A Method of Video Shot-Boundary Detection based on Grey Modeling for Histogram Sequence

Xin Liu[1,*] and Jin Dai[2]

Affiliation :[1,2] Chongqing University of Posts and Telecommunications, School of Software Engineering, Chongqing, China, 400065
[1]liuxin@cqupt.edu.cn, [2]daijin@cqupt.edu.cn

Abstract

Video Shot-Boundary Detection is important for video analysis, segmentation and retrieval. The detection results of Shot-Boundary are not only relation with two nearest frames of shot cuts, but also their several neighboring frames. Video stream is multi-stream time sequence. Grey model based on set sequence (SGM) is proposed to analyze and process multi-stream time sequences constructed with sampled histograms in this paper. Consequently, a novel Shot-Boundary Detection method is put forward based on SGM and color histogram, which includes the following steps: (1) When a new video frame coming into the detection system, the image is transformed into intensity image under HSI color model; (2) Histogram of the video frame can be obtained. Then, taking samples for the histogram and constructing set sequence with proximate sampled histograms is implemented; (3) SGM is introduced to simulate the sampled histogram sequence in this detection method. Moreover, absolute mean error (AME) and regulative AME (R_{AME}) with thresholds are applied to make a detection judgment. Finally, the experiments illuminate that the proposed method is an effective Shot-Boundary Detection method, and it is superior to method of Histogram disparity significantly, especially when it is used to detect cuts with some complex video scenes.

Keywords: Video Shot-Boundary Detection, Histogram, video stream sequence, grey modeling.

1. Introduction

Video Shot-Boundary Detection is important for video analysis, video segmentation and video retrieval. A video shot is a sequence of frames taken by one camera during a single continues action in time and space [1, 2]. In real-time shot segmentation, the detection system must make a judgment automatically whether it is the beginning of another shot when a new video frame is coming. So, the Shot-Boundary Detection has fundamental significance for video analysis. There are mainly two transitions between Shots: abrupt shot cut, and gradual transition. However, it makes Shot-Boundary Detection difficult when the following situations appeared, such as, the flashing light and special effects in the shot, or shot quickly moving from one scene to another scene.

There are many related works. Lee et al. proposed a unified scheme of shot boundary detection and anchor shot detection in news video story parsing [3]. A graph partition based video scene boundary detection approach, in which multiple features extracted from the video, was proposed to determine the video scene boundaries through an unsupervised clustering procedure [4]. The top-down search scheme and Frame similarity via local features were used in paper [5]. Objects within shots were detected via local image features and used for revealing visual discontinuities among shots. A video shot segmentation scheme based on a dual-

---

* Corresponding Author: Xin Liu, E-mail: liuxin@cqupt.edu.cn

detection model was proposed by Jiang et al., which included the pre-detection and re-detection processes [6]. The concepts of uneven blocked color histogram difference and uneven blocked pixel value difference based on human visual features were introduced, which were used as the main descriptors of the pre-detection process to enlarge the importance of central areas and to reduce the noises of background movements and logos. In the re-detection round, the scale invariant feature transform was applied to re-detect boundaries so as to improve the detection precision rate. In paper [7], the raw news videos were firstly split into shots by a four-threshold method, and the key frames were extracted from each shot. After that, the anchor person detection was conducted from these key frames by using a clustering-based method based on a statistical distance of Pearson's correlation coefficient. Smeaton et al. presented an overview of the TRECVid shot boundary detection task, a high-level overview of the most significant of the approaches taken, and a comparison of performances, focusing on one year (2005) as an example [8]. Kucuktunc et al. presented a fuzzy color histogram-based shot-boundary detection algorithm specialized for content-based copy detection applications [9]. In paper [2], authors focused on the problem of hard cut detection and proposed an automatic algorithm in order to accurately determine abrupt transitions from video. They suggested a fuzzy rule-based scene cut identification approach in which a set of fuzzy rules was evaluated to detect cuts. Priya et al. proposed a new method for detecting shot boundaries in video sequences by extracting edge strength using orthogonal vectors from blocks of the frames, and transition (boundary) identification procedure which was capable of detecting the shot or non shot boundaries in the video sequences [10]. The adaptive measuring method based on Traditional Color Histogram (TCH) and Histogram of Gradient Directions (HGD) was proposed by Lin et al. for the threshold of distance as the basis of identifying shot changes between frames [11]. In paper [12], the propose algorithm extracted structure features from each video frame by using dual-tree complex wavelet transform. Then, spatial domain structure similarity was computed between adjacent frames. The declaration of shot boundaries was decided based on carefully chosen thresholds. Mendi et al. presented a system for video shot boundary detection and key-frame extraction from video sequences based on color histograms. Computations of color histogram differences, self-similarity modeling, unsupervised k-means clustering had been performed in segmentation process [13]. In paper [14], several local indicators were extracted from MPEG macro-blocks, and Ada-Boost was employed for feature selection and fusion. The selected features were used in classifying candidate cuts into five sub-spaces via pre-filtering and rule-based decision making. Following that, global indicators of frame similarity between boundary frames of cut candidates were examined using phase correlation of dc images. In paper [15], adaptive local threshold was adopted to classify non-boundary segments and candidate segments that may contain shot boundaries. The candidate segments were refined using bisection-based comparisons to eliminate non-boundary frames. Then, refined candidate segments are preserved for further detections. Frame transition parameters and frame estimation errors based on global and local features were used for boundary detection and classification [16]. Lu et al. adopted a candidate segment selection and singular value decomposition (SVD) to speed up the SBD, and color histograms of all frames were used in this paper [17]. Fu et al. proposed an adaptive shot boundary detection algorithm based on HSV space, which considered the characteristics of two shot transitions, used different data processing in cut and gradual transition detection and used partitioning technology combing with the relationship between spatial and color information [18]. In paper [19], the main decision sources were image color histograms, object recognizer results, motion comparators, audio pattern analyzers, keypoint extractors

