3D Face recognition by ICP-based shape matching

Boulbaba Ben Amor[1], Karima Ouji[1], Mohsen Ardabilian[1], Liming Chen[1]
[1] LIRIS Lab, Lyon Research Center for Images and Intelligent Information Systems, UMR 5205 CNRS
Centrale Lyon, France

***Abstract*—In this paper, we propose a novel face recognition approach based on 2.5D/3D shape matching. While most of existing methods use facial intensity image, we aim to develop a method using three-dimensional information of the human face. This is the main innovation of our technology. In our approach, the 3D dimensional information is introduced in order to overcome classical face recognition problems which are pose, illumination and facial expression variations. The paradigm is to build a 3D face gallery using a laser-based scanner: the off-line stage. In the on-line stage, the recognition, we capture one 2.5D face model at any view point and with any facial expressions. Our processing allows the identification of the presented person by performing the captured model with all faces from the database. Here, the Iterative Closest Point-based matching algorithm provides the pose of the probe whereas the region-based metric provides a spatial deviation between the probe and each face from the gallery. In this metric, we calculate the global recognition score as a weighted sum of region-based distances already labelled as mimic or static regions. For automatic 3D face segmentation, we use an immersion version of watershed segmentation algorithm. This paper also presents some experiments in order to shown illumination, pose and facial expression compensations.**

I. INTRODUCTION AND MOTIVATION

Over the past few decades, biometrics and particularly face detection **[1]**, analysis, measurement and description have been applied widely in several applications such as recognition, video surveillance, access control, production of personal documents (i.e. passport, national identity card **[2]**, etc). However, as described in the Face Recognition Vendor Test (FRVT) **[3]** and other face recognition evaluations reports, most commercial face recognition technologies (FRT) suffer from two kinds of problems. The first one concerns inter-class similarity such as twins' classes, fathers and sons' classes. Here, the people have similar appearances which make their discrimination difficult. The second and the more important is the intra-class variations causes by significant changes in lighting conditions, pose variations (i.e. three-dimensional head orientation), and facial expressions (see figure 1). In fact, lighting conditions dramatically change the face appearance, so approaches based on intensity images are not sufficient for overcoming this problem. Pose variations also present immense problem for recognition and handicap for performing comparisons between frontal face images (from dataset) and changed viewpoint images (the probe). Furthermore, compensation of the facial expressions is also a difficult task in 2D-based approaches, because it significantly changes the texture image of the face.

Current state-of-the-art in face recognition is rich in works which aim to resolve problems regarding the challenge. The majority of these researches use intensity images of the face, called 2D model-based techniques, for recognition. In contrast, the second paradigm to recognition is the 3D model-based techniques in which researches exploit, in addition to textural and silhouette information, the three-dimensional shape of the face in order to mitigate some variations. It is in this category of techniques that the proposed approach belongs. Typically, the 3D faces of interest are saved in a library during an offline phase. During the online recognition phase, a single captured 2.5D model of the face present in the scene is matched with the models in the library in order to find the identity and the pose of the person.

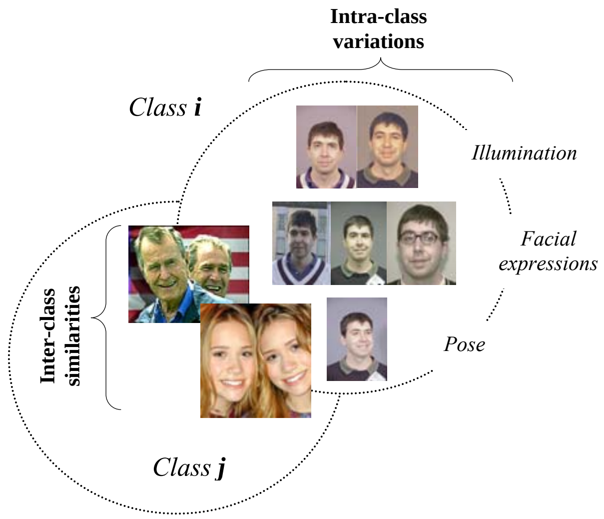


Figure 1. Inter-class similarities and intra-class variations: face recognition problems

The remainder of the paper is organized as follows: Section (2) presents a state-of-the-art in 3D face recognition techniques particularly model-based ones. Section (3) describes an overview of the proposed approach. In Sections (4) we focus on developed works for 2.5D and 3D face acquisition for recognition. The section (5) describes the database collection procedure. In sections (6) and (7), we emphasize our recognition algorithms using the 2.5D/3D face matching coupled with a region-based metric. Finally, a conclusion and future works are presented in section (8).

