



SEMINÁRIO

Visão Computacional

Aluno: Afonso Pinheiro

Detecting Pedestrians Using Patterns of Motion and Appearance

*Paul Viola *Michael J. Jones *Daniel Snow

Por que detectar pedestres?



<http://conexaopenedo.com.br/2016/02/sistema-consegue-detectar-pedestres-em-tempo-real/>

Introdução

Utilizar informações de movimentos e informações de aparência em um único detector.

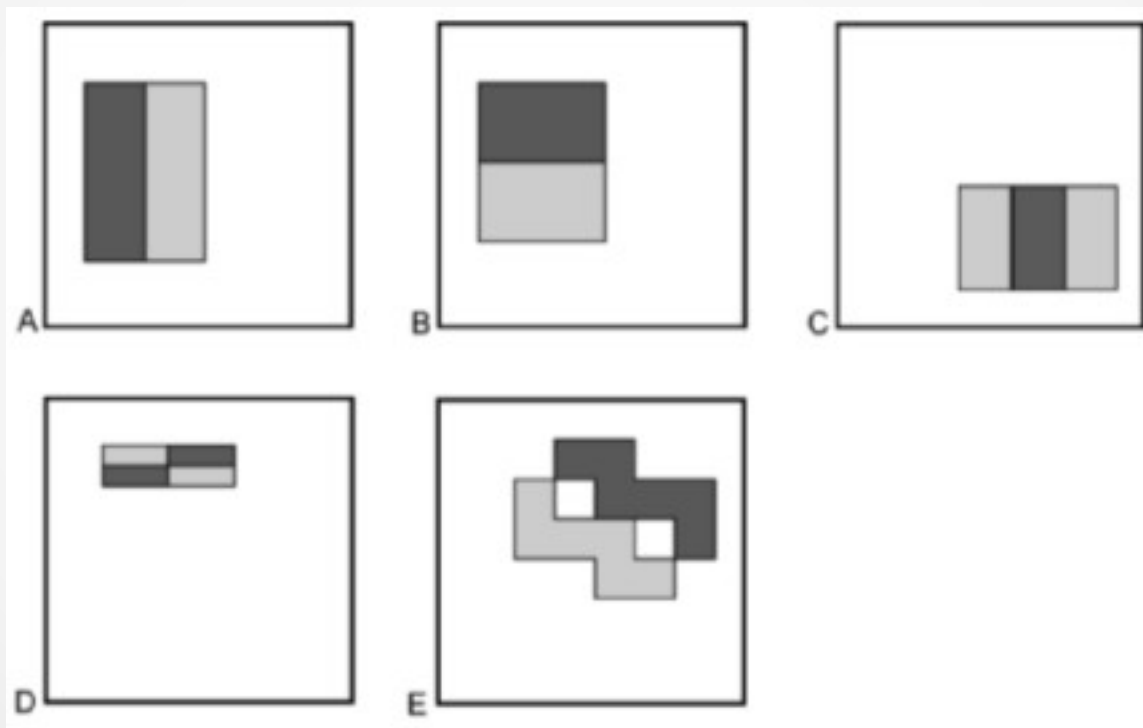
AdaBoost

Nessa implementação não há necessidade de mecanismo separados para rastreamento, segmentação, alinhamento e registro.

Extração de padrão de movimento a curto prazo

Técnicas utilizadas

Filtros de retângulo



P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In IEEE Conference on Computer Vision and Pattern Recognition, 2001.

Técnicas utilizadas

Filtros de movimento

$$\Delta = \text{abs}(I_t - I_{t+1})$$

$$U = \text{abs}(I_t - I_{t+1} \uparrow)$$

$$L = \text{abs}(I_t - I_{t+1} \leftarrow)$$

$$R = \text{abs}(I_t - I_{t+1} \rightarrow)$$

$$D = \text{abs}(I_t - I_{t+1} \downarrow)$$



Técnicas utilizadas

Filtros

$$f_i = r_i(\Delta) - r_i(S)$$

$$f_j = \phi_j(S)$$

$$f_k = r_k(S)$$

$$f_m = \phi(I_t)$$

Técnicas utilizadas

A classifier, C , is a thresholded sum of features:

$$C(I_t, I_{t+1}) = \begin{cases} 1 & \text{if } \sum_{i=1}^N F_i(I_t, \Delta, U, L, R, D) > \theta \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

