

ESTIMATIVA DE MOVIMENTO E DETECÇÃO DE PELE APLICADAS AO PROBLEMA DE DETECÇÃO DE FACE EM VÍDEO NO ESPAÇO DE CORES RGB

VALESKA DE SOUSA UCHÔA*, KELSON RÔMULO TEIXEIRA AIRES*, CORNÉLIA JANAYNA PEREIRA PASSARINHO*, FRANCISCO WENDER PEREIRA DOS REIS*, RODRIGO DE MELO SOUZA VERAS*

*Universidade Federal do Piauí - UFPI
Departamento de Computação
64049-550, Teresina, Piauí, Brasil

Emails: valeskauchoa@live.com, kelson@ufpi.edu.br, janaynapassarinho@gmail.com, franciscowenderreis@gmail.com, rveras@ufpi.edu.br

Abstract— Face detection in unconstrained scenarios is one of the most researched tasks in Computer Vision due to its wide applicability. In this paper, we propose a face detection method that allows the creation of a real-time surveillance system. Our method consists of three stages: motion estimation, skin detection and face detection. The motion estimation method is based on a Gaussian Mixture Model. Model and skin detection is based on RGB information. The face detection stage uses the Viola and Jones *framework*. Experiments have shown bigger precision and less false positive rates than the method that inspired this work. Lastly, the proposed method presents processing time that is compatible to the methods which are already described in the literature that have similar quality.

Keywords— face detection, motion estimation, skin detection, surveillance.

Resumo— Detecção de faces em ambientes não-controlados é uma das tarefas mais estudadas em Visão Computacional, principalmente por sua vasta aplicabilidade. Neste trabalho, propomos um método de detecção de face que possibilite a criação de um sistema de vigilância em tempo real. Nossa proposta é dividida em três etapas: estimação de movimento, detecção de pele e detecção de face. A estimação de movimento se dá através de um modelo de mistura de gaussianas. A detecção de pele é feita com base em informações do espaço de cores RGB. A etapa de detecção de faces ocorre através do *framework* de Viola e Jones. Os resultados mostram uma melhoria, em relação ao método que o inspirou, no que tange ao número de falsos positivos e na precisão. Por fim, o método apresenta um tempo de processamento compatível ao de métodos com qualidade similar existentes na literatura.

Palavras-chave— detecção de face, estimação de movimento, detecção de pele, vigilância.

1 Introdução

O problema de detecção de faces em ambientes não-controlados é uma tarefa de Visão Computacional de grande aplicabilidade, especialmente em sistemas de vigilância (Kalal et al., 2010). A detecção de faces caracteriza-se pela localização de faces em quadros de um vídeo. Esse processo possibilita o rastreamento e o reconhecimento de faces de indivíduos em determinadas cenas de interesse. Apesar dos avanços realizados nesse campo, ainda há desafios a serem enfrentados no que tange à detecção de face. Dentre os principais desafios estão a posição da face e a oclusão por meio de outros objetos ou de outras faces. Além disso, um aspecto que pode influenciar no desempenho de um sistema de detecção é o fundo da cena captada (Verma et al., 2003). Essas dificuldades, quando bem contornadas, conferem ao sistema a robustez desejada, uma vez que mudanças na orientação, expressão facial, escala e iluminação, podem descaracterizar a face de um indivíduo a ser detectado (Xianbao et al., 2005).

Dentre as diversas abordagens utilizadas para resolver o problema de detecção de faces, destacam-se as soluções que usam *deformable parts-based model* (DPM) (Zafeiriou et al.,

2015) e as que utilizam *templates* rígidos com aprendizagem baseada em *boosting*¹ (Freund e Schapire, 1999). Um exemplo de utilização de *boosting* pode ser encontrado no método clássico de detecção proposto por Viola e Jones (2004). Em se tratando de tempo de processamento e qualidade de detecção, essas duas soluções são distintas. O ideal para um sistema complexo de vigilância em tempo real que realize as tarefas de detecção, reconhecimento e rastreamento é que haja um equilíbrio entre essas duas métricas. Tal fato é importante para a corretude das detecções realizadas, mas também para que o tempo de detecção de face não prejudique as demais atividades a serem realizadas pelo sistema de vigilância. Por isso, neste trabalho, deseja-se atingir um compromisso entre taxa de acerto e tempo de processamento, que possibilite o funcionamento do sistema em tempo real.

