsito (com o tamanho de 16×16 pixels) preenche os campos receptivos dos neurônios de TSD com o maior fator de escala (ver figura 25).

Métrica

O objetivo principal do sistema proposto é resolver o problema de detecção de placas de trânsito em uma imagem bidimensional extraída de ambientes reais, onde a resposta deve ser binária: ou a placa foi encontrada, ou não foi encontrada. Para mensurarmos estatisticamente os resultados obtidos, utilizamos como métrica a Tabela de Contingência [53]. A Tabela de Contingência classifica entradas booleanas (verdadeiro ou falso – placa ou não placa) de duas classes: classe verdadeira (informações da base original – *ground truth*) e classe de hipóteses (saída gerada pelo sistema), gerando quatro possíveis saídas:

- Verdadeiro Positivo (VP): se existir uma placa na imagem e o sistema detecta esta placa;
- Falso Positivo (FP): se não existir uma placa na imagem e o sistema detecta uma placa;
- Falso Negativo (FN): se existir uma placa na imagem e o sistema não detecta esta placa;
- Verdadeiro Negativo (VN): se não existir uma placa na imagem e o sistema não detecta esta placa.

A tabela 3 representa graficamente as quatro possíveis saídas geradas pela Tabela de Contingência.

Tabela 3 - Tabela de Contingência com seus quatro possíveis estados

Resultados		Esperado	
		Placa	Não Placa
Realizado	Placa	VP	FP
	Não Placa	FN	VN

Como sugerido por Fawcett [53], temos dois conjuntos à serem classificados: o conjunto de dados Esperado – representado pelos dados do arquivo de referência base (*ground truth*) – e o conjunto Realizado – representado pelas saídas do sistema desenvolvido. As células coloridas na tabela 3 representam a Tabela de Contingência. A diagonal principal (destacada em verde) representam os valores corretos e a diagonal secundária (destacada em vermelho) representa os erros.

Após a classificação dos dados, através da Tabela de Contingência, utilizamos as seguintes equações para medir o desempenho do sistema:

$$Precision \text{ (Taxa de Positivos Previstos)} = \frac{VP}{VP + FP} \quad (16)$$

$$Recall \text{ (Taxa de Verdadeiros Positivos)} = \frac{VP}{VP + FN} \quad (17)$$

Onde *Precision* corresponde à taxa de valores verdadeiramente positivos previstos para classificação, ou seja, é a razão do total de placas de trânsito detectadas com sucesso por este mesmo total, somado ao total de placas de trânsito detectadas pelo sistema, mas que não existem na base referência (*ground truth*). *Recall* corresponde à taxa de valores legitimamente positivos, ou seja, é a razão do total de placas de trânsito detectadas com sucesso por este mesmo total, somado ao total de placas de trânsito não detectadas pelo sistema mas que existem na base referência (*ground truth*).

Para avaliar as contribuições de cada um dos módulos do nosso sistema de detecção de placas de trânsito, foram analisados os seguintes desempenhos:

1. Considerando apenas o módulo TSD;
2. Considerando os módulos TSD e BI;
3. Considerando os módulos TSD, BI e TSC.

A figura 37 mostra o desempenho do nosso sistema em uma curva de *precision x recall*, para o conjunto de dados de validação de GTSDb nos cenários 1, 2 e 3, representados pelas cores vermelha, verde e azul, respectivamente.

No gráfico da figura 37 podemos observar três curvas, uma para cada um dos cenários. Tal como o gráfico da figura 37 mostra, a curva gerada em 3, onde consideramos todo o sistema (TSD + BI + TSC), apresenta o melhor desempenho, ou seja, a área sob a curva de *precision × recall* é maior quando se utiliza o sistema inteiro. No entanto, o desempenho do sistema usando os cenários 1 e 2 são, à primeira vista, imprevisíveis.

O gráfico da figura 37 mostra que a área sob a curva de *precision × recall*, para o cenário 2 (TSD + BI), é menor do que a do cenário 1 (TSD), gerando um resultado oposto ao esperado. Isto acontece devido ao módulo BI descartar a maioria das tuplas que recebe, mesmo utilizando uma baixa probabilidade como limiar. Por outro lado, no cenário 1 (TSD) muitas tuplas são classificadas como placas de trânsito válidas e algumas delas contribuem para o *recall* em nosso sistema. Apesar desta avaliação, a versão final de todo o sistema não deveria conter o módulo de Inferência Bayesiana. No entanto, uma grande carga de dados seria enviada para o módulo TSC, uma vez que no cenário 1 (TSD) uma grande quantidade de tuplas são submetidas.

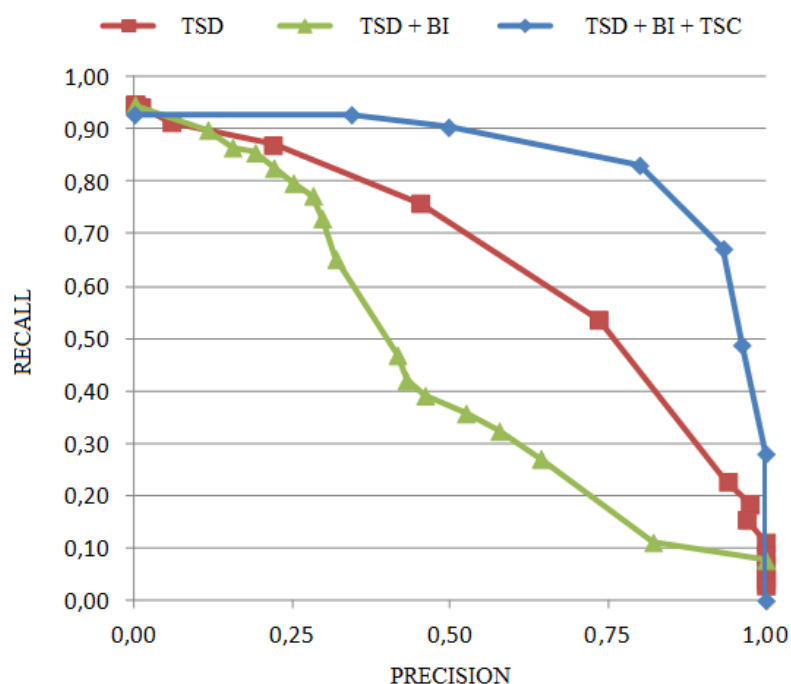


Figura 37 - Desempenho do sistema para os cenários 1, 2 e 3

A diferença entre as áreas dos cenários 1 (TSD) e 3 (TSD + BI + TSC) demonstra os benefícios da utilização do módulo para a realização de simetria (TSC) que corrige a maioria das tuplas que seriam consideradas como falsos positivos, contribuindo para um bom *recall*, mostrado na curva 3 (TSD + BI + TSC) representado pela tendência de cor azul.

Os resultados do nosso sistema de detecção placas de trânsito foram submetidos ao *website* de avaliação do GTSDb em 28 de fevereiro de 2013, em meio à conferência internacional de detecção de placas de trânsito (IJCNN 2013). Nosso sistema foi classificado entre os 16 primeiros melhores métodos – em um total de 53 métodos submetidos entre as 18 equipes participantes – para a categoria Proibitivo, mostrando uma área de 85,12% abaixo da curva de *precision × recall*. O gráfico da Figura 38 mostra os dados coletados a partir da página, na *web*, do GTSDb. A tendência de cor vermelha, com símbolos quadrados nos limiares, mostra a curva de desempenho do sistema para o conjunto de dados de avaliação (teste). Já a tendência em azul, com pontos em losango, mostra a curva de desempenho para o conjunto de dados de validação (observe que esta tendência é a mesma apresentada na figura 37, também em azul). O gráfico da Figura 38 também mostra que o desempenho do sistema para o conjunto de validação é bem próximo ao conjunto de avaliação (teste).

