International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-4 Number-1 Issue-14 March-2014

The Comparative study of Automated Face replacement Techniques for Video

Harmesh Sanghvi[1], D.R. Kasat[2], Sanjeev Jain[3], V.M. Thakare[4]

Abstract

For entertaining purposes, a computerized special effect referred to as "morphing" has enlarged huge attention and face replacement is one of the interesting tasks. Face replacement in video is a useful application in the amusement and special effect industries. Though various techniques for face replacement have been developed for single image and generally applied in animation and morphing, there are few mechanisms to spread out these techniques to handle videos automatically. Face replacement in video automatically is not only a fascinating application, but a challenging problem. For face replacement in video, the frame-by-frame manipulation process using the software is often time consuming and labor-intensive. Hence, the paper compares numerous latest Automatic face replacement techniques in video to understand the various problems to be solved, their shortcomings and benefits over others.

Keywords

Face replacement, Face detection, Face alignment, Face relighting, Blending, Morph-able Model

1. Introduction

Artists have long used the techniques of changing human face for comical and caricature purposes. With the beginning of digital technology, such editing has become common place through tools like Adobe Photoshop. Though, such editing was awkward and required exhaustive manual work by the artist so as to provide a plausible image. Certain methods allowed for automatic face replacement of people in single image [2, 3]. For example, in 2004 the method by V. Blanz et al. [2] fits a Morphable model to faces in both the source and target images and renders the source face with the parameters assessed from the target image. Finally, it replaces the target face with source face in the target image. Morphable model is assembled from a factual investigation of human faces, got from an expansive database of 3D images, which might be changed by changing parameters. It can gauge the 3D state of a human face, its positioning in the space, and brightening conditions in the scene. Thus the reconstructed face extracted from 2D image can be manipulated in 3D [1].

In 2008, D. Bitouk et al. [3] described another system for automatic face swapping using a large database of faces. Though it is hard for user's to find a candidate face to match the target face in appearance and pose from their images, the system allowed de-identification automatically by selecting candidate face images from a large face library that is similar to the target face in appearance and pose. Lastly, it replaces the target candidate with selected candidate from the library image using image based method.

In the first method, during an initialization step, the shape parameters of the model are fitted to both human faces. Then, these parameters are varied to obtain the structure of the face in a deformed shape, but in the same pose. To show this deformation in a video, it has to be performed in each and every frame. So it is clear that this methodology can be applied, but have to handle some challenging problem to perform an automatic face replacement in video and to maintain spatial and temporal coherence.

Certain methods have been proposed which allows replacing face in video automatically by solving the various challenging problems [4-8]. Essentially, there are three basic steps to replace face in video as shown in [Figure. 1].

A face detection algorithm should locate the positions of the character human precisely. However, it is not enough for knowing where the face appears in a video. To replace the face, the system has to detect not only the face, but also the outlines of the facial

Manuscript received February 18, 2014.

Sanghvi Harmesh, Computer Engineering, SCET, Surat, India.
D.R. Kasat, Associate Professor, Computer Engineering Department, SCET, Surat, India.
Sanjeev Jain, Director, MITS, Gwalior.
V.M. Thakare, Professor and Head, Computer Science and Engineering Department, Amravati University, Amravati, India.

profile and structures. Thus, a face alignment algorithm plays a critical role in these systems. The eventual result is ended by pasting up the synthesized sequence onto the target character's face. A visually convincible output video is done by blending the synthesized sequence and the target video together seamlessly. This paper presents few latest techniques providing face replacement in video with various scenarios.