Para atingir os objetivos, propõe-se um método composto por uma etapa de estimativa de plano de fundo, seguida por uma detecção de pele e pelo módulo de detecção de face. A estimativa de plano de fundo é baseada em um modelo de mistura de gaussianas (Lima et al., 2015),

¹Boosting é um procedimento que combina classificadores “fracos”, objetivando uma melhora nos resultados da classificação.

enquanto a detecção de pele é fundamentada nas informações das bandas do espaço de cores RGB (Passarinho et al., 2015). O método de detecção de faces usa *haar-like templates* em uma abordagem de classificação em cascata (Viola e Jones, 2004). Os resultados foram obtidos a partir da avaliação do método proposto em dois vídeos coloridos, comumente utilizados na avaliação de algoritmos de detecção e seguimento de indivíduos (Ross et al., 2007; Babenko et al., 2009a).

2 Trabalhos relacionados

A detecção de faces já foi objeto de estudo de diversos trabalhos na área de Visão Computacional. Esses trabalhos podem ser divididos em dois grupos de soluções (Zafeiriou et al., 2015). Um grupo de algoritmos é baseado em *templates* rígidos enquanto outro grupo emprega modelos baseados em partes deformáveis (DPM).

Um exemplo do primeiro grupo é o *framework* Viola-Jones, mais utilizado para detecção de face (Viola e Jones, 2001). Desde a sua apresentação, surgiram inúmeros trabalhos inspirados em sua metodologia (Lienhart e Maydt, 2002; Li et al., 2002; Wu e Ai, 2008; Heydarzadeh e Haghighat, 2011).

O trabalho de Viola e Jones (2001) se destacou por três características principais: o conceito de imagem integral, a aprendizagem com AdaBoost, e uma cascata de classificadores. O referido trabalho representou um marco na tarefa de detecção de face pela combinação desses atributos, que tornaram possível um sistema robusto em tempo real. No entanto, muitas faces de perfil não são detectadas. Além disso, não há garantia de que a face detectada é humana, por exemplo, focinhos e caricaturas são detectados.

Um recente representante do segundo grupo de algoritmos de detecção de faces é o algoritmo DPM, proposto por Mathias et al. (2014). As abordagens DPM se baseiam no fato de que é mais fácil modelar aparência local do que global. Além disso, combinam estimação de variáveis para alinhamento e agrupamento durante o treinamento. Os modelos conseguem expressar mudanças significativas na aparência de objetos (Felzenszwalb et al., 2010). Entretanto, apesar desse método atingir taxa de acerto próxima de 100%, o algoritmo é consideravelmente lento.

O trabalho publicado por Luh (2014) utiliza o método Viola-Jones combinado com detecção de pele. O autor propôs um modelo de cor híbrido chamado $RGB - C_b C_r C_g$ no qual a classificação de pele é realizada.

Ban et al. (2014) propuseram um método de detecção de face baseado na probabilidade de

presença de pele através de um algoritmo *boosted*. Um modelo estocástico calcula a similaridade entre uma região de cor e a cor de pele. Um classificador em cascata foi construído utilizando características *haar-like* (Papageorgiou et al., 1998) e LBP (*Local Binary Pattern*) (Ojala et al., 1996).

Trabalhos recentes encontrados na literatura sugerem que classificar região de pele contribui para melhores resultados de detecção de face (Liulu e Mingyang, 2016; M. e Vadivel, 2016; Dahal et al., 2016). Isso se deve ao fato do espaço de busca por face se tornar mais restrito, o que contribui para um tempo menor de processamento. Além disso, essa restrição ajuda a garantir que faces detectadas sejam, de fato, faces humanas. Assim sendo, na próxima seção, propomos um novo método de detecção de faces que utiliza detecção de pele combinada com estimação de movimento. Ainda, o método proposto utiliza informações de cor. O modelo proposto consegue executar em um tempo aceitável para sistemas em tempo real.

