Real time video segmentation for MPEG-4.

Per-Erik Björk
Department of Computer Science / Centre for Distance-spanning Technology
Luleå University of Technology, 971 87 Luleå, Sweden
Per-Erik.Bjork@cdt.luth.se

Abstract

This paper describes and discusses different methods of video segmentation for use in MPEG-4. The methods discussed are motion detection, “blue screen”, laser and multiple cameras. The paper then specifies some requirements for the video segmenting method to make it as easy to produce MPEG-4 media as any other media format. This paper is a survey of some of the different methods for video segmentation and some conclusions on which ones are usable for real time MPEG-4 production.
1 Introduction

It becomes increasingly an issue how to generate MPEG-4 media automatically and easy, especially for video-conferencing and streaming media. Currently it is difficult to produce MPEG-4 media that uses all the features that the standard [1] offers. This paper concentrates on one of those features namely video segmentation. Video segmentation means that one can split the video into different segments and send them independent of each other, for instance a weather report, where we have a computer generated background and a presenter in front. There already exist methods that can be used for this kind of segmentation, but has so far mainly been used for computer vision and motion detection. We will then describe what the advantages and disadvantages are with different methods based on the requirements that we uphold. These requirements are constructed to make the production of MPEG-4 media as easy and effortless as possible. To produce MPEG-4 media should not be more difficult then any other media.

2 Requirements

For MPEG-4 to be accessible by the majority of the population, there are some requirements that must be met. It should not be more difficult to produce then any other media format. For this to be true we here state some requirements:

1. Productions should be able to be made “live” or in real time, which usually means 25-30 frames/sec.
2. Cheap. The hardware required should not be too expensive.
3. Should be able to use it for any kind of production, from weather reports to action movies.

These requirements are of coarse very demanding to meet, but must be fulfilled if everyone should be able to use this technology.

3 Methods of segmentation.

In this section we will discuss different methods of video segmentation and in which situation they are useful. The general idea of segmenting video is to provide a higher compression by splitting the video into different areas and encode them separately. By doing this we can optimize the encoding better for that object, it also makes it possible for people with low bandwidth to only decode the area relevant to them. We will discuss some different approaches to segmenting video, such as how objects move and how far away from the camera different objects are.

3.1 Motion detection

Motion detection [6] is based on finding objects that moves differently then the rest of the scene. Lets say we are trying to segment a video clip of a football match. When we watch this clip, we can see that the players and the football move around on the football field. This is what motion detection is about, to try to make a computer pick out the moving players and football and segment them from the football field.
This method is difficult to apply on the fly since it is based on a sequence of images, but it is possible with the penalty of some delay. However, it is not applicable on all situations, for example a talk show where the participants are sitting in front of a window and there is people moving around outside, it would then be close to impossible to separate the foreground from the background when there only is a limited bandwidth available.

3.2 “Blue screen”

The use of “blue screens” is common in the TV and movie industries and is based on the idea to consider everything of a certain color to be transparent. One example is the weather report where the presenter stands in front of a blue or green screen and every thing that is of that color is replaced with the weather map. A disadvantage with this approach is that the foreground objects or presenter cannot consist of any of this color since it will be considered transparent and masked out. With this method different video segments are created separately and then joined together before transmitting them. This method is of coarse not very applicable on the football scenario since one would have to produce the foreground and background separately.

3.3 Laser

This method is very good at generating a close to reality 3D approximation. It is based on the idea to let a laser dot wander over the scene and measure the distance to each point and from this create a 3D representation of the scene. This method is expensive and it is not applicable in all scenarios. An example is the football scenario, where it would be impossible to scan the entire football field for every video frame. This approach is mainly used when one wants a high precision approximation of a scene such as scanning a 3D object or navigating a robot.

3.4 Multiple cameras

When we position multiple cameras [3,4] focused on the same scene we can by area matching triangulate the distance to different points in the scene with fair accuracy. The concept is the same as the human eye, but can be enhanced by adding an extra “eye”. However, this approach is computationally intensive and hence difficult to do in real time. So far, this method has mainly been used to obtain a good 3D approximation of the scene and then mapping the texture on the obtained scene.

4 Conclusions/Discussion

The requirements where not met by any of the methods discussed, but with alterations and/or combinations to the different methods the requirements should be possible to fulfill. Motion detection is a good way of segmenting a video but it is not applicable on all scenarios as mentioned before. The usage of “blue screens” is not applicable on all kind of scenarios either and hence not very practical in general since it would require a single colored background. The laser method has a very high accuracy, which is not required to make simple video segmentation, but it might then be possible to lower the accuracy, cost and time to scan a scene to better suit the requirements.
With the multiple cameras approach we almost fulfill the requirements, but there are some other problems with this approach such as a robust area matching [2] that is very compute intensive. There are some other methods that are not mentioned here because they are not suitable for real-time video segmentation but could improve the result in combination with others. Such as out of focus blur [5] in combination with multiple cameras could prove beneficial. This would approximate the human eye, which uses at least these methods in addition with motion detection and “camera” motion [3].

5 Future work

The next step would be to test the different methods for video segmentation and see which ones or which combinations are most suitable for MPEG-4. When a good scheme has been found it would be interesting to try it in a conference application to create new ways of visualize participants. One example would be to only display the foreground objects (the heads of the participants) next to each other without a window frame around (arbitrary shaped), anywhere on the desktop, which of coarse would save lots of space on the desktop. And if the accuracy of the scheme is sufficient, it could also be used for MPEG-4 face animation or direct wireframe and texture mapping creation.

References


