Analysis of Shot Boundary Based on Color and Texture Features of Frame

To cite this article: Chuen-Horng Lin et al 2013 IOP Conf. Ser.: Mater. Sci. Eng. 53 012004

View the article online for updates and enhancements.
Analysis of Shot Boundary Based on Color and Texture Features of Frame

Chuen-Horng Lin¹, Muh-Don Hsiao² and Li-Jung Fu³

¹Department of Computer Science and Information Engineering, ²,³Department of Information Management, National Taichung University of Science and Technology, No. 129, Sec. 3, Sanmin Rd., Taichung, Taiwan, R.O.C.

E-mail:¹linch@nutc.edu.tw,²paul@nutc.edu.tw and ³j0916637996@gmail.com

Abstract. This paper proposed a shot boundary detection method based on color and texture features of frames. The shot boundary detection is employed in order to build a highly efficient method for video description. The research methods include feature extraction using color and texture feature, and shot boundary detection offering an adaptive measuring method. In this paper, the color and texture features are Traditional Color Histogram (TCH) and Histogram of Gradient Directions (HGD), respectively. The adaptive measuring method was proposed for the threshold of distance as the basis of identifying shot changes between frames. In order to validate the shot boundary detection method proposed herein, the video database provided by Carleton University website was used for the experiment. The experimental and comparison results of shot boundary detection were validated, and the experimental results were analyzed. This study carried out a series of comparisons and analyses for the above video database, and proved that the shot boundary detection method proposed could effectively detect the shot boundary.

Keywords: shot boundary detection, texture feature, color feature, adaptive measuring.

1. Introduction

Under the rapid development of information technology and multimedia technology, many electronic imaging equipments have become popular, resulting in the growing amount of multimedia information. Facing such massive multimedia information, how to create a mechanism to search for multimedia information effectively is a topic worth exploring. As most multimedia information can be obtained from the internet, it is very convenient for users; however, in the case of fuzzy keyword definition and different classification rules when users are selecting the type of information, the search result may not be satisfactory, and the search efficiency is low.

At present, there are three common methods to search for multimedia information (including word, excel, power point, pdf). The first method uses keywords to search for e-articles, using keywords to search for the related information; the second method uses keywords to search for images, comparing the description of video with the keywords; the third method uses keywords to search for videos. These three search methods can result in thousands of data, images or videos as search results in a very short time. However, the keywords cannot explicitly describe the content of the data to be searched, so the users have to screen the data.

¹ To whom any correspondence should be addressed
If the keyword is replaced by video description, the problems in the keyword-based search can be reduced. The video description mode detects the features from the video content to represent the video; however, as the technology advances, the video can generate diversified shot change automatically by software, so that the video cut becomes more convenient. The video production modes have no specific standard, they vary with the video characteristics; thus, how to define the video content is not defined, as long as the user requirements are met.

A video is composed of a series of consecutive frames, and a frame is a static image. As the relation between consecutive frames is very complex, and the video is composed of different consecutive frames, general studies on video retrieval are divided into feature extraction, video cut/shot boundary detection, key feature frame selection and video retrieval/indexing.

Previous studies on image retrieval and video retrieval are mostly based on the image processing and computer vision theory. In terms of image retrieval, many studies use the content based image retrieval (CBIR) method to replace the keyword retrieval mode. At present, the CBIR method is mostly used, and the technology has been mature. In terms of video retrieval, Deng et al.[2,3] focused on the content-based image and video retrieval (CBIVR) field. The features in the frame data are extracted and compared with the frame features in the image database for the similarity to find out the most similar frame to be retrieved. This method no longer requires manual annotation.

The CBIVR system extracts the features from the frames directly, and then identifies the boundary of frame conversion by comparing the frames. The mostly used frame features are color[1,4], texture[1,5-6], spatial relation[7], shape[8], space-time feature[9], key point feature[10] and motion mode feature[11].

This paper proposes an efficient video description mode, and creates a new content-based video retrieval algorithm. The main processing procedure of video content description in this paper is to extract the frame features, detect different shots, filter the shot noise with fuzzy content definition, cluster the shots, select the representative key feature set of shot, and classify the inconsecutive shots with similar contents. The shot description of video is then established based on the key feature set for redescribing the entire video.

