

State of The Art

In

3D Face Recognition

Index

1	<u>FROM 2D TO 3D</u>	3
2	<u>SHORT BACKGROUND</u>	4
2.1	THE MOST INTERESTING 3D RECOGNITION SYSTEMS	4
2.1.1	FACE RECOGNITION USING RANGE IMAGES [1]	4
2.1.2	FACE RECOGNITION BASED ON DEPTH AND CURVATURE FEATURES [5]	6
2.1.3	A NEW ATTEMPT TO FACE RECOGNITION USING 3D EIGENFACES [8]	7
2.1.4	FACE RECOGNITION BASED ON FITTING A 3D MORPHABLE MODEL [3]	9
2.1.5	3D FACE MODELING USING TWO ORTHOGONAL VIEWS AND A GENERIC FACE MODEL [2]	11
2.1.6	ASYMMETRIC 3D/2D PROCESSING FOR FACE RECOGNITION [22]	12
2.1.7	COMPONENT-BASED FACE RECOGNITION USING 3D MORPHABLE MODELS [10]	13
2.1.8	ADAPTIVE RIGID MULTI-REGION SELECTION (ARMS) [12]	15
2.1.9	THE PARTIAL ICP (ITERATIVE CLOSEST POINT) ALGORITHM [13]	17
2.1.10	3D FACE RECOGNITION USING POINT SIGNATURE [14]	19
2.2	THE MOST INTERESTING 2D+3D MULTI-MODAL SYSTEMS	21
2.2.1	FEATURES EXTRACTION USING GABOR FILTERS FOR 2D AND POINT SIGNATURE FOR 3D [9]	21
2.2.2	PCA-BASED MULTI-MODAL 2D+3D FACE RECOGNITION [4]	21
2.2.3	3D+2D FUSION AT BOTH FEATURE AND DECISION LEVELS USING LOCAL BINARY PATTERNS [15]	22
2.2.4	HIERARCHICAL MATCHING USING ICP FOR 3D AND LDA FOR 2D [17]	24
2.2.5	FACE RECOGNITION FROM 2D AND 3D IMAGES USING 3D GABOR FILTERS [18]	25
3	<u>DISCUSSION</u>	28
4	<u>REFERENCES</u>	29

Index of Figures

FIG. 1 THE REGULATED MESH MODELS IN DIFFERENT LEVELS. (A) BASIC MESH. (B) LEVEL ONE. (C) LEVEL TWO. (D) LEVEL THREE. (E) LEVEL FOUR.....	8
FIG. 2 (SOURCE: ARTICLE [2]): DERIVED FROM A DATABASE OF 200 LASER SCANS, THE 3D MORPHABLE MODEL IS USED TO ENCODE GALLERY AND PROBE IMAGES. FOR IDENTIFICATION, THE MODEL COEFFICIENTS A AND B OF THE PROBE IMAGE ARE COMPARED WITH THE STORED COEFFICIENTS OF ALL GALLERY IMAGES.....	10
FIG. 3 FRONT AND PROFILE VIEW OF A SUBJECT.....	11
FIG. 4 GENERIC 3D MODEL.....	12
FIG. 5 GENERATION OF THE 3D MODEL.....	14
FIG. 6 EXAMPLES OF THE COMPONENTS EXTRACTED FROM A FRONTAL VIEW AND HALF-PROFILE VIEW OF A FACE.....	14


State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	


FIG. 7 ROC CURVES FOR THE COMPONENT-BASED AND THE GLOBAL FACE RECOGNITION SYSTEMS.....	15
FIG. 8 IMAGES OF THE SAME PERSON WITH DIFFERENT EXPRESSIONS RENDERED.....	16
FIG. 9 PROBES WITH NEUTRAL EXPRESSION.....	17
FIG. 10 PROBES WITH NON-NEUTRAL EXPRESSION.....	17
FIG. 11 DISCARDED AREA IN FACIAL SURFACE WITH DIFFERENT $P-RATE=0.9, 0.7, 0.2$	18
FIG. 12 RANK-ONE RATE: PCA VS. PARTIAL ICP.....	19
FIG. 13 DEFINITION OF POINT SIGNATURE.....	20
FIG. 14 SINGLE- VERSUS MULTI-MODAL BIOMETRICS.....	22
FIG. 15 CALCULATION OF LBP CODE FROM 3X3 SUBWINDOW.....	22
FIG. 16 CUMULATIVE MATCH CURVES FOR 3D AND 2D.....	23
FIG. 17 CUMULATIVE MATCH CURVES FOR 3D+2D FUSION.....	24
FIG. 18 CUMULATIVE MATCHING PERFORMANCE.....	25
FIG. 19 RANK-ONE MATCHING ACCURACY ($A = 1$) WITH AND WITHOUT HIERARCHICAL STRUCTURE.....	25
FIG. 20 A CONVOLUTION EXAMPLE BASED ON THE 3D SGF. (A) ORIGINAL IMAGE; (B).....	26
FIG. 21 RECOGNITION RATES OF FRONTAL TEST FACES. (A) GABOR-BASED (B) EIGENFACE.....	27
FIG. 22 RECOGNITION RATES OF NON-FRONTAL TEST FACES. (A) GABOR-BASED (B) EIGENFACE.....	27

INDEX OF TABLES

TABLE 1 RESULTS OF THE RANGE BASED METHOD IN TERMS OF RECOGNITION RATE.	5
TABLE 2 RESULTS OF THE DEPTH AND CURVATURE BASED FACE RECOGNITION METHOD.	7
TABLE 3 CCR IN MANUAL DB (USING NN AND KNN).	8
TABLE 4 CCR IN THE FIRST 30 PERSONS OF THE AUTOMATIC DB (WITH AND WITHOUT NON-FACE MESHES, USING NN AND KNN).	8
TABLE 5 CCR IN THE AUTOMATIC DB (WITH AND WITHOUT NON-FACE MESHES, USING NN AND KNN).	9
TABLE 6 MEAN AVERAGE IDENTIFICATION RATE ON THE CMU-PIE DATA SET, AVERAGED OVER ALL LIGHTENING CONDITIONS FOR FRONT, SIDE AND PROFILE VIEW GALLERIES.	10
TABLE 7 GALLERY IMAGES ARE FRONTAL VIEWS EXTRACTED FROM BA. ONLY CONDITION BK HAS DIFFERENT ILLUMINATION CONDITION THAN THE OTHERS.	11
TABLE 8 VOTING RATE OF EACH MODEL USING THE DIFFERENT POINT SIGNATURES OF SCENE (FACE 1 TO FACE 6)	21
TABLE 9 SUMMATION OF THE OUTLINED ALGORITHMS, IN THEIR ORDER OF APPEARANCE.	29

1 From 2D to 3D

Evaluations such as the FERET Tests and Face Recognition Vendor Test 2002 [6], have underlined that the current state of the art in 2D face recognition is not yet sufficient for use in biometric applications. Indeed, although recent algorithms show fairly high accuracy under tightly constrained conditions, their performances degrade significantly when these constraints are relaxed. Thus, variations due to pose, facial expression, aging, occlusion of parts of the face and illumination of the scene bring to the face recognition community new problems to solve. Of course, some efforts

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

have already been made in order to overcome these difficulties, but obtained results, in terms of recognition rate confirmed that these types of variations are a real challenge.

Since the beginning of the nineties, 3D has been believed to be a good means to tackle some of the cited problems, as it is intrinsically pose and illumination invariant. But this technique has long been left aside by the community due to the cost and inaccuracy of 3D sensors. In the recent years, these sensors have become cheaper, faster and more accurate, allowing works of face recognition using 3D models of faces to be carried out.

Unlike in the 2D case, only a few databases of 3D face model are publicly available to date and in most cases the number of subjects and the quality of the 3D models are quite low. This leads in a fairly high number of works to testing promising techniques on ad-hoc, small 3D face databases and makes the real efficiency of the techniques difficult to evaluate. . The 3D_RMA is an example of a database of 3D face models represented by clouds of points. For a long time it has been the only publicly available database, and was of rather low quality. On the other hand 3D meshes (wrl, 3ds, ...) are available today from the newest technologies, but in most cases are proprietary databases. Nevertheless, the constant progress of 3D capturing technologies is influencing the quality of the recognition techniques. Indeed, the first algorithms applied directly on clouds of points [8], after a suitable triangulation, while more recent ones work directly on meshes, considering in some cases the information provided both by the 3D shape and texture.

2 Short Background

To date, the majority of research works in the field of face recognition, and all of the major commercial face recognition systems use intensity images of the face. This paradigm is referred to as 2D face recognition. On the other hand, 3D face recognition takes into consideration the shape of the head, and in particular brings information about the depth, which is lost in 2D.


For both 2D and 3D, the word *identification* refers to a one-to-many matching where the best matched person of the database is sought. *Verification* is means a one-to-one matching, through which is verified a claimed identity. Finally, *multi-modal* refers to a strategy that combines 2D and 3D information in order to achieve the recognition task.

A brief description of the most interesting recognition schemes available in the literature is given in the following.

2.1 The most interesting 3D recognition systems

2.1.1 Face Recognition Using Range Images [1]

Some of the first approaches to the 3D face recognition worked on range data directly obtained by range sensors, due to the low costs of this hardware, respect to the laser scanners used for example by Gordon in [5]. In fact, in [1] the range images are acquired by means of a structured light approach. The most important disadvantage of this choice are the missing data due to the occlusions or improperly reflected regions. This problem is then avoided using two sensors rather than one, and applying a merging step for integrating the obtained 3D data. The initial step, consists in calibrating the sensors, so that such parameters as projection matrix, camera direction, ... are computed. Then merged images are computed, that is for every original 3D data point, the coordinates in the merged range image are calculated based on

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

the parameters of the virtual sensor. If 3D points of two different surfaces have to be mapped onto the same pixel, a sort of z-buffering is applied to disambiguate the mapping. The template images obtained from this acquisition process are then used as training and testing set for two different approaches. The first one are the eigenfaces. The dimension of the space of face templates is reduced applying the Principal Component Analysis both for training and testing, so that for each testing image the nearest one in terms of euclidian distance is searched. Another method is also tested on the template images, the HMMs (Hidden Markov Models). As this technique is only applicable on one-dimensional signals, the template images are first transformed in a mono-dimensional signal by means of a slide window, that move on the image from the top to the bottom and from the left to the right. The HMM has five states. For every person in the database, the parameters of the HMM are calculated in a training phase. When a test image is presented, the probability of producing this image is computed by means of the Viterbi algorithm. All images in the database have a size of 75×150 pixels. Since both the methods require a training phase, the dataset has been partitioned in two subsets of 120 training and 120 test images. Results are shown in Table 1. For the experiments reported under the category *smoothing*, no rotation was done, but an additional smoothing step was applied, with $\sigma=0.5$ and $\sigma=1.5$. On the contrary, the rotations are controlled, so that the rotation around the y axis is constantly 30° and the rotation around the x axis is 20°.

