The source code for this work is made publicly available at: https://github.com/joelb92/Smart-Pose-Facial-Matching

EXPLORING FACIAL REGIONS IN UNCONSTRAINED SCENARIOS: EXPERIENCE ON ICB-RW

Ester Gonzalez-Sosa
Ruben Vera-Rodriguez
Julian Fierrez
Javier Ortega-Garcia
Universidad Autonoma de Madrid

Previous works have studied the potential of using facial regions instead of the whole face in biometrics for unconstrained scenarios.[16-17] In Bonnen, Klare, and Jain,[16] four facial regions (eyebrows, eyes, nose, and mouth) were used, conducting some of the experiments with the ARFace, a database with fixed occlusions in a constrained scenario (high resolution with controlled illumination and pose). In Tome et al.,[17] additional face regions were considered (up to 15) using the SCFace database. This database simulates a forensic scenario, including mugshot and CCTV images. This database, though, is not completely realistic, as users cooperate with the system (controlled pose) and the illumination is also controlled. In the present work, we explore face recognition through facial regions on the QUIS-CAMPI dataset. This database is one of the most challenging forensic databases in the literature, as it comprises mugshot images and CCTV images acquired in fully unconstrained scenarios without any cooperation from the users. The CCTV images have variations in pose, occlusions, illumination, distance, expression, etc. Please notice that our intention in this work is not to beat state-of-the-art approaches, but to give an insight into the potential use of facial regions in unconstrained scenarios. Our main objective in this line of work is to devise general face methods to exploit region-based face processing applicable to existing matches. We really believe this region-based processing will benefit even the most advanced face recognition approaches (e.g., based on deep learning) when confronted by challenging scenarios such as the one represented in the ICB-RW competition. With this vision in mind, the present work presents an example of how a robust face matcher based on SIFT can be improved by also considering frontalized facial regions.

Preprocessing, Feature Extraction, and Matching

Preprocessing

Figure 1 shows the general scheme followed in this work. The face is detected using the bounding box information provided as metadata by ICB-RW organizers. The preprocessing stage involves grayscaling, illumination normalization,[18] and resizing (320 x 320). As facial region extraction highly depends on the subject pose, we frontalize the face using the software provided by Hassner et al.[2] The frontalization process involves the estimation of a projection matrix between a query image and a standard 3D reference.

Feature Extraction

Two different features are computed for each image: (i) local binary patterns (LBP) of 9 facial regions and (ii) scale invariant feature transform descriptors (SIFT) of the whole face.

1. Local Binary Patterns of Facial Regions (LBP): In this work, 9 facial regions are extracted from the frontalized face: right eye, left eye, left eyebrow, right eyebrow, nose, mouth, chin, eyes, eyebrows, and face. First, a set of 68 landmarks are extracted through active shape modeling (ASM). Each facial region is extracted from the location of some landmark points as described in Gonzalez-Sosa et al.[19] Then, the facial region is divided into 10 x 10 blocks. The histogram of LBPs (59 uniform patterns) is

computed per each block. The final feature vector of a facial region is the concatenation of the different histograms of LBP computed per block.

2. Scale Invariant Feature Transform Descriptors (SIFT): While local binary patterns highly depend on the spatial correlation between images, SIFT features are more robust against changes in scale and rotations; therefore, they may be more suitable for comparing images without frontalization. In our implementation, SIFT descriptors are computed using cells of 6 x 6 pixels around keypoints and 16 orientations.

Matching

For SIFT descriptors, the similarity between two single images is defined as the number of matched keypoints between the two images, given a certain threshold. The dissimilarity between two LBP descriptors of two facial regions is computed using the Euclidean distance, followed by a normalization by the dimension of the particular facial region feature, to assure that all facial regions contribute similarly.

Experimental Protocol

The QUIS-CAMPI training set is composed of 3 mugshot images and 5 CCTV images per user. In our submitted approach, we only use the frontal mugshot image and the 5 CCTV images as the training images of a particular watch-list subject. At the evaluation phase, we have a test CCTV image, which is preprocessed and frontalized as described earlier. We apply one to one comparisons between the test CCTV image and all the training images belonging to the same watch-list subject before estimating the final score. If frontalization succeeds, these comparisons are carried out using LBP descriptors extracted from 9 facial regions; SIFT descriptors are used otherwise. The final score between a test CCTV image and a watch-list subject derives from the combination of the individual scores that result from the comparisons of the test CCTV image with each of the training images. This combination function depends on the specific face recognition system employed:

1. SIFT-based system: The final similarity score is the maximum of the 6 individual similarities.
2. Frontalized Region-based system: When attempting to compute a final similarity score, we address a N x 9 matrix of similarities, where N is the number of training images from a particular watch-list subject that have been successfully frontalized, and 9 is the number of facial regions considered in each individual comparison. The final score is the sum of the best 5 facial region similarities, having previously chosen the maximum similarity of each facial region.

