Face Recognition for Video Security Applications

Ngoc-Son VU[1], Siméon SCHWAB[1,2], Pierre BOUGES[2], Xavier NATUREL[1], Christophe BLANC[1,2], Thierry CHATEAU[2], Laurent TRASSOUDAINE[2]

[1]VESALIS, 8 allée Evariste Galois, 63000 Clermont-Ferrand, France

[2]Université Blaise Pascal, 24 avenue des Landais, 63177 Aubière, France

Ngoc-Son.Vu@vesalis.com, simeon.schwab@univ-bpclermont.fr, xavier.naturel@vesalis.fr,
christophe.blanc@vesalis.com, laurent.trassoudaine@univ-bpclermont.fr, thierry.chateau@univ-bpclermont.fr

Résumé – La reconnaissance faciale devient un outil important pour améliorer la sécurité, avec de nombreuses applications possibles. Il reste cependant de nombreuses difficultés, et les résultats peuvent toujours être de mauvaise qualité dans des environnements non contraints. L'article présente un système complet de reconnaissance faciale, destiné à fonctionner dans un environnement réaliste. Nous présentons des systèmes, allant de la détection à la reconnaissance, avec des outils additionnels nécessaires dans des scénarios réels, et améliorer les performances de la reconnaissance pour les environnements non contraints. Nous concluons par présenter une méthode alternative d'identification de personnes en utilisant le clustering de visages.

Abstract – Face recognition is an important tool for improving security, with many possible applications. It is however a difficult task in uncontrolled environments, where results can be dramatically reduced compared to controlled conditions. This article presents face recognition in a realistic setup, a complete processing chain from face detection to actual recognition, with additional tools that are necessary in real-life scenarios. State-of-the-Art face detection and recognition methods are detailed, as well as methods to improve recognition accuracy in uncontrolled environments. We conclude by presenting an alternative method for identification using face clustering.

1. Introduction

Face recognition is emerging as a promising tool for enforcing security in various situations. The biometric passport is perhaps the widest diffused application, but access control, fraud detection, identity check, detection of missing persons and general video-surveillance are among other possible security applications.

This article presents a complete processing chain. Some methods are also presented, that can greatly help recognition in real life situations. The paper is structured as follows. An overview of the methods and the general system are given in section 2. Face detection is detailed in section 3. Section 4 presents some methods that can improve general performance in difficult situations, while our technique for face recognition is explained in section 5. Eventually, section 6 provides a radically different approach for identification using face clustering, which is suitable in difficult environments.

2. Overview

Face recognition for unconstrained video-surveillance environments is a highly demanding task, and needs several pre-processing to be usable. We present a complete system that can cope with difficult situations. Fig. 1 gives a detailed workflow of the processing. The first and mandatory step is face detection, detailed in Section 4.

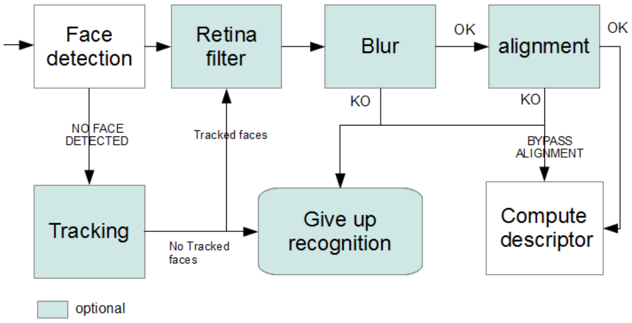

FIG. 1: Detailed workflow from face detection to recognition

Tracking can be used to recover from misses from face detection. The retina filter is used as *photometric* normalization to increase both lighting invariance and recognition performance. The blur filter can be used to detect whether an image is too blurry to be usable for recognition. Performed just before recognition, alignment is for *geometrically* normalizing the detected face so that facial features (eyes, mouth) are at canonical positions in image. These processes are optional and described in Section 6

3. Face detection

Face detection is a well-known and studied topic, which has gained widespread attention and usage, especially since the work of Viola and Jones [1]. However, there are still issues with non-frontal poses and occlusions, especially in a video-surveillance context. We present here a technique based on the popular boosted cascade of Viola and Jones, with a probabilistic formulation that can handle these issues, from Bouges et al. [2].

In a cascade of weak classifiers, occlusions and poses that are different from the learned pose cause some weak classifiers to answer negatively, corrupting the final decision, as in Figure 2. These classifiers are considered here as missing data, thus avoiding classical solutions based on multiple training.

