One-Shot 3D-Gradient Method Applied to Face Recognition

J. Matías Di Martino(✉), Alicia Fernández, and José Ferrari

Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
{matiasdm,alicia,jferrari}@fing.edu.uy

Abstract. In this work we describe a novel one-shot face recognition setup. Instead of using a 3D scanner to reconstruct the face, we acquire a single photo of the face of a person while a rectangular pattern is been projected over it. Using this unique image, it is possible to extract 3D low-level geometrical features without the explicit 3D reconstruction. To handle expression variations and occlusions that may occur (e.g. wearing a scarf or a bonnet), we extract information just from the eyes-forehead and nose regions which tend to be less influenced by facial expressions. Once features are extracted, SVM hyper-planes are obtained from each subject on the database (one vs all approach), then new instances can be classified according to its distance to each of those hyper-planes. The advantage of our method with respect to other ones published in the literature, is that we do not need and explicit 3D reconstruction. Experiments with the Texas 3D Database and with new acquired data are presented, which shows the potential of the presented framework to handle different illumination conditions, pose and facial expressions.

Keywords: 3D face recognition · Differential 3D reconstruction

1 Introduction

Face recognition is one of the most popular and challenging problems in the field of pattern recognition and computer vision [1]. It has many applications such as security control and prevention, medical and biometrical analysis or gesture understanding. In the last decade, lot of research included three-dimensional (3D) face information to improve recognition rates and make the methods more robust to pose, gesture and illumination variations [3]. Bronstein et al. [4] used 3D facial scanners and achieved a robust recognition framework by modeling facial expressions as surface isometries, and constructing an expression-invariant face representation using the canonical forms approach. The work of Chang et al. [7] was one of the first that combined scores obtained from matching multiple overlapping regions around the nose. A similar method was proposed by Faltemier et al. [11] where 28 different regions around the face were selected and a score-based fusion approach was followed. Kakadiaris et al. [15] presented a 3D deformable model approach where the face was parametrized by the annotated

A. Pardo and J. Kittler (Eds.): CIARP 2015, LNCS 9423, pp. 176–183, 2015.
DOI: 10.1007/978-3-319-25751-8_22

face model (AFM). Mahoor et al. [18] used the principal curvature to represent the face image as a 3D binary image, then Hausdorff distance and iterative closest points (ICP) [2] were used for matching. More recently, Li et al. [17] used facial curves to form a rejection classifier and produce a facial deformation mapping from which adaptively select different face regions for matching. In Y. Zhang et al. [20] each 3D facial surface was mapped onto a 2D lattice that represents local 3D geometrical or textural properties, then traditional 2D face recognition techniques were applied. Lei et al. [16] presented a 3D face recognition approach based on low-level geometric features collected from the eyes-forehead and nose regions, and Support Vector Machine (SVM) algorithm was used to separate different subjects' representation.

In the present work instead of using a 3D scanner we acquire a standard color photo of the face while a rectangular pattern is been projected. We show that from this unique image, it is possible to extract low-level geometrical features without the explicit 3D reconstruction, as well as texture information. To handle expressions variations and occlusions that may occur by wearing a scarf or a bonnet, we extract information just from the eyes-forehead and nose regions which tend to be less influenced by facial expressions, as discussed by Lei et al. [16].

The rest of this work is organized as follows. Section 2 starts with a general description of the proposed framework and then some important specific steps are detailed. Section 3 presents some experimental results, firstly using the Texas 3D Face Recognition Database [12], and secondly, using our own database where all the steps of the proposed framework are involved. In section 4 we present the conclusions and discuss some future work.

2 Description of the Proposed Method

We will start by presenting an overview of the proposed framework, highlighting the main steps and properties. After that, some important individual steps will be described in detail.

The main steps of the proposed technique can be identified as: (i) acquisition; (ii) extraction of geometric and texture information; (iii) localization of the eyes and nose position; (iv) extraction of features from the rigid and semi-rigid regions of the face, and finally; (v) train/classify.