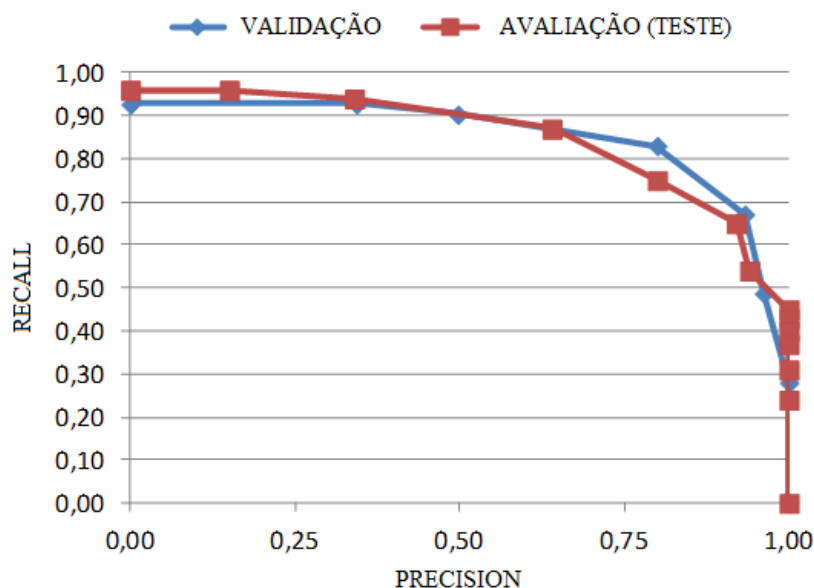


Figura 38 - Desempenho do sistema considerando TSD + BI + TSC (cenário 3) nas bases de dados de validação (tendência de cor azul) e de avaliação - teste (tendência de cor vermelha)

Também foi avaliado o desempenho dos módulos do sistema em função do tempo. O módulo de detecção das placas de trânsito, TSD, executa o treino das 12 placas de trânsito em 3,85 segundos, e executa uma sacada em 0,42 segundo, em média, enquanto o módulo TSC realiza todo o processo necessário para a detecção e correção de simetria em 0,18 segundo, também em média. Portanto, se todo o sistema detecta uma placa de trânsito em uma única sacada em TSD, o tempo de detecção total de uma placa de trânsito é de $0,60\text{ s} = 0,42\text{ s} + 0,18\text{ s}$ (o tempo consumido pelo módulo de inferência bayesiana, BI, é insignificante). No entanto, o sistema pode não detectar uma placa de trânsito com uma única sacada de TSD. No pior caso de uma detecção de placa de trânsito durante a sondagem de uma única imagem são realizadas 780 sacadas, pois são utilizados vários fatores de escala – para o maior fator de escala utilizado por TSD, 162 sacadas são necessárias. Entretanto o conjunto de fatores de escala pode ser reduzido, por exemplo, na execução de TSD com o menor fator de escala, apenas 6 sacadas são necessárias. Neste caso, utilizando computadores atuais, todo o sistema pode funcionar a uma taxa de cerca de 3,6 s para uma imagem.

5.3. Calibração

Com o objetivo de obter um melhor resultado nos experimentos executados neste capítulo, realizamos a calibração dos parâmetros da RNSP VG-RAM utilizada nos módulos TSD e TSC. Inicialmente, o principal interesse foi alcançar a configuração paramétrica que obtivesse a melhor precisão na tarefa de detecção de placas de trânsito. Para isso, utilizamos o mesmo conjunto de imagens de validação do experimento realizado no capítulo 5.2, durante a Competição Alemã de Detecção de Placas de Trânsito, etapa da Conferência Internacional em Redes Neurais (*International Joint Conference on Neural Networks – IJCNN 2013*). Um segundo passo seria obter esta configuração ótima que também apresentasse um excelente resultado relativo à latência do sistema, minimizando custos computacionais durante a execução da atividade em questão.

A calibração do nosso sistema foi executada utilizando todo o conjunto das 600 imagens de validação do conjunto de imagens de GTSDb – considerando apenas as placas de trânsito da categoria Proibitiva – da seguinte maneira:

1. O sistema obtém uma imagem do conjunto de imagens de validação do GTSDb;
2. Um movimento sacádico é executado para esta imagem, direcionando o centro de atenção da sacada para o centro da placa de trânsito, calculado de acordo com as informações contidas no arquivo de *ground truth* (fornecido pelo GTSDb juntamente com o conjunto de imagens) relativas a imagem e a placa de trânsito em questão;
 - a. Neste momento, a imagem contida na janela de atenção da sacada é recortada, filtrada e escalada conforme explicado nas seções anteriores.
3. O sistema inicia o treinamento da RNSP e, após o mapeamento log-polar de N em ϕ , a rede neural é treinada com o centro de atenção da sacada para o centro da função log-polar (seção 4.2);
4. Após o treinamento da placa, o sistema executa um movimento sacádico direcionando o centro de atenção para uma região fora da placa de trânsito, mas dentro da janela de atenção de forma que seja possível a visualização do que foi treinado na janela de atenção atual. A sacada gera uma tupla, para um arquivo, contendo informações em um padrão explicado a seguir;

5. Uma nova sacada é executada. Nesta situação, o seu centro de atenção está bastante afastado do centro de atenção da placa de trânsito, de modo que seja impossível a visualização da região que está sendo monitorada pelos neurônios da rede. Assim como no passo anterior, a sacada também gera uma tupla, para um arquivo, contendo informações em um padrão explicado a seguir;
6. Toda a memória dos neurônios da rede é removida.

Estes 6 passos são executados para todas as imagens, que contém imagens de placas de trânsito do tipo Proibitivo do conjunto de imagens de validação de GTSDb, para cada possível combinação entre os parâmetros da rede. Abaixo está descrito o padrão da tupla retornado em cada sacada:

$$(ImgId, m, s, x_c, y_c, m \times n, w, \sigma, \alpha, ij)$$

Tabela 4 - Descrição dos valores retornados pelas tuplas durante as sacadas na etapa de calibração

Código	Descrição
<i>ImgId</i>	número da imagem
<i>m</i>	probabilidade da sacada encontrar uma placa de trânsito
<i>s</i>	fator de scala
x_c, y_c	coordenadas do alvo sacádico
$m \times n$	dimensão da entrada da camada neural
<i>w</i>	número de sinapses por neurônio
σ	desvio padrão da distribuição normal
α	fator de log utilizado na função log-polar
<i>ij</i>	cálculo do índice de similaridade de Jaccard

Métrica

Nós empregamos duas medidas de precisão para avaliar os resultados gerados durante a calibração do sistema:

- Índice de similaridade de Jaccard (IJ):

$$IJ = \frac{a}{a + b + c} \quad (18)$$

Onde:

a = Área onde há interseção entre o retângulo delimitador (*bounding box*) da possível imagem de placa de trânsito obtida pela sacada e a delimitação real da placa no arquivo de *ground truth*.

b = Área total onde o movimento sacádico possivelmente encontrou uma imagem de placa de trânsito.

c = Área total da delimitação retangular da placa de trânsito, informada através do arquivo de *ground truth*.

