YCVIU 1197
21 September 2005 Disk Used
ARTICLE IN PRESS
No. of Pages 15, 4C : 4 Model 5+
Vimala (CE) / Anitha (TE)

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computer Vision and Image Understanding xxx (2005) xxx–xxx

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition

Kevin W. Bowyer *, Kyong Chang, Patrick Flynn

Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

Received 27 August 2004; accepted 13 May 2005

Abstract

This survey focuses on recognition performed by matching models of the three-dimensional shape of the face, either alone or in combination with matching corresponding two-dimensional intensity images. Research trends to date are summarized, and challenges confronting the development of more accurate three-dimensional face recognition are identified. These challenges include the need for better sensors, improved recognition algorithms, and more rigorous experimental methodology.

Keywords: Biometrics; Face recognition; Three-dimensional face recognition; Range image; Multi-modal

1. Introduction

Evaluations such as the Face Recognition Vendor Test (FRVT) 2002 [46] make it clear that the current state of the art in face recognition is not yet sufficient for the more demanding applications. However, biometric technologies that currently offer greater accuracy, such as fingerprint and iris, require much greater explicit cooperation from the user. For example, fingerprint requires that the subject cooperate in making physical contact with the sensor surface. This raises issues of how to keep the surface clean and germ-free in a high-throughput application. Iris imaging currently requires that the subject cooperate to carefully position their eye relative to the sensor. This can also cause problems in a high-throughput application. Thus there is significant potential application-driven demand for improved performance in face recognition. One goal of the Face Recognition Grand Challenge program [45] sponsored by various government agencies is to foster an order-of-magnitude increase in face recognition performance over that documented in FRVT 2002.

The vast majority of face recognition research and commercial face recognition systems use typical intensity images of the face. We refer to these as "2D images." In contrast, a "3D image" of the face is one that represents three-dimensional shape. A recent extensive survey of face recognition research is given in [60], but does not include research efforts based on matching 3D shape. Our survey given here focuses specifically on 3D face recognition. This is an update and expansion of earlier versions [8,9], to include the initial round of research results coming out of the Face Recognition Grand Challenge [16,23,33,41,44,50], as well as other recent results [42,28,29,20,32,31]. Scheenstra et al. [51] give an alternate survey of some of the earlier work in 3D face recognition.

We are particularly interested in 3D face recognition because it is commonly thought that the use of 3D sensing has the potential for greater recognition accuracy than 2D. For example, one paper states—"Because we are working in 3D, we overcome limitations due to viewpoint and lighting variations" [34]. Another paper describing a different approach to 3D face recognition states—"Range images have the advantage of capturing shape variation irrespective of illumination variabilities" [22]. Similarly, a third paper states—"Depth and curvature features have several advantages over more traditional intensity-based

* Corresponding author. Fax: +1 574 631 9260.
E-mail addresses: kwb@cse.nd.edu (K.W. Bowyer), kchang@cse.nd.edu (K. Chang), flynn@cse.nd.edu (P. Flynn).

1077-3142/$ - see front matter © 2005 Published by Elsevier Inc.
doi:10.1016/j.cviu.2005.05.005

features. Specifically, curvature descriptors: (1) have the potential for higher accuracy in describing surface-based events, (2) are better suited to describe properties of the face in a areas such as the cheeks, forehead, and chin, and (3) are viewpoint invariant'' [21].

2. Background concepts and terminology

The general term ''face recognition'' can refer to different application scenarios. One scenario is called ''recognition'' or ''identification,'' and another is called ''authentication'' or ''verification.'' In either scenario, face images of known persons are initially enrolled into the system. This set of persons is sometimes referred to as the ''gallery.'' Later images of these or other persons are used as ''probes'' to match against images in the gallery. In a recognition scenario, the matching is one-to-many, in the sense that a probe is matched against all of the gallery to find the best match above some threshold. In an authentication scenario, the matching is one-to-one, in the sense that the probe is matched against the gallery entry for a claimed identity, and the claimed identity is taken to be authenticated if the quality of match exceeds some threshold. The recognition scenario is more technically challenging than the authentication scenario. One reason is that in a recognition scenario a larger gallery tends to present more chances for incorrect recognition. Another reason is that the whole gallery must be searched in some manner on each recognition attempt.

While research results may be presented in the context of either recognition or authentication, the core 3D representation and matching issues are essentially the same. In fact, the raw matching scores underlying the *cumulative match characteristic* (CMC) curve for a recognition experiment can readily be tabulated in a different manner to produce the *receiver operating characteristic* (ROC) curve for an authentication experiment. The CMC curve summarizes the percent of a set of probes that is considered to be correctly matched as a function of the match rank that is counted as a correct match. The rank-one recognition rate is the most commonly stated single number from the CMC curve. The ROC curve summarizes the percent of a set of probes that is falsely rejected as a tradeoff against the percent that is falsely accepted. The equal-error rate (EER), the point where the false reject rate equals the false accept rate, is the most commonly stated single number from the ROC curve.

The 3D shape of the face is often sensed in combination with a 2D intensity image. In this case, the 2D image can be thought of as a ''texture map'' overlaid on the 3D shape. An example of a 2D intensity image and the corresponding 3D shape are shown in Fig. 1, with the 3D shape rendered in the form of a range image, a shaded 3D model and a mesh of points. A ''range image,'' also sometimes called a ''depth image,'' is an image in which the pixel value reflects the distance from the sensor to the imaged surface. In Fig. 1, the lighter values are closer to the sensor and the

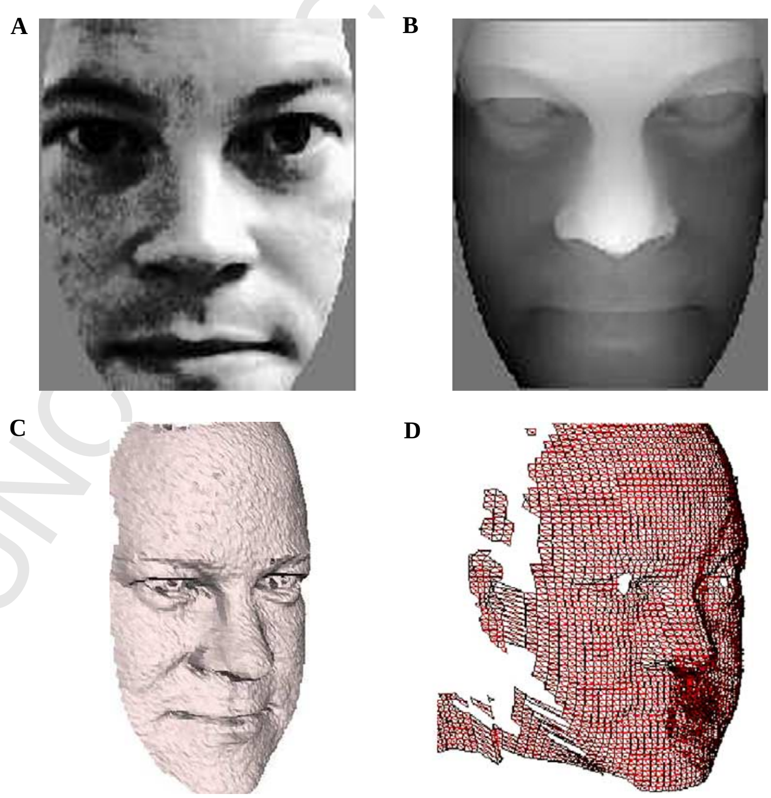

Fig. 1. Example of 2D intensity and 3D shape data. The 2D intensity image and the 3D range image are representations that would be used with ''eigenface'' style approaches. (A) Cropped 2D intensity image. (B) 3D rendered as range image. (C) 3D rendered as shaded model. (D) 3D rendered as wireframe.

darker values are farther away. A range image, a shaded model, and a wire-frame mesh are common alternatives for displaying 3D face data.

