FAST 3D FACE ALIGNMENT AND IMPROVED RECOGNITION THROUGH PYRAMIDAL NORMAL MAP METRIC

Andrea F. Abate, Michele Nappi, Stefano Ricciardi, Gabriele Sabatino

Università degli Studi di Salerno, Dipartimento di Matematica e Informatica
84084 Fisciano (Salerno), Italy
{abate, mnappi, sricciardi, gsabatino}@unisa.it

ABSTRACT

Face's tri-dimensional shape represents a highly discriminating yet challenging biometric identifier due to different issues, some of which related to capture, alignment and normalization. This paper presents an improved normal map based face recognition approach, which relies on a novel method to automatically align a captured 3D face mesh to a reference template, allowing a more precise face comparison. The alignment algorithm exploits pyramidal-normal-map metric, a coarse to finer measurement of angular distance between two surfaces computed through normal maps with progressively increasing resolution. After the registration has been performed, the normalized face can be rapidly compared to any other template in the gallery database for authentication or identification purposes using standard normal map metric. The alignment approach avoids the need for a rough or manual face pre-alignment and maximizes recognition precision, requiring a fraction of the time needed by the Iterative Closest Point (ICP) method to operate. We show preliminary experimental results on a 3D dataset featuring 235 different subjects.

Index terms - Face recognition, image processing

1. INTRODUCTION

Many of the challenges in face recognition are directly or indirectly related to the modalities in which this biometric is captured and pre-processed, and this applies for 2D and 3D recognition approaches as well. To this regard, Bowyer, Chang and Flynn in their recent survey on the state of the art in 3D face recognition [1] show that face acquisition in three dimensions by means of range image is not free from some of the posing issues typical of 2D face capture. In fact, subject posing within the capture field randomly affects position, rotation and scaling of resulting point cloud.

Consequently, a registration of each enrolled face is required for many recognition methods to perform optimally. Moreover, the wide range of variations in face shape due to facial expressions and to the effects of age, represent additional issues which need to be addressed to allow the diffusion of this biometric in many applicative contexts. The advantage of 3D data is that they represent the actual face shape and not its bi-dimensional projection, so the normalization task usually implies to find the right combination of rotation, traslation and scaling of acquired data to match a reference template. So, despite the greater complexity and cost of 3D shape acquisition there is a growing interest in the research community to exploit the potential of this approach [2-4]. According to literature, 3D based methods can exploit a plurality of metrics, some of which, like Principal Component Analysis (PCA) [5, 6] and Hausdorff distance [7], have been originally proposed for 2D recognition and then extended to range images. Other approaches instead, have been developed specifically to operate on 3D shapes, like those exploiting Extended Gaussian Image [8], the Iterative Closest Point (ICP) method [9, 10], curvature analysis [11] or normal map [12]. One line of work is represented by multi-modal approaches, which typically combine 2D (intensity or colour) and 3D (range images or geometry) facial data and in some case different metrics, to improve recognition accuracy and/or robustness over conventional techniques [13, 14]. This work is focused on improving precision and reliability in normal map based face recognition by means of a fast alignment of captured face to a reference template, before the matching is performed. Normal map, indeed, has proved to be a very fast metric when applied to face recognition [15], but it requires a precise registration of position and rotation between probe and gallery to achieve maximum recognition accuracy. The proposed approach exploits the sensibility of this metric to misalignment when matching surfaces, to measure the distance between two faces and to progressively reduce it to a minimum in the enrolment stage with the smallest possible overhead. As extensions of normal map based face recognition have already been proposed [16] to address issues like facial expression and the presence of beard, they are not included in this work but they could be easily integrated in the approach presented.

This paper is organized as follows. In section 2. the proposed face recognition method is presented in detail. In section 3. experimental results are shown and briefly commented. The paper concludes in section 4. exposing current and future directions for this research.