A feature, F , is simply a thresholded filter that outputs one of two votes.

$$F_i(I_t, I_{t+1}) = \begin{cases} \alpha & \text{if } f_i(I_t, \Delta, U, L, R, D) > t_i \\ \beta & \text{otherwise} \end{cases} \quad (2)$$

where $t_i \in \mathcal{R}$ is a feature threshold and f_i is one of the motion or appearance filters defined above. The real-valued α and β are computed during AdaBoost learning (as is the filter, filter threshold t_i and classifier threshold θ).

Técnicas utilizadas

pixels. The scale invariance of the detection is achieved by operating on image pyramids. Initially the pyramids of I_t and I_{t+1} are computed. Pyramid representations of $\{\Delta, U, L, R, D\}$ are computed as follows:

$$\Delta^l = \text{abs}(I_t^l - I_{t+1}^l)$$

$$U^l = \text{abs}(I_t^l - I_{t+1}^l \uparrow)$$

$$L^l = \text{abs}(I_t^l - I_{t+1}^l \leftarrow)$$

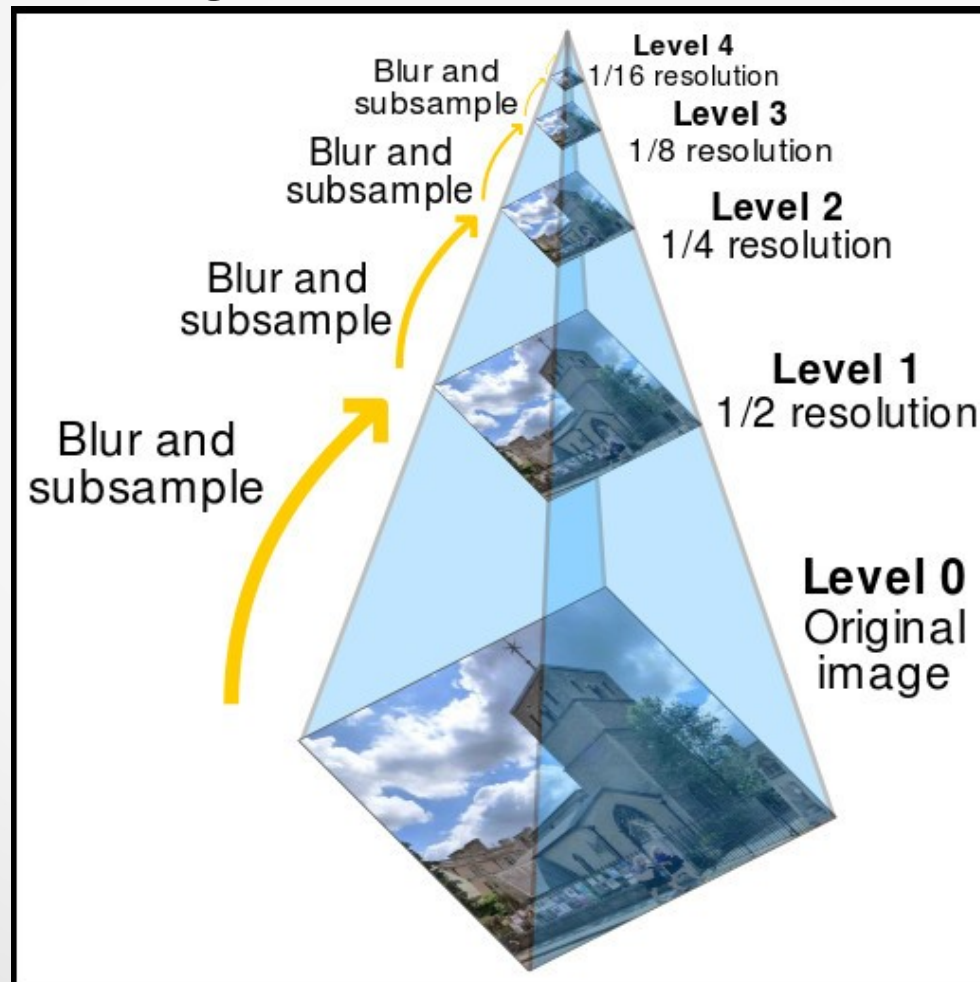
$$R^l = \text{abs}(I_t^l - I_{t+1}^l \rightarrow)$$

