



Universidade de São Paulo

Biblioteca Digital da Produção Intelectual - BDPI

Departamento de Ciências de Computação - ICMC/SCC

Comunicações em Eventos - ICMC/SCC

2015-10

A visual attention approach for the tracking of vehicles through UAV

Workshop de Visão Computacional, XI, 2015, São Carlos.

<http://www.producao.usp.br/handle/BDPI/49459>

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo



Laboratório de Visão Computacional



XI Workshop de Visão Computacional
October 05th - 07th 2015 - São Carlos (SP) - Brazil



**XI Workshop de Visão Computacional
WVC'2015**

October 05th – 07th, 2015

São Carlos – SP - Brazil

Proceedings

**University of São Paulo
São Carlos School of Engineering**

A visual attention approach for the tracking of vehicles through UAV

Raphael Montanari, Daniel C. Tozadore, Eduardo S. Fraccaroli and Roseli A. F. Romero

Universidade de São Paulo

Email: {rmontana, eduardo, rafrance}@icmc.usp.br, tozadore@usp.br

Alcides X. Benicasa

Universidade Federal de Sergipe

Email: alcides@ufs.br

Resumo—Rotary-wing unmanned aerial vehicles (UAV) are being widely used in different applications due to several features, such as, mobility, lightweight, embedded processing, flying in different height levels. Among the possible applications, they are being used in surveillance tasks, agriculture environments monitoring, power lines inspections and detection of diseases. In all applications, it is necessary to analyse the images captured by the UAV. This analysis is usually done by experts, but a system that process the captured images in a autonomous way can speed up the results and more images could be processed in a lesser period of time, which is the ideal for real time applications. This work presents a system for vehicle detection aiming the surveillance task. This system is evaluated in terms of precision of detection of a particular vehicle. The image data has been collected by using a monocular camera and a quadrotor developed in our Lab. The experimental results demonstrate that the proposed system can be used to the task of tracking of cars under controlled situations.

I. INTRODUÇÃO

A abordagem robusta para a detecção de veículos em tempo real a partir de um veículo aéreo não tripulado (VANT) é uma meta importante no âmbito da implantação de um VANT totalmente autônomo para reconhecimento e vigilância aérea, sobretudo em missões de resgate ou de vigilância. Este tipo de sistema tem crescente popularidade em aplicações civis e militares, devido à sua capacidade de operar em terreno amplo, de difícil acesso e não estruturado. Em missões de busca e salvamento em áreas remotas, onde o acesso é dificultado por montanhas, vastas áreas de terra sem redes de estradas ou em áreas atingidas por desastres naturais (terremoto, inundação) pode ser auxiliada e melhorado por sistemas VANT autônomos. Em tais situações, informando o local dos objetos detectados de interesse (por exemplo, veículos, pessoas) podem eliminar a necessidade de análise de imagem manual das imagens de vídeo. A análise das imagens obtidas a partir da plataforma aérea coloca muitos desafios devido ao movimento da plataforma, a instabilidade da imagem e do tamanho relativamente pequeno do objeto de interesse. Dependendo da orientação, altitude de vôo da câmera, e da aparência dos objetos dentro do ambiente observado, a detecção automática dos objetos muda drasticamente, tornando a detecção dos objetos uma tarefa desafiadora. Nesse trabalho utilizamos imagens capturadas por uma câmera instalada na base inferior de um quadrotor para identificar veículos. Os principais desafios do

sistema consistem em lidar com a orientação da imagem 3D, imagens borradas devido à vibração e variações nas condições de iluminação e ambientes dinâmicos.

Existe uma vasta literatura sobre detecção de veículos a partir de imagens aéreas. O trabalho de Zhao e Nevatia [14] explorou um método de reconhecimento de veículos a partir de imagens aéreas de baixa resolução. Gleason [5] discute um sistema de detecção de veículos que tenta corresponder veículos contra um modelo em malha de triângulos 3D de forma adaptativa "top-down".

