Shot Boundary Detection Method for News Video

Ming Jiang[1,2] , Jingcheng Huang[1,2] , Xingqi Wang[1,2] , Jingfan Tang[1,2] , Chunming Wu[3]

(1. Institute of Software and Intelligent Technology, Hangzhou Dianzi University, Hangzhou, 310018, China)
(2. Zhejiang Provincial Engineering Center on Media Data Cloud Processing and Analysis, Hangzhou Dianzi University, Hangzhou, 310018, China)
(3. College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China)

Email: jmzju@163.com, zjcnxghjc@126.com, xqiwang@163.com, tangjf@hdu.edu.cn, wuchunming@zju.edu.cn

***Abstract*—It is very important to detect shot boundary accurately and quickly in a large number of news video data. Therefore, we proposed a new method with dual-detection model. The method is divided into two stages, i.e. pre-detection and re-detection. In the pre-detection stage, the uneven blocked differences based on the feature of human vision are presented and used in adaptive binary search to detect shot boundaries. In the re-detection round, Speeded Up Robust Features (SURF) method is applied to exclude false detections. The experimental results show that this method can detect abrupt boundaries of news video quickly and accurately.**

***Index Terms*—Shot segmentation, Uneven blocking, Sliding window, Adaptive binary search, SURF.**

I. Introduction

With the development of multimedia technology and computer network, digital video is produced and used widely. People come into contact with video data at an unprecedented growth rate, so how to retrieval the required information conveniently, fast, and accurately in the face of a large number of video information has been the focus of attention. The artificial video classification method is too time-consuming and susceptible to the impact of human factors. Commonly, the first step for most content-based video analysis techniques is to segment a video into some shots. Shot segmentation is the basis of video analysis etc. Upon shot segmentation, we can further develop the key frame extraction and content-based video retrieval technology.

News video is a video that a lot of people are concerned about. For example, China News Broadcast. We want to know the main news in the world recently according to the news video, which are closely related to us. But we just want to know the main content of news, rather than spend a lot of time to watch every detail of news. That needs to extract key frame of news video, and we should segment video in order to extract the key frame. From this point of view, we can see the significance of video segmentation.

The transition from one shot to the next may be of various types: broadly categorized as abrupt change and gradual change. Abrupt change, also known as cut, denotes instantaneous transition from one shot to another. Gradual change is usually obtained by incorporating photographic effect through editing, and it can be classified as fade-out, fade-in, dissolve, wipe, etc.

Due to the promptness of news video, abrupt change accounts for 90% or more, and the gradual shot generally appears at the beginning and end of the news video. Therefore, this paper only states the abrupt change of news video.

News video has relatively fixed structure: beginning with one anchorperson or two, that broadcast news headlines, and each piece of news has reporters to report, anchorperson will also comment for particularly important news. In every news shot, the news caption reflecting the main content of this shot will appear at a specific location, like the bottom of the shot.

Scholars have done a lot of research on shot segmentation in the field of video analysis, mainly including the following methods: pixel differences based method, histogram comparisons based method [5], edge differences based method, etc. These methods have limitations. For instance, pixel differences based method is simple and easy to implement, but it is very sensitive to noise and movement of objects. Most of the algorithms detect shot boundary only once, and is unable to get a result with high accuracy. In this paper, we propose a video shot segmentation method with dual-detection model. Adaptive binary search and Speeded Up Robust Features (SURF) are introduced into this method. Experimental results show that this algorithm has higher accuracy and efficiency, especially for news video data.

II. Our Method

We propose a new method with dual-detection model. First, the input video of each video frame is converted into HSV color space, and then use uneven blocking as a determination factor, combined with adaptive binary search to detect shot boundaries. Second, Speeded Up Robust Features (SURF) method is applied to exclude

doi:10.4304/jcp.8.12.3034-3038

false detections, which can improve the accuracy of the algorithm.

A. HSV Color Space

In image processing, it is very important to choose appropriate color model. HSV color space is more closely to human perspective feeling and independent on the display device, therefore we choose HSV color space to extract boundary feature in this paper. The value (V) only reflects luminance information, unrelated to the color information of the image; hue (H) and saturation (S) indicates the subjective feeling of the person on the color, unrelated to value (V). If we use one component only, we will surely lose a lot of useful information, which is not desirable. We combine various components to extract the information available in the image.

