Exploiting representation plurality for robust and efficient face recognition

Vitomir Štruc[1]**, Jerneja Žganec Gros**[2]**, Simon Dobrišek**[1]**, Nikola Pavešić**[1]

[1]*Faculty of Electrical Engineering, University of Ljubljana*
Tržaška 25, SI-1000 Ljubljana, Slovenia
[2]*Alpineon d.o.o., Ulica Iga Grudna 15, SI-1000 Ljubjana, Slovenia*
E-mail: {vitomir.struc, simon.dobrisek, nikola.pavesic}@fe.uni-lj.si, jerneja.gros@alpineon.si

Abstract

The paper introduces a novel approach to face recognition that exploits plurality of representation to achieve robust face recognition. The proposed approach was submitted as a representative of the University of Ljubljana and Alpineon d.o.o. to the 2013 face recognition competition that was held in conjunction with the IAPR International Conference on Biometrics and achieved the best overall recognition results among all competition participants. Here, we describe the basic characteristics of the submitted approach, elaborate on the results of the competition and, most importantly, present some general findings made during our development work that are of relevance to the broader (face recognition) research community.

1 Introduction

Despite the progress made during the last decades, the existing face recognition technology still suffers from some drawbacks that relate mainly to the use of the technology in uncontrolled and unconstrained environments. In such environments the technology commonly achieves much lower recognition rates than under controlled and supervised conditions.

Two dominant research directions have emerged to address this problem. The first focuses on extracting invariant yet discriminative facial features which can be used with robust classification approaches to cope with facial appearance changes caused, for instance, by illumination, expression or pose changes. Examples of techniques from this research direction can be found in [1] or [2]. The second research direction with respect to robust face recognition focuses on combining several facial representations and overcoming externally induced facial appearance changes by considering information in different feature spaces. Examples of techniques from this direction are presented in [3] or [4].

This work has been supported in parts by the national research program P2-0250(C) Metrology and Biometric Systems, the postdoctoral project BAMBI (ARRS ID Z2-4214), by the European Union, European Regional Fund, in scope of the framework of the Operational Programme for Strengthening Regional Development Potentials for the Period 2007-2013, contract No. 3211-10-000467 (KC Class) and European Union's Seventh Framework Programme (FP7-SEC-2011.20.6) under grant agreement number 285582 (RESPECT). The authors additionally appreciate the support of COST Actions IC 1106 and IC1206.

This paper introduces an novel approach to face recognition that can be classified among the techniques from the second research direction, as it exploits representation plurality to achieve robust recognition. The proposed approach crops the facial region from each input image in three different ways. It then maps each of the three crops into one of four color representations and finally extracts several feature types from each of the twelve facial representations. As will be shown, the described procedure results in a total of thirty facial representations that are combined at the matching score level using a fusion approach based on linear logistic regression (LLR).

The presented approach was recently enlisted[1] among the participants of the 2013 face recognition competition that was held in conjunction with the IAPR International Conference on Biometrics, which is one of the premier conferences on the subject of biometrics. The competition was conducted on the MOBIO database [5], which features facial images taken in uncontrolled conditions using various mobile devices. The enlisted approach achieved the best overall performance among all participants and demonstrated highly robust performance [6]. The goal of this paper is two-fold: *i)* to describe the system enlisted in the 2013 ICB face recognition competition, and, more importantly, *ii)* to present some general findings made during our development work.

The rest of the paper is structured as follows: Section 2 introduces the novel plurality-based approach, but adjusts the description to the specifics of the competition. In Section 3 experiments conducted during the development work are presented. These experiments assess different aspects of the proposed approach and are important to the wider face recognition community as well. Comparative results with other participants are also presented in this section. The paper concludes with some final comments in Section 4.

2 Proposed methodology

2.1 Overview

Before we start with the description of the novel approach, please note that the eye-coordinates of all facial images were provided with the MOBIO database by the compe-

[1]As a representative of the University of Ljubljana and Alpineon d.o.o.

