Computer Vision and Image Understanding 154 (2017) 137–151

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

3D-2D face recognition with pose and illumination normalization

Ioannis A. Kakadiaris[a,1,*], George Toderici[a,2], Georgios Evangelopoulos[a,3], Georgios Passalis[a,b,4], Dat Chu[a,5], Xi Zhao[a,6], Shishir K. Shah[a,1], Theoharis Theoharis[a,b,7]

[a] Computational Biomedicine Laboratory (CBL), Department of Computer Science, Univ. of Houston, 4800 Calhoun, Houston, TX 77204, USA
[b] Department of Informatics, Univ. of Athens, TYPA Buildings, Panepistimiopolis, Ilisia, 15784, Athens, Greece

ARTICLE INFO

Article history:
Received 28 April 2015
Revised 24 February 2016
Accepted 28 April 2016
Available online 21 May 2016

Keywords:
Face and gesture recognition
Biometrics
Physically-based modeling
3D-2D face recognition
Illumination normalization
Model-based face recognition
3D-2D model fitting
Object recognition
Computer vision

ABSTRACT

In this paper, we propose a 3D-2D framework for face recognition that is more practical than 3D-3D, yet more accurate than 2D-2D. For 3D-2D face recognition, the gallery data comprises of 3D shape and 2D texture data and the probes are arbitrary 2D images. A 3D-2D system (UR2D) is presented that is based on a 3D deformable face model that allows registration of 3D and 2D data, face alignment, and normalization of pose and illumination. During enrollment, subject-specific 3D models are constructed using 3D+2D data. For recognition, 2D images are represented in a normalized image space using the gallery 3D models and landmark-based 3D-2D projection estimation. A method for bidirectional relighting is applied for non-linear, local illumination normalization between probe and gallery textures, and a global orientation-based correlation metric is used for pairwise similarity scoring. The generated, personalized, pose- and light- normalized signatures can be used for one-to-one verification or one-to-many identification. Results for 3D-2D face recognition on the UHDB11 3D-2D database with 2D images under large illumination and pose variations support our hypothesis that, in challenging datasets, 3D-2D outperforms 2D-2D and decreases the performance gap against 3D-3D face recognition. Evaluations on FRGC v2.0 3D-2D data with frontal facial images, demonstrate that the method can generalize to databases with different and diverse illumination conditions.

1. Introduction

Face recognition (FR) has been a key topic in computer vision, pattern recognition, and machine learning research, with extensions to perceptual, behavioral, and social principles. In parallel, FR technology has been advancing in terms of sensors, algorithms, databases, and evaluation frameworks. This increasing interest is driven partly by the difficulty and challenges of the task (i.e., a complex, intra-class object recognition problem) and partly by a wide variety of applications involving identity management. Research challenges include (i) separating intrinsic from extrinsic appearance variations; (ii) developing discriminative representations and similarity metrics; and (iii) discovering performance invariants across heterogeneous data and conditions. Application-wise, face is emerging as a powerful biometric, a high-level semantic for content-based indexing and retrieval, and a natural and rich communication modality for human-computer interaction. The existing frameworks for face recognition vary across approaches (e.g., data-driven, model-based, perceptual) or facial data domains (e.g., images, point clouds, depth maps).

Methods for image-based FR have been pushing performance boundaries on nearly-frontal-view faces and constrained illumination conditions (Abate et al., 2007). However, the appearance of 2D images, under real-life and realistic acquisition conditions, is affected by extrinsic and identity-independent factors, such as variations in pose/viewpoint, illumination, facial expressions, time-lag, and occlusions (partial-data). In challenging, image "in the wild" benchmarks (Wolf et al., 2011), state-of-the-art performance depends on face pre-alignment, combining representations, or large-scale training. To alleviate extrinsic variability, increase the

