Pattern Recognition 47 (2014) 659–671

Contents lists available at ScienceDirect

Pattern Recognition

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

Continuous Generalized Procrustes analysis

Laura Igual [a,b,*], Xavier Perez-Sala [b,c,d], Sergio Escalera [a,b], Cecilio Angulo [d], Fernando De la Torre [e]

[a] Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain
[b] Computer Vision Center, Universitat Autònoma de Barcelona, Building O, Barcelona, Spain
[c] Fundació Privada Sant Antoni Abat, Rambla de l'Exposició, 59-69, Vilanova i la Geltrú, Spain
[d] Universitat Politècnica de Catalunya, Av. Víctor Balguer 1, Vilanova i la Geltrú, Spain
[e] Carnegie Mellon University, Robotics Institute, 5000 Forbes Avenue, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 23 July 2012
Received in revised form
23 July 2013
Accepted 3 August 2013
Available online 13 August 2013

Keywords:
Procrustes analysis
2D shape model
Continuous approach

ABSTRACT

Two-dimensional shape models have been successfully applied to solve many problems in computer vision, such as object tracking, recognition, and segmentation. Typically, 2D shape models are learned from a discrete set of image landmarks (corresponding to projection of 3D points of an object), after applying Generalized Procustes Analysis (GPA) to remove 2D rigid transformations. However, the standard GPA process suffers from three main limitations. Firstly, the 2D training samples do not necessarily cover a uniform sampling of all the 3D transformations of an object. This can bias the estimate of the shape model. Secondly, it can be computationally expensive to learn the shape model by sampling 3D transformations. Thirdly, standard GPA methods use only one reference shape, which can might be insufficient to capture large structural variability of some objects.

To address these drawbacks, this paper proposes continuous generalized Procrustes analysis (CGPA). CGPA uses a continuous formulation that avoids the need to generate 2D projections from all the rigid 3D transformations. It builds an efficient (in space and time) non-biased 2D shape model from a set of 3D model of objects. A major challenge in CGPA is the need to integrate over the space of 3D rotations, especially when the rotations are parameterized with Euler angles. To address this problem, we introduce the use of the Haar measure. Finally, we extended CGPA to incorporate several reference shapes. Experimental results on synthetic and real experiments show the benefits of CGPA over GPA.

1. Introduction

Procrustes analysis (PA) [1–3] is a form of statistical shape analysis used to analyze the distribution of a set of shapes. Given two shapes PA “superimposes” both shapes by optimally translating, rotating and scaling one shape towards the other. If more than two shapes are registered, the problem is typically known as generalized Procrustes analysis (GPA). GPA has been typically used in computer vision as a first step to build 2D models of shape or appearance of objects. These 2D models have been applied to solve problems such as object recognition [4,5], facial feature detection and tracking [6,7] and image segmentation [8,9]. In particular, Point distribution models (PDMs) and active shape models (ASMs) [11] are among the most popular techniques to learn 2D objects models. PDMs and ASMs build the shape models from a 2D training set of image landmarks. In PDMs and ASMs, first GPA is used to remove rigid transformations and, then principal component analysis (PCA) is applied to construct a subspace that models the variation of the normalized shapes [11].

Fig. 1 (left) illustrates the GPA process of building shape models for PDM or ASM: given a set of 2D views of one or several 3D rigid or non-rigid objects under several configurations, the shape of the object is represented by several landmarks that are consistently labeled across view-points. Observe that if the object is rigid and the projection is orthographic, all views can be represented using a three-dimensional subspace [10]. Given the set of shapes (2D projections across views, objects or non-rigid transformations of 3D objects), GPA aligns the shapes using a rigid transformation (e.g., Euclidean or affine) to a 2D reference shape such that it minimizes the least-squares error. Although GPA has been extensively used, it suffers from three main limitations when modeling non-rigid transformations of a 3D object or a class of 3D objects: (i) 2D training samples do not necessarily cover a uniform sampling of all 3D transformations of an object, thereby biasing the estimate of the 2D models towards some particular

* Corresponding author at: Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain. Tel.: +34 934 020 854; fax: +34 934 021 601.
E-mail addresses: ligual@ub.edu, lauraigual@gmail.com (L. Igual), xavier.perez-sala@upc.edu, xavips@gmail.com (X. Perez-Sala), sergio@maia.ub.es (S. Escalera), cecilio.angulo@upc.edu (C. Angulo), ftorre@cs.cmu.edu (F. De la Torre).

