Computer Vision and Image Understanding 120 (2014) 59–69

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Real-time facial shape recovery from a single image under general, unknown lighting by rank relaxation ☆

Minsik Lee *, Chong-Ho Choi

Department of Electrical and Computer Engineering, and ASRI, Seoul National University, Republic of Korea

ARTICLE INFO

Article history:
Received 23 August 2012
Accepted 23 December 2013
Available online 31 December 2013

Keywords:
Facial shape recovery
Shape from shading
Statistical face model
Bilinear model
Rank relaxation

ABSTRACT

Statistical shape from shading under general light conditions can be thought of as a parameter-fitting problem to a bilinear model. Here, the parameters are personal attributes and light conditions. Parameters of a bilinear model are usually estimated using the alternating least squares method with a computational complexity of $O((n_s + n_\phi)^2 n_p)$, where n_s, n_ϕ, and n_p are the dimensions of the light conditions, personal attributes, and face image features, respectively, for each iteration. In this paper, we propose an alternative algorithm with a computational complexity of $O(n_s n_\phi)$ for each iteration. Only the initial step requires a computational complexity of $O(n_s n_\phi n_p)$. This can be accomplished by reformulating the problem to that of a linear least squares problem, with a search space limited to a set of rank-one matrices. The rank-one condition is relaxed to obtain a possibly full-rank matrix. The algorithm then finds the best rank-one approximation of that matrix. By the Eckart–Young theorem, the best approximation is the outer product of the left and right singular vectors corresponding to the largest singular value. Since only this pair of singular vectors is needed, it is better to use the power iteration method, which has a computational complexity of $O(n_s n_\phi)$ for each iteration, than calculating the full singular value decomposition. The proposed method provides accurate reconstruction results and takes approximately 45 ms on a PC, which is adequate for real-time applications.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Over the past two decades, model-based approaches in shape from shading (SFS) [1–11] have successfully accommodated the need for an efficient and accurate 3-D reconstruction method without any specialized hardware device. Unlike the early SFS methods [12], these approaches impose model constraints to rule out the ambiguity [13] caused by the ill-posedness of the SFS problems and guarantee a unique solution. Except for a few alternatives, such as the method based on light attenuation [14], the model-based approaches have formed the mainstream of the 3-D reconstruction methods that use a single image, and they are applied in many areas [5,15,16].

However, there are some drawbacks to the model-based approaches. Often they are either too computationally expensive or too restrictive in their application. The most prominent work that falls into the former case is the morphable model (MM) [2], which synthesizes a 3-D face shape by minimizing the difference between a face image and a pre-built statistical model of face shape and texture. The MM is an analysis-by-synthesis approach and is actively studied in the area of computer vision [5,15]. The MM-based methods can simultaneously estimate the pose and appearance of a face in an image, but they usually require heavy computation (e.g., they can take a few *minutes* to reconstruct a face). Many works belong to the group that are too restrictive, such as the work in [1], which was the first attempt to utilize a model-based approach for 3-D face reconstruction, but failed to achieve high reconstruction accuracy. These methods are applied to fixed-pose images under a single, known light source or a fixed light condition and are relatively faster than MM. However, they are too restrictive to use and some of them perform poorly. Reiter et al. [3] and Lei et al. [6] both find a linear mapping between image and depth, based on the features extracted using canonical correlation analysis (CCA) [17]. Lei et al. used near infrared images as input images to mitigate the effects of illumination changes. They applied N-mode singular value decomposition (SVD) [18], which is a generalization of SVD for multi-dimensional data, as an alternative feature extraction method. These methods cannot handle light variations and so require special light conditions. A statistical model of azimuthal equidistant projection of surface normals is proposed in [7] and is combined with Lambert's law for optimization. It requires the light source to be known and is vulnerable to large albedo variations. A non-stationary stochastic filtering framework [19] is applied to estimate the albedo and illuminance of a face image in [8] as a

☆ This paper has been recommended for acceptance by Edwin R. Hancock.

