KEYFRAME ANIMATION FROM VIDEO

Timo Stich, Marcus Magnor

Computer Graphics Lab
TU Braunschweig, Germany
{stich,magnor}@cg.cs.tu-bs.de

ABSTRACT

This paper proposes a method for analyzing and synthesizing video sequences, specifically suited for image sequences of natural phenomena. We combine a low-dimensional representation of arbitrary image sequences and an image morphing technique to create realistic in-between images. The visualization based on the Isomap algorithm allows users to easily select parts of the video that have periodic character. From these segments, new sequences can be synthesized in real-time. To smooth transition artifacts between the reordered subsequences a Monge-Kantorovich based image morphing method is applied to interpolate in-between images. Our approach is useful for e.g. video key-frame animation, automatic looping or creating slow-motion sequences.

***Index Terms*—** Image sequence analysis, Animation

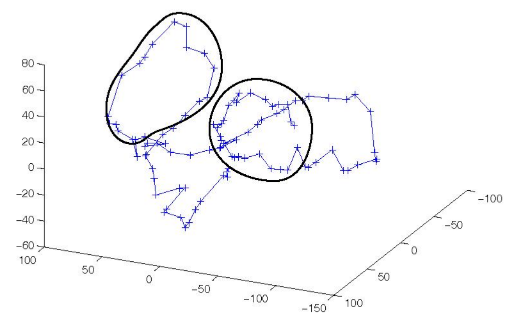

Fig. 1. Low-dimensional representation of a burning fire sequence. The Isomap algorithm represents all images as points in a k-dimensional space (here k=3) so that the distances between the images are preserved up to a minimal residual. Connecting the images according to their temporal order results in a video trajectory. The loops that can be seen in this view identify periodic subsequences.

1. INTRODUCTION

Despite recent advances in computer graphics to achieve photo-realisim from synthetic scenes, real-world footage is still superior. Image based approaches are also advantageous to use if synthetic scene models are time-consuming to create. Examples are dynamic natural phenomena like fire, smoke and water. However, the drawback of image sequences is the lack of manipulative freedom.

In this work we are combining two interesting and powerful recent approaches to tackle this problem. In a first step, we address the problem of selecting snippets having a periodic character by using a low-dimensional representation. This view of the data intuitively reveals these parts of the input sequences (see Figure 1). In a second step, these subsequences are then in turn used to synthesize new sequences with the desired properties. Users can easily create looping versions of the input sequence, script the appearance over time or adjust the temporal resolution.

Utilizing an image morphing method especially suitable for images of natural phenomena, we are able to create smooth but detail preserving transitions between the newly sequenced video blocks. This method is also suitable for creating in-between images for temporal super resolution without further user interaction.

2. RELATED WORK

The idea of reordering subsequences of videos to produce new output sequences was first proposed in [1]. An input video of a talking person was segmented into phoneme sequences which were then used to produce new sequences of the same person saying different things. A closely related work are *Video Textures* [2] introduced by Schoedl et al. Here, quasi-periodic video sequences are analyzed and transition points are found that are used to create looping versions of the input sequence. In addition image morphing techniques are used that rely on optical flow estimations to smoothen the transitions. However, especially for natural phenomena, the assumption of constant pixel intensity is often violated, resulting in loss of detail in in-between images.

Another approach for editing video sequences of natural phenomena is proposed in [3]. Here the user paints flow lines into the sequence, which are then used to learn textured particles from the input sequence. Although this is an interesting approach to create new image sequences of natural phenomena, the results variety is limited.

Using the Isomap low-level embedding to explore image sequences was addressed in [4]. However, the work does not include image morphing techniques to create new images but restricts to the nearest neighbor solution. Thus the applicability for video editing is limited.