As commonly used, the term *multi-modal biometrics* refers to the use of multiple imaging modalities, such as 3D and 2D images of the face. The term "multi-modal" is perhaps imprecise here, because the two types of data may be acquired by the same imaging system. In this survey, we consider algorithms for multi-modal 3D and 2D face recognition as well as algorithms that use only 3D shape. We do **not** consider here the family of approaches in which a generic, "morphable" 3D face model is used as an intermediate step in matching two 2D images for face recognition. This approach was popularized by Blanz and Vetter [5], its potential was investigated in the FRVT 2002 report [46], and variations of this type of approach are already used in various commercial face recognition systems. However, this type of approach does not involve the sensing or matching of 3D shape descriptions. Rather, a 2D image is mapped onto a deformable 3D model, and the 3D model with texture is used to produce a set of synthetic 2D images for the matching process.

3. Recognition based solely on 3D shape

Table 1 gives a comparison of selected elements of algorithms that use only 3D shape to recognize faces. The works are listed chronologically by year of publication, and alphabetically by first author within a given year. The earliest work in this area was done over a decade ago [12,21,26,39]. There was relatively little work in this area through the 1990s, but activity has increased greatly in recent years.

Most papers report performance as the rank-one recognition rate, although some report equal-error rate or verification rate at a specified false accept rate. Historically, the experimental component of work in this area was rather modest. The number of persons represented in experimental data sets did not reach 100 until 2003. And only a few works have dealt with data sets that explicitly incorporate pose and/or expression variation [38,30,44,16,11]. It is therefore perhaps not surprising that most of the early works reported rank-one recognition rates of 100%. However, the Face Recognition Grand Challenge program [45] has already resulted in several research groups publishing results on a common data set representing over 4000 images of over 400 persons, with substantial variation in facial expression. Examples of the different facial expressions present in the FRGC version two dataset are shown in Fig. 2. As experimental data sets have become larger and more challenging, algorithms have become more sophisticated even if the reported recognition rates are not as high as in some earlier works.

Table 1
Recognition algorithms using 3D shape alone

Author, year, reference	Persons in dataset	Images in dataset	Image size	3D face data	Core matching algorithm	Reported performance
Cartoux, 1989 [12]	5	18	?	Profile, surface	Minimum distance	100%
Lee, 1990 [26]	6	6	256×150	EGI	Correlation	None
Gordon, 1992 [21]	26 train 8 test	26 train 24 test	?	Feature vector	Closest vector	100%
Nagamine, 1992 [39]	16	160	256×240	Multiple profiles	Closest vector	100%
Achermann, 1997 [3]	24	240	75×150	Range image	PCA, HMM	100%
Tanaka, 1998 [52]	37	37	256×256	EGI	Correlation	100%
Achermann, 2000 [2]	24	240	75×150	Point set	Hausdorff distance	100%
Chua, 2000 [17]	6	24	?	Point set	Point signature	100%
Hesher, 2003 [22]	37	222	242×347	Range image	PCA	97%
Lee, 2003 [27]	35	70	320×320	Feature vector	Closest vector	94% at rank 5
Medioni, 2003 [34]	100	700	?	Point set	ICP	98%
Moreno, 2003 [38]	60	420	2.2K points	Feature vector	Closest vector	78%
Pan, 2003 [42]	30	360	3K points	Point set, range image	Hausdorff and PCA	3–5% EER, 5–7% EER
Lee, 2004 [28]	42	84	240×320	Range, curvature	Weighted Hausdorff	98%
Lu, 2004 [30]	18	113	240×320	point set	ICP	96%
Russ, 2004 [49]	200 FRGC v1	468	480×640	Range image	Hausdorff distance	98% verification
Xu, 2004 [57]	120 (30)	720	?	Point set + feature vector	Minimum distance	96% on 30, 72% on 120
Bronstein, 2005 [11]	30	220	?	Point set	"canonical forms"	100%
Chang, 2005 [16]	466 FRGC v2	4007	480×640	Point set	multi-ICP	92%
Gökberk, 2005 [20]	106	579	?	Multiple	Multiple	99%
Lee, 2005 [29]	100	200	Various	Feature vector	SVM	96%
Lu, 2005 [31]	100	196 probes	240×320	Surface mesh	ICP, TPS	89%
Pan, 2005 [41]	276 FRGC v1	943	480×640	Range image	PCA	95%, 3% EER
Passalis, 2005 [44]	466 FRGC v2	4007	480×640	Surface mesh	Deformable model	90%
Russ, 2005 [50]	200 FRGC v1	398	480×640	Range image	Hausdorff distance	98.5%

Fig. 2. Example images in 2D and 3D with different expressions. The seven expressions depicted are: neutral, angry, happy, sad, surprised, disgusted, and "puffy."

Cartoux et al. [12] approach 3D face recognition by segmenting a range image based on principal curvature and finding a plane of bilateral symmetry through the face. This plane is used to normalize for pose. They consider methods of matching the profile from the plane of symmetry and of matching the face surface, and report 100% recognition for either in a small dataset.

Lee and Milios [26] segment convex regions in a range image based on the sign of the mean and Gaussian curvatures, and create an extended Gaussian image (EGI) for each convex region. A match between a region in a probe image and in a gallery image is done by correlating EGIs. The EGI describes the shape of an object by the distribution of surface normal over the object surface. A graph matching algorithm incorporating relational constraints is used to establish an overall match of probe image to gallery image. Convex regions are asserted to change shape less than other regions in response to changes in facial expression. This gives some ability to cope with changes in facial expression. However, EGIs are not sensitive to change in object size, and so two similar shape but different size faces will not be distinguishable in this representation.

Gordon [21] begins with a curvature-based segmentation of the face. Then a set of features are extracted that describe both curvature and metric size properties of the face. Thus each face becomes a point in feature space, and nearest-neighbor matching is done. Experiments are reported with a test set of three views of each of eight faces and recognition rates as high as 100% are reported. It is noted that the values of the features used are generally similar for different images of the same face, "except for the cases with large feature detection error, or variation due to expression" [21].

Nagamine et al. [39] approach 3D face recognition by finding five feature points, using those feature points to standardize face pose, and then matching various curves or profiles through the face data. Experiments are performed for 16 subjects, with 10 images per subject. The best recognition rates are found using vertical profile curves that pass through the central portion of the face. Computational requirements were apparently regarded as severe at the time this work was performed, as the authors note that "using the whole facial data may not be feasible considering the large computation and hardware capacity needed" [39].

Achermann et al. [3] extend eigenface and hidden Markov model (HMM) approaches used for 2D face recognition to work with range images. They present results for a dataset of 24 persons, with 10 images per person, and report 100% recognition using an adaptation of the 2D face recognition algorithms.

Tanaka et al. [52] also perform curvature-based segmentation and represent the face using an extended Gaussian image (EGI). Recognition is performed using a spherical correlation of the EGIs. Experiments are reported with a set of 37 images from a National Research Council of Canada range image dataset [48], and 100% recognition is reported.

Chua et al. [17] use "point signatures" in 3D face recognition. To deal with facial expression change, only the approximately rigid portion of the face from just below the nose up through the forehead is used in matching. Point signatures are used to locate reference points that are used to standardize the pose. Experiments are done with multiple images with different expressions from six subjects, and 100% recognition is reported.

Achermann and Bunke [2] report on a method of 3D face recognition that uses an extension of Hausdorff distance matching. They report on experiments using 240 range images, 10 images of each of 24 persons, and achieve 100% recognition for some instances of the algorithm.

Hesher et al. [22] explore principal component analysis (PCA) style approaches using different numbers of eigenvectors and image sizes. The image data set used has six different facial expressions for each of 37 subjects. The performance figures reported result from using multiple images per subject in the gallery. This effectively gives the probe image more chances to make a correct match, and is known to raise the recognition rate relative to having a single sample per subject in the gallery [36].

Medioni and Waupotitsch [34] perform 3D face recognition using an iterative closest point (ICP) approach to match face surfaces. Whereas most of the works covered here use 3D shapes acquired through a structured-light sensor, this work uses 3D shapes acquired by a passive stereo sensor. Experiments with seven images each from a set of 100 subjects are reported, with the seven images sampling different poses. An EER of "better than 2%" is reported.

