A Linear Approach of 3D Face Shape and Texture Recovery using a 3D Morphable Model

Oswald Aldrian

William A. P. Smith

Department of Computer Science
The University of York

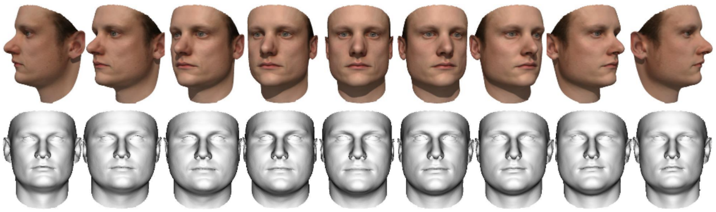

Figure 1: Top row shows 2D renderings of face No. 4 in different rotation angles ($-70^\circ : 70^\circ$). Second row shows the corresponding 3D shape reconstructions.

In this paper, we present a robust and efficient method to statistically recover the full 3D shape and texture of faces from single 2D images. We separate shape and texture recovery into two linear problems. For shape recovery, we learn empirically the generalization error of a 3D morphable model [1] using out-of-sample data. We use this to predict the 2D variance associated with a sparse set of 2D feature points. This knowledge is incorporated into a parameter-free probabilistic framework which allows 3D shape recovery of a face in an arbitrary pose in a single step. Under the assumption of diffuse-only reflectance, we also show how photometric invariants can be used to recover texture parameters in an illumination insensitive manner. We present empirical results with comparison to the state-of-the-art analysis-by-synthesis methods [4] and show an application of our approach to adjusting the pose of subjects in oil paintings. The closest work in spirit to what we present here is that of Romdhani *et al.* [5]. They present linear solutions to computing an incremental update to the shape and texture parameters given dense measurements of residual errors provided by optical flow. However, their iterative approach requires nonlinear optimisation of pose parameters and illumination terms. In this paper, we propose a shape estimation method which incorporates an empirically measured model of variance into a linear objective function. By doing so, we do not need a weight factor that trades off between the model and the data. We recover texture using a linear error based on a photometric invariant which is unaffected by illumination conditions. At the expense of assuming diffuse-only reflectance and that the location of sparse feature points are known [2], we are able to accurately fit a 3DMM at greatly reduced computational expense whilst closely competing with the accuracy of much more sophisticated methods. We present the first quantitative comparative evaluation of 3DMM fitting algorithms.

In order to obtain a linear solution for shape recovery, we decompose the problem into two steps which can be iterated and interleaved: 1. estimation of a camera projection matrix using known 3D-2D correspondences, and 2. estimation of 3D shape parameters using a known camera projection matrix. With this to hand, shape parameters can be recovered using only matrix multiplications. By using the recovered shape to re-estimate the camera matrix, we can iterate the process which typically converges in only 3 iterations. We model two sources of variance which can be used to explain the difference between observed and modelled feature point positions in the image. By having an explicit model of this variance, we negate the need for an ad hoc regularisation weight parameter. The first source of variance is the generalisation error of the morphable model. This describes how feature points deviate from their true position in 3D when the optimal model parameters are used to describe a face. The second source of variance is the 2D pixel noise, this is related to the accuracy with which the feature points can be marked up or detected in 2D. The total variance of a feature point is the sum of the 3D variance projected to 2D and the 2D variance.

Our statistical surface texture model captures variations in diffuse albedo. This forms one parameter of a number of possible parametric reflectance models (e.g. Phong) which in turn determines the appearance of a face [1]. This is the approach used in analysis-by-synthesis model fitting. We take a different approach. By making some assumptions about the surface reflectance and illumination, we are able to arrive at a photometric invariant which can be measured directly from the image and used to fit the texture model in an illumination-insensitive manner. Moreover, the resulting solution is linear in terms of the observed image intensities and can therefore be executed efficiently. We make the assumption that surface reflectance is diffuse only and that illumination is provided by any combination of directional and ambient white sources. Our experiments are based on the *Basel Face Model* [4]. The model is accompanied by 10 out of sample faces. Each face is rendered in 9 poses and 3 lighting conditions per pose giving 270 renderings in total. We use a subset of the Farkas feature points [3] to reconstruct face shape. Figure (1) shows shape reconstructions of a sample face in 9 different pose angles. In figure (2) we show the shape and texture reconstruction for face No. 7 in frontal pose under the 3 different lighting conditions. Note the presence of cast shadows in the image which do not effect our method. Our experiments also showed that the number of feature points alone is not the significant factor. On average, the shape reconstruction error is lower for close to profile views compared to front views, even though nearly half as many feature points are visible. This implies that the pose of a face effects the information content in a feature point observation.

Figure 2: Top row shows renderings of Face No. 7 in frontal pose and 3 different illumination conditions. Second row shows the corresponding shape and texture reconstructions.

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

[2] V. Blanz, A. Mehl, T. Vetter, and H-P. Seidel. A statistical method for robust 3d surface reconstruction from sparse data. In *Proc. 3DPVT*, pages 293–300, 2004.

[3] L. Farkas. *Anthropometry of the Head and Face*. Raven Press, New York, 1994.

[4] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. In *Proc. IEEE Intl. Conf. on Advanced Video and Signal based Surveillance*, 2009.

[5] Sami Romdhani, Volker Blanz, and Thomas Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *in European Conference on Computer Vision*, pages 3–19, 2002.