Enhancing 3D Face Recognition By Mimics Segmentation

Boulbaba Ben Amor, Mohsen Ardabilian, and Liming Chen
MI Department, LIRIS Laboratory, CNRS 5205
Ecole Centrale de Lyon, 36 av. Guy de Collongue, 69134 Lyon , France
{Boulbaba.Ben-Amor, Mohsen.Ardabilian, Liming.Chen}@ec-lyon.fr

Abstract

In this paper, we present a novel approach for face recognition and authentication based on dimensional surface matching. While most of existing methods use facial intensity images, a recent set of approaches focus on introducing depth information to surmount some of challenging problems such as pose, illumination, and facial expression variations. The presented matching algorithm is based on ICP (Iterative Closest Point) which aligns one presented probe model to a 3D face model from the gallery data set and provides perfectly its posture. Recognition score is given by a region-based similarity metric which takes into account regions' labels. Here, a specific study in facial expression analysis is done for the labelling purpose. We aim by means of this study at a best way to split and segment the 3D gallery's faces. A new multi-view registered 3D face database including significant variations (pose, illumination, and particularly facial expressions), is collected in order to achieve experiments on our proposal.

1. Introduction and recent work

Over the past few decades, biometrics and particularly face recognition and authentication have been applied widely in several applications such as recognition inside video surveillance systems, and authentication within access control devices. However, as described in the Face Recognition Vendor Test Report [13], and as in other reports, most commercial face recognition technologies suffer from two kinds of problems. The first one concerns inter-class similarity such as twins' classes, and fathers and sons' classes. Here, people have similar appearances which make their discrimination difficult. The second problem, which is more complex, is due to intra-class variations caused by changes in lighting conditions, pose variations (i.e. three-dimensional head orientation), and facial expressions. On the one hand, lighting conditions change dramatically the face appearance; consequently ap-

proaches only based on intensity images are insufficient to employ. On the other hand, pose variations present a considerable handicap for recognition while performing comparisons between frontal face images and changed viewpoint images. In addition, compensation of facial expressions is a difficult task in 2D-based approaches, because it significantly changes the appearance of the face in the texture image.

Current state-of-the-art in face recognition is interesting since it contains works which aim at resolving these problems. The majority of these works use intensity faces' images for recognition or authentication, called 2D model-based techniques. A second family of recent works, known as 3D model-based, exploits three-dimensional face shape in order to mitigate some of these variations. Where some of them propose to apply subspace-based methods, others perform shape matching algorithm. As described in [12, 9, 11], classical linear and non-linear dimensional reduction techniques such as PCA and LDA are applied to range images from data collection in order to build a projection sub-space. Further, the comparison metric computes distances between the obtained projections. Shape matching-based approaches rather use classical 3D surface alignment algorithms that compute the residual error between the surface of probe and the 3D images from the gallery as proposed in [4] and [10]. In [6], authors present a new proposal which considers the facial surface (frontal view) as an isometric surface (length preserving). Using a global transformation based on geodesics, the obtained forms are invariant to facial expressions. After the transformation, they perform one classical rigid surface matching and PCA for sub-space building and face matching. A good reviews and comparison studies of some of these techniques (both 2D and 3D) are given in [7] and [14]. Another interesting study which compares ICP and PCA 3D-based approaches is presented in [8]. 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". Nice reviews and comparison stud-

ies of these techniques (both 2D and 3D) are given in [7], [14], and recently [5].

In the present paper, first we perform experiments on the ICP-based algorithm for both authentication and recognition tasks. Second, we present a novel improved region-based approach. Obtained results show improvements in recognition rates (for recognition) and Error trade-off curves (for authentication).

2. The proposed overall approach

Our 3D face recognition technique is based only on 3D face surface. It consists of two stages: the off-line and the on-line stages. In the off-line stage, full 3D face database with neutral expressions is built (the gallery dataset). Each full 3D face model is obtained via the association of three partial models acquired from different viewpoints (frontal view, left profile view, and right profile view). This 3D face formation process also undergoes with a manual segmentation in different regions which are labelled via different colors. In the on-line stage, the partial model is captured and only the face region is segmented. Then, this probe model is performed to all full 3D faces in the gallery dataset, for identification, or compared to the genuine model, in the authentication case. The core of the second stage consists in aligning surfaces via ICP and then matching them with the designed region-based similarity metric. In fact, final similarity score is the weighted sum of elementary scores given by each region. The sub-sections below detail both off-line and on-line phases of the proposed recognition/authentication paradigm.