* Corresponding author.
E-mail addresses: ikakadia@central.uh.edu, ioannisk@uh.edu (I.A. Kakadiaris), george.toderici@gmail.com (G. Toderici), gevang@mit.edu (G. Evangelopoulos), passalis@di.uoa.gr (G. Passalis), dattanchu@gmail.com (D. Chu), zhaoxi1@gmail.com (X. Zhao), shah@cs.uh.edu (S.K. Shah), theotheo@idi.ntnu.no (T. Theoharis).
[1] CBL, University of Houston.
[2] Google Inc., was with CBL, University of Houston when this work was performed.
[3] the Laboratory for Computational and Statistical Learning, MIT and Istituto Italiano di Tecnologia was with CBL, University of Houston when this work was performed.
[4] Accenture, was with CBL, University of Houston and University of Athens, Greece when this work was performed.
[5] CBL, University of Houston when this work was performed.
[6] School of Management, Xian Jiaotong University, 710049, Xian, P.R. China, was with CBL, University of Houston when this work was performed.
[7] IDI, NTNU, Norway, University of Athens, Greece, and CBL.

http://dx.doi.org/10.1016/j.cviu.2016.04.012
1077-3142/

discriminative ability, and boost the performance of conventional, image-based methods, alternative facial modalities, and sensing devices have been considered.

Three-dimensional recognition, from depth images, depth point clouds, or 3D meshes, has emerged as a distinct principle in biometrics and face recognition research (Abate et al., 2007; Bowyer et al., 2006), driven by improved 3D sensors, publicly available databases, and systematic evaluation benchmarks like the Face Recognition Grand Challenge (FRGC) (Phillips et al., 2005) and Face Recognition Vendor Test (FRVT) (Phillips et al., 2010). In these frameworks, which explored the feasibility of using 3D data both for enrollment and recognition, the 3D-based algorithms demonstrated a potential for very high recognition rates. For example, on FRGC v2.0, the 3D-3D face recognition system by Kakadiaris et al. (2007) reported a 97.5% rank-1 recognition and an average verification rate of 97.1% at 0.001 false acceptance rate (Ocegueda et al., 2011b), and the system of (Wang et al., 2010) 98.3% and 98.13%, respectively. Ocegueda et al. (2011a) achieved state-of-the-art performance both in FRGC v2.0, with 99% identification and 98% verification rate, and in the challenging 3D Twins Expression database (3DTEC) (Vijayan et al., 2011).

As an alternative, an asymmetric recognition system may involve 3D data for enrollment and 2D for verification or identification (3D-2D) or, the converse, 2D data for gallery and 3D data for probes (2D-3D). In the former, the need for 3D acquisition hardware is restricted to enrollment only and can facilitate the acquisition, storage, and distribution of high-quality databases of 3D models. In the latter, the abundance of existing face databases, composed primarily of 2D images, can provide reference enrollment sets for matching new 3D data. Independently, 3D model-based facial signatures are more discriminative and robust to condition variations. In this work, we propose a 3D-2D recognition framework which makes use of 3D data for enrollment, while requiring only 2D data for recognition, and which can be readily applied to the 2D-3D case also.

From the 3D gallery data, we build subject-specific, non-parametric 3D facial models by fitting a deformable Annotated Face Model (AFM) (Kakadiaris et al., 2007). The model surface parametrization defines a canonical 2D representation, the geometry image, that enables texture values assignment to corresponding 3D model points (Theoharis et al., 2008). A probe 2D image is mapped onto a subject-specific gallery model by explicitly accounting for relative pose and camera parameters using point-landmark correspondences (*pose estimation*). The estimated 3D-2D projection transformation is employed to generate pose-normalized texture images from the 2D image data and 3D model points (*texture lifting*). For matching, probe and gallery textures are lifted using the same 3D model. Their lighting conditions are further normalized using an illumination transfer method based on an analytical reflectance model (*texture relighting*). The final matching score between relit gallery and probe textures is a global similarity value obtained from low-level local orientation features.

