# Unconstrained 3D Face Reconstruction

Joseph Roth, Yiying Tong, Xiaoming Liu
Department of Computer Science and Engineering. Michigan State University.

This paper presents an algorithm for "unconstrained" 3D face reconstruction from a 2D photo collection of face images of a subject captured under a diverse variation of poses, expressions, and illuminations, without meta data about the cameras, timing, or light conditions. The output of our algorithm is a true 3D face surface model represented as a watertight triangulated surface with albedo data or texture information. This is certainly a very challenging problem, as we do *not* have access to stereo imaging [8] or video [3]. Motivated by the success of the state-of-the-art method [6], we developed a novel photometric stereo-based method with two distinct novelties. First, working with a true 3D model allows us to enjoy the benefit of using images from all possible poses, including profiles, without warping them to a frontal image. Second, by leveraging emerging face alignment techniques and our *novel normal field-based Laplace editing*, a combination of landmark constraints and photometric stereo-based normals drives our surface reconstruction.

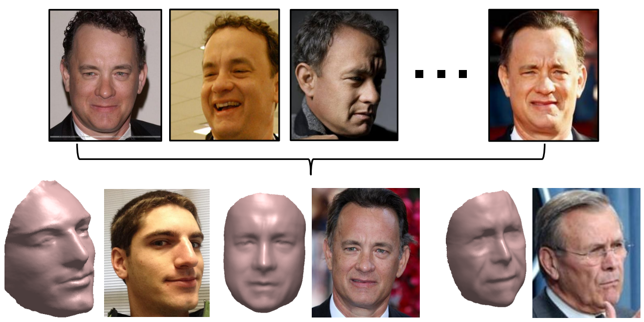

Figure 1: Given Tom Hanks' photo collection with pose, expression, and illumination variations, our system performs surface reconstruction, shown along with a real photo at the same viewpoint.

Obtaining a user-specific 3D face surface model is a well studied task in computer vision literature; as it has a variety of applications ranging from 3D-assisted face recognition [1], 3D expression recognition [9], facial animations [2], and more. Despite the emergence of 3D sensors to directly acquire 3D face models, accurately *reconstructing the 3D surface model from 2D images* is a long-standing computer vision problem. There is also no shortage of 2D images continuing to be captured today.

Prior techniques reconstruct from a range of different *constrained* inputs including: stereo imaging [8], photometric stereo [4], and video sequences [3]. These techniques are relatively easier because they can make assumptions about the face based on knowledge of the environment, camera positions, optical flow, or lighting. When the images are captured in an *unconstrained* fashion, there is one state-of-the-art approach using an impressive photometric stereo-based method to produce high-quality face models from photo collections [6], where the recovering of a locally consistent shape is intelligently achieved by using a different subset of images for each part of the face. However, there are still two limitations in [6]. One, mainly near-frontal images are selected to contribute to the reconstruction, while the consensus is that non-frontal, especially profile, images are highly useful for 3D reconstruction. Two, due to surface reconstruction on a 2D grid, a 2.5D height field, rather than a full 3D model, is produced.

Motivated by the state-of-the-art results and amendable limitations of [6], this paper proposes a novel approach to 3D face reconstruction. Our approach is also enabled by the recent explosion of face alignment techniques [5, 7], where the precision of 2D landmark estimation has been substantially improved. Specifically, given a collection of unconstrained face images, we first perform 2D landmark estimation [10] of each image. In order to prepare an enhanced 3D template for the photometric stereo, we deform a generic 3D face template such that the projections of its 3D landmarks are consistent with the estimated 2D landmarks on all images, and the surface normals are maintained. With the enhanced 3D template, 2D face images at *all poses* are projected onto the 3D surface, where the collection of projections will form a data matrix spanning all vertices of the template. Since there are inevitably missing elements in the data matrix due to varying poses, matrix completion is employed and followed by the estimation of surface normal and lighting decomposition via SVD. We further deform the 3D shape so that its updated surface normals will be similar to the estimated ones, under the landmark constraint and an additional boundary constraint.

To illustrate the strength of our approach, we perform experiments on several large collections of celebrities, as well as one subject where the ground truth 3D model is collected. Both qualitative and quantitative experiments are conducted and compared with the state-of-the-art method.

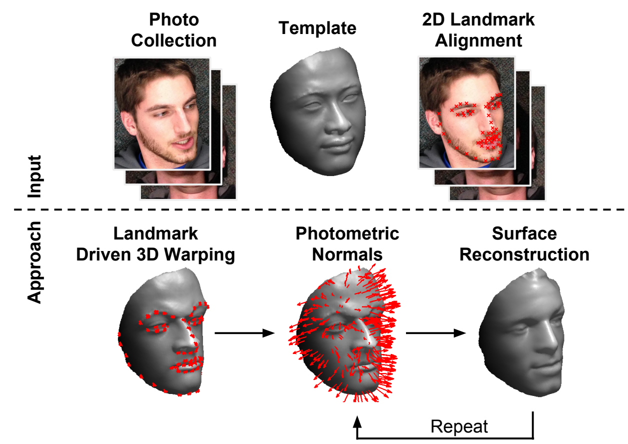


Figure 2: Overview of the proposed 3D face reconstruction technique.

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

[1] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE T-PAMI*, 25(9):1063–1074, 2003.

[2] Chen Cao, Qiming Hou, and Kun Zhou. Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Trans. Graph.*, 33(4):43, 2014.

[3] Pablo Garrido, Levi Valgaerts, Chenglei Wu, and Christian Theobalt. Reconstructing detailed dynamic face geometry from monocular video. *ACM Trans. Graph.*, 33(6):158, 2013.

[4] Mark F Hansen, Gary A Atkinson, Lyndon N Smith, and Melvyn L Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *CVIU*, 114(8):942–951, 2010.

[5] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *CVPR*, 2014.

[6] Ira Kemelmacher-Schlizerman and Steven M Seitz. Face reconstruction in the wild. In *ICCV*, 2011.

[7] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 FPS via regressing local binary features. In *CVPR*, 2014.

[8] Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, and Christian Theobalt. Lightweight binocular facial performance capture under uncontrolled lighting. *ACM Trans. Graph.*, 31(6):187, 2012.

[9] Jun Wang, Lijun Yin, Xiaozhou Wei, and Yi Sun. 3D facial expression recognition based on primitive surface feature distribution. In *CVPR*, volume 2, pages 1399–1406. IEEE, 2006.

[10] Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li. Learn to combine multiple hypotheses for accurate face alignment. In *ICCVW*, pages 392–396, 2013.