Image morphing techniques are popular in computer graphics since the mid 80ies. There are many interesting applications like view interpolation of static scenes [5]. However most of the techniques depend on user interaction [6, 7]. One example of a morphing technique especially suitable for sequences of natural phenomena is presented in [8]. However, a drawback of the method is that the images have to be normalized before the warping can be computed. This sometimes results in unwanted image alterations.

A recent improvement in computing the optimal image warping in the Monge-Kantorovich sense is published in [9]. The authors suggest a very fast local approach to solve the problem. In our work, we are using an implementation of this method to create smooth transitions and new in-between images for temporal super resolution motion sequences.

3. VIDEO ANALYSIS

For analyzing image sequences, we are relying on a recent approach for visualizing the internal structure of the data[4]. The idea follows an approach where the images of the sequence are analyzed based on an image similarity measure independent of the temporal order. Combined with dimensionality-reduction methods like the Isomap algorithm [10], insightful new representations of the video sequences are obtained. A representation in 3 dimensions of a sequence depicting burning fire is shown in Figure 1. Although the space spanned by the points found with the Isomap algorithm has no interpretation per se, it serves as an embedding for the images so that the L2-distances between these points in this space, approximate the used similarity measure between the images. As can be seen in this figure, the periodic character of the input sequence is intuitively revealed by the loops of the trajectory. This allows users to easily identify quasi-periodic parts of the input sequence which we denote as *video blocks*. They are the building blocks for newly created sequences in the synthesis step.

3.1. Low-Dimensional Representation

In general, the Isomap algorithm can be applied to find a low-dimensional representation of an image sequence $I_t, t \in 1 \ldots N$ which preserves a metric distance $D(I_i, I_j)$ defined between the images. Thus the first step is to choose an image metric that faithfully represents the similarity and dissimilarities between the images of the sequence. In practice it suffices to use the simple L_2 distance metric between images which is also advantageous in the sense that it can be computed very fast .

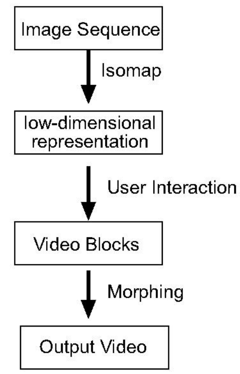

Fig. 2. Overview of our method. From an input video sequence, first a low dimensional representation is obtained based on the Isomap algorithm. Using this representation of the video the user can identify quasi-periodic snippets. These subsequences are then reordered either automatically or user-driven to create new sequences. Transitions between the blocks are smoothly interpolated using Monge-Kantorovich based image morphing.

The Isomap algorithm finds an embedding of the images in an arbitrary dimension $1 \leq k \leq N$. For each dimension there is an residual error which reduces with increasing k. Since we use the embedding to create a visualization of the image sequences, we are only interested in solutions for $k \leq 3$. Together with the temporal order of the images the sequence is represented by a trajectory in this space. For the natural phenomena sequences we analyzed, this representation intuitively revealed the quasi-periodic subsequences.

Other distance measures, like the Earth Movers distance [11], only marginally improve the resulting embedding since the Isomap algorithm depends only on the most similar images. For these, different measures have similar values [4]. For this work we used the Matlab implementation provided by Tenenbaum et al.[1] A detailed description of the Isomap algorithm can be found in [10].

3.2. Video Blocks (VB)

In the low-dimensional representation described in the previous section, the user can easily identify periodic parts of the input video. These parts form loops in the trajectory, which are not loops in the strong sense but points on the trajectory that are neighbors in this space (see Figure 1). Selecting the neighboring points defines a start and end image of a quasi-periodic subsequence of the video. We are denoting these subsequences as video blocks (VB) to express their role as building blocks in the synthesis step. A VB, B, is thus a tem-

[1] http://isomap.stanford.edu/

porally ordered set of images $B = \{I_i : s \leq i \leq e\}$ defined by a start and end image of the subsequence.