3 Método Proposto

O algoritmo de detecção de faces proposto visa sua integração com métodos de seguimento e reconhecimento facial. Por isso, é fundamental que a tarefa de detecção seja realizada de modo a disponibilizar tempo para a execução das tarefas seguintes. Atender a esse requisito possibilita a implementação de um sistema de vigilância em tempo real.

Para tal, propõe-se um método de detecção de faces em conjunto com um algoritmo estimador de movimento e um detector de pele. O detector de faces é baseado no detector em cascata desenvolvido por Viola e Jones (2001). A estimação de movimento é feita por um modelo de mistura de gaussianas (Lima et al., 2015). A detecção de pele utiliza informações das bandas do espaço de cores RGB com os limiares encontrados em (Passarinho et al., 2015). Um esquema mostrando uma visão geral da proposta é apresentado na Figura 1.

Detectores de face que utilizam características *haar* combinadas com *boosting classifiers* são uma escolha satisfatória para sistemas em tempo real por serem uma solução bastante rápida. No entanto, foi observado nos experimentos que em imagens e vídeos de ambientes não-controlados, um grande número de falsos positivos é encontrado (ver Seção 4). As fases de detecção de movimento e pele têm o propósito de reduzir esse número de falsos positivos, além de restringir a área de busca para diminuir o esforço computacional.

A detecção de movimento se dá através de um mecanismo de segmentação por estimativa de

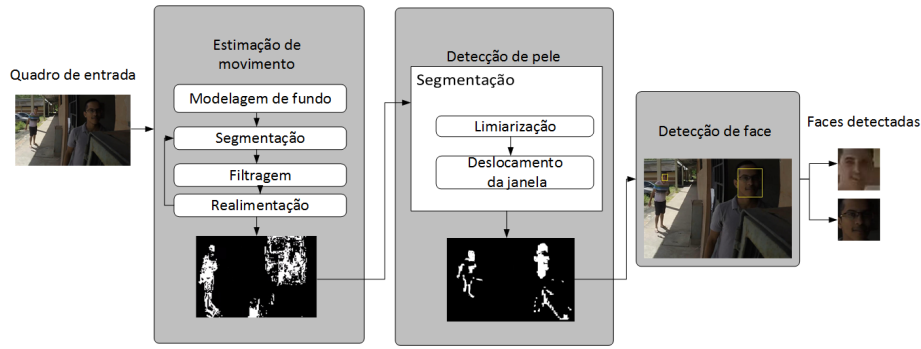


Figura 1: Esquema do método proposto.

fundo com realimentação, utilizando a abordagem apresentada em (Lima et al., 2015). A modelagem possibilita o agrupamento dos dados em duas classes: plano de fundo (*background* - *BG*) e primeiro plano (*foreground* - *FG*). O primeiro plano corresponde à região com indícios de movimento onde espera-se encontrar face. Esse processo é dividido em dois estágios: um estimador de fundo, que classifica os *pixels* de acordo com os valores do modelo de cor RGB; e um detector de sombras, que valida as classificações feitas no estágio anterior.

Ao final dessa etapa, tem-se como resultado uma imagem binária correspondendo a cada quadro do vídeo. Os *pixels* brancos indicam regiões de primeiro plano onde foi detectado movimento, e os pretos correspondem ao plano de fundo.

