Timo Stich ; Marcus Magnor

Image Morphing for Space-Time Interpolation

Auch erschienen als:
Technical Report 2007-4-2. - Computer Graphics Lab, TU Braunschweig

TIMO STICH
stich@cg.tu-bs.de

Prof. Dr. Ing. MARCUS MAGNOR
magnor@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Image Morphing for Space-Time Interpolation

Technical Report 2007-4-2

April 16, 2007

Computer Graphics Lab, TU Braunschweig

Rendering convincing transitions between individual pictures is the main challenge in image-based rendering and keyframe animation as well as the prerequisite for many stunning visual effects. We present a perception-based method for automatic image interpolation, achieving psycho-visually plausible transitions between real-world images in real-time. Based on recent discoveries in perception research, we propose an optical flow-based warping refinement method and an adaptive nonlinear image blending scheme to guarantee perceptional plausibility of the interpolated in-between images. Conventional, uncalibrated photographs suffice to convincingly interpolate across space, time, and between different objects, without the need to recover 3D scene geometry, actual motion, or camera calibration. Using off-the-shelf digital cameras, we demonstrate how to continuously navigate the viewpoint between camera positions and shutter release times, how to animate still pictures, create smooth camera motion paths, and how to convincingly morph between depictions of different objects.

0.1 Introduction

If what we perceive with our eyes changes with time our brain automatically attempts to interpret the changing visual input in terms of plausible motion of the viewpoint and/or of the observed object or scene [Ellis 1938; Graham 1965; Giese and Poggio 2000; Giese and Poggio 2003]. In the physical world, the rules that define plausible motion are set by temporal coherence, parallax, and perspective projection. Our brain, however, refuses to feel constrained by the unrelenting laws of physics in what it deems plausible motion. Image metamorphosis experiments, in which unnatural, impossible in-between images are interpolated, demonstrate that our brain, under certain circumstances, willingly accepts chimeric images as plausible transition stages between images of actual, known objects [Beier and Neely 1992; Seitz and Dyer 1996; Wolberg 1998]. Another example are cartoon animations which for the longest time were hand-drawn pieces of art that didn't need to succumb to physical correctness. The goal of our work is to exploit this freedom of perception for space-time interpolation, i.e., to generate transitions between still images that our brain accepts as plausible motion.

The notion of using interpolation techniques to synthesize images at intermediate viewpoints is not new. Chen and Williams already proposed to use image interpolation to synthesize in-between views from a set of images[1993]. To ensure perceptional plausibility, however, their method relies on physical consistency by enforcing epipolar constraints for which the input images must be calibrated. All subsequent image-based rendering (IBR) techniques rely on physically consistent interpolation: besides the images, additional camera calibration parameters, and frequently also scene geometry, are needed to compute new views according to the physical laws of image formation [Levoy and Hanrahan 1996; Gortler et al. 1996; Debevec et al. 1998; Wood et al. 2000; Isaksen et al. 2000; Buehler et al. 2001; Snavely et al. 2006]. For time-varying scenes the acquisition cameras must all be synchronized to be able to relate images of the same instant across cameras in addition to calibration [Matusik et al. 2000; Carranza et al. 2003; Matusik and Pfister 2004; Zitnick et al. 2004; Vedula et al. 2005]. Clearly, the need for calibrated, synchronized acquisition is highly inconvenient as it implies time-consuming recording preparations as well as expensive acquisition hardware. Instead, an image interpolation approach that takes into account how our visual brain processes image sequences is able to provide plausible interpolation results across space and time from nothing more than a collection of unsynchronized, uncalibrated images.

Besides view interpolation and keyframe animation, convincing space-time interpolation is also the key to stunning visual effects [Wolf 2006]. Slow motion, frozen moments, multiple exposures, time, space or motion blur, and many more effects [Inc. 2007] are straight-forward to create with convincing space-time interpolation.

In this paper, we describe how to perform image-based rendering of arbitrary time-varying scenes based on perception-aware image morphing. In contrast to previous work done in IBR, we do not enforce physical correctness but optimize

for perceptional plausibility. In particular, this paper contributes a framework to render convincing in-between views and in-between time instants from nothing more than a set of uncalibrated, unsynchronized images.

0.2 Related Work

Image-based rendering (IBR) methods achieve highly realistic rendering results of natural objects or scenes from a collection of calibrated photographs. While some IBR methods rely solely on image number to minimize aliasing artifacts [Levoy and Hanrahan 1996; Matusik and Pfister 2004], most IBR approaches make additional use of epipolar constraints [Chen and Williams 1993; McMillan and Bishop 1995; Seitz and Dyer 1996; Matusik et al. 2000], scene depth [Gortler et al. 1996; Isaksen et al. 2000; Buehler et al. 2001; Zitnick et al. 2004], or full 3D geometry information [Debevec et al. 1998; Wood et al. 2000; Carranza et al. 2003; Vedula et al. 2005; Snavely et al. 2006].

For many potential application scenarios, the crucial handicap of IBR is the need for accurate camera calibration, additional geometry modeling, and/or synchronized acquisition. These limitations make data acquisition for IBR a time-consuming and delicate endeavour which typically requires a controlled environment and expensive equipment.

