EXPRESSION MORPHING FROM DISTANT VIEWPOINTS

Tao Fu
Dept. of Electrical & Computer Engineering
University of Central Florida, FL, USA
taf@bruce.engr.ucf.edu

Hassan Foroosh
School of Computer Science
University of Central Florida, FL, USA
foroosh@cs.ucf.edu

ABSTRACT

In this paper, we propose an image-based approach to photo-realistic view synthesis by integrating field morphing and view morphing in a single framework. We thus provide a unified technique for synthesizing new images that include both viewpoint changes and object deformations. For view morphing, we relax the requirement of monotonicity along epipolar lines to piece-wise monotonicity, by incorporating a segmentation stage prior to interpolation. This allows for dealing with occlusions and visibility issues, and hence alleviates the "ghosting effects" that typically occur when morphing is performed between distant viewpoints. We have particularly applied our approach to the synthesis of human facial expressions, while allowing for wide change of viewing positions and directions.

1. INTRODUCTION

A central goal in image-based modeling and rendering is view synthesis using a small number of reference images. This can have wide range of applications from the obvious cases of generating powerful visual effects to more low-level applications such as video compression. Generally speaking, image morphing is a coupling of image warping with color interpolation [1]. After specifying corresponding features on reference images (such as points, lines, and mesh nodes), a pixel mapping function is defined to interpolate the geometry between the reference images. In addition to geometric interpolation, a color interpolation process (mostly cross-dissolve) is then applied to generate in-between images. The most compelling results of image morphing are those involving transformations from one person to another, and morphing between different expressions of the same person.

One of the earliest methodologies used for image morphing was mesh warping [2], which is based on interpolating between corresponding mesh nodes as image features. Field morphing, which was developed in [3], reduced the complexity by considering correspondences between a sparse set of feature lines and defining a simple mapping function. The mapping functions have also been studied in the past based on scattered data interpolation [4-6]. An important class of view synthesis techniques that are also purely 2D are the methods that are based on the plenoptic function [7-12]. The aim is to model the chromacity of the light observed from every position and direction in the 3D space by a 5D function, and hence synthesize arbitrary images by sampling this function.

More recently, view synthesis based on modeling the underlying 3D structure has gained popularity. These methods can be broadly divided into two categories. Methods that explicitly model the 3D structure [13,14], and those that only implicitly include the 3D information [15]. Our work, which belongs to the latter category, focuses on two main issues:

- Image synthesis in the presence of both non-rigid object deformations and rigid camera/object motions, e.g. synthesizing facial expression of a moving head.
- Realistic synthesis of new views from distant locations and orientations of cameras with occluded or partially viewed regions.

The approach that we propose integrates field morphing [3] and view morphing [15] in a single framework. It takes advantage of field morphing's ability to morph one expression to another for the same viewpoint, and view morphing's ability to morph same expression from different viewpoints. Given four reference images, we generate a realistic morphing from one viewpoint with one expression to another viewpoint with a different expression.

2. PROBLEM DESCRIPTION

Suppose we have four images of the same person: one pair (I_{lc}, I_{rc}) from two different directions but with the same facial expression (closed mouth), and another (I_{lo}, I_{ro}) with similar camera poses but a different facial expression (open mouth). Our goal is to synthesize a realistic video sequence from these images using only 2D information, so that it appears that the person's head moves from left to right as the facial expression gradually changes.

This work was partially supported by ONR grant #N000140410512 and Sun Microsystem's grant #EDUD-7824-030482-US

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Unlike the problem of view morphing, where only the camera moves from left to right, in this case the object also undergoes deformations, i.e. the object motion is not completely rigid and the amount of motion varies for different parts of the object. So the problem is out of the range of classical view morphing techniques. On the other hand, applying field morphing cannot solve the problem either, since we have distant viewpoints for each pair. Field morphing is typically suitable for expression changes in the same viewpoint. Changing the position and the orientation of the camera will cause severe distortions in the in-between synthesized image frames.

We propose an approach based on integrating field morphing and view morphing. Our technique takes advantage of field morphing's ability to morph one expression to another from the same viewpoint, and view morphing's ability to morph same expression between different viewpoints. The steps involved are described below.

