3D Face Recognition with Reconstructed Faces from a Collection of 2D Images

João Baptista Cardia Neto[1]([✉]) and Aparecido Nilceu Marana[2]

[1] São Carlos Federal University - UFSCAR, São Carlos, SP 13565-905, Brazil
`joao.cardia@fatec.sp.gov.br`
[2] UNESP - São Paulo State University, Bauru, SP 17033-360, Brazil
`nilceu@fc.unesp.br`

Abstract. Nowadays, there is an increasing need for systems that can accurately and quickly identify a person. Traditional identification methods utilize something a person knows or something a person has. This kind of methods has several drawbacks, being the main one the fact that it is impossible to detect an imposter who uses genuine credentials to pass as a genuine person. One way to solve these kinds of problems is to utilize biometric identification. The face is one of the biometric features that best suits the covert identification. However, in general, biometric systems based on 2D face recognition perform very poorly in unconstrained environments, common in covert identification scenarios, since the input images present variations in pose, illumination, and facial expressions. One way to mitigate this problem is to use 3D face data, but the current 3D scanners are expensive and require a lot of cooperation from people being identified. Therefore, in this work, we propose an approach based on local descriptors for 3D Face Recognition based on 3D face models reconstructed from collections of 2D images. Initial results show 95% in a subset of the LFW Face dataset.

Keywords: Biometrics · 3D face recognition · 3DLBP ·
Face reconstruction

1 Introduction

The several drawbacks of the traditional methods have stimulated the research on biometric identification methods, which are based on biological or behavioral traits of the individuals. Because biometric identification systems use something that persons are, they are more difficult to circumvent [11].

In surveillance systems, the utilization of biometric characteristics can drastically improve the system performance and, among several other characteristics, face has several advantages [9]. Face is the biometric feature that best suits the covert identification, since the current technology is able to provide high resolution 2D face images captured by low cost cameras, in a secret way, at a distance and without cooperation from the people being identified [9]. However,

in general, biometric systems based on 2D face recognition perform very poorly in unconstrained environments, common in covert identification scenarios, since the input images present variations in pose, illumination and facial expressions.

One way to increase the automated facial recognition system accuracy is to utilize sensing devices such as 3D scanners [5]. However, it would be impossible to utilize such scanners in unconstrained environments for covert identification, since they need cooperation from the people being identified, in order to capture the 3D face models.

One alternative for obtaining 3D facial data from a person, without collaboration, in unconstrained scenarios, would be to reconstruct a 3D face model from a gallery of 2D face images of such a person. The reconstruction method proposed by [13] can be utilized to such end.

In this work we intend to develop a new approach based on 3D face models obtained from collections of 2D face images, in order to identify people in unconstrained or very harsh environments.

2   Unconstrained 3D Face Reconstruction

The method proposed by [13] has the following three main steps (for more details refer to the original work):

First step: Given a collection of unconstrained face images, the method performs 2D landmark estimation and enhances a 3D template by deforming a generic 3D face template such that the projection of its 3D landmarks are consistent with the estimated 2D landmarks;

Second step: The method estimates the person-specific face normals via photometric stereo. It takes 2D face images at all poses and project them onto the enhanced 3D face template to establish a dense correspondence across the images. Then, it jointly estimates the lighting and surface normals via SVD (Singular-Value Decomposition);

Third step: The method deforms the 3D shape so that its updated surface normals become similar to the estimated ones, under the landmark constrain and an additional boundary constraint. This process iterates until convergence.

Figure 1 illustrates the result of the 3D face model reconstruction of a person from a collection of 2D faces from this person by using the method proposed in [13].

3   3D Local Binary Pattern

In [4] the 3D Local Binary Patterns (3DLBP) was proposed. This variation of the original operator considers not only the signal of the difference, but also the absolute depth difference. In the original work was stated that, for face, more than 93% of all depth differences (DD) with $R = 2$ are smaller than 7. Due to this property the absolute value of the DD is stored in three binary units ($i_2 i_3 i_4$). Therefore, it is possible to affirm:

$$|DD| = i_2 \cdot 2^2 + i_3 \cdot 2^1 + i_4 \cdot 2^0 \tag{1}$$
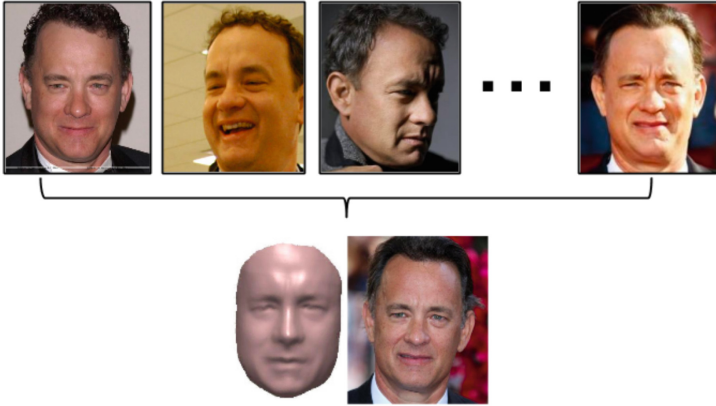
Fig. 1. Example of the 3D face model reconstructed by using the method in [13]. Adapted from: http://cvlab.cse.msu.edu/project-face-recon.html