This paper proposed a shot boundary detection method based on color and texture features of frames. Section 2.1 presents the Video Assumption. Feature Extraction will be introduced in section 2.2. Section 2.3 presents the Distance Measurement of Feature. Experiments and comparisons with other approaches are presented in Section 3. Conclusions will be offered in Section 4.

2. The Proposed Method
The query processing in a video retrieval is as shown in Figure 1, including videos video assumption, feature extraction, distance measurement of feature, shot boundary detection, shot noise filter, shot cluster, key feature set of cluster and shot, shot classification and video retrieval.
2.1. Video Assumption
A video is composed of multiple different scenes, a scene is composed of multiple shots, a shot is composed of a series of consecutive frames, and a frame is a static image, as shown in Figure 2.

First, each shot is detected from the frame sequence and expressed as $S_i, i=1,2,...,M$, there are $M$ shots, the shot with fuzzy content definition is expressed as $S^{\text{F}}_i, i=1,2,...,\tilde{M}$, the total number is $\tilde{M}$. After shot noise filter, there are $N (=M-\tilde{M})$ effective shots. Video is expressed as follows:

$$video = V \cup \tilde{V}, \quad V = \bigcup_{i=1}^{M} S_i, \quad \tilde{V} = \bigcup_{i=1}^{\tilde{M}} S^{\text{F}}_i,$$

(1)

Then, the shots are clustered according to their characteristics, if No. $n$ shot is classified as $L_n$ cluster, expressed as $C_n L_n$. The key feature set of cluster is taken out of each cluster. If the key feature
set of No. \( L_n \) cluster of No. \( n \) shot is expressed as \( k(C_n, L_n) \), and the key feature set of shot of No. \( n \) shot is expressed as \( k(S_n) \), the shots of the video are classified finally, No. \( R \) class is expressed as \( CS_k \).

2.2. Feature Extraction

A video is composed of a series of consecutive frames, and the processing time for extracting complex features may be too long. Therefore, this paper uses universally calculated and simple features, which are Traditional Color Histogram (TCH) and Histogram of Gradient Directions (HGD).

2.2.1. Traditional Color Histogram (TCH). This paper uses the color histogram as one of frame features, as the frame has three components of RGB. The histogram of three components of \( j \)-th frame can be expressed as follows:

\[
H_R^j = (H_{R,0}^j, H_{R,1}^j, \ldots, H_{R,255}^j),
H_G^j = (H_{G,0}^j, H_{G,1}^j, \ldots, H_{G,255}^j),
H_B^j = (H_{B,0}^j, H_{B,1}^j, \ldots, H_{B,255}^j),
\]

where \( H_{R,i}^j \), \( H_{G,i}^j \), and \( H_{B,i}^j \) denote the total number of pixels of \( R \), \( G \) and \( B \) components of \( j \)-th frame on \( i \)-th grey respectively, and there are 768 features.

2.2.2. Histogram of Gradient Directions (HGD). The color histogram difference method is very sensitive to the variance in pixel tone scale. There may be excessive detection, so the texture feature is included. Considering the texture of frame, the color frame is converted into gray scale frame, RGB to YCbCr conversion is the most commonly used color coordinate system for image processing. Y is the luminance component and \( C_b \) and \( C_r \) are the chrominance components. The luminance edge is determined using Y information to perform the edge detection method described above. RGB of the pixel \((x, y)\) was transformed into the YCbCr domain using the following formula:

\[
\begin{bmatrix}
Y \\
C_b \\
C_r
\end{bmatrix} = \begin{bmatrix}
0.209 & 0.587 & 0.114 \\
-0.168 & -0.331 & 0.5 \\
0.5 & -0.418 & -0.081
\end{bmatrix} \times \begin{bmatrix}
R \\
G \\
B
\end{bmatrix} + \begin{bmatrix}
128 \\
128 \\
128
\end{bmatrix}
\]

Sobel edge detection was employed to detect the gradient feature. The gradient magnitude \( \nabla Y \) and direction \( \theta(x, y) \) for \( Y(x, y) \) of pixel at coordinates \((x, y)\) are given by:

\[
\nabla Y = g(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)} \quad \text{and} \quad \theta(x, y) = \tan^{-1}\left(\frac{g_y(x, y)}{g_x(x, y)}\right)
\]

In most cases, the edge variation can be expressed as horizontal direction \( g_x \) and vertical direction \( g_y \). A 3\times3 block was used to compute the variations in the horizontal and vertical directions \( g_x \) and \( g_y \).

