3D Face Recognition using the Surface Interpenetration Measure: a Comparative Evaluation on the FRGC database

Chauã C. Queirolo, Luciano Silva, Olga R. P. Bellon, Maurício P. Segundo *
IMAGO Research Group - Universidade Federal do Paraná
P.O. Box 19092, 81531-980, Curitiba - PR, Brasil
Email: {chaua, luciano, olga, mauricio}@inf.ufpr.br

Abstract

This paper focuses a comparative evaluation of our framework for 3D face recognition and state-of-the-art systems. Our method uses a Simulated Annealing-based approach (SA) for range image registration with the Surface Interpenetration Measure (SIM) as the similarity measure, in order to match two face images. The authentication score is obtained by combining the SIM values corresponding to the matching of four different face regions. Experiments were performed on the FRGC v2 database simulating both verification and identification systems and the obtained results were compared to those reported in the literature. By using all the images in the database, a verification rate of 95.9% was achieved, at a False Acceptance Rate (FAR) of 0.1%. In the identification scenario, a rank-one accuracy of 99.5% was obtained. To our knowledge, this is the best rank-one score obtained on the FRGC v2 database, as compared to previously published results.

1. Introduction

Several methods have been proposed to perform 3D face recognition, such as Log-Gabor filter [5], Hierarchical Graph Matching (HGM) [8], Annotated Deformable Model [9] and Fusion Summation [12]. An extensive survey of works related to 3D face recognition is presented in [3]. Nowadays, a common approach is to employ image registration techniques to perform range image matching [4], [7], [9], [13]. The Iterative Closest Point (ICP) [2] method or one of its variants is usually applied to accomplish this task. The Mean Squared Error (MSE), minimized during the ICP process, is then used to compute similarity between two face images.

*The authors gratefully acknowledge CNPq, CAPES and FINEP for the financial support.

In this paper we present a complete framework for face recognition which can be employed in both verification and identification systems. Initially, the face is extracted and segmented into four regions from the input image. Then, each segmented region is registered with its corresponding from the database using a Simulated Annealing-based approach (SA) [10]. The Surface Interpenetration Measure (SIM) [19] is computed during the registration process, to perform a precise matching between two face images. The SIM was recently presented for 3D face recognition and the results showed the potential for robustness of this new measure [17]. This capability was confirmed by comprehensive experiments, which is the focus of this work.

The authentication score is computed by combining the resulting SIM values. Two approaches were evaluated for 3D face authentication: sum rule [11], and hierarchical model [6]. Also, in this work we present an improved hierarchical model which includes other evaluation metrics. Concerning to evaluate our method, exhaustive experiments were performed on the Face Recognition Grand Challenge database (FRGC) [16]. Furthermore, we present a summary analysis of results, including comparisons against the most relevant related works [5], [7]–[9], [12], [14], [15].

This paper is organized as follows. The proposed framework stages, (1) preprocessing, (2) matching, and (3) authentication, are described in Sections 2, 3 and 4, respectively. Section 5 presents the experimental results, including a comparison with related works, followed by the final remarks in Section 6

2. Preprocessing

The segmentation process uses our approach based on depth of range maps, that correctly segments 99.9% of faces from the FRGC v2 database [18]. This method automatically extracts the face from the input images.

The face extraction process is composed basically by two main stages: (1) locating homogeneous regions in the input image by using clustering combined with edge data, and (2) identifying candidate regions that belong to the face region by an ellipse detection method based on the Hough Transform. More details about this approach are described in [18].

The face image is segmented into four regions: (a) circular area around the nose, (b) elliptical area around the nose, (c) upper head, including the eyes, nose and forehead, and (d) the entire face region. Also some facial feature points are detected to be used during the matching process [17], [18].

3. Matching

Earlier results, presented in [1], have shown that the SA produces more accurate alignments and a higher recognition rate as opposed to ICP. A brief explanation of the SIM and the SA approach for range image registration are presented as follows.

3.1. Surface Interpenetration Measure (SIM)

This measure is based on the interpenetration effect resulting from the nature of real range data. The data presents slightly rough surfaces with small local distortions caused by limitations of the acquiring system.

By quantifying interpenetration, one can more precisely evaluate the registration results and provide a highly robust control. Registrations of two range images presenting good interpenetration have high SIM values, and erroneous alignments produce low SIM values and that small differences in MSE can yield significant differences in SIM.

Furthermore, alignments with high SIM present a very low interpoint distance between the two surfaces. That is, SIM is a far more sensitive indicator of alignment quality when comparing "reasonable" alignments. In this paper, we present the SIM pseudocode in Algorithm 1. For further details about this measure, the reader should refer to [19].