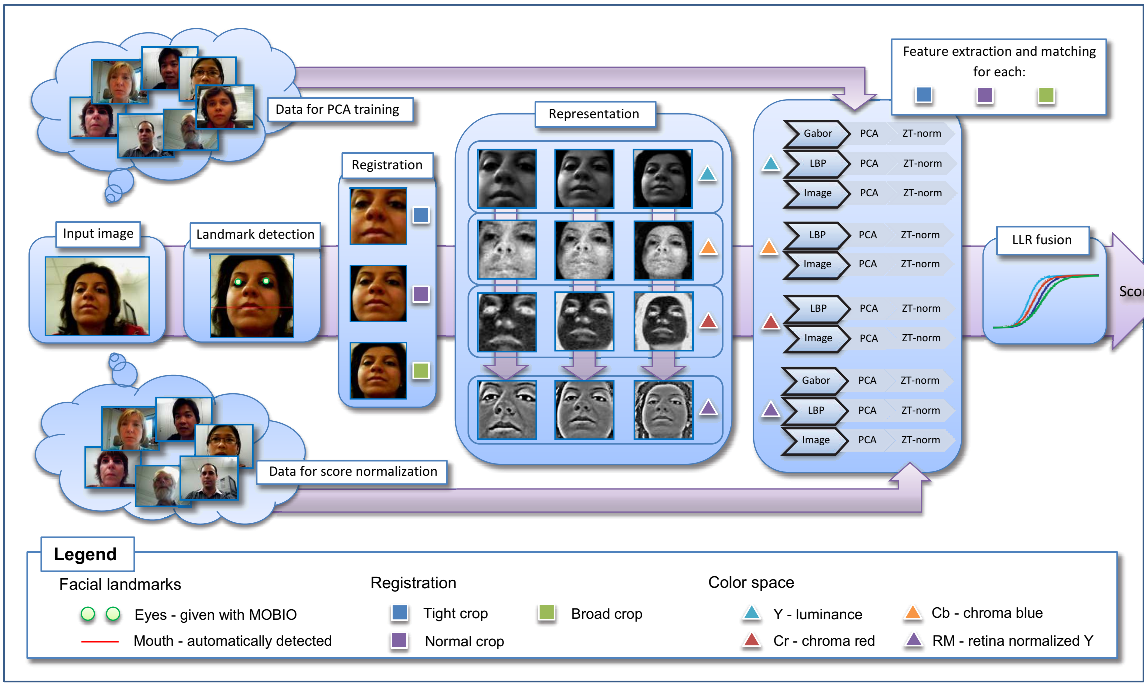


Figure 1: A simplified diagram of the proposed approach

tition organizers. A face detection and eye localization procedure was, therefore, unnecessary. An overview of our approach is shown in Fig. 1. Clearly, its most basic characteristic is plurality of facial representations.

In the first step, the proposed approach uses the manually marked eye-coordinates and an automatically detected mouth location to geometrically normalize each given facial image. Once the image is normalized, the system crops the facial area from the input image based on bounding boxes of three different sizes. The cropping procedure results in three distinct facial areas (a tightly, a normally and a broadly cropped one), each of which is then represented in the $YCbCr$ color space. In the next step, the luminance component of the cropped images (i.e., Y) is subjected to a *retina-modeling*-based photometric normalization procedure [7]. This processing step produces a normalized version of the facial image (denoted as Pn in the remainder) and together with the remaining color components (i.e., Y, Cb, and Cr components) forms the basis for feature extraction. In the feature extraction step, various image descriptors, such as intensity values, Gabor features and LBP features are first computed from the image representations and then subjected to PCA for dimensionality reduction. In the last step, all (subspace) feature vectors corresponding to the given input test image are matched against the corresponding enrollment feature vectors to produce similarity scores, which are normalized using a special type of ZT normalization and ultimately combined using a variant of the recently proposed *temporal fusion* [8], [6].

