Face Morphing using 3D-Aware Appearance Optimization

Fei Yang[1]* Eli Shechtman[2]† Jue Wang[2]‡ Lubomir Bourdev[2]§ Dimitris Metaxas[1]¶

[1]Rutgers University [2]Adobe Systems

Figure 1: Our system can generate fully automatically high quality face morphing animation between faces of different pose and expression. **Top:** morphing between images of the same subject. **Bottom:** morphing between different subjects. The input images are the first and last column (highlighted in red).

Abstract

Traditional automatic face morphing techniques tend to generate blurry intermediate frames when the two input faces differ significantly. We propose a new face morphing approach that deals explicitly with large pose and expression variations. We recover the 3D face geometry of the input images using a projection on a prelearned 3D face subspace. The geometry is interpolated by factoring the expression and pose and varying them smoothly across the sequence. Finally we pose the morphing problem as an iterative optimization with an objective that combines similarity of each frame to the geometry-induced warped sources, with a similarity between neighboring frames for temporal coherence. Experimental results show that our method can generate higher quality face morphing results for more extreme pose, expression and appearance changes than previous methods.

1 Introduction

Image morphing is a special visual effect in which one image is smoothly transformed into another. It has been extensively explored and widely used in motion pictures and animations. Image morphing between two images usually begins with extracting features from both images and building correspondence between the two feature sets. A pixel-wise mapping function is then derived from the sparse feature correspondence, which is used to warp both input images into desired alignment at each interpolation position. Finally, color interpolation is performed to generate each transition frame [21].

In this paper we study the problem of image morphing between two face images, either from the same or different individuals. This is a challenging task, since human faces are highly non-rigid and could perform large 3D shape deformation under expression and pose variations. Moreover, human perception is sensitive to even a small amount of artifacts in faces.

Generic image morphing methods do not employ a 3D face model and, therefore, they are unable to accurately factor the differences between the two images due to pose and expression variations. As we will demonstrate later, without such accurate models of pose and expression, the interpolation results become unnatural. Furthermore, some previous morphing methods require tedious manual labeling on input images, which is undesirable in many applications.

We propose a novel approach for generating high quality face morphing animations. Our system is fully automatic and requires only the two end images as input. We first extract facial landmarks from each input image, and project them on a subspace learned from an external face shape dataset to recover the 3D face geometry. The geometry is factored into the pose and the expression of the input face. These are interpolated independently to create realistic intermediate shapes of the input face. For expression interpolation, we combine the expression flow [22] derived from the interpolated 3D models. Finally, our system employs an iterative appearance optimization framework where each intermediate frame is required to have similar geometry to the two corresponding interpolated models (up to a small residual optical flow to capture subtle geometric changes), as well as similar appearance to its neighbor frames. This optimization results in a sharp and temporally coherent interpolation sequence, as shown in examples in Figure 1.

To demonstrate the effectiveness of the proposed algorithm we

*e-mail: feiyang@cs.rutgers.edu
†e-mail:elishe@adobe.com
‡e-mail:juewang@adobe.com
§e-mail:lbourdev@adobe.com
¶e-mail:dnm@cs.rutgers.edu

compare it against a number of commonly used morphing approaches. Our experiments indicate that the proposed approach can generate higher quality face morphing results. As an additional application, we show how the system can be used to delete an undesired expression from a video sequence.

2 Related Work

User-assisted morphing. Most previous image morphing approaches require the user to manually specify feature correspondences between input images [20]. Mesh morphing methods [21, 10] define the spatial transformation at mesh points or snake curves. The mesh is then deformed with the constraint to maintain topological equivalence. The field morphing method [1] uses corresponding lines in the source and target images to define the mapping function between the two images, which simplifies the user input. These methods have already been implemented in commercial systems. However they require tedious annotation and do not work well for large variations in pose, as the 2D transformations used in these methods do not preserve 3D facial shape.

