VIDEO COMPRESSION VIA MINIMUM FRAME DIFFERENCE LOCALIZATION ADAPTED FOR MOBILE COMMUNICATIONS

KHATTAB M. ALI ALHEETI

1 Department of Information Systems, College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq [co.khattab.alheeti@uoanbar.edu.iq]

MUZHIR SHABAN AL-ANI

2University of Human Development, College of Science and Technology, Department of Information Technology, Sulaimani, KRG, Iraq [muzhir.al-ani@uhd.edu.iq]

ABSTRACT

Huge amount of multimedia data are circulated via communications media (wired and wireless), Internet, and many other media, so these data need to be compressed to achieve storage space and speed. Many method are used for image and video compression such as discrete cosine transform, discrete wavelet transform and discrete curvelet transform. Some of these methods are efficient and have wide range of applications. The aim of this paper is to implement an efficient approach of video compression based on frame difference localization. The implemented approach based on the mixing of two methods; firstly, calculate the difference between the two near frames to avoid the similarity and secondly, applying discrete wavelet transform to reduce the amount of similar pixels as possible. The implemented approach is tested for many types of image, in which a good performance results are obtained. This approach is adapted for mobile communications in order to save speed and space.

Keywords: Image Compression, Video Compression, Frame Compression, Frame Extraction 2D-DWT.

1. INTRODUCTION

The advanced of mobile and wireless technologies leading wide range of applications in various areas [1,2]. The wide expand of Internet all over the world that leading to huge circulation of text, audio, images and videos in each moment [3,4,5]. These data occupy huge amount of space and lead speed down the transmission step by step [6,7]. Many years ago researchers and industries try to find methods and systems to overcome these problems, but these problems continue to grow according to huge usability of Internet and communications media and electronic devices [8,9,10]. The large amount of bandwidth required for the transmission or storage of digital videos is the main objective for researchers to develop algorithms for compressing images and videos while maintaining the highest possible quality [11,12].

The earlier versions of the standard Moving Picture Experts Group (MPEG) are implemented discrete cosine transform, then the advanced versions are used discrete wavelet transform and discrete curvelet transform [13,14,15]. The gray scale image using 8 bits per pixel, so it generates up to 256 gray level of intensities [16,17]. On the other hand, color image can be represented by three bytes each of 8 bits, 8 bits for red, 8 bits for green and 8 bits for blue. Digital videos consist of a stream of images called frames, and these frames are a part of time interval from each other [18,19,20]. The video signal is sampled into frames and fields, in which a complete frame is completed in specified time that is equal to real time and the fields are including either odd frames or even frames [21,22,23].

There are thousands of articles and books published in the field of image and video compression [24,25]. In addition there are several methods and algorithms investigated this subject [26,27]. This paper try to implement and efficient approach for video compression via minimum frame difference localization.

2. DATA COMPRESSION

Digital audio and video signals integrated with computers, telecommunication networks, and consumer products are faced to the information revolution [28,29]. This revolution is concentrated
on the advanced technologies of storage and transmission media and one of the important issue is the digital compression of audio and video signals [30,31]. Most sensory signals contain a substantial amount of redundant or superfluous information. The first three standard TV transmissions are PAL, SECAM and NTSC [32,33]. Phase Alternating Line (PAL) is a standard color encoding system for analog television used in broadcast television systems in most countries broadcasting at 625 line and 25 frames [34,35]. Sequential color with memory (SECAM) is a standard analog color television system first used in France [36,37]. National Television System Committee (NTSC) is the analog television system that is used in the USA, Philippines, South Korea, Taiwan and Japan [38,39]. Most countries that using analog television standards, have switched to newer High definition TV and digital television standards [40,41].

Thousands of artificial satellites circulating the orbit Earth, some take pictures of earth, planets, sun, black holes, and dark matter [42,43]. These images help scientists better understand the solar system and the universe. Satellites are mainly used for communications, such as television signals and telephone calls around the world [44,45]. A group of more than 20 satellites constitutes the global positioning system.

The television camera that captures 25 or 30 frames per second from a stationary scene produces similar frames one after the other [46]. Compression is the process that attempts to remove the redundancy information as possible to represent frames by a smaller amount of data, or by a lower data rate [47]. One of the big advantage of compression is in data rate reduction [48]. Data rate reduction reduces transmission costs with the fixed transmission capacity that leading to a better quality of multimedia presentation [49].

MPEG is short for the Moving Pictures Experts Group who is responsible for the video encoding standards that we often use [50,51]. Both MPEG2 and MPEG4 codec are widely used audio/video compression standards in many multimedia applications, ranging from mobile devices to DVD players [52,53,54].