Some constraints were applied to the SIM to avoid incorrect corresponding points and to obtain more precise alignments. It was defined a constraint $m = 5$ as the maximum angle allowed between normal vectors at points c and p, $\vec{n}_c$ and $\vec{n}_p$, respectively. Another constraint was defined to eliminate the corresponding points on surfaces boundaries. In this case, $p \in C$ if $c \notin D$, with D the set of boundary points in B, with boundary thickness defined as $b = 1$. The window centered in p was defined as $n = 5$, as suggested in [19].

Algorithm 1: Surface Interpenetration Measure.

Input: Two range images A and B
foreach *point* $p \in A$ **do**
    Define a neighborhood N_p with size $n \times n$ centered in p ;
    Search the corresponding point c of p in image B, with $c \notin D$;
    Compute angle θ between normal vectors $\vec{n_p}$ and $\vec{n_c}$, regarding points p and c ;
    if $\theta < m$ **then**
        foreach $q_i, q_j \in N_p$, *with* $q_i \neq q_j$ **do**
            if $[(q_i - c) \cdot \vec{n_c}] \cdot [(q_j - c) \cdot \vec{n_c}] < 0$ **then**
                $C_{(A,B)} \Leftarrow C_{(A,B)} \bigcup p$;
            end
        end
    end
end
return $\frac{|C_{(A,B)}|}{|A|}$

3.2. Simulated Annealing (SA)

Simulated Annealing is a stochastic algorithm for local search [11]. The main difference between SA and other local search algorithms, *e.g.* Hill Climbing, is that SA can accept a worse solution than the current one in the iterative process. Due to this, SA does not remain "tied" to local minima and it has better chances to reach a solution close enough to the global one.

To apply SA on registration of two range images, six parameters (three parameters each for rotation and translation relative to a 3D coordinate system) are needed to define the candidate solutions as a "transformation vector" that, when applied to one image, can align it with the other. The SA-based approach has three stages: (1) initial alignment, (2) coarse alignment, and (3) fine alignment.

Firstly, an initial solution is obtained by aligning the centers of mass of the two face images. The coarse alignment is performed using a SA-based searching procedure to minimizes a robust evaluation measure, based on the MSAC robust estimator combined with the MSE of the corresponding points between the two face images. Then, a precise alignment is obtained by using a SA-based searching procedure with SIM as the evaluation measure.

Also, an enhanced SA-based approach was used to handle facial expression effects during face matching [17]. The main idea is to guide matching of a neutral face with another one with expression to the face most invariant areas (*e.g.* forehead and nose regions) [17].

One main concern about employing SA in an authentication system regards to time constraints. We performed an experiment using a dataset from FRGC v2 database to evaluate the SA computational time. The obtained average time to match two face regions were: (1) 1.3s for nose regions, (2) 2.0s for upper head, and (3) 3.1s for entire face region. These times are equivalent to those reported using ICP [7], [9].

4. Authentication

Two approaches were evaluated for 3D face authentication: sum rule [11], and hierarchical model [6]. Also, we present an improved version of the hierarchical model which includes other evaluation metrics.

4.1. Sum Rule

After the matching stage, the similarity score is achieved by combining the resulting SIM values using the sum rule [11]. This approach is more discriminative than using the SIM value computed from each region by itself [17]. Also, this score is suitable for both identification and verification systems.

4.2. Hierarchical Evaluation Model

The hierarchical evaluation model was proposed for verification systems at a False Acceptance Rate (FAR) of 0% [6]. In this approach, one region is analyzed at each level of the hierarchy. Successive regions are only analyzed if the matching score of the previous one was not sufficient to determine whether images belong to the same subject or not. This approach can boost verification rate because sometimes one single region can lead to the correct result while the combination of all regions cannot. This particular situation is observed on images that have hair occlusion, noise or facial expression [6].

4.3. Improved Hierarchical Evaluation Model

In this paper, we present an improvement to the hierarchical evaluation model. At each step of the hierarchy, instead of evaluating only the matching score obtained at that level, we also evaluated the sum of all levels computed so far. The matching hierarchy was defined as: (C) circle nose, (E) elliptical nose, (U) upper head, (F) face region, and (M) face region using the enhanced SA approach, which induces the alignment to the facial expression invariant regions. Totally, 28 measures are computed, and the sum of the five SIM values is the last evaluated one. Table 1 shows the scores computed at each hierarchy level.

Table 1. Scores computed at each level of the hierarchy.

Hierarchy Level	**Evaluation Metrics**
C	C
E	E, C+E
U	U, C+U, C+E, C+E+U
F	F, C+F, E+F, U+F, C+E+F, C+U+F, E+U+F, C+E+U+F
M	M, C+M, E+M, U+M, F+M, C+E+M, C+U+M, C+F+M, E+U+M, E+F+M, U+F+M, C+E+U+M, C+E+U+F+M

5. Experimental Results

In our experiments, we used the FRGC v2 database, which is the largest available database of 3D face images. This database is composed of 4,007 images. The experiments evaluated two types of authentication systems: verification and identification. We performed a comparison of each image with all other remaining ones in the database, totaling more than forty million combinations. Once we had all the matching results, they were split into several datasets for a detailed analysis.

