UNIVERSIDADE FEDERAL DE SERGIPE
CAMPUS PROF. ALBERTO CARVALHO
DEPARTAMENTO DE SISTEMAS DE INFORMAÇÃO

THIAGO DA SILVA ALMEIDA

Modelo de Detecção e Reconhecimento de Semáforos Baseado em Atenção Visual e Inteligência Artificial

ITABAIANA
2015

UNIVERSIDADE FEDERAL DE SERGIPE

CAMPUS PROF. ALBERTO CARVALHO

DEPARTAMENTO DE SISTEMAS DE INFORMAÇÃO

THIAGO DA SILVA ALMEIDA

Modelo de Detecção e Reconhecimento de Semáforos Baseado em Atenção Visual e Inteligência Artificial

Trabalho de Conclusão de Curso submetido ao Departamento de Sistemas de Informação da Universidade Federal de Sergipe - DSII-TA/UFS, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Alcides Xavier Benicasa

ITABAIANA

2015

THIAGO DA SILVA ALMEIDA

Modelo de Detecção e Reconhecimento de Semáforos Baseado em Atenção Visual e Inteligência Artificial

Trabalho de Conclusão de Curso submetido ao Departamento de Sistemas de Informação da Universidade Federal de Sergipe - DSIITA/UFS, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Itabaiana, 24 de Fevereiro de 2015.

BANCA EXAMINADORA:

Prof. Dr. Alcides Xavier Benicasa
Orientador
DSIITA/UFS

Prof. Msc. Adolfo Pinto Guimarães
DSIITA/UFS

Prof. Msc. Andrés Ignácio Martinez Menéndez
DSIITA/UFS

Aos meus pais.

AGRADECIMENTOS

Este trabalho não teria se tornado realidade sem a participação de algumas pessoas especiais. Gostaria de agradecer primeiramente à minha família. Ao meu pai, Rivaldo, por sempre estar pronto a me ajudar no que quer que eu precise, e à minha mãe, Givaneide, por todo o carinho e cuidado que sempre tem comigo. Aos meus irmãos, Bruno pelas brincadeiras e conversas que só a gente entende, Thaynara por todo o apoio e carinho.

Ao meu orientador, Prof. Dr. Alcides Xavier Benicasa, pela confiança, paciência, e oportunidade de aprendizado que me proporcionou nos projetos que pudemos trabalhar juntos durante a graduação. Deixo aqui o meu agradecimento e admiração pelo seu exemplo profissional.

Agradeço aos companheiros de trabalho, em especial a Jeferson, pela compreensão nos momentos mais corridos deste trabalho. Foi essencial para que eu pudesse concluí-lo.

Aos demais professores do Departamento de Sistemas de Informação da Universidade Federal De Sergipe - Campus Prof. Alberto Carvalho pelo ensino de qualidade, e, em especial ao Prof. Msc. Marcos Dósea, pelos ensinamentos e conselhos durante o tempo em que fiz parte da Empresa Júnior do Departamento.

Agradeço também àqueles amigos que sempre estiveram presentes, pessoal ou virtualmente, me incentivando quando eu estive cansado, especialmente a Manoela; Nathanael, que trabalhou comigo em tantos projetos; e Rafael, parceiro pra qualquer hora.

Por fim, agradeço a todos aqueles que, de alguma forma, contribuíram para o meu crescimento profissional e pessoal.

RESUMO

É notável a quantidade de informação visual presente no trânsito, o que aumenta a possibilidade do sinal de trânsito passar despercebido, o que pode causar um acidente grave. Sendo também crescente o número de pesquisas relacionadas a sistemas de transporte inteligentes, no qual veículos possam se auto-guiar, tendo provavelmente que coexistir com veículos manualmente guiados. Observando esses dois fatos, um detector e reconhecedor de semáforos se faz bastante útil.

Este trabalho de conclusão de curso tem como objetivo propor um mecanismo de detecção e reconhecimento de semáforos que se baseia em conceitos biológicos de inteligência artificial, mais especificamente atenção visual, e de processamento de imagens.

O mecanismo proposto utiliza informações obtidas do histograma de cores da área detectada para classificar determinada cena como possuindo um semáforo vermelho ou verde, e obteve bom funcionamento nos experimentos apresentados no trabalho. Os experimentos foram realizados nos períodos noturnos e diurnos.

PALAVRAS-CHAVES: Detecção de semáforo; Reconhecimento de semáforo; Trânsito; Sistemas de Transporte Inteligentes; Atenção Visual; Inteligência Artificial; Processamento de Imagens; Reconhecimento baseado em histograma de cores.

ABSTRACT

It its known that there is a large amount of visual information present in the traffic, which increases the possibility of traffic signals go undetected, which could cause a serious accident. Besides this, there is a increasing the number of research related to intelligent transport systems, in which vehicles can be self-guided, probably having to coexist with manually guided vehicles. Watching these two facts, a detector and recognizer of traffic lights becomes very useful.

This course conclusion work aims to propose a detection mechanism and recognition of traffic lights based on biological concepts of artificial intelligence, specifically visual attention, and image processing.

The proposed mechanism uses information from the color histogram of the detected area for classifying scene as having a red or green light, and achieved good functioning in the experiments presented in the work.. The experiments were performed in the daytime and nighttime periods.

KEYWORDS: Detection of traffic light; Recognition of traffic light; traffic; Intelligent Transport Systems; Visual attention; Artificial Intelligence; Image Processing; Recognition based on color histogram.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A quantidade de informação visual e sonora disponível para ser processada pelos seres vivos é quase sempre muito grande. A capacidade de selecionar consciente ou inconscientemente determinados estímulos, sejam visuais, sonoros ou outros, dentre uma grande variedade de estímulos é essencial e intrínseca à maioria dos seres vivos. Essa capacidade biológica de atenção visual, ou sonora, nos faz capazes de reagir rapidamente a alterações no ambiente.

No caso de estímulos visuais, Benicasa (2013) afirma que alguns estímulos são naturalmente conspicuosos ou salientes em um determinado contexto. Um bom exemplo disso é como uma jaqueta vermelha posicionada entre vários ternos pretos receberá automaticamente e involuntariamente a atenção de quem visualiza o conjunto.

Tendo-se como base estudos em seres humanos e macacos, pode-se afirmar que o processo de seleção visual seleciona apenas um subconjunto da informação sensorial disponível, na forma de uma região circular do campo visual, conhecida como foco de atenção (BENICASA, 2013).

Desta forma, segundo Shic e Scassellati (2007) a atenção auxilia na redução da quantidade de informação que resulta de todas as combinações possíveis dos estímulos sensoriais pertencentes a uma cena, pois apenas informações que estão dentro da área da atenção são processadas, enquanto que o restante é suprimido (CAROTA; INDIVERI; DANTE, 2004). Considerado isso, Itti (2005) define atenção visual como um eficiente mecanismo para reduzir tarefas complexas, como análise de uma cena, em um conjunto de sub-tarefas menores.

Importante destacar que já existem sistemas computacionais baseados no mecanismo de atenção visual. Alguns modelos de atenção visual computacional são apresentados por Benicasa (2013), Itti L. (1998), Itti (2000) e Walther (2006).

Uma atividade que necessita muito da atenção visual é a direção automobilística. É notável a quantidade de atenção que um motorista precisa dispor para dirigir bem, especialmente em grandes cidades, nas quais o trânsito se torna cada vez mais caótico e, um dos artifícios utilizado mundialmente para ajudar a controlar e organizar o trânsito nas cidades é o semáforo. No entanto, o semáforo não é o único componente no trânsito que requer a atenção visual do condutor do veículo. A quantidade de itens que exigem a atenção do motorista no trânsito é imensa, por exemplo, outro veículo o ultrapassando, placas, endereço a ser procurado, pedestres atravessando a rua, algumas vezes animais na pista, motoristas imprudentes, o que pode fazer com que o semáforo passe despercebido.

Apesar de algo relativamente simples, desobedecer ao sinal fechado do semáforo pode levar a conseqüências desastrosas. Neste cenário, um reconhecedor automático de semáforos poderia alertar o motorista sobre o estado do semáforo, diminuindo assim a chance de esquecimento, que poderia vir a causar um acidente grave.

Ainda outro cenário de aplicação de reconhecedor automático de semáforos é o de uso de veículos autoguiados. Há um grande número de pesquisas nessa área, e embora diversos pesquisadores destaquem o uso de sensores entre os carros autoguiados que dispensem o uso dos semáforos, algo mais realista, de acordo com Diaz-Cabrera, M., Cerri, P., Medina-Sanchez, J. (2012), seria um cenário híbrido, com veículos autoguiados e outros não, necessitando ainda de semáforos. Os veículos autoguiados necessitariam então de um bom detector e reconhecedor de semáforos para não vir a causar acidentes.

No entanto, implementar um detector e reconhecedor de semáforos não é uma tarefa trivial, uma vez que há diversos problemas a serem vencidos para se detectar e reconhecer o estado de um semáforo de forma confiável. Entre os desafios estão a condição do tempo que altera a iluminação do ambiente e dificulta a identificação de qual sinal está ativo no semáforo, semáforos de tipos diferentes (horizontal ou vertical, suspenso ou em poste), além de outros componentes que se confundem com o semáforo por terem características comuns como a cor. O objetivo do trabalho é, então, propor um mecanismo de detecção e reconhecimento de semáforos baseado em atenção visual, que funcione bem em diferentes cenários, inicialmente diurno e noturno. Serão utilizadas técnicas de Processamento de Imagens (PI) e de Inteligência Artificial (IA) no mecanismo proposto.

O presente trabalho está organizado como segue. No Capítulo 2 são introduzidos conceitos de Processamento de Imagens e de Atenção Visual pertinentes à pesquisa, são apresentados também os trabalhos relacionados, em que são abordados métodos de detecção de semáforos baseados em atenção visual e reconhecimento de objetos utilizando histograma de cores. No Capítulo 3 o modelo proposto no trabalho é detalhado. No Capítulo 4 é possível observar e analisar diversos experimentos que validam o mecanismo proposto. Por fim, no Capítulo 5 são realizadas as conclusões e considerações finais.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo serão apresentados conceitos importantes para o entendimento do mecanismo proposto. Na primeira seção conceitos básicos de Processamento de Imagens serão explicados. Na seção seguinte o modelo de Atenção Visual, no qual o mecanismo proposto foi baseado, será detalhado e, por fim, serão apresentados trabalhos relacionados.

2.1 Processamento de Imagens

O uso de câmeras para obtenção de imagens importantes em determinado ambiente é comum, e em conjunto com uso da câmera, algoritmos de processamento de imagens podem ser utilizados para o tratamento e obtenção de informações relevantes da cena.

A área de processamento de imagens digitais tem atraído grande interesse nas últimas décadas. A evolução da tecnologia digital, aliada ao desenvolvimento de novos algoritmos, capazes de processar sinais bidimensionais, vem permitindo uma gama de aplicações cada vez maior (MORALES; CENTENO; MORALES, 2003), como por exemplo, na medicina principalmente na ajuda de diagnósticos, cirurgia guiada por computador, em geo-processamento, radares de trânsito, sensoriamento remoto na visualização do clima de uma determinada região, na arquitetura e nas engenharias (elétrica, civil, mecânica), entre outros (MORGAN, 2008).

Para Gonzalez R. C. e Woods (2010), o interesse em métodos de processamento de imagens digitais decorre de duas áreas principais de aplicação: melhoria da informação visual para a interpretação humana e processamento de dados para percepção automática através de máquinas. Segundo Grando (2005), na abordagem de Gonzáles e Woods, a primeira categoria concentra-se em técnicas para melhora de contraste, realce e restauração de imagens danificadas. A segunda categoria concentra-se em procedimentos para extrair de uma imagem informação de forma adequada, para o posterior processamento computacional. É na segunda categoria, ou seja, na percepção automática por máquinas, que se enquadra o trabalho aqui descrito.

2.1.1 Passos Fundamentais em Processamento de Imagens

Uma imagem pode ser definida como uma forma compacta de representar muitas informações. Em um sistema de processamento de imagens, essas informações podem passar por diversas etapas, as quais descrevem o fluxo das informações com um dado obje-

tivo definido pela aplicação (GONZALEZ R. C. E WOODS, 2010). A estrutura funcional completa de um sistema de processamento de imagens é ilustrada na Figura 1.

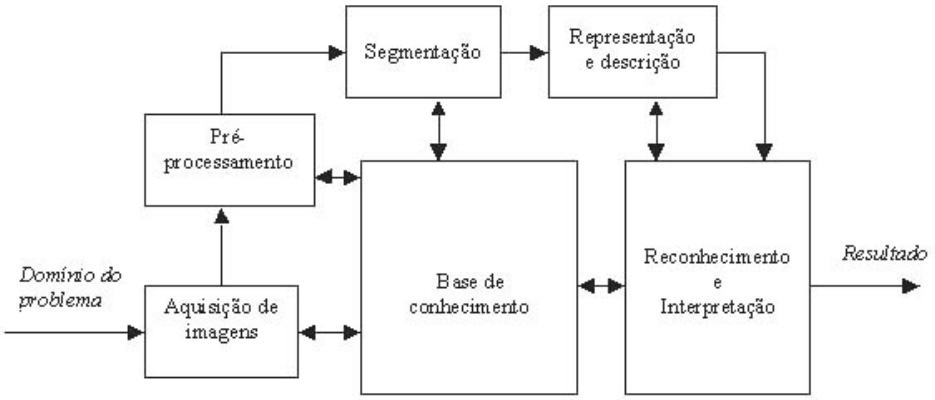


Figura 1: Estrutura funcional completa de um sistema de processamento de imagens, adaptada de Gonzalez R. C. e Woods (2010).

