CS231A Course Project Proposal – Automatic Cinemagraphs

Bo Xian See
Stanford University
bsee@stanford.edu

Walter Fei Li
Stanford University
feili@stanford.edu

Abstract

We propose a method of automating the creation of cinemagraphs from an input video using image segmentation and tracking. The user will select a target object in a single frame of the video. The video is stabilized to normalize the coordinates of the chosen frame. Then the algorithm will track the object in subsequent frames. We expect the input video to have highly similar frames, since the video should be taken from a single scene. After extracting the tracked object, we propose to use Laplacian Pyramid to blend the tracked object onto the first frame. Finally, the scene is looped using video texturing and converted to a small GIF image format.

1. Introduction

Cinemagraphs are short, repeating animated images (usually GIF format) of a mainly static scene with slight motions. Currently, artists must manually create cinemagraphs. Previous related works to automate the process use video texturing to continuously loop through a video at selected regions [1]. However, their method does not track objects in motion, but instead consider regions containing dynamic objects. This loses the context of the scene as the regions are considered independent of each other.

In this paper, we aim to overcome this by using image segmentation and tracking. Instead of considering the image as independent regions, we consider movement based on the object. We track the movement of the object using multiple frames of the video to improve the accuracy of our segmentation and tracking.

1.1. Method

We aim to use segmentation based on simple clustering or inconsistency with background flow [2-4] to segment and track the target object across all the frames. This is possible because we expect mainly translational movement in the scene.

After segmenting and tracking the object, we obtain a sequence of frames that only contains the moving object. These frames are then blended onto the first frame using Laplacian Pyramids. Infinite looping can then be applied using video texturing [5].

1.2. Data

We intend to record video using a digital camcorder. Our sample size should be at least five videos of different scenes. Sample data will include approximately five-second clips of mostly static scenery. For each video, we manually select an object moving in the scene. Due to the expected limitations of image segmentation and tracking, scenes must be generally static with only a few dynamic objects.

1.3. Evaluation

The results of our process will be evaluated by the accuracy of object tracking through multiple frames and the accuracy of blending objects into the background scene over videos with 100 sample frames. Quantitatively, the accuracy of object tracking can be measured by checking if each frame has the correct segmentation and tracking. To alleviate this tedious process, a difference image between the current frame and the segmented object is used for human evaluation. A frame with significant error (i.e. area above a threshold) is considered a bad frame. The accuracy is simply the number of good frames over total number of frames compared.

To evaluate the efficiency of the blending, we take a random sample of 8 frames which are correctly segmented. Then, we include 2 frames with no blending, and place these images in random order. We then ask 5 to 10 users to determine if the frame has blending or not. The efficiency of the blending would be the number of frames the users correctly labeled as blended.

2. References

[1] Tompkin, James, Fabrizio Pece, Kartic Subr, and Jan Kautz. "Towards Moment Imagery: Automatic Cinemagraphs." *CVMP Conference* (2011): *Towards Moment Imagery: Automatic Cinemagraphs*. Web. 20 Oct. 2011.

[2] A. Meygret and M. Thonnat. Segmentation of optical flow and 3D data for the interpretation of mobile objects. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 238--245, 1990.

[3] W.B. Thompson. Combining motion and contrast for segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(6):543--549, 1980.

[4] W.B. Thompson and T.-C. Pong. Detecting moving objects. In *Proc. 1st Int. Conf. on Computer Vision*, pages 201--208, 1987.

[5] Arno Schodl, Richard Szeliski, David H. Salesin, and Irfan Essa. Video textures. *Proceedings of SIGGRAPH 2000*, pages 489-498, July 2000.

Future Distribution Permission