Shot Change Detection Using Global and Local Information with Adaptive Threshold

Tsung-Han Tsai, Chih-Lun Fang and Yung-Chien Chen
Electrical Engineering Department, National Central University, Jhongli City, Taoyuan County 32001
Taiwan, R. O. C., E-mail: {han, allen, tracey}@dsp.ee.ncu.edu.tw

Abstract: In this short paper, a novel approach that combines global and local information to detect abrupt and gradual shot changes is proposed. In this approach, abrupt shot change is detected by an adaptive threshold computed from DC frame histogram difference with a sliding window. To detect gradual shot changes, this work performs an analysis of block-wise histogram difference with a statistical model. The proposed method increases the application of histogram information to abrupt and gradual shot change detection. The experiments and comparisons on various video sequences demonstrate good detection results.

Key words: Shot Change Detection, Adaptive Threshold, Histogram, Statistical model.

1. INTRODUCTION

In recent years, content-based video retrieval has been developed to obtain the desired videos from a large video database. For accurate video retrieval, videos have to be segmented into shots which consist of interrelated consecutive frames. The interrelated consecutive frames are taken contiguously by a single camera and denote a continuous action in the time and space domain. However, the similar content in consecutive frames makes it difficult to segment videos into shots. Hence, a shot change detection approach [1]-[6] is required to perform video segmentation. In general, two kinds of shot change exist in videos: abrupt and gradual. Gradual shot change consists of fade in, fade out, and dissolves. Gradual shot changes are more difficult to detect than abrupt ones due to the special effect by the film-makers. Additionally, camera operations and object movements often cause false alarm in gradual shot change detection. Therefore, the aim of this study is to provide a solution which tends to enhance the performance of abrupt shot change detection and gradual shot change detection.

To detect the shot changes, the similarity between two frames must be well defined. The similarity is derived from a feature. This feature is computed from pixels [1] and histograms [2], [3].

The pixel-based methods are highly sensitive to object motions meanwhile the histogram-based methods are sensitive to abrupt boundary and invariant to image rotation. The gradual shot change detection is realized by color coherence change in [4]. Additionally, gradual shot change detection is treated as a linear transition form in [5]. The detection benefits in [5] are not applicable for local variations. In [6], it reports that histogram differences methods in YUV color space achieve the best overall performance among different color spaces. Some previous histogram-based methods provide a desirable performance for abrupt shot change detection [3], but few are suitable for gradual shot changes detection. To overcome this difficulty, a robust histogram-based detection method (HDM) is proposed not only for abrupt but also for gradual shot change detection. Abrupt shot change detection is achieved by DC histogram difference and an adaptive threshold. For gradual shot change detection, Wavelet smoothing is firstly manipulated to eliminate the camera motion. This work then performs a block-wise histogram method which divides a frame into a fixed number of blocks and records its spatial variations. With the recorded spatial variations, a statistical model is manipulated to detect gradual shot changes. The propose algorithm not only applies histogram difference to abrupt and gradual shot change detection but also avoids false alarm caused by camera operations and object movements.
This short paper is organized as follows: Section 2 illustrates the extraction of similarity. Section 3 proposes the shot change detection algorithm. Section 4 presents the experimental results and discussion. Section 5 draws our conclusions.

2. SIMILARITY EXTRACTION

The overview of the HDM is illustrated in Fig. 1. This work utilizes a back processing order to cover the past frames in a time interval. The DC frame histogram difference and block-wise histogram difference are derived from each of $L$ frame pairs between frame $t$ and $t-1$, up to frame $t-L$ where $L$ is the sliding window size. The classification of a frame into normal shot frame or shot change frame is accomplished by an adaptive threshold ($T_a$) derived from a specific feature. Details of the HDM are proposed in the following sections.

Moreover, the effectiveness of detecting the shot changes depends on the suitable choice of the similarity between two frames. Since histogram differences are insensitive to object motion and invariant to image rotation, this work adopts the whole frame and block-wise histogram difference of the DC image as the global and local information respectively. Referring to the similarity, we exploit several types of measurement algorithms. Since $X^2$ test [2] generally has better performance with respect to other measurements, the $X^2$ test [2] is manipulated to obtain the similarity. Thus, the similarity is computed from the sum of square differences of corresponding bins as follows:

$$D_c(X,Y) = \sum_{j=1}^{K} \left( \frac{(H^X(j) - H^Y(j))^2}{\max(H^X(j),H^Y(j))} \right)$$

if $(H^X(j)  \neq 0 \cup H^Y(j)  \neq 0)$, otherwise

$$D_b(X,Y) = \sum_{j=1}^{P} \left( \frac{(H^X(j) - H^Y(j))^2}{\max(H^X(j),H^Y(j))} \right)$$