### 2.1. Off-line processing of the gallery

As described above, the overall framework requires 3D full face models in the gallery dataset in order to perform matching from any viewpoint. The full 3D face model is obtained by, initially, aligning three 2D 1/2 models followed by merging their meshes and textures. This process is illustrated by Figure 1(a) and 1(b) where we show the partial models and the final 3D full face model after registration and fusion.This full 3D face production process coupled with a manual segmentation stage establishes the off-line gallery processing phase. More details on segmentation process are given in section 4.

2.2. On-line 3D face matching

Within the recognition stage, the presented probe 2D 1/2 model is compared to all 3D models in the gallery dataset. It is done through two elementary processes: coarse and fine alignment. A coarse alignment consists in approximating the rigid transformations between the probe and the full 3D

(a) 2D 1/2 face models with texture images
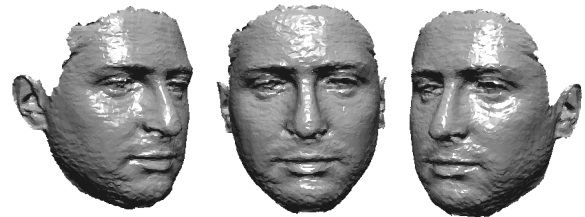
(b) Obtained 3D full face model after registration

Figure 1. 3D full face model after partial model registration.

face. Whereas, a fine alignment uses the ICP (Iterative Closest Point) algorithm [3] to minimize the global deviation between the overlapped surfaces. ICP is an iterative procedure minimizing the MSE (Mean Square Error) between points in partial model and the closest points in the 3D full model. The recognition/authentication decision is made based on the obtained distribution of distances between the surfaces. The weight of one distance depends on the region to which the corresponded vertex belongs.

3. ICP-based 3D face matching

Many solutions are developed to resolve the 3D surface matching task, especially in range image registration and 3D object recognition. The well-known algorithm is the Iterative Closest Point developed by Besl et al. and published in [3]. In our approach we consider first a coarse alignment step, which approximates the rigid transformation between models and brings them closer. Then we perform our variant of ICP algorithm which computes the minimal distance and converges to a minima starting from the latter solution. We use, in our initial experiments, a manual pre-annotation step

IEEE
COMPUTER
SOCIETY

in which some feature points are landmarked in the probe model and in the 3D face model from the gallery. The rigid transformation (R_{init}, t_{init}), including rotation R_{init}, and translation t_{init}, is computed by using the selected points. This initial stage presents a good initialization before the second fine alignment stage using ICP. Figure 2 illustrates this process and shows the result of the rigid transformation applied to the probe model.

There are at least two advantages of the coarse alignment: (a) good initialization which guarantees the convergence to global minima; and (b) reducing time consuming for computing optimal solution.

At each iteration of the algorithm, the geometric transformation that best aligns the probe model and the 3D model from the gallery is computed. 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 target of ICP consists in 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 (1), is calculated and applied to the each point of P:

$$e(R, t) = \frac{1}{N} \sum_{i=0}^{N} \|(Rp_i + t) - y_i\|^2 \qquad (1)$$

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: $R = R_k.R$ and $t = t + t_k$. The criterion to be minimized in the iteration k becomes (2):

$$e(R_k, t_k) = \frac{1}{N} \sum_{i=0}^{N} \|(R_k(Rp_i + t) + t_k - y_i\|^2 \qquad (2)$$


Algorithm 1 Iterative Closest Point algorithm

Require: $P = \{p_i\}$ (from gallery), $X = \{y_i\}$ (probe)
Ensure: (R, T) which minimise the MSE, matched points, spatial deviation between matched points.
  1: Find closest point pairs.
  2: Compute best transform which minimizes the MSE.
  3: Apply transform to the probe model $X = \{y_i\}$.
  4: Iterate (1), (2),and (3) until numerical convergence, compute at each iteration (R_k, t_k).
  Return: (R, t); (p_i, y_i): matched points: (d_i): spatial deviation.