$$D^l = \text{abs}(I_t^l - I_{t+1}^l \downarrow)$$

where X^l refers to the the l -th level of the pyramid. Classi-

Técnicas utilizadas

Pirâmide de imagens



Processo de treinamento

O processo de treinamento usa o AdaBoost para selecionar um subconjunto de características e construir o classificador. Em cada etapa o algoritmo de aprendizado escolhe a partir do conjunto heterogêneo de filtros, incluindo filtros de aparência, filtros de direção de movimento, filtros de cisalhamento de movimento, filtros de magnitude de movimento. O algoritmo AdaBoost também seleciona o limiar ótimo de cada característica bem como os votos α e β de cada característica.

Processo de treinamento

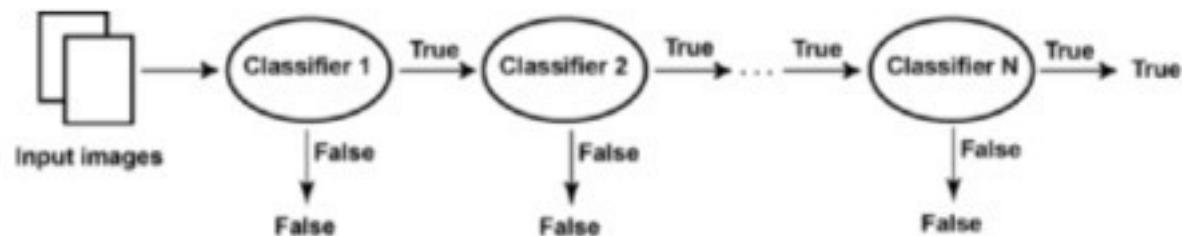


Figure 3: Cascade architecture. Input is passed to the first classifier with decides true or false (pedestrian or not pedestrian). A false determination halts further computation and causes the detector to return false. A true determination passes the input along to the next classifier in the cascade. If all classifiers vote true then the input is classified as a true example. If any classifier votes false then computation halts and the input is classified as false. The cascade architecture is very efficient because the classifiers with the fewest features are placed at the beginning of the cascade, minimizing the total required computation.

Processo de treinamento



Experimentos

DataSet



Experimentos

Training set

- Foram utilizadas 6 sequências de vídeo para o treinamento de um detector dinâmico de pedestre e um detector estático de pedestre. O detector dinâmico foi treinado utilizando pares de frames e o detector estático foi treinado usando apenas filtros de aparência.
- Cada Estágio da cascata é um classificador treinado melhorado usando 2250 exemplos positivos e 2250 exemplos negativos.

Processo de treinamento

Training set

- O limiar de detecção do classificador adicionado recentemente na cascata é ajustado para que a taxa de falso negativo seja muito baixa. O limiar é definido usando um conjunto de validação com 200 pares de imagens com exemplos positivos marcados.
- O limiar é ajustado de modo que pelo menos 99,5% dos pedestre continue sendo detectados corretamente, enquanto pelo menos 10% dos falso positivo após da última etapa são eliminados. Se este objetivo não pode ser atingido então mais características são adicionadas para o atual classificador.