Preprocessing	Eigenfaces	HMM
No preprocessing	97.50 %	90.83 %
Smoothing $\sigma=0.5$	98.33 %	90.00 %
Smoothing $\sigma=1.5$	98.33 %	76.67 %
Rotation	100.00 %	89.17 %

Table 1 Results of the range based method in terms of recognition rate.

2.1.2 Face Recognition Based on Depth and Curvature Features [5]

In past 2D approaches, the features used in describing faces have been limited to eyes, nose, mouth and face boundaries, neglecting the additional information provided by low contrast area, such as jaw boundary, cheeks and forehead. Then is clear that an approach based on range and curvature data has several advantages over intensity image approaches by virtue of the more available information. Furthermore curvature has the valuable characteristic of being *viewpoint invariant*. This method defines a set of high level features, which are eyes, nose and head, and includes the following features: *Nose bridge* (nasion), *Nose base* (base of septum), *Nose ridge*, *Eye corner cavities* (inner and outer), *Convex center of the eye* (eyeball), *Eye socket boundary*, *Boundary surrounding nose*, *Opposing positions on the cheeks*. Each of these regions on the face image is described in terms of a set of relationships of depth and curvature values. Since several region can respect a set of constraints, this set is designed in order to reduce the search to a single definition of the feature. The set of constraints is given by:

- sign of Gaussian and mean curvature,
- absolute extent of a region on the surface,
- distance from the symmetry plane,
- proximity to a target on the surface,
- protrusion from the surrounding surface,
- local configuration of curvature extrema.

The high level features and regions are used to compute a set of low level features, where the most basic scalar features correspond to measurements of distances. The set of low level descriptors is given by: *left and right eye width*, *eye separation*, *total span of eyes*, *nose height/width/depth*, *head width*, *maximum Gaussian curvature on the nose ridge*, *average minimum curvature on the nose ridge*, *Gaussian curvature at the nose bridge and base*. For each face image, this set of features is computed, placing it in the space of all possible faces, while the Euclidean distance is used as measure in the scaled feature space. Two different training sets are used for feature detection and recognition. The former consists of 26 subjects, while the latter includes 8 faces with 3 views each for a total of 24 faces. Two different characteristics of the method are assessed in the experimentation. The first is a classification of the low level features on the basis of two main properties: *robustness in detection* (their measurements have to be consistent for the same face also when pose and/or expression change), *discriminating power* (their values must vary distinctly over the range of different individuals). The second is the recognition rate obtained for different sets of low level features. The results are shown in Table 2. For each of the targets there are two faces with the same identity remaining in the database. Table 2 shows for each feature set the percentage of targets for which the best match was correct (top), and the percentage of the targets for which the second best match was also correct (bottom). The basic set denoted (I), includes the best 4 features head width, nose height, depth and width, while the other three set include increasing numbers of features added according to their discriminating power.

Feature set considered	Recognition Rates
I	75.0 %
	70.8 %

II	91.7 %
	83.3 %
IV	95.8 %
	79.2 %
III	100.00 %
	79.2 %

Table 2 Results of the depth and Curvature based face recognition method.

2.1.3 A New Attempt To Face Recognition Using 3D Eigenfaces [8]

This method apply on the face models of the 3D_RMA database, in which models are represented by scattered point clouds. So the first problem to be addressed consists in building the mesh from the clouds of points. This is done by means of an iterative algorithm. At first the nose tip is localized as the most prominent point in the point cloud. Then a basic mesh is aligned with the point cloud, subdivided and tuned step by step as shown in Fig. 1. Four step are considered enough for the refinement process. Point clouds have different orientations, and resulting meshes preserve this orientation, so an average model is computed and all the meshes are aligned with it, tuning six parameters for the rotations and six for the translations. Due to the noise, some built mesh models cannot describe the geometric shape of the individual. These mesh models are called non face models. Each mesh contains 545 nodes and it is used as a bi-dimensional intensity image, in which the intensity of the pixel is the Z-coordinate of the corresponding node. The eigenfaces technique is applied to these intensity images. A subset of the computed images is used for the training. Let M_1, M_2, \dots, M_n the mesh images in the training set and let be M_{aver} the average mesh image and $\Phi_i = M_i - M_{aver}$. Then the covariance matrix is computed as:

$$C = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T = AA^T$$

The eigenvalues and the eigenvectors of the matrix C are computed, keeping only the $e < n$ orthogonal eigenvectors u_1, u_2, \dots, u_e , which correspond to the $e < n$ largest eigenvalues. Both training and testing mesh images are projected on this space, while 20 most dominant 3D eigenfaces are considered, so that a vector $V \in \mathbb{R}^{1 \times 20}$ is associated to each mesh model. The database used for the experiments is the 3D_RMA, which consists of subjects acquired in two different sessions. From these sessions two databases are built: automatic DB (120 persons) and manual DB (30 persons). After computing the similarity differences between test samples and the training data, the nearest neighbor classifier (NN) and the k-nearest neighbor classifier (KNN) are used for recognition.

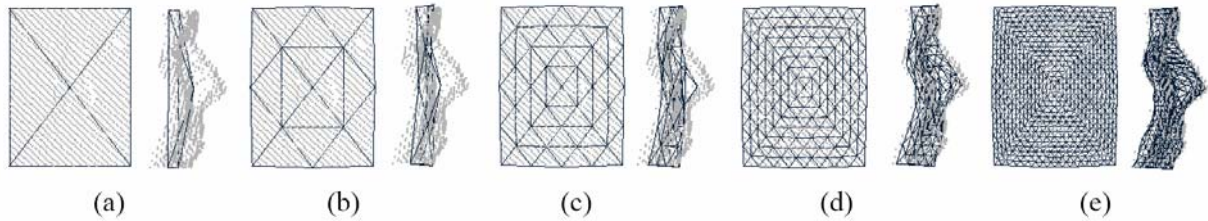


Fig. 1 The regulated mesh models in different levels. (a) Basic mesh. (b) Level one. (c) Level two. (d) Level three. (e) Level four.

The identification accuracy is evaluated on different subsets of the 3D_RMA database. Because of the limited quantity of samples, the Leave-one-out Cross Validation method is used (i.e. each time a mesh image is left out as a test sample and the training is done on the remainder), while the CMS (Cumulative Match Score) is used to evaluate the identification performances. Results are reported in the following tables which show the CCR (Correct Classification Rate) in manual DB (using NN and KNN), the CCR in the first 30 persons of the automatic DB (with and without non-face meshes, using NN and KNN) and the CCR in the automatic DB (with and without non-face meshes, using NN and KNN). Table 3, Table 4 and Table 5

Database	NN	KNN
Manual DB, session 1 (3 instances for each)	92.2 %	92.2 %
Manual DB, session 2 (3 instances for each)	84.4 %	84.4 %
Manual DB, sessions 1 and 2 (6 instances for each)	93.9 %	93.9 %

Table 3 CCR in manual DB (using NN and KNN).

Database	First 30 Models		Non-face meshes removed (22 persons)	
	NN	KNN	NN	KNN
Automatic DB, session 1 (3 instances for each)	71.1 %	73.3 %	83.3 %	83.3 %
Automatic DB, session 2 (3 instances for each)	80.0 %	80.0 %	89.4 %	89.4 %
Automatic DB, session 2 (3 instances for each)	80.6 %	82.2 %	92.4 %	92.4 %

Table 4 CCR in the first 30 persons of the automatic DB (with and without non-face meshes, using NN and KNN).

Database	All Models (120 persons)		Non-face meshes removed (91 persons)	
	NN	KNN	NN	KNN
Automatic DB, session 1 (3 instances for each)	59.2 %	60.3 %	71.1 %	71.8 %
Automatic DB, session 2 (3 instances for each)	59.2 %	61.1 %	67.4 %	68.5 %
Automatic DB, session 2 (3 instances for each)	69.4 %	71.1 %	79.3 %	80.2 %

Table 5 CCR in the automatic DB (with and without non-face meshes, using NN and KNN).

2.1.4 Face Recognition Based on Fitting a 3D Morphable Model [3]

This face recognition system combines deformable 3D models with a computer graphics simulation of projection and illumination. Given a single image of a person the algorithm automatically estimates 3D shape, texture and all relevant 3D scene parameters. The morphable face model is based on a vector space representation of faces. This space is constructed, such that any convex combination of the examples S_i and T_i belonging to the space, describes a human face:

$$S = \sum_{i=1}^m a_i S_i \text{ and } T = \sum_{i=1}^m a_i T_i$$

In order to assure that continuous changes on a_i and b_i represent a transition from one face to another, avoiding artifacts a dense point-to-point correspondence constraint has to be guaranteed. This is done by means of a generalization of the optical flow technique on gray-level images to the three-dimensional surfaces. Vectors S and T are directly extracted from the 3D model, where S is the concatenation of the Cartesian coordinates (x, y, z) of the 3D points and T is the concatenation of the corresponding texture information (R, G, B) . Furthermore the PCA is applied to the vectors S_i and T_i of the example faces $i=1, 2, \dots, m$, while the Phong's model is used to describe the diffuse and specular reflection of the surface. In this way an average morphable model is derived from scans (obtained with a Cyberware™ 3030PS laser scanner) of 100 males and 100 females, from 18 to 45 years old. Then, by means of a cost function, the fitting algorithm optimizes a set of shape coefficients and texture coefficients along with 22 rendering parameters concatenated in a feature vector ρ , such as pose angles, 3D translations, focal length, Two paradigms have been used in order to test the method. In the first one all gallery images are analyzed by the fitting algorithm, and the shape and texture coefficients are stored. In the same way, for a probe image all the coefficients are computed and compared with all gallery data, in order to find the best match, a graphical representation is given in Fig. 2. In the second one, the three-dimension face reconstruction is used in order to generate synthetic views of the subjects in a 2D face database, which are then transferred to a second viewpoint-dependent recognition system.

