

Shot Boundary Detection Algorithm Based on HSV Histogram and HOG Feature

Hong Shao^{1, a*}, Yang Qu^{2, b} and Wencheng Cui^{3, c}

¹ School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110023, China

² School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110023, China

³ School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110023, China

^a276446462@qq.com, ^bquyang0919@163.com, ^c13654067256@qq.com

Keywords: HSV color histogram, HOG feature, shot boundary detection.

Abstract. Spatial information can not be contained in the statistical feature of hue-saturation-value(HSV) color histogram, moreover, which is not robust for the videos with the similarity of background information and affected by illumination. Therefore, a new video shot segment method based on HSV histogram and Histogram of Gradient (HOG) feature is proposed. Firstly, HSV color histogram is used to detect the difference between two adjacent frames which is then compared with the adaptive threshold. After the initial detection of the shot boundary is finished, HOG feature of the video is adopted to make further shot boundary detection, which eliminates the wrong shot boundaries and adds missed shot boundaries. Experimental results indicate that the proposed method achieves higher accuracy and recall.

Introduction

As the basis of video processing and content analysis, shot boundary detection plays an important role in video information management. Video shot transform is generally divided into abrupt and gradual shot transform. Abrupt shot transform means that the previous shot directly connects the next shot without any editor. Gradual shot transform refers to a shot transform method through series of editing means, where next shot gradually replaces previous shot. Abrupt shot detection of video is discussed here. Abrupt shot detection method generally extract the low-level features of the image, such as pixels, sub block statistical characteristics, color histogram, edge, etc. to measure the frame difference. Pixel difference and color histogram algorithm are the two typical methods. In recent years, some new methods are proposed to solve this problem. Vasileios C. et al. and Zhang C. et al. propose a method of dividing the shot by the training support vector machine (SVM) classifier[1,2]. Zhang et al. present a method based on the weighted color histogram and adopt an improved sliding window method for the second test which eliminates the impact of the video flash phenomenon in the detection of shot boundary [3]. Zhe-Ming Lu et al. extract color histograms in the HSV space to form a frame feature matrix and singular value decomposition (SVD) is then performed on the frame feature matrices of all candidate segments to reduce the feature dimension [4]. Hansung et al. adopt singular value to decompose the color histogram feature vector and discard the small singular value and then use the clustering method to find video shot

boundaries [5]. Jiang et al. carry out the non-uniform block according to the different underlying features of the video frames and then the shot boundary is detected by using the feature matching of speed-up robust features (SURF) [6]. However, the performance of the algorithm may be quickly decreased in the methods above when there are camera movement, object movement and bright sudden change in the scene. Shot boundary detection algorithm based on HSV color histogram and HOG feature is proposed in this paper.

Shot Boundary Detection Algorithm Based on HSV Histogram and HOG Feature

Due to the color histogram's simplicity in calculation, clearly separating capacity and good robustness in the video image translation, rotation and camera pulling far or closer to the movement, color histogram of video is adopted to finish preliminary shot boundary detection. HSV model is much closer to people's perception of color than RGB color model, so HSV model is used in color histogram. Because HOG feature can keep good invariance in geometric and optical deformation, moreover, it also can reflect certain edge characteristic and position information. Therefore, after HSV color histogram is used for preliminary detection of video shot, HOG feature is adopted for secondary detection, which improves the algorithm's precision and recall.

Flow chart of the algorithm is shown in Fig.1.