4. VIDEO SYNTHESIS

After the input sequence has been segmented into VBs, these blocks are used to synthesize new image sequences. However, by rearranging the blocks, temporal discontinuities at transitions might occur. We propose to smoothen these transitions by creating in-between images that preserve visual continuity. Unlike other low-dimensional embeddings obtained e.g. by PCA, in-between images are not unambiguously defined by their coordinates. So it is not clear how to obtain new in-between images to achieve smooth transitions. A straightforward solution is to use the nearest neighbor image for a given point in the embedded space. However, new connections between video blocks may have no in-between images in the original sequence. Our solution is to create new in-between images using image morphing techniques. Specifically, we use a Monge-Kantorovich based (aka Earth-Movers Distance) image morphing approach that is especially suitable to morph images depicting natural phenomena.

The synthesis step is divided into two parts. First we sequence the previously found VBs. In the second step, new in-between images are synthesized to create smooth transitions or even temporal super-resolution.

4.1. Sequencing

To create new image sequences the VBs from the video analysis step are rearranged. We propose several ways to create new sequences from the video blocks. An automatic approach is to specify the probability for each VB. This could be e.g. an uniform distribution if no preferences are given. A more controlled output is achieved by specifying transition probabilities $P(i, j)$ between blocks. This way a probability matrix P is defined which can be interpreted as a first-order Markov process. This process can be sampled to obtain new orderings. Another way is to sequence the blocks manually. Once the VBs are identified and labeled, the user can script the order of the blocks by defining an ordered set. The advantage is that the user has control while working with a high-level description of the image sequence.

The path connecting two blocks is defined by the connection between the two images that are the nearest neighbors in the VBs.

4.2. Image Morphing

Generally, an image morph u between two images I_i and I_j, is a mapping so that $I_j = I_i \circ u$. We restrict ourselves to automatic morphing techniques that need no user interaction to identify matching features between source and target image. One possible solution are optical flow based morphing methods, e.g. [13]. However, for image sequences depicting natural phenomena like fire, the assumption of constant pixel intensity over time is strongly violated leading to unsatisfactory results. Instead, we are proposing to use an image morphing method based on the Monge-Kantorovich problem. Originally stated by Monge [14] for solving the problem of moving piles of dirt in an optimal sense, recent work has translated this to the problem of image morphing [8, 9]. The image model assumption is, unlike the optical flow model, not pixel- but volume-based. Source and target image are assumed to have the same total pixel intensity $\int I_i = \int I_j$. The Monge-Kantorovich solution is then a transport mapping that redistributes the mass in a minimal distance sense. Small bright regions can be mapped to larger dimmer regions, thus better reflecting the physical properties of diffusion processes.

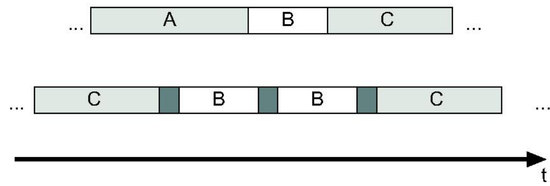

Fig. 3. Resequencing the VBs labeled A,B and C. To smooth the transition between the blocks new in-between images between the closest images are created using image morphing (dark grey blocks). The duration of the blocks can also be adjusted to create slow motion versions of the subsequences.

Fig. 4. Monge-Kantorovich image morphing between two images of a flame sequence. The left image is the source and the right the target image. In between is a newly created in-between image with morphing halfway from source to target.

Figure 4 shows the morphing results between two images of a fire sequence. To create smooth transitions, the forward and backward morphs between the source and the target images are blended. Let u_{ij} denote the warping from I_i to I_j, then we can write

$$I_{ij}(t) = (1-t)(I_i \circ (t\, u_{ij})) + t(I_j \circ ((1-t)u_{ji}))$$

where $I_{ij}(t)$ is the in-between image with $I_{ij}(0) = I_i$ and

$I_{ij}(1) = I_j$. The blending improves the result since in general $u_{ij} \neq u_{ji}$ because of the local approach to the solution. This gives smooth transitions while preserving details and avoiding double exposure effects.