and edge descriptors which were selectively employed in a cascaded manner. In paper [20], a new SBD method was proposed using color, edge, texture, and motion strength as vector of features (feature vector). Features were extracted by projecting the frames on selected basis vectors of Walsh-Hadamard transform (WHT) kernel and WHT matrix. After extracting the features, based on the significance of the features, weights were calculated. The weighted features were combined to form a single continuity signal, used as input for Procedure Based shot transition Identification process (PBI). In paper [21], color histogram and mutual information were used together to measure the difference between frames, and corner distribution of frames is utilized to exclude most of false boundaries. Vila et al. proposed two different information-theoretic approaches to detect the abrupt shot boundaries of a video sequence which were based on Tsallis mutual information and Jensen-Tsallis divergence used to quantify the similarity between two frames [22]. Hue, Saturation, Value (HSV) histogram was introduced into the pre-processing of the video to output the candidate gradual shot segments, and speeded up robust features (SURF) was used to detect the gradual shot boundaries by Wu et al. [23]. In paper [24], a video shot segmentation scheme with dual-detection model was proposed. In the pre-detection round, the Uneven Blocked differences were presented and used in Adaptive Binary Search (ABS) to detect shot boundaries. In the re-detection round, the Scale Invariant Feature Transform (SIFT) method was applied to exclude false detections. An algorithm of random learning of p-adic neural networks was applied to the problem of detection of changes in streams of video information such as shot boundaries by Benois-Pineau [25].

Color histogram is a common feature used to detect Shot-Boundary in above literatures and other research work. In this paper, color histogram is also applied in the proposed detection method, and grey model for set sequence is introduced to model the histogram information. So, the rest of this paper is organized as follows: the proposed Shot-Boundary Detection method will be expounded in the second part, including the reason and procedure about the proposition of grey model based on set sequence, and the steps of the proposed detection method. The experiments and some comparisons will be expatiated in the third part.

2. The Proposed Method

2.1. Video Frames Processed under HSI Color Model

In this paper, 'I' of every frame under HSI color model is processed for Shot-Boundary Detection, and 'I' is the simplified form of Intensity component. HSI is a color model proposed by an American color-scientist H.A.Munseu in 1915, which reflects the way that human visual system perceives color through three basic components Hue, Saturation and Intensity. The frame will be transformed into intensity image under HSI when it entered into the detection system, and IHSI ('I' under HSI) is used in the proposed method. An example shown as Fig. 1, there is an anchor in this shot and IHSI of this frame is shown in Fig. 1(b).

Figure 1. An Example: (a) a Video Frame in RGB; (b) IHSI of (a)

2.2. Histogram and Sampled Histogram

When to detect the Shot-Boundary, histograms of video frames could be utilized because they are great distinguishing between two different shots. Histogram is sampled for reducing computation in this method. Take two frames from a shot cut as Fig. 2 for example, their histograms are totally dissimilar. Of course, the suitably sampled histograms do not have a great impact on the detection results.
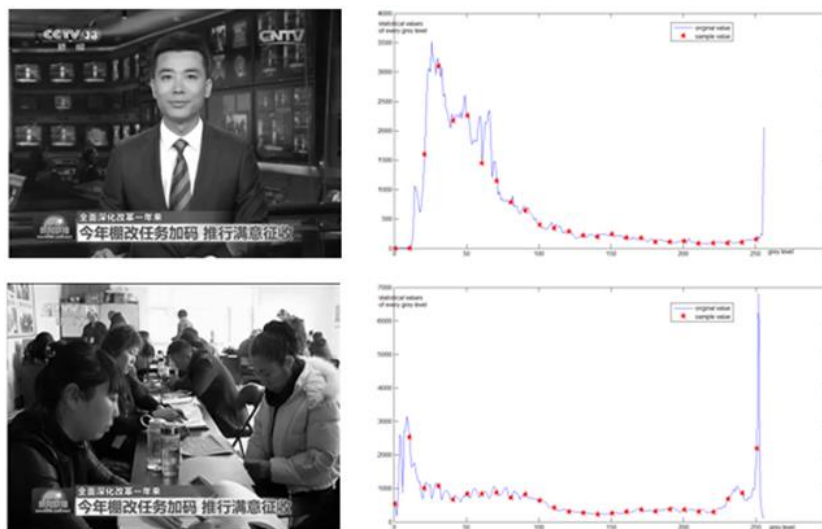

Figure 2. Histograms and Sampled Histograms of a Shot Cut

The detection results of Shot-Boundary are not only relation with two nearest frames of shot cuts, but also their several neighboring frames. So it is difficult to make an accurate detection just using the disparity of two nearest histograms. Grey model is introduced to analyze and process sampled histogram sequences constructed by several neighboring video frames here. Grey modeling is a process of building an approximate differential equation model based on sequences, and the approximate differential equation model is called as Grey Model [26], which is an effective simulation and prediction method for limited time sequence.

2.3. Grey Model for Sampled Histogram

2.3.1 Traditional grey model

Traditional grey modeling is proposed by Deng [27], and some principles and definitions are given here. The original sequence is $x^{(0)}$: $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, $\forall x^{(0)}(k) \in x^{(0)}$, $k \in \{1, 2, \ldots, n\}$. Set $x^{(1)}$ be the AGO generation sequence of $x^{(0)}$, $x^{(1)} = AGO\ x^{(0)}$ can be expressed as $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n))$, $x^{(1)}(k) = x^{(1)}(k-1) + x^{(0)}(k)$, $k \in \{2, 3, \ldots, n\}$, $x^{(1)}(1) = x^{(0)}(1)$. That is, $x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m)$.

