

Formulários de Extração de Dados

Título do Artigo: Achieving Robust Face Recognition from Video by Combining a Weak Photometric Model and a Learnt Generic Face Invariant.

Autores: Ognjen Arandjelovic, Roberto Cipolla.

Data de Publicação: Janeiro 2013.

Veículo de Publicação: Pattern Recognition, Volume 46, Issue 1, Pages 9–23.

Fonte: ACM

Link: <http://www.sciencedirect.com/science/article/pii/S0031320312002944>

Abstract: In spite of over two decades of intense research, illumination and pose invariance remain prohibitively challenging aspects of face recognition for most practical applications. The objective of this work is to recognize faces using video sequences both for training and recognition input, in a realistic, unconstrained setup in which lighting, pose and user motion pattern have a wide variability and face images are of low resolution. The central contribution is an illumination invariant, which we show to be suitable for recognition from video of loosely constrained head motion. In particular there are three contributions: (i) we show how a photometric model of image formation can be combined with a statistical model of generic face appearance variation to exploit the proposed invariant and generalize in the presence of extreme illumination changes; (ii) we introduce a video sequence “re-illumination” algorithm to achieve fine alignment of two video sequences; and (iii) we use the smoothness of geodesically local appearance manifold structure and a robust same-identity likelihood to achieve robustness to unseen head poses. We describe a fully automatic recognition system based on the proposed method and an extensive evaluation on 323 individuals and 1474 video sequences with extreme illumination, pose and head motion variation. Our system consistently achieved a nearly perfect recognition rate (over 99.7% on all four databases).

Resumo: De acordo com o texto, o reconhecimento facial automático têm se estabelecido como uma das maiores áreas de pesquisa ativa em computação visual. Em ambientes com condições controladas foi relatada uma taxa de erro de 10% ou mais, enquanto em ambientes não controlados o desempenho é ainda pior. Os autores

acreditam que a principal razão desta diferença está principalmente na maioria dos algoritmos que até funcionam em ambientes controlados, mas não são satisfatórios em ambientes não controlados. Comparado com o reconhecimento em apenas uma imagem, o reconhecimento facial em sequência de imagens (video) é uma área de pesquisa relativamente nova (2013). Variação de luminosidade possivelmente é o desafio mais significativo para o reconhecimento facial automático em sequências de imagens. Uma das ideias-chaves do projeto é a “*re-iluminação*” de sequências de video. O objetivo para isto é, pegar duas sequências de faces de entrada e produzir uma terceira, “sintética”, que contem as mesmas poses da primeira com a iluminação da segunda. Métodos de avaliação foram testados em 4 bases de dados, contendo um total de 323 indivíduos e 1.474 sequências, totalizando 117.271 faces detectadas automaticamente.

Contribuições: Mostrar como um modelo fotométrico de formação de imagem pode ser combinado com um modelo estatístico de face genérica com variação de aparência e introduzir um algoritmo de “*re-iluminação*” em sequências de imagens.

Técnicas Utilizadas: Algoritmo proposto pelo trabalho, Algoritmo de “*re-iluminação*” proposto, Floyd's Algorithm, Expectation Maximization Algorithm, Gaussian Mixture Model (GMM), cascaded detector of Viola and Jones.

Base(s) de Imagens Utilizada(s): *CamFace*, *ToshFace*, *FaceVideo* (gratuita) e *Faces96* (gratuita).

Resolução das imagens faciais obtidas: 45x45, 60x60 e 80x80 pixels.

Variáveis Consideradas: Variação de pose, variação de iluminação, ambiente real e imagens de baixa resolução.

Resultados Encontrados: O algoritmo proposto superou outros métodos do estado da arte com um reconhecimento facial quase perfeito (aproximadamente 99.7%) em todas as 4 bases de imagens de teste. Esta é uma taxa de reconhecimento facial extremamente alta considerando os problemas enfrentados, tais como variação de pose e de iluminação. Assim concluiu-se que o algoritmo proposto é muito promissor para o

ambiente proposto pelo trabalho.

Comentários Adicionais: Apesar de o trabalho tratar de variáveis relevantes como variação de iluminação e de pose, as bases de dados utilizadas para treinamento e testes parecem ser geralmente de imagens frontais apesar de ter alguma variação de pose geralmente as faces se encontram de frente para a câmera, o que difere bastante do trabalho proposto, sendo que serão utilizadas câmeras de segurança e os indivíduos geralmente não estarão de frente para as mesmas.

Referências Relevantes:

- [1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, *ACM Computing Surveys* 35 (4) (2004) 399–458.
- [7] G. Shakhnarovich, J.W. Fisher, and T. Darrel. Face recognition from long-term observations, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 3, May–June 2002, pp. 851–868.
- [14] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 19 (7) (1997) 721–732.
- [16] M. Everingham, A. Zisserman, Automated person identification in video, in: *Proceedings of the IEEE International Conference on Image and Video Retrieval CIVR*, 2004, pp. 289–298.
- [28] O. Arandjelovic´, R. Cipolla, An illumination invariant face recognition system for access control using video, in: *Proceedings of the IAPR British Machine Vision Conference (BMVC)*, September 2004, pp. 537–546.
- [34] M. Bichsel, A.P. Pentland, Human face recognition and the face image set's topology, *Computer Vision, Graphics and Image Processing: Image Understanding* 59 (2) (1994) 254–261.
- [36] O. Arandjelovic´, A. Zisserman. Automatic face recognition for film character retrieval in feature-length films, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, June 2005, pp. 860–867.
- [48] O. Arandjelovic´, R. Cipolla, A pose-wise linear illumination manifold model for face recognition using video, *Computer Vision and Image Understanding (CVIU)* 113 (January(1)) (2009) 113–125.
- [54] P. Viola, M. Jones, Robust real-time face detection, *International Journal of*

Computer Vision (IJCV) 57 (May(2)) (2004) 137–154.

[62] H. Wang, S.Z. Li, Y. Wang, Face recognition under varying lighting conditions using self quotient image, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR), May 2004, pp. 819–824.

Título do Artigo: A pose-wise linear illumination manifold model for face recognition using video.

Autores: Ognjen Arandjelovic, Roberto Cipolla.

Data de Publicação: Janeiro 2009.

Veículo de Publicação: Computer Vision and Image Understanding, Volume 113, Issue 1, Pages 113–125.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S107731420800115X>

Abstract: The objective of this work is to recognize faces using video sequences both for training and novel input, in a realistic, unconstrained setup in which lighting, pose and user motion pattern have a wide variability and face images are of low resolution. There are three major areas of novelty: (i) illumination generalization is achieved by combining coarse histogram correction with fine illumination manifold-based normalization; (ii) pose robustness is achieved by decomposing each appearance manifold into semantic Gaussian pose clusters, comparing the corresponding clusters and fusing the results using an RBF network; (iii) a fully automatic recognition system based on the proposed method is described and extensively evaluated on 600 head motion video sequences with extreme illumination, pose and motion pattern variation. On this challenging data set our system consistently demonstrated a very high recognition rate (95% on average), significantly outperforming state-of-the-art methods from the literature.

Resumo: Segundo os autores o reconhecimento facial automático está crescendo muito em diversas áreas de pesquisa, porém vários algoritmos desenvolvidos não alcançam uma performance satisfatória em um ambiente real. O fator da variação de iluminação é um dos fatores mais desafiadores no reconhecimento facial em ambientes reais. Pode-

se observar que em ambientes reais a movimentação e variação de poses não são lineares. No método proposto pelo trabalho variações de movimentação de face são representadas por conjuntos de imagens aglomeradas, isto é vantajoso para normalizar os “frames” de entrada o máximo possível. Em um primeiro passo são removidas partes da imagem de entrada que não contém características de interesse, ao final é removido o fundo da imagem, restando apenas o rosto, o qual é redimensionado para uma escala uniforme de tamanho 30x30 pixels. Foi utilizado o Gamma Intensity Correction (GIC) para compensar mudanças de brilho. A iluminação de cada imagem do “cluster” é normalizada. Para estabelecer uma performance o algoritmo proposto foi comparado com: Kernel Principal Angles (KPA) de Wolf e Shashua [12], Mutual Subspace Method (MSM) de Fukui e Yamaguchi [11], “KL divergence-based algorithm” de Shakhnarovich et al. (KLD) [10] e “Majority vote across all pairs of frames using Eigenfaces” (MVE) de Turk e Pentland [42]. Os resultados encontrados no reconhecimento facial nos testes foi muito satisfatório, com uma taxa média de 95,2%.

Contribuições: Generalização da iluminação e um sistema automático de reconhecimento facial utilizando métodos propostos no trabalho.

Técnicas Utilizadas: Gaussian pose clusters, utilizam um detector de pupila e narina proposto por Fukui and Yamaguchi, k-means clustering algorithm, Mahalanobis distance, Principal Component Analysis (PCA), Euclidean distance, Radial Basis Functions (RBF).

Base(s) de Imagens Utilizada(s): FaceDB60.

Resolução das imagens faciais obtidas: 30x30 pixels.

Variáveis Consideradas: Variação de iluminação, variação de pose e padrão de movimento e baixa resolução das imagens faciais.

Resultados Encontrados: O método proposto se mostrou bastante eficiente nos testes, alcançando uma taxa média de 95,2%, com desvio padrão médio de apenas 4,7%. Através disto os autores concluíram que o método proposto foi bem sucedido no reconhecimento facial em um ambiente com variação de luminosidade e movimentação.

Comentários Adicionais: São citadas duas abordagens para o problema de variação de luminosidade em uma única imagem, são elas, “illumination cones” e “3D morphable model”. Na seção 2.3 são citados algumas classes de algoritmos de reconhecimento facial. Apesar de as imagens utilizadas para treinamento e testes possuírem variação de pose e luminosidade, pode-se dizer que as mesmas ainda são relativamente boas para realizar o reconhecimento facial.

Referências Relevantes:

- [4] W.A. Barrett, A survey of face recognition algorithms and testing results, *Systems and Computers* 1 (1998) 301–305.
- [5] T. Fromherz, P. Stucki, M. Bichsel, A survey of face recognition, MML Technical Report, (97.01), 1997.
- [27] X. Liu, T. Chen, Video-based face recognition using adaptive hidden Markov models, in: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2003, pp. 340–345.
- [42] M. Turk, A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3 (1) (1991) 71–86.
- [45] O. Arandjelovic', R. Cipolla, A new look at filtering techniques for illumination invariance in automatic face recognition, in: *Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, April 2006, pp. 449–454.

Título do Artigo: Face Recognition from Video Using the Generic Shape-illumination Manifold.

Autores: Ognjen Arandjelovic e Roberto Cipolla.

Data de Publicação: 2006.

Veículo de Publicação: Proceedings of the 9th European Conference on Computer Vision - Volume Part IV (Berlin, Heidelberg, 2006), 27–40.

Fonte: ACM.

Link: http://link.springer.com/chapter/10.1007%2F11744085_3

Abstract: In spite of over two decades of intense research, illumination and pose

invariance remain prohibitively challenging aspects of face recognition for most practical applications. The objective of this work is to recognize faces using video sequences both for training and recognition input, in a realistic, unconstrained setup in which lighting, pose and user motion pattern have a wide variability and face images are of low resolution. In particular there are three areas of novelty: (i) we show how a photometric model of image formation can be combined with a statistical model of generic face appearance variation, learnt offline, to generalize in the presence of extreme illumination changes; (ii) we use the smoothness of geodesically local appearance manifold structure and a robust same-identity likelihood to achieve invariance to unseen head poses; and (iii) we introduce an accurate video sequence “reillumination” algorithm to achieve robustness to face motion patterns in video. We describe a fully automatic recognition system based on the proposed method and an extensive evaluation on 171 individuals and over 1300 video sequences with extreme illumination, pose and head motion variation. On this challenging data set our system consistently demonstrated a nearly perfect recognition rate (over 99.7% on all three databases), significantly out-performing state-of-the-art commercial software and methods from the literature.

Resumo: Após duas décadas de pesquisa intensa, iluminação e pose continuam sendo aspectos desafiadores para a maioria das aplicações de reconhecimento facial. O objetivo deste trabalho é reconhecer faces em sequências de vídeos, em um ambiente realístico e não controlado, com variação de luminosidade, pose e padrão de movimento, e onde as imagens são de baixa resolução. Através disto, os autores descrevem um sistema de reconhecimento totalmente automático, com base no método proposto. A abordagem consiste em usar um modelo fotométrico fraco de formação de imagem com uma máquina de aprendizagem para a modelagem de múltiplas faces. Especificamente os autores mostram que os efeitos combinados da face e iluminação pode ser efetivamente aprendido usando PCA probabilístico (PPCA) a partir de um pequeno conjunto de sequências de vídeos, com diferentes condições de iluminação, enquanto um algoritmo de 're-iluminação' é utilizado para fornecer mais robustez. Através dos dados testados, o sistema demonstrou uma taxa de reconhecimento quase perfeita (acima de 99,7%) superando os softwares e métodos do estado da arte.

