An Abstraction of Low Level Video Features for Automatic Retrievals of Explosion Scenes

Jongho Nang[1], Jinguk Jeong[1], Sungyong Park[1], and Hojung Cha[2]

[1] Dept. of Computer Science, Sogang University, 1 Shinsoo-Dong, Mapo-Ku
Seoul 121-742, Korea
jhnang@ccs.sogang.ac.kr
[2] Dept of Computer Science, Yonsei University, Seoul 120-749, Korea

Abstract. This paper proposes an abstraction mechanism of the low-level digital video features for the automatic retrievals of the explosion scenes from the digital video library. In the proposed abstraction mechanism, the regional dominant colors of the key frame and the motion energy of the shot are defined as the primary low-level visual features of the shot for the explosion scene retrievals. The regional dominant colors of shot are selected by dividing its key frame image into several regions and extracting their regional dominant colors, and the motion energy of the shot is defined as the edge image differences between key frame and its neighboring frame. Upon the extensive experimental results, we could argue that the recall and precision of the proposed abstraction and detecting algorithm are about 0.8, and also found that they are not sensitive to the thresholds.

1 Introduction

Recently, there have been a limited number of research efforts to retrieve a high level information automatically from the digital video for a specific purpose. A noticeable example of this research is the MoCa system [1,3,4] developed at Universität Mannheim. In this system, the genres of the film are automatically classified as news, tennis, animation, and advertisement by analyzing the low level video features such as motion energy and scene length. Furthermore, a violence scene is automatically detected by analyzing the motion of the objects in the video. All of these researches extract some high level semantic information from the multimedia data by analyzing and comparing the low level video features such as color histogram, object motion, and shot length to the predefined low level features of interesting (or willing to detect) events.

In order to summarize or extract the highlights of the long action movies, the scene with interesting events such as the explosion, the car racing, and the gun fighting should be identified first. Among these noticeable events, the explosion of building or car might be the most interesting events and usually the highlight of the movies that every user wants to retrieve. This paper proposes an abstraction mechanism of the low level features of digital video for the automatic retrieval of explosion scenes from a large video archive. Since the explosions of building or car are always accompanied

Y.-C. Chen, L.-W. Chang, and C.-T. Hsu (Eds.): PCM 2002, LNCS 2532, pp. 200-208, 2002.

with a yellow-tone flame that is changed rapidly, these features could be used to abstract the explosion events. In the proposed abstraction mechanism, the regional dominant color of the key frame, the motion energy of the shot, and the simplicity of the edge image of the shot are selected as the abstraction of shot for automatic explosion retrievals. The proposed automatic explosion scene retrieval algorithm declares a scene has an explosion event if it contains a shot whose regional dominant colors include a yellow-tone color, its motion energy is higher than that of other shots in the scene, and the edge image of its key frame is relatively simple compare to that of other neighboring frame. Upon the extensive experimental results while changing the thresholds used in retrieval algorithm, we could argue that the recall and precision are more than 0.8, and these values are robust to the thresholds. The proposed explosion scene retrieval algorithm could be used to build a digital video library with a high level semantic query capability, and summarize and abstract the digital movies automatically.

2 Abstracting and Retrieval of Explosion Scene

2.1 Characteristics of Explosion Shot

We could extract three characteristics of explosion events by analyzing several explosion scenes in the various movies as shown in Figure 1. First, if there is an explosion of building, car, or bomb in the shot, the frames in the shot contains a lot of yellow-tone pixels because there are always strong yellow-tone flames in the explosion. The whole image or part of the image could be spread with the flames. Secondly, since the flames are changed dynamically while the explosion is progressed, there are always a lot of motions in the shot with explosions. Finally, since the flames veil all other objects in the frame, they are not visible precisely. Although there would be other events that meet above three characteristics or there is an explosion event that does not have above three characteristics exactly, they could be used to effectively identify the explosion events as shown in our experimental results.