The view morphing method [16] preserves the 3D shape during morphing without the explicit use of 3D models. It works by pre-warping the two images prior to computing a morph, and then post-warping the interpolated images. This method works well for rigid objects. However, it cannot accurately estimate the projection parameters for a deformable object like a face exhibiting change in expression. Furthermore, it still requires substantial manual correspondence. In contrast, our approach is fully automatic and can handle both pose and expression variations.

Automatic morphing. Bichsel [2] proposes a fully-automatic morphing technique using a Bayesian framework and maximizing the penalized likelihood of the spatial and color transformations given the input images. Zanella et al. [23] and some other commercial face morphing packages such as the Face Morpher[1] use Active Shape Model (ASM) [5] to find corresponding points and then linearly interpolate their locations for mostly frontal view morphing. Our method also uses ASM for initial correspondence, however this is followed by projection on a 3D subspace and further appearance optimization. Therefore our method is more robust to the inaccuracies in point locations, and can handle large 3D pose variations.

Mahajan et al. proposed the Moving Gradients [12] method for automatic image interpolation which handles occlusions explicitly. It finds an optimal path for gradients at every pixel from one image to the other, and get impressive results. However, the motions were roughly linear and thus mostly confined to small 2D changes. Another recently proposed automatic method is Regenerative Morphing (RM) [17] in which the output sequence is regenerated from small pieces of the two source images in a patch-based optimization framework. The method generates appealing automatic morphs between radically different images and can produce impressive image interpolation results with additional point correspondences (either manual or based on automatic feature matching). Our optimization algorithm was inspired by RM, but it operates at the frame-level, as opposed to the patch-level. We report comparisons to RM, using the same ASM correspondences as our method, as well as to other general morphing methods (Sec. 6). The results show that our method is better suited for high-quality realistic face morphing.

3D face animation. There has been extensive work on creating face animations based on 3D face models [6][8][19]. However, accurate 3D face reconstruction from a 2D image is a challenging task by itself. We reconstruct a rough shape of the face that suffices for applying small changes in pose and expression [22]. Blanz et al. [3] proposed a face animation system by using a 3D morphable model of face shape and texture. The system has to model all facial components accurately (e.g. eyes, teeth and ears) and is computationally expensive.

[1] http://www.facemorpher.com/

3 Algorithm Overview

The framework of the proposed face morphing approach is shown in Figure 2. Given the two input images A and B, we first fit a 3D shape to each of them, as described in Section 4. A 3D shape contain two sets of parameters: external parameters describing the 3D pose of the face, and intrinsic parameters describing the facial geometry of the person under the effect of facial expression. We then linearly interpolate both the intrinsic and external parameters of the two input faces, resulting in a series of interpolated 3D face models, as shown in the bottom of Figure 2.

We use the dataset of 3D face models proposed by [18]. Since all the 3D models have one-to-one dense vertex correspondence, by projecting the 3D models to the image plane, we obtain warping functions from each input image to all interpolated frames. By warping the input image A with its corresponding warping functions, we can generate a series of deformed images $A_1, A_2, ..., A_T$, where T is the number of intermediate frames to be generated. Similarly, a series of images $B_1, B_2, ..., B_T$ can be obtained by warping image B.

A weighted averaging between A_t and B_t would give us an interpolated face C_t as the morphing result, as shown in Figure 2. However this approach is not optimal, as each A_t and B_t pairs is processed individually, thus they are not necessarily well aligned and temporal coherence is not guaranteed. To further improve the quality of the morphing sequence, inspired by the recent Regenerative Morphing approach [17], we define a morphing energy function and employ an iterative optimization approach to minimize it, as described in Section 5.

4 Fitting 3D shapes

We first describe how to fit a 3D face shape to a single face image. We follow the method described in the Expression Flow system [22], which is an efficient method for fitting 3D models to near-frontal face images. This method first localizes facial landmarks using the Active Shape Model (ASM) [13], then fits the 3D model based on these landmarks.