There is also i_1, a binary unit defined by:

$$i_1 = \begin{cases} 1 \text{ if DD } \geq 0; \\ 0 \text{ if DD } < 0. \end{cases} \tag{2}$$

Those four binary units are divided into four layers and, for each of those layers, four decimal numbers are obtained: P_1, P_2, P_3, P_4. The value of the P_1 has the same value as the original LBP. For matching, the histograms of the local regions (P_1, P_2, P_3, P_4) are concatenated. The Fig. 3 shows the process for the generation of the 3DLBP, given an image (Fig. 2).
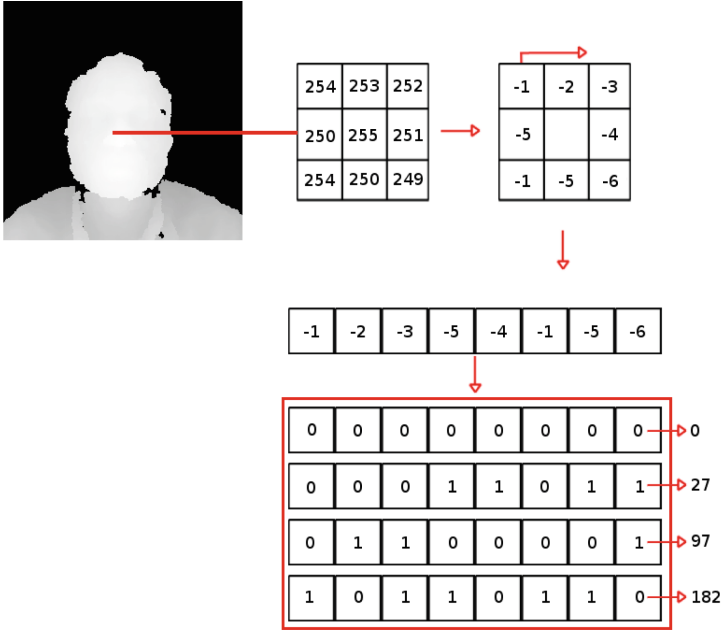
4   Proposed Method

Our method receives a collection of 2D face images and, utilizing the reconstruction method [13], builds a 3D face model. After constructing the model our method extracts a cloud point and utilizes it as input to the 3DLBP module.

The 3DLBP module was inspired in [2] but, after the construction of the new depth maps, the 3DLBP is applied in 64 micro regions and the histograms of each of those regions are concatenated. The final image descriptor is the concatenation of the histograms from each region. Figure 3 shows a block diagram of our proposed method for 3D face recognition with reconstructed faces from a collection of 2D images.

The pre-processing steps are the following:

Symmetric Filling: The Symmetric Filling technique, proposed in [7], utilizes the left side of the face to increase point density by including the set of mirrored points from the right side of the face, and vice-versa. However, not all the mirrored points are useful because the goal is to fill only in the missing

Fig. 2. The process of the 3DLBP proposed by [4]. Each of the differences is encoded into the layers (layer 2, 3 and 4) and the signal into the layer 1.

data (occluded regions, for instance). Likewise in [7], the strategy to be used in our method is to add the mirrored point only if there is no neighboring point at that location. During this process, if the Euclidean distance from a mirrored point to its neighbors in the original point cloud is greater than a threshold value δ, then that point will be added to the original face data.

Iterative Closest Point: The Iterative Closest Point (ICP) [1], is a very well known solution for the problem of registration. It tries to find a rigid transformation that minimizes the least-square distance between two points. Given two 3D point set (A and B), ICP performs the following three basic steps:

1. Pair each point of A to the closest point in B;
2. Compute the motion that gives the lowest Mean Squared Error (MSE) between the points;
3. Apply the motion to the point set A and update the MSE.

The three aforementioned steps are performed until the MSE is lower than a threshold τ. A complete description of this method can be found in [1,3].

