A study of identification performance of facial regions from CCTV images

T. Ali[a], P. Tome[b], J. Fierrez[b], R. Vera-Rodriguez[b], L.J. Spreeuwers[a] and R.N.J. Veldhuis[a].

[a]University of Twente,
Faculty of EEMCS, P.O. Box 217, 7500 AE Enschede, The Netherlands
{t.ali, L.J.Spreeuwers, R.N.J.Veldhuis}@ewi.utwente.nl

[b]Biometric Recognition Group - ATVS,
Escuela Politecnica Superior - Universidad Autonoma de Madrid,
Avda. Francisco Tomas y Valiente, 11 - Campus de Cantoblanco 28049 Madrid, Spain
{pedro.tome, julian.fierrez, ruben.vera}@uam.es

Abstract. This paper focuses on automatic face identification for forensic applications. Forensic examiners compare different parts of the face image obtained from a closed-circuit television (CCTV) image with a database of mug shots or good quality image(s) taken from the suspect. In this work we study and compare the discriminative capabilities of different facial regions such as eye, eyebrow, mouth, etc. It is useful because it can statistically support the current practice of forensic facial comparison. It is also of interest to biometrics as a more robust general-purpose face recognition system can be built by fusing the similarity scores obtained from the comparison of different individual parts of the face. For experiments with automatic systems, we simulate a very challenging recognition scenario by using a database of 130 subjects each having only one gallery image. Gallery images are frontal mug shots while probe set consist of low quality CCTV camera images. Face images in gallery and probe sets are first segmented using eye locations and recognition experiments are performed for the different face regions considered. We also study and evaluate an improved recognition approach based on AdaBoost algorithm with Linear Discriminant Analysis (LDA) as a week learner and compare its performance with the baseline Eigenface method for automatic facial feature recognition.

1 Introduction

The difficulty of automatic face recognition mainly depends on the type of facial images we want to compare. A lot of research has been carried out to perform automatic face recognition and as a result several systems are available [1–6]. Problems such as different facial expressions, illumination conditions and poses have been studied and to certain extent some solutions have been proposed [2, 7, 8]. A relatively less investigated problem is the automatic face recognition from low quality images taken using CCTV camera. To date, there is no automatic system available which can reliably compare CCTV images with high quality images in mug shot database or image(s) taken from the suspect. This task is manually performed by forensic examiners where instead of following a holistic approach they use a "feature-based" approach Each part such as nose, eyes, mouth, etc. is compared separately and a conclusion is reached by observing similarities and differences. Finally conclusions based on the different facial features along with the relative importance of each is used to state an opinion in the form of a ratio of how likely is that the two images being compared are obtained from the same person to the ratio of how likely is that the two images being compared are obtained from different persons [8, 9].

The task of facial feature comparison is very challenging when one or both images under consideration are taken using CCTV camera because of the low quality. An automatic system comparing individual facial features is highly desirable as it will not only make the manual comparison of forensic examiners faster but will also help standardize this process. It is not possible with current state-of-the art recognition technologies to replace the manual comparison process in forensic face recognition; however, an automatic system can reduce, to a great extent, the manual effort. This can be, for instance, displaying top 10 candidate matches from a database of thousands of images based on a facial feature extracted from a criminal face image taken at a crime scene from a CCTV camera. Individual facial feature recognition is also important in cases such as having partial occlusion of the face and when only one facial feature is visible. In such cases even state-of-the art commercial face recognition systems such as [6] fail to work. Studies like the one presented in this paper are also necessary to scientifically support and help to establish procedures to assign relative weights to the opinions that can be inferred from different parts of the face.

In this paper we study the recognition performance of different facial features using two automatic recognition systems. The first system is the baseline Eigenface approach [4] while the second system is based on AdaBoost algorithm where we use LDA as a weak learner. The remaining of this paper is organized as follows: section 2 reviews the protocol followed by forensic examiners to carry out the facial comparison which is the main motivation for this work. Section 3 describes the database, evaluation protocol and the segmentation of face images. Section 4 briefly describes the improved boosting-based LDA approach. Experimental results based on the Eigenface method and the boosting approach are presented in section 5. Finally, in section 6 we draw conclusions and mention future research directions.