Este artigo está organizado da seguinte maneira: Na seção II são apresentados alguns trabalhos relacionados. Na seção III, encontra-se a abordagem utilizada para tratar objetos em imagem. Na seção IV são descritos os hardwares e os dados coletados para avaliar a performance do sistema. Na seção V são apresentados os experimentos realizados e finalmente na seção VI discutidos os resultados obtidos

II. TRABALHOS RELACIONADOS

As técnicas baseadas em visão são uma das abordagens mais comuns para analisar veículos a partir de imagens ou vídeos. A visualização de veículos pode variar de acordo com as posições da câmara fotográfica, das condições de iluminação e do fundo da imagem. Já para detecção de veículos, as soluções existentes baseiam-se numa grande variedade de técnicas - algumas delas brevemente comentadas neste artigo.

No levantamento bibliográfico de Ramakrishnan et al. [10], os autores listam trabalhos na área e comparam os métodos utilizados em cada um deles. Dentre eles destacam-se os trabalhos Hinz e Baumgartner [7] propõem um modelo hierárquico que descrevem diferentes níveis de detalhes em características de veículos e um método de detecção com base em cascata de classificadores. [3] usa um algoritmo de detecção de veículo com base em propriedade de formas simétricas. O algoritmo *mean shift* é geralmente usado, mas apresenta alta complexidade computacional.

Lavigne et al. [11] usam SVM (*Support Vector Machine*) para classificar estes características-chave SIFT (*Scale Invariant Feature Transform*).

Xuelong Li et al. [2] usam classificadores AdaBoost como extensão do Histograma de Gradientes Orientados (HOG) para

melhorar as características HOG e testaram essa estrutura em uma plataforma aérea de baixa altitude.

III. SELEÇÃO BASEADA EM OBJETO

A abordagem proposta para selecionar objetos salientes é composta pelos seguintes módulos: um módulo de extração de características visuais; um modelo enviesado *top-down* baseado em *features*; uma rede LEGION para segmentação de imagens; uma rede MLP para reconhecimento de objetos, uma rede de integra e dispara, que cria o mapa objeto-saliência e, por fim, um módulo de seleção de objeto, que destaca os objetos mais importantes na cena.

A Figura 1 mostra um fluxograma do modelo proposto. A primeira etapa do modelo de atenção visual é responsável por extrair, em paralelo, as características visuais iniciais em toda a cena. Os resultados desta fase são os seguintes mapas de conspicuidade: cores, a intensidade e orientação.

Para uma descrição completa de como os mapas de conspicuidade são computados, consulte [9]. A próxima etapa do modelo é a combinação dos resultados a partir dos mapas de conspicuidade com pesos específicos, para o enviesamento *top-down* da segmentação da rede LEGION, que segue o algoritmo proposto em [12].

A saída desses módulos alimenta os seguintes outros módulos: a rede de reconhecimento de objetos e a rede de integra e dispara, que cria o mapa saliência de objetos. Estes dois últimos módulos e a enviesamento *top-down* da rede LEGION são apresentados nas seções seguintes.

A. Enviesamento *Top-down*

O enviesamento *top-down* proposto neste modelo é definido pela associação de pesos da saída dos mapas de conspicuidade (C_k). Os valores de saliência para todos os mapas de são ponderados e combinados em um único mapa de saliência S_m definido como:

$$S_m = \frac{1}{n^k} \sum_k W_k C_k, k \in \{\text{intensidade, cor, orientação}\} \quad (1)$$

no qual n^k representa o numero de mapas de conspicuidade escolhido do enviesamento e W_k determina o peso dos respectivos mapas C_k .

Segundo [12], o processo de segmentação na rede LEGION é baseado na ideia de que um segmento deve conter pelo menos um oscilador, denotado como um líder, que se encontra no centro de uma grande região homogênea. Líderes são osciladores i com potencial lateral $p_i \geq \theta$ no qual θ é um *threshold* [12]. A fim de gerar o enviesamento *top-down* do modelo proposto, um oscilador i definido como líder, somente irá pulsar se seu valor de saliência $S_{m_i} \geq \theta_{bias}$.