The ICP algorithm (Algorithm 1) presented above always converges monotonically to a local minimum [3]. However, we can hope for a convergence to a global minimum if initialization is good. For this reason, we perform
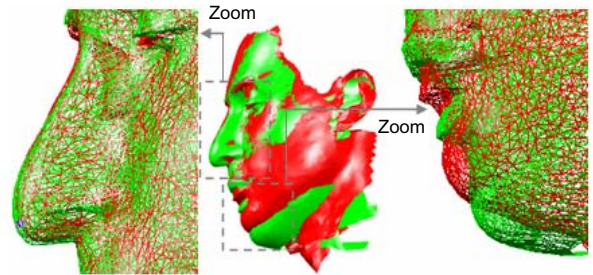
Figure 2. Fine alignement of 3D face surfaces via ICP algorithm.

the previous coarse alignment procedure before the fine one. Figure 2 zoom in on some regions in aligned models; here the 3D model is one neutral 3D model from the gallery whereas the 2D 1/2 model is a scan of the same subject with facial expressions. The fine alignment contributes to minimizing the distance between the points.

In our variant of ICP, we use a set of features selected based on the tolerance level T of spatial deviation. This allows a rapid convergence of the algorithm which processes only these points and cancels points which presents spatial deviation value greater than T. In contrast, deviation between surfaces concerns all points in overlapped region.

4. 3D facial split and segmentation

Our first experiments based on global-ICP [2], without considering the region-based similarity metric, show that the defect of the rigid matching process is linked to its sensitivity to 3D shape deformation. We have concluded that the less the deformation is, the invariant the ICP-based approach is [2]. In the present approach, we suggest to associate different weights to different areas of the face. In fact, there are only some muscles and parts of the face which are responsible to these remarkable deformations. Consequently, we propose to split and segment the full 3D face to different regions labelled from the less to the most static. This can be interesting to compensate facial expression variations.

In order to perform in a best way this segmentation process, we analyze the facial expressions according to both anatomical theory and some empirical tests. Next, we classify muscles according the degree of deformations that they cause. Finally, the 3D full face model is segmented to different patches, each of them is colored with a specific color. In practice, the segmentation step of full 3D faces is inserted in the off-line stage and probe models remain unchanged.
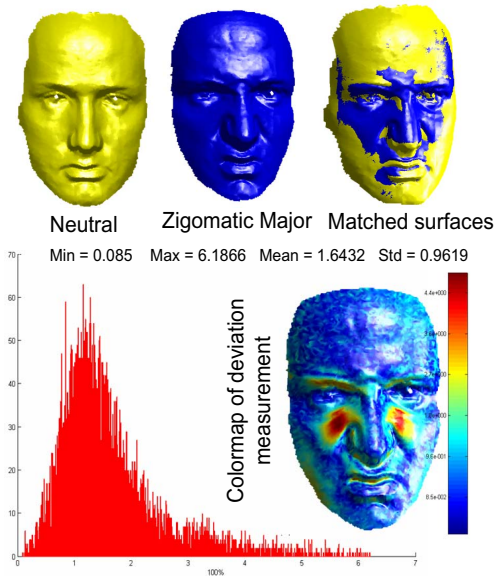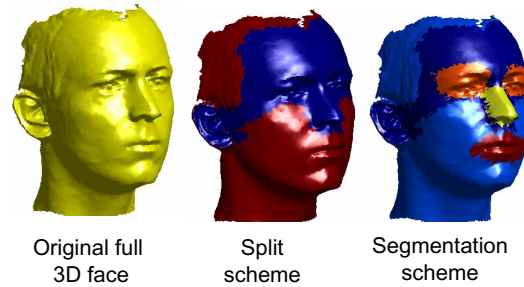
Figure 3. Muscle effects measurements on 3D shape of face.