Contribuições: Mostrar como um modelo fotométrico de formação de imagem pode ser combinado a um modelo estatístico de variação de aparência facial, com o objetivo de

generalizar na presença de alterações extremas de iluminação. Apresentar um algoritmo de 're-iluminação' para sequências de vídeos, com o objetivo de aumentar a robustez.

Técnicas Utilizadas: PCA probabilístico (PPCA), Euclidean distance, Gaussian Mixture Model (GMM).

Base de Imagens Utilizada: FaceDB100, FaceDB60 e FaceVideoDB.

Resolução das imagens faciais obtidas: 60 pixels, 60 pixel e 45 pixel (de acordo com as bases de imagens)

Variáveis Consideradas: Variação de luminosidade, de pose e movimentação.

Resultados Encontrados: O algoritmo proposto superou significativamente outros métodos do estado da arte, alcançando uma taxa de reconhecimento quase perfeita (acima de 99,7%) em todos os 3 bancos de dados, contendo variação de luminosidade, pose e movimento.

Comentários Adicionais: O trabalho é muito útil para o estado da arte, pois apresenta um algoritmo de 're-iluminação', o qual poderá ser utilizado futuramente no trabalho proposto.

Referências Relevantes: Nenhuma referência relevante nova a declarar.

Título do Artigo: Thermal and reflectance based personal identification methodology under variable illumination.

Autores: Ognjen Arandjelovic, Riad Hammoud e Roberto Cipolla.

Data de Publicação: Maio de 2010.

Veículo de Publicação: Pattern Recognition.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S0031320309004427>

Abstract: The objective of this work is to authenticate individuals based on the appearance of their faces. This is a difficult pattern recognition problem because facial appearance is generally greatly affected by the changes in the way a face is illuminated, by the camera viewpoint and partial occlusions, for example due to eye-wear. We describe a fully automatic algorithm that systematically addresses each of these challenges. The main novelty is an algorithm for decision-level fusion of two types of imagery: one acquired in the visual and one acquired in infrared electromagnetic spectrum. Specifically: we examine: (i) the effects of preprocessing of data in each domain, (ii) the fusion of holistic and local facial appearance, and (iii) propose an algorithm for combining the similarity scores in visual and thermal spectra in the presence of prescription glasses and significant pose variations, using a small number of training images (5–7). Our system achieved a high correct identification rate of 97% on a freely available data set containing extreme illumination changes.

Resumo: Variação de poses da cabeça e iluminação se mostram os maiores desafios no reconhecimento facial. Neste trabalho os autores consideram o paradigma de harmonização em grupos de imagens. Neste caso, o contraste de pose e mudanças de iluminação são muito difíceis de tratar por algoritmos de reconhecimento facial. Espectro de imagens térmicas são muito úteis pois são praticamente insensíveis a mudanças de iluminação, por outro lado, não tem muitas informações faciais do indivíduo. Nestes sentidos, a ideia principal deste trabalho é que a fusão destas duas modalidades irá melhorar a robustez do reconhecimento facial com a presença de variação de luminosidade. Um grande número de algoritmos foram propostos para resolver o problema de variação de iluminação, os quais são apresentados em dois grupos: métodos baseado em modelo e baseado em aparência. Muitos estudos sugerem que o reconhecimento facial em espectros térmicos oferece algumas vantagens distintas, como por exemplo a invariância para mudanças de iluminação. Muitas tentativas bem sucedidas foram feitas para fundir modalidades visuais e térmicas para melhorar a performance do reconhecimento facial. O sistema proposto consiste em três módulos: (i) processamento e registro de dados, (ii) detecção de óculos e (iii) fusão da representação das faces holística e local usando modalidades visuais e térmicas. Primeiramente são detectados os olhos e a boca das faces nas imagens, e então as mesmas são “recortadas” restando apenas uma imagem facial de tamanho 80x80 pixels. Os óculos de

lente são apresentados em imagens termais como 'manchas' escuras, no sistema proposto isto é utilizado para auxiliar no reconhecimento, detectando a presença de óculos na face. Ao final o sistema demonstrou uma taxa alta (97%) de reconhecimento facial.

Contribuições: Algoritmo para a fusão de imagens, examinar efeitos do pré-processamento de dados no domínio do trabalho, fusão de holística e local e propor um algoritmo para combinar pontuações de similaridade visual e térmica na presença de óculos e variação significativa de pose.

Técnicas Utilizadas: Algoritmo proposto, Análise de componentes principais (PCA).

Base de Imagens Utilizada: Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS) (gratuita).

Resolução das imagens faciais obtidas: 80x80 pixels.

Variáveis Consideradas: Variação de pose, variação de iluminação e uso de óculos.

Resultados Encontrados: Como proposto o sistema utiliza as duas modalidades que são fundidas, onde uma complementa a outra, e proporcionar uma boa invariância de iluminação. O método proposto atingiu uma taxa de reconhecimento alta (97%) na base de dados utilizada.

Comentários Adicionais: Citam também um sistema de reconhecimento facial, Facelt, utilizado no estado da arte. Levam em consideração o uso de óculos, o que pode afetar significativamente o reconhecimento facial.

Referências Relevantes:

[1] O. Arandjelovic, R. Cipolla, Face recognition from video using the generic shape-illumination manifold, in: Proceedings of European Conference on Computer Vision (ECCV), vol. 4, 2006, pp. 27–40.

[10] J.R. Price, T.F. Gee, Towards robust face recognition from video, in: Proceedings of the Applied Image Pattern Recognition Workshop, Analysis and Understanding of Time

Varying Imagery, 2001, pp. 94–102.

[13] O. Yamaguchi, K. Fukui, K. Maeda, Face recognition using temporal image sequence, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR), vol. (10), 1998, pp. 318–323.

Título do Artigo: Part-based spatio-temporal model for multi-person re-identification.

Autores: A. Bedagkar-Gala e Shishir K. Shah.

Data de Publicação: Outubro 2012.

Veículo de Publicação: Pattern Recognition Letters, 33, 14, 1908–1915.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S0167865511002777>

Abstract: In this paper we propose an adaptive part-based spatio-temporal model that characterizes person's appearance using color and facial features. Face image selection based on low level cues is used to select usable face images to build a face model. Color features that capture the distribution of colors as well as the representative colors are used to build the color model. The model is built over a sequence of frames of an individual and hence captures the characteristic appearance as well as its variations over time. We also address the problem of multiple person re-identification in the absence of calibration data or prior knowledge about the camera layout. Multiple person re-identification is a open set matching problem with a dynamically evolving and open gallery set and an open probe set. Re-identification is posed as a rectangular assignment problem and is solved to find a bijection that minimizes the overall assignment cost. Open and closed set re-identification is tested on 30 videos collected with nine non-overlapping cameras spanning outdoor and indoor areas, with 40 subjects under observation. A false acceptance reduction scheme based on the developed model is also proposed.

Resumo: O rastreamento de pessoas em uma ou mais câmeras é um problema importante e relevante a muitas aplicações. O rastreamento em múltiplas câmeras requer a habilidade de 're-identificar' pessoas que por alguma razão possam ter saído do 'campo de visão' da câmera. Os autores propõe um modelo que combina cor e características faciais. O modelo inclui ou exclui características faciais com base na

disponibilidade das partes da imagem. Na ausência de boas imagens da face, o modelo se baseia apenas nos recursos de cores. A primeira vez que uma pessoa é visualizada, o seu modelo de aparência é registrado no conjunto de galerias, o qual é um conjunto de IDs de pessoas já visualizadas anteriormente. As cores são uma sugestão muito expressiva e poderosa para reconhecer objetos e isto é utilizado no trabalho para caracterizar a aparência. Além dos outros modelos corporais extraídos da imagem, como braços e pernas, por exemplo, a imagem da cabeça também é extraída da imagem e através desta são extraídas as características faciais. Todas as imagens de cabeça são redimensionadas para 24x20 pixels. As imagens de região facial são então vetorizadas e utilizadas para a extração das características faciais. Os testes são realizados comparando o método de reconhecimento com cores, com o reconhecimento facial e com o SDALF model (estado da arte).

Contribuições: Propor um modelo adaptativo espaço-temporal, com base na cor e características faciais, para resolver o problema de 're-identificação' com múltiplas câmeras.

Técnicas Utilizadas: Histogram of Oriented Gradients (HOG), Active Color Model (ACM), Munkres algorithm, SDALF model (comparação).

Base de Imagens Utilizada: Uma rede de câmeras próprias instaladas em um edifício geraram as imagens utilizadas.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Variação de luminosidade, pontos de visão das câmeras, variação de pose e oclusões parciais.

Resultados Encontrados: Os experimentos foram realizados em um ambiente real, tanto externo quanto interno, com a utilização de 9 câmeras. Os modelos apresentados baseado em cores e características faciais se mostraram mais eficazes em comparação com o modelo SDALF (estado da arte), tanto em ambiente externo quanto interno.

Comentários Adicionais: Não utiliza apenas reconhecimento facial, mas também

detecta outras partes do corpo, como cabeça, braços e pernas. Achei este trabalho muito interessante, pois tem vários pontos semelhantes ao trabalho que pretendo desenvolver, o qual envolve o reconhecimento facial com múltiplas câmeras, com várias pessoas, em um ambiente real, com variação de luminosidade e de pose.

Referências Relevantes:

Ao, M., Yi, D., Lei, Z., Li, S., 2009. Face recognition at a distance: System issues. In: Tistarelli, M., Li, S., Chellappa, R. (Eds.), Handbook of Remote Biometrics. Springer, London, Chapter 6.

Belhumeur, P., Hespanha, J., Kriegman, D., 1997. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence Special Issue on Face Recognition.

Título do Artigo: Distinctiveness of Faces: A Computational Approach.

Autores: Manuele Bicego, Enrico Grosso, Andrea Lagorio, Gavin Brelstaff, Linda Brodo e Massimo Tistarelli.

Data de Publicação: Maio 2008.

Veículo de Publicação: Transactions on Applied Perception.

Fonte: ACM.

Link: <http://dl.acm.org/citation.cfm?doid=1279920.1279925>

Abstract: This paper develops and demonstrates an original approach to face-image analysis based on identifying distinctive areas of each individual's face by its comparison to others in the population. The method differs from most others—that we refer as unary — where salient regions are defined by analyzing only images of the same individual. We extract a set of multiscale patches from each face image before projecting them into a common feature space. The degree of “distinctiveness” of any patch depends on its distance in feature space from patches mapped from other individuals. First a pairwise analysis is developed and then a simple generalization to the multiple-face case is proposed. A perceptual experiment, involving 45 observers, indicates the method to be fairly compatible with how humans mark faces as distinct. A quantitative example of face

authentication is also performed in order to show the essential role played by the distinctive information. A comparative analysis shows that performance of our n-ary approach is as good as several contemporary unary, or binary, methods, while tapping a complementary source of information. Furthermore, we show it can also provide a useful degree of illumination invariance.

Resumo: Alguns autores citam que características anatômicas, como nariz, olhos e boca, são as áreas mais distintas, ou seja, geralmente as mais utilizadas para o reconhecimento facial. Porém, o autor do artigo cita que o espaço ao redor dos olhos, nariz e boca, podem ser como, ou até mais característicos. Através disto, os autores propõe uma nova abordagem, computando as áreas faciais de um indivíduo que parecem distintas de outras faces selecionadas de uma população. Esta abordagem é conceitualmente diferente das demais existentes, que detectam e analisam partes específicas de uma face para o propósito de autenticação ou reconhecimento. Basicamente os autores, selecionam áreas de uma face que parecem mais distintas de outras, e então comparam estas áreas distintas com outras áreas distintas extraídas de outras imagens, podendo assim realizar o reconhecimento facial. Primeiramente são extraídas e codificadas, em “vetores de características”, as possíveis áreas distintas de cada face. Posteriormente é indicado como os pares de faces são analisados. Finalmente são explicados alguns exemplos que motivam refinamentos no processo. Os autores começaram considerando apenas 2 faces apenas, para após isto, examinar como expandir e processar isto para o caso de múltiplas faces em uma imagem. As imagens utilizadas foram todas em escala de cinza, e de dimensões 320x200 pixels, e recortadas para reduzir a influência do fundo da imagem.