De posse das estimativas de primeiro plano, inicia-se a detecção de pele. Esse processo se dá através de um bloco de $n \times n$ *pixels* que é deslocado por todo o quadro da imagem colorida, na região correspondente ao primeiro plano. Nos experimentos utilizamos $n = 25$. Cada pixel dentro do bloco tem seus valores nas três bandas do espaço de cores RGB analisados a fim de se determinar se o pixel pertence a uma região de pele. O deslocamento da janela tem comportamento semelhante aos filtros morfológicos básicos, no entanto, cada *pixel* é analisado somente uma vez. Essa análise é feita na forma da Equação 1,

$$px = \begin{cases} 1, & \text{se } R > 20, G > 30, B < 100, \\ & R > G, R > B \\ 0, & \text{c.c.} \end{cases}, \quad (1)$$

onde R, G e B são os valores dos *pixels* nas bandas vermelha, verde e azul, respectivamente. O valor na banda R deve ser maior que nas outras duas, pois tons de pele geralmente são cores quentes, onde o vermelho predomina. A imagem Im' resultante desse passo é uma imagem binária que é usada como entrada na avaliação através da janela deslizante. Se mais de cinquenta por cento dos *pixels* de Im' dentro da janela

forem iguais a 1, os valores de toda a janela são copiados nos *pixels* correspondentes em Im . Caso contrário, as posições equivalentes em Im recebem 0. Essa estratégia visa corrigir os valores de *pixels* possivelmente mal classificados. O processo se resume na Equação 2,

$$Im'(x, y) = \begin{cases} 0, & \text{se } \sum_{\substack{i=a \\ j=b}}^{i \leq a+n-1 \\ j \leq b+n-1} Im(i, j) \geq 0.50 * n^2 \\ Im(x', y'), & \text{c.c.,} \end{cases}, \quad (2)$$

onde Im representa a imagem de entrada desta etapa, Im' é a imagem resultante, os pares (x, y) e (x', y') são pontos correspondentes nas duas imagens, n a dimensão da janela, e (a, b) as coordenadas do *pixel* superior esquerdo da janela. A avaliação ocorre desse modo para evitar que pequenos objetos com cores semelhantes aos tons de pele sejam erroneamente classificados como pele. A imagem Im resultante da fase de detecção de pele é uma imagem binária em que os *pixels* brancos indicam as regiões com presença de pele, onde o detector de face atuará. Em seguida, utilizamos o algoritmo de Viola e Jones (2001) para detectar as faces na imagem em escala de cinza correspondente à região de pele.

4 Experimentos e resultados

Experimentos foram feitos com dois vídeos coloridos comumente utilizados na validação de sistemas de seguimento (Ross et al., 2007; Babenko et al., 2009b; Babenko et al., 2011; Kim et al., 2008). O primeiro vídeo, *DavidIndoor*, contém 1538 faces enquanto o segundo vídeo, *Trellis* contém 1136 faces. A dimensão de cada quadro, nos dois vídeos, é 320×240 *pixels*. Os testes foram executados em um processador *Intel Core i5 2.8 GHz*. Os vídeos foram submetidos à implementação padrão do *framework* Viola-Jones, e ao algoritmo DPM (Mathias et al., 2014) para fins de comparação. O método Viola-Jones se destaca por ser uma possibilidade em tempo real enquanto o método DPM atinge altas taxas de acerto, embora utilize um elevado tempo de

processamento. O método proposto procura alcançar um equilíbrio entre essas duas variáveis, sendo uma solução factível para um sistema em tempo real. Os resultados dos três métodos podem ser visualizados na Figura 2. As Figuras 3 e 4 exibem os resultados de cada etapa em um quadro de cada vídeo. É importante observar que a Figura 4a é a estimativa de movimento de uma cena capturada com a câmera em movimento.

As métricas que permitem a comparação dos métodos e que mostraremos aqui são a sensibilidade e a precisão. A sensibilidade, também conhecida como *recall*, mede a capacidade do sistema de detectar faces, e determinada pela Equação 3,

$$recall = VP/(VP + FN), \quad (3)$$

onde VP é o número de verdadeiros positivos, ou seja faces detectadas corretamente; e FN representa os falsos negativos, ou seja, as faces que não foram detectadas. A precisão (*precision*), ou o valor preditivo positivo, corresponde à confiabilidade da detecção da face, sendo definida como na Equação 4,

$$precision = VP/(VP + FP), \quad (4)$$

onde FP é a quantidade de falsos positivos, que são objetos classificados como face mas que não o são;

De acordo com os dados das Tabelas 1 e 2, a sensibilidade do Viola-Jones, assim como a do sistema proposto, foi inferior à da solução DPM. Observando a precisão, percebe-se que a proposta no presente trabalho é mais confiável que a do algoritmo Viola-Jones para o vídeo *Trellis*. No entanto, a abordagem DPM se mostrou bastante precisa atingindo o valor máximo 1 em ambos os vídeos, o que significa que não apresenta falsos positivos. As taxas VP e FP podem ser observadas nas Tabelas 3 e 4.