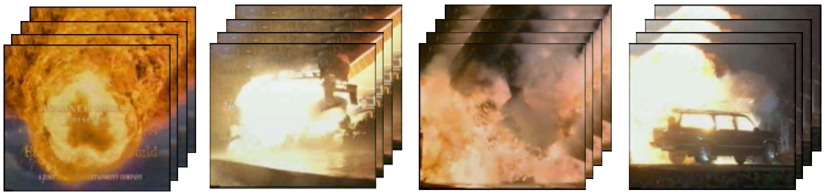

Fig. 1. Some example of explosion shots

2.2 Abstraction Mechanism of Explosion Shot

The low level visual features of shot with explosion event should be carefully selected and properly abstracted in order to retrieve the explosion shots precisely. The low level video features proposed in this paper for automatic explosion scene retrievals are *the regional dominant color* of the shot which reflects the color of the flames, *the* *motion energy* of the shot which reflects the rapid spread of the flames, and *the simplicity of the edge image* of the shot which reflects the phenomenon that other objects in the shot are hidden by flames.

(1) Abstraction of Color Information
Let us first present how to abstract the color information of the shot for the automatic explosion retrievals. Since to make a color histogram [5] of the all frames requires a lot of the computations, we only check the color histogram of the key frame (1^{st} frame) of the shot. Since the flames in the explosion could be appeared in the whole or the part of the frame image, we divide the key frame image to several regions and check the dominant color of each region separately. The dominant color of region is the highest ranked color in the color histogram of that region. If the number of the regions of the key frame whose dominant color is the yellow-tone exceeds a certain threshold, we suspect that the key frame (or shot) may have an explosion event. Of course, the range of the yellow-tone color is defined as the 48 yellow-tone colors among the 512 quantized RGB color space. Let us formalize the proposed color abstraction mechanism. Assume that the key frame of the shot *i*, K_i, is divide into *m* regions. Let *Y* be a set of 48 yellow-tone colors extracted from the 256 quantized RGB color space, and d_i^j be the dominant color of the *j*-th region of K_i. We declare that the shot may have an explosion event when its key frame K_i satisfies following condition;

$$\left|\{d_i^j \in Y \mid 1 \le j \le m\}\right| < \alpha \cdot m \tag{1}$$

where α is a threshold ($0 < \alpha < 1$).

(2) Abstraction of Motion Information
The most interesting characteristic of the explosion shot is that there are more object motions than the previous and successive shots in the same scene because the flames are rapidly spread into whole frame in the explosion shot. One way to abstract this object motion information is to compute the total amount of motions in the shot. Let F_j^i and F_{j+k}^i be the *j*-th and (*j+k*)-th frame of the shot *i*, and E_j^i and E_{j+k}^i be their binary edge images respectively. Then, the motion energy of the *i*-th shot, M_i, is defined as follows;

$$D_{j,j+k}^i = MF(E_j^i - E_{j+k}^i) \tag{2}$$

$$M_i = \sum_m \sum_n D_{j,j+k}^i(m,n) \tag{3}$$

where $D_{j,j+k}^i$ is the edge image difference between E_j^i and E_{j+k}^i, $D_{j,j+k}^i(m,n)$ is the pixel value at coordinate (*m*,*n*) of $D_{j,j+k}^i$, and *MF* is a median filter [2] to remove the noise. We declare that the shot may have an explosion event when the motion energy of the shot is β times higher than the average motion energy of the shots in the same scene of *n* shots. Let M_i be the motion energy of the *i*-th shot, then above condition could be represented as follows;

$$M_i \ge \beta \cdot \frac{\sum_{j=1}^{n} M_j}{n} \tag{4}$$

If there are at least δ shots that satisfy this condition, we declare that the scene may have an explosion event. This abstraction mechanism is shown in Figure 2 graphically, in which the j-th and ($j+k$)-th frames of the shot are extracted from the shot, and the edge image difference of these frames is computed by subtracting one edge image from the other edge image. The resulting image passes a median filter to remove the noise. The number of remaining pixels in the edge image difference is defined as the abstraction of the motion energy of the shot.

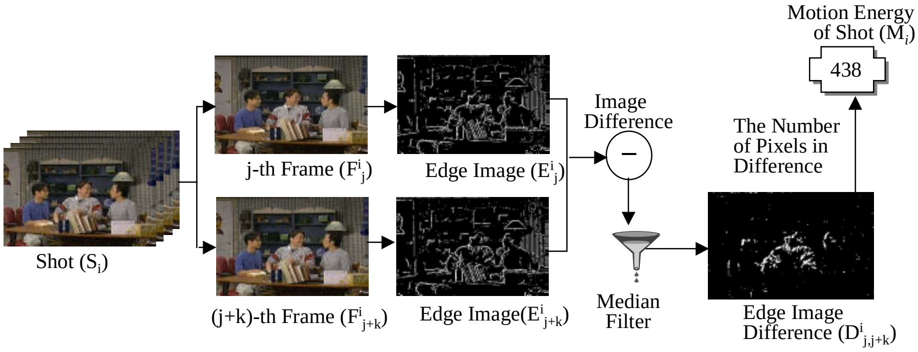
Motion Energy of Shot (M_i)
438
Image Difference
The Number of Pixels in Difference
j-th Frame (F^i_j)
Edge Image (E^i_j)
Shot (S_i)
(j+k)-th Frame (F^i_{j+k})
Edge Image(E^i_{j+k})
Median Filter
Edge Image Difference ($D^i_{j,j+k}$)