Como afirma Gonzalez R. C. e Woods (2010) o diagrama não significa que todo processo se aplique a uma imagem, pois as metodologias podem ser aplicadas a imagens para diferentes propósitos, com diferentes objetivos. Sendo assim, segue uma descrição das etapas mais comuns. As descrições destas etapas foram norteadas principalmente por Facon (2002) e Gonzalez R. C. e Woods (2010), sendo as seguintes:

- Aquisição da imagem: consiste em adquirir uma imagem através de um sensor e transformá-la em uma imagem digital, sobre a forma de uma tabela de valores discretos inteiros chamados de *pixel* (FACON, 2002). Dentre os aspectos envolvidos neste passo pode-se mencionar: a escolha do tipo do sensor, o conjunto de lentes a utilizar, as condições de iluminação da cena, os requisitos de velocidade da aquisição, a resolução e o número de níveis de cinza da imagem digitalizada, entre outros (GONZALEZ R. C. E WOODS, 2010);

- Pré-processamento: a imagem resultante do passo anterior pode apresentar diversas imperfeições, tais como presença de pixels ruidosos, contraste e/ou brilho inadequado, regiões interrompidas ou indevidamente conectadas, entre outras. Assim, a função do pré-processamento é melhorar a imagem de forma a aumentar as chances para o sucesso dos processos seguintes. Este passo envolve técnicas para filtragem e realce, remoção de ruído, compressão, e etc. (GONZALEZ R. C. E WOODS, 2010). O pré-processamento não é indispensável, mas necessário na maioria dos casos (FACON, 2002);

- Segmentação: consiste em dividir uma imagem em partes ou objetos constituintes, ou seja, nos objetos de interesse que compõem a imagem. A segmentação é efetuada

pela detecção de descontinuidades (contornos) e/ou de similaridades (regiões) na imagem (FACON, 2002). Em geral, a segmentação automática é uma das tarefas mais difíceis no processamento de imagens digitais. Por um lado, um procedimento de segmentação de imagens bem-sucedido aumenta as chances de sucesso na solução de problemas que requerem que os objetos sejam individualmente identificados. Por outro lado, algoritmos de segmentação inconsistentes quase sempre acarretam falha no processamento (GONZALEZ R. C. E WOODS, 2010);

- Representação e descrição: o alvo da representação é elaborar uma estrutura adequada, agrupando os resultados das etapas precedentes (FACON, 2002). A representação pode ser por fronteira e/ou regiões. A representação por fronteira é adequada quando o interesse se concentra nas características externas (cantos ou pontos de inflexão). A representação por região é adequada quando o interesse se concentra nas propriedades internas (textura ou forma do esqueleto) (GONZALEZ R. C. E WOODS, 2010). O processo de descrição, também chamado de seleção de características, procura extrair características que resultam em informação quantitativa ou que sejam básicas para a discriminação entre classes de objetos (GONZALEZ R. C. E WOODS, 2010);

- Reconhecimento e interpretação: reconhecimento é o processo que atribui um rótulo a um objeto, baseado na informação fornecida pelo descritor. A interpretação envolve a atribuição de significado a um conjunto de objetos reconhecidos (GONZALEZ R. C. E WOODS, 2010). É o passo mais elaborado do processamento de imagens digitais, pois permite obter a compreensão e a descrição final do domínio do problema, fazendo uso do conhecimento a priori e do conhecimento adquirido durante as fases precedentes (FACON, 2002);

- Base de conhecimento: o processamento de imagens digitais pressupõe a existência de conhecimento prévio sobre o domínio do problema, armazenado em uma base de conhecimento, cujo tamanho e complexidade variam dependendo da informação. Embora nem sempre presente, a base de conhecimento guia a operação de cada módulo do processamento, controlando a interação entre os módulos (GRANDO, 2005).

É possível perceber, à medida que se passa por níveis crescentes de abstração, que ocorre uma redução progressiva da quantidade de informações manipuladas. Na aquisição da imagem e no pré-processamento, os dados de entrada são *pixels* da imagem original e os dados de saída representam propriedades da imagem na forma de valores numéricos associados a cada pixel. Na segmentação, representação e descrição, esse conjunto de valores produz como resultado uma lista de características. O reconhecimento e a in-

terpretação produzem, a partir dessas características, uma interpretação do conteúdo da imagem (FACON, 2002).

2.1.2 Fundamentos de Imagens Digitais

Para Gonzalez R. C. e Woods (2010), uma imagem pertencente a uma cena pode ser definida como uma função bidimensional $f(x, y)$, composta por um determinado número de *pixels*, de modo que cada *pixel* deva possuir coordenadas de localizações x e y, associadas a um valor específico. É importante notar que o processamento de uma imagem depende diretamente destes valores. O termo *pixel* é uma abreviatura do inglês *picture element* que significa elemento da figura, corresponde a menor unidade de uma imagem digital, onde são descritos a cor e o brilho específico de uma célula da imagem (MORGAN, 2008).

Normalmente, uma imagem capturada por uma câmera de vídeo é apresentada em cores, assim, cada pixel da imagem deve ser formado por um conjunto de valores, ou também conhecido como canais de cores, geralmente de tamanho três ou quatro, podendo pertencer aos padrões de cores RGB (*red*, *green* e *blue*) ou $CMYK$ (*cyan*, *magenta*, *yellow* e *key=black*), respectivamente. Cada canal de cor, em geral, possui uma variação que vai de 0 a 255.

A combinação dos canais de cores pode representar uma grande variedade de cores, no entanto, muitas vezes essa quantidade de informação pode ser desnecessária para os objetivos de determinadas aplicações, de modo que seu processamento possa levar a desperdício de recurso. Uma forma para a resolução deste problema é a utilização de uma técnica de transformação para tons de cinza, considerado um processo simples sob os canais de cores. Aqui consideramos o padrão RGB, uma vez que este é o padrão de cores utilizado neste trabalho. A Equação 2.1, apresentada a seguir, descreve a transformação dos canais de um pixel i, pertencente ao padrão RGB, para tons de cinza, como segue:

$$I_i = \frac{R_i + G_i + B_i}{3}, \tag{2.1}$$

onde é calculada a média aritmética dos três canais de cores, R, G, e B do *pixel* i. I_i é o valor do tom de cinza obtido, também conhecido como valor de intensidade, representando o *pixel* i por um único valor.

Como afirma Grando (2005) uma imagem digital f(x,y) pode ser representada por uma matriz, cujos índices de linha e coluna identificam um ponto *(pixels)* da imagem e representam o conjunto de valores (canais de cores). Por exemplo, a Figura 2 representa uma imagem digital de 4 pixels de largura por 4 pixels de altura, cujos elementos são dados pelas intensidades dos pixels nas posições correspondentes.

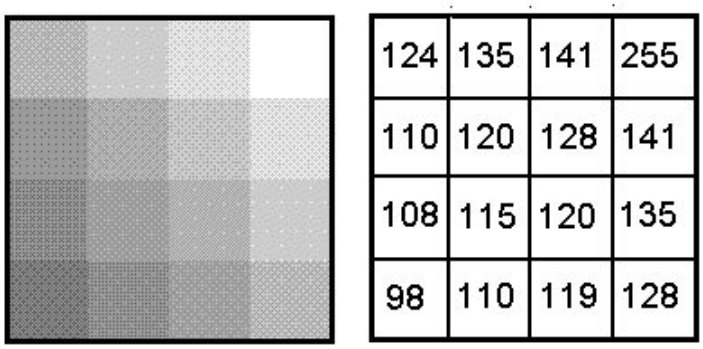


Figura 2: Imagem digital e respectiva Representação matricial, obtida de Almeida, A. B. (1998).

2.1.2.1 Limiarização - *Thresholding* e Histograma

De acordo com Marques O. F. e Vieira (1999), o objetivo da limiarização é identificar duas classes distintas na imagem, por meio do uso de um limiar para dividir a imagem em dois conjuntos de *pixels*. Considerando um limiar T, qualquer ponto x, y na imagem, tal que $f(x, y) > T$, será chamado de ponto do objeto, caso contrário, o ponto será chamado ponto de fundo (GONZALEZ R. C. E WOODS, 2010). Para Marques O. F. e Vieira (1999), esse processo também é conhecido como binarização, pois tem como resultado uma imagem binária, composta por *pixels* brancos e pretos. O processo de limiarização é descrito como segue:

$$lim(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & f(x, y) < T \end{cases} \tag{2.2}$$

Segundo Grando (2005), os métodos de limiarização têm duas abordagens distintas, uma global e outra local. O método global utiliza um único limiar T para toda imagem, já o método local têm como princípio dividir a imagem em sub-regiões, cada região tem seu limiar. Além disto estes métodos são classificados em dois grupos: manual e automático. O método manual é baseado na disposição dos níveis de cinza no histograma, sendo a escolha do limiar feita de forma empírica por um operador humano. No método automático, também baseado no histograma, não há necessidade da escolha do valor de limiar, uma vez que os próprios algoritmos retornam esse valor. A Figura 3 ilustra a influência do valor do limiar sobre a qualidade da limiarização.

Para Marques O. F. e Vieira (1999), o histograma de uma imagem é composto por um conjunto de números, indicando o percentual de *pixels* contidos na imagem, que apresentam um determinado nível de cinza. Estes valores são normalmente representados por um gráfico de barras que fornece, para cada nível de cinza, o número, ou o percentual, de *pixels* correspondentes na imagem. A informação obtida através do cálculo do histograma de cores de uma imagem pode ser útil para a indicação de sua qualidade quanto ao nível

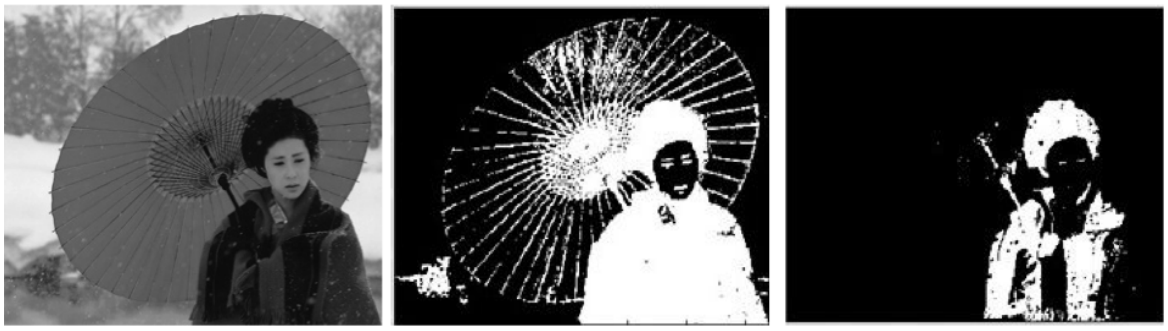


Figura 3: Influência do valor do limiar sobre a qualidade da limiarização, da esquerda para a direta: imagem original, com limiar 30, com limiar 10, obtida de Melo, N. (2009).

de contraste, brilho médio, ou demais informações a serem utilizadas para fins específicos. De acordo com Gonzalez R. C. e Woods (2010), o cálculo do histograma de cores pode ser descrito como:

$$h(r_k) = n_k \tag{2.3}$$

onde r_k é o k-ésimo nível de cinza e n_k é o número de *pixels* da imagem contendo o nível de cinza r_k. Na Figura 4 é apresentado um exemplo da aplicação das Equações 2.1 e 2.3.

Gonzalez R. C. e Woods (2010) afirma que histogramas são fáceis de serem calculados utilizando-se um aplicativo computacional, inclusive implementações econômicas de *hardware* podem ser usadas para este fim, e por esse motivo o histograma se torna uma ferramenta popular para o processamento de imagens em tempo real.

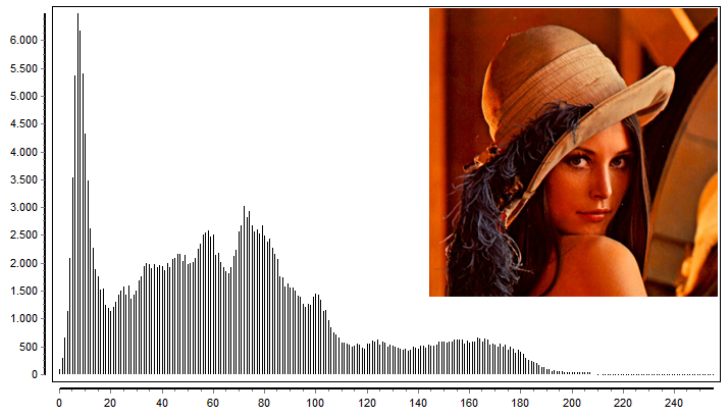


Figura 4: Exemplo de transformação para tons de cinza e cálculo do histograma de cores. Imagem Lena (512x512*pixels*) obtida de Gonzalez R. C. e Woods (2010).

2.2 Atenção Visual

A quantidade de informação visual e sonora disponível para ser processada pelos seres vivos é quase sempre muito grande. A capacidade de selecionar consciente ou incons-

cientemente determinados estímulos, sejam visuais, sonoros ou outros, dentre uma grande variedade de estímulos é essencial e intrínseca à maioria dos seres vivos. Essa capacidade biológica de atenção visual, ou sonora, nos faz capazes de reagir rapidamente a alterações no ambiente.

No caso de estímulos visuais, Benicasa (2013) afirma que alguns estímulos são naturalmente conspicuosos ou salientes em um determinado contexto. Para exemplificar, pode-se imaginar uma jaqueta vermelha posicionada entre vários ternos pretos. A jaqueta receberá, automática e involuntariamente, a atenção de quem visualiza o conjunto.

Tendo-se como base estudos em seres humanos e macacos, pode-se afirmar que o processo de seleção visual seleciona apenas um subconjunto da informação sensorial disponível, na forma de uma região circular do campo visual, conhecida como foco de atenção (BENICASA, 2013).

Desta forma, segundo Shic e Scassellati (2007) a atenção auxilia na redução da quantidade de informação que resulta de todas as combinações possíveis dos estímulos sensoriais pertencentes a uma cena, pois apenas informações que estão dentro da área da atenção são processadas, enquanto que o restante é suprimido (CAROTA; INDIVERI; DANTE, 2004). Considerado isso, Itti (2005) define atenção visual como um eficiente mecanismo para reduzir tarefas complexas, como análise de uma cena, em um conjunto de sub-tarefas menores.

Wolfe e Horowitz (2004) demostraram que algumas características como cor, orientação ou tamanho dos objetos em uma imagem são responsáveis por guiar o mecanismo biológico de atenção visual. Para o entendimento do processo de atenção visual, é importante observar que a busca por um ponto de maior atenção ou saliência pode ser simples e eficiente em alguns casos, porém não tão simples para outros (BENICASA, 2013).

A Figura 5 mostra algumas destas características. Na Figura 5(a), o contraste entre o azul e o vermelho ressalta a existência de um numeral 5 (cinco) de cor diferente dos demais. No entanto, perceber um número cinco azul e maior é um pouco mais complicado. A Figura 5(a) também é um exemplo da importância de conhecimento a priori para executar determinadas buscas visuais, pois dificilmente é possível identificar o número dois existente nesta imagem sem que alguém tenha dito que há um número dois. As Figuras 5(b) e 5(c) demonstram a importância da orientação e do contraste de cores para ressaltar objetos diferentes em imagens. Na Figura 5(b) é difícil encontrar os pares de triângulos horizontais, mas esta tarefa é simplificada devido ao contraste de cores entre os retângulos azuis e os retângulos rosas. Na Figura 5(d), a busca por cruzes é ineficiente devido ao fato de que aqui a informação de intersecção não guia a atenção.

Como afirma Pereira (2007), em uma visão didática, podem ser identificados dois métodos principais para obtenção da atenção visual. Os métodos *top-down* e *bottom-up*.

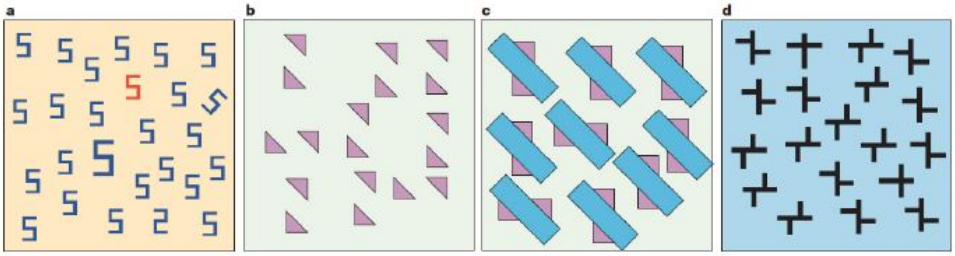


Figura 5: Exemplos de busca visual Wolfe e Horowitz (2004).

