Journal of Computational Science 18 (2017) 1–11

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Performance enhanced spatial video compression using global affine frame reconstruction

D. Raveena Judie Dolly [a], G. Josemin Bala [a], J. Dinesh Peter [b,*]

[a] Department of Electrical Technology, Karunya University, Coimbatore, India
[b] Department of Computer Sciences Technology, Karunya University, Coimbatore, India

ARTICLE INFO

Article history:
Received 20 August 2016
Received in revised form 15 October 2016
Accepted 13 November 2016
Available online 15 November 2016

Keywords:
Video compression
Group of pictures (GoP)
Adaptive frame determination
Affine motion parameters
Optimization
Affine warping

ABSTRACT

The modern era of information technology suffers a serious loss due to the lack of a cutting edge methodology for storing mega sized videos. It is at this juncture, video compression makes a mark for its necessity. There have been several research outcomes where almost all researchers have followed a particular methodology of adopting GoP (Group of pictures) for video compression, focusing on I (Intra), B (Bi-directional) & P (Predicted) frame determination. These frames remain fixed throughout the process of GoP regardless of the camera motion. Moreover, it also leads to buffering of memory within the past and future thereby consuming more computational time for B-frames. These vital issues are handled by an adaptive framework of determining frames based on a matching criteria rather than utilizing fixed GoP pattern. NSEW affine translation (NAT) is introduced for replacing B-frames with either I or P frame. The framework involves VCAME (Video Compression using Affine Motion Estimation) & VDAW (Video Decompression using Affine Warping) methodologies for compressing and decompressing a video sequence, based on the resulted I & P frames. The methodology was investigated over four vital parameters, the file size, computational time, SSIM (Structural Similarity Index) & PSNR (Peak Signal to Noise ratio), which proved the superiority of the proposed technique. Further, the methodology was also investigated with optimizing the affine motion parameters (AMP) using nonlinear least squares, BFGS (Broyden–Fletcher–Goldfarb–Shanno) and Limited-memory BFGS which yet again proved to be far more superior than conventional techniques yielding an average PSNR gain of **2.52 dB**.

1. Introduction

The fast movement of still images constitute a video. These still images are referred as frames. Successive frames in a video sequence expose the scene variations, which may be a drastic change or a small change depending on the camera motion. Any video sequence considered for compression are categorised into Intra ('I'), Inter ['P (Prediction)' and 'B (Bi-directional)'] frames respectively. I frames constitute the key reference frame for successive ones. Prediction of P-frames depends on the former I or P frames. The B-frames are interpolated from forward and backward frames. The choice of the frame type decides the quality and the compression ratio of the compressed video. The more number of intra frames, better the quality of video. Conversely, the more number of inter frames, better the compression ratio and poorer the quality of video. The frames in a GoP can either follow IP, IBP or IBBP frame patterns. Prediction analysis is emphasized in conventional algorithms such as MPEG-4 and H.264 to reduce data between a series of frames. Differential coding technique is employed for comparing the reference frame with the succeeding frames, where the changed pixels alone with respect to the reference frame are encoded. Realisation is performed by motion estimation and motion compensation. The motion vector is decided upon the correlation of motion between two frames. By adopting this technique the encoded and the transmitted number of pixel values are considerably reduced. For real time applications, motion estimation is not well suited due to its computationally intensive operation. Conventionally, if the file size is lowered by raising the compression level the visual quality gets affected. A significant reduction in file size without sacrificing the visual quality can be achieved by using several efficient compression techniques as mentioned in the literature. The proposed video compression method involves affine parameter estimation for motion estimation and affine warping for motion compensation where the motion parameters are estimated and stored as compressed data.

* Corresponding author.
E-mail addresses: dollydinesh@karunya.edu (D.R.J. Dolly), josemin@karunya.edu (G.J. Bala), dineshpeter@karunya.edu (J.D. Peter).

http://dx.doi.org/10.1016/j.jocs.2016.11.003
1877-7503/

2. Related works

All conventional video coding standards follow I, B & P frame choices for the purpose of compression. The GoP has one intra frame with subsequent inter frames. It is vital to decide the dissemination of these three frame types so as to improvise the coding efficiency. The GoP structure was emphasized by Wang et al. [24] where it was determined in real time by the information obtained from the Mean of absolute difference (MAD) and sum of absolute difference (SAD). Generally, the conventional video compression methods have the key reference frame as I-frame as in the work of Paul et al. [13] which insisted on adapting multiple reference frames (MRFs) for improving the performance of the H.264 video coding standard over other contemporary video coding standards. MRFs provide better predictions than single reference frame for a video with repetitive motion, dynamic background and illumination changes. The existence of many I-frames significantly decreases the coding performance for a video sequence which does not contain any modification in the scene while comparing the previous frames. If significant scene changes are detected in between frames, the concerned frame is marked as I-frame as recommended by Ohm et al. [8]. Therefore, optimal number of I-frames have to be selected carefully in order to get better compression ratio without compromising the quality of decompression. Zatt et al. [26] has adopted the usage of adaptive GoP where, video segmentation is employed first to generate GoP according to the video content. Aghamaleki et al. [29] has adopted a strategy of varied GoP to detect inter frame forgeries. Al-Ani et al. [27] suggested frame difference approach to calculate '*frame near distance*' for efficient compression using the wavelet technique. Mukherjee et al. [28] has employed video compression algorithms for HDR (High dynamic range videos). Xu et al. [30] uses video compression algorithms for data hiding, where motion vector of a macro block is embedded in the same frame in another macro block for replacing the corrupted motion vector.