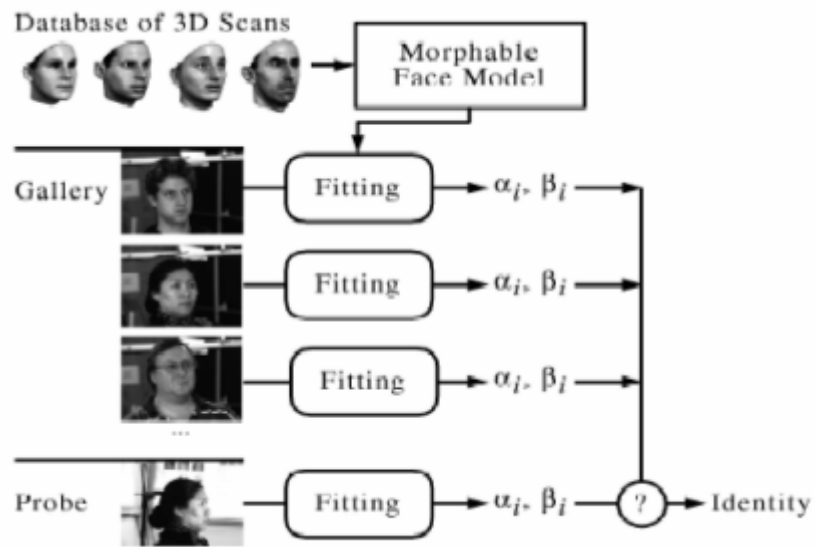


Fig. 2: Derived from a database of 200 laser scans, the 3d morphable model is used to encode gallery and probe images. For identification, the model coefficients α and β of the probe image are compared with the stored coefficients of all gallery images.

Model fitting and identification have been tested on two publicly available databases of images: CMU-PIE et FERET. The individuals in these databases are not contained in the set of 3D scans that form the morphable face model. The reconstruction algorithm is run on all 4,488 PIE and 1,940 FERET images. For all images, the starting condition is the average face at a front view and with frontal illumination. On each image, between six and eight feature points are manually defined, while the following distance function has been used to compare two different faces c_i :

$$d_w = \frac{\langle c_1, c_2 \rangle_w}{\|c_1\|_w \cdot \|c_2\|_w} \text{ with } \langle c_1, c_2 \rangle_w = \langle c_1, C_w^{-1} c_2 \rangle \text{ and where } C_w \text{ is the covariance matrix.}$$

Table 6 and Table 7 show some detailed results on the PIE and FERET databases:

Probe view	Gallery view		
	Front	Side	Profile
Front	99.8 %	99.5 %	83.0 %
Side	97.8 %	99.9 %	86.2 %
Profile	79.5 %	85.7 %	98.3 %
Total	92.3 %	95.0 %	89.0 %

Table 6 Mean average identification rate on the CMU-PIE data set, averaged over all lightening conditions for front, side and profile view galleries.

Probe view	Correct identification
<i>Ba</i>	(gallery)
<i>Bb</i>	94.8%
<i>Bc</i>	95.4%
<i>Bd</i>	96.9%
<i>Be</i>	99.5%
<i>Bf</i>	97.4%
<i>Bg</i>	96.4%
<i>Bh</i>	95.4%
<i>Bi</i>	90.7%
<i>Bk</i>	96.9%
Total	95.9%

Table 7 Gallery images are frontal views extracted from *Ba*. Only condition *Bk* has different illumination condition than the others.

2.1.5 3D Face Modeling Using Two Orthogonal Views and a Generic Face Model [2]

This method uses the 3D coordinates of a set of facial feature points, calculated from two images of a subject, in order to deform a generic 3D face model. Images are grabbed by two cameras with perpendicular optical axes. The 3D generic model is centered and aligned by means of the procrustes analysis, which models the global deformation, while local deformation are described by means of the 3D spline curves. The front and profile view of a subject are shown in Fig. 3. They are used in order to locate facial features, such as eyes and mouth, by means of a probabilistic approach. An example of the obtained 3D model is given in Fig. 4. Twenty-nine vertices are kept on this model, divided in two subsets, 15 principal vertices and 14 additional vertices.

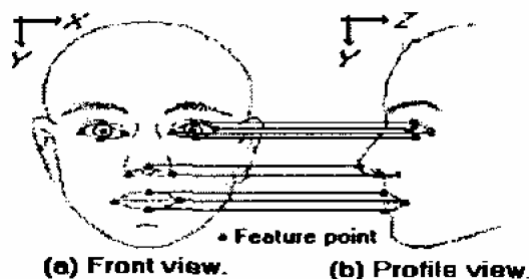


Fig. 3 Front and profile view of a subject.

The algorithm has been applied to 26 subjects, giving a gallery of 26 models, each one characterized by a feature vector of 29 vertices coordinates. Then given the two views of a test face, the coordinates of the 29 feature points are

computed and compared with all the models in the gallery, in order to find the best match. It results that 25 people of the 26 subjects are classified correctly, with a recognition rate of 96.2%.

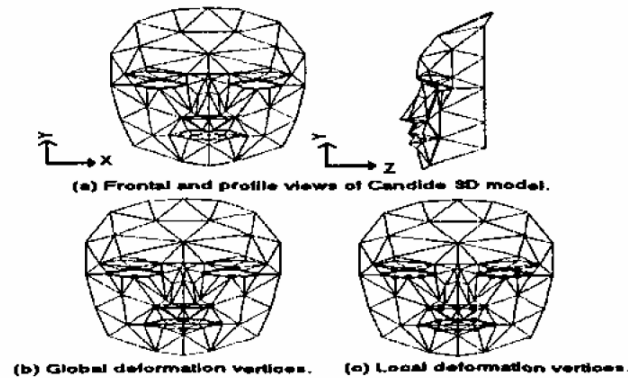


Fig. 4 Generic 3D model

2.1.6 Asymmetric 2D/3D Processing for Face Recognition [22]

The introduced method addresses the problem of pose and illumination variations. For that purpose, the authors propose an asymmetric 2D/3D scheme based on geometric invariants. Here, "asymmetric" means that enrolment is a 2D+3D-based process (geometric invariants are computed using both 2D and 3D data), while recognition is a 2D only process (a single probe image is used). The choice for such a paradigm is based on the observation that, although 3D acquisition is becoming cheaper, the problem of 3D sensitivity to environment conditions (such as illumination) in the acquisition phase still exists. Furthermore, acquiring a 3D model from someone needs his/her cooperation while this is not necessarily the case for acquiring an image. Thus this paradigm appears to be more realistic than a scheme fully based on 3D.

One of the problem underlined in this paper referring to the use of 3D geometric invariants is that previous works on the subject addressed the problem of rigid and clearly distinct object recognition. Obviously, faces don't belong to this class of 3D objects. Therefore, the 3D invariant has been chosen based on its robustness to noise regarding the position on the control points used in its computation. The one proposed by Weinshall [23] best meets this requirement: the invariant computation is performed on the 2D projection of the 3D model, but doesn't depend on the projection. The 19 control points were chosen as a subset of the feature points defined in MPEG-4.

When a new user has to be enrolled, the system acquires both the 3D shape and the 2D texture of his/her face. The control points are then manually located on the 2D texture and the corresponding 3D points are automatically retrieved on the 3D shape. Ratios of distances (2D-based invariants) are computed using the (x,y) coordinates of the control points. Subsequently, the 3D-based invariants described in [23] are computed based on the acquired 3D model. The results of both steps are concatenated into a single feature vector.

The recognition process is carried out as follows: let F be a query image submitted to the system. First, the 19 control points are manually located on F. The 2D invariants (distance ratios) are computed and used to retrieve the K -best subjects from the database. Then, the computation of the 3D invariant takes place and permits to identify the individual on the probe image.

Experiments were carried out on a proprietary database of 50 individuals. There were 3 models per individual. In the first experiment the goal was to investigate how much the discriminative power of the 2D invariants drops with respect to K . In the second experiment, 50 models have been considered to assess the performances of the system with respect to the accuracy in locating the control points. Noise was randomly generated and added to the coordinates of the control points.

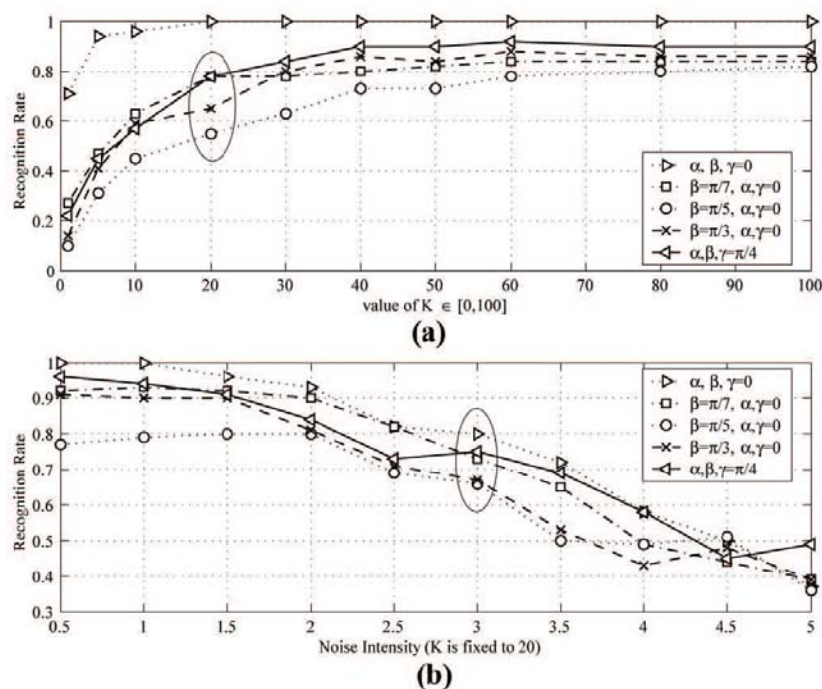


Fig. 5 (a) Recognition rate of the system, varying pose and the K parameter. (b) The probability that the correct subject is the first answer or is in the first 5 answers, when increasing random noise is added and $K = 20$.

Results show (cf. ellipses) that manual selection of the control points introduced an uncertainty equivalent to an additive noise of 3 pixels. Despite this manual selection step, recognition rates obtained are encouraging and call for experiments using an automatic feature points detection method.

2.1.7 Component-based Face Recognition using 3D Morphable Models

[10]

This system is an attempt to solve the problem of pose and illumination variations. The authors' solution uses a view-based approach to the problem, where a set of synthetic images with different pose and illumination conditions are artificially created for each enrolled person.