- Razão entre as probabilidades (R):

$$R = \frac{Sp - Sl}{Sp} \quad (19)$$

Onde:

Sp = Probabilidade de a rede ter encontrado uma placa de trânsito na sacada próxima a placa.

Sl = Probabilidade de a rede ter encontrado uma placa de trânsito na sacada distante da placa.

Avaliamos também o desempenho da etapa de calibração em relação ao tempo de processamento desta atividade. Para tal, contabilizamos o número total de movimentos sacádicos realizados e sua latência. O total de execuções foi calculado em relação a quantidade de imagens que contém placas de trânsito do tipo proibitivas, contidas no conjunto de imagens de validação de GTSDb. Neste conjunto, a quantidade de imagens contendo placas de interesse é 131. Com estas informações foi possível calcular o tempo aproximado durante todo o processo de calibração, detalhado a seguir:

- Parâmetros x Total de Imagens x (1Sp + 1Sl): $875 \times 131 \times 2 = 229250$
- Tempo médio de cada execução (passo 1 ao 6): $\cong 0.5 s^*$
- Tempo total: $229250 \times 0.5 \cong 14625 s \cong 1910.42 \text{ min} \cong 79.6 \text{ hs}$

* Este valor possivelmente sofrerá alterações dependendo das configurações de hardware da máquina utilizada no procedimento, assim como o tamanho da camada de entrada da rede neural.

Conforme mencionado na seção anterior, nosso sistema possui 5 parâmetros de entrada para a RNSP VG-RAM sendo que deste grupo, 4 variáveis foram utilizadas no processo de calibração resultando em um total de 875 configurações distintas (ver tabela 5). Para gerar os

gráficos estatísticos, utilizamos a dimensão da rede como referencial no comparativo entre as demais variáveis do conjunto de parâmetros. A seguir segue a descrição do conjunto de valores escolhidos para cada parâmetro, na realização da calibração da rede.

Tabela 5 - Quantidade de execuções do programa para cada configuração

Parâmetro	Variações
$m \times n$	7
w	5
σ	5
α	5
Total de configurações	875

Para executarmos os testes nesta etapa de calibração do sistema, implementamos scripts *bash* – interpretador de comandos amplamente utilizado em diversas plataformas – visando automatizar os passos descritos acima.

O gráfico da figura 39 apresenta as mudanças ocorridas no desempenho da RNSP ao longo do crescimento da dimensão da rede relativo às configurações do conjunto de parâmetros mencionados anteriormente. É possível perceber que há um ganho significativo no desempenho da rede quando a dimensão da rede ($m \times n$) alcança 33×25 neurônios, independente da configuração dos demais parâmetros. Após este ápice qualitativo e quantitativo, à medida que a dimensão da rede continua a crescer algumas configurações apresentam resultados imprevisíveis e inesperados. Entretanto, a grande maioria das combinações paramétricas continua a melhorar seu desempenho.

A grande área, em azul claro, representa o comportamento geral das tendências geradas a partir dos resultados desta etapa de calibração. É possível perceber uma área de cor violeta representando o melhor desempenho da RNSP obtido pela calibração, a partir da dimensão 65×49 . Este resultado ótimo foi alcançado com a seguinte configuração paramétrica:

- 256 sinapses por neurônio;
- 2,5 de desvio padrão da distribuição normal (distribuição das sinapses);
- 2,0 do fator de escala da função log-polar (fator de ampliação [41]).

Logo abaixo desta área violeta, há uma área em azul bem claro que acompanha o melhor resultado descrito acima, sendo o segundo melhor resultado alcançado. Esta configuração é a mesma utilizada nos resultados obtidos na seção 5.2, no qual os parâmetros da RNSP foram definidos de forma empírica.

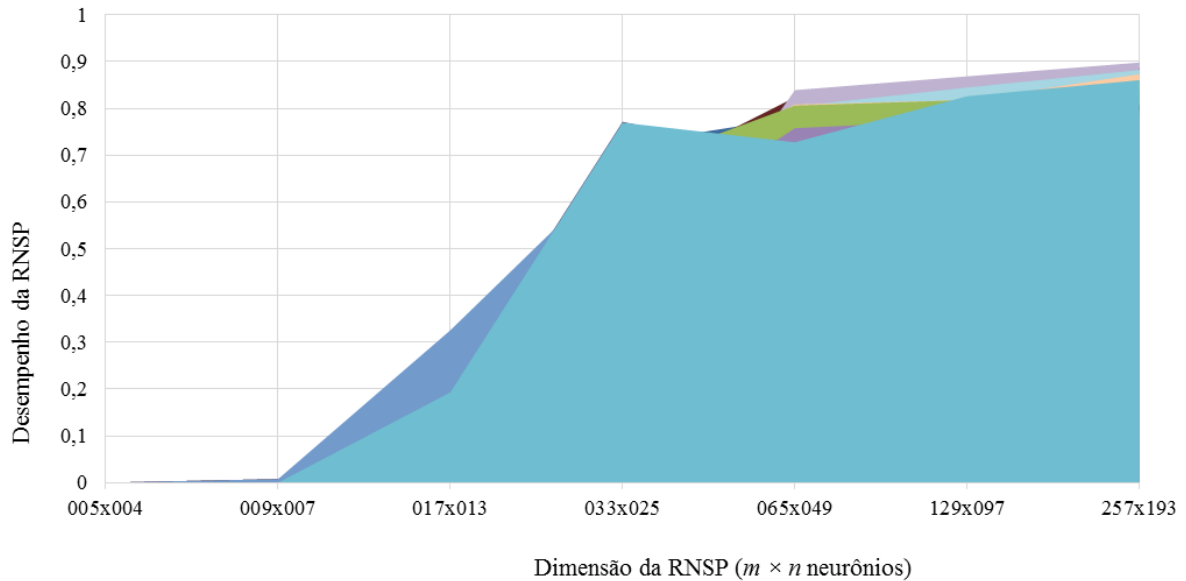


Figura 39 - Gráfico gerado após a calibração utilizando todo o conjunto de parâmetros

5.3.1. Variação do Número de Sinapses

Nesta seção apresentamos os resultados da variação dos possíveis valores do número de sinapses ($|W|$) por neurônio em função da quantidade de neurônios na rede.

O gráfico da figura 40 ilustra as tendências da RNSP VG-RAM para o conjunto de variações das sinapses dos neurônios durante a calibração. Cada item da legenda, à direita do gráfico, representa a cor da curva no gráfico, a variação das sinapses (W), o desvio padrão da distribuição normal (σ) e o fator de escala da função log-polar (α), respectivamente. Os vértices presentes em cada curva representam o desempenho alcançado (eixo Y) para cada dimensão da camada neural (eixo X).

