Research Article

A Fast and Robust Key Frame Extraction Method for Video Copyright Protection

Yunyu Shi,[1] Haisheng Yang,[2] Ming Gong,[2] Xiang Liu,[1] and Yongxiang Xia[1]

[1]*School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China*
[2]*Shanghai Media Group, Shanghai, China*

Correspondence should be addressed to Xiang Liu; morningcall@sues.edu.cn

The paper proposes a key frame extraction method for video copyright protection. The fast and robust method is based on frame difference with low level features, including color feature and structure feature. A two-stage method is used to extract accurate key frames to cover the content for the whole video sequence. Firstly, an alternative sequence is got based on color characteristic difference between adjacent frames from original sequence. Secondly, by analyzing structural characteristic difference between adjacent frames from the alternative sequence, the final key frame sequence is obtained. And then, an optimization step is added based on the number of final key frames in order to ensure the effectiveness of key frame extraction. Compared with the previous methods, the proposed method has advantage in computation complexity and robustness on several video formats, video resolution, and so on.

1. Introduction

Video data has been increased rapidly due to rapid development of digital video capture and editing technology. Therefore, video copyright protection is an emerging research field and has attracted more and more attention. Digital video watermark is a traditional method for video copyright protection. However there are some faults about the above method and it is not suitable for huge video data on the Internet.

Key frame extraction is a powerful tool that implements video content by selecting a set of summary key frames to represent video sequences. Most of the existing key frames extraction methods are not suitable for video copyright protection, as they do not meet specific requirements.

Generally, key frame extraction techniques can be roughly categorized into four types [1], based on shot boundary, visual information, movement analysis, and cluster method. And then sometimes it could be completed in compressed domain [2]. Nowadays, cluster-based methods are mostly applied in video content analysis research. In these methods, key frame extraction is usually modeled as a typical clustering process that divides one video shot into several clusters and then one or several frames are extracted based on low or high level features [3–6]. The methods in compressed domain usually are not suitable for diverse formats of videos from the Internet. Transcoding may increase time complexity and inaccuracy.

How to achieve a meaningful key frame is an important problem in various communities. The focus of the work is to represent the video content adequately and fast [7, 8]. In this paper, an active detection method is proposed. First, the key frame is defined for video copyright protection. And then, a key frame extraction algorithm based on two-step method with low level features is proposed.

The distinct features of our algorithm are as follows. (1) The definition of key frame is specific for video copyright protection. (2) The method is with lower computation complexity. (3) The method is robust for online videos regardless of video formats, video resolution, and so on.

The rest of the paper is organized as follows. The proposed key frame extraction method is presented in Section 2, while experimental results are listed in Section 3. Finally the conclusions are drawn in Section 4.
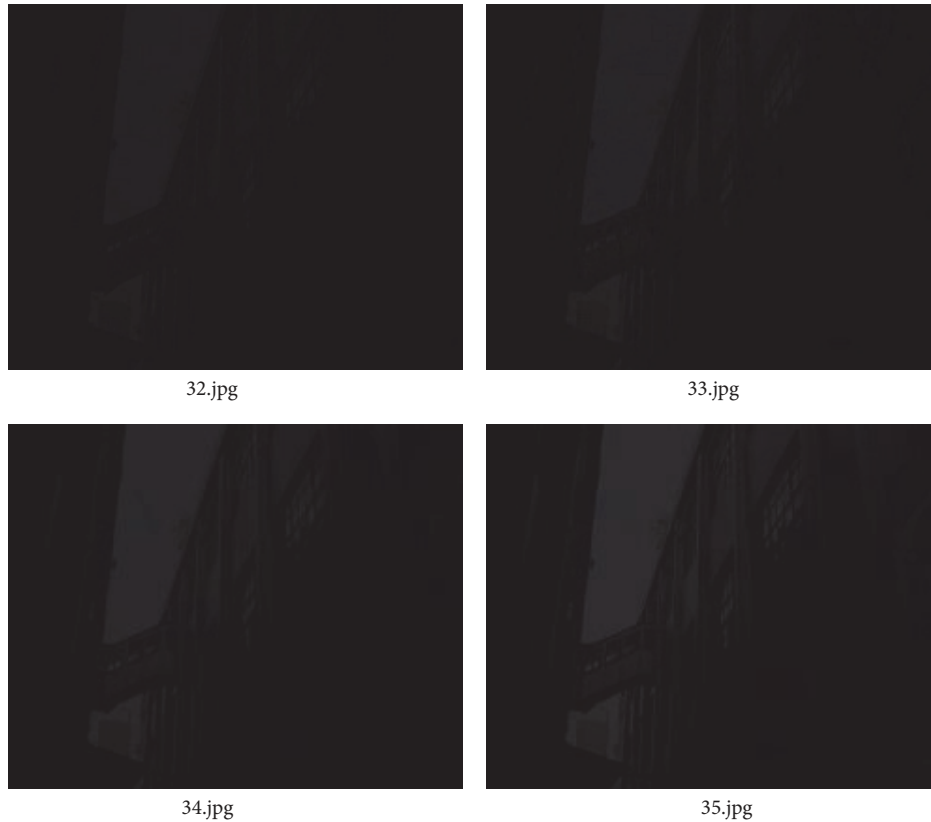
FIGURE 1: Four frames with low gray value extracted from a tested video.

