Computer Vision and Image Understanding 117 (2013) 158–169

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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

meshSIFT: Local surface features for 3D face recognition under expression variations and partial data ☆

Dirk Smeets *, Johannes Keustermans, Dirk Vandermeulen, Paul Suetens

K.U. Leuven, Faculty of Engineering, ESAT/PSI, IBBT-K.U. Leuven Future Health Department Medical Imaging Research Center, University Hospitals Gasthuisberg, Herestraat 49-bus 7003, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 22 September 2011
Accepted 5 October 2012
Available online 1 November 2012

Keywords:
meshSIFT
3D face recognition
Local features
Expressions
Missing data
Partial data
Pose normalisation
Symmetry plane

ABSTRACT

Matching 3D faces for recognition is a challenging task caused by the presence of expression variations, missing data, and outliers. In this paper the *meshSIFT* algorithm and its use for 3D face recognition is presented. This algorithm consists of four major components. First, salient points on the 3D facial surface are detected as mean curvature extrema in scale space. Second, orientations are assigned to each of these salient points. Third, the neighbourhood of each salient point is described in a feature vector consisting of concatenated histograms of shape indices and slant angles. Fourth, the feature vectors of two 3D facial surfaces are reliably matched by comparing the angles in feature space. This results in an algorithm which is robust to expression variations, missing data and outliers.

As a first contribution, we demonstrate that the number of matching meshSIFT features is a reliable measure for expression-invariant face recognition, as shown by the rank 1 recognition rate of 93.7% and 89.6% for the Bosphorus and FRGC v2 database, respectively. Next, we demonstrate that symmetrising the feature descriptors allows comparing two 3D facial surfaces with limited or no overlap. Validation on the data of the "SHREC'11: Face Scans" contest, containing many partial scans, resulted in a recognition rate of 98.6%, clearly outperforming all other participants in the challenge. Finally, we also demonstrate the use of meshSIFT for two other problems related with 3D face recognition: pose normalisation and symmetry plane estimation. For both problems, applying meshSIFT in combination with RANSAC resulted in a correct solution for ±90% of all Bosphorus database meshes (except ±90° and ±45° rotations).

1. Introduction

Although research in automatic face recognition has been conducted since the 1960s [1], it is still an active research area. Since 2D, image-based, face recognition is still hampered by pose variations and varying lighting conditions, recent research has shifted from 2D to 3D face representations. This shift is demonstrated by the establishment of large evaluation studies of 3D face recognition algorithms. In 2006, the Face Recognition Grand Challenge (FRGC) [2] was the first large comparison, followed by the Shape Retrieval Contest (SHREC) in 2007 [3], 2008 [4] and 2011 [5].

Three-dimensional face recognition in real case scenarios is becoming affordable due to technological improvements in 3D surface acquisition devices for security purposes. However, some important challenges inherent to 3D face recognition as well as related to acquisition issues remain. Inherent challenges are mainly due to intra-subject deformations, often caused by changes in facial expressions [6]. Facial muscle contractions cause the soft tissue of the face to deform during expression variations, affecting automatic recognition.

The second challenge is posed by the limited field of view of most 3D scanners, impeding the scanning of the entire face. As a result, 3D face recognition is still pose dependent. In realistic situations, such as for uncooperative subjects or uncontrolled environments, no assumption can be made on the pose. Therefore, 3D face recognition methods should be able to match partials scans with little or even no overlap. Fig. 1 shows an example of such partial scans, again of the same individual.

1.1. Related work

Since excellent surveys exist summarising the extensive work in 3D face recognition [6,7], we will only review the work on expression-invariant face recognition and on face recognition not requiring overlap.

1.1.1. Expression-invariant 3D face recognition

Expression-invariant 3D face recognition methods can be subdivided into three classes, depending on the way these methods handle expressions.

☆ This paper has been recommended for acceptance by L.S. Davis.

* Corresponding author. Fax: +32 16349001.
E-mail addresses: dirk.smeets@uz.kuleuven.be (D. Smeets), johannes.keustermans@uz.kuleuven.be (J. Keustermans), dirk.vandermeulen@uz.kuleuven.be (D. Vandermeulen), paul.suetens@uz.kuleuven.be (P. Suetens).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.10.002

Fig. 1. An example of partial scans of the same individual with partial or even small overlap (images from the SHREC 2011 data [5]).

Historically, the first face recognition methods dealing with expression variations were region-based. These methods rely on parts of the face that remain unaffected during expression variations. The first and most used strategy is to select well-defined, anatomic regions based on observations or on literature such as the region around the nose [8,9], cheek [10], chin [10], eyes [8], forehead [8,11] and the region above the mouth [12]. A second strategy to determine expression-invariant regions, is the use of local features. Hereby regions, defined as local neighbourhoods around points of interest, are selected and matched automatically. If a local neighbourhood is small enough, it is assumed to be stable under expression variations. Convex regions [10], Gabor features [13–15], matched local invariant range images [16,17], Haar and Pyramid wavelet features [18], local shape pattern (LSP) features [19], local binary patterns (LBPs) [20] appear to be less affected by expressions. The algorithm presented in this paper belongs to this type of strategy. The third strategy is based on the automatic determination of the parts unaffected by expression variations as determined after alignment/registration as in [21]. Points with a low registration error are considered to belong to an unaffected and thus more rigid part of the face, whereas points with a high registration error are more likely to belong to a part of the face that is affected by expression variations. Alternatively, these regions can be learned using a training database [20]. Related to learning expression-robust regions is the subdivision of the face in small regions. By fusing the results of these different regions (suppressing those affected by expression variations), a high recognition accuracy is achieved [22,23].

The second major class of expression-invariant face recognition methods uses statistical models. A multivariate Gaussian (principal component analysis (PCA) based) point distribution model can deal with expressions by including faces with expression in the training data as in [24–26]. Expression induced deformations can also be modelled explicitly using PCA-decompositions, leading to 'principal warps' as is done by [27,28]. The former linearly combined this expression model with a PCA shape model for identity, assuming that it is possible to transfer expressions from one face to another. When this assumption is considered to be false, it is necessary to combine the expression model and identity model into a bilinear model as in [29]. However, model fitting becomes computationally more demanding. Statistical models different from PCA have been suggested as well: independent component analysis (ICA) [24], linear discriminant analysis (LDA) [25] or simply pointwise mean and standard deviation [30].

The third class of algorithms makes use of an isometric deformation model in which facial surface changes due to expression variations are modelled as isometric deformations. The most used isometric deformation invariant representations are iso-geodesics, curves containing points on an equal geodesic distance to a reference point (nose tip), as in [31–35]. A computationally more demanding representation is the geodesic distance matrix, containing the geodesic distance between each pair of points as in [36–39] or between a limited number of points as in [40,41].

An comparative study of 3D recognition methods dealing with expression variations is given in [42], elaborating more on the advantages and disadvantages of the different classes. It also provides a meta-analysis in an attempt to compare the classes more quantitatively.

1.1.2. 3D face recognition for partial data

The general strategy to handle partial data is to fit a full face model to the partial scan. In literature, the Morphable Model