2 Forensic Examiners' Facial Comparison

In this section we briefly review the forensic experts way of facial comparison which is the main motivation behind our work. The discussion is based on the guidelines set forward by the workgroup on face comparison at Netherlands Forensic Institute (NFI) [10, 11] which is a member of European Network of Forensic Science Institutes [12]. The facial comparison is based on morphologicalanthropological facial features. In most cases the pictures are obtained or processed to be in the same posture. The comparison mainly focuses on:

- Shape of mouth, eyes, nose, ears, eyebrows, etc.
- Relative distance among different relevant facial features
- Contour of cheek- and chin-lines
- Lines, moles, wrinkles, scars, etc. in the face

When comparing faces manually, it should be noted that differences can be invisible due to underexposure, overexposure, resolution too low, out-of-focus and distortions in imaging process, specifically when considering information from surveillance camera. Furthermore, similar facial features can result in different depictions due camera position regarding the head, insufficient resolution, difference in focusing of two images, and distortion in imaging process.

Due to the aforementioned effects, which usually make the comparison process difficult, the anthropological facial features are visually compared and classified as: similar in details, similar, no observation, different, different in details. Apparent similarities and differences are further evaluated by classifying facial features as: weakly discriminative, moderately discriminative, and strongly discriminative. A conclusion based on this comparison process is a form of support for either the prosecutions or defense hypothesis and can be stated as no support, limited support, moderate support, strong support, and very strong support. The process is subjective to great extent and the conclusion of one expert can be different than other. The final result is based on the combination of the comparison results of different individual features. This is in contrast to automatic biometric face recognition systems where the whole face image is usually considered as a single entity [2, 4].

3 Database description and face segmentation

We use SCFace database [13] in our experiments which consists of 130 subjects each having one frontal mug shot image and 5 surveillance camera images. This database presents novel and challenging tests for automatic face recognition systems due to the very low quality images taken by surveillance cameras. A few examples of mug shot and surveillance camera images used in our experiments are shown in figure 1. There are five different surveillance cameras used each with three different distances from the subjects. For simplicity in our experiments we consider only one surveillance camera with the closest distance to the subjects.

All of the frontal mug shot and surveillance camera images are segmented using the ground truth locations of the eyes. Segmentation of the face image into different parts is based on standard facial proportions [14]. An example of the set of segments into which a face image is divided is shown in figure 2. As shown in figure 2, pixels outside the region of interest are masked by setting them to zero. Given a probe patch of a facial feature extracted from a surveillance camera image, it is matched with each of the 130 patches extracted from the frontal mug shot images.

Fig. 1. A few sample gallery (first row) and probe images (second row) used in our experiments.

4 Facial Feature Recognition

To handle the complex nature of individual facial feature recognition from low quality CCTV images we use LDA [15] as a weak learner in Adaboost.M2 [16] for feature [1] extraction while classification is performed using simple Euclidean distance. The performance of traditional LDA-based approach [3] is improved by incorporating it in the boosting framework. Since both LDA and AdaBoost are well known algorithms we only provide a brief description of our employed recognition system highlighting the way LDA is integrated in AdaBoost.M2. Each round of boosting generates a new LDA subspace particularly focusing on examples which are misclassified in previous LDA subspace. The final feature extractor module is an ensemble of several specific LDA solutions. In order to incorporate LDA in boosting framework, slight modifications are introduced in the way the within-class and the between-class scatter matrices are constructed at the end of each boosting iteration by incorporating the weight associated with each sample. Please refer to [5] for a detailed description of using LDA as a weak leaner in AdaBoost algorithm.

[1] Here the term "feature" refers to a vector of values describing the characteristics of an image patch. This is the common use of the term "feature" in pattern recognition. In order to avoid ambiguity we always use the term "facial features" for referring to the parts of the face such as eye, eyebrow, nose, etc.