The gradient direction histogram\[12\] calculates the cumulative distribution of gradient size in gradient direction as one of frame features. In order to avoid too sensitive gradient direction, the gradient direction is selected in a region, the gradient direction is between 0 and 360°, \( \hat{\delta} \) is one unit. The gradient sizes of pixels in this direction range are calculated, a frame can establish the histogram of values in \( \frac{360°}{\hat{\delta}} \) gradient direction regions (\( \frac{360°}{\hat{\delta}} \) regions), expressed as follows:
\[ H_k^j(\delta) = \sum_{i=1}^{N^j} HGD(H_k^j, g_i^j, \theta_i^j), \quad (5a) \]

and

\[ HGD(H_k^j, g, \theta) = \begin{cases} H_k^j + g_i^j, & \text{if } (k-1) \times \delta \leq \theta_i^j \leq k \times \delta, \\ 0, & \text{otherwise} \end{cases}, \quad (5b) \]

\( H_k^j \) denotes the total gradient size in \( j \)-th frame and gradient direction in \( k \)-th region, \( N^j \) is the total number of pixels in \( j \)-th frame, \( g_i^j \) and \( \theta_i^j \) are the gradient size and direction on \( i \)-th pixel of \( j \)-th frame respectively, the range value of \( k \) is \( 0 < k \leq \frac{360}{\delta} \). \( H_k^j \) is regarded as one feature of the frame in this study, so there are 12 features.

### 2.3. Distance Measurement of Feature

TCH and HGD are useful to describe the relationship between color and texture of an image. Due to significantly complementary, these two features are integrated to establish a color-texture and color-histogram based frame matching between frames. TCH of the \( j-1 \) frame and the \( j \) frame as \( (H_{R_k}^{j-1}, H_{G_k}^{j-1}, H_{B_k}^{j-1}) \) and \( (H_{R_k}^j, H_{G_k}^j, H_{B_k}^j) \) are obtained from equation (6). The frame matching distance \( \Delta_{j}^{TCH} \) between frames based on TCH is formulated as the following:

\[ \Delta_{j}^{TCH} = \sum_{i=1}^{255} \left( \frac{H_{R_k}^{j-1} - H_{R_k}^j}{\max(H_{R_k}^{j-1} + H_{R_k}^j, V)} + \frac{H_{G_k}^{j-1} - H_{G_k}^j}{\max(H_{G_k}^{j-1} + H_{G_k}^j, V)} + \frac{H_{B_k}^{j-1} - H_{B_k}^j}{\max(H_{B_k}^{j-1} + H_{B_k}^j, V)} \right), \quad (6) \]

where \( V \) is any small number that avoids denominator is 0.

Considering HGD \( H_k^{j-1}(\delta) \) and \( H_k^j(\delta) \) of the \( j-1 \) and \( j \) frame are obtained from equation (7). The definition of the frame matching distance \( \Delta_{j}^{HGD} \) between frames based on the HGD is shown as the following equation:

\[ \Delta_{j}^{HGD} = \sum_{i=1}^{360} \left( \frac{H_k^{j-1} - H_k^j}{\max(H_k^{j-1} + H_k^j, V)} \right), \quad (7) \]

The frame matching combines the TCH and HGD to quantize the similarity of the \( j-1 \) and \( j \) frame. Using such frame matching, one can define the frame matching distance \( \Delta_{j}^{TH} \) between the \( j-1 \) and \( j \) frame as:

\[ \Delta_{j}^{TH} = \Delta_{j}^{TCH} + \Delta_{j}^{HGD} \quad (8) \]

where \( \Delta_{j}^{TH} \) decreases with increasing similarity of \( Q \) and \( D \).