Once the texture and the 3D geometrical information is retrieved, we perform a curvature analysis to localize the nose tip and eyes corners.

From the corners of the eyes and nose tip position, we extract two different areas of the face. The first is a rectangular region that contains both eyes, and the second one is a trapezoidal region that contains the nose. As it was demonstrated in several works, these are the portions of the face more rigid and less affected by facial expressions (see e.g. [3,16] and reference therein).

The next step consists on defining features using the information available in the selected portions of the face. We analyze features computed from the geometrical 3D information and from the texture information. Then using a training dataset we define boundaries in the m-dimensional feature space using the Support Vector Machine (SVM) algorithm.

Figure 1 illustrates the main steps of the proposed framework and the kind of data processed at each step. Before presenting the experimental results, additional explanations and details of some key steps are covered in the following subsections.

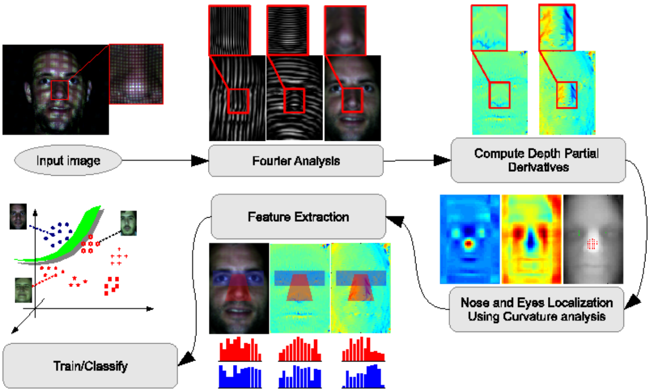

Fig. 1. Illustration of the main steps of the proposed framework. Note: due to image resolution some artifacts can appear, the full images, the code and a demo can be found at authors' web page.

2.1 Nose and Eyes Localization

To achieve a robust localization of the eyes and nose regions, we follow the procedure presented in [9] to estimate those candidate pixels. In our previous work [9] we address the problem of detecting the position of the nose and the eyes using face gradient information. Once candidates points for the nose and eyes are obtained, we remove those false positive nose and eyes detections by follow 2 basic steps. Firstly we estimate from the training set (in which eyes and nose position was manually marked) the likelihood distribution for the distance between the eyes and the distance between each eye and the nose (this step is performed once and in the training step). Secondly, we keep from all the nose and eyes candidates (obtained by curvature inspection) those who present inter-distances with higher likelihood.

2.2 Feature Extraction

Two different regions of the face will be considered for local feature extraction (as illustrated in Fig. 1). The first patch (named PatchA) corresponds to the regions around eyes and the second one (PatchB) corresponds to the region around the nose. The region of the nose, is the portion of the face less distorted by facial expressions, thus is called the rigid region. The area of the eyes is called semi-rigid region [16], and finally, the rest of the face is much more affected due to facial expression, and therefor was not considered in the present work.

Following the approach of [8], we calculate the partial derivatives D_x and D_y of the face depth (D); $D_{x,y}$ are directly obtained (from the acquired image) by measuring the deformation of the projected pattern. Then, using these partial derivative, over the PathA we have to calculate the directional derivatives D_v and $D_{v_\perp}$, where v is a direction parallel to the largest side of PatchA. Analogously D_u and $D_{u_\perp}$ are calculated over PatchB, where u is a direction parallel to the largest side of PatchB. Histograms for these quantities are computed to construct the 3D low level features descriptor, in addition, the LBP description over the Texture image is also taken into account.