Processo de treinamento

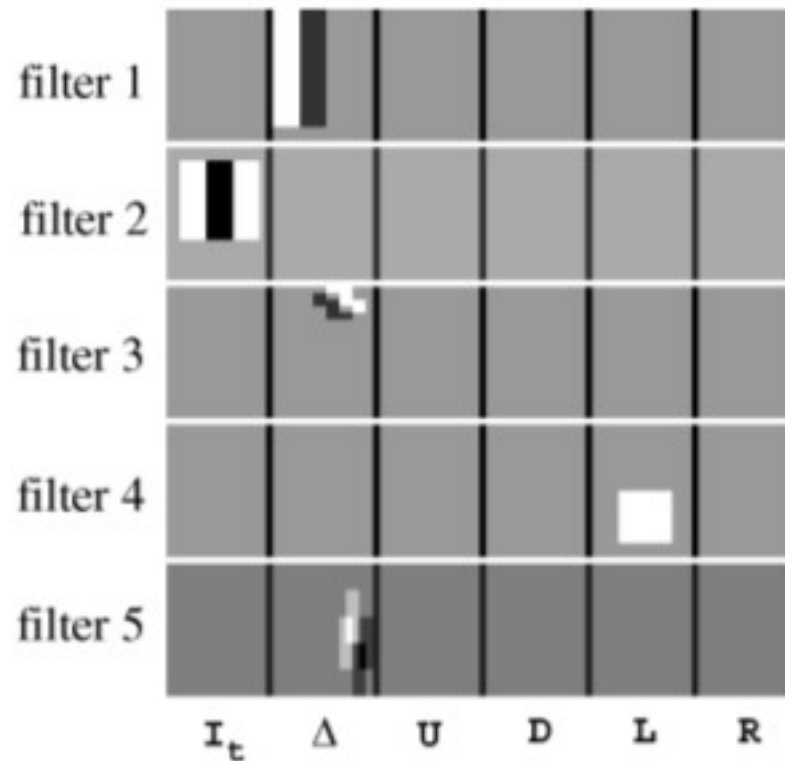
Training set

- O detector de pedestre dinâmico foi treinado usando 54,624 filtros que foram uniformemente sub amostrados a partir do conjunto muito grande de todos os filtros que coube em uma janela de 20x15 pixels.
- O detector estático foi treinado usando 24,328 filtros também uniformemente sub amostrados a partir do conjunto total possível.

Processo de treinamento

Training set

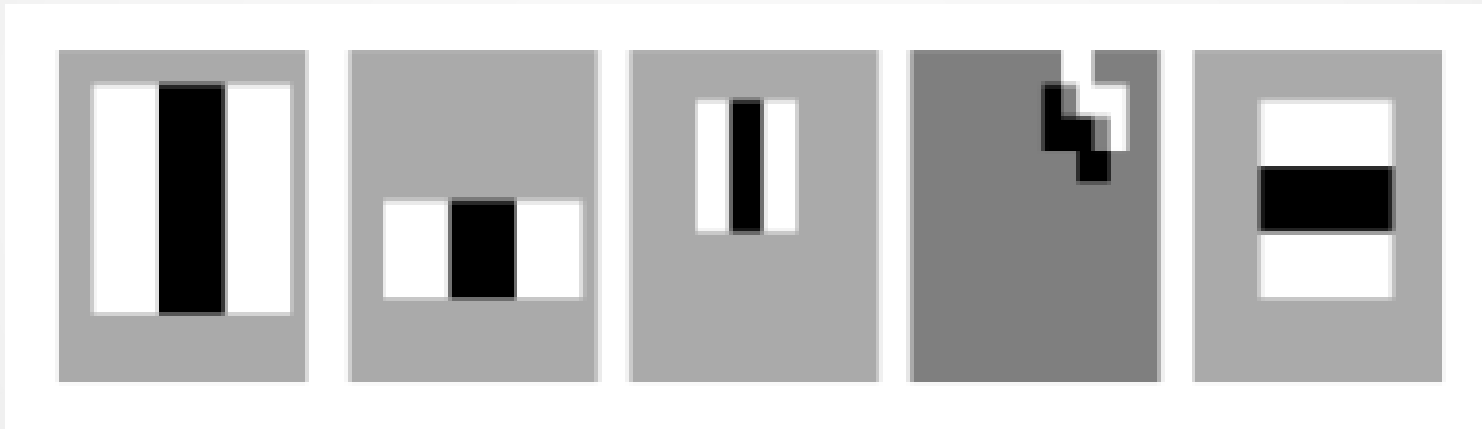
- Os 5 primeiros filtros aprendidos para o detector dinâmico.



Processo de treinamento

Training set

- Os 5 primeiros filtros aprendidos para o detector estático.