0031-3203/$ - see front matter
http://dx.doi.org/10.1016/j.patcog.2013.08.006

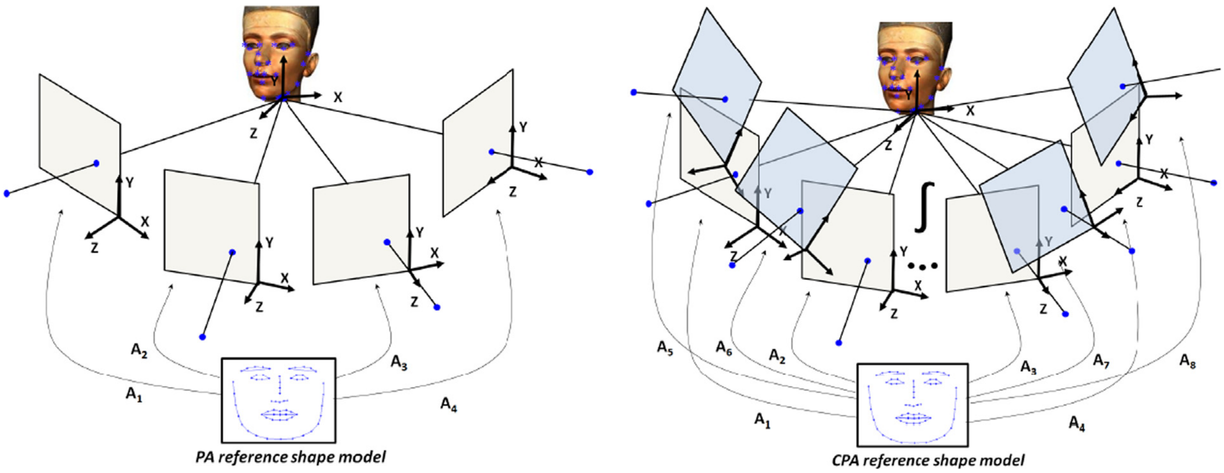

Fig. 1. Illustration of GPA (left) and CGPA (right) to construct 2D shape models from 3D objects.

configuration; (ii) it is computationally expensive to compute a rich set of 2D projections from all possible 3D transformations of a set of objects; and (iii) the large variability of the object class cannot necessarily be well registered with only one reference shape.

In order to deal with these limitations, we propose *continuous generalized Procrustes analysis* (CGPA). CGPA generalizes GPA using a *continuous* formulation that avoids the need to generate 2D projections from 3D configurations and uniformly covers the space of 3D transformations. Fig. 1 (right) illustrates the main idea behind CGPA, CGPA integrates over the space of 3D rotations avoiding the need to compute 2D projections. The continuous approach proposed in this paper is efficient in space and time, and is not biased to non-uniform sampling of the input space. A requirement of CGPA is to have access to a 3D mesh of several configurations of one or more 3D object, which is a realistic assumption in several computer vision problems. It is important to notice that building 2D models from 3D samples is a problem that has been relatively unexplored in computer vision [13,27].

A major challenge of the proposed work is to integrate 3D objects over the special orthogonal group in 3D: $S0(3)$. While there are many possible parameterizations of $S0(3)$, we have chosen Euler angles because it is easy to determine the relation between the rotation limits and the integration domain (unlike other parameterizations such as quaternions). However, Euler angles suffer from well-known problems such the lack of uniform integration over the space of rotations [12] or the gimbal lock effect. In this paper, to address these problems we use the Haar measure in the definition of the integral and uniformly integrate over the space of rotations. In addition, in some cases a simple mean in the case of GPA is not enough to model the variability of objects across view-points, and we propose a multi-reference CGPA by using several reference shapes. Experimental results in several synthetic and real experiments show the benefits of CGPA over GPA. A preliminary version of this work was presented in [13].

