IDENTIFICATION AND DETECTION OF THE SAME SCENE BASED ON FLASH LIGHT PATTERNS

Masao Takimoto, Shin'ichi Satoh and Masao Sakauchi

University of Tokyo, National Institute of Informatics

ABSTRACT

A method has been developed to identify video shots of the same scene where camera flash lights are observed, and the method has been tested by using it to detect such shots from a large TV video archive. Camera flashes are often used in impressive scenes, such as interviews of important persons. Because such scenes are broadcasted repeatedly on various TV programs, a method for detecting them is a promising approach for semantic video indexing. The proposed identification method is invariant to the differences in viewpoint, illumination or any other visual environment because it depends on comparison between temporal occurrence patterns of flash lights. Furthermore, because each flash pattern is represented with a binary array, the comparison requires low computing cost. These advantages mean that the proposed method can be considered to provide semantic and efficient video analysis.

1. INTRODUCTION

Environments to accumulate large amounts of video data are becoming popular. For example, in our laboratory at the National Institute of Informatics (NII), TV broadcasting videos are accumulated in MPEG format. However the lack of an efficient and convenient way to use them for various purposes is still an important problem. As regards TV videos, people are able to choose video segments from TV program guides, but such information is created by humans and is not always satisfactory for the audience.

One promising approach to solving this problem is the "shot identification" method. This identifies pairs of shots of the same scene: these shots are called "identifiable shots" in this paper. Moreover, if identifiable shots are detected from a large video archive, the existence of semantic relationship between them is simultaneously clarified. This provides useful information for semantic recognition of video archives. For example, scenes broadcasted many times can be considered as containing some important contents.

We call shots in which camera flash lights are observed "flash shots". We introduce a method to detect identifiable flash shots. Although this method can deal only with flash shots, it is a good approach to getting semantically meaningful relationships from video archives because flash shots are often impressive for the audience.

The rest of the paper is organized as follows. The background of the research is outlined in Section 2. In Section 3, the algorithm used in the identification is given. Section 4 describes our experiments and evaluations. The conclusion and future work are discussed in Section 5.

2. BACKGROUND

Compared with other existing methods of video analysis, the method proposed in this paper has three characteristics.

Identification of shots not frames

Many studies have defined similarity between images. Most of them are based on feature values in images, such as color histogram and feature points. However, our goal is the identification of shots not images. For this purpose, our method uses not only image features themselves, but also how they changes through a shot. The "flash pattern" which means the occurrence pattern of frames where flash light is used, is the criterion for identification in our method. This approach enables the identification by the shot rather than by the frame.

Availability in huge video archive

Our method consists of two operations: flash detection and pattern comparison. They can be formed with simple calculations by using physical properties of flash lights which are described in detail in Section 3. These simple calculations mean that our method has rather low computing cost and can be applied to huge video archives. This is a big advantage over other video analysis algorithms that use complicated parameter estimation.

Invariance to cameras, temporal offsets and added captions

Identifiable shots are detected even if they are taken by different cameras(Fig.1) because the identification is based on temporal flash patterns as mentioned above. Furthermore, if there is temporal offset between shots or different captions are added to each shot, the identification works still accurately. This invariance enables identification over different TV programs and broadcasting stations. Some researchers have been tackling similar problem which is called "image near duplicate detection". By using this expression, our method can be

1-4244-0367-7/06/$20.00 ©2006 IEEE

said as “shot near duplicate detection”.

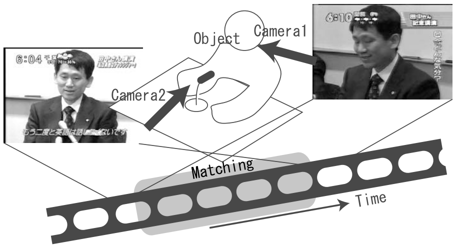

Fig. 1. Identifiable shot

3. ALGORITHM

The identification method consists of two steps: flash detection and pattern comparison.