2.3 Training and Classifying

The last step of the proposed framework consists on training a classifier for each subject we want to recognize. We decide to use Support Vector Machine (SVM) because it shows to be an efficient and robust algorithm for the sake of face recognition[16]. Recall that for a binary classification problem, where we assume that are known m training samples $x_k \in \mathbb{R}^N$ ($k = 1..m$), with labels $y_k \in \{-1, 1\}$, SVM finds the hyper-plane with largest margin by solving the optimization problem:

$$\begin{aligned} &\min_{\omega,b,\xi} \left(\frac{1}{2}\omega^T\omega + C\sum_i \xi_i \right) \\ &\text{s.t. } y_i\left(\omega^T x_i + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0. \end{aligned} \tag{1}$$

The parameter C is a penalty parameter of the error term and must be set; ω is a vector orthogonal to the hyper-plane, b is a constant that sets the location of the hyper-plane and ξ_i are auxiliary variables that allow to handle non separable problems. Depending on the problem, we may replace the constraint $y_i\left(\omega^T x_i + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0$ by $y_i\left(\omega^T \phi(x_i) + b\right) \geq 1 - \xi_i, \quad \xi_i \geq 0$ (Kernel SVM) which allows us to find the hyper-plane in a higher dimensional space. Once the SVM hyper-plane is obtained, one can classify new instances according to its positions with respect to the hyper-plane, i.e. by measuring the signed distance defined as:

$$d(x) = \frac{\omega^T x + b}{\|\omega\|}. \tag{2}$$

For a detailed explanation of the SVM algorithm we refer the reader to [5,10,14,19].

As in the work of Y. Lei et all. [16] we opted by a *one versus all* approach. That means, that for each subject in the database, we will solve a two class problem in which we try to separate those feature vectors that belong to a given subject (named *positive* class) from the rest of the feature vectors (*negative* class). To find each hyper-plane, we used the implementation of SVM given in [6] and we tried both linear SVM and Kernel-SVM [with a Radial Basis Function kernel (SVM-RBF)]. The cost parameter of SVM algorithm (C) and the kernel parameter (γ) (when the kernel was considered) were estimated performing 5-fold cross validation.

Table 1. Accuracy (percentage) over Train and Test Sets. The superscript (A or B) recalls the patch from which features were extracted. Recall that the patch A corresponds to the region of the eyes while patch B corresponds to the region of the nose. The first two columns shows the results obtained just considering the geometrical information (represent by histograms of depth partial derivatives values). The last three columns shows the accuracy obtained by considering all the features extracted from the path A, patch B and the union of these, respectively.

(Acc %)	D_v^A $D_{v_\perp}^A$	D_u^B $D_{u_\perp}^B$	$LBP^A_{(Tex.)}$	$LBP^B_{(Tex.)}$	All feat. PatchA	All feat. PatchB	All feat.
Train	99.3 ± 0.4	98.4 ± 0.5	77.0 ± 4.2	51.6 ± 3.6	99.8 ± 0.2	99.4 ± 0.3	$\mathbf{100 \pm 0.0}$
Test	91.9 ± 1.7	93.6 ± 1.3	64.9 ± 3.8	44.3 ± 3.8	95.2 ± 0.8	95.1 ± 1.8	$\mathbf{99.4 \pm 0.2}$

3 Experiments and Evaluation

In this section, we perform two different set of experiments; firstly, experiments with the Texas 3D Face Recognition Database [12,13], and secondly, experiments with acquired images where the whole framework is tested.

3.1 Evaluation over the Texas 3D Face Recognition Database

This database contains 1149 pairs of high resolution, pose normalized, preprocessed, and perfectly aligned color and range images of 118 subjects. Additionally, it includes the locations of 25 manually marked anthropometric facial fiducial points. From those 25 fiducial points, just three are used (the nose tip and the corner of the eyes).

Texas database was split in two sets: one used for training (estimating SVM hyper-plane parameters) and the other was reserved for evaluation. Train and test sets contain (each) 486 samples obtained from scans of 25 different subjects. Once the 25 hyper-planes were obtained, the signed distance to each hyper-plane is measured for all the samples in train and test datasets. To each pair of range and color images, the assigned class is the one associated to the hyper-plane that has the higher distance to the respective feature vector.