2.1 Prewarping

The basic idea in this stage is to rectify the image pairs so that as described in [15] the image planes become parallel (i.e. with parallel epipolar lines). There are many existing techniques that would allow us to rectify a pair of images in either case of calibrated and uncalibrated cameras[16]. For completeness, we describe an approach similar to [15]. In practice, the goal is to make the epipolar lines parallel to the x-axis by mapping the epipoles to the infinite point $\begin{bmatrix}1 & 0 & 0\end{bmatrix}^T$. The resulting fundamental matrix of rectified (prewarped) images will then be of the form

$$\mathbf{F}_w = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

In order to obtain such configuration, we need to determine two homographies $\mathbf{H}_l^{-1}$ and $\mathbf{H}_r^{-1}$ so that when applied to the left and right images respectively would yield a rectified pair. Based on the Longuet-Higgins equation [16], this implies that the two homographies should satisfy

$$\mathbf{H}_l \mathbf{F} \mathbf{H}_r = \mathbf{F}_w \tag{2.2}$$

where $\mathbf{F}$ is the fundamental matrix prior to rectification, which can be found using the well-known 8-point algorithm [17].

Now, let Π be a plane parallel to the line through the camera projection centers, i.e. $\mathbf{C}_0\mathbf{C}_1$. Suppose Π intersects the image plane I_i at $\mathbf{d}_i$, then the rotation of I_i about any line parallel to $\mathbf{d}_i$ will make the two image planes parallel. If we choose the line that passes through the camera center i.e. $\mathbf{d}_i = \begin{bmatrix} -d_i^x & d_i^y & 0 \end{bmatrix}^T$, then the homography that will rectify I_i is given up to an affine transformation by

$$\mathbf{R}_{\theta_i}^{d_i} = \begin{bmatrix} (d_i^x)^2 + (1-(d_i^x)^2)\cos\theta_i & d_i^x d_i^y (1-\cos\theta_i) & d_i^y \sin\theta_i \\ d_i^x d_i^y (1-\cos\theta_i) & (d_i^y)^2 + (1-(d_i^y)^2)\cos\theta_i & -d_i^x \sin\theta_i \\ -d_i^y \sin\theta_i & d_i^x \sin\theta_i & \cos\theta_i \end{bmatrix} \tag{2.3}$$

which corresponds to the rotation of the image plane about $\mathbf{d}_i$ by an angle θ_i. To make sure that this results in a rectified pair the rotation must be chosen such that the epipoles $\mathbf{e}_i = \begin{bmatrix} e_i^x & e_i^y & 1 \end{bmatrix}^T$ are mapped to infinite points $\mathbf{e}_i = \begin{bmatrix} e_i^x & e_i^y & 0 \end{bmatrix}^T$. It therefore follows that

$$\theta_i = \tan^{-1}\left(\frac{e_i^z}{d_i^y e_i^x - d_i^x e_i^y} \right) \tag{2.4}$$

The epipoles themselves are readily obtained by solving the following homogeneous equations using singular value decomposition

$$\mathbf{F}\mathbf{e}_l = 0 \quad \text{and} \quad \mathbf{F}^T \mathbf{e}_\mathbf{r} = 0 \tag{2.5}$$

In order to minimize the rotation angle, we choose the line $\mathbf{d_0}$ = [-d_0^y d_0^x 0]^T = α[e_0^x e_0^y 0].

Once the image planes are aligned, the next step is to rotate the image planes about the z-axis so that the epipolar lines are horizontal.

$$\mathbf{R}_{\phi \mathrm{i}} = \begin{bmatrix} \cos\phi_i & -\sin\phi_i & 0 \\ \sin\phi_i & \cos\phi_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.6}$$

Where $\phi_i = -\tan^{-1}(e_i^y / e_i^x)$.