O método *top-down* usa conhecimentos obtidos a priori para detectar regiões de maior interesse numa imagem. Esses conhecimentos podem ser obtidos de várias formas. Geralmente, utilizam-se ferramentas de aprendizagem baseadas em modelos geométricos/relacionais (como redes semânticas ou grafos relacionais) ou modelos estatísticos (como redes neurais e máquinas de vetores de suporte). Porém, esses conhecimentos também podem ser fornecidos por um ser humano, selecionando-se manualmente regiões de maior interesse numa imagem. A atenção visual *bottom-up* é guiada por características primitivas da imagem como cor, intensidade e orientação. Além disso, ela atua de modo inconsciente, ou seja, o observador é levado a fixar sua atenção em determinadas regiões da imagem devido aos estímulos causados pelos contrastes entre características visuais presentes na imagem.

O sistema de atenção visual *bottom-up* proposto por Itti L. (1998) é um dos mais conhecidos e utilizados atualmente para seleção de regiões salientes em imagens. A seguir serão descritos os principais aspectos desse modelo.

2.2.1 Modelo de Saliência

Na Figura 6, uma adaptação de Itti L. (1998), é apresentado o modelo do mapa de saliências. O modelo pode ser descrito nas seguintes etapas: extração de características, filtragem linear, diferenças centro-vizinhança, normalização e soma dos mapas de características. A imagem de entrada é decomposta em três mapas de características: intensidade, cor e orientação. Os mapas de características são criados através de pirâmides de Gauss e Gabor, através de sucessivas filtragens e sub-amostragens da imagem de entrada (ITTI L., 1998).

Para o entendimento da geração de um mapa de saliência, serão descritos a seguir os principais aspectos do modelo proposto em Itti e Koch (2001).

2.2.1.1 Extração de Características Visuais Primitivas

Para gerar um mapa de saliência, três tipos de características visuais primitivas são extraídas: cor, intensidade e orientação. Quatro canais de cores são criados (R para vermelho, G para verde, B para azul e Y para amarelo). Sendo r, g, b os canais vermelho,

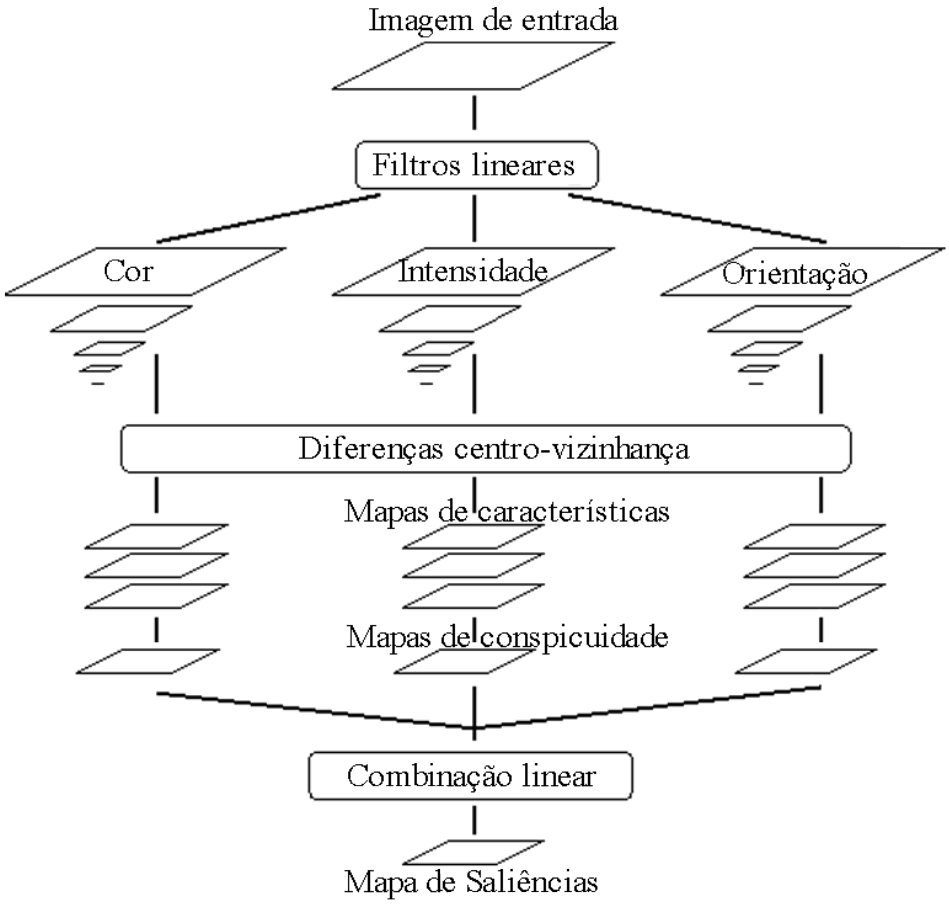


Figura 6: Arquitetura do modelo do mapa de saliências, adaptada de Itti L. (1998).

verde e azul da imagem de entrada, os canais de cores são representados por (BENICASA, 2013):

$$R = \frac{r - (g + b)}{2} \tag{2.4}$$

$$G = \frac{g - (r + b)}{2} \tag{2.5}$$

$$B = \frac{b - (r + g)}{2} \tag{2.6}$$

$$Y = \frac{(r + g)}{2} - \frac{|r - g|}{2} - b \tag{2.7}$$

A imagem de intensidades é representada por $I = (r+g+b)/3$, que define a imagem em tons de cinza. A Figura 7 apresenta um exemplo de extração das características.

Os canais de cores e a imagem de intensidades são submetidos a um processo de filtragem linear. Este processo é realizado por meio da geração de Pirâmides Gaussianas

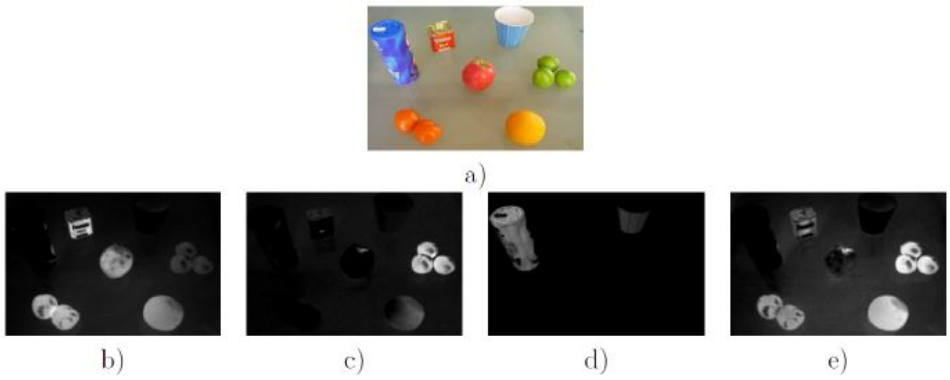


Figura 7: Extração de 4 canais de cores. a) Imagem de Entrada, b) Extração do canal vermelho, c) Canal verde, d) Canal azul e e) Canal amarelo (SIKLOSSY, 2005 apud BENICASA, 2013)

e Pirâmides Direcionais. A Pirâmide Gaussiana é composta por versões filtradas passa-baixa da convolução Gaussiana aplicada à imagem de entrada. A Pirâmide Direcional é uma decomposição multi-escala e multi-orientação de uma imagem. Nesta decomposição linear, uma imagem é subdividida em um conjunto de sub-bandas localizadas em escala e orientação. A representação piramidal é usada para a obtenção de amostras da imagem sem detalhes indesejáveis. A seguir, os processos de geração das Pirâmides Gaussianas e Direcionais são detalhados.

2.2.1.2 Pirâmide Gaussiana

As Pirâmides Gaussianas são geradas utilizando um algoritmo proposto por Burt e Adelson apud Benicasa (2013), a imagem de entrada é representada por uma matriz G_0, essa matriz contém c colunas e r linhas de pixels. Para cada nível da pirâmide é gerada uma imagem em uma escala menor que a escala no nível superior. A imagem de entrada é a base ou nível zero da Pirâmide Gaussiana. Cada nível inferior da pirâmide contém uma imagem que é uma redução ou uma versão filtrada passa-baixa da imagem da base da pirâmide (BENICASA, 2013). Dessa forma é possível obter versões filtradas da imagem original, de uma maneira que embora se perca informação para gerar cada nível da pirâmide, a informação realmente importante e que se destaca na imagem permanecerá. Uma pirâmide Gaussiana G_θ pode ser definida recursivamente como segue:

$$G_\theta(x,y) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m+2,n+2)\ G(x,y), \quad para\ \theta = 0 \tag{2.8}$$

$$G_\theta(x,y) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m+2,n+2)\ G_{\theta-1}(2x+m,2y+n), \quad para 0 < \theta \leq 8 \tag{2.9}$$

onde $w(m, n)$ são os pesos gerados a partir de uma função Gaussiana, utilizados para gerar os níveis da pirâmide para todos os canais. A Figura 8 mostra um exemplo da Pirâmide Gaussiana.

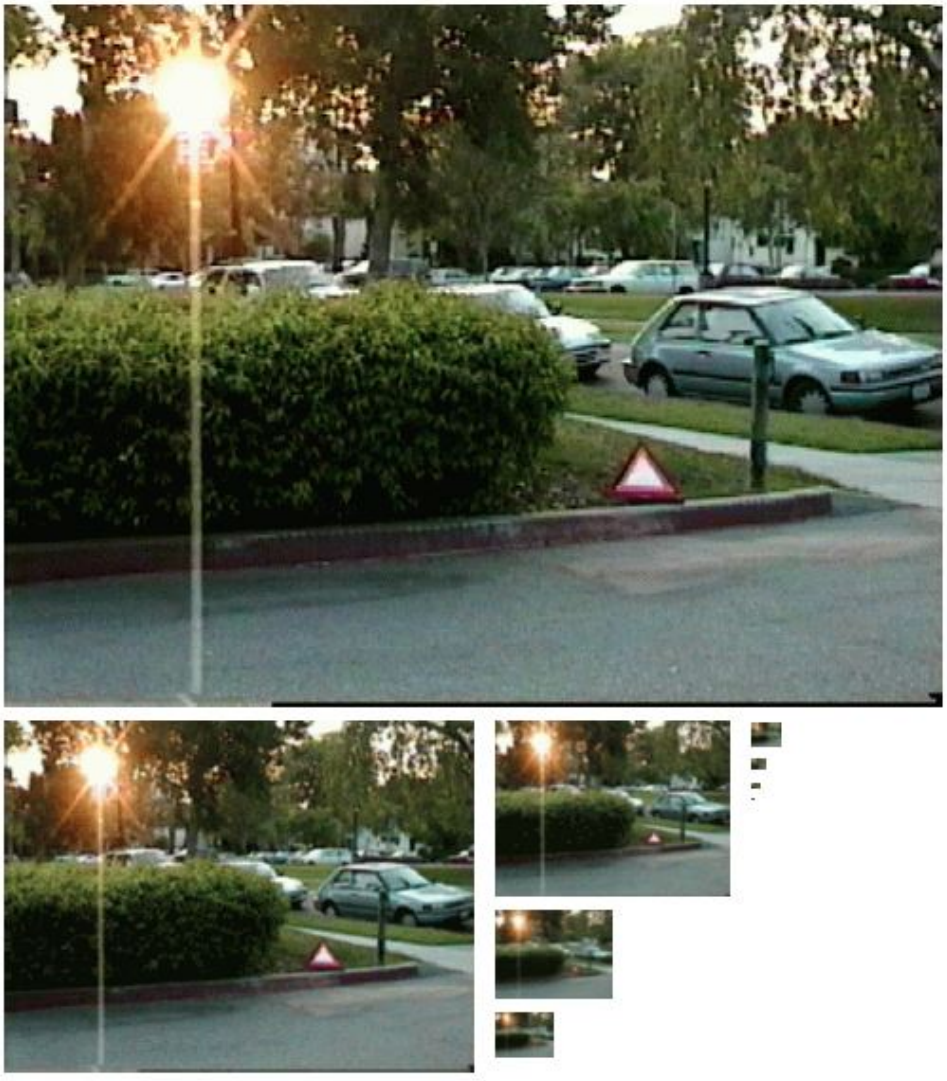


Figura 8: Representação piramidal, usada para a obtenção de amostras da imagem sem detalhes indesejáveis, obtida de Itti (2000)

2.2.1.3 Pirâmide Direcional

O modelo de (ITTI L., 1998) também considera informações sobre orientações locais como uma característica importante para o desenvolvimento da atenção visual. Na Figura 9 é apresentado um exemplo em que o contraste na orientação pode guiar a atenção visual.

Os mapas de orientações $O_\theta(\theta)$ são criados através da convolução do mapa de intensidades I_θ, com filtros direcionais de Gabor para quatro orientações $\theta \in 0°, 45°, 90°, 135°$. A aplicação destes filtros visa identificar barras ou bordas em uma determinada direção, para isso utiliza-se de uma função gaussiana (BENICASA, 2013).

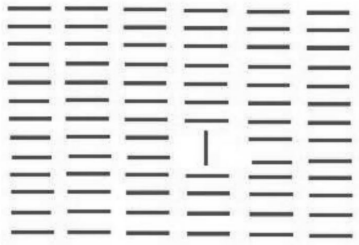


Figura 9: Exemplo de orientação, barra vertical inserida em um ambiente com barras horizontais torna-se o elemento mais saliente devido a grande diferença de orientação (SIKLOSSY, 2005 apud BENICASA, 2013)

2.2.1.4 Diferenças Centro-Vizinhança

Os mapas de características são obtidos por meio da diferença entre canais de cores em diferentes escalas, este processo é conhecido como diferença centro-vizinhança. Nesta subtração de imagens, o centro é um *pixel* da imagem em uma escala $c \in \{2, 3, 4\}$ e a vizinhança é o *pixel* correspondente de outra imagem em uma escala $s = c + \sigma$ com $\sigma \in \{3, 4\}$ da pirâmide (PEREIRA, 2007).

A subtração destas duas imagens é obtida pela interpolação das imagens para a escala fina, seguida da subtração ponto a ponto (BENICASA, 2013). O primeiro conjunto de mapas é construído a partir do contraste de intensidades, definido como segue:

$$\mathcal{I}(c, s) = |I(c) \ominus I(s)| \tag{2.10}$$

que apresenta inspiração biologicamente baseada nos mamíferos, onde o contraste de intensidade é detectado por neurônios sensíveis a centros escuros com vizinhança clara e por neurônios sensíveis a centros claros com vizinhança escura (ITTI; KOCH, 2001). O segundo conjunto de mapas é similarmente construído a partir dos canais de cores, definidos como:

$$\mathcal{RG}(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \tag{2.11}$$

$$\mathcal{BY}(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))|, \tag{2.12}$$

onde a inspiração biológica para a construção desse conjunto de mapas é a existência, no córtex visual, do chamado sistema de cores oponentes: no centro de seus campos receptivos, neurônios são excitados por uma cor e inibidos por outra e vice-versa. Tal sistema existe para vermelho=verde, verde=vermelho, azul=amarelo, amarelo=azul (ITTI; KOCH, 2001). O terceiro conjunto de mapas é construído a partir de informações de orientação local, de acordo com as seguintes equações:

$$\mathcal{O}(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|, \tag{2.13}$$

onde $\theta \in 2, 3, 4$. Neste caso, a inspiração biológica para a construção dos mapas de orientação é a propriedade de neurônios do sistema visual de responder apenas a uma determinada classe de estímulos, como por exemplo barras orientadas verticalmente (ITTI; KOCH, 2001).