2. The Proposed Key Frame Extraction Method

2.1. Definition of Key Frame for Video Copyright Protection. There are some distinct features about the key frame for video copyright protection. So, the key frame for video copyright protection is defined firstly before video preprocessing and key frame extracting.

The key frames should meet the following three conditions.

(1) The gray value of a key frame is within a certain range to allow viewers to have subjective perception about the video content. Four images with low gray value in Figure 1 are extracted from a single video, which is difficult for almost viewers to recognise the content.

(2) The final key frame sequence must be arranged in chronological order consistent with original video sequence, in order to satisfy temporal features and to be different from the short promotion trailer.

(3) Appropriate redundancy of some key frames is allowed to ensure the periods or intervals along the processing of video content. Figure 2 indicates the condition by selecting four images from a tested video, which are with similar content, that is to say, one judge in the show every once in a while.

In general, radio and television programs need to convey certain visual content; that is, video images that are too dark or too bright do not meet these subjective feelings. Four images in Figure 1 are extracted form a tested video, which are always too dark for viewers to perceive the video content. The phenomenon is sometimes with gradual transitions of shots. In order to distinguish and program trailers and other programs, the intervals between extracted key frames must be consistent with the frames from the original video. As online video piracy is often divided into smaller video files for playback, thus mastering the key frame extraction should allow appropriate redundancy to ensure a period of time. Taking the talent show as an example, the moderator reviewing screen may arise for every player in a game situation, as shown in Figure 2; then the time of video frames' critical information is reserved for the key frame extraction processing.

2.2. Two-Stage Method for Key Frame Extraction. Figure 3 is the key frame extraction overall flowchart for digital video copyright protection. First, a digital video is decomposed into video frames. The downloaded video from the network includes several video formats, such as f4v, flv, and mp4. In order to improve the universality of video key extraction algorithm, the present method does not consider the specific format and video stream structure, and the video is decoded before the processed video frame decomposition. It is seen from Figure 3 that the program to extract key frame is divided into two steps. Firstly, alternative key frame sequence based on the color characteristics of the original difference

25.jpg

669.jpg

1061.jpg

1778.jpg

FIGURE 2: Four frames with similar content extracted from a tested video.
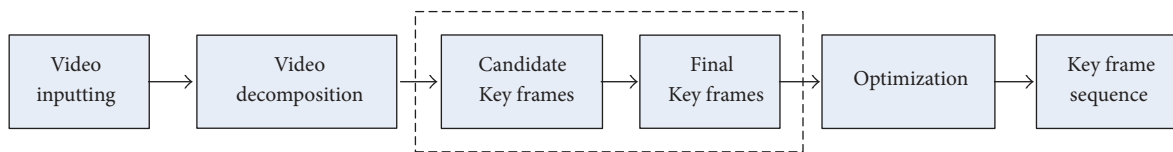

FIGURE 3: Flowchart of the proposed key frame extraction method.

between video frames is obtained; then key frame sequence is got according to the structure characteristic differences between alternative key frames sequence, and finally it is determined by the number of key frames in order to ensure the effectiveness of key frames.

