International Journal of Scientific Engineering and Technology(ISSN:Applied)
Volume No.1,Issue No.1 pg:23-30 Dec.2011

Key Frame Extraction Based on Block based Histogram Difference and Edge Matching Rate

Kintu Patel Oriental ***Institute of Science & Technology, Bhopal , MP, India***

Kintu125@gmail.com

Abstract— This paper presents a new approach for key frame extraction based on the block based Histogram difference and edge matching rate. Firstly, the Histogram difference of every frame is calculated, and then the edges of the candidate key frames are extracted by Prewitt operator. At last, the paper makes the edges of adjacent frames match. If the edge matching rate is up to 50%, the current frame is deemed to the redundant key frame and should be discarded. The experimental results show that the method proposed in this paper is accurate and effective for key frame extraction, and the extracted key frames can be a good representative of the main content of the given video.
Keywords- Block based Histogram Difference, edge matching rate; key frame extraction; Threshold Point.

I. INTRODUCTION

With the development of multimedia information technology, the content and the expression form of the ideas are increasingly complicated. How to effectively organize and retrieve the video data has become the emphasis of the study. The technology of the key frame extraction is a basis for video retrieval. The key frame which is also known as the representation frame represents the main content of the video. Using key frames to browse and query the video data greatly reduces the amount of processing data. Moreover, key frames provide an organizational framework for video retrieval. In general, the key frame extraction follows the principle that **[1]** the quantity is more important than the quality and removes redundant frames in the event that the representative features are unspecific. Currently, key frame extraction algorithms **[2]** can be categorized into following four classes: 1. Content based approach. Key frames are extracted according to the changes of the color, the brightness, the texture and other information between adjacent frames. 2. Unsupervised clustering based approach **[3]**, the method firstly calculates the distance between the current frame and the center of each cluster which has already existed, then compares the distance with pre-set threshold and assigns the current frame to its nearest cluster. Finally, the frames with the minimum distance from the center of their respective clusters are taken as key frames. 3. Motion based approach **[4]**. This method calculates the amount of exercise in the lenses and selects the frames whose amount of exercise takes the local minimum value as key frames. 4. Compressed video streams based approach, key frames are extracted with the DC coefficients and motion vectors in the MPEG compressed video streams. In order to overcome the shortcomings of the above algorithms, this paper proposes a new approach for key frame extraction based on the image Histogram difference and edge matching rate, which calculates the histogram difference of two consecutive frames, then chooses the current frame as a candidate key frame whose histogram difference is above the threshold point, finally matches the edges between adjacent candidate frames to eliminate redundant frames. The experimental results show that the key frames extracted by the method reflect the main content of the video and the method is good approach to determine key frames.

II. KEY FRAME EXTRACTION

The method for key frame extraction consists of three steps:

Input a video and calculate the block based histogram difference of each consecutive frame.

Choose the current frame as a candidate key frame whose histogram difference is above the threshold point.

Extract the edges of the candidate key frames and calculate the edge matching rate of adjacent frames. If the edge matching rate is up to 50%, the current frame is considered as a redundant frame and should be eliminated from the candidate key frames.

Procedure of the key frame extraction is shown in Fig.1.

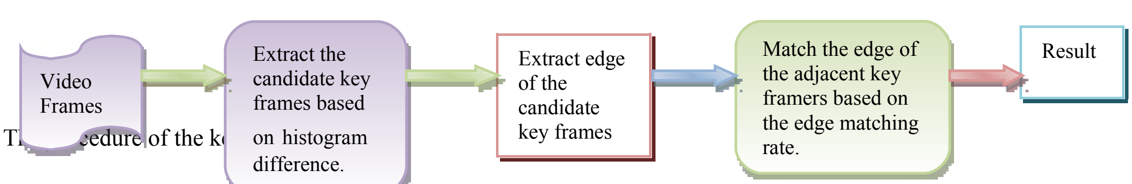

Figure 1. T cedure of the k

A) Make use of the block based histogram difference to Extract Candidate Key Frames