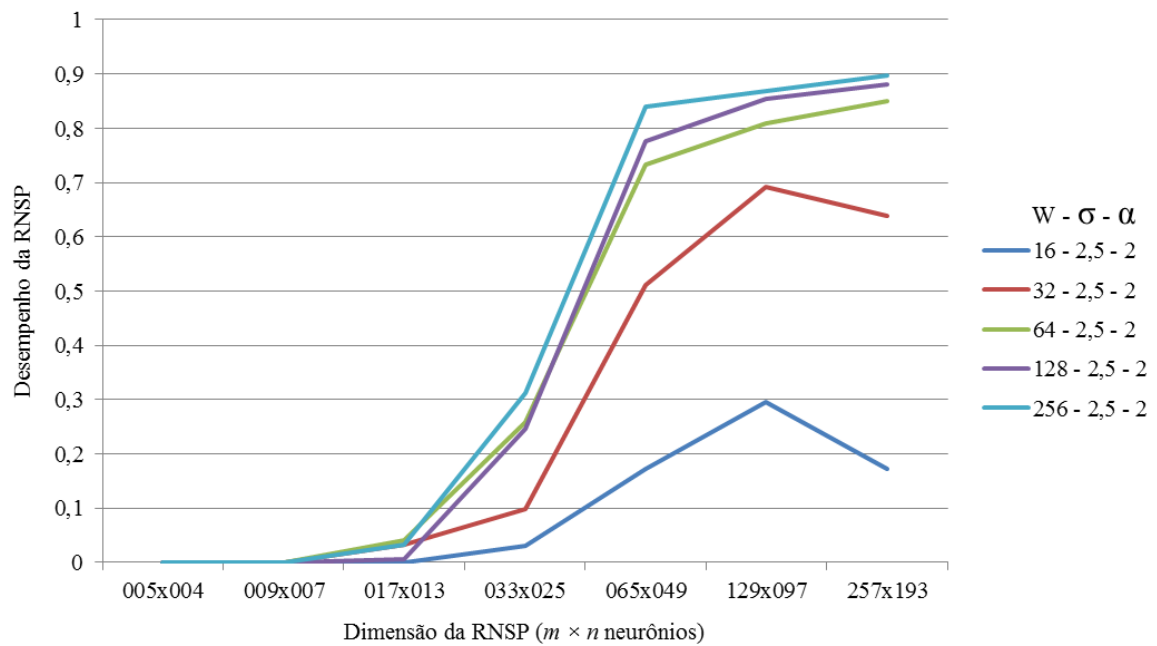


Figura 40 - Gráfico de tendências de desempenho da variação das sinapses em função da dimensão da RNSP

É possível observar que o desempenho aumenta conforme a dimensão da rede. Este ganho é ainda maior nas configurações onde há um maior número de sinapses (W) por neurônio. Nota-se ainda um comportamento imprevisível em duas configurações com menor número de sinapses e com a maior dimensão utilizada na calibração: 257×193 neurônios com 16 sinapses e 257×193 neurônios com 32 sinapses.

O gráfico da figura 40 mostra também que a melhor configuração obtida pelo sistema alcançou uma taxa de desempenho de 90% para a configuração de 257×193 neurônios com 256 sinapses por neurônio.

A tabela 6 representa numericamente os valores utilizados para representar o gráfico da figura 40. Conforme mencionado, cada curva no gráfico representa a tendência de uma configuração onde a variação é a quantidade de sinapses por neurônio. O valor de cada posição da tabela representa o vértice da sua respectiva configuração do tamanho da rede pelo número de sinapses.

Para gerar este gráfico comparativo utilizamos, estaticamente, os melhores resultados observados para os demais parâmetros de entrada da rede. Neste caso, os melhores desempenhos dos demais parâmetros foram 2,5 (σ) para o desvio padrão da distribuição normal e 2,0 para o fator de escala da função log-polar (α).

Tabela 6 – Matriz de desempenho do número de sinapses utilizado na calibração em função da dimensão da RNSP

		Número de Sinapses				
		16	32	64	128	256
Dimensão da Camada Neural	005x004	0,00	0,00	0,00	0,00	0,00
	009x007	0,00	0,00	0,00	0,00	0,00
	017x013	0,00	0,03	0,04	0,01	0,03
	033x025	0,03	0,10	0,26	0,25	0,31
	065x049	0,17	0,51	0,73	0,78	0,84
	129x097	0,30	0,69	0,81	0,85	0,87
	257x193	0,17	0,64	0,85	0,88	0,90

5.3.2. Variação do Desvio Padrão da Distribuição Normal

Nesta seção apresentamos os resultados da calibração do sistema variando o parâmetro σ , referente ao desvio padrão da distribuição normal (gaussiana) aplicada para o padrão de interconexão das sinapses dos neurônios, em função da dimensão da rede.

O gráfico da figura 41 mostra as tendências da rede para o conjunto de variações do desvio padrão (σ) da distribuição normal das sinapses. Cada item da legenda, à direita do gráfico, representa a cor da curva no gráfico, seguido do ótimo valor obtido para as sinapses (W), seguido pela variação dos valores do desvio padrão da distribuição normal (σ) e então seguido pelo fator de escala da função log-polar (α). É possível perceber que a RNSP com poucos neurônios, mas com elevados valores do desvio padrão, atinge maior desempenho em relação a valores mais baixos. Entretanto, este comportamento torna-se inversamente proporcional à medida que a dimensão da rede aumenta, exceto quando o valor do desvio padrão da distribuição normal é o 1,25 (o menor utilizado na calibração).

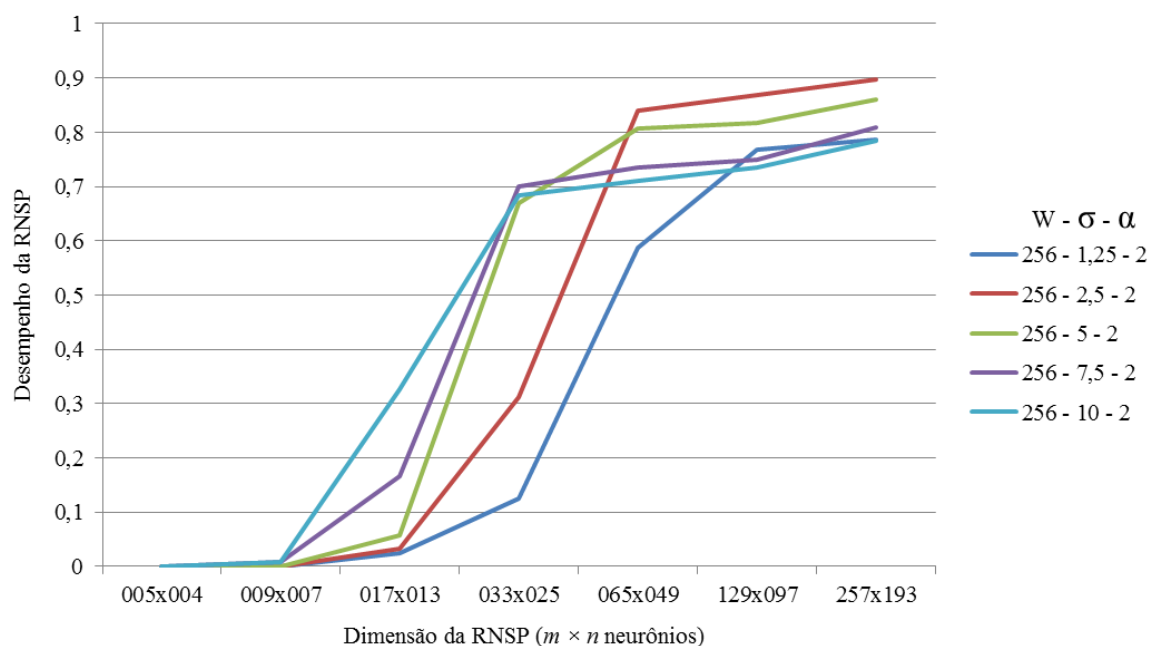


Figura 41 - Gráfico de tendências de desempenho da variação do desvio padrão da distribuição normal das sinapses em função da dimensão da RNSP

Com o aumento do desvio padrão da distribuição normal ocorre uma diminuição da taxa de acertos em consequência direta do espalhamento excessivo das sinapses em torno da posição central dos neurônios no plano da camada de entrada $m \times n$. Este comportamento ocorre dado que uma grande parte das sinapses dos neurônios que monitoram o objeto de interesse recairá sobre regiões fora deste objeto. Consequentemente as sinapses produzirão informações ambíguas a respeito do objeto, o que levará a rede neural a possíveis perdas.