A informação *top-down* sobre um alvo pode ser aprendida automaticamente de acordo com a tarefa, porém neste trabalho os parâmetros W_{col} , W_{int} , W_{ori} e o valor de *threshold* θ_{bias} são escolhidos pelo usuário.

B. Reconhecimento *Top-Down*

O modelo proposto usa ambas as abordagens bottom-up e top-down. Características visuais iniciais, ou seja, contraste intensidade, contraste de cor e contraste de orientação, definem o sinal bottom-up. Por outro lado, informações sobre objetos previamente memorizadas e as suas características (top-down) são responsáveis por orientar o processo de seleção. Assim, a fim de aplicar o modelo proposto para selecionar os objetos salientes de uma determinada cena, a rede MLP deve ser treinada com um conjunto de objetos que representam os alvos desejados da cena. Sendo que para proporcionar a atenção de novos objetos, uma nova fase de treinamento deve ser realizada.

O conjunto de objetos usados para treinar a rede de reconhecimento de objetos MLP é composto de um conjunto de imagens binárias extraídas da cena. A fim de fazer isso, a cena de entrada é apresentada para a rede LEGION que retorna um conjunto de objetos segmentados. Esses objetos são rotulados manualmente e usadas para treinar a rede MLP através do algoritmo de *backpropagation* [6].

Após o processo de treinamento, a rede MLP é capaz de reconhecer um conjunto de segmentos (objetos). Assim, a dinâmica global do sistema pode ser compreendida como segue. Cada vez que um segmento é destacado (pulsção) na rede LEGION, ele é apresentado à rede MLP. A saída da MLP indica se o objeto está entre aqueles memorizados pelo sistema de reconhecimento. Se o objeto é reconhecido, o valor de saída da rede é utilizado para fixar o atributo de reconhecimento $R_{i,j}$, sendo i e j representam a posição dos *pixels* dentro de cada segmento. Inicialmente, $R_{i,j} = 0$ para todos os neurônios. No final deste processo, todos os neurônios relacionados com objetos que devem receber atenção (*top-down*) irá ser atribuído a um valor de reconhecimento ($R_{i,j} = [0, 1]$) que modulará o processo de atenção.

Segmentos que representam objetos desconhecidos também podem apresentar valores diferentes de zero de reconhecimento. Para evitar esses objetos de receberem modulação top-down, um *threshold* para o valor de reconhecimento (θ_r) é adotado. Assim, os segmentos inferiores a este limite não são consideradas. Assim, o valor de reconhecimento $R_{i,j}$ é definido por:

$$R_{i,j} = \begin{cases} 1, & \text{if } R_{i,j} \geq \theta_r \\ 0, & \text{caso contrário} \end{cases}, \quad (2)$$

C. Seleção prévia de objetos salientes

Para cada pixel da cena da entrada, os seguintes descritores foram extraídos: contraste de intensidade (I), diferença espacial em cores RG e RY (C), orientações O_θ com $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ (O), localização de pixel $[i, j]$ (L), e o valor de reconhecimento $R_{i,j}$ (R) estabelecido pelo módulo de reconhecimento MLP.

Seja l um neurônio pertencente a um segmento ativo no modelo de LEGION e k seu respectivo índice das características, denotado como $l_k = [C, I, O, L, R]$. Após concluir o processo de segmentação são calculados os valores de saliência de todos

