3D Face Recognition in the presence of 3D model degradations

Pierre Lemaire[1] Di Huang[1] Joseph Colineau[2] Mohsen Ardabilian[1] Liming Chen[1]

[1]Université de Lyon, CNRS, Ecole Centrale Lyon
LIRIS UMR 5205
69134, Ecully, France
{first_name.last_name}@ec-lyon.fr

[2]Thales Research & Technology France
1, Av Augustin Fresnel
91767 PALAISEAU Cedex, France
joseph.colineau@thalesgroup.com

Abstract: The problem of 3D face recognition has received a growing interest in the past decades. While proposed approaches have proven their efficiency over renowned databases as FRGC, little work has been conducted on the robustness of such algorithm to the quality of 3D models. In this work, we present a study of the robustness of our 3D face recognition algorithm, namely MS-ELBP+SIFT, to face model degradations. Those degradations include Gaussian noise, decimation, and holes. Degradations are generated on a subset of the FRGC database, hence enabling us to compare the robustness of our approach to them. Results are provided through a comparative study with the baseline ICP method.

1 Introduction

While studies have proven the effectiveness of current approaches to 3D face recognition, few of them have evaluated the impact of 3D model degradations. Those degradations may have several origins. At first, some degradation may be issued from the 3D acquisition itself. 3D laser scanners have been reported to be sensitive to the wavelengths of artificial lightings. In some cases, a high reflectance of the acquired surface can also generate holes or spikes on the resulting model. Since they are electronic devices, binocular sensors also suffer from thermal noise (Nyquist-Johnson noise), Schottky noise, etc. We shall count the acquisition conditions as a source of degradations also, since occlusions, self-occlusions or movements resulting in a blur might occur during the acquisition process. A short study of those model degradations for 3D Face Recognition within the FRGC database [Ph05] was conducted in [Fa07], as well as in [Ka07]. Compression might also be useful for storage capacity matters, as well as sampling (resolution reduction of the 3D model). The distance of the subject to the 3D scanner also has an impact on the resolution of the 3D model. Those considerations are to be taken into account especially in the case of asymmetric biometric scenarios, in which the enrolment step can benefit from high quality acquisition conditions, while the verification step is often to be conducted with light gear, in rather unconstrained conditions. Analyzing the behavior of recognition algorithms in the presence of such degradations is a necessary step before using such methods into a real environment application.

A majority of authors in the domain of 3D face analysis seem to have, often implicitly, admitted the existence of such issues in their work; hence the use of pre-processing steps [Fa07][DHu10][Dr10]. Those pre-processing steps include peaks, and more generally noise removal with the help of mean, median or Gaussian filters over the Z coordinates. Resampling, remeshing techniques are also used [Dr10]. Pre-processing methods also include hole filling algorithms by the mean of interpolation techniques. Authors have also dedicated works, and developed techniques relative to occlusion handling [Co06], which is in some cases equivalent to holes handling.

Despite 3D model degradations are generally admitted to have an impact over face recognition algorithms performances [Ro09][Po09], to our knowledge little to none work has been presented to measure it. In this paper, we propose a study on the performances of our state of the art face recognition approach, MS-ELBP + SIFT [DHu11], under various canonical model degradations such as noise, decimation and holes. Results were compared with ICP [Be92], which is generally considered as a baseline algorithm for the problem of 3D face recognition.

This paper is organized as follows. The second section exposes our method for 3D Face Recognition, providing results on the FRGC database. The third section exposes our method for measuring the robustness of 3D Face Recognition algorithms to degradations. Experimental are provided. Section 4 concludes the paper.

2 Our 3D Face Recognition Algorithm: MS-ELBP + SIFT.

In this section, we describe our method for 3D face recognition, also detailed in [DHu11]. The first step is based on a novel approach for representing the 3D face depth-map, called Multi Scale Extended Local Binary Patterns (MS-ELBP). The second step is using the rather classical Scale Invariant Feature Transform (SIFT) approach.

In the state-of-the-art, block-based 3DLBP histograms [YHu06] and LBP based range faces [DHu10] were investigated for 3D facial representation, while they do not provide accurate description of local shape variations. In this section, we first recall the basics of LBP. We then introduce Extended Local Binary Patterns (ELBP) and Multi-Scale strategy to generate the new 3D geometric facial representation, called MS-ELBP Depth Faces (MS-ELBP-DFs), which accurately encodes local shape variations of range faces. Results are eventually exposed and prove the effectiveness of our approach.

2.1 LBP and its descriptive power of local shape variations

LBP [Oj02] is a non-parametric algorithm, and was first proposed to describe local texture of 2D images. The most important properties of LBP are its tolerance to monotonic illumination variations and computational simplicity. So, it has been extensively adopted for 2D face recognition in the last several years [Ah04].