A tabela 7 representa os valores gerados para plotar o gráfico da figura 41. Cada posição da tabela representa o vértice da sua respectiva configuração do tamanho da rede em função do valor do desvio padrão da gaussiana.

Para gerar este gráfico comparativo, utilizamos os melhores resultados observados para os demais parâmetros de entrada da rede. Conforme vimos na seção 5.3.1, temos a configuração com melhor desempenho quando os neurônios da rede possuem 256 sinapses. O outro parâmetro com melhores resultados é o fator de escala da função log-polar com seu valor igual a 2,0.

Tabela 7 - Matriz de desempenho dos valores utilizados para o desvio padrão da distribuição normal na calibração em função da dimensão da RNSP

		Desvio Padrão da Gaussiana				
		1,25	2,5	5	7,5	10
Dimensão da Camada Neural	005x004	0,00	0,00	0,00	0,00	0,00
	009x007	0,00	0,00	0,00	0,01	0,01
	017x013	0,03	0,03	0,06	0,17	0,33
	033x025	0,13	0,31	0,67	0,70	0,68
	065x049	0,59	0,84	0,81	0,74	0,71
	129x097	0,77	0,87	0,82	0,75	0,74
	257x193	0,79	0,90	0,86	0,81	0,78

5.3.3. Variação do Fator de Escala

Nesta seção apresentamos os resultados da calibração do sistema variando o parâmetro α , que é o fator de escala (fator de ampliação ou *log-factor*) da transformação de redimensionamento sofrida pela imagem na RNSP.

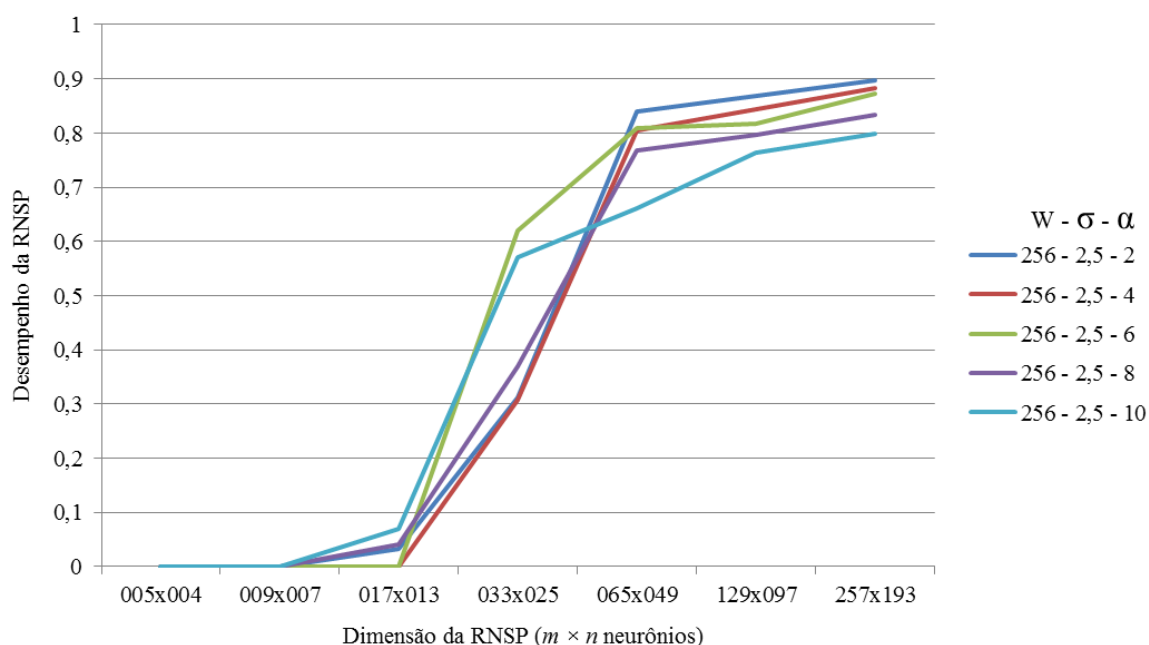


Figura 42 - Gráfico de tendências de desempenho da variação do fator de escala (fator de ampliação) em função da dimensão da RNSP

A figura 42 ilustra as tendências de desempenho para o conjunto de variações do fator de escala (α) relativas ao fator de escala, utilizadas na função log-polar, responsável pelo mapeamento retinotópico da imagem de entrada para a rede.

Cada item da legenda, à direita do gráfico, representa a cor da curva no gráfico, a variação das sinapses (W), o desvio padrão da distribuição normal do padrão de interconexão sináptica (σ) e o fator de escala da função log-polar (α). É possível observar que em redes menores dimensões e elevados valores para o fator de escala, contribuem para um resultado melhor. Este comportamento torna-se inversamente proporcional em redes com dimensões maiores. Isto é justificado pela ampliação de escala da imagem (zoom), fazendo com que mais neurônios monitorem regiões da imagem das quais fazem parte o objeto de interesse (ver figura 25). Conforme apresentado neste estudo, quanto maior o fator de escala da função log-polar maior é o afastamento do centro de atenção da sacada. Este afastamento é ruim quando há um número alto de neurônios pelo fato de que uma grande quantidade de neurônios irá monitorar uma região fora do objeto de interesse.

A tabela 8 representa os valores gerados para plotar o gráfico da figura 42. Cada posição da tabela representa o vértice da sua respectiva configuração do tamanho da rede em função do fator de escala da função log-polar.

Para gerar este gráfico comparativo, utilizamos os melhores resultados observados para os demais parâmetros de entrada da rede. Conforme vimos na seção 5.3.1, temos a configuração com melhor desempenho quando os neurônios da rede possuem 256 sinapses. O outro parâmetro com melhores resultados é o desvio padrão da distribuição normal de valor igual a 2,5.

Tabela 8 - Matriz de desempenho dos valores utilizados para o fator de escala (fator de ampliação) na calibração em função da dimensão da RNSP

		Fator de Escala da Função Log-polar				
		2	4	6	8	10
Dimensão da Camada Neural	005x004	0,00	0,00	0,00	0,00	0,00
	009x007	0,00	0,00	0,00	0,00	0,00
	017x013	0,03	0,00	0,00	0,04	0,07
	033x025	0,31	0,31	0,62	0,37	0,57
	065x049	0,84	0,80	0,81	0,77	0,66
	129x097	0,87	0,84	0,82	0,80	0,76
	257x193	0,90	0,88	0,87	0,83	0,80

5.4. Resultados Após Calibração

Nesta seção apresentaremos o resultado obtido na tarefa de detecção de placas de trânsito, utilizando as imagens do conjunto de dados do modelo referência da Competição Alemã de Detecção de Placas de Trânsito (*The German Traffic Sign Detection Benchmark - GTSDDB*) disponibilizado no portal eletrônico do Instituto de Neuroinformática da Universidade de Ruhr (<http://benchmark.ini.rub.de>) [11], com a configuração de melhor desempenho da RNSP obtida na etapa de calibração (seção anterior). Abaixo estão enumerados os parâmetros relativos a esta configuração de melhor desempenho:

1. Número de neurônios da rede igual a 65×49 ;
2. Número de sinapses por neurônio igual a 256;
3. Tamanho da entrada de rede igual a 201×201 ;
4. Desvio padrão igual a 2,5;
5. Fator de escala igual a 2.