2.2.1.5 Saliência

Segundo Benicasa (2013), a maioria dos modelos de atenção *bottom-up* inspirados biologicamente segue a hipótese de Koch and Ullman (1985), onde vários mapas de características alimentam um único mapa mestre ou mapa de saliência.

O mapa de saliência é um mapa escalar bidimensional de atividade representado topograficamente pela conspicuidade ou saliência visual (ITTI; KOCH, 2001). Uma região ativa em um mapa de saliência codifica o fato desta região ser saliente, não importando se esta corresponde, por exemplo, a uma bola vermelha meio a bolas verdes, ou a um objeto que se move para a esquerda enquanto outros se movem para a direita (BENICASA, 2013).

Para a construção de um único mapa de saliência, os mapas de características são individualmente somados ($\oplus$) nas diversas escalas, gerando três mapas deconspicuidades: $\widetilde{I}$ para intensidade, $\widetilde{C}$ para cor e $\widetilde{O}$ para orientação. Entretanto, um fator importante a ser notado é que, previamente à somatória dos mapas de cada característica, Itti L. (1998) propõem sua normalização, denotada por *N(.)*, com o objetivo de que uma região que apresente um nível de saliência contrastante com as demais seja amplificada e, por outro lado, regiões salientes não contrastantes sejam mutuamente inibidas. A Figura 10 demonstra a função da normalização *N(.)*.

Após o processo de normalização, os mapas de características são então combinados em três mapas de conspicuidades, conforme descrito anteriormente, definidos como segue:

$$\bar{\mathcal{I}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c, s)), \tag{2.14}$$

$$\bar{\mathcal{C}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c, s)) + \mathcal{N}(\mathcal{BY}(c, s))], \tag{2.15}$$

$$\bar{\mathcal{O}} = \sum_{\theta \in \{0°, 45°, 90°, 135°\}} \mathcal{N}\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c, s, \theta)) \right) \tag{2.16}$$

De acordo com Itti L. (1998), a motivação para a criação dos três canais separados ($\bar{\mathcal{I}}$,$\bar{\mathcal{C}}$,$\bar{\mathcal{O}}$) é a hipótese de que características similares competem pela saliência,

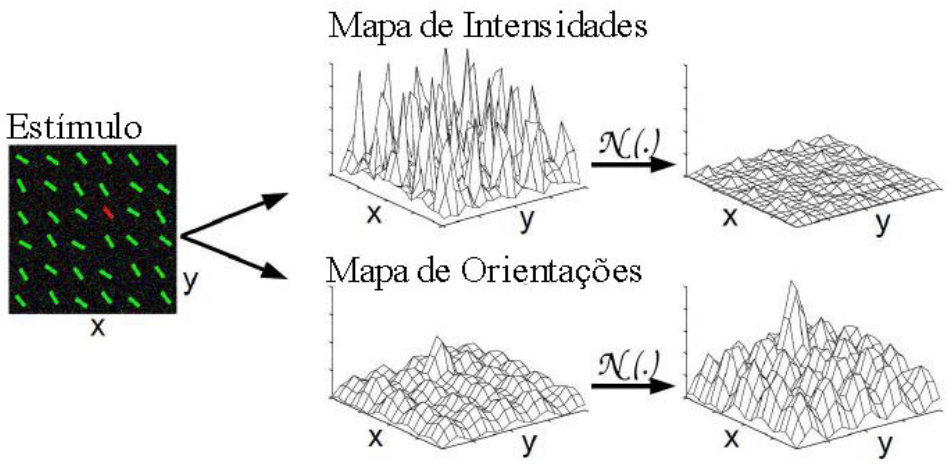


Figura 10: Exemplo do comportamento do operador de normalização *N(.)*, obtido de Benicasa (2013)

enquanto modalidades diferentes contribuem independentemente para o mapa de saliência.Finalmente, os três mapas de conspicuidades são normalizados e somados, resultando em uma entrada final para o mapa de saliência s, como segue:

$$\mathcal{S} = \frac{1}{3}(\mathcal{N}(\bar{\mathcal{I}}) + \mathcal{N}(\bar{\mathcal{C}}) + \mathcal{N}(\bar{\mathcal{O}})) \tag{2.17}$$

2.3 Trabalhos Relacionados

Modelos de atenção visual tem sido amplamente utilizados na literatura. Por exemplo, Jacob, H. (2013) apresenta um modelo de atenção visual para sumarização automática de vídeos de programas televisivos. O objetivo é que o modelo de atenção visual baseado no sistema visual humano e em visão computacional proposto pelo autor permita a identificação de trechos de vídeo mais relevantes em termo de conteúdo.

Atenção visual também tem sido fortemente aplicada na área de robótica. Santana, P. (2011) e Pais, G., Munhoz, V., Policastro, C. (2009) apresentam suas pesquisas sobre robôs inspirados biologicamente que utilizam visão computacional, o primeiro mais focado na interação do robô com o ser humano, e o segundo na identificação de obstáculos e caminhos.

Outra área de aplicação de atenção visual é de detecção e reconhecimento de placas de trânsito conforme trabalhos de Rodrigues, F. (2002) e Poffo, F. (2010), entre outros.

A área de Sistemas de Transporte Inteligentes (*Intelligent Transport Systems* - ITS, sigla em inglês), na qual o presente trabalho está inserido, é bastante vasta e é composta de subáreas como: redes veículares, comunicação entre veículos e componentes da via, monitoração de veículos e de vias, entre várias outras. Existem vários centros de pesquisa,

alguns inclusive governamentais, como por exemplo U.S. Department of Transportation (US DOT) (2015), European Comission of Mobility and Transport (2015), Australian government - Department of Infraestructure and Regional Development (2015) e LaRA (2015).

Diversos tipos de reconhecedores têm sido sugeridos na literatura, e embora todos utilizem alguma técnica de processamento de imagens, nenhum dos estudados até o momento aborda diretamente atenção visual. Alguns dos trabalhos mais relevantes serão resumidos a seguir.

Em Chung Y. (2002), foi proposto um método de detecção e reconhecimento utilizando informação de cor do semáforo, seu trabalho é um dos mais citados entre os trabalhos mais recentes. No método proposto pelos autores uma vez que a performance de sistemas de visão computacional é muito afetada num ambiente aberto por conta da variação de iluminação, inicialmente uma série de imagens do ambiente são utilizadas pelo sistema para estimar parâmetros de iluminação. Após uma filtragem por cor, um método *fuzzy* é aplicado à imagem filtrada, juntamente com parâmetros de iluminação gerados no início do processo. Ainda outro método *fuzzy* é aplicado após isso para eliminar ruídos e destacar áreas que podem ser semáforos. Para eliminar áreas que foram destacadas na fase anterior, mas que são falsos positivos, Chung Y. (2002) utilizam informação relativa temporal e espacial, pois, de acordo com os autores, os sinais verde, amarelo e vermelho possuem relação de espaço bem definidas entre si, bem como de sequência de mudança de sinal (verde, após o verde sempre o amarelo, após o amarelo sempre o vermelho, e após o vermelho sempre o verde).

Outro método que utiliza informação de cor é o proposto por Yang X. (2008). Para a fase de detecção, os autores aplicam ajuste na imagem de entrada a fim de obter o que eles chamam de puros vermelho, verde e azul. A partir daí um filtro de cor verde é aplicado com o objetivo de remover todos os componentes que não forem verdes. Após isso é aplicado um algoritmo de limiarização, e em seguida um filtro de mediana para remover ruídos.

Para o reconhecimento do semáforo, Yang X. (2008) utilizam um algoritmo de correlação cruzada normalizada, que, de acordo com os autores, é bastante efetivo em imagens com grande variação na escala de cinza. Após isso, *templates* são utilizados para reconhecer definitivamente a forma da área destacada e escolhida pelos algoritmos anteriores. Importante ressaltar que o trabalho citado reconhece sinais circulares, em forma de seta e sinais do semáforo para pedestres. Os testes foram realizados em horário diurno e noturno, bem como em dias nublados e com neve.

Um método de detecção sem uso de informação de cor foi proposto por Charette R. (2009). Em vez de usar a cor para isso, os autores detectam a luz do semáforo utilizando o algoritmo *Top-Hat*, e filtram o resultado da aplicação do algoritmo usando as propriedades

conhecidas das luzes dos semáforos, tais como: raio e a luz do semáforo. Num processo seguinte de eliminação de falsos positivos, para cada luz detectada é encontrada sua posição na imagem original, e daí aplica-se um algoritmo de crescimento na região. Se a área identificada for similar a área destacada pelo Top-Hat, a luz detectada provavelmente é real. Para reconhecimento final do semáforo, templates geométricos adaptativos são aplicados nas luzes candidatas remanescentes. Interessante notar que os templates foram projetados de forma a se implementar facilmente diversos formatos de semáforo com pouca mudança no código. Os autores disponibilizam também um benchmark de imagens que outros pesquisadores podem usar para testar em suas pesquisas, e se desejarem, comparar os resultados.

Gong J. (2010) desenvolveram um método de detecção e reconhecimento utilizando segmentação de cor. Num primeiro momento os autores utilizam 150 imagens de diferentes condições de iluminação, brilho e ambiente de fundo para calcular valores estatísticos de azul, vermelho e verde. A partir daí uma imagem binária é obtida e um processo morfológico é utilizado para diminuir o ruído na imagem binária, o processo de redução de ruído ainda é completado com algoritmos de erosão e dilatação da imagem e repetido diversas vezes a depender da resolução da imagem. Num processo semelhante ao de Charette R. (2009), os locais detectados na imagem como possíveis candidatos a serem semáforos são destacados na imagem original, e daí as regiões das imagens originais podem ser identificadas com base num aprendizado de máquina feito previamente com diversas imagens de semáforos.

Apesar de em número consideravelmente menor, já se encontra na literatura métodos capazes de identificar semáforos à noite. Um desses métodos foi proposto por Kim H. (2013), utilizando informação de cor para selecionar regiões candidatas a semáforo, por meio do isolamento de áreas verdes e vermelhas, utilizando uma transformada de cor, após, falsos candidatos são eliminados com base nas características conhecidas de uma luz de semáforo (raio, por exemplo). Um classificador SVM (*Suport Vector Machine*) é utilizado em conjunto com três outros algoritmos com o objetivo de identificar o significado do semáforo. Este classificador é específico para o horário noturno e só funciona com semáforos horizontais.

Classificação baseada em histograma de cores tem sido bastante usada na literatura pela facilidade e rapidez de se calcular o histograma. Casati e Rodrigues (2010) apresentam um trabalho no qual utilizam histograma de cores quantizado por misturograma para reconhecimento facial. Motoki (2006) também faz uso de informações do histograma de cores para classificação de rochas ornamentais com base nas suas cores. Histogramas também são utilizados para classificação de texturas no trabalho de Liu (2006), e Chapelle (2010) utiliza informações de histograma associadas a algoritmo SVM com o objetivo de classificar imagens.

Na Tabela 1, a seguir, observa-se os principais trabalhos relacionados citados anteriormente, com algumas das suas características em destaque. Na última linha da tabela, encontra-se o que se pretende alcançar com relação a estas características no trabalho proposto.

Tabela 1: Comparação de algumas características importantes dos trabalhos relacionados com o trabalho proposto

Autor	Ano	Usa Cor	Horário	Tipo de Semáforo
CHUNG et al.	2002	SIM	Diurno e noturno	Horizontal
YANG et al.	2008	SIM	Diurno e noturno	Horizontal
CHARETTE et al.	2009	NÃO	Diurno	Horizontal e Vertical
GONG et al.	2010	SIM	Diurno	Horizontal
KIM et al.	2013	SIM	Noturno	Horizontal
ALMEIDA	2015	SIM	Diurno e noturno	Horizontal e Vertical

O mecanismo que se propõe neste trabalho utiliza informação de cor tanto para detecção como para o reconhecimento. Durante essa fase, o passo inicial consiste em criar o mapa de saliência da imagem com o fim de destacar áreas vermelhas e verdes. Na fase de reconhecimento, o mecanismo proposto utiliza informações do histograma de cores calculado usando o resultado da fase de detecção. O processo detalhado será explicado no capítulo a seguir.

3 PROPOSTA DE MODELO DE DETECÇÃO E RECONHECIMENTO DE SEMÁFOROS

Com o objetivo de detectar e reconhecer o semáforo em uma cena, o presente trabalho propõe um mecanismo, construído com base nos modelos e conceitos de processamento de imagens e atenção visual, apresentados no capítulo anterior.

Este capítulo apresentará detalhadamente o mecanismo desenvolvido. Inicialmente serão abordados os passos de criação do mapa de saliência, e como o mapa de saliência apresentado difere do mapa de saliência proposto por Itti (2005). Em seguida, serão apresentados os passos referentes à classificação com base em histogramas.

O modelo proposto pode ser dividido em duas partes, a detecção de área saliente e, em seguida, o processo de classificação dessa área. Ambos os processos são norteados pela Atenção Visual Computacional. O diagrama da Figura 11 apresenta o fluxo do mecanismo proposto, que pode ser descrito brevemente como: o processo se inicia com a obtenção da imagem de entrada, provinda de uma câmera posicionada dentro de um veículo e direcionada para a parte dianteira do veículo; após isso a imagem obtida é tratada e, em seguida, processada com o algoritmo de atenção visual com fim de obter o mapa de saliência; no passo seguinte o mapa de saliência é limiarizado; esse mapa limiarizado é utilizado para ponderar o histogramas de verde e vermelho da imagem original; e por fim, a classificação se dá com base nestes histogramas. Nas seções a seguir o processo será mais profundamente detalhado.

3.1 Obtenção e Pré-processamento de Imagem

O primeiro passo do mecanismo proposto é a obtenção da imagem de entrada para o processamento. Neste caso, uma câmera é utilizada e um conjunto de fotos ou vídeos são obtidos. A Figura 12 apresenta um exemplo de imagem de entrada.

Na fase seguinte, de pré-processamento, os dados de entrada são tratados. Com o objetivo de melhorar a velocidade de processamento faz-se necessário redimensionar a mídia obtida caso a resolução seja muito alta, sendo que essa perda de qualidade não influencia no resultado do processamento posterior. A resolução utilizada foi de 640 *pixels* de largura por 480 *pixels* de altura. No caso de vídeos ainda é necessário realizar a extração de quadros (*frames*), uma vez que o procedimento se dá foto a foto.