Based on the above considerations, the frame difference method is used to extract key frames by analyzing the presence of spatial redundancy and temporal redundancy. In order to improve operational efficiency, it is worth mentioning that this method is different from the traditional shot segmentation method [9], for that the traditional approach is to conduct a video shot segmentation, then to extract key frames from each shot, and finally to compose key frame sequence of the video. In this method, the segmentation is not considered and then to extract key frames directly from the video.

2.2.1. Alternative Key Frame Sequence Based on Color Features. Color is one of the important properties of the image and is often used to characterize the statistics of the image [10, 11], and even for some specified domain video, color information can be expressed directly semantics, such as soccer video, usually on behalf of green grass. In addition, different color space of the sensory perception of visual effects is inconsistent. In order to achieve an effective balance between the key frame extraction efficiency and the speed,

the RGB color space is used and the color histogram for each frame is calculated. Then the color histogram difference between adjacent frames is adopted in the present method, as shown in Figure 4.

Based on the number of key frames, color feature extraction method for video sequence obvious video content conversion has a good ability to judge, but to little effect, or change the gradient color; light detection effect is not ideal, because the color histogram for pretty gradients and lighting effects such as gradients are very sensitive to the frame between a few dozen frames of video content; despite little change between adjacent frames, the significant changes between color histogram features are occurring. As previously stressed, in order to quickly and effectively perform key frame extraction, the video shot segmentation will not be adopted directly. Although motion estimation, optical flow analysis, and motion modeling method are effective in the previous method, the time complexity is also too high; these problems have a serious impact on the practical application of copyright in video monitoring.

2.2.2. Final Key Frame Sequence Based on Structure Features. Figure 5 is a key frame sequence optimization based on structural features. The program uses the first frame extraction based on color features alternate key and then extracted key frames to optimize based on structural features; that is, the
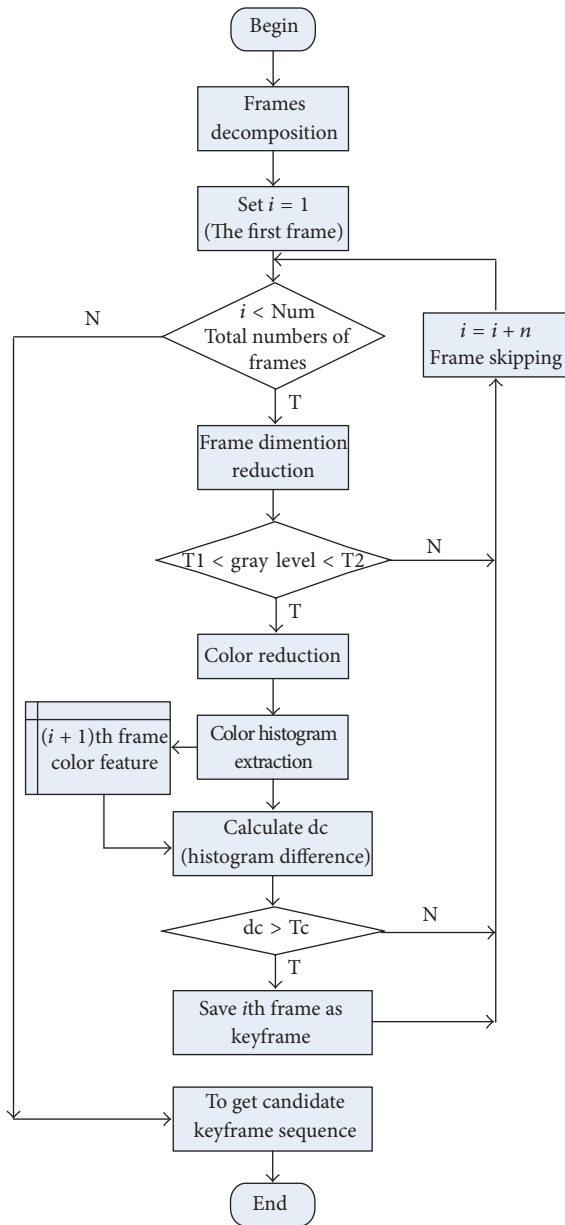
FIGURE 4: Flowchart of generating candidate key frame sequence.