The face shape is defined using a shape vector s concatenating the X, Y, Z coordinates of all vertices. The deformable face model is constructed by running Principal Component Analysis (PCA) [4] on the training dataset from Vlasic et al. [18]. A new shape can be formed as a linear combination of eigenvectors v_i and the mean shape $\bar{s}$:

$$s = \bar{s} + \sum \beta_i v_i = \bar{s} + \mathbf{V}\beta. \quad (1)$$

The 3D fitting is performed by varying the coefficients β in order to minimize the error between the projections of the pre-defined landmarks on the 3D face geometry and the 2D feature points detected by ASM. The fitting error for the kth landmark is defined as:

$$E_k = \frac{1}{2}||P \cdot L_k \cdot (\bar{s} + \mathbf{V}\beta) - X_k||^2 \quad (2)$$

where P is a 2×3 projection matrix, L_k is a selection matrix that selects the vertex corresponding to the kth landmark and X_k are the X, Y coordinates of the kth ASM landmark. The total energy E, which is the total fitting error of all landmarks, is minimized with respect to the projection matrix P and the shape coefficients β using iterative optimization approach. In the first step the projection matrix P is optimized to align the current 3D shape to the 2D features. In the second step the shape coefficients β are optimized to deform the 3D shape for better fitting. The two steps are repeated until convergence.

4.1 The projection matrix P

In the 3D fitting algorithm described above the projection matrix P is computed to align the current 3D shape to the 2D landmarks. The

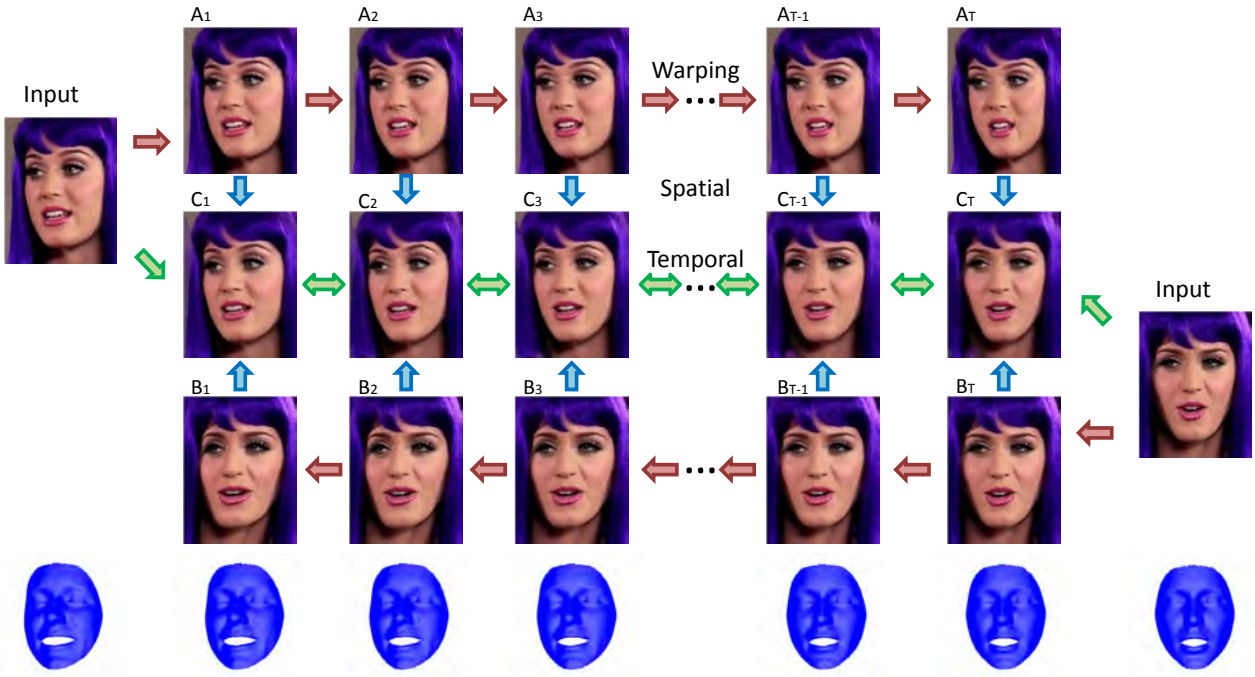

Figure 2: Overview of our framework. We first fit 3D shapes to both input images, and interpolate the 3D models for intermediate frames. The faces are warped using the difference between the 3D models. In each frame, the warped faces are blended together.