Image metamorphosis, or image morphing, denotes interpolation between images of different objects from (user-defined) correspondences alone, i.e., without any additional information such as geometry or camera calibration. Well-known is the line-based morphing method proposed by Beier and Neely [1992] from its use in Michael Jackson's music video "Black & White". Lerios et al. [1995] extended the approach to 3D voxels and addressed ghosting artifacts by correcting the warp field. Other warping techniques have been discussed by Wolberg [1998], including the popular thin-plate spline interpolation which is based on point correspondences. A computationally more complex method based on line features was recently proposed by Schaefer et al. [2006]

Image morphing algorithms are based on a dense 2D vector field of correspondences along which both images are warped and linearly blended to obtain in-between images. This simplistic motion model, however, does not allow one to properly handle, e.g., occlusions and disocclusions. Recent advances in perceptio research give clues on how non-linear blending can be employed to perceptionally conceal otherwise annoyingly visible inconsistencies of in-between images [Giese and Poggio 2000; Giese and Poggio 2003].

The optical flow plays a major role in perceptional motion analysis [Giese and Poggio 2003]. Since the pioneering work on local and global optical flow reconstruction by Lucas and Kanade [1981] and Horn and Schunck [1981], respectively,

a multitude of computational approaches have been devised and various fields of application have been discovered [Barron et al. 1994; Baker and Matthews 2004].

2D optical flow cannot represent occlusions or disocclusions. It allows, however, to refine the warping field during image morphing to ensure perceptional consistency with respect to our brain's visual processing pipeline [Giese and Poggio 2003].

Image blending as a processing step in image compositing is traditionally realized as the linear combination of images [Burt and Adelson 1983; Porter and Duff 1984]. Only recently, Grundland et al. [2006] proposed different non-linear blending functions to preserve image contrast, color, or salient regions. When applied to image morphing, however, preserving any of these characteristics can prove detrimental since occlusion artifacts could actually become amplified if, e.g., a more salient foreground vanishes into a less salient background. Nevertheless, by adapting the blending function perceptionally, non-linear blending can also conceal (dis)occlusions during image morphing.

In the following, we give a brief overview of recent psychophysics research results that give clues on how to generate perceptionally plausible in-between images. In Sect. 0.4, we describe our perception-aware framework for space-time interpolation based on image morphing. After giving implementation details, we demonstrate in Sect. 0.6 that our framework yields convincing transitions across space and time from uncalibrated still images obtained using eight off-the-shelf digital still cameras. We conclude by pointing out promising applications of our framework in movie production, visual media, and telecommunications.

0.3 Visual Motion Perception

A wealth of research has been published on visual perception, and it is possible here to give only a very incomplete overview of the topic. One source we draw on is Gestalt theory [Ellis 1938] which asserts that human perception of motion depends on the intensity as well as spatial and temporal distance between visual stimuli [Graham 1965]. By presenting the eye with still images at a high enough rate, consistent movement is perceived. This is the basic principle of all movies since Lumiere's invention of the Cinematographe in 1895. We also take into account that contours and image regions of high spatial frequency play a key role in motion interpretation [Movshon et al. 1986; Weiss et al. 2002; McDermott and Adelson 2004]. A comprehensive overview paper on biological motion perception has been compiled recently by Giese and Poggio [2003]. The authors describe a hierarchical neural model for motion perception that postulates two separate visual processing pipelines for motion, the *form pathway* and the *motion pathway*.

To perceive plausible motion, we assume that both pathways must separately interpret the incoming image sequence as being consistent. The *form pathway* recognizes (biological) motion as a sequence of individual 'snapshots'. For these

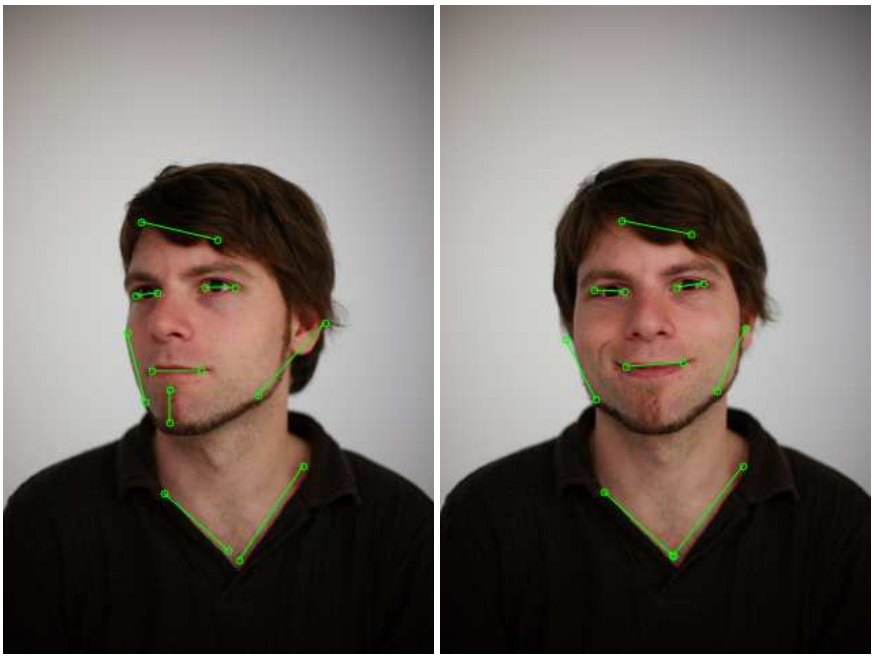

Figure 1: Corresponding line features between source and target image can be found automatically or specified manually, allowing large numbers of images to be matched in a short time.

snapshots to be perceived consistent, they have to be individually recognizable by the form pathway as plausible objects. Hence, each snapshot must exhibit low-level visual features typical of objects, such as contours. Research on monkeys also indicates that part of a 3D object recognition consists of neurons that represent explicit 2D views of objects [Logothetis et al. 1995], while other neurons capture perceptual invariances and fire for any view of the object [Riesenhuber and Poggio 2002]. Consequently, each snapshot must bear sufficient resemblance to known objects, and the perceptional invariances must not vary too much over the image sequence.