In this paper, we extracted hue, saturation, and value information from three channels of H, S, V, and then through being combined with each of the frame difference information, frame difference of two images are obtained.

B. Uneven Blocking

According to the human visual feature and news video feature, we propose the concept of uneven blocking. Human vision has the auto-focusing characteristics that a person will focus on the central area, foreground and moving objects of a picture. Humans are concerned generally middle portion of the video frame and the video information surrounding the central portion is contained relatively less.

According to the characteristics discussed above, we divide video frame into three sub-blocks. Figure 1 shows the uneven blocking method where Group 1 is the focus of human visual system that contains more information, while Group 3 in the corners will be paid less attention to. The importance of Group 2 is between Group 1 and Group 3. Thus the importance of each group can be judged by different weighting coefficients. Group 1 is the biggest, followed by Group 2, Group 3 is the minimum. The coordinates in the top left corner of Group 1 is (W*0.15, H*0.15), the coordinates in the lower right corner is (W*0.85, H*0.85). W and H are the width and height of frame.

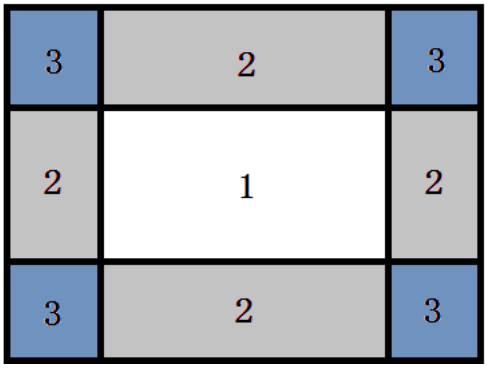

Figure 1. Uneven blocking.

C. Adaptive Binary Search

Divide a video window into left sub window and right sub window, where the middle frame belongs to both sub-windows. Calculate and compare the difference between two sub-windows. If the difference is too large, the sub-window whose discontinuity is larger may have an abrupt transition. This binary search process is continued until the window size is 2. Otherwise there exists no boundary in this window, skipping the window and sliding to the next one. Detecting video whose length of frame is n, the complexity of binary search computation is $O(\log_2^n)$, while the complexity of frame-by-frame comparison algorithm is $O(n)$. If n is large, the search time efficiency has improved significantly.

D. The Re-detection Based on SURF Feature Matching

SURF algorithm have better results in the field of image matching, which is widely used for object recognition, motion tracking. SURF algorithm has better robustness and real-time compared with SIFT algorithm.

Figure 2 shows SURF algorithm. The shot segmentation principle is as follows: Extracting the SURF feature points of two frames, if they are very different, then the two frames are very different and there may be a shot boundary. Otherwise match the points of two frames. If the matching ratio is low, then it confirms the shot boundary between the two frames, otherwise it labels the result in previous period as false alarm.

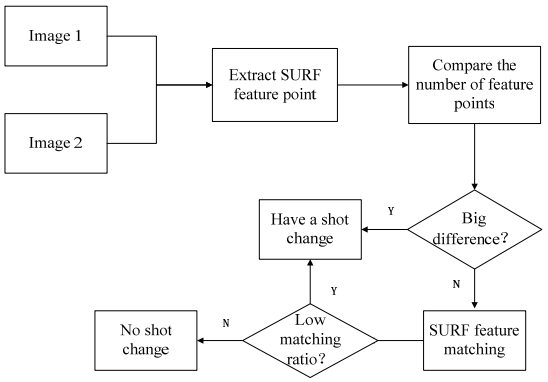

Figure 2. SURF algorithm.

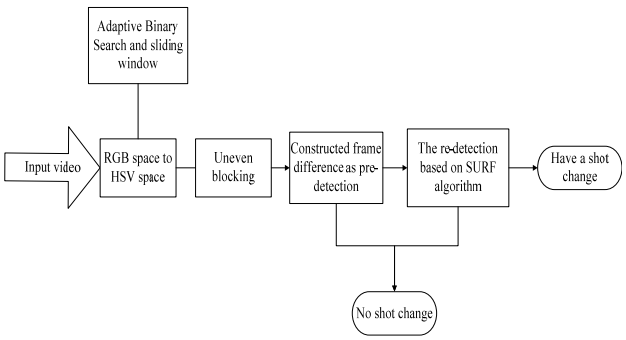

Figure 3. Our method.

