International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 5 Issue 11 ǁ Nov. 2017 ǁ PP.71-76

Efficient Video Coding Techniques for the Optimal Repeated Frame Compensation

Soumendra Prasad Rout[1], Pallavi Priyadarshini[2]
[*1] *Department of Electronics & Communication Engineering,, Gandhi Engineering College, Odisha, India*
[2] *Department of Electronics & Communication Engineering, Gandhi Institute For Technology, Odisha, India*

ABSTRACT*: A large amount of storage and bandwidth is needed when data transmission is done using uncompressed digital video. This results in a large amount of data that demands the need of video compression. Many of the existing video compression standards share common or similar coding tools, and there is currently no explicit way to exploit such commonalities at the level of implementations. Moreover, the possibility of taking advantage of the continuous improvements in coding is only possible by replacing the existing approach with a new one. This paper proposes an efficient video content representation by exploiting the temporal redundancies using optimal extraction of repeated frames and scenes. A new standard for video coding called Optimal Repeated Frame Compensation (ORFC) is used, in which the repeated frames are combined together to form a single frame in order to reduce the total number of frames.*
Keywords: *Compression, fidelity, key frame extraction, pixel prediction, video coding*

I. INTRODUCTION

A video is said to be a collection of several images. Each image is called frames and the amount of images shown per second is called Frames Per Second (FPS). A video can be represented by numerous consecutive frames, each of which corresponds to a constant time interval. Two or more sequences in a video represent the same scene with different view point. Therefore, the two sequences are related with each other though they are different. However, such a representation is not adequate for new emerging multimedia applications, such as content-based indexing, retrieval, and video browsing. Moreover, tools and algorithms for effective organization and management of video archives are still limited. As the video data originate from the same scene, the inherent similarities of the image are exploited for efficient video compression. Key frame extraction, is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing and retrieval. The use of key frames reduces the amount of data required in video indexing and provides the framework for dealing with the video content in an efficient way. Key frame is the frame which can represent the salient content and information of each shot. The key frames extracted must summarize the characteristics of the video, and the image characteristics of a video can be tracked by all the key frames in time sequence. The efficient representation of video content at key frames level is crucial, for video copy detection as they don't meet specific requirements. As a result, simple approaches of low complexity are usually preferred. Generally, key frame extraction techniques can be roughly categorized into Sequential and Cluster-based methods [1]. In sequential methods, consecutive frames are compared in a sequential manner. The key frames are detected depending on the similarity with either the previous frames or the previously detected key frame. The main disadvantage is that, this method computes only the similarity between adjacent frames and ignores the overall change trend in the shot range. In cluster-based methods, the frames are grouped into a finite set of clusters in the selected feature space [2], and then the key frame set is obtained by collecting the representatives of each cluster group. In this method, key frames are selected regardless of the temporal order of each frame. If key frames are extracted for each shot independently and the scenery changes slowly in each shot, cluster-based methods are able to provide an understanding of the overall visual content of a video. In this paper, a novel key frame extraction method is proposed. Compared with other existing methods, the proposed approach has two main characteristics: (1) it extracts the key frames using a simplified algorithm called 'Optimal Repeated Frame Compensation (ORFC) (2) the ORFC uses an adaptive key frame extraction method in which the repeated frames are combined together to form a single frame in order to reduce the number of frames. It then employs the local features to further refine key frame candidates, which helps the system to get high quality key frames. Rather than coding each entire picture repeatedly, video can be represented more effectively by coding only the changes present in the video content. This ability to use the temporal-domain redundancy to improve the coding efficiency is what fundamentally distinguishes the proposed system from other existing methods.