Esta configuração é muito semelhante à configuração utilizada na Competição Alemã de Detecção de Placas de Trânsito, GTSDDB, (seção 5.2) diferenciando-se apenas no valor do desvio padrão da distribuição normal bidimensional, que na ocasião o valor utilizado para este parâmetro foi 5.0. A metodologia utilizada para avaliar e gerar os resultados para esta configuração, selecionada pela etapa de calibração com o melhor desempenho alcançado, foi a mesma utilizada para a Competição Alemã de Detecção de Placas de Trânsito, GTSDDB, explanada na seção 5.2.

O gráfico da figura 43 mostra o resultado obtido com a melhor configuração após a etapa de calibração e o resultado obtido durante a competição GTSDDB (seção 5.2), a partir da página do GTSDDB na *web*. A tendência que representa esta melhor configuração obtida na etapa de calibração é representada pela cor vermelha. Em segundo plano colocamos a tendência de cor azul, representando o resultado obtido durante a competição GTSDDB. Para a categoria Proibitivo, a melhor configuração da etapa de calibração alcançou uma área de 84,72% abaixo da curva de *precision* $\times$ *recall*, apresentando um desempenho 0,40% menor do que a configuração utilizada na competição. É possível notar que as tendências se comportam de forma bastante semelhantes, apresentando algumas poucas variações. Assim como nos gráficos gerados pelos resultados da competição GTSDDB (seção 5.2), os símbolos ao longo

das tendências representam os limiares utilizados para seleccionar as tuplas (seção 4) que possivelmente indicariam uma placa de trânsito detectada.

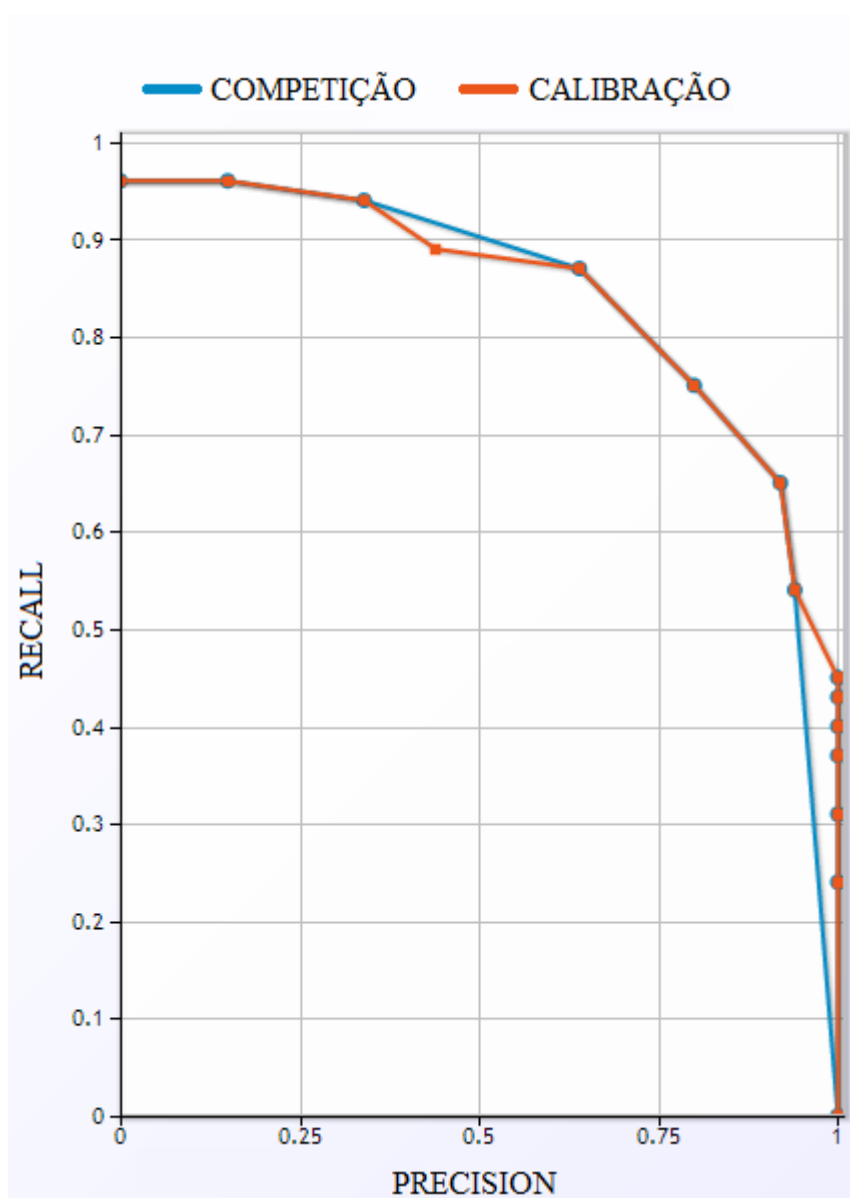


Figura 43 – Comparativo entre as tendências de desempenho com a melhor configuração paramétrica (em vermelho) e a configuração utilizada na competição (em azul)

Fonte: adaptado de: <http://benchmark.ini.rub.de/index.php?section=gtsdb&subsection=submissions>

Quanto menor for o limiar maior é o valor do *recall*, dado que este indicador avalia a quantidade máxima de placas de trânsito detectadas independente da quantidade de tuplas que o sistema detectou como placa, mas que de fato, não era – falsos positivos (ver equação (17)). Apesar do bom desempenho em ambos os resultados, não foi possível detectar todas as placas

de trânsito contidas nas 300 imagens do conjunto de avaliação da Competição Alemã de Detecção de Placas de trânsito. É possível perceber também que, para limiares mais altos, a tendência relativa à configuração de melhor desempenho obtida na etapa de calibração (ver seção 5.3) apresentou melhores resultados em comparação com a tendência relativa a configuração utilizada na competição.

6. CONCLUSÃO

6.1. Sumário

O desafio da detecção de placas de trânsito em imagens bidimensionais vem atraindo a atenção de muitos pesquisadores nos últimos anos. Recentemente a IJCNN anunciou seu desafio e lançou uma competição para a atividade em questão. Pesquisadores de diversas regiões do mundo participaram da competição proposta e então foi constatada uma variedade de técnicas, fundamentadas em metodologias já apresentadas na literatura para detecção de objetos de interesse, utilizadas pelos participantes durante o desafio para resolver o problema proposto.