Fig. 2. Abstraction of motion information in shot

(3) Simplicity of the Edge Image

If there is a shot that has a lot of yellow-tone objects with a dynamic movement or camera operations such as panning, above two abstraction mechanisms alone could not retrieve the explosion shot precisely. Fortunately, other characteristic of explosion event could help us to solve this ambiguity. It is the simplicity of the edge image of the frames in the explosion shot. Since the flames are usually spread rapidly in the explosion shot, the other objects in the shot are being hidden rapidly as shown in Figure 1. It means that the binary edge images of the frames in the explosion shot come to be simple as the explosion is being progressed. We could use this information to distinguish the other yellow-tone shots with high motion energy from the explosion shots. In the case of the explosion shot, since the flames come to hide all other objects in the frame rapidly, the difference of the number of edge pixels between frames in the same shot is large enough in addition to a high motion. However, in the case of a dynamic shot without explosion, since all objects are always visible clearly in all frames of the shot, the difference of the number of edge pixels between the frames in the shot is small enough although the motion energy is. This characteristic of explosion could be used to filter out the shots with a relatively high motion energy but not an explosion shot. This condition is represented as follows;

$$G^{i}_{j,j+k} = \frac{\sum_{m}\sum_{n}\left|E^{i}_{j+k}(m,n)\right|}{\sum_{m}\sum_{n}\left|E^{i}_{j}(m,n)\right|} \geq \gamma \tag{5}$$

where $G^{i}_{j,j+k}$ is the edge pixel difference between F^{i}_{j} and $F^{i}_{j,j+k}$ in S^{i} , $E^{i}_{j}(m,n)$ is the pixel value at coordinate (*m*,*n*) of the edge image E^{i}_{j}, and γ is a threshhold. Eq. (5) needs to compute only ’differences in the number of edge pixels’ because the motion energy has already been considered in Eq.(2).

2.3 Retrieval Algorithm

Let us briefly explain an overall explosion shot retrieval algorithm shown in Figure 3. First, we index an MPEG video stream into shots, and grouping them into the scenes automatically or manually. The key frame (or first frame) of shot and its neighbor frame are selected to compute the regional dominant colors and their binary edge images. With these binary edge images, the motion energy of each shot could be computed with Eq. (2) and (3). If the regional dominant colors of the key frame include a yellow-tone color and the motion energy of the shot is higher than the average motion energy of the scene (Eq. (4)), we could declare that the shot contains an explosion event. Since there is also a shot that has a high motion energy without any explosions, we use Eq. (5) to filter out these shots. If above three conditions are satisfied simultaneously, we declare that the shot has an explosion event. These conditions are sufficient and robust enough to retrieve almost all explosion events as shown in the following experiments.

3 Experimental Results and Analyses

We have implemented the proposed abstraction and retrieval algorithm on the top of the digital video library that we are currently building. The performance of the retrieval algorithm presented in Figure 3 depends on the several threshold values[1] such as ;

1. the β in Eq. (4) which represents the number of times that exceeds the average motion energy of the scene to declare an explosion shot,
2. the δ which represents the minimal number of shots which hold the condition in Eq.(4) to declare an explosion scene,
3. the γ value in Eq. (5) which represents the ratio of the number of edge pixels in current and neighboring binary edge images to declare an explosion shot.

[1] We divided the frame into 4x4 regions when computing the regional dominant color of the key frames in the experiments. We declare a shot may have an explosion event if more than two regions of the key frame have a yellow regional dominant color. It means that we fix the α = 0.125 (2/16) in our experiments.