method in the Expression Flow system [22] uses a weak perspective projection model, and solves P using a least squares approach. However, we find this approach not optimal, since the variations of facial shapes are also captured by P, resulting in an inaccurate solution of β. For instance, an elongated face in the image can be explained either by a large scale in the Y direction of the projection matrix, or by an elongated 3D face geometry.

To avoid this ambiguity, we add two constraints to the projection matrix: the X and Y directions have the same scale, and there is no skew between them. Such a camera model is usually referred to as a restricted camera model [7]. The projection matrix P is parameterized with 6 variables, which can be estimated iteratively using the Levenberg-Marquardt algorithm [14].

5 Optimization of Shape and Appearance

We pose morphing as an optimization problem with an objective function that requires each frame in the output sequence of faces to be similar in shape (expression and pose) and appearance to a linear interpolation of these from the two sources. In addition, we want the shape and appearance to change smoothly from frame to frame.

Given the 3D shapes for the faces in input images A and B, we can easily interpolate their expression and 3D pose parameters to any intermediate view using an image warp induced from the 3D change. After that we employ an iterative optimization that maximizes the appearance similarity of the morphed sequence to the shape interpolated views while maintaining temporal coherence.

5.1 Prewarping the sources

The 3D face shapes contain two sets of parameters. The intrinsic parameters are the shape coefficients β, describing the facial geometry of a person exhibiting a facial expression. The external parameters include the rotation angles $\theta_x, \theta_y, \theta_z$, the scale c, and the 2D translations x_0, y_0, describing the 3D pose of the face. We linearly interpolate the intrinsic and external parameters of two input faces, resulting in a series of interpolated 3D face models. The interpolation of the each parameter p in frame t is defined as:

$$p_t = k_A p_A + k_B p_B, \quad (t = 1, ..., T) \tag{3}$$

where $k_A = (1 - \frac{t}{1+T})$ and $k_B = \frac{t}{1+T}$. Then we reconstruct the 3D shapes for all intermediate frames, as shown in Fig. 2 (bottom row).

To warp the input faces to the desired pose and expression, we apply "Expression Flow" [22]. As shown in Fig. 3, given two 3D shapes, we compute the difference between the projections of each corresponding vertex, and get a flow map. Applying the flow warps the original faces A and B to images A_t and B_t correspondingly with an interpolated pose and expression.

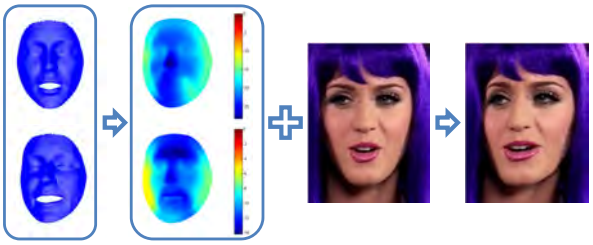

Figure 3: Warping the input image using expression flow. The flow map is computed by comparing the difference between two 3D models. The resulting image is the face warped to new pose and expression.

5.2 Appearance Optimization

A simple weighted averaging between A_t and B_t would give us an interpolated face C_t, as shown in Fig. 2. However this approach is not optimal, as each A_t and B_t pair is prewarped individually, thus temporal coherence is not guaranteed. Furthermore, C_t may be blurry due to misalignment between A_t and B_t. To further improve the quality of the morphing sequence, inspired by the Regenerative

Morphing method [17], we define a morphing energy function and employ an iterative optimization to minimize it.

