3D Computer Vision
Final Project Proposal

Sungmin Hong

April 9, 2015

1 Introduction

The main theme of the final project is reconstructing 3D face from multiple sources. The first part of the project follows a framework suggested in the 'Face reconstruction from the wild' [4]. The authors proposed the 3D face reconstruction method from large unstructured image collections, such as, images from google search or personal photo collections.

2 Algorithm Overview

Normalize Face Images

Since images from internet or personal collections are unstructured, we need to normalize each image to a canonical face, which is frontal view of a face. In [4], they used a fiducial-based approach to normalize faces. They first detect a face in the image by using the object detection method using the soft cascade method [1]. Fiducial points of a face, such as, corners of eyes, nostrils, tip of a nose, and etc. are extracted from the detected region by using the fiducial points detection method [2]. The matlab codes for the face detection method is provided by the matlab package and the fiducial points detection method is provided in the matlab Computer Vision package.

Let q and Q be the fiducials in 2D images and the 3D template shape. They assumed 2D points in each image are the weak perspective projection of 3D points on the template surface.

$$q = P_w Q = [sR|t]Q \tag{1}$$

To estimate P_w, they subtracted the centroid and calculate the pseudo-inverse of relative positions of 3D points.

$$\tilde{q} = q - \bar{q} \qquad \tilde{Q} = Q - \bar{Q} \tag{2}$$

$$A = \tilde{q}\tilde{Q}^T(\tilde{Q}\tilde{Q}^T)^{-1} \tag{3}$$

where $\bar{q}$ and $\bar{Q}$ are the mean of q and Q. And the translation t is defined as follows.

$$t = \bar{q} - A\tilde{Q} \tag{4}$$

e s and R are recovered by estimating the singular values of A. Since the R and s is 3D translation and scaling, we add a row, which is the cross product of A_1 and A_2, and do the singular value decomposition.

$$A' = \left[\frac{A}{A_1 \times A_2}\right] = U_A D_A V_A^T \tag{5}$$

Then the rotation $R = U_A V_A^T$ and the two identical singular value becomes s.

Initial Lighting and Shape Estimation

Let I_i be the i th canonical face image of size N and M is the concatenated matrix of K images. Then M is $N \times K$ matrix which contain all canonical face images with different lighting conditions. The uncalibrated photometric stereo reconstruction method using the singular value decomposition method [6] take rank-4 approximation to get $M = LS$, where $n \times 4$ matrix $L = U_M\sqrt{D_M}$ and $4 \times K$ matrix $S = \sqrt{D_M}V_M^T$. L and S represent the light source directions and the surface properties, which are albedo and the surface normals, respectively.

Surface Refinement using Local Shape Estimation

The surface S reconstructed from the rank-4 approximation lose details by nature. To overcome the over smoothing problem, the local patches of an image are used to update L and S. Thus the distance at pixel j is represented as follows,

$$d_j = |M_j - LS_j|^2 \tag{6}$$

where M_j is a $n \times 1$ vector representing the intensities of a pixel j in all K images. Then normalize the distance d_j and select $k > 4$ images under the threshold. With k iamges the surface structure at j th pixel is estimated by minimizing

$$\min_{S_j} ||M_{k\times 1} - L_{k\times 4}S_j|| + S_j^T G S_j \tag{7}$$

where $G = diag(-1, 1, 1, 1)$ for Tikhonov regularization.

Then S_j has a closed form solution,

$$S_j = (L_{k\times 4}^T L_{k\times 4} + G)^{-1} L^T M_{k\times 1} \tag{8}$$

Ambiguity Recovery

The rank-4 approximation of the surface properties and the lighting condition has 4×4 ambiguity since $M = LS$ can be $M = LA^{-1}AS$ with an invertible 4×4 matrix A. The ambiguity may lead to the surface to be non-integrable. The ambiguity can be resolved by minimizing the energy between the template surface and the estimated surface property.

$$\min_{A} ||S_t - AS||^2 \tag{9}$$

where A is 4×4 ambiguity and S_t is the surface property of the template shape. Then L is updated to $L' = LA^{-1}$ and $S' = AS$.

Integration

In theiry paper, they used the simple integration method, but it looks possible to use Chellappa's method [3]. Then we have an updated surface S_{t+1} and use the updated surface as a new template surface and iterate the algorithm until convergence.

Discussion

The proposed final project is about 3D face reconstruction from an unstructured image collection of the person. The framework follows the method proposed in [4]. The project will be extended to 3D face video reconstruction as suggested in [5] which reconstructs the 3D face in a single video frame if time allows. Also, the work may be extended to reconstruction of an unseen part of a face in an image or a video using matrix factorization method and surface registration method.

References

[1] Lubomir D. Bourdev and Jonathan Brandt. Robust Object Detection via Soft Cascade. In *CVPR (2)*, pages 236–243. IEEE Computer Society, 2005.

[2] Mark Everingham, Josef Sivic, and Andrew Zisserman. Hello! My name is... Buffy" – Automatic Naming of Characters in TV Video. In Mike J. Chantler, Robert B. Fisher, and Emanuele Trucco, editors, *BMVC*, pages 899–908. British Machine Vision Association, 2006.

[3] Robert T. Frankot and Rama Chellappa. A Method for Enforcing Integrability in Shape from Shading Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(4):439–451, 1988.

[4] Ira Kemelmacher-Shlizerman and Steven M. Seitz. Face reconstruction in the wild. In Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc J. Van Gool, editors, *ICCV*, pages 1746–1753. IEEE, 2011.

[5] Suwajanakorn S., Kemelmacher-Shilzerman I., and Seitz S. M. Total Moving Face Reconstruction. *ECCV*, 8692:796–812, 2014.

[6] Alan L. Yuille, Daniel Snow, Russell Epstein, and Peter N. Belhumeur. Determining Generative Models of Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability. *International Journal of Computer Vision*, 35(3):203–222, 1999.