### 2.4. Shot Boundary Detection

This paper takes the distance measurement \( \Delta_{j}^{TH} \) between two frames as the basis of identifying shot change. The distance measurement between two similar frames is small, and the distance measurement of shot change is large.
The measured distance value between two frames determines the shot change. If the measured distance is greater than the threshold $T_H$, it is regarded as shot change, as shown in Figures 3 and 4; otherwise, the two frames are similar.

![Figure 3. The measured distance value between two frames.](image)

![Figure 4. The shot boundary between two frames [13].](image)

As the conversion of frame is not always sudden conversion, the conversion of some frames is slow. If only the difference between two consecutive frames is used for detection, the detection of shot will be influenced. Therefore, the shot boundary detection proposed in this paper can be divided into three steps to detect the boundary of shot change.

**Step 1:** if the measured distance value of $j$-th frame $\Delta^{TH}_{j} > w_{1}\mu$, the inter frame is called candidate shot change boundary; otherwise, it is called non-shot change boundary, $\mu$ is the measured mean value of distance of all the frames in the video, expressed as follows:

$$\mu = \frac{1}{N-1} \sum_{j=2}^{N} \Delta^{TH}_{j}, \quad (9)$$
where $N$ the total number of frames of the video, $\Delta_j^{TH}$ is the difference between $j$-th and $(j-1)$-th, $w_j$ is the weight. For the candidate shot change boundary, considering the problems resulted from slow change, a frame difference value correction method is proposed. The difference value of frame is not merely the difference between two consecutive frames. Thus, this paper proposes three levels of thresholds to make four intervals, as shown in Figure 5.

![Figure 5. Three levels of thresholds.](image)

If the difference value is between $T_M$ to $T_C$, the difference values are in bell-shaped distribution and expressed as low width, as shown in Figure 6 (a), the frame range value of this interval is set as three frames ($d = 3$) in front and behind respectively. If the difference value is between $T_C$ to $T_H$, the difference values are in normal distribution and the bell shape is expressed in medium size, as shown in Figure 6 (b). The frame range value of this interval is set as two frames ($d = 2$) in front and behind respectively. If the difference value is greater than $T_H$, the difference values are in steep shape and the bell shape is expressed in high narrow, as shown in Figure 6 (c). The frame range value of this interval is set as one frame ($d = 1$) in front and behind respectively.

![Figure 6. Three distributions of the difference value.](image)

Three difference thresholds $T_M$, $T_C$ and $T_H$ are defined as high difference threshold $T_H$, medium high difference threshold $T_C$ and mean difference threshold $T_H$ respectively, expressed as follows:

$$T_M = w_2 \mu, T_C = w_3 \mu \text{ and } T_H = w_4 \mu,$$

where $w_2, w_3$ and $w_4$ are the weights. The maximum difference value of the region is recalculated according to the designed frame range, expressed as follows:

$$\Delta_j^{max_L} = \max[\Delta_j^{TH}, \Delta_{j+d-1}^{TH}, \Delta_{j+d}^{TH}].$$

Finally, the difference value $\Delta_j^{TH}$ on candidate shot change boundary must meet the following conditions:
where $w_5$ is the weight of basic difference value and $\Delta_j^{TH} = \frac{w_6 \Delta_j^{max}}{10}$ and $\Delta^{max} = \max(\Delta_j^{TH})$, $j = 2, ..., N$. If the difference value on the candidate shot change boundary does not meet the aforesaid conditions, the candidate condition is cancelled.