To extract robust frame difference from consecutive frames, we used verified x^2 test which shows good performance comparing existing histogram based algorithm and to increase detection effect of color value subdivision work, color histogram comparison using the weight of brightness grade. Also to reduce the loss of spatial information and to solve the problem for two different frames to have similar histogram, we used local histogram comparison. Color histogram comparison ($d_{r,g,b}$ (fi,fj)) is calculated by histogram comparison of each color space of adjacent two frame (fi,fj) and it is defined as Equation 1.1. [**5**]

$$d_{r,g,b}(f_i, f_j) = \left(\left|H_i^r(k) - H_j^r(k)\right| + \left|H_i^g(k) - H_j^g(k)\right| + \left|H_i^b(k) - H_j^b(k)\right|\right) \quad \ldots(1.1)$$

Where $H_i^r(k), H_i^g(k), H_i^b(k)$ represent the number (N) of bean (k) of each color space (r,g,b) in frame fi.. ssssss

Using the weight for brightness grade change of each color space from (1), we can redefine it as Equation 1.2.

$$d_{r,g,b}(f_i, f_j) = \left(\left|H_i^r(k) - H_j^r(k)\right| \times \alpha + \left|H_i^g(k) - H_j^g(k)\right| \times \beta + \left|H_i^b(k) - H_j^b(k)\right| \times \gamma\right) \quad \ldots(1.2)$$

α ,β ,γ shows the constants to change the brightness grade according to NTSC standard and it is defined as α = 0.299, β = 0.587, γ = 0.114.

Among static analysis method for emphasizing the difference of two frames, X^2 test comparison (d_{wx^2}(fi , fj)) is efficient method to detect Candidate key frames by comparison change of the histogram and it is defined as Equation 1.3.

$$d_{wx^2}(f_i, f_j) = \begin{cases} \sum \dfrac{\left(H_i(k) - H_j(k)\right)^2}{\max\left(H_i(k) - H_j(k)\right)} & , if(H_{i,j} \neq 0) \\ 0 & , otherwise \end{cases} \quad \ldots(1.3)$$

The histogram based method may have a problem to detect two different with similar color distribution as same image as it doesn't use the spatial information. This problem can be solved by the method of comparing local histogram distribution as dividing frame area. The value of frame difference through color histogram comparison of each area according to the area division and its accumulation is given by Equation 1.4.

$$d(f_i, f_j) = \sum_{bl=1}^{m} DP(f_i, f_j, bl) \quad \text{Where} \quad DP(f_i, f_j, bl) = \sum_{k=1}^{N} \left|H_i(k, bl) - H_j(k, bl)\right| \quad \ldots(1.4)$$

H_i (k,bl) is the histogram distribution of k position of the frame (f_i) block(bl) and m is the number of total blocks. Using the merits of subdivided local histogram comparison applying weight to each color space in above Equation (1.2), value of difference expansion using statistical method of Equation (1.3) and use of spatial information of the frame by local histogram as Equation (1.4), in this report, The value of difference extraction formula, is given in Equation (1.5) by combining above formulas, will be used for robustness of value of difference extraction.

$$d(f_i, f_j) = \sum_{bl=1}^{m} D_{x^2}(f_i, f_j, bl)$$

$$D_{x^2}(f_i, f_j, bl) = \sum_{k=1}^{N} \left[\frac{\left(H_i^r(k) - H_j^r(k)\right)^2}{\max\left(H_i^r(k), H_j^r(k)\right)} \times \alpha\right] + \left[\frac{\left(H_i^g(k) - H_j^g(k)\right)^2}{\max\left(H_i^g(k), H_j^g(k)\right)} \times \beta\right] + \left[\frac{\left(H_i^b(k) - H_j^b(k)\right)^2}{\max\left(H_i^b(k), H_j^b(k)\right)} \times \gamma\right] \quad \ldots(1.5)$$

In above formula, H_i^r (k), H_i^g(k), and H_i^b(k) is histogram distribution of each space *r, g, b* owned by number *i* frame *where*, *N* is total number of bean of histogram and *m* is the total number of the blocks . Here, the value of difference was created from Equation (1.5) by histogram comparison of each block after dividing the frame into same block areas.

Histogram based comparison methods are highly preferred because they are robust to detrimental effects such as camera and object motion and changes in scale and rotation. However, such methods sometimes fail to identify changes between shots having similar color content or intensity distribution. On the other hand, pixel-wise comparison methods can well identify changes between shots having a similar color content or intensity distribution, but they are very sensitive to movements of cameras or objects. Since the adopted pixel difference feature is extracted from DC images, it becomes less sensitive to small object and camera motions. However, it still is not enough for reliable shot change detection.