Moreno and co-workers [38] approach 3D face recognition by first performing a segmentation based on Gaussian curvature and then creating a feature vector based on the segmented regions. They report results on a dataset of 420 face meshes representing 60 different persons, with some sampling of different expressions and poses for each person. Rank-one recognition of 78% is achieved on the subset of frontal views.

Lee et al. [27] perform 3D face recognition by locating the nose tip, and then forming a feature vector based on contours along the face at a sequence of depth values. They report 94% correct recognition at rank five, but do not report rank-one recognition. The recognition rate can change dramatically between ranks one and five, and so it is not possible to project how this approach would perform at rank one.

Pan et al. [42] experiment with 3D face recognition using both a Hausdorff distance approach and a PCA-based approach. In experiments with images from the M2VTS database [35] they report an equal-error rate (EER) in the range of 3–5% for the Hausdorff distance approach and an EER in the range of 5–7% for the PCA-based approach.

Lee and Shim [28] consider approaches to using a "depth-weighted Hausdorff distance" and surface curvature information (the minimum, maximum, and Gaussian curvature) for 3D face recognition. They present results of experiments with a data set representing 42 persons, with two images for each person. A rank-one recognition rate as high as 98% is reported for the best combination method investigated, whereas the plain Hausdorff distance achieved less than 90%.

Lu et al. [30] report on results of an ICP-based approach to 3D face recognition. This approach assumes that the gallery 3D image is a more complete face model and the probe 3D image is a frontal view that is likely a subset of the gallery image. In experiments with images from 18 persons, with multiple probe images per person, incorporating some variation in pose and expression, a recognition rate of 97% was achieved.

Russ et al. [49] present results of Hausdorff matching on range images. They use portions of the dataset used in [14] in their experiments. In a verification experiment, 200 persons were enrolled in the gallery, and the same 200 persons plus another 68 imposters were represented in the probe set. A probability of correct verification as high as 98% (of the 200) was achieved at a false alarm rate of 0 (of the 68). In a recognition experiment, 30 persons were enrolled in the gallery and the same 30 persons imaged at a later time were represented in the probe set. A 50% probability of recognition was achieved at a false alarm rate of 0. The recognition experiment uses a subset of the available data "because of the computational cost of the current algorithm" [49].

Xu et al. [57] developed a method for 3D face recognition and evaluated it using the database from Beumier and Acheroy [4]. The original 3D point cloud is converted to a regular mesh. The nose region is found and used as an anchor to find other local regions. A feature vector is computed from the data in the local regions of mouth, nose, left eye, and right eye. Feature space dimensionality is reduced using principal components analysis, and matching is based on minimum distance using both global and local shape components. Experimental results are reported for the full

120 persons in the dataset and for a subset of 30 persons, with performance of 72 and 96%, respectively. This illustrates the general point that reported experimental performance can be highly dependent on the dataset size. Most other works have not considered performance variation with dataset size. It should be mentioned that the reported performance was obtained with five images of a person used for enrollment in the gallery. Performance would generally be expected to be lower with only one image used to enroll a person.

Bronstein et al. [11] present an approach to 3D face recognition intended to allow for deformation related to facial expression. The idea is to convert the 3D face data to an "eigenform" that is invariant to the type of shape deformation that is modeled. In effect, there is an assumption that "the change of the geodesic distances due to facial expressions is insignificant." Experimental evaluation is done using a dataset containing 220 images of 30 persons (27 real persons and 3 mannequins), and 100% recognition is reported. A total of 65 enrollment images were used for the 30 subjects, so that a subject is represented by more than one image. As already mentioned, use of more than one enrollment image per person will generally increase recognition rates. The method is compared to a 2D eigenface approach on the same subjects, but the face space is trained using just 35 images and has just 23 dimensions. The method is also compared to a rigid surface matching approach. Perhaps the most unusual aspect of this work is the claim that the approach "can distinguish between identical twins."

Gökberk et al. [20] compare five approaches to 3D face recognition using a subset of the data used by Beumier and Acheroy [4]. They compare methods based on extended Gaussian images, ICP matching, range profile, PCA, and linear discriminant analysis (LDA). Their experimental dataset has 571 images from 106 people. They find that the ICP and LDA approaches offer the best performance, although performance is relatively similar among all approaches but PCA. They also explore methods of fusing the results of the five approaches and are able to achieve 99% rank-one recognition with a combination of recognizers. This work is relatively novel in comparing the performance of different 3D face recognition algorithms, and in documenting a performance increase by combining results of multiple algorithms. Additional work exploring these sorts of issues would seem to be valuable.

Lee et al. [29] propose an approach to 3D face recognition based on the curvature values at eight feature points on the face. Using a support vector machine for classification, the report a rank-one recognition rate of 96% for a data set representing 100 persons. They use a Cyberware sensor to acquire the enrollment images and a Genex sensor to acquire the probe images. The recognition results are called "simulation" results, apparently because the feature points are manually located.

Lu and Jain [31] extend previous work using an ICP-based recognition approach [30] to deal explicitly with variation in facial expression. The problem is approached as a rigid transformation of probe to gallery, done with ICP, along with a non-rigid deformation, done using thin-plate spline (TPS) techniques. The approach is evaluated using a 100-person dataset, with neutral-expression and smiling probes, matched to neutral-expression gallery images. The gallery entries are whole-head data structures, whereas the probes are frontal views. Most errors after the rigid transformation result from smiling probes, and these errors are reduced substantially after the non-rigid deformation stage. For the total 196 probes (98 neutral and 98 smiling), performance reaches 89% for shape-based matching and 91% for multi-modal $3D + 2D$ matching [32].

Russ et al. [50] developed an approach to using Hausdorff distance matching on the range image representation of the 3D face data. An iterative registration procedure similar to that in ICP is used to adjust the alignment of probe data to gallery data. Various means of reducing space and time complexity of the matching process are explored. Experimental results are presented on a part of the FRGC version 1 data set, using one probe per person rather than all available probes. Performance as high as 98.5% rank-one recognition, or 93.5% verification at a false accept rate of 0.1%, is achieved. In related work, Koudelka et al. [24] have developed a Hausdorff-based approach to prescreening a large dataset to select the most likely matches for more careful consideration [24].

Pan et al. [41] apply PCA, or eigenface, matching to a novel mapping of the 3D data to a range, or depth, image. Finding the nose tip to use as a center point, and an axis of symmetry to use for alignment, the face data are mapped to a circular range image. Experimental results are reported using the FRGC version 1 data set. The facial region used in the mapping contains approximately 12,500–110,000 points. Performance is reported as 95% rank-one recognition or 2.8% EER in a verification scenario. It is not clear whether the reported performance includes the approximately 1% of the images for which the mapping process fails.

Chang et al. [16] describe an "multi-region" approach to 3D face recognition. It is a type of classifier ensemble approach in which multiple overlapping subregions around the nose are independently matched using ICP, and the results of the multiple 3D matches fused. The experimental evaluation in this work uses essentially the FRGC version 2 data set, representing over 4000 images from over 400 persons. In an experiment in which one neutral-expression image is enrolled as the gallery for each person, and all subsequent images (of varied facial expressions) are used as probes, performance of 92% rank-one recognition is reported.

Passalis et al. [44] describe an approach to 3D face recognition that uses annotated deformable models. An average 3D face is computed on a statistical basis from a training set. Landmark points on the 3D face are selected based on descriptions by Farkas [18]. Experimental results are presented using the FRGC version 2 data set. For an identification experiment in which one image per person is enrolled in the gallery (466 total) and all later images (3541) are used as probes, performance reaches nearly 90% rank-one recognition.

4. Multi-modal algorithms using 3D and 2D data

While 3D face recognition research dates back to before 1990, algorithms that combine results from 3D and 2D data did not appear until about 2000. Most efforts to date in this area use relatively simplistic approaches to fusing results obtained independently from the 3D data and the 2D data. The single most common approach has been to use an eigenface type of approach on each of the 2D and 3D independently, and then combine the two matching scores. However, more recent works appear to take a variety of quite different approaches. Interestingly, several commercial face recognition companies already have capabilities for multi-modal 3D + 2D face recognition.