**Step 2:** The range of frame will be expanded in this stage, and three levels of thresholds are proposed to distinguish four intervals, as shown in Figure 7.

$$d = 6 \quad d = 3 \quad d = 1$$

![Figure 7. Three levels of thresholds.](image)

If the difference value is between $\Delta_j^{TH}$ and $\Delta_j^{max}$, the difference values are in bell-shaped distribution and expressed in low width, as shown in Figure 6 (a), the frame range value of this interval is increased to six frames ($d = 6$) in front and behind respectively. If the difference value is in $\Delta_j^{TH}$ and $\Delta_j^{TH}$, the difference values are in normal distribution and the bell shape is expressed in medium size, as shown in Figure 6 (b), and the frame range value of this interval is increased to three frames ($d = 3$) in front and behind respectively. If the difference value is greater than $\Delta_j^{TH}$, the difference values are in steep shape and the bell shape is expressed in high narrow, as shown in Figure 6 (c). The frame range value of this interval is also one frame ($d = 1$) in front and behind respectively.

Three difference thresholds $T_{M2}$, $T_{C2}$ and $T_{H2}$ are defined as Stage II high difference threshold $T_{H2}$, Stage II medium high difference threshold $T_{C2}$ and Stage II mean difference threshold $T_{M2}$ respectively, expressed as follows:

$$T_{M2} = w_2 \mu, T_{C2} = w_6 \mu, \text{and} \ T_{H2} = w_6 \mu$$

where $w_7$, $w_9$ and $w_6$ are the weights of parameters.

The final step is to check whether the difference value on the candidate shot change boundary is the maximum value in the frame range. If yes, it is defined as candidate shot change boundary; otherwise, the candidacy is cancelled.

**Step 3:** This stage is the last stage eliminating the candidate shot change boundary, considering the removal of non-shot change boundary. The $j$-th candidate shot change boundary must meet the following conditions:

$$|\Delta_j^{TH} - \Delta_{j-1}^{TH}| \leq w_{10} \mu \text{ or } |\Delta_j^{TH} - \Delta_{j+1}^{TH}| \leq w_{10} \mu$$

where $\Delta_j^{TH}$ and $\Delta_{j-1}^{TH}$ are the absolute difference of difference of $j$-th and $(j-1)$-th frame, $\Delta_j^{TH}$ and $\Delta_{j+1}^{TH}$ are the absolute difference of difference of $j$-th and $(j+1)$-th frame, and $w_{10}$ is the weight. If the candidate shot change boundary meets the above conditions, the candidacy is cancelled, and the rest are the shot change boundaries in this paper.

## 3. Experiments

Two types of video database are used for this experiment. The first type is the video database provided by Carleton University[13] website, for validating the shot boundary detection proposed in this paper. This type of video database contains ten videos, as shown in Table 1, and the objects in all the videos move very fast. The colors of No. 2, 3 and 5 videos are plain and relatively dark. Although the content of No. 10 video is relatively dark, the colors are multiple. The quantity of consecutive frames of most
shots is less than 10, and some dissolution effects are usually added in shot change, so that the detection of shot boundary becomes more difficult.

**Table 1.** The first types of video database.

| No. | No.1 | No.2 | No.3 | No.4 | No.5 |
|-----|------|------|------|------|------|
| 1   | ![Image](image1) | ![Image](image2) | ![Image](image3) | ![Image](image4) | ![Image](image5) |
| 2   | ![Image](image6) | ![Image](image7) | ![Image](image8) | ![Image](image9) | ![Image](image10) |

As the first type of video database provides diversified videos, including cartoons, dramas and black and white videos, the threshold parameter of each video is adjusted slightly to conform to different types of videos. Table 2 shows the setting of parameter weights $w_1$ and $w_6$. The content of No. 10 video is relatively dark and the shot change is implemented by dissolve. Thus, all the feature differences are relatively small, and $w_1$ is reduced to 1. As for the setting of parameter weights $w_7$ to $w_9$, No. 1, 2, 4, 6, 7 and 9 videos have the same setting. As No. 3 video is of gray image, the content of No. 5 and 8 videos is relatively dark, and the background is black. The objects in No. 8 video move fast, the set values must be adjusted. As for the setting of parameter weight $w_{10}$, No. 3, 7 and 10 are adjusted down properly for slow frame conversion, relatively dark video content and black background.