The main assumption for candidate key frame detection is as follows:

Within a single shot, inter-frame variations are small, which results in a slowly varying feature signal.

However, an abrupt change in histogram difference causes a sharp peak in a feature signal.

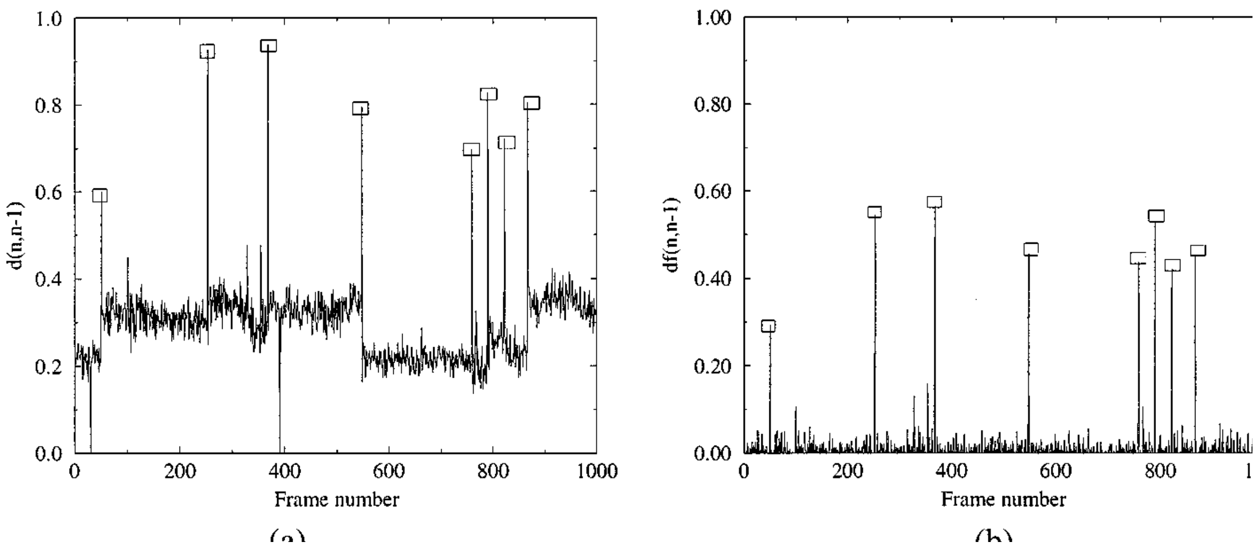

Figure 2. (a) Original 1-D frame difference signal $d(n,n-1)$ and
(b) Its filtered signal $d_f(n, n-1)$. Boxes point out the cuts.

So we can detect candidate key frames by recognizing these peaks. However, the sensitivity of these features to camera motion, object motion, and other noises strongly influences detection performance. In order to remove this phenomenon, a filtering scheme to reduce feature signal values at high activity regions while minimizing effects on those at actual shot changes, is needed . In this report, we choose an unsharp masking technique, i.e.

$$d_f(n,n-1) = \begin{cases} d(n,n-1) - \tilde{d}(n,n-1) & if\ d(n,n-1) > \tilde{d}(n,n-1) \\ 0 & otherwise \end{cases} \quad ...(1.6)$$

Here, the 1-D frame difference signal $d(n; n-1)$ can either be $d^h(n; n-1)$ or $d^p(n; n-1)$. $\tilde{d}(n; n-1)$ denotes the low-pass filtering and/or median filtering result of $d(n; n-1)$, and $d_f(n; n-1)$ denotes the unsharp masking output, respectively. After sequentially applying unsharp masking to both histogram difference and pixel difference features, we obtain the filtered signal $d_f(n,n-1)$ as shown in Figure 2.

Flow chart for detecting the candidate key frames from video is shown in Figure 3. first of all the original video is converted in to frames. Then each frame is decomposed in to three components (Red, Green and Blue images).then each component of every frame is divided in to nine blocks. Now histogram of each block is computed and finds the histogram difference between successive frames block by block wise using equation 5.if histogram difference is greater than threshold value candidate key frame will be detected.

