VIDEO FACE REPLACEMENT

Secondary Paper

ACM Transactions on Graphics, Vol. 30, No. 6, Article 130, Publication date: December 2011

The method presented in the paper is for replacing facial performances in video. The approach mentioned accounts for the variability in identity, visual appearance, speech, and timing between source and target videos. Not similar to the prior work, it does not require substantial manual operation or complex acquisition hardware, only single-camera video. This method needs to use a 3D multilinear model to track the facial performance in both videos. Using the corresponding 3D geometry, They warp the source to the target face and retime the source to match the target performance. Then compute an optimal seam through the video volume that maintains temporal consistency in the final composite. method showcase the use of our method on a variety of examples and present the result of a user study that suggests our results are difficult to distinguish from real video footage.

Some cases have the need that the timing of the facial performance matches precisely in the source and the target footage. However, it might be very tedious to match these timings exactly as demonstrated by the numerous takes that are typically necessary to obtain compelling voiceovers Instead, method only requires a coarse synchronization between source and target videos and automatically retime the footage to generate a precise match for the replacement. After tracking and retiming, the method blends the source performance into the target video to produce the result. This blending makes use of gradient-domain compositing to merge the source actor's face into the target video. While gradient domain compositing can produce realistic seamless results, the quality of the composite is often tied to the seam along which the blend is computed. Using an arbitrary seam is known to lead to bleeding artifacts. To minimize these artifacts, the method automatically computes an optimal spatiotemporal seam through the source and target that minimizes the difference across the seam on the face mesh and ensure that the regions being combined are compatible. In the second stage the method uses this seam to merge the gradients and recover the final composite video.

This paper presents a function for face exchange in video that achieves quality end-product using an easy acquisition process. Not like the previous work, the approach imagines cheap hardware and requires minimal supervision. Just using a solo camera and basic reflection, the paper capture source video that will be inserted into a target video. Then the method demands us to track the face in both the source and target videos using a 3D multilinear model. wrap the video from the actual source in both space and time to align it to the target. Finally, blend the videos by computing an optimal patio-temporal seam and a novel mesh-centric gradient domain blending technique. The system replaces all or part of the face in the target video with that from the source video. Source and target can have the same person or two different subjects. They can contain similar performances or two very different performances. And either the source or the target can be existing footage, as long as the face poses are approximately the same. This leads to a handful of unique and useful scenarios in film and video editing where video face replacement can be applied. For example, it is common for multiple takes of the same scene to be shot in close succession during a television or movie shoot. While the timing of performances across takes is very similar, subtle variations in the actor's inflection or expression distinguish one take from the other. Instead of choosing the single best take for the final cut, our system can combine, e.g., the mouth performance from one take and the eyes, brow, and expressions from another to produce a video montage

This paper show results for a number of different subjects, capture conditions, and replacement scenarios. Some of the figures show multi-take video montage examples, both shot outdoors with a handheld camera. Another figure shows dubbing results of a translation scenario, where the source and target depict the same subject speaking in different languages, with source captured in a studio setting and target captured outdoors. A figure in the top shows a replacement result with different source and target subjects and notably different performances. Figure in the bottom shows a retargeting result with different subjects, where the target was used as an audiovisual guide and the source retimed to match the target. User interaction Although most of our system is automatic, some user interaction is required. This includes placing markers in FaceGen, adjusting markers for tracking initialization, and specifying the initial blending mask. Interaction in FaceGen required 2-3 minutes per subject. Tracking initialization was performed in less than a minute for all videos used in our results; the amount of interaction here depends on the accuracy of the automatic face detection and the degree to which the subject's expression and viseme differ from closed-mouth neutral. Finally, specifying the mask for blending in the first frame of every example took between 30 seconds and 1 minute. For any given result, total interaction time is therefore on the order of a few minutes, which is significantly less than what would be required using existing video compositing methods.

The authors conclude a system for producing face replacements in video that requires only single-camera video and minimal user input and is robust under significant differences between source and target. They have shown with a user study that results generated with this method are perceived as realistic. Their method is useful in a variety of situations, including multi-take montage, dubbing, retargeting, and face replacement. Future improvements such as inpainting for occlusions during large pose variations, 2D background warping for vastly different face shapes, and lighting transfer between source and target will make this approach applicable to an even broader range of scenarios.