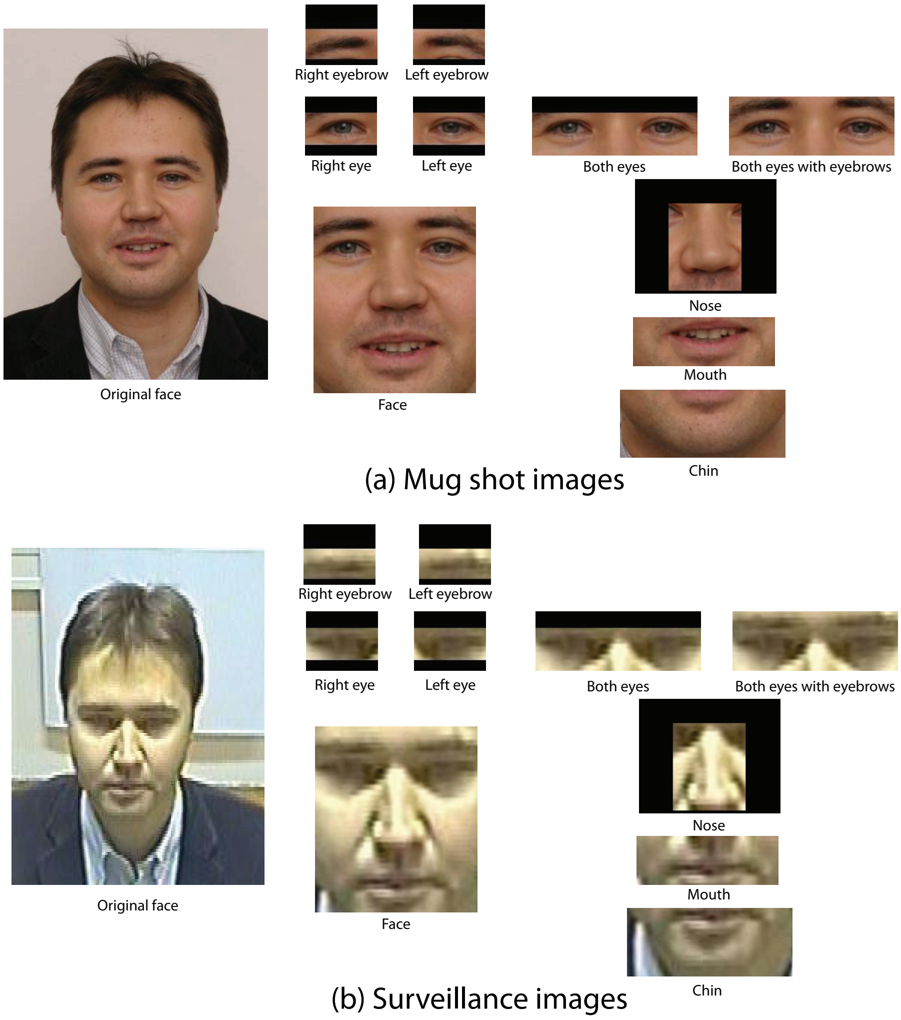
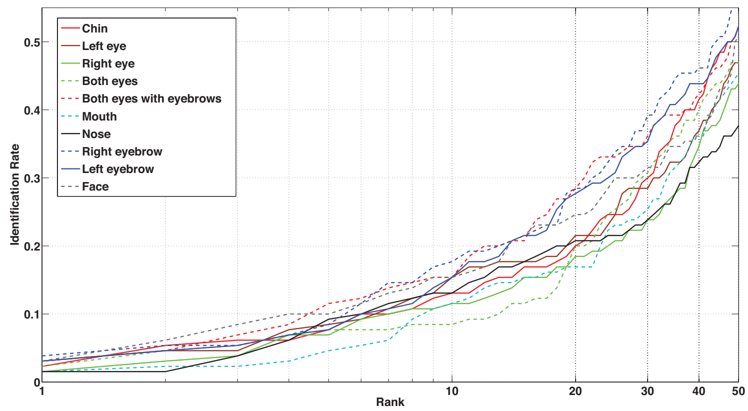


Fig. 2. (a) Mug shot images (b) Surveillance camera images.