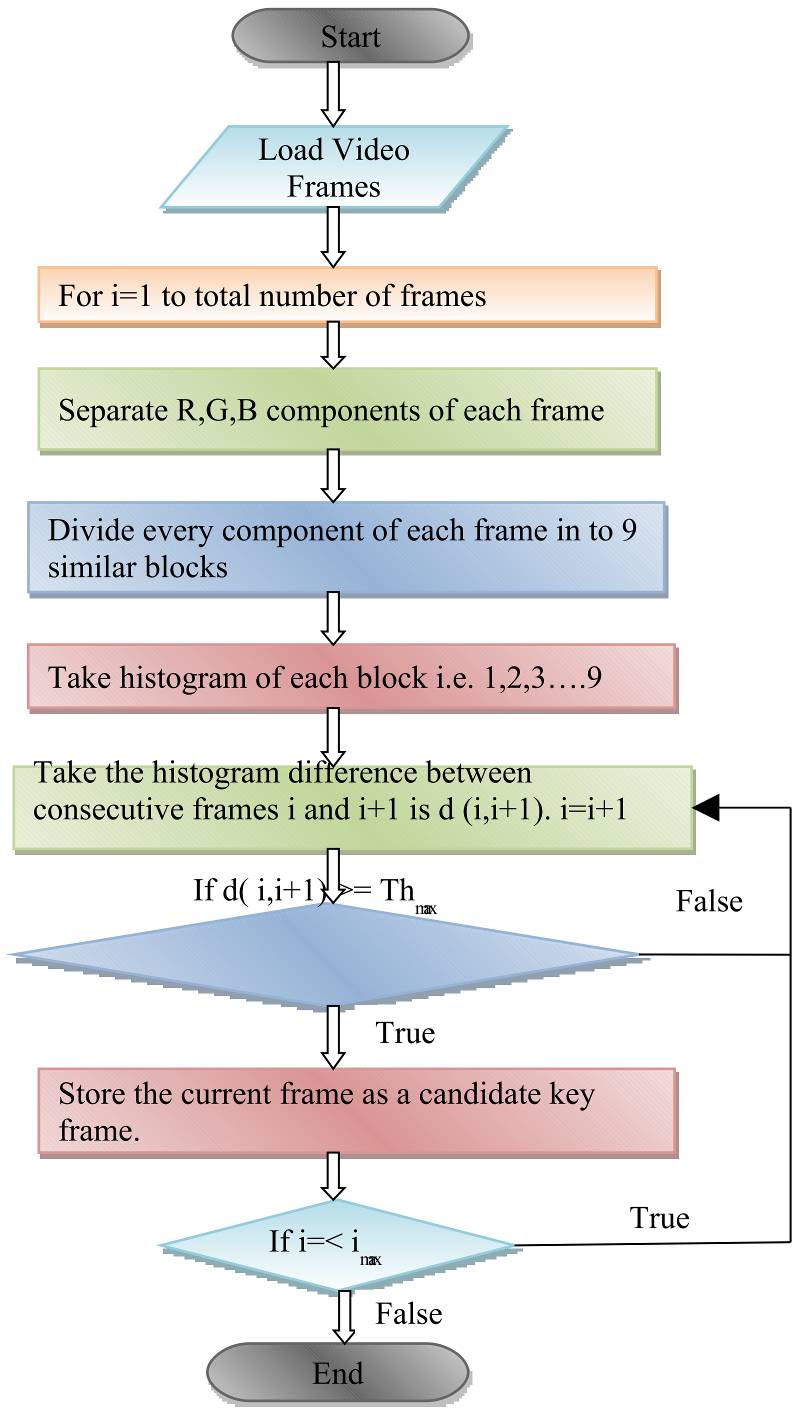
Start
Load Video Frames
For i=1 to total number of frames
Separate R,G,B components of each frame
Divide every component of each frame in to 9 similar blocks
Take histogram of each block i.e. 1,2,3….9
Take the histogram difference between consecutive frames i and i+1 is d (i,i+1). i=i+1
If d(i,i+1) >= Th max
False
True
Store the current frame as a candidate key frame.
If i=< i max
True
False
End

Figure 3. Flow chart for key frame detection in video transition.