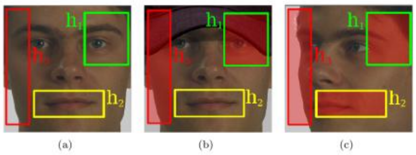

FIG. 2: sub-windows of weak classifiers learned on an upright face (a). Answers from these classifiers in (b) and (c) may be wrong (in red) because of occlusions and different poses

The face detector is using the covariance feature proposed in [3], and the logitboost variant of the boosting algorithm is used to build the cascade. To handle missing classifiers, thresholds in the classification process are modified by taking into account the probabilities at all previous stages. Results are presented in Figure 3.

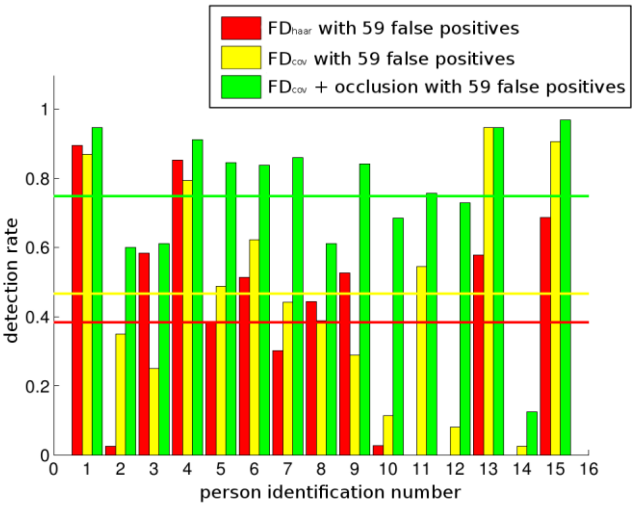

Fig. 3: Detection results on a realistic video-surveillance scene. Fifteen different persons are filmed in an unconstrained environment. Fdhaar is the basic method of Viola and Jones. Fdcov is the standard cascade with the covariance feature, and Fdcov-occlusion is the proposed method.

4. Pre and post-processing

Video surveillance data is very challenging, because of low quality video, low frame rate, small sizes, varying lighting conditions, or motion blur. We present here some techniques to enhance the data that is presented to the face recognition process.

- Tracking can help to recover from face detection misses and thus present more examples.
- Blurry images can confuse the recognition. Blur detection is thus helpful so that only sharp images are processed. It is based on the blur metric defined in [4].
- Retina filtering: illumination variations that often occur on face images significantly degrade the recognition performance. Inspired by the fact that our eyes, due to their retinas, are capable to see and identify objects in different lighting conditions, we proposed to use an efficient method based on the retina modeling for illumination normalization [5]. This algorithm consists of two adaptive nonlinear functions and a Difference of Gaussians filter, followed by a simple truncation. This ***real-time*** method not only removes the illumination variations but also enhances both the image contour and contrast (see Fig. 4), and thus, fortunately, improves significantly the recognition rates in both cases whether there are lighting changes on facial images or not.

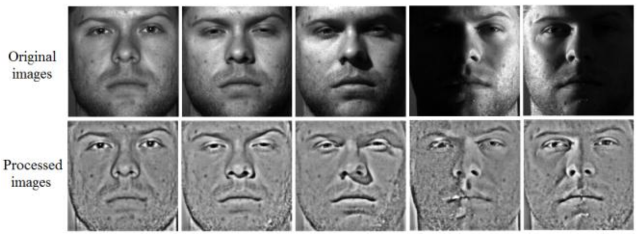

Fig. 4: Examples of retina filtering

- Alignment is a necessary step that normalizes geometrically the input face features to canonical positions. It can be done with a simple geometric scheme based on some face component positions (eyes, mouth), given by local detectors. A more holistic approach can be used, such as in [6]. It is an unsupervised method, where the minimization of a sum of entropy function gives the transformation parameters for a set of images (see result example in Fig. 5).
- Temporal smoothing is a post-processing method that averages the recognition results on some time-window. It can use the best result (hard decision) or a sum of the recognition scores (soft decision)

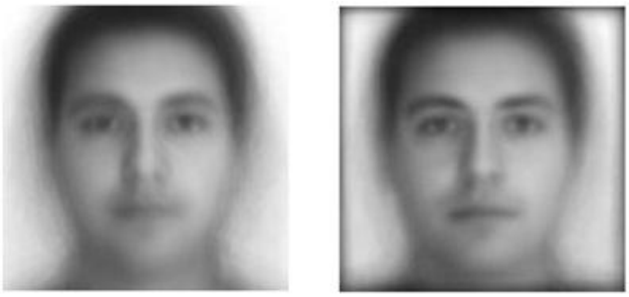

Fig. 5: Average images of a 200-image set before (left) and after (right) the alignment. Since the goal of alignment is to transform faces into a standard pose, the average of the considered image set should have clear facial features after alignment. Here the right image is much clearer than the left one.