II. PROPOSED METHOD

This approach proposes a method to identify and visualize repetitive structures in frames using ORFC. The repeated frames are combined together to form a single frame in order to reduce the number of frames. It is mainly used to identify repetitions in a frame. For extracting limited and meaningful information of video data, an adaptive video content framework is proposed in this paper. Fig.2.1 illustrates the block diagram of the proposed method, consisting of four modules: video detection, video sequence analysis, classification, and key frame extraction. The first three modules are mainly used to produce a feature vector representation of the video sequence by first segmenting it into distinct video shots [3]. Such a representation provides a more meaningful description of the video content and, therefore, key frame extraction can be implemented more efficiently. The frame difference is calculated mainly on the basis of two classifications. One is enhanced pixel prediction and other one is temporal prediction. Both help to extract frames in an efficient way by exploiting the inherent similarities among adjacent frames.

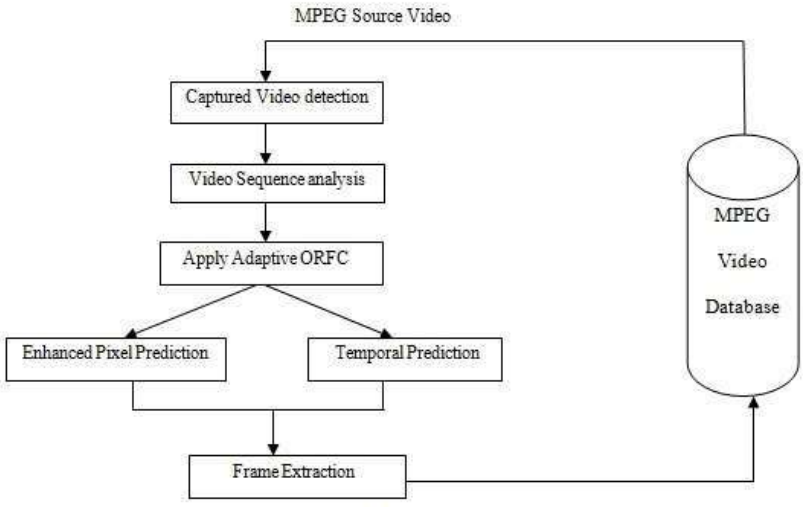


Fig.2.1. Block diagram of proposed method

Two similarities are mainly examined for key frame extraction here. The first one is based on the inter-view similarity among still image frames. In second, temporal similarity is noticed between temporally successive image frames based on minimization of a cross-correlation criterion. A new algorithm of key frame extraction from compressed video data is presented in this paper. For any video, the first common step is to segment the videos into temporal "shots". A shot represents a sequence of frames captured from a unique and continuous record from camera as shown in Fig.2.2.

Fig.2.2. Video sequences partitioned into adjacent frames

Once a video sequence is temporally partitioned into video shots, the next step of the analysis is extraction of unrepeated frames of each shot. The goal of frame extraction is to determine the presence of a set of frames representing different shots. This, however, is a difficult problem, since it involves a priori knowledge about the shots to be detected and thus can only be solved for a limited range of application contexts [9, 18]. As the video data originate from the same scene, two similarities are mainly examined for key frame extraction here. The first one is based on the inter-view similarity among still image frames. In second, temporal similarity is noticed between temporally successive image frames based on minimization of a cross-correlation criterion. This classification corresponds to the natural arrangement of video images into a Matrix Of Pictures (MOP) [4]. In MOP, each row holds temporally successive pictures of one view, and each column consists of spatially neighbouring views captured at the same time instant. All video sequences are still arranged into the rows of the MOP. Fig.2.3 depicts a matrix of pictures for N = 4 image sequences, each form a Group of Views (GOV), and K = 4 temporally successive pictures a temporal Group Of Pictures (GOP) [5]. For example, the images of the first view sequence are denoted by x1, k with k = 1, 2,...,K. MOPs with NK images to estimate the compression efficiency of coding schemes are discussed here. It mainly exploits all similarities among these images. Here, the idea is to distinguish between inter-view similarity and temporal similarity. Therefore, further sub-

classification of inter-view similarities is not intended. These inherent similarities of the image are exploited for efficient video compression. All existing video coding standards developed so far states that temporal similarities can be exploited with motion compensation techniques that are well known from single-view video compression. It results in unreliable transmission. This paper takes the temporal predictor in [6] as a reference. Also an enhanced pixel predictor [7] is used which exploits the inherent similarities among adjacent frames

using ORFC. Almost in all the existing methods adjacent frames are compared but not in an optimal way. So in order to minimise the number of frames in an optimal way, each and every frames obtained are compared and minimised using ORFC. The proposed scheme mainly exploits the redundancies among the images.

