Video Summarization for Copyright Protection

Basma Mustafa M. H.*, Suhad A. Ali, Majid Jabbar Jawad
The University of Babylon, College of Science for Women, Computer Science Dept.

bsma.myh@gmail.com

Abstract. Nowadays, a summarization of videos for copyright application is an interested research area in computer vision. Key frames extraction is crucial phase in video summarization. This work proposed a method for key frames extraction. The method consists of two operations for covering the content of the whole video sequence. Firstly, a set of candidates key frames are extracted based on adaptive thresholding of difference of histogram of adjacent frames of video data. Secondly, the extraction of a final key frames is done depending on computing Euclidian distance between two adjacent frames in candidates key frames set. The proposed method was tested with a set of standard and online television videos. The experimental results show the efficiency of the proposed system.

1. Introduction

Rapidly development in information technologies and networks make a high revolution in the digital era. Currently, a people are highly depending on network technology. Accordingly, the digital video content is shared rapidly on the unsecure networks. The sharing of digital video opened a new challenge in terms of copyright violation detection. Therefore, video copyright protection field has attracted more attention. Digital watermarking technique is one of some techniques can be used for solving the above challenge [1].

The extracted key frame from video is very important feature which represents a summary of video sequences.

In general, a key frame extraction technique can be classified four methods namely, shot boundary, Content Analysis, Motion Analysis, and clustering [8]. In shot boundary method, two approaches can be used which frame average and histogram average. In content analysis method, key frames are extracted depending on the several features namely color, texture and other valuable information related to each frame in video, where each frame in the video which has information that changed significantly are chose a key frames. In the motion analysis method, the extracting of key frames depends on movement in the video shots can easily be detected or analyzed by analyzing the optical flow of the video sequence. Finally, a local minimum in the movement is considered a key frame. In cluster method, three steps are being done. All the cluster center is initialized in the first one, where the current frame is classified as the class of this kind or as a new class by computing the distance between the current frame and the cluster center in the second step. Finally, all the frames which are close to the cluster center is considered as the key frame [2].

The rest of the paper is organized in four sections. The first one introduces the related works. The second one illustrates the proposed method. The third one explains the experimental results and tests related to the proposed system. Finally, the conclusions are given in the last section.
2. Related Works

Many schemes related to key frame extraction has been proposed. Herein, some of these schemes which can be illustrated as follows:

In 2012, Naveed Ejaz [3] presented adaptive key frame extraction scheme for purpose of video summarization. The aggregation mechanism is used in this scheme. In this scheme a technique called Video Summarization Using Key Frame Extraction (VSUKFE). The VSUKFE uses inter-frame differences calculated depending on the correlation of RGB color, color histogram and moments of inertia. These features are used for the automated generation of video summaries. In this method, visual features extracted from the correlation of RGB color channels, color histogram, and moments of inertia is combined for extracting a key frames from the video. The combination operation is done by using aggregation mechanism. Experimental results show that the proped work generates summaries same as summaries created by humans.

In 2012, S. T. Dhagdi and P.R. Deshmukh [4] propsed a key frame based video summarization depending on the uutomatic threshold and edge matching rate method. In this method, the shot boundary of video is detected for segmenting thecamera shots. The suggested technique based on the block based difference of histogram difference and rate of edge matching. The histogram difference of every frame is calculated firstly. By Prewitt operator, the edges of the candidate key frames are extracted. Then, the edges of adjacent frames are matched. When the edge matching rate is laeger than average edge matching rate, the current frame is considered as redundant key frame and should be discarded.

In 2014, Shaoshuai Lei et al. [5] proposed extracting of a key frame scheme for purpose of video summary and video index. Firstly, a dynamic distance separability algorithm is advanced for dividing a shot into subshots based on semantic structure. Then, by using SVD decomposition, appropriate keyframes are extracted in each subshot. Experimental illustrates results, the proposed work achieves good semantic structure for semantics-based video index and meanwhile produces video summary consistent with human perception.

In 2015, an extracting of Key frame scheme by analysis of histograms of video frames depending on the statistical methods is presented by Sheena C V and N. K. arayanan [6]. In this work, a key frame extraction using thresholding of absolute difference of histogram of consecutive frames of video data for the purpose video summarization. Measurements such as compression ratio and fidelity value are used for evaluating the performance of the proposed method.

In 2015, Lei Pan, Xin Shu and Ming Zhang proposed a key frame extraction algorithm depending on unsupervised clustering and compressive sensing is proposed [7]. This method consists of four steps. In the first step a high dimensional feature is constructed based on the three types of filters. In the second step, a low dimensional feature is generated based on spare transition matrix. A sub-shot segmentation is done by using unsupervised clustering. Finally, a key frame is extracted by using bhattacharyya coefficient. The experimental results illustrate that the proposed work could extract key frames efficiently and properly.