The *motion pathway* recognizes (biological) motion by analyzing optical-flow patterns. Local motion-detecting neurons provide what is essentially the optical flow between subsequent images. The optical flow is analyzed with respect to translational flow as well as with respect to motion edges. For plausible motion, both optical flow components must be consistent and may not change abruptly between images. Similar to the confirmed “snapshot neurons” of the form pathway, Giese and Poggio postulate the existence of optical-flow pattern neurons along the motion pathway [2003]. These are selective to learned optical flow patterns, e.g., of 3D rotation or human gait. Consequently, the optical flow must conform to general learned motion patterns, such as translation and rotation. If the form pathway recognizes, e.g., a person, the motion must additionally conform to the learned

optical flow pattern of human gait.

To synthesize complex motion patterns, Giese and Poggio propose to employ a morphing approach [2000]. In contrast to our work, however, their contribution concentrates solely on motion synthesis and is not concerned with image synthesis. In the following, we propose a framework to generate transitions between images that are to be perceived as plausible motion.

0.4 Perception Based Image Morphing

The inputs to our system are a set of uncalibrated images and corresponding line features between these images which are specified manually or found automatically [Anonymous 2007] (cf Fig. 1). Image morphing is the combination of image warping and image blending to interpolate between two arbitrary images $I^j : \mathbb{R}^2 \to \mathbb{R}$. Given a warping function $W : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ and a blending function $b : \mathbb{R} \to \mathbb{R}$, we can write the morphing function $M : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$

$$M(x,t) = b(t)\ (I^A \circ W)(x,t) + (1-b(t))\ (I^B \circ W)(x,1-t) \tag{1}$$

where $I^j \circ W$ denotes the image I^j warped by W.

Schaefer et al. [2006] demonstrate that simple motion can – to some extent – be created from a single image by means of image warping alone. Complex motion such as rotations involving lighting change, occlusion and object change, however, can only be created by interpolating between several images. The reason is that warping alone cannot account for lighting/color changes, nor does it resolve inconstitencies due to occlusion/object change, both of which are necessary to create perceptionally plausible motion.

In what follows, we propose a system for perceptionally motivated space-time image interpolation, addressing the potential artifact sources. First, we introduce a warp field refinement to address small mismatches in the input warping. We then develop a robust classification for the remaining image discrepancies. Based on this classification, we put forward an adaptive non-linear image blending scheme, significantly increasing the plausibility of the perceived motion in the image interpolation results.

0.4.1 Warp Field Refinement

The warping defined by line features as proposed by [Beier and Neely 1992] still produces small but noticeable artifacts, especially at object borders, Fig. 2. This is due to the nature of the warp field construction which smoothly interpolates the linear transformations of line features, producing errors when matching curved contours. A naïve solution is to use more line features to better approximate the contour at the pace of more time-consuming user interaction.

Figure 2: Comparison of the morphing results of the same in-between image: without (left) and with (right) warping field refinement. Artifacts due to contour mismatches are efficiently reduced.

Instead, we propose to refine the warping function W at each time step by an additive warping correction term $W_{Correction}$

$$\hat{W}(t) = W(t) + W_{Correction}(t), \tag{2}$$

which can be computed automatically using optical flow estimation.

In general, optical flow estimation performs very well in finding a spatially close match of two images. It runs into problems when correspondences over large spatial distances have to be established and produces unpredictable results in the presence of occlusions. Starting from a warped pair of images, the mentioned problems of optical flow estimation are effectively circumvented because a first approximation of pixel correspondences is already given. This is also the theoretical foundation for the most sophisticated optical flow estimation algorithms to date [Papenberg et al. 2006].

$W_{Correction}$ is hence defined as a linear combination of optical flows $F : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ between the warped images at t:

$$\begin{aligned} W_{Correction}(x,t) = t\, F((I^A \circ W)(x,t), (I^B \circ W)(x,1-t)) + \\ (1-t)\, F((I^B \circ W)(x,1-t), (I^A \circ W)(x,t)) \end{aligned} \tag{3}$$

Substituting W by $\hat{W}$ in (1), we can significantly reduce artifacts due to small image mismatches, Fig. 2. Summarizing, the warping defined by the line features can be seen as a pre-warping of the images to establish a first approximation of pixel correspondences, which is then refined to sub-pixel accuracy using optical flow estimation.

0.4.2 Classifying Image Differences

Given the refined warp field, we now need to deal with the two remaining sources of error, i.e., changes in color or brightness and mismatches due to occlusion or object changes. By object changes we mean actual variations in appearance of the object over time, for example opening or closing of the eyes. We first have to decide on a per-pixel basis which class of error is responsible for a perceived image difference. Since we already have dense correspondence between the images, this distinction can be based on color difference.

As we are interested in the *perceived* color and brightness distance, we use the CIE Lab color space [CIE 1976] to measure the difference between the warped images. We propose a weighted color difference, emphasizing color change and reducing the influence of brightness changes:

$$D_{Color}(I^A, I^B) = ||I_a^A - I_a^B||^2 + ||I_b^A - I_b^B||^2 + \alpha \cdot ||I_L^A - I_L^B||^2 \tag{4}$$

where I_a and I_b denotes the color channels a and b, I_L is the L channel from Lab space and α is a weighting constant. We found a value of 0.25 for α to produce good results in our experiments.