Both Joint Photographic Expert Group (JPEG) and Moving Pictures Experts Group (MPEG) are two different types of compressing formats, JPEG is mainly used for image compression, while MPEG is used for both audio and video compression [55,56,57]. The aim of MPEG was to generate standards for audio and video compression and transmission, this is extended to be a big group [58,59,60]. The generation standards are [61,62,63]:

- MPEG-1 (1993): Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s (ISO/IEC 11172) [64].
- MPEG-3: Dealt with standardizing scalable and multi-resolution compression and was intended for HDTV compression but was found to be redundant and was merged with MPEG2 [66].
- MPEG-4 (1999): Coding of audio-visual objects. Includes compression of AV data for web and CD distribution, voice and broadcast television applications [67].
- MPEG-7 (2002): Multimedia content description interface. Not a standard which deals with the actual encoding of moving pictures and audio, like MPEG1, MPEG2 and MPEG4 [68].

3. RELATED WORK

Image and video compression is a wide subject and big amount of papers are published in this subject. This section will be concentrated on the new and advanced works.

Xiaolin Chen (2012) implemented a lossless video compression system via a new estimate of the Backward Adaptive pixel-based fast Predictive Motion Estimation based on a pixel adaptable to the rear. This scheme predicts motion on a pixel-by-pixel basis by comparing a group of observed pixels passed between two adjacent images, eliminating the need to transmit lateral information. Combined with prediction and a fast search technique, the proposed algorithm achieved better entropy results and a significant reduction in computation than the full-pixel search for a set of standard test sequences. The experimental results also show that this approach overcomes the complete block search in terms of velocity and entropy. This approach offered the integration of Backward Adaptive pixel-based fast Predictive
Motion Estimation into a complete lossless video compression system [71].

Jonathan Fabrizio et al. (2012) present a new algorithm for motion compensation using a motion estimation method based on the tangent distance. This method is compared to a blocking approach in various common situations. While blocking algorithms usually only predict block positions over time, this method also predicts the evolution of pixels in these blocks. The prediction error is then greatly reduced. The method is implemented in theory of encoding by proving this algorithm to improve the performance of the video codec. This approach provides a new motion compensation algorithm for video compression. This method offered improving in the compression ratio compared with the classic block comparison strategy [72].

F. Racape et al. (2013) implemented the use of conjunction with current and future standard compression systems. The introduced a texture analysis tools for segmentation and characterization, in order to locate candidate regions for synthesis: motion compensation or texture synthesis. The decoder fills them using texture synthesis. The remaining regions of the images are conventionally encoded. They can serve as input for the synthesis of textures. The chosen tools are developed and adapted to ensure the consistency of the whole scheme. Therefore, a texture characterization step provides the required parameters to the texture synthesizer. Two texture synthesizers, including a pixel-based and patch-based approach, are used in different types of texture to complement each other [73].

Rupesh Gupta et al. (2013) guided a video compression algorithm by visual visibility at macro-block level. This approach is modeled as a two-step process namely, detection of the protruding region and the cleavage of the structure. Visual effectiveness is modeled as a low-level combination, as well as high-level features that become important for the top-level visual cortex. A vector machine of relevance is formed in vectors of specific three-dimensional characteristics of the global, local and visibility scarcity measurements, in order to produce probabilistic values that form the map of wages. These efficiency values are used for non-uniform bit allocation in video images. To achieve these goals, this approach also offered a new video compression architecture, which integrates the advantage, to save a lot of computing. This architecture is based on the threshold of the mutual information between successive frames to flagging frames that require a new calculation of the salinity and the use of motion vectors for the propagation of the values of salience [74].

Shiping Zhu (2014) explored a new encoding of fractal video sequences with automatic region-based functionality, to improve the performance of the fractal video encoding. In order to increase the quality of the decoding image, the intra-frame coding, the unlock loop filter and the sub-pixel coupling are applied to the codec. They used an efficient search algorithm to increase the compression ratio and coding rate. The automatic encoding of fractal video sequences based on the region greatly reduces the encoding stream. Experimental results indicated that the proposed algorithm is more robust and provides much less coding and bit rate time while maintaining decompression image quality than the conventional methods [75].

Nijad Al-Najdawi (2014) developed a rapid search motion estimation algorithms to reduce the computational cost required by the full search algorithms. They presented a new fast search algorithm based on the hierarchical search approach, where the number of sites sought is reduced compared to the complete search. The original image is sub-sampled in two additional levels. The complete search is performed at the highest level where the complexity is relatively low. The improved three-stage search algorithm and a proposed new search algorithm are used on the two consecutive levels. The obtained results show that when using the standard precision measurements and the standard set of video sequences, the performance of the proposed hierarchical search algorithm approximates the complete search with a reduction in complexity of 83.4% and a quality corresponding to more than 98% [76].