3.1. Flash detection

Before flash detection, videos are divided into shots. Shot changes are detected by examining the difference between color histograms between frames, but shot change detection is not the concern of this paper. Proper shot detection is presupposed.

First, the average luminosity of every pixel is calculated from every frame and if it rises and falls rapidly in few frames, it is considered that flash light is taken at a frame of local maximum average luminosity (Fig.2). Such a frame is called a “flash frame” in this paper.

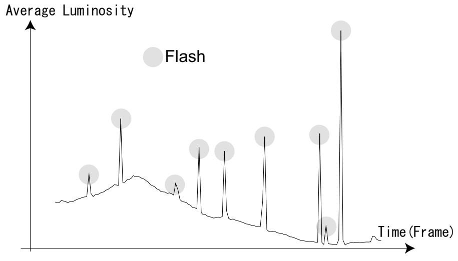

Fig. 2. Average luminosity changing through the flash shot

This method seems very simple but it can detect almost all flash frames. However, many false-positives remain. Fig.3 shows examples of them.

In the shot of Fig.3(a), a fluorescent light blinks regularly. In the shot of Fig.3(b), snowflakes fall and a light turns on and off. In the shot of Fig.3(c), a fire flickers. In each of these examples, the average luminosity changes similarly to that in true flash shots. In order to avoid these false-positives, two further operations are applied. These are detailed in next two paragraphs.

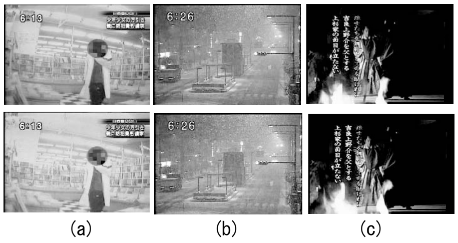

Fig. 3. Examples of false-positives

Validation by matching frames (frame validation)

It is assumed that when the frame rate is high enough, the motions of objects between adjacent frames must be very small. To the contrary, snowflakes or fires, which might cause false-positives can change their locations or forms so rapidly that the difference in pixel value between frames tends to be large. Because of that, by matching the neighboring frames of each flash frame, such false-positives can be differentiated from true flash shots.

First, in order to eliminate the effect of camera motion such as panning or zooming, optical flow values are estimated and the neighboring frames of the candidate flash frame are translated according to these values. Next, new two images are created by subtracting each translated image from the flash frame in pixel intensity. (In Fig.4 which shows this operation, these images are called “luminosity difference images”.) Ideally, pixel values in these images should correspond to only the effect of flashlights. Therefore if a pixel in one image has high intensity, that is to say, the pixel is affected by flash light, the intensity of the corresponding pixel in another image should also be high.

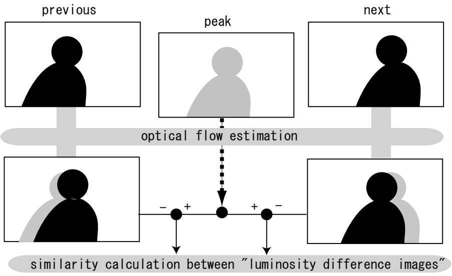
previous
peak
next
optical flow estimation
−
+
+
−
similarity calculation between "luminosity difference images"

Fig. 4. Frame validation

Based on this supposition, the distance between the two created images is defined:

$$d_{frame} = \frac{n}{N}$$

where n denotes the number of pairs of corresponding pixels in which only one pixel intensity is higher than some threshold and N denotes the number of pairs of corresponding pixels in which at least one pixel intensity is higher than the threshold. In this definition, if d_{frame} is low, it is considered that flash lights is used in that frame.

