

Detecção de Pessoas por meio de Cascata de Classificadores e Descritor Hog para Dispositivos Móveis

Cleiton Silva Tavares¹

Marco Túlio A. N. Rodrigues²

César Augusto de Oliveira Soares³

Resumo: Este trabalho apresenta a construção de um método destinado a ajudar condutores de veículos utilizando dispositivos móveis. O objetivo é fazer a detecção de pedestres e, conseqüentemente, contribuir para a diminuição de acidentes. A solução pode ser utilizada em diferentes cenários, a partir de um sistema de semáforo, locomotiva, veículos autônomos e veículos comuns. O método se baseia em cascata de classificadores e o descritor Hog para realizar a detecção de pedestres. A definição da resolução auxiliaram na construção dos protótipos. Os resultados apresentados demonstram a viabilidade do método o que foi validado pelos resultados alcançados na detecção.

Palavras-chave: Detecção de objetos, Classificadores, Descritores de característica

1. Ciência da Computação, Bacharelado, Universidade de Itaúna, cleitonsilvatavares@gmail.com.
2. Ciência da Computação, Doutorado, UIT, tulio.rodrigues@gmail.com.
3. Ciência da Computação, Mestrado, UIT, FAPAM, cesaroliveira18@hotmail.com

Marque a forma de apresentação desejada: (X) oral; () pôster



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

 Universidade de Itaúna

Introdução

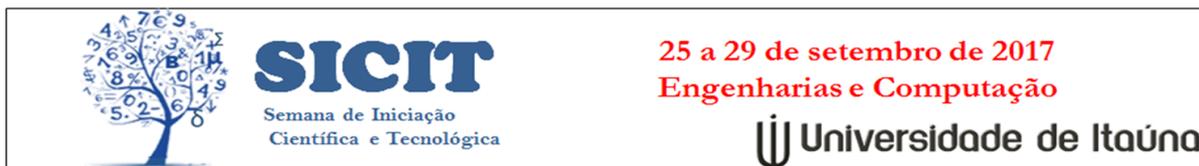
Mesmo não se tratando de um tema novo, a visão computacional vem trazendo muitas contribuições com suas bases de pesquisas, dentre suas diversas aplicações, este trabalho terá como foco o sistema de detecção de pessoas, sendo abordado dois algoritmos capazes de realizar a detecção, sendo eles o Haar Cascade e o Hog Descriptor.

[Vedaldi and Fulkerson 2010] apresenta que a maioria dos algoritmos de visão computacional são relativamente novos e quando uma aplicação que utiliza esses algoritmos fica disponível eles se limitam a serem utilizados apenas em ambientes muito específicos. E de acordo com [Viola et al. 2005], os algoritmos de visão computacional que utilizam detecção, demonstra que um detector pode ser treinado para obter informações de movimento, tentando rastrear objetos em movimento ao longo de muitos quadros e, em seguida, analisar o movimento para procurar periodicidade, ou outras pistas, mas também pode ser implementado um detector que seja capaz de utilizar como base de informações o movimento e a intensidade, executando a busca exaustiva sobre a imagem inteira em todas as escalas, usando grandes conjuntos de dados para alcançar a detecção e com taxas de falsos positivos muito baixos.

[Gavrila and Philomin 1999] demonstram em seu trabalho que, pode ser necessário o uso de vários sensores para dar ao condutor de um veículo, informações relevantes sobre o ambiente e, se desejar, poder executar tarefas simples de controle do veículo.

Diante dessas informações, pode-se notar que para a utilização de uma dessas tecnologias só seria possível em um ambiente específico, onde seria necessário abandonar recursos que são utilizados hoje, e investir uma quantia significativa para a aquisição de câmeras ou sensores adequados para poder utilizar esses recursos tecnológicos, ou até mesmo realizar a compra de um novo veículo que já possua componentes tecnológicos que permitam a utilizam desses recursos.