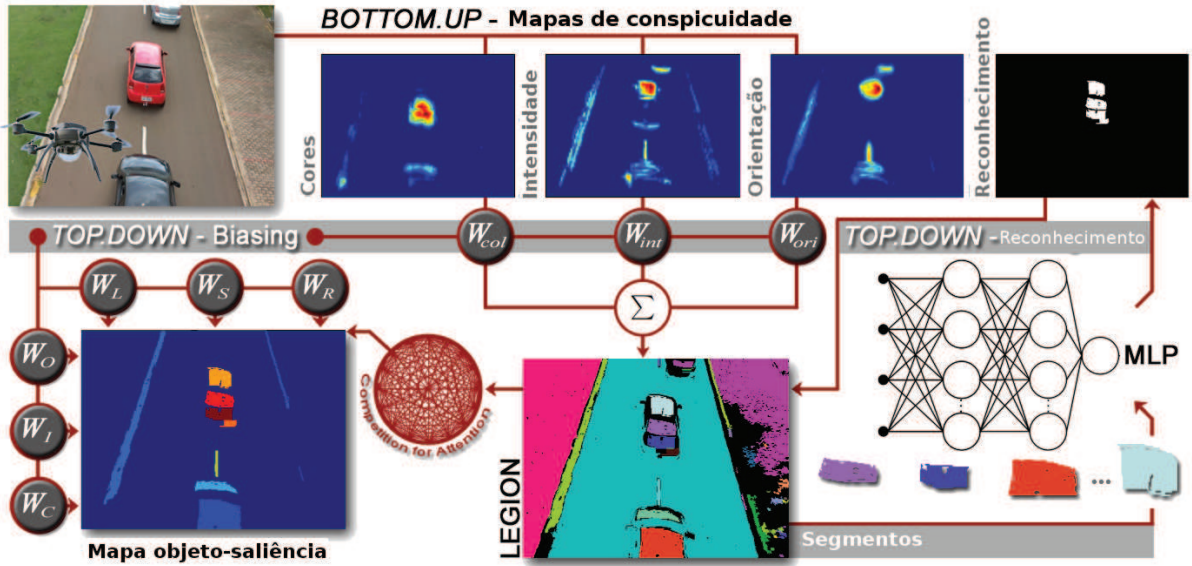


Figura 1: Um fluxograma do modelo de seleção de objetos usado.

os pixels que pertencem à imagem de entrada, a média de funcionalidade de cada saliência de k do segmento j pode ser definida como:

$$S_k^j = \frac{1}{n^j} \sum_{i \in n^j} l_{ki}^j, \quad (3)$$

no qual n^j representa o número de neurônios no segmento j e l_{ki}^j é o valor do mapa de saliência no neurônio i pertencendo à característica k no segmento j . Cada objeto é representado permanentemente por k características, preservando a informação do segmento.

Segundo [13], [1], [4], uma outra característica importante que pode orientar a implantação de atenção é o tamanho do objeto, o que, neste trabalho, é representado por n^j i.e. o tamanho do objeto é incorporado no valor saliência S_k^j de acordo com o número de neurônios que representa o segmento j . Portanto, o vetor da característica é redefinido e normalizado, tal como $l_k = [C, I, O, L, S, R] \in [0, 1]^6$.

Note-se que os segmentos não podem ter valores significativos de saliência. Com isto em mente, a utilização de um valor de *threshold* da saliência (θ_s) é indicado. Assim, os segmentos com valor de saliência abaixo do valor de *threshold* não vão competir por atenção. A seleção prévia de saliência dos objetos S^j é dada por:

$$P_{priorselection}(S^j) = \begin{cases} 1, & \text{if } \left(\frac{1}{n^k} \sum_{i \in n^k} S_i^j \right) \geq \theta_s \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

sendo que n^k denota o número de *features* responsáveis por guiar a atenção no caso deste experimento.

D. Mapa Objeto-Saliência

O mapa objeto-saliência é definido como uma rede composta de objetos, com dois tipos de conexões: excitatórias e

inibitórias. Conexões excitatórias representam um mecanismo de cooperação responsável por sincronizar grupos de objetos que representam bem os padrões de semelhanças (objetos com características semelhantes). Por outro lado, as ligações de inibição são concebidas para inibir a objetos relativos a objetos da cena de fundo, permitindo que o objeto relacionado com o objeto mais saliente da cena seja o selecionado.

