Shot Boundary Detection Using Co-occurrence of Global Motion in Video Stream

Yosuke Murai, Hironobu Fujiyoshi
Department of Computer Science, Chubu University, Aichi 487-8501 Japan
yoo@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

We propose a method of shot boundary detection based on the co-occurrence of global motion in video stream. In addition to the conventional features based on appearance and local motion, we apply ST (Space-Time) patch analysis for detecting global motion in video stream. And then we perform shot boundary detection by constructing AdaBoost classifiers which represent the co-occurrence of global motion and the conventional features. Experimental results show that our method had 3.8% higher F-measure value than that of the conventional method for gradual shot boundary detection.

1. Introduction

The shot boundary detection from video stream is an important task for video summarizing and handling technology. Various automatic shot boundary detection algorithms have been proposed [1][4][5][6]. The appearance and local motion based features obtained by color information [1] and motion vector [6] are used to detect the "cut" shot boundary. However, the conventional shot boundary detection algorithms are not effective for "gradual" shot transitions. This is because that the gradual shot transition such as dissolve and wipe is very similar to the transition caused by camerawork.

In this paper, we propose a method of shot boundary detection based on the co-occurrence of global motion in video stream. Our approach uses the ST (Space-Time) patch analysis to extract global motions, which is effective to classify a transition into gradual shot boundary or that caused by camerawork. The AdaBoost classifier is used to represent the co-occurrence of the global motion and other various features. The use of co-occurrence of global motion obtained by the ST-patch analysis makes detection more effective.

2 Global motion extraction by ST-patch

The major techniques that have been used for shot boundary detection algorithm are pixel differences, statistical differences, histogram comparisons, edge differences, compression differences, motion vectors, and etc,. These appearance and local motion based features sometimes fail to detect gradual transition because of similar to that caused by camerawork. To overcome this problem, we apply the ST-patch analysis [3] to extract global motion caused by gradual shot transition.

2.1 ST-patch analysis

The ST-patch P is extracted from a small domain of a spatio-temporal image which extend the image in the direction of time [3]. The rank-increase Δr that is one of the ST-patch features describe motion consistency within the ST-patch. The case of $\Delta r \approx 0$ is consistent motion, and when $\Delta r \approx 1$ is inconsistent motion. Another ST-patch feature is the motion similarity. The motion similarity between two different ST-patches which is called P1 and P2 is calculated from continuous rank-increase Δr. Let Δr_1 be the continuous rank-increase of P1. Let Δr_2 be the continuous rank-increase of P2. Moreover, let Δr_{12} be the continuous rank-increase of ST-patch which is combined P1 and P2. Then, the motion similarity is calculated from the following equation:

$$m_{12} = \frac{min(\Delta r_1, \Delta r_2)}{\Delta r_{12}}. \quad (1)$$

The case of $m_{12} \approx 1$ is high similarity, and when $m_{12} \approx 0$ is low similarity. We define detected features Δr and m_{12} as the feature which is based on global motion. In this paper, we compute m_{12} and Δr from the down-sampled images as a ST-patch, as shown in Figure1.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

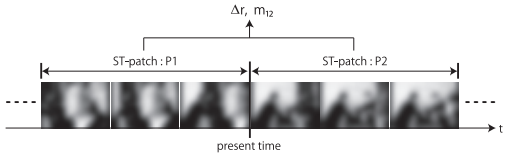

Figure 1. ST-patch analysis

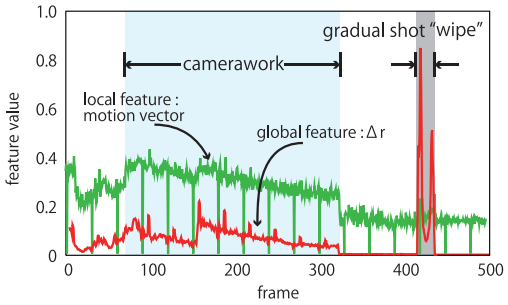

Figure 2. Example of extracted features

2.2 Global motion by ST-patch analysis

We compute the ST-patch features from the whole images. The Procedure for our global feature extraction method is described as follows:

1. Input images are down-sampled to 32×24 pixels (this size is one tenth of an original image).
2. In order to reduce the small motion, a input image is smoothed with a Gaussian filter.
3. ST-patch is computed from the frame buffer of the smoothed low-resolution images.
4. The motion consistency Δr and the motion similarity m_{12} are computed from obtained ST-patchs.