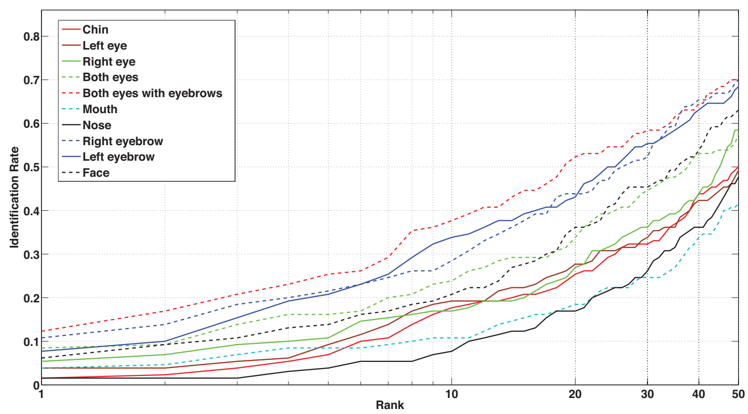
This kind of ensemble based approach takes advantage of both LDA and boosting and outperforms simple LDA based systems in complex face recognition tasks. This is particularly important where a small number of training samples for each subject are available (1 image patch per facial feature in this case) compared to the number of dimensions of the samples i.e., the small-sample-size problem [17] and when non-linear variations are present in facial images. Our employed face recognition system is more robust when performing recognition of low resolution face images. This result is also verified by the authors in [5] where they use similar approach for face recognition.

5 Experimental Results

There are 130 subjects each having only one image both in gallery and probe sets. Each face image is segmented and as a result we have 130 patches for each facial feature both in gallery and probe set. Figure 3 (a) shows the Cumulative Match Characteristics (CMC) curves of different facial feature when the Eigenface [4] method is applied to this close-set identification task. Only components whose eigenvalues are equal to or greater than 1 are retained. Simple Euclidean distance is used for classification. Very low identification results are observed mainly due to very low quality probe patches obtained from surveillance camera images, only one training sample, and relatively high size of gallery. For the same identification scenario, we see improved identification rates for all facial features using the AdaBoost approach discussed in section 4 (figure 3 (b)).



(a) Eigenface approach



(b) Adaboost approach

Fig. 3. Identification performance of different facial features.

	Left eye	Right eye	Left EB	Right EB	Mouth	Nose	Eyes with EB	Eyes	Chin	Face
Eigenface approach	2.31	1.54	3.08	3.85	1.54	1.54	3.08	2.31	2.31	3.84
AdaBoost approach	3.85	5.39	7.69	10.77	3.85	1.54	12.31	8.46	2.31	6.15

Table 1. Rank 1 identification rate (%) (EB stands for eyebrow).

	Left eye	Right eye	Left EB	Right EB	Mouth	Nose	Eyes with EB	Eyes	Chin	Face
Eigenface approach	11.54	11.54	15.38	17.69	11.54	13.08	15.38	8.46	13.08	15.38
AdaBoost approach	19.23	16.92	33.85	28.46	10.77	7.69	37.69	23.85	17.69	20.77

Table 2. Rank 10 identification rate (%) (EB stands for eyebrow).

Table 1-2 list the rank 1 and rank 10 identification rate of each facial feature. It can be concluded that different automatic systems might rank different facial feature differently with respect to their discriminative capabilities. It is important to note that since the segmentation process is based on eyes location, the eye regions are expected to be better aligned than other regions. However, it is a standard practice in automatic face recognition to locate eyes positions and normalize face based on eyes positions.

Besides identification performance, it is also important to consider performance in verification scenario. In forensic facial comparison, a simple verification situation happens when an image from a suspect is compared with an image obtained from a crime scene. This is a one-to-one comparison for which different evaluation metrics such as Area under Receiver Operating Characteristics (ROC) curve (verification rate vs. false acceptance rate) and Equal Error Rate (EER) are used. In our experiments we use area under the ROC to summarize the verification performance of both systems for different facial features. Higher value of area under the ROC implies better verification performance of a system and vice versa. Table 3 summarizes results of verification experiments using the area under ROC curve metric. In table 4 we rank facial feature according to their verification performance using each method.