2. IMPROVING RECOGNITION PRECISION THROUGH PYRAMIDAL NORMAL MAP BASED ALIGNMENT

The overall workflow of proposed method can be resumed as follows. After a face is captured and converted in a 3D surface, it is aligned to a reference template with regard to position. Then, a measurement of its orientation relative to the template is computed, by means of an iterative process until a threshold value is reached.

When this happens, the overall angular difference between the template and the acquired face is used to rotate it, thus completing the alignment. At this point the face is registered and the comparison to each template in the dataset can be performed, rapidly matching the corresponding normal maps.

We discuss the whole recognition process in greater detail throughout the following subsections 2.1. and 2.2..

2.1. Basic normal map based face comparison

As the main contribution of this work is the alignment algorithm which relies on normal map comparison, we briefly expose how this metric works. The idea behind normal map metric is to compare two 3D surfaces by means of their corresponding 2D normal maps, whereas a normal map is a bitmap storing normals to each polygon on the surface as RGB coded pixels. More precisely, to represent these normals by a colour image, we first need to project each 3D vertex onto a 2D space using a spherical projection (opportunely adapted to mesh size). At this point, we can store normals of a given mesh M in a bi-dimensional array N to represent face geometry using the previously 2D-projected vertex coordinate and quantizing the length of the three scalar components of each normal. We refer this resulting array as the Normal Map N of the mesh M and this is the face descriptor we intend to use for one-to-one face comparison. A colour depth of 24 bit per pixel (8 bit for each R,G and B channel) is adequate for biometric applications, providing an angular resolution below one degree for a 180 degrees wide range of values. To compare the normal map N_A from input subject to another normal map N_B previously stored in the reference database, we compute the angle included between each pairs of normals represented by colours of pixels with corresponding mapping coordinates, and store it in a new *Difference Map D* with components r, g and b opportunely normalized from spatial domain to colour domain, so $0 \le r_{N_A}, g_{N_A}, b_{N_A} \le 1$ and $0 \le r_{N_B}, g_{N_B}, b_{N_B} \le 1$. The value θ, with $0 \le \theta < \pi$, is the angular difference between the pixels with coordinates (x_{N_A}, y_{N_A}) in N_A and (x_{N_B}, y_{N_B}) in N_B and it is stored in D as a grey-scale image (see Fig. 1.). At his point we estimate the similarity score between N_A and N_B by means of histogram function H. On the X axis we represent the resulting angles between each pair of comparisons (sorted from 0° degree to 180° degree), while on the Y axis we represent the total number of differences found. The curvature of H represents the angular distance distribution between mesh M_A and M_B, thus two similar faces feature very high values on small angles, whereas two unlike faces have more distributed differences. We define a similarity score through a weighted sum between H and a Gaussian function, as in:

$$similarity\_score = \sum_{x=0}^{k} \left(H(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \right) \qquad (2)$$

where varying σ and k is possible to change recognition sensibility.