More precisely, at enrolment, the use of a morphable model allows the computation of a 3D model from face images using an analysis by synthesis method [11]. To create the 3D face model, only three images of a person's face are needed. Once the 3D face models of all the subjects in the training database are computed, arbitrary synthetic face

images under varying pose and illumination conditions are generated to train a component-based recognition system (the components of a face are characteristic parts of this face).

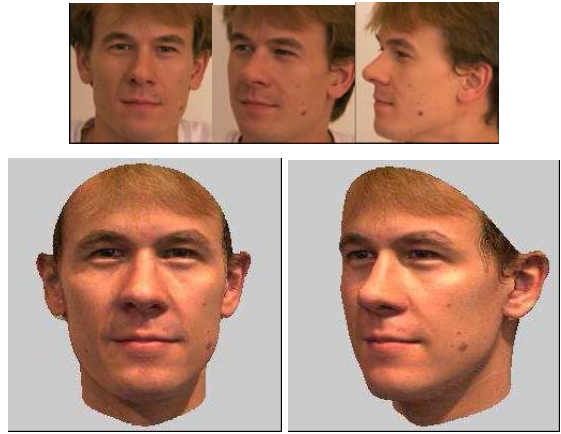


Fig. 6 Generation of the 3D model.

Before training the component-based recognition system, a component-based detector is used, which takes as input a 2D image of a face. It detects the face in the image and extracts the facial components. This part of the system is first used to extract the components of the synthetic face images generated from the 3D models. Histogram equalization is performed on the extracted components to later improve recognition accuracy. The grey pixels values of each component are then taken from the histogram-equalized image and combined into a single feature vector. Feature vectors were constructed for each enrolled person, and corresponding classifiers were trained.



Fig. 7 Examples of the components extracted from a frontal view and half-profile view of a face

The face recognition system consisting of second-degree polynomial SVM (Support Vector Machine) classifiers is trained on these feature vectors in a one vs. all approach (the SVM is trained to separate a subject from all the other subjects in the database).

At runtime, the component-based detector is reused to extract the components of the 2D probe image. The resulting components are then used as inputs to the SVM classifiers. To determine the identity of a person, the normalized

outputs of the SVM classifiers are compared. The identity associated with the face classifier with the highest normalized output is taken as the identity of the face.

A test set was created by taking images of six people. They were asked to rotate their faces and the lighting conditions were changed by moving a light source around the subject. The test set consisted of 200 images of each person with varying pose and illumination conditions. The component-based face recognizer was compared to a global face recognizer; both systems were trained and tested on the same images. ROC curves are given for each system.

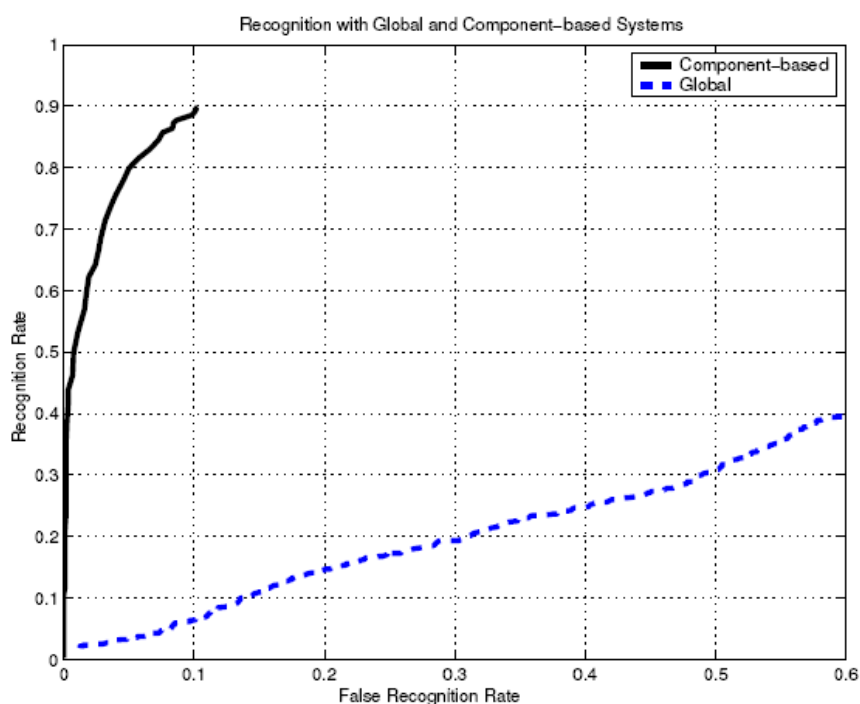



Fig. 8 ROC curves for the component-based and the global face recognition systems

This discrepancy can be explained by the fact that the components of a face are less sensitive to rotation than the whole face pattern. Furthermore, the backgrounds in the test images were non-uniform. The global system occasionally contained distracting background parts.

2.1.8 Adaptive Rigid Multi-region Selection (ARMS) [12]

With this system, the authors address the problem of face recognition under varying expressions between probe and gallery images. They compare their approach with the most commonly used approaches in 3D face recognition, namely PCA (eigenfaces) algorithm and ICP (Iterative Closest Point) algorithm. Their experimental results show that the performance of either approach degrades substantially in the case of expression variation between gallery and probe. Indeed, based on a big dataset of 449 persons over 4,000 3D images, they report an average rank-one rate of 91.0% for PCA baseline and 77.7% for ICP baseline when matching neutral probe images to the neutral gallery images. When matching non-neutral probe images to neutral gallery images, the results report average rank-one rates for ICP baseline and PCA baseline of 61.5% and 61.3% respectively.

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

Their approach (ARMS: Adaptive Rigid Multi-region Selection) is based on finding a relatively rigid region in the high curvature area on a face (such as the nose). This rigid region is then matched to the same region in the gallery and similarity is measured using RMS (Root Mean Square) error reported by ICP. As several of those high curvature areas can be used, voting or fusion rules can be considered to determine identity during the decision process.

The detection of the regions of interest (high curvature) of a face is realized through the following steps: first, a group of skin region is located by a skin detection method using the corresponding 2D color image (both 2D images and 3D scans are available in the dataset). Pixels are used in the skin detection method only if they have a valid corresponding 3D point. Valid 3D points found in regions detected by the skin detection are subject to 3D geometrical feature computation to classify an observed facial surface. Gaussian curvature and mean curvature are computed and geometrical shape can be identified by surface classification (see Fig. 8).

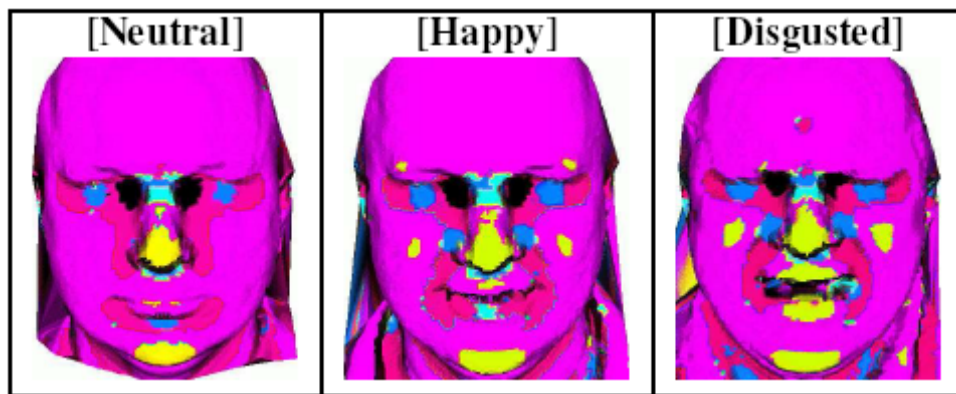


Fig. 9 Images of the same person with different expressions rendered based on surface types. The surface type of a region of the face may change as deformation is introduced. As cheeks are lifted, shown in *happy* and *disgusted*, peaks are detected at the upper cheeks, in both sides or in lips.

Once 3D surface classification is complete, the following regions are detected: nose tip (peak region), eyes cavities (pit region) and nose bridge (saddle region). The last step involves surface registration to measure the similarity of shape between a gallery and a probe surface. As explained before, the similarity score is based on the RMS error reported by ICP: given a pair of surfaces to be matched, the initial registration is performed by translating the centroid of the probe surface to the centroid of the gallery surface. Iterative alignment based on point difference between two surfaces is performed. At the end of each iteration, the RMS difference between the two surfaces is computed. The iteration ends when there is little or no change.

Experiments of the authors have shown that the use of two different regions around the nose yielded the best results, and thus are used in the final results. Regarding the fusion, after experiments, the product rule (which takes the product of the RMS error values) have been selected for giving the best performances.

The same dataset as the one used for ICP baseline and PCA baseline performance evaluation was used in the reported results.

In order to isolate the performances of the automatic feature-finding in the ARMS algorithm (curvature-based nose detection), performances for an ARMS-manual version of the algorithm are shown. In this version, manually identified

landmark points are used to extract the regions for matching, and so, recognition errors resulting from errors in the segmentation of the 3D are eliminated.

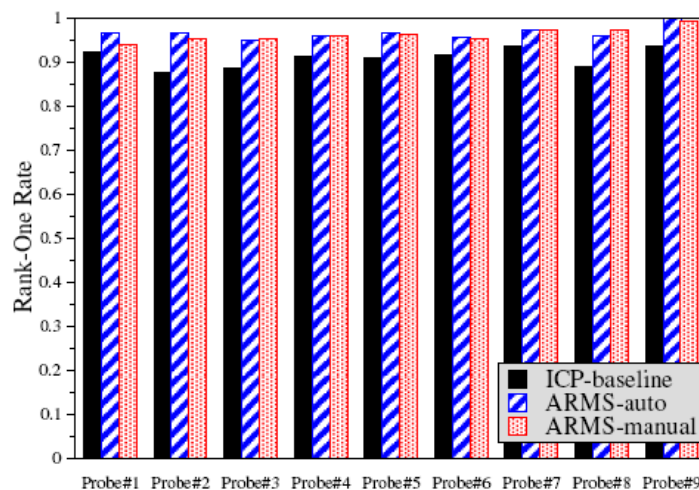


Fig. 10 Probes with neutral expression

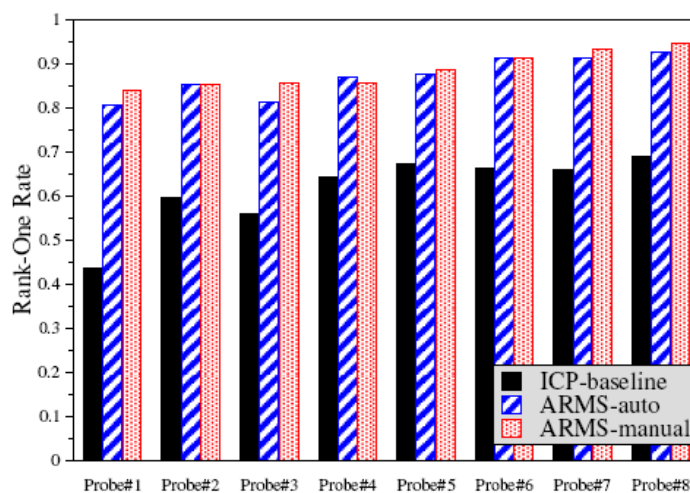


Fig. 11 Probes with Non-Neutral Expression

One surprising element of this work is that such good performances can be achieved using only a small portion of whole face by the ARMS method.