Para não ficarmos limitados a ambientes específicos onde existe a demanda da aquisição de equipamentos específicos para a utilização de recursos disponibilizados pela visão computacional, podemos buscar uma forma alternativa para que consigamos atingir um número maior de usuários, onde conforme é apresentado por [Felt et al. 2011] a partir de 2011, o Android Market já incluiu mais aplicações do que a Apple App Store e apresentado por [Android a] que o Android é um sistema operacional que move mais de um bilhão de



dispositivos ao redor do mundo, desde smartphones e tablets a relógios, Tvs, carros, voltaremos nosso foco em tentar desenvolver nossa aplicação para esta plataforma, podendo assim tentar atingir o maior número de usuários possíveis.

O problema consiste em dada uma sequência de vídeo como entrada, detectar pessoas dentro de um raio de detecção fornecendo a saída esperada em um tempo eficiente.

Com a realização de pesquisas, análises e avaliação dos métodos utilizados para detecção de pessoas e conhecendo o funcionamento da biblioteca OpenCV, este trabalho pretende desenvolver um sistema de detecção de pessoas em tempo real que poderá ser utilizado pelo próprio dispositivos móvel de forma a atender a população sem a necessidade de realizar grandes investimentos financeiros, de forma a auxiliar o condutor de um veículo podendo assim impactar positivamente para a diminuição do número de acidentes de trânsito.

O trabalho tem como principal contribuição o estudo na área de Visão Computacional, propondo um protótipo que utilize um método de detecção de pessoas em tempo real para dispositivos móveis.

Principais Trabalhos Relacionados

Esta seção descreve os principais trabalhos utilizados como referência para o desenvolvimento deste projeto.

Detecting Pedestrians Using Patterns of Motion and Appearance

[Viola et al. 2005] combinam detectores com base de informação de movimento com detectores baseados em informações de aparência(intensidade) utilizando um algoritmo de detecção treinado (usando AdaBoost para selecionar um subconjunto de recursos e construir o classificador). Com o uso de um único classificador para a detecção exigiria muitos recursos e, assim, ser muito lenta para operação em tempo real, eles propuseram uma arquitetura em cascata para fazer o detector eficiente.

Pode ser observado nas Figuras 1 e 2 alguns exemplos de detecção realizadas pelos detectores de pedestres dinâmico e estático, respectivamente.

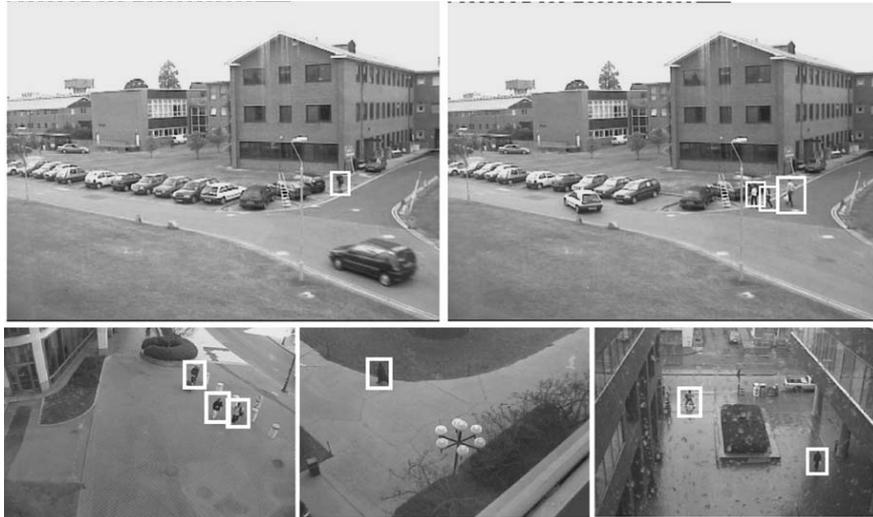


Figura 1. Exemplos de detecções para a detector dinâmico [Viola et al. 2005].