Em muitos casos, o objeto procurado em uma determinada cena se encontrará sempre em uma determinada localidade da mesma, conforme observado na Figura 13, que

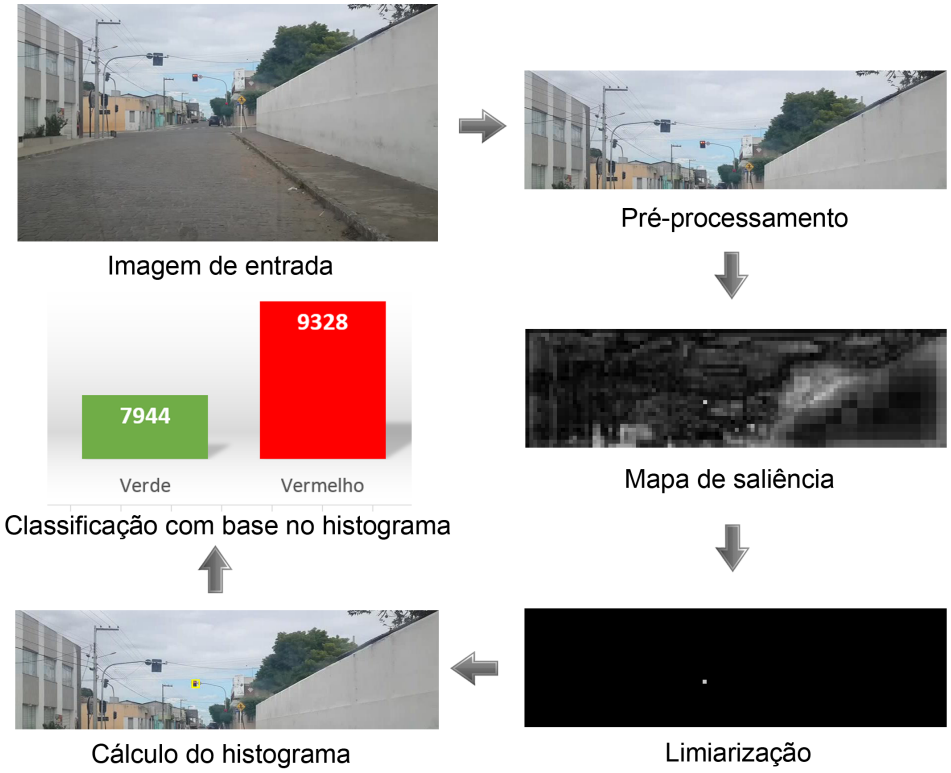


Figura 11: Diagrama que representa o fluxo do mecanismo proposto

Figura 12: Exemplo de imagem de entrada

apresenta o semáforo sempre na parte superior da imagem, dessa forma é possível definir uma altura ou largura de corte.

Uma vez que o semáforo sempre aparecerá na metade superior da cena foi utilizada uma altura de corte chamada de θ_{cut}, esse valor varia de 0 a 1, sendo que 1 representa a imagem inteira e 0 não realiza processamento algum pois desconsidera a imagem. Na Figura 14 pode-se observar a imagem de entrada apresentada anteriormente após a fase de pré-processamento. Na Figura 14(a) observamos a imagem pré-processada com $\theta_{cut} =$

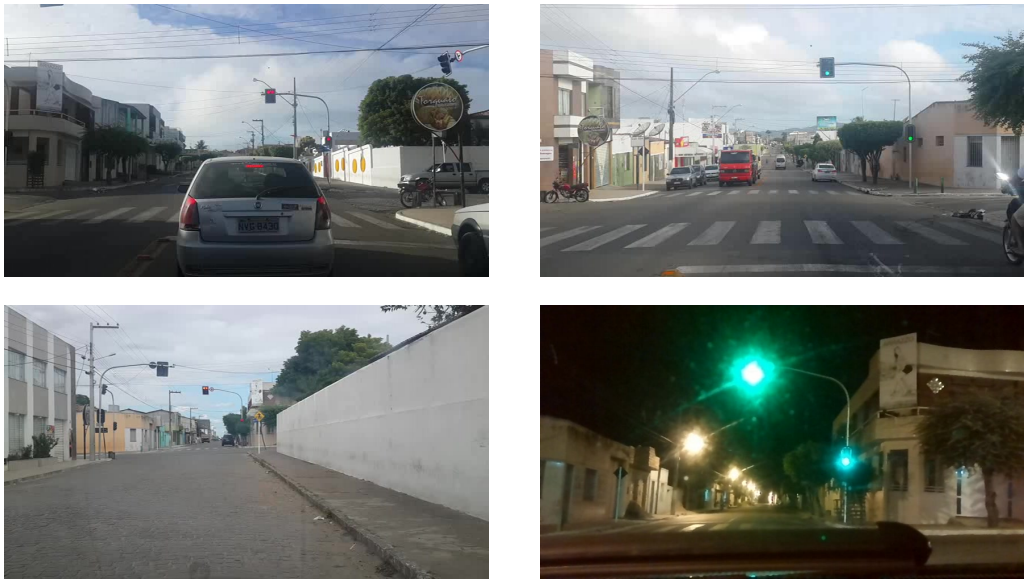


Figura 13: Exemplos de imagens de semáforos obtidas do interior de um veículo

0.7, e na Figura 14(b) a mesma imagem pré-processada com $\theta_{cut} = 0.5$.

(a) $\theta_{cut} = 0.7$

(b) $\theta_{cut} = 0.5$

Figura 14: Exemplos de imagens pré-processadas com diferentes valores de θ_{cut}

3.2 Detecção de Semáforo

3.2.1 Criação do Mapa de Saliência

Para detectar os pontos mais salientes da imagem de entrada, que são os pontos vermelhos ou verdes da luz do semáforo, foi utilizado o algoritmo de mapa de saliência proposto por Itti L. (1998), este foi explicado com maiores detalhes no capítulo anterior. Algumas adaptações foram feitas no algoritmo implementado, para que este se encaixasse melhor às necessidades da pesquisa corrente.

O primeiro passo do algoritmo é a extração dos canais de cores, seguido da aplicação de uma pirâmide gaussiana, conforme proposto por Itti L. (1998). Os canais de cores R (vermelho) e G (verde) extraídos da imagem apresentada na Figura 14(b) utilizando as Equações 2.4 e 2.5, respectivamente, podem ser observados na Figura 15. Foram utilizados apenas os canais R e G pois são as cores de semáforo que se procura como pontos salientes na cena. Essa diferença entre o modelo proposto e o modelo de Itti L. (1998), permite obter maior velocidade de processamento, uma vez que a manipulação da imagem é diminuída pela metade ao desconsiderar dois dos quatro canais disponíveis.

(a) Canal R

(b) Canal G

Figura 15: Canais de cores R e G extraídos da imagem apresentada na 14b

A pirâmide gaussiana utilizada nesta pesquisa possui 5 níveis e foi calculada utilizando-se uma matriz de pesos de tamanho 4x4, gerados por uma função gaussiana. Os valores da matriz de pesos utilizada podem ser vistos na Tabela 2.

A pirâmide Gaussiana G_θ utilizada pode ser definida recursivamente como segue:

Tabela 2: Matriz de pesos utilizados para calcular a pirâmide gaussiana

1	1	1	1
1	10	10	1
1	10	10	1
1	1	1	1

$$G_\theta(x,y) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m+2,n+2)\ G(x,y), \quad para\ \theta = 0 \tag{3.1}$$

$$G_\theta(x,y) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m+2,n+2)\ G_{\theta-1}(2x+m,2y+n), \quad para 0 < \theta \leq 4 \tag{3.2}$$

onde $w(m,n)$ são os pesos apresentados na Tabela 2, utilizados para gerar os níveis da pirâmide para todos os canais.

Na Figura 16 observa-se as pirâmides gaussianas dos canais de cores R e G apresentados na Figura 15. Observa-se que a imagem perde qualidade, e ao mesmo tempo em que parte da informação é perdida, as regiões mais relevantes da imagem permanecem e ganham destaque. Esse passo é fundamental a fim de separar apenas regiões que se sobressaiam na imagem.

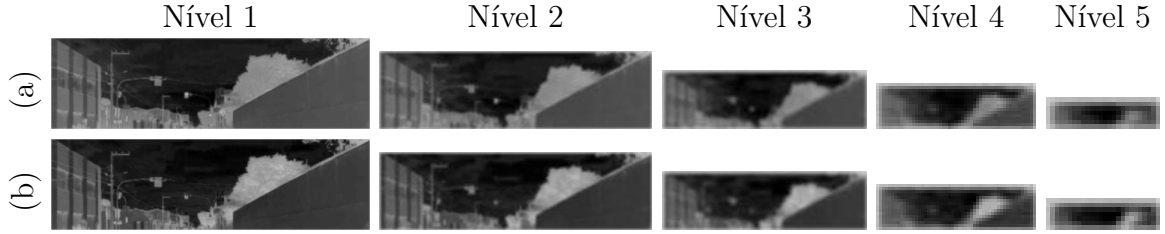


Figura 16: Pirâmides gaussianas da imagem apresentada na 14(b). a) Canal R; b) Canal G.

Aplicando a Equação 2.11 aos níveis das pirâmides gaussianas geradas, é possível calcular os mapas de características de cor para o canal de cor RG (vermelho e verde combinados). A Figura 17, a seguir, permite observar os quatro mapas de características RG gerados a partir da imagem da Figura 14(b).

De posse dos mapas de características faz-se necessário calcular o mapa de conspicuidade RG, que constitui na soma e normalização dos mapas de características. A normalização, proposta por Itti L. (1998), é realizada previamente à somatória dos mapas de cada característica, com o objetivo de amplificar regiões que apresentem um nível de saliência que a contraste das demais, bem como inibir regiões salientes não contrastantes. A Figura 10, apresentada no capítulo anterior, demonstra a função da normalização realizada.

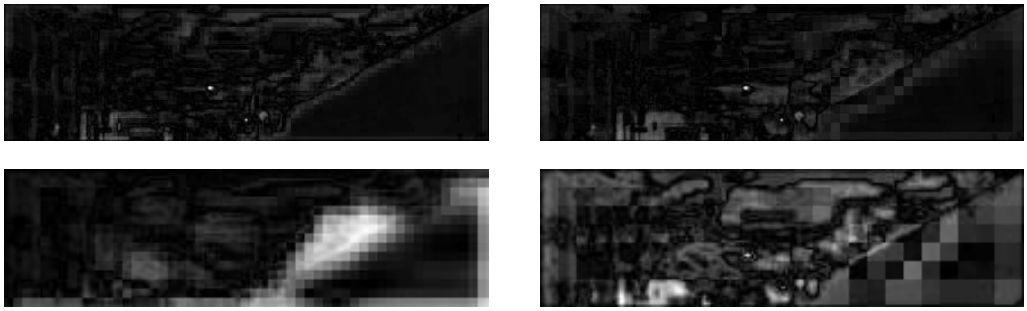


Figura 17: Mapas de características RG extraídos da imagem da Figura 14(b)

O mapa de saliência proposto por Itti L. (1998) é constituído por três mapas que combinados destacam pontos salientes aos olhos humanos, seguindo um comportamento semelhante ao biológico, estes são: mapa de orientações, mapa de cores e mapa de intensidade. No entanto, as regiões que são procuradas na imagem de entrada se destacam apenas por cor, e não por orientação (vertical, horizontal, diagonal) ou por intensidade da região, por esse motivo, o mapa de saliência final utilizado na pesquisa é formado apenas pelo mapa de conspicuidade RG. Esse mapa de cores destaca as regiões vermelhas e verdes da imagem de entrada, conforme observa-se na Figura 18.

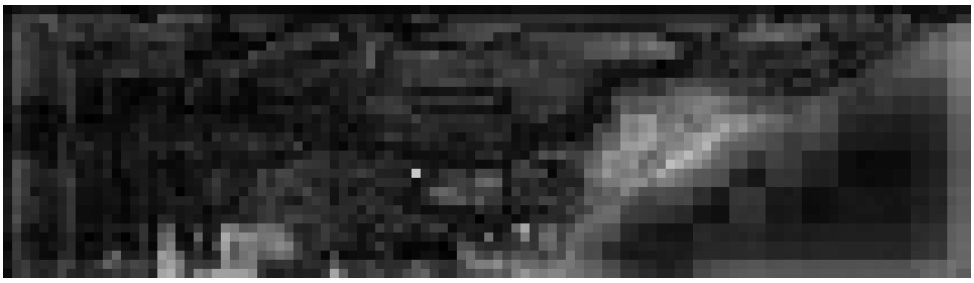


Figura 18: Mapa de saliência da imagem da Figura 14(b).

Dessa forma, pode-se definir o mapa de saliência $\bar{S}$ proposto neste trabalho como:

$$\bar{S} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c,s))] \tag{3.3}$$

3.2.2 Limiarização do Mapa de Saliência

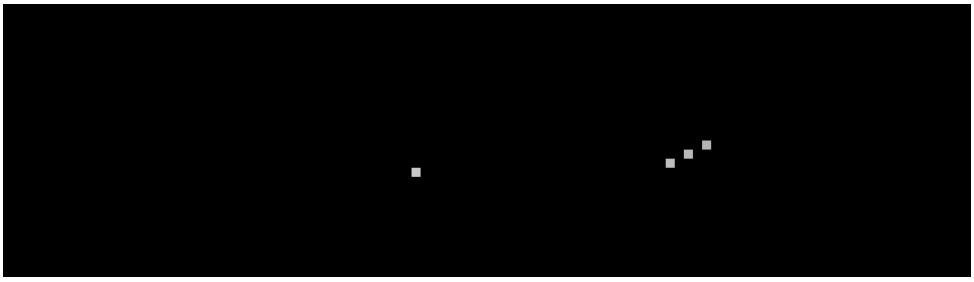
O passo seguinte é a limiarização do mapa de saliência, com o fim de destacar o ponto mais saliente na cena. Para tanto se faz necessário definir um limiar, que foi denominado θ_{sal}. O mapa de saliência é gerado em tons de cinza, sendo que as regiões mais claras do mapa são os pontos mais salientes. O θ_{sal} varia de 0 a 1, sendo que 0 representa a ausência de limiar e o mapa não sofre alteração alguma, e 1 representa o limiar mais alto, ocasionando na total perda de informação da imagem. Sendo o θ_{sal} um valor não-inteiro, é necessário multiplicá-lo pelo valor 255, a fim de obter o valor exato

do limiar em valor de pixel, que varia de 0-255. A limiarização utilizada pode ser definida por:

$$lim(x,y) = \begin{cases} \text{f(x,y)} & f(x,y) \geq \theta_{sal} * 255 \\ 0 & f(x,y) < \theta_{sal} * 255 \end{cases} \tag{3.4}$$

Na Figura 19 pode ser observado o mapa de saliência referente à imagem de entrada apresentada na Figura 18, sendo limiarizado com diferentes valores de θ_{sal}.

(a) $\theta_{sal} = 0{,}50$



(b) $\theta_{sal} = 0{,}70$

(c) $\theta_{sal} = 0{,}78$

Figura 19: Limiarização do mapa de saliência da Figura 18 com diferentes valores para θ_{sal}.

3.3 Reconhecimento com Histogramas

Considerando o mapa de saliência limiarizado, é possível aplicá-lo à imagem original pré-processada, de forma que os valores dos *pixels* do ponto saliente são realçados enquanto os valores dos *pixels* do restante da cena são inibidos. Isso é feito por multiplicar *pixel* a *pixel* o mapa de saliência limiarizado pela imagem pré-processada original. Como

exemplo deste processo, na Figura 20 é apresentado o resultado da multiplicação do mapa de saliência limiarizado da Figura 19(c) à imagem original processada na Figura 14(b).