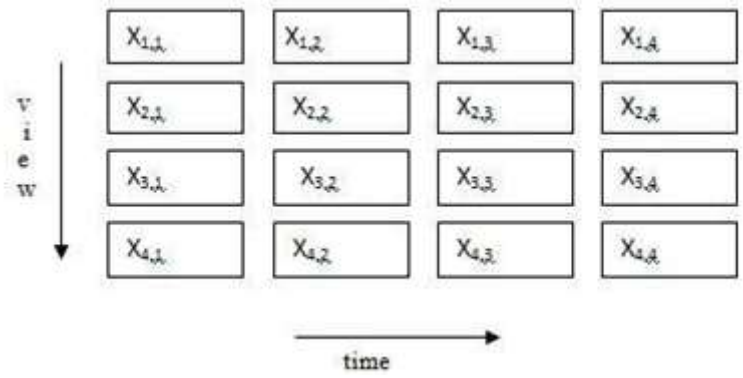


Fig. 2.3. Matrix of pictures (MOP) for N = 4 image sequences, each comprising of K = 4 temporally successive pictures.

In this paper, an efficient video content representation is proposed using optimal extraction of repeated frames and scenes. For performing the repeated frame/scene extraction, a new approach for video coding is proposed in this paper which is called 'Optimal Repeated Frame Compensation'(ORFC) which provides a powerful representation of video shots for the problem of key frame extraction. Here, comparison between adjacent frames is carried out. If the adjacent frames are equal then ignore the repeated frame. The ORFC approach compares the obtained minimised frame with next frame. Thus the maximum number of repeated frames is ignored here. In Fig.2.2.the number of frames is 8. After applying ORCP the frames are minimised to 2 as shown in Fig.2.4.

Fig.2.4. Key frames extracted using ORFC

III. OPTIMAL FRAME BASED BLOCK-MATCHING MOTION COMPENSATION

The procedure to estimate the adaptive pixel-based prediction [8] among consecutive frames based on intraframe coding modes is described here. Here, each image frame is divided into a fixed number of square blocks as shown in Fig. 3.1.