Texas Instruments application report by Mathew [12] stated that B-frame requires a reference from the previous and future frames. Therefore, conventional standards follow a method of encoding the frame that follows a B-frame before encoding the B-frame. Moreover, the increase in the number of B-frames introduce delay in hierarchical-B coding. Paul et al. [14] suggested that, when motions in the video source is complex, avoiding B-frames results in better performance. Hence, replacement of B-frames was suggested. Apart from that, MPEG-2 code was also implemented and tested with and without the presence of B-frames. The compression time taken for *Rhinos* video using MPEG-2 in *Matlab* was about 60.5 s with B-frames. After converting the B-frames to P-frames, the time consumed was found to be 40.5 s. For the purpose of testing, *Vipmosaicking* video was considered and it was observed that it took 57 s with B-frames and 41 s without B-frames.

Tabatabai et al. [4] suggested that affine transformation can be used for co-ordinate transformation in video compression. Lu et al. [23] proposed a new *Gestalt factor*, "*shift and hold*" for 2D object extraction. This ignited the authors in this proposed work for performing shifting and correlation for the replacement of B-frames. Wiegland et al. [18] emphasized the utilization of affine motion model for traditional motion-compensated coders incorporating transformations like shear and rotation, allowing greater flexibility to the underlying image.

Although plethora of approaches are available, the proposed VCAME methodology is focused with adaptive frame determination. Once all the frames are adaptively recognised as either I, P or B-frames, the proposed NSEW affine translation method described in Section 3.2 is imposed to substitute all the B-frames with either I-frame or P-frame. A novel methodology named VCAME (Video Compression using Affine Motion Estimation) is proposed, where the affine motion parameters (AMPs) are found between first I-frame and the adjacent P-frames until another I-frame is encountered. This procedure continues by considering the succeeding I-frame as the reference frame. The AMPs along with I-frames are stored in a file. This file represents the compressed data. While decoding, VDAW (Video Decompression using Affine Warping) method utilizes these stored AMPs and I-frames. The AMPs are applied to I-frame using warping operation to obtain corresponding P-frames. Thus the complete set of frames is obtained for a particular video sequence. Various standard optimization methods are used to refine the AMPs.

3. Proposed methodology

Video compression is enhanced by determining the frames adaptively during the initial stage instead of randomly selecting I, B and P-frames, as portrayed in Fig. 1. For an input video sequence, the B-frames are considered for replacement in order to reduce the buffering between the past and future frames by using the proposed NSEW affine translation (NAT). The proposed VCAME methodology is applied to obtain affine motion estimation parameters between the successive frames. Though the Block matching strategies are easy to implement, it could produce poor results when numerous moving objects exist in one block. The affine motion estimation model estimates not only translational motions but also the non-translational motions such as zooming, rotation etc. Affine motion estimation model is performed between I and P-frames until another I-frame is encountered. These parameters are stored in a file with corresponding frame indexes along with jpeg encoded I-frames which is considered to be the compressed data. The resulted affine motion parameters are applied to the I-frame in order to get back the P-frame using the proposed VDAW approach from which the decompressed video sequence is obtained.

3.1. Adaptive frame determination (AFD)

The investigation was carried out using adaptive frame determination (AFD) to improve the quality of video coding as suggested

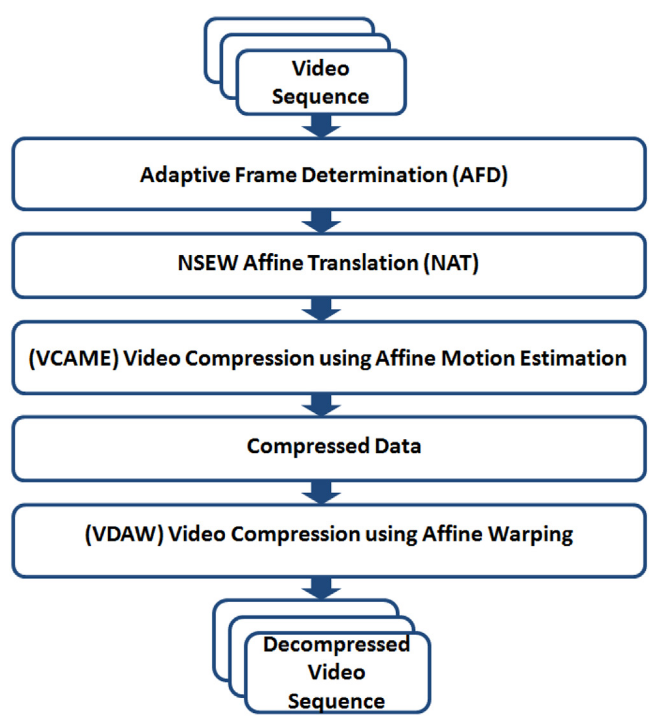

Fig. 1. Process flow of VCAME and VDAW.