Figure2 shows the value of Δr over time as global motion, calculated from P1 and P2 as shown in Figure1, and motion vectors as a local motion which is one of the conventional features used in shot boundary detection. In the use of local motion, there is no big changes between transitions caused by camerawork and gradual transition such as dissolve and wipe effect. On the other hand, we see that the use of global motion by Δr makes it difference.

3 Overview of proposed method

Our method for using the co-occurrence global motions and the conventional features is diagrammed in Figure3. First, we classify a transition into shot boundary or not at each frame by AdaBoost classifiers. Then, the final classifier outputs the frame number of the first and the last frames of a shot segment.

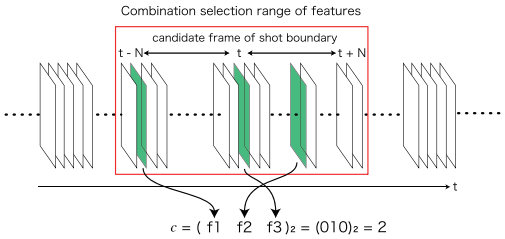

Figure 4. Feature selection by AdaBoost

3.1 AdaBoost classifier

We construct two AdaBoost classifiers for detecting each "cut" and "global" transitions. Our method uses the representation method proposed by Mita *et al.* [9] to express the co-occurrence between the different kinds of features. To improve the generalization performance, we use weak classifiers that observe multiple features. Feature co-occurrence makes it possible to classify difficult examples that are misclassified by weak classifiers using a single feature. We represent the statistics of feature co-occurrence using their joint probability. To calculate the joint probability, we quantize the feature value v_i to two levels. By doing so, each feature value is represented by a binary variable s, which is 1 or 0, specifying shot boundary or non shot boundary respectively. The variable s for an example v_i is calculated by

$$s = \begin{cases} 1 & P(C_p|v_i) > P(C_n|v_i) \\ 0 & otherwise \end{cases}, \tag{2}$$

where s is classified by Bayes theorem:

$$P(C_k|v_i) = \frac{P(v_i|C_k)P(C_k)}{P(v_i)} \quad (k = p, n), \tag{3}$$

where $P(v_i|C_k)$ is the probability that is obtained from the probability density function, $P(C_k)$ is the prior probability, $P(v_i)$ is the existence probability, and C_p is the positive class and C_n is the negative class. The probability density function $P(v_i|C_k)$ is approximated through a smoothed 1D histogram that obtained the training data. The features c are represented by combining the binary variables computed from multiple features. When the variables are 0, 1 and 0, the feature c is computed by

$$c = (010)_2 = 2. \tag{4}$$

When the AdaBoost classifier is trained, computed feature c is automatically selected from a frame buffer ($2N$ frames) as shown in Figure4, as if the error is minimum. So, each weak classifier represents the co-occurrence of various features at various frames.