$x^{(0)}(k) + az^{(1)}(k) = b$ is definition of grey model GM(1,1), where GM(1,1) is a one-order, one-variable Grey Model. a is the development coefficient, the value and symbol reflect the development trend of $x^{(0)}$ and $x^{(1)}$. b is grey action, obtaining by computing. It is an equivalent action with grey information overlay. $z^{(1)}(k) \in z^{(1)}$, $z^{(1)} = (z^{(1)}(2), \ldots, z^{(1)}(n))$, $z^{(1)}(k) = 0.5\ x^{(1)}(k) + 0.5\ x^{(1)}(k-1)$, $k \in \{2, \ldots, n\}$. $z^{(1)}$ is called as whiten sequence of background value. Any $z^{(1)}(k)$ is mean value of $x^{(1)}(k)$ and $x^{(1)}(k-1)$.

Then, the whiten model of GM(1,1) is: $\dfrac{dx^{(1)}}{dt} + ax^{(1)} = b \cdot$

Its response equation is: $\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \dfrac{b}{a})e^{-ek} + \dfrac{b}{a}$, $\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$.

$[a\ b]^T$ is parameter vector of the model. Set $P_1 = [a\ b]^T$, and the matrix equation of P_1 under least square criterion is shown as follows: $P_1 = [a\ b]^T = (B^T B)^{-1} B^T y_n$,

where, $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$, $y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$.

The sequence constructed by sampled histograms of video frames is also a multi-stream time sequence. Theoretically, GM(1,1) could be used to simulate a histogram sequence through modeling multiple times. For example, if there are m samples in every histogram, GM(1,1) will be modeled m times. However, when the values of one stream among the multi-stream sequence are all equal or have the similar situation as shown in Fig. 3, there will be some abnormal simulation by GM(1,1), and even GM(1,1) cannot be modeled. So, a grey model for multi-stream time sequence is required. In this paper, grey model based on set sequence is proposed to simulate sampled histogram sequences.

2.3.2 Grey model based on set sequence

Let $Xs^{(0)}$ be original set sequence, $Xs^{(0)} = (Xs^{(0)}(1), Xs^{(0)}(2), \ldots, Xs^{(0)}(n))$, where, $Xs^{(0)}(k)$ is a set with m elements, $k \in \{1, 2, \ldots, n\}$. Set $Xs^{(1)}$ be the accumulated generating of $Xs^{(0)}$, then $Xs^{(1)} = $ sAGO $Xs^{(0)}$, $Xs^{(1)} = (Xs^{(1)}(1), Xs^{(1)}(2), \ldots, Xs^{(1)}(n))$, $\forall\ Xs^{(1)}(k) \in Xs^{(1)}$, $k \in \{1, 2, \ldots, n\}$, $Xs^{(1)}(k)$ is a m-elements set, also be denoted as $Xs^{(1)}(i, k)$, $i \in \{1, \ldots, m\}$.

$Xs^{(1)}(i, k) = Xs^{(1)}(i, k-1) + Xs^{(0)}(i, k)$, $\forall\ k \in \{2, 3, \ldots, n\}$, $Xs^{(1)}(i, 1) = Xs^{(0)}(i, 1)$.

For $Xs^{(1)}(i, k)$, $k \in \{1, 2, \ldots, n\}$ expresses that the set sequence is extendable in the time axis direction, and $i \in \{1, 2, \ldots, m\}$ expresses the position of elements in every set. So, sAGO is accumulating for corresponding elements of sets on the set sequence. That is,

$Xs^{(1)}(k) = \sum\limits_{q=1}^{k} Xs^{(0)}(q) \cdot$

Grey model based on set sequence (SGM) is defined as follows:

$Xs^{(0)}(k) + a\,Zs^{(1)}(k) = S$,
(1)

Where, $S = \{a1, a2, \ldots, a_m\}$. This model is denoted as SGM(1, 1+m). SGM(1,1+m) is a one-order grey model based on set sequence with 1+m parameters; a is development coefficient, and S is grey action which is constructed by m unknown parameters. $Zs^{(1)}$ is whiten sequence of $Xs^{(1)}$. $Zs^{(1)} = (Zs^{(1)}(2), \ldots, Zs^{(1)}(n))$, $Zs^{(1)}(k) \in Zs^{(1)}$, $Zs^{(1)}(k) = 0.5\ Xs^{(1)}(k) + 0.5\ Xs^{(1)}(k-1)$, that is, $Zs^{(1)}(i, k) = 0.5\ Xs^{(1)}(i, k) + 0.5\ Xs^{(1)}(i, k-1)$, $\forall\ k \in \{2, 3, \ldots, n\}$, $i \in \{1, \ldots, m\}$.

The whiten model of SGM(1,1+m) is:

$$\frac{dXs^{(1)}}{dt} + aXs^{(1)} = S,$$
(2)

$[a, a_1, a_2, \ldots, a_i, \ldots, a_m]^T$ is the parameter vector of SGM(1,1+m), denoted as Ps_m. The matrix equation of parameter vector Ps_m under least square criterion is shown as follows:

$$Ps_m = [a, a_1, a_2, \ldots, a_i, \ldots, a_m]^T = (Bs^T Bs)^{-1} Bs^T ys_n,$$
(3)

Where, $ys_n = \begin{bmatrix} Xs^{(0)}(1,2) \\ Xs^{(0)}(2,2) \\ \vdots \\ Xs^{(0)}(i,k) \\ \vdots \\ Xs^{(0)}(m,n) \end{bmatrix}$, $Bs = \begin{bmatrix} -Zs^{(1)}(1,2) & 1 & 0 & \cdots & 0 \\ -Zs^{(1)}(2,2) & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ -Zs^{(1)}(i,k) & & & & \\ \vdots & & & 1 & \\ -Zs^{(1)}(m,n) & & & \cdots & 1 \end{bmatrix}$,

$Zs^{(1)}(i, k)$, $Xs^{(0)}(i, k)$ are respectively the i-th value of $Zs^{(1)}(k)$, $Xs^{(0)}(k)$, ($i = 1, \ldots, m$, $k = 2, \ldots, n$). ys_n is a 1-column matrix, whose rows are $m \times (n-1)$. Bs is a matrix, whose rows are $m \times (n-1)$ and columns are $m+1$. The last m columns in matrix Bs are composed of 0 and 1, in which the first element in the first column is 1, and the posterior ones are 1 every other m; …; the k-th element in the k-th column is 1, and the posterior ones are 1 every other m; the m-th element in the m-th column is 1, and the posterior ones are 1 every other m; the others are all 0.