**Table 2.** The weight values in this paper.

| No. | $w_1$ | $w_2$ | $w_3$ | $w_4$ | $w_5$ | $w_6$ | $w_7$ | $w_8$ | $w_9$ | $w_{10}$ |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 1     |
| 2   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 1     |
| 3   | 2     | 1     | 2     | 3     | 0.1   | 3     | 1     | 2     | 3     | 0.5   |
| 4   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 1     |
| 5   | 2     | 1     | 2     | 3     | 0.1   | 3     | 1     | 2     | 3     | 1     |
| 6   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 1     |
| 7   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 0.5   |
| 8   | 2     | 1     | 2     | 3     | 0.1   | 3     | 1.5   | 3     | 5     | 1     |
| 9   | 2     | 1     | 2     | 3     | 0.1   | 3     | 2     | 4     | 6     | 1     |
| 10  | 1     | 1     | 2     | 3     | 0.1   | 3     | 0.5   | 1     | 2     | 0.1   |

The shot boundary detection results in this study are divided into Fuzzy Logic Approach[14], Edge Tracking[15], Histogram[16] and Pixel Localization[17] for comparison and analysis. The Fuzzy Logic Approach applies color, texture, edge difference and motion compensation to video shot boundary. The Edge Tracking uses texture feature and edge tracking to generate different frames to detect the boundary of slow conversion. The Histogram calculates the difference between two consecutive frames. Pixel Localization uses texture feature and the inter frame in image compression to calculate the difference between regional pixels to detect the shot boundary.

Three efficiency evaluation standards are used in this paper for the efficiency evaluation of experimental results of shot boundary detection, which are precision, recall[17] rate and F1 index[14],
expressed as follows:

\[
\text{precision} = \frac{\text{correct}}{\text{correct} + \text{falsepositive}}, \quad \text{recall} = \frac{\text{correct}}{\text{correct} + \text{missed}},
\]

\[
F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}},
\]

where correct is the ground truth of shot boundary, falsepositive is not a shot boundary, missed refers to the shot boundary is not detected. The result of shot boundary and the efficiency evaluation are shown in Table 3. The percentage sign of numerical values is omitted, and the red bold numerical values represent better results. The precision, recall rate and comprehensive F1 index of No. 1, 2, 4, 6, 7 and 9 videos detected using this method have reached 100%. The precision, recall rate and comprehensive F1 index of No. 3, 5, 8 and 10 videos are lower than that by other methods. The factors influencing the precision and recall rate and the correction of the ground truth are described below.

| No. | Present method | Fuzzy logic | Edge tracking | Histogram | Pixel localization |
|-----|----------------|-------------|---------------|-----------|-------------------|
|     | P   | R   | F1 | P   | R   | F1 | P   | R   | F1 | P   | R   | F1 | P   | R   | F1 | P   | R   | F1 | P   | R   | F1 |
| 1   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 37.5 | 54.5 | 82.5 | 82.5 | 82.5 | 82.5 | 82.5 | 82.5 |
| 3   | 90.7 | 86.0 | 88.3 | 88.1 | 100 | 93.7 | 59.5 | 87.0 | 70.7 | 93.6 | 53.6 | 68.2 | 76.4 | 77.8 | 77.1 | 82.5 | 82.5 | 82.5 |
| 4   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 94.1 | 97.0 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5   | 90.0 | 96.4 | 93.1 | 93.8 | 100 | 96.8 | 93.8 | 100 | 96.8 | 95.5 | 70.0 | 80.8 | 86.7 | 86.7 | 86.7 | 86.7 | 86.7 | 86.7 |
| 6   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 7   | 100 | 100 | 100 | 95.0 | 100 | 97.4 | 81.0 | 94.4 | 87.2 | 100 | 66.7 | 80.0 | 70.8 | 99.4 | 82.7 | 82.7 | 82.7 |
| 8   | 87.8 | 90.0 | 88.9 | 90.9 | 100 | 95.2 | 89.5 | 89.5 | 89.5 | 97.1 | 89.5 | 93.0 | 92.7 | 100 | 96.2 | 96.2 | 96.2 |
| 9   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 50.0 | 75.0 | 100 | 100 | 100 | 100 | 100 | 100 |
| 10  | 72.2 | 72.2 | 72.2 | 92.9 | 89.7 | 91.3 | 49.7 | 89.7 | 64.0 | 85.0 | 39.5 | 54.0 | 62.3 | 54.0 | 57.9 | 57.9 | 57.9 | 57.9 | 57.9 |

Table 3. The precision of these methods.