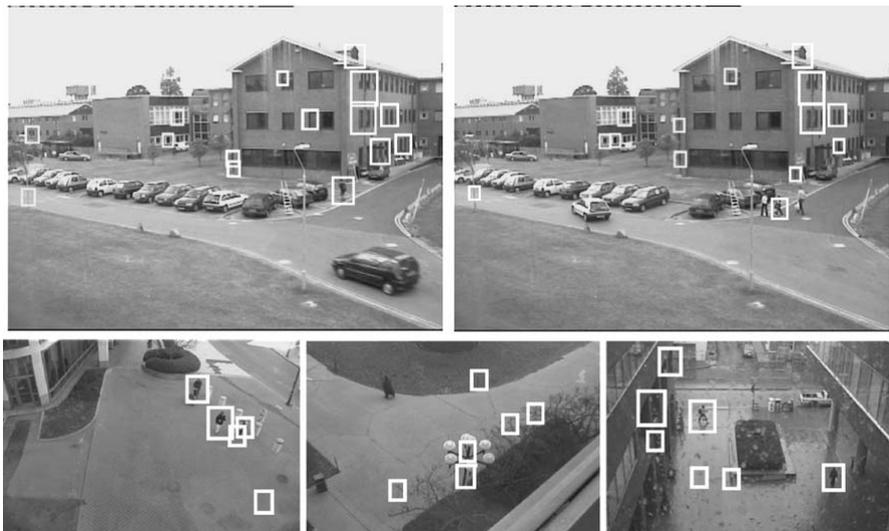


Figura 2. Exemplos de detecções para o detector estático [Viola et al. 2005].

Real-time Pedestrian Detection with Deformable Part Models

[Cho et al. 2012] descrevem um sistema de detecção de pedestres em tempo real que utiliza uma análise de restrições geométricas para pesquisar de forma eficiente pirâmides de recursos. O sistema de detecção baseia-se em uma simplificação na estrutura de algoritmo de cascata em estrela para um modelo deformável à base de parte. Utilizando o modelo baseado no HOG em parte deformável como parte de um sistema de rastreamento de detecção que se destina ao uso com um veículo autônomo, bem como um sistema de alerta precoce para a



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

Universidade de Itaúna

segurança do condutor. Exemplos de resultado de detecção do sistema podem ser observados pela Figura 3.



Figura 3. Resultados de detecção do sistema de detecção de pedestres. Com um modelo deformável baseado em parte, o sistema detecta pedestres até 25m de forma confiável [Cho et al. 2012].

Metodologia

Esta sessão descreve a metodologia utilizada para o desenvolvimento dos protótipos responsáveis em realizar a detecção de pessoas.

O sistema de detecção de pedestres em tempo real que será utilizado consiste basicamente na utilização de três etapas: (a) aquisição do vídeo (b) pré-processamento (c) verificação de características (d) marcação de detecção.

O processo do sistema inicia-se pela arquitetura do detector de pessoas, na definição de qual método será analisado. O vídeo sempre estará capturando informações em tempo real, por meio de uma câmera de celular. Enquanto o vídeo está sendo capturado, os frames são processados em escala de cinza para servir de entrada para o detector de pessoa.

As características utilizadas para definir os critérios de detecção são os valores padrões disponibilizados pelo próprio método escolhido, onde os valores de detecções realizadas são armazenadas.

Os resultados de detecção realizados servem de entrada para o sistema de marcação de detecção, que no resultado final consiste em apresentar em todo momento uma marcação de um retângulo de cor verde em todos os objetos que foram detectados.



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

 Universidade de Itaúna

Resultados e Validações

Esta seção descreve os resultados obtidos através dos experimentos realizados com os protótipos responsáveis em realizar a detecção de pessoas.

Para o desenvolvimento dos protótipos foram utilizados a ferramenta Android Studio 2.2.2.0, o OpenCV Manager 2.20, biblioteca versão 2.4.11.0, ARM v7a with Neon and VFPv4.

As imagens utilizadas foram obtidas através de vídeos capturados através do aplicativo AZ Screen Recorder disponibilizado pela Google Play que permite gravar a tela do Android.

Foram desenvolvidos 2 protótipos utilizando dois métodos de detecção de pessoas utilizando o conjunto de características do corpo de pessoas padrões, disponibilizadas pelo próprio método da biblioteca Opencv.

O processo de detecção de pessoa ocorre em um programa móvel, podendo ser utilizado em qualquer lugar, desde que haja luminosidade o suficiente para que uma pessoa seja detectada. O protótipo se apresenta em forma de um aplicativo, desenvolvido para aparelhos com o sistema operacional Android, podendo ser instalado em qualquer dispositivo móvel que possua os requisitos mínimos para que haja um bom funcionamento.