Results are summarized in Table 1 for different subsets of features. As we can see the area of the nose is the more robust region of the face for the sake of 3D face recognition when we have expression variations. This fact is in agreement with recent research in this field (see e.g. [7,11,16,17]). Furthermore, the geometrical information seems to be more effective than the texture analysis, this is an expectable result as texture is easily affected, e.g. by illumination conditions or gestures. The best results were achieve by the fusion of all the features.

3.2 Evaluation of the Entire System

In a second series of experiments our own one-shot database was used. This database was generated by illuminating the subjects with structured light, as described in section 2. Images were collected over different days, and therefore, illumination conditions, pose and face expressions present significant variations.

The database contains pictures of approximately 120 different subjects, and in most cases there was available only one picture for each person. As the main objective of this second series of experiment is to evaluate the robustness of the proposed framework under different pose and facial expressions, we will focus on the two class problem where multiples images of one of the authors play the role of the positive class and the rest of the subjects represents the negative class.

As training set we used 219 images, with multiples images of the target subject (positive class) and the rest of images from different subjects representing the negative class. The test set is composed of 48 new negative instances (new pictures of people that were not present in the training set) plus 46 new positive instances (pictures of the target subject). In this test dataset, a wide range of facial expressions and pose variations are included as well as some pictures of the subjects with the face partially occluded (i.e. wearing a scarf or a bonnet as illustrated in Fig. 2). Figure 2 shows red/green histograms obtained by measuring the signed distance to the hyper-plane of each negative/positive sample on the test set.

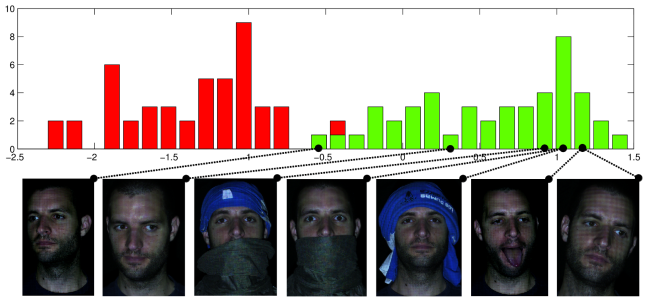

Fig. 2. Number of test instances versus its signed distance to SVM hyperplane. Red and green histograms were obtained by considering the negative and positive samples respectively. Under the histogram, some examples of positive samples are shown. The dashed lines shows the particular distance obtained for each of the example images.

By considering the signed distance of each test sample to the hyper-plane, one can classify each sample as positive if the signed distance [Eq. 2] is higher than certain threshold and negative in the other case. Figure 3 shows the Precision (portion of the samples labeled as positive that actually belong to the positive class), Recall (portion of samples of the positive class correctly classified) and the Precision-Recall curve for different threshold values.

A larger database is required to perform more exhaustive experiments (e.g. by repeating the previous two-class experiment for several subjects) before extracting quantitative conclusions. Despite this, some interesting aspects of the proposed technique can be addressed. Firstly, the proposed approach shows to be

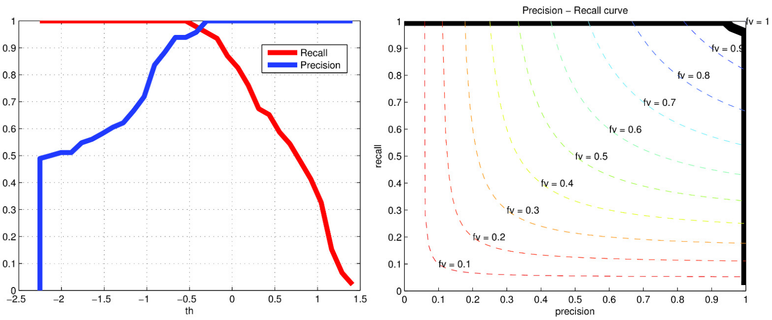

Fig. 3. Left: Recall and Precision for different threshold values. Right: Precision-Recall curve obtained varying the threshold value used for classification. The colored dashed lines shows the lines of constant F-measure (which represents the geometrical mean between Recall and Precision).

robust to facial expression (even exaggerated ones). Secondly, also promising results were achieved when the areas of the face (other than the nose or forehead) were occluded (e.g. by a scarf or bonnet), which shows that the approach is also robust to some intentional variations that a subject may produce such as variations of facial hair (e.g. shave his beard off).