Neste trabalho, foi estudada e desenvolvida uma solução computacional para realizar a tarefa de detecção de placas de trânsito baseada na habilidade do ser humano em localizar objetos de interesse em uma cena, utilizando seu sistema visual. Para isso, empregamos a arquitetura de uma Rede Neural Sem Peso (RNSP) baseada em neurônios do tipo VG-RAM (acesso compartilhado à memória) em conjunto com um modelo biologicamente inspirado nos movimentos oculares sacádicos, explorando cada imagem utilizada nesta pesquisa em busca do objeto de interesse (placas de trânsito). A avaliação do sistema proposto evidencia que a abordagem é apropriada para a tarefa em questão, permitindo detectar precisamente placas de trânsito em imagens reais.

6.2. Análise Crítica

O propósito principal deste trabalho foi desenvolver uma arquitetura computacional baseada no modelo fisiológico da visão dos seres humanos, com o objetivo de detectar objetos de interesse (placas de trânsito) em imagens de ambientes reais. Assim como no sistema visual biológico, onde as informações visuais projetadas na retina são enviadas ao Córtex Visual Primário para mapear e processar as informações obtidas e posteriormente enviadas ao Colículo Superior onde são executados os movimentos oculares sacádicos, esta arquitetura foi capaz de identificar um ponto em um objeto de interesse (alvo sacádico) previamente conhecido para então realizar movimentos sacádicos em direção a este ponto, a partir de uma sequência de imagens bidimensionais (a exemplo de imagens monoculares projetadas em uma das retinas do olho humano).

O desempenho do nosso sistema para resolver o problema da detecção de placas de trânsito é dado por dois fatores. Primeiro cada sinapse dos módulos TSD e TSC recolhe o resultado de uma comparação entre dois pixels, monitorados por duas sinapses de seu respectivo neurônio, que é executada pela célula *Minchinton* deste par de sinapses. Nossas configurações para ambos os sistemas têm centenas de sinapses por neurônio e milhares de neurônios. Portanto, durante os testes, centenas de milhares destas comparações são realizadas para cada imagem de entrada e os resultados são comparados com resultados equivalentes, previamente aprendidos, a partir de imagens do conjunto de validação utilizadas para treinar a rede. Esta quantidade de comparações de pixels permite não só uma alta capacidade de discriminação, mas também de generalização. Outro fator é que, através do padrão de interconexão sináptica e da arquitetura log-polar, cada neurônio de TSD ou TSC monitora uma região específica da placa de trânsito, o que reduz o impacto global da oclusão, variação de cor e condições de iluminação no desempenho destes sistemas [20].

6.3. Conclusões

Nesta dissertação, apresentamos uma nova abordagem para a detecção de placas de trânsito com base em Redes Neurais Sem Peso com neurônios do tipo VG-RAM (VG-RAM WNN). As nossas experiências mostraram que RNSP VG-RAM podem ser empregadas para a

detecção de placas de trânsito com boa precisão, utilizando um número bastante pequeno de amostras de treino (neste trabalho foi utilizada apenas uma amostra de cada placa de interesse para o treino). A principal vantagem da utilização da arquitetura das RNSP VG-RAM, comparada com outras abordagens de redes neurais utilizadas para detecção de placas de trânsito, é sua simples implementação e rápido treino, sem necessidade de um grande conjunto de amostras para treinar a rede.

Foi realizada uma série de experimentos de calibração dos parâmetros do sistema de detecção de placas de trânsito, visando obter a configuração com melhor desempenho. Apesar de o melhor resultado obtido na etapa de calibração não ter sido o mesmo utilizado na competição, quando adotamos uma configuração empiricamente, as duas configurações são bastante semelhantes distinguindo-se apenas pelo valor do desvio padrão da distribuição normal (função densidade probabilidade), σ . A configuração adotada na competição, escolhida empiricamente, teve o segundo melhor desempenho nos resultados obtidos na etapa de calibração.

Embora os resultados obtidos sejam iniciais, sendo necessário ainda bastante estudo nesta área, tanto no contexto biológico quanto nas abordagens computacionais, concluímos que foram bastante satisfatórios. Aplicando os conceitos biológicos aprendidos durante o estudo feito no desenvolvimento do trabalho apresentado, o sistema desenvolvido realizou a detecção de placas de trânsito de um conjunto de 300 imagens utilizando apenas 12 imagens para treinar a rede neural.

6.4. Trabalhos Futuros

Dentre os trabalhos futuros que pretendemos realizar está o desenvolvimento de outro módulo, integrado ao módulo TSD, capaz de realizar movimentos microssacádicos em uma região fora da região da fóvea, conhecida como região periférica da retina [51]. Acreditamos que com a implementação deste trabalho, a performance da rede será ainda melhor e seu custo computacional cairá bastante, pelo fato de que não será necessário sondar toda a imagem para alcançar os objetos de interesse. Movimentos microssacádicos atuarão no redirecionamento do alvo sacádico em uma imagem, diminuindo o tempo computacional para detectar placas de trânsito em imagens bidimensionais.

Outra abordagem futura seria a implementação destes módulos utilizando uma arquitetura paralela através de GPUs – *Graphics Processing Units* – em linguagem CUDA – *Compute Unified Device Architecture* – visando diminuir o tempo de processamento das sacadas. Uma limitação identificada no modelo seria a necessidade de executar sacadas em toda a imagem em diferentes níveis de zoom, de forma sequencial. Com a proposta das microssacadas, direcionando o centro de atenção para locais onde possivelmente existem placas de trânsito e paralelizando o processamento sacádico em nível de hardware, acreditamos que o desempenho computacional do nosso sistema aumentará bastante.

Conforme citado, a rede é treinada com apenas 12 imagens. Aumentar o número de exemplos de treinamento poderia melhorar significativamente o seu desempenho do sistema, principalmente para alcançar uma precisão de 100% na detecção das placas. Nós também gostaríamos de avaliar o sistema proposto em placas de trânsito do sistema rodoviário brasileiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] S. Escalera, X. Baró, O. Pujol, J. Vitrià, P. Radeva, *Traffic Sign Recognition Systems*, Springer Series: SpringerBriefs in Computer Science, 2011.
- [2] D. Gavrilă, *Traffic sign recognition revisited*, Proceedings of the Mustererkennung, DAGM-Symposium, p. 86-93, 1999.
- [3] Y. Li, *Real-time traffic sign detection: an evaluation study*, Proceedings of the 20th International Conference on Pattern Recognition (ICPR), p. 3033-3036, 2010.
- [4] Y.-Y. Nguwî, A. Z. Kouzani, *Detection and classification of road signs in natural environments*, Neural Computing and Applications, vol. 17, no. 3, p. 265-289, 2008.
- [5] T. M. Nguyen, S. S. Ahuja, Q. M. J. Wu, *A real-time ellipse detection based on edge grouping*, IEEE International Conference on Systems, Man and Cybernetics, p.3280-3286, 2009.
- [6] N. Barnes, A. Zelinsky, L. S. Fletcher, *Real-time speed sign detection using the radial symmetry detector*, IEEE Transactions on Intelligent Transportation Systems, vol.9, no.2, p.322-332, 2008.
- [7] Y. Aoyagi, T. Asakura, *A study on traffic sign recognition in scene image using genetic algorithms and neural networks*, Proceedings of the 22nd International Conference on Industrial Electronics, Control, and Instrumentation, p. 1838-1843, 1996.
- [8] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, F. Lopez-Ferreras, *Road-sign detection and recognition based on support vector machines*, IEEE Transactions on Intelligent Transportation Systems, vol.8, no.2, p.264-278, 2007.
- [9] X. Baro, S. Escalera, J. Vitria, O. Pujol, P. Radeva, *Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification*, IEEE Transactions on Intelligent Transportation Systems, vol.10, no.1, p.113-126, 2009.
- [10] I. Aleksander, *WISARD to MAGNUS: A family of weightless virtual neural machines*, RAM-Based Neural Networks, p. 18-30, 1998.