Quando um objeto é pulsado na LEGION, seu sinal de saliência é comparado com todos os objetos selecionados, nos quais os estados são atualizados por:

$$\dot{v}_i = -v_i + E_i - W_Y Y_i + \sum_{k=1}^6 W_k S_k^i, \quad i = 1, \dots, n^s \quad (5)$$

sendo n^s o número de objetos pulsantes. A variável v_i representa um estado de ativação do segmento i , E_i é o termo de acoplamento excitatório e W_Y é o peso da inibição do acoplamento inibitório Y_i . A equação (5) representa um neurônio da rede integra-e-dispara.

Considere S_k^j um objeto pertencente a um segmento ativo no modelo LEGION e k seu respectivo índice de características, o termo de acoplamento excitatório E_i e o termo de acoplamento inibitório Y_i são definidos por:

$$E_i = Y_i = d(S_k^i, S_k^j), \quad (6)$$

sendo E_i atualizado se e somente se o valor de E_i contém o valor máximo de ativação do objeto i , S_k^i representa cada objeto pulsante e k é o índice da característica. A similaridade entre as características que representam o objeto S_k^j ara os outros objetos é definido por:

$$d(m, l) = \exp \left(-\sqrt{\sum_{k=1}^6 W_k (m_k - l_k)^2} \right), \quad (7)$$

sendo W_k o peso associado a cada característica de k . Ajustando os pesos de W_k é possível guiar a atenção para as características desejadas. Assim, se $W_k = 0$ para toda a informação primitiva da imagem de entrada, o modelo proposto torna-se estritamente um modelo *top-down*, e se $W_k = 0$ para informações relacionadas à reconhecimento de objeto, torna-se um modelo *bottom-up*.

As conexões inibidoras são determinadas com base no contraste entre as características.

Deste modo, se dois objetos são alimentados por características semelhantes, isto é, o contraste entre eles é pequeno ou zero, o termo Y_i , na equação (6), se aproxima de 1.0 (ver equação (7)) e, portanto, o termo de acoplamento Y_i na equação (5) assume um valor alto.

Por outro lado, quando o sinal de tais objetos é definido por características diferentes, o inibidor de acoplamento entre eles é pequeno ou mesmo nulo. Em consequência, os objetos com características semelhantes são mutuamente inibidos devido à competição gerada pelas conexões inibidoras. Um objeto que tem um elevado contraste em relação aos outros não é inibido e permanece oscilante. Assim, ele representa o objeto sob o foco de atenção do sistema.

IV. DESCRIÇÃO DE HARDWARE E COLETA DE DADOS

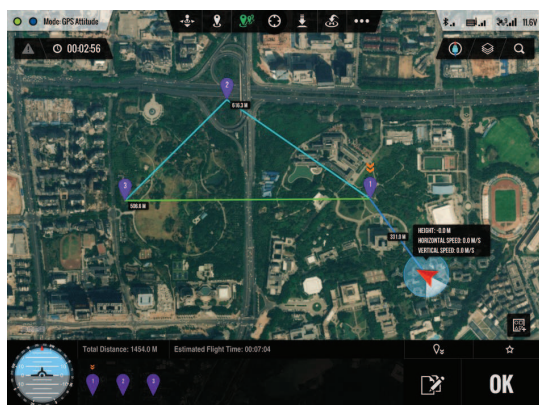


Figura 2: Software da estação terrestre DJI [8].

Uma câmera GoPro foi utilizada para aquisição de imagens. Essa câmera tem como característica o baixo custo, alta resolução (12MB), correção de ruído/vibração, porém por conta da sua grande lente angular ocorre distorção, mas permite um ângulo de visão de 170 graus. Além da elevada resolução, o uso da câmera GoPro apresenta outras vantagens que facilitam sua aplicação nas condições desejadas, como conexão por Wi-fi e caixa de proteção.

A Figura 2 mostra a tela do software localizado na base terrestre DJI.