FIGURE 5: Flowchart of generating final key frame sequence and optimization based on key frames number.

alternative key frame structure similarity between adjacent frames is determined to further reduce key frames.

The method is derived from the structural similarity evaluation method for image quality evaluation [12] and is a measure of the similarity of the two images; the value closer to 1 indicates that the two images' quality is more similar. Structural similarity theory states that natural image signal is highly structured and that there is a strong correlation between pixels, especially airspace closest pixels; this correlation contains important information visual objects in the scene structure. Human visual system (HVS) main function is to extract structured information from view; it can be used as a measure of structural information perceived image quality of approximation. In this scenario, the structure is similar to
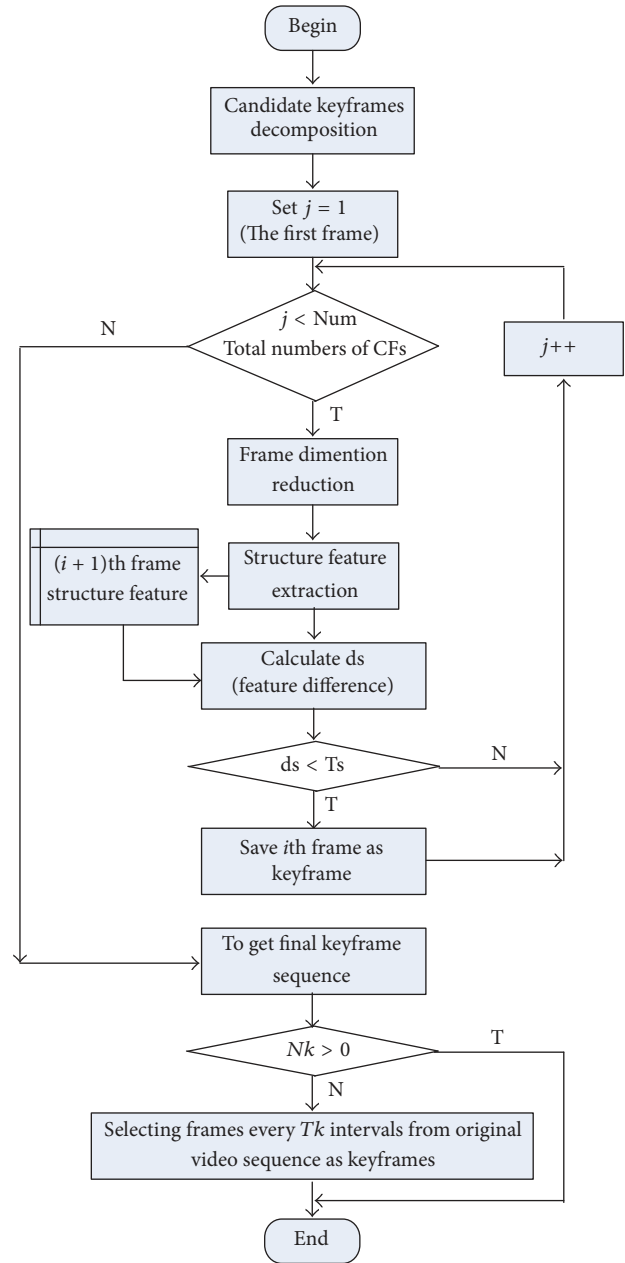
the concept introduced to the key frame optimization process; thereby removing the extraction of the frame structure information is not sensitive to this problem based on color feature key.