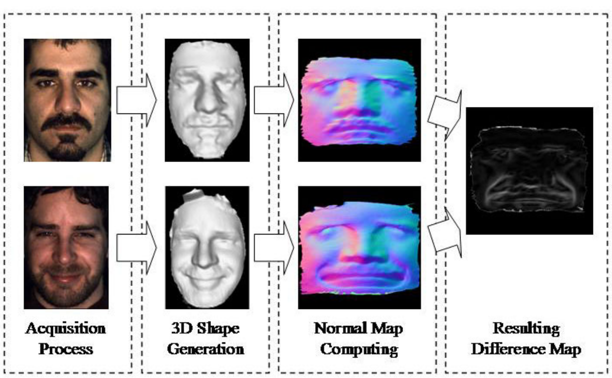

Fig. 1. From capture to comparison exploiting normal map metric

2.2. Pyramidal-normal-map based alignment algorithm

As claimed in the introduction, a precise registration of captured face is required by normal map based comparison to achieve the best recognition performance. So, the obvious choice could be to use the most established 3D shape alignment method, the Iterative Closest Point (ICP), to this aim. Unfortunately ICP is a time expensive algorithm. The original method proposed by Chen and Medioni [17] and Besl and McKay [18] features a $O(N^2)$ time complexity, which has been lowered to $O(N\log(n))$ by other authors [19] and further reduced by means of heuristic functions or by shape voxelization and distance pre-computing in its most recent versions [20]. Nevertheless the best performance in face recognition applications are in the range of many seconds to minutes, depending on source and target shape resolution. This overhead still represents a limit to identification application where the number of comparison can be in the range of many thousands and more. As the whole normal-map based approach is aimed to maximize the overall recognition speed reducing the comparison time, we considered ICP not suited to fit well into this approach.

So we introduce pyramidal-normal-map based face alignment. A pyramidal-normal-map is simply a set of normal maps relative to the same 3D surface ordered by progressively increasing size (in our experiments each map differs from the following one by a factor of 2). Our purpose is to exploit this set of local curvature descriptors to perform a fast and precise alignment between two 3D shapes, measuring the angular distance (on each axis) between an unregistered face and a reference template and reducing it to a point in which it does not significantly affect recognition precision.

The template is a generic neutral face mesh whose centroid corresponds to the origin of the reference system. To achieve a complete registration the captured face has to match position and rotation of reference template. Scale matching, indeed, is not needed as the spherical projection applied to generate the normal map is invariant to object size. The first step in the alignment procedure is therefore to compute face's centroid which allows to match reference template position simply offsetting all vertices by the distance from centroid to the axis origin. Similarly, rotational alignment can be obtained through a rigid transformation of all vertices once the angular distance between the two surface has been measured.

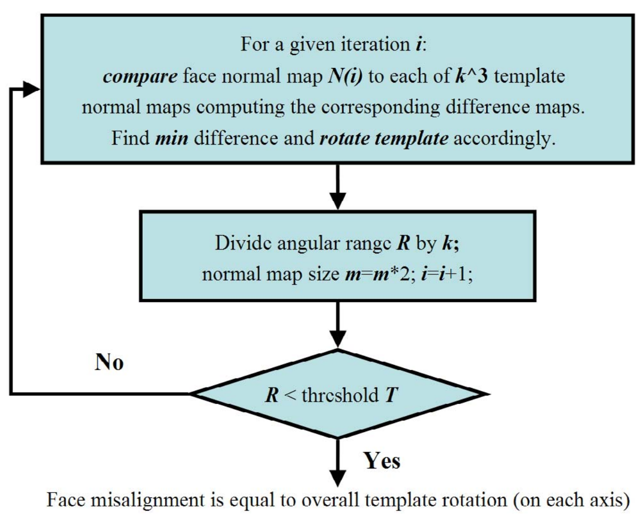

Fig. 2. Main loop of pyramidal-normal-map alignment algorithm

As we intend to measure this distance iteratively and with progressively greater precision, we decide to rotate the reference template instead of captured face. The reason is simple: because the template used for any alignment is always the same, we can pre-compute for every discrete step of rotation the relative normal map once and offline, drastically reducing the time required for alignment. Before the procedure begins, a set-up is required to compute a pyramidal normal map for the captured face (a set of four normal maps with size ranging from 16x16 to 128x128 has proved to be adequate in our tests). At this time the variables controlling the iteration are initialised, like the initial size *m* of normal maps, the angular range reduction factor *k*, and *R*, the maximum angular range for the algorithm to operate, i.e. the maximum misalignment allowed between the two surfaces. We found that for biometric applications a good compromise between robustness and speed is reached setting this value within the $[-90°,+90°]$ range, with *k*=4, but the full $[-180°,+180°]$ range can be reached if required. The main body of the iterative procedure can be described as follows (see Fig. 2). With the first iteration, the smallest normal map in the pyramid is compared to each of *k*^3 pre-computed normal maps of the same size relative to the coarsest rotation steps of template. The resulting difference maps are evaluated to find the one with the highest similarity score, which represent the better approximation to alignment (on each axis) for that level of pyramid. Then the template is rotated according to this first estimate. The next iteration starts from this approximation comparing the next normal map in the pyramid (with size *m*=*m**2) to every template normal map of corresponding size found within a range which has now its centre on the previous approximation and whose width has been reduced by a factor *k*. This scheme is repeated for *i* iterations until the range's width fall below a threshold value *T*. At this point the sum of all *i* approximations found for each axis is used to rotate the captured face, thus resulting in its alignment to the reference template (see Fig. 3.). Using the above mentioned values for initialisation, four iterations (*i*=4) with angular steps of 45°, 11,25°, 2.8° and 0.7° are enough to achieve an alignment adequate for recognition purpose.