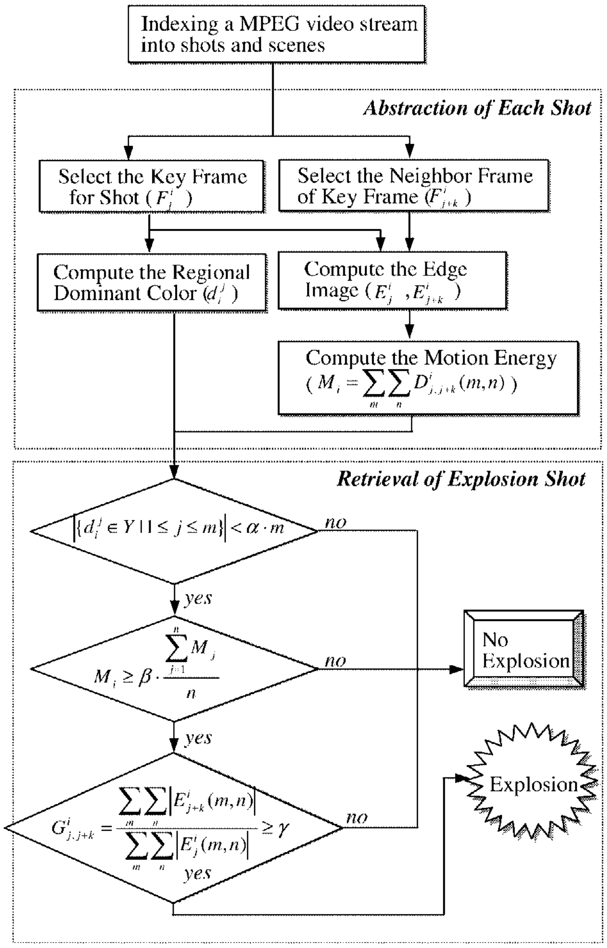
Indexing a MPEG video stream into shots and scenes
Abstraction of Each Shot
Select the Key Frame for Shot (F_j^i)
Select the Neighbor Frame of Key Frame (F_{j+k}^i)
Compute the Regional Dominant Color (d_i^j)
Compute the Edge Image (E_j^i ,E_{j+k}^i)
Compute the Motion Energy ($M_i = \sum_m \sum_n D_{j,j+k}^i(m,n)$)
Retrieval of Explosion Shot
$\left|\{d_i^j \in Y \mid 1 \le j \le m\}\right| < \alpha \cdot m$
no
yes
$M_i \ge \beta \cdot \frac{\sum_{j=1}^{n} M_j}{n}$
no
No Explosion
yes
$G_{j,j+k}^i = \frac{\sum_m \sum_n \left|E_{j+k}^i(m,n)\right|}{\sum_m \sum_n \left|E_j^i(m,n)\right|} \ge \gamma$
no
yes
Explosion

Fig. 3. An explosion shot retrieval algorithm

We have experimented proposed retrieval algorithm with several action movies such as *Lethal Weapon-4*, *Terminator-1*, and *Platton* each of which has a lot of explosion scenes. Since each movie is too long (about two hours) to store as a single MPEG file, we divide it into two segments so that totally 6 segments of one hour long are used in our experiments. Table 1 shows the performance of the proposed retrieval algorithm with the threshold set that produces the best performance. The threshold values producing the best performances are obtained via an extensive experimental analysis. As shown in this table, the proposed algorithm produces more than 90% performance except *Platoon* movie. Since it is a war movie in Vietnam, the explosion sometimes raises a cloud of dust so that the explosion does not always accompany with the yellow-ton flames. That makes the proposed algorithm sometimes fails to detect the explosion event in *Platoon.* Furthermore, since it is a combat movie at a Vietnam jungle so that there might be a rapid action on a jungle of trees with the yellow leaves, the algorithm sometimes detects it as an explosion event. That is the reason why proposed algorithm produces a false detection in *Platoon*. Except these special cases, the proposed algorithm could produce a very high performance as shown in the "Total (2)" field of the Table 1 that represents the recall and precision values of the experiments excluding these cases. Of course, this relatively high performance might be influenced by the threshold values. However, their effects are not so much as explained in the following experiments.

In order to investigate the effects of the threshold values on the performance of the proposed algorithm, we have tested the algorithm with the same movies while varying the threshold values α, δ, and γ. Table 2 shows the summary of other experiments while varying the threshold values presented in Eq. (1), Eq. (4), and Eq. (5). If we neglect the color information of shot, the recall could be higher since all shots with a high motion energy are extracted, but the precision is lower as shown in the first three rows in Table 2. On the other hand, if we consider the color information, the recall would be somewhat lowered, but the precision would be raised as shown in the last three rows in Table 2.