II. RELATED WORK ON 3D FACE RECOGNITION

Most face recognition algorithms belong to one of the two main families: appearance-based and model-based. The appearance-based methods, also called view-based, use statistical techniques to analyze the distribution of the image vectors in the vector space, and derive an efficient and effective representation (feature sub-space) according to different applications. In fact, the face image is consequently considered as a high-dimensional vector (i.e. a point in a high-dimensional vector) and the dimensional reduction mathematical tools (linear and non-linear) are applied to these vectors. In contrast, the first model-based approaches are based on 2D feature points which are located on the face. Indeed, they compute some distances between these features, considered as a biometric signature. These systems are enhanced, thereafter, by the introduction of the Active Appearance Model **[4]** which aims to cancel the facial expression on a given face and synthesize novel expressions in the same face. A more advanced 3D morphable face model presented in **[17, 16]** is explored to capture the true 3D structure of human face surface. Both morphable model methods come under the framework of “interpretation through synthesis”. These techniques use a generic 3D face model obtained by PCA as the middle step for fitting the 3D morphable model to the presented probe image. The main advantage of model-based methods is that the model, which integrates prior human knowledge, has intrinsic physical relationship with real faces.

As a recent sub-category of the model-based approaches, we find the 3D face matching methods. The uses of 3D images allow the overcoming of limitations due to pose and illumination variations. The appearance of the face is more sensitive than its 3D shape to facial expressions. In addition, the depth image presents several important topological descriptors, especially curvature which is very interesting for feature face localization. All these advantages, as mentioned in **[15]**, encourage the following of these methods in the face recognition challenges. We can group these approaches into two families: statistical-based methods and shape matching-based methods. The works, which belong to the first family, apply the classical linear and non-linear dimensional reduction (PCA, LDA…) techniques to the range images **[9, 12, 13]** from the data collection in order to build the sub-space. The second kind of approaches uses classical shape matching algorithms in order to compute the spatial deviations between the probe and the 3D images from the gallery **[18]**. The test image is typically a 2.5D face model with unknown orientation or a complete 3D face model. These techniques require good initialization for good matching and provide corresponding points, spatial deviation between them and pose parameters of the test image. One of the most important works is described in **[11,7,8]** in which the authors combine a surface matching algorithm with an appearance synthesis approach for a best recognition rate. The final recognition result is a weighted sum of the two scores. In **[14]**, the authors present a novel idea which represents the facial surface as an isometric surface (length preserving). Using a global transformation based on geodesic distance, the obtained forms are independent of facial expressions. After this transformation, they perform one classical rigid surface matching and PCA for sub-space building and face matching. A good review and comparison of these techniques is given in **[10]**. Another interesting study which compares ICP and PCA based approaches is given in **[6]**. Here, the authors show a baseline performance between these approaches and conclude that ICP-based method performs better than a PCA-based method. Their challenge is expression changes particularly, “eye lip open/closed” and “mouth open/closed”. In the present work, we propose a novel approach for face matching using 2.5D partial model in any viewpoint and with any facial expressions. Already, the 3D face database is built in an offline phase. The shape matching is performed by the well-known Iterative Closet point matching algorithm and the recognition is done by a weighted sum of spatial deviation in different regions of the reference face.

III. PROPOSED APPROACH: OVERVIEW

Our people identification approach is based on the face, the most important information to recognize someone. Since 1960, researchers have worked on this popular challenge in order to provide an efficient solution for many applications such as video surveillance, immigration control, access control, etc. The proposed solution can be used for authentication or for enrollment processes (figure 2). In an offline phase, we build our 3D face database with neutral expressions, the faces inside are composed of the 3D meshes and the associated texture images both saved in the same VRML(Virtual Reality Modeling Language) file. In the online step, we first capture one 2.5D face and conserve only the skin region. Then we match the given partial model with all faces in the dataset. The core of our algorithm consists of the alignment step then the matching step of the given surfaces. For the first task, approximating the transformations between the views, we apply a coarse alignment, and then we perform a fine alignment by ICP algorithm. This algorithm is an iterative procedure minimizing the mean square error (MSE) between points in one view and the closest points, respectively, in the other.