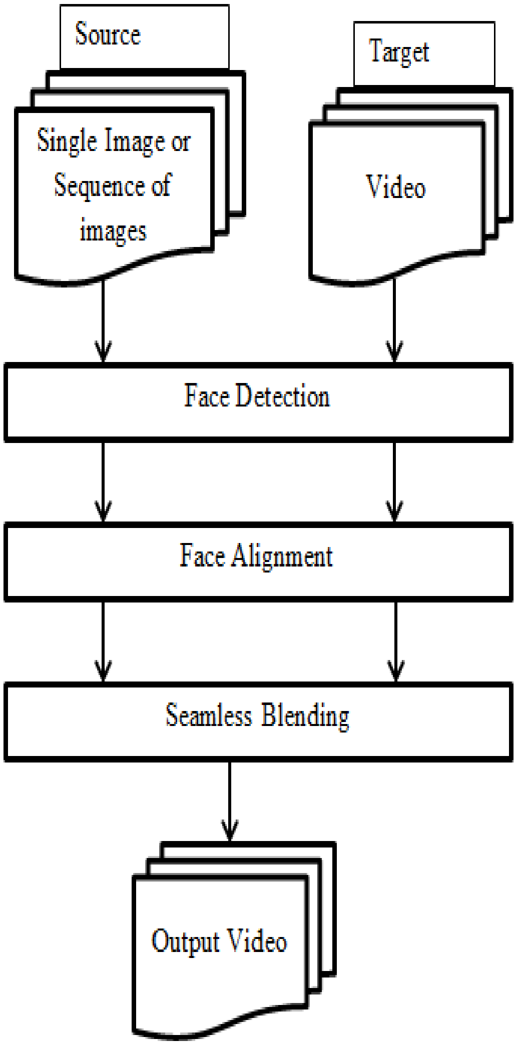

Figure 1: Basic steps for face replacement

2. Face Replacement Techniques

A) Image Based Face Replacement in Video:

In 2009, Y. Liang [4] proposed the system which provides plausible face replacement in video. This system allows replacing target human face from target video with source face in source video. This replacement algorithm has three main stages. First, given an input video, it detects all faces that are present, and align such detected faces. Second, it analyses facial expressions of each detected faces and select candidate face images from source video that are most similar to the target face in pose and expression. Finally, it blends candidate replacements to target video.

Methodology Overview:

For face detection here Bayesian Tangent Shape Model (BTSM) is used. At that moment system contains two models. One indicates the prior shape distribution in tangent shape space and the other is a likelihood model in image shape space. Based on these two models, the posterior distribution of model parameters can be derived. To replace the face, it requires outlines of the facial profile and structures. For face alignment 2D landmark points are used from BTSM (as mentioned above). To identify appropriate source pose and expression, system allows user to select the best candidate for each frame independently in the target video. To achieve this, system splits the target video into several screens. This process is referred to as clustering. After clustering, user needs to select appropriate frame for each cluster. The best candidate face must have the most similar pose and expression.

Once the user selected the best candidate face for each cluster, it will synthesize in-between frames. To concatenate the clusters, it requires interpolating the frames between clusters. A new smooth image interpolation may be accomplished through warping two photos towards same positions then dissolve the particular image textures in concert. After producing the smooth image sequence, the final step is to warp the facial features to appropriate positions to match the expressions exactly. It warps the smoothly interpolated image sequence according to the difference vectors. The actual warped images tend to be going to end up being mixed up into target video clip.

Sometimes, the size of the target character's face and the source character's face does not correspond with each other due to scale. Face concealment can be easily done by clearing the gradient field to zero in the facial region then reconstructed by integrating the modified gradient. It could receive an even as well as de-identification encounter by simply assigning null gradient for the pixels from the facial region. While system inserts the source face on the target face, the pasting area may perhaps surpass this border of the target face. The irrelevant background would be blended together such that the smudge effect will occur. In order to seamlessly clone the source spot into a target image, the operation is often completed simply by fixing a large linear process. Instead of solving the large linear system of Poisson equation, it uses image cloning.

B) 3D Morphable Model Based Face Replacement in Video:

In 2009, Y. T. Cheng et al. [5] proposed the system that changes the target subject face in the target video with the source subject face, beneath similar pose, illumination and expression. This approach is based on 3D morph-able model [1] and an expression model database to deal with expressions and the input information of the source subject face is reduced to one to two images. The system has a targeted video the other supplier photograph since feedback, as well as the productivity will be the video with the targeted subject face replaced with the source subject face.