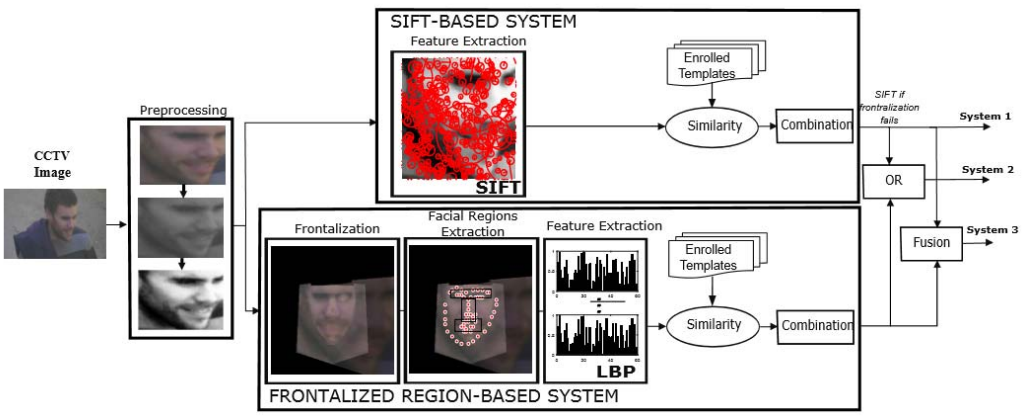

Figure 1. General scheme of the different systems considered in this work: system 1 (baseline), system 2 (submitted) to the ICB-RW Competition, and system 3 (improved).

Results

Results are reported in terms of identification task with rank-1, rank-5, and Area Under the Curve (AUC) between 0 and 1 for the QUIS-CAMPI dataset. Figure 2 shows the cumulative match characteristic curves for the three different systems considered:

- System 1 (baseline): Using only SIFT descriptors (R1=20.0; R5=34.0, AUC=0.69).
- System 2 (submitted): Based on LBP facial regions or SIFT descriptors, depending on the frontalization (R1=24.0; R5=39.1, AUC=0.73).
- System 3 (improved): Based on the fusion of SIFT descriptors and LBP facial regions or only SIFT descriptors, depending on the frontalization (R1=34.2; R5=48.6, AUC=0.80).

The submitted approach improves the baseline system from 0.69 to 0.73 in terms of AUC, and also improves rank-1 and rank-5 rates. The frontalization and the possibility of using similarities of facial regions coming from different training images of the subject may be the reason for this improvement. A big performance improvement is seen with the improvised fusion in which an AUC of 0.80 is obtained, yielding a 15.94 percent relative improvement with respect to the baseline system. Concerning rank-1 rates, there is an absolute improvement of 14.2 with respect to the baseline system. This is due to the complementary information coming from the fusion of SIFT descriptors and the LBP facial regions (when frontalization is possible).

