평균오차를 이용한 멀티미디어 동영상 데이터를 위한 효율적인 장면전환 검출

정창렬* · 고진광* · 이 준**

Effective Scene Change Detection Method for Multimedia Data as Video Images using Mean Squared Error

Chang-Ryul Jung* · Jin-Gwang Koh* · Joon Lee*

요 약

동영상데이터의 대용량 검색을 할 때 사용자가 원하는 검색점을 추출하여 인덱싱과 재생하기 위해서는 동영상 데이터의 프레임 리스트 제공이 필요하다. 평균오차를 적용하여 프레임의 대각선으로부터 픽셀값을 추출한다. 각각의 프레임으로부터 추출된 픽셀의 RGB값은 행렬구조로 저장하고, 저장된 프레임은 두 지점을 비교하여 장면전환점을 추출한다. 또한 동영상의 전체적인 구조를 파악할 수 있는 알고리즘과 장면전환점을 구현하였다. 본 연구에서는 제안한 방법의 성능 우수성을 기존연구와 비교하여 분석하고 증명했다.

ABSTRACT

When retrieving voluminous capacity of video image data, it is necessary to provide synopsized frame lists of video image data for indexing and replaying at the exact point where the user want to retrieve. We apply Mean Squared Error method to extract certain pixel value from diagonal direction of a frame. The RGB value of a pixel extracted from each frame is saved in a matrix form, and this frame is retrievedas a scene change point if the compared value of two points met the certain condition.

Also implement the algorithm and provide a way to seize entire structure of video image and the point of scene changes. Finally, we analyze and prove that our method has better performance compared with the others.

키워드

Vedeo Image Data, Scene Change Detection, Multimedia Data

Ⅰ. Introduction

As multimedia communication technology and popularization of internet growth, the demand of optimized multimedia data and processing environment is also increased.

Especially, the high quality indexing method of multimedia data is required to provide a quality of service.

In general, the video image retrieval system provides a frame list to retrieve exact scene change point for preparing a representative frame list. There are several scene change detection algorithms for video image data. The

* 순천대학교 공과대학 컴퓨터과학과

** 조선대학교 전자정보공과대학 컴퓨터공학부

접수일자 : 2002. 10. 19

Color-Pair algorithm represented by Nagasaka[2] used a histogram value of color component that has changed at inter-gap between image boundary.

Ham pa pur [6] and Shaharay[4] propose a model-based scene change detection method. Nagasaka's method, however, has problems on changed histogram threshold by distribution of image values. And the method represented by Hampapur and Shaharay takes much processing time and memory because of this method requires a compressed video image data to apply.

In the scene change detection method using Mean Squared Error(Mean Squared Error), however, user can grab the whole situation in a lump with voluminous capacity of video image data by sampling the diagonal points of image data from each frame. Using this method, the voluminous capacity of video image data can be transformed into a small portion of stillness image from extracting certain value of pixel from diagonal direction of each frame. And the RGB value of a pixel extracted from each frame is saved in m by n matrix A, and this frame is selected as a scene change point if the compared pixel value met certain threshold using Mean Squared Error. It makes easier to user to seize entire structure of video image and the point of scene changes, and we also analyze and prove that our method has better performance compared with the others.

II. Related work

2.1 Structure of video image data

The hierarchical structure of video image data is composed with four terms: frame, shot, scene, and cut. As shown in the fig. 1, a 'frame' is the smallest unit of video image data that is composed with many pixels in stillness form. The world 'shot' is the basic unit of video image data that is composed with a set of frames in sequence, and a 'scene' is a set of sequence of shots which is composed in significance of spatial and temporal segmentation. Finally, the 'cut' is a scene change point between shots or scenes.

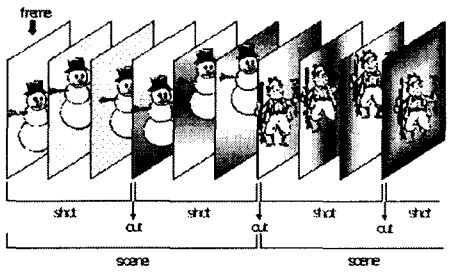


Fig. 1 Structure of a video image

2.2 Scene Change Detection Algorithm

In general, the scene change detection is divided into two categories: the gradual transition[9] and camera actions like zooming and panning although it is not a scene change. Also, the video segmentation is classified as non-compressed video stream and compressed video stream defense on the substitution of video image. The scene change detection method for non-compressed video stream is a pixel comparison[2]. The pixel comparison method has a vulnerable point on the change of some portions, and to improve this vulnerable point the luminosity histogram comparison method has been proposed[5][7]. The methods, however, can not be reflected into the content information as shape, color, or texture, since it depends on only luminosity value of an image.