Compared to asymmetric or heterogeneous recognition methods that map features across different modalities, the developed 3D-2D framework (termed UR2D) relies on a modality synergy, in which a 3D model is used for registration, alignment, and pose-light normalization of 2D image and texture data. Compared to previous approaches for 3D-2D registration and fitting (Gu and Kanade, 2006), UR2D employs the 3D shape information for relighting (using surface normal information) and score computation (extracting signatures in the geometry image space). Compared to existing multimodal 2D+3D methods (Jahanbin et al., 2011; Mian et al., 2007), UR2D integrates facial data across modalities and across enrollment/recognition phases in a subject-specific manner. In addition, unlike existing 3D-aided 2D recognition methods that use a 2D image to infer a 3D gallery model (Romdhani et al., 2006), UR2D is based on personalized gallery models constructed by fitting a model on the actual 3D facial data.

Our contributions can be summarized as follows: (i) we describe a conceptual framework for 3D-2D (or 2D-3D) face recognition; (ii) we propose a novel 3D-2D system for face image verification and identification from 3D datasets; (iii) we advocate the use of 2D+3D data to build subject-specific 3D gallery models that allow for personalized, texture-based similarity scores; (iv) we propose a method for model-based texture representation and a relighting algorithm for illumination normalization that improves recognition under lighting variations; and (v) we demonstrate empirically that 3D-2D recognition surpasses 2D-2D on challenging 2D+3D data with pose and illumination variations, and can approximate 3D-3D, shape-based similarity methods.

2. Related work

3D-2D and 3D-aided 2D face recognition: Recognition with 3D data spans an extensive body of work in 3D, 2D+3D (Bowyer et al., 2006), and 3D-aided 2D (Abate et al., 2007) FR. Rama et al. (2006) presented a method for simultaneous pose estimation and 2D face recognition that uses 3D data for training, though as a cylindrical texture image representation and not a full shape model. Riccio and Dugelay (2007) proposed using geometric invariants on the face to establish a correspondence between the 3D gallery face and a 2D probe image, disregarding the texture registered with the 3D data. Yin and Yourst (2003) used frontal and profile 2D images to construct models for 3D-based recognition. Blanz and Vetter (2003) introduced the 3D morphable model that captures face geometry and texture from 2D images. Using a statistical 3D face model and point correspondences, a gallery model is built from a 2D image. This is in principle different from our work, which uses real 2D and 3D data to build a 3D subject-specific gallery model. The morphable model framework was adopted for 3D-model-based 2D face recognition under illumination and pose variations (Romdhani et al., 2006), in fitting, synthesis, or normalization approaches (Zhang et al., 2014). Wang *et al.* extended it for a spherical harmonic representation (Wang et al., 2009). In contrast, methods for asymmetric 3D-2D FR learn a mapping between 3D and 2D data. Huang et al. (2010) map features extracted from gallery range images (2.5D) to 2D for 2D-matching of texture probe images. An extension, based on pose-light normalization and a mid-level representation, is reported to attain 95.4% recognition rate on FRGC v2.0 (Zhang et al., 2012).

3D face models from 2D images: In 2D-based methods, facial surface reconstruction from single or multiple images has been approached through stereo, structure-from-motion (Bregler et al., 2000), photometric-stereo (Georghiades et al., 2001), and shape-from-shading methods (Atick et al., 1996; Kemelmacher-Shlizerman and Basri, 2011) by estimating depth values from geometric, photometric, and gradient properties. Given a 3D prototype, reconstruction from images is obtained by fitting the model to 2D (i.e., estimating model parameters from image and geometric constraints) (Lee and Ranganath, 2003; Levine and Yua, 2009; Park et al., 2005). A model can be constructed from facial sample collections or prior knowledge on facial properties and physiology, and may be sparse to the number of points (i.e., point distribution models) or dense (i.e., vertex points with surface parametrization). Examples include shape-subspace projections, such as active appearance models (Matthews et al., 2007) and 3D morphable models (Blanz and Vetter, 2003), statistical deformable models (Kakadiaris et al., 2007; Mpiperis et al., 2008), elastic models (Prabhu et al., 2011), or reference samples (Kemelmacher-Shlizerman and Basri, 2011). Gu and Kanade (2006) fit a sparse set of surface points and the associated texture patches and simultaneously estimate deformation parameters and pose, requiring a