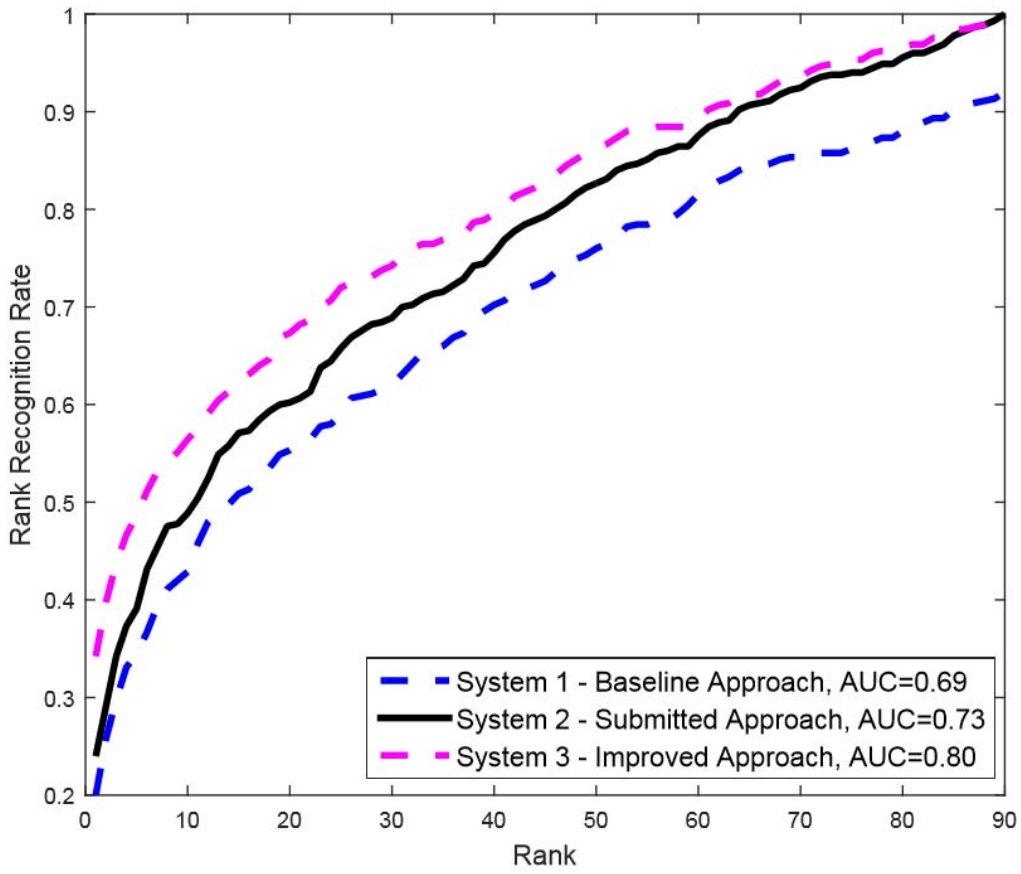

Figure 2. Cumulative Match Characteristic curves for system 1 (baseline), system 2 (submitted), and system 3 (improved).

Conclusion

This work explores the problem of face recognition in real unconstrained scenarios using a facial region approach. Our approach aims to be robust against challenging scenarios, either by using descriptors robust to rotations and changes in scales, or using texture information from different facial regions extracted from a frontalized face. It also introduces a combination function to estimate the best final score among a test CCTV and the training images. Finally, we propose an improved system based on the combination of complementary information coming from SIFT and LBP descriptors that outperformed significantly the submitted approach.

Acknowledgments

This work has been partially supported by project CogniMetrics TEC2015-70627-R (MINECO/FEDER). E. GonzalezSosa is supported by a PhD scholarship from Universidad Autonoma de Madrid.

UNSUPERVISED FACE RECOGNITION IN THE WILD

Michele Nappi
University of Salerno

Daniel Riccio
University of Naples Federico II

Luigi de Maio
Biometric and Imaging Processing Laboratory (BIPLab)

The soaring number of video surveillance cameras that are installed in public places makes face recognition from video surveillance an increasingly important task. In this contribution we present a new unsupervised face identification framework that searches faces extracted from video frames among a set of enrolled identities, which represent a gallery of known persons.

Face recognition from video is attracting ever increasing attention from both academic laboratories and industries, due to its high potential in many real world security applications. Most of the present methods deal with face recognition from videos that supply face images with high-resolution and favorable conditions in terms of pose and illumination.

This scenario is quite far from that, characterized by real video-surveillance applications, where low resolution cameras acquiring unaware people often provide low-quality face images, which are affected by large distortions in terms of non-frontal pose and/or uneven illumination. The main goal of researchers in this field is filling this gap.

As classifying faces acquired in uncontrolled settings is a complex task, most of the present methods are supervised approaches. They rely on a preliminary training stage on labeled faces to learn the structure of the feature space aiming to optimize the separation among different classes.

However, unsupervised methods show the advantage of classifying faces without any previous knowledge of the class distribution. This represents a desirable property when dealing with a large number of clusters with little labeled data.

This contribution proposes a complete framework, namely Unsupervised Face Recognition in the Wild (UFRW) for face recognition in video-surveillance applications, where few pictures per person are provided as enrolled identities that must be identified in single video frames that are submitted to the system as probes.

The UFRW biometric system has been tested in the ICB-RW 2016 challenge, where the goal was to identify persons appearing in video-surveillance frames (still images). Objects to be identified were also provided with high quality images that have been used for enrolling them into the system.

The whole pipeline of UFRW is shown in Figure 1.