The program uses only similarity index structure similar to the structure of the components. From the perspective of an image composition, structural information is defined as an independent component from brightness and contrast in the theory of structural similarity index. And it could reflect the properties of objects in the scene. The covariance is selected as a structural similarity metric. The main calculation is as follows.

FIGURE 6: Examples of alternative key frames.

Covariance as a structural similarity measure for the image block x, y of the correlation coefficient, namely, the covariance of x and y, is calculated as

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_i)(y_i - \mu_i), \tag{1}$$

where N is the number of the patches and μ_i is the average value.

In the alternative key frame sequence, the front frame could be as the original image, and the adjacent frame is set as the test image. According to the two corresponding image blocks at the same position x (in the original image) and y (in the test image), the structure similarity component between the two image blocks is calculated as (x, y):

$$s(x, y) = \frac{\sigma_{xy} + C}{\sigma_x \sigma_y + C}, \tag{2}$$

where $C = ((KL)^2/2)$, $K \ll 1$, $L \in (0, 255]$ and σ_x, σ_y are x and y variance, respectively.

If the component values of $s(x, y)$ are small, then the distinction between the contents of the information is not; at the same time they do not have to be retained as a key frame, which can be extracted only as a key frame is optimized.

2.3. Optimization Based on the Number of Key Frames. After extracting alternative key frames based on color features and key frames based on structural features, the number of key frames will be determined to meet the demand. If no key frame is extracted from a video, then it will extract the appropriate number of key frames from the original video, in accordance with isochronous interval. Usually this occurs in the lens without the division, such as newscasts broadcast of a piece with only anchor shot. There are no significant changes in color and structural features between video frames.

3. Experiments and Analysis

The method is applied to a lot of online videos downloaded from several video websites and the digital linear tapes are from Shanghai Media Group. The algorithm was implemented in C++ and OpenCV 2.0, and then the experiments were conducted on a Windows 7 system with an Intel i7 processor and 16 GB RAM.

Firstly, we took television show "SUPER DIVA" to verify the effectiveness and robustness of the proposed method. More than 20 versions of the copies or near-duplicates were downloaded, which may be different in video formats (.mp4, .rm, .wmv, .flv, etc.), spatial resolutions (1920 ∗ 1080, 1080 ∗ 720, 720 ∗ 576, etc.), video lengths (such as short clips cut from a full video), and so on. The results which are got from the downloaded video with mp4 format are partly shown in Figures 6 and 7.

From Figure 6, we could see that most key frames are covering the video content exactly. There are also some frames similar with content, such as the three frames in the 2nd and 3rd row. The difference among these frames is color background, especially the bubble lights. So the final key frames are extracted based on the structural difference from the alternative key frames, as shown in Figure 7. In general, these final key frames meet the three conditions mentioned in Section 2. The frame content could be viewed definitely and their order consisted with the original video, and there

FIGURE 7: Examples of final key frames extracted from Figure 6.

FIGURE 8: Three sets of key frames from different versions of the tested video.

is appropriate redundancy, such as the third frame in the 1st row and the last frame in the 3rd row.

Secondly, three different versions of SUPER DIVA were tested to get the final key frames. They are different in formats or resolutions and are noted in V1 (.mp4, 640 ∗ 352), V2 (.flv, 608 ∗ 448), and V3 (.avi, 512 ∗ 288). The results are partly shown in Figure 8. It should be noted that the frames listed in Figure 8 are cropped to the same size for the appearance. Generally, each set of key frames are consistent with others, especially with almost the same video content and the same time line. The reason for the different key frames may be because of the same feature difference thresholds, Tc and Ts.

Thirdly, the optimization step based on the number of key frames was tested and the results are listed in Figure 9. The original video is a short promotion trailer about a famous

movie. There's almost no feature difference among these original frames, only the mouth movements and few hand movements of the introducer. So no key frames are extracted based on the color and structure information. Therefore, the optimization based on the a fixed time interval is needed in order to satisfy the key frame demand and ensure the following processes for video copyright detection.

