Automatic Construction Of Robust Spherical Harmonic Subspaces

Patrick Snape, Yannis Panagakis, Stefanos Zafeiriou
Department of Computing, Imperial College London.

The recovery of dense 3D shape from images is a highly underdetermined problem with many practical uses. Despite the increased availability of accurate 3D scanning technology, traditional colour photographs still outnumber 3D data by many orders of magnitude. In this paper, we recover 3D shapes from a large collection of unconstrained images by using shading cues as is typical in algorithms such as Photometric Stereo (PS) and Shape-from-Shading (SFS). In particular, we demonstrate that by using coarse piecewise warping algorithms we can build a shape model that has applications in a number of areas.

Given a single image, 3D recovery under arbitrary illumination is extremely challenging. For this reason, we concentrate on jointly recovering shape from a large collection of images. In particular, we seek to exploit the redundant information across the images by imposing a low rank constraint on the recovered shape representation. To demonstrate this, we use facial images due to their abundance and their ability to be represented by a low-dimensional subspace [4]. To recover shape from the images, we exploit this low-dimensional subspace and perform *class-specific uncalibrated photometric stereo*. This depends on two key assumptions about the images (1) that the objects reflectance properties are approximately lambertian (2) that the class of objects within the images (i.e. faces) can be compactly represented by a low-dimensional subspace. Both of these assumptions are valid for faces: first-order spherical harmonic functions are able to describe up to 95% of facial illumination variance [4] and the low-dimensional properties of faces has been exploited in works such as 3D Morphable Models (3DMM)[2] and Active Appearance Models (AAM) [6].

Given multiple images of a single object under unknown varying illumination, uncalibrated PS can be performed to recover a representation of shape. Assuming the object is convex, first-order spherical harmonics (SH) can be recovered which can be directly related to the objects surface normals [1, 4]. First order SH can be represented by a low-dimensional linear subspace, and thus the recovery of the SH from the image pixels is performed via a matrix decomposition:

$$\mathbf{X} = \mathbf{B}\mathbf{L} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{d \times n}$ is the matrix of observations, each of the n columns represents a vectorised image of the object with d total pixels, $\mathbf{B} \in \mathbb{R}^{d \times 4}$ contains the first order SH basis images and $\mathbf{L} \in \mathbb{R}^{4 \times n}$ is the matrix of lighting coefficients.

We generalise the classical uncalibrated PS problem to a specific class of objects, namely faces. In this case, we expect as input a collection of unconstrained images of faces, such as those available to download from the Internet. Given the unfiltered nature of these images, it is probable that there will be a large variety of illumination variance across all of the images. Thus, given that both faces and SH can be expressed by low-dimensional subspaces, we seek to perform a matrix factorisation that will yield a statistical model of SH harmonics for faces. Surface normals can be recovered from the SH and then integrated to recover dense 3D shape [3]. More formally, we seek to perform the decomposition

$$\mathbf{X} = \mathbf{B}(\mathbf{L} * \mathbf{C}) \tag{2}$$

where $\mathbf{B} \in \mathbb{R}^{d \times 4k}$ is the first order SH linear basis, $\mathbf{L} \in \mathbb{R}^{4 \times n}$ is the matrix of first order SH lighting coefficients, $\mathbf{C} \in \mathbb{R}^{k \times n}$ is the matrix of shape coefficients and $(\cdot * \cdot)$ denotes the Khatri-Rao product. Here k denotes the number of shape coefficients in the model. In order to calculate k, we cast (2) as a low-rank factorisation problem. Formally we propose to solve the following non-convex optimisation problem:

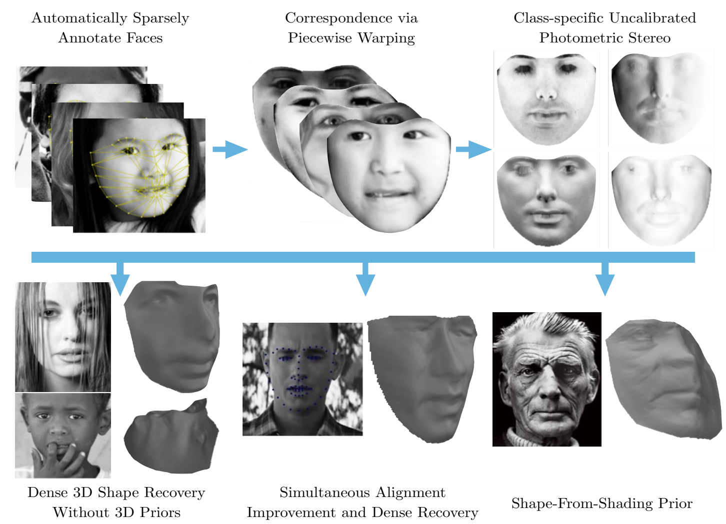

Figure 1: Illustrates the pipeline for construction and the uses of the model. The first row shows the rough model construction pipeline. The second row shows three uses demonstrated in the paper of the SH model.

$$\begin{aligned} \underset{\mathbf{A},\mathbf{E},\mathbf{B},\mathbf{L},\mathbf{C}}{\text{argmin}} \quad & \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + \frac{\mu}{2}\|\mathbf{A} - \mathbf{B}(\mathbf{L} * \mathbf{C})\|_F^2 \\ \text{subject to} \quad & \mathbf{X} = \mathbf{A} + \mathbf{E},\ \mathbf{B}^T\mathbf{B} = \mathbf{I}. \end{aligned} \tag{3}$$

Although the above problem is non-convex, an accurate solution can be obtained by employing the Alternating Directions Method (ADM). The specifics of this optimisation are discussed in detail in the paper.

To form the data matrix $\mathbf{X}$, the input images must be in correspondance. To tackle this alignment problem, we propose using a coarse, but efficient, piecewise warping algorithm such as Piecewise Affine. Although coarse, piecewise warping functions are controlled by a sparse set of keypoints, which enables the use of very efficient facial alignment methodologies. Due to this, our model building algorithm is very efficient, taking just 12 minutes to build a model from over 2000 images. We also show that, due to the use of sparse landmarks for alignment, our model can be used in a number of areas including: dense 3D shape recovery, simultaneous facial alignment and SH harmonic recovery and as a statistical prior in SFS. This is in contrast to other state-of-the-art techniques that utilise more accurate, but time consuming correspondence methods such as optical flow [5].

In conclusion, we demonstrate that dense facial shape can be efficiently recovered from a large collection of unconstrained facial images without prior knowledge of 3D facial shape.

[1] R Basri and D W Jacobs. Lambertian reflectance and linear subspaces. *IEEE T-PAMI*, 25(2):218–233, 2003.

[2] V Blanz and T Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, pages 187–194, 1999.

[3] R Frankot and R Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE T-PAMI*, 10(4):439–451, 1988.

[4] A S Georghiades, P N Belhumeur, and D J Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE T-PAMI*, 23(6):643–660, 2001.

[5] I. Kemelmacher-Shlizerman. Internet based morphable model. In *ICCV*, pages 3256–3263, Dec 2013. doi: 10.1109/ICCV.2013.404.

[6] I Matthews and S Baker. Active appearance models revisited. *IJCV*, 60 (2):135–164, 2004.

This is an extended abstract. The full paper is available at the CV Foundation webpage.