if $(H^X(j)  \neq 0 \cup H^Y(j)  \neq 0)$, otherwise

where $D_c(X,Y)$ and $D_b(X,Y)$ denote the similarity, $H^X(j)$ is the $j$-th bin of the histogram of DC image $X$, and $H^Y(j)$ is the $j$-th bin of the histogram of DC image $Y$. $K$ is the total number of bins of the histogram of DC image $X$. $H^{X,Y}(j)$ is the $j$-th bin of the histogram of the block-wise of DC image $X$. $H^{X,Y}(j)$ is the $j$-th bin of the histogram of the block-wise of DC image $Y$. $P$ is the total number of bins of the histogram of the block-wise of DC image $X$.

3. SHOT CHANGE DETECTION

A decision function is required to perform the shot change detection after generating the similarity. However, with a consideration that shot change is a local activity in the temporal domain, it is more appropriate to set the threshold to match the local activity. The difficulty with the use of threshold is that it should be applicable to detect most types of shot changes. Without manual parameter specification, an adaptive threshold is evaluated as follows:

$$T_a(t) = \frac{1}{L} \sum_{k=1}^{L} D_{c,b}(t, t-k)$$
where $D'_t(X, t-k)$ is the DC frame histogram difference ($D'_t(X, Y)$ in (2)) and block-wise histogram difference ($D'_c(X, Y)$ in (3)). This threshold may suffer from noise due to lightness vibration or self-in noise. Thus $D'_t(t, t-k)$ in the sliding window are applied to calculate the adaptive threshold except the maximum and minimum $D'_t(t, t-k)$. Furthermore, an abrupt shot change exists if one of the following conditions is satisfied:

1. The maximum of DC frame and block-wise histogram difference ($D'_t(X, Y)$) occurs at $(t, t-k)$, for $0 \leq k \leq L$.
2. The DC frame and block-wise histogram difference of $(t, t-k)$ is larger than the adaptive threshold ($T_{a[t]}$ in (4)), and a significant increase which relates to the previous DC frame histogram difference and block-wise histogram difference exists.

### Table 1
Recall and Precision of Shot Change Detection

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Number of frames</th>
<th># of abrupt shot</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F1</th>
<th># of gradual shot</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movies</td>
<td>20562</td>
<td>226</td>
<td>99.12</td>
<td>95.58</td>
<td>97.32</td>
<td>7</td>
<td>85.71</td>
<td>57.14</td>
<td>68.57</td>
</tr>
<tr>
<td>Music video</td>
<td>8712</td>
<td>71</td>
<td>97.18</td>
<td>97.18</td>
<td>97.18</td>
<td>47</td>
<td>89.36</td>
<td>76.60</td>
<td>82.49</td>
</tr>
<tr>
<td>News</td>
<td>6231</td>
<td>43</td>
<td>97.67</td>
<td>98.82</td>
<td>93.03</td>
<td>46</td>
<td>86.96</td>
<td>78.26</td>
<td>82.38</td>
</tr>
<tr>
<td>Sports programs</td>
<td>12452</td>
<td>36</td>
<td>94.44</td>
<td>91.67</td>
<td>96.99</td>
<td>11</td>
<td>100</td>
<td>73.64</td>
<td>84.82</td>
</tr>
<tr>
<td>Challenge at Glen Canyon (TRECVID 2001)</td>
<td>48451</td>
<td>231</td>
<td>97.32</td>
<td>96.67</td>
<td>96.99</td>
<td>11</td>
<td>100</td>
<td>73.64</td>
<td>84.82</td>
</tr>
<tr>
<td>The Great Web of Water (TRECVID 2001)</td>
<td>50569</td>
<td>380</td>
<td>99.01</td>
<td>95.83</td>
<td>97.39</td>
<td>151</td>
<td>95.46</td>
<td>72.73</td>
<td>82.56</td>
</tr>
<tr>
<td>Spaceworks Episode 3 (TRECVID 2001)</td>
<td>52927</td>
<td>183</td>
<td>99.23</td>
<td>97.54</td>
<td>98.38</td>
<td>116</td>
<td>78.45</td>
<td>64.54</td>
<td>70.82</td>
</tr>
<tr>
<td>Average</td>
<td>28558</td>
<td>167</td>
<td>98.04</td>
<td>96.02</td>
<td>97.02</td>
<td>55</td>
<td>90.85</td>
<td>71.13</td>
<td>79.62</td>
</tr>
</tbody>
</table>

Condition 1 is no doubt the shot boundary because shot changes result in visual content significant change. Moreover, condition 2 is based on the analysis of shot changes. Without considering the pervious DC frame histogram difference, it may lead to a lot of false alarms. Significantly, with the measurement of the similarity at several consecutive frames, the proposed algorithm could avoid the false alarms of abrupt shot change detection due to camera motion such as panning or zooming.