Figure 4. Common final face split and segmentation schemes.

4.1. Facial expression analysis for labelling

In order to provide a common segmentation scheme for all faces in the 3D gallery face dataset, we studied 3D facial deformations caused by facial expression. Firstly, based on the head anatomy theory, we distinguish some elementary muscles which contribute to the global face deformation within a particular facial expression. This theory shows that there are 14 possible elementary deformations of the 3D face caused by only 11 muscles and the lower jaw. Secondly, an empirical study consists in sorting these areas from static to mimic using the rigid matching approach. In this empirical study, we asked some people to make move only one muscle or part for each emotion. Then, we compute the global distance between the current model and the neutral corresponded common model via ICP. A color map, which represents distance between patches, is also provided in order to locate sensitive regions as shown in Figure 3.

We also sort these face areas by their distance of the corresponded model to the neutral one. According these studies, we conclude that most important deformations are due to the lower jaw, then to the cheeks regions and finally to eyes and to frontal areas. We propose, then, to classify these deformations into some classes according to their degree of deformations. These classes will be considered in the common splitting and segmentation schemes.

This classification allows us to design an efficient segmentation scheme. First, two different regions can be discriminated: lower and higher parts of the face. Lower part

includes the most mimic regions (lower jaw, mouth, chin, and cheeks), whereas the higher part includes the most static regions (nose, ears, front, eyes, and eyebrows). This is the first splitting scheme illustrated in Figure 4. In each part (lower, higher) there are some regions which are more sensitive than others. In the higher part, we can observe that nose and ear regions are more static than eyes and front regions. In the lower part, we can also distinguish the mouth region (lips, teeth and tongue) which causes significant variations for a speaking face. This defines the common segmentation scheme of the 3D full face as illustrated by Figure 4.
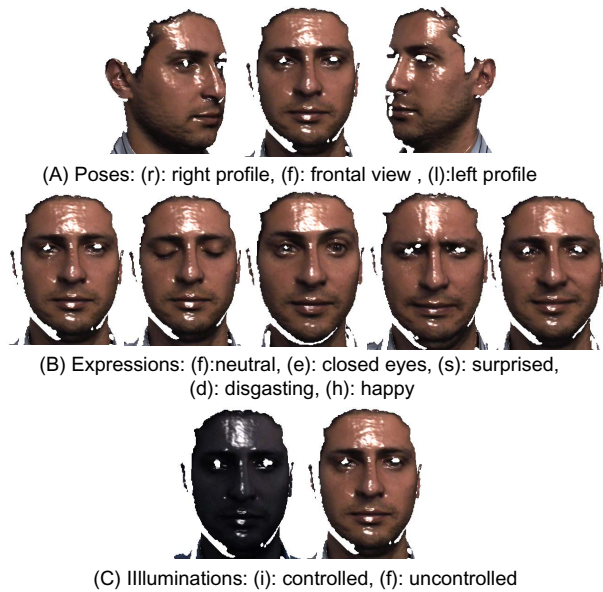
4.2. Region-based similarity metric

As illustrated in Figure 3, the rigid matching process presents some limits to facial expression variations. Therefore, our proposed approach focus on regions of the face in which shape changes are smallest. The new similarity metric is based on the distance between overlapped surfaces outputs of ICP algorithm. However, this distance does not have a uniform weight. In fact, depending on region's label in the 3D full face already split and segment, this metric consists in attributing various weights to various areas. Indeed, the more static the region is, the greater its weight is. Equation (3) gives the global recognition score according to region-based scores. Here α_i represents weights, Ψ_i is the individual region score and Ψ is the recognition score.

$$\Psi = \sum_{i=0}^{N} \alpha_i \Psi_i \qquad (3)$$

5. Experiments and discussions

We have acquired a new 3D face database with significant changes in pose, illumination and especially facial expressions. It consists of 3D faces in the gallery dataset and

(A) Poses: (r): right profile, (f): frontal view , (l):left profile

(B) Expressions: (f):neutral, (e): closed eyes, (s): surprised,
(d): disgasting, (h): happy

(C) Illluminations: (i): controlled, (f): uncontrolled

Figure 5. Variations in probe images for recognition and authentication experiments.