The problem can now be seen as a labeling problem where each pixel is labeled according to its error class. This is similar to background subtraction and seamless texture combination, and we choose a graph-cut approach to find a solution. Each node in the graph corresponds to one pixel in the image and is connected to the source and sink node. The weights are defined as the color distance between I^A and its warped counterpart $I^B \circ W$

$$w_{source} = \beta - D_{Color}(I^A, I^B \circ W) \tag{5}$$

and

$$w_{sink} = D_{Color}(I^A, I^B \circ W) - \beta \tag{6}$$

where β denotes the classification threshold. For each neighboring pixel, an edge between the corresponding nodes with constant flow Φ is added to regularize the problem. The minimal cut on the so constructed graph then gives a labeling of each pixel to one of the two classes, where all nodes connected to the source are classified as brightness changes, and the nodes connected to the sink are classified as artifacts due to occlusion or object change. To get the final mask we combine the labeling results of the pixels of I^A and the labeling of I^B computed by analogy.

An example of a mask derived from the labeling is depicted in Fig. 3. Note that the mask has to be computed only once and is warped in parallel to the images for rendering.

0.4.3 Non-linear Image Blending

Psycho-visual optimality in the context of image morphing aims at finding the blending function $b(t)$ in (1) that yields the perceptually most plausible motion

between two images. Using our previous classification of the image pixels, we can rephrase this problem into how to blend the two different color discrepancy classes so that a visually optimal motion is perceived.

Differences in luminance can be very easily addressed by simple linear interpolation. This is the basis of most previous image morphing approaches and has proven well in the case of non-occluding surfaces and static scenes. In the case of known camera calibration Seitz and Dyer [1995] could even prove that the interpolated views are physically correct. Therefore, the number of plausibly interpolated images is unlimited as all of them appear correct to the human observer as they are very close to physical correctness.

In the case of object changes and occlusions however, these regions cannot be addressed in the same way. As no match in the other image is given, linear blending causes artifacts as can be seen in Fig. 3. Although there is no physical information available about how these regions should transform into each other, the brain can still detect plausible motions between the images. As long as the perceived visual impulses are similar to the predicted visual impulses, consistent motion is perceived, while perceiving a deviating motion impulse leads to distracting errors. Using a linear blending in these regions hence leads to a fade-in fade-out motion, which is often a significant deviation from the predicted motion and results in an interrupted motion experience. Using this observation and the basic definition of motion perception, we conclude that presenting *wrong* motion information in the blending step is worse than avoiding motion perception through blending in these regions. Specifically, if we use a linear warping scheme combined with a non-linear blending scheme for problematic regions, we achieve improved motion results with less artifacts while obtaining as-smooth-as-possible motion between two static images.

To meet this behavior we propose a scaled standard logistic function $b_s(t)$ in the blending step with steepness parameter s

$$b_s(t) = \frac{C_s}{1 + e^{-t\,s}} \tag{7}$$

where C_s is a normalization constant dependent on s so that $b_s(0) = 0$ and $b_s(1) = 1$. For $s \to 1$ we have a linear blending function while for $s \to \infty$ results in a stepping function with the transition at $t = 0.5$. Using the previously classification mask we simply set $s = 1$ for pixels with differences in luminance and $s = 10$ for pixels in occluding or changing regions. A smooth transition is achieved by using a low-pass filtered version of the steepness mask, thereby avoiding visual artifacts when blending edges of the steepness mask.

0.4.4 Plausible Feature Animation

The work of Beier and Neely [1992] proposes to use linear point interpolation on each point of the line feature for animation purposes. This however causes unrealistic deformations when the feature rotates between the images as can be seen

in Fig. 4. Instead, we propose to use an as rigid as possible transformation of the line features to interpolate the position of the features during animation. Following [Alexa et al. 2000], we find a rotation and stretching to map each source feature to its corresponding target feature . Then we interpolate these transformations in the respective domain to obtain a plausible feature animation over time, preserves scale and resulting in natural motion (cf. fig. 4).

0.4.5 Motion Layers

Image morphing is per se defined as an interpolation of the whole image lattice, based on smooth warping of the lattice. For some applications however, this is not flexible enough especially as boundaries between differently moving objects can not be represented accurately due to velocity discontinuity at these borders. Also, for view morphing, background and foreground interpolation usually cannot be faithfully represented by a single image morph.

To overcome this limitation, we propose to use a common concept in image and video editing, *warping layers* [Wang and Adelson 1993]. First, a segmentation of the foreground and background is obtained by background subtraction techniques or blue screen methods. Then, each foreground object is morphed separately using our perceptually driven method on each layer. For spatial morphing between different cameras the background is additionally morphed. The final result is obtained by combining the individual layers via alpha blending in a predefined order. Additional layers can be used if the foreground objects further exhibit motion discontinuities that cannot be addressed by one layer alone [Liu et al. 2005].

0.5 Implementation

The implementation of our method is divided into a pre-processing and a real-time rendering part. The preprocessing step for image morphing, once line correspondences have been established, is to compute the mask which is used during non-linear blending. This is done by computing the graph cut as described in section 0.4.2 using the implementation of Boykov and Kolmogorov [].

Rendering is implemented as a multi-pass rendering on graphics hardware. For each layer, the following steps are computed:

1: **for all** Layers **do**
2: Animation of features (Section 0.4.4)
3: Computation of feature weights ([Anonymous 2007])
4: Compute optical flow to refine the warping (Section 0.4.5)
5: Warping of source and target image
6: Adaptive non-linear blending (Section 0.4.3)
7: **end for**
8: Alpha blend all layers

To achieve real time performance, all steps are implemented as GLSL fragment shaders on GPU. The most complex operation is hereby the computation of the optical flow. We implemented the algorithm of Horn and Schunck [1981] as two fragment shaders, where the first computes the energy term for each pixel of the pre-warped images using the warping defined by the line features. The energy is then successively minimized during ping pong multi pass rendering with the second shader. For the presented results we used a fixed iteration length of 50 iterations to compute the optical flow which showed stable convergence during our experiments. Overall performance of the rendering part is dependent on the number of line features, the number of layers and the image size. In case of the Capoeira dance sequence with NTSC resolution, rendering the foreground layer with 9 line features takes 2 ms per frame on a GeForce 7900 GTX and an AMD Athlon64 X2 4800+ Dual Core Processor.