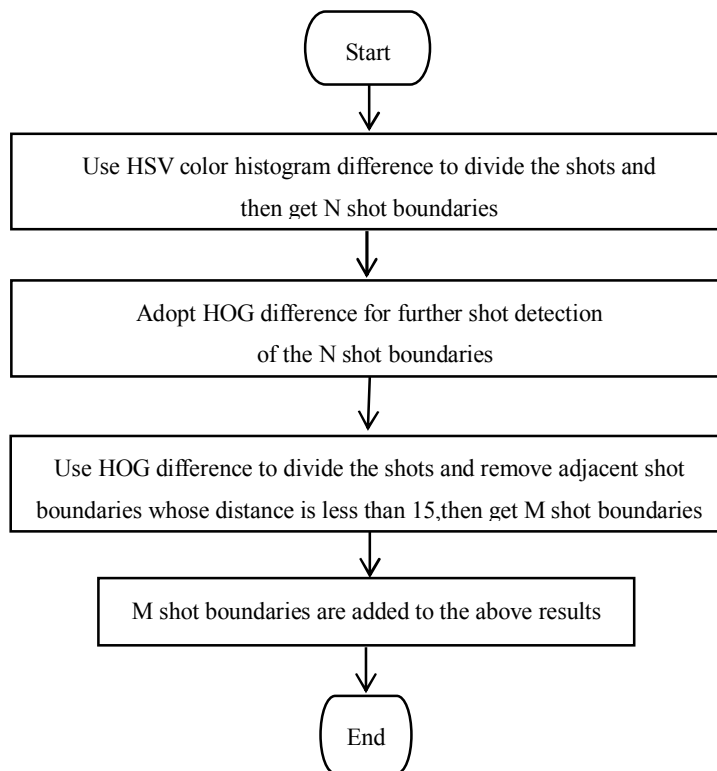


Fig.1 Flow chart of algorithm proposed

Preliminary Detection Based on HSV Color Histogram

HSV Model

The HSV model constitutes a uniform color space which adopts a linear scale. HSV space coordinate system can be represented by six pyramid as shown in Fig.2. Wherein, H is on behalf of hue, ranging from 0° to 360° , S is on behalf of saturation, ranging from 0.0 to 1.0, V is on behalf of value, ranging from 0 to 255. Moreover, 0° corresponds to the red, 120° corresponds to the green, and 240° corresponds to the blue. At the apex of six pyramid, $V = 0$, H and S have no definition for the black. At the top surface center of six pyramid, $S = 0$, $V = 1$, H, on behalf of white, is undefined.

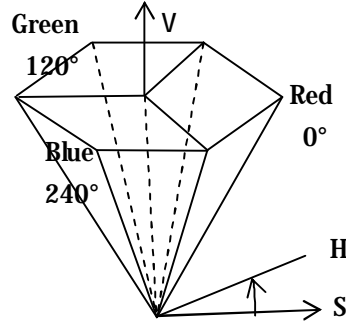


Fig. 2 HSV space coordinate system

Preliminary Detection Algorithm

In order to reduce the computational complexity, HSV color space is firstly quantized in general, and then pixel number of per quantization unit is counted which forms color histogram, moreover, compare video frame difference with adaptive threshold, where video frame difference is represented by the distance of HSV color histogram difference. If the video frame difference is greater than the threshold value, there is a shot boundary between two adjacent frames among which the shot boundary is set as the starting frame for the next shot.

Specific steps are as follows:

Step1: The video is extracted into a series of frame images based on RGB mode, and then frame images are converted into HSV mode according to the following 3 formulas.

$$H = \begin{cases} \arccos \frac{(R-G)+(R-B)}{2\sqrt{(R-B)^2+(R-B)(G-B)}} & B \leq G \\ 2\pi - \arccos \frac{(R-G)+(R-B)}{2\sqrt{(R-B)^2+(R-B)(G-B)}} & B > G \end{cases} \quad (1)$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (2)$$

$$V = \frac{\max(R,G,B)}{255} \quad (3)$$

Step2: According to the human visual discrimination, the sensitivity of hue is higher than the brightness and saturation. The H is quantized to 16 levels, and S is quantized to 8 levels, and V is quantized to 8 levels.

Step3: Pixel number of per quantization unit is counted which consists of color histogram with 1024 bin's one-dimensional vector.

Step4: Because Chi-square measure has high accuracy in matching, it is used to calculate the color histogram of adjacent two frame difference $d_j(H_j, H_{j+1})$ where j represents frame number, H_j and H_{j+1} represent two adjacent frames.

$$d_j(H_j, H_{j+1}) = \sum_{i=1}^{1024} \frac{(H_j(i) - H_{j+1}(i))^2}{H_j(i) + H_{j+1}(i)} \quad (4)$$

Step5: Calculate the whole video's adjacent frame difference's mean m and variance σ , then adaptive threshold T is calculated.

$$m = \frac{\sum_{i=2}^{N-1} d_i(H_i, H_{i+1})}{N-1} \quad (5)$$

$$\sigma = \sqrt{\sum_{i=1}^{N-1} \frac{(d_i(H_i, H_{i+1}) - m)^2}{N-1}} \quad (6)$$

$$T = m + S \times \sigma \quad (7)$$

Where N represents the number of frame and through a lot of experiments S is assumed to be 7.

Step6: The inter-frame difference is successively compared with the threshold value T. If it is greater than the threshold, shot boundary serial number will be stored in the text.