5. Face recognition

This important step consists of two main sub-steps: (1) feature extraction where *normalized* face region is represented by a feature vector (also called descriptor), and (2) classification where different descriptors are compared and matched in order to give a final recognition result.

Generally, facial feature description is the most important aspect, if inadequate features are used, even the best classifier will fail to achieve an accurate result. Good face descriptors are desired to have the following properties: *robustness*, *distinctiveness*, and *low computational cost* in terms of both time and space. These allow the system to quickly deliver high accurate results to the end user. Unfortunately, most of existing face descriptors do not balance these criteria: the features producing high-quality recognition results are computationally intensive, whereas low complexity algorithms do not perform reliably enough. We propose here to use innovate descriptors with complementary strength, called soft-POEM (Pattern of Oriented Edge Magnitudes) and POD (Patterns of Orientation Difference)[7,8].

For every pixel, the POEM feature is built by applying a self-similarity based structure on oriented magnitudes, calculated by accumulating a local histogram of gradient orientations over all pixels of image cells, centered on the considered pixel. Contrary to POEM which considers the relationships between edge *magnitude* distributions, POD encodes the relationships between *orientations* of local image patches. Once the gradient orientation is obtained for all pixels, the rest of the POD algorithm consists of two steps: smoothing which acts as information incorporation, and encoding the incorporated orientation pixel-by-pixel.

The POEM and POD images are then divided into patches where histograms are estimated, all estimated histograms are concatenated in order to get a descriptor. To reduce the feature dimension, a PCA is applied, followed by a Whitening process.

By associating this face representation with the *simple* “nearest neighbor” *classifier*, the state-of-the-art face recognition results are established on several common databases. Presented in Table 1 are the recognition results obtained on the comprehensive FERET face database which gathers 3000+ images of 1200 subjects with different conditions of lighting, expression, and expression. It is worth noting that: (1) this algorithm is ***real-time***; (2) other algorithms with similarly high recognition results are all based on Gabor filters; as feature extraction time, compared to only the first step of those Gabor based methods, ours is already at least 20 times faster.

Tab. 1: Recognition rate comparisons with state-of-the-art algorithms tested with standard FERET evaluation protocol. Note that, the time for extracting our whole representation is at least 20 times smaller than that of convolutions of an image with 40 Gabor kernels

Method	Rate	Note
HOG	73.3	
LBP	75.4	
LGBP	82.2	Processing on 40 Gabor magnitude images.
HGPP	89.4	Processing on 90 Gabor images
LGBP+ LGXP	96.9	LGXP = LXPs on Gabor phases
Ours	97.7	

6. Face clustering

Face recognition can fail, especially with low resolution and quality videos/frames as in video-surveillance condition. A different approach to identification is proposed here, which cluster similar faces, so that an individual is easily identifiable by its cluster rather than a single face detection.

Our method focuses on challenging situations encountered in actual cases: dense scenes (large number of people in a short period of time) with especially small faces (in resolution) and erratic pedestrian movements. To efficiently cluster face detections occurring in these types of videos, we use all the available information: time stamp, position in the frame and appearance provided by the video. Our approach [9] is global, it takes into account all the face detections to group them in a trajectory. This is not done in standard multi-face tracking.

The probabilistic framework used to represent our problematic, is based on recent works on tracking-by-detection [10]. Even if the combinatorial complexity is high, the optimal clustering is found in a reasonable computing time, thanks to a network-flow representation of the problem.

We also elaborate a sequential approach [11] to deal with long video and real time scenario.

Although our method has not reached the required quality for visual surveillance applications, we present a starting point for a video face summarization system, in scenes where automatic face recognition remains a challenging issue.