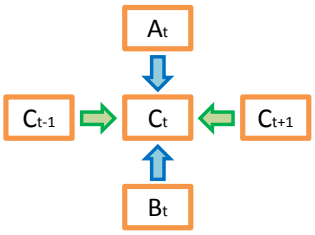

Figure 4: Appearance optimization.

First, every frame should be similar to the source images with an interpolated pose and expression. This prevents the appearance of the sequence from deviating too much from the source images. Second, the changes between every frame and its interpolated neighbor frames should be small. These two requirements give the following energy function in frame t.

$$\begin{aligned} E_t &= k_t^{(A)}||C_t - W_{f_{A_t}}(A_t)||^2 + k_t^{(B)}||C_t - W_{f_{B_t}}(B_t)||^2 \\ &+ k^{(C)}||C_t - W_{f_{C_{t-1}}}(C_{t-1})||^2 + k^{(C)}||C_t - W_{f_{C_{t+1}}}(C_{t+1})||^2 \end{aligned}$$

where A_t and B_t are input faces warped by expression flow and C_t is the interpolated face at frame t. In order to handle small residual misalignments between the warped sources, we apply a residual warp W induced by an optical flow [11] between A_t (or B_t, C_{t-1}, C_{t+1}) and C_t. We set $k^{(C)}$ to 1, and then $k_t^{(A)}$ and $k_t^{(B)}$ are interpolated between 0 and 1.

This quadratic energy function is minimized by a simple weighted sum of four images. Hence the face at frame t is updated as follows:

$$\begin{aligned} C_t &= k_t^{(A)} W_{f_{A_t}}(A_t) + k_t^{(B)} W_{f_{B_t}}(B_t) \\ &+ k^{(C)} W_{f_{C_{t-1}}}(C_{t-1}) + k^{(C)} W_{f_{C_{t+1}}}(C_{t+1}) \end{aligned} \quad (4)$$

We run typically two iterations for each frame and sweep over all frames in consecutive order for three times back and forth for a convergence of the entire morph sequence. Our algorithm is summarized in Algorithm 1.

Algorithm 1 *Appearance Optimization*

```
Fit 3D shapes to two input images A and B.
Compute interpolated 3D face shapes.
for all frames t do
    Warp input images A and B to interpolated states A_t and B_t.
end for
Initialize C_t as weighted sum of A_t and B_t.
repeat
    for all frames t do
        repeat
            Optimize C_t using Eq. 4.
        until frame converges
    end for
until sequence converges
```

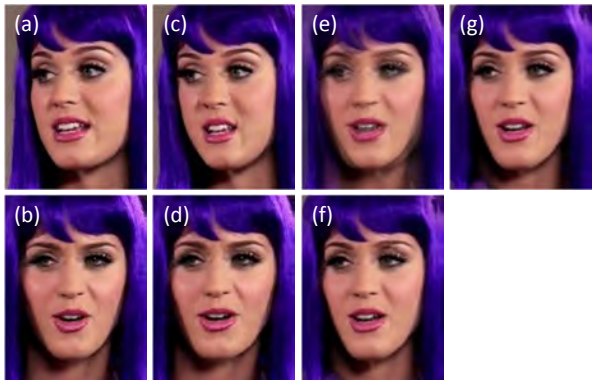

Figure 5: The result of disabling appearance optimization. (a) and (b) and two input images. (c) and (d) are acquired by warping the input images using 3D models. (e) is the result when using only similarity to the warped sources A_t and B_t (a weighted sum of (c) and (d)). (f) is the result when using only temporal smoothness (C_{t-1} and C_{t+1}). (g) is the result of the full system.

5.3 Warping background

The optimization above generates a morph sequence for the face region only. In addition, we apply optical flow based interpolation to warp the background outside the face region, by interpolating the flow for each frame. We use the Moving Least Squares [15] method to smoothly blur the difference between the two flows at the boundary between the regions.