The rest of the document is organized as follows: Section 2 reviews previous work in GPA and functional data analysis (FDA), Section 3 gives the mathematical background necessary for CGPA formulation and Section 4 motivates and derives CGPA. Section 5 reports our experimental results and Section 6 presents the conclusions and outlines future lines of research. Finally, in Appendix A we review the GPA fitting algorithm.

2. Previous work

This section reviews previous work within the field of computer vision on Procrustes analysis and functional data analysis (FDA).

2.1. Generalized Procrustes analysis (GPA)

Let $\mathbf{D} = [(\mathbf{D}_1^{(2)})^T, \ldots, (\mathbf{D}_m^{(2)})^T]^T$ be a set of m shape samples that we wish to align. Note that the super-script $^{(2)}$ indicates that the shapes are 2D. Shape samples are represented as ℓ 2D landmarks embedded in an $\mathbb{R}^{2\times\ell}$ matrix $\mathbf{D}_i^{(2)}$ (see footnote[1] for notation)

$$\mathbf{D}_i^{(2)} = \begin{pmatrix} x_{i1} & \ldots & x_{i\ell} \\ y_{i1} & \ldots & y_{i\ell} \end{pmatrix}.$$

GPA optimizes over the 2D geometric transformation $\mathbf{T}_i$ (e.g., affine, Euclidean) that aligns each sample with respect to the reference shape, by minimizing the energy of the *reference-space model* (see Fig. 2 (right)) [14]

$$E_R(\mathbf{M}, \mathbf{A}) = \sum_{i=1}^{m} \|\mathbf{T}_i\mathbf{D}_i^{(2)} - \mathbf{M}\|_F^2, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{2\times\ell}$ represents the reference shape, and $\mathbf{T}_i$ in $\mathbf{T} = [\mathbf{T}_1^T, \cdots, \mathbf{T}_m^T]^T \in \mathbb{R}^{2n\times 2}$ corresponds to the rigid transformation for the shape sample $\mathbf{D}_i^{(2)}$. GPA can also be optimized using the *data-space model* (see Fig. 2 (left)) in the following way [14]:

$$E_D(\mathbf{M}, \mathbf{A}) = \sum_{i=1}^{m} \|\mathbf{D}_i^{(2)} - \mathbf{A}_i\mathbf{M}\|_F^2, \tag{2}$$

where $\mathbf{A}_i$ is the inverse transformation of $\mathbf{T}_i$ and $\mathbf{A} = [\mathbf{A}_1^T, \ldots, \mathbf{A}_m^T]^T \in \mathbb{R}^{2n\times 2}$ corresponds to the rigid transformation for the reference shape $\mathbf{M}$.

Recall that the error function of the reference-space model minimizes the difference between the reference shape and the registered shape data; in the data-space model, the error function compares the observed shape points with the transformed reference shape, i.e., shape points predicted by the model and based on the notion of average shape [15]. This difference between the two models leads to different properties. Since the reference-space cost is a sum of squares and it is linear in the optimization parameters, it can be optimized via alternated least square methods. In contrast, the data-space cost is a bilinear problem and non-convex (in general). If there are no missing data, the data-space model can be solved using singular value decomposition (SVD). A major advantage of the data-space model is that it is

[1] $\mathbb{N}$ and $\mathbb{R}$ denote the set of natural and real numbers, respectively, and $\mathbb{R}^d$ denotes the set of real vectors of dimension d. We assume that $m, d, l, n, p, i \in \mathbb{N}$. A bold capital letter denotes a matrix, $\mathbf{D}$; a bold lower-case letter a column vector, $\mathbf{d}$. $\mathbf{D}_i$ represents the ith block matrix of the matrix $\mathbf{D}$. All non-bold letters denote scalar variables. $\|\mathbf{D}\|_F^2 = \text{Tr}(\mathbf{D}^T\mathbf{D})$ designates the square of the Frobenius norm of a matrix. The set operation $\mathbf{\Omega}\backslash\mathbf{\Gamma}$ stands for the set difference of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$. $\nabla_{\mathbf{u}}F$ is the gradient operator with respect to u of the function F.