A Tool for Automatic Cinemagraphs

Mei-Chen Yeh and Po-Yi Li
Department of Computer Science and Information Engineering
National Taiwan Normal University, Taipei, Taiwan
myeh@csie.ntnu.edu.tw, lby0203@gmail.com

ABSTRACT
A cinemagraph is a new type of medium that infuses a static image with the dynamics of one or a few particular regions. It is in many ways intermediate between a photograph and a video, and provides a simple, yet expressive way to mix static and dynamic elements from a video clip. However, the process of creating cinemagraphs is usually tedious for end users and requires serious photo editing skills. In this demonstration we show a tool that creates cinemagraphs in a fully automatic manner. The technique should enable new features such as "intelligent cinemagraph mode" for digital cameras that provides an alternative method to capture the moment.

Categories and Subject Descriptors
I.4.9 [**Image Processing and Computer Vision**]: Applications.

General Terms
Algorithms, Experimentation

Keywords
Cinemagraph, Photography

1. INTRODUCTION
The ubiquity of digital cameras and smartphones, coupled with advances in computer vision and graphics, are bringing about a change in the way how a scene or event can be captured and manipulated. For example, a user can now easily take a panoramic picture by simply moving the camera as he/she shoots and the camera would automatically combine shots into a panorama image. Moreover, some cameras even feature an intelligent scene recognition mode where the camera can automatically detect various types of scenes and selects the appropriate camera settings.

Although new technologies have enabled new, creative ways to capture a scene or event, the form of digital photography is still limited to either a still image or a video. Figure 1 shows a waterfall picture. Please imagine that if water flows, this picture could become much more dynamic and interesting.

In this demonstration, we present a fully automatic approach for creating cinemagraphs. A cinemagraph is a new form of digital photography, recently exploded by a fashion photographer Jamie Beck and a visual designer Kevin Burg [5]. Unlike an animation in the Graphics Interchange Format (GIF) that simply loops or repeats a short video, cinemagraphs isolate a particular part of a video while leaving the rest static like a photograph. By freezing most of the moving elements and animating one or just a few, cinemagraphs are able to draw attention to a certain subject in a creative and effective manner. Cinemagraphs have been becoming popular, but the creation of cinemagraphs still requires professional image processing and editing skills.

Figure 1. A waterfall picture. The photograph may become more interesting if water flows.

A general guideline to create cinemagraphs using image processing software such as Photoshop and After Effects includes the following steps: (1) frame the input video—import the entire or a portion of a video and make every frame into its own separate layer; (2) catch the moving moment—select the frames that capture the desired movement, and choose one layer that shows the consistent, non-moving elements of the cinemagraph; (3) mask—edit the static layer by using a vector mask that filters out the moving elements; (4) synthesize—place a adjustment mask over all layers and generate the animation. For end users, the manual process of creating cinemagraphs is usually tedious. One of the major challenges is the generation of masks and layers, in which a user has to carefully select image parts and video frames to animate, and a still image to be used as the background. To automate the process, the following problems need to be addressed: (1) How to detect and analyze the dynamic characteristics of a scene or event? (2) How to localize the moving regions that give an interesting cinemagraph? (3) How to sample video frames to generate a seamless loop? These questions are nontrivial as input video clips, especially those captured with a smartphone, may have very complex motion patterns combined from ego-motion and scene motion.

Although a few tools have been recently launched to render cinemagraphs — such as Kinotopic [6], iCinegraph [7], and Cliplets [1]—, these tools are semi-automatic and the movement selection process is somewhat cumbersome. In this demonstration, we present a fully automatic tool for creating cinemagraphs from short video clips that requires no user intervention. In the following we describe in detail our system design, including video processing and motion analysis methods, and the demonstration setup.

MM'12, October 29–November 2, 2012, Nara, Japan.
ACM 978-1-4503-1089-5/12/10.

2. SYSTEM OVERVIEW

As depicted in Fig. 2 (a), the proposed system works as follows. Given a video clip, the task of automatic cinemagraph creation is to identify both where (the spatial location) and when (the temporal location) the interesting dynamic patterns are present. For example, the video shown in Fig. 2 (a) contains a spinning chair. It would be interesting to retain only the motions of the chair while freezing all the others. We first apply a video stabilization technique [8] to remove ego-motions. Next, we perform motion analysis and localize the interesting moving patterns based on the approach proposed in [3]. More specifically, we devise an approach that seeks the best selection strategy by maximizing the cumulative local motion magnitudes within a time period, with the conditions that the sampling points are determined to ensure a seamless loop. Finally, we edit the static layer by using a vector mask that filters out the moving elements and place an adjustment mask over frames to render the cinemagraph.