For the original set sequence $Xs^{(0)}$ and its sAGO generation sequence $Xs^{(1)}$, the response equation of whiten model SGM(1,1+m) can be obtained:

$$\hat{X}s^{(1)}(k+1) = (Xs^{(0)}(1) - \frac{S}{a})e^{-ak} + \frac{S}{a},$$
(4)

$$\hat{X}s^{(0)}(k+1) = \hat{X}s^{(1)}(k+1) - \hat{X}s^{(1)}(k).$$
(5)

According to the two response equations, the set sequence could be simulated and predicted. The elements of every set on the prediction sequence can be obtained by $Xs^{(0)}(1)$ and the values of corresponding position in set S.

There is a simulation example about a histogram sequence which is obtained by neighboring video frames (from 2548-th to 2552-th ones) of CCTV news in February 6, 2015. The histogram sequence is respectively simulated by SGM and GM(1,1) modeled 256 times as shown in Fig. 3. In Fig. 3, Values(k) is values of the k-th histogram in which there are 256 values denoted as i where $i=0\ldots255$. For ease of observation, let Values(k) = ($k-1$) × 5000 + Values(k), $k=1...5$. There are some abnormal simulations by GM(1,1) which made the simulation values greatly less than 0 and it is impossible for a histogram. There is also some simulation error by SGM, that is due to the shot switching between 2551 and 2552 frame from broadcast room to the exterior which makes the abrupt change of histograms. Two criteria [28, 29], *MSE* (the mean square error) which measures the overall performance of the model and *AME* (the absolute mean error) which measures the average error, are used to evaluate the simulation performance by SGM for histogram sequences of videos. As shown in Table 1, the *AME* and *MSE* values of 2547-2551 frames are much less than 2548-2552 frames because 2547-2551 frames are neighboring ones in the same shot. It is for this reason that SGM is used to detect the video cuts here.

$$AME = \frac{1}{mn}\sum_{k=1}^{n}\sum_{i=1}^{m}\left|Xs^{(0)}(k)(i) - \hat{Xs}^{(0)}(k)(i)\right|$$

(6)

$$MSE = \frac{1}{mn}\sum_{k=1}^{n}\sum_{i=1}^{m}(Xs^{(0)}(k)(i) - \hat{Xs}^{(0)}(k)(i))^2$$

(7)

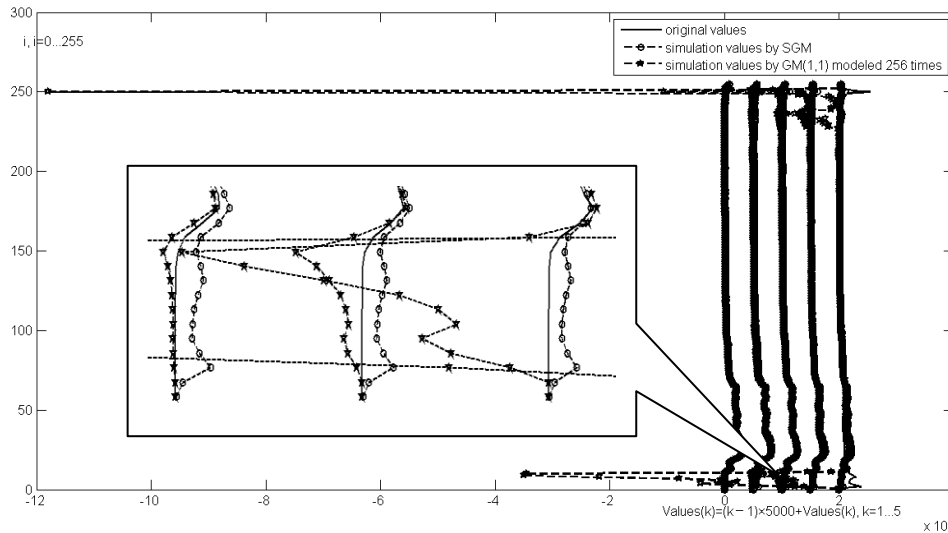Where, $Xs^{(0)}(k)(i)$ and $\hat{Xs}^{(0)}(k)(i)$ are respectively the original value and simulation value.

Figure 3. The Simulation Results of Histograms from 2548 to 2552 Frame by two Grey Models

Table 1. The Performance by SGM for Histograms from 2547 to 2552 Frame

frames	AME	MSE
2547-2551	15.4017	726.3065
2548-2552	163.2115	1.1301e+05

2.3.3 Grey model for sampled Histogram and Shot-Boundary judgment rules

The sampling interval is *h* when to sample the histograms. That means there are *m* sample values which constitute a set about static information of gray level (*m=256/h*, there are 256 gray levels for intensity images). The set of *m* samples is represented as $Xs^{(0)}(k)$ in $Xs^{(0)}$. $Xs^{(0)}$ is set sequence constructed by sampled histograms. In experiments of this paper, length of $Xs^{(0)}$ is 5, that is *k=1,…,n*, *n=5*, and *h=10*, *m=26*. Once a new video frame comes into the detection system, the system will do grey modeling with SGM to compute *AME* values, and the set sequence in the modeling process is constructed by consecutive 5 sampled histograms which is obtained by the current frame and the previous four ones. Then, another criteria, R_{AME} (regulative *AME*), will be computed at the same time as formula (8).

$$R_{AME} = \frac{100 * m * AME}{\sum_{i=1}^{m}\left|\max Xs^{(0)}(k) - \min Xs^{(0)}(k)\right|}$$

(8)

Thresholds T and $T1$ are used to make a judgment whether a frame is the beginning of another shot. T and $T1$ relate to the intensity and size of video frames. The steps of proposed Shot-Boundary detection method are shown as Fig. 4. In Fig. 4, $frame_i$ is the i-th frame of detected video, and i is frame number here.