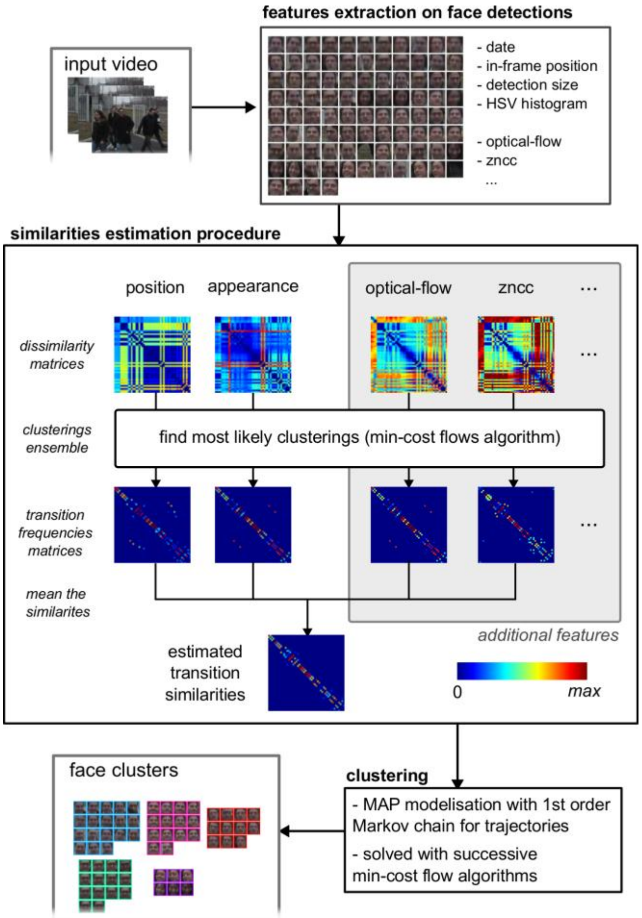
features extraction on face detections
input video
- date
- in-frame position
- detection size
- HSV histogram
- optical-flow
- zncc
...
similarities estimation procedure
position
appearance
optical-flow
zncc
...
dissimilarity matrices
clusterings ensemble
find most likely clusterings (min-cost flows algorithm)
transition frequencies matrices
mean the similarites
estimated transition similarities
additional features
0
max
face clusters
clustering
- MAP modelisation with 1st order Markov chain for trajectories
- solved with successive min-cost flow algorithms

Fig. 6: Different steps of our video-based face detections clustering method. We first detect faces in all the frame and extract features. Then, inter-detections similarity matrices are estimated in an appropriate way considering the different features and detections repartition in the video. With these similarities we finally cluster the face detections using a network-flow algorithm to find an optimal solution.

7. Conclusion

We have presented a complete system of face recognition, suitable for video-surveillance. In such difficult conditions, robustness is needed at all stages. A improved face detection is presented, that allows to detect faces with occlusions and at multiple poses, with little extra cost. A set of tools to improve the recognition results in difficult situations is also presented, to deal with illumination conditions, motion blur, and to recover from face detection and recognition misses. Despite these specific tools, face recognition for video surveillance is still very challenging, and a very different method based on face clustering is presented, that can be used in difficult cases.

While good results have been obtained on specific databases and applications, improvements made by face recognition systems are however still far from their expected performance in unconstrained environments. Issues like pose and orientation, large databases, are still important issues for face recognition systems in a security context.

8. Acknowledgements

This work is supported by the Vesalis company and the Biorafale project, funded by OSEO.

Références

[1] P. Viola and M. Jones, *Robust Real-time Object Detection,* International Journal of Computer Vision, 2001

[2] P. Bouges, T. Chateau, C. Blanc, and G. Loosli, *K-nearest neighbors to handle missing weak classifiers in a boosted cascade. In Proceedings of IEEE ICPR* 2012, Japan

[3] O. Tuzel, F. Porikli, and P. Meer: *Region Covariance: A Fast Descriptor for Detection and Classification*. In ECCV 2006.

[4] F. Crete, T. Dolmiere, and P. Ladret, *The blur effect: perception and estimation with a new no-reference perceptual blur metric*, In: Proceedings of SPIE Vol. 6492

[5] W. Ni, N.-S. Vu, and A. Caplier, *Lucas-Kanade based entropy congealing for joint face alignment*, Image and Vision Computing, Elsevier 2012

[6] N.-S. Vu and A. Caplier, *Illumination-robust face recognition using retina modeling*, In: IEEE ICIP 2009, Egypt.

[7] N.-S. Vu and A. Caplier. *Enhanced Patterns of Oriented Edge Magnitudes for Face Recognition and Image Matching*. IEEE Transactions on Image Processing, Vol. 21, no.3, March 2012.

[8] N.-S. Vu, *Exploring Patterns of Gradient Orientations and Magnitudes for Face Recognition*. IEEE Transactions on Information Forensics and Security, 2013, To Appear.

[9] S. Schwab, T. Chateau, C. Blanc, L. Trassoudaine, *A multi-cue spatio-temporal framework for automatic frontal face clustering in video sequences*. Eurasip Journal on Image and Video Processing 2012

[10] L. Zhang, Y. Li, R. Nevatia, *Global data association for multi-object tracking using network flows*, IEEE, CVPR 2008

[11] S. Schwab, T. Chateau, C. Blanc, L. Trassoudaine, *Suivi de visages par regroupement de détections: traitement séquentiel par blocs*, RFIA 2012