5. RESULTS

We have applied our approach to several real-world video sequences depicting natural phenomena.

Fig. 5. Examples of our test sequences.

The example sequences consist of 70 to 200 images in the input sequence. We used a three dimensional representation of the sequence to identify the VBs. The results found on our project web page[2] show the application of random resequencing and temporal super resolution to create slow motion effects. Sequences showing a single phenomena, like the fire sequence, are best suited for our approach. Once the warpings between the images are pre-computed the creation of new sequences is possible in real time. Image sequences showing spatial parts that have independent temporal periodicity, like the fireside example, sometimes cause noticeable transitions.

6. DISCUSSION AND FUTURE WORK

With the proposed approach we can synthesize new image sequence from convenient video footage. The powerful low-dimensional representation generated by the Isomap algorithm enables users to easily identify subsequences having a periodic character that we call video blocks (VBs). These blocks are then arbitrarily concatenated to synthesize new image sequences to suit the users need. In the synthesis step we further applied Monge-Kantorovich image morphing which is especially suitable for morphing images depicting natural phenomena like fire or smoke. This creates smooth but detail preserving transitions between the VBs but can also be used to create temporal super resolution sequences. Example applications are automatic looping of the input sequences or high level scripted sequencing of the VBs.

For future work we will concentrate on improving the image morphing step. Although the method so far works well for similar images, distinctly different images are not always morphed satisfyingly. We are confident that a multi-frequency morphing technique will yield convincing results. Separating the images of the input sequence into patches instead of using the whole image in the analysis step would also improve the results. Looping the different parts separately should account for the problems sometimes encountered for image sequences that are divided into spatial parts having different temporal periodicity.

[2] `http://www.cg.cs.tu-bs.de/people/stich/video_editing/`

7. REFERENCES

[1] C. Bregler, M. Covell, and M. Slaney, "Video rewrite: driving visual speech with audio," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97)*, 1997, pp. 353–360.

[2] A. Schoedl, R. Szeliski, D. Salesin, and I. Essa, "Video textures," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '00))*, 2000, pp. 489–498.

[3] K. Bhat, S. Seitz, J. Hodgins, and P. Khosla, "Flow-based video synthesis and editing," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 360–363, 2004.

[4] R. Pless, "Image spaces and video trajectories: Using isomap to explore video sequences," in *Ninth IEEE International Conference on Computer Vision (ICCV'03)*, 2003, vol. 2, pp. 1433–1441.

[5] S. Seitz and C. Dyer, "View morphing," in *Proceedings of the 23th annual conference on Computer graphics and interactive techniques (SIGGRAPH '96))*, 1996, pp. 21–30.

[6] G. Wolberg, "Image warping: A survey," *Visual Computer*, vol. 14, pp. 360–372, 1998.

[7] T. Beier and S. Neely, "Feature-based image metamorphosis," *Computer Graphics*, vol. 26, no. 2, pp. 35–42, 1992.

[8] R. Chartrand, K. Vixie, B. Wohlberg, and E. Bollt, "A gradient descent solution to the Monge-Kantorovich problem," Preprint: LA-UR-04-6305, 2005.

[9] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal on Computer Vision*, vol. 60, no. 3, pp. 225–240, 2004.

[10] J. Tenenbaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[11] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Proceedings of the Sixth International Conference on Computer Vision (ICCV '98)*, 1998, pp. 59–66.

[12] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman & Hall, London, 1994.

[13] W. Hinterberger and O. Scherzer, "Models for image interpolation based on the optical flow," *Computing*, vol. 66, pp. 231–247, 2001.

[14] G. Monge, "Mémoire sur la théorie des déblais at des remblais," *Histoire de l'Académie Royale des Sciences de Paris*, pp. 666–704, 1781.