ROBUST IDENTIFICATION OF GRADUAL SHOT-TRANSITION TYPES

Ankush Mittal

Computer Science Department
National University of Singapore 117543

Loong-Fah Cheong, Leung Tung Sing

Electrical and Computer Engg.
National University of Singapore

ABSTRACT

Many video related applications require identification of shot transition and its type. Several algorithms for detecting both abrupt and gradual transitions exist. However, there is no integrated and robust framework for detecting transition types accurately. We show in this paper that combination of individual works by different researchers for identifying different transition types yields poor performance. An approach is presented using several algorithms working on patterns of effective average gradient, double chromatic difference etc. such that the approach is robust to false detection or false alarms. A comparison with the standard approaches over 3 hours of experimental data on non-trivial video classes such as commercials, MTV, sports etc. shows the superiority of our approach.

1. INTRODUCTION

Video segmentation into the individual shots is the first step for most of the video applications such as automatic video indexing, retrieval, editing and abstracting. Generally, both abrupt (i.e., flat cut) and gradual transitions (such as wipes, dissolves etc.) are used in narrative films and videos to convey the story structure. The works on identifying different media types by several researchers such as Mittal et al. [1], Eickeler [2] etc. rely on proper identification of shot-transition types.

Although there are many algorithms available for the shot transition detection, since different shot transitions have different characteristics, it is hard to use one single feature and a single algorithm to capture all the characteristics of shot transitions efficiently. The majority of previous works have focused on the detection of shot transitions but they fail to identify the transition types accurately. This motivated us to address the problem of shot transition type detection.

In the following sections, a robust framework of shot transition type detection is presented. Different features such as mean, variance and double chromatic differencing are used with different strategies to capture the characteristics of cut, fade, dissolve, and wipe transitions, respectively.

The paper is organized as follows. Section 2 presents an initial approach by combining a few of existing algorithms that identify different transition types. The poor performance of this approach is also presented in this section. Section 3 presents our approach and the integrated framework. Experiment results and comparison are presented in Section 4, followed by the conclusion in Section 5.

2. SHOT DETECTION BY COMBINATION OF ALGORITHMS: SIMPLER APPROACH

Recently, many researchers have worked on identifying shot transitions. The Pixel-wise comparison [3] approach detects if two frames are significantly different for identifying abrupt shot change. A modification to that is histogram comparison method. Zhang et al. [4] used two thresholds to detect gradual transitions such as wipes and dissolves. Fernando et al. [5] use statistical features and structural properties of the images are used to identify wipe transition region by applying Hough transform.

Developing an algorithm to identify four different types of shot transition (cut, fade-in/out, wipe and dissolve) can be straightforward by integrating any four different algorithms that are described above to deal with each transition independently. One of the combinations is to use histogram comparison for cut [3], variance curve method [6] for dissolve and fade-in/out and statistical approach [7] for wipe transition. Only three features are required from each frame: histogram, mean and variance of each pixel. The approach with combination is described here.

2.1. Results of the combination of the simplest algorithms

The experimental results show that this combination generates many false detections with dissolve and wipe transitions. The overall precision rate was as low as 18.3%. The main problem is caused by the dissolve and the wipe detection. If we examine the variance curve carefully, two major causes of the false detection can be identified easily.

Problem 1: When the variance magnitude of the frame is low, the valley of parabola that is caused by the dissolve becomes less distinct. The dissolves in certain portions of the sequence generate a larger parabolic
shape than those in other portions. It becomes very difficult to detect the valley based on a set of global thresholds.

Problem 2: Large object motion and camera motion (e.g., pan or zoom) will introduce similar parabola shape. These types of motion are very common during commercials and climax.

Fig. 1. (a) The effect of different transition on the variance sequence. (b), (c) & (d) are the corresponding DCD sequence of the first, second and third dissolve region in (a).

3. OUR APPROACH

The approach with combination of algorithms did not give acceptable results. We present our approach in this section. The dissolve detector based on variance magnitude is improved by an approach based on Double chromatic difference. Another problem that we need to overcome is the low precision rate introduced by the wipe detector.