Figura 20: Resultado da aplicação do mapa de saliência limiarizado (Figura 19(c)) à imagem original processada (Figura 14(b))

Por fim, calcula-se o histograma da imagem obtida após a multiplicação, apresentada na Figura 20. Considerando que o mapa de saliência destaca uma área verde ou vermelha, é preciso identificar a cor predominante da área. Para tanto, o somatório de verde do histograma (Equação 3.5) é calculado, bem como o somatório de vermelho do histograma (Equação 3.6). Os valores de $\sigma_{vermelho}$ e σ_{verde} são então definidos como segue:

$$\sigma_{verde} = \sum_{i=0}^{255} h_G(r_k) \tag{3.5}$$

$$\sigma_{vermelho} = \sum_{i=0}^{255} h_R(r_k) \tag{3.6}$$

É importante notar que, o tom de verde e vermelho do semáforo é particularmente forte, o que permite supor que, entre os pontos mais salientes da cena, sempre estará a luz do semáforo, e este será o ponto de maior atenção. Notado isso, o cálculo de σ_{verde} e $\sigma_{vermelho}$ do histograma da imagem pré-processada, ponderada pelo mapa limiarizado, permite classificar a cena com a presença de um semáforo sinalizando 'verde' ou 'vermelho', a depender do maior valor encontrado ao comparar σ_{verde} e $\sigma_{vermelho}$.

A Figura 21 apresenta os valores de σ_{verde} e $\sigma_{vermelho}$ encontrados ao calcular o histograma da aplicação do mapa de saliência limiarizado da Figura 19(c) à imagem original pré-processada da Figura 14(b). Observa-se que o valor de $\sigma_{vermelho}$ foi de 9832, bastante superior ao de σ_{verde}, que foi de 7944, resultando assim em uma classificação correta da cena como possuindo um semáforo vermelho.

O Algoritmo 1 apresenta o resumo do fluxo do modelo proposto e apresentado

(a) σ_{verde} (b) $\sigma_{vermelho}$

Figura 21: Resultado do somatório do histograma resultante da aplicação do mapa de saliência limiarizado da Figura 19(c) à imagem original pré-processada da Figura 14(b).

neste capítulo.

Algoritmo 1: Fluxo do modelo de detecção e reconhecimento proposto.

```
Obter imagem de entrada;
Pré-processar imagem de entrada;
Criar mapa de saliência;
Limiarizar mapa de saliência;
Calcular o histograma ponderado da imagem de entrada pré-processada;
Calcular σ_vermelho e σ_verde;
if σ_vermelho > σ_verde then
    sinal vermelho;
else
    if σ_vermelho < σ_verde then
        sinal verde;
    else
        impossível classificar;
    end
end
```

No próximo capítulo serão apresentados experimentos que validam o mecanismo proposto apresentado neste capítulo.

4 EXPERIMENTOS

Com o objetivo de testar e validar o mecanismo proposto uma aplicação foi construída. Esta aplicação possui como entrada um conjunto de imagens que são processadas e classificadas, exibindo como resultado se na cena possui um semáforo verde ou vermelho.

Para implementar a aplicação foi utilizada a linguagem Java, que é uma linguagem de programação e plataforma computacional lançada pela primeira vez pela Sun Microsystems em 1995 (JAVA, 2014). A linguagem Java segue o paradigma orientado a objetos, ou seja, de acordo com Javafree (2014) um sistema construído em Java é composto por um conjunto de classes e objetos bem definidos que interagem entre si, de modo a gerar o resultado esperado.

A linguagem foi escolhida por possuir vasta documentação disponível na internet, IDEs poderosas e gratuitas, bem como diversas bibliotecas que podem ser necessárias para o funcionamento da aplicação.

A seguir serão apresentados uma série de experimentos utilizando essa aplicação. Para todos os experimentos foram utilizados os parâmetros $\theta_{cut} = 0{,}40$ e $\theta_{sal} = 0{,}30$. Estes parâmetros foram escolhidos com base em conjuntos de testes e se mostraram os mais adequados para o conjunto de experimento apresentados neste trabalho.

As imagens de entrada foram obtidas de vídeos gravados durante o dia e durante a noite, da câmera traseira de um celular posicionado dentro de um carro. A câmera utilizada possui resolução de 8 megapixeis, e os vídeos foram gravados na seguinte dimensão: 1920 pixeis de largura por 1080 pixeis de altura. Os vídeos foram redimensionados para 640 pixeis de largura por 480 pixeis de largura e foram extraídos quadros dos vídeos em intervalos de 1 segundo.

Os experimentos foram executados num notebook de processador Intel i3, primeira geração, com 4 gigabytes de memória RAM, e foram realizados considerando-se apenas semáforos verticais, encontrados na região. O custo computacional foi de 300-400 milissegundos para processar cada imagem, considerando os parâmetros utilizados.

O primeiro experimento, constitui-se de um conjunto de 28 quadros, dos quais 18 possuem o semáforo vermelho, e 10 possuem o semáforo verde. Foi realizado em ambiente arborizado, que pode confundir o modelo a depender dos parâmetros utilizados. Na Figura 22 é possível observar um trecho detalhado do experimento, onde estão exibidas diversas linhas, que apresentam, nesta ordem, as imagens de entrada, os mapas de saliência, os mapas de saliência limiarizados, e por fim as classificações encontradas ao calcular os

histogramas ponderados. A mesma organização é utilizada nas outras figuras de trechos dos experimentos realizados.

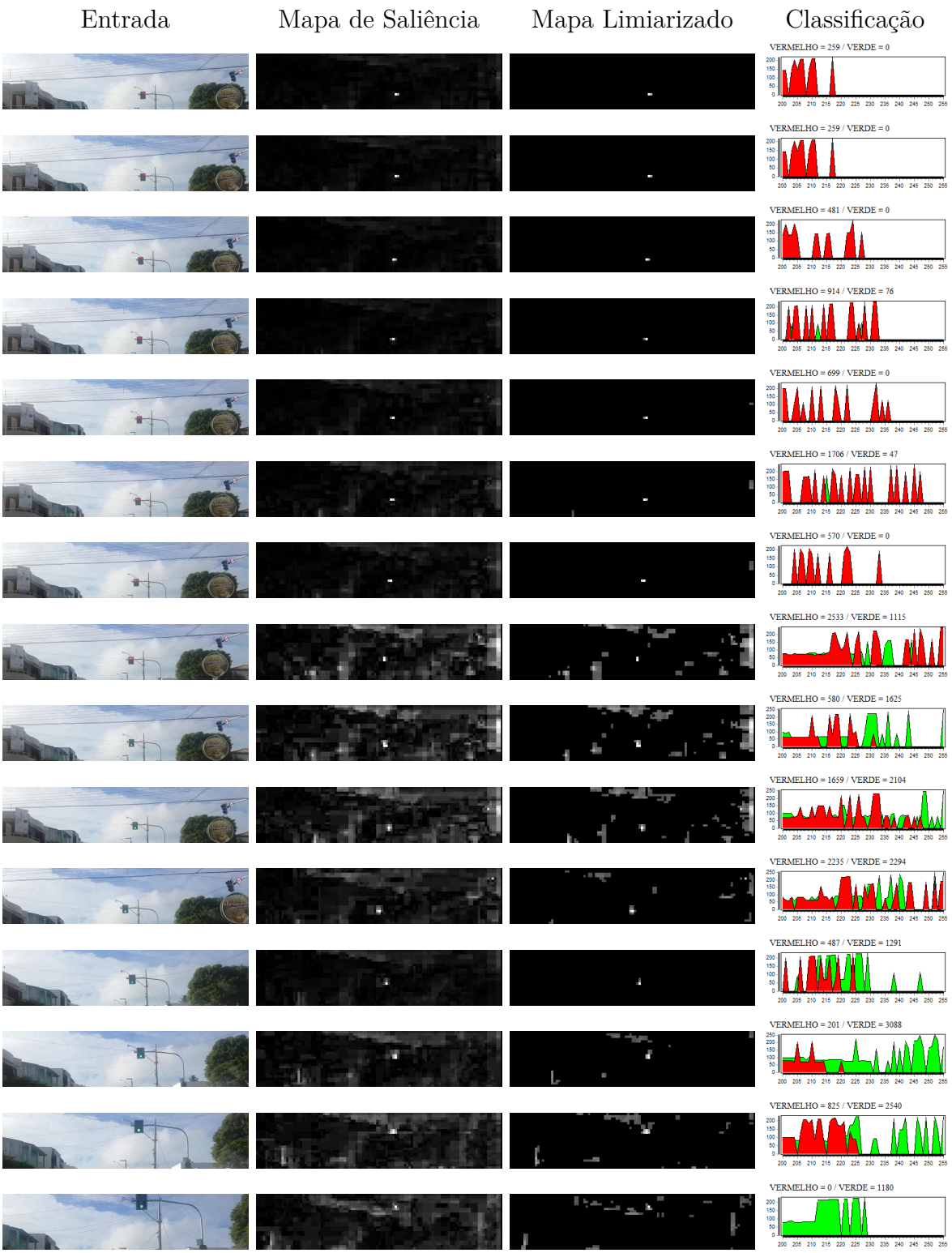


Figura 22: Experimento 1 - trecho de 15 imagens das 28 imagens obtidas dia 01/12/2014, em Ribeirópolis - SE, às 6 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

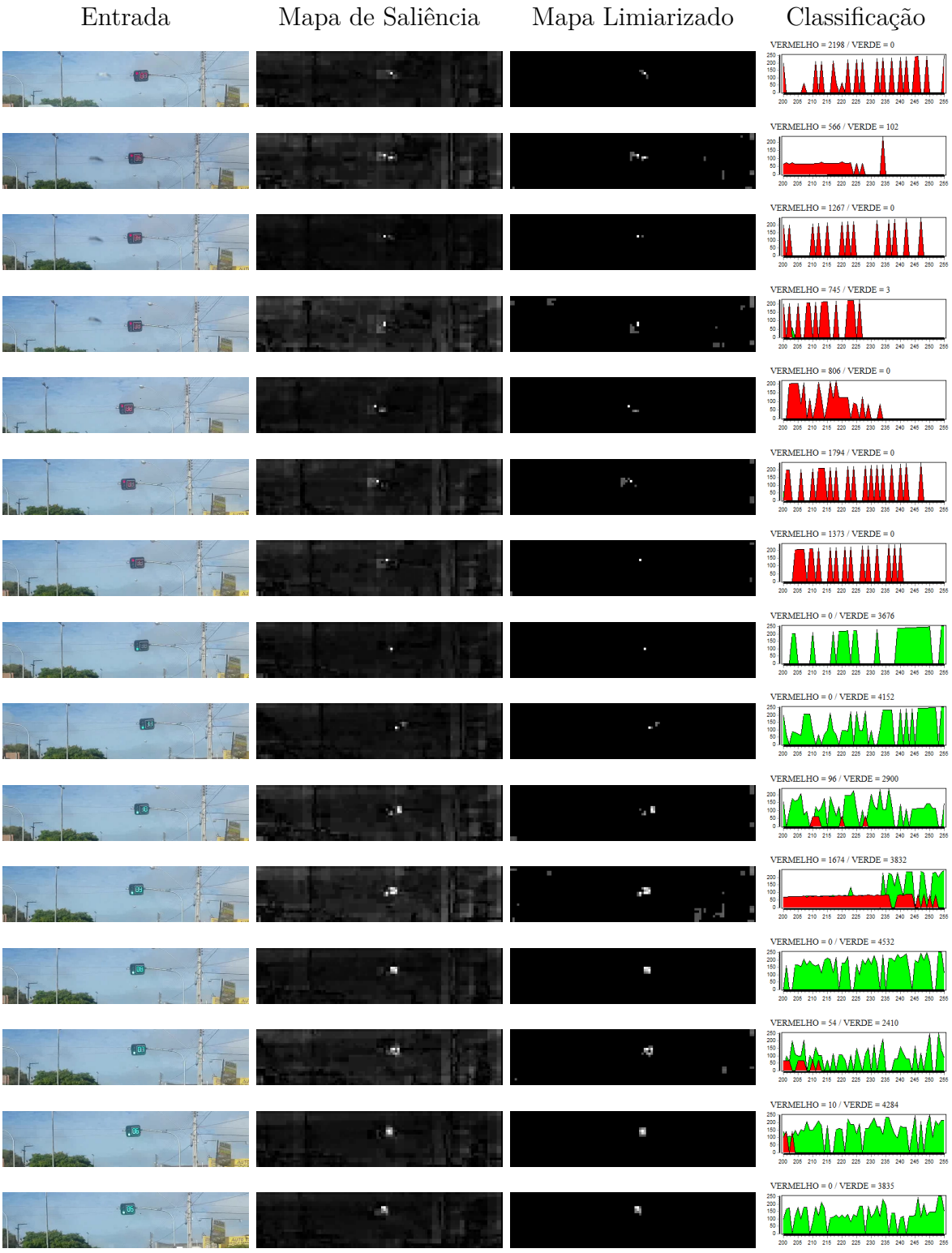


Figura 23: Experimento 2 - trecho de 15 imagens das 51 imagens obtidas dia 01/12/2014, em Ribeirópolis - SE, às 6 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

O segundo experimento, constitui-se de um conjunto de 51 quadros, dos quais 37 possuem o semáforo vermelho, e 14 possuem o semáforo verde. Na Figura 23 é possível observar um trecho detalhado do experimento. O experimento foi realizado em ambiente

arborizado, sem luz forte, com semáforos grandes e visivelmente destacados. Embora a presença de árvores seja grande, a maior parte some ao aplicar o valor de θ_{cut}