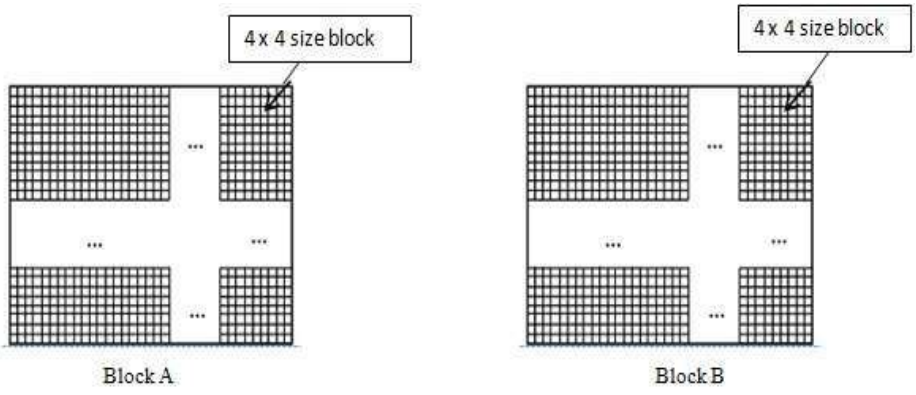


Fig.3.1. Adaptive Pixel-based prediction of two consecutive blocks

For each block in the frame, a search is made for the best matching block, to give the least prediction error, usually minimising either mean square difference which is easier to compute. A good match during the search means that a good prediction can be made, but the improvement in prediction must outweigh the cost of transmitting the motion vector. A good match requires that the block should not overlap objects in the image that have different degrees of motion. This reduces the work and transmission costs of subsequent correction stages but with greater cost for the motion information itself. Therefore, an adaptive coding scheme is applied to the pixel by comparing the temporal and spatial variations around the pixel in block A and block B. The flowchart is given in fig.3.3. In enhanced pixel based prediction, it is important to note that ORFC utilize intraframe compression coding mode selection to calculate the data which specifies differences in specific pixels from one frame to the next, rather than simply redrawing an entire frame each time. It compares only the differences from one frame to the next, and saves bandwidth by only processing significant changes in particular pixels. The difference of two patterns is calculated with cross-correlation coefficient which is given in eqn.1as :

$$C_{ij} = \frac{Area\ (fragment_i \cap fragment_j)}{\sum_{i=1}^{N} Area\ fragment_i \sum_{j=1}^{M} Area\ (fragment_j)} \quad (1)$$

where {$fragment_i$} in (i=1,.....,N) denotes a fragment set in pattern I and {$fragment_j$} in (j=1,...M) denotes a fragment set in pattern J as shown in fig 3.2.. The Area ($fragment_i \cap fragment_j$) is the overlapping area of pattern I and J. It is measured only for calculating cross-correlation coefficient [9]. For all candidate fragments J, the cross-correlation coefficients are calculated. The candidate fragment which gives the maximum value of C_{ij} is decided as the identified fragment of fragment I. Thus it helps to find out the best search point for pattern matching.

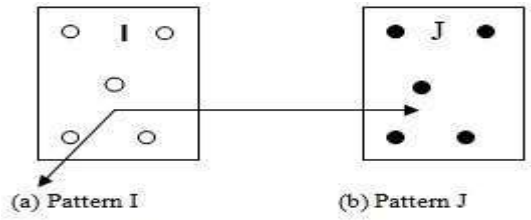


Fig 3.2 Representation of patterns

Enhanced Temporal Prediction

The temporal prediction predicts the color of each pixel based on the values of those pixels located in the same pixel neighbourhood in the previous two frames. This approach is adaptive and simple, and has the advantage of not requiring the transmission of any side information. The proposed enhanced temporal predictor aims to find the accuracy of location prediction between the current frame[i] and reference frame[i-1]. It provides a great way to view images based on the accuracy of location by examining each pixel very carefully, in order to find out very minute difference in frame[i] and frame[i-1].

Let $p_{i(a,b)}$ be the symbol to be encoded. The proposed temporal predictor aims to find the best matched symbol in reference frame[i-1], which is denoted as the temporal predictor $p_i{}^T\ a, b$. The temporal predictor of symbol $p_i(a,b)$ searches and achieves the minimum cumulative absolute difference (CAD) [10] within the search range, where denotes the target window, pi(a,b) and pi-1(a,b) denote the symbol values of the current frame[i] and the reference frame[i-1] respectively. It is given in eqn.2 as:

$$CAD(T_W) = \sum_{(m,n)\in T} |pi(a, b) - pi - 1(a + m, b + n)| \quad (2)$$

where T_w denotes the target window, $p_i(a,b)$ and $p_{i-1}(a,b)$ denote the symbol values of the current frame[i] and the reference frame[i-1] respectively. After calculating CAD, next the accuracy of location is calculated. The accuracy of location prediction f_a is the ratio of predicted probability of actual next location to f_c. fa of stay- duration s_i in S_i is defined in eqn.3 as follows:

$$f(s_i, S_i) = \frac{p_{ij} \int_{s_i-\psi}^{s_i+\psi} \epsilon_{ij}\, t\, dt}{f_c(s_i, S_i)} \qquad (3)$$

where the numerator is provision probability derived from stay-duration patterns to actual next location lj . fa is related directly within the given location prediction to the given departure. The optimal solution generates fa=1, which means that the mobility model correctly predicts the next location at the actual departure moment. This will support enhanced data exploitation tools, including real-time display of target coordinates, digital data archiving, DVR playback capability, and image mosaicking.