One way to improve the procedure is to confirm each potential wipe region from the statistical wipe detector with some other algorithms. After trying different methods, the Plateau algorithm [8] was found to be able to work best with the statistical wipe detector. Although it is designed to detect gradual transitions, it provides high accuracy in verifying the potential wipe regions as well.

3.1. Double chromatic difference (DCD)

Consider an ideal dissolve, which means that there is no local motion or camera motion in the sequence and the change of intensity is linear; then there exist a frame at \( t_k \) with its intensity \( g(x, y, t_k) \) equal to the average intensity of the starting and ending frames of the dissolve. That is,

\[
g(x, y, t_k) = \frac{g(x, y, t_0) + g(x, y, t_N)}{2} \quad (1)
\]

where the dissolve sequence start at \( t_0 \) and end at \( t_N \). Based on Eq.(1), the DCD sequence [9] is defined as,

\[
DCD(t) = \sum_{x,y} F \left( \frac{g(x, y, t_0) + g(x, y, t_N)}{2} - g(x, y, t) \right)
\]

such that \( t_0 \leq t \leq t_N \), where \( t_0 \) and \( t_N \) are any 2 points within the dissolve period \([0, T]\) such that \( 0 \leq t_0 < t_N \leq T \). \( g(x, y, t) \) is a frame within the possible dissolve sequence. \( F(\cdot) \) is a thresholding function. For convenience, we set the threshold to zero.

As shown in Figure 1a, the second dissolve causes a small valley in a variance sequence, but the DCD sequence of the second dissolve region shown in Figure 1c form a valley as large as that of the first and the third dissolve region in Figure 1b & 1d. Furthermore, it can be showed that DCD(t) shows approximate parabolic shape only during a dissolve, and some other shape during other editing effects as well as other camera motion such as pan, zoom-in as shown in Figure 2c & d. From Figure 2a and b, no valley is formed in both fade-in/out and wipe transitions. However, the main limitation in using DCD is that the approximate boundaries of the dissolves must be known before DCD can be computed. Fortunately, the values of \( t_0 \) and \( t_N \) can be approximated by locating the starting and ending point of the parabolic variance sequences.

Although using variance sequence is not very accurate in locating the dissolve boundaries, it can be shown that DCD values will still form a parabolic shape by substituting \( t_0 \) and \( t_N \) with the staring and ending point of each downward parabolic regions in the variance sequence. From Eq.(2),

\[
DCD(t) = \sum_{x,y} \left\{ \frac{\alpha(t_0) + \alpha(t_N) - 2\alpha(t)}{2} \right\} \times |p - q|
\]

(3)

From Eq.(3), we can see that for any \( t_0 \), \( t_N \) satisfying \( 0 \leq t_0 \leq t_N \leq T \), DCD(t), will always show approximate parabolic shape. That means that the positions of the starting point and the ending point of a dissolve are not essential in DCD calculation and they can be approximated by the fast variance curve algorithm.

3.2. The new dissolve detector

The variance sequence for downward parabolic curves is examined. Confirmation of each of potential dissolve detected is done by substituting the starting point \( t_0 \) and the ending point \( t_N \) of each of the downward parabolic variance sequence into Equation (2) to compute the DCD(t). A dissolve is declared if the DCD(t) shows a parabolic shape as well.
3.3. Integrating Plateau algorithm

From the experimental results, this plateau algorithm tends to detect the beginning portion of a wipe as a gradual transition while the statistical wipe detector detects the latter portion. The combined wipe detection consists of two passes in which the statistical wipe detector is first applied. For each potential wipe region, reduce the starting point by 10 frames, if any of this extended region overlap with any gradual transition region detected by the Plateau algorithm, a wipe is declared the union of both the region (see Figure 3).

3.4. Abrupt scene change

The plateau algorithm can be used to detect an abrupt scene change in a sequence of DC images by setting $k = 1$ [8]. Therefore the histogram comparison algorithm with this Plateau approach was used so that the abrupt scene change detection can be performed with a DC image sequence without fully decompressing the video.

4. EXPERIMENTAL EVALUATION

The experiments verify that the unified algorithm can detect and identify different types of gradual transitions. Since the abrupt scene change is easy to detect with high accuracy, therefore the identification results for gradual transitions are separated away from the overall results.