The result of false shot boundary detection of No. 3 video is shown in Figure 8. As the content of consecutive frames is relatively dark or relatively bright, the detection of shot boundary is influenced, as shown in Figure 8 (a) and Figure 8 (b). The red frame is the false boundary detected by this method. The traveling speed of objects in the video results in falsely detected shot boundary, as shown in Figure 8 (c) and Figure 8 (d).
Different users have different opinions on the shot boundary, so that the recall rate in the experimental results is reduced. As a result, the videos with problems are proposed and described. The real shot data of No. 3 video are compared with the data after practical viewing for analysis. It is found that the difference between the first frame and the second frame is larger than that between the third frame and the fourth frame. Therefore, it is appropriate to change the fourth frame of the ground truth to the second frame. The yellow frame is changed the ground truth as shown in Figure 9.

The ground truth of No. 5 video are compared with the data after practical viewing for analysis. As shown in Figure 10 (a), there is a short dissolution, and the ground truth of the shot boundary define the frame after dissolution (No. 249). This paper argues that the frame (No. 248) in the change
between shots should be defined. Figure 10 (b) has the same situation, namely the first and the second frame can be identified as the same shot; thus, this paper suggests that the third frame is appropriate. The yellow frame in Figure 10. is the changed the ground truth of the shot boundary.

![Figure 10](image)

**Figure 10.** The frames between two shot.

The ground truth of the shot boundary in No. 10 video are compared with the data after practical viewing for analysis. It is found that the consecutive frames are the course of dissolution, as shown in Figure 11 (a). It is appropriate to change the ground truth to yellow frame. Many boundaries in No. 10 video are converted by dissolution, as the ground truth have not indicated these dissolutions. This paper selects a set of consecutive frames as indication. The yellow frame in Figure 11 (b). is the increased ground truth.

![Figure 11](image)

**Figure 11.** The frames between the different shot.

According to the aforesaid modification of real shot data, the precision, recall rate and comprehensive F1 index of the 10 videos are recalculated, as shown in Table 4. The numerical values of No. 3, 5 and 10 videos are still increased after the real shot data are modified, thus proving that this method can detect the short dissolution of sudden change and slow change.
Table 4. The precision of these methods for redefined the ground truth.

| No. | Present method | Fuzzy logic | Edge tracking | Histogram | Pixel localization |
|-----|----------------|-------------|---------------|-----------|-------------------|
| 1   | 100 100 100    | 100 100 100 | 100 100 100   | 100 100 100| 100 100 100       |
| 2   | 100 100 100    | 100 100 100 | 100 100 100   | 100 37.5 54.5| 82.5 82.5 82.5   |
| 3   | 92.6 87.7 90.1 | 88.1 100 93.7| 59.5 87.0 70.7| 93.6 53.6 68.2| 76.4 77.8 77.1   |
| 4   | 100 100 100    | 100 100 100 | 100 100 100   | 100 94.1 97.0| 100 100 100       |
| 5   | 93.3 100 96.5 | 93.8 100 96.8| 93.8 100 96.8 | 95.5 70.0 80.8| 86.7 86.7 86.7   |
| 6   | 100 100 100    | 100 100 100 | 100 100 100   | 100 100 100| 0.0 0.0 0.0       |
| 7   | 100 100 100    | 95.0 100 97.4| 81.0 94.4 87.2| 100 66.7 80.0| 70.8 99.4 82.7   |
| 8   | 87.8 90.0 88.9 | 90.9 100 95.2| 89.5 89.5 89.5| 97.1 89.5 93.0| 92.7 100 96.2   |
| 9   | 100 100 100    | 100 100 100 | 100 100 100   | 100 50.0 75.0| 100 100 100       |
| 10  | 88.9 96.7 92.6 | 92.9 89.7 91.3| 49.7 89.7 64.0| 85.0 39.5 54.0| 62.3 54.0 57.9   |

4. Conclusions
This paper proposed a shot boundary detection method based on color and texture features of frames, and performed a series of analyses. The experimental results of comparison and analysis of this shot boundary detection method were compared with Fuzzy Logic Approach, Edge Tracking, Histogram and Pixel Localization. The analysis of experimental results showed that the ground truth provided by website has obvious inaccuracy in recognition, so that the analysis result is influenced. The boundary of shot was detected effectively, proving that the proposed method can be used for shot filter, cluster, classification and retrieval in the future. This method can also be developed into a rapid and convenient video retrieval system.