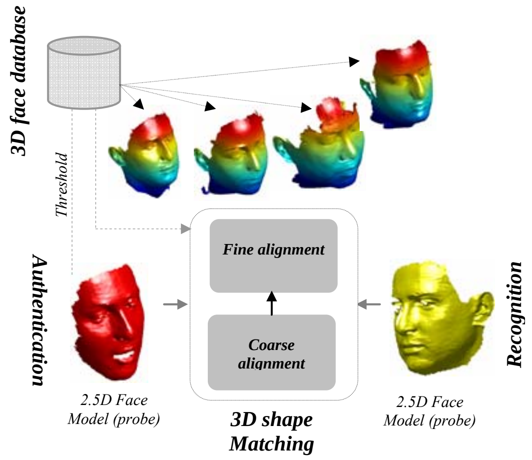


Figure 2. Proposed approaches for face recognition (enrollement) and authentication based on 2.5D/3D face shape matching

The results are two matched sets of points in the 2.5D probe model and the 3D face model from the database. Furthermore, global spatial deviation between each pair of points is provided. The recognition process is based on the obtained distribution of this distance produced by regions.

IV. 2.5D AND 3D FACE PHOTOGRAPHY FOR RECOGNITION

For recognition, we have already developed a complete 2.5D and 3D human face acquisition framework based on a stereo sensor coupled with a structured lighting source. In **[5]**, we propose an accurate and, at the same time, a low-cost solution dedicated to the 3D model-based face recognition techniques (3D-FRT). In our approach, we first calibrate the stereo sensor in order to extract its optical characteristics and geometrical parameters. Second, epipolar geometry coupled with a projection of special structured light on a face, improves the resolution of the stereo matching problem, by transforming it into a one-dimensional search problem and a sub-pixel features matching. Next, we apply our adapted and optimized dynamic programming algorithm to pairs of features which are already located in each scanline. Finally, 3D information is found by computing the intersection of optical rays coming from the pair of matched features. The final face model is produced by a pipeline of four steps: (a) Spline-based interpolation, (b) Partial models' alignment then integration, (c) Mesh generation, and (d) Texture mapping. Figure 3 illustrates some stages of this approach.

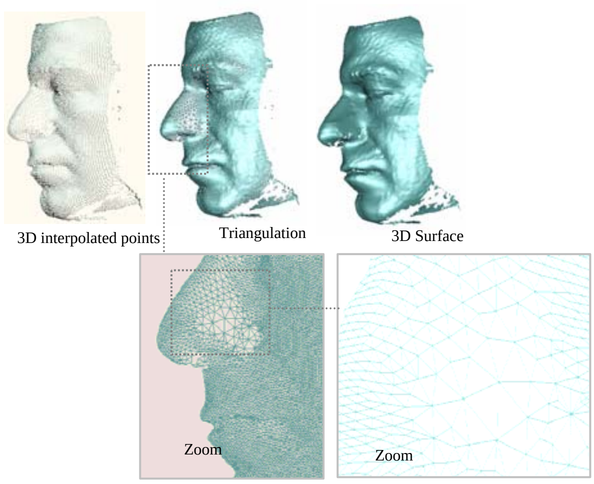
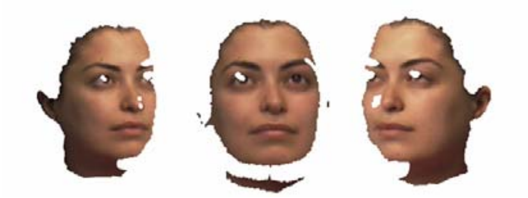


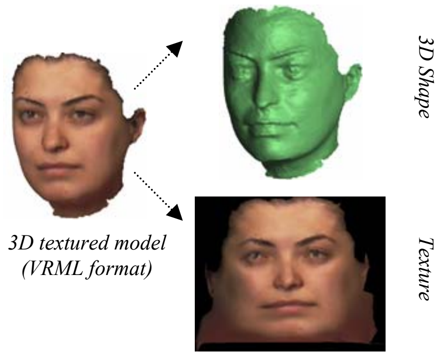
Figure 3. 2.5D face acquisition using stereo sensor assisted structured light

V. DATABASE COLLECTION

As described on the system overview section, we must have a 3D face database for performing recognition from any viewpoint. The complete 3D face is obtained by merging three partial models as shown in figure 4 (a). First, we photograph the subject in three directions then we register the 2.5D partial meshes and merge the texture images. All reconstructed models must have neutral facial expressions as shown by figure 4(b). They have also the same number of vertices in their mesh.