For applying the content information of video image to scene change detection, the outlined pixel comparison method using edge retrieval filter is used[1]. Also from calculating the

motion vector from the block segmentation of a frame and comparing the differences with current frame, the motion continuity can be calculated by substituting the number of motion vectors[3].

(1) Scene change detection using intensity of pixel.

This method compares intensity of pixel from the same area of two frames, and detects changes between two frames. This algorithm counts the number of changed pixels and if the ratio of changed pixels exceeds the threshold t, the scene change detection is made and, thus, the scene change point is extracted[8].

Since this comparison method is sensitive to movement of camera and object's motion, it makes scene change detection very difficult. Because of a camera movement as pan and zoom causes lots of pixel changes, these changes are recognized as scene change. Also, the fast moving objects causes effectiveness of pixel changes. These scene changes caused from pixel changes can be reduced by comparing pixel value with the pixels around. The differences of pixel value between frames are used to detect scene change points. However, method is sensitive to noise and it has weakness point on gradual scene changes.

(2) Scene change detection using luminosity histogram.

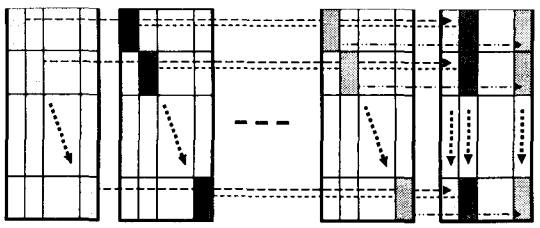
Histogram represents number of pixels within the pixel's luminosity value in a frame, and the difference of pixel's luminosity value in a frame is compared. Then the scene change point can be detected if the difference exceeds some threshold. This method is less sensitive than using differences of pixel value for noise. However, it also has some weakness point on gradual scene changes, and this method is less effective than color or gray histogram comparison method. Nagasaka and Tanaka[2] use a square root for definite difference between two histograms.

Sensitivity of camera and moving objects are reduced by comparing luminosity histogram of two frames by enlarging the differences occurred from small changes caused by camera and objects movement[2][7]. The coordinate of color histogram is more stable because it doesn't need to change the color value to gray level.

III. The Mean Squared Error Retrieval Syste

3.1 Domain sampling of video image data

Generally, the video image is constructed with sequence of frames that has similar background and images. Thus, the retrieval point of video image is a point where the continuity of video image is broken, and that image is used as a scene change point. The main focus of this paper is to retrieve the scene change point that appears gradually. For this purpose, the video image is expressed as a m by n matrix A(where, m = height of image, n = number of frames). The columns of matrix A is a set of sampling of diagonal elements of each frame. From matrix A, a_{ij} and $a_{i(j+1)}$ are compared and the column is chosen for a scene change if the difference is greater. The fig. 2 shows the relationship between a frame sampling and a matrix expression.



frame 1 frame 2 frame n Matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \Lambda & a_{1n} \\ a_{21} & a_{22} & a_{23} & \Lambda & a_{2n} \\ a_{31} & a_{32} & a_{33} & \Lambda & a_{3n} \\ M & M & M & M & M \\ a_{m1} & a_{m2} & a_{m3} & \Lambda & a_{mn} \end{bmatrix}$$

Fig. 2 Frame sampling and a matrix expression

The data retrieved from each frame of video image is sampled as small amount of stillness images, and the sampling result is shown in fig. 3. From the sampling image, the scene change points are distinguished as shown in vertical lines.

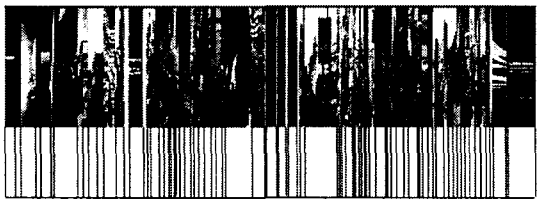


Fig. 3 Sampling image from diagonal pixel value.

The size of this stillness images are reduced by sampling the interval of many frames. Also, it provides a way to seize entire structure of video image and the point of scene changes.

3.2 Sampling the Mean Squared Error of vide image data

The values extracted from the diagonals of the each frame in video image are the scene change point where the difference of the column a_i and a_{i+1} of the matrix A met the certain value. To define the difference between column a_{ij} and $a_{i(j+1)}$ of matrix A, we use mean square error.