Resultados

Detector dinâmico



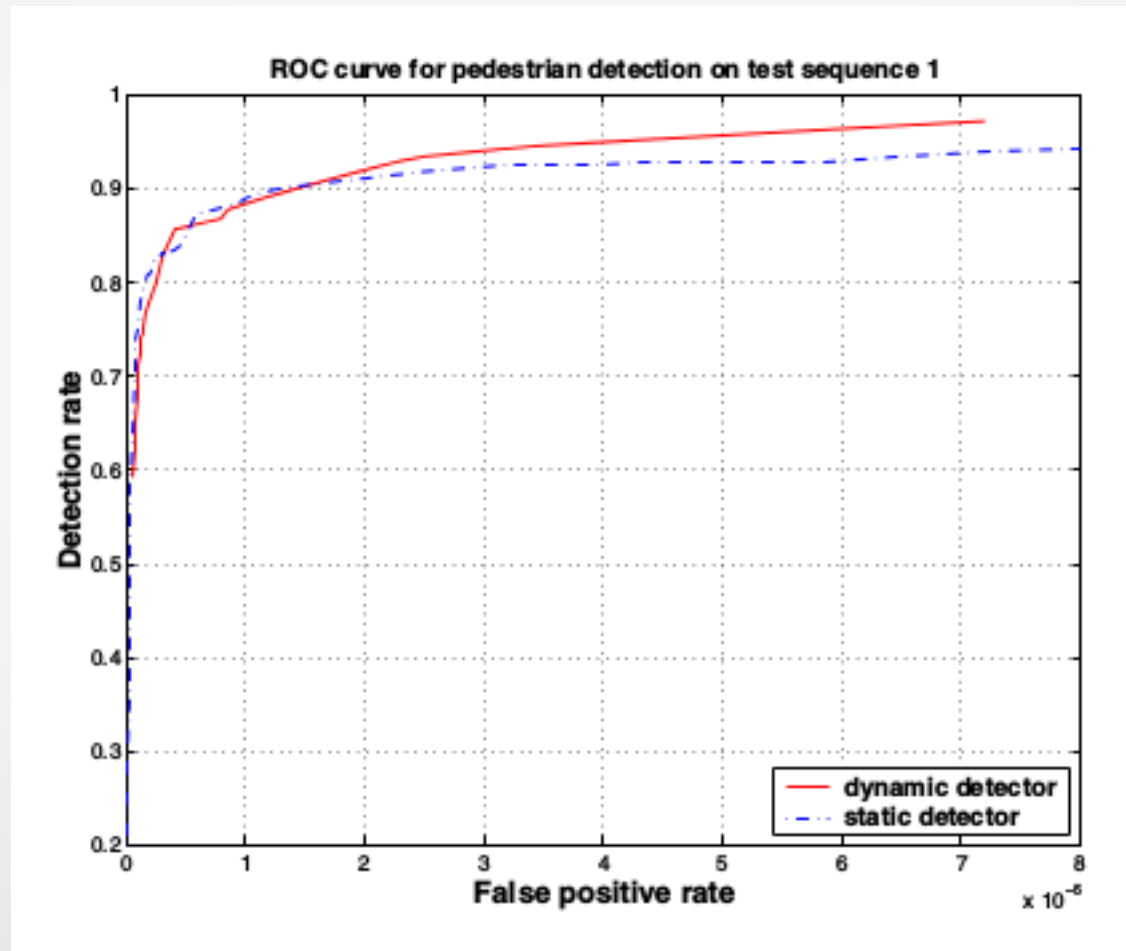
Resultados

Detector estático



Resultados

Curva ROC



Conclusão

- Tomando por base abordagens anteriores, essa é a primeira a combinar ambas informações de movimento e aparência em um único detector. Essa abordagem conseguiu detectar pedestres a partir de vários pontos de vistas com uma baixa taxa de falso positivo.
- O sistema exige cerca de 0.25 segundos para detectar todos os pedestres em uma imagem de 360x240 pixels rodando em um processador 2.8GHz.

Conclusão

- Com essa abordagem foi desenvolvido uma representação de imagem em movimento que é extremamente eficiente. Além de um sistema que detecta pedestres em imagens de baixa resolução com condições de dificuldades (tais como chuva e neve).
- A combinação de informações de movimento e aparência pode ser bem aplicável a outros problemas como classificação de expressão facial e possível leitura labial.



DÚVIDAS ?