E. Our Proposed Method

Figure 3 shows the details of proposed method in this paper. In this method,

1) Read and decode a video window, divide the window into left sub window (LW) and right sub window (RW), where the middle frame belongs to both sub-windows. We assume a video shot sustains 0.5s at least, National Television Standards Committee (NTSC) standardized video frame rate is about 30 frames/sec. Therefore, a video shot lasts at least 15 frames. We set each sub window 8 frames.

2) Calculate the difference between first and last frame of each sub-window, marked as DL, DR. The process of calculation for the uneven blocking for each video frame showing in Figure 2, calculate the difference of hue, saturation and brightness of three groups based on HSV space. Then calculate total difference of hue, saturation and brightness based on HSV space, marked as $FdH(i,j), FdS(i,j), FdV(i,j)$. The calculation formula is as follows:

$$FdH(i,j) = (\sum_{m=1}^{3} a_m FdH(i,j,m))/MN . \quad (1)$$

$$FdS(i,j) = (\sum_{m=1}^{3} a_m FdS(i,j,m))/MN . \quad (2)$$

$$FdV(i,j) = (\sum_{m=1}^{3} a_m FdV(i,j,m))/MN . \quad (3)$$

m is the Group index, i and j are frame indexes, respectively. M is the width and N is the height. Meanwhile $FdH(i,j,m), FdS(i,j,m), FdV(i,j,m)$ represent the difference of hue, saturation, and brightness between i and j frame in the Group m. a_m is the weighting coefficients of Group m. In this paper, we set $a_1 = 0.7, a_2 = 0.2, a_3 = 0.1.$ We get the frame difference of two images by multiplying $FdH(i,j,m), FdS(i,j,m), FdV(i,j,m)$. The calculation formula is as follows:

$$D = FdH(i,j) * FdS(i,j) * FdV(i,j) . \quad (4)$$

3) The calculation formula of pre-detection by adaptive binary search is as follows:

$$D_L > b * D_R . \quad (5)$$

$$D_R > b * D_L . \quad (6)$$

b is the threshold of pre-detection. In this paper, we set $b = 2$. If the DL and DR meet (5) or (6), then we consider that an abrupt transition exists in corresponding to the left or right sub-window. Go to step 1) and search the probable sub window until the window size is 2.

4) If there may be a shot boundary through pre-detection, then we extract the SURF feature points of the two frames, marked as N_1 and N_2 .

$$N_1 < T . \quad (7)$$

$$N_2 < T . \quad (8)$$

$$|N_1 - N_2| > TP . \quad (9)$$

T is the threshold of feature point. TP is the compare threshold of feature point. In this paper, we set $T = 10, TP = 200$.

If N_1 and N_2 meet one of (7), (8), (9), we consider that there exists an abrupt transition. Otherwise match SURF feature points of two frames, calculating the matching ratio as follows:

$$R = \frac{N}{N_1 + N_2 - N} \times 100\% . \quad (10)$$

Where R is the matching rate, N is the number of correctly matched feature points. If R is lower than the threshold (8%), then it confirms the shot boundary between two frames, otherwise it denies the result of pre-detection round.

5) Slide window. If shot boundary has been detected, then considers the last frame of the shot as the first frame of next window. Otherwise the 15th frame of the current window as the first frame of the next window. Repeat the steps above until the end of the video. Finally, if the number of right window frames are less than the left window, we only compare the frame difference between first and last of the left window. And the same as right window. If the difference is obvious, we consider that there exists shot boundary. Also the left of the window frames may also be no more than 8 frames, we only compare the frame difference between first and last of the left window.