4 Conclusions

A novel one-shot 2D+3D face recognition approach was presented. Instead of using a 3D scanner, we acquire a single photo of the face while a rectangular pattern is been projected over it; from this unique image, it is possible to extract low-level geometrical features without the explicit 3D reconstruction as well as texture information. Also as the projected pattern is static, the proposed method can be applied to dynamic scenes and can be trivially extended to video face analysis. On the other hand, the proposed framework is likely limited to indoor scenarios, and requires a set of images of the subjects we want to be able to identify (for the negative class one single image of many different subjects is enough) as a SVM algorithm must be trained.

Experiments shows the potential of the proposed framework to handle pose and facial expression variations, while requiring very low hardware requirements (as we just need e.g. a rectangular static pattern, a lens, and a led light source).

Acknowledgments. The authors thank PEDECIBA, CSIC and ANII for their financial support.

References

1. Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: 2d and 3d face recognition: A survey. Pattern Recognition Letters **28**(14), 1885–1906 (2007)
2. Besl, P., McKay, N.D.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992)
3. Bowyer, K.W., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. Computer Vision and Image Understanding **101**(1), 1–15 (2006)
4. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Three-Dimensional Face Recognition. International Journal of Computer Vision (IJCV) **64**(1), 5–30 (2005)
5. Chang, C., Lin, C.: LIBSVM: a library for support vector machines (2001)
6. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2**, 27:1–27:27 (2011). Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
7. Chang, K.I., Bowyer, K.W., Flynn, P.J.: Multiple nose region matching for 3D face recognition under varying facial expression. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(10), 1695–1700 (2006)
8. Di Martino, J.M., Ayubi, G., Fernández, A., Ferrari, J.A.: Differential 3d shape retrieval. Optics and Laser in Engineering **58C**, 114–118 (2014)
9. Di Martino, J.M., Fernández, A., Ferrari, J.A.: 3d curvature analysis with a novel one-shot technique. In: IEEE International Conference on Image Processing (ICIP2014) (2014)
10. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley, New York (2001)
11. Faltemier, T.C., Bowyer, K.W., Flynn, P.J., Member, S.: A Region Ensemble for 3-D Face Recognition **3**(1), 62–73 (2008)
12. Gupta, S., Castleman, K., Markey, M., Bovik, A.: Texas 3d face recognition database. In: 2010 IEEE Southwest Symposium on Image Analysis Interpretation (SSIAI), pp. 97–100, May 2010
13. Gupta, S., Markey, M., Bovik, A.: Anthropometric 3d face recognition. International Journal of Computer Vision **90**(3), 331–349 (2010)
14. Hsu, C., Chang, C., Lin, C.: A practical guide to support vector classification. National Taiwan University, Taipei 106, Taiwan (2010)
15. Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(4), 640–649 (2007)
16. Lei, Y., Bennamoun, M., El-Sallam, A.A.: An efficient 3D face recognition approach based on the fusion of novel local low-level features. Pattern Recognition **46**(1), 24–37 (2013)
17. Li, X., Da, F.: Efficient 3D face recognition handling facial expression and hair occlusion. Image and Vision Computing **30**(9), 668–679 (2012)
18. Mahoor, M.H., Abdel-Mottaleb, M.: Face recognition based on 3d ridge images obtained from range data. Pattern Recognition **42**, 445–451 (2009)
19. Steinwart, I., Christmann, A.: Support Vector Machines, 1st edn. Springer, USA (2006)
20. Zhang, Y.N., Guo, Z., Xia, Y., Lin, Z.G., Feng, D.D.: 2D representation of facial surfaces for multi-pose 3D face recognition. Pattern Recognition Letters **33**(5), 530–536 (2012)