- [11] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, C. Igel, *Detection of traffic signs in real-world images: the German Traffic Sign Detection Benchmark*, Proceedings of the International Joint Conference on Neural Networks, 2013.
- [12] E. R. Kandel, J. H. Schwartz, T. M. Jessell, *Principles of Neural Science*, 4th ed. McGraw-Hill, 2000.
- [13] N. Tabareau, D. Bennequin, A. Berthoz, J. Slotine, B. Girard, *Geometry of the superior colliculus mapping and efficient oculomotor computation*, Biological Cybernetics, vol. 97, p. 279-292, 2007.
- [14] R. A. Marino, T. P. Trappenberg, M. Dorris, D. P. Munoz, *Spatial interactions in the superior colliculus predict saccade behavior in a neural field model*, Journal of Cognitive Neuroscience, vol. 24, no. 2, p. 315-336, 2012.
- [15] I. Aleksander, *Self-adaptive universal logic circuits*, IEEE Electronic Letters, vol. 2, no. 8, p. 231-232, 1966.
- [16] T. B. Ludermir, A. C. P. L. F. Carvalho, A. P. Braga, M. D. Souto, *Weightless neural models: a review of current and past works*, Neural Computing Surveys, vol. 2, p. 41-61, 1999.
- [17] R. J. Mitchell, J. M. Bishop, S. K. Box, J. F. Hawker, *Comparison of some methods for processing grey level data in weightless networks*, RAM-based Neural Networks, p. 61-70, 1998.
- [18] F. Sebastiani, *Machine learning in automated text categorization*, ACM Computing Surveys, vol. 34, no. 1, p. 1-47, 2002.
- [19] World Health Organization, *Global status report on road safety 2013: supporting a decade of action*, p 1-9, 2013.
- [20] A. Souza, C. Fontana, F. Mutz, T. Oliveira, M. Berger, A. Forechi, J. Neto, E. Aguiar, C. Badue, *Traffic Sign Detection with VG-RAM Weightless Neural Networks*, IJCNN, 2013.
- [21] D. H. Hubel, *Eye, Brain, and Vision*, W. H. Freeman, 1995.

- [22] E.R. Kandel, J.H. Schwartz, T.M. Jessel, *Principles of Neural Science*, 4th Ed. Prentice-Hall International Inc., 2000.
- [23] Retina, Scholarpedia, Disponível em <<http://www.scholarpedia.org/article/Retina>>, acesso em 11/2013.
- [24] Fovea Centralis, Wikipedia, Disponível em <http://en.wikipedia.org/wiki/Fovea_centralis>, acesso em 11/2013.
- [25] J. Otero-Millan, S. L. Macknik, S. Martinex-Conde, *The Journal of Neuroscience*, April 25, 2012.
- [26] W. C. Hall, A. MKoschovakis, *The Superior Colliculus, New Approaches for studying sensorimotor integration*, 2004.
- [27] J. Morén, T. Shibata, K. Doya, *The Mechanism of Saccade Motor Pattern Generation Investigated by a Large-Scale Spiking Neuron Model of the Superior Colliculus*, V. 08, I 02, 2013.
- [28] Z. M. Hafed, L. Goffart, R. J. Krauzlis, *A neural mechanism for microsaccade generation in the primate superior colliculus*, Science 323: 940–943, 2009.
- [29] K. L. Gerhard, *Ophthalmology: A Pocket Textbook Atlas*, 2000.
- [31] N. J. Ghandi, E. L. Keller, *Comparison of Saccades Perturbed by Stimulation of the Rostral Superior Colliculus, the Caudal Superior Colliculus, and the Omnipause Neuron Region*, Journal of Neurophysiology, 82:3236-3253, 1999.
- [32] K. Rayner, *Eye movements in reading and information processing*, Psychological Bulletin, 85: 618-660, 1978.
- [33] K. Rayner, *Eye movements in reading and information processing: 20 years of research*, Psychological Bulletin, 85, 618-660 1998.
- [34] M. I. Posner, *Orienting of attention*, Quarterly Journal of Experimental Psychology, 32, 3-25, 1980.
- [35] S. P. Liversedge, K. Rayner, S. J. White, J. M. Findlay, E. McSorley, *Binocular coordination of the eyes during reading*, Current Biology, 16, 1726-1729, 2006.

- [36] S. P. Liversedge, S. J. White, J. M. Findlay, K. Rayner, *Binocular coordination of eye movements during reading*, Vision Research, 46, 2363-2374, 2006.
- [37] C. W. Oyster, *The Human Eye: Structure and Function*, Sinauer Associates, 01/01/1999.
- [38] Parvocellular cell, Wikipedia, Disponível em <http://en.wikipedia.org/wiki/Parvocellular_cell>, acesso em 11/2013.
- [39] Visual Cortex, Wikipedia, Disponível em <http://en.wikipedia.org/wiki/Visual_cortex>, acesso em 11/2013.
- [40] Area V1, Scholarpedia, Disponível em <http://www.scholarpedia.org/article/Area_V1>, acesso em 11/2013.
- [41] Visual Map, Scholarpedia, Disponível em <http://www.scholarpedia.org/article/Visual_map>, acesso em 11/2013.
- [42] Log-Polar Mapping, Instituto Superior Técnico Lisboa, Disponível em <<http://users.isr.ist.utl.pt/~alex/Projects/TemplateTracking/logpolar.htm>>, acesso em 10/2013.
- [43] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2 ed. 1999.
- [44] J. Austin, *RAM-Based Neural Network, A Short Hystory*, 1998.
- [45] J. M. Wolfe, *Guided Search 4.0: A guided search model that does not require memory for rejected distractors*, Journal of Vision, 1 (3), 349, 2001.
- [46] C. Quaia, P. Lefèvre, L. M. Optican, *Model of the control of saccades by superior colliculus and cerebellum*, Journal of Neurophysiology, 82, 999-1018, 1999.
- [47] R. Walker, E. McSorley, *The parallel programming of voluntary and reflexive saccades*. Vision Research, 46, 2082-2093, 2006.
- [48] R. Walker, D. G. Walker, M. Husain, C. Kennard, *Control of voluntary and reflexive saccades*, Experimental Brain Research, 130 (4), 540-544, 2000.
- [49] S. B. Hutton, E. M. Joyce, T. R. E. Barnes, C. Kennard, *Saccadic distractibility in first-episode schizophrenia*. Neurpsychologia, 40(10), 1729-1736, 2002.

- [50] D. P. Munoz, S. Everling, *Look away: the anti-saccade task and the voluntary control of movement*, Nature Neuroscience, 5, 217-228, 2004.
- [51] A. Kingstone, R. M. Klein, *What are human express saccades?*, Perception and Psychophysics, 54 (2), p. 260-273, 1993.
- [52] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, “*Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark*,” in International Joint Conference on Neural Networks, 2013.
- [53] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters, v. 27, n. 8, p. 861–874, 2006.
- [54] International Joint Conference on Neural Networks *official website*, Disponível em <<http://www.ijcnn2013.org/>>, acesso em 11/2013.