Contribuições: Demonstrar o poder discriminativo de informações extraídas da face. Propor uma alternativa para a análise facial baseada em identificar áreas faciais distintas em comparação com outras da população.

Técnicas Utilizadas: Laplacian of Gaussian (LoG), Standard Histogram Equalization, Eigenfaces, “Whitened Angle Distance”, Principal Component Analysis (PCA).

Base de Imagens Utilizada: BANCA database e XM2VTS database.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Variação de iluminação.

Resultados Encontrados: Através do teste de autenticação realizado e comparado com outros 4 métodos desenvolvidos por outros autores, o método proposto pelo artigo teve uma média da taxa de erro ponderada relativamente boa (10.93%) em comparação com os demais métodos que ficam entre as taxas de 10.15% e 11.87%. Assim sendo, isto confirma que a metodologia apresentada é válida e complementa como uma alternativa para as técnicas já existentes.

Comentários Adicionais: Achei este trabalho muito interessante, pois utiliza áreas distintas das faces para comparar com outras faces, diferente de outros métodos que utilizam geralmente olhos, boca e nariz, o modelo proposto também utiliza as regiões em torno dos olhos, nariz e boca, bem como outros pontos do rosto e do até mesmo do cabelo.

Referências Relevantes:

B ICEGO , M., G ROSSO , E., AND T ISTARELLI , M. 2005. On finding differences between faces. In Audio- and Video-Based Biometric Person Authentication, T. Kanade, A. Jain, and N. Ratha, Eds. Vol. LNCS 3546. Springer, New York. 329–338.

G ROSS , R. AND B RAJOVIC , V. 2003. An image preprocessing algorithm for illumination invariant face recognition. In Audio- and video-based biometric person authentication, J. Kittler and M. Nixon, Eds. Vol. LNCS 2688. 10–18.

M ING -H SUAN , Y., K RIEGMAN , D., AND A HUJA , N. 2002. Detecting faces in images: a survey. IEEE Trans. on Pattern Analysis and Machine Intelligence 24, 3458.

Título do Artigo: Boosted human re-identification using Riemannian manifolds.

Autores: Sławomir Bąk, Etienne Corvée, Francois Brémont e Monique Thonnat.

Data de Publicação: Junho 2012.

Veículo de Publicação: Image and Vision Computing 30 (2012) 443–452.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S0262885611001065>

Abstract: This paper presents an appearance-based model to address the human re-identification problem. Human re-identification is an important and still unsolved task in computer vision. In many systems there is a requirement to identify individuals or determine whether a given individual has already appeared over a network of cameras. The human appearance obtained in one camera is usually different from the ones obtained in another camera. In order to re-identify people a human signature should handle difference in illumination, pose and camera parameters. The paper focuses on a new appearance model based on Mean Riemannian Co-variance (MRC) patches extracted from tracks of a particular individual. A new similarity measure using Riemannian manifold theory is also proposed to distinguish sets of patches belonging to a specific individual. We investigate the significance of MRC patches based on their reliability extracted during tracking and their discriminative power obtained by a boosting scheme. Our method is evaluated and compared with the state of the art using benchmark video sequences from the ETHZ and the i-LIDS datasets. Re-identification performance is presented using a cumulative matching characteristic (CMC) curve. We demonstrate that the proposed approach outperforms state of the art methods. Finally, the results of our approach are shown on two further and more pertinent datasets.

Resumo: A “re-identificação” humana pode ser definida como identificar novamente uma pessoa que já foi identificada por outra câmera em uma rede de câmeras. Os autores citam a “re-identificação” em ambientes de câmeras de segurança, em grandes sistemas, como por exemplo, aeroportos, metrô, entre outros. Através disto uma boa opção para a “re-identificação” seria as roupas que a pessoa está usando. Quando a “re-identificação” diz respeito a um grande número de pessoas é preciso fornecer uma assinatura distinta e invariável. Modelos baseados em aparência são divididos em dois grandes grupos: abordagem baseada em uma imagem e abordagem baseada em múltiplas imagens. Através disto, os autores propõem um novo modelo para o problema da “re-identificação” humana baseado na média de covariância Riemannian. Os experimentos realizados com as bases de dados i-LIDS e ETHZ mostram que é possível montar uma assinatura robusta que pode resolver o problema da “re-identificação”.

Contribuições: Propõe um novo modelo de aparência baseado na média de covariância Riemannian, propõe uma nova medida de similaridade usando a teoria de Riemannian para distinguir conjuntos de correções pertencentes a um indivíduo específico, e extrair novos conjuntos de indivíduos dos dados i-LIDS para investigar com mais cuidado vantagens da utilização de resultados de rastreamento na construção das assinaturas.

Técnicas Utilizadas: Mean Riemannian Co-variance (MRC), Histogram of Oriented Gradient (HOG), Histogram Equalization, Pearson Product–Moment Correlation Coefficients (PMCC).

Base de Imagens Utilizada: i-LIDS dataset e ETHZ database.

Resolução das imagens faciais obtidas: 16x16 e 32x32 pixels.

Variáveis Consideradas: Iluminação, poses e parâmetros das câmeras.

Resultados Encontrados: Os experimentos mostraram que a abordagem construiu uma assinatura humana robusta, que pode ser relevante mesmo com diferenças nos parâmetros das câmeras.

Comentários Adicionais: Apesar de o trabalho tratar de identificação em múltiplas câmeras, a identificação e/ou re-identificação não utiliza o reconhecimento facial.

Referências Relevantes:

[3] M. Bauml, K. Bernardin, M. Fischer, H.K. Ekenel, R. Stiefelhagen, Multi-pose face recognition for person retrieval in camera networks, Proceedings of the 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS, 2010.

[5] R. Chellappa, A.K. Roy-Chowdhury, A. Kale, Human identification using gait and face, Proceedings of the 20th Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, 2007, pp. 1–2.

Título do Artigo: Dynamic multi-objective evolution of classifier ensembles for video face recognition.

Autores: Jean-Francois Connolly, Eric Granger e Robert Sabourin.

Data de Publicação: Junho 2013.

Veículo de Publicação: Applied Soft Computing 13 (2013) 3149–3166.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S1568494612003961>

Abstract: Due to a limited control over changing operational conditions and personal physiology, systems used for video-based face recognition are confronted with complex and changing pattern recognition environments. Although a limited amount of reference data is initially available during enrollment, new samples often become available over time, through re-enrollment, post analysis and labeling of operational data, etc. Adaptive multi-classifier systems (AMCSs) are therefore desirable for the design and incremental update of facial models. For real time recognition of individuals appearing in video sequences, facial regions are captured with one or more cameras, and an AMCS must perform fast and efficient matching against the facial model of individual enrolled to the system. In this paper, an incremental learning strategy based on particle swarm optimization (PSO) is proposed to efficiently evolve heterogeneous classifier ensembles in response to new reference data. This strategy is applied to an AMCS where all parameters of a pool of fuzzy ARTMAP (FAM) neural network classifiers (i.e., a swarm of classifiers), each one corresponding to a particle, are co-optimized such that both error rate and network size are minimized. To provide a high level of accuracy over time while minimizing the computational complexity, the AMCS integrates information from multiple diverse classifiers, where learning is guided by an aggregated dynamical niching PSO (ADNPSO) algorithm that optimizes networks according both these objectives. Moreover, pools of FAM networks are evolved to maintain (1) genotype diversity of solutions around local optima in the optimization search space and (2) phenotype diversity in the objective space. Accurate and low cost ensembles are thereby designed by selecting classifiers on the basis of accuracy, and both genotype and phenotype diversity. For proof-of-concept validation, the proposed strategy is compared to AMCSs where incremental learning of FAM networks is guided through mono- and multi-objective optimization. Performance is assessed in terms of video-based error rate and resource requirements under different

incremental learning scenarios, where new data is extracted from real-world video streams (IIT-NRC and MoBo). Simulation results indicate that the proposed strategy provides a level of accuracy that is comparable to that of using mono-objective optimization and reference face recognition systems, yet requires a fraction of the computational cost (between 16% and 20% of a mono-objective strategy depending on the data base and scenario).

Resumo: O reconhecimento facial está sujeito a variações consideráveis, como por exemplo, iluminação, pose, expressão facial, orientação e oclusão, o que podem diminuir muito o desempenho durante o reconhecimento facial. O trabalho aborda desafios relacionados a concepção de sistemas adaptativos multi-classificadores para o reconhecimento facial em video, onde os modelos podem ser criados e atualizados ao longo do tempo, tornando disponíveis novos dados de referência. O trabalho tem o foco em aplicações de reconhecimento facial baseadas em video, em que podem ocorrer dois tipos de cenários: “enrollment (initial design)” e atualização de modelos faciais. O principal problema abordado é o projeto de sistemas adaptativos precisos e eficientes para realizar o reconhecimento facial em video, onde sequencias de video são utilizadas para montar o modelo facial de cada indivíduo durante a fase de aprendizagem/treinamento. Um problema bastante desafiador no reconhecimento facial em video são a baixa qualidade das imagens e tamanho que é geralmente muito pequeno. Geralmente para reduzir a ambiguidade e prover um alto nível de acurácia, aplicações de reconhecimento facial combinam informações temporais e espaciais contidas em sequencias de video. Primeiramente o sistema executa a segmentação para localizar e isolar as regiões de interesse (ROI) nas faces localizadas na imagem. A partir das regiões de interesse são extraídas características utilizadas para o rastreamento e classificação. Já as características de rastreamento podem ser a posição, velocidade, aceleração e número de rastreio atribuídos a cada ROI na cena, de modo que é possível seguir o movimento ou expressão das faces através das imagens da sequencia de video. Por fim o módulo de decisão pode integrar as respostas do módulo de rastreamento e classificação ao longo de uma sequencia de video.

Contribuições: Propor uma estratégia de aprendizagem gradual (ADNPSO) para evoluir conjuntos de classificadores em resposta a novos dados de referência em reconhecimento facial através de video.

Técnicas Utilizadas: Bayesian inference, Aggregated Dynamical Niching PSO (ADNPSO) algorithm, Adaptive multi-classifier systems (AMCSs), Viola–Jones algorithm.

Base de Imagens Utilizada: Base do Instituto de Tecnologia da Informação do Conselho Nacional de Pesquisas do Canadá (IIT-NRC), Motion of Body (MoBo - Carnegie Mellon University).

Resolução das imagens faciais obtidas: 24x24 pixels.

Variáveis Consideradas: Baixa resolução, oclusão, movimentação, expressão facial.

Resultados Encontrados: Os resultados gerais indicam que o uso de informações no espaço de busca de cada objetivo, permite a criação de classificadores que são mais precisos e com menor custo computacional. Isto resulta em conjuntos que dão uma precisão comparável ao obtido com otimização.

Comentários Adicionais: Apesar do trabalho ser muito interessante, acredito que não deveria ter sido incluso na revisão sistemática, pois foge um pouco do assunto principal do trabalho que pretendo desenvolver, não trazendo tantos fatores relevantes a ele.

Referências Relevantes:

- [10] L.-F. Chen, H.-Y. Liao, J.-C. Lin, Person identification using facial motion, in: Proceedings on Image Processing, vol. 2, 2001, pp. 677–680.
- [22] D.O. Gorodnichy, Video-based framework for face recognition in video, in: Second Workshop on Face Processing in Video in Proceedings of the Conference on Computer and Robot Vision, Victoria, Canada, May, 2005, pp. 325–344.
- [37] F. Matta, J.-L. Dugelay, Person recognition using facial video information: a state of the art, Journal of Visual Language and Computing 20 (3) (2009) 180–187.
- [53] J. Stallkamp, H. Ekenel, R. Stiefelhausen, Video-based face recognition on real-world data, in: IEEE 11th International Conference on Computer Vision 2007, ICCV 2007, vol. 1, October, 2007, pp. 1–8.
- [63] X. Zhang, Y. Gao, Face recognition across pose: a review, Pattern Recognition 42 (11) (2009) 2876–2896.

Título do Artigo: Performance Analysis for Automated Gait Extraction and Recognition in Multi-camera Surveillance.

Autores: Michela Goffredo, Imed Bouchrika, John N. Carter e Mark S. Nixon.

Data de Publicação: Outubro 2010.

Veículo de Publicação: Multimedia Tools Appl. 50, 1 (Oct. 2010), 75–94.

Fonte: ACM.