B) Extract Edges of the Candidate Key Frames

The candidate key frames obtained from the above treatment do well in reflecting the main content of the given video, but exist a small amount of redundancy, which need further processing to eliminate redundancy. As the candidate key frames are mainly based on the Histogram difference which depends on the distribution of the pixel gray value in the image space, there may cause redundancy in the event that two images whose content are the same exist great difference from the distribution of the pixel gray value. For example, the substance content of images a and b in Fig.4 don 't change, but the two images are both identified as key frames as a result of the different gray value distribution, resulting in redundancy.

Figure 4. original images a and b respectively.

$$p(f_i, f_{i+1}) = s/n$$

In the formula,

$$n = max\,(n_{f_i}, n_{f_{i+1}})$$

$$s = \sum_{i}^{m}\sum_{j}^{n} h(i,j)$$

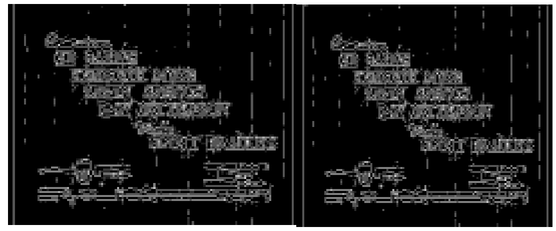

Figure 5. Edge Images of A and B.

As edge detection can remove the irrelevant information and retain important structural properties of the image, we can extract the edges of objects in the image to eliminate redundancy. At present, there are many edge detection algorithms, which are mainly based on the differentiation and combined with the template to extract edges of images. Edge detection operators that are commonly used are: Roberts [6] operator, Sobel operator, Prewitt operator and the Laplace operator etc. Here we extract edges of frames by Prewitt operator.

C) Eliminate Redundant Frames Based on the Edge Matching Rate

The edge images have no difference in the distribution of the gray value. For example, the images shown in Fig.5 are the edge images of the images shown in Fig.4 with Prewitt operator and both of them are remarkably similar. So we use the edge matching rate to match the edges of adjacent frames to eliminate redundant frames. The formula for calculating the edge matching rate is as follows:

$$h(i,j) = \begin{cases} 1, v_{fk}(i,j) = v_{fk+1}(i,j) \\ 0, \quad otherwise \end{cases} \qquad (1.7)$$

$where\ v_{fk}(i,j)$ and $v_{fk+1}(i,j)$ are the pixel values of the position (i, j) in the frame f_k and the frame f_{k+1}, respectively. *m* and *n* indicate the height and the width of the image, n_{fi} and n_{fi+1} represent the number of the pixels on the edge of the frame f_i and the frame f_{i+1} respectively. Assume the key frame sequence as $\{f_1, f_2, f_3, \ldots\ldots f_k\}$ (the total number of the candidate key frames is k), we make use of the following steps to eliminate redundant frames:

1). Use the Prewitt operator to extract edges of the candidate key frames and obtain their corresponding edge images.

2). Set j=2.

3). Calculate the edge matching rate $p(f_{j-1}, f_j)$ between the current frame f_j and the previous frame f_{j-1} with the formula (1.7). If $p(f_{j-1}, f_j)$ is up to 50%, the current frame f_j will be marked as a redundant frame.

4). j=j+1, if j>k, go to (5). Otherwise, return to (3) and continue processing the remaining frames.

5). Remove the frames which have been marked as redundant frames from the candidate key frames. The remaining candidate key frames are the ultimate key frames. With the edge detection and edge matching, we eliminate redundant key frames, improve the accuracy rate of the key frame extraction and reduce the redundancy.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A NBA video is used in experiment. The video actually contains 11 shots through analysis. With our algorithm the first frame of the sub-shot and key-frame extracted are shown in Table 1. As can be seen from Table I, the more fiercely the shot changes, the more sub-shot and key-frame will be extracted. On the contrary, the less fiercely the shot changes, the fewer sub-shot and key-frame will be obtained. The number of key frames extracted is closely related to the intensity of changes, but has nothing to do with the length of the shot.

