Popular Science Summary of Segmentation of Image Sequence into Scene-Coherent Parts

Johan Sjöberg

January 9, 2015

The Narrative Clip is a small wearable camera that takes an image every 30 seconds. By wearing the clip a whole day a user captures the main events of the day, a sort of personal photographic memory. As useful as this can be, wearing the clip a full day results in well over 1500 images and few people have the time to process that amount of images each day.

In this thesis the main objective was to construct an algorithm that automatically divided a full day of images into smaller segments. The goal is that each segment should describe a certain event or situation that is comprehensible by a human. For a typical example, consider your are driving to eat dinner with your parents. In this case you would want the algorithm to segment the driving part into one segment and the dinner into another.

The problem is not very easy, for example how do you decide what a moment is? Consider the previous example, but the transportation consist of walking to the car, driving the car and walking to the dinner. Should these three events be treated individually or combined? In fact, as shown in this thesis, the answer to a question like this depends on how you ask.

This problem is solved by using multiple humans to segment a sequence and then only choose the points where most of the humans agree should be a segmentation. This removes the subjective input from a single user since multiple users have to agree on a specific segmentation point.

The actual algorithm is made up of two parts, an indoor/outdoor segmentation and a mean change detection method.

The indoor/outdoor segmentation is based on the fact that when we move from indoors to outdoors (and vice versa) it is a high probability that we start something new and we should start a new segment. Consider, once again, the dinner example, we are outside before we arrive to the dinner which is most likely indoors. We simple have to find the point where we transition from outdoor to indoor to find the point of interest.

The mean change detection method comes from the idea that a certain situation should have similar colors, textures and other features. For example, consider sitting at the beach and then walk over to a grass field to play soccer. It should be a lot more blue colors and wave like textures at the beach compared to the soccer field. By finding when these changes take place we should be able to detect the relevant segments.

These algorithms were tested on a number of image sequences collected throughout the thesis, the results were not promising. 70 % of all the possible segmentation points weren't found and about 50 % of the points that were detected weren't actual segmentation points. It might be possible to improve the result by refining certain parts but it is still rather likely to not achieve a satisfactory results.