Lao et al. [25] perform 3D face recognition using a sparse depth map constructed from stereo images. Iso-luminance contours are used for the stereo matching. Both 2D edges and iso-luminance contours are used in finding the irises. In this specific limited sense, this approach is multi-modal. However, there is no separate recognition result from 2D face recognition. Using the iris locations, other feature points are found so that pose standardization can be done. Recognition is performed by the closest average difference in corresponding points after the data are transformed to a canonical pose. Recognition rates of 87–96% are reported using a dataset of 10 persons, with four images taken at each of nine poses for each person.

Beumier and Acheroy [4] approach multi-modal recognition by using a weighted sum of 3D and 2D similarity measures. They use a central profile and a lateral profile, each in both 3D and 2D. Therefore they have a total of four classifiers, and an overall decision is made using a weighted sum of the similarity metrics. A data set representing over 100 persons imaged on multiple sessions, with multiple poses per session, is acquired. Portions of this data set have been used by several other researchers [57,20]. In this paper, results are reported for experiments on a subset of the data, using a 27-person gallery and a 29-person probe set. An equal-error rate as low as 1.4% is reported for multi-modal 3D + 2D recognition that merges multiple probe images per subject. In general, multi-modal 3D + 2D is found to perform better than either 3D or 2D alone.

Wang et al. [56] use Gabor filter responses in 2D and "point signatures" in 3D to perform multi-modal face recognition. The 2D and 3D features together form a feature vector. Classification is done by support vector machines with a decision directed acyclic graph (DDAG). Experiments are performed with images from 50 subjects, six images per subject, with pose and expression variations. Recognition rates exceeding 90% are reported.

Bronstein et al. [10] use an isometric transformation approach to 3D face analysis in an attempt to better cope with variation due to facial expression. One method they propose is effectively multi-modal 3D + 2D recognition using eigen decomposition of flattened textures and canonical images. They show examples of correct and incorrect recognition by different algorithms, but do not report any overall quantitative performance results for any algorithm.

Tsalakanidou et al. [55] report on multi-modal face recognition using 3D and color images. The use of color rather than simply gray-scale intensity appears to be unique among the multi-modal work surveyed here. Results of experiments using images of 40 persons from the XM2VTS dataset [35] are reported for color images alone, 3D alone, and 3D + color. The recognition algorithm is PCA-style matching, followed by a combination of the results for the individual color planes and range image. Recognition rates as high as 99% are achieved for the multi-modal algorithm, and multi-modal performance is found to be higher than for either 3D or 2D alone.

Chang et al. [14] report on PCA-based recognition experiments performed using 3D and 2D images from 200 persons. One experiment uses a single set of later images for each person as the probes. Another experiment uses a larger set of 676 probes taken in multiple acquisitions over a longer elapsed time. Results in both experiments are approximately 99% rank-one recognition for multi-modal 3D + 2D, 94% for 3D alone, and 89% for 2D alone. The multi-modal result was obtained using a weighted sum of the distances from the individual 3D and 2D face spaces.

Godil et al. [19] present results of 3D + 2D face recognition using 200 persons worth of data taken from the CAESAR anthropometric database. They use PCA for matching both the 2D and the 3D, with the 3D represented as a range image. The 3D face data from this database may be rather coarse, with approximately 4000 points reported on the face. Multiple approaches to score-level fusion of the two results are explored. Performance as high as 82% rank-one recognition is reported.

Papatheodorou and Rueckert [43] perform multi-modal 3D + 2D face recognition using a generalization of ICP based on point distances in a 4D space (x, y, z, intensity). This approach integrates shape and texture information at an early stage, rather than making a decision using each mode independently and combining decisions. They present results from experiments with 62 subjects in the gallery, and probe sets of varying pose and facial expression from the images in the gallery. They report 98–100% correct recognition in matching frontal, neutral-expression probes to frontal neutral-expression gallery images. Recognition drops when the expression and pose of the probe images is not matched to those of the gallery images, for example to the range of 73–94% for 45° off-angle probes, and to the range of 69–89% for smiling expression probes.

Tsalakanidou and a different set of co-workers [54] report on an approach to multi-modal face recognition based on an embedded hidden Markov model for each modality. Their experimental data set represents a small number of different persons, but each has 12 images acquired in each of five different sessions. The 12 images represent varied pose and facial expression. Interestingly, they report a higher EER for 3D than for 2D in matching frontal neutral-expression probes to frontal neutral-expression gallery images, 19% versus 5%, respectively. They report that "depth data mainly suffers from pose variations and use of eyeglasses" [54]. This work is also unusual in that it is based on using five images to enroll a person in the gallery, and also generates additional synthetic images from those, so that a person is represented by a total of 25 gallery images. A longer version of this work appears in [53].

Hüsken et al. [23] describe the Viisage approach to multi-modal recognition. The 3D matching follows the style of hierarchical graph matching already used in Viisage's 2D face recognition technology. This is felt to allow greater speed of matching in comparison to techniques based on ICP or similar iterative techniques. Fusion of the results from the two modalities is done at the score level. Multi-modal performance on the FRGC version 2 data set is reported as 93% verification at 0.01 FAR. In addition, it is reported that performance of 2D alone is only slightly less than multi-modal performance, and that performance of 3D alone is substantially less than that of 2D alone. In this context, it may be interesting to note that results from a group (Geometrix) that originally focused on 3D face recognition show that 3D alone outperforms 2D alone, whereas results from a group (Viisage) that originally focused on 2D alone show that 2D alone outperforms 3D alone.

Lu et al. [32] build on earlier work with ICP style matching of 3D shape [30] to create a 3D + 2D multi-modal system. They use a linear discriminant analysis approach for the 2D matching component. Their experimental data set consists of multiple scans of each of 100 persons. Five scans with a Minolta Vivid 910 system are taken in order to create a 3D face model for enrolling a person. Enrollment is done with neutral expression. Six scans are taken of each person, three with neutral expression, and three with smiling expression, to use as individual probes for testing. They report better performance with 3D matching alone than with 2D matching alone. They also report 98% rank-one recognition for 3D + 2D recognition on neutral expressions alone, and 91% on the larger set of neutral and smiling expressions.

Maurer et al. [33] describe the Geometrix approach to multi-modal 3D + 2D face recognition. The 3D matching builds on the approach described by Medioni and Waupotitsch [34], whereas the 2D matching uses the approach of Neven Vision [40]. A weighted sum rule is used to fuse the two results, with the exception that "when the shape score is very high, we ignore the texture score" [33]. Experimental results are presented for the FRGC version two data set. The facial expression variations in this dataset are categorized into "neutral," "small," and "large" and results are presented separately for these three categories. Multi-modal performance for the "all versus all" matching of the 4007 images reaches approximately 87% verification at 0.01 FAR. They also report that 3D + 2D outperforms 3D alone by a noticeable increment, and that the verification rates for 2D alone are below those for 3D alone.

5. Trends in research directions

The recognition rates reported by the various works listed in Tables 1 and 2 should be interpreted with extreme caution. A number of factors combine to make direct comparisons problematic in most cases. Among these factors are different sizes of data set, different inherent levels of difficulty of the dataset, and different methods of experimental design. The results reported by Xu et al. [57] give a example of how dramatically the size of a dataset can affect reported performance. They found 96% rank-one recognition using a 30-person dataset, but this fell to 72% when using a 120-person dataset. Chang [16] documented a smaller decrease in performance with increasing size of