Although the shot boundary detection algorithm based on the HSV color histogram is robust to the local motion of the object to a certain extent, it is not suitable for very fast movement of the object and also not robust to brightness change and images with similar background. Therefore, HOG feature is used for secondary testing.

Secondary Detection Based on HOG

HOG Feature

The appearance and shape of the local object can be well described by the gradient or the edge of the direction density distribution. Meanwhile, in terms of image geometry and optical distortion, HOG feature can keep good invariance. Therefore, HOG feature is calculated for further shot detection by using the method in [7]. The specific steps of HOG feature extraction are as follows:

(1) Gray color image and then use Gamma to check image which reduces the impact of local image shadows and lighting changes while suppresses noise interference.

(2) Compute the gradient size and gradient direction of each pixel to obtain edge information.

(3) Divide the image into several cells of the same size and the size of the cell is set to 16×16.

(4) Divide the gradient direction into 9 bin and count the gradient direction histogram of the cell, and then HOG feature of the cell is obtained.

(5) Each 2×2 cell forms a block and put features of all blocks in the image together which constitutes final HOG feature.

Secondary Detection Algorithm

Firstly, assume that shot boundary set is $\{S_1, S_2, \dots, S_N\}$ after preliminary detection, calculate HOG feature of each frame in the video and then calculate the two adjacent frames' HOG feature difference for feature matching. The HOG feature difference of two adjacent frames is shown as follows where the HOG feature difference between i frame and i-1 frame is $Fh(i)$, N represents the number of blocks and H_i represents HOG feature of i-frame.

$$Fh(i) = \sum_{n=1}^N \sum_{m=1}^{86} |H_i - H_{i-1}| \quad (8)$$

The HOG difference of obtained shot boundary from Section 2 and previous frame is $Fh(S_i)$

which is respectively compared with the set threshold T' which is set as 5 through experiments. If

$Fh(S_i)$ is much smaller than T' , brightness transformation may have occurred in S_i shot boundary

where the border is wrong division, which need to be excluded from the results after the preliminary shot division. Then boundary $\{B_1, B_2, \dots, B_M\}$ which are suspected shots will be obtained after

comparing $Fh(i)$ with adaptive threshold. In order to fully express video content, usually a shot will

persist for some time, we assume that a video shot lasts for a minimum of 0.5s and according to the National Television Standards Committee (NTSC) standard video frame rate is 29.97 per second, so a video shot lasts at least 15 frames [8]. In the results of the shot boundaries obtained above, if the

separation distance between two adjacent shot boundaries is less than 15, the shot boundaries, which may contain object moving quickly and flash effects, are removed and the resulting shot boundary set $\{B_1, B_2, \dots, B_M\}$ is added to resulting shot boundaries in Section 2. Moreover, it is also used for HSV color histogram detection.

Experiment Results and Analysis

Due to the diversity and complexity of video and other reasons, the shot detection has not been established a set of internationally recognized evaluation standard system. Therefore, two standards are used as shot segmentation detection evaluation criteria and their definitions are as follows:

$$\text{Precision} = \frac{\text{Detects}}{\text{Detects} + \text{FA}} \quad (9)$$

$$\text{Recall} = \frac{\text{Detects}}{\text{Detects} + \text{MD}} \quad (10)$$

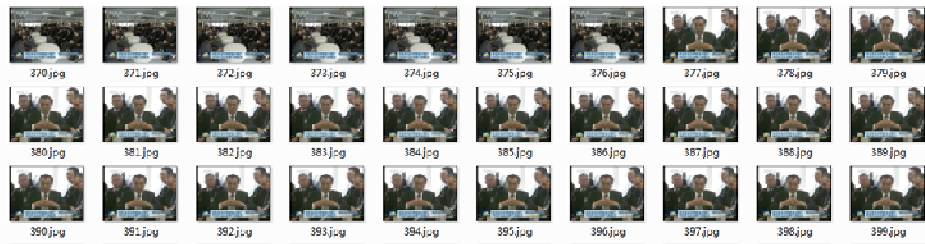
Where Detects represents the number of correctly detected shots, MD represents the number of missing shots and FA represents the number of false detection shots. Missing detection refers to some actual shot boundaries in the process of shot detection which have been judged to be the frames of the same shot. False detection means misjudging the frame which belongs to the same shot as a shot boundary and dividing a shot into two or more shots.