After applying these image plane rotations, the fundamental matrix has the form:

$$\mathbf{F_w} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & 1 & b \end{bmatrix} \tag{2.7}$$

In order to make $\mathbf{F_w}$ of the form in (2.1), the second image should be scaled and translated by the matrix

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -a & -b \\ 0 & 0 & 1 \end{bmatrix} \tag{2.8}$$

So the prewarping transforms are given by

$$\mathbf{H}_l^{-1} = \mathbf{R}_{\phi l} \mathbf{R}_{\theta l}^{dl} \tag{2.9}$$

$$\mathbf{H}_r^{-1} = \mathbf{T}\mathbf{R}_{\phi r} \mathbf{R}_{\theta r}^{dr} \tag{2.10}$$

2.2 Generating In-Between Images

Once the images are rectified, we can generate the in-between images to synthesize a video sequence. This is where our approach departs from the classical view morphing techniques [15]. In the classical approach in-between frames are synthesized by simply performing linear interpolation between feature points along the rectified epipolar lines (i.e. the scan lines). This has two major shortcomings:

- First and foremost it makes the strong assumption of monotonicity of feature points, i.e. assumes that the order along the epipolar lines is preserved. A frequently occurring situation where this assumption fails is caused by occlusions that often happen when rendering from distant viewpoints. In the presence of occlusions, a highly undesirable effect that occurs is the so called "ghosting effect".
- A second important shortcoming is that it assumes the color remains invariant between two feature points along the epipolar lines. This not only fails for non-Lambertian surfaces, but also limits the morphing to rigid motions of the object. For instance, when morphing face images with different expressions, the morphing process assigns inaccurate shape and color values to non-rigid regions.

In order to tackle these problems we propose to relax the monotonicity assumption to piece-wise monotonicity. This implies that we segment the rectified images into sub-regions so that no sub-region can cause self-occlusion. Of course, since local deformations (due to for instance facial expressions) can cause self-occlusion, each deformable part of the object needs to be at least segmented as a separate region. Given the segment boundaries as features and other additional feature lines, we then apply field morphing [] for interpolation within a segment along the epipolar lines. If a segment is visible in both images then we can use cross-dissolve, otherwise color will only be contributed by the image in which the region is visible.

Although the above approach alleviates the problems caused by occlusions, it generates an artifact that needs to be handled, i.e. the region boundaries appear as seams in the synthesized views after post-warping. However, fortunately this problem is less severe compared to occlusions and ghosting effects and can be readily solved using a blending (feathering) technique similar to commonly used ones in image mosaicing.

The final results for expression morphing are generated by using the following two steps

- First generate two synthesized video sequences, i.e. one for each expression with camera moving from one position to another, using the approach described above.
- Then generate a third video sequence by applying segment-wise field morphing between corresponding frames in the above synthesized sequences.

The result is a compelling effect of photo-realistic expression morphing.

3. RESULTS & CONCLUDING REMARKS

We have performed an extensive set of experimentations with an example of a face image shown below. Figure 1 illustrates our motivation to develop the technique described above for eliminating ghosting artifacts. Two of the four reference images used as the input data are illustrated together with prewarped and interpolated images using both classical view morphing and our approach. Figure 1-(d) shows the segmentation used for imposing piece-wise monotonicity. Accordingly, occluded and non-occluded regions are handled by morphing the corresponding segmented regions separately along the epipolar lines. Figure 2 shows the output of the algorithm. Three video sequences are generated from the four reference images using the approach described above. The top sequence is between left and right reference images with closed mouth. The bottom sequence is the morphing between the open-mouth images. And the middle sequence is the field morphing between the two sequences.

Comparing the results with classical morphing, we obtain a superior performance, which avoids ghosting effects due to occlusions and wide changes of viewing angles, while allowing for local non-rigid deformations such as facial expressions.


(a) (b) (c) (d) (e)

Figure 1: (a) & (b) reference images for closed mouth pair, (c) example of a prewarped and interpolated image using classical view morphing, (d) a segmented and labled prewarped reference image, (e) an example of a prewarped and interpolated image using our approach.