Methodology Overview:

Given the source image, it reconstructs the 3D model of the source subject face using 3D Morphable model [1]. The 3D face synthesizer derives a Morphable face to fit the input image, and map the texture from the image to the derived 3D face model. A face alignment algorithm is is usually placed on the target video to identify the exhaustive facial features and skeletons of the target subject face [4].

A pose estimator exploits the face alignment results to calculate the head pose parameters of the target subject face. Here method employs a 3D face expression database to clone the expressions to the source face model. To fit the expressions to the target video, Y. T. Cheng et al. [5] proposed an algorithm to extract the expression parameters. In some videos, directly rendering the source subject face model onto the target frame results in illumination inconsistency. A relighting algorithm relights the rendered source subject face for illumination consistency. Finally, it seamlessly composites the rendered source model with the target frame using Poisson equation, proposed in 2003 by P. P´Erez, et al. [10].

The output is a video with the target face replaced by the source face, with similar pose, expression, and lighting.

C) Automatic Face Replacement in Video Based on 2D Morphable Model:

In 2010, F. Min, et al. [6] proposed an automatic face replacement approach in video based on 2D Morphable model. This approach includes three main modules: face alignment, face morph, and face fusion. Given a source image and target video, the Active Shape Models (ASM) is applied to source image and target frames for face alignment. After that, the source face shape can be warped correspond the target face shape by a 2D Morphable model. The color and lighting of source face are adjusted to keep consistent with those of target face, and seamlessly blended in the target face. This approach is fully automatic without user interference, and generates natural and realistic results.

Methodology Overview:

In face alignment, the ASM algorithm is used to the target frame and source image to identify the exhaustive facial features. In face morph, a 2D morphable model is fitted to the target and source face. Adjusting the parameter of the morphable model, the shape of source face is warped to match that of target face. In face fusion, a relighting and recolor algorithm is applied to the source face to keep illumination and color consistency with the target face [6]. The source face is seamlessly blended in the target face using Poisson equation, proposed in 2003 by P. P´Erez, et al. [10].

D) Video Face Replacement:

In 2011, K. Dale, et al. [7] proposed the method which allows replacing facial performances in video. It also provides face replacement in target video from source video. The system tracks both the faces in source and target video using multilinear model [9].

Using this tracked 3D Geometry, source face is warped to target face in every frame of video. It is sometimes important that the timing of the facial performance matches exactly in the source and target video; this is done by retiming algorithm. After tracking and retiming, it blends the source performance in the target video to produce the final result. They computed optimal seam through the video volume that maintains temporal consistency in the final composite.

Methodology Overview:

This System consists of following Major steps: 1) Face Tracking 2) Face alignment 3) Blending

1) Face tracking:

The method of D. Vlasic et al. [9] computes the pose and parameters of the multilinear face model to track a face across a video. Initialization is critical, as errors in the initialization will be propagated throughout the sequence. Therefore, System provides a simple user interface that can ensure good initialization and can correct tracking for troublesome frames. The user can adjust points of markers on the eyes, eyebrows, nose, mouth, and jawline through the interface, from which the best-fit pose and model parameters are computed. This default face mesh is generated from the multilinear model [9] using a user-specified set of initial points corresponding to the most appropriate expression, viseme, and identity.

2) Face Alignment:

To align the source face in the target frame, it uses face geometry from the source sequence and pose from the target sequence. It also takes texture from the source image. Automatic retiming algorithm then compares the average minimum Euclidean distance between the first partial derivatives with respect to time that is used to match the facial performance in video. Retiming applied on sequence of images till first partial derivatives with respect to time will match in both videos.