Validation of temporal occurrence patterns (pattern validation)

In order to discriminate between true flash shots and other types of false-positives, the temporal occurrence pattern of the flash frame (flash pattern) is validated. For example, a blinking fluorescent light gives a visual effect that is similar to that of flash light, so frame validation cannot deal with such false-positives. Pattern validation, however, depends on the assumption that the occurrence of flash lights follows the form of a Poisson arrival because they are produced by humans. Therefore the observed flash pattern is compared with the Poisson arrival model.

$$d_{pattern} = \sum \frac{(f_i - np_i)^2}{np_i}.$$

This is the distance in a histogram of intervals of flash frames, where f_i denotes the observed number of intervals whose length is i frames, p_i denotes theoretical probability, and n is the total number of intervals. If $d_{pattern}$ is higher than some threshold, the pattern should be a false-positive.

3.2. Pattern comparison

To compare any pair of patterns, we defined a method to compare two patterns of the same length. When two patterns are given, every flash frame in one pattern is checked to see whether the corresponding (or adjacent) frame in another pattern is also a flash frame. If both frames are flash frames, they are considered to be "matched". Then the ratio between the number of matched flash frames and the number of all flash frames is defined as the similarity between patterns. This similarity can be considered as a kind of edit distance. However, our matching criteria do not require that two flash frames be in exactly corresponding frames. If one is found in an adjacent frame, the two frames are also considered to be matched.

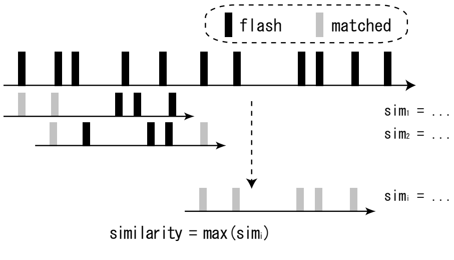

Fig. 5. Calculation of similarity between shots

This local similarity is calculated with every temporal offset (Fig.5). Global similarity between two patterns is defined as the maximum local similarity. This value is used in comparison between shots: pairs that have high similarity are considered to be identifiable shots. Thresholds of shot length and the number of flash frames are applied for accurate comparison. These parameters also guarantee the uniqueness of each pattern.

4. EVALUATION

First, flash frames were detected from 80 hours of video extracted from our video archive (Section 1) by applying the method described in 3.1. Ground truth is examined manually. Based only on average luminosity, 928 shots were detected as flash shots, and 709 shots among them were confirmed to be true-positives. Because the threshold of the change in average luminosity was set to a rather low value, recall was almost 100%. The total length of these flash shots was about 3 hours, which is about 3.5% of all of the examined videos.

4.1. Flash validation

The two validation methods described in 3.1 were applied to the detected flash shots. Figs.6 and 7 are the results, which are histograms of true-positives and false-positives about each validation value.

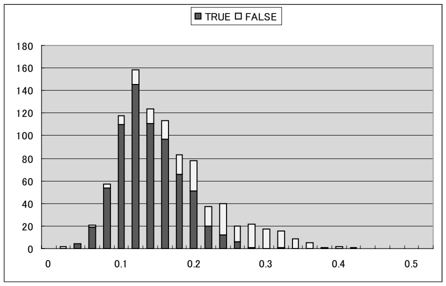

Fig. 6. Histogram of true-positives and false-positives after frame validation

Table.1 is the result of flash detection from another 80 hours of videos. Based on the results of the previous experiment, shown in Figs.6 and 7, the threshold of each validation value was set to 0.25 (frame validation) and 80.0 (pattern validation). Precision improved to over than 90% due to validations, and such high precision in flash detection also makes the accuracy in the detection of identifiable shots much higher.

