SPECIAL EFFECTS IN FILM/VIDEO MAKING: A NEW MEDIA INITIATIVE PROJECT

Chun-hao Wang[1], *Yongjin Wang*[1], *Meifeng Lian*[1], *Bruce Elder*[2], *Xiaoou Tang*[3], *Ling Guan*[1]

[1] Ryerson Multimedia Laboratory, Ryerson University, Toronto, Canada
[2] Graduate Programme in Communication and Culture, Ryerson University, Toronto, Canada
[3] Microsoft Research Asia

ABSTRACT

We present a system and a set of tools for producing special effects in film/video making by applying image processing and human centered computing techniques. A combination of shot detection, object segmentation, background generation, and image warping techniques are used. The user selects the image frame or the object of interest, and the image warp transformation to be used from the GUI. Wrapping can be performed either on a whole image sequence or on an object of interest in the sequence. In the latter, the object is first segmented and its motion tracked. Object segmentation is achieved either by snakes or graph cuts. Steerable Pyramid background generation is then used to fill in the portion cut from the foreground.

1. INTRODUCTION

Film and video makers have long relied on video processing techniques to create special effects for movies, television and advertising. The techniques have evolved over the years ranging from the use of blue screens to computer generated imagery. Supported by a New Media Initiative Grant jointly sponsored by Natural Science and Engineering Research Council of Canada and Canada Council for the Arts, a multidisciplinary research team from the Department of Electrical and Computer Engineering and School of Image Arts at Ryerson University, Canada, started the quest of raising the art of making special effects to a new and systematic level with the state-of-the-art innovations and advancement in image processing, computer vision and machine learning.

This work was initially motivated by John Cage's writings [1] on the aesthetics of music and the resolute challenge they posed to conventional ideas about art-making, which have inspired many composers. Cage believed that the creative process should imitate nature in its manner of operation and strived to find creative methods that would accord nature a role in shaping the work. The richness of Cage's writing helped make the use of aleatory techniques common among composers. However, the direct inspiration for this project was drawn from James Tenney's book *Meta+Hodos* [2], in which the author made extensive use of measures of similarity in the analysis of music structures in subsequent composers applied those methods to generating series of musical events. The artists on the Ryerson team conceived the possibility of developing analogous compositional procedures for working with video sequences, and soon began a collaboration with scientists and engineers, aimed at applying aleatory processes constrained by machine learning methods to generate special effects in film and video making.

This project has focused on three main objectives: 1) using measures of similarity to select scenes of interest and constrain random processes in the processing; 2) performing a warping effect appropriate to the selected video sequence or a particular object of interest in the sequence; and 3) modeling film/video makers working methods via machine learning and expert systems. In this paper, we present our work to fulfill the second objective: creating object warping special effects.

Experimental results will be presented to demonstrate the performance of the system. Combined with the image retrieval techniques we recently developed [3], which can be used to automatically select scenes of interest via similarity matching, we have started working on fulfilling the third objective - modeling film and video makers working methods via machine learning and expert systems, and completing the project - Special Effects in Film/Video Making.

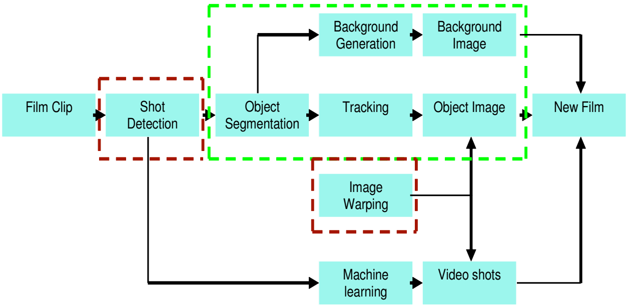


Fig. 1. The special effects system

2. OVERVIEW OF THE SYSTEM

The framework of our system is illustrated in Fig. 1. Shot detection is first applied to the film clip to obtain shots with different content or semantic meaning. Different warping algorithms are applied to produce different special effects for each shot. For a specific shot, a machine learning process is implemented to determine a set of transformations that might be appropriately applied, and the transformations are ranked according to their suitability.

