Real-time 3D Face Super-resolution From Monocular In-the-wild Videos

Patrik Huber*, William Christmas, Adrian Hilton, Josef Kittler
Centre for Vision, Speech and Signal Processing
University of Surrey, UK

Matthias Rätsch
Image Understanding and Interactive Robotics
Reutlingen University, Germany

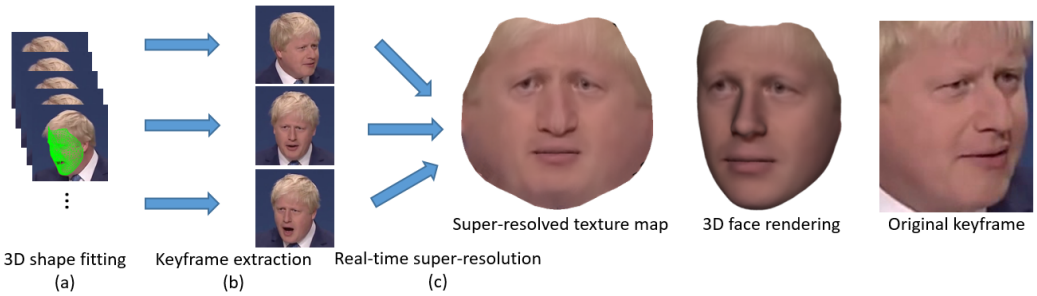

Figure 1: *(a) 3D shape fitting in each video frame. (b) Keyframe extraction based on pose and frame quality. (c) Median-based super-resolution fusion of the keyframes. The inter-eye-distance of the input frames is around 45 pixels. (Video from the 300-VW dataset)*

Abstract

We present a fully automatic approach to real-time 3D face reconstruction from monocular in-the-wild videos. We use a 3D Morphable Face Model to obtain a semi-dense shape and combine it with a fast median-based super-resolution technique to obtain a high-fidelity textured 3D face model. Our system does not need prior training and is designed to work in uncontrolled scenarios.

Keywords: 3D Face Reconstruction, Face Super-Resolution, 3D Morphable Face Model, Real-time, Open Source Software

Concepts: •**Computing methodologies** → **Reconstruction;** *Appearance and texture representations;*

1 Introduction

Reconstructing a face in 3D from monocular video sequences is still a challenging task on videos that are recorded in in-the-wild scenario, which contain low resolution, motion blur, and large pose variations. Previous approaches have tackled the problem in controlled conditions, with applications in animation, motion capture and games. For example, [Ichim et al. 2015] track from hand-held video input, but their data is collected in rather controlled scenarios. Further, they require subject specific training and manual labelling by an experienced labeler, taking 1 to 7 minutes per subject, which is a tedious process, and their resulting model is person-specific. [Cao et al. 2015] don't require user specific training, but present only results in controlled, frontal and high image resolution and require CUDA to achieve real-time performance. Additionally, none of these approaches focuses on reconstructing the texture of the tracked person, as their main use case is avatar creation.

*Contact: http://www.patrikhuber.ch

SIGGRAPH '16 Posters, July 24-28, 2016, Anaheim, CA
ISBN: 978-1-4503-4371-8/16/07
DOI: http://dx.doi.org/10.1145/2945078.2945145

In this work, we propose an approach that uses a semi-dense shape fitting with keyframe extraction and combine super-resolution texture fusion with a model-based approach. We build upon the 3D Morphable Model fitting approach of [Huber et al. 2016; Aldrian and Smith 2013] and adopt the super-resolution approach of [Maier et al. 2015] to work on monocular video sequences and in an incremental fashion. In comparison to existing approaches, we sacrifice detailed shape for real-time capabilities, require no user-specific training, and focus on texture reconstruction.

2 Our approach

Given a set of facial landmarks in each frame, we fit the shape of a 3D Morphable Face Model to these points using a closed-form solution for shape and pose fitting. On top of the PCA model, we use a set of expression blendshapes that model the 6 universal Ekman emotions. This initial fit is used to calculate a frame quality score using the variance of Laplacian measure. Together with the estimated pose, we extract keyframes. We divide the pose range into 20° intervals and add a frame to the set of keyframes if that particular bin is currently empty or the score of the frame higher than the current keyframe. Additionally, we penalise frames with pronounced expressions.

In a second step, we merge the keyframes using a median-based super-resolution approach. The texture of each keyframe is remapped to a common reference of higher resolution using the semi-dense registration obtained from the model fitting. For each pixel, we then compute the weighted median, using the previously calculated weights, and, additionally, a weighting based on the view-angle of its vertex to the camera and the distance from the camera. This super-resolved texture map is recomputed whenever a new keyframe is added.

The contributions of this work are as follows: We combine the semi-dense shape fitting with a real-time super-resolution approach to obtain a super-resolved 3D face texture. We demonstrate the robustness of our approach on the challenging 300 Videos in the Wild dataset ([Shen et al. 2015]) that contains challenging conditions like motion blur, pose and low resolution. In future work, we plan to improve the quality of the shape fitting, which would further improve the quality of the super-resolved texture.

Our 3D face model and fitting library are available from www.4dface.org.

Acknowledgements

Support from the EPSRC Programme Grant FACER2VM (EP/N007743/1) is gratefully acknowledged.

References

ALDRIAN, O., AND SMITH, W. A. P. 2013. Inverse rendering of faces with a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence 35*, 5, 1080–1093.

BLANZ, V., AND VETTER, T. 1999. A Morphable Model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM Press/Addison-Wesley Publishing Co., 187–194.

CAO, C., BRADLEY, D., ZHOU, K., AND BEELER, T. 2015. Real-time high-fidelity facial performance capture. *ACM Trans. Graph. 34*, 4 (July), 46:1–46:9.

HUBER, P., HU, G., TENA, R., MORTAZAVIAN, P., KOPPEN, W. P., CHRISTMAS, W., RÄTSCH, M., AND KITTLER, J. 2016. A multiresolution 3D Morphable Face Model and fitting framework. In *International Conference on Computer Vision Theory and Applications (VISAPP)*.

ICHIM, A. E., BOUAZIZ, S., AND PAULY, M. 2015. Dynamic 3D avatar creation from hand-held video input. *ACM Trans. Graph. 34*, 4 (July), 45:1–45:14.

MAIER, R., STÜCKLER, J., AND CREMERS, D. 2015. Super-resolution keyframe fusion for 3D modeling with high-quality textures. In *2015 International Conference on 3D Vision, 3DV 2015, Lyon, France, October 19-22, 2015*, IEEE, 536–544.

SHEN, J., ZAFEIRIOU, S., CHRYSOS, G. G., KOSSAIFI, J., TZIMIROPOULOS, G., AND PANTIC, M. 2015. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In *2015 IEEE International Conference on Computer Vision Workshop, ICCV Workshops 2015, Santiago, Chile, December 7-13, 2015*, IEEE, 1003–1011.