Pattern Recognition 57 (2016) 21–30

Contents lists available at ScienceDirect

Pattern Recognition

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

A Grassmann framework for 4D facial shape analysis

Taleb Alashkar [a], Boulbaba Ben Amor [a,*], Mohamed Daoudi [a], Stefano Berretti [b]

[a] Institute Mines-Télécom/Télécom Lille; CRIStAL (UMR CNRS 9189), France
[b] Department of Information Engineering, University of Florence, Italy

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form
5 March 2016
Accepted 9 March 2016
Available online 25 March 2016

Keywords:
4D face recognition
Curvature-maps
Grassmann manifold
Dictionary learning
Sparse coding

ABSTRACT

In this paper, we investigate the contribution of dynamic evolution of 3D faces to identity recognition. To this end, we adopt a subspace representation of the flow of curvature-maps computed on 3D facial frames of a sequence, after normalizing their pose. Such representation allows us to embody the shape as well as its temporal evolution within the same subspace representation. Dictionary learning and sparse coding over the space of fixed-dimensional subspaces, called Grassmann manifold, have been used to perform face recognition. We have conducted extensive experiments on the BU-4DFE dataset. The obtained results of the proposed approach provide promising results.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years automatic face analysis has attracted increasing interest in the field of computer vision and pattern recognition due to its inherent challenges and its potential in a wide spectrum of applications, including security surveillance [1,2] and diagnostic of facial pathology [3]. Despite the great progress, 2D face analysis approaches that depend on color or gray-scale image analysis, still suffer from illumination and pose variations, which often occur in real-world conditions. With the rapid innovation of 3D cameras, the 3D shape is regarded as a promising alternative to achieve robust face analysis [4,5]. Very recently, the advent of 4D imaging systems capable of acquiring temporal sequences of 3D scans (i.e., 4D is regarded as 3D over the time) made possible comprehensive face analysis by introducing the temporal dimension, where the temporal behavior of 3D faces is captured by adjacent frames [6,7]. Note that such temporal information is crucial for analyzing the facial deformations. Despite the large amount of work on static and dynamic 3D facial scans analysis, temporal modeling is still almost unexplored for identity recognition. Moving from shape analysis of static 3D faces to dynamic faces (4D faces) gives rise to new challenges related to the nature of the data and the processing time – which static and dynamic shape representations are most suited to 4D face analysis? How the temporal dimension can contribute to face analysis? Is it possible to compute statistical summaries on dynamic 3D faces? From a perspective of face classification, which relevant features and classification algorithms can be used?

In this paper, we aim to answer the above questions by proposing a comprehensive framework for modeling and analyzing 3D facial sequences (4D faces), with an experimental illustration in face recognition from 4D sequences.

Recently, works addressing face analysis from temporal sequences of 3D scans start to appear in the literature, encouraged by the advancement in 3D sensors' technology, with some of them restricted to RGB-D Kinect-like sensors. In [8], Berretti et al. investigated the impact of 3D facial scans' resolution on the recognition rate by building super resolution 3D models from consumer depth camera frames. Experimental studies using the new 3D super resolution method validate the increase of recognition performance with the reconstructed higher resolution models. Hsu et al. [9] showed that incorporating depth images of the subjects in the gallery can improve the recognition rate, especially in the case of pose variations, even though there are only 2D still images in the testing. In the last few years, some works addressed face recognition from dynamic sequences of 3D face scans as well like in [6], where Sun et al. proposed a 4D-HMM based approach. In this work, a 3D dynamic spatio-temporal face recognition framework is derived by computing a local descriptor based on the curvature values at vertices of 3D faces. Spatial and temporal HMM are used for the recognition process, using 22 landmarks manually annotated and tracked over time. As an important achievement of this work, it is also evidenced that 3D face dynamics provides better results than 2D videos and 3D static scans.

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2016.03.013
0031-3203/© 2016 Elsevier Ltd. All rights reserved.

Subspace representation for dynamic facial information either for image sets or for image sequences (videos) showed a great success. Shigenaka et al. [10] proposed a Grassmann distance mutual subspace method (GD-MSM) and Grassmann Kernel Support Vector Machine (GK-SVM) comparison study for the face recognition problem from a mobile 2D video database. In [11], Lui et al. proposed a geodesic distance based algorithm for face recognition from 2D image sets. Turaga et al. [12] presented a statistical method for video based face recognition. These methods use subspace-based models and tools from Riemannian geometry of the Grassmann manifold. Intrinsic and extrinsic statistics are derived for maximum-likelihood classification applications. More recently, Huang et al. [13] proposed learning projection distance on Grassmann manifold for face recognition from image sets. In this work, an improved recognition is obtained by representing every image set using a Gaussian distribution over the manifold.