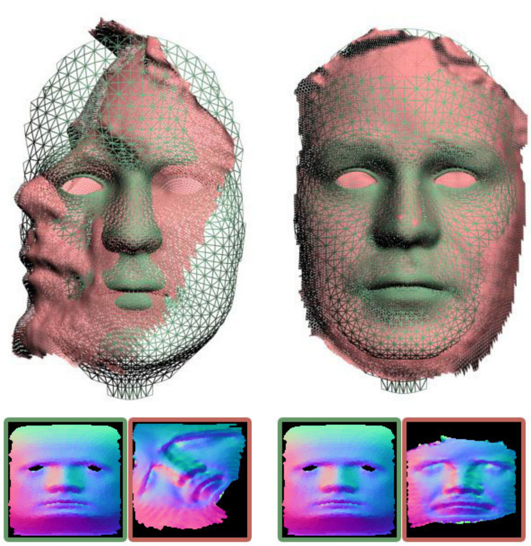

Fig. 3. Face captured (shaded) and reference template (wireframe) before and after alignment. Above: 3D shape, below: normal maps

As the number of angular steps is constant for each level of iteration, the total number of template normal maps generated offline for *i* iterations is *ik*^3, and the same applies to the total number of comparisons. It has to be noted that the time needed for a single comparison (difference map computing), is independent by mesh resolution but it depends on normal map size instead, whereas the time needed to pre-compute each template's normal map depends on its polygonal resolution. The template does not need to have the same resolution and topology of captured face, it is sufficient it has a number of polygons at least greater than the number of pixels in the largest normal map in the pyramid and a roughly regular distribution of vertices. Finally, another advantage of proposed algorithm is that no preliminary rough alignment is needed by the method to converge if the initial face misalignment (for each axis) is within *R*.