A aquisição das imagens foi realizada utilizando um quadrotor, equipado com um sistema de controle comercializado pela empresa DJI, NAZA V2 e uma câmera GoPro. O sistema de controle compreende em uma IMU (Inertial Measurement Unit), um GPS, um gerenciador de energia e um data link. Esse sistema realiza o controle de estabilidade de acordo

com a posição da aeronave e se comunica por meio do data link (receptor/quadrotor) com a estação terrestre (transmissor/terra). A estação terrestre se comunica via bluetooth com um iPad, no qual está instalado o software proprietário. O software disponibiliza diversas funções e configurações, como percorrer uma rota pré-determinada autônoma, decolar e pousar de forma autônoma, controlar via joystick virtual (no iPad), iniciar, alterar e abortar a missão, definir altitude, velocidade, tempo de planagem, etc.

As imagens foram coletadas durante um voo do quadrotor de 30 minutos de duração a uma altura de 30 metros. Um total de 287 imagens foram selecionadas manualmente a partir do vídeo gravado e usado para testar o sistema proposto.

O sistema visual proposto foi testado para detectar veículos pertencentes a cinco classes diferentes. São elas: carros, caminhões, motos, picapes e ônibus, como pode ser visto na Figura 3.

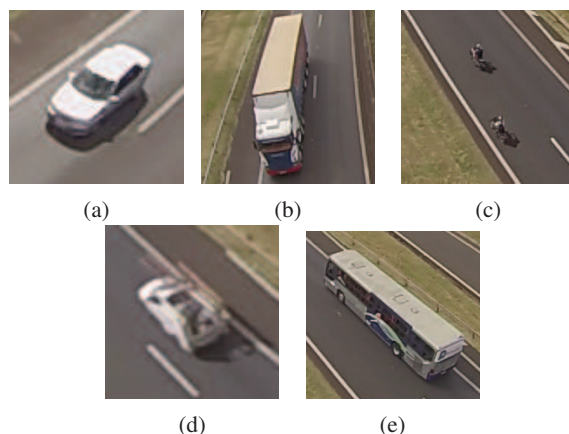


Figura 3: Os cinco tipos de veículos considerados: (a) carro, (b) caminhão, (c) moto, (d) caminhonete e (e) ônibus.

V. EXPERIMENTOS

A calibração da câmera é o processo de corrigir e remover a distorção das lentes da câmera, este processo é feito apenas uma vez. Nesta etapa, alguns dos quadros de vídeo captados são usados para detectar um padrão de tabuleiro de xadrez e determinar a matriz de distorção e a matriz de parâmetros intrínsecos da câmera. Após o processo de calibração as imagens são retificadas para obter uma imagem com menos distorção.

A captura de vídeo é feita em 60 quadros por segundo, mas como o software não pode processar toda essa quantidade de informação, forma selecionados quadros-chave a serem processados. Os quadros-chave são selecionados com base no movimento dos objetos na cena. Se a quantidade de pixels diferentes entre duas imagens exceder um limite, em seguida, uma nova imagem é selecionada como quadro-chave.

A imagem apresentada na Figura 4 foi utilizada para testar a performance do sistema proposto. Entretanto, outras imagens foram extraídas do conjunto de dados, constituído por 287

imagens no total, para também serem testadas, podendo conter um ou mais veículos na cena.



Figura 4: Imagem de entrada do sistema.

Para o reconhecimento de veículos, a rede MLP foi treinada para cada tipo de classe de veículo. Um exemplo de imagens utilizado para treinar a classe de caminhões é mostrando na Figura 5.

Foi considerado um total de 30 imagens para cada um dos cinco tipos de veículos para o treinamento da MLP. A MLP utilizada é constituída por 900 neurônios na camada de entrada (imagens de 30x30 pixels), também 900 neurônios na camada escondida e 5 neurônios na camada de saída, um por cada classe. Foi utilizada a função tangente hiperbólica para ativação dos neurônios.