Link: <http://link.springer.com/article/10.1007%2Fs11042-009-0378-5>

Abstract: Many studies have confirmed that gait analysis can be used as a new biometrics. In this research, gait analysis is deployed for people identification in multi-camera surveillance scenarios. We present a new method for viewpoint independent markerless gait analysis that does not require camera calibration and works with a wide range of walking directions. These properties make the proposed method particularly suitable for gait identification in real surveillance scenarios where people and their behaviour need to be tracked across a set of cameras. Tests on 300 synthetic and real video sequences, with subjects walking freely along different walking directions, have been performed. Since the choice of the cameras' characteristics is a key-point for the development of a smart surveillance system, the performance of the proposed approach is measured with respect to different video properties: spatial resolution, frame-rate, data compression and image quality. The obtained results show that markerless gait analysis can be achieved without any knowledge of camera's position and subject's pose. The extracted gait parameters allow recognition of people walking from different views with a mean recognition rate of 92.2% and confirm that gait can be effectively used for subjects' identification in a multi-camera surveillance scenario.

Resumo: "Gait analysis" é um estudo sistemático da locomoção humana. No trabalho "gait analysis", também traduzido como análise de marcha, é utilizado para a identificação de pessoas em ambientes de múltiplas câmeras. O objetivo de um sistema de segurança inteligente é rastrear pessoas automaticamente através de uma rede de câmeras e detectar comportamentos anormais. É proposto então uma solução para a identificação de sujeitos em redes de câmeras, utilizando a análise de marcha.

Reconhecer indivíduos enquanto eles caminham é um desafio particular da área de pesquisa. O método proposto para a visão independente da análise de marcha, é baseado em duas fases consecutivas: estimação dos marcadores de juntas para a sequencia de imagens e retificação de ponto de vista. Combinando os resultados dos testes com os diferentes ângulos da câmera, foi alcançado uma taxa de reconhecimento de 92,2%.

Contribuições: Apresentar um novo método de análise de ponto de vista independente, que não requer calibração da câmera e funciona com uma ampla variedade de direções, propor uma solução para a identificação de sujeitos através de redes de câmeras, utilizando a análise de marcha.

Técnicas Utilizadas: Adaptive Sequential Forward Floating Selection (ASFFS) search algorithm, Euclidean distance, K-nearest neighbour (KNN).

Base de Imagens Utilizada: CASIA-B database.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Ângulo de vista da câmera.

Resultados Encontrados: Com base em um conjunto de 275 sequencias de video de sujeitos caminhando, uma alta taxa de reconhecimento de 92,2% foi alcançada por todos os pontos de vista combinados. Isto mostra que a análise de marcha pode ser utilizada de forma eficiente para a identificação independente do ponto de vista.

Comentários Adicionais: Acredito que este trabalho não trás muitas informações relevantes ao trabalho proposto, pois não trata especificamente de reconhecimento facial.

Referências Relevantes: Nenhuma referência relevante nova.

Título do Artigo: Performance Analysis for Gait in Camera Networks.

Autores: Michela Goffredo, Imed Bouchrika, John N. Carter e Mark S. Nixon.

Data de Publicação: 2008.

Veículo de Publicação: Proceedings of the 1st ACM Workshop on Analysis and Retrieval of Events/Actions and Workflows in Video Streams (New York, NY, USA, 2008), 73–80.

Fonte: ACM.

Link: <http://dl.acm.org/citation.cfm?doid=1463542.1463555>

Abstract: This paper deploys gait analysis for subject identification in multi-camera surveillance scenarios. We present a new method for viewpoint independent markerless gait analysis that does not require camera calibration and works with a wide range of directions of walking. These properties make the proposed method particularly suitable for gait identification in real surveillance scenarios where people and their behaviour need to be tracked across a set of cameras. Tests on 300 synthetic and real video sequences, with subjects walking freely along different walking directions, have been performed. Since the choice of the cameras' characteristics is a key-point for the development of a smart surveillance system, the performance of the proposed approach is measured with respect to different video properties: spatial resolution, frame-rate, data compression and image quality. The obtained results show that markerless gait analysis can be achieved without any knowledge of camera's position and subject's pose. The extracted gait parameters allow recognition of people walking from different views with a mean recognition rate of 92.2% and confirm that gait can be effectively used for subjects' identification in a multi-camera surveillance scenario.

Resumo: O trabalho implementa a análise de marcha para identificação de sujeitos em cenários de vigilância multi câmera. O método proposto torna-se particularmente adequado para a identificação da marcha em cenários reais de vigilância, onde é possível ter um conjunto de câmeras. O método proposto é baseado em duas fases consecutivas: estimativa das articulações de marcadores de uma sequência de imagens e a retificação do ponto de vista. Foram realizados testes em 300 sequências de vídeo sintéticos e reais. Os resultados obtidos mostram que a análise de marcha pode ser alcançada sem qualquer conhecimento da posição da câmera e postura do sujeito. Os

parâmetros extraídos permitem visualizar uma taxa média de reconhecimento de 92,2%, e confirma que a marcha pode ser utilizada efetivamente para identificação de sujeitos em um cenário de vigilância multi-câmera.

Contribuições: Um novo método de ponto de vista independente de análise de marcha que não requer calibração da câmera.

Técnicas Utilizadas: Adaptive Sequential Forward Floating Selection (ASFFS), Euclidean distance, K-nearest neighbour (KNN).

Base de Imagens Utilizada: CASIA database.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Múltiplas câmeras.

Resultados Encontrados: Testes nas sequências de vídeo permitiram avaliar o desempenho da abordagem proposta. Os parâmetros extraídos permitiram uma taxa média de reconhecimento de 92,2%. Os resultados obtidos mostram que a análise de marcha pode ser usada de forma eficiente para a identificação independente de vista dos sujeitos com câmeras de vídeo disponíveis no mercado.

Comentários Adicionais: O trabalho trata apenas de reconhecimento de face e não de identificação.

Referências Relevantes: Nenhuma referência relevante nova a declarar.

Título do Artigo: Combine Image Quality Fusion and Illumination Compensation for Video-based Face Recognition.

Autores: Chao Wang e Yongping Li.

Data de Publicação: Março 2010.

Veículo de Publicação: Neurocomputing 73 (2010) 1478–1490.

Fonte: ACM.

Link: <http://www.sciencedirect.com/science/article/pii/S0925231209004160>

Abstract: Illumination variation is a critical factor affecting the performance of face recognition especially in video-based face recognition. This paper addresses this problem in two aspects: (i) a novel approach is proposed to classify the illumination orientation so that the uneven illumination on face can be compensated with pertinence. Illumination direction map and plane-fit method are put forward to determine illumination orientations. Experimental results on several public available face databases show that our approach achieved considerable performance gain in contrast to other state-of-the-art methods. (ii) Image quality fusion rule is designed to reduce the influence which is caused by the degradation of facial image quality due to illumination compensation. The degradation can lower the recognition performance. Motivated by human cognitive process and combined with video features, the rule fuses the recognition result of every face video frame to opt best result. This quality fusion rule exhibits effectiveness for illumination compensation, and experimental results on the face video databases with varied illuminations demonstrated that the proposed approach achieved satisfactory recognition rate.

Resumo: Como sabe-se a variação de iluminação é um fator crítico que afeta a performance do reconhecimento facial especialmente baseado em vídeo. O trabalho apresenta o problema em dois aspectos: uma nova abordagem é proposta para classificar a orientação da luminosidade, para então poder compensar esta variação; e uma regra de fusão da qualidade de imagens é projetada para reduzir a influência causada pela degradação da qualidade da imagem facial dada pela compensação da iluminação. Muitas vezes o reconhecimento facial em sistemas reais não é muito satisfatório, geralmente pela iluminação do ambiente, pose e expressão da face, mas principalmente pela questão da variação da iluminação. Assim sendo, o foco do trabalho é eliminar os efeitos da desigualdade de iluminação das faces em vídeos. Os autores consideram que há dois passos principais para amenizar o problema da variação de iluminação: (i) eliminar ou compensar a desigualdade de iluminação na face; (ii) após o processo (i) haverá uma degradação na qualidade da imagem facial, para isso pretende-se aplicar uma fusão sobre a qualidade da imagem. Para categorizar a fonte de luz nas imagens os autores utilizam um modelo baseado no modelo de iluminação Lambert's e

na teoria do sub espaço linear de iluminação. Imagens faciais com a iluminação já compensada podem ser classificadas de acordo com a qualidade da imagem, onde é atribuído um peso a cada imagem. Quanto maior o peso, significa que a imagem tem maior contribuição no reconhecimento. Após os experimentos com bases de imagens e de vídeos foi possível comprovar que as abordagens propostas são relevantes e tem um desempenho relevante em comparação com o estado da arte.

Contribuições: A principal contribuição é amenizar/compensar o problema da variação de luminosidade no reconhecimento facial em imagem e video.

Técnicas Utilizadas: Lambert's illumination model, CANDIDE-3 model, Histogram Equalization, Gaussian low pass filter, Principal Component Analysis (PCA), Cosine Distance (CSD), Active Appearance Model (AAM), Bidimensional Regression (BDR).

Base de Imagens Utilizada: YaleB database, Yale database, AR database, IIT-NRC face video database, UO face video database e outra base de dados própria obtidas de vídeos de segurança do metrô.

Resolução das imagens faciais obtidas: 64x64, 160x120, 320x240 e 360x288 pixels.

Variáveis Consideradas: Variação da iluminação.

Resultados Encontrados: Foram comparados o algoritmo de compensação de iluminação proposto, com outros métodos do estado da arte, incluindo, LogAbout, Wavelet e outros, em todas as bases de imagens e vídeos citadas acima. De acordo com a tabela 3, nas bases de imagens de teste, o algoritmo proposto mostrou melhorar significativamente o reconhecimento facial, em comparação com os outros métodos comparados. As regras de fusão de qualidade de imagem, em bases de video com iluminação variada, mostrou-se eficaz e teve um ganho de desempenho considerável em comparação com o estado da arte.

Comentários Adicionais: Apesar de o trabalho ter o foco principal na variação da iluminação, acredito que é bastante relevante ao trabalho que pretendo desenvolver, sendo que o mesmo também terá que lidar com essa variável.

Referências Relevantes:

- [1] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: a survey, Pattern Recognition Letters 28 (14) (2007) 1885–1906.
- [3] B. Globe, Face recognition fails in Boston airport, July 2002.
- [22] Xudong Xie, Kin-Man Lam, Face recognition under varying illumination based on a 2D face shape model, Pattern Recognition 38 (2) (2005) 221–230.

Título do Artigo: Incremental Learning Patch-based Bag of Facial Words Representation for Face Recognition in Videos.

Autores: Chao Wang, Yunhong Wang, Zhaoxiang Zhang e Yiding Wang.

Data de Publicação: Outubro 2014.

Veículo de Publicação: Multimed Tools Appl (2014) 72:2439–2467.

Fonte: ACM.

Link: <http://link.springer.com/article/10.1007%2Fs11042-013-1562-1>

Abstract: Video-based face recognition is a fundamental topic in image processing and video analysis, and presents various challenges and opportunities. In this paper, we introduce an incremental learning approach to video-based face recognition which efficiently exploits the spatiotemporal information in videos. Face image sequences are incrementally clustered based on their descriptors, and the representative face images of each cluster are picked out. The incremental algorithm of creating facial visual words is applied to construct a codebook using the descriptors of the representative face images. Continuously, with the quantization of the facial visual words, each descriptor extracted from patches is converted into codes, and codes from each region are pooled together into a histogram. The representation of the face image is generated by concatenating the histograms from all regions, which is employed to perform the categorization. In the online recognition, a similarity score matrix and a voting algorithm are employed to judge a face video's identity. Recognition is performed online while face video sequence is continuous and the proposed method gives nearly realtime feedback. The proposed method achieves a 100% verification rate on the Honda/UCSD database and 82 % on

the YouTube database. Experimental results demonstrate the effectiveness and flexibility of the proposed method.

Resumo: Neste trabalho os autores introduzem uma abordagem de aprendizagem incremental para o reconhecimento facial baseado em vídeo. Sequências de imagens faciais são agrupadas com base em seus descritores, e as imagens representativas de face de cada grupo são selecionadas. Geralmente algoritmos de reconhecimento facial em vídeo fazem uso de sequências de imagens fixas para realizar o reconhecimento. Nestes sistemas, a extração de características faciais são aplicados de forma independente para cada quadro de imagem. Recentemente mais técnicas tentam explorar a informação temporal entre a consulta e bases de dados de vídeos. A informação temporal possibilita a análise das informações temporais faciais como um identificador biométrico para o reconhecimento de pessoas, e assim, tem a vantagem de ser menos sensíveis a variação na aparência facial, por exemplo, devido a diferentes condições de iluminação e alterações de expressão. Para explorar estas informações alguns sistemas aprendem continuamente para a adaptabilidade de amostras de treinamento. Os autores propõe um dicionário de “palavras visuais”, com características extraídas de uma determinada imagem facial. Através disto, os autores propõe uma aprendizagem incremental através de um algoritmo de “palavras visuais” para o reconhecimento facial em vídeo. Primeiramente, um algoritmo de aprendizado incremental é utilizado para selecionar as imagens faciais mais representativas do grupo. Um “livro” das “palavras visuais faciais” é gerado utilizando os representantes já selecionados. Imagens faciais são convertidas em códigos com a quantização das “palavras visuais faciais”. As representações finais das imagens faciais para identificação são geradas por concatenamento de todas as regiões. Deste modo é aumentada a taxa de utilização das informações espaço-temporais.