(a) Three partial models for the registration: left, frontal, and right



(b) Texture to shape association procedure

Figure 4. 3D face database collection

For our primary experiments, our database contains 20 3D faces and 180 2.5D test models. For each person the test dataset contains 9 partial models (1 frontal, 2 profiles and 6 with expressions), as illustrated by figure 5. Both 3D and 2.5D faces have about 7000 vertices and 30000 triangles which is a appropriate resolution for representing 3D shape and performing in short time the alignment steps.

(a) right, face, left

(b) neutral, exclamation, unhappy, smiling, content, very happy

Figure 5. (a)Face pose variations (b) facial expression variations

The main goal of this database collection process is to evaluate the robustness of our developed techniques and others to illumination, pose and facial expression variations.

VI. ICP-BASED FACE MATCHING STRATEGY

The main contribution of our approach is the use of dimensional information which is lost by projection in the two-dimensional photos. The intuitive way for recognition is the shape matching process. Many solutions are developed for this task especially for range image registration. The basic algorithm is "Iterative Closest Point" developed by Bessel and al. and published in **[19]**. In our approach we consider one coarse alignment step, which approximates the rigid transformation between models, then we perform our optimized ICP algorithm variant which rapidly converges to a global minima resulting from this initial solution.

A. Coarse alignement

Currently we are working on full-automatic coarse alignment procedure based on 3D face features located by curvature analysis. However, we present here a manual step in which the user must select more than two corresponded 3D points in the probe model and the 3D face model from the gallery. Rigid transformation (R, T, s) including rotation R, translation T and scale s is computed using the selected points. This procedure presents a good initialization stage before the fine alignment stage using ICP. Figure 3 illustrates this process and shows the result of the rigid transformation applied to one of the given models. There are two advantages of the coarse alignment: good initialization which guarantees the convergence to global minima and reduces the convergence time of the ICP algorithm.

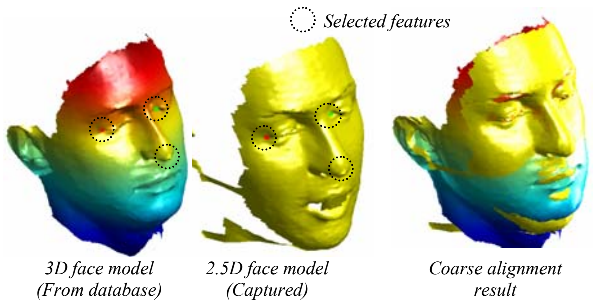


Figure 6. The coarse alignment step (initilisation for ICP)

B. Fine alignment

Our fine alignment algorithm is based on the well-known Iterative Closet Point (ICP) algorithm **[19]** which is an iterative procedure minimizing the mean square error (MSE) between points in one view and the closest points, respectively, in the other. At each iteration of the algorithm, the geometric transformation that best aligns the probe model and the database model is calculated. Intuitively, starting from the two sets of points $P = \{p_i\}$, as a reference data, and $X = \{y_i\}$, as a test data, the goal is to find the rigid transformation (R,t) which minimizes the distance between these two sets of points. The principle of ICP consists of determining for each point p_i of the reference set P the nearest point in the second set X within the meaning of the Euclidean distance. The rigid transformation, minimizing a least square criterion (3), is calculated and applied to the each point of P:

$$e(R,t) = \frac{1}{N}\sum_{i=1}^{N}\left\|(Rp_i + t) - y_i\right\|^2 \quad (3)$$

This procedure is alternated and iterated until convergence (i.e. stability of the minimal error). Indeed, total transformation (R,t) is updated in an incremental way as follows: for each iteration k of the algorithm, $R=R_kR$ and $t=t+t_k$. The criterion to be minimized in the iteration k becomes (4):

$$e(R_k,t_k) = \frac{1}{N}\sum_{i=1}^{N}\left\|(R_k(Rp_i + t) + t_k - y_i\right\|^2 \quad (4)$$