0.6 Results and Discussion

Our acquisition setup consists of eight off-the-shelf Canon 5D still cameras which feature 12 megapixel resolution and maximally 4 frames per second. The cameras are equipped with 28mm lenses to capture any scene at relatively wide angle. The shutter release on all cameras can be triggered collectively by wire which, however, does not perfectly synchronize shutter release times. For acquisition, the cameras are mounted on tripods which are set up roughly equally spaced around the scene. Neither intrinsic nor extrinsic calibration is performed. All sequences we present here are captured using this system. In the accompanying video, we present examples for space, time, and joint space-time interpolation. Since perceptionally plausible motion cannot be assessed from still images, we abstain from including additional figures here.

The first example in our video shows a capoeira dancer during training in the gym. Setting up the recording equipment on-site is uncomplicated and quick since we do not need to pay much attention to camera positioning or orientation. The video shows time and space interpolation results for the dancer while she performs a head- and then a handstand. Concering interpolation across time, the example represents about the fastest complex motion that can still be plausibily interpolated. To achieve convincing interpolation results for even faster, complex movements, recording at higher frame rates would be necessary. For simpler motion that does not exhibit so many contrary movements, even faster scenes can be managed. Because the cameras are space roughly 10 degrees apart, interpolation across space (frozen moment effect) is not critical. Note the faithfully interpolated hair of the dancer which would be problematic or often impossible using geometry based methods.

Another sequence in the video shows a second capoeira dancer performing a flik-flak. Again, we freeze time and show space interpolation by moving along the path defined by the camera centers. Although some asynchronism between

camera release times is noticeable for this very fast motion, it is not an issue for our framework since it is independent of the underlying motion, in contrast to other view interpolation methods. Note also that our method still performs well in the presence of strong motion blur, as is apparent at hand and foot of the dancer. In Fig. 5, we show one example for a visual effect (multi-exposure rendering) that can be directly created with our framework.

We also recorded a dancing couple performing a slow waltz. We interpolate across time while the dancers start in a forward motion followed by a turn. During the turn, the female dancer completely vanishes behind the male in one frame, resulting in missing motion information. The overall impression of the interpolated motion, however, is maintained even in this very complex object scenario. Note also that she is wearing a ball gown, showing complex, non-Lambertian reflection patterns which are plausibly interpolated.

Another example in our video depicts facial expression interpolation from still photos. As there is no motion encoded in the sequence per se, the images are used as key frames to plot the shown interpolated expression changes. Since our method does not rely on any model for the interpolation, we can animate more extreme expression changes in exchange for loss of overall flexibility as each "target" expression and view perspective in the desired output must be recorded. Note, that the plausible interpolation results in the case of closing/opening eyes and closing/opening mouth which can also be addressed simultaneously to a view change.

Besides space-time interpolation, our framework is kept general enough to also allow for convincing interpolation between different objects. Our video example shows an animated version of human evolution by interpolating between excavated skulls from different times and places (another kind of space-time interpolation). Although we did not explicitly address this scenario in our method, the perceived motion is nonetheless very plausible. Even transitions between skulls with missing fragments or parts and more complete counterparts are perceptually more plausible compared to simple linear blending since our method successfully suppresses disturbing fade-in fade-out motion.

The fundamental limitation of image morphing is, of course, that the viewpoint always lies between camera recording positions, i.e., in the (triangulated) surface spanned by the camera positions during acquisition. Other image-based rendering techniques are able to render the scene from any arbitrary viewpoint. However, image information is recorded only at camera positions; moving much closer towards the object results in blurred rendering results, while moving away from the object is almost equivalent to rendering the object at a smaller scale.

0.7 Conclusion

We have presented a framework for space-time image interpolation based on image morphing. Instead of enforcing physical correctness, our approach is geared towards synthesizing perceptually plausible transitions. We show that the warp field

can be automatically refined by making use of the optical flow, classify remaining image differences, propose non-linear blending to conceal (dis)occlusions, argue that as-rigid-as-possible feature interpolation yields superior results, and we advocate separate warping of motion layers. Our contributions enable smooth, convincing interpolation across space and time from arbitrary, uncalibrated still images.

With our framework, we can continuously navigate between camera positions and shutter release times without the need for time-consuming camera calibration, error-prone geometry reconstruction, or an expensive synchronized multi-video acquisition system. We hope that by having eliminated these acquisition constraints, image-based rendering will find viable new application fields, e.g., in special effects production and new visual media such as 3D-TV. Besides computer graphics, our framework may also make an impact on video coding research. Perception-based morphing allows for encoding substantially less (so-called I-)frames because many more in-between (B-)frames can be convincingly interpolated over much longer time spans. Advancing research into visual motion perception will help to improve the proposed framework in the future and to further refine image morphing for space-time interpolation.

