Face Reconstruction in the Wild

Fangfang Li

June 18, 2018

Abstract

We address the problem of reconstructing 3D face models from large unstructured photo collections, obtained by Google image search or from personal photo collections in iPhoto. This problem is extremely challenging due to the high degree of variability in pose, illumination, facial expression, non-rigid changes in face shape and reflectance over time and occlusions. In light of this extreme variability, no single reconstruction can be consistent with all of the images. Instead, we define as the goal of reconstruction to recover a model that is locally consistent with the image set, each local region of the model is consistent with a large set of photos, resulting in a model that captures the dominant trends in the input data for different parts of the face. Our approach leverages multi-image shading, but unlike traditional photometric stereo approaches, allows for changes in viewpoint and shape. We optimize over pose, shape, and lighting in an iterative approach that seeks to minimize the rank of the transformed images. This approach produces high quality shape models for a wide range of celebrities from photos available on the Internet.

1. Introduction

An Internet image search for a celebrity yields t-housands of photos. Similarly, our personal photo collections contain thousands of photos of faces. In this paper we consider the problem of computing 3D face reconstructions from such collections.

2. Pose normalization

To account for variations in face orientation over the image set, we warp each image to a canonical, frontal pose. To this end, we first estimate pose by detecting fiducial points on the face, and use the positions of these fiducials from a template 3D face to recover 3D rotation, translation, and scale for each photo.

The relation between points q on the image and points on the template Q is given by

$$q = sRQ + t. \tag{1}$$

To recover s, R, and t, we first subtract the centroid from both point sets to get $p = q - \overline{q}$ and $P = Q - \overline{Q}$, then estimate a 2×3 linear transformation $A = pP^T(PP^T)^(-1)$ and translation $t = \overline{q} - A\overline{Q}$. To recover an estimate of the rotation and scale we let the third row of A be the cross product between the first two rows and by taking its SVD, $A' = UDV^T$, we estimate the closest rotation in terms of Frobenius norm $R = UV^T$. Two of the singular values of A' are identical [1], and this is our estimate of scale. We thenes timate the yaw, pitch and roll angles from the rotation matrix. Given the estimated pose we transform the template to the orientation of the face in the image, the image is back-projected onto the shape, and then a frontal view of the face is rendered [2]. This results in a collection of faces where every face is in approximately frontal position as can be seen in Fig. 1

Figure 1. Expression normalization by low-rank approximation. The first row shows the warped images, the 2nd row shows the the low rank approximated images. Note how the lighting is mostly preserved, but the facial expression is normalized.

3. Initial lighting and shape estimation

The local selection of images works as follows. For each point on the face we first calculate how well the images fit the initial shape estimate, we calculate the distance where is a vector representing the intensities of a pixel in all images (column of M), S_j is 4×1 and L is $n \times 4$. We then normalize the distance and choose a subset of images for which the distance is less than a threshold, making sure that the number of images is larger than 4 and that the condition number of $L_k \times 4$is not high (k represents the number of chosen images) to prevent degenerate lighting conditions [3]. The resulting set of images is then used to recover S_j by minimizing the following functional

$$min\,||M_m \times 1 - L_k \times 4S_j|\,| + s_j^T G S_j \qquad (2)$$

The first term represents the lighting consistency relation and the second term acts as a Tikhonov regularization term which avoids poor conditioning.

References

[1] K. S. Ira and S. M. Seitz. Face reconstruction in the wild. In *ICCV*, pages 1746–1753, 2011. 1

[2] Y. Lin, G. Medioni, and J. Choi. Accurate 3D face reconstruction from weakly calibrated wide baseline images with profile contours. In *CVPR*, pages 1490–1497, 2010. 1

[3] F. Liu, D. Zeng, Q. Zhao, and X. Liu. Joint face alignment and 3D face reconstruction. In *ECCV*, pages 545–560, 2016. 2