Sparse representation and dictionary learning attracted a lot of attention recently, due to their success in many computer vision problems. In [14], a sparse coding framework was presented for face recognition from still images. In this work, Wright et al. showed that using sparse coding the role of feature extraction on the performance is not so important, and the sparse coding is more tolerant with face occlusion. Yang et al. [15] proposed a robust sparse coding (RSC) approach for face recognition. In this work, the sparse coding problem is solved as a constrained robust regression, which makes the recognition more robust against occlusion, change of lighting and expression variation in still images. Elhamifar et al. [16] presented the Sparse Subspace Clustering (SSC) algorithm that classifies linear subspaces after finding their sparse coding. A generalization of sparse coding and dictionary learning was proposed by Xie et al. [17], which permits its application on subspace data representations that do not have a linear structure, like the Riemannian manifold. Mapping points from a non-linear manifold to tangent spaces shows good classification results on texture and medical images' classification.

In [18], Harandi et al. proposed an extrinsic solution to combine sparse coding and dictionary learning with nonlinear subspaces, like the Grassmann manifold. Embedding the Grassmann manifold into the symmetric matrices' sub-manifold makes the sparse coding on the induced manifold possible, faster, and more coherent than intrinsic embedding on one or more tangent spaces. Application to 2D video face datasets shows the efficiency of this approach against other learning solutions.

2. Methodology and contributions

In this paper, we investigate the contribution of 3D face dynamics in face recognition. To this end, after a preprocessing step, we compute surface curvature from each 3D static mesh of a sequence, and project it to a 2D map (call edcurvature map). A sequence of curvature maps is then cast to a matrix form by re shaping the 2D maps to column vectors, Singular Value Decomposition (SVD) is used to reduce the subspace spanned by the matrix to that of the first k-singular-vectors, which in turn is regarded as a point on a Grassmann manifold. Recognition using extrinsic methods based on sparse coding and dictionary learning on the manifold achieved the best performance. An overview of the proposed approach is shown in Fig. 1.

In summary, the main contributions of this paper are:

- A fully automatic and computationally cheap face recognition approach using 4D data. To the best of our knowledge, this is the first study in the literature, which brings the subspace modeling methodology with advanced geometric and learning tools to 3D face sequences.
- An in-depth investigation of the contribution of the 3D shape dynamics to face recognition.
- An extensive experimental analysis, involving the BU-4DFE dataset and three classification schemes based on intrinsic and extrinsic methods on the manifold.

The rest of the paper is organized as follows: in Section 3, the methodology of modeling 4D faces on Grassmann manifold as well as essential elements on the geometry of these manifolds is presented; Section 4 discusses sparse representation and dictionary learning on the Grassmann manifold; our 3D dynamic face recognition framework is presented in Section 5; Experimental results and their discussion are given in Section 6; finally, our conclusions and future work are drawn in Section 7.

3. Modeling sequences of 3D faces on Grassmann manifold

The idea of modeling multiple-instances of visual data, like set of images or video sequences, as linear subspaces for classification and recognition tasks has revealed its efficiency in many computer vision problems [12,19,20]. This compact low-dimensional data representation has the main advantage in its robustness against noise or missing parts in the original data. Besides, the availability of computational tools from differential geometry makes working on non-linear data (e.g., the space of k-dimensional subspaces) possible, and allows managing the non-Euclidean nature of these spaces. Accordingly, in this work, we adopt the subspace representation solution for analyzing 4D facial sequences. To our knowledge, this is one of the first investigations on modeling the temporal evolution of 3D facial shapes with application to face recognition. Studying the effects of these two aspects together is still an open problem in computer vision applications.

In the remaining of this section, we will describe the static 3D shape representation using mean curvature computed on 3D facial surfaces as well as the associated subspace representation to capture their temporal dynamics (Section 3.1). In addition, since the subspace learning approach that we propose lies on the Grassmann manifold, we will also recall essential background on its geometry, and related definitions including metrics and distances (Section 3.2) and sample mean computation (Section 3.3).

3.1. Static and dynamic 3D shape representation

In the proposed solution, we consider 3D scans of the face acquired continuously via a dynamic 3D scanner (3D plus time, also called 4D), thus producing a temporal 3D sequence with the dynamic evolution of the 3D face. Using these data, the proposed approach is designed to exploit the spatio-temporal information. To achieve this goal, a subspace modeling technique is applied as follows: (i) the 3D scans are preprocessed by cropping the facial region from the rest of the scan, then pose normalization, denoising via smoothing, and holes filling are performed; (ii) the mean curvature on 3D surfaces is computed, so that a flow of curvature-maps is produced by projection; (iii) the k-SVD orthogonalization procedure is applied to subsequences of the curvature-maps, so as to obtain an orthonormal basis spanning an optimized subspace. This subspace represents an element on the Grassmannian manifold $\mathcal{G}_k(\mathbb{R}^n)$, being n the dimension of curvature maps.

The shape information of every 3D scan is captured first by computing, as 3D local descriptor, the mean curvature $H=(k_1+k_2)/2$, where k_1 and k_2 are the two principle curvatures. The mean curvature values are computed at every vertex, then they are visualized and saved as a 2D map using a blue-red color

Download English Version:

https://daneshyari.com/en/article/530032

Download Persian Version:

https://daneshyari.com/article/530032

Daneshyari.com