Adaptive Region of Interest for Video

Collection Editor:
Dan Nelson

Adaptive Region of Interest for Video

Collection Editor:
Dan Nelson

Authors:
Erik Funkhouser
J Daniel Hays
Dan Nelson

Online:
< http://cnx.org/content/col11256/1.1/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: December 14, 2010
PDF generated: October 29, 2012
For copyright and attribution information for the modules contained in this collection, see p. 35.

Table of Contents

Chapter 1

Intro: Widescreen vs Fullscreen[1]

1.1 Widescreen vs Fullscreen

Many movies today are filmed with an aspect ratio of 16:9 (width:length) to allow for wider shots. Movies filmed in this format are generally referred to as **widescreen**. However, most older televisions have an aspect ratio of 4:3, and most television programming is presented in the 4:3 format.

This discontinuity in aspect ratio presents a problem to TV programming stations as well as the movie industry when trying to display movies on standard TVs. The two most popular solutions are:

1.1.1 1) Letterbox

Include the entire 16:9 frame inside the 4:3 frame, resulting in black bars at the top and bottom of the screen. This method preserves all of the information from the original frame, but results in a large portion of the 4:3 display being used to display no information.

1.1.2 2) Crop

Crop the sides of the 16:9 frame to generate a 4:3 frame that completely fills the standard TV frame. This method results in a **fullscreen** movie that completely fills the 4:3 display.

However, this method loses a substantial portion of the information from the original movie frame, potentially eliminating important scene elements.

The following modules will focus on developing an improved version of solution number 2 by computationally finding the region of interest in the scene and preserving this information when cropping.

[1] This content is available online at <http://cnx.org/content/m36482/1.1/>.

Chapter 2

Problems with Cropping[1]

2.1 Widescreen to Fullscreen Conversion Loses Information

When cropping a widescreen image down to a fullscreen aspect ratio, information will be lost from the edges of the scene. When performing this cropping, there are generally two approaches that are common in the movie and television industries:

2.1.1 1) Center Cut

Using this method, a 4:3 region is defined at the center of the widescreen scene, and the sides are cropped. This results in equal amounts of information being lost from the left and right edges of the scene.

This would not be a huge problem if everything of interest in the movie took place in the direct center of of the scene. However, as you may notice when watching your favorite flick, oftentimes the **Region of Interest (ROI)**is skewed towards the left or right of the scene. When this happens, the center cut method will result in important parts of the scene being cropped out. As seen below in a scene from the film Punch Drunk Love (directed by P.T. Anderson and starring Adam Sandler), a simple center cut approach results in the main character's being partially cropped out while the majority of the preserved scene contains nothing all that interesting.

[1]This content is available online at <http://cnx.org/content/m36557/1.1/>.