O primeiro passo foi a definição de qual tipo de resolução a ser utilizada para a realização dos experimentos. Para isso foram analisados dois itens, a não ocorrência de perda de frames e qual tornaria o conteúdo capturado o mais próximo da realidade.

Para esse passo foram utilizados 2 aparelhos para a realização dos testes, sendo eles:

Asus Zenfone 2 Laser - CPU: Qualcomm Snapdragon 410 MSM8916 Cortex-A53 Quad-Core, 1.2 GHz, 2GB DDR RAM, câmera traseira de 13 MP com capacidade de gravar vídeos com uma resolução de 1920x1080 pixels;

Samsung Galaxy S6 - CPU: SAMSUNG Exynos 7420, Quad-core 1.5 GHz Cortex-A53 + Quad-core 2.1 GHz Cortex-A57, 3GB DDR RAM, câmera traseira de 16 MP com capacidade de gravar vídeos com uma resolução de 3840x2160 pixels.

Os testes iniciais utilizados com a resolução padrão de ambos os smartphones não foi detectado grandes problemas na utilização do método de Haar Cascade, mas não apresentaram resultados satisfatórios com o método de Hog Descriptor, por apresentar



inúmeras perdas de frames que estavam sendo capturados em tempo real por apresentar uma alta resolução.

Os testes realizados com ambos os aparelhos apresentaram perda de frames somente quando utilizado o método de Hog Descriptor com todas as configurações default, onde a resolução de captura do Galaxy S6 é 3840x2160 pixels e a do Asus Zenfone 2 Laser que é 1920x1080 pixels.

De acordo com testes realizados por [Silva 2015] com diferentes resoluções de vídeos, levando em consideração o tempo médio de processamento e a taxa de sucesso na detecção dos pedestres, que para os testes realizados encontrou a resolução 1280x520 pixels como sendo a com melhor eficiência.

Foram realizados testes com os 4 grupos de resoluções apresentadas por [Silva 2015] em ambos os dispositivos com o método de Hog Descriptor, pois o método de Haar Cascade não apresentou perda de frames ao executar com alta resolução:

1. Resolução de 1920x1080 pixels - Apresentou perda de frames;
2. Resolução de 1280x720 pixels - Apresentou perda de frames;
3. Resolução de 1280x520 pixels - Não apresentou perda de frames;
4. Resolução de 640x480 pixels - Não apresentou perda de frames;

Para tentar tornar o conteúdo capturado com o mais próximo da realidade, sem a perda de frames no decorrer da captura, escolhi utilizar a resolução de 1280x520 pixels, os demais parâmetros foram definidos de acordo com a utilização demonstrada por [Silva 2015]. Para as demais etapas dos testes, foi adotado o aparelho Asus Zenfone 2 Laser.

Com todos os parâmetros definidos foram realizados os testes de captura de vídeo em tempo real com os protótipos responsáveis pela detecção de pessoas, o resultado da detecção pode ser observado nas figuras 4 e 5 onde é demonstrada a detecção de pessoas com o método de Haar Cascade e na figura 6 utilizando o método de Hog Descriptor.

Vários fatores podem contribuir para a detecção correta ou a não detecção de pessoas, como a posição da pessoa, a quantidade de iluminação, a cor da roupa e do meio onde a pessoa se encontra, entre outros.

Com isso, foi comum a ocorrência de falsos positivos, onde eram detectados outros objetos que não fosse uma pessoa, como pode ser observado na figura 7 com Haar Cascade e na figura 8 com Hog Descriptor.