The following datasets were defined: (1) *gold*, composed of 933 neutral and noiseless images, (3) *small*, 2,637 neutral and small expression images, (4) *fall*, 1,893 images taken during fall 2003, (5) *spring*, 2,114 images taken during spring 2004, (6) *first*, includes only the first image of each one of the 466 subjects, and (7) *all*, the entire database with 4,007 images.

The experiment *Fall vs. Spring* was suggested by the FRGC program because the time interval between the two datasets increases its difficulty [16]. Although we have performed a large number of experiments, some details have been omitted here because of space limits. A complete description of these datasets and experiments can be verified at IMAGO Research Group homepage: *www.imago.ufpr.br/3D_Face_Recognition*.

5.1. Verification Experiment

Each image from the probe dataset was compared with the ones present in the gallery. Also, we detected an impostor subject in FRGC v2 database that was removed from the 0% FAR experiment. For experiment using 0.1% FAR, all images available from the datasets were used. Results using 0.1% FAR are presented in Table 2. The first two columns are the gallery and probe datasets, respectively, followed by the verification rate.

Table 2. Verification rate at 0.1% FAR.

Gallery	Probe	Verification rate
Gold	*Gold*	99.9%
Gold	*Small*	99.5%
Gold	*All*	97.7%
Fall	*Spring*	96.4%
All	*All*	95.9%

Results for the verification experiments at 0% FAR are presented in Table 3. The first two columns are the gallery and probe datasets, respectively. The third column is the performance using the sum rule as evaluation measure. Column $H1$ shows the results using the hierarchical evaluation model as proposed in [6]. Finally, column $H2$ shows the performance using the improved hierarchical model approach (see Section 4.3).

Table 3. Verification rate at 0% FAR.

Gallery	Probe	Sum	H1	H2
Gold	*Gold*	98.4%	98.7%	99.2%
Gold	*Small*	94.2%	94.8%	95.8%
Gold	*All*	84.0%	85.4%	87.2%
Fall	*Spring*	68.5%	73.8%	77.6%
All	*All*	70.5%	75.3%	77.9%

By analyzing the experimental results it can be noticed that when faces with expression and noise are added to the datasets, the verification rate is affected in both 0% FAR and 0.1% FAR experiments. This behavior is expected since we perform matching assuming that the face is a rigid object. Also, by using the hierarchical evaluation model it is possible to boost recognition performance, and our improved approach produced better results when compared with the other two.

5.2. Identification Experiment

For the identification experiment, images from one dataset were defined as gallery and the remaining images were included in the probe dataset. In the gallery datasets, subjects that did not have a corresponding image in probe were not included. Results achieved for rank-one are presented in Table 4.

Table 4. Rank-one recognition rate.

Gallery	Probe	Rank-one
Gold	*Others*	99.5%
First	*Others*	99.5%
Fall	*Spring*	99.3%
All	*All*	99.6%

5.3. Comparison with other methods

Some authors have published their results using the FRGC v2 database. We reproduced the same experiments, to make a comparison with the performance of these methods. Table 5 shows a verification experiment using *All vs. All* dataset. Table 6 presents results for rank-one using *First vs. Others* dataset, and finally the results from *Fall vs. Spring* are available in Table 7. All verification results are reported at 0.1% FAR.

Table 5. Verification: All vs. All.

Method	Verification Rate
Mian *et al.* [15]	86.6%
Maurer *et al.* [14]	87.0%
Cook *et al.* [5]	92.3%
Faltemier *et al.* [7]	93.2%
Our approach	**95.9%**

Table 6. Identification: First vs. Others.

Method	Rank-one
Cook *et al.* [5]	92.0%
Kakadiaris *et al.* [9]	97.0%
Faltemier *et al.* [7]	97.2%
Our approach	**99.6%**

Table 7. Verification: Fall vs. Spring.

Method	Verification Rate
Husken *et al.* [8]	86.9%
Lin *et al.* [12]	90.0%
Faltemier *et al.* [7]	94.8%
Kakadiaris *et al.* [9]	**97.0%**
Our approach	96.4%

We also performed the same experiment *Fall vs. Spring* described in [9], which uses two different datasets, one with only neutral images, and the other only with images with facial expression. These datasets were obtained using the classification provided by the FRGC v2 database. These results are presented in Table 8. Column $R1$ shows the results reported in [9] and the column $R2$ presents our performance.

Table 8. Verification: Fall vs. Spring using FRGC datasets.