Figure 1: Original, widescreen format scene from Punch Drunk Love

Figure 2.1

Figure 2: Result of center cut cropping

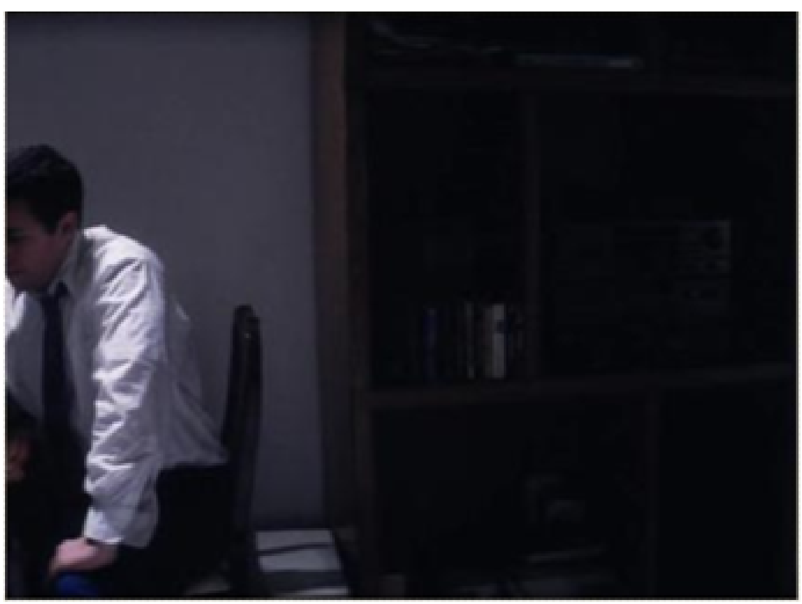

Figure 2.2: Note how part of the character's face and body has been cropped out

2.1.2 2) Pan-and-Scan

In the pan-and-scan method, an editor goes through the movie and moves around the ROI such that the important elements of each scene are not cropped out. This is a time-intensive and subjective process, and the results of it will vary depending on who determines the ROI. However, this is the preferred method because it ensures that embarrassing results such as in Figure 2 above do not ocurr.

2.2 Better Solution is Needed

What is needed is an automated, quantitative Pan-and-Scan system that can analyze a movie, determine where the important scene elements are, and adjust the ROI so that these important elements are not cropped out.

To develop such a system, we first need to know what the "important scene elements" are so that we can find a suitable method of quantifying them.

Chapter 3

What is Important in a Movie Scene?[1]

3.1 Motion, Edges, and Focus Determine ROI

Three major indicators of what is important in a movie scene are the amounts of motion, edges, and focus in different regions of the screen.

3.1.1 Motion

Motion is perhaps the best indicator of where the "action" is taking place in a scene. In an uncropped scene, objects that are in motion tend to draw the eyes of the audience. This makes motion detection a good scene element to quantify for our adaptive ROI system.

3.1.2 Edges

Sharp differences in an image tend to indicate the boundaries of separate objects in a scene. Generally, objects that are in focus tend to have more clearly defined edges while objects that are not in focus will have less defined edges. Sharp edges draw the eyes of the audience in a scene, as they are the delineators of separate objects. Thus, by detecting edges in a scene, we can begin to identify which objects stand out from their surroundings and should be included in the ROI.

3.1.3 Focus

What is in focus and what is out of focus in a scene could be said to be the most important way in which a movie director will tell the audience what is important and what they should be looking at. Generally speaking, areas of the scene that are in focus are the areas in which the important action in a scene will be taking place, so detecting the relative focus in different regions of the scene will be an important factor in determining the ROI.

3.2 Quantifying the Scene Elements

Now that we have decided on which scene elements we want to include in the ROI, we must develop systems for calculating the "amount" and location of each element in a scene.

[1]This content is available online at <http://cnx.org/content/m36549/1.1/>.

Chapter 4

Motion Detection via Frame Subtraction[1]

4.1 Change in Pixel Values Indicates Motion

A simple method of subtracting one movie frame from another will provide information about which parts of the scene have changed (generally due to motion). This method was performed on each frame of the movie, with consecutive frames being subtracted from each other.

4.1.1 Frame Subtraction

First, the scene is converted to an array of pixel values. These pixel values are the averaged **Red, Green, and Blue (RGB)** values for each pixel. The pixel values of the previous frame are then subtracted from the current frame's pixel values, and the absolute value of the values is taken. The result is an array of values that represent how much each pixel has changed between the two frames, with higher values representing more change. The amount of change in a region of pixels can be interpreted as the amount of motion that is taking place in that region. These data can then be used to determine where in the scene the most motion is taking place.

4.1.2 Illustrative Example

The images below show the results of subtracting two frames from Punch Drunk Love. Note that these are **NOT** consecutive frames, as the changes between consecutive frames can be very small. The frames presented below were chosen to clearly illustrate the results of frame subtraction. In the "difference" frame, higher values (more change) are represented by white.

[1]This content is available online at <http://cnx.org/content/m36555/1.1/>.

Figure 1: The current frame

Figure 4.1

Figure 1: The previous frame (not consecutive frames)

Figure 4.2