III. Experimental and Comparative results

All experiments are carried out in the Visual Studio 2008 platform, using C + + programming language and additional open source library of video processing: OpenCV library. To test the algorithm, we use 5 clips of the program of China Central Television (CCTV) news broadcast. We use recall and precision as metrics. The recall and precision are defined as:

$$R = \frac{N_c}{N_c + N_m} \times 100\% . \quad (11)$$

$$P = \frac{N_c}{N_c + N_f} \times 100\% . \quad (12)$$

N_c is the number of correctly detected shot boundaries, N_m is the number of missed shot boundaries, N_f is the number of false detections.

Experimental results with our algorithm are showed in Table I. In 5 clips of News video, this method correctly detected the 1799 shot boundaries, including 72 missed and 39 false detections. The precision is achieved as 97.8% and recall 96.1%. The reason of missed detection is that the matching ratio is high caused by few feature points. And most of the false detections caused by the irregular movement of the camera.

TABLE Ⅰ
EXPERIMENTAL RESULTS WITH OUR ALGORITHM

News video	Detected	Missed	False	Recall (%)	Precision (%)
News 1	382	16	9	95.9	97.6
News 2	423	13	5	97.0	98.8
News 3	377	16	8	95.8	97.9
News 4	368	16	11	95.7	97.0
News 5	288	11	6	96.2	97.9
Average	1838	72	39	96.1	97.8

A. Compare with the Twin Comparison Method

We compare our method with the Twin Comparison [5] method. Their detection results are presented in Table Ⅱ. We can see that the recall increased by 5.4% and precision increased by 14.3%. So our algorithm has been greatly improved in precision.

TABLE Ⅱ
COMPARISON BETWEEN OUR PROPOSED METHOD AND TWIN COMPARISON METHOD

News video	Proposed method		The twin comparison method	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)
News 1	95.9	97.6	91.0	84.3
News 2	97.0	98.8	91.5	78.1
News 3	95.8	97.9	93.4	89.0
News 4	95.7	97.0	87.3	79.7
News 5	96.2	97.9	90.1	86.5
Average	96.1	97.8	90.7	83.5

B. Compare with the Algorithm in Ref. [2]

We compare our method with the algorithm in Ref. [2]. Their detection results are presented in Table Ⅲ. Compared with the Ref. [2], the recall and precision of our algorithm has been greatly improved. Especially the precision increased of 12.8%.

TABLE Ⅲ
COMPARISON BETWEEN OUR PROPOSED METHOD AND THE ALGORITHM IN REF.[2]

News video	Proposed method		The method in Ref.[2]	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)
News 1	95.9	97.6	93.8	85.3
News 2	97.0	98.8	94.2	87.1
News 3	95.8	97.9	95.1	85.8
News 4	95.7	97.0	94.6	84.7
News 5	96.2	97.9	93.2	82.1
Average	96.1	97.8	94.2	85.0

According to the experimental results in method of Twin Comparison and in Ref. [2] , we can see that Twin Comparison method is simple, but with higher rate of false detection; although the method in Ref. [2] has some improvement, the precision is not very high; our method uses uneven blocking according to the human visual feature, and reduces the false detection rate through the re-detection based on SURF algorithm. So our method achieves better results than the other two.

IV CONCLUSION

The shot segmentation is the first and important step of news video retrieval. In order to complete the shot segmentation, we must determine the measurement method of the frame difference first. In this paper, we proposed a novel shot segmentation algorithm based on adaptive binary search and SURF. In the pre-detection round, the uneven blocking mechanism and the adaptive binary search method are proposed. In the re-detection round, the SURF feature matching algorithm is used to exclude false detections of the pre-detection round and improves the detection precision.

Experimental and comparative results have indicated that our method performances well in news video. Compared with other algorithms, our method can improve the accuracy of detection and reduce the computational complexity. Certainly, there is still much room for shot boundary detection to improve. In the future, we will combine multiple features to overcome missed shots and remove false detection.

ACKNOWLEDGMENT

This work is supported by the National High Technology Development 863 Program of China (No.2011AA01A107) and the Zhejiang Provincial Technical Plan Project (No. 2011C13008).