Tabela 1: Sensibilidade e precisão resultantes em cada método para o vídeo *DavidIndoor*.

Método	Sens.	Prec.
Viola-Jones (Viola e Jones, 2001)	0,85	0,96
DPM (Mathias et al., 2014)	0,90	1
Método proposto	0,85	0,96

Tabela 2: Sensibilidade e precisão resultantes em cada método para o vídeo *Trellis*.

Método	Sens.	Prec.
Viola-Jones (Viola e Jones, 2001)	0,55	0,77
DPM (Mathias et al., 2014)	0,76	1
Método proposto	0,51	0,98

Outro fator que deve ser avaliado é o tempo no qual cada método é capaz de realizar a

Tabela 3: Número de verdadeiros positivos e falsos positivos resultantes em cada método para o vídeo *DavidIndoor*.

Método	VP	FP
Viola-Jones (Viola e Jones, 2001)	1312	55
DPM (Mathias et al., 2014)	1392	0
Método proposto	1303	49

Tabela 4: Número de verdadeiros positivos e falsos positivos resultantes em cada método para o vídeo *Trellis*.

Método	VP	FP
Viola-Jones (Viola e Jones, 2001)	622	181
DPM (Mathias et al., 2014)	870	0
Método proposto	578	13

detecção. A Tabela 5 mostra o tempo médio demandado no processamento de um quadro do vídeo. Observa-se que não há significativa diferença entre o tempo de processamento obtido no Viola-Jones e no método proposto, dada a proximidade do tempo médio. Em contrapartida, a solução DPM executou em um tempo cerca de cinquenta vezes maior que as outras duas.

Tabela 5: Tempo médio de processamento por quadro (em segundos) para cada método.

Método	Tempo médio	Desv. Pad.
Viola Jones	0,1503	0,05
DPM	7,3624	0,12
Método proposto	0,1685	0,06

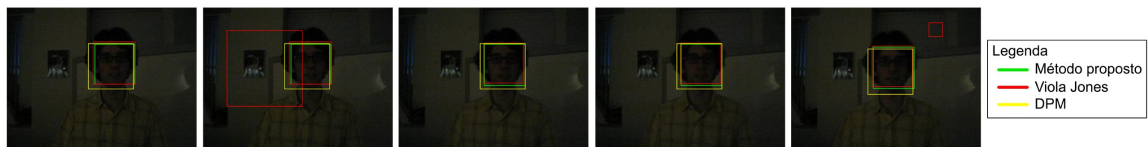
A partir dos experimentos realizados, é possível perceber que o sistema implementado com as alterações propostas apresenta resultados bastante promissores. A redução do espaço de busca através da inserção de uma etapa de regiões de movimento e de pele mostrou ser uma escolha satisfatória.

O fato de nem todas as faces serem detectadas no decorrer dos quadros se justifica pela movimentação arbitrária dos indivíduos nos vídeos, e pelas condições não ideais de iluminação e pose.

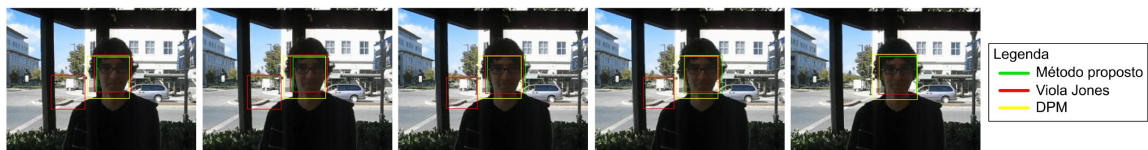
A abordagem DPM apresenta taxas de acertos superiores e ligeiramente mais confiáveis que o método proposto. No entanto, tem como desvantagem o tempo de execução, o que a torna uma estratégia inviável em um sistema de vigilância em tempo real.