Specifically, the original LBP operator labels each pixel of a given image by thresholding in a 3×3 neighborhood. Each of the 256 LBP code can be regarded as a microtexton. LBP can describe local shape structures, such as flat, concave, convex etc. when operated on depth images. However, the direct application of LBP on range images potentially results in unexpected confusion to similar but different local shapes. To address this problem, we considered two complementary solutions. The first one aims at improving the discriminative ability of LBP with Extended LBP coding approach, and the other one focuses on providing a more comprehensive geometric description of the neighborhood by exploiting Multi-Scale strategy. Both solutions are discussed in the two following subsections respectively.

2.1.1 Extended Local Binary Patterns

The reason why LBP cannot be competent to recognize similar shapes is caused by its operational mode. It only compares relative differences between the central pixel and its neighbors. Instead of LBP, ELBP not only extracts relative gray value difference between the central pixel and its neighboring pixels, but also focuses on their absolute difference. ELBP is a generalized version of 3DLBP [YHu06].

Specifically, the ELBP code consists of several LBP codes at multiple layers which encode the exact gray value difference *(GD)* between the central pixel and its neighboring pixels. The first layer of ELBP is actually the original LBP code, encoding the sign of *GD*. The following layers of ELBP then encode the absolute value of *GD* (Fig. 1). The first layer of the ELBP simply encodes the sign of *GD*. The absolute values of *GD*, i.e. 1, 5, 3, 2, 1, 2, 3, 0, are then encoded in their binary numbers: $(001)_2$, $(101)_2$, $(011)_2$, $(010)_2$, etc. Using a LBP scheme on all the binary bits, we finally generate the corresponding ELBP code for each layer. As a result, the information encoded in the additional layers can be used to distinguish shapes described as similar by the first layer (original LBP).

Theoretically, in one image, the maximum value of *GD*s is 255 (between 0 and 255), which means that 8 additional layers are required to encode *GD*s. However, range faces are pretty smooth; and the *GD*s in a local surface generally do not vary dramatically. It was also proven by the work in [YHu06] that more than 93% of *GD*s are smaller than 7 between points within two pixels. Hence, we set the number of additional layers to 3.

2.1.2 Multi-Scale Strategy

LBP facial representation can be achieved in two ways: one is LBP histogram [Sh08]; another is LBP face. The latter approach, as investigated in our method, regards the corresponding decimal number of the LBP binary code as the intensity value of each pixel, and generates the LBP face.

MS-ELBP-DFs of range face image can be achieved by varying the neighborhood size of LBP operator, by first down-sampling the range image and then adopting the LBP operator with a fixed radius (Fig. 2). The number of sampling points is 8, and the value

of radius varies from 1 to 8. As we can see in Figure 2, the original range image is very smooth, while the resulting MS-ELBP-DFs contain many more details of local shapes.

2.2 SIFT BASED LOCAL FEATURE MATCHING

Local feature extraction operated directly on smooth 3D face images leads to few local features with low distinctiveness. Meanwhile, the produced MS-ELBP-DFs contain many more details of local shapes, and thus enhance their distinctiveness. Once the MS-ELBP-DFs are achieved, the widely-used SIFT based features [Lo04] are extracted from them for similarity score calculation and final decision.

We use the SIFT operator on each MS-ELBP-DF separately. Because MS-ELBP-DFs highlight local shape characteristics of smooth range images, many more SIFT-based keypoints can be detected for the following matching step than those in the original range images. From statistical work done on the FRGC database, the average number of feature points extracted from each of MS-ELBP-DFs is 553, while that of each range image is limited to 41.

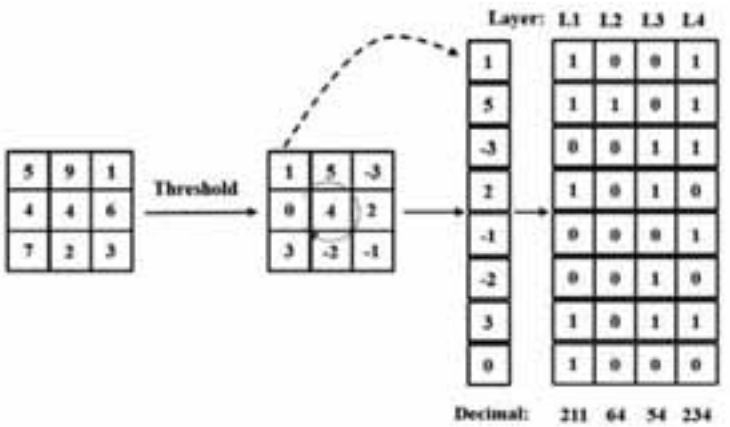

Figure 1. An example of the ELBP Operator.