Table 2
Recognition algorithms combining use of 3D and 2D data

Author, year, reference	Persons in dataset	Images in dataset	Image size	3D face data	Core matching algorithm	Reported performance
Lao, 2000 [25]	10	360	480 × 640	Surface mesh	Minimum distance	91%
Beumier, 2001 [4]	27 gallery 29 probes	81 gallery, 87 probes	?	Multiple profiles	Minimum distance	1.4% EER
Wang, 2002 [56]	50	300	128 × 512	Feature vector	SVM, DDAG	>90%
Bronstein, 2003 [10]	157	?	2250 points	Range, point set	PCA, "eigen"	Not "eigen"
Chang, 2003 [14]	200 (275 train)	951	480 × 640	Range image	PCA	99% 3D + 2D, 93% 3D only
Tsalakanidou, 2003 [55]	40	80	100 × 80	Range image	PCA	99% 3D + 2D, 93% 3D only
Godil, 2004 [19]	200	400	128 × 128	Range image	PCA	82% rank 1
Papatheodorou, 2004 [43]	62	806	10,000 points	Point set	ICP	100–66%
Tsalakanidou, 2004 [54]	50	3000	571 × 752	Range image	EHHM per mode	4% EER
Hüsken, 2005 [23]	466	4,007 FRGC v.2	480 × 640	hier. graph	graph match	93% verification at 0.01 FAR
Lu, 2005 [32]	100	598	320 × 240	Point set	ICP, LDA	91%
Maurer, 2005 [33]	466	4007 FRGC v.2	480 × 640	Surface mesh	ICP, Neven	87% verification at 0.01 FAR

dataset, and found that the decrease was larger for the component of the dataset containing expression variation than it was for the component of the dataset with all neutral expressions. This points out that there is no simple rule of thumb to adjust reported performance for the size of dataset. The reported performance is also greatly dependent on the inherent difficulty of the data. The presence of expression variation is one element of increased difficulty, but pose variation, time lapse between gallery and probe, presence of eyeglasses, and other factors are also important. The design of the experiment also influences the reported performance. For example, we have noted that using more than one image of a person in the enrollment data generally increases performance. This type of enrollment can be done with essentially any approach. Comparing reported results between studies that differ in just this one element of methodology is problematic. The "biometric experimentation environment" associated with the Face Recognition Grand Challenge is a significant attempt to address these issues of comparable methodology and dataset [45].

One trend that can be noted concerns the variety and sophistication of algorithmic approaches explored. Rather than converging on some one or two standard algorithmic approaches, it appears that the variety and sophistication of algorithmic approaches explored is expanding. While the eigenface style of approach was popular initially, it seems less popular currently. ICP-style approaches also have been popular, and they appear to be evolving in potentially useful directions. For example, Papatheodorou and Rueckert [43] use a "4-D" version of ICP to fuse the intensity result with the 3D shape result. And Chang et al. [16] use a classifier ensemble type of approach to combining multiple ICP results. However, approaches that use ICP or Hausdorff distance are computationally demanding, and so one attractive line of research involves methods to speed up the 3D matching. For example, Russ et al. [50] have looked at a number of ways to speed up the computation of an earlier Hausdorff matching approach [49]. Also, Yan and Bowyer [59] have looked at trading off space of the enrollment data structure to speed up computation of ICP style matching in biometrics.

One clear trend is toward increasingly challenging experimental evaluation. Historically, much of the work in this area was evaluated using datasets representing a few tens of people, and the first studies to report results on datasets representing 100 or more persons appeared just in the last three years. But the field has moved quickly to reporting results on datasets consisting of thousands of images of hundreds of people. Also, a variety of approaches have been proposed to handle expression variation, and newer experimental data sets facilitate this line of research [45]. 3D face recognition is perhaps now entering an experimental phase similar to what 2D face recognition entered a decade ago with the FERET evaluations [47]. The days when reporting 100% recognition on a dataset of images involving less than 100 persons could be considered serious experimental evaluation are likely passed. It seems likely that the trend toward more challenging experimental results will continue in the near future, as researchers in 3D face recognition strive to develop more generally competent systems.

Several observations can be made with regard specifically to multi-modal 3D + 2D face recognition. All results that we are aware of show that multi-modal performs better than 3D alone or 2D alone. However, these comparisons generally do not control for the same number of image samples, and when this is done the apparent performance difference between 3D + 2D and 2D is greatly reduced. For example, Chang et al. [13] looked at this issue in the context of using an eigenface approach for each of 3D and 2D in a multi-modal recognition study. Using a single 2D image for enrollment and for recognition, the rank-one recognition rate was approximately 91%, and a single 3D image gave approximately 89%. Multi-modal 3D + 2D gave a recognition rate of approximately 95%. This seems to be a reasonable-sized increase in performance. However, it results from comparing the use of two image samples to represent a person to the use of one image sample. It is possible to use two different 2D images to represent a person for enrollment and for recognition. This results in performance of approximately 93%, implying that half the apparent gain in going to multi-modal recognition may be due simply to using two image samples to represent a person.

The literature appears split on whether using a single 3D example outperforms using a single 2D example. Some researchers have found that it does [14,33] and some researchers have found the opposite [54,23]. There is probably more feeling that 2D currently allows better recognition performance. However, even when it is acknowledged that 2D currently appears to offer better recognition performance, this is often thought to be a temporary situation—"Although 2D face recognition still seems to outperform the 3D face recognition methods, it is expected that this will change in the near future" [51].

6. Challenge for 3D face recognition: improved sensors

Current 3D sensing technologies used for face recognition fall into three basic categories. One category can be labeled passive stereo. The Geometrix system is one example of this approach [34]. In the passive stereo approach, two cameras with a known geometric relationship are used to image the subject, corresponding points are found in the two images, and the 3D location of the points can be computed. Another approach can be labeled pure structured light. The Minolta sensor used in [14,30] would be a straightforward example of this. This approach uses a camera and a light projector with a known geometric relationship. A light pattern is projected into the scene, detected in an image acquired by the camera, and the 3D location of points can then be computed. A third approach is best considered a hybrid of passive stereo and structured lighting. In such techniques, a pattern is projected onto the scene and then imaged by a stereo camera rig.

The projected pattern simplifies the selection of, and can improve the density of, corresponding points in the multiple images. The 3Q "Qlonerator" system is one example of this type of sensor [1].

Even under ideal illumination conditions for a given sensor, it is common for artifacts to occur in face regions such as oily regions that appear specular, the eyes, and regions of facial hair such as eyebrows, mustache, or beard. The most common types of artifacts can generally be described subjectively as "holes" or "spikes." A "hole" is essentially an area of missing data, resulting from the sensor being unable to acquire data. A "spike" is an outlier error in the data, resulting from, for example, an inter-reflection in a projected light pattern or a correspondence error in stereo. An example of "holes" in a 3D face image sensed with the Minolta sensor is shown in Fig. 3. Artifacts can and do occur with essentially all range sensors. They are typically patched up by interpolating new values based on the valid data nearest the artifact.

Another limitation of current 3D sensor technology, especially relative to use with non-cooperative subjects, is the depth of field for sensing data. The depth of field for acquiring usable data might range from about 0.3 m or less for a stereo-based system to about 1 m for a structured-light system such as the Minolta Vivid 900 [37]. Increased depth of field would lead to more flexible use in application.

Also, the image acquisition time for the 3D sensor should be short enough that subject motion is not a significant issue. Acquisition time is generally a more significant problem with structured-light systems than with stereo systems. It may be less of an issue for authentication type applications, in which the subjects can be assumed to be cooperative, than it is for recognition type applications.

6.1. The myth of "illumination invariance"

As noted earlier, it is often asserted that 3D is, or should be, inherently better than 2D for purposes of face recognition [22,34,10,51]. One reason often asserted for the superiority of 3D is that it is "illumination independent" whereas 2D appearance can be affected by illumination in various ways. It is true that 3D shape per se is illumination independent, in the sense that a given 3D shape exists the same independent of how it is illuminated. However, the sensing of 3D shape is generally not illumination independent—*changes in the illumination of a 3D shape can greatly affect the shape description that is acquired by a 3D sensor*.

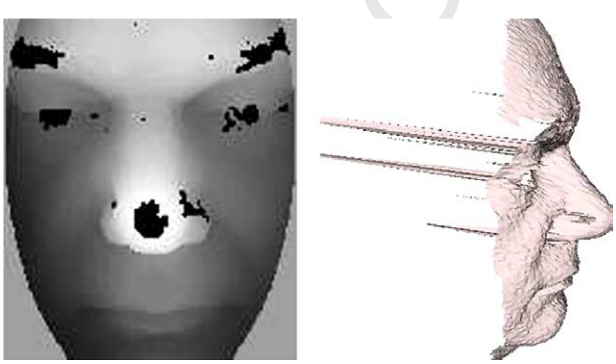

Fig. 3. Example of "hole" and "spike" artifacts in sensed 3D shape. The 3D data are rendered as a cropped, frontal view, range image on the left. The black regions are "holes" of missing data. The data is rendered as a side view of a shaded shape model on the right. Noise points in the data are readily apparent as "spikes" away from the face surface. Essentially all 3D sensors are subject to some level these sorts of artifacts in the raw data.