Figure 6. DET curves for all probes with different value of α_2.

2D 1/2 test scans in the probe dataset. In this paper, we provide experiments which pertains to both identification and authentication problems.
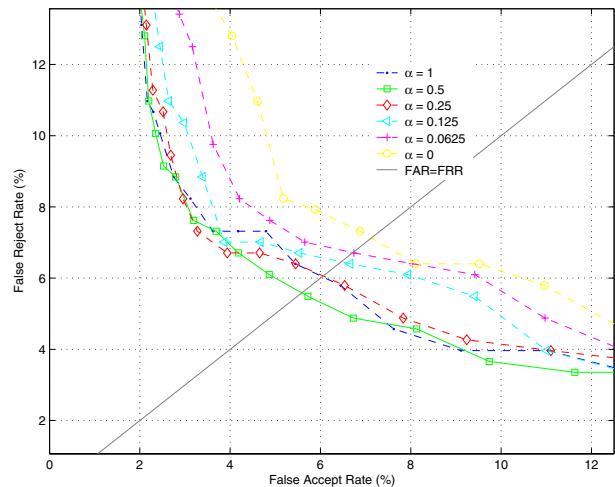
5.1. Database collection

The ECL-3D-face database includes 50 full 3D faces in the gallery and 400 2D 1/2 test models as probes. For each subject, the test dataset contains 8 partial models (1 frontal, 2 profiles, 1 with controlled illumination and 4 with different expressions), as illustrated in Figure 5. Only one acquisition session is considered. Both full 3D and probe models have about 7000 vertices and 13000 faces which is a detail preserving resolution shape and allows the convergence of the ICP algorithm in short time (about 5 s in matlab implementation).

For each face image (2D 1/2 or 3D), a set of meta-data, containing 3D landmarks, pose, illumination and expression values, is stored in XML format. The main target of this new database is to evaluate the robustness of the ICP-based technique and the region-based technique to illumination, pose and facial expression variations.

5.2. Experimental results

Figure 6, 7,and 8 present the obtained results regarding verification (one-to-one matching) and recognition (one-to-many matching) tests using the proposed paradigm. As illustrated in Figure 5, 8 experiments are developed for dif-

ferent variations. In all experiments, similarity matrices, rank-one recognition rates and Error trade-off curves (DET) are produced. Further,for each experiment, 6 other sub-experiments are conducted. Each sub-experiment consider a different value of the regions' weight α. In fact, after the splitting step, we attribute different weights to higher and lower parts of the 3D face. For the higher part, considered as static, α remains still constant ($\alpha_1 = 1$), while for the lower part, considered as mimic, α is variable $\alpha = \alpha_2$. Here, the goal is to find the optimal value of α_2 for which the EER and HTER are close to zero. Figure 6 shows the DET curves performed on the entire probe images including {(d), (s), (h), (e), (i), (l), (r), (f)} variations with different value of α_2.

The best result is obtained for $\alpha_2 = 0.5$, which gives us EER=5.5%.

The DET curves illustrated in Figure 7 and 8 are obtained for two facial expression experiments: (s) and (h) with different values of α_2. According these curves, we can easily note the improvements of the proposed paradigm via the region-based proposed similarity metric for high security applications. In fact, for FAR=0.1%, the accuracy of the system is increased, for α_2 between 0.5 and 0.25, while the FRR is decreased.

In identification experiments (one-to-many matching) we note that the rank-one recognition rate increases from 97.25% for ($\alpha_2 = 1$) to 97.86% for ($\alpha_2 = 0.5$).