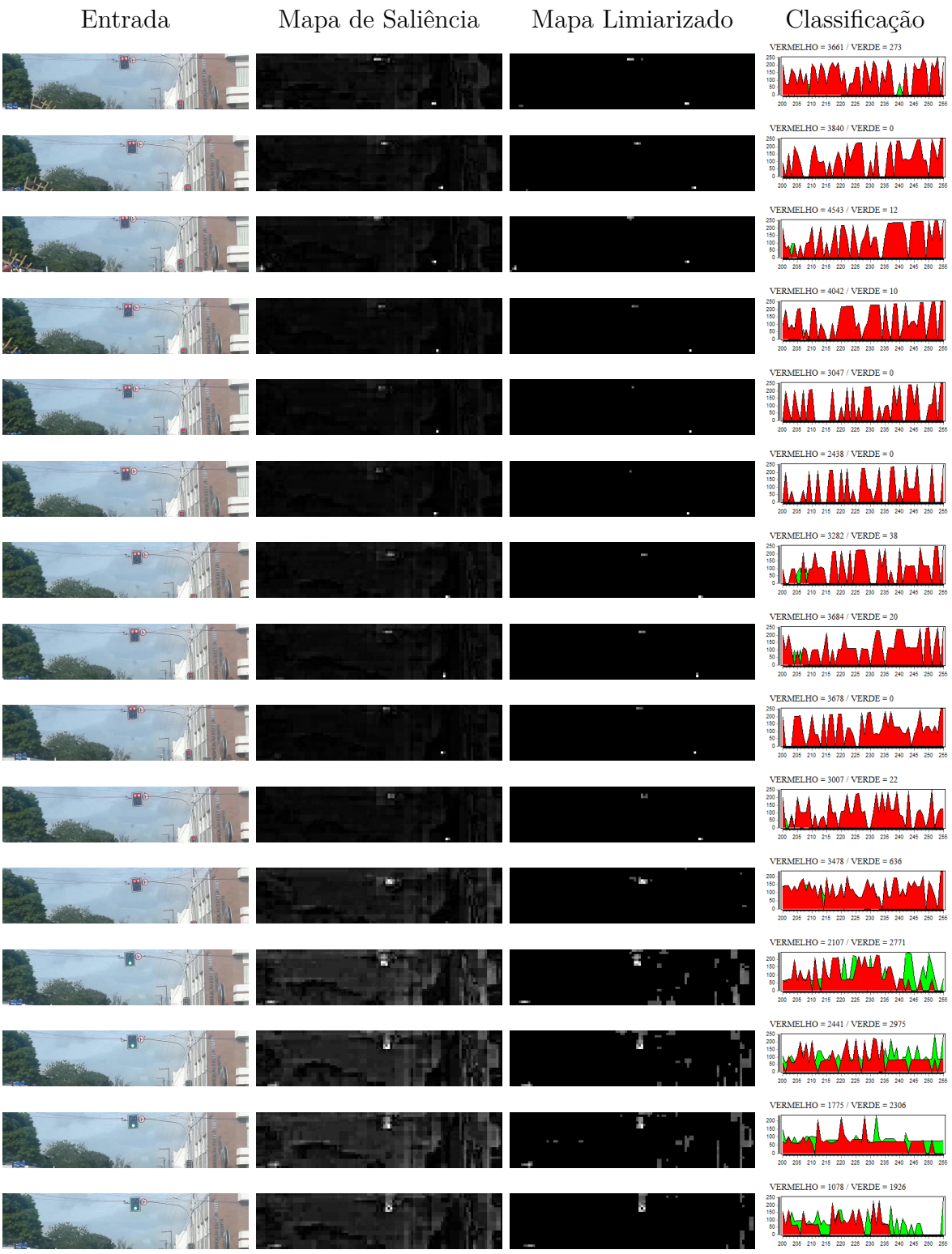


Figura 24: Experimento 3 - trecho de 15 imagens das 35 imagens obtidas dia 01/12/2014, em Ribeirópolis - SE, às 6 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

O terceiro experimento, constitui-se de um conjunto de 35 quadros, dos quais 25 possuem o semáforo vermelho, e 10 possuem o semáforo verde. Na Figura 24 é possível observar um trecho detalhado do experimento. A maior parte das cenas possui dois semáforos, um superior e um lateral, sendo que o superior apresenta duas luzes vermelhas. Interessante notar que apesar da presença de uma placa vermelha ao lado do semáforo superior, este sempre se destaca, validando assim o bom comportamento do modelo.

O quarto experimento, constitui-se de um conjunto de 35 quadros, dos quais 24 possuem o semáforo vermelho, e 11 possuem o semáforo verde. Na Figura 25 é possível observar um trecho detalhado do experimento. Este experimento apresenta incidência de luz forte na cena, o que dificulta a detecção e posterior classificação. Apesar disto, observa-se que o modelo obteve um bom comportamento com os parâmetros utilizados.

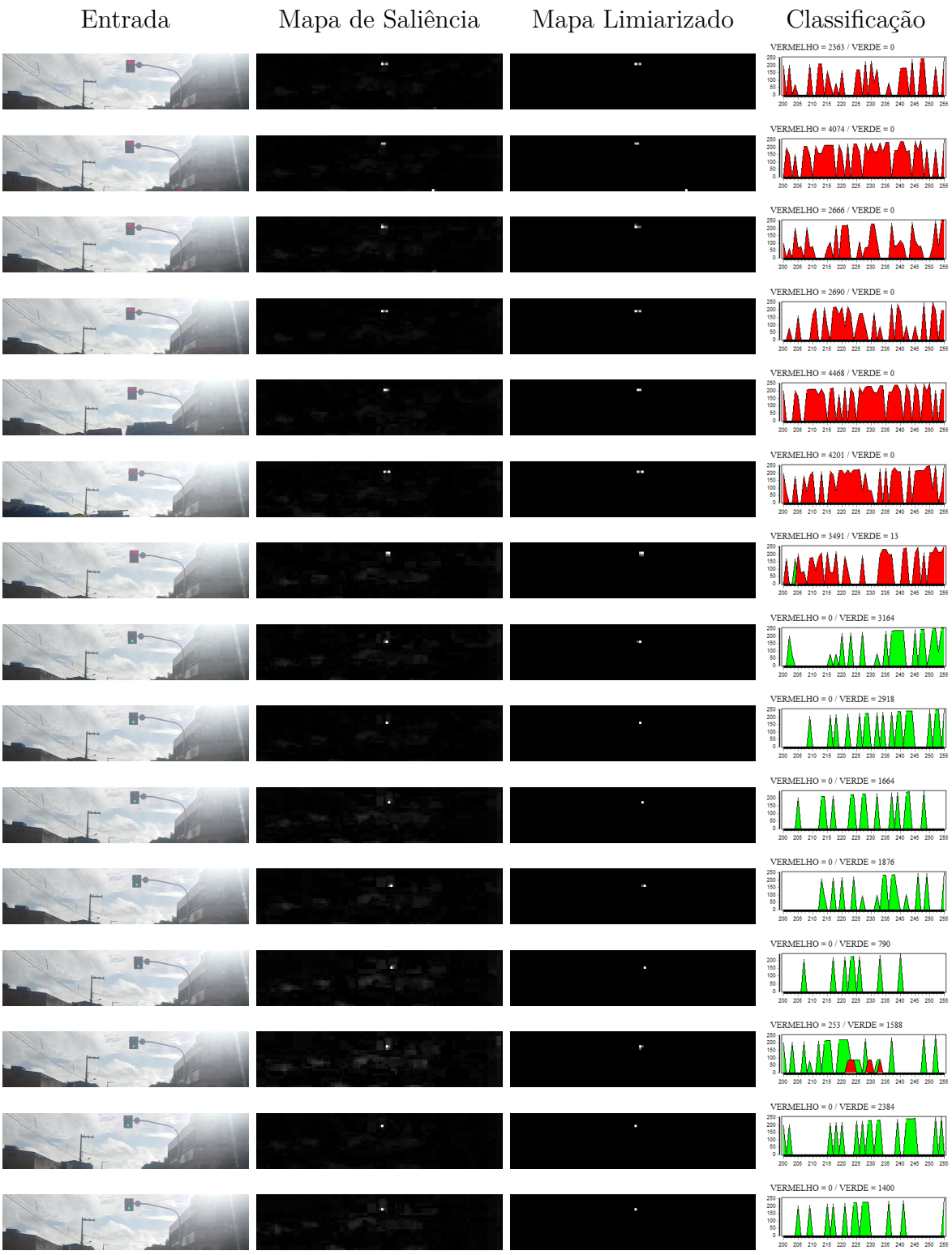


Figura 25: Experimento 4 - trecho de 15 imagens das 35 imagens obtidas dia 01/12/2014, em Ribeirópolis - SE, às 6 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

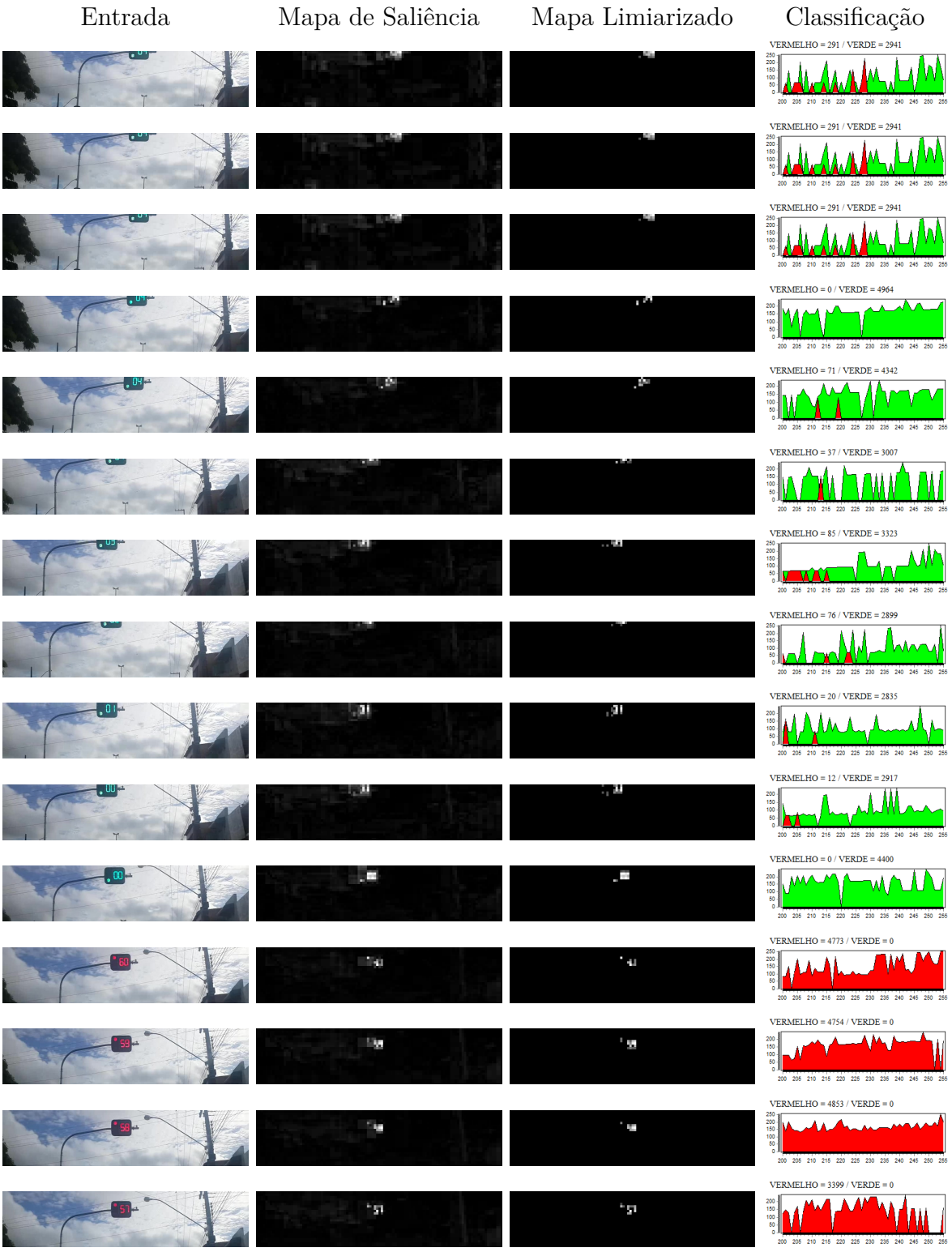


Figura 26: Experimento 5 - trecho de 15 imagens das 28 imagens obtidas dia 17/12/2014, em Ribeirópolis - SE, às 15 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

O quinto experimento, constitui-se de um conjunto de 16 quadros, dos quais 4 possuem o semáforo vermelho, e 12 possuem o semáforo verde. O experimento apresenta grande variação de posição do semáforo entre os quadros, e apesar disto, o modelo detecta

corretamente a saliência.

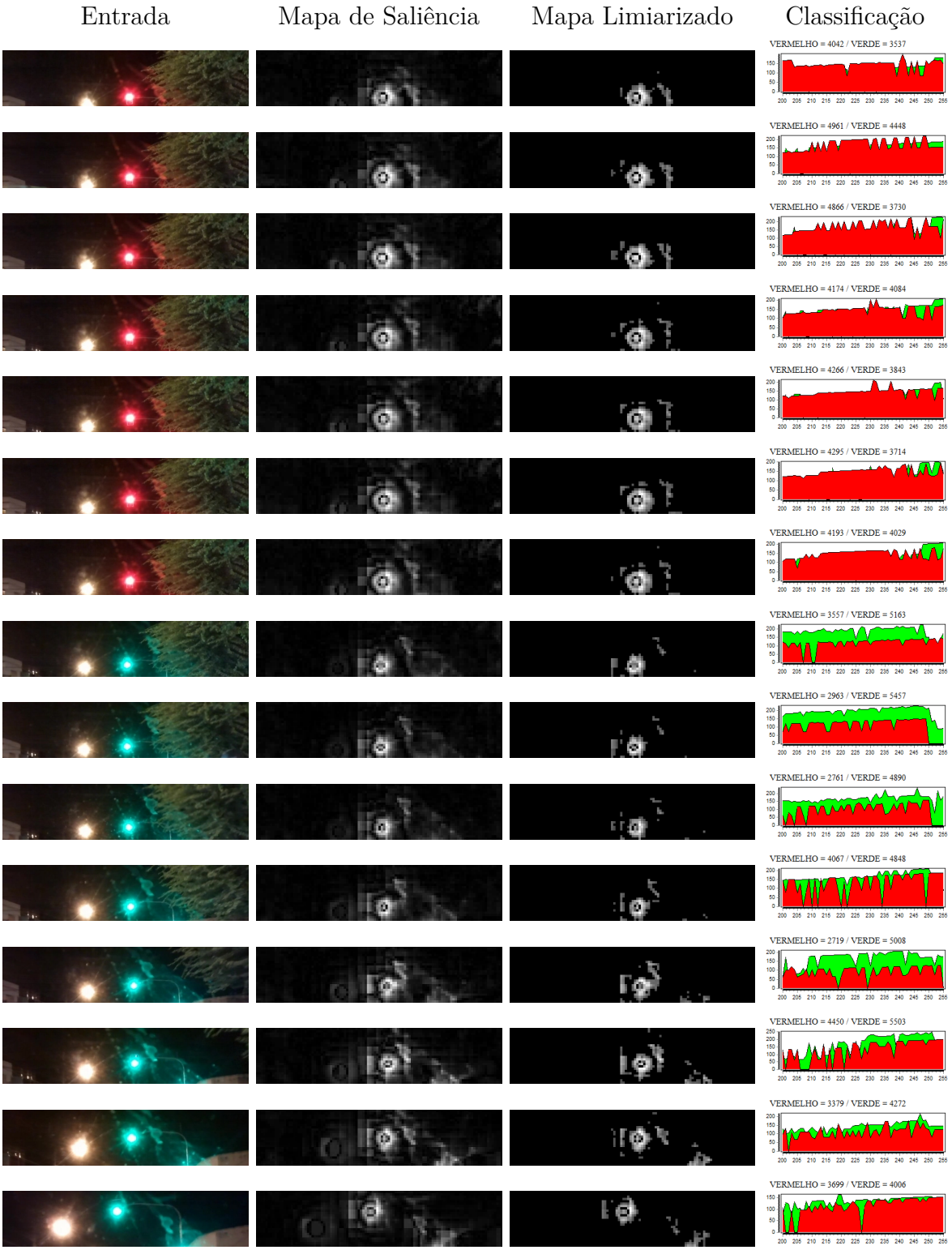


Figura 27: Experimento 6 - trecho de 15 imagens das 57 imagens obtidas dia 22/12/2014, em Ribeirópolis - SE, às 21 horas. Utilizou-se $\theta_{cut} = 0,40$ e $\theta_{sal} = 0,30$

O sexto experimento, realizado à noite, constitui-se de um conjunto de 57 quadros, dos quais 47 possuem o semáforo vermelho, e 10 possuem o semáforo verde. Na Figura 27

é possível observar um trecho detalhado do experimento. À noite, o modelo comportou-se particularmente bem, uma vez que os semáforos tornam-se mais luminosos devido a ausência de luz solar. Considerou-se um ambiente livre de outras luzes vermelhas ou verdes, como letreiros por exemplo.