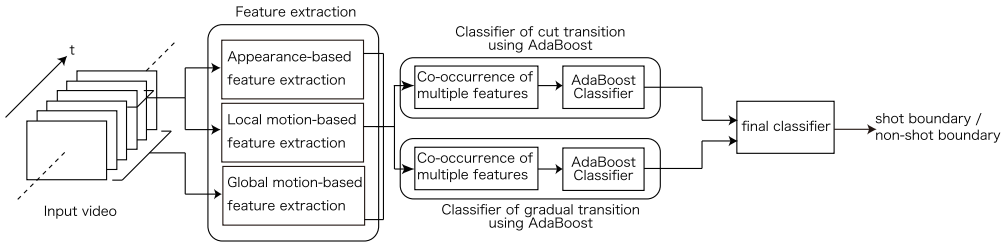

Figure 3. Flow of the proposed method

The final strong classifier of AdaBoost $H(x)$ is a linear combination of K weak classifiers $h_t(x)$:

$$H(x) = sign\left(\sum_{t=1}^{K} \alpha_t h_t(x)\right), \tag{5}$$

where α_t is the weight of training data. A weak classifier, $h_t(x)$, is described by the following equation for the discriminate function based on conditional probability:

$$h_t(x) \begin{cases} +1 & P_t(y=+1|c) > P_t(y=-1|c) \\ -1 & otherwise \end{cases}, \tag{6}$$

where $P_t(y=+1|c)$ and $P_t(y=-1|c)$ are joint probabilities observing feature co-occurrence represented by feature c and the class label $y_i \in \{+1, -1\}$. They are evaluated with respect to weights $D_t(i)$ of the training sample of AdaBoost:

$$P_t(y=+1|c) = \sum_{i:C_t(x_i)=c \wedge y_i=+1} D_t(i), \tag{7}$$

$$P_t(y=-1|c) = \sum_{i:C_t(x_i)=c \wedge y_i=-1} D_t(i). \tag{8}$$

3.2 The final classifier

The final classifier is rule based decision to detect continuity of classification results sequence obtained by the AdaBoost classifier at previous stage. We calculate a distance between color histograms at the start frame and the last frame of a detected segment. If the distance is above the threshold, the detected segment is extracted as a shot boundary. This procedure works to reduce a false positive.

4 Experiment

4.1 Evaluation

We use 85,000 frames for training and 150,000 frames for test from the VDB(Video Data Base) [7] as shown in Table1, which contain video programs of drama, news, comedy and etc. In this evaluation, we used three information retrieval measures [2]: Recall, is the proportion of correct retrievals compared to all possible correct retrievals. Precision, is the proportion of correct retrievals among all retrieval results.

Table 1. Data set

Data	# of frames	# of cuts	# of graduals
Training	85,000	400	1,300
Test(cut)	130,000	600	1,800
Test(grad)	20,000	120	780

$$r = \frac{correct}{correct + missed},\ p = \frac{correct}{correct + false} \tag{9}$$

The F-measure summarizes both into one number:

$$F(r,p) \quad = \quad \frac{2rp}{r+p} \tag{10}$$

4.2 Experimental results

We compare two feature pattern methods, "A,L" (A,L means the use of appearance and local motion feature as the conventional method [5]), and "A,L,G" (A,L,G means conventional feature (A,L) plus global motion (G) obtained by ST-patch analysis). Table2 shows the experiment results of "cut" shot boundary detection. We see that our method has almost same performance comparing to the conventional approach for "cut" shot boundary detection. Table3 shows the experiment results of "gradual" shot boundary detection. Our method had 3.8% higher F-measure value than that of the conventional method. This indicates that using global motion obtained by the ST-patch analysis is effective for shot boundary detection.

4.3 Discussion

Each round of AdaBoost chooses from the total set of the various features at various frames. The final classifier balances the various features in order to maximize classification performance. The weight α, the selected feature, and the selected frame are chosen at each