Both authentication and identification experiments show robustness of the first approach (global-ICP) to posture and
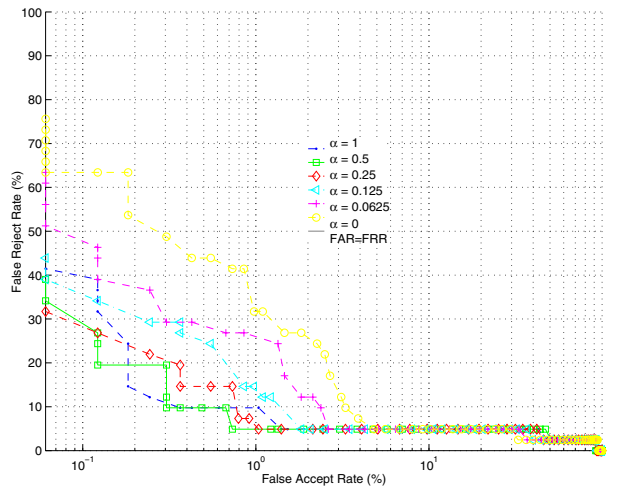
Figure 7. DET curves for Experiment (s) with different value of α_2.
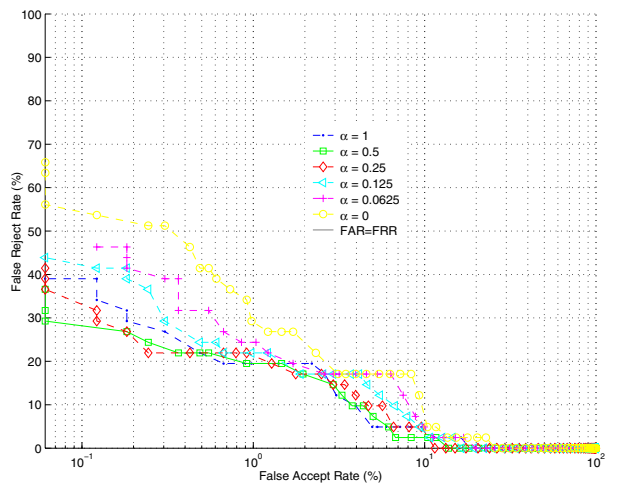
Figure 8. DET curves for Experiment (h) with different value of α_2.

illumination changes. However, algorithm's accuracy decreases with increasing the face shape deformations. An enhanced solution to this problem is given by the region-based technique. This approach gives best recognition and EER rates by the concentration of weights in static regions of the 3D face. In our future work, we plan to work on automatic coarse alignement based on automatic landmarking in face models. Another interesting area is to integrate shape and texture channels in order to increase the accuracy by fusion. Furthermore, the development of our database is to be

continued within the IV2 French project[1] to perform more significant experiments.

References

[1] http://lsc.univ-evry.fr/techno/iv2/.

[2] B. B. Amor, M. Ardabilian, and L. Chen. New experiments on icp-based 3d face recognition and authentication. In *International Conference on Pattern Recognition*, Hong-Kong, 2006.

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

[4] C. Beumier and M. Acheroy. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.

[5] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Comput. Vis. Image Underst.*, 101(1):1–15, 2006.

[6] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision (IJCV)*, 64(1):5–30, August 2005.

[7] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Transactions on PAMI*, 27(4):619–624, 2005.

[8] K. J. Chang, K. W. Bowyer, and P. J. Flynn. Effects on facial expression in 3D face recognition. In *Proceedings of the SPIE, Volume 5779, pp. 132-143 (2005).*, pages 132–143, Mar. 2005.

[9] T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition: An eigensurface approach. In *Proc. IEEE International Conference on Image Processing*, pages 1421–1424, Singapore, 2004. poster.

[10] X. Lu and A. K. Jain. Integrating range and texture information for 3d face recognition. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 156–163, 2005.

[11] S. Malassiotis and M. G. Strintzis. Pose and illumination compensation for 3D face recognition. In *Proc. IEEE International Conference on Image Processing*, pages 91–94, Singapore, 2004. oral.

[12] G. Pan, Z. Wu, and Y. Pan. Automatic 3D face verification from range data. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 193–196, Hong Kong, 2003.

[13] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone. Frvt 2002: Evaluation report. Technical report, NIST, March 2003.

[14] C. Xu, Y. Wang, T. Tan, and L. Quan. Depth vs. intensity: Which is more important for face recognition? In *Proc. 17th International Conference on Pattern Recognition*, Cambridge, UK, 2004. poster.

IEEE
COMPUTER
SOCIETY