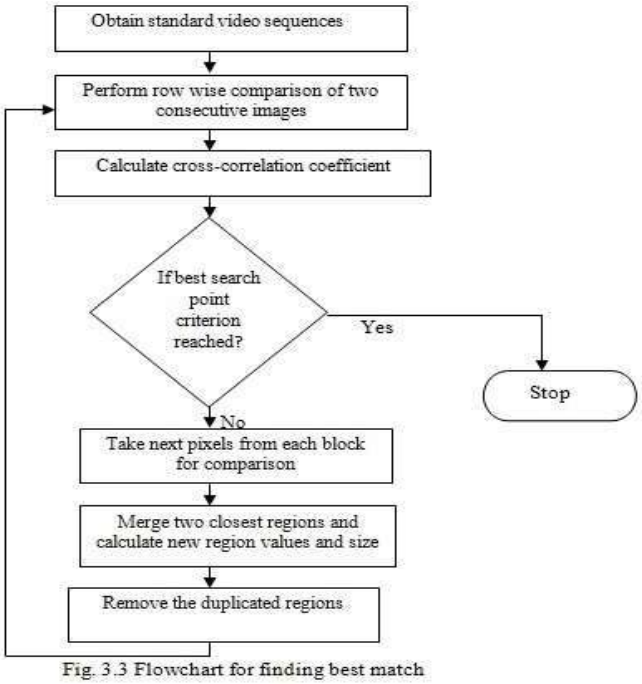
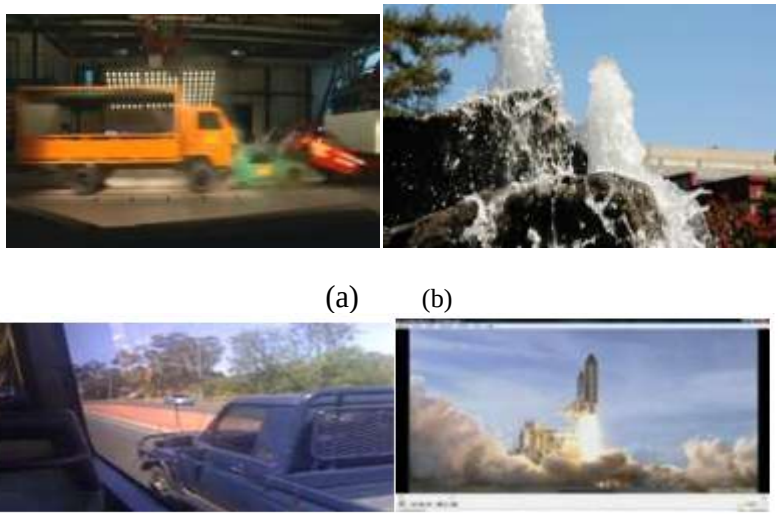


Fig. 3.3 Flowchart for finding best match

IV. RESULTS AND ANALYSIS

`The key frames mainly reduce the amount of data required in video indexing and preserve the overall contents of the original video. The number of key frames accurately extracted measure the validity of the algorithm. Four segments of video are taken for doing the experiment as shown in fig. 5



(a) (b)

(c) (d)

Fig.5. Four video sequences with different typical motion characteristics.
(a)adac-crash(b) midnikon-D300,(c) cycorder-640,(d) svc1

The two important parameters used to measure the validity of the algorithm is compression ratio and fidelity. The compression ratio is used to quantify the reduction in data representation size whereas; fidelity is to measure the correlation degrees of the sets in the image classifications. Fidelity is defined as a standard Semi-Hausdorff distance between the key frame and shot frame produced as described in [11]. It is given in eqn. 4 as:

$$\text{Fidelity} = 1 - \frac{dsh}{\max_{i} \max_{j}(d_{ij})} \quad (4)$$

where d_{sh} is the Semi-Hausdorff distance and d_{ij} denotes the dissimilarity of the shot.
As the fidelity increases, the more accurate key frames generated. The results are shown in Table 1.