3. COMMENTS TO EXPERIMENTS

Because the proposed method works on 3D polygonal meshes, we need to acquire actual faces and to represent them as polygonal surfaces. 3D scanners are typically used to this purpose, capturing range data, converting them in a point cloud and through a further processing in a mesh of connected triangles. Mesh filtering is often required, as the 3D points sampled over the surface may result offset in the form of holes and spikes. In this work the experiments have been conducted using a structured light scanner, the Mega Capturor II from Inspeck Corp., to acquire each subject in an indoor environment. Since this kind of scanner projects a controlled beam of light during capturing, the resulting mesh is not significantly affected from lighting conditions. After the high resolution scanning a 1 out of 4 sub-sampling has been performed to reduce mesh complexity and to filter the surface. We built a 3D database including 235 different individuals (138 males and 97 females, age ranging from 19 to 40) featuring eight expressive or posing variations (neutral, closed eyes, talking, 25° up, 25° down, smile, 25° right, roll right). The enrolled subjects were asked to look at some coloured markers positioned behind the camera, resulting in a rough head orientation for the 25° up, down and right and for the roll poses. Each resulting face has an average number of 60-80.000 polygons, corresponding to a minimum detail of about 1.5 millimetres. For the first group of experiments we wanted to assess the overall face recognition accuracy of proposed method, using all the variations as probes aligned via the proposed algorithm and all the neutral faces as gallery. This test resulted in 96,6% of successful recognitions. The second group of tests is meant to measure alignment accuracy, using the neutral faces for gallery and talking, closed eyes and smile variations as probes. Moreover, the probes have been rotated of known angles on the three axis to stress the algorithm. The results are shown on Fig. 4. where after four iterations, 95.1% of probes have been re-aligned with a tolerance of less than two degree and for 73.1% of them the alignment error is below one degree. The purpose of the third group of experiments is to measure the effect of posing variations and probe misalignment on recognition performance without the alignment step. Also in this case we used the neutral faces for gallery and talking, closed eyes and smile variations, additionally rotated of known angles, as probes. The results in Fig. 5. show that for a misalignment within one degree the recognition rate is 98.1%, which drops to 94.6% if misalignment reaches two degrees. For greater misalignments the performance rapidly get worse (75% at four degrees) , proving that a good probe alignment is important to achieve an high recognition accuracy by normal map metric.

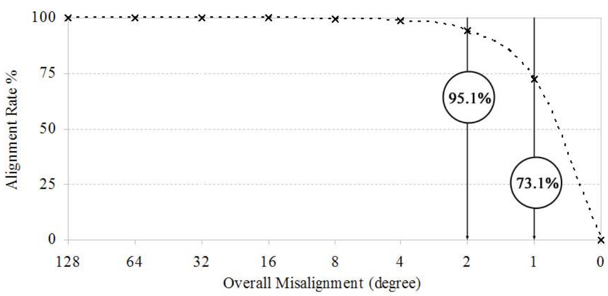

Fig. 4. Error distribution for alignment algorithm

As the average computational cost of a single comparison (128x128 sized normal maps) is about 3 milliseconds for a Dual Amd Opteron 2,4 GHz based PC, the total time needed to alignment is slightly more than 0.3 seconds, allowing an almost real time response. The overall memory requirement to completely store the template's precomputed normal maps is a mere 4 Mbytes (for 4 iterations and normal map ranging from 16x16 to 128x128), as only 4^3 maps for iterations are needed. Concluding, a real biometric identification system based on the proposed methodology could allow to register an enrolled face and to compare it to a gallery of 235 subjects in about 1 second.

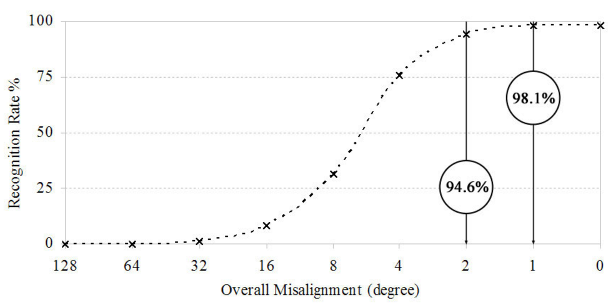

Fig. 5. Effects of posing variations and probe misalignment on recognition rate

4. CONCLUSIONS AND ONGOING RESEARCH

We presented a face recognition method based on normal map metric and improved through a face registration procedure exploiting a novel iterative pyramidal-normal-map based alignment algorithm. The proposed approach maximizes the discriminating potential of normal map metric whose main drawbacks are related to the need for a precisely registered face enrolment, while preserving the high comparison speed typical of this method. The alignment algorithm has proved to be suited to biometric applications and a valid alternative to ICP for this specific task, featuring time complexity not depending from mesh resolution and with an average execution time of about 0,3 seconds on a PC class hardware. The normal map metric is a perfect candidate to vector processor computing, as proved by the partial implementation proposed in [15]. So we are currently working on a fully GPU (Graphics Processing Unit) based implementation of this approach which could perform the alignment and comparison tasks on the latest generation graphics boards, thus allowing to operate efficiently on very large biometric databases. We are also in the process to experiment the proposed method on Face Recognition Grand Challenge dataset to test its performance on a reference gallery.