For a general expression, from the matrix A shown in fig. 2,the mean square error of kth and $k+1$ th columns is expressed as E_k as shown in equation 1.

```
for (j=1 to frame x axis value)
 for ( k=1 to frame y axis value) {
 store A[i][j]=image[j][i]; // diagonal data value of i frame;
 image [i][j]=image[j][i]; // diagonal data sampling of i - frame;
}
; evaluating Mean Squred Error
for (f=1 to frame_end) {
  for (g=1 to j) {hap=hap + ( store[g][f]  storeA[g][f+1])^2 );
  MSE_Tab[f] = SQRT(hap); } // Store MSE of Ef;
  hap=0 );
; Comparing with threshold
 for (f=1 to frame_end) {
  Chang[f]=MSE_Tab[f] MSE_Tab[f+1]; // error of each frames
  if(Chang[f] >= threshold)
   Chang_Image[] = f; // save video image data
   K=k+1 );
```

$$E_1 = \sqrt{(a_{11}-a_{12})^2+(a_{21}-a_{22})^2+\Lambda+(a_{m1}-a_{m2})^2}$$

$$E_k = \sqrt{(a_{1k}-a_{1(k+1)})^2+(a_{2k}-a_{2(k+1)})^2+\Lambda+(a_{mk}-a_{m(k+1)})^2}$$

-------- (1)

From the E_k calculated using Mean Squared Error, the scene change point k can be found where the condition shown in equation 2 satisfies.

$$|E_k - E_{k+1}| \geq t \quad \text{---------- (2)}$$

In equation (2), t is a threshold for detecting scene change point and E_k is the frame number where the scene change occurred. Especially, the Mean Squared Error method provides great performance and it also responses on the gradual scene changes sensitively by using normalized threshold.

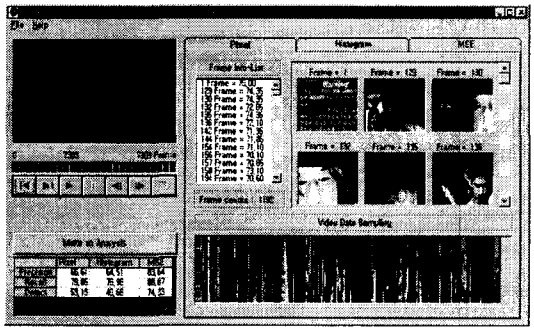


Fig. 4 Result of scene change detection for a Music video

Table 1. Analytical result of scene change detection

	Pixel		Histogram		Mean Squared Error	
	Precision	Recall	Precision	Recall	Precision	Recall
Movie	82.95%	83.89%	88.50%	83.93%	92.23%	97.89%
Drama	82.85%	60.07%	88.92%	76.99%	95.23%	81.83%
News	80.60%	89.22%	88.77%	77.02%	91.83%	97.23%
Advertisement	63.13%	81.72%	61.04%	77.61%	84.21%	88.72%
Music Video	66.61%	79.85%	64.51%	76.98%	83.64%	88.84%

As a practical example, the fig. 4 show the result of scene change detection screen that shows total detected number of frames. From fig. 4, 1059 frames are detected out of 1192 total frames in a part of a movie by using Mean Squared Error.

For the same video image, 917 frames and 952 frames are detected by using histogram and pixel value respectively. This results 79.85%, 76.98%, and 88.84% of recall. Recall: detected scene change points out of total scene change points. for Mean Squared Error, histogram, and pixel value respectively.

3.3 Analytical result of experimental tests

There are many kinds of video images around the world. We specially consider the video images as movie, drama, news and advertisement in this paper. Especially the advertisement video image is a good sample of video image data for testing the performance of detecting method since the advertisement video image often uses special effectiveness. The news has characteristics on fast scene changes. The movie and drama video image data has characteristics on sensitivities and accuracy of image data. In this paper, therefore, four different data as advertisement, news, movie and drama are tested under the same environmental condition. In table 1. we represent the result of comparison with precision and recall Recall: accuracy of scene change point detection. Filtering detected scene change points to find out the number of correct scene change points of four different types of video image data.

In fig. 5 and 6, we also represent the graphical expression of the result of precision and recall respectively. From fig. 6, the precision calculated by using Mean Squared

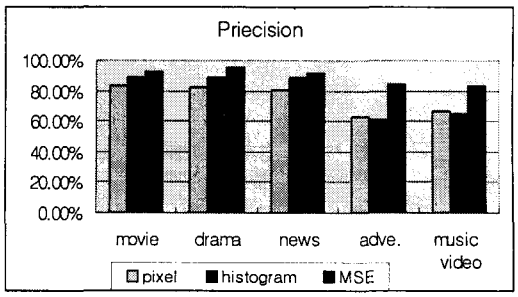


Fig. 5 Comparison result of precision

error results 5% to 10% of more detections are made. In the other word, using the Mean Squared Error method, there will be 5%~10 % of better detection is occurred.