Contribuições: Os autores propõe uma aprendizagem incremental através de um algoritmo de “palavras visuais” para o reconhecimento facial em vídeo.

Técnicas Utilizadas: Adaboost algorithm, Camshift algorithm, Scale-invariant feature transform (SIFT), Multi-Block LBP (MB-LBP), Viola and Jones, Camshift.

Base de Imagens Utilizada: Honda/UCSD Video database, Youtube database.

Resolução das imagens faciais obtidas: 222x214 pixels.

Variáveis Consideradas: Variação de pose, iluminação e expressões faciais.

Resultados Encontrados: O algoritmo proposto se mostrou bastante eficiente, se mostrando superior a outros métodos populares para explorar informações espaço-temporais, até porque não coloca restrições de poses e expressões fixas sobre os sujeitos. O algoritmo atingiu uma taxa de verificação de 100% na base de dados Honda/UCSD e 82% na base de dados do Youtube.

Comentários Adicionais: As bases de dados utilizadas nos experimentos são muito interessantes, a base Honda/UCSD por ser uma base já utilizada em vários outros trabalhos e a base do Youtube, pois até o momento não foi encontrado nenhum outro trabalho que utilize esta base, e parece ser um tanto inusitada, porém interessante pelas diversas possibilidades de testes que ela pode proporcionar.

Referências Relevantes:

1. Aggarwal G, Chowdhury A, Chellappa R (2004) A system identification approach for video-based face recognition. In: Proc. ICPR, pp 175–178.
15. Kim M, Kumar S, Pavlovic V, Rowley H (2008) Face tracking and recognition with visual constraints in real-world videos. In: Proc. CVPR, pp 1–8.

Título do Artigo: Face Detection in Low-resolution Color Images.

Autores: Jun Zheng, Geovany A. Ramirez e Olac Fuentes.

Data de Publicação: 2010.

Veículo de Publicação: Proceedings of the 7th International Conference on Image Analysis and Recognition - Volume Part I (Berlin, Heidelberg, 2010), 454–463.

Fonte: ACM.

Link: http://link.springer.com/chapter/10.1007%2F978-3-642-13772-3_46

Abstract: In low-resolution images, people at large distance appear very small and we may be interested in detecting the subject's face for recognition or analysis. However, recent face detection methods usually detect face images of about 20×20 pixels or 24×24 pixels. Face detection in low-resolution images has not been explicitly studied. In this work, we studied the relationship between resolution and the automatic face detection rate with Modified Census Transform and proposed a new 12-bit Modified Census Transform that works better than the original one in low-resolution color images for object detection. The experiments show that our method can attain better results than other methods while detecting in low-resolution color images.

Resumo: O trabalho trata da detecção facial em imagens de baixa resolução, pois ainda não há estudos explícitos quem envolvem isto. Os autores propõe uma nova abordagem para o modelo Modified Census Transform, que segundo eles será melhor que o original, em imagem de baixa resolução e coloridas. A detecção facial é uma parte importante para aplicações em computação visual, pois pode envolver diversas áreas. Porém a detecção facial pode ser uma tarefa muito difícil, pois há vários fatores que podem influenciar, tais como variação de tamanho, orientação, pose, expressão facial, oclusão e condições de luminosidade. Vários trabalhos nesta área tem sido desenvolvidos, porém a maioria trata de imagens maiores que 20×20 pixels ou 24×24 pixels, ou seja, não há nenhum trabalho que trate explicitamente de detecção de faces em imagens de baixa resolução (menores que 20×20 pixels). Porém em ambientes reais de segurança, geralmente as pessoas aparecem muito pequenas, devido a distância entre a câmera e a pessoa, sendo o tamanho da resolução da face em média de 8×8 pixels. Acredita-se que esta foi a principal motivação do trabalho. Ao final o trabalho mostrou ter uma melhor taxa de detecção de faces em imagens de baixa resolução, se comparado com o modelo 9-bit Modified Census Transform, validando o objetivo do trabalho.

Contribuições: Estudar a relação entre a resolução das imagens e a detecção automática de faces avaliado com o "Modified Census Transform" e propor uma nova abordagem "12-bit Modified Census Transform", que funciona melhor que a original em imagens coloridas de baixa resolução. O "9-bit Modified Census Transform" é baseado em imagens cinza, o que funciona bem detectando faces de tamanho maior ou igual a 24×24 pixels, porém não tem um bom desempenho em imagens de mais baixa resolução, como por exemplo, 8×8 pixels. Para imagens coloridas, os autores

adicionaram 3-bits a mais para descrever informações de cor RGB para cada pixel, e com isso proporam o “12-bit Modified Census Transform” para a detecção em imagens de faces de baixa resolução.

Técnicas Utilizadas: Modified Census Transform (MCT), AdaBoost algorithm.

Base de Imagens Utilizada: Georgia Tech color frontal face database.

Resolução das imagens faciais obtidas: De 6x6 até 24x24 pixels e 150x150 pixels.

Variáveis Consideradas: Baixa resolução.

Resultados Encontrados: O modelo proposto 12-bit MCT foi testado com resoluções diferentes de imagens faciais, sendo elas, 24x24, 16x16, 8x8 e 6x6. A partir dos testes e das Tabelas 3 e 4 apresentadas no texto foi possível concluir que o modelo proposto teve uma melhor taxa de detecção com as resoluções mais baixas 6x6 e 8x8 e também teve menos “alarmes falsos”. Porém nas resoluções de 16x16 e 24x24, apesar de o modelo proposto ter menos alarmes falsos ele apresentou uma taxa de detecção um pouco mais baixa do que o 9-bit MCT. Através disto pode-se dizer que o modelo proposto alcançou seus objetivos e que é realmente válido para imagens de baixa resolução.

Comentários Adicionais: Os autores citam trabalhos de Shinji Hayashi e Osamu Hasegawa que proporam novos detectores de face que melhoraram a avaliação de 39% para 71% em imagens de tamanho 6x6 pixels. Apesar do trabalho não tratar da identificação de pessoas, mas apenas da detecção de faces em imagens, acredito que ainda assim é bastante válido como estado da arte para o trabalho que pretendo desenvolver, pois trata da detecção de rostos em imagens de baixa resolução, ou seja, ainda assim faz parte do trabalho que irei desenvolver.

Referências Relevantes:

2. Hayashi, S., Hasegawa, O.: Robust face detection for low-resolution images. Journal of Advanced Computational Intelligence and Intelligent Informatics 10 (2006) 93–101.

Título do Artigo: Colour Invariants for Machine Face Recognition.

Autores: Ognjen Arandjelovic e Roberto Cipolla.

Data de Publicação: 2008.

Veículo de Publicação: 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08, 1–8.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4813306>

Abstract: Illumination invariance remains the most researched, yet the most challenging aspect of automatic face recognition. In this paper we investigate the discriminative power of colour-based invariants in the presence of large illumination changes between training and test data, when appearance changes due to cast shadows and non-Lambertian effects are significant. Specifically, there are three main contributions: (i) we employ a more sophisticated photometric model of the camera and show how its parameters can be estimated, (ii) we derive several novel colour-based face invariants, and (iii) on a large database of video sequences we examine and evaluate the largest number of colour-based representations in the literature. Our results suggest that colour invariants do have a substantial discriminative power which may increase the robustness and accuracy of recognition from low resolution images.

Resumo: Os autores estão interessados em representações faciais à base de cores para “máquinas” de reconhecimento facial. Na área de reconhecimento facial, foi dada pouca atenção a questão das cores, porém acredita-se que é um fator relevante e que pode auxiliar no reconhecimento facial com imagens de baixa resolução. Através do uso das cores da imagem, a mudança de luminosidade pode causar efeitos dramáticos na performance do reconhecimento. Simplificando a abordagem dos autores, nas imagens faciais, regiões sem informação são consideradas e excluídas na construção do modelo de aparência, pois elas não contribuem para a pontuação de semelhança entre as sequências. Os autores classificaram os pixels não informativos caso a luminosidade fosse menor do que 3% ou maior do que 97% entre a iluminação máxima da imagem. Apesar de não encontrarem melhorias estatisticamente significativas quando utilizando o

espaço de cor RGB, outros resultados encontrado argumentam que não seria correto concluir que a partir disto as informações de cor não tem nada a acrescentar ao poder discriminativo de iluminação. A taxa de reconhecimento alcançada utilizando componentes cromáticos individuais ultrapassou significativamente o de Hue e quase igualando o desempenho em escala de cinza.

Contribuições: (i) empregar um modelo fotométrico de câmera mais sofisticado e mostrar como estes parâmetros podem ser estimados; (ii) derivar várias novas invariantes de faces à base de cores; (iii) em um grande banco de dados de sequências de vídeos os autores examinam e avaliam o maior número de representações à base de cor na literatura. Resultados obtidos utilizando imagens 'brutas' são úteis como *benchmark* para quantificar a gravidade da variação de iluminação no banco de dados. Em comparação com outro trabalho do estado da arte [12], os dados utilizados para o trabalho são muito mais desafiadores, e teve uma taxa de reconhecimento aproximadamente 25% inferior, utilizando escalas de cinza.

Técnicas Utilizadas: Viola-Jones detector, Canonical Correlations (CC), PCA.

Base de Imagens Utilizada: Banco de dados de sequências de vídeos cedidos pela Universidade de Cambridge.

Resolução das imagens faciais obtidas: 40x40 pixels.

Variáveis Consideradas: Iluminação.

Resultados Encontrados: Segundo os resultados encontrados, os autores afirmam que os efeitos das cores são realmente um fato substancial. O desempenho em um grande banco de dados com variação de iluminação, sugere que o uso de cores pode melhorar significativamente algoritmos baseados em escala de cinza.

Comentários Adicionais: Acredito que este trabalho foge um pouco do foco principal do trabalho que pretendo desenvolver, pois trata principalmente da questão das cores na imagem, através da iluminação, além disto, achei o trabalho um tanto quanto complexo.

Referências Relevantes: Nenhuma referência relevante nova.

Título do Artigo: Target-based evaluation of face recognition technology for video surveillance applications.

Autores: Dmitry Gorodnichy e Eric Granger.

Data de Publicação: Dezembro 2014.

Veículo de Publicação: IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM), 110–117.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7015451>

Abstract: This paper concerns the problem of real-time watch-list screening (WLS) using face recognition (FR) technology. The risk of flagging innocent travellers can be very high when deploying a FR system for WLS since: (i) faces captured in surveillance video vary considerably due to pose, expression, illumination, and camera interoperability; (ii) reference images of targets in a watch-list are typically of limited quality or quantity; (iii) the performance of FR systems may vary significantly from one individual to another (according to so-called “biometric menagerie” phenomenon); (iv) the number of travellers drastically exceeds the number of target people in a watch-list; and finally and most critically, (v) due to the nature of optics, images of faces captured by video-surveillance cameras are focused and sharp only over a very short period of time if ever at all. Existing evaluation frameworks were originally developed for spatial face identification from still images, and do not allow one to properly examine the suitability of the FR technology for WLS with respect to the above listed risk factors intrinsically present in any video surveillance application. This paper introduces the target-based multi-level FR performance evaluation framework that is suitable for WLS. According to the framework, Level 0 (face detection analysis) deals with the system’s ability to process low resolution faces. Level 1 (transaction-based analysis) deals with the ability to match faces in open-set problems, where target vs. non-target distributions are unbalanced. Level 2 (subject-based analysis) deals with robustness of the system to different types of target individuals. Finally, Level 3 (spatio-temporal analysis) allows one to examine the overall

FR system discrimination by means of accumulating the recognition decision confidence over a face track, which can be used for developing more robust intelligent decision-making schemes including face triaging. The results from testing a commercial state-of-art COTS FR product on a public video data-set are shown to illustrate the benefits of this framework.