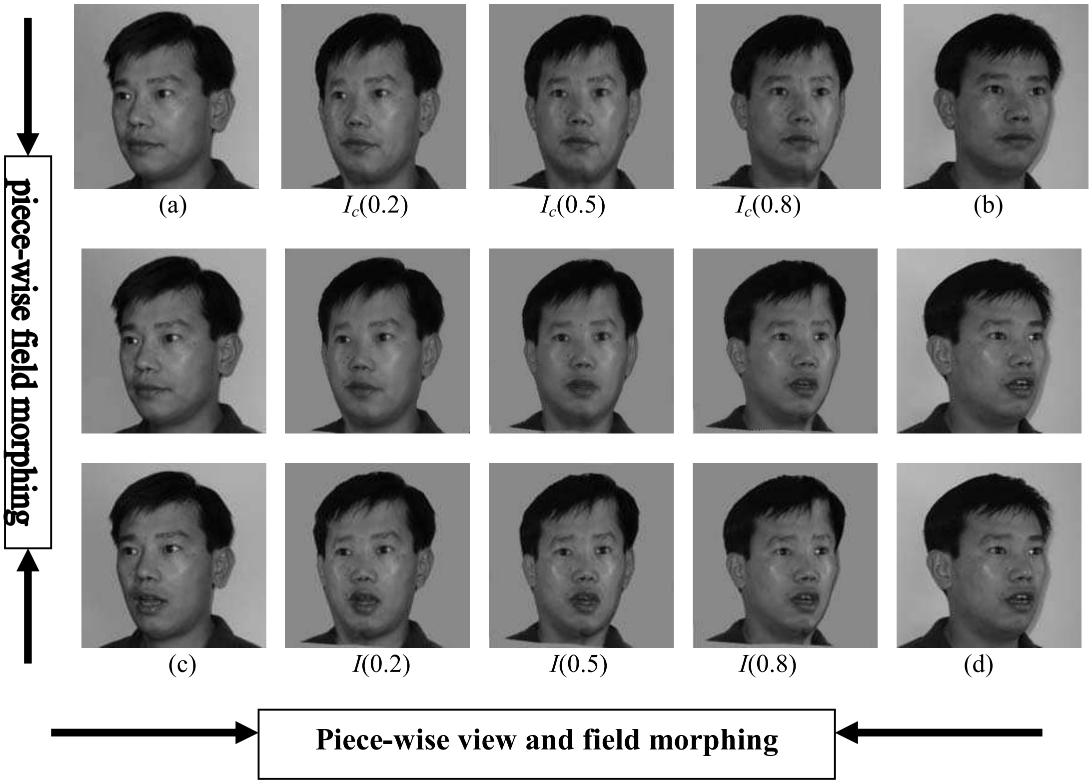

Figure 2: (a), (b), (c) and (d) reference images with different expressions and viewing directions. Top row synthesized in-between views for closed mouth sequence, bottom row: synthesized sequence for open mouth, middle row: synthesized sequence for both viewpoint and expression changes.

5. REFERENCES

[1] G. Wolberg. Recent advances in image morphing. *Computer graphics international,* 1996

[2] George Wolberg. *Digital image warping*. IEEE Computer Society Press, 1990

[3] T. Beier and S. Neely. Feature-based image metamorphosis. In *Proc. SIGGRAPH 92*, 35 – 42, 1992.

[4] D. Ruprecht and H. Muller. Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications,* 1995

[5] N. Arad, N. Dyn, D. Reisfeld, and Y. Yeshurun. Image warping by radial basis functions: Applications to facial expressions. *CVGIP: Graphical Models and Image Procession.* 1994

[6] S.-Y. Lee, K.-Y. Chwa, J. Hahn, and S.Y. Shin. Image morphing using deformation techniques. *The Journal of Visualizatin and Computer Animatin,* 1996

[7] E.H. Adelson, J.R. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing* The MIT Preess, 1991

[8] L. McMillan, G. Bishop Plenoptic modeling: An image-based rendering system. *SIGGRAPH,* 1995

[9] M. Levoy, P. Hanrahan. Light field rendering. *Proc. SIGGRAPH 96,*1996

[10] S.J. Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen. The lumigraph. *Proc. SIGGRAPH 96*, 1996

[11] Heung-Yeung Shum, Li-Wei He. Rendering with concentric mosaics. *Proc. SIGGRAPH 99*, 1999

[12] Shmuel Peleg and Joshua Herman. Panoramic mosaics by manifold projection. *Proc CVPR 97,* 1997

[13] P.E. Debevec, C.J. Taylor, J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *SIGGRAPH,* 1996

[14] K.N. Kutulakos and S.M. Seitz. *A theory of shape by space carving*, IJCV, 38(3):197-216, 2000.

[15] S.M. Seitz and C.R. Dyer, Toward Image-Based Scene Representation Using View Morphing *Proc. ICPR, vol. I*, pp. 84-89 1996

[16] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000

[17] R.I. Hartley. In defense of the 8-point algorithm. *Proc. Int. Conf. On Computer Vision*, pp. 1064 – 1070 1995