The Open Video Project and news broadcast clips in May 22, 2015 have been selected as experimental data and the experiment is done on the platform of Microsoft Visual Studio 2010. The experimental results are shown in Table 1.

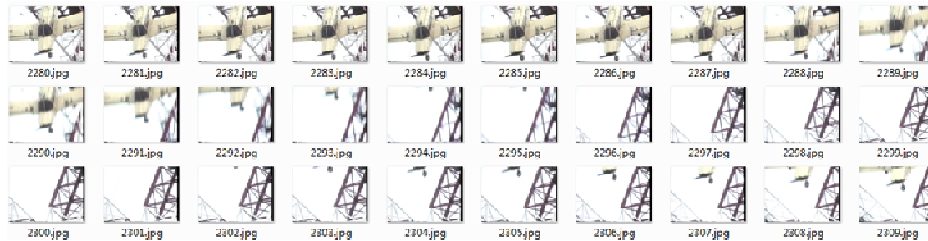
Table 1 Testing results of the algorithm

Video	frames	shots	Proposed algorithm		HSV histogram method	
			precision	recall	precision	recall
News1	4048	31	1.000	0.968	0.533	0.484
News2	4173	43	1.000	0.953	1.000	0.767
News3	4198	45	1.000	0.978	1.000	0.733
Openvideo1	4000	16	1.000	0.938	0.923	0.750
Openvideo2	4273	13	1.000	1.000	1.000	1.000
Openvideo3	4201	13	1.000	1.000	1.000	1.000

Six video clips are selected to verify the algorithm adopted where there are 161 actual shots, 5 undetected shots. Overall, the precision is 100% and the recall is 96.9% where the average recall of the Open Video Project's video clips is up to 97.7%. At the same time, the precision is 75.2% and the recall is 89.0% when HSV color histogram method is used. Compared to using only HSV color histogram to detect shot boundary, precision and recall of proposed algorithm have been significantly improved.



(a) News video sequence



(b) open project video sequence

Fig. 3 Video sequences for testing

The 377th frame shown in Fig.3(a) shows a shot boundary which can be detected by proposed method. Fig.3(b) shows fast moving video sequence and the proposed algorithm can effectively avoid the false detection caused by fast moving objects like Fig.3(b).

Experimental results show that the effect of the shot detection algorithm adopted is better able to correctly detect shot boundary, and basically check the whole shot boundary, and have some robustness to the object of the fast motion and illumination changes.

Conclusion

A shot boundary detection method based on HSV color histogram and HOG feature is proposed. The experimental results show that the proposed method can achieve high precision and recall and reduce the effects of the image brightness changes and the movement. The next work is to study how to improve the speed of the algorithm, and detect the effectiveness of the algorithm in other kinds of videos such as film, cartoon, advertising, etc.

Acknowledgments

This work was sponsored by Natural Science Foundation of Liaoning Province (No.201202162) and Program for Liaoning Excellent Talents in University (No.LJQ2013013).

References

- [1] Vasileios Chasanis, Aristidis Likas, Nikolas P.: Galatsanos. Simultaneous detection of abrupt cuts and dissolves in videos using vector machines. *Pattern Recognition Letters*, 30 (2009) 55-65.
- [2] Zhang C, Wang W Q. A robust and efficient shot boundary detection approach based on fisher criterion. In: 20th ACM International Conference on Multimedia, ACM Press, USA, 2012, pp.701-704.
- [3] ZHANG Jing, WANG Xuejun, JIANG Enyuan. News Video Shot Detection Based on Weighted Color. *Journal of Jilin University (Information Science Edition)*, 33 (2015) 39-44. (In Chinese)
- [4] Zhe-Ming Lu, Yong Shi. Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 22 (2013) 5136 -5154.

- [5]HANSUNG LEE, JAEHAK YU, YOUNGHEE IM, et al. A Unified Scheme of Shot Boundary Detection and Anchor Shot Detection in News Video Story Parsing. *Multimedia Tools and Applications*, 51(2011)1127-1145.
- [6]JIANG Ming, HUANG Jingcheng, WANG Xing, et al. Shot Boundary Detection Method for News Video. *Journal of Computers*, 8 (2013) 3034-3038.
- [7]Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Press, USA, 2005, pp.886–893.
- [8]CHAO Juan, SUN Tanfeng, JIANG Xinghao. A Video Shot Segmentation Algorithm Based on Dual-detection Model. *Journal of Shanghai Jiaotong University*, 45 (2010) 1542-1546. (In Chinese)