6 Experiments

6.1 Face Morphing

We first apply the proposed method in face morphing between images of the same subject. The results on two pairs of images are shown in Fig. 9, and Fig. 8 with closed-up views. We compare our method with four previous morphing methods: registration-based crossfading, mesh morphing, optical flow, and regenerative morphing. For cross-fading, we first register the two input images by estimating a similarity transform between them using the detected feature points. This is similar to the frame transition method used in the Photobios system [9]. For mesh morphing, we triangulate the face image using the 68 feature points detected by the ASM method. We use the recently proposed optical flow method [11] for comparison. For regenerative morphing, we use the authors' implementation.

As shown in the figures, the face morphing results generated by previous methods contain noticeable artifacts, mainly due to the large pose and expression changes between input images. The results generated by our system have fewer artifacts and are of higher quality. We also noticed that results generated by regenerative morphing contain some temporal jittering, which is not visible in still images. Please refer to the supplementary video for temporal coherence comparison, as well as more results. We also apply face morphing between different subjects. One example is shown in Fig. 1. Another example is shown in Fig. 6, where we also compare the quality of interpolated faces with previous approaches.

To further evaluate the effectiveness of the proposed method, we compare our results against those generated by disabling appearance optimization. The 3D models are still used to warp the input images to the desired 3D pose and expression. As shown in Fig. 5(e), we take the weighted sum of the warped faces. The results are blurry because the faces are not aligned well. Fig. 5(f) shows the result when using only temporal smoothness (C_{t-1} and C_{t+1}),

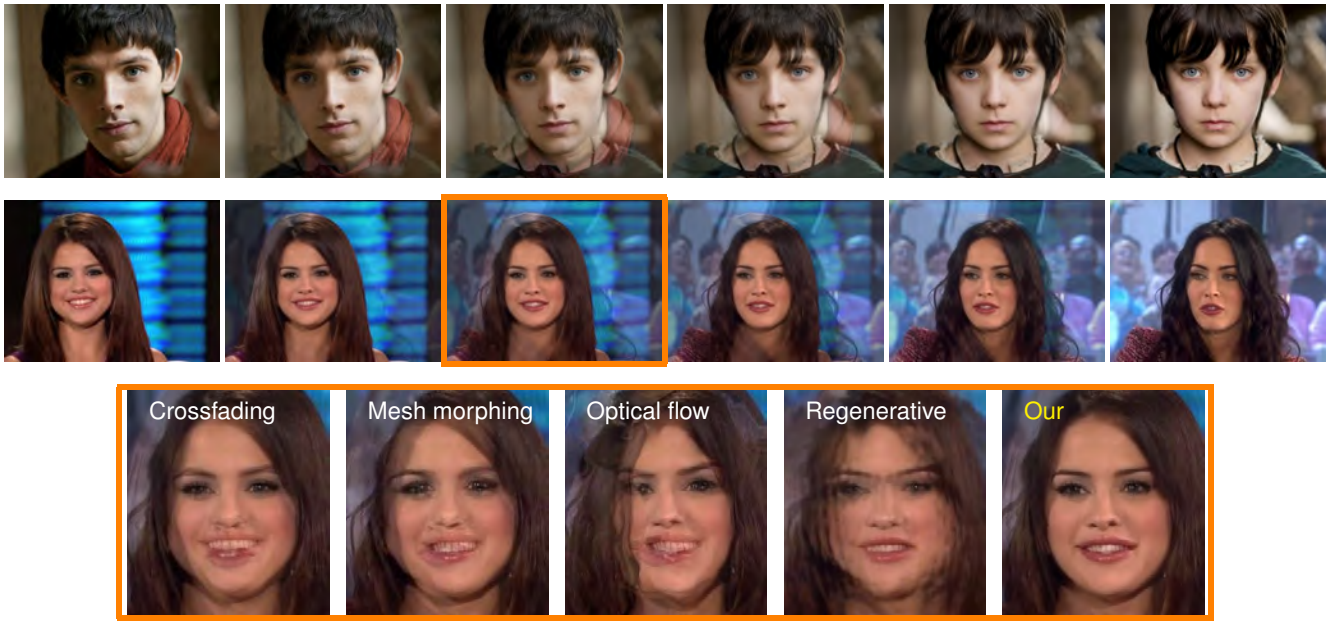

Figure 6: Morphing between different subjects. **Top two:** Our morphing result. **Bottom:** Comparison of different algorithms on the third frame.