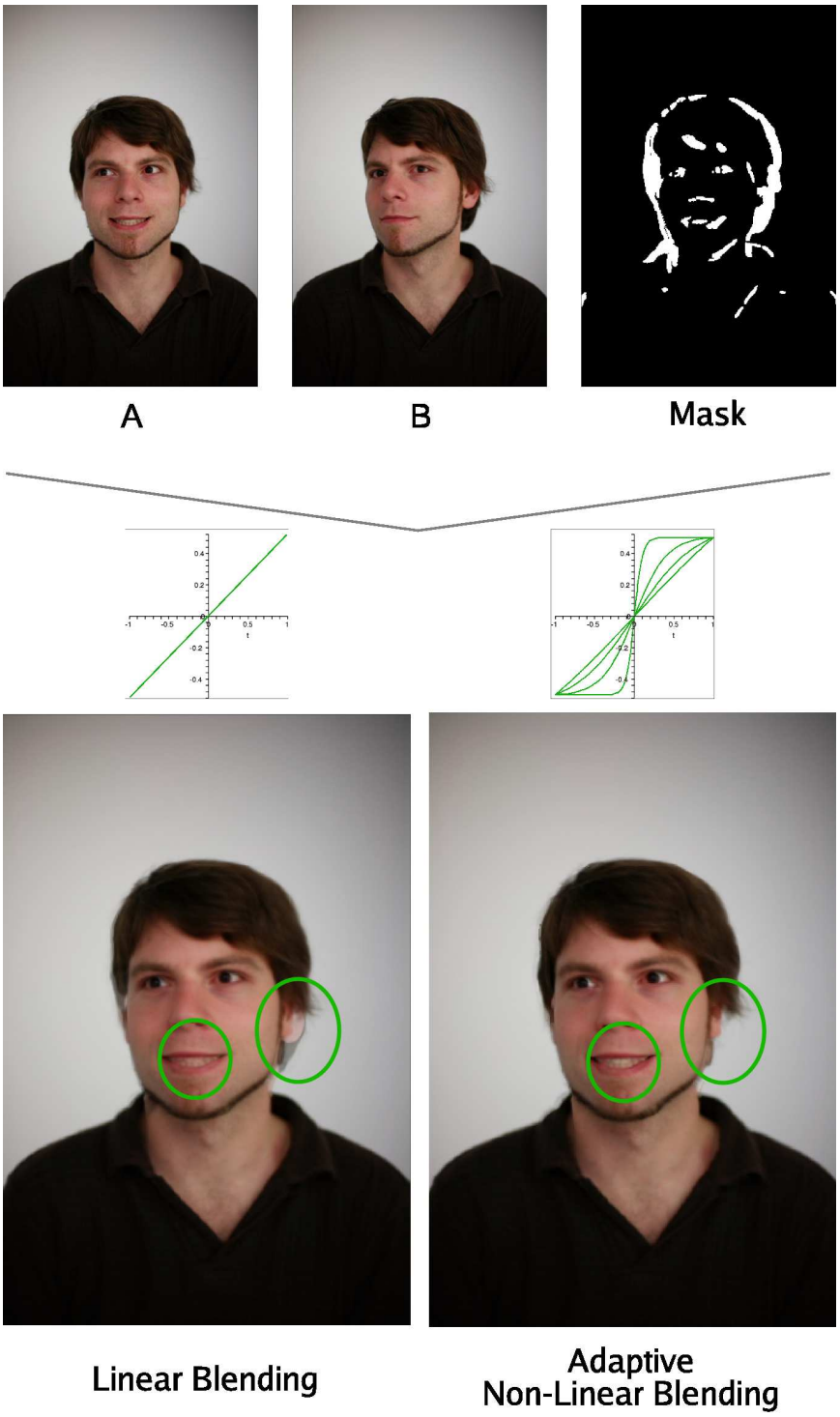

Figure 3: The first line shows the two input images and the classification mask computed using graph-cut optimization applied to the perceptual color distance. The bottom line compares between linear blending (bottom left) and our adaptive non-linear image blending method (bottom right) results. (Dis)occluding object regions are blended faster to avoid ghosting artifacts (green circles). In combination with image warping, the resulting image sequence is perceived as plausible motion.

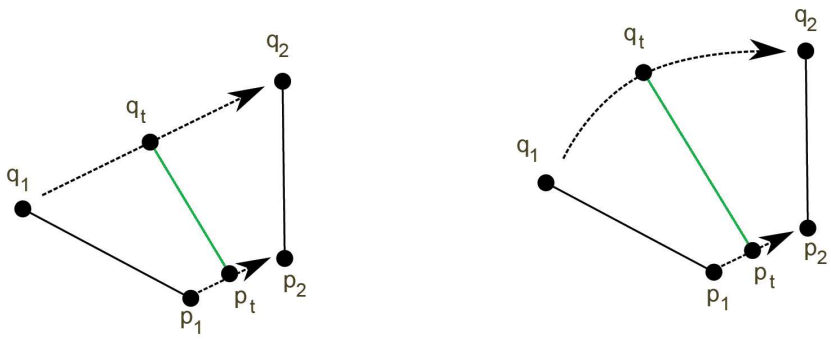

Figure 4: Linear interpolation between line feature points causes unrealistic scaling of the features during animation (left). As-rigid-as-possible deformation (right) preserve scale during animation and improves form plausibility of the interpolated images.