5. REFERENCES

[1] K.W. Bowyer, K. Chang, P. A. Flynn, Survey of 3D and Multi-Modal 3D+2D Face Recognition, ICPR 2004.

[2] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, N. Murtuza, Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE Workshop on Face Recognition Grand Challenge Experiments, San Diego, CA, June 2005.

[3] X. Lu, D. Colbry, A. K. Jain, Three-dimensional model based face recognition, in: 7th IEEE Workshop on Applications of Computer Vision, 2005, pp. 156–163.

[4] B. Gokberk, A. A. Salah, L. Akarun, Rank-based decision fusion for 3D shape-based face recognition, LNCS 3546: International Conference on Audio- and Video-based Biometric Person Authentication (July 2005) 1019–1028.

[5] Zhang J., Yan Y., And Lades M. 1997. Face Recognition: Eigenface, Elastic Matching, and Neural Nets. in Proc. of the IEEE, vol. 85, no. 9, pp. 1423–1435.

[6] C. Hesher, A. Srivastava, and G. Erlebacher, A novel technique for face recognition using range images, Seventh Int'l Symposium on Signal Processing and Its Applications, 2003.

[7] B. Achermann and H. Bunke, Classifying range images of human faces with Hausdorff distance, 15-th International Conference on Pattern Recognition, pages 809-813, September 2000.

[8] H. T. Tanaka, M. Ikeda, and H. Chiaki, Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition, Third International Conference on Automated Face and Gesture Recognition, pages 372-377, 1998.

[9] G. Medioni and R. Waupotitsch, Face recognition and modeling in 3D. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003), pages 232-233, October 2003.

[10] Lu, X., Jain, A., and Colbry, D.. Matching 2.5d face scans to 3d models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 28, no 1, 2006.

[11] Colombo, A., Cusano, C., and Schettini, R. 2006. 3d face detection using curvature analysis. Pattern Recognition, Vol 39, no 3, 2006

[12] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Fast face recognition based on normal map, in Proceedings of ICIP 2005, IEEE International Conference on Image Processing, Genova, Italy, July 2005

[13] F. Tsalakanidou, D. Tzovaras, M. G. Strintzis, "Use of depth and color eigenfaces for face recognition" , in Pattern Recognition Letters, vol. 24, No. 9-10, pp. 1427-1435, Jan-2003.

[14] T. Papatheodorou, D. Rueckert, "Evaluation of Automatic 4D Face Recognition Using Surface and Texture Registration" , in Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 321-326, May 2004, Seoul, Korea.

[15] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Ultra fast GPU assisted face recognition based on 3d geometry and texture data, in Proceedings of ICIAR 2006, IEEE International Conference on Image Analysis and Recognition, September 2006

[16] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Multi-modal face recognition by means of augmented normal map and pca, in Proceedings of ICIP 2006, IEEE International Conference on Image Processing, July 2006

[17] Y. Chen, G.Medioni, Object modeling by registration of multiple range images, Image and Vision Computing 10 (1992) 145–155.

[18] P. Besl, N. McKay, A method for registration of 3-D shapes, IEEE Transaction on Pattern Analysis and Machine Intelligence 14 (1992) 239–256.

[19] T. Jost, H. H¨ugli, Multi-resolution ICP with heuristic closest point search for fast and robust 3D registration of range images, in: Fourth International Conference on 3-D Digital Imaging and Modeling, October 06 - 10, 2003, pp. 427–433.

[20] P. Yan, K. Bowyer, "A Fast Algorithm for ICP-Based 3D Shape Biometrics" , in Proceedings of the ACM Workshop on Multimodal User Authentication, pp. 25-32, December 2006.