Engineering Applications of Artificial Intelligence 69 (2018) 127–136

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Wavelet-based frame video coding algorithms using fovea and SPECK

J.C. Galan-Hernandez [a,*], V. Alarcon-Aquino [a], O. Starostenko [a], J.M. Ramirez-Cortes [b], Pilar Gomez-Gil [b]

[a] Department of Computing, Electronics, and Mechatronics, Universidad de las Americas Puebla, Sta. Catarina Martir, Cholula, Puebla, C.P. 72810, Mexico
[b] Department of Electronics and Computer Science, Instituto Nacional de Astrofisica, Optica y Electronica Tonantzintla, Puebla, Mexico

ARTICLE INFO

Keywords:
Video compression
Fovea
SPECK
Wavelet compression
HEVC

ABSTRACT

Current video coding standards such as the H.264 or Advanced Video Coding (AVC) and the H.265 or High Efficiency Video Coding (HEVC) are based on the Integer Discrete Cosine Transform (IDCT). However, it has been shown that wavelet-based algorithms have better performance than IDCT-based algorithms for still images. Also, exploiting the human visual system characteristics in stationary point of view image and video transmissions like the fovea aliasing can improve the quality of the reconstructed image. In this paper, two wavelet-based video coding approaches called SPECK-based Codec (SP-Codec) and Adaptive Wavelet/Fovea-based Codec (AWFV-Codec) are proposed. The proposed SP-Codec approach applies the Set Partitioned Embedded Block Codec (SPECK) wavelet-based compression algorithm in order to increase the intra-frame coding quality. The second proposed approach AWFV-Codec uses the Adaptive Fovea Set Partitioned Embedded Block Codec (AFV-SPECK) fovea wavelet-based compression algorithm for intra-frame coding. Fovea based compression allows to increase the quality of the reconstructed frames over regions of interest (ROI). A comparison of the proposed algorithms against the HEVC based on IDCT approach shows how the proposed wavelet-based schemes achieve higher compression ratios and higher reconstruction quality. When fixation point of view is stationary, the proposed wavelet/fovea-based algorithm achieves better quality at lower compression ratios than the wavelet-based only proposed algorithm.

1. Introduction

Data compression, also called source coding (Dragotti and Gastpar, 2009), is the process of creating a representation of a data source using a smaller data stream. Data compression limits are provided by the main theorem proved by Shannon (2001). A source can be compressed down to Shannon's rate–distortion limit without loss of information (Salomon and Motta, 2010). The efficiency of a compression algorithm is measured by its data compressing ability (Shukla and Prasad, 2011) and how close is the compressed data stream to the rate–distortion limit. If a compression algorithm compresses a data source further down its rate–distortion limit, it is said that such algorithm is a lossy compression algorithm because some data is lost and the original data cannot be recovered. Because lossy compression can achieve high compression ratios, it has been of general interest to research lossy compression algorithms for big data sources where the loss of data can be isolated. Some common big data sources are image and video data. There are several methods of lossy compression for images based on transform functions such as the JPEG (Wallace, 1991) and the JPEG2000 (Acharya and Tsai, 2004) standards or stochastic methods (Song et al., 2016). Video sources present higher challenge because it can contain more information than still images. Also, there is no ideal compression algorithm that works for every video source (Salomon and Motta, 2010).

A video stream can be seen as two separated sets of data, audio and image. Video coding techniques operate over the series of images called frames and audio coding over the audio stream. Modern video coding frameworks contains both lossy and lossless compression algorithms that are combined in order to achieve high compression ratios. Modern video coding frameworks are also called hybrid codecs. Current video coding standards such as HEVC and Display Stream Compression (DSC) (Walls and MacInnis, 2016) used in High-Definition Multimedia Interface (HDMI) and Display Port connections are based on a general hybrid video coding scheme (Wien, 2015) shown in Fig. 1.

In Fig. 1, blocks are defined as follows:

- Spatial Coding

* Corresponding author.
E-mail addresses: juan.galan@udlap.mx (J.C. Galan-Hernandez), vicente.alarcon@udlap.mx (V. Alarcon-Aquino).

https://doi.org/10.1016/j.engappai.2017.12.008
Received 15 June 2017; Received in revised form 21 October 2017; Accepted 12 December 2017

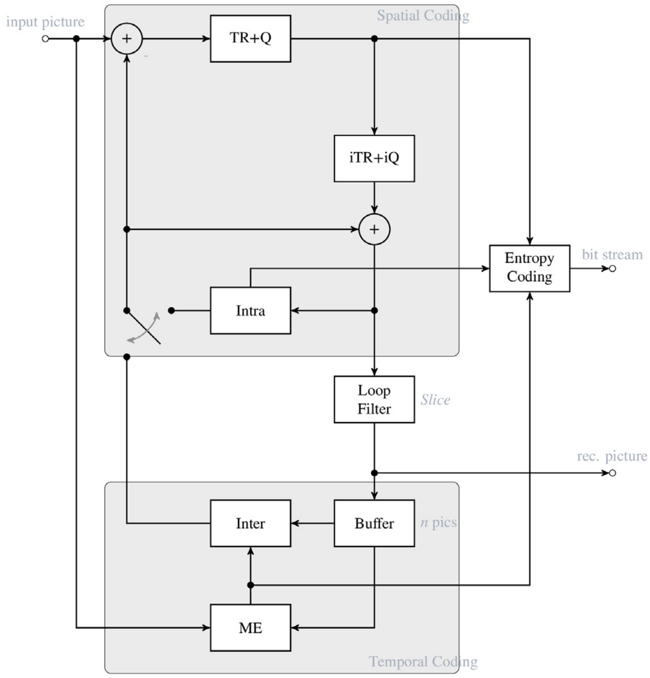

Fig. 1. Block diagram of the hybrid video framework.