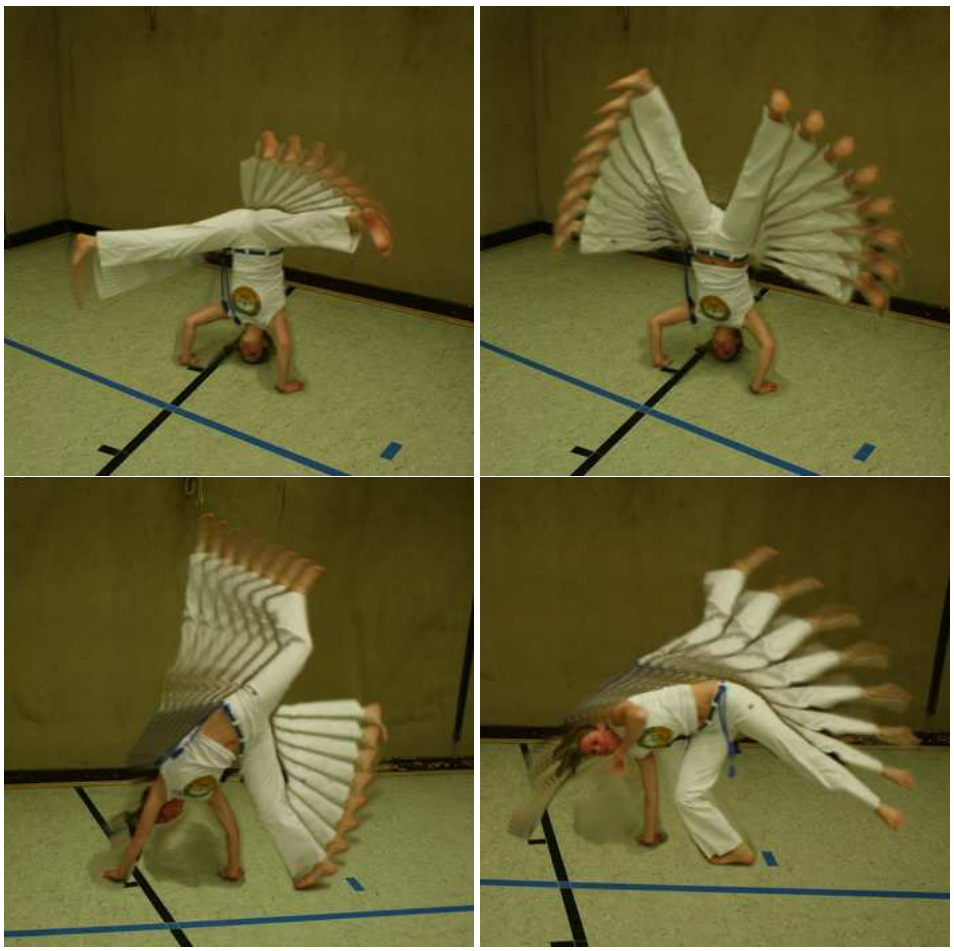

Figure 5: Image morphing for visual effects: multi-exposure images created from two consecutive capoeira photos. Several discrete in-between time instants are interpolated and overlaid.

References

[2000]ALEXA, M., COHEN-OR, D., AND LEVIN, D. 2000. As-rigid-as-possible shape interpolation. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH)*, New Orleans, 157–164.

[2007]ANONYMOUS. 2007. A probabilistic approach to automated image metamorphosis. Submitted to IEEE CVPR.

[2004]BAKER, S., AND MATTHEWS, I. 2004. Lucas-Kanade 20 Years On: A unifying framework. *International Journal of Computer Vision 56*, 3, 221–255.

[1994]BARRON, J., FLEET, D., AND BEAUCHEMIN, S. 1994. Performance of Optical Flow Techniques. *International Journal of Computer Vision 12*, 1, 43–77.

[1992]BEIER, T., AND NEELY, S. 1992. Feature-based image metamorphosis. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'92)*, Chicago, ACM, 35–42.

[] BOYKOV, Y., AND KOLMOGOROV, V. The MAXFLOW algorithm. http://www.cs.cornell.edu/People/vnk/software.html.

[2001]BUEHLER, C., BOSSE, M., MCMILLAN, L., GORTLER, S., AND COHEN, M. 2001. Unstructured lumigraph rendering. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'01)*, Los Angeles, ACM, 425–432.

[1983]BURT, P., AND ADELSON, E. H. 1983. A multiresolution spline with application to image mosaics. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'83)*, Detroit, ACM, 217–236.

[2003]CARRANZA, J., THEOBALT, C., MAGNOR, M., AND SEIDEL, H. P. 2003. Free-viewpoint video of human actors. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'03)*, San Diego, ACM, 569–577.

[1993]CHEN, S., AND WILLIAMS, L. 1993. View interpolation for image synthesis. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'93)*, Anaheim, ACM, 279–288.

[2000]CHENNEY, S., AND FORSYTH, D. A. 2000. Sampling plausible solutions to multi-body constraint problems. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'00)*, New Orleans, ACM, 219–228.

[1976]CIE, 1976. Colorimetry, publication no.15, supplement no. 2.

[1998]DEBEVEC, P., BORSHUKOV, G., AND YU, Y. 1998. Effi cient view-dependent image-based rendering with projective texture-mapping. In *Proc. Eurographics Rendering Workshop (EGRW'98)*, 105–116.

[1977]DEMPSTER, A. P., LAIRD, N. M., AND RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistics Society B 39*, 185–197.

[1938]ELLIS, W., Ed. 1938. *A Source Book of Gestalt Psychology*. Kegan Paul, Trench, Trubner & Co. Ltd.

[2002]EZZAT, T., GEIGER, G., AND POGGIO, T. 2002. Trainable videorealistic speech animation. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'02)*, San Antonio, ACM, 388–398.

[2000]GIESE, M., AND POGGIO, T. 2000. Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision 38*, 59–73.

[2003]GIESE, M., AND POGGIO, T. 2003. Neural mechanisms for the recognition of biological movements. *Nature Reviews – Neuroscience 4* (Mar.), 179–192.

[1996]GORTLER, S., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. 1996. The Lumigraph. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'96)*, New Orleans, ACM, 43–54.