The type of transformation we are interested in is image warping. Warping can be categorized into two kinds: for an entire scene or for a particular object of interest in the scene. In the former, the whole image is warped by a predefined transformation such as sinusoidal waveform or B-spline mapping. In the latter, the object is first segmented and its motion tracked. The image warping algorithms are applied only to objects that user has selected and that are isolated by our image segmentation methods. Based on the information from previous frames, the position of the object can be identified from the following frames by tracking the movement of the object and in each frame, image warping is only applied to the identified object. However, because the warping algorithms change the shape and position of the pixels of the object, the original background can not cover all the pixels and thus some regions are left blank. To fix this problem, a background generation technique is used to generate the missing pixels.

For object based warping, the first step requires human interaction to roughly localize the object. Either a snakes based or a graph cuts based segmentation technique is then used to obtain the boundary. Snakes require the user to outline the boundary of the object, and graph cuts requires the user to mark example pixels of the object and pixels of the background (labeling foreground and background).

3. TOOL DEVELOPMENT

3.1. Shot Detection

There are three types of transitions: sharp transitions, which film and video-makers generally call "cuts"; a fade-in or fade-out; and a dissolve. A cut is an abrupt transition between two shots. The frames on either side of the cut are markedly different. In a fade-out, the shot that is coming to an end, gradually turns darker, though sometimes it turns lighter by increments. An fade-in shot that follows generally begins in black and gradually turns lighter in increments until it reaches average brightness, though, if the previous shot had ended with a fade to white, the shot begins in whiteness and turns darker in increments, until it reaches average brightness. A dissolve is composed of overlapping fades, generally with the outgoing shot fading to black, and the incoming shot beginning in black and turning lighter.

To detect transitions between shots, the frame distance of DC and color coefficients of the DCT for each frame is calculated and thresholded. The energy histogram of a DCT coefficient is created by counting the number of times an energy level appears in the DCT coefficient blocks of a DCT encoded frame. Frame distance is calculated by using the following city block distance function

$$D(n) = \frac{1}{T}\sum_{t=1}^{M} |h_{n-1}(t) - h_n(t)| \qquad (1)$$

where $h_n(t)$ is the DCT coefficient histogram at frame n and block t.

The presence of a sharp transition causes a sharp peak in frame distance, and the presence of a dissolve transition is indicated by a small but non-zero peak. To enhance the detection of dissolve transitions, Twin Window Amplification Method (TWAM) [4] was applied in the uncompressed domain. It amplified dissolve transitions and slightly reduced sharp transitions, providing overally more accurate shot detection results.

3.2. Object Segmentation

Object based warping requires segmentation of the object as the first step. Two segmentation tools were developed. The first implementation of object segmentation required the user to roughly localize the object and the boundary using control points. A snake based technique is then used to converge to the actual boundary of the object. A snake is an active contour that iteratively moves toward the boundary of the actual object using energy minimization [5]. The active contour method produced sufficiently accurate results, as shown in Figs. 2(a) and (b). However, it is time consuming and involving intensive user interaction.

The second implementation of object segmentation uses the graph cuts method introduced by Boykov et al. [6, 7]. The graph cuts method is an energy minization technique to detect the best contour given a small sample of the subject and the background. Figs. 2(c) and (d) show the procedure of applying our segmentation technique and one example of the segmentation results. The user "seeds" or indicates the pixels in the object of interest (foreground) and some background pixels in the image. The number of pixels required is small, and can be fine-tuned later, making it an interactive process. The foreground and background pixels act as a hard constraint. Then the min-cut/max-flow algorithm is used to calculate the optimal cut to segment the image, satisfying the hard constraints. Two factors are taken into account in the process of segmentation: 1) The relative intensity between pairs of adjacent pixels and 2) The relative value of a pixel matching the histogram of the foreground and the histogram of the background pixels.

To implement this algorithm, we construct a graph using each pixel as a graph node, with known foreground and background pixels labeled. Each node is connected to its nearest neighbor with a weighted edge to represent the difference in intensity. Each node is also connected to a foreground and a background node with a weighted edge, with the weights corresponding to how the pixel matches the histogram of seeded foreground and background. The segmentation will minimize the cost of the segmentation according to the function