In 2016, Hana Gharbi et al. [8] proposed a scheme for extracting a key frame. The interest points and a points matching method is based for extracting operation, where this method also based on local description around each interest point and spatial constraints coupled with geometric invariants. A shot detection is the first phase if key frame extraction. The histogram matching is applying in shot detection by applying in the first phase. In order to extrat the key frame, three steps must be done with shot. The first one is applying the SIFT detector to extract interest points for all images in the shot. The second one is building a repeatability table for describing the repeatability inter-frame. In the third one, the Principal Component Analysis PCA and the Hierarchical Ascendant Clustering algorithm HAC is applied on the repeatability table. The PCA and HAC is used for extracting the similar images and the corresponding key frames groups. The redundancy in the extracted key is minimized while the rest of image cover cover all the significant events (this illustarted in the experimental results).

In 2017, for video copyright protection, a fast and robust key frame extraction method is presented by Yunyu Shi et al. [9]. The extraction of key frames operation is done depending on difference of frame philoshopy. In order to cover the all events of the video sequence, a method called two-stage is done for extracting the perfect key frames. In the first, alternative sequence is found depending on the different characteristic of color between adjacent frames from original sequence. In the second,
analyzing between adjacent frames from the alternative sequence for obtaining the final key frames. At the end, number of final key frames is added supporting the effectiveness for video copyright protection purpose. The experimental results illustrate that presented work reduces the computation complexity and increases the robustness on numerous video formats and video resolution.

3. Proposed Method
According to [8] there are three requirements must be fulfilled for extracting a key frame from video for protection of copyright of video purpose. These requirements are:

1. **Key frame gray value range:** it means that the gray levels of key frames should be in certain range to enable the viewers to recognize frame content.

2. **Arrangement of key frames:** it means that the sequence of final key frames must be arranged in chronological order consistent with original video sequence.

3. **Key frames redundancy:** in copyright protection purpose, the redundancy of key frames is acceptable. It means that the same frame may be selected as key frames if it found in different shots.

The three conditions above must take in our consideration when design the proposed method.

The proposed method consists of two main procedures these are:

A. Extraction of a candidate key frame procedure.

B. Extracting of a final key frame procedure.

Figure 1 shows the general two procedures for extracting of a key frame.

**Figure 1.** Steps of the proposed method (a) Procedure of candidate key frames extraction (b) Procedure of final key frames extraction
A. Procedure of candidate key frame extraction

The color feature in the video plays a crucial role for characterizing the statics of it. The presented method depends on color properties of the original difference between histogram of video frames. Where the difference between histogram of two adjustment frames \((i, j)\) is evaluated via the equation (1).

\[
D(F_i, F_j) = \frac{|hist_i - hist_j|^2}{hist_j} \tag{1}
\]

Where \(F_i\) and \(F_j\) are two adjustment frames.

Then, the standard deviation of the difference array \(D\) is computed according to the following equation:

\[
stdDF = \frac{1}{N} \sqrt{\sum_{i=1}^{N}(D(i) - DM)^2} \tag{2}
\]

Where \(N\) is the length of difference array \(D\) and \(DM\) is the mean of \(D\).

Finally, the threshold \((T)\) which is used for extracting the key frames is calculated by equation (3).

\[
T1 = \frac{\alpha}{2} \times (maxstdDF + minstdDF) \tag{3}
\]

Where

\(\alpha\) : Constant its value in the range \(0 \geq \alpha \leq 1\)

Algorithm (1) shows the steps of candidate key frames extraction.

| Algorithm (1): candidate key frames extraction |
|-----------------------------------------------|
| **Input:** Video                              |
| **Output:** Candidate Key Frames              |
| **For each frame I from 1 to N do**           |
|   **Step 1:** Convert RGB color frame into gray scale color |
|   **Step 2:** reduce the dimension of each frame into \(200 \times 200\) |
|   **Step 3:** Calculate the histogram of each pixel in the frame. |
|   **Step 4:** Calculate the differences between the pixels of each two adjust frames according the Eq. 1: |
|   **Step 5:** Find the standard deviation of the pixel differences according to equation (2) |
|   **Step 6:** Compute Adaptive threshold value of the camera frames depends on the normalized summations as follow equation (3). |
|   **Step 7:** Extracting Candidate key frames that have standard deviation of difference greater than the threshold value (T1) |
| **End for**

B. Final Key Frame Extraction Procedure

Figure (1.b) shows the steps of obtaining the final key frames. The input is candidate key frames, where the histogram is computed for each frame. Then normalization is applied for each gray level \((g)\) in the frame by dividing number of occurrence gray level \((Ng)\) to the total number of pixels as following:

\[
N_{hist}(g) = \frac{Ng}{N \times N} \tag{4}
\]

The Euclidian distance is computed between two consecutives frames \((i, j)\) using equation (5):

\[
F_{dis}(i, j) = |N_{hist}(i, g) - N_{hist}(j, g)| \tag{5}
\]