Figure 2. MS-ELBP-DFs of a range face image with different radii from 1 to 8 (from left to right).

Given the local features extracted from each MS-ELBP-DF pair of the gallery and probe face scan respectively, two facial keypoint sets can be matched. Matching one keypoint to another is accepted only if the matching distance is less than a predefined threshold t times the distance to the second closest match. In this research, t is empirically set at 0.6 as in [Lo04]. A bigger number of matched keypoints means a better matching relationship. A face in the probe set is matched with every face in the gallery. The multi-scale analysis is then achieved by fusing the matching scores of all scales by using a basic weighted sum rule. More details about that step are provided in [DHu11].

2.3 Performances

We have conducted experiments over the FRGC v2.0 database [Ph05]. This database is one of the most popular datasets, containing 4007 3D face models of 466 different subjects. Table I provides comparative results of the Rank One experiment with other state of the art approaches.

TABLE I. RANK-ONE RECOGNITION RATES ON FRGC V2.0.

	Rank-One Recognition Rate
ICP	72.2%
Mian et al. [Mi07]	96.2%
Kakadiaris et al. [Ka07]	97.0%
MS-ELBP-DFs	**97.2%**

Extended results and analysis are provided in [DHu11].

In the following section, we are looking forward to analyze the behavior of our approach in the presence of original model degradations.

3 Robustness to Degradations

As stated in the introduction, 3D models can suffer from various types of degradations, from various origins. It seems difficult to acquire ground-truth over model degradations or to model degradations precisely. Hence, in this experiment, we decided to generate degraded data from original 3D scans. That choice allows us to keep mastery over the parameters of the degradations. The degradations we applied may be considered as canonical degradations, namely Gaussian noise, Decimation and Holes.

Our experimental protocol is the following. We first randomly picked 100 different subjects within the FRGC v2.0 database. For each subject, one model is randomly picked as a gallery model, which expression is neutral. Also 100 different models were randomly picked for each subject, as probe models. From there, we are looking forward to analyze the performance loss in our algorithm under model degradations. For that purpose we applied basic degradations to our probe models set to generate new, degraded sets. At that point, the first concern we should have is that the original data we are working with

(gallery, probe) are as little degraded as possible. We chose to apply to both gallery and probe sets a pre-processing inspired by the method exposed in [Fa07]. More precisely, we cropped the faces based on the nose-tip within a sphere of diameter 100mm. Nose-tips were manually located on every face. Then, we applied a median filtering within a 5x5 square for every pixel in the depth image. The probe face set was then altered to some extent to create new, degraded sets, according to the following degradations:

- Gaussian noise corresponds to the injection of an error within a Gaussian distribution on the Z coordinates on the depth image. This tends to emulate the behavior of electronic noise of acquisition devices, albeit a simplistic manner. In our experiments, we set the RMS value of the error to 0.4mm.
- Decimation corresponds to removing vertices from the original data. In this experiment, vertices are picked randomly and removed from a ratio of x4. That means that the decimated model includes 4x fewer vertexes than the original ones.
- Holes are generated at random locations on the face. At first, we pick a random vertex on the surface of the face. Then, we crop the hole according to a 1 centimeter radius sphere centered on the latter vertex.

Figure 3 shows examples of those degradations.

We judged several measures to be interesting. The Rank-One Recognition Rate; the Rank-Five Recognition Rate; the Equal Error Rate (EER) appeared to us like good indicators of the decrease in performances relative to models degradations. Experiments were conducted with both the baseline algorithm ICP [Be92] and our MS-ELBP+SIFT method. Results are exposed in Tables II and III. In those tables, “Rank-1 RR” lines stand for Rank One Recognition Rate, “Rank-5 RR” stand for “Rank Five Recognition Rate” and “EER” for Equal Error Rate.

TABLE II. BEHAVIOUR OF THE BASELINE ICP ALGORITHM IN THE PRESENCE OF DEGRADATIONS

	Original	**Noise**	**Decimation**	**Holes**
Rank-1 RR	0.712895	0.693431	0.669100	0.720195
Rank-5 RR	0.824818	0.817518	0.793187	0.836983
EER	0.1350	0.1392	0.1977	0.1353

Figure 3. An example of degradations applied to one model. From left to right: the Original face, Noise applied, Decimation applied, Holes applied. On the decimation picture, no interpolation was conducted. Then, blank parts correspond to a lack of pixels.

TABLE III. BEHAVIOUR OF OUR MS-ELBP + SIFT ALGORITHM IN THE PRESENCE OF DEGRADATIONS.