2.1.9 The Partial ICP (Iterative Closest Point) Algorithm [13]

This is another system attempting to address the problem of facial expressions in face recognition. The partial ICP method is designed to implicitly and dynamically extract the rigid parts of facial surfaces. The authors, acknowledging the fact that facial surface is a non-rigid object, assume that there are regions of the face that will keep their shape and position among different expressions. According to them, if these regions can be identified, the 3D non-rigid face recognition can be reduced to the rigid case. Their method is designed with this objective in mind. Its performances are compared to the ones of PCA baseline.

Let's consider the algorithm in more details. Given two surfaces, at each iteration, the closest points from the first surface to the second are searched. According to the result of this step, a transformation of the first surface is computed

in order to reduce the distance (more precisely the Root Mean Square (RMS) error) between the two sets of points. This loop is iterated until the RMS error stops changing or the change is below a threshold. While the traditional ICP-based method uses all point pairs to compute the transformation and the RMS error, the authors select only a part of the points pairs. After sorting the RMS errors of pairs of points in increasing order, they reject the worst $n\%$ of pairs (the pairs with the biggest errors). Discarding $n\%$ of pairs means removing those points in non-rigid region of facial surface. $(1-n\%)$ is called the *p-rate*.

Figure 11 shows some deformation images in which the red areas indicate the regions removed by setting different *p*-rates. When *p-rate* equals 0.7, most non-rigid parts are discarded.

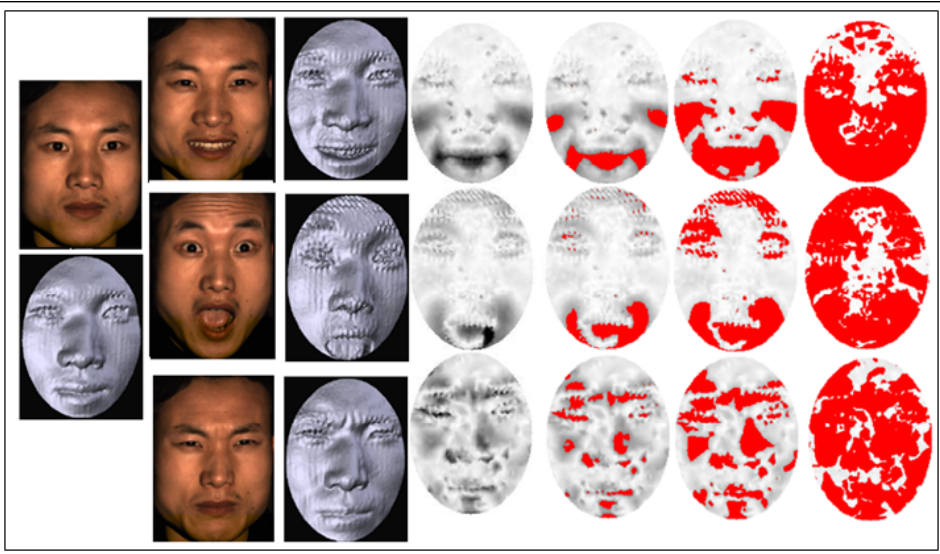


Fig. 12 Discarded area in facial surface with different *p-rate*=0.9, 0.7, 0.2 (5th, 6th, 7th columns respectively). Regions in red indicate the removed parts.

To carry out the experiments, the ZJU-3DFED 3D facial database, collected by the authors, is used. This database consists in 360 models with 40 subjects, and 9 scans with four different kinds of expression for each subject. Here, the neutral expression face models are used as gallery images, and the other 320 scans are classified into 4 probe sets: *Smile, Surprise, Sad, and Neutral*.

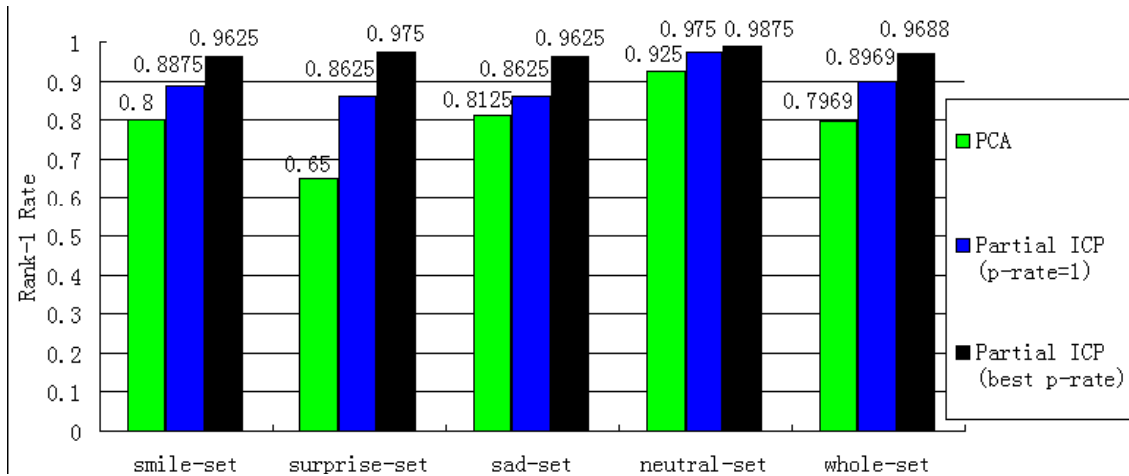


Fig 13 Rank-one rate: PCA vs. partial ICP

The partial ICP method outperforms PCA-based method on rank-one performance among all probe sets: PCA-based method get average rank-one rate of 75.41%. Partial ICP with p-rate=1 (which is equivalent to traditional ICP) gets a rank-one recognition rate of 89.69%. Partial ICP with best p-rate obtains an average rank-one of 96.88% on all probe sets.

2.1.10 3D Face Recognition Using Point Signature [14]

This 3D face recognition scheme also addresses the problem of face recognition with expression variations, and is based on the same assumption as the scheme in 2.1.8: the face is a non-rigid surface but there are some regions which don't vary (too much) regarding shape and position under expression changes. The authors propose point signature to carry out the recognition process based on the extraction of the rigid parts of the face.

The definition of Point Signature is summarized here, for details, the reader may refer to [21].

For a given point p , we place a sphere of radius r , centered at p . The intersection of the sphere with the object surface is a 3D space curve C , whose orientation can be defined by an orthonormal basis formed by a normal vector, \mathbf{n}_1 , a "reference" vector \mathbf{n}_2 , and the vector cross-product of \mathbf{n}_1 and \mathbf{n}_2 . A new plane P' is defined by translating the plane fitted through the space curve C (and used to define the vector \mathbf{n}_1) in a direction parallel to \mathbf{n}_1 . The perpendicular projection of C to P' forms a new planar curve C' forming a signed distance profile. Every point on C may now be characterized by:

1. The signed distance from itself to the corresponding point in C' .
2. A clockwise rotation angle θ about \mathbf{n}_1 from the reference direction \mathbf{n}_2 .

We refer to this profile of distances, $d(\theta)$, with $0 \leq \theta \leq 360$ degrees, as the signature at point p . In practice, the distance profile is represented by a discrete set of values $d(\theta_i)$, $0 \leq \theta_i \leq 360$.

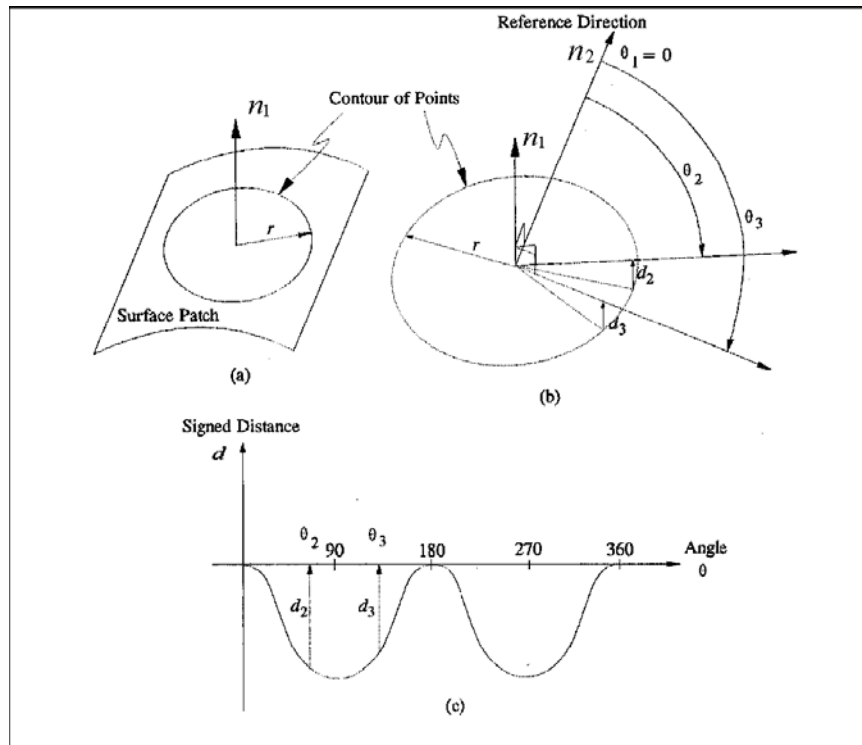


Fig 14 Definition of point signature

The similarity matching between a given signature $d_s(\theta_i)$ and a candidate signature $d_m(\theta_i)$ is as follows: if $d_s(\theta_i)$ has a clear global maximum, the matching is computed as:

$$|d_s(\theta_i) - d_m(\theta_i)| \leq \epsilon_{tol}(\theta_i), \text{ for all } i=1, \dots, n_\theta$$

where the tolerance band $\epsilon_{tol}(\theta_i)$ is used to handle the errors in computing Point Signature and achieve a better acceptance and rejection of candidate signatures.