There is an example about the analysis of AME and R_{AME} as shown in Fig. 5 and Fig. 6. Fig. 5 illustrates AME values of a video segment computed by the detection system, and the video segment comes from 3000 to 4000 frame of the first video sets (described in experiments part). Fig. 6 illustrates R_{AME} values corresponding to Fig. 5. In this example, T is 85, and $T1$ is 11.5. AME values of frames from 3540 to 3546 are greater than threshold value T, but they are not the cuts and only somebody thumbing through the book as shown in Fig. 7. If only using AME as a judgment condition, these frames would be detected as shot-boundary. However, they are weeded out after adding R_{AME} as the procedure in Fig. 4. Fig. 8 shows another frame sequence in which there are 5 cuts, and the detection results by the proposed method are shown in Fig. 9.

```
p←i, c←0, lab←0, maxA←0, j←1
While (frameᵢ)
    (1) obtain (I_HSI)ᵢ of Iᵢ, and histogram of (I_HSI)ᵢ
    (2) Xs⁽⁰⁾(i)←histogram of (I_HSI)ᵢ be sampled
    (3) construct set sequence Xs⁽⁰⁾=(Xs⁽⁰⁾(i-4), Xs⁽⁰⁾(i-3), Xs⁽⁰⁾(i-2), Xs⁽⁰⁾(i-1), Xs⁽⁰⁾(i))
    (4) AMEⱼ←grey modeling based on Xs⁽⁰⁾ and compute AME
    (5) (S_AME)ⱼ←compute S_AME,
    (6) if (AMEⱼ>T) then
            maxA←find the local maximum
          else
            if maxA !=0 then
                lab←1
            end if
        end if
    (7) if (lab==1) then
        (7-1) c←c+1
        (7-2) if (c ==3) then
            for q from j-3 to j step 1
                if ((S_AME)ⱼ<T1) then
                    frame_{p+q-4} is the beginning of another shot
                    break
                end if
            end for
            c←0, lab←0, maxA←0
        end if
    end if
    (8) j←j+1
End while
```

Figure 4. The Procedure of the Proposed Detection Method
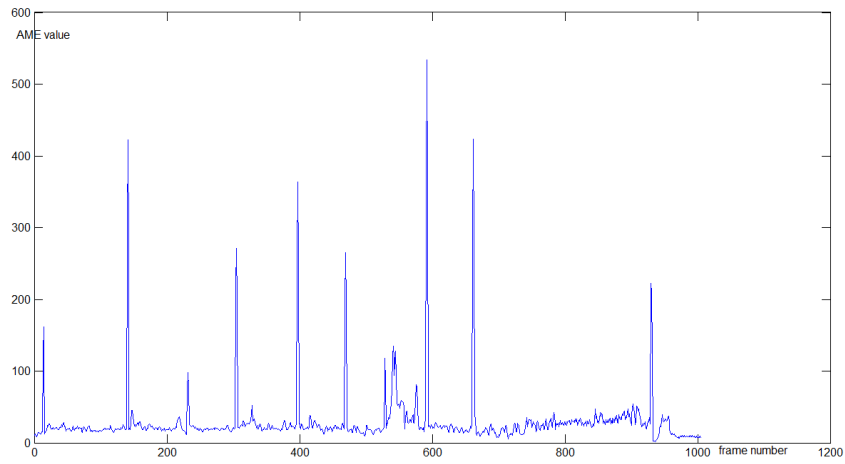
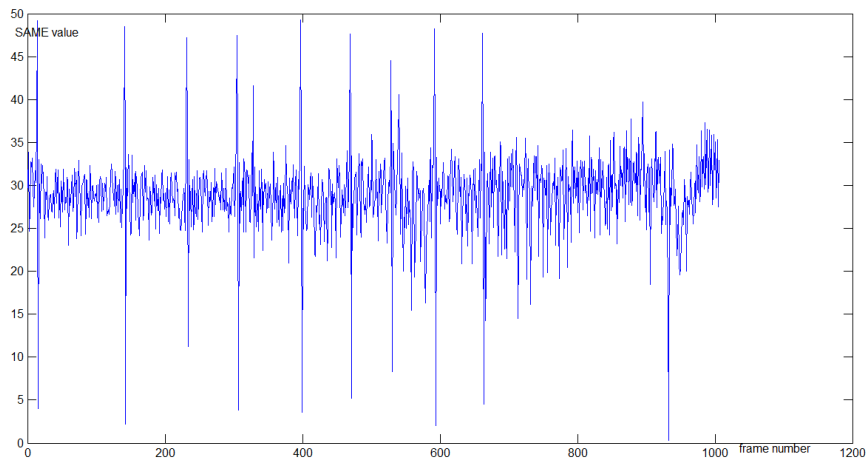**Figure 5. AME Values of CCTV news in February 6, 2015 from 3000 to 4000 Frame**

Figure 6. SAME Values Corresponding to Fig. 5

Figure 7. Video Frames about Somebody Thumbing through the Book

Figure 8. Another Frame Sequence in the First Data Sets

Figure 9. Five Pairs Cuts Detected in Figure 8

3. Experiment and Analysis

The performance of proposed method is tested and proved on several various videos. Experiments were carried out on videos from different categories containing movies, TV plays, documentary and news including entertainment shows, military maneuver, conferences and so on. The details of some test videos have been presented in Table 2, in which include complex video effects, object rapid motions, sharp illumination, strong camera flashes and dense smoke. The comparison method is histogram disparity (Dh) between two of nearest frames I_k and I_{k+1}, shown in formula (9). Threshold value Th is used to judge the cuts.