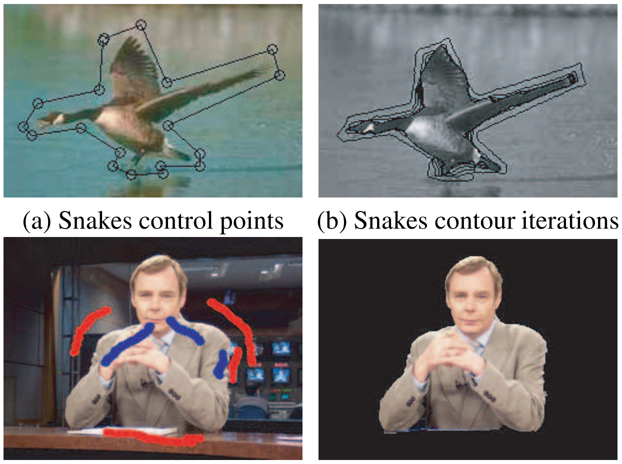
$$E = \mu \sum_{p \in \mathcal{P}} R_p(A_p) + \sum_{(p,q) \in \mathcal{N} : A_p \neq A_q} B_{p,q} \qquad (2)$$

$\mathcal{N}$ is the set of neighboring pixels. The total cost of the segmentation E is determined by the first summation representing the cost in assigning a pixel as foreground or background based on existing histograms, and the second summation representing the cost between adjacent pairs of pixels in the segmentation. $R_p(A_p) = -\log Pr(I|back)$ is the relative cost of assigning a pixel a foreground or background based on their histograms ($A_p = 0$ for background, $A_p = 1$ for foreground). $B_{p,q}$ is the edge cost between all adjacent pair of pixels in the given boundary.

The edge cost is calculated using a Gaussian function

$$B_{\{p,q\}} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)}. \qquad (3)$$

The difficulty lies in assigning an edge cost that correctly represents the likelihood of an actual boundary. Eq. 3 assigns a high cost value for pixels of similar intensities, and low costs for different pixels. Therefore, the algorithm is likely to segment the image across boundaries of high intensity value differences. A Gaussian function can model the noise of pixel values and thus eliminate camera and device induced noise.


(a) Snakes control points (b) Snakes contour iterations

(c) Graph cuts original image (d) Graph cuts result

Fig. 2. Object segmentation using snakes and graph cuts

3.3. Optical Flow Tracking

After segmenting the object from the first frame, the position of the object in the following frames can be determined by tracking the movement of the object. The system uses a combination of optical flow and snakes segmentation for tracking. The optical flow method is integrated with snakes segmentation to segment the objects from each frame. The optical flow method provides a rough estimation of the position of the object, while graph cuts approach uses a 3D image object segmentation to segment other frames. We use a pyramid method to give a course-to-fine estimation of the optical flow field. Fig. 3 plots the optical flow field that is generated from two continuous frames.

As for graph cuts, the optical flow tracking can be integrated as well, but it is easier to build a 3D node graph as the graph cuts algorithm can be extended to N-dimensional images.

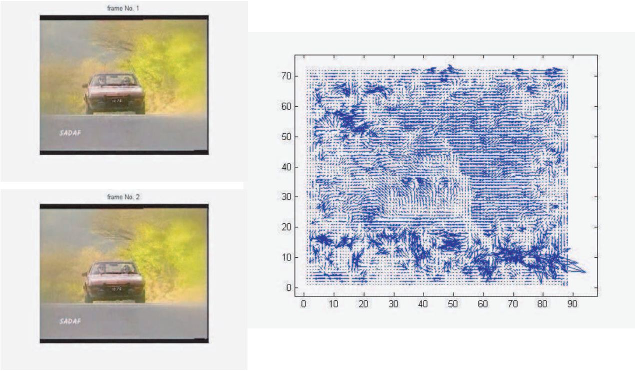


Fig. 3. Optical flow field

3.4. Image Warping

With the interested image frame identified or the object of interested extracted from the background, image warping is applied to create different special effects. In our experiments, special effects used include two-pass mesh warping, inversion circle, B-spline mapping to warp the objects. Fig. 4 shows a sinusoidal wave warped image using the program's user interface. We tested various image warping algorithms by Wolberg [8]. The artists can choose which image warping algorithm to apply and vary the parameters via the GUI. A batch process can be programmed to use gradual changes in the parameters frame by frame.

3.5. Background Generation using Steerable Pyramid

The segmented and tracked object in each frame are applied an image warping transformation to generate special effects. In this case, the position of the pixels is changed and the original background in each frame can not cover missing object pixels.