If $d_s(\theta_i)$ has several similar local maxima, as we don't know from which one to start the matching (from which θ_i), we have to phase shift $d_s(\theta_i)$ for each position of local maximum before matching computation is carried out at this position.

To find the rigid parts of the face surfaces, an adaptive threshold is computed to distinguish between rigid and non-rigid parts, taking into account the mean and variance of the distances (as written above) between the points of the two surfaces. Points with distances above the threshold will be discarded and won't be used for identification or verification. To select the most likely model and verify, the probe image is pre-processed by computing the point signature of each range point. Every pre-processed point is used for a "vote": given a point and its signature, points with a similar signature are searched among the points of the models of the enrolled persons. Each person who have a similar signature receives a vote. The models are then ranked according to the number of votes they received. Then, verification can be carried out starting with the most likely candidate.

Experiments were carried out using a database of 6 persons. Four face images with different expressions were captured for each person.

	voting rate of each model					
	Scene Face 1	Scene Face 2	Scene Face 3	Scene Face 4	Scene Face 5	Scene Face 6
Model 1	82.58	60.61	36.36	46.21	34.09	28.79
Model 2	57.02	83.97	23.97	48.76	37.19	28.93
Model 3	73.44	70.94	93.75	31.25	56.25	46.88
Model 4	40.00	63.57	20.71	80.00	41.43	33.57
Model 5	60.14	76.08	47.55	62.94	90.21	49.65
Model 6	76.92	60.19	42.31	70.19	71.15	78.85

Table 8 Voting rate of each model using the different point signatures of scene (face 1 to face 6)

2.2 The most interesting 2D+3D multi-modal systems

Multi-Modal systems have the aim to improve the recognition rate of the existing 2D approaches, combining results obtained from both a 3D and a 2D recognition system. Hereafter are listed some of the most interesting 2D+3D multi-modal systems.

2.2.1 Features Extraction using Gabor Filters for 2D and Point Signature for 3D [9]

Wang et al: the method apply on both range data and texture. In the 3D domain, the Point Signature is used in order to describe the shape of the face, while the Gabor filters are applied on the texture in order to localize and characterize ten control points (corners of the eyes, nose tip, corner of the mouth, ...). The PCA analysis is then applied separately to the obtained 3D and 2D feature vectors, and then the resulting vectors are integrated to form an augmented vector which is used to represent each facial images. For a given test facial image, the best match in the gallery is identified according to similarity function or SVM (Support Vector Machine). The experiments are conducted on a database of 50 subjects, with different poses.

2.2.2 PCA-based Multi-modal 2D+3D Face Recognition [4]

Chang et al: this work is a report on PCA-based recognition experiments performed using shape and texture data from 200 persons. A total of 278 subject have been acquired with a Minolta Vivid 900 range scanner. The probe set consists of 166 individuals, while the training set contains 278 subjects, including those whose images are used for testing. The scanned images are normalized, in order to fill holes and remove picks. The 3D and 2D data are treated separately. Different fusion techniques and different distances (Euclidian and Mahalanobis) are tested. The experiments shown in Fig. 15. confirm that the Multi-modal overcomes both the 2D and 3D, when considered independently.

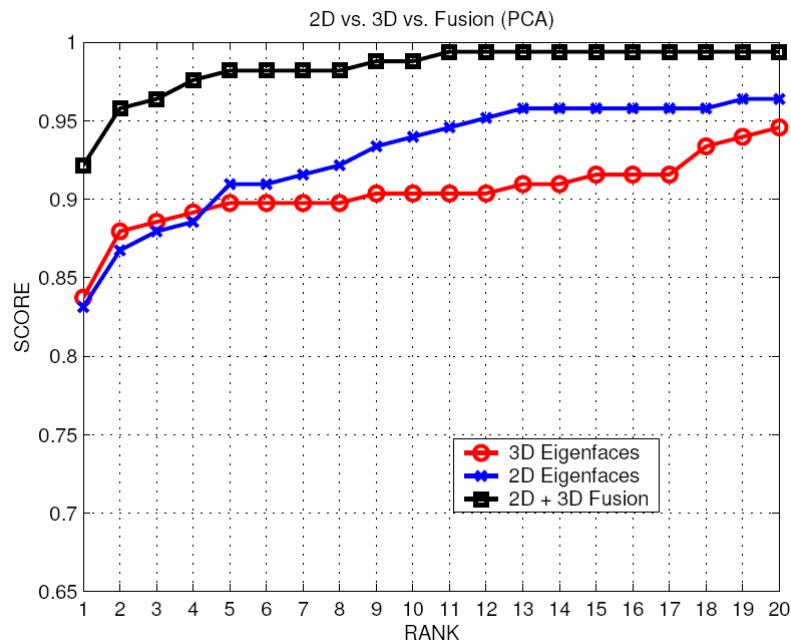


Fig. 15 Single- versus multi-modal biometrics.

2.2.3 3D+2D Fusion at Both Feature and Decision Levels Using Local Binary Patterns [15]

In most papers about 2D+3D, the fusion of the two modalities has been done at the decision level. This paper presents a framework for fusing 2D and 3D face recognition at both feature and decision levels. The authors choose to use Local Binary Patterns (LBP), which was originally proposed as a descriptor for textures.

Feature extraction is carried out as follows: first, face images are preprocessed so that they are aligned in a predefined way. LBP features are then extracted from the cropped and preprocessed images. The basic form of an LBP operator labels the pixels of an image by thresholding the 3x3 neighborhood of each pixel with the center value and considering the result as a binary number:

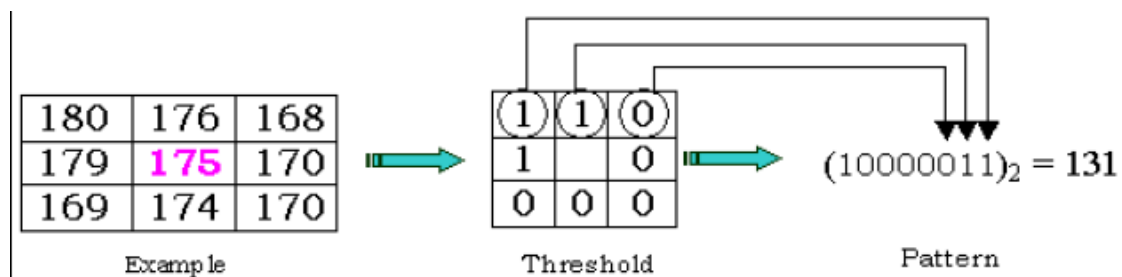


Fig. 16 Calculation of LBP code from 3x3 subwindow

The LBP can be extended to use neighborhood of different sizes. An LBP is called uniform if it contains at most two bitwise 0-1 or 1-0 transitions. There are 58 uniform LBP code patterns for 8-bits LBP code, and 256-58=198 non-uniform LBP patterns. Denoting all the non-uniform LBP patterns with a single bin, then there are a set of $L+1 = 59$ possible LBP code types for the 8-bit LBP code. Let's denote this set by $\mathcal{L}=\{0,1,\dots,L\}$ such that $LBP_{(x,y)}$ is in \mathcal{L} , and the

local LBP histogram over a block $S_{(x,y)}$ centered at (x,y) by $H_{(x,y)} = (H_{(x,y)}(0), H_{(x,y)}(1), \dots, H_{(x,y)}(L))$. The histogram can be defined as:

$$H_{(x,y)}(\ell) = \sum_{(x',y') \in S(x,y)} I\{LBP(x',y') = \ell\}, \quad \ell \in \mathcal{L}$$

where $I(\cdot)$ is in $\{0,1\}$ and is an indication function of a boolean condition, and $S_{(x,y)}$ is a local region centered at (x,y) . The histogram $H_{(x,y)}$ contains information about the distribution of the local micropatterns, such as edges, spots and flat areas over the block $S_{(x,y)}$. Individual LBP labels contain information about the patterns at the pixel-level, whereas the frequencies of the labels in the histogram produce information on the regional level. The collection of the histograms at all possible pixels $\{H_{(x,y)}, \text{ for all } (x,y)\}$ provides the global level description.

Separate learning for 3D and 2D face recognition is carried out as follows: every blocks centered at each pixel position are considered. Each block is given a weight. The weights are derived using an AdaBoost learning method (see [16] for more details). The learning also produces the final classifier. To dispense the need for a training process for faces of a newly added person, a large training set describing intra-personal or extra-personal variations is used and a universal two-class classifier is trained. Thus classifiers training is carried out using intra and inter-class histogram differences. Classification is carried out on the difference between the probe histogram and the gallery histogram.

To fuse 2D and 3D at feature level, the same AdaBoost learning procedure is used: it selects the most effective features from the complete 2D+3D difference feature set.

A large 2D+3D database is created using a Minolta 3D digitizer. The images are taken with varying pose, expression and lighting changes. The database contains 2305 images.

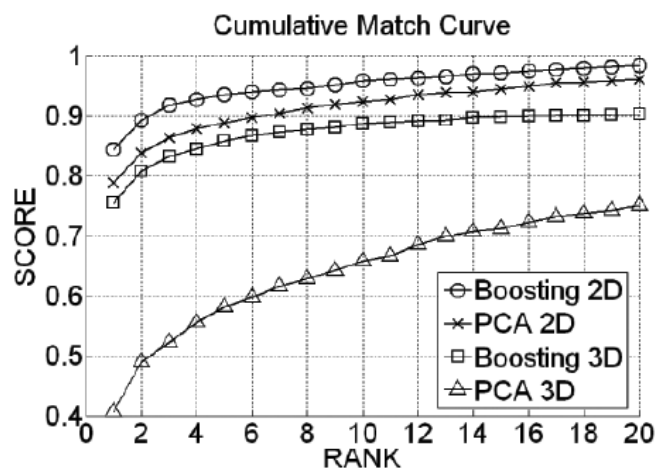


Fig. 17 Cumulative Match Curves for 3D and 2D

To contrast with the proposed AdaBoost learning fusion scheme (feature-level fusion), two non-boosting fusion schemes are included: the first is the PCA-based 3D+2D score fusion (called "CBF"). The second uses a sum rule to fuse the two AdaBoost classification scores (decision-level fusion).