The ICP algorithm presented above always converges monotonically to a local minimum **[19]**. However, we can hope for a convergence to a global minimum if initialization is good. For this reason, we perform the above coarse alignment procedure before the fine one. Figure 7 illustrates zooms of some regions in aligned models; here the 3D model is one neutral 3D model from the gallery whereas the 2.5D model is a scan of the same face with facial expressions (opened mouth and eyes). It's clear that fine alignment contributes to minimizing the distance between the points. In our algorithm we use, for performing ICP, a set of features selected based on the tolerance level of spatial deviation. This allows a rapid convergence of the algorithm which processes only these points and cancels points which presents spatial deviation value superior to the tolerance value. In contrast, correspondence concerns all intersecting points. The steps of the algorithm are given as follows:

Algorithm: *ICP-based matching*

Inputs: *P = {pi} (model from database), X = {yi} (scan model)*
Outputs: *(R,T) which minimise MSE, matched points, spatial deviation*
- *Find closest point pairs (init.: coarse alignment solution)*
- *Compute best transform which minimizes the MSE,*
- *Apply transform to the scan model,*
- *Iterate until numerical convergence, compute at each iteration (R_k,t_k),*

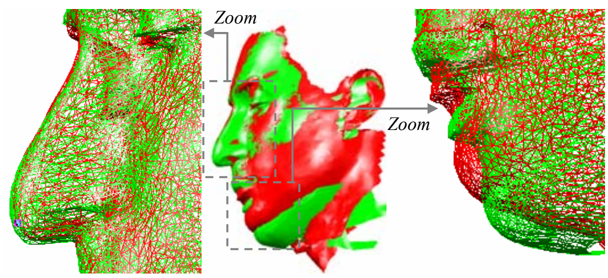


Figure 7. Iterative Closest Point for fine alignement

VII. FACE RECOGNITION METRIC

Performing rigid matching based on the ICP alignment algorithm provides good recognition results. However, it is sensitive to significant facial expression changes. For this reason and based on some empirical observations of face anthropometry, we partition the face model into regions labeled as mimic and static. The mimic regions show mainly variations in the face shape (especially chin, mouth and eyes) whereas static areas present small deformations (the rest of the face). In the recognition process we take into account this classification in order to associate different weights with static and mimic regions. The segmentation process is done automatically by the watershed algorithm which outputs 3D homogenous areas based on the gradient of the range image.

In the sub-sections below we describe, firstly, the global distance for recognition. Secondly, we propose an anthropometric study of the face and an automatic segmentation step. Finally, we propose a novel region-based metric which bears in mind deformable regions and static regions and consequently more robust to facial expressions.

A. Watershed-based face segmentation

3D surface segmentation involves partitioning the surface into groups of subsets of meshes which are homogeneous with respect to some criteria. To understand watershed-based segmentation we consider shape of face as a topographic surface. If we flood this surface from its minima and, if we prevent the merging of the waters coming from different sources, we partition the image into two different sets: the catchment basins and the watershed lines. This is closely the principal of the immersion version of this algorithm proposed by L. Vincent and P. Soille in **[20]** which is used in our technique. The main problem of this approach is the over-segmentation phenomena due to the small variations which exist in the surface of the face. Our solution is to apply a Gaussian then a median filter on the range image in order to cancel the small variations which appear in the 3D surface and to eliminate the undesirable picks. As illustrated by figure 8, this pretreatment allows us to overcome this well-known problem in watershed based segmentation.

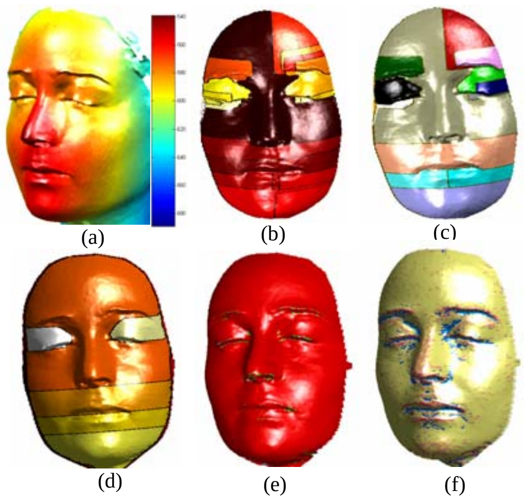


Figure 8. (a) 3D face (b) watershed segmentation without filtering (c) watershed with gaussian filter (d) watershed with median filter (e) Gaussian curvature (f) mean curvature