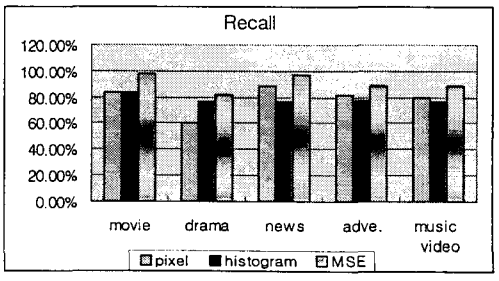


Fig. 6 Comparison result of recall

From fig. 6, the scene change point detection using Mean Squared Error method produces about 6% to 20% of better detection ratio compared with the other methods as pixel value and histogram methods.

V. Conclusion

In general, video image retrieval system requires repetition of retrieval for searching specific frames on the video image data. However, the Mean Squared Error method can change the video image as stillness image from extracting certain value of pixel from diagonal direction of a frame, and it can also provide a way to seize entire structure of video image and the point of scene changes. The color value of a pixel extracted from each frame is saved as a m by n matrix A (where, m = height of image, n = number of frames), and this frame is marked as a scene change point if the value dose not reach certain threshold using Mean Squared Error.

Four different types of video images are employed to prove the performance of Mean Squared Error method. Comparing movie, news, drama and advertisement with three different scene change detection method including Mean Squared Error results the ratio of 5% to 10 % better precision detection and 6% to 20% better recall detection is made by using Mean Squared Error method. Especially, the pixel value method and the histogram method has some vulnerable points on gradual scene change, however, the Mean Squared Error method provides great performance on the gradual scene changes. Therefore, the Mean Squared Error method can be applied into the video image data of having many gradual scene changes like news and movie.

Reference

[1] A. Akutsu et al, "Video Indexing using motion vectors" Visual Communications and Image Processing '92, pp. 1522-1530, Boston, MA, November 1992, SPIE.

[2] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full Motion Search for Object Appearance", Visual Database System, II, Eds, E. Knuth, and L.M Wegner, Elsevier Science Publishers B..V. IFIP, pp. 113-127, 1999.

[3] B.-L Yeo and B. Liu, "Rapid Scene Analysis on Compressed Video", IEEE Tr. on CSVT, Vol. 5, No. 6, pp. 533-544, Dec. 1995.

[4] B. Shahraray, "Scene Change Detection and Content-Based Sampling of Video Sequences", Proc. IS&T/SPIE, Vol. 2419, CA, February 1995.

[5] C.L. Fennema, and W.B. Thompson, "Velocity Determination in Scene Containing Several Moving Objects", Computer Graphics and Image Processing, Vol. 9, No. 4, 1979, pp. 310-315.

[6] Hampapur, A., Jain, R., and Weymouth, T., "Digital Video Segmentation", Proc. ACM Multimedia 94, San Francisco, CA, pp.357-364, October, 1994,.

[7] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic Partitioning of Full-Motion Video", ACM/Springer Multimedia Systems, Vol. 1, No. 1, 1993, pp. 10-28.

[8] S. Berretti, A. Del Bimbo, P. Pala, "Sensations and Psychological Effects in Color Image Database", IEEE International Conference on Images Processing, pp.560-563, 1997.

[9] W. Niblack et al. The Qbic project Querying Images, by content using color, texture, and shape. In SPIE 1908, Storage and Retrieval for Image and Video Databases, Feb. 1993.

[10] Gonzalez and Woods, Digital Image Processing, 2nd edition, Prentice Hall, 2001.

[11] William K. Pratt, Digital Image Processing, 3rd Edition, John Wiley, 2001.

저자소개

정창렬(Chang-Ryul Jung)

1995년 광주대학교 전자계산학과 졸업(학사)

1999년 순천대학교 교육대학원 컴퓨터교육학과 졸업(석사)

2001년 순천대학교 대학원 컴퓨터과학과 박사수료

※관심분야: Information Security, Image Processing, Mobile Agent, Authentication,

고진광(Jin-Gwang Koh)

1982년 홍익대학교 컴퓨터공학과 졸업(학사)

1984년 홍익대학교 대학원 컴퓨터공학과 졸업(석사)

1997년 홍익대학교 대학원 컴퓨터공학과 졸업(박사)

1984. 3.~1998. 2. 송원전문대학 전자계산과 전임강사

1988. 3.~현재 순천대학교공과대학 정보통신공학부 정교수

1997. 8.~1998. 2. 미국 Oregon State Univ. 컴퓨터공학과 방문교수

2000. 12.~2001. 2. 일본 류큐대학 정보공학과 방문교수

2001. 3.~현재 순천대학교 정보전산원장

※관심분야: Database, Electronic Commerce Image processing, EC Security

이 준(Joon-Lee)

1976년 조선대학교 정자공학과(공학사)

1981년 조선대학교 대학원 전자공학과(공학석사)

1997년 숭실대학교 대학원 전자계산학과(공학박사)

1982년~현재 조선대학교 전자정보공과대학 컴퓨터공학부 교수

※관심분야: 분산 운영체제, 정보보호, 병렬처리, 프로그래밍 관련