Burcu Karasoy, Fatih Kamisli (2015) investigated alternative transformations for multi-view video coding. This approach analyzed the spatial characteristics of the predicted residuals offset by the disparity and the results of the analysis show that many regions have similar 1-D signal characteristics to the previous results for the motion-compensated prediction residuals. Signals with such characteristics can be processed more efficiently with transforming and adapted to these characteristics, it is advised using 1-D
transformations in the compression prediction residuals offset by disparity in multi-view video coding. To show the attainable compression gains when using these transformations, they modified the reference software of the video multi-view modification of H.264 / AVC coding so that each residual block can be transformed with either 1-D transformation or with conventional 2-D discrete cosine transformation. The experimental results show that coding gains can be obtained ranging from about 1 to 15% of the Bjontegaard-Delta bit saving [77].

Mark Q. Shaw et al. (2015) studied a method for selectively modifying a video sequence using a color contrast sensitivity model based on the human visual system. The model identified regions of high variance with frame-to-frame differences that are visually imperceptible to a human observer. The model based on color difference formula, taking advantage of the nature of progressive image-based video coding. The use of a color contrast sensitivity model alone was not enough. Therefore, it was important to integrate information on the perceptual health and spatial activity of the scene. The method was implemented in the reference software of the JM 18.0 H.264 / AVC encoder and resulted in an average gain of 37% in data compression without noticeable degradation in video quality [78].

R. Mukherjee et al. (2016) achieved objective and subjective global assessment carried out with six published video compression algorithms. The objective evaluation was performed on a large set of 39 high dynamic range video sequences using seven digital error measures: PSNR, logPSNR, puPSNR, puSSIM, Weber MSE, HDR-VDP and HDR-VQM. The subjective evaluation related to six pre-selected sequences and two subjective experiments based on the ranking with a hidden reference to two different output rates with 32 participants each, who were responsible for classifying the distorted high dynamic range images in an uncompressed version of the same sequence. The obtained results indicated a strong correlation between objective and subjective evaluation. In addition, noncompliant compatible compression algorithms seem to work better at lower output speeds than the backwards compatible algorithms in the parameters used in this evaluation [79].

S.V.N. Murthy, B.K. Sujatha (2016) designed a technique to achieve an effective balance between video compression using H.265 protocol and 8K resolution retention. The study implemented multilevel optimization in the coding process using H.265 where JPEG2000 standards play a crucial role. The study also applied a new concept of orthogonal projection that manages the required pixel metadata at each frame transition followed by motion compensation. Using multiple file formats of 30 video datasets, the obtained result of the study is around 49% improvement in data quality and about 59% improvement in video compression [80].

Fernanda A. Andaló et al. (2017) discussed a simple effective framework to generate highly compressible videos that contain only relevant information for specific tasks of computer vision, such as faces for facial expression recognition, license plates for the optical recognition task of characters, and others. This approach takes advantage of the latest computer vision task required to compose video images with only the required data. Video images are compressed and can be stored or transmitted to powerful servers where long and long tasks are performed. Experiments explore the trade-offs between distortion and bitrate for a wide range of compression levels and the impact generated by compression artifacts on the accuracy of the desired vision task. They showed that for two computer vision tasks implemented by different methods, it is possible to considerably reduce the amount of data that needs to be stored or transmitted without compromising accuracy [81].

Engin Şenol, Nükhet Özbek (2017) presented a new methodology of subjective evaluation that offers better efficiency over time and more accurate scores for the same or even less assessors, using new objective measurements for multi-view compressed video quality of experience measurement. The research objective was concentrated on two aspects. First, the new subjective test methodology is developed and the quality of experience of the different multi-view video compression technologies are evaluated. Then an effective solution over time is proposed as multiple stimulation plus an extension of the simultaneous presentation of the conventional method with continuous quality double stimulation. Second, objective quality measures based on maps of structural similarity and depth are proposed. The experiments are carried out on a self-stereoscopic screen, while HVC-based encoders and H.264 / MPEG-4 high-resolution video encoding are used for compression [82].
M. Kiruba, V. Sumathy (2018) proposed a new architecture model of discrete tchebichef transform with register pre-allocation based folded architecture (RPFA) for image compression. Due to cross-connection of the folded architecture, the registration number is reduced. A novel method is introduced in the proposed discrete tchebichef transform architecture. This multiplier concept involves the cross-function of the multiplier with the split pattern of the binary multiplication stream. The optimum design of the discrete tchebichef transform architecture produces a minimum number of Flip Flop accounts, latency and power consumption. The proposed approach achieves a better peak signal to noise ratio, better structural similarity, slower delay, power consumption, and better mean square error than existing multiplier architectures [83].

Eduardo Cermeño et al. (2018) explained a smart video solution for perimeter protection should select and display the cameras that are most likely to bear witness to a relevant event, but systems based on background modeling tend to give importance to problematic situations independently whether or not an intrusion occurs. They proposed a module based on the learning machine and the global functionalities, providing an adaptability to the video surveillance solution so that problematic situations can be recognized and granted to the correct priority. Tests with thousands of hours of video show how an intruder detector can work, but also how a simple failure in a camera can flood a surveillance center with alerts. The new proposal is able to learn and recognize events so that alerts of problematic environments can be handled properly [84].

