Demo of Face2Face: Real-time Face Capture and Reenactment of RGB Videos

Justus Thies[1] Michael Zollhöfer[2] Marc Stamminger[1] Christian Theobalt[2] Matthias Nießner[3]
[1]University of Erlangen-Nuremberg [2]Max Planck Institute for Informatics [3]Stanford University

Figure 1: *Our facial reenactment in a real-time setup: a monocular target video sequence (e.g., from Youtube) is reenacted based on the expressions of a source actor who is recorded live with a commodity webcam. Image courtesy of https://www.whitehouse.gov.*

Abstract

We present a novel approach for real-time facial reenactment of a monocular target video sequence (e.g., Youtube video). The source sequence is also a monocular video stream, captured live with a commodity webcam. Our goal is to animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion. To this end, we first address the under-constrained problem of facial identity recovery from monocular video by non-rigid model-based bundling. At run time, we track facial expressions of both source and target video using a dense photometric consistency measure. Reenactment is then achieved by fast and efficient deformation transfer between source and target. The mouth interior that best matches the re-targeted expression is retrieved from the target sequence and warped to produce an accurate fit. Finally, we convincingly re-render the synthesized target face on top of the corresponding video stream such that it seamlessly blends with the real-world illumination. We demonstrate our method in a live setup, where Youtube videos are reenacted in real time.

Keywords: face capture, facial reenactment, expression transfer

Concepts: •**Computing methodologies** → **Image manipulation;**

1 Introduction

We employ a new dense markerless facial performance capture method based on monocular RGB data, similar to state-of-the-art methods. However, instead of transferring facial expressions to virtual CG characters, our main contribution is monocular *facial reenactment* in real-time. In contrast to previous reenactment approaches that run offline, our goal is the *online* transfer of facial expressions of a source actor captured by a commodity webcam to a target actor. The target sequence can be any monocular video; e.g., legacy video footage downloaded from Youtube with a facial performance. We aim to modify the target video in a photo-realistic fashion, such that it is virtually impossible to notice the manipulations. Faithful photo-realistic facial reenactment is the foundation for a variety of applications; for instance, in video conferencing, the video feed can be adapted to match the face motion of a translator, or face videos can be convincingly dubbed to a foreign language.

In our method, we first reconstruct the shape identity of the target actor using a new global non-rigid model-based bundling approach based on a prerecorded training sequence. As this preprocess is performed globally on a set of training frames, we can resolve geometric ambiguities common to monocular reconstruction. At runtime, we track both the expressions of the source and target actor's video by a dense analysis-by-synthesis approach based on a statistical facial prior. Although our focus is not on (geometrically) accurate face tracking, we demonstrate that our RGB tracking accuracy is on par with the state of the art, even with online tracking methods relying on depth data. In order to transfer expressions from the source to the target actor in real-time, we propose a novel transfer functions that efficiently applies deformation transfer directly in the used low-dimensional expression space. For final image synthesis, we re-render the target's face with transferred expression coefficients and composite it with the target video's background under consideration of the estimated environment lighting.

We demonstrate highly-convincing transfer of facial expressions from a source to a target video in real time. We show results with a live setup where a source video stream, which is captured by a webcam, is used to manipulate a target Youtube video. In addition, we compare against state-of-the-art reenactment methods, which we outperform both in terms of resulting video quality and runtime. In the end, we are able to demonstrate the first real-time reenactment method which requires only a single, commodity webcam.

2 User Experience

As our facial reenactment approach runs in real time, we demonstrate it in a live setup. In our booth, visitors are able to reenact the facial expressions of Youtube videos. To this end, we prepared a variety of videos that can be freely selected by the user. We show the reenactment results on a large screen that provides immediate feedback to the participants; we hope this will be a great user experience and also fun to watch for everybody standing around the booth. In addition, we plan to keep track of the best and funniest reenactment results; and share them with the world via a social media platform (i.e. Facebook). This live demo illustrates our *Face2Face* approach

[Thies et al. 2016], which can also be used for video dubbing applications or real-time translation in a skype call.

References

THIES, J., ZOLLHÖFER, M., STAMMINGER, M., THEOBALT, C., AND NIESSNER, M. 2016. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. In *Proc. CVPR 2016*.