2.2 Preprocessing

The preprocessing procedure of the proposed approach comprises two distinct steps as already outlined in the previous section, i.e., mouth detection and geometric normalization of the facial area. The mouth detection step of the approach is conducted based on the technique proposed by Pozne in [9]. Using the provided eye-coordinates and the detected mouth position, the facial area is then geometrically normalized in accordance with the procedure illustrated on the left side of Fig. 2. The image is first rotated in such a way that the line connecting the eyes is in a horizontal position, while the facial area is then cropped in respect with the inter-occular distance and rescaled to a fixed size of 128×128 pixels. Adjusting the value of the coefficients k_1, k_2, k_3, and k_4 results in differently cropped facial areas. For our approach we use: $k_1 = k_2 = 0.8, k_3 = 0.4, k_4 = 0.3$ for the tight crop, $k_1 = k_2 = 1.0, k_3 = 0.6, k_4 = 0.7$ for the normal crop and $k_1 = k_2 = 1.25, k_3 = 1.0, k_4 = 1.0$ for the broad crop. Some examples of the cropped images are shown on the right side of Fig. 2. Note that all regions of the bounding box that are outside of the image area are padded with zeros.

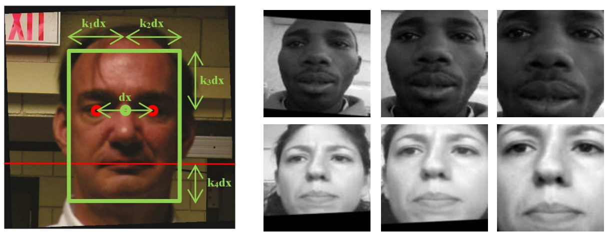


Figure 2: Geometric normalization: illustration of the procedure (left), sample cropped intensity images (right)

2.3 Representation plurality

After the preprocessing procedure the face is represented in three different ways - with a tightly, normally and broadly cropped facial area. To exploit representation plurality at all levels, we transform the initial RGB color

Figure 3: Face representation with color components (from left to right): Y, Cb, Cr components, Pn

Table 1: Summary of feature representations for each crop type

	Y	Cb	Cr	Pn
Intensity values	•	•	•	•
Gabor	•	-	-	•
LBP	•	•	•	•

space of the facial images into the chromatic $YCbCr$ color space, which was found previously to be highly suitable for the task of face recognition [10], and represent each facial crop using each of the three color components of the $YCbCr$ color space. Additionally, we adopt the photometrical normalization procedure from [7] to normalize the Y component with respect to illumination and use the result Pn as our fourth color representation. The presented procedure results in 12 "color" representations (3 facial crops $\times$ 4 color channels), which form the basis for the feature extraction procedure (see Fig. 3).

To capture as much of the discriminative information of the facial representations as possible we derive three types of feature vectors from the facial areas, namely, feature vectors comprised of intensity values, Gabor magnitude responses [11] and local binary patterns (LBP-s) [12]. Note that we compute all three types of features only from the grey-scale and photometrically normalizes color channels, while the chromatic color components Cb and Cr are used only to derive feature vectors of intensity values and LBPs (see Fig. 1). This feature extraction procedure results in ten feature space representations for each type of facial crop, as shown in Table 1. In the final step of the feature extraction procedure, all feature representation are mapped into a PCA subspace to reduce the dimensionality of the feature vectors.

2.4 Matching, score normalization and fusion

The thirty feature representations (10 feature representations $\times$ 3 facial crops) obtained with the procedure described in the previous section are used individually in the matching procedure. Once an identity claim is made[2], the feature representations derived from the probe image are matched against the corresponding feature representations of the target image resulting in thirty matching (or similarity) scores. Here, the whitened cosine similarity measure is used to compute the matching scores. Prior to the fusion procedure, all scores are also normalized based on some background data using a special variant of $zt-$score normalization [13]. To combine the 30 $zt-$normalized matching scores into a single one that can form the basis for identity inference, we adopt the recently proposed temporal fusion procedure, proposed by Poh and Kittler in [8].

[2]Note that the approach was tested in verification mode.