The acquisition of 3D shape by either stereo or structured-light involves taking one or more standard 2D intensity images. The 2D images are typically taken with commercially available digital cameras. The camera can receive light of an intensity that saturates the detector, and can also receive light levels too low to produce high-quality images. The 2D image can have artifacts due to illumination, and the artifacts in the 2D images can lead to artifacts in the 3D images. The types of artifacts that can arise in the 2D and the 3D are of course different, but are often related. The determination of which type of image inherently has more frequent or more important artifacts due to illumination is not clear, and is possibly sensor and application dependent.

Fig. 4 makes the point that the shape models acquired by currently available 3D sensors can be greatly affected by changes in illumination. Two 3D shape models of the same face are shown, rendered as smooth-shaded 3D meshes without any superimposed texture map. Models were converted to VRML format and then rendered as a shaded image. One shape model is acquired under ambient lighting conditions appropriate to the particular sensor, and the other is acquired at the same session but with an extra studio spotlight turned on, located about 1.5 m in front of and slightly above the person. The glaring artifacts in the second shape model are due to the change in the lighting conditions. The particular manufacturer and model of sensor are not important to this example, as it is not our point to argue for or against any particular 3D sensor. In our experience, similar problems can occur for any of the 3D sensors currently used in the face recognition research community, whether they operate on a stereo or a structured-light basis. Current 3D sensors take various approaches to the problem of coping with changes in illumination. The Cyberware sensor is one extreme example. It requires that the subject be positioned accurately and quite close to the sensor, and uses its own strong illumination. The illumination is so strong that most subjects find it difficult not to blink during a scan. Thus the Cyberware controls the conditions of acquisition strongly enough that ambient light is nearly unimportant. The Minolta Vivid 900 has a relatively narrow range of ambient lighting in which it will function. The quality of the sensed 3D shape can degrade with variation in lighting, but large changes in lighting simply cause the system to be unable to acquire 3D shape. Our view is that no particular technology or manufacturer has yet solved this problem in a general way with respect to surveillance applications.

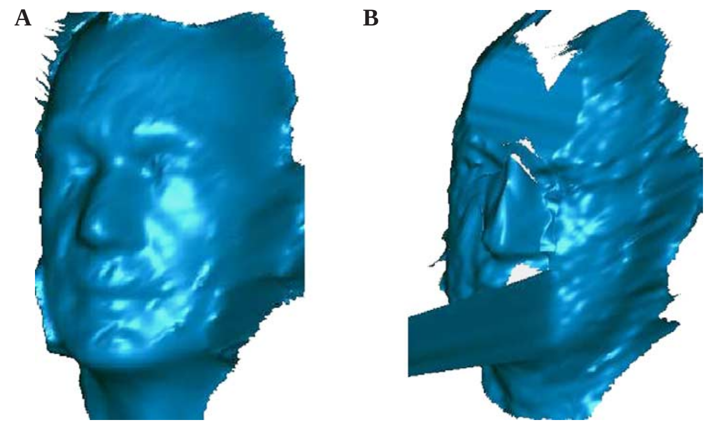

Fig. 4. Example shape models of same person under different lighting conditions. (A) With lighting appropriate to sensor. (B) With additional studio spotlight 1.5 m away.

Creating a sensor that automatically adapts to variations in illumination is certainly a major practical area for advance in 3D sensor technologies.

A related point is that evaluation of 3D shape should only be done when the color texture is *not* displayed. When a 3D model is viewed with the texture map on, the texture map can hide significant artifacts in the 3D shape. This is illustrated by the pair of images shown in Fig. 5. Both images represent the same 3D shape model, but in one case it is rendered with the texture map on and in the other case is rendered as a shaded view of the shape model. The shape model clearly has major artifacts that are related to the lighting highlights in the image.

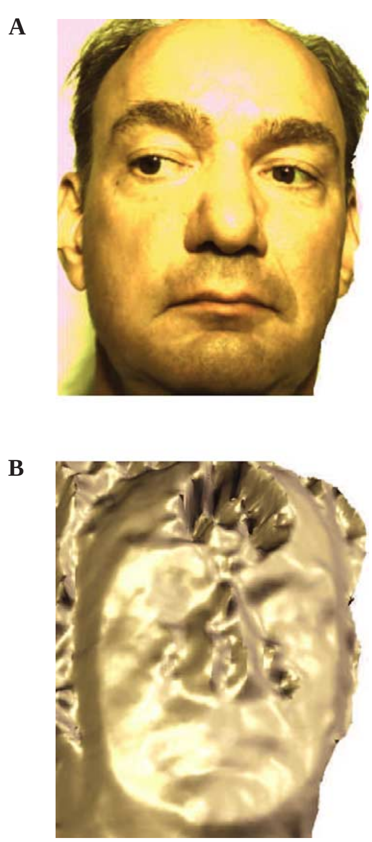

Fig. 5. Example of a 3D shape errors masked by viewing with texture map on. (A) A view of a 3D model rendered with the texture map on. (B) The same 3D model as in (A) but rendered as shaded model without the texture map on.

6.2. Tradeoffs in "active" versus "passive" acquisition

One important issue is whether or not the sensor is an "active" one; that is, whether it projects light of some form onto the scene. If it projects coherent light, then there are potential eye safety issues. If it does not project coherent light, then issues of depth-versus-accuracy tradeoff become more important. If the sensor projects a sequence of light stripes or patterns and acquires an image of each, then the effective acquisition time increases. In general, shorter acquisition times are better than longer acquisition times, in order to minimize artifacts due to subject motion. The shortest image acquisition time possible would seem to be that of a single image, or multiple images taken truly simultaneously. In this regard, a stereo-based system would seem to have an advantage. However, stereo-based systems can have trouble getting a true dense sampling of the face surface. Systems that depend on structured-light typically have trouble in regions such as eyebrows, and often generate spike artifacts when light undergoes multiple reflections. Systems that depend on stereo correspondence often have sparse sampling of points in regions where there is not much natural texture, and may generate surfaces that are too smooth in such cases.

6.3. Sampling and accuracy of 3D points

There is currently no clear concept of what sampling density and depth accuracy of 3D points is truly needed for 3D face recognition. Experimental results in the literature come from data where the number of sample points on the face may range from a few hundred to a few tens of thousands. The accuracy of the depth data likely varies over a similar broad range. There are some results suggesting that depth accuracy of less than 1 mm is useful [14]. However, this is based on experiments with a particular data set and a particular (eigenface style) algorithm. Since the cost of range sensors can increase dramatically with increases in the number of sample points or the accuracy of the depth value, more work is needed to determine what is truly required for face recognition applications. Boehnen and Flynn [6] performed an experimental evaluation of the depth accuracy of five current 3D sensors in a face sensing context. We are not aware of any other such comparison in the literature.

Considering all of the factors related to current 3D sensor technology, it seems that the optimism sometimes expressed for 3D face recognition relative to 2D face recognition may be premature. Existing 3D sensors are certainly capable of supporting advanced research in this area, but are far from ideal for practical application. An ideal 3D sensor for face recognition applications would combine at least the following properties: (1) image acquisition time similar to that of a typical 2D camera, (2) a large depth of field; e.g, a meter or more in which there is essentially no loss in accuracy of depth resolution, (3) robust operation under a range of "normal" lighting conditions, (4) no eye safety issues arising from projected light, (5) dense sampling of depth values; perhaps 1000×1000, and (6) depth resolution of better than 1 mm. Evaluated by these criteria, we do not know of any currently available 3D sensor that could be considered as ideal for use in face recognition.