which has artifacts above the right eye. After applying appearance optimization, our system can get good intermediate frames.

6.2 Replacing Undesired Expression

Our method can be used to stitch the video after removing a portion of it. This can be used to remove an undesired expression, such as a tic or a yawn. As shown in Figure 7, the frames highlighted in red are manually identified and removed. Our face morphing system could generate a smooth transition and stitch the video after removing a clip.

6.3 Timing

Our system is implemented in Matlab. It takes about 10 minutes to create 8 intermediate frames for face regions about 200 by 200 pixels, on a Intel CPU of 2.40 GHz. A large portion of the time is taken by the repeated optical flow computations. With latest GPU based optical flow method, the running time may be significantly reduced.

6.4 Limitations and Future Work

To be able to perform face morphing fully automatically, our system must rely on fully automatic performance of its components. Specifically, we rely on ASM for localizing facial components, and current ASM implementations can fail under large viewpoint variations or occlusions. We use optical flow to capture subtle geometric changes and build a residual warp. However, it might break when there are large differences in facial appearance, skin tones or illumination between two inputs. Therefore the current method is not effective for morphing between radically different people.

In this paper we focus on the problem of morphing faces and we do not have a sophisticated model for the background. Another direction for future work is in improving the model of the facial structure, such as explicitly modeling the locations of the pupils and ensuring they properly interpolate when the input images have large differences in gaze direction.

7 Conclusion

We address the problem of generating high quality face morphing animation given two faces of difference pose and expression. We show that a 3D subspace model learned from a small collection of human faces exhibiting realistic expression, constrains well the space of possible face deformations for interpolating casual face photos. Unlike traditional warping methods used for morphing that require accurate correspondence between the two source faces, we warp the two faces *independently* and only *roughly* using an 3D model based flow. Thus, we avoid traditional warping artifacts like fold-overs and "holes". The appearance optimization can recover small misalignment and other small changes not covered by the shape model and so traditional blurriness and hosting artifacts are suppressed. Future direction include handling partial occlusion (like sunglasses) using a more robust appearance optimization, face extrapolation, non-linear interpolation and applying this approach to other classes of objects for which effective 3D subspace models can be learned.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful feedback. Fei Yang and Dimitris Metaxas are partially supported by the following grants to Dimitris Metaxas: NSF-0549115, NSF-0725607, NASA-NSBRI-NBTS01601, NASA-NSBRI-NBTS004, ONR-N000140910104 and Adobe systems.

References

[1] T. Beier and S. Neely. Feature-based image metamorphosis. In *Proceedings of ACM SIGGRAPH*, pages 35–42, 1992.

[2] M. Bichsel. Automatic interpolation and recognition of face images by morphing. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 1996.

[3] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. *Comput. Graph. Forum*, 22(3):641–650, 2003.

[4] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of ACM SIGGRAPH*, pages 187–194, 1999.

[5] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models: Their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[6] Z. Deng and U. Neumann. *Data-Driven 3D Facial Animation*. Springer, 2008.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

Figure 7: Our method can stitch the video after removing a clip. **Top:** the original video. The frames to be removed are highlighted in red. **Bottom:** the result. The new frames interpolated are highlighted in green. In this example we align the before/after frames for clarity, but the padding need not have the same duration as the removed clip.