$$Dh = \frac{1}{256}\sum_{i=0}^{255}\left|I_{k+1}(q) - I_k(q)\right|$$

(9)

The following metrics is used to evaluate the performance of the proposed cut detection methods [2, 30]: (1) Precision to measure the quality expressed as the percentage of correct detections; (2) Recall to measure the quantity expressed as the percentage of detected true cuts; (3) F is a measure combination of Precision and Recall.

$$Precision = \frac{\text{Number of correct detections}(CD)}{\text{Number of all detections}(AD)} \times 100$$

(10)

$$Recall = \frac{\text{Number of correct detections}}{\text{Number of true cuts}} \times 100$$

(11)

$$F = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \times 100$$

(12)

Table 2. Description of the Experimental Video Sets

Video ID	Category	Filename	Frames	Frame size
N1	news	CCTV news 20150206.avi	2000-35000	360×480
S1	Sport news	CCTV sports news 20150810.avi	2000-16000	272×488
D1	documentary	Roof of the World 01.avi	2000-14000	624×1112
D2	documentary	A Bite of China Ⅱ 06. avi	2000-3000	272×488
TP1	TV plays	The.Big.Bang.Theory.S08E22.avi	3000-4000	576×1024
M1	movie	The Croods.avi	5000-7000	544×1280
C1	Cartoon	Kekexiaoai 01.avi	100-860	878×1280
C2	Commercial	Evian.avi	500-1500	544×968

Table 3. The Detection Results and Thresholds of the Two Methods

VideoID	Ture cuts	DH method			Proposed method		
		CD	*AD*	*Th*	*CD*	*AD*	*(T, T1)*
N1	273	256	318	200	244	268	(85, 11.5)
S1	81	77	92	230	79	79	(85, 11.5)
D1	189	179	200	1000	183	185	(300,11.5)
D2	38	27	36	300	33	33	(85,16)
TP1	18	18	18	500	18	18	(85,11.5)
M1	14	15	14	1000	14	14	(600,11.5)
C1	13	13	22	2000	10	11	(1000,11.5)
C2	34	34	34	800	34	34	(300,11.5)

Table 4. The Performance Comparison of the Two Methods

VideoID	DH method			Proposed method		
	Precision	*Recall*	*F*	*Precision*	*Recall*	*F*
N1	80.5	93.77	86.63	91.04	89.38	90.2
S1	83.7	95.06	89.02	100	97.53	98.75
D1	89.5	94.71	92.03	98.92	96.83	97.86
D2	75	71.05	72.97	100	86.84	92.96
TP1	100	100	100	100	100	100
M1	93.3	100	96.53	100	100	100
C1	59.1	100	75.05	90.91	76.9	83.32
C2	100	100	100	100	100	100

Detection results, thresholds and performance comparison of these video sets in Table 2 are demonstrated in Table 3 and Table 4 with the proposed method and method of Histogram disparity. It was observed that the performance of proposed method is superior to method of Histogram disparity. And the result is analyzed as follows:

In video N1, there are some complex video scenes, for example, scenes with video effects, hard light and flash lamp, shot switching under similar scenes, moving light in dark, rapid shaking and moving of camera, object fast appearing or disappearing in video scenes. As shown in Table 3 and Table 4, a satisfactory detection result is still obtained with the proposed method. We made another experiment for video N1 with the proposed method under the same condition. The only difference is to get rid of the lightest and darkest gray level when sampled histograms, that is, to take sample of gray level from 10 to 246. We obtained better detection results than ones in Table 3 and Table 4 (the number of correct detections is 250, the number of all detections is 254, then, *Precision*=98.4, *Recall*=91.58, *F*=94.87). Fig. 10 shows the frame pairs detected by method of Histogram disparity from 32240 to 32324 which is affected badly by smoke and light of cannonball. There only have 4 cuts in this frame segment which correctly detected by proposed method as shown in Fig. 11.
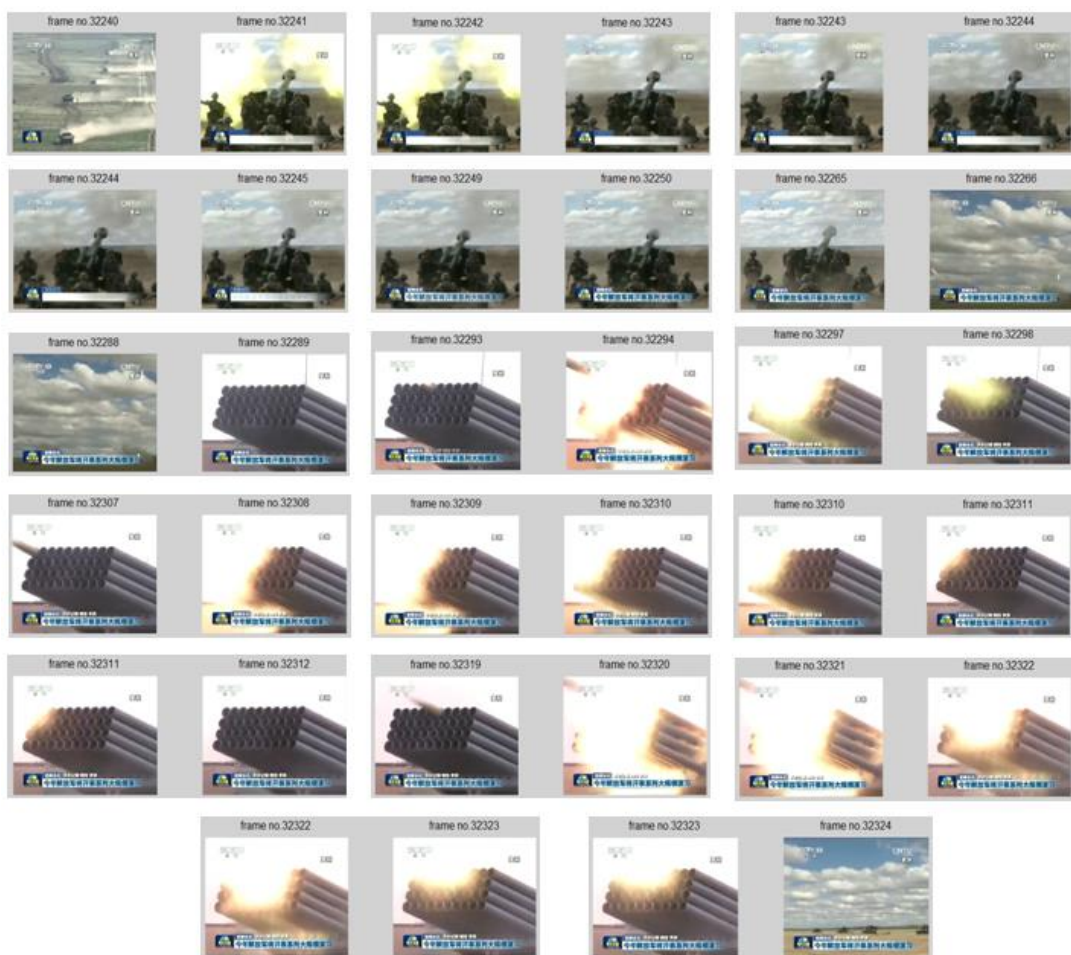