Dataset	R1 [9]	R2
Neutral Expression	98.5%	99.2%
Non-Neutral Expression	95.6%	93.6%

By observing the results, our method produced the best score for the first two experiments when compared with the other works. In *Fall vs. Spring* experiment, our method achieved a score very close to [9]. When using datasets classified by expressions, our method has the best performance in the neutral dataset, but slightly lower performance compared to the non-neutral dataset. One of the main reasons is that we perform the matching procedure assuming that the face is a rigid object.

Other authors who employ similar approaches to perform matching also report this limitation, and to reduce the expression impact, they focus on small and invariant face areas [4], [7]. Kakadiaris *et al.* [9] can deal better with the expression effects because their framework includes deformable face models in the matching process, which improves their method's efficiency. Nevertheless, when using only neutral faces our method still achieves the best results.

6 Final Remarks

In this paper, we presented a method for 3D face recognition. Our method uses a SA-based approach for range image registration with the SIM. We performed extensive experiments on the FRGC v2 database to evaluate our method's performance. By using this novel approach one can distinguish if two face images with neutral expression belong to the same subject or not with a verification rate of 99%, at a FAR of 0.1%.

From the experimental results we observed that, when comparing a neutral face with other faces with expressions, our method has a slightly lower performance. When using all images from the database, in the *All vs. All* experiment, faces still can be recognized with 95.9% accuracy, at 0.1% FAR.

We also applied an improved approach of the hierarchical evaluation method to perform a verification experiment at 0% FAR [6]. By including partial sum rules during hierarchy analysis, we improved the verification rate from 70.5% to 78.5%, in the *All vs. All* experiment. In the *Fall vs. Spring* experiment, we achieved a verification rate of 96.4%, at 0.1% FAR, and a rank-one accuracy of 99.3%. Although our method suffers when images contain facial expression and noise, we achieved results very close to the one reported by Kakadiaris *et al.* [9], which employs deformable face models to deal with the facial expression effects.

In an identification scenario, our method achieves 99.5% recognition rate in the *First vs. Others* experiment. For all the experiments in the identification mode, our method achieved rank-one accuracy greater than 99%. To our knowledge, these are the best results for this experiment reported in the literature.

References

[1] O. Bellon, L. Silva, C. Queirolo, S. Drovetto Jr., and M. Segundo. 3D face image registration for face matching guided by the surface interpenetration measure. In *Proc. IEEE ICIP*, pages 2661–2664, 2006.

[2] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE PAMI*, 14(2):239–256, 1992.

[3] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *CVIU*, 101:1–15, 2006.

[4] K. Chang, K. Bowyer, and P. Flynn. Multiple nose region matching for 3D face recognition under varying facial expression. *IEEE PAMI*, 28(10):1695–1700, 2006.

[5] J. Cook, C. McCool, V. Chandran, and S. Sridharan. Combined 2D/3D face recognition using log-gabor templates. In *Proc. IEEE AVSS*, page 83, 2006.

[6] S. Drovetto Jr., L. Silva, and O. Bellon. Hierarchical evaluation model for 3D face recognition. In *Proc. VISAPP*, volume 2, pages 67–74, 2008.

[7] T. Faltemier, K. Bowyer, P. Flynn. A region ensemble for 3D face recognition. *IEEE IFS*, 3(1):62–73, 2008.

[8] M. Husken, M. Brauckmann, S. Gehlen, and C. V. der Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. In *Proc. CVPR*, page 174, 2005.

[9] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, T. Theoharis. Three-dimensional face recognition in the presence of facial expression: An annotated deformable model approach. *IEEE PAMI*, 20:pages 640–649, 2007.

[10] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680,

[11] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE PAMI*, 20(3):226–239, 1998.

[12] W. Lin, K. Wong, N. Boston, and Y. H. Hu. 3d face recognition under expression variations using similarity metrics fusion. In *Proc. ICME*, pages 727–730, 2007.

[13] X. Lu, A. K. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE PAMI*, 28(1):31–43, 2006.

[14] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of geometrix activeid 3D face recognition engine on the FRGC data. In *Proc. CVPR*, page 154, 2005.

[15] A. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE PAMI*, 29(11):1927–1943, 2007.

[16] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. CVPR*, pages 947–954, 2005.

[17] C. Queirolo, M. Segundo, O. Bellon, and L. Silva. Noise versus facial expression on 3d face recognition. In *Proc. ICIAP*, pages 171–176, 2007.

[18] M. Segundo, C. Queirolo, O. Bellon, and L. Silva. Face and facial features detection for 3d face recognition under facial expression variations. In *Proc. ICIAP*, pages 431–436, 2007.

[19] L. Silva, O. Bellon, and K. Boyer. Robust range image registration using the surface interpenetration measure and enhanced genetic algorithms. *IEEE PAMI*, 27:762–776, 2005.