O método de validação *10-fold validation* foi empregado para obter a média de classificações corretas e foi obtida uma acurácia de 71,33%.

Na Tabela I, pode ver-se a matriz de confusão obtida para a classificação das cinco classes de veículos consideradas. Houve um erro de classificação maior entre caminhonetes e caminhões e entre caminhões e ônibus. Como esperado,



Figura 5: Subconjunto das imagens consideradas para o treinamento de caminhões.

motos têm o mínimo erro de classificação devido seu tamanho, enquanto outras classes têm mais semelhança entre si. Na Figura 6d observa-se o resultado para a imagem apresentada na Figura 4.

Nas Figuras 6(a) a (c), podem ser vistas as intensidades de cores e características de orientação extraídos pelo processo *bottom-up* aplicado à imagem original.

Na Figura 6(d), é mostrada a posição do objeto de maior atenção. Na Figura 7, é mostrada a imagem correspondente à segmentação pela rede LEGION e na Figura 8 o objeto saliente de maior atenção aparece na cor vermelha. Finalmente, na Figura 9 é mostrado um gráfico que contém os valores da saliência dos objetos que competiram pela atenção. Podemos ver que a região do asfalto foi a de maior valor saliência, mas considerando o enviesamento *top-down* de objetos conhecidos pelo MLP, o objeto com valor igual a 0,35 saliência é o preferido e destacado pelo sistema visual

O sistema de atenção visual detectou os veículos em cenas reais com taxa de acerto de 83%. Isto é devido à necessidade de lidar com diferentes condições de luz (não consideradas nas imagens utilizadas para o treinamento) e também devido ao ruído nas imagens durante o processo de captura de imagens.

Tabela I: A matriz de confusão da classificação.

Classe real	Carros	Caminhões	Motos	Camionete	Ônibus
Carros	22	3	2	3	0
Caminhões	1	20	0	2	7
Motos	3	0	27	0	0
Camionete	8	5	0	17	0
Ônibus	0	8	0	1	21

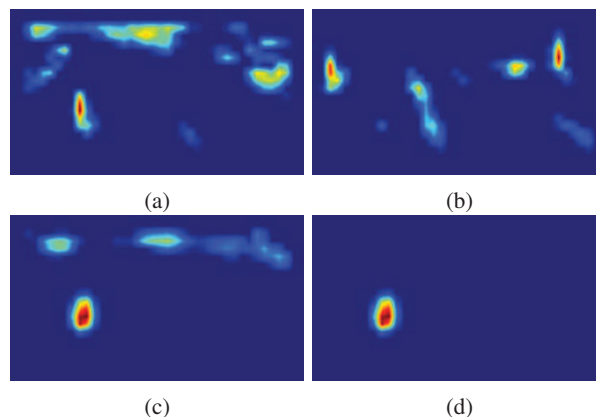


Figura 6: Mapas de conspicuidade de (a) intensidades, (b) cores e (c) orientações e o (d) reconhecimento.

VI. CONCLUSÃO

Neste trabalho, o sistema de atenção visual *bottom-up* e *top-down* foi estendido para ser usado para entrada de vídeo, obtido por uma câmera monocular montada em um quadrotor desenvolvido pelo laboratório com o objetivo de detectar veículos de cinco classes diferentes em rodovias. Para a detecção de características das imagens, considerou-se um modelo de rede LEGION combinada com uma rede de



Figura 7: Imagem segmentada.

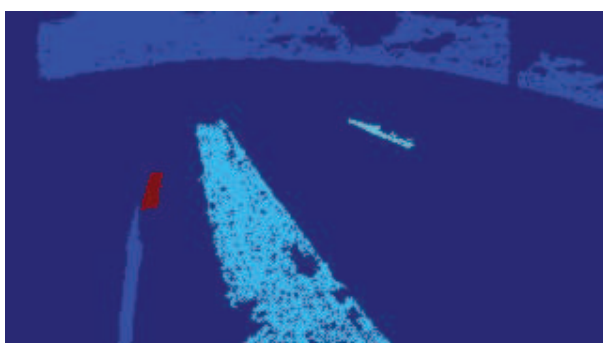


Figura 8: Imagem destacando o objeto saliente.