4. Conclusions

A key frame extraction method based on frame difference with low level features is proposed for video copyright protection. Exactly, a two-stage method is used to extract accurate key frames to cover the content for the whole video sequence. Firstly, an alternative sequence is obtained

FIGURE 9: Optimized key frames based on the frame number.

based on color characteristic difference between adjacent frames from original sequence. Secondly, the final key frame sequence is obtained by analyzing structural characteristic difference between adjacent frames from the alternative sequence. And thirdly, an optimization step based on the number of final key frames is added in order to ensure the effectiveness for video copyright protection processes. Tested with several television videos with different content, formats, and resolutions, it is shown that the proposed method has advantages in computation complexity and robustness on several video formats, video resolution, and so on. In the future the adaptive threshold is the primary research point.
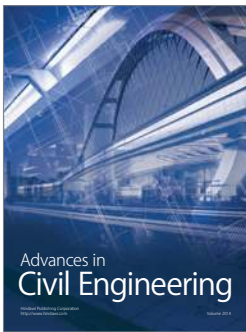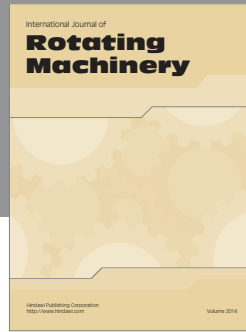
Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. M. Zhang, H. Y. Liu, and S. M. Sun, "Keyframe extraction based on improved ant algorithm and agglom erative," *Computer Engineering and Applications*, vol. 49, no. 3, pp. 222–225, 2013.

[2] X. Guo and F. Shi, "Quick extracting keyframes from compressed video," in *Proceedings of the 2nd International Conference on Computer Engineering and Technology (ICCET '10)*, pp. V4163–V4165, Chengdu, China, April 2010.

[3] S. Angadi and V. Naik, "Entropy based fuzzy C means clustering and key frame extraction for sports video summarization," in *Proceedings of the 5th International Conference on Signal and Image Processing (ICSIP '14)*, pp. 271–279, January 2014.

[4] Q. Guo, C. Zhang, Y. Zhang, and H. Liu, "An efficient SVD-based method for image denoising," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2016.

[5] J. Peng and Q. Xiao-Lin, "Keyframe-based video summary using visual attention clues," *IEEE Multimedia*, vol. 17, no. 2, pp. 64–73, 2010.

[6] L. Liu and G. Fan, "Combined key-frame extraction and object-based video segmentation," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 15, no. 7, pp. 869–884, 2005.

[7] M. Chatzigiorgaki and A. N. Skodras, "Real-time Keyframe extraction towards video content identification," in *Proceedings of the 16th International Conference on Digital Signal Processing (DSP '09)*, pp. 68–73, grc, July 2009.

[8] S. Lei, G. Xie, and G. Yan, "A novel key-frame extraction approach for both video summary and video index," *The Scientific World Journal*, vol. 2014, Article ID 695168, 9 pages, 2014.

[9] J. Li, Y. Ding, Y. Shi, and Q. Zeng, "DWT-based shot boundary detection using Support Vector Machine," in *Proceedings of the 5th International Conference on Information Assurance and Security (IAS '09)*, pp. 435–438, September 2009.

[10] Z. Sun, K. Jia, and H. Chen, "Video key frame extraction based on spatial-temporal color distribution," in *Proceedings of the 4th International Conference on Intelligent Information Hiding and Multiedia Signal (IIH-MSP '08)*, pp. 196–199, August 2008.

[11] J. Zhang, X. Jiang, G. Li, and L. Jiang, "Key frame extraction based on particle swarm optimization," *Journal of Computer Applications*, vol. 31, no. 2, pp. 358–361, 2011.

[12] Y. Shi, Y. Ding, R. Zhang, and J. Li, "Structure and hue similarity for color image quality Assessment," in *Proceedings of the International Conference on Electronic Computer Technology (ICECT '09)*, pp. 329–333, February 2009.