5 Conclusão e trabalhos futuros

Este trabalho propõe um método de detecção de face em ambientes não controlados, auxiliado por etapas prévias de estimação de movimento e

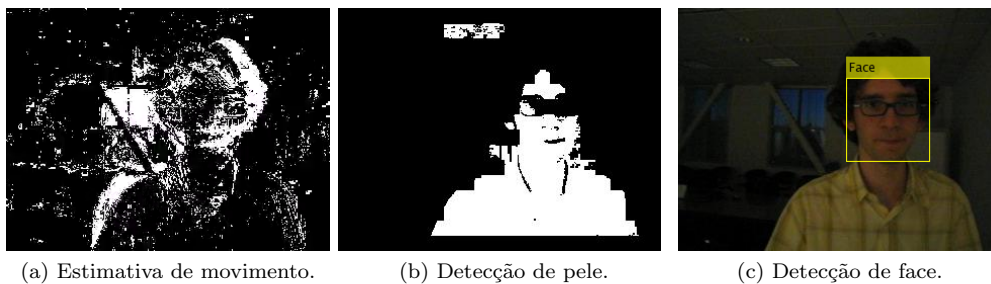


(a) Resultados da detecção de face em uma sequência de cinco quadros consecutivos do vídeo *DavidIndoor*.



(b) Resultados da detecção de face em uma sequência de cinco quadros consecutivos do vídeo *Trellis*.

Figura 2: Resultados de cada método.

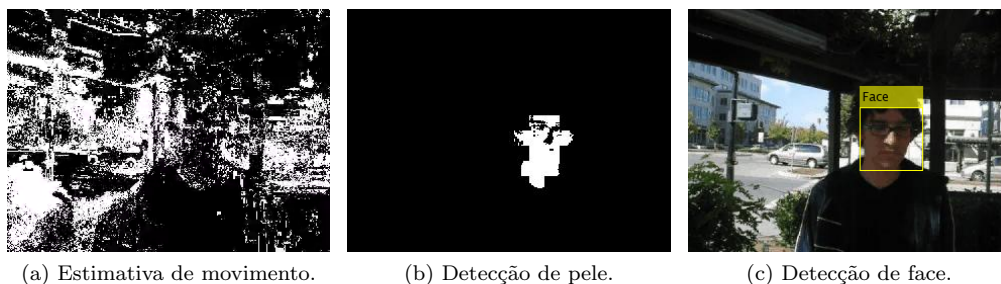


(a) Estimativa de movimento.

(b) Detecção de pele.

(c) Detecção de face.

Figura 3: Resultados de cada etapa para o vídeo *DavidIndoor*.



(a) Estimativa de movimento.

(b) Detecção de pele.

(c) Detecção de face.

Figura 4: Resultados de cada etapa para o vídeo *Trellis*.

detecção de pele. Os experimentos mostraram que a proposta é eficaz em reduzir o número de falsos positivos em comparação ao *framework* Viola-Jones. Ainda, é mais adequada para sistemas em tempo real do que a abordagem DPM.

Os resultados aqui apresentados são frutos de uma pesquisa inicial, a ser aperfeiçoada de modo a tornar o sistema mais robusto. Em trabalhos futuros serão utilizadas técnicas de predição. O objetivo de tais técnicas será recuperar faces que não estejam em uma pose adequada para a detecção frontal feita pelo algoritmo. Isso possibilitará que indivíduos não encontrados pelo método de detecção atual sejam localizados.

Referências

Babenko, B., Yang, M. H. e Belongie, S. (2009a). Visual tracking with online multiple

instance learning, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 983–990.

Babenko, B., Yang, M. H. e Belongie, S. (2009b). Visual tracking with online multiple instance learning, *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 983–990.

Babenko, B., Yang, M. H. e Belongie, S. (2011). Robust object tracking with online multiple instance learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(8): 1619–1632.