4.1. Set up

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Duration (m/s°)</th>
<th>D</th>
<th>W</th>
<th>F1</th>
<th>FO</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTV</td>
<td>13'8&quot;</td>
<td>44</td>
<td>0</td>
<td>38</td>
<td>42</td>
</tr>
<tr>
<td>Commercials</td>
<td>1'40&quot;</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Sports</td>
<td>13'38&quot;</td>
<td>45</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Action Movie</td>
<td>3'49&quot;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>News</td>
<td>3'43&quot;</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>36'00&quot;</td>
<td>99</td>
<td>16</td>
<td>47</td>
<td>49</td>
</tr>
</tbody>
</table>

Table 1. Video database used for experiments

Table 1 gives the description of the test database. The action movie sequence is made up of the climax portions of a few action movies.

Fig. 4. Different wipe patterns are included in the testing sequence to evaluate the performance of the integrated algorithm

4.2. Performance parameters

For the purpose of comparison, twin comparison algorithm [4] was applied to the same testing video sequence. Since the combined wipe detector make use of the Plateau algorithm, we would compare its performance as well.

To test the integrated algorithm, the following criteria is used:

Criteria #1 A gradual transition is considered as correctly detected if at least one of its frames has been detected and identified correctly.

Criteria #2 If the gradual transition is detected but misclassified, this is counted as both miss and false detection.

Criteria #3 If a long dissolve transition is detected and identified as the multiple dissolve transitions, then only one will be counted as correct detection and the rest as false detections even through the identifications are correct.
Criteria #4 When two dissolves are close to each other and only a single dissolve is detected and identified, only the first dissolve is counted as correct detection while the second one as miss detection.

Since both of the twin comparison and the plateau algorithm cannot identify the individual gradual transition, therefore criteria #2 is not applicable, and these algorithms are evaluated based on criteria #1.3 & 4 without the identification restriction. That means, as long as they can detect the gradual transition, a correct detection will be counted.

Table 2. Recall rate and Precision rate for various algorithms

<table>
<thead>
<tr>
<th></th>
<th>Integrated</th>
<th>Twin</th>
<th>Plateau</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C_i</td>
<td>F_i</td>
<td>M_i</td>
</tr>
<tr>
<td>sport</td>
<td>47</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>comm.</td>
<td>13</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>mtv</td>
<td>96</td>
<td>24</td>
<td>28</td>
</tr>
<tr>
<td>news</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>movie</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>157</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>REC</td>
<td>75.4%</td>
<td>56.9%</td>
<td>46.9%</td>
</tr>
<tr>
<td>PRE</td>
<td>74.7%</td>
<td>64.5%</td>
<td>38.8%</td>
</tr>
</tbody>
</table>

Table 2 lists the correct, miss and false detections and identification of the integrated algorithms under columns C_i, M_i, F_i; subscript i used to indicate the accuracy in identification is also taken into account. From Table 2, it can be clearly seen that the integrated algorithm provides the best performance in both detection and identification. The twin comparison was the second best in detection. The sport sequences which consist of only wipe transitions were used to test the improvement of the new wipe detector. As shown in the row "Sport" in Table 2, the new approach removed 46 false detections and recovered 12 miss detections from the original plateau algorithm.

In addition, the integrated algorithm can handle the short transitions very well in a commercial sequence. It does not generate many false detections with the flashlight in the News sequence. With the action motive sequences, the integrated algorithm is less sensitive to the object and camera motion than the other two. This improvement is mainly contributed by the DCD verification process.

5. CONCLUSIONS

The integrated algorithm not only can detect but also identify cuts, fades, dissolves and wipes automatically based on only three easily extractable features from each frame: mean, variance and DCD. The entire algorithm works with DC images, so that the shot detection and identification process can be carried out without fully decompressing the video stream. Experimental results show that the algorithm can be considered as a reliable way to detect and identify different kind of transitions automatically with acceptable accuracy.

Acknowledgments

Heartfelt thanks are due to Professor Krishnan, Indian Institute of Technology, for giving the inspiration for this work. The help received from CIT Multimedia Lab. at NUS is greatly appreciated.

6. REFERENCES