6 Conclusions and Future Work

Comparing individual facial features of two or more faces is a common practice that forensic examiners carry out during their investigation of a crime when there

	Left eye	Right eye	Left EB	Right EB	Mouth	Nose	Eyes with EB	Eyes	Chin	Head
Eigenfaces	57	56	61	63	55	53	60	56	60	59
AdaBoost approach	63	66	75	77	55	57	79	69	61	72

Table 3. Verification performance using percentage of area under ROC (EB stands for eyebrow).

Eigenface method	**Proposed method**
Right eyebrow	Both eyes with eyebrow
Left eyebrow	Right eyebrow
Both eyes with eyebrows	Left eyebrow
Chin	Face
Face	Eyes
Left eye	Right eye
Eyes	Left eye
Right eye	Chin
Mouth	Nose
Nose	Mouth

Table 4. Ranking facial feature based on verification performance.

are facial images from a crime scene. In this paper we presented preliminary experiments to compare and evaluate the discriminative capabilities of different facial features. We studied a boosting based LDA approach and compared its performance with the standard Eigenface method for individual facial feature recognition. The studied method has shown improved performance, however, still it is far from the point where it can be used in real applications. It is however important to study and understand the recognition performance of different facial features by recognition algorithms. This can lead to future research such as building more robust face recognition systems by the weighted sum of all facial features recognition results. Also it is more important in cases where crime scene images are partially occluded or only a few facial features are visible. Our future research will include improving the recognition performance as well as combining evidence from different facial feature comparison to single evidence for forensic face recognition.

Acknowledgments.

This work is carried out during the secondment of the first author at ATVS, Autonomous University of Madrid. The research is funded by the European commission as a Marie-Curie ITN-project (FP7-PEOPLE-ITN-2008) under grant agreement number 238803 and the Spanish Direccion General de la Guardia Civil (DGGC). The authors would like to thank Nicomedes Exposito, Francisco J. Vega, Patricio Leston and Pedro A. Martinez for their valuable comments.

References

1. Abatea, A.F., Nappi, M., Riccioa, D., Sabatino, G.: 2D and 3D face recognition: a survey. Pattern Recogn. Lett. 28, 1885-1906 (2007)
2. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Survey. 35, 399-458 (2003)
3. Belhumeur, P. N., Hespanha, J.P., Kriegman, D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence 19, 711-720 (1997)
4. Turk, M., Pentland, A. Eigenfaces for recognition. J. Cogn. Neurosci. 3, 72-86 (1991)
5. Lu, J., Plataniotis, K. N., Venetsanopoulos, A. N. Boosting linear discriminant analysis for face recognition. Proc. IEEE Int. Conf. Image Process., vol. 1, pp.657-660 (2003)
6. http://www.cognitec-systems.de/FaceVACS-SDK.19.0.html
7. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognition-a review. Comput. Vis. Image Underst. 97, 103-135 (2005)
8. Ali, T., Spreeuwers, L.J., Veldhuis, R.N.J.: Forensic Face Recognition: A Survey. Face Recognition: Methods, Applications and Technology, Computer Science, Technology and Applications. Nova Publishers, 9, ISBN 978-1-61942-663-4 (2012)
9. Ali, T., Spreeuwers, L.J., Veldhuis, R.N.J.: Towards automatic forensic face recognition. ICIEIS, Malaysia, CCIS, Springer Verlog 252, pp. 47-55 (2011)
10. Netherlands Forensic Institute, http://www.forensicinstitute.nl
11. General method of facial comparison, Addendum 1, internal document, Netherlands Forensic Institute
12. European Network of Forensic Science Institutes, http://www.enfsi.eu/
13. Grgic, M., Delac, K., Grgic, S. SCface - surveillance cameras face database. Multimedia Tools and Applications Journal. 51, 863-879 (2011)
14. Tome, P., Blazquez, L. Face regions segmentation based on facial proportions. Technical Report, ATVS - Biometric Recognition Group, Universidad Autonoma de Madrid, (2012)
15. Duda, R. O., Hart, P. E., Stork, D. H. Pattern Classification (2nd ed.). Wiley Interscience, ISBN 0-471-05669-3, MR 1802993 (2000)
16. Freund, Y., Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55, 119-139 (1997)
17. Raudys, S.J., Jain, A.K. Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners. IEEE Trans. Pattern Analysis and Machine Intelligence 13, 252-264 (1991)