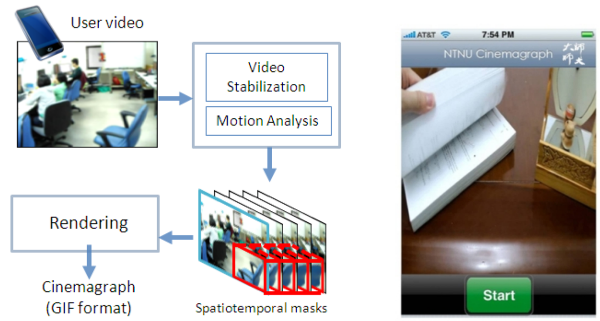
2.1 Motion Analysis

For motion analysis, we characterized the input video with a set of frame-to-frame optical flow fields. A pixel is described by the magnitude of the optical flow. We subtracted a constant from each optical flow magnitude, introducing positive and negative magnitude values. Thus, we obtained a representation where positive values imply relatively large motions. Now, individual pixel can vote for the spatial mask that may contain interesting moving patterns. We considered the score of a mask is the summation of the shifted optical flow magnitudes. The 3D array of positive and negative values is then passed to the next step for localizing a good spatial mask.

We adopt the branch-and-bound scheme to efficiently search for a sub-image that has the maximization of a quality function [2]. The choice was made because this method always examines the most promising solutions in terms of its quality bound, and does not impose restrictions on the values that the mask coordinates can take. In the implementation, we further speed up the search process by decomposing the 6D parameter space into a 4D spatial parameter space and a 2D temporal parameter space [4]. More specifically, we first employ the branch-and-bound strategy to search the spatial parameter space. Once the spatial window is determined, we then search for the optimal temporal segment. In general, the process takes a few seconds for a short (e.g. less than 10 seconds) video clip with a conventional PC.

2.2 Cinemagraph Rendering

Once the motion regions and the frame segment have been determined, the final step is to generate a seamless video loop. We compute the sum of squared difference (SSD) between pairs of moving regions—a small SSD is required to obtain a video loop that has unnoticeable temporal artifacts. In case the footage does not lend itself to loop well, we interpolate frames at the end of the looping cinemagraph to smoothly return to the starting frame. The final step is to composite the moving region onto the background layer, per frame in the identified interval, and save the file in the Graphics Interchange Format (GIF).



(a) (b)

Figure 2. (a) System framework; (b) User interface.

3. DEMONSTRATION SETUP

The demonstration has the form of stand-alone software with a graphical user interface shown in Fig. 2 (b). The system streams and displays the video on the screen when the application is activated. When the "Start" button is pressed, the system takes a short video and starts the cinemagraph construction process. The construction process is fully automatic. We will demonstrate the system's ability to generate creative cinemagraphs.

4. ACKNOWLEDGMENTS

This work was supported in part by the National Science Council, Taiwan, under Grants NSC 101-2221-E-003-023.

5. REFERENCES

[1] N. Joshi, S. Metha, S. Drucker, E. Stollnitz, H. Hoppe, M. Uyttendaele, and M. Cohen, "Cliplets: juxtaposing still and dynamic imagery," Technical Report, Microsoft Research, 2012.

[2] C. H. Lampert, M. B. Blaschko, and T. Hofmann, "Efficient subwindow search: a branch and bound framework for object localization," *TPAMI*, 31(12):2129-2342, 2009.

[3] M. Yeh and P. -Y. Li, "An approach to automatic creation of cinemagraphs," in *Proc. of ACM MM*, 2012.

[4] J. Yuan, Z. Liu, and Y. Wu, "Discriminative subvolume search for efficient action detection," in *Proc. of IEEE CVPR*, 2009.

[5] Cinemagraph. http://cinemagraphs.com/. [07/27/2012]

[6] Kinotopic. http://kinotopic.com/. [07/27/2012]

[7] iCinegraph. http://www.icinegraph.com/. [07/27/2012]

[8] VirtualDub. http://www.virtualdub.org/. [07/27/2012]