Figure 10. The Frame Pairs Detected by DH Method from 32240 to 32324 in Video N1

Figure 11. The Correct Detection by Proposed Method N1 from 32240 to 32324 in Video N1

In video S1, there is a missed detection by proposed method as shown in Fig. 12, but it will be detected correctly when $T1=15$. In video C1, there are two missed detection and error detection because of four gradual transition (GT) between normal frames and completely dark frame (this error detection could be counted as a correct detection, and it is just a rearward frame pair in the gradual transition). However, the gradual transition shot could be correctly detected by proposed method. For example, there is an effect processing in an inner shot from 29009 to 29131 and a gradual transition to 29132 in video N1, which is correctly detected by proposed method as shown in Fig. 13.

Figure 12. False Detection by Proposed Method for Video S1

Figure 13. Correct GT Detection by Proposed Method for Video N1

In addition, the research work makes a comparison and analysis for the processing time of these two methods. Grey model is applied to simulate and predict limited time sequence. Compared to method of Histogram disparity (512 (2×256) values computed), several sampled histograms about 5 frames are just modeled in the proposed method (only 130 (5×26) values computed), it makes the time performance of the whole method acceptable. Meanwhile, the selection of threshold values is simpler than method of Histogram disparity. The detection performance by method of Histogram disparity is be serious affected by different threshold Th, but different thresholds (T, $T1$) have not had a great impact on the detection performance by the proposed method. According to the intensity and size of video frames, Thresholds T and $T1$ can be obtained easily.

4. Conclusion

This paper proposes a video Shot-Boundary Detection method combining grey model based on set sequence with color histogram. Firstly, the traditional grey model is introduced and analyzed. When modeling for multi-stream sequence, there would be some abnormal simulation even failure simulation by GM(1,1). To analyze and process video stream, grey model based on set sequence (SGM) is proposed in this paper. Then, the reason SGM applied in Shot-Boundary Detection is illustrated carefully. The proposed Shot-Boundary Detection method including the following steps: (1) When a new video frame coming into the detection system, the image is transformed into intensity image under HSI color model; (2) Histogram of the video frame can be obtained. Then, taking samples for the histogram and constructing set sequence with proximate sampled histograms is implemented; (3) SGM is introduced to simulate the sampled histogram sequence in this detection method. Moreover, absolute mean error (AME) and regulative AME (R_{AME}) with thresholds are applied to make a detection judgment. Finally, some experiments are use to illuminate that the proposed method is an effective Shot-Boundary Detection method and is significantly superior to method of Histogram disparity. Generally, the proposed method is able to detect cuts in various videos including videos with some complex scenes, for instance, scenes with hard light, flash lamp, shot switching under similar scenes, moving light in dark, rapid shaking and moving of camera, object fast appearing or disappearing in video scenes. In future work, we will go on research and optimize video Shot-Boundary Detection method based on grey modeling.

Acknowlegdements

References

[1]   S. Das, S. Sural, A. K. Majumdar, "Detection of hard cuts and gradual transitions from video using fuzzy logic", International Journal of Artificial Intelligence and Soft Computing, Vol.1, No.1, **(2008)**, PP. 77-98.

[2]   R. Dadashi, H. R. Kanan, "AVCD-FRA: A novel solution to automatic video cut detection using fuzzy-rule-based approach", COMPUTER VISION AND IMAGE UNDERSTANDING, Vol.117, No.7, **(2013)**, pp. 807-817.

[3]   H. Lee, J. Yu, Y. Im, "A unified scheme of shot boundary detection and anchor shot detection in news video story parsing", MULTIMEDIA TOOLS AND APPLICATIONS, Vol. 51, No.3, **(2011)**, pp. 1127-1145.

[4]   U. Sakarya, Z. Telatar, "Video scene detection using graph-based representations", SIGNAL PROCESSING-IMAGE COMMUNICATION, Vol. 25, No.10, **(2010)**, pp. 774-783.

[5]   M. Birinci, S. Kiranyaz, "A perceptual scheme for fully automatic video shot boundary detection", SIGNAL PROCESSING-IMAGE COMMUNICATION, Vol. 29, No. 3, **(2014)**, pp. 410-423.

[6]   X. H. Jiang, T. F. Sun, J. Liu, "An adaptive video shot segmentation scheme based on dual-detection model", Neurocomputing, Vol.116, **(2013)**, pp. 102-111.

[7]   P. Ji, L. J. Cao, X. G. Zhang, "News videos anchor person detection by shot clustering", Neurocomputing, Vol.123, **(2014)**, pp. 86-99.