Resumo: O presente trabalho introduz um *framework* de avaliação de performance para o reconhecimento facial multinível e baseado em alvo, adequado para lista de triagem. De acordo com o *framework*, o nível 0 (análise de detecção de face) envolve a habilidade de processar faces em imagens de baixa resolução. Nível 1 (análise baseada em transações) lida com o problema de corresponder faces em conjuntos de problemas abertos, onde a distribuição de 'alvo' vs. 'não alvo' não são equilibradas. Nível 2 (análise baseada em sujeito) lida com a robustez do sistema para diferentes tipos de indivíduos 'alvos'. E por fim, o nível 3 (análise espaço-temporal) permite analisar o sistema de reconhecimento facial por meio de acumulação da confiança de decisão, que pode ser utilizado para o desenvolvimento de outros sistemas de tomada de decisão inteligentes e mais robustos. Os autores citam o uso em aeroportos, onde apresentam o termo 'criminosos' para os 'alvos', e 'viajantes' para os 'não alvo'. Neste caso se o sistema foi feito para processar apenas imagens que estão em foco, então, a chance de perder um alvo individual é muito grande. Em outro caso, se o sistema utiliza todas as imagens faciais, até as que estão fora do foco e muito pequenas, então o risco de corresponder a falsamente a pessoas 'não alvo' aumenta. Na avaliação de tecnologias para aplicação em câmeras de vigilância, é proposto categorizar todas as possibilidades de cenários de acordo com o triângulo "who-what-where", através disto, os autores citam alguns tipos de cenários, tais como, controle de passaporte, corredor de uma só 'mão', salões de aeroportos, entre outros. Através de testes em diversos tipos de cenário é possível avaliar a performance do sistema de reconhecimento facial e até mesmo apresentar e tratar as possíveis vulnerabilidades destes sistemas. Duas metodologias de reconhecimento facial são possíveis para a solução do problema proposto, são elas *Cohort-based* e *Target-based*. Para os experimentos, os autores utilizam um software comercial de reconhecimento facial, segundo eles bem conhecido, o qual o nome não é divulgado.

Contribuições: Os autores propõe uma metodologia de avaliação de reconhecimento

facial multi-nível para lidar com as principais fontes de risco que estão intrinsecamente presentes em aplicações de vigilância de vídeo.

Técnicas Utilizadas: Não mencionado, até por que os experimentos são realizados com um software comercial.

Base de Imagens Utilizada: “Faces in Action” dataset, Chokepoint dataset, S2V dataset.

Resolução das imagens faciais obtidas: De 12 a 64 pixels.

Variáveis Consideradas: As seguintes variáveis são apresentadas, porém não são 'tratadas' durante o desenvolvimento do trabalho: Pose, expressão, iluminação e problemas de operabilidade interna da câmera.

Resultados Encontrados: A aplicação da metodologia foi apresentada em uma avaliação de um produto comercial de reconhecimento facial. Os testes realizados mostraram que o produto apresenta um pequeno desvio de desempenho acerca do reconhecimento de um indivíduo para outro, o que é um importante indicador da robustez do sistema, ao mesmo tempo, foram detectados algumas vulnerabilidades no sistema que precisam ser melhoradas. E ao final os autores apresentam algumas direções para melhorar o desempenho dos sistemas de reconhecimento facial em aplicativos de vigilância de vídeo.

Comentários Adicionais: O trabalho foge um pouco do foco principal do trabalho que pretendo desenvolver, o qual envolve como fim o reconhecimento facial, ao meu ver este trabalho utiliza o reconhecimento facial como meio para aplicar a metodologia, então acredito que não é tão relevante para o estado da arte.

Referências Relevantes:

J. R. Barr, K. W. Bowyer, P. J. Flynn, S. Bismas, “Face recognition from video: a review. International Journal of Pattern Recognition and Artificial Intelligence, April 20, 16:2, 2012.

D. Gorodnichy, E. Granger, and P. Radtke, “Survey of commercial technologies for face

recognition in video”, CBSA Science and Engineering Directorate, Division Report 2014-06 (TR).

Título do Artigo: Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera.

Autores: Mohammad A. Haque, Kamal Nasrollahi, Thomas B. Moeslund.

Data de Publicação: 2013.

Veículo de Publicação: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 443–448.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6636680>

Abstract: Traditional still camera-based facial image acquisition systems in surveillance applications produce low quality face images. This is mainly due to the distance between the camera and subjects of interest. Furthermore, people in such videos usually move around, change their head poses, and facial expressions. Moreover, the imaging conditions like illumination, occlusion, and noise may change. These all aggregate the quality of most of the detected face images in terms of measures like resolution, pose, brightness, and sharpness. To deal with these problems this paper presents an active camera-based real-time high-quality face image acquisition system, which utilizes pan-tilt-zoom parameters of a camera to focus on a human face in a scene and employs a face quality assessment method to log the best quality faces from the captured frames. The system consists of four modules: face detection, camera control, face tracking, and face quality assessment before logging. Experimental results show that the proposed system can effectively log the high quality faces from the active camera in real-time (an average of 61.74ms was spent per frame) with an accuracy of 85.27% compared to human annotated data.

Resumo: O trabalho apresenta um sistema de aquisição de imagens faciais de alta qualidade de tempo real e baseado em câmera, que utiliza parâmetros *pan-tilt-zoom* da câmera para focar na face de sujeitos em uma cena, e então emprega um método de

avaliação da qualidade da face para registrar as melhores faces dos *frames* capturados da câmera de vídeo. O sistema consiste de quatro módulos: detecção de face, controle da câmera, rastreamento da face e avaliação da qualidade da face. A performance de uma aplicação de reconhecimento facial depende muito da qualidade das imagens obtidas. Por isto, é proposto uma forma de captura, em tempo real, de sequências de imagens faciais de alta qualidade. O módulo de detecção de face é responsável por detectar as faces em uma sequências de imagens (*frames*). O módulo de controle de câmera utiliza informações do módulo de detecção de face, como por exemplo, pontos x e y e altura e largura da face, para controlar automaticamente o *zoom* da câmera para focar na face do sujeito. Pode acontecer de o sujeito se mover rapidamente e sair do *zoom* da câmera. Para resolver este problema foi implementado uma 're-detecção' da face, ou seja, caso a câmera não esteja capturando nenhuma face, ela volta o *zoom* ao estado original retorna para o ponto inicial de detecção de face. Módulo de rastreamento de face rastreia a face e dá a região mais exata correspondente a região da face nos quadros ampliados (com *zoom*), empregando um algoritmo de rastreamento. O módulo de avaliação da qualidade da face é responsável por avaliar a qualidade das faces extraídas das sequências de vídeo capturadas pela câmera.

Contribuições: Propor um sistema, em tempo real, de aquisição de imagem facial de alta qualidade baseado em câmera. Sendo a principal contribuição o tempo de processamento.

Técnicas Utilizadas: Face Quality Assessment (FQA), Viola and Jones method, Camshift, Implementações em C++ utilizando as bibliotecas OpenCV (computer vision) e POCO (network package).

Base de Imagens Utilizada: Para os testes foi utilizada uma câmera própria, com suas configurações descritas no trabalho.

Resolução das imagens faciais obtidas: 10x10 e 70x70 pixels.

Variáveis Consideradas: Utilizam variação de pose, nitidez, brilho e resolução. Mas também citam como problemas: distância entre a câmera e o sujeito, movimentação constante, expressões faciais, variação de iluminação, oclusão e ruído.

Resultados Encontrados: A partir dos experimentos observou-se que, em comparação com a percepção humana, a abordagem proposta pelo trabalho determinou efetivamente a qualidade das faces extraídas com precisão de 85,7%. Além disto, os resultados do sistema proposto a respeito do custo computacional também foram excelentes. Com o conjunto de dados experimental o sistema registrou faces de boa qualidade em uma média de 16.20 fps (*frames por segundo*).

Comentários Adicionais: O trabalho é bem completo, apresenta bem os módulos do sistema, e será muito útil como estado da arte.

Referências Relevantes:

S. Z. Li, and A. K. Jain, Handbook of Face Recognition, 2 nd Edition, DOI 10.1007/978-0-85729-932-1, 2011.

X. Cheng, R. Lakemond, C. Fookes, and S. Sridharan, "Efficient Real-Time Face Detection For High Resolution Surveillance Applications," Proc. Of the 6 th Int. Conf. On Signal Proce. and Comm. Systems, pp.1-6, 2012.

F. W. Wheeler, R. L. Weiss, and P. H. Tu, "Face Recognition at a Distance System for Surveillance Applications," Proc. Of the 4 th IEEE Int. Conf. on Biometrics, pp. 1-8, 2010.

Título do Artigo: Fusing face recognition from multiple cameras.

Autores: Josh Harguess, Changbo Hu e J. K. Aggarwal.

Data de Publicação: 2009.

Veículo de Publicação: Workshop on Applications of Computer Vision (WACV), 1–7.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5403055>

Abstract: Face recognition from video has recently received much interest. However, several challenges for such a system exist, such as resolution, occlusion (from objects or self-occlusion), motion blur, and illumination. The aim of this paper is to overcome the problem of self-occlusion by observing a person from multiple cameras with uniquely

different views of the person's face and fusing the recognition results in a meaningful way. Each camera may only capture a part of the face, such as the right or left half of the face. We propose a methodology to use cylinder head models (CHMs) to track the face of a subject in multiple cameras. The problem of face recognition from video is then transformed to a still face recognition problem which has been well studied. The recognition results are fused based on the extracted pose of the face. For instance, the recognition result from a frontal face should be weighted higher than the recognition result from a face with a yaw of 30° . Eigenfaces is used for still face recognition along with the average-half-face to reduce the effect of transformation errors. Results of tracking are further aggregated to produce 100% accuracy using video taken from two cameras in our lab.

Resumo: O trabalho proposto trata do problema de “auto-occlusão”, ou seja, quando um indivíduo de alguma forma oculta parte do próprio rosto, seja com a mão, com cabelo, etc. O objetivo principal é capturar o indivíduo de várias câmeras diferentes, com pontos de visão diferente, e fundir para alcançar um resultado de reconhecimento significativo. Vários pesquisadores tem o foco no reconhecimento facial em apenas uma câmera, mas cenários mais realísticos de vigilância incluem múltiplas câmeras. Os autores propõe um método de rastrear o sujeito usando um “modelo de cabeça de cilindro” (CHM). Sabendo a pose do CHM é possível produzir uma vista frontal da face do sujeito, mesmo faltando algumas informações do rosto, por causa da “auto-occlusão”. Ao fundir os resultados de múltiplas câmeras, o reconhecimento facial melhorou de 67,4% para 94,4% nos vídeos de teste. Através dos testes pode-se notar claramente que a ideia de fundir os resultados é mais bem sucedido do que o reconhecimento de forma independente. Este método por si só mostrou um aumento na precisão de 4% a 25% a mais nos testes realizados, em comparação com o reconhecimento de forma independente. Os dois métodos que se mostraram mais bem sucedidos para fundir os resultados parecem ser a distância mínima e o método de “Gaussian weights”.

Contribuições: O principal objetivo deste trabalho é superar o problema da “auto-occlusão”, observando uma pessoa de diferentes câmeras, com diferentes pontos de visão exclusivamente para o rosto da pessoa, e fundir os resultados do reconhecimento de uma forma significativa.

Técnicas Utilizadas: Eigenfaces, Gaussian weights, distância mínima, melhor pose, Euclidean distance, Principal Components Analysis (PCA), average-half-face (AHF), Nearest-neighbors (NN).

Base de Imagens Utilizada: Sequência de vídeos próprias, obtidas de duas câmeras, onde continham apenas um sujeito em cada vídeo.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Resolução, oclusão, auto-occlusão.

Resultados Encontrados: Ao fundir os resultados de múltiplas câmeras, o reconhecimento facial melhorou de 67,4% para 94,4% nos vídeos de teste. O método por si só mostrou um aumento na precisão de 4% a 25% a mais, em comparação com o reconhecimento de forma independente. Ou seja, a ideia de fundir os resultados é claramente mais bem sucedido do que o reconhecimento de forma independente

Comentários Adicionais: O trabalho é muito interessante, mas tem como foco apenas resolver ou amenizar o problema da “auto-occlusão”, utilizando uma abordagem diferente, o CHM.

Referências Relevantes:

[2] C. Hu, J. Harguess, and J. K. Aggarwal. Patch-based face recognition from video. In International Conference on Image Processing (ICIP), pages 1–4, November 2009.

Título do Artigo: Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters.

Autores: Jingu Heo, Marios Savvides e B.V.K. Vijayakumar.

Data de Publicação: 2005.

