Volume 101, Number 1, January 2006

ISSN 1077-3142

ELSEVIER

Computer Vision and Image Understanding

Computer Vision and Image Understanding

Volume 101, Number 1, January 2006

Available online at www.sciencedirect.com

SCIENCE DIRECT

Computer Vision and Image Understanding 101 (2006) 1–15

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
Available online 11 October 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), jin.chang@philips.com (K. Chang), flynn@cse.nd.edu (P. Flynn).

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

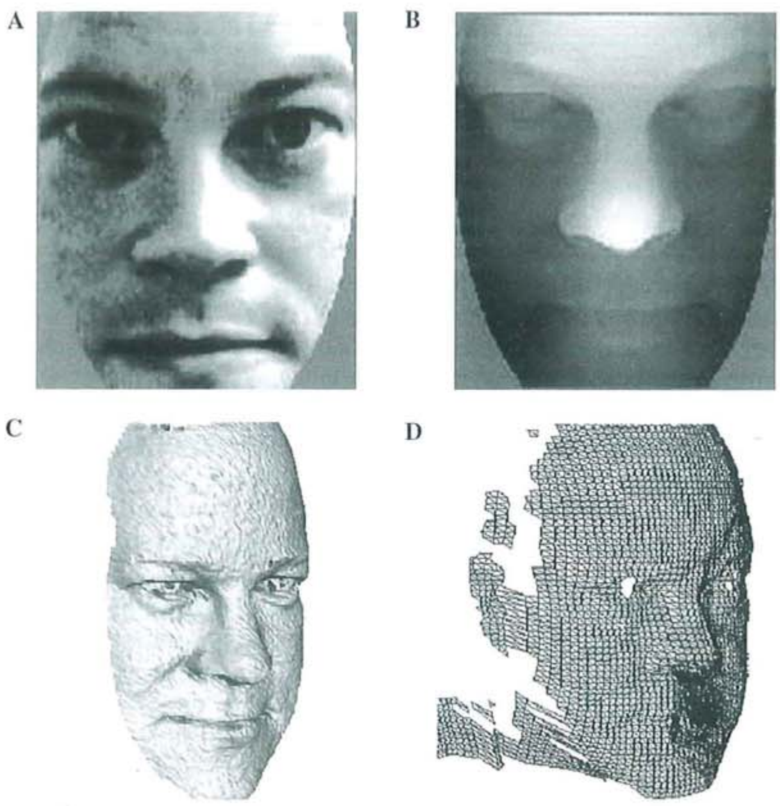

Fig. 1. Example of 2D intensity and 3D shape data. The 2D intensity image and the 3D range image are representations that would be used with

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	Not available	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	Not available	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	Not available	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	Not available	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	Not available	Point set + feature vector	Minimum distance	96% on 30, 72% on 120
Bronstein, 2005 [11]	30	220	Not available	Point set	"canonical forms"	100%
Chang, 2005 [16]	466 FRGC v2	4007	480×640	Point set	multi-ICP	92%
Gökberk, 2005 [20]	106	579	Not available	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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.