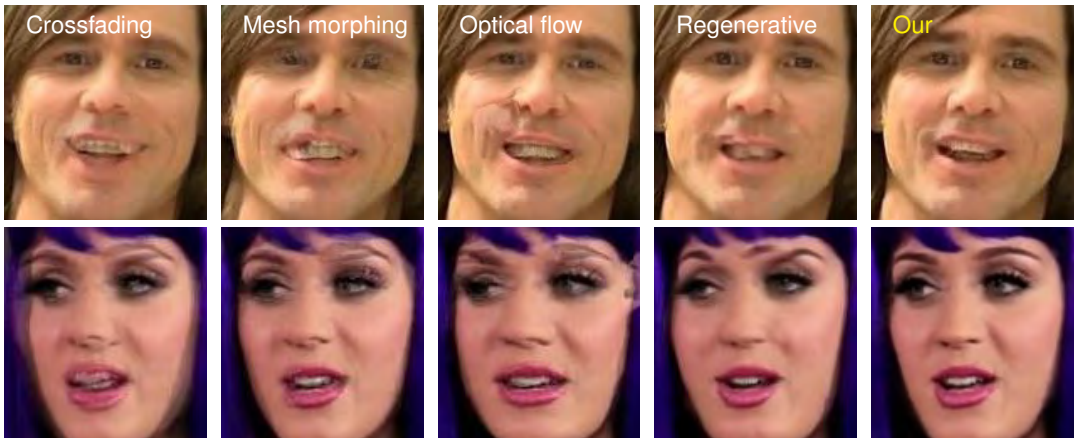

Figure 8: The close-up views of the middle column of Figure 9.

[8] P. Joshi, W. C. Tien, M. Desbrun, and F. H. Pighin. Learning controls for blend shape based realistic facial animation. In *ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003.

[9] I. Kemelmacher-Shlizerman, E. Shechtman, R. Garg, and S. M. Seitz. Exploring photobios. In *Proceedings of ACM SIGGRAPH*, volume 30, page 61, 2011.

[10] S. Lee, G. Woberg, K.-Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Trans. Visualization and Computer Graphics*, 2(4):337–354, 1996.

[11] C. Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. Doctoral Thesis, Massachusetts Institute of Technology, 2009.

[12] D. Mahajan, F.-C. Huang, W. Matusik, R. Ramamoorthi, and P. Belhumeur. Moving gradients: a path-based method for plausible image interpolation. *ACM Trans. Graphics*, 28:42:1–42:11, 2009.

[13] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Proceedings of ECCV*, pages 504–513, 2008.

[14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, second edition, July 2006.

[15] S. Schaefer, T. Mcphail, and J. Warren. Image deformation using moving least squares. In *Proceedings of ACM SIGGRAPH*, pages 533–540, 2006.

[16] S. M. Seitz and C. R. Dyer. View morphing. In *Proceedings of ACM SIGGRAPH*, pages 21–30, 1996.

[17] E. Shechtman, A. Rav-Acha, M. Irani, and S. Seitz. Regenerative morphing. In *Proceedings of CVPR*, San-Francisco, CA, 2010.

[18] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. In *Proceedings of ACM SIGGRAPH*, volume 24, pages 426–433, 2005.

[19] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3-d facial expressions. In *Proceedings of EuroGraphics*, pages 677–686, 2004.

[20] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA, 1990.

[21] G. Wolberg. Recent advances in image morphing. In *Computer Graphics International*, 1996.

[22] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas. Expression flow for 3D-aware face component transfer. In *Proceedings of ACM SIGGRAPH*, 2011.

[23] V. Zanella, H. Vargas, and L. V. Rosas. Active shape models and evolution strategies to automatic face morphing. In *Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms, Part II*, pages 564–571, Berlin, Heidelberg, 2007.

Figure 9: Face morphing results for two subjects. For each subject: **Top row**: crossfading. **2nd row**: mesh morphing. **3rd row**: simple optical flow. **4th row**: regenerative morphing. **Bottom row**: our method. The close-up views of the middle columns are shown in Figure 8.