Ban, Y., Kim, S.-K., Kim, S., Toh, K.-A. e Lee, S. (2014). Face detection based on skin color likelihood, *Pattern Recogn.* **47**(4): 1573–1585.

- Dahal, B., Alsadoon, A., Prasad, P. W. C. e Elchouemi, A. (2016). Incorporating skin color for improved face detection and tracking system, *2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pp. 173–176.
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D. e Ramanan, D. (2010). Object detection with discriminatively trained part-based models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(9): 1627–1645.
- Freund, Y. e Schapire, R. (1999). A short introduction to boosting, *Journal of Japanese Society for Artificial Intelligence*.
- Heydarzadeh, Y. e Haghighat, A. T. (2011). An efficient face detection method using adaboost and facial parts, *Computer, IT and Electronic department Azad University of Qazvin Tehran, Iran*.
- Kalal, Z., Mikolajczyk, K. e Matas, J. (2010). Face-tld: Tracking-learning-detection applied to faces, *2010 IEEE International Conference on Image Processing*, pp. 3789–3792.
- Kim, M., Kumar, S., Pavlovic, V. e Rowley, H. (2008). Face tracking and recognition with visual constraints in real-world videos, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8.
- Li, S. Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H. e Shum, H. (2002). Statistical learning of multi-view face detection, *In Proceedings of the 7th European Conference on Computer Vision*, pp. 67–81.
- Lienhart, R. e Maydt, J. (2002). An extended set of haar-like features for rapid object detection, *Image Processing. 2002. Proceedings. 2002 International Conference on*, Vol. 1, pp. I-900–I-903 vol.1.
- Lima, K. A. B., dos Reis, F. W. P. e Aires, K. R. T. (2015). Aprimoramento do modelo de mistura de gaussianas para segmentação de veículos utilizando abordagem de limiarização local, *Simpósio Brasileiro de Automação Inteligente*.
- Liuliu, W. e Mingyang, L. (2016). Multi-pose face detection research based on adaboost, *2016 Eighth International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 409–412.
- Luh, G. C. (2014). Face detection using combination of skin color pixel detection and viola-jones face detector, *2014 International Conference on Machine Learning and Cybernetics*, Vol. 1, pp. 364–370.
- M., P. V. e Vadivel, S. (2016). Face extraction using skin color and pca face recognition in a mobile cloudlet environment, *2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pp. 41–45.
- Mathias, M., Benenson, R., Pedersoli, M. e Van Gool, L. (2014). Face detection without bells and whistles, *ECCV*.
- Ojala, T., Pietikinen, M. e Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* **26**(1): 51–59.
- Papageorgiou, C., Oren, M. e Poggio, T. (1998). A general framework for object detection, *International Conference on Computer Vision*.
- Passarinho, C. J. P., Salles, E. O. T. e Filho, M. S. (2015). Face tracking in unconstrained color videos with the recovery of the location of lost faces, *IEEE Latin America Transactions* **13**(1): 307–314.
- Ross, D. A., Lim, J., Lin, R. S. e Yang, M. S. (2007). Incremental learning for robust visual tracking, *International Journal of Computer Vision*.
- Verma, R. C., Schmid, C. e Mikolajczyk, K. (2003). Face detection and tracking in a video by propagating detection probabilities, **25**.
- Viola, P. e Jones, M. (2001). Robust real-time object detection, *International Journal of Computer Vision* **4**.
- Viola, P. e Jones, M. J. (2004). Robust real-time face detection, *International Journal of Computer Vision* **57**(2): 137–154.
- Wu, Y. e Ai, X. (2008). Face detection in color images using adaboost algorithm based on skin color information, *Knowledge Discovery and Data Mining, 2008. WKDD 2008. First International Workshop on*, pp. 339–342.
- Xianbao, W., Wenming, C. e Guojun, L. (2005). A method of face location in complex background, *International Conference on Neural Networks and Brain*, Vol. 3.
- Zafeiriou, S., Zhang, C. e Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future, *Computer Vision and Image Understanding* **138**: 1 – 24.