5. Acknowledgements
This work was supported in part by National Science Council, Taiwan, under GRANT NO. NSC 101-2221-E-025-009.

References
[1] C. H. Lin, R. T. Chen, and Y. K. Chan 2009 A smart content-based image retrieval system based on color and texture feature *Image and Vision Computing* vol 27, pp 658-665
[2] A. F. Smeaton, P. Over, and A. R. Doherty 2010 Video shot boundary detection: Seven years of TRECVID activity *Computer Vision and Image Understanding* Vol 114, pp 411-418
[3] H. Lee, J. Yu, Y. Im, J. M. Gil, and D. Park 2011 A unified scheme of shot boundary detection and anchor shot detection in news video story parsing *Journal Multimedia Tools and Applications* Vol 51, pp 1127-1145
[4] M. H. Hung, C. H. Hsieh, C. M. Kuo, and J. S. Pan 2011 Generalized playfield segmentation of sport videos using color features *Pattern Recognition Letters* Vol 32, pp 987-1000
[5] A. Hafiane and B. Zavidovique 2008 Local relational string and mutual matching for image retrieval *Information Processing and Management* Vol 44, pp. 1201-1213
[6] C. H. Lin and Y. J. Syu 2010 Fast segmentation of porcelain images based on texture features *Journal of Visual Communication and Image Representation* Vol 21, pp 707-721
[7] C. H. Lin, Y. K. Chan, K. H. Chen, D. C. Huang, and Y. J. Chang 2011 Fast color-spatial feature based image retrieval methods *Expert Systems with Applications* Vol 38, pp 11412-11420
[8] C. H. Wei, Y. Li, W. Y. Chau, and C. T. Li 2009 Trademark image retrieval using synthetic features for describing global shape and interior structure *Pattern Recognition* Vol 42, pp 386-394

[9] W. Chen and Y. J. Zhang 2008 Parametric model for video content analysis *Pattern Recognition Letters* Vol 29, pp 181-191

[10] Y. G. Jiang, J. Yang, C. W. Ngo, and A. G. Hauptmann 2010 Representations of keypoint-based semantic concept detection: A comprehensive study *IEEE Transactions on Multimedia* Vol 12, pp 42-53

[11] A. M. Amel, B. A. Abdessalem, and M. Abdellatif 2010 Video shot boundary detection using motion activity descriptor *Journal of Telecommunications* Vol 2, pp 54-59

[12] S. R. Fountain and T. N. Tan 1998 Efficient rotation invariant texture features for content-based image retrieval *Pattern Recognition* Vol 31, pp 1725-1732

[13] P. Bose, R. Laganiere, and A. Whitehead, VIVA Research Lab, University of Ottawa, School of Computer Science, Carleton University, Some results in video segmentation, Retrieved on January, 2012, from http://www.site.uottawa.ca/~laganier/videoseg/.

[14] H. Fang, J. Jiang, and Y. Feng 2006 A fuzzy logic approach for detection of video shot boundaries *Pattern Recognition* Vol 39, pp 2092-2100

[15] A. Whitehead, P. Bose, and R. Laganiere 2004 Feature based cut detection with automatic threshold selection *Image and Video Retrieval* pp 2064-2065

[16] S. Pfeiffer, R. Lienhart, G. Kühne, and W. Effelsberg 1998 The MoCA project-movie content analysis research at the university of mannheim *GI Jahrestagung* pp 329-338

[17] J. S. Boreczky and L. A. Rowe 1996 Comparison of video shot boundary detection techniques *Journal of Electronic Imaging* Vol 5, pp 122-128