Shot	First Frame of Sub-shot	Length of sub -shot	Key Frames
Shot1(1-55)	1,24,41	23,18,12	23,28,52
Shot 2(97-369)	97,142,199,313	46,58,115,57	139,168,236,336
Shot 3(370-437)	370	37	401
Shot 4(438-711)	438,581,643,657,681	144,63,15,25,30	563,639,647,660,691
Shot 5(723-916)	723,762,788, 813,836,855,869	40,27,26,24, 20,15	747,774,803, 825,847,860,878
Shot 6(917-997)	917	80	964
Shot 7(998-1324)	998,1050,1086, 1143,1159,1178, 1199,1251	53,37,58,17, 19,22,53,76	1030,1081,1088, 1155,1166,1190, 1204,1261
Shot 8(1325-1501)	1325,1398	74,104	1376,1473
Shot 9(l557-1912)	1557,1712,1766,1791,1833,1851, 1912	176,55,26, 43,19,62	1710,1745,1788, 1799,1825,1847, 1854
Shot 10(1913-1951)	1913	39	1945
Shot 11(1952-1979)	1952,1961	10,19	1954,1964

Table I Data of Key Frame Extraction

Fig.6 (a) is the key-frames extracted from the NBA basketball game using this algorithm. Fig.6 (b) shows the content of the video by extracting the first frame of each shot. Fig.6(c) indicates the content of the video by extracting the final frame of every shot. The key-frame sequence of Fig. 6 (a) gives a more comprehensive showcase for the whole video content than Fig.6(b) and Fig.6(c). The first frames and last frames in Fig. 6(b) and Fig.6(c) is not a good representation of the content of the entire shot, while the key-frames extracted in Fig.6(a) highlight more important details of the contest video. For example, five key-frames and eight key-frames are extracted for the tight defense and fast attack plot respectively, which are a complete representation of the main content of the shot.

(a)

(b)

(c)

Figure 6. Key-frame extraction results

Key-frame extraction algorithm based on histogram difference and edge matching can better extract key-frames of the shot, as shown in Fig.7. Fig.7(a) is shot I which starts with frame I while ending with frame 55, and includes 55 frames in total. Three key-frames extracted indicate the special editing effects of the beginning of the whole video. Fig.7(b) is shot 4 which starts with frame 474 while ending with frame 704, and includes 231 frames in total. Five key-frames extracted better describe the defensive situation of the basketball games. Fig.7(c) is shot 7 which starts with frame 1024 while ending with frame 1289, and includes 266 frames in total. Eight key frames extracted completely show the entire process of the counter-attack after a steal. The results show that our key frame extraction algorithm determines the number of key frames according to the content within a shot, that is to extract frames with a maximum histogram difference value from the sub-shot as the key-frame, which overcomes the shortcomings of the traditional key-frame extraction method that the number of each shot key-frame is fixed or the key-frame cannot express video content very well.

(a)

(b)

(c)

Figure 7. Key-frame extraction of different shots

IV. CONCLUSION

The paper presents a method based on the image histogram difference and edge matching rate for key frames extraction. The approach avoids shot segmentation and has a good adaptability. The entire experimental results show that the algorithm has the high accuracy to detect key frames and the extracted key frames represent the main content of the given video. Meanwhile, the algorithm provides a good basis for video retrieval. However, in the condition of the complicated background of objects ’ motion there exists errors in the extracted edges of objects, leading to a certain redundancy, and further study is needed.

REFERENCES

[1] Yingying Zhu and Dongru Zhou, “An Approach of Key Frame Extraction from MPEG Compressed Video, ” Computer Engineering, Vol. 30, pp.12-13, 2004.

[2] Min Zhi, “Key Frame Extraction from Scenery Video, ” IEEE International Conference on Wavelet Analysis and Pattern Recognition, 2007.

[3] Jiawei Rong, Wanjun Jin and Lide Wu, “Key Frame Extraction Using Inter-Shot Information,” IEEE International Conference on Multimedia and Expo, 2004 .

[4] A. Divakaran, R. Radhakrishnan and K.A. Peker, “Motion activity based extraction of key-frames from video shots, ” Proc. of IEEE ICIP, Vol.1, pp.932-935, September 2002.

[5] Priyadarshinee Adhikari, Neeta Gargote, Jyothi Digge, and B.G. Hogade, Abrupt Scene Change Detection, World Academy of Science, Engineering and Technology 42 2008.