* Corresponding author.
E-mail addresses: mlee.paper@gmail.com (M. Lee), chchoi@csl.snu.ac.kr (C.-H. Choi).

1077-3142/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2013.12.010

preprocessing step for SFS algorithms. However, the original SFS algorithms [12] suffer from the ambiguity mentioned above and require that there be a single light source in an image. Kemelmacher-Shlizerman and Basri [9] proposed a novel approach that adjusts the details of a reference face shape to fit a new image. Despite the advantage that it requires only a single reference shape, its performance is poor.

To overcome these disadvantages, we proposed a tensor-based optimization approach in [10]. We formulated a bilinear model [20–23] of spherical harmonics [24] by using tensor algebra techniques, such as N-mode SVD, to handle general, unknown light conditions. The method offers high reconstruction accuracy, and takes less than half a second on a PC. As is usual for bilinear models, the fitting problem was solved by the alternating least squares (ALS) method [25]. Since each iteration is a linear least square problem in ALS, the computational complexity for each iteration is in the order of $O((n_s+n_\phi)^2 n_p)$, where n_s, n_ϕ, and n_p are the dimensions of light conditions, personal attributes, and face image features, respectively.

In this paper, we propose a novel algorithm with a computational complexity $O(n_s n_\phi)$ for each iteration and $O(n_s n_\phi n_p)$ for the initial step. This is accomplished by reformulating the fitting problem, which becomes a linear least squares problem of a rank-one matrix. By relaxing the rank constraint on the search space, the solution matrix, which may be full rank, can be easily found, and the best rank-one approximation is calculated afterwards. This is somewhat similar to the semidefinite relaxation techniques [26], except that the search space is not limited to the set of symmetric matrices. According to the Eckart–Young theorem [27], the best rank-one approximation is given by the outer product of the right and left singular vectors corresponding to the largest singular value. Since we only need this pair of singular vectors, finding only the largest singular value and its corresponding singular vectors is preferable to calculating the full SVD. This can be done by the power iteration method [28], which is a numerical algorithm that finds the greatest absolute eigenvalue or singular value of a matrix. The computational complexity of the power iteration method is in the order of $O(n_s n_\phi)$ for each iteration, which makes the algorithm more efficient. Only the initial step, which solves the least squares problem by relaxing the constraint, requires a computational complexity of $O(n_s n_\phi n_p)$.

After finding the solution, a linear mapping is applied to calculate the corresponding depth map. The proposed method provides accurate reconstruction results and performance similar to that of the method in [10], although the computation time of the proposed method is much lower. The average reconstruction time is approximately 45 ms on a PC, which is sufficient for real-time applications.

The remainder of the paper is organized as follows. The basics of tensor algebra are introduced in Section 2, and the proposed algorithm is explained in Section 3. The performance is evaluated in Section 4, and finally, the paper is concluded in Section 5.

2. Tensor algebra notations

In this paper, we use tensor algebra and explanations for standard operations, such as mode-k flattening $\mathbf{A}_{(k)}$, mode-k product $\mathcal{A}\times_k\mathbf{M}$, and N-mode SVD, can be found in [6,18]. In this section, we explain non-standard operations that we use in the rest of the paper.

The reshaping of a tensor is denoted as

$$\mathcal{B} = R(\mathcal{A}; n'_1, n'_2, \ldots, n'_{N'}),$$

where $\mathcal{A} \in \mathbb{R}^{n_1\times\cdots\times n_N}, \mathcal{B} \in \mathbb{R}^{n'_1\times\cdots\times n'_{N'}}$, and $\prod_l^N n_l = \prod_l^{N'} n'_l$. Through the reshaping, $\mathcal{B}_{i'_1\ldots i'_{N'}} = \mathcal{A}_{i_1\ldots i_N}$ for

$$1+\sum_{j=1}^{N'}(i'_j-1)\prod_{l<j}n'_l = 1+\sum_{j=1}^{N}(i_j-1)\prod_{l<j}n_l.$$