Finally, extract the final key frames that have distance \((F_{dis})\) smaller than threshold \((T2)\). Algorithm (2) shows the steps of final key frames extraction.
4. Experimental Results
The proposed method is tested with several videos; some of these are online television videos to verify the effectiveness of the method download from websites. The tested videos with different formats such as (avi, mp4,...).  The method was implemented in MATLAB, windows 7 system with Intel I processor and 16 GB RAM. Figure (2) shows examples of candidate key frames of standard video. The extracting of key frames operation aims to reduce or summarize the video frames while preserving the overall contents of video. Compression Ratio (CR) is used to evaluate the summarized video. CR is computed as follows:

$$CR = 1 - \frac{kfs}{tkf} \quad (6)$$

Where

Kfs represents the number of key frames in summery.
Tkf: represents the total number of key frames in original video.

---

**Algorithm (2): Final key frames extraction**

**Input:** group of Candidate Key Frames (CKF), Similar Frame Index (SIF=0)

**Output:** group of Final Key Frames

**Step1:** For each two consecutive frames (I ,I+1) in (CKF) to number of candidates frame (cno.) do

**Step 2:** Convert RGB color frame into gray scale color

**Step 3:** reduce the dimension of each frame into 200 × 200

**Step 4:** Calculate the histogram of each pixel in the frame.

**Step 5:** Normalize the histogram according to equation (4)

**Step 6:** Compare the frames Euclidian distance difference (Fdis) with threshold (T2)

**Step7:** if Fid<T2

Save frame I in similar frame array (SFA)

SFI=SFI+1

Go to step1

Else

If SFI<1

Save frame I as Final Key Frames

Else

Compute Average frame of SFA

Save frame I in SFA as Final Key Frames which have minimum distance to the average Frame.

SFI=0;

Go to step1

End for
Figure (2) shows that the extracted frames covered most video content. Also, it can be noted that some extracted frames from frame (2) to frame (7) are similar in their content. The difference among the similar extracted frames is the movement of object, color, and background. To extract final key frames the algorithm (2) is applied.

Figure (3) shows that the similar frames from frame number (2) to frame number (7) give one final key frame. The number of original video frames is (53) while the number of summarized frames is (5) with clearly video content. So, the compression ratio value is equal to (CR = 90.5700%).

In order to illustrate the performance of the proposed method, another example which is online videos selected from a television show “Super Diva”. Figure (4) and figure (5) show examples of candidate and final key frames of standard video. The number of original video frames is (1000) while the number of summarized frames is (15) with clearly video content. So, the compression ratio value is equal to (CR = 98.5000%).
5. Conclusions
A method for key frame extraction is proposed in this work. The purpose of extracting is for video copyright protection. The method consists of two stages to extract the key frames that covered the total content of video. In the first stage the candidate key frames are extracted based on differences of frame histogram standard deviation and adaptive threshold. In the second stage, represents the extracting of final key frames basing on computing Euclidian distance between frames to process the similar frames in their contents. For the future works the extracted key frames will be used for copyright protection based on watermarking.
References

[1] H. Yassine and A. Boukabou, “A robust and secure key-frames based video watermarking system using chaotic encryption”, Multimedia Tools and Applications, Volume 77, Issue 7, April 2018.

[2] K. Sahu and S. Verma, “Key Frame Extraction From Video Sequence: A Survey”, International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 05, May 2017.

[3] N. Ejaz, T. B. Tariq and S. W. Baik, "Adaptive key frame extraction for video summarization using an aggregation mechanism”, Journal of Visual Communication and Image Representation, Volume 23 Issue 7, October, 2012.

[4] S. T. Dhtagdi and P.R. Deshmukh, “Keyframe Based Video Summarization Using Automatic Threshold & Edge Matching Rate”, International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012.

[5] S. Lei, G. Xie, and G., “A Novel Key-Frame Extraction Approach for Both Video Summary and Video Index”, Hindawi, scientific journal, Volume 2014.

[6] S. C V, N.K. Narayanan, “Key-Frame extraction by analysis of histograms of video frames using stastical method”, 4th International Conference on Eco-Friendly Computing and Communication Systems, Procedia Computer Science pp. 36-40, 2015.

[7] L. Pan, X. Shu and M. Zhang, “A Key Frame Extraction Algorithm Based on Clustering and Compressive Sensing”, International Journal of Multimedia and Ubiquitous Engineering Volume 10, No.11, 2015.

[8] H. Gharbi, S. Bahroun and E. Zagrouba, “A Novel Key Frame Extraction Approach for Video Summarization”, International Joint Conference on Computer Vision Theory and Applications, Rome (2016) 2016.

[9] Y. Shi, I H. Yang, M. Gong, X. Liu and Y. Xia1, “A Fast and Robust Key Frame Extraction Method for Video Copyright Protection”, Journal of Electrical and Computer Engineering, Volume 2017.