3 Experiments and results

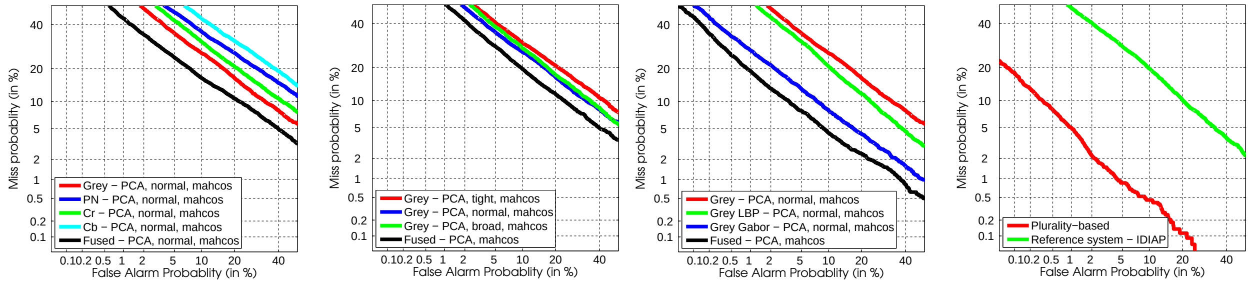
All of our experiments presented in the remainder of this section are conducted on the MOBIO database, which was also the database of choice for the competition. The database was recorded with mobile devices (mobile phones and notebooks) in uncontrolled conditions and, therefore, represents quite a challenge to the existing face recognition technology. The database was partitioned into three distinct sets for the competition: *i) the training set*, which was reserved for training background models (such as PCA or LDA transformation matrices, UBM models, etc.) and potentially for score normalization, *ii) the development set* , which was used to train any hyper-parameters of the adopted recognition approach (e.g., feature space dimensionality, fusion parameters, etc.), and *iii) the evaluation set*, which was employed to assess the final performance of the submitted approach. Note that only for images from the training and development sets identity information was available, while the evaluation set was shipped to the competition participants in an anonymized form. The final performance metrics and graphs were, therefore, generated by the organizers of the competition.

The experiments were conducted on subsets of images belonging to either female or male subjects. A total of approximately 95000 (client and impostor) verification attempts were conducted on the development set and approximately 190000 (client and impostor) verification attempts were conducted on the evaluation set.

To measure the performance of the submitted approaches the false acceptance (FAR) and false rejection rates (FRR) were used. These rates were also adopted to produce Detection Error Trade-off (DET) curves of the experiments. As quantitative evaluation metrics, the Equal Error Rate (EER) on the development set and the Half Total Error Rate (HTER) on the evaluation set were used as well. For a detail explanation of the employed metrics and performance graphs the reader is referred to [6].

During the development work we assessed the impact of various factors on the verification performance of our approach. The results, which were generated using images of female and male subjects together, are shown in Fig. 4. As we can see in Fig. 4 (a), different color representations result in different verification performance, while all representations combined (using LLR fusion) ensure the best results. Similar findings can also be observed in Fig. 4 (b) for the cropping type. The normally cropped facial area ensures the best results, followed in order by the broadly and tightly cropped areas. This indicates that information about the shape of the face is also of high importance to the recognition task. Last but not least, if we look at the performance of different feature types (Fig. 4 (c)), we can see that the best verification results were achieved with Gabor magnitude features, followed by LBPs and pure intensity values. Note that the combined plurality-based approach significantly outperformed the PCA+LDA baseline system provided by IDIAP (Fig. 4 (d)).