TABLE 1. THE RESULTS OF THE EXPERIMENT

Video sequences	# of shots	# of frames	# of KF	Compression ratio(%)	Fidelity
A: adac-crash	12	1656	20	99.7	0.7329
B: midnikon-D300	16	2040	28	99.3	0.7432
C: cycorder-640	9	1405	14	99.6	0.7634
D: svc1	18	2280	22	99.5	0.7783

From the results, the average compression ratio of the new algorithm is 99.525% and the average fidelity is 0.75. The result shows that the proposed algorithm is valid to segment the shot and to extract key frames in an optimal way. It gives good feasibility and strong robustness.

V. CONCLUSION

In this paper, a mechanism for automatic extraction of the most representative frames in video databases is proposed. A minimization of the frames repeated has been done here using an optimization technique called ORCP. This technique is used for indicating the most characteristic frames within each selected block. To accomplish the optimal extraction, first a detailed analysis of frames extracted has been studied in order to obtain an image representation more suitable for classification. It has been seen that the proposed method is valid to segment the shot and to extract key frames in an optimal way. It gives good feasibility and strong robustness

REFERENCES

[1] H. Kim, J. Lee, H. Liu, and D. Lee, "Video Linkage: Group based copied video detection," in Proc of CIVR'08, Niagara Falls, Canada, July 7–9, 2008, pp. 397-406.

[2] X. Zeng, W. Hu, W. Li, X. Zhang, and B. Xu, "Keyframe extraction using dominant-set clustering," in Proc Int. Conf. Multimedia and Expo (ICME'08), Hannover, Germany, June 2008, pp. 23-26.

[3] Yannis S. Avrithis, Anastasios D. Doulamis, Nikolaos D. Doulamis, and Stefanos D. Kollias, "A Stochastic Framework of Optimal Key Frame Extraction from MPEG Video Databases," in Proc. Int. Symp. Computer Vision and Image Understanding., Vol. 75(1), pp. 3–24, July 1999.

[4] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multi-view videos ," IEEE Transactions on Circuits and Systems for Video Technol., vol. 17(11), pp. 1454-1484, Nov 2007.

[5] Markus Flierl and Bernd Girod, "Multiview Video Compression-Exploiting Inter-Image Similarities," IEEE Signal Processing Magazine, Special Issue on Multiview Imaging and 3DTV, Vol. 24(6), pp. 66-76, Nov 2007.

[6] Z. Ming-Feng, H. Jia, and Z. Li-Ming, "Lossless video compression using combination of temporal and the spatial prediction," in Proc. IEEE. Int. Conf. Neural Netwoks Signal Processing, pp. 1193–1196, Dec 2003.

[7] K. H. Yang and A. F. Faryar, "A context-based predictive coder for lossless and near-lossless compression of video," in Proc. Int. Conf. Image Processing., vol. 1, pp. 144–147, Sep. 2000.

[8] Sung-Eun Kim, Jong-Ki Han, and Jae-Gon Kim, "An Efficient Scheme for Motion Estimation Using Multi-reference Frames in H.264/AVC," IEEE Transactions on Multimedia.,Vol. 8(3), pp. 457-466., June 2006.

[9] Ruan Xiaodong, and Song Xiangqun, "Research On The Algorithm of the Binary Image Cross- Correlation For Unsteady Flow Field Measurement," ACTA MECHANICA SINICA (English Series), Vol. 15(1), Feb 1999.

[10] K.Dinesh and T. Arumuga Maria Devi, " Motion Detection and Object Tracking in Video frame sequence on IWT of Adaptive Pixel Based Prediction Coding," in Proc. IRACST-Engineering Science Technology: An International Journal (ESTIJ), ISSN: 2250-3498,Vol.2, No. 4, August 2012.

[11] Guozhu Liu and Junming Zhao, "Key Frame Extraction from MPEG Video Stream," in Proc. Int. Symp. Computer Science and Computational Technology., Huangshan , pp. 007-011, Dec 2009.