Veículo de Publicação: IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, 2005. CVPR Workshops, 9–9.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1565304>

Abstract: This paper presents the face recognition performance evaluation using visual and thermal infrared (IR) face images with correlation filter methods. New correlation filter designs have shown to be distortion invariant and the advantages of using thermal IR images are due to their invariance to visible illumination variations. A combined use of thermal IR image data and correlation filters makes a viable means for improving the performance of face recognition techniques, especially beyond visual spectrum. Subset of Equinox databases are used for the performance evaluation. Among various advanced correlation filters, minimum average correlation energy (MACE) filters and optimum trade-off synthetic discriminant function (OTSDF) filters are used in our experiments. We show that correlation filters perform well when the size of face is of significantly low resolution (e.g. 20 x 20 pixels). Performing robust face recognition using low resolution images has many applications including performing human identification at a distance (HID). The eyeglass detection and removal in thermal images are processed to increase the performance in thermal face recognition. We show that we can outperform commercial face recognition algorithms such as Facelt ® based on Local Feature Analysis (LFA).

Resumo: O uso combinado de dados de imagens termais e filtros de correlação torna-se um meio viável para melhorar o desempenho de técnicas de reconhecimento facial. O reconhecimento facial em imagens de baixa resolução tem muitas aplicações, incluindo o identificação humana a distância (HID). Os autores mostram que os filtros de correlação tem um bom desempenho quando o tamanho da imagem é significativamente pequeno, ou de baixa resolução (por exemplo: 20x20 pixels). Uma face é essencialmente um objeto 3D, ou seja, fontes de luminosidade de diferentes direções podem mudar drasticamente a aparência de uma face. O reconhecimento facial perde precisão rapidamente se a luminosidade é fraca ou não uniforme. Imagens termais infravermelho não dependem de iluminação, pois os sensores medem a energia térmica emitida pelos objetos. Assim, imagens termais tem muita vantagem no reconhecimento facial em ambientes de baixa luminosidade. Porém as “assinaturas” termais podem ser significativamente alteradas de acordo com diferentes temperaturas do corpo, geradas por exercícios físicos ou temperaturas de ambiente. Através disto, pode-se dizer que a combinação de imagens visuais e termais podem minimizar problemas, no

reconhecimento facial, causados por apenas uma das modalidades. Resultados apresentaram que sobre variação de iluminação e expressões faciais, o reconhecimento facial a partir de imagens termais pode ter uma performance melhor do que o reconhecimento facial visual. Os resultados também mostram que o filtro OTSDF tem uma performance boa quando as imagens são de resolução extremamente baixa, mostrando que este método é muito atrativo para aplicações que utilizam baixas resoluções. Entretanto, imagens termais de sujeito usando óculos podem perder muitas informações ao redor dos olhos, pois o óculos bloqueia uma grande porção de energia termal emitida pelos olhos. Para minimizar este problema os autores optaram por remover a região dos óculos, o que demonstrou um aumento de performance no reconhecimento facial com imagens termais, se comparado com o mesmo reconhecimento utilizando a região dos óculos.

Contribuições: Avaliação de performance de imagens visuais e termais, utilizando filtros de correlação, incluindo também a comparação com outros algoritmos de reconhecimento facial.

Técnicas Utilizadas: minimum average correlation energy (MACE) filter, optimum trade-off synthetic discriminant function (OTSDF) filter, Local Feature Algorithm (LFA), Principal Component Analysis (PCA), synthetic discriminant function (SDF) filter. Nos testes também é utilizado o sistema de reconhecimento facial comercial FaceIt.

Base de Imagens Utilizada: Equinox database.

Resolução das imagens faciais obtidas: 12x12 pixels até 128x128 pixels.

Variáveis Consideradas: Uso de óculos e luminosidade.

Resultados Encontrados: Os resultados encontrados mostram que filtros de correlação mostram uma melhor performance em imagens de baixa resolução (exemplo 32x32 pixels), tanto para imagens visuais quanto termais. Em comparação entre imagens visuais e termais, o reconhecimento facial termal mostrou uma performance maior sobre várias condições de iluminação e expressão facial. Outro ponto apresentado é que o uso de óculos pode dificultar bastante o reconhecimento facial, e que a substituição da

região dos olhos melhorou muito a performance do reconhecimento em imagens termais. Concluiu-se que a combinação das imagens visuais e termais podem melhorar significativamente a performance do reconhecimento facial se comparado com cada método individualmente.

Comentários Adicionais: O trabalho trata também de imagens termais, o que não faz parte do escopo do trabalho que pretendo desenvolver, pois necessitaria também de sensores de infravermelho. Porém trata também do uso de óculos e de imagens de baixa resolução, o que são fatores bastante relevantes para o desempenho do reconhecimento facial.

Referências Relevantes:

I. Craw, N. Costen, T. Kato, and S. Akamatsu, "How should we represent faces for automatic recognition?," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, pp.725-736, 1999.

R. Chellappa, C.L. Wilson, S. Sirohey, "Human and machine recognition of faces: a survey," Proceedings of the IEEE, Vol.83 (5), pp705-741 (1995).

Título do Artigo: Frontal face detection for surveillance purposes using dual Local Binary Patterns features.

Autores: Wael Louis e K.N. Plataniotis.

Data de Publicação: 2010.

Veículo de Publicação: 2010 17th IEEE International Conference on Image Processing (ICIP), 3809–3812..

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5653543>

Abstract: Face detection in video sequence is becoming popular in surveillance applications, but the usage of large number of features and the long training time are persistent problems. This paper integrates two types of Local Binary Patterns (LBP) features in order to achieve a high detection rate with a high discriminative power face

detector. First LBP feature is a novel way of using the Circular LBP, in which the pixels of the image are targeted; it is a non-computationally expensive feature extraction. The second LBP feature is the LBP Histogram, in which regions in the image are targeted; it is more computationally expensive than Circular LBP features but has higher discriminative power. The proposed detector is examined on real-life low-resolution surveillance sequence. Conducted experiments show that the proposed detector achieves 98% detection rate in comparison to 91% for the Lienhart detector. The proposed detector tolerates wide range of illumination changes.

Resumo: Este trabalho integra dois tipos de características Padrões Binários Locais (Local Binary Patterns – LBP), para atingir uma alta taxa de reconhecimento com um alto poder discriminativo em detectores de face. A primeira característica LBP é uma nova abordagem de utilização do LBP circular, onde os pixels da imagem são o alvo. A segunda é a LBP histograma, em que regiões da imagem são o alvo, e é computacionalmente mais caro que a LBP circular, mas tem um maior poder discriminativo. O detector proposto é examinado em um ambiente de vigilância real e de baixa resolução. O método proposto no artigo é para implementar um detector de face frontal com uma alta taxa de performance (ex.: $\geq 90\%$). Onde o detector é treinado com um pequeno número de características discriminativas e em um curto espaço de tempo. O detector tem o objetivo de ser usado em sistemas de vigilância reais, utilizando imagens 2D de baixa resolução, de câmeras estáticas colocadas em posições que capturam geralmente a face frontal dos sujeitos. O detector proposto é baseado no LBP circular, que tem poder discriminativo simples de extrair, o que custa pouco poder computacional e o LBP histograma, que tem um alto poder discriminativo, porém tem também um alto custo computacional.

Contribuições: Integrar dois tipos de características (LBP circular e LBP histograma) para melhorar o poder discriminativo delas, e em seguida integrá-los levando a um detector facial de alto poder discriminativo.

Técnicas Utilizadas: AdaBoost algorithm, GentleBoost algorithm.

Base de Imagens Utilizada: Viola and Jones dataset, Ole Jensen dataset e outra base de dados disponibilizada pela Universidade de Toronto.

Resolução das imagens faciais obtidas: 24x24 pixels e 60x80 pixels.

Variáveis Consideradas: Até citam algumas variáveis que podem afetar a detecção, mas não especificam de tratá-las durante o trabalho.

Resultados Encontrados: O detector LBP foi testado em vídeos reais de uma câmera de vigilância, e os resultados experimentais mostram que ele supera o desempenho do detector Lienhart por atingir uma taxa de detecção de 98,22% comparado a 91,05% para o detector de Lienhart. Além disto, o detector LBP apresentou robustez para uma vasta gama de mudanças de iluminação.

Comentários Adicionais: O trabalho é muito interessante pois trata da detecção de faces em vídeos de vigilância, ou seja, pode ser utilizado futuramente como estado da arte, já que o trabalho proposto envolve primeiramente detectar as faces no vídeo.

Referências Relevantes:

O.H. Jensen and R. Larsen, "Implementing the Viola-Jones Face Detection Algorithm," M.S. thesis, Technical University of Denmark, 2008.

V. Popovici, J.P. Thiran, Y. Rodriguez, and S. Marcel, "On performance evaluation of face detection and localization algorithms," in Proc. of ICPR, 2004, vol. 1, pp. 313–317.

Título do Artigo: Enhanced weakly trained frontal face detector for surveillance purposes.

Autores: Wael Louis, K.N. Plataniotis e Yong Man Ro.

Data de Publicação: 2010.

Veículo de Publicação: 2010 IEEE International Conference on Fuzzy Systems (FUZZ), 1–8.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5584450>

Abstract: Face detection is becoming popular in surveillance applications; however, the need of enormous size face/non-face dataset, large number of features, and long training time are persistent problems. This paper claims that only a subset of the total number of features conserves the major power to detect faces; hence, this subset is capable to detect faces with high detection rate. The proposed detector fuses the results of two classifiers where one is trained with only 40 Haar-like features and the other is trained with only 50 LBP Histogram features. A pre-processing stage of skin-tone detection is applied to reduce the false positive rate. The detector is examined on real-life low-resolution surveillance sequence. Conducted experiments show that the proposed detector can achieve a high detection rate and a low false positive rate. Also, it outperforms Lienhart detector and tolerates wide range of illumination and blurring changes.

Resumo: Os autores acreditam que apenas um subconjunto do total de características conserva grande potência para a detecção de faces. Assim, é proposto um detector de faces que combina os resultados de dois classificadores, um treinado com 40 características Haar-like e o outro com 50 características LPB 'histogram'. É aplicada também uma fase de pré-processamento de tom de pele para reduzir a taxa de 'falso positivos', ou seja, imagens que não são realmente faces de sujeitos. Detectores faciais baseados em imagem requerem uma enorme quantidade de tempo para alcançar uma taxa desejada, bem como um longo período de tempo para treinamento. Muitos modelos gastam um enorme tempo também coletando imagens 'falsas', ou seja, que não correspondem a imagens faciais. Muitos estudos foram realizados sobre a técnica de detecção de Viola e Jones que utiliza características Haar-likes, porém estas características tem um poder discriminativo limitado. Outros modelos foram propostos, como por exemplo, Local Binary Patterns e Modified Census Transform (MCT), porém a maioria necessita de uma grande base de dados de treinamento e bastante tempo de treinamento. Portanto, o método proposto pelos autores consiste em implementar um detector facial frontal que é treinado com um pequeno número de características extraídas de uma pequena base de dados. Podendo ser treinado em um curto período de tempo, com o propósito de ser utilizado em sistemas de vigilância com informações em 2 dimensões (2D) de uma câmera estática posicionada onde geralmente captura-se a face frontal dos indivíduos. Combinando as decisões de cada um dos dois classificadores, é possível preservar as detecções verdadeiras e reduzir drasticamente

as detecções falsas. Os testes foram conduzidos baseando-se na taxa de detecção de faces e o número de detecções de “falso positivos”. A detecção da face é considerada correta quando a distância Euclidiana entre a face e a face detectada for menor do que 30% da largura do tamanho da detecção.

Contribuições: Os autores propõe um método que consiste em implementar um detector facial frontal que é treinado com um pequeno número de características extraídas de uma pequena base de dados, com o propósito de ser utilizado em sistemas de vigilância, com informações em 2 dimensões, com uma câmera estática posicionada onde geralmente captura-se a face frontal dos indivíduos. Através disto, o trabalho apresenta duas contribuições principais: a primeira consiste em provar que é possível extrair características de apenas um pequeno subconjunto dos recursos totais, e isso contribui para a maior parte dos resultados. A segunda contribuição consiste na utilização de uma fase de pré-processamento usando um detector de tons de pele para eliminar os pixels que tem uma alta probabilidade de não ser pele.

Técnicas Utilizadas: LBP Histogram, Enhanced Parallel Detector (EPD), Haar-like features, Viola e Jones, AdaBoost algorithm, Euclidean distance, GentleBoost algorithm.

Base de Imagens Utilizada: Viola and Jones faces dataset, Ole Jensen dataset, foi utilizado também outra base que corresponde a um ambiente real, e foi disponibilizado pela Universidade de Toronto, para fins de pesquisa.

Resolução das imagens faciais obtidas: 24x24 pixels e 60x80 pixels.

Variáveis Consideradas: Iluminação.