[8] A. F. Smeaton, P. Over, A. R. Doherty, "Video shot boundary detection: Seven years of TRECVid activity", COMPUTER VISION AND IMAGE UNDERSTANDING, Vol.114, No.4, (2010), pp. 411-418.

[9] O. Kucuktunc, U. Gudukbay, O. Ulusoy, "Fuzzy color histogram-based video segmentation", COMPUTER VISION AND IMAGE UNDERSTANDING, Vol.114, No.1, (2010), pp. 125-134.

[10] L. G. G. Priya, S. Domnic, "Edge Strength Extraction using Orthogonal Vectors for Shot Boundary Detection", Proceedings of the 2nd International Conference on Communication, Computing and Security, Rourkela, INDIA, (2012).

[11] C. H. Lin, M. D. Hsiao, L. J. Fu, "Analysis of Shot Boundary Based on Color and Texture Features of Frame", Proceedings of the 5th International Conference on Mechatronics, Kuala Lumpur, MALAYSIA, (2013).

[12] K. K. Warhade, S. N. Merchant, U. B.Desai, "Shot boundary detection in the presence of illumination and motion", SIGNAL IMAGE AND VIDEO PROCESSING, Vol. 7, No.3, (2013), pp. 581-592.

[13] E. Mendi, C.Bayrak, "Shot boundary detection and key-frame extraction from neurosurgical video sequences", IMAGING SCIENCE JOURNAL, Vol. 60, No. 2, (2012), pp. 90-96.

[14] J. H. Ren, J. M. Jiang, J. Chen, "Shot Boundary Detection in MPEG Videos using Local and Global Indicators", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, Vol.19, No.8, pp. (2009), 1234-1238.

[15] Y. N. Li, Z. M. Lu, X. M. Niu, "Fast video shot boundary detection framework employing pre-processing techniques", IET Image Processing, Vol.3 No.3, (2009), pp.121-134.

[16] P. P. Mohanta, S. K. Saha, B. Chanda, "A Model-Based Shot Boundary Detection Technique Using Frame Transition Parameters", IEEE TRANSACTIONS ON MULTIMEDIA, Vol.14, No.1, (2012), pp. 223-233.

[17] Z. M. Lu, Y. Shi, "Fast Video Shot Boundary Detection Based on SVD and Pattern Matching", IEEE TRANSACTIONS ON IMAGE PROCESSING, Vol.22, No.12, (2013), pp. 5136-5145.

[18] Q. F. Fu, Y. C. Zhang, L. Y. Xu, "A Method of Shot-Boundary Detection based on HSV Space", Proceedings of the 9th International Conference on Computational Intelligence and Security, PEOPLES R CHINA, (2013), pp. 219-223.

[19] M. Guder, N. K. Cicekli, "Dichotomic Decision Cascading for Video Shot Boundary Detection", Proceedings of the 15th IEEE International Symposium on Multimedia, Anaheim CA, (2013), pp. 227-230.

[20] L. G. G. Priya, S. Domnic, "Walsh-Hadamard Transform Kernel-Based Feature Vector for Shot Boundary Detection", IEEE TRANSACTIONS ON IMAGE PROCESSING, Vol. 23, No.12, (2014), pp. 5187-5197.

[21] G. Y. Gao, H. D. Ma, "To accelerate shot boundary detection by reducing detection region and scope", MULTIMEDIA TOOLS AND APPLICATIONS, Vol. 71, No. 3, (2014), pp. 1749-1770.

[22] M. Vila, A. Bardera, Q. Xu, "Tsallis entropy-based information measures for shot boundary detection and keyframe selection", SIGNAL IMAGE AND VIDEO PROCESSING, Vol.7, No.3, (2013), pp. 507-520.

[23] Z. L. Wu, P. Xu, "A Fast Gradual Shot Boundary Detection Method Based on SURF", Proceedings of the 8th International Conference on Intelligent Systems and Knowledge Engineering, Shenzhen, PEOPLES R CHINA, (2013).

[24] X. H. Jiang, T. F. Sun, J. Liu, "An Video Shot Segmentation Scheme Based on Adaptive Binary Searching and SIFT", Proceedings of the 7th International Conference on Intelligent Computing, Zhengzhou, PEOPLES R CHINA, (2011).

[25] J. Benois-Pineau, A. Khrennikov, "Significance Delta Reasoning with p-Adic Neural Networks: Application to Shot Change Detection in Video", COMPUTER JOURNAL, Vol.53, No. 4, (2010), pp. 417-431.

[26] J. L. Deng, "Control problems of grey systems", Systems and Controls Letters, No.5, (1982), pp. 288-294.

[27] J. L. Deng, "The Base of Grey Theory", Press of Huazhong University of Science and Technology, Wuhan, (2002) (In Chinese).

[28] Y. H. Lin, P. C. Lee, T. P. Chang, "Adaptive and high-precision grey forecasting model", Expert Systems with Applications, Vol.36, (2009), pp.9658-9662.

[29] Z. S. He, X. Liu, Y. Chen, "Secondary-diagonal mean transformation Partial Grey Model based on matrix series", SIMULATION MODELLING PRACTICE AND THEORY, Vol.26, **(2012)**, pp. 168-184.

[30] V. Chasanis, A. Likas, N. Galatsanos, "Simultaneous detection of abrupt cuts and dissolves in videos using support vector machines", Pattern Recognition Letters, Vol.30, No. 1, **(2009)**, pp. 55-65.

Authors

Xin Liu, She is an associate professor at the Chongqing University of posts and telecommunications. She received her MS and PhD degrees in Computer Science from Chongqing University in 2005 and 2012, respectively. Her research domain includes video processing, image processing, grey system.

Jin Dai, He is an associate professor at the Chongqing University of posts and telecommunications. He received his PhD degrees in Computer Science from Chongqing University in 2011. His research domain includes intelligent information processing and natural language processing techniques.