Once our approach was trained and all hyper-parameters were set on the development image set of the MO-



(a) Color channels (b) Crop type (b) Feature choice (c) Our approach vs. baseline

Figure 4: DET curves of the experiments. The results show the effect of color, crop type, feature choice on the verification performance of the proposed approach. The final plot compares the combined approach to the baseline approach from the competition

Table 2: Summary of participants

Institution	Short	System/approach
Baseline	Baseline	PCA+LDA
Centre de Développement des Technologies Avancées	CDTA	LBP-based system
University of Campinas - Harvard University	UC-HU	Hierarch. covolutional neural network
Tempere University of Technology	TUT	Gabor-LBP histogram sequence
Idiap Research Institute	Idiap	Gabor+LBP
University of Technology Sydney	UTS	Gabor phase + local phase quantization
Galician R&D Center in Advanced Telecommunicat.	GRAD	Gabor+oriented edge magnitudes
CPqD	CPqD	Four types of LBPs + SVM

Table 3: Results on the evaluation set (taken from [6])

Instit.	EER (f)	HTER (f)	EER (m)	HTER (m)
Baseline	14.7 %	20.9 %	14.8 %	17.1 %
CDTA	10.7 %	28.5 %	7.7 %	11.9 %
UC-HU	4.7 %	10.8 %	3.5 %	**6.2** %
TUT	8.6 %	13.9 %	7.3 %	11.5 %
IDIAP	6.2 %	12.5 %	6.6 %	10.3 %
UTS	7.5 %	13.6 %	6.1 %	12.0 %
GRAD	5.4 %	12.3 %	3.1 %	9.5 %
CPqD	6.3 %	11.2 %	5.5 %	7.7 %
Ours	**2.8** %	**10.5** %	**1.7** %	7.5 %

BIO database, we conducted experiments on the evaluation image set and submitted the results to the organizers. Seven other institutions submitted their results as well. A brief summary of the participating institutions and their systems is shown in Table 2. For a more detail description have a look at [6].

The results of the competition are shown for images of female subjects (marked with (f)) and images of male subjects (marked with (m)) in Table 3. Note that on the development set, where the EERs were computed, our approach clearly outperforms all other participants. On the evaluation set, where the HTERs were computed, our approach still performs the best for the experiments with images of female subjects and ranks in second on the male images. In retrospective, the fusion process overfitted to the development image set. A simpler fusion process would probably have resulted in better generalization over different image sets.

4 Conclusion

We have presented a novel approach to face recognition exploiting plurality of representation to achieve robust face recognition. We have described a recent face recognition competition, where the presented approach achieved the overall best performance and highlighted the importance of color, shape and feature type choice for the task of face recognition.

References

[1] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust Face recognition via sparse representation," *IEEE TPAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[2] M. Yang, L. Zhang, S. Shui, and D. Zhang, "Robust Kernel Representation With Statistical Local Features for Face Recognition," *IEEE TNNLS*, vol. 24, no. 6, pp. 900–912, 2013.

[3] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE TIP*, vol. 18, no. 8, pp. 1885–1896, 2009.

[4] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE TIP*, vol. 19, no. 6, pp. 1635–1650, 2010.

[5] C. McCool *et al.*, "Bi-modal person recognition on a mobile phone: using mobile phone data," in *IEEE ICME Workshop on Hot Topics in Mobile Multimedia*, 2012.

[6] M. Günther *et al.*, "The 2013 face recognition evaluation in mobile environment," in *The 6th IAPR International Conference on Biometrics*, Jun. 2013.

[7] N. Vu and A. Caplier, "Illumination-robust face recognition using retina modeling," in *Proceedings of ICIP*, vol. 1, 2009, pp. 3289–3292.

[8] N. Poh, J. Kittler, and F. Alkoot, "A discriminative parametric approach to video-based score-level fusion for biometric authentication," in *Proceedings of ICPR 2012*, vol. 1, 2012, pp. 2335–2338.

[9] A. Pozne, "Pridobivanje vizualnih značilk za samodejno razpoznavanje govora," PhD Thesis, Univerza v Ljubljani, Fakulteta za elektrotehniko, 2005.

[10] J. Yang, C. Liu, and J. Yang, "What Kind of Color Spaces Is Suitable for Color Face Recognition?" *Neurocomputing*, vol. 73, pp. 2140–2146, 2010.

[11] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for robust face recognition," *EURASIP JASP*, 2010.

[12] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *In Proc. of ECCV*, 2004, pp. 469–481.

[13] R. Auckenthaler, M. Careya, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *DSP*, vol. 10, no. 1–3, pp. 42–54, 2000.