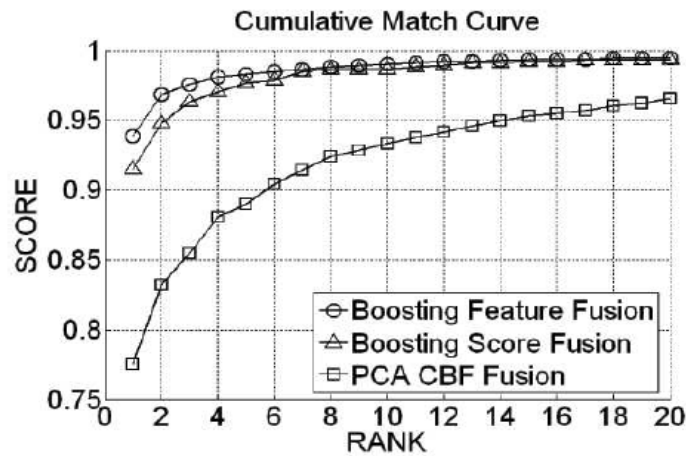


Fig. 18 Cumulative Match Curves for 3D+2D fusion

2.2.4 Hierarchical Matching using ICP for 3D and LDA for 2D [17]

The authors of this paper have developed a system combining 3D ICP registration and 2D LDA classification for more robustness against view, lighting and facial appearance variations. They propose a hierarchical matching scheme which first matches a probe 2.5D scan to 3D models and then uses the M best matched models to perform a LDA classification.

A 2.5D scan is a simplified 3D (x,y,z) surface representation that contains at most one depth value for every point in the (x,y) plane. Each scan can only provide a single viewpoint of the object, instead of the full 3D view.

The enrollment phase for the 3D part of the system consists in taking several 2.5D scans from different viewpoints, and construct the 3D model from these scans.

Surface matching is performed following a coarse-to-fine strategy. A feature based coarse alignment is employed, where three points are needed to compute the rigid transformation between the probe scan and the 3D model. In the current implementation these points are manually selected. The fine registration process follows the ICP framework. The root mean square distance minimized by the ICP algorithm is used as the primary matching distance of face scans. Then, linear discriminant analysis (LDA) is applied only to a small list of candidates, generated dynamically by the surface matching stage for each test scan. In the experiments, the top M=30 candidates are selected. The training samples used in LDA are derived from the 3D models (with their texture information). Indeed, from the registered (pose-normalized) 3D models are generated several synthetic 2D images with lighting variation and minor pose shifts. Two 2D-3D integration methods are actually tested: the hierarchical matching and the weighted sum rule.

For the weighted sum rule the equation is as follows:

$$MD_{\text{comb}} = MD_{\text{ICP}} + \alpha \cdot MD_{\text{LDA}}$$

Where $MD_{\text{LDA}} = (1 - MS_{\text{LDA}}) / 2$, MS_{LDA} is the matching score generated by the appearance-based matching components.

For the hierarchical matching, if MD_{ICP} is below a given threshold, then it is considered a good surface matching and no further step is applied. Else, the LDA algorithm is applied using the best registered models.

The database used for the experiments contains 100 subjects. 5 scans with neutral expression for each subject were captured to construct the 3D model. Another 6 scans are captured for testing with different expression (neutral and smiling).

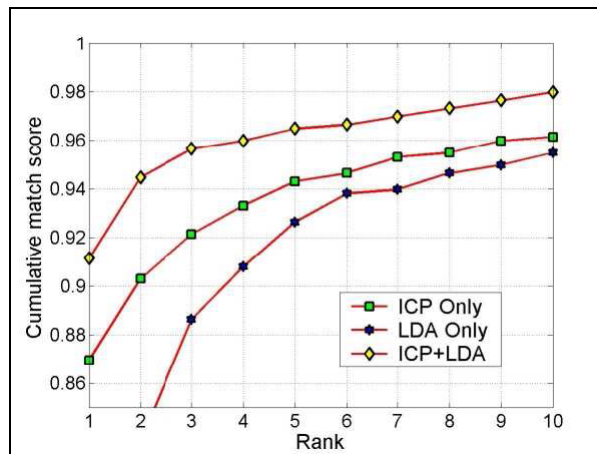


Fig. 19 Cumulative matching performance

Scheme	w/o hierarchical structure	w/ hierarchical structure
Surface matching	87%	88%
Surface matching + Appearance-based	91%	92%


Fig. 20 Rank-one matching accuracy ($\alpha = 1$) with and without hierarchical structure

2.2.5 Face Recognition from 2D and 3D Images using 3D Gabor Filters

[18]

Based on both 2D and 3D facial information, this paper focuses on extracting invariant features that can be used to recognize faces, with different facial expressions or extracted from varying viewpoints, from only one stored prototype face (frontal view with neutral expression) per person. The authors, based on the results in [19] which show that Gabor responses are robust against variations caused by changes in facial expression, head pose and lighting conditions, propose a modified version of the Gabor filters adapted to the 3D face recognition paradigm: 3D Spherical Gabor Filters (3D SGF). Indeed, traditional 3D Gabor filters are only robust against slight view variations. The 3D SGF is designed to cope with extensive view variations. To solve the missing point problem, caused by self-occlusion under large rotation angles, a 2D Gabor histogram, rather than the widely used integral operation is used in the computation of the distance between images.

Furthermore, the LTS-HD (Least Trimmed Square Hausdorff Distance) [20] is employed to compute the distance between the images resulting from the convolution of the original images with Gabor filters. Indeed, this distance is

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

known to perform well even if the object is occluded or degraded by noise or distortions. Given two finite points sets $A = \{a_1, \dots, a_p\}$ and $B = \{b_1, \dots, b_q\}$, the LTS-HD is defined as:

$$h_{LTS}(A, B) = \frac{1}{H} \sum_{i=1}^H D_B(a_i) \quad \text{With } D_B(a_i) = \min_{b_j \in B} \|a_i - b_j\|$$

Let's see in more details what is a 3D SGF. A 3D SGF is defined as:

$$g(\mathbf{x}, F) = \hat{g}(x, y, z) \exp\left(j2\pi F \sqrt{x^2 + y^2 + z^2}\right)$$

where F is the center frequency of the 3D SGF and

$$\hat{g}(x, y, z) = \frac{1}{(2\pi)^{3/2} \sigma^3} \exp\left[-\frac{(x^2 + y^2 + z^2)}{(2\sigma^2)}\right]$$

The 3D SGF is spherically symmetric. This rotation invariant characteristic makes the 3D SGF responses feasible for face recognition despite of different viewpoints. Given the image $I(x,y,z)$, the Gabor responses of the 3D SGF at point $\mathbf{x}=(x,y,z)$ are defined as:

$$J_n(\mathbf{x}) = \int I(\mathbf{x}') g(\mathbf{x} - \mathbf{x}', F_n) d\mathbf{x}'$$

by using different discrete center frequencies F_n .

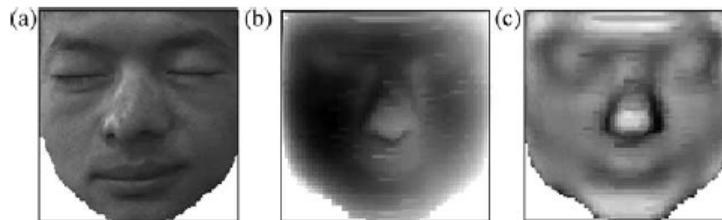


Fig. 21 A convolution example based on the 3D SGF. (a) Original image; (b) Amplitude response; (c) Phase response.

But the use of the integral is very sensitive to the missing points problem. To tackle it, a 2D histogram (less sensitive to that issue) is defined using:

$$\text{Re}(\mathbf{x}') = I(\mathbf{x}') \hat{g}(\mathbf{x} - \mathbf{x}') \cos(2\pi F_n \|\mathbf{x} - \mathbf{x}'\|)$$

$$\text{Im}(\mathbf{x}') = I(\mathbf{x}') \hat{g}(\mathbf{x} - \mathbf{x}') \sin(2\pi F_n \|\mathbf{x} - \mathbf{x}'\|)$$

where \mathbf{x}' is a point in the neighboring sphere of \mathbf{x} . Each of the 2 dimensions of the histogram correspond to a discretization of $\text{Re}(\mathbf{x}')$ and $\text{Im}(\mathbf{x}')$ respectively between their minimum and maximum values.

Not all the points of a given image are used to compute the 3D SGF response. In the experiments, feature points are sampled evenly and densely from each face. For classification, instead of involving all these points, only the matching pairs which have much smaller feature distances are employed: they are more robust to expressions and varying views.

The experiments are carried out with a face database involving 80 individuals. Each person provides 12 facial images, 6 frontal images with expression variations, and 6 non-frontal images with different view angles. The frontal view image of each individual with neutral expression is utilized to construct the model library.