An intuitive method to generate background is by interpolation. The black areas can be filled in by interpolating the pixel value of its nearest neighbors. This method generally can not provide good performance if the background is complex. In this project, we utilize a steerable pyramid method to generate textures. The steerable pyramid is a linear multi-scale multi-orientation image decomposition method. For texture representation, it decomposes the image first, and then measures the statistics for each pair of coefficients at nearby positions, orientations, and scales. For texture synthesis, it starts with an image of Gaussian white noise, constructing the steerable pyramid and forcing the sample statistics of each subband to match those of a reference texture image.

The Steerable Pyramid method [9] works well for texture images, which are spatially homogeneous, and typically contain repeated structures. Its performance degrades when theses assumptions can not be met. However, it provides smoother effects than interpolation and thus can be applied in cases where the region of generation is small.

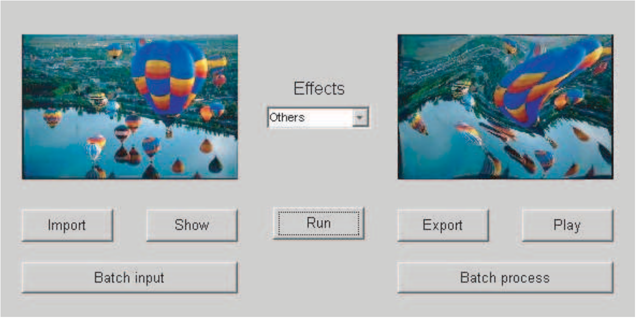
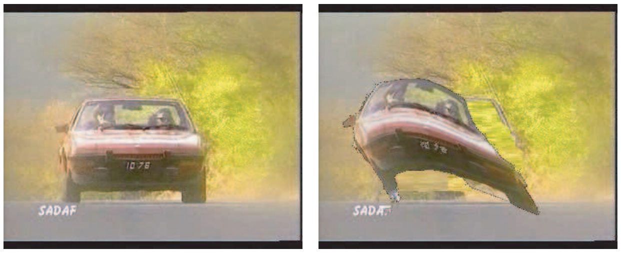


Fig. 4. Program's user interface



(a) Original frame

(b) After image warp

Fig. 5. Image warp example

3.6. Experiments and Results

The special effects system were implemented and tested on a 2.4Ghz, 1GB RAM Pentium 4 platform. Using snakes segmentation it takes 4 to 5 minutes to process a 40 frame (352 x 288 pixels) film clip. Using graph cuts segmentation on the same clip the total process takes about 8 seconds. A GUI allows the user to select the frames, mark objects of interest and the background for segmentation, and select the special effect to use on the object. Artists on the Ryerson team have been testing the software to create sample film clips with special effects for making their film - *The Young Prince*.

4. CONCLUSION AND FUTURE RESEARCH

In this paper, we presented a system and a set of tools for processing special effects in film using shot detection, object segmentation, image warping, and background generation. We have shown that it proves to be valuable in the ongoing research in experimental filmmaking.

The results of this project could be applied more widely than the field of experimental filmmaking. They could be adopted for use throughout the multimedia industry, especially for creating special effects for television and advertising.

There are two main directions for future research: 1) Incorporating machine learning to capture the working methods of film-makers. This will enable an expert system to provide analysis on the best special effects to use given the video shot. 2) Expand the capability of the video processing application by enhancing the object segmentation and background generation algorithms.

5. REFERENCES

[1] J. Cage, *Silence: Lectures and Writings by John Cage*, Wesleyan University Press, 1961.

[2] J. Tenney, *Meta-hodos and Meta Meta-Hodos: A Phenomenology of 20th Century Musical Materials and an Approach to the Study of Form*, Frog Peak Music, Santa Fe, N.M., 1998.

[3] K. Jarrah, I. Lee, M. Kyan, and L Guan, "Application of image visual characterization and soft feature," *to present at IS&T/SPIE Symposium on Electronic Imaging*.

[4] O. Bao and L. Guan, "Scene change detection using dct coefficients," *International Conference on Image Processing*, 2002.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.

[6] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *International Conference on Computer Vision*, vol. I, pp. 105–112, 2001.

[7] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *Proceedings of ACM SIGGRAPH*, pp. 303–308, 2004.

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

[9] E. Simoncelli and W. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," *IEEE Second International Conference on Image Processing*, 1995.