Figura 4. Detecção realizado por Haar Cascade

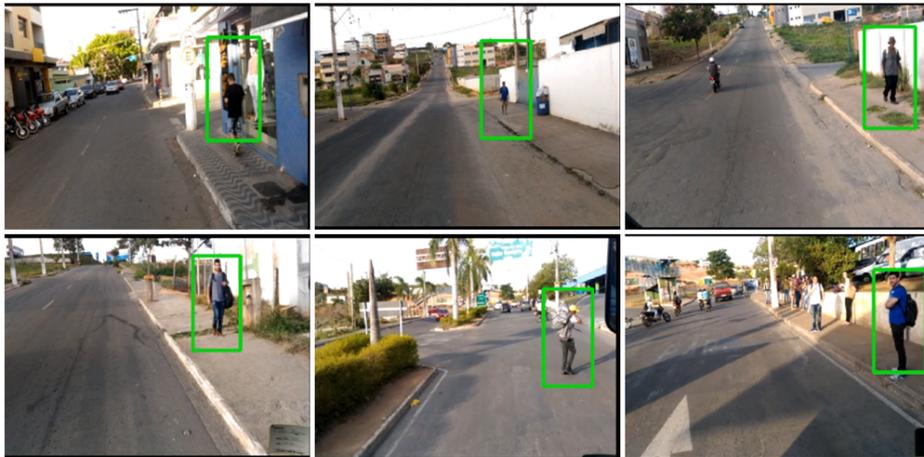


Figura 5. Detecção realizado por Haar Cascade

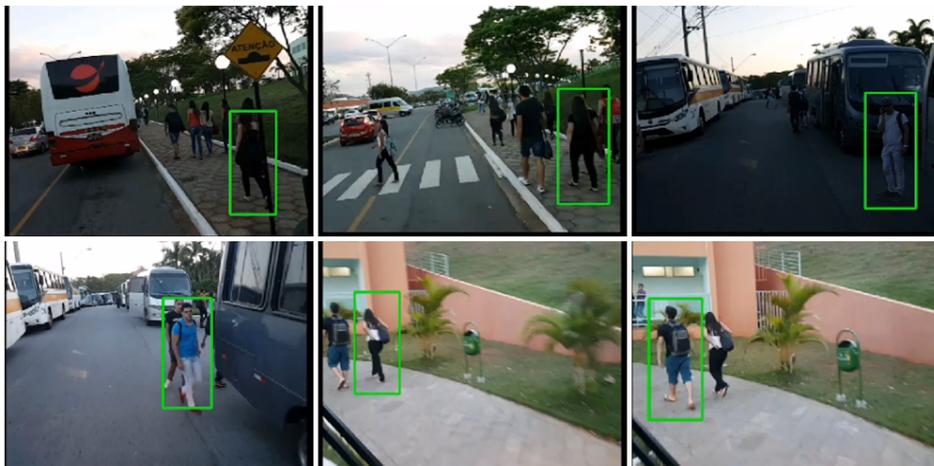


Figura 6. Detecção realizado por Hog Descriptor

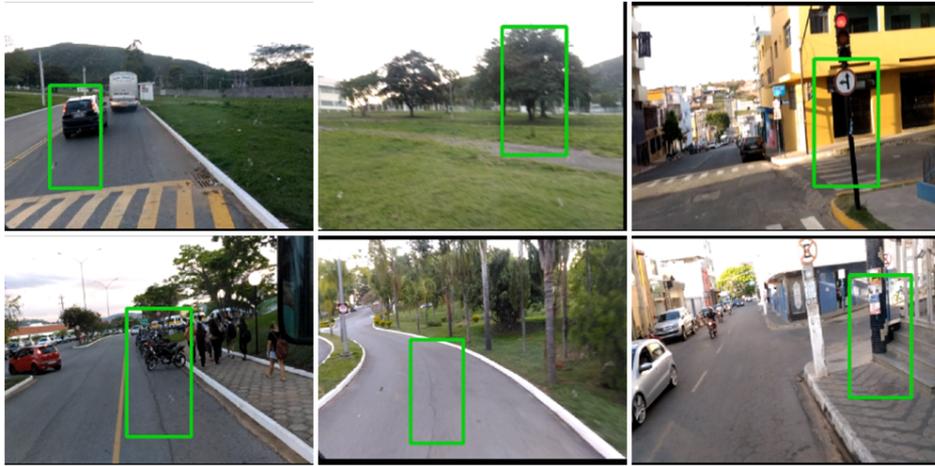


Figura 7. Falso positivo em Haar Cascade



Figura 8. Falso positivo em Hog Descriptor



Figura 9. Falso negativo em Haar Cascade



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

Universidade de Itaúna



Figura 10. Falso negativo em Hog Descriptor

Outro detalhe identificado foi a existência de falsos negativos, onde existiam pessoas que não foram detectadas pelos algoritmos conforme figura 9 que foi utilizada com Haar Cascade e na figura 10 com Hog Descriptor.