3) Blending:

It is possible to generate a truly photo-realistic composition of two images by the Poisson blending algorithm proposed in 2003 by P. P´Erez, et al. [10] but it needs specifying the area from the aligned video that needs to be mixed into the target video. In addition, this seam needs to be specified in every frame of the composite video, making it very tedious for the user to do. This method incorporates these requirements in a novel graph cut framework that estimates the optimal seam on the face mesh. For every frame in the video, It computes a closed polygon on the having constructed this graph. The construction of the graph confirms that, in every frame, the graph-cut seam forms a closed polygon that separates the target vertices from the source vertices. Lastly, it fusions the source and target videos using gradient-domain fusion.

E) Face Replacement in Video from a Single Image:

In 2012, A. Niswar, et al. [8] proposed a system to replace the face in a video with the face of a different person from a single image, where the face is not limited to be in a specific pose, e.g. exactly frontal. This system is able to replace the target face with another face of an entirely different pose and animate the new face based on the original speech in the video.

Methodology Overview:

This System consists of following Major steps: 1) 3D face Reconstruction 2) 3D Face animation 3) Feature Point Tracking 4) 2D projection with Blending.

1) 3D Face Reconstruction:

Initially, a 3D face model is built from an image by deforming a generic 3D face model to appropriate the face in the image. At first, the face pose in the image is determined by computing the optimal affine transformation parameters to transform the 3D face to the pose in the image. Unlike previous works, the parameters are computed using non-linear optimization, which increases the accuracy of the face pose determination. The 3D face is then projected into 2D space and deformed to fit the face in the image. Finally, the deformed 2D face is merged with the depth information from the transformed 3D face to obtain the new 3D face model.

2) 3D Face Animation:

It reuses the speech in the original video to animate the 3D face with animation parameters computed from a viseme dataset. The speech animation consists in computing the trajectory of the viseme parameter for the whole speech, based on the phonemes and their durations extracted from the original speech.

3) Feature Points Tracking:

In order to project the animated 3D face to the original video correctly, it uses a robust, real-time, automatic monocular tracking method. It provides the position of some facial feature points in the video.

4) 2D Projection with Blending:

The animated 3D face is finally projected to the original video frames to replace the face by computing the optimal projection matrix for each frame based on the tracked feature points and the corresponding vertices of the 3D face. The projected face is then seamlessly embedded to the video frame using Poisson image blending.

Table 1: Performance Analysis

Techniques	Pros	Cons
A. Image Based Face replacement in Video [4]	Less time complexity because of Image based method; It can be used for entertainment purpose;	Facial expression and pose would be same; It requieres to shoot target video with facieal expression and pose similar to the source video; It requieres maual inputs in clustering process; The tolerance to pose variance is limited by the robustness of face alignment algorithm;
B. 3D Morphable Model Based Face Replacement in Video [5]	Here source is reduced to single Image; It takes care of facial expession and pose of target face;	Time comsuming process because of 3d model based method; The tolerance to pose variance is still limited by the robustness of face alignment algorithm;
C. Automatic Face Replacement in Video Based on 2D Morphable Model [6]	This approach is fully automatic without user interference; Less time comsuming process because of 2d model based method;	It does not take care about facial expression; The tolerance to pose and expression variance is limited by the robustness of ASM; Sharp lighting and violent movement in videos may affect the final result;
D. Video Face Replacement [7]	It gives plausible results; This approach is fully automatic with less user interference; It takes care about facial expression	Tracking is based on optical flow, which requires that the lighting change slowly over face; Tracking often degrades beyond the range of poses outside 45^0 from frontal; Lighting must also be similar between source and target;
E. Face Replacement in Video from a single image [8]	It works very well for non-frontal face; This system is able to replace the target face with another face of an entirely different pose and animate the new face based on the original speech in the video;	High complexity

3. Conclusion

In this paper we have surveyed the growth of Video face replacement and described recent advances in the field. Table I above depicts the performance analysis of the different automatic face replacement techniques surveyed in the paper. Face replacement techniques in video share the following components: face detection, face alignment, and Blending. The ease with an artist can effectively use these techniques to replace the face in existing videos for caricature purpose. We surveyed various Face replacement techniques in video, including those image based warping and model based. Image based techniques do not take care of facial expression and 3D model based techniques are time consuming and suffer from robustness of face alignment algorithm.