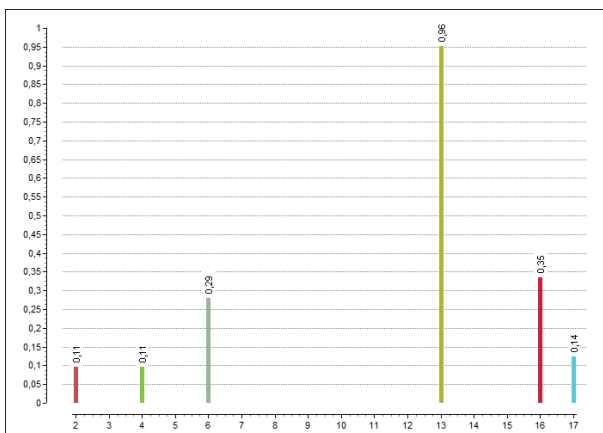


Figura 9: Gráfico dos valores salientes dos objetos que competiam pela atenção.

neurônios integra-e-dispara e para a classificação dos veículos foi utilizada uma rede *multi layer perceptron*. Resultados positivos foram encontrados para a tarefa de detecção de veículos em uma rodovia, uma vez que o sistema proposto conseguiu detectar 83% dos veículos e classificar 71% dos veículos detectados. Como trabalho futuro, pretendemos usar outra técnica para a classificação de objetos, para aprimorar o conjunto de dados de imagem e também para comparar o sistema proposto com os outros utilizados para a detecção de imagens de objetos, tais como, SIFT e SURF.

AGRADECIMENTOS

Os autores agradecem a FAPESP (processo 2012/14725-4) e o CNPq.

REFERÊNCIAS

- [1] T. A. and G. S., "Feature analysis in early vision: evidence from search asymmetries," *Psychol. Rev.*, vol. 95, pp. 15–48, 1988.
- [2] X. Cao, C. Wu, P. Yan, and X. Li, "Linear svm classification using boosting hog features for vehicle detection in low-altitude airborne videos," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2421–2424.
- [3] J.-Y. Choi and Y.-K. Yang, "Vehicle detection from aerial images using local shape information," in *Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology*. Springer-Verlag, 2008, pp. 227–236.
- [4] S. D., "The combination of spatial frequency and orientation is effortlessly perceived," *Percept. Psychophys.*, vol. 43, pp. 601–603, 1988.
- [5] J. Gleason, A. Nefian, X. Bouysounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 2065–2070.
- [6] S. Haykin, *Redes Neurais - Princípios e Práticas*. Bookman, 2001.
- [7] S. Hinz and A. Baumgartner, "Vehicle detection in aerial images using generic features, grouping, and context," in *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*. Springer-Verlag, 2001, pp. 45–52.
- [8] N. D. Innovationshor, "Ipad ground station," <http://http://www.dji.com/product/ipad-ground-station>, 2015 (accessed January 7, 2015).
- [9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [10] V. Ramakrishnan, A. K. Prabhavathy, and J. Devishree, "A survey on vehicle detection techniques in aerial surveillance."
- [11] S. Sahli, Y. Ouyang, Y. Sheng, Lavigne, and D. A. Lavigne, "Robust vehicle detection in low-resolution aerial imagery," pp. 76 680G–76 680G–8, 2010.
- [12] D. Wang and D. Terman, "Image segmentation based on oscillatory correlation," *Neural Computation*, vol. 9, pp. 805–836, 1997.
- [13] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it ?" *Nature Review Neuroscience*, vol. 5, pp. 495–501, 2004.
- [14] L. Zhao and E. E. N. Macau, "A network of dynamically coupled chaotic maps for scene segmentation," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1375–1385, 2001.