Conclusão

Esta sessão trata das conclusões e incorporações futuras que poderão ser realizadas com base neste trabalho. Este trabalho é um ponto de partida para o estudo de técnicas mais eficientes para detecção de pessoas, tema de grande importância para a sociedade nos tempos atuais devido ao grande número de automóveis existentes.

A captura de vídeos com tráfego de pedestres se fez necessário para que os resultados fossem os mais reais possíveis. A manipulação desses vídeos foi de extrema importância para o andamento do projeto, impactando diretamente nos resultados dos experimentos

O principal foco deste trabalho foi tornar a rotina para detecção dos pessoas o processo mais simples possível, para que qualquer usuário possa utilizar o mesmo.

Mesmo com a existência de um resultado específico, a rotina de detecção tinha a possibilidade de variar o seu resultado drasticamente com a presença de falsos positivos e falsos negativos, que variam de acordo com o conjunto de parâmetros utilizado em sua chamada. Devido a essa circunstância, seguidos testes foram repetidos em busca de uma variedade nos resultados.



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

 Universidade de Itaúna

Nos experimentos de detecção apresentou uma grande variação nos resultados apresentados entre ambos os algoritmos. Ambos realizaram o que era proposto, a detecção de pessoas, mas também apresentaram falsos positivos e falsos negativos.

Para os testes realizados houve ocorrências em ambos os métodos, em que a detecção ocorria porém, também apresentavam situações de falso positivo ou falso negativo no mesmo frame, ou a ocorrência de falso positivo e falso negativo no mesmo frame em que foi detectado uma pessoa.

Trabalhos Futuros

Após analisar os resultados foi possível encontrar alguns pontos que podem ser melhorados em relação a solução utilizada, buscando a produção de resultados melhores dos que foram obtidos nos protótipos desenvolvidos.

Para que se obtenha um melhor desempenho do algoritmo de detecção de pessoas é necessária uma otimização do mesmo, principalmente ao tempo de processamento, para que a utilização do mesmo em tempo real seja eficiente e confiável para ser utilizado. O desenvolvimento deste experimento utilizando apenas os exemplos default de dois métodos, demonstram que são válidos e podem ser aplicados, desde que seja realizada uma fase de treinamento com uma boa quantidade de exemplos para aumentar a precisão no ato da detecção.

Referências

[Android a] Android. Revolucionando telas de todos os tamanhos. <https://www.android.com/>. Online; accessed 13 May 2016.

[Cho et al. 2012] Cho, H., Rybski, P. E., Bar-Hillel, A., and Zhang, W. (2012). Real-time pedestrian detection with deformable part models. In Intelligent Vehicles Symposium (IV), 2012 IEEE, pages 1035–1042. IEEE.

[Felt et al. 2011] Felt, A. P., Chin, E., Hanna, S., Song, D., and Wagner, D. (2011). Android permissions demystified. In Proceedings of the 18th ACM conference on Computer and communications security, pages 627–638. ACM.



SICIT
Semana de Iniciação
Científica e Tecnológica

25 a 29 de setembro de 2017
Engenharias e Computação

 Universidade de Itaúna

[Gavrila and Philomin 1999] Gavrila, D. M. and Philomin, V.(1999). Real-time object detection for "smart" vehicles. The Proceedings of the Seventh IEEE International Conference on.

[Silva 2015] Silva, E. D. L. (2015). Detecção de pedestres no trajeto a ser percorrido pelo automóvel.

[Vedaldi and Fulkerson 2010] Vedaldi, A. and Fulkerson, B. (2010). Vlfeat: An open and portable library of computer vision algorithms. In Proceedings of the 18th ACM international conference on Multimedia, pages 1469–1472. ACM.

[Viola et al. 2005] Viola, P., Jones, M. J., and Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision, 63(2):153–161.