References

[1] V. Blanz and T. Vetter, “A Morphable Model For The Synthesis Of 3D Faces”, In SIGGRAPH '99: Proceedings of The 26th Annual Conference On Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley Publishing Co. Pages 187-194, New York, USA, 1999.

[2] V. Blanz, K. Scherbaum, T. Vetter and H.P Seidel, “Exchanging Faces in Images”, Computer Graphics Forum (Proc. EUROGRAPHICS) 23, 3, Pages 669-676, 2004.

[3] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, “Face Swapping: Automatically Replacing Faces in Photographs”, In SIGGRAPH '08: ACM SIGGRAPH 2008 Papers, Pages 1-8, New York, USA, 2008.

[4] Y. Liang, “Image Based Face Replacement in Video”, Master's Thesis, CSEI Department, National Taiwan University, 2009.

[5] Y. T. Cheng, V. Tzeng, Y. Liang, C. C. Wang, B. Y. Chen, Y. Y. Chuang, and M. Ouhyoung,“3D Model Based Face Replacement in Video”, In SIGGRAPH 2009 Poster, ACM.

[6] F. Min, N. Sang, Z. Wang, “Automatic Face Replacement in Video Based On 2D Morphable model”, Proceeding ICPR '10 Proceedings Of The 2010 20th International Conference On Pattern Recognition, IEEE Computer Society Washington, Pages: 2250-2253, DC, USA 2010.

[7] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister, "Video Face Replacement", ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 30, 6, 2011.

[8] A. Niswar, E. P. Ong and Z. Huang, "Face Replacement in Video from a Single Image", In SIGGRAPH Asia 2012 Posters, ACM.

[9] D. Vlasic, M. Brand, H. P fister and J. Popovi´C, "Face Transfer with Multilinear Models", ACM Trans. Graphics (Proc. SIGGRAPH) 24, 3, Pages 426–433, 2005.

[10] P. P´Erez, M. Gangnet and A. Blake, "Poisson Image Editing. ACM Trans", Graphics (Proc. SIGGRAPH) 22, 3, Pages 313–318, 2003.

Sanghvi Harmesh K was born in Idar, Gujarat on 16th May, 1991. He has completed his B.E. degree in Computer Engineering from HGCE, Ahmedabad in 2012. Currently he is pursuing M.E. in Computer Engineering from SCET, Surat, Gujarat.

Prof. Dipali Kasat is working as Associate Professor, Computer Engg. Dept., Sarvajanik College of Engg. & Tech, Surat. She has completed her M.E.(CSE-IT) from VIT Pune and is pursuing PhD in Computer Science & Engg. from SGB Amravati University, Amravati. She has published various national & international papers. She is specialized in Image/Video Morphing, Steganography, Steganalysis and Information Security.

Dr. Sanjeev Jain is working as a Dierctor of MITS a Grant in aid Autonomus Engineering college of Government of Madhya Pradesh. Madhav Institute of Technology and Scinece is one of the oldest engineering college of central India, established in 1957. He had completed his Master of Technology (M.Tech.), Computer Engineering from Indian Institute of Technology, Delhi and PhD from Amravati University. He has published various national & international papers.

Dr. Vilas Thakare is working as Professor and Head, in Computer Science, Faculty of Engineering & Technology, Post Graduate Department of Computer Science, SGB Amravati University, Amravati. He had worked as a Member, Expert Committee for AICTE (WR), CEDTI, and YCMOU. Also as Member, Advisory Committee, IICC, Nagpur University. He had been the Member, BOS, at SRTM University, Nanded, Nagpur University, Dr. BAMU, Aurangabad and is a Recognized PhD Supervisor in Computer Science, Computer Engg. , Electronics Engg. etc. He has to his credit, many Technical Papers that have been published in national & international Journals & Conference Proceedings. He has also provided consultancy for election data processing.