4.2. Pattern comparison

Similarity as defined in 3.2 was calculated for every pair of flash shots found in the 80 hours of videos. Fig.8 is the his-

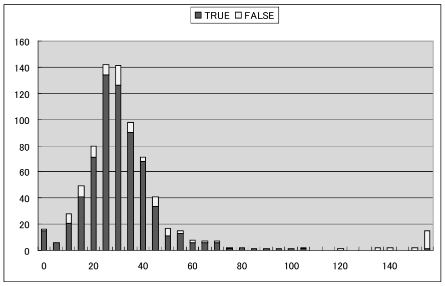

Fig. 7. Histogram of true-positives and false-positives after pattern validation

method	recall	precision
Detection by average luminosity	100%	76.4%
Frame validation	93.9%	87.9%
Pattern validation	93.1%	90.4%

Table 1. Precision-recall of each flash validation

togram of the similarities. There are 213 true shot pairs in this data set. In this histogram, a gap between true-positives and false-positives exist at 0.7-0.9. This shows the fact that a pair of different flash scenes hardly ever has identical or similar flash patterns. In fact, if the threshold of similarity is set to 0.8, about 90% of positive pairs are true-positives.

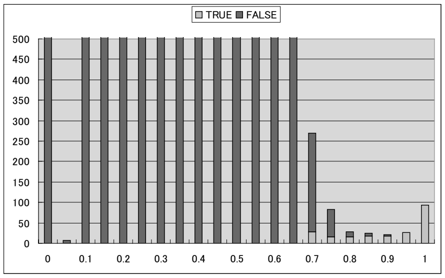

Fig. 8. Histogram of similarities from every pair of flash patterns

In this experiment, it took about 13 minutes to calculate the similarities of all pairs. The time complexity of this search is $O(M^2)$ (M denotes the number of flash shots). In that sense, removing false-positives in flash detection by flash validations is highly effective.

4.3. Detection results

Fig.9 is examples of detected identifiable flash shots. For each pair of these shots, two shots were actually taken by different cameras, different captions were added, and the shots have different temporal offsets.

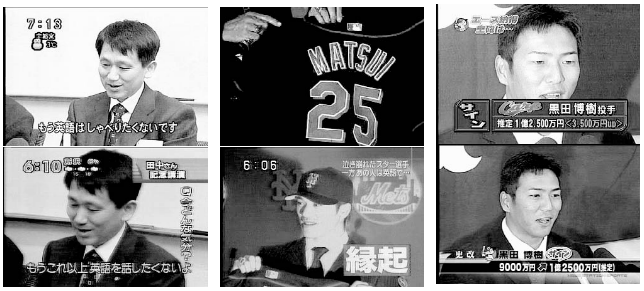

Fig. 9. Examples of detected pairs

5. SUMMARY

The proposed method has been used to detect shots that contain identical flash scenes from a large amount of video footage. Because the method depends on comparison in temporal patterns, the calculations of similarity were not affected by differences of cameras, captions, temporal offset, or any other environmental factors. This dependency on temporal information is the characteristics of this method. However, no application that uses the information detected by this method has yet been constructed. This will be addressed in future work.

6. REFERENCES

[1] Yaron Caspi and Michal Irani, "Spatio-Temporal Alignment of Sequences" IEEE transaction on pattern analysis and machine intelegence, Vol.24, No.11, November, 2002.

[2] P.H.S.Torr and A.Zisserman, "Feature Based Methods for Structure and Motion Estimation" Proc. Vision Algorithms Workshop, pp.279-295, 1999.

[3] Josef Sivic and Andrew Zisserman. "Video Google: A Text Retrieval Approach to Object Matching in Videos" Proc. ICCV, pp.1470-1477, 2003.

[4] Nozha Boujemaa, Francois Fleuret, Valerie Gouet and Hichem Sahbi. "Automatic Textual Annotation Of Video News Based on Semantic Visual Object Extraction" Proc. SPIE, Vol.5307, pp.329-339, 2003.

[5] G.Piriou, P.Bouthemy and J.F.Yao. "Extraction of Semantic Dynamic Content from Videos with Probabilistic Motion Models". Proc ECCV, vol.3, pp.145-157, 2004.

[6] Dong-Qing Zhang and Shin-Fu Chang. "Detecting Image Near-Duplicate by Stochastic Attributed Relational Graph Matching with Learning". Proc ACM Multimedia, pp.877-884, 2004.