Os experimentos apresentados possuem um total de 194 quadros, sendo 137 obtidos pelo dia e 57 pela noite, todos realizados na cidade de Ribeirópolis - SE. O modelo comportou-se bem em todos os ambientes detalhados anteriormente, sendo estes: com árvores e objetos vermelhos, como placas, presentes na cena; com incidência de luz solar forte; à noite; e com grande variação de posição entre os quadros. Apesar do bom comportamento em diversos cenários, o modelo comportou-se de forma instável em alguns casos, apresentados na Figura 28.

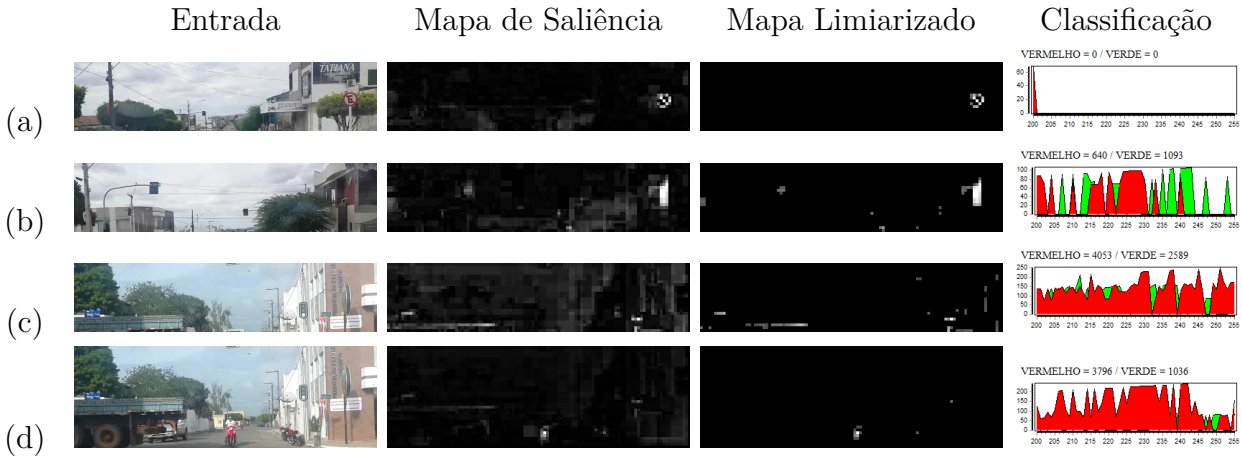


Figura 28: Casos em que o modelo não se comportou bem. Todas imagens obtidas no período diurno. (a), (b) e (c) utilizam $\theta_{cut} = 0,40$, e (d) utiliza $\theta_{cut} = 0,60$ para demonstrar como obter informação irrelevante pode ser prejudicial. As imagens foram obtidas dias 01 e 17 de dezembro de 2014, às 6 horas e às 15 horas respectivamnte

No primeiro caso (Figura 28(a)), é possível notar que o semáforo não foi encontrado na cena, a única saliência encontrada refere-se à placa de trânsito que apresenta uma quantidade considerável de vermelho. Isso acontece pois, em decorrer da distância em que o semáforo se encontra, sua quantidade de vermelho/verde pode não ser suficiente para que se destaque na cena.

Nos segundo e terceiro casos (Figura 28(b) e (c)), embora o semáforo seja detectado como uma saliência mais fraca, ao contrário do caso anterior, seu valor de saliência é baixo, e não é suficiente para garantir a classificação do semáforo como verde. Dessa forma um outro ponto mais saliente qualquer define a classificação.

O quarto caso (Figura 28(d)), apresenta um erro causado por informação irrelevante na cena que poderia ser eliminada por diminuir o θ_{cut} utilizado. Neste caso específico,

com o fim de exemplificar, foi utilizado $\theta_{cut} = 0{,}60$. Observa-se que o ponto mais saliente na cena foi o veículo vermelho, que não apareceria na imagem caso o θ_{cut} fosse menor, permitindo a correta classificação, uma vez que observando o mapa de saliência percebe-se que o semáforo está representado pelo segundo ponto mais saliente da cena.

Uma maneira de resolver os problemas apresentado nos três últimos casos seria utilizar um mecanismo de reconhecimento, dessa forma seria possível excluir o falso positivo encontrado primariamente e direcionar a atenção para o próximo ponto saliente, reconhecendo-o em seguida.

Alguns dos mecanismos de reconhecimento que poderiam ser utilizados são com algoritmos SVM ou redes neurais artificiais, e o reconhecimento poderia se basear tanto informações do histograma de cores como em outras características do semáforo e do mapa de saliências, como por exemplo o tamanho da saliência encontrada, sendo que, quanto mais informações se utiliza em conjunto para o reconhecimento, mais confiável este se torna.

5 CONCLUSÃO

O modelo proposto obteve um funcionamento satisfatório ao descobrir o sinal do semáforo em ambientes diurnos e noturnos. Embora os experimentos tenham sido realizado apenas com semáforos verticais, entende-se que o mecanismo apresentaria comportamento semelhante em outros tipos de semáforo, uma vez que o mecanismo realiza o reconhecimento com base no histograma da área mais saliente, e não há relação com a forma ou outras características do objeto semáforo. Com o fim de provar isso, futuramente serão realizados experimentos com outros tipos de semáforo.

Importante enfatizar o desempenho do modelo, que implementado obteve velocidade de processamento de 300-400 milissegundos por quadro, utilizando os parâmetros definidos. Isso mostra a viabilidade de uso do modelo em tempo real. Experimentos em tempo real serão realizados futuramente.

O modelo proposto implementado precisa de algumas melhorias para que possa ser utilizado, entre elas diminuir os casos em que ocorre erro na classificação. A falha na detecção e posterior reconhecimento em alguns casos deu-se, em sua maioria, por haver um outro objeto que, por alguma variação do ambiente recebeu o maior foco de atenção.

Dessa forma, um objetivo a ser alcançado em trabalhos futuros é refinar o reconhecimento, utilizando mecanismo que escolha entre as áreas mais salientes da cena, e não apenas a mais saliente de todas. Esse reconhecimento pode ser feito com base em algoritmos de inteligência artificial, utilizando informações do semáforo ou do mapa de saliência para classificação.

Outra informação importante a ser definida é a distância mínima necessária para obter uma boa detecção e reconhecimento, constitui então um trabalho futuro obter essa informação, bem como calcular a distância aproximada em que o semáforo se encontra.

Com o objetivo de situar melhor o modelo proposto na academia, é necessário realizar comparações com outros trabalhos que não utilizam atenção visual, para dessa forma mostrar porque a abordagem é interessante de ser utilizada. Este ponto constitui um dos trabalhos futuros.

Referências

Almeida, A. B. *Usando o Computador para Processamento de Imagens Médicas*. 1998. Acesso em: 15 fev. 2015. Disponível em: <http://www.informaticamedica-.org.br/informaticamedica/n0106/imagens.htm>. Citado 2 vezes nas páginas 7 e 17.

Australian government - Department of Infraestructure and Regional Development. *How are Intelligent Transport Systems being used?* 2015. Acesso em: 15 fev. 2015. Disponível em: <http://www.infrastructure.gov.au/transport/its/its_use.aspx>. Citado na página 27.

BENICASA, A. Sistemas computacionais para atenção visual top-down e bottom-up usando redes neurais artificiais. *USP São Carlos*, 2013. 2013. Citado 9 vezes nas páginas 7, 11, 19, 21, 22, 23, 24, 25 e 26.

CAROTA, L.; INDIVERI, G.; DANTE, V. A softwarehardware selective attention system. 2004. Neurocomputing, 647653, 2004. Citado 2 vezes nas páginas 11 e 19.

CHARETTE R., N. F. Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. *IEEE Intelligent Vehicles Symposium*, 2009. p. 358–363, 2009. Citado 2 vezes nas páginas 27 e 28.

CHUNG Y., W. J. C. S. A vision-based traffic light detection system at intersections. *Journal of National Taiwan Normal University: Mathematics, Science Technology*, 2002. v. 47, p. 67–86, 2002. Citado na página 27.

Diaz-Cabrera, M., Cerri, P., Medina-Sanchez, J. *Suspended Traffic Lights Detection and Distance Estimation Using Color Features*. 2012. Acesso em: 17 set. 2014. Disponível em: <http://www.ce.unipr.it/people/bertozzi/pap/cr/itsc2012.semafori.pdf>. Citado na página 12.

European Comission of Mobility and Transport. *Inteligent Transportation Systems*. 2015. Acesso em: 15 fev. 2015. Disponível em: <http://ec.europa.eu/transport/themes/its-/index_en.htm>. Citado na página 27.

FACON, J. Processamento e analise de imagens. 2002. PUCP, 2002. Citado 3 vezes nas páginas 14, 15 e 16.

GONG J., J. Y. X. G. G. C. T. G. C. H. The recognition and tracking of traffic lights based on color segmentation and camshift for intelligent vehicles. *IEEE Intelligent Vehicles Symposium, Universidade da Califórnia*, 2010. 2010. Citado na página 28.

GONZALEZ R. C. E WOODS, R. E. *Processamento Digital de Imagens - 3ª ed.* [S.l.]: São Paulo: Pearson, 2010. Citado 7 vezes nas páginas 7, 13, 14, 15, 16, 17 e 18.

GRANDO, N. Segmentação de imagens tomográficas visando a construção de modelos médicos. 2005. Dissertação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. CEFET-PR Centro Federal de Educação Tecnológica do Paraná, 2005. Citado 4 vezes nas páginas 13, 15, 16 e 17.

ITTI, L. Models of bottom-up and top-down visual attention. 2000. California Institute of Technology, 2000. Citado 3 vezes nas páginas 7, 11 e 23.

ITTI, L. Models of bottom-up attention and saliency. 2005. Neurobiology of Attention, Chapter 94. Elsevier, Oxford, 2005. Citado 3 vezes nas páginas 11, 19 e 30.

ITTI, L.; KOCH, C. Computational modelling of visual attention. 2001. Nature Reviews Neuroscience 2, 194203, 2001. Citado 3 vezes nas páginas 20, 24 e 25.

ITTI L., K. C. N. E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11)*, 1998. p. 1254–1259, 1998. Citado 9 vezes nas páginas 7, 11, 20, 21, 23, 25, 33, 34 e 35.

Jacob, H. *Desenvolvimento de um Modelo de Atenção Visual para Sumarização de Vídeos de Programas Televisivos*. 2013. Acesso em: 17 set. 2014. Disponível em: <http://www-.files.scire.net.br/atrio/cefet-mg-ppgmmc_upl/THESIS/174/hugodrummondjacob.pdf>. Citado na página 26.

JAVA, . Java. disponível em: $http : //www.java.com/pt_br/download/faq/whatis_java.xml/..$ 2014. Acesso em setembro de 2014, 2014. Citado na página 39.

JAVAFREE, . Javafree. disponível em: $http : //javafree.uol.com.br/artigo/871497/tutorial - java - 3 - orientacao - a - objetos.html\#ixzz3dtfalg4f..$ 2014. Acesso em setembro de 2014, 2014. Citado na página 39.

KIM H., S. Y. K. S. P. J. J. H. Night-time traffic iht detection based on svm with geometric moment features. *World Academy of Science, Engineering and Technology*, 2013. v. 7, 2013. Citado na página 28.

LaRA. *LaRA - La Route Automatisée*. 2015. Acesso em: 15 fev. 2015. Disponível em: <http://www.infrastructure.gov.au/transport/its/its_use.aspx>. Citado na página 27.

MARQUES O. F. E VIEIRA, H. N. *Processamento Digital de Imagens*. [S.l.]: Rio de Janeiro: Brasport, 1999. Citado na página 17.

Melo, N. *Abordagens do processo de Segmentação: Limiarização, Orientada a Regiões e Baseada em Bordas*. 2009. Acesso em: 15 fev. 2015. Disponível em: <http://www.dsc-.ufcg.edu.br/~pet/jornal/setembro2011/materias/recapitulando.html>. Citado 2 vezes nas páginas 7 e 18.

MORALES, D.; CENTENO, T. M.; MORALES, R. C. Extração automática de marcadores anatômicos no desenvolvimento de um sistema de auxílio ao diagnóstico postural por imagens. 2003. III Workshop de Informática aplicada à Saúde CBComp, 2003. Citado na página 13.

MORGAN, J. Técnicas de segmentação de imagens na geração de programas para máquinas de comando numérico. 2008. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Santa Maria, 2008. Citado 2 vezes nas páginas 13 e 16.

Pais, G., Munhoz, V., Policastro, C. *Descrição da arquitetura multimodal para controle de estímulos em robôs sociais*. 2009. Acesso em: 17 set. 2014. Disponível em: <http:/-/www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_342.pdf>. Citado na página 26.

PEREIRA, E. T. Atenção visual bottom-up guiada por otimização via algoritmos genéticos. 2007. Campina Grande, 2007. Citado 2 vezes nas páginas 19 e 24.

Poffo, F. *Visual Autonomy Protótipo para Reconhecimento de Placas de Trânsito*. 2010. Acesso em: 17 set. 2014. Disponível em: <http://www.bc.furb.br/docs/mo/2010-/344092_1_1.pdf>. Citado na página 26.

Rodrigues, F. *Localização e Reconhecimento de Placas de Sinalização Utilizando um Mecanismo de Atenção Visual e Redes Neurais Artificiais*. 2002. Acesso em: 17 set. 2014. Disponível em: <http://docs.computacao.ufcg.edu.br/posgraduacao-/dissertacoes/2002/Dissertacao_FabricioAugustoRodrigues.pdf>. Citado na página 26.

Santana, P. *Visual attention and swarm cognition for off-road robots*. 2011. Acesso em: 17 set. 2014. Disponível em: <http://hdl.handle.net/10451/4228>. Citado na página 26.

SHIC, F.; SCASSELLATI, B. A behavioral analysis of computational models of visual attention. *International Journal of Computer Vision*, 2007. 2007. Citado 2 vezes nas páginas 11 e 19.

U.S. Department of Transportation (US DOT). *Inteligent Transportation Systems Joint Program Office*. 2015. Acesso em: 15 fev. 2015. Disponível em: <http://www.its.dot.gov-/>. Citado na página 27.

WALTHER, D. Interactions of visual attention and object recognition: Computation modeling, algorithms, and psychophysics. *California Institute of Technology, Pasadena, California*, 2006. 2006. Citado na página 11.

WOLFE, J. M.; HOROWITZ, T. S. What attributes guide the deployment of visual attention and how do they do it? 2004. Nature Review Neuroscience 5, 495501, 2004. Citado 3 vezes nas páginas 7, 19 e 20.

YANG X., Z. Z. K. Y. A real time traffic light recognition system. *WSPC-IJIA*, 2008. 2008. Citado na página 27.