	Original	**Noise**	**Decimation**	**Holes**
Rank-1 RR	0.934307	0.905109	0.927007	0.927007
Rank-5 RR	0.963504	0.944039	0.956204	0.958637
EER	0.0546	0.0638	0.0559	0.0592

Interestingly enough, in our case holes did not produce any performance decrease on the baseline ICP algorithm (almost no change in EER, slight improvement in recognition rates). While both performances are consistently lowered, the ICP algorithm seems to suffer more from decimation than from Gaussian noise. While our algorithm is affected by every type of degradation, this experiment shows its robustness to decimation and holes, for both the EER and recognition rates. Our algorithm is slightly less robust to Gaussian noise, while maintaining rather high performance rates and EER. Overall, this experiment shows the overall good robustness of our algorithm to various kinds of model degradations.

4 Conclusion and Perspectives

In this paper, we presented a novel experiment, aimed at analyzing the behavior of state of the art techniques in the field of 3D face recognition techniques in the presence of model degradations. More precisely, we compared the impact of canonical degradations such as noise, decimation and holes, on both our state of the art approach, namely MS-ELBP + SIFT, and the baseline algorithm ICP. We show that, while providing far stronger performances than the baseline ICP algorithm, our method also proves to be globally more robust to degradations such as noise, decimation and holes. Interestingly, our algorithm showed some sensitivity to noisy data. This study proves that, for real applications, pre-processing tools specifically designed for dealing with noise, or noise detection algorithms may be required. This study also proves that, depending on the algorithm used, data sampling may be employed to reduce the computational time, while affecting only slightly the overall performances. We are currently working on enhancing the results of this experiment by varying the parameters of each type of degradations.

Acknowledgement

This work has been supported in part by the French National Research Agency (ANR) through the FAR 3D project under the grant ANR-07-SESU-003.

References

[YHu06] Y. Huang, Y. Wang, T. Tan: Combining statistics of geometrical and correlative features for 3D face recognition, BMVC, 2006.

[DHu10] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen: 3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching, BTAS, 2010.

[Oj02] T. Ojala, M. Pietikäinen, T. Maenpaa: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, PAMI, vol. 24, no. 7, pp. 971-987, 2002.

[Ah04] T. Ahonen, A. Hadid, M. Pietikäinen: Face recognition with local binary patterns, ECCV, 2004.

[Sh08] C. Shan and T. Gritti: Learning discriminative LBP-histogram bins for facial expression recognition, BMVC, 2008.

[Lo04] D.G. Lowe et al.: Distinctive image features from scale-invariant key-points, IJCV, vol. 60, no. 4, pp. 91-110, 2004.

[Ph05] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, K. I. Chang, K. Hoffman, J. Marques, J. Min, W. Worek: Overview of the face recognition grand challenge. CVPR, vol. I, pp. 947-954, 2005.

[Ka07] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis: Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach, PAMI, vol. 29, no. 4, pp. 640-649, 2007.

[Mi07] A. S. Mian, M. Bennamoun, and R. Owens, An efficient multimodal 2D-3D hybrid approach to automatic face recognition, PAMI, vol. 29, no. 11, pp. 1927-1943, 2007.

[Fa07] T. C. Faltemier, Flexible and Robust 3D Face Recognition Dissertation, University of Notre Dame, M.S., 2007.

[Be92] P. Besl and N. McKay., A method for Registration of 3-D Shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 14(2):239 - 256, February 1992.

[Ro09] R. N. Rodrigues et al.: Robustness of multimodal biometric fusion methods against spoof attacks, Journal of Visual Language and Computing, doi:10.1016/j. jvlc.2009.01.010, 2009.

[Po09] N. Poh et al., Benchmarking Quality-dependent and Cost-sensitive Score-level Multimodal Biometric Fusion Algorithms, IEEE Transactions on Information Forensics and Security archive Volume 4 , Issue 4 (December 2009), Pages: 849-866, 2009.

[Dr10] Hassen Drira, B. Ben Amor, Mohamed Daoudi, Anuj Srivastava: Pose and expression-invariant 3d face recognition using elastic radial curves. In Proc. BMVC, pages 90.1-11, 2010.

[Co06] A. Colombo, C. Cusano, R. Schettini: Detection and Restoration of Occlusions for 3D Face Recognition, IEEE International Conference on Multimedia and Expo, pp. 1541-1544, 2006.

[Ch05] K. I. Chang, K. W. Bowyer, P. J. Flynn: Adaptive rigid multi-region selection for handling expression variation in 3D face recognition, IEEE workshop on FRGC Experiments, 2005.

[DHu11] D. Huang, M. Ardabilian, Y. Wang, L. Chen: A Novel Geometric Facial Representation based on Multi-Scale Extended Local Binary Patterns, Proc of IEEE International Conference on Automatic Face and Gesture Recognition, 2011.