7. Challenge for 3D face recognition: improved algorithms

One important area for improved algorithms is to better handle expression variation between gallery and probe images. Significant effort has begun to be put into this problem in the last few years. The FRGC data set is the most challenging data set supporting research on this topic at the time of this writing [45]. Approaches that treat the face as a rigid object, such as standard eigenface or ICP approaches, do not perform well in the presence of expression variation. There are at least three general methods that one might employ to attempt to deal with varying facial expression. One approach would be to simply concentrate on regions of the face whose shape changes the least with varying facial expression. For example, one might ignore the lips and mouth region, since their shapes vary greatly with expression. Or one might select feature points on the face where the shape changes relatively little with expression. Of course, there is no large subset of the face that is perfectly shape invariant across all expression changes, and so this approach will not be perfect. Another approach would be to enroll a person into the gallery by intentionally sampling a set of different facial expressions, and to match a probe against the set of shapes representing a person. This approach requires the set of different facial expressions for enrollment, and it may be difficult to acquire or generate the needed data. This approach also runs into the problem that, however large the set of facial expressions sampled for enrollment, the probe shape may represent an expression different from any of those sampled. Thus this approach also does not seem to allow the possibility of a perfect solution. A third approach would be to have a general model of 3D facial expression that can be applied to any person's image(s). The search for a match between a gallery and a probe shape could then be done over the set of parameters controlling the particular instantiation of the shape. There likely is no general model to predict, for example, how each person's neutral-expression image is transformed into their smiling image. A smile means different things to different persons' facial shapes, and different things to the same person at different times and in different cultural contexts. Thus this approach seems destined to also run into problems.

Chang et al. [16] explore an approach that tries to use regions of the face that change relatively little with common expressions. They use two different shape regions around the nose area, perform an ICP-based matching independently for each region, and combine the results of the two matches. They call this an Adaptive Rigid Multi-region Selection (ARMS) approach. They evaluate this approach on version two of the Face Recognition Grand Challenge data set [45]. They report that using smaller regions of face shape data from around the nose actually improves performance even in the case of matching neutral-expression probe to neutral-expression gallery. The ARMS approach results in 96% rank-one recognition when matching neutral expression to neutral expression, and 87% when matching varied expression to neutral expression. While the 87% performance is a substantial improvement over the performance of the standard ICP algorithm, there is clearly still room for further improvement.

In addition to a need for more sophisticated 3D recognition algorithms, there is also a need for more sophisticated multi-modal combination. Those studies that suggest that 3D allows greater accuracy than 2D also suggest that multi-modal recognition allows greater accuracy than either modality alone. And a 2D camera is typically already present as a part of a 3D sensor, so it seems that 2D can generally be acquired along with 3D. Thus the more productive research issue may not be 3D versus 2D, but instead the best method to use to combine 3D and 2D. Multi-modal combination has so far generally taken a fairly simple approach. The 3D recognition result and the 2D recognition result are each produced without reference to the other modality, and then the results are combined in some way. It is at least potentially more powerful to exploit possible synergies between the two modalities in the interpretation of each modality. For example, knowledge of the 3D shape might help in interpreting shadow regions in the 2D image. Similarly, regions of facial hair might be easy to identify in the 2D image and help to predict regions of the 3D data which are more likely to contain artifacts.

While this survey has only dealt with multi-modal biometrics in the sense of 3D + 2D face, there are other interesting possibilities to be explored. For example, the use of 2D images of the face has the potential to provide data that might be used for iris recognition or ear recognition [15] as well. And the use of 3D data of the face has the potential to provide data that might be used for 3D ear recognition [58] as well. Thus there appear to be several opportunities to exploit multi-biometric approaches other than 3D + 2D face.

8. Challenge for 3D face recognition: improved methodology

One barrier to experimental validation and comparison of 3D face recognition is lack of appropriate datasets. Desirable properties of such a dataset include: (1) a large number and demographic variety of people represented, (2) images of a given person taken at repeated intervals of time, (3) images of a given person that represent substantial variation in facial expression, (4) high-spatial resolution, for example, depth resolution of 1 mm or better, and (5) low frequency of sensor-specific artifacts in the data. Expanded use of common datasets and baseline algorithms in the research community will facilitate the assessment of the state of the art in this area. It would also improve the interpretation of research results if the statistical significance, or lack thereof, was reported for observed performance differences between algorithms and modalities.

Another aspect of improved methodology would be the use, where applicable, of explicit and distinct training, validation, and test sets. For example, the "face space" for a PCA algorithm might be created based on a training set of images, the number of eigenvectors used and the distance metric used then selected based on a validation set, and finally the performance estimated on a test set. The different sets of images would be non-overlapping with respect to the persons represented in each.

A more subtle methodological point is involved in the comparison of multi-modal results to results from a single modality. Multi-modal 3D + 2D performance is always observed to be greater than the performance of 2D alone. However, as explained earlier, this comparison is generally biased in favor of the multi-modal result. A more appropriate comparison would be to a 2D recognition system that uses two images of a person both for enrollment and for recognition. When this sort of controlled comparison is done, the differences observed for multi-modal 3D + 2D compared to "multi-sample" 2D are smaller than those for a comparison to simple 2D [13]. This suggests that the research issue of how to select the best set of multiple samples of a given modality is one that could be important in the future.

9. Summary

Face recognition has many potential applications of great significance to our society [7]. The use of 3D sensing is an important avenue to be explored for increasing the accuracy of biometric recognition. It is clear from this survey that research involving 3D face recognition is in a period of rapid expansion. New work is appearing often, and in a wide variety of journals and conferences. We have attempted to be comprehensive and current in this survey, but this is a difficult goal, and we have likely inadvertently omitted some important recent work. We apologize to the authors of any work that we have omitted.

Three-dimensional face recognition faces a number of challenges if research achievements are to transition to successful use in major applications. The quality of 3D sensors has improved in recent years, but certainly even better 3D sensors are needed. In this case, "better" means sensing that is less sensitive to ambient lighting, has fewer artifacts, and requires less explicit user cooperation. A sensor that provides greater accuracy, but does so by requiring that the person remain motionless for several seconds at a relatively precise distance from the sensor, will likely not help to move 3D face recognition closer to broad application.

Similarly, three-dimensional face recognition needs better algorithms. Here, "better" means more tolerant of real-world variety in the pose, facial expression, eye-glasses, jewelry and other factors. At the same time, "better" also means less computationally demanding. Three-dimensional face recognition in general seems to require much more computational effort "per match" than does 2D face recognition.

The field also needs to mature in its appreciation of rigorous experimental methodology for validating improvements to the state of the art. The larger and more challenging public data sets that are now available to the research community are only one element of this. These data sets will facilitate comparisons between approaches, but data sets alone do not guarantee sound comparisons. For example, a comparison of a proposed new approach to an eigenface approach that uses a clearly too-small training set is a "straw person" sort of comparison. Ideally, researchers would compare directly to the results achieved by other researchers on the same data set. Also, as mentioned earlier, the interpretation of the size or importance of reported improvements would be aided by the use of appropriate tests of statistical significance.

If all of these challenges are addressed, then some of the optimistic expressions about the potential of 3D face recognition will have a chance to come true.

Acknowledgments

This work is supported by National Science Foundation Grant CNS-0130839, by the Central Intelligence Agency, and by Department of Justice Grant 2004-DD-BX-1224.

References

[1] 3DMD Systems. 3q qlonerator. <http://www.3q.com/offerings_prod.htm/>.

[2] B. Achermann, H. Bunke, Classifying range images of human faces with Hausdorff distance, in: 15-th International Conference on Pattern Recognition, September 2000, pp. 809–813.

[3] B. Achermann, X. Jiang, H. Bunke, Face recognition using range images, International Conference on Virtual Systems and MultiMedia (1997) 129–136.

[4] C. Beumier, M. Acheroy, Face verification from 3D and grey level cues, Pattern Recognition Letters 22 (2001) 1321–1329.

[5] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 1063–1074.

[6] C. Boehnen, P.J. Flynn, Accuracy of 3D scanning technologies in a face scanning context, in: Fifth International Conference on 3D Imaging and Modeling (3DIM 2005), June 2005.

[7] K.W. Bowyer, Face recognition technology and the security versus privacy tradeoff, IEEE Technology and Society (2004) 9–20.

[8] K.W. Bowyer, K. Chang, P.J. Flynn, A survey of 3D and multi-modal 3D + 2D face recognition, in: 17-th International Conference on Pattern Recognition, August 2004, pp. 358–361.