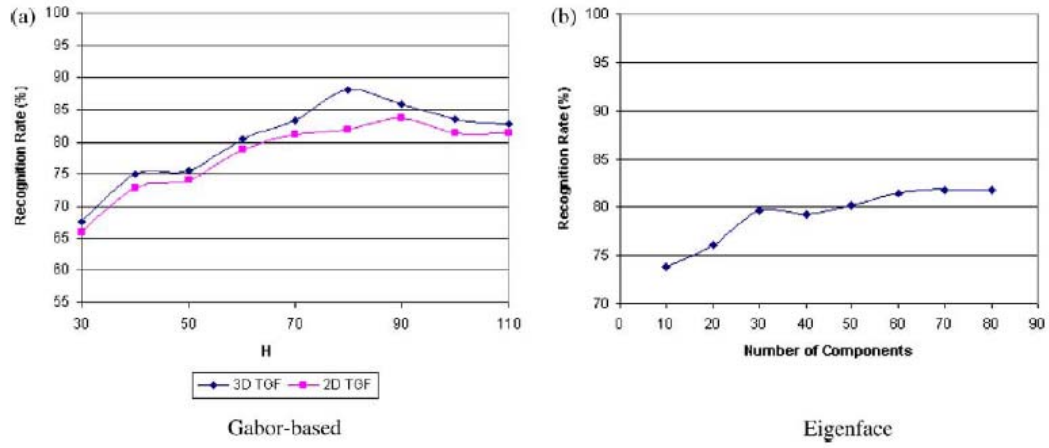


Fig. 22 Recognition rates of frontal test faces. (a) Gabor-based (b) Eigenface

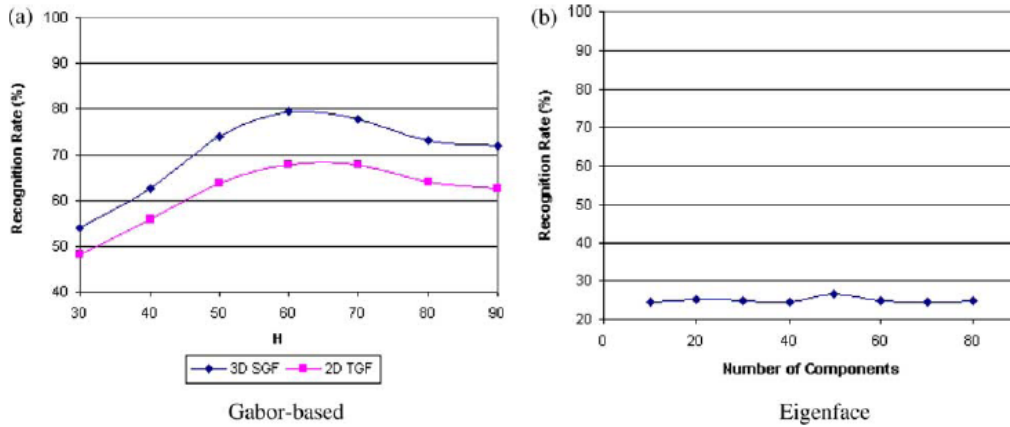



Fig. 23 Recognition rates of non-frontal test faces. (a) Gabor-based (b) Eigenface

The parameter H corresponds to the number of best matching pairs used to determine the LTS-HD directed distance $h_{LTS}(A,B)$.

3 Discussion

Here follows a summation table of the outline algorithms:

Face recognition method and year of publication of the paper	3D only	Multi-modal 2D+3D	Explicitly studied variabilities			Database
			Pose	Illumination	Expression	
Face Recognition Using Range Images (1997)	X		X			-Proprietary -240 range images
Face Recognition Based on Depth and Curvature Features (1992)	X		X			-Proprietary
A New Attempt To Face Recognition Using 3D Eigenfaces (2004)	X		X			-3D_RMA
Face Recognition Based on Fitting a 3D Morphable Model (2003)	X		X	X		-Proprietary (200 3D scans) +CMU-PIE +FERET
Face Modeling Using Two Orthogonal Views and a Generic Face Model (2003)	X		X			-Proprietary -26 individuals -1 model/ind.
Asymmetric 2D/3D Processing for Face Recognition (2005)	X		X	X		-Proprietary -50 individuals -3 models/ind.
Component-based Face Recognition Using 3D Morphable Models (2003)	X		X	X		-Proprietary -6 individuals -200 images/ind. - pose + illumination variations
Adaptive Rigid Multi-region Selection (2006)	X				X	-Proprietary -449 individuals -4000 3D images -expression variations
Partial ICP Algorithm (2006)	X				X	-ZJU-3DFED -40 individuals -9 scans/ind. -expression variations
3d Face Recognition Using Point Signature (2000)	X				X	-Proprietary -6 individuals -4 images/ind. -expression variations
Gabor Filters (2D) and Point Signature (3D) (2002)		X	X			-Proprietary -50 individuals -pose variations
PCA-based Multi-modal 2D+3D Face Recognition (2003)		X				-Proprietary -278 individuals
3D+2D Fusion at Both Feature and Decision Levels using Local Binary Patterns (2005)		X	X	X	X	-Proprietary -2305 images -pose + expression + illumination variations
Hierarchical Matching Using ICP for 3D and LDA for 2D (2005)		X	X	X	X	-Proprietary -100 individuals -11 scans/ind.

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

Face Recognition from 2D and 3D Images using 3D Gabor Filters (2005)		X	X		X	-Proprietary -80 individuals -24 images/ind. -pose + expression variations
--	--	---	---	--	---	---


Table 9 Summation of the outlined algorithms, in their order of appearance.

One limitation to some existing approaches to 3D face recognition involves sensitivity to size variation. Approaches that use a purely curvature based representation, such as extended Gaussian images, are not able to distinguish between two faces of similar shape but different size. On the contrary, approaches based on PCA or ICP (Iterative closest Point), avoid this problem but their performances throw down when changes in expression are present between gallery and probe images. Many recent papers acknowledge this issue, but only few of them explicitly include in the design of their algorithm a solution to the problem. Among these ones exist three main trends: one is the use of Gabor filters, which are known to be one of the most robust feature extraction scheme against pose, expression and illumination variations. The other important trend is to try to detect (dynamically or not) and discard face areas which vary “too much” with expression changes. The last one, the component-based method, prefer weighting the different components of the face (which can be for example rectangular regions) according to their capacity to accurately describe a face, including in presence of the adverse conditions already cited.


Many researchers believe that there are more possibilities to progress in the accuracy of face recognition using combined 2D + 3D scheme than any of them alone. In this category of work, we can observe that the majority of papers to date treats the combination of 2D and 3D as the fusion of the results (the scores) of two distinct and uncorrelated problems. Obviously, there is a need for more sophisticated combination. The most recent papers try to link both 2D and 3D paradigm at the feature extraction level, which seems a more effective way to improve the results: clearly, it is at least potentially more powerful to exploit possible synergies between the two modalities. Another interesting problem is the absence of an appropriate standard dataset, with a large number and demographic variety of people, which images have been taken at repeated intervals of time, with meaningful changes in expression, pose and lighting as it exists in 2D.

4 References

- [1] Achermann B., Jiang X., Bunke H., “*Face Recognition Using Range Images*”, Proc. Of the International Conference on the Virtual Systems and Multimedia (VSMM97), pp: 129-136, September 1997.
- [2] Ansari A., Abdel-Mottaleb M., “*3D Face Modeling Using Two Orthogonal Views and a Generic Face Model*”, Proc. Of the International Conference on Multimedia and Expo (ICME '03), Vol. 3, pp.: 289-292, July 2003.
- [3] Blanz V., Vetter T., “*Face Recognition Based on Fitting a 3D Morphable Model*”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp.: 1063-1074, September 2003.
- [4] Chang K. I., Bowyer K. W. and Flynn P. J., “*Multi-Modal 2D and 3D Biometrics for Face Recognition*”, Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), pp. 187–194, October 2003.

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

- [5] Gaile G. “*Face Recognition Based on Depth and Curvature Features*”, Proc. of IEEE Computer Society Conference on Computer Vision & Pattern recognition, pp: 808-810, June 1992.
- [6] Phillips P. J., Grother P., Michaels R. J., Blackburn D. M., Tabassi E. and Bone J., “*FRVT 2002: overview and summary*”. Available at www.frvt.org, March 2003.
- [7] Phillips P. J., Wechsler H., Huang J. and Rauss P., “*The FERET Database and Evaluation Procedure for Face Recognition Algorithms*”, Image and Vision Computing J., vol. 16, no. 5, pp.: 295-306, 1998.
- [8] Xu C., Wang Y., Tan T., Long Q., “*A new Attempt to Face Recognition Using 3D Eigenfaces*”, the 6th Asian Conference on Computer Vision (ACCV), Vol. 2, pp.884-889, 2004.
- [9] Wang Y., Chua C. and Ho Y., “*Facial Feature Detection and Face Recognition from 2D and 3D images*”, Pattern Recognition Letters, 23:1191-1202, 2002.
- [10] Huang J., Heisele B., and Blanz V., “*Component-based Face Recognition with 3D Morphable Models*”, 4th Conference on Audio- and Video-Based Biometric Person Authentication, 2003.
- [11] Blanz V. and Vetter T., “*A morphable model for synthesis of 3D faces*”, Computer Graphics Proceedings SIGGRAPH, pages 187-194, Los Angeles, 1999.
- [12] Chang K. I., Bowyer K. W., and Flynn P. J., “*Adaptive rigid multi-region selection for handling expression variation in 3D face recognition*”, to appear in IEEE Workshop on Face Recognition Grand Challenge Experiments.
- [13] Wang Y., Pan G., Wu Z., Wang Y., “*Exploring Facial Expression Effects in 3D Face Recognition using Partial ICP*”, The 7th Asian Conference on Computer Vision (ACCV'06), Lecture Notes in Computer Science, vol. 3851, pp.581-590, Hyderabad, India, January 13-16, 2006.
- [14] Chua C. S., Han F., and Ho Y. K., “*3d human face recognition using point signature*,” in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, March 2000, pp. 233--238.
- [15] Li S. Z., Zhao C., Zhu X., Lei Z., “*3D+2D Face Recognition by Fusion at Both Feature and Decision Levels*”, In Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures. Beijing. Oct 16, 2005.
- [16] Schapire R., “*A brief introduction to boosting*”, In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence. (1999)
- [17] Lu X. and Jain A. K., “*Integrating Range and Texture Information for 3D Face Recognition*”, Proc. 7th IEEE Workshop on Applications of Computer Vision (WACV'05), pp. 156-163, Breckenridge, CO, 2005
- [18] Wang Y. and Chua C.-S., “*Face recognition from 2D and 3D images using 3D Gabor filters*”, In Image and Vision Computing, Volume 23, Issue 11, 1 October 2005, Pages 1018-1028
- [19] Wiskott L., Fellous J., Kruger N., and von der Malsburg C., “*Face recognition by elastic bunch graph matching*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):775-779, 1997.
- [20] Sim D.G., Kwon O.K., Park H., “*Object matching algorithm using robust hausdorff distance measures*”, IEEE Transactions on Image Process 8 (3) (1999) 425–429.
- [21] Chua C. S. and Jarvis R., “*Point signature: A new representation for 3d object recognition*”, International Journal on Computer Vision, 25(1):63–85, 1997.

State of the Art in 3D Face Recognition  RR 06 160	Draft - Version: 2.0 (revised) Date: 10 April 2006
Authors: Remy Etheve , Daniel Riccio, Jean-Luc Dugelay	

- [22] Riccio D. and Dugelay J.-L., "*Asymmetric 3D/2D processing: a novel approach for face recognition*", ICIAP 2005, 13th International Conference on Image Analysis and Processing, September 6-8, 2005, Cagliari, Italy- Also published as Lecture Notes in Computer Science, Volume 3617
- [23] Weinshall D., "*Model-based invariants for 3D Vision*", In International Journal of Computer Vision, vol. 10, no. 1, pp. 27–42, 1993.