For Nth-order tensor $\mathcal{A}$, the mode-1 or mode-N flattenings can be defined by the reshaping operator.

$$\mathbf{A}_{(1)} = R\left(\mathcal{A}; n_1, \prod_{l>1} n_l\right),$$

$$\mathbf{A}_{(N)} = R\left(\mathcal{A}; \prod_{l<N} n_l, n_N\right)^T.$$

The vectorization operator can be defined in the following way.

$$\mathrm{vec}(\mathcal{A}) = R\left(\mathcal{A}; \prod_l n_l\right).$$

The mode-k product of a tensor can be expressed in terms of the reshaping operator and the Kronecker product. Let

$$\mathcal{B} = \mathcal{A}\times_1\mathbf{W}_1\times_2\mathbf{W}_2,$$

where $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3\times\cdots\times n_N}, \mathcal{B} \in \mathbb{R}^{n'_1\times n'_2\times n_3\times\cdots\times n_N}, \mathbf{W}_1 \in \mathbb{R}^{n'_1\times n_1}$, and $\mathbf{W}_2 \in \mathbb{R}^{n'_2\times n_2}$. Then it can be shown after some manipulation that the following relation is satisfied.

$$R\left(\mathcal{B}; n'_1 n'_2, \prod_{2<l} n_l\right) = (\mathbf{W}_2\otimes\mathbf{W}_1)R\left(\mathcal{A}; n_1 n_2, \prod_{2<l} n_l\right). \quad (1)$$

3. Rank relaxation for facial shape recovery

In this section, we explain the detailed procedure of our algorithm. We model a face image as a bilinear model having two parameter vectors, one for personal attributes and the other for light conditions, and the bases of the model are trained based on samples beforehand. The parameter vector for personal attributes contains abstract information about the appearance of an individual face, which is independent of light condition. After estimating the parameter vectors from an input image using our efficient rank relaxation technique, we perform a linear mapping from the personal attribute vector to the corresponding 3D depth. This linear mapping is also trained beforehand with a set of face images and depth samples. The reconstruction phase is simple enough to allow real-time reconstruction on a regular PC, even though the algorithm achieves the state-of-the-art reconstruction accuracy.

3.1. Bilinear model for face images

A vector corresponding to a face image in a fixed pose under general lighting can be expressed as

$$\mathbf{I} = \mathbf{H}\mathbf{s},$$

where $\mathbf{I} \in \mathbb{R}^{m_p}$, $\mathbf{H} \in \mathbb{R}^{m_p\times n_s}$, and $\mathbf{s} \in \mathbb{R}^{n_s}$ are the image vector, basis matrix, and light condition vector, respectively. Let us assume that $\mathbf{H}$ is expressed as a linear model of the personal attribute vector $\boldsymbol{\phi} \in \mathbb{R}^{n_\phi}$ as

$$\mathbf{H} = \mathcal{H}\times_3\boldsymbol{\phi}^T, \quad (2)$$

where $\mathcal{H} \in \mathbb{R}^{m_p\times n_s\times n_\phi}$ is a third-order tensor. Then, the overall image model is expressed as

$$\mathbf{I} = \left(\mathcal{H}\times_3\boldsymbol{\phi}^T\right)\mathbf{s} = \mathcal{H}\times_2\mathbf{s}^T\times_3\boldsymbol{\phi}^T.$$

This is a bilinear model [20] of the light condition and the personal attribute.

In order to find the personal attribute vector $\boldsymbol{\phi}$ corresponding to an image $\mathbf{I}$, the following optimization problem should be solved.

Download English Version:

https://daneshyari.com/en/article/525622

Download Persian Version:

https://daneshyari.com/article/525622

Daneshyari.com