After abrupt shot change detection, the residual frames are sent into gradual shot change detection. First, camera and object motions always introduce a larger variation than gradual shot change. This variation may cause false alarm of gradual shot change detection. Although the applied histogram difference is insensitive to object motion, gradual shot change detection remains sensitive to camera motion. To eliminate this effect, an automatic de-noise processing on each frame by Wavelet smoothing is performed. The transformed points with high frequency are removed. Second, the block-wise histogram difference (in (3)) and its spatial variation are extracted to detect the gradual shot changes. Gradual shot change consists of dissolve, fade in and fade out. Theoretically, a dissolve operation from scene $X$ to scene $Y$ in time duration $T$ can be represented as...
\[
\frac{t}{T} Y + \left(1 - \frac{t}{T}\right) X, \quad 0 \leq t \leq T
\] 

(5)

In (5), a fade in is a particular case with \(X = 0\), and fade out is another case with \(Y = 0\). This means that the content of consecutive frames in gradual shot change has increasing difference. To detect the gradual shot change, a statistical model is applied as depicted in Fig. 2. The block-wise information is manipulated. If there exists increasing block-wise histogram difference in the consecutive frames, these frames are detected as gradual shot change. With this model, the HDM assesses not only the global changes but also the local variations. Once the feature is extracted and matched with the model, a gradual shot change is detected.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Abrupt and gradual shot changes are separately evaluated because of their different performance measures. Simulated videos consist of TRECVID 2001 dataset, movie, music, news and sport videos. These videos have 1170 abrupt shot changes and 382 gradual shot changes. The performance of a shot change detection algorithm is evaluated by recall and precision. Recall denotes the percentage of correct detected shot changes which is performed by the detection algorithm with respect to the ground-truth shot changes. The precision is the percentage of correct detected shot changes with respect to the detected shot changes. The recall and precision are defined as

\[
\text{Recall} = \frac{N_c}{N_c + N_m} \times 100\% \quad (6)
\]

\[
\text{Precision} = \frac{N_c}{N_c + N_f} \times 100\% \quad (7)
\]

where \(N_c\) is the number of correct detected shot changes, \(N_m\) is the number of miss detected shot changes, \(N_f\) is the number of false detected shot changes, \(N_c + N_m\) is the number of the ground-truth shot changes, and \(N_c + N_f\) is the number of detected shot changes. Additionally, to evaluate recall and precision both, we use a common metric F1 defined as follows

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)
\]

F1 combines recall and precision. Thus, F1 can objectively reflect the value of precision and recall. If the value of precision and recall are high, the value of F1 is high.

The experimental results are presented in Table 1. For the abrupt shot change detection, the average recall is about 98.04%, the average precision is about 96.02%, and the average F1 is 97.02. However, most of the miss detections are due to the great variation. The false alarm results from the camera motion, such as multi-angular camera shooting. Fig. 3 shows a detected abrupt shot change in our experiment.

For the gradual shot change detection, we use the statistical model proposed in Section 3 to detect the gradual shot changes. Once the feature is extracted and matched with the model, we declare that the gradual shot change occurs. As shown in Table 1, the average recall of gradual shot changes reaches 90.85%, and the average F1 is 79.62. In actual, the gradual shot change detection is more difficult than abrupt shot change detection. With the benefit of our proposed approach, the performance of gradual shot change detection can be greatly enhanced. The false alarm and miss detection has been decreased by our proposed approach. Fig. 4 depicts a detected gradual shot change. Furthermore, our algorithm is objectively compared with the method in [3]. The comparison results are illustrated in Table 2. Our recall and precision are all higher than that of [3]. These data demonstrate the competitive performance of our proposed algorithm.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrupt Shot Change Detection Comparison in TRECVID 2001 Dataset</td>
</tr>
<tr>
<td>Our method</td>
</tr>
<tr>
<td>Recall (%)</td>
</tr>
<tr>
<td>Precision (%)</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

A shot change detection algorithm (HDM) is proposed in this paper. This study applies the global and local information as the extracted features and proposes an adaptive threshold algorithm with a statistical model to detect shot change. The experiment results show the relative high detection precision (96.02%). Thus, our proposed algorithm could benefit the shot change detection for video segmentation applications.

REFERENCES