REFERENCES

[1] Xinghao Jiang, Tanfeng Sun, Jin Liu, Wensheng Zhang, Juan Chao.“ an video shot segmentation scheme based on adaptive binary searching and SIFT,” Lecture Notes in Computer Science (including subseries Lecture Notes in

Artificial Intelligence and Lecture Notes in Bioinformatics) , 6839 LNAI, pp.650-655, 2011.
[2] Hua Zhang, Ruimin Hu, Lin Song, “A shot boundary detection method based on color feature,” Proceedings of 2011 International Conference on Computer Science and Network Technology, ICCSNT 2011, VOL.4, pp.2541-2544, 2011.
[3] J. Yuan, H. Wang, Xiao，L. , Z. Wu，J. Li, et al. , "A formal study of shot boundary detection, " IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 2, pp. 168-186, February 2007.
[4] J.S. Boreczky, and L.A. Rowe, "Comparison of video shot boundary detection techniques," Journal of Electronic Imaging, Vol. 5(2), pp. 122-128, April 1996.
[5] H. Zhang, A. Kankanhalli and S.W. Smoliar, "Automatic partitioning of full-motion video," Multimedia Systems, Vol. 1993, Springer-Verlag, pp.10-28, 1993.
[6] D. Zhang, W. Qi and H. Zhang, "A new shot boundary detection algorithm," Advances in Multimedia Information Processing - PCM2001 Lecture Notes in Computer Science, Volume 2195, pp. 63-70,2001.
[7] FU Chang-jian, LI Guo-hui, HU Jun-tao.Video Hierarchical Structure Mining [J]. Computer Engineering and science, 2006，26(6):159-162.
[8] Boreczky J S, Rowe L A. Comparison of Video Shot Boundary Detection Techniques [A]. In SPIE Conf Storage & Retrieval for Image & Video Databases[C]. San Jose: SPIE, 1996, 170-179.
[9] Ralph M. Ford, Craig Robson, Daniel Temple, Michael Gerlach. Metrics for shot boundary detection in digital video sequences [J]. Multimedia Systems, 2000,8:37-46.
[10] ZHU Xi, LIN Xing-gang. Survey on Video Temporal Segmentation [J]. Chinese Journal of Computer. 2004.27(8): 1027-1035.
[11] Yun Liu, Xueying Liu, Chao Huang. A new Method for Shot Identification in Basketball Video [J]. Journal of Software. Vol 6, No 8 (2011), 1468-1475, Aug 2011.
[12] Peng Qiangqiang, Zhao Long. A Modified Segmentation Approach for Synthetic Aperture Radar Images on Level Set [J]. Journal of Software. Vol 8, No 5 (2013), 1168-1173, May 2013.
[13] Wen-hui Li, Bo Fu, Lin-chang Xiao, Ying Wang, Pei-xun Liu. A Block-based Video Smoke Detection Algorithm [J]. Journal of Software. Vol 8, No 1 (2013), 63-70, Jan 2013.

Ming Jiang He received the B.S. degree and M.S. degree in science in 1996 and 2001 respectively, and Ph.D. degree in Computer Science in 2004, all from Zhejiang University, China. He is currently a Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, Internet QoS provisioning, and network multimedia processing.

Jingcheng Huang He received his BSc in Software Engineering from Hangzhou Dianzi University in 2011. Currently he is a master student in this university. His primary research area focuses on image and video processing, image segment.

Xingqi Wang He received his Bachelor and Master degree from Harbin Institute of Technology in 1997 and 1999, respectively, and Ph.D degree from Zhejiang University in 2002. He is an associate professor in college of Computer Science, Hangzhou Dianzi University, China. All his major are Computer Science. As a researcher, he visited CERCIA, University of Birmingham, UK from 2005 to 2006. His research interests include machine learning, data mining and multimedia content analysis.

Jingfan Tang He received the Ph.D. degree in Computer Science in 2005 from Zhejiang University, China. He is currently an Associate Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, quality assurance, process improvement and legacy system re-engineering..

Chunming Wu He received the B.S. degree, M.S. degree and Ph.D. degree in Computer Science from Zhejiang University, China, in 1989, 1992 and 1995 respectively. He is currently a Professor in college of Computer Science, Zhejiang University, and the director of NGNT laboratory. His research fields include Network Multimedia processing, reconfigurable network technology, network virtualization and artificial intelligence