BRINGING PORTRAITS TO LIFE

Primary Paper

ACM Transactions on Graphics, Vol. 36, No. 4, Article 196. Publication date: November 2017

Some objects convey as big a range of depth and meaning as the human face. Face expressions in humans convey not only the major emotions, but also the subtle variations, a rather small varied nuanced view into the emotional state of a person, for example, a sad smile, blushing. In this work, the main picture will keep in revolting around animating faces in human portraits and controlling their expressions. To avoid crossing into the "uncanny valley", previous facial animation techniques usually assume the availability of a video of the target face, which exhibits variation in both pose and expression. A video imputed, or even an image collection of the target face allows for an accurate 3D face reconstruction, over which face textures are mapped and manipulated. If you are comparing this work to the previous work, you will find a different improvement the authors did in this they used as input only a single image of a target face to animate it. This makes the method more widely used and applicable in various situations and contexts than the previous work which other authors had did on this topic.

The method presented to us in this paper can be summarized by saying that this is a technique to automatically animate a still portrait, making it possible for the subject in the photo to come to life and express various emotions. The approach used a driving video and develop means to transfer the expressiveness of the subject in the driving video to the target portrait. In comparison to previous work that requires an input video of the target face to reenact a facial performance, the technique in this paper uses only a single target image. The technique animates the target image through 2D warps that imitate the facial transformations in the driving video. As warps alone do not carry the full expressiveness of the face, we add fine-scale dynamic details which are commonly associated with facial expressions such as creases and wrinkles. Furthermore, the authors imagined regions that are hidden in the input target face, most notably in the inner mouth. This technique gives rise to reactive profiles, where people in still images can automatically interact with their viewers. The method demonstrates our technique operating on numerous still portraits from the internet.

The Technique takes a solo target photo of a neutral face in frontal pose and generates a video that expresses various emotions. Previous works addressed face manipulation from a single image, but do not focus on generating an animated video. Other works like the one did by Thies et al in 2016 and also the work did by the famous Vlasic et al in 2005 manipulate or reenact a facial performance, but these assume the availability of a video of the target face. In what follows, we elaborate on the most closely related works. Most prior works require more than a single target image of a face to automatically manipulate it. A video-to-image facial retargeting application was previously introduced in Cao et al. [2014], but their method is not automatic and requires some user interaction. Facial editing using deep networks was introduced in Yeh et al in 2016. The method of Liu et al used in 2001 enables transferring the fine-scale details of one person’s changed expression to a neutral target image. In the work proposed in this paper after making so much of changes by analyzing the previous work by the very intelligent authors, we extend their technique to accommodate more general image pairs and a stable video output. Recently, fried et al. [2016] presented a technique to manipulate the camera viewpoint from a single input image. Their method enables modifying the apparent relative pose and the distance between the camera and the subject. Other works, such as Hassner in 2015], specifically address the problem of face fractalization, as it is extremely beneficial for facial recognition [Ding and Tao 2016].

The authors have presented a method that brings life to a single still portrait in the sense that the image is animated to imitate the facial expressions of a driving video. The use of a single input image captured in casual settings is particularly challenging, but it offers a wide applicability of the method. Technically, the main challenge was to preserve the identity of the target face, while manipulating it with warps and features taken and transferred from frames of some arbitrary driving video. They built on the fact that there is a significant commonality in the way humans "warp" their faces to make an expression. Thus, transferring local warps between aligned faces succeeds in hallucinating facial expressions. they transfer both geometric warps, consisting of 2D offsets, and photometric changes, consisting of illumination ratios.

SUMMARY

The video face replacement and brining portraits to life are related in the element of changing the taken picture from it’s static position to some other modified position. In the video face replacement, we replace the image or the static image in the video into some other similar image which is the same approach used in bringing a portrait to life by just changing the static image to a portrait by changing the image like we do in video face replacement. In the second one we replace the face or the object with different emotions.