Savitzky-Golay Filter: The Savitzky-Golay filter is a lowpass filter based on least-squares polynomial approximation [15]. The Savitzky-Golay filter tries to find filter coefficients c_n that preserves the higher moments of the filtered data. It utilizes the idea of a window moving through the function data and approximate it by a polynomial of higher order, quadratic or quartic [12]. More details about this filter can be found in [14,15].
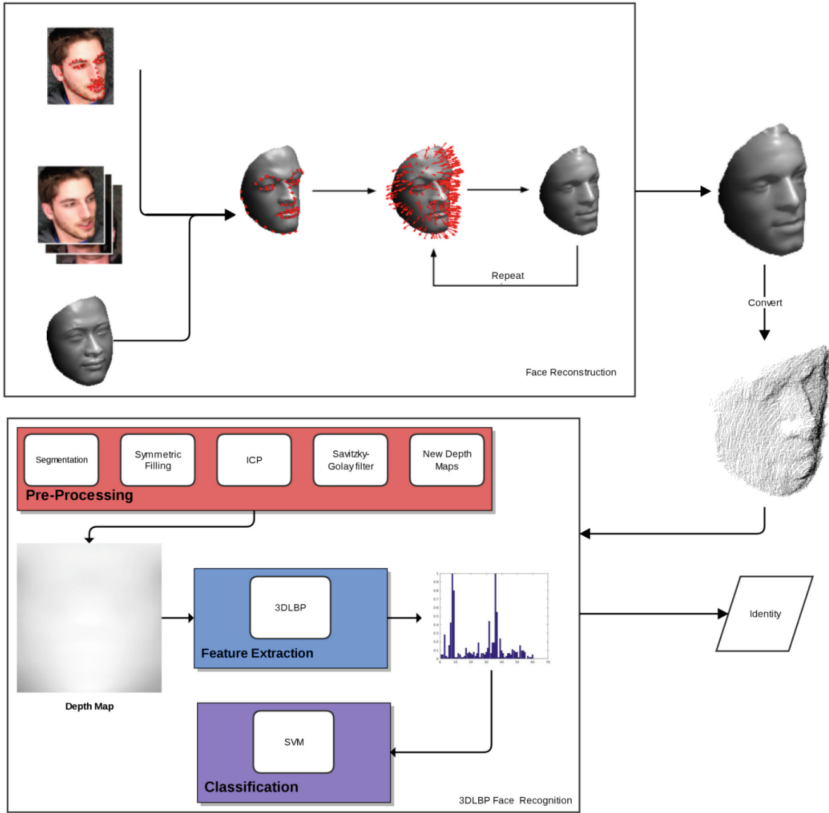
Fig. 3. A block diagram that describes the proposed method. Inspired in [2].

Generation of Depth Map: In the proposed method, depth maps from cloud points must be generated. Cloud points will be obtained from the reconstructed 3D face models. In order to generate depth maps from cloud points, a circular region with radius R is cropped centered at the nose tip. Then, the cropped image goes through the symmetric filling process. Finally, the resulting face image is fitted to a smooth surface using an approximation method, implemented by an open source code[1] written in Matlab. The result of this process is a 100×100 matrix, as illustrated in Fig. 3.

5   Databases

For evaluating the performance of the proposed approach we utilized the Labeled Faces in the Wild (LFW).

---

[1] http://mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit.

The Labeled Faces in the Wild [6] is a dataset composed of 13,000 images collected on the Web, to serve as means for studying the problem of unconstrained face recognition. The dataset contains 1,680 subjects with a varying quantity of images, each folder of a subject is labeled with its name. The Fig. 4 shows some examples of the LFW face dataset.

**Fig. 4.** Examples from a set of images from the LFW Face Database [6].

As it is possible to see the base can be quite challenging and, utilizing the face reconstruction, can help to attenuate pose and illumination problem.

6   Experiments

For the LFW databases the experiments were made following the protocol:

– A subset with thirty subjects were separated in order to be possible to get subjects with at least twelve face images;
– For each subject two 3D face model with half the images were created and;
– For each 3D face model, they were rotated from -30 to $30°$ in the Y and X axis;
– The classification results were evaluated with a 10-fold cross-validation.