4. METHODOLOGY

The methodology of this approach is concentrated on design and implementation of video compression via minimum frame difference localization. This methodology is performed in in the following items.

4.1 Implemented Approach

First thing that to be understand that any video at the origin it is a stream of image or frames. So to start any process on videos it is important to return back to the reference images that are the fundamental of the video. This approach of video compression is implemented via the following stages (figure 1):

- Preprocessing in which introduced cleaning, filtering, resizing and adapted the video to be standard for processing.
- Frames extraction in which extract the video into a number of specific frames within a certain time.
- Differences measures in which measure the difference value between frames.
- Threshold localization in which fixed the frame passing a certain threshold.
- Frames compression in which pass the localized selected frames into the compression step.
- Video format in which reconstruct the video depending on the new compressed frames.

4.2 Frame Difference Measures

This is an important stage of the overall compression approach in which mixing both frame selection operation and threshold measure operation to achieve an accurate decision for the compression process. This stage can be implemented via many steps (figure 2):

- Frame resizing in which resize the frames into a certain specified size of the original video.
• Histogram measures in which applying histogram in each frame to find the distribution of pixels.
• Frame difference measures in which it calculate the difference between each two neighboring frames.
• Threshold measures in which calculate the absolute difference according to a certain threshold.
• Avoid unwanted frames in which removes the frames with the difference equal to zero or below the average.
• Frame re-extracting and video format in which re-extract the frames with difference greater than zero then reconstruct the video format.

The process of extraction and re-extraction of video is shown in figure 3 in which starts to separates the video into separated frames in frame extraction and then after processing of frames, collects the frames to reconstruct the new video.

4.3 Two Dimensional Discrete Wavelet Transform
Two dimensional discrete wavelet transform (2D-DWT) is applied in this approach for compression of selected frames. 2D-DWT is implemented on frames via a combination of both low pass filter (LPF) and high pass filter (HPF) for both rows and columns respectively. LPF is used to maintain the details of image and HPF is uses to maintain the edges of image. 2D-DWT is implemented via the following steps (figure 4):
• Step1: Applying low pass filter (LPF) on columns to generate low band (L-band).
• Step2: Applying low pass filter (HPF) on columns to generate high band (H-band).
• Step3: Applying high pass filter (LPF) on rows of step1 to generate low low band (LL-band).
• Step4: Applying high pass filter (LPF) on rows of step2 to generate high low band (HL-band).
• Step5: Applying high pass filter (HPF) on rows of step1 to generate low high band (LH-band).
• Step6: Applying high pass filter (HPF) on rows of step2 to generate high high band (HH-band).

Figure 5 shows the implementation of 2D-DWT for first level, second level and third level, this procedure is done via considering the LL band as an original image for the next step and so on.

5. RESULTS AND DISCUSSION

The implementation of this video compression approach depends on many steps starting from frames extraction and ending to frames re-
extraction. The most important stage of this approach is concentrated on the threshold measures and frames difference. Two type of frame selection and thresholding are applied in this approach. These methods are zero difference measures method and mean value measures.

At zero difference method it removes the frames where the distance between any two consecutive frames is zero, and the zero difference occurs in many videos that have small changes between frames and when the frame extracted above twenty frames per second, in addition this case will minimize the number of frames to be re-extracted. Figure 6 illustrate the graph of zero measures to find difference between frames.

Mean value measures in which the mean value of the frames difference is calculated. The mean value is obtained by dividing the sum of the observed values of frames difference by the number of observations. Then when the frames difference between any consecutive frames is lower than the mean value of the frames difference it will be removed. Figure 7 illustrate the graph of mean measure to find difference between frames.

6. ASPECTS AND ADVANTAGES

Most of the presented literatures are studies image and video compression based on direct techniques. This approach deals with video compression via minimum frame difference localization adapted for mobile communications. The difference of this approach is its compensation between the compression and the differences between frames.

7. CONCLUSIONS

Recently, image and video compression techniques are widely used in various applications. The applied video compression approach is implemented comparing a pair of frames to measure the difference value between frames. A minimum frame difference localization is applied to select the frames that have significant values of difference. This approach based on two fundamental methods; zero measures and mean measures to archive the difference between frames. 2D-DWT is applied in this approach to achieve a maximum compression with reasonable quality. A good performance of the compressed video is achieved via the implementation of this approach that achieved the main objective of this work.

The limitation of this approach deals with the adaptation of this method with different types of mobile technologies that is considered as a future research direction in this field.
REFERENCES:


Representation, Volume 27, February 2015, Pages 7-27.