[1965]GRAHAM, C. 1965. *Vision and Visual Perception*. New York: Wiley, ch. Perception of movement.

[2006]GRUNDLAND, M., VOHRA, R., WILLIAMS, G. P., AND DODGSON, N. A. 2006. Cross dissolve without cross fade: preserving contrast, color and salience in image compositing. In *Proc. Eurographics*, 577–586.

[1981]HORN, B., AND SCHUNCK, B. 1981. Determining Optical Flow. *Artificial Intelligence 17*, 185–203.

[2005]IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. 2005. Spatial keyframing for performance-driven animation. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 107–116.

[2007]INC., D. A., 2007. Digital air techniques. http://www.digitalair.com/techniques/index.html.

[2000]ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. 2000. Dynamically reparameterized light fi elds. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'00)*, New Orleans, ACM, 297–306.

[1973]JOHANSSON, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, 201–211.

[2003]KWATRA, V., SCHÖDL, A., ESSA, I., TURK, G., AND BOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'03)*, San Diego, ACM, 277–286.

[1995]LERIOS, A., GARFINKLE, C. D., AND LEVOY, M. 1995. Feature-based volume metamorphosis. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH)*, Los Angeles, 449–456.

[1996]LEVOY, M., AND HANRAHAN, P. 1996. Light fi eld rendering. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'96)*, New Orleans, ACM, 31–42.

[1994]LITWINOWICZ, P., AND WILLIAMS, L. 1994. Animating images with drawings. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH)*, Orlando, 409–412.

[2005]LIU, C., TORRALBA, A., FREEMAN, W. T., DURAND, F., AND ADELSON, E. H. 2005. Motion magnifi cation. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH)*, Los Angeles, 519–526.

[1995]LOGOTHETIS, N., PAULS, J., AND POGGIO, T. 1995. Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5, 552–563.

[2004]LOWE, D. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2, 91–110.

[1981]LUCAS, B., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, 674–679.

[2004]MATUSIK, W., AND PFISTER, H. 2004. 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'04)*, Los Angeles, ACM, 814–824.

[2000]MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S., AND MCMILLAN, L. 2000. Image-based visual hulls. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'00)*, New Orleans, ACM, 369–374.

[2004]MCDERMOTT, J., AND ADELSON, E. 2004. The geometry of the occluding contour and its effect on motion interpretation. *Journal of Vision 4*, 944–954.

[1995]MCMILLAN, L., AND BISHOP, G. 1995. Plenoptic modeling: An image-based rendering system. *Proc. ACM Conference on Computer Graphics (SIGGRAPH'95)*, Los Angeles (Aug.), 39–46.

[1986]MOVSHON, J., ADELSON, E., GIZZI, M., AND NEWSOME, W. 1986. The analysis of moving visual patterns. *Experimental Brain Research 11*, 117–152.

[2006]PAPENBERG, N., BRUHN, A., BROX, T., DIDAS, S., AND WEICKERT, J. 2006. Highly accurate optic fbw computation with theoretically justifi ed warping. *International Journal of Computer Vision 67*, 2, 141–158.

[1984]PORTER, T., AND DUFF, T. 1984. Compositing digital images. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'84)*, ACM, 253–259.

[1999]RIESENHUBER, M., AND POGGIO, T. 1999. Hierarchical models of object recognition in cortex. *Nature Neuroscience 2*, 11, 1019–1023.

[2002]RIESENHUBER, M., AND POGGIO, T. 2002. Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12, 162–168.

[2006]SCHAEFER, S., MCPHAIL, T., AND WARREN, J. 2006. Image deformation using moving least squares. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'06)*, Boston, ACM, 533–540.

[1995]SEITZ, S., AND DYER, R. 1995. Physically-valid view synthesis by image interpolation. In *Proc. Workshop on Representation of Visual Scenes*, 18–25.

[1996]SEITZ, S. M., AND DYER, C. R. 1996. View morphing. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'96)*, New Orleans, ACM, 21–30.

[2002]SHECHTMAN, E., CASPI, Y., AND IRANI, M. 2002. Increasing space-time resolution in video. In *Proc. European Conference on Computer Vision (ECCV'02)*, 753–768.

[2006]SNAVELY, N., SEITZ, S., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3d. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'06)*, Boston, ACM, 835–846.

[2005]VEDULA, S., BAKER, S., AND KANADE, T. 2005. Image based spatio-temporal modeling and view interpolation of dynamic events. *ACM Transactions on Graphics 24*, 2 (April), 240–261.

[1993]WANG, J., AND ADELSON, E. 1993. Layered representation for motion analysis. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'93)*, 361–366.

[2002]WEISS, Y., SIMONCELLI, E., AND ADELSON, E. 2002. Motion illusions as optimal percepts. *Nature Neuroscience 5*, 6, 598–604.

[1998]WOLBERG, G. 1998. Image morphing: A survey. *Visual Computer 14*, 360–372.

[2006]WOLF, M. 2006. Space, time, frame, cinema: Exploring the possibilities of spatiotemporal effects. *New Review of Film and Television Studies* (Dec.), 369–374. www.digitalair.com/techniques/STFC.pdf.

[2000]WOOD, D., AZUMA, D., ALDINGER, K., CURLESS, B., DUCHAMP, T., SALESIN, D., AND STUETZLE, W. 2000. Surface light fields for 3D photography. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'00)*, New Orleans, ACM, 287–296.

[2004]ZITNICK, C., KANG, S., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. In *Proc. ACM Conference on Computer Graphics (SIGGRAPH'04)*, Los Angeles, ACM, 600–608.