[9] K.W. Bowyer, K. Chang, P.J. Flynn, A survey of 3D and multi-modal 3D + 2D face recognition, Face Processing: Advanced Modeling and Methods, to appear.

[10] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Expression-invariant 3D face recognition, in: International Conference on Audio- and Video-Based Person Authentication (AVBPA 2003), LNCS, vol. 2688, 2003, pp. 62–70.

[11] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Three-dimensional face recognition, International Journal of Computer Vision (2005) 5–30.

[12] J.Y. Cartoux, J.T. LaPreste, M. Richetin, Face authentication or recognition by profile extraction from range images, in: Proceedings of the Workshop on Interpretation of 3D Scenes, 1989, pp. 194–199.

[13] K. Chang, K. Bowyer, P. Flynn, An evaluation of multi-modal 2D + 3D face biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4) (2005) 619–624.

[14] K. Chang, K. Bowyer, P. Flynn, Face recognition using 2D and 3D facial data, in: Multimodal User Authentication Workshop, December 2003, pp. 25–32.

[15] K. Chang, K.W. Bowyer, S. Sarkar, B. Victor, Comparison and combination of ear and face images for appearance-based biometrics, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1160–1165.

[16] K.I. Chang, K.W. Bowyer, P.J. Flynn, Adaptive rigid multi-region selection for handling expression variation in 3D face recognition, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

[17] C. Chua, F. Han, Y.K. Ho, 3D human face recognition using point signature, IEEE International Conference on Automatic Face and Gesture Recognition (2000) 233–238.

[18] L. Farkas, Anthropometry of the Head and Face, Raven Press, New York, 1994.

[19] A. Godil, S. Ressler, P. Grother, Face recognition using 3D facial shape and color map information: comparison and combination, in: Biometric Technology for Human Identification, SPIE, vol. 5404, April 2005, pp. 351–361.

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

[21] G. Gordon, Face recognition based on depth and curvature features, Computer Vision and Pattern Recognition (CVPR) (June) (1992) 108–110.

[22] C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range imaging, in: Seventh International Symposium on Signal Processing and Its Applications, 2003, pp. 201–204.

[23] M. Husken, M. Brauckmann, S. Gehlen, C. von der Malsburg, Strategies and benefits of fusion of 2D and 3D face recognition, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

[24] M.L. Koudelka, M.W. Koch, T.D. Russ, A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

[25] S. Lao, Y. Sumi, M. Kawade, F. Tomita, 3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision, in: International Conference on Pattern Recognition (ICPR 2000), 2000, pp. II:911–916.

[26] J.C. Lee, E. Milios, Matching range images of human faces, in: International Conference on Computer Vision, 1990, pp. 722–726.

[27] Y. Lee, K. Park, J. Shim, T. Yi, 3D face recognition using statistical multiple features for the local depth information, in: 16th International Conference on Vision Interface, June 2003. Available at <www.visioninterface.org/vi2003/>.

[28] Y. Lee, J. Shim, Curvature-based human face recognition using depth-weighted Hausdorff distance, in: International Conference on Image Processing (ICIP), 2004, pp. 1429–1432.

[29] Y. Lee, H. Song, U. Yang, H. Shin, K. Sohn. Local feature based 3D face recognition, in: International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005), LNCS, vol. 3546, July 2005, pp. 909–918.

[30] X. Lu, D. Colbry, A.K. Jain, Matching 2.5D scans for face recognition, in: International Conference on Pattern Recognition (ICPR 2004), 2004, pp. 362–366.

[31] X. Lu, A.K. Jain, Deformation analysis for 3D face matching, in: 7th IEEE Workshop on Applications of Computer Vision (WACV '05), 2005, pp. 99–104.

[32] X. Lu, A.K. Jain, Integrating range and texture information for 3D face recognition, in: 7th IEEE Workshop on Applications of Computer Vision (WACV '05), 2005, pp. 155–163.

[33] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, G. Medioni, Performance of geometrix activeidtm 3D face recognition engine on the frgc data, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

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

[35] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, XM2VTSDB: the extended M2VTS database, in: Second International Conference on Audio- and Video-based Biometric Person Authentication, 1999, pp. 72–77.

[36] J. Min, K.W. Bowyer, P. Flynn, Using multiple gallery and probe images per person to improve performance of face recognition, Notre Dame Computer Science and Engineering Technical Report (2003).

[37] Minolta Inc. Konica Minolta 3D digitizer. <http://www.minoltausa.com/vivid/>.

[38] A.B. Moreno, Ángel Sánchez, J.F. Vélez, F.J. Díaz, Face recognition using 3D surface-extracted descriptors, in: Irish Machine Vision and Image Processing Conference (IMVIP 2003), September 2003.

[39] T. Nagamine, T. Uemura, I. Masuda, 3D facial image analysis for human identification, in: International Conference on Pattern Recognition (ICPR 1992), 1992, pp. 324–327.

[40] Neven Vision, Inc. Nevenvision machine vision technology. <http://www.nevenvision.com/>.

[41] G. Pan, S. Han, Z. Wu, Y. Wang, 3D face recognition using mapped depth images, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

[42] G. Pan, Z. Wu, Y. Pan, Automatic 3D face verification from range data, in: International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2003, pp. III:193–196.

[43] T. Papatheodorou, D. Reuckert, Evaluation of automatic 4D face recognition using surface and texture registration, in: Sixth International Conference on Automated Face and Gesture Recognition, May 2004, pp. 321–326.

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

[45] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, Computer Vision and Pattern Recognition (CVPR) (2005).

[46] P.J. Phillips, P. Grother, R.J. Michaels, D.M. Blackburn, E. Tabassi, J. Bone, FRVT 2002: overview and summary. Available at <www.frvt.org/>.

[47] P.J. Phillips, H. Moon, P.J. Rauss, S. Rizvi, The FERET evaluation methodology for face recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000).

[48] M. Rioux, L. Cournoyer, Nrcc three-dimensional image data files, National Research Council of Canada, NRC 29077, June 1988.

[49] T.D. Russ, K.W. Koch, C.Q. Little, 3D facial recognition: a quantitative analysis, in: 45-th Annual Meeting of the Institute of Nuclear Materials Management (INMM), July 2004.

[50] T.D. Russ, M.W. Koch, C.Q. Little, A 2D range Hausdorff approach for 3D face recognition, in: IEEE Workshop on Face Recognition Grand Challenge Experiments, June 2005.

[51] A. Scheenstra, A. Ruifrok, R.C. Veltkamp, A survey of 3D face recognition methods, in: International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005), LNCS, vol. 3546, July 2005, pp. 891–899.

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

[53] F. Tsalakanidou, S. Malassiotis, M. Strintzis, Face authentication and authentication using color and depth images, IEEE Transactions on Image Processing 14 (2) (2005) 152–168.

[54] F. Tsalakanidou, S. Malassiotis, M. Strintzis, Integration of 2D and 3D images for enhanced face authentication, in: Sixth International Conference on Automated Face and Gesture Recognition, May 2004, pp. 266–271.

[55] F. Tsalakanidou, D. Tzocaras, M. Strintzis, Use of depth and colour eigenfaces for face recognition, Pattern Recognition Letters 24 (2003) 1427–1435.

[56] Y. Wang, C. Chua, Y. Ho, Facial feature detection and face recognition from 2D and 3D images, Pattern Recognition Letters 23 (2002) 1191–1202.

[57] C. Xu, Y. Wang, T. Tan, L. Quan, Automatic 3D face recognition combining global geometric features with local shape variation information, in: Sixth International Conference on Automated Face and Gesture Recognition, May 2004, pp. 308–313.

[58] P. Yan, K.W. Bowyer, Empirical evaluation of advanced ear biometrics, in: IEEE Workshop on Empirical Evaluation Methods in Computer Vision (EEMCV 2005), June 2005.

[59] P. Yan, K.W. Bowyer, A fast algorithm for ICP-based 3D shape biometrics, in: Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005), October 2005 (to appear).

[60] W. Zhao, R. Chellappa, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (December) (2003) 399–458.