Table 1. Experimental Results when $\alpha = 0.125$, $\beta = 5$, $\delta = 2$, $\gamma = 200$

Movie Title	Number of Scenes	Number of Explosions	Recall	Precision
Lethal Weapon-4 (1)	16	3	1.00	1.00
Lethal Weapon-4 (2)	16	3	1.00	1.00
Terminator-1 (1)	32	3	1.00	0.75
Terminator-1 (2)	19	6	1.00	1.00
Platoon (1)	18	4	0.00	0.00
Platoon (2)	18	4	0.75	0.75
Total (1)	119	23	0.78	0.86
Total (2)	101	19	0.95	0.90

Table 2. Experimental Results on the Effect of Thresholds to the Performance

Checked Conditions		Recall	Precision
Color is neglected	Eq.(4)	0.98	0.31
	Eq.(5)	0.87	0.33
	Eq.(4) + Eq.(5)	0.87	0.38
Color is considered	Eq.(1) + Eq.(4)	0.78	0.78
	Eq.(1) + Eq.(5)	0.74	0.81
	Eq.(1) + Eq.(4) + Eq.(5)	0.74	0.81

From these experiments, we found that the regional color of the key frame greatly contributes to the precision of the proposed algorithm since it filter out the other shots with rapid changes. The performance of the proposed algorithm is not so sensitive to α if it is in the range of $0.125 \leq \alpha \leq 0.25$. Furthermore, the checking of the differences of edge images (*i.e.* checking the ratio of the number of edge pixels in the binary edge images of two neighboring frames in the same shot) could contribute to the precision of the proposed algorithm because this condition filters out the other shots in which the yellow objects are moved rapidly but not an explosion shot. This phenomenon that the flames rapidly hide the other objects in the frame is a characteristic of the explosion shot that other dynamic shots do not have. However, its effect to the performance of the proposed algorithm is somewhat small.

4 Concluding Remarks

The extracting of a high level semantic information from the digital movie automatically is an important task to build a useful digital video library. However, it has been a very difficult task without a lot of sophisticated artificial intelligence techniques that would not be available in the near future. Recently, there have been some researches to extract a limited number of high level information from the digital video, and these researches usually try to extract a specific information such as dialogues, action, and violence. The explosion shot abstracting and retrieval algorithm proposed in this paper is the one of these research efforts. This paper analyzes the characteristics of the explosion events, and find that, in the explosion shot, the yellow-tone flames are spread into the whole frame rapidly and eventually hide almost all other objects in the frame. Upon these characteristics of the explosion shots, this paper proposes a scheme to abstract them, in which some low level video features such as the regional dominant colors of the key frame, the motion energy of shot, and the binary edge image differences between the neighboring frames in the shot are selected as the abstraction for the explosion shot retrievals. Also an algorithm to automatically retrieve a scene with explosion shots are proposed and experimented. Upon the experimental results, we could argue that the proposed abstraction and retrieval algorithm could find the explosion events from the digital video archives with about 80% recall and precision. Furthermore, its performance is robust to thresholds that are usually dependent on the contents of the movies.

The performance of the proposed scheme could be improved if we also use the audio information of the digital video since the explosion events usually accompany with a very loud sound. The proposed abstraction and automatic retrieval algorithm could be used to summarize the long movies because the explosion event would be the part of the highlight of the movies, and used to build a meta database which contains a high level semantic information.

Acknowledgments. This work was supported from the Korea Science & Engineering Foundation in 2002 (Project Title: *Contents-aware Media Streaming QoS Technologies for Wireless Communication Networks*).

References

[1] S. Fischer, “Automatic Violence Detection in Digital Movies,” *Proceeding of SPIE Multimedia Storage and Archiving Systems*, 1996, pp.212-223.

[2] R.C. Gonzalez, *Digital Image Processing*, Addison Wesley, 1993.

[3] http://www.informatik.uni-mannheim.de/informatik/pi4/projects/MoCA/ .

[4] S. Pfeiffer, R. Lienhart, S. Fischer and Wolfgang Effelsberg, "Abstracting Digital Movies Automatically,” *Journal of Visual Communication and Image Representation*, Vol.7, No.4, 1996, pp.345-353.

[5] M. Stricker and M. Orengo, "Similarity of Color Images,” Proceeding of *SPIE Conference on Storage and Retrieval for Image and Video Databases III*, Vol. 2670, 1996, pp. 381-391.