It is clear that the number of regions in Fig. 8(b) and 8(c) is more significant than in fig. 8(d). This is done by filtering pretraitement applied to the surface of the shape before the segmentation process. We also present results of Gaussian 8(e) and mean curvature 8(f) to discriminate some important features which will be used to automate the coarse alignment stage. In (5) K is the Gaussian curvature and H is the mean curvature.

$$K = \frac{eg - f^2}{EG - F^2}, \qquad H = \frac{eG + Eg - 2fF}{EF - G^2} \qquad (5)$$

where E, F, and G are coefficients of the first fundamental form and e, f, and g are coefficients of the second fundamental form **[21]**.

B. Global distance

Our first approach for recognition using 2.5D/3D shape matching is based on spatial deviation distribution between matched points in probe and the 3D faces from the gallery. The global distance is measured for each pair of corresponding points obtained by ICP. Figure 6 shows two histograms of distribution and a color map which presented by colors the distance under the request (the 2.5D model). For the first request, with changed facial expressions, as shown in figure 9 that the higher parts of the models are better matched than the lower parts. This is because the higher parts are more static than the lower ones. This is the object of the second approach presented in sub-section C, in which a region-based distance is proposed in order to take in account these variations. In the second given example, the request (left profile) is perfectly matched to the 3D face and the pose and the identity of the person is given correctly.

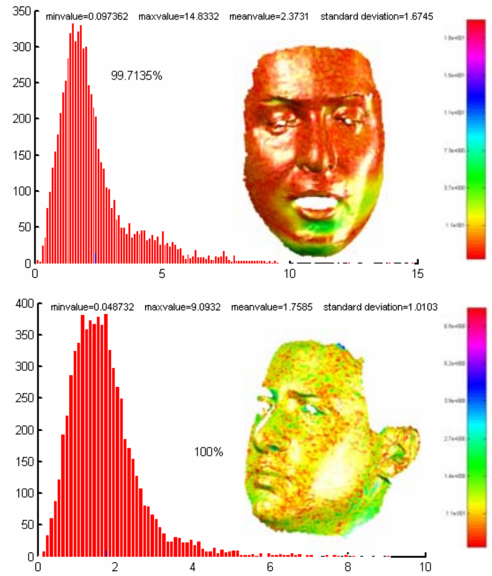


Figure 9. Examples of spatial deviations and colormaps between the 2.5D scan of face and the 3D face from the database

C. Region-based distance

As illustrated in figure 9, the rigid matching process presents some limits to facial expression variations. Our solution is to partition the face model to two kinds of regions: the mimic and the static regions. This human face segmentation model is the result of empirical and anthropometric studies. For the static regions, the rigid matching is more significant for computing recognition score. After surface segmentation we attribute to the obtained regions, different weights and the global distance takes into account the label of each region. This metric is more robust than the global distance calculation. Equation (6) gives the global recognition score function of region-based scores. Here λ_i represent the weights, ψ_i is the individual region score and ψ is the recognition score.

$$\psi = \sum_i \lambda_i \psi_i \qquad (6)$$

This defines a novel face matching metric which is robust to facial expressions by segmenting the face model to static and mimic regions and concludes with a region-based matching score. Here the 2.5D/3D matching allows the pose problem compensation and the region-based score mitigates the facial expression variations. Texture information is not integrated in this work so illumination variations are also compensated. Moreover, it will be integrated in order to enhance the recognition/authentication decision. After probe/gallery models matching, we can extract corresponded texture and compute similarity based on one of the traditional methods (correlation, PCA, etc.) since the pose is already identified.

VIII. Conclusion and Future Work

In this paper we have presented a novel face recognition method which is based on the Iterative Closest Point matching algorithm and a novel region-based metric. As mentioned in section 5, we are currently concentrating on developing a full automatic procedure for the coarse alignment stage. This is based on curvature analysis of the 3D face surface. Our future work concerns the integration of intensity information in our recognition process after the 3D matching. Indeed, the texture image presents the appearance of the human face and provides complementary information for recognition. After ICP alignment we have corresponding 3D points in the probe model and each face belonging to the dataset. In addition, face models are composed of textural and dimensional information, so after ICP processing we can match texture information in probe and gallery models. This can enhance the recognition results and ensure the automatic decision process. Moreover, the presented spatial deviation is the result of point-to-point distance so we will use point-to-plane distance which is more precise for computing spatial deviation between models.