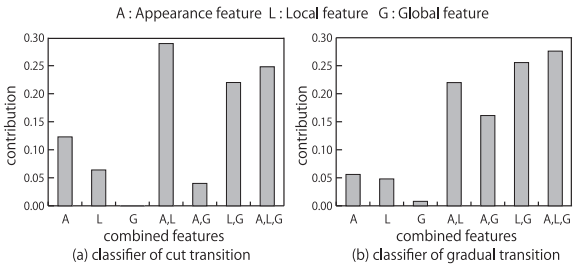

Figure 5. Contribution of combined featuers

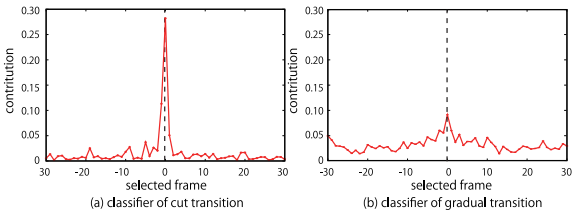

Figure 6. Contribution of frames

Table 2. Experimental result of "cut" boudnary

		Conventional method (A,L)	Proposed method (A,L)	Proposed method (A,L,G)
cut	R	98.7	97.1	98.2
	P	96.6	98.4	98.6
	F	97.6	97.7	98.3

Table 3. Experimental result of "gradual" boundary

		Conventional method (A,L)	Proposed method (A,L)	Proposed method (A,L,G)
grad	R	74.0	75.4	80.1
	P	92.0	91.7	92.4
	F	82.0	82.7	85.8

round are very important factors for classification performance. So, we evaluate which feature is effective for shot boundary detection by calculating the sum of weight α for each feature set c as a contribution of features. Figure5 shows the contribution of combined feature c (cf., Section 3.1). In the case of the "cut" shot classifier, the contribution of the combined feature A,L is higher than others. The feature of global motion does not contribute to the "cut" shot boundary detection. On the other hand, in the case of the "gradual" shot classifier, the contribution of A,G is higher than that in the "cut" shot classifier. Moreover, L,G and A,L,G, which combined global motion feature and the conventional features, are higher than those in the case of the cut classifier. This result indicates that the global motion feature is effective for "gradual" boundary detection.

Second, we evaluate which frame is effective for shot boundary detection by calculating the sum of weight α for each frame as a contribution of frame. From Fig.6(a), we see that selected frames in the case of the "cut" classifier contribute around the 0 frame. This is because the period of a "shot" segment is short. On the other hand, the selected frames contribute overall in the case of gradual classifier. This is because gradual transitions such as dissolve and wipe occur over multiple frames. These results show that the co-occurrence of various features at various frames is important for gradual boundary detection.

5 Conclusion

In this paper, we propose a method of shot boundary detection based on the co-occurrence of global motion and the conventional features. The experimental results show that our method had 3.8% higher F-measure value than that of the conventional method. We confirm that using the co-occurrence of global motion by ST-patch analysis is effective for gradual shot boundary detection.

References

[1] A. Nagasaka and Y. Tanaka. Automatic video indexing and full-video search for object appearances. *Transactions of Information Processing Society of Japan*, 33(4):543–550, 1992.

[2] C. J. van Rijsbergen. Information retrieval. *Butterworths*, 1797.

[3] E. Shechtman and M. Irani. Space-time behavior based correlation. *Computer Vision and Pattern Recognition*, 1:405–412, 2005.

[4] H. J. Zhang, A. kankanhalli, and S. W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1:10–28, 1993.

[5] K. Suzuki, H. Sakano, S. Ohtsuka. Gradual shot boundary detection from video sequence using edge difference features and discriminant analysis. *Technical report of IEICE. PRMU*, 103(295):31–36, 2003.

[6] K. Suzuki, M. Nakajima, H. Sakano. Abrupt shot boundary detection from video sequence using motion direction histogram feature. *Proc. of the IEICE General Conference*, 2003.

[7] N. Babaguchi, et al. Video database for evaluating video processing. *Technical report of IEICE. PRMU*, 102(155):69–74, 2002.

[8] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying production effects. *Multimedia Systems*, 7:119–128, 1999.

[9] T. Mita, T. Kaneko and O. Hori. Joint haar-like features for face detection. *IEEE International Conference on Computer Vision*, pages 1619–1626, 2005.