Resultados Encontrados: O detector proposto usa dois classificadores fracamente treinados em cascata e em paralelo, onde cada um é treinado com um conjunto pequeno de características discriminativas de face. O detector é aplicado a um cenário real e é comparado com o Lienhart (estado da arte). A robustez do sistema foi examinada, adicionando artificialmente ruído as imagens. Foi adicionado um ruído de mudança de iluminação entre -100% e +100%, outro ruído que foi aplicado foi o filtro Gaussian para desfocar as imagens. Combinando os resultados dos dois classificadores o detector

provou a sua capacidade de detecção mesmo lidando com uma alta gama de mudanças de iluminação.

Comentários Adicionais: Acredito que este trabalho seja bastante importante para o estado da arte, pois trata de um método de detecção de rostos robusto e que pode lidar bem com variação de luminosidade.

Referências Relevantes:

A. Hadid, M. Pietikainen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in IEEE Proc. of CVPR, vol. 2, 2004.

C. Shen, S. Paisitkriangkrai, and J. Zhang, "Face detection from few training examples," in IEEE Proc. of ICIP, 2008, pp. 2764–2767.

Título do Artigo: Face detection for video summary using enhancement-based fusion strategy under varying illumination conditions.

Autores: Richa Mishra and Ravi Subban.

Data de Publicação: 2014.

Veículo de Publicação: 2014 International Conference on Science Engineering and Management Research (ICSEMR), 1–8.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7043648>

Abstract: A biometric-based techniques emerge as the promising approach for most of the real-time applications including security systems, video surveillances, human-computer interaction and many more. Among all biometric methods, face recognition offers more benefits as compared to others. Diagnosing human faces and localizing them in images or videos is the priori step of tracking and recognizing. But the performance of face detection is limited by certain factors namely lighting conditions, pose variation, occlusions, low resolution images and complex background. To overcome the problems, this paper examines a fusion strategy in the enhancement-based skin-color segmentation approach that can improve the performance of face detection algorithm. The method is

robust against complex background, ethnicity and lighting variations. The method consists of three steps. The first step receives spatial transform techniques in parallel to enhance the contrast of the image, change the color space of the enhanced images to YcbCr, apply skin segmentation technique and yield the binary segmented images. The second step ascertains the weight of accuracy (WoA) of each of the segmented image and fed it into the fusion strategy to get the final skin detected region. Finally, the last step localizes the human face. The methodology is not constrained to just frontal face identification. However, it is invariant with the diverse head postures, enlightenment condition and size of faces. The experimental result demonstrates the improvement in the accuracy and precision along with the reduction in FPR as compared to other enhancement classifiers.

Resumo: Localizar e diagnosticar faces de sujeitos em imagens ou vídeos é o primeiro passo para o rastreamento e reconhecimento facial. Porém a performance do reconhecimento facial está ligado a condições de luminosidade, variação de pose, oclusão, baixa resolução e fundo da imagem complexo. Para superar estes problemas o trabalho examina uma estratégia de fusão baseada em segmentação de tom de pele e realce, que podem melhorar o desempenho do algoritmo de detecção facial. O método consiste de três passos: o primeiro passo recebe técnicas de transformação espacial em paralelo para aumentar o contraste das imagens; o segundo passo verifica o peso de precisão de cada uma das imagens segmentadas e utiliza a estratégia de fusão para obter a região final de pele detectada; e finalmente o último passo detecta a face humana. O processo de detecção facial inicia com a localização de faces humanas seguido pela extração e verificação destas. Os atributos selecionados pelos autores foram, a cor da pele, área da região facial e ponto central. A estratégia de fusão foi adaptada para reduzir a taxa de 'falso positivos' e aumentar a taxa de 'verdadeiro positivo'. O algoritmo proposto corrige o problema da deslocalização, identificando corretamente a cor da pele com o mínimo de 'falso positivo', para localizar a face humana.

Contribuições: Os autores propõe um algoritmo de detecção de face eficiente e robusto apresentado através da detecção de cor da pele usando uma estratégia de fusão baseada em realce e processamento morfológico para superar as limitações mencionadas no resumo.

Técnicas Utilizadas: Piecewise linear decision boundary algorithm, entre outros propostos.

Base de Imagens Utilizada: Caltech face database, Indian face database, e imagens coletadas da internet.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Mudanças de iluminação, gestos variados, poses e oclusão.

Resultados Encontrados: Os experimentos indicaram que a acurácia esta por volta de 90,9% na base de imagens Caltech, com uma taxa de apenas 7,9% de 'falso positivos'. Já na Indian face database ficou com uma taxa de 'falso positivos' de apenas 3,2% e acurácia de 95,9%. E por fim, com as imagens obtidas da internet foi possível chegar a uma taxa ainda menor de 'falso positivos', apenas 2,6%, e uma acurácia de 95%. A estratégia de fusão foi adaptada para reduzir a taxa de 'falso positivos' e aumentar a taxa de 'verdadeiro positivo'. O algoritmo proposto corrige o problema da deslocalização, identificando corretamente a cor da pele com o mínimo de 'falso positivo', para localizar a face humana. A principal vantagem da abordagem é a simplicidade e baixa complexidade.

Comentários Adicionais: O trabalho é muito interessante e teve resultados realmente bons, com baixa taxa de 'falso positivos', e pode ser muito útil como estado da arte para o trabalho proposto.

Referências Relevantes:

E. Hjelm, B.K. Low, "Face detection: A survey", J. Computer Vision and Image Understanding, vol. 83, pp. 236–274, 2001 .

Título do Artigo: Pre-Attentive Face Detection for Foveated Wide-Field Surveillance.

Autores: S.J.D. Prince, J.H. Elder, Y. Hou e M. Sizinstev.

Data de Publicação: 2005.

Veículo de Publicação: Seventh IEEE Workshops on Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1, 439–446.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4129515>

Abstract: Conventional surveillance sensors suffer from an unavoidable tradeoff between image resolution and field of view. This problem may be overcome by combining a fixed, pre-attentive, low-resolution wide-field camera with a shiftable, attentive, high-resolution narrow-field camera. Here we present techniques for orienting the attentive camera to faces detected in the pre-attentive wide-field image stream. Unfortunately, the low image resolution of the wide-field sensor precludes the use of most conventional face detection algorithms. Instead, we argue that reliable performance can best be achieved by accurate probabilistic combination of multiple cues: skin detection, motion detection and foreground extraction. Fast sampling of scalespace over all three modalities is achieved using integral images and parametric models of response distributions are derived using supervised learning techniques. Log likelihood ratios for each modality are combined with spatial priors incorporating tracking and novelty objectives to yield a posterior map indicating the probability of a face appearing at each image location. The result is a real-time attentive visual sensor which reliably fixates faces over a 130 deg field of view, allowing high-resolution capture of facial images over a large dynamic scene.

Resumo: A detecção de pessoas em cenas com amplo campo de visão são complicadas por muitos fatores, como por exemplo, baixa resolução, diversidade de poses e oclusão. Os autores apresentam técnicas para orientar a câmera para os rostos detectados em um campo amplo de imagem. A baixa resolução da imagem do sensor impede o uso da maioria dos algoritmos de detecção facial convencionais. Em contraponto, argumenta-se que o desempenho de confiança pode ser melhorado com a combinações de algumas pistas, como: detecção de pele, detecção de movimento e

extração de primeiro plano. Os sensores utilizados são fixos, o que simplifica a análise do movimento visual. O objetivo deste trabalho é desenvolver e avaliar um método para a detecção rápida e confiável de cabeças de sujeitos em um fluxo de vídeo de campo amplo. O que permite que o sistema possa fixar e analisar as faces de vários sujeitos que realizam diversas atividades dentro do ambiente.

Contribuições: Desenvolver e avaliar um método para a detecção rápida e confiável de cabeças de sujeitos em um fluxo de vídeo de campo amplo sobre uma gama muito ampla de condições. E como resultado apresentar um sensor visual em tempo real, que se fixa ao longo das faces dos sujeitos, em um ângulo de 130 graus de campo de visão, permitindo a captura em alta resolução de imagens faciais em uma cena ampla e dinâmica.

Técnicas Utilizadas: Viola & Jones, EM algorithm e outros métodos próprios.

Base de Imagens Utilizada: Testes realizados com câmeras próprias.

Resolução das imagens faciais obtidas: Não especificado.

Variáveis Consideradas: Cor da pele (para detecção) e movimento. Apenas citam: variação de pose, baixa resolução e oclusão.

Resultados Encontrados: Durante os testes 97% dos rostos detectados eram realmente rostos de indivíduos. O resultado é um sensor visual em tempo real que localiza de forma confiável faces nas imagens, sobre um campo de visão de 130 graus, permitindo a captura de imagens faciais em alta resolução e em uma ampla cena.

Comentários Adicionais: O artigo foge um pouco do tema proposto pelo trabalho que pretendo desenvolver, e deveria ser revisto pelos outros integrantes do projeto para uma reavaliação.

Referências Relevantes:

H. Kruppa, M.C. Santana, and B. Schiele, "Fast and robust face finding via local context," in Proc. VS- PETS, 2003, pp. 157–164.

Título do Artigo: A novel image fusion scheme for robust multiple face recognition with light-field camera.

Autores: R. Raghavendra, Kiran B Raja, Bian Yang e Christoph Busch.

Data de Publicação: 2013.

Veículo de Publicação: 2013 16th International Conference on Information Fusion (FUSION), 722–729.

Fonte: IEEE.

Link: <http://ieeexplore.ieee.org>

Abstract: Accurate face recognition in wide area surveillance application is a challenging problem because of multiple pose, non-uniform illumination, low resolution and out-of-focus face images that are recorded with conventional surveillance cameras (Closed-Circuit TeleVision or Pan-Tilt-Zoom). In this paper, we address the problem of face recognition in wide area surveillance with a light-field camera. The main advantage of a light-field camera is that, it can provide different focus (or depth) images in a single exposure (capture) which is not possible with a conventional 2D camera. In this work, we propose a novel weighted image fusion scheme to combine different depth (or focus) images rendered by a light-field camera. The proposed image fusion scheme is not only dynamic in handling number of depth (or focus) images but also adaptive in assigning higher weights to the best focus image as compared to the out-of-focus image. Extensive experiments are carried out on our newly acquired face dataset captured using Lytro light-field camera to bring out the merits and demerits of the proposed weighted image fusion scheme for face recognition in wide area surveillance applications.

Resumo: Os autores abordam o problema do reconhecimento facial em uma área de vigilância com uma câmera com campo de luz. A principal vantagem de uma câmera com campo de luz é que ela pode fornecer um foco diferente para imagens de uma única exposição, o que não é possível com câmeras 2D convencionais. Assim, os autores propõe um esquema de fusão de imagem ponderada para combinar diferentes profundidades de imagens renderizadas por uma câmera de campo de luz. O método proposto pode ser estruturado em três fases: (1) detecção e extração da face; (2)

seleção da imagem e esquema de fusão ponderada; (3) extração e identificação de características. Cada fase é detalhadamente explicada nas seções do trabalho. O conjunto de dados foi obtido pelos próprios pesquisadores no laboratório onde o projeto foi desenvolvido, e envolve ambientes internos e externos.

Contribuições: Os autores propõe uma abordagem de um esquema de avaliação de pesos para então realizar um esquema de fusão de imagens ponderadas para combinar diferentes profundidades de uma imagem para melhorar a precisão do reconhecimento facial.

Técnicas Utilizadas: Viola-Jones face detector, 2D- Discrete Wavelet Transformation (DWT), Algoritmo proposto para avaliação de pesos, Inverse Discrete Wavelet Transform (IDWT), Local Binary Patterns (LBP), Log-Gabor filters, Sparse Reconstruction Classifier (SRC).

Base de Imagens Utilizada: O conjunto de dados para testes foi obtido pelo próprio laboratório onde o projeto foi desenvolvido.

Resolução das imagens faciais obtidas: 120x120 pixels.

Variáveis Consideradas: Apenas citam alguns desafios, como: multiplas poses, variação de luminosidade, baixa resolução, face fora de foco e distância.

Resultados Encontrados: Através dos testes desenvolvidos com a base de dados própria, em ambientes internos e externos, avaliou-se o detector de face em cada amostra e obteve-se uma taxa de detecção de 94,8%.

Comentários Adicionais: Este trabalho se mostrou bastante interessante pois trata da profundidade dos sujeitos em uma imagem, ponto este, que provavelmente precisará ser tratado no desenvolvimento do trabalho proposto.

Referências Relevantes:

Ross.A, Nandakumar.K, and J. A.K. Handbook of Multibiometrics. Springer-Verlag edition, 2006.