References

[1] Tsishkou D., Hammami M., Chen L., “Face Detection in Video Using Combined Data-mining and Histogram based Skin-color Model”, IEEE Third International Symposium on Image and Signal Processing and Analysis, ISBN 953-184-062-8, Rome, Italy, September 2003. pp. 500-503.

[2] K. Peng, L. Chen and S. Ruan, “Face Image Encryption and Reconstruction for Smart Cards Using Fourrier-Mellin Transform”, Proc. of MEDIANET’04, Tozeur, Tunisie, 2004, pp.

[3] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E Tabassi, and J.M. Bone, “FRVT 2002: Evaluation Report”,. March 2003

[4] T.F. Cootes, G.J. Edwards, and C.J. Taylor, Active appearance models, Proc. ECCV, vol. 2, pp. 484–498, 1998.

[5] B. Ben Amor, M. Ardabilian, L. Chen, “3D Face Modeling Based on Structured-light Assisted Stereo Sensor”. Proceeding of ICIAP 2005, Cagliari, Italia, 6-8 September 2005.

[6] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn. Effects on facial expression in 3D face recognition. In *SPIE Conference on Biometric Technology for Human Identification*, volume 5779 of *SPIE Proceedings*, Orlando, FL, 2005.

[7] Xiaoguang Lu and Anil K. Jain. Integrating range and texture information for 3D face recognition. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 156-163, Breckenridge, CO, 2005.

[8] Xiaoguang Lu and Anil K. Jain. Deformation analysis for 3D face matching. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 99-104, Breckenridge, CO, 2005.

[9] Gang Pan and Zhaohui Wu. 3D face recognition from range data. *International Journal of Image and Graphics*, 5(3):1-21, 2005.

[10] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619-624, 2005.

[11] Xiaoguang Lu, Rein-Lien Hsu, Anil K. Jain, Behrooz Kamgar-Parsi, and Behzad Kamgar-Parsi. Face recognition with 3D model-based synthesis. In *Proc. International Conference on Biometric Authentication*, volume 3072 of *Lecture Notes in Computer Science*, pages 139-146, Hong Kong, 2004. Springer-Verlag.

[12] Thomas Heseltine, Nick Pears, and Jim Austin. Three-dimensional face recognition: An eigensurface approach. In *Proc. IEEE International Conference on Image Processing*, pages 1421-1424, Singapore, 2004.

[13] Sotiris Malassiotis and Michael G. Strintzis. Pose and illumination compensation for 3D face recognition. In *Proc. IEEE International Conference on Image Processing*, pages 91-94, Singapore, 2004.

[14] Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. Expression-invariant 3D face recognition. In *Proc. International Conference on Audio- and Video-based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*, pages 62-70, Guildford, UK, 2003.

[15] Chenghua Xu, Yunhong Wang, Tieniu Tan, and Long Quan. Depth vs. intensity: Which is more important for face recognition? In *Proc. 17th International Conference on Pattern Recognition*, Cambridge, UK, 2004.

[16] Mun Wai Lee and Surendra Ranganath. Pose-invariant face recognition using a 3D deformable model. *Pattern Recognition*, 36(8):1835- 1846, 2003.

[17] Volker Blanz and Thomas Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063- 1074, 2003.

[18] Charles Beumier and Marc Acheroy. Face verification from 3D and grey level clues. *Pattern Recognition Letters*, 22:1321- 1329, 2001.

[19] P. Besel and N. Mckay: “A method for registration of 3D-shapes”. IEEE trans. Pattern analysis and Machine intelligence, 14(2):239-256, 1992.

[20] L. Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", IEEE Transactions on PAMI, Vol. 13, No. 6, June 1991, pp. 583-598.

[21] Gray, A. "The Gaussian and Mean Curvatures" and "Surfaces of Constant Gaussian Curvature." §16.5 and Ch. 21 in Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd ed. Boca Raton, FL: CRC Press, pp. 373-380 and 481-500, 1997.