The results achieved a Rank 1 of 94.74% of recognition rate, Fig. 5 shows them average Cumulative Match Characteristics (CMC) curve obtained in the aforementioned experiments on part of the LFW database. Comparing this result with the results of the literature one can observe, currently, the recognition rate from most of them have around 97% to 99% Rank 1 in their hardest cases [8,10, 16–19]. But it is not a fair comparison since the protocols are different and the current work is in progress so, it is expected, to improve the recognition rate, utilizing CNNs and more appropriate fusion strategies of characteristics.
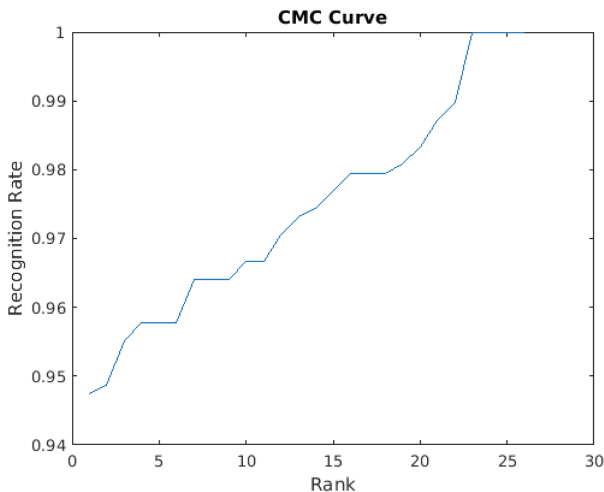
Fig. 5. Average CMC curve obtained on part of LFW Face Database.

7 Conclusion

With the obtained so far it is already possible to conclude that 3D Face Recognition can be done utilizing reconstructed 3D models, this can be applied to several scenarios, such as surveillance systems. It is possible to improve the recognition rate, for instance, by using a CNN for subject classification and fusing the decision of other methods.

For the LFW database, it is needed to increase the number of face images for some subjects, this can be done by developing a web-crawler which can search the internet for more faces for particular subjects.

All those improvements are being considered as future work.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992). https://doi.org/10.1109/34. 121791
2. Cardia Neto, J.B., Marana, A.N.: Utilizing deep learning and 3DLBP for 3D face recognition. In: Mendoza, M., Velastín, S. (eds.) CIARP 2017. LNCS, vol. 10657, pp. 135–142. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75193-1_17
3. Chetverikov, D., Stepanov, D., Krsek, P.: Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm. Image Vis. Comput. **23**(3), 299–309 (2005). https://doi.org/10.1016/j.imavis.2004.05.007. http://www.sciencedirect.com/science/article/pii/S0262885604001179

4. Huang, Y., Wang, Y., Tan, T.: Combining statistics of geometrical and correlative features for 3D face recognition. In: Proceedings of the British Machine Vision Conference, pp. 90.1–90.10. BMVA Press (2006). https://doi.org/10.5244/C.20.90

5. Kakadiaris, I.A., et al.: Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 640–649 (2007). https://doi.org/10.1109/TPAMI.2007.1017

6. Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H., Hua, G.: Labeled faces in the wild: a survey. In: Kawulok, M., Celebi, M.E., Smolka, B. (eds.) Advances in Face Detection and Facial Image Analysis, pp. 189–248. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-25958-1_8

7. Li, B., Mian, A., Liu, W., Krishna, A.: Using kinect for face recognition under varying poses, expressions, illumination and disguise. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV), pp. 186–192 (2013). https://doi.org/10.1109/WACV.2013.6475017

8. Liu, J., Deng, Y., Bai, T., Huang, C.: Targeting ultimate accuracy: face recognition via deep embedding. CoRR abs/1506.07310 (2015). http://arxiv.org/abs/1506.07310

9. Nguyen, V., Do, T., Nguyen, V.-T., Ngo, T.D., Duong, D.A.: How to choose deep face models for surveillance system? In: Sieminski, A., Kozierkiewicz, A., Nunez, M., Ha, Q.T. (eds.) Modern Approaches for Intelligent Information and Database Systems. SCI, vol. 769, pp. 367–376. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76081-0_31

10. Parkhi, O.M., Vedaldi, A., Zisserman, A., et al.: Deep face recognition. In: BMVC, vol. 1, p. 6 (2015)

11. Prabhakar, S., Pankanti, S., Jain, A.: Biometric recognition: security and privacy concerns. IEEE Secur. Priv. **1**, 33–42 (2003). https://doi.org/10.1109/MSECP.2003.1193209

12. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd edn. Cambridge University Press, New York (2007)

13. Roth, J., Tong, Y., Liu, X.: Unconstrained 3D face reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015

14. Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**(8), 1627–1639 (1964). https://doi.org/10.1021/ac60214a047

15. Schafer, R.W.: What is a savitzky-golay filter? [lecture notes]. IEEE Signal Process. Mag. **28**(4), 111–117 (2011). https://doi.org/10.1109/MSP.2011.941097

16. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015

17. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015

18. Wen, G., Chen, H., Cai, D., He, X.: Improving face recognition with domain adaptation. Neurocomputing **287**, 45–51 (2018). https://doi.org/10.1016/j.neucom.2018.01.079. http://www.sciencedirect.com/science/article/pii/S0925231218301127

19. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31