– Transform and Quantization (**TR + Q**). A video frame is taken from the video source and encoded using a spatial transform and quantized.
– Inverse Transform and Inverse Quantization (**iTR + iQ**). The compressed frame is reconstructed in order to calculate the prediction error with the same information available by the decoder.

• Temporal Coding

– Buffer. A buffer of size n is kept for calculating motion vectors.
– Inter. Inter-prediction of frames using estimated motion vectors and a chosen frame (key frame).
– ME or Motion Estimation. Movement Vectors are estimated using previous or future frames if available.

• Entropy Coding. A lossless entropy algorithm is used over the data stream either the quantized coefficients or the estimated motion vectors.
• Loop Filter. Frames are sliced in multiple blocks that are processed independently.

In spatial coding, also called Intra-frame coding, the still image that represents a frame of a video source is compressed using lossy image compression algorithms. Spatial coding uses the information of previous or future frames in order to predict the movement of different sections of the frame to be coded. Motion prediction algorithms are not likely to find a perfect match of the different sections of the encoded frame. Because of that the encoder must calculate the differences between the predicted frame and the original frame. The remaining difference is an image known as the prediction error (Bovik, 2009). The prediction error can also be encoded using lossy compression. An improvement over any of the main blocks should improve the overall quality of the reconstructed video stream. While most video coding standards such as the H.264 (Bovik, 2009), HEVC, and the DSC rely on the Discrete Cosine Transform (DCT) for lossy intra-frame coding, improvements made by the HEVC standard increases the overall quality of the compression. However, it has been shown (Boopathi and Arockiasamy, 2012) that the Discrete Wavelet Transform (DWT) can achieve better image reconstruction that the DCT at high bit compression ratios as in the Dirac codec (Özenli, 2016). Furthermore, because current standards focus on general use cases, intra-frame algorithms can be improved on specific implementations. On video sources where the fixation point of the user is stationary, coding schemes are able to exploit the Human Visual System (HVS) for improving the quality of the reconstruction using fovea coding (Lee and Bovik, 2003). Some applications with stationary point of view are wide field of view content (Sacha et al., 2017) and current virtual reality devices such as the Oculus Rift (Ross and Lenton, 2016) or the HTC Vive (Dempsey, 2016).

In this paper we propose the use of the Lifting Wavelet Transform (LWT), which is an improved implementation of the DWT (Goswami and Chan, 2011) for intra-frame coding. Also, a video coding approach called SP-Codec is proposed that uses the SPECK compression algorithm (Pearlman and Said, 2008) for quantization in order to increase the quality of the reconstructed frames and residual errors. A second approach called AWFV-Codec is proposed. AWFV-Codec is based on the LWT and a new compression algorithm proposed in Galan-Hernandez et al. (2013) based on SPECK and fovea coding called AFV-SPECK that increases the quality of the reconstructed video sequence over selected areas centered around a fovea center.

The remainder of this paper is organized as follows. In Section 2, an overview of the SPECK algorithm is given. In Section 3 an overview of fovea coding is reported. Section 4 presents the proposed approach. Section 5 presents results, and Section 6 reports conclusions and future work.

2. SPECK

The SPECK algorithm proposed in Pearlman et al. (2004a) exploits the hierarchical structure and the energy clustering of wavelet decomposition. The SPECK algorithm is an extension of the Embedded Zero Tree Wavelet (EZW) (Shapiro, 1993) algorithm. SPECK was proposed as an improvement of the Set Partitioning Hierarchical Tree (SPIHT) algorithm that is the standard benchmark on wavelet based image coding (Pearlman et al., 2004a). SPECK was designed to be easy to implement and use less resources than the SPIHT algorithm. Image quality obtained by the SPECK algorithm is not higher than the one achieved by the SPIHT algorithm. However, SPECK presents a fast and low resource algorithm that shows a good trade off between quality and performance. SPECK can also compete with standard algorithms such as the Verification Model (VM) algorithm used in Joint Photographic Experts Group 2000 (JPEG2000) (Acharya and Tsai, 2004). SPECK defines several sets and lists. Let X be a set of pairs (i, j) that contains all positions of all coefficients a wavelet decomposition with level Λ and wavelet ψ. SPECK defines several coefficient sets. When a wavelet decomposition of one level is applied to an image, four bands are obtained: an **LL** band or approximation coefficients band, and three detail coefficients bands called **HL**, **HH**, **LH**. Higher levels of decomposition over the **LL** sub band applied recursively yield into more detail coefficients. The coefficient sets of the different sub bands are $\mathbf{HL}_l$, $\mathbf{LH}_l$ and $\mathbf{HH}_l$ where $l \leq \Lambda$ is the level of decomposition to which the sub band belongs to. Also, let Θ^q be a set of quantized coefficient of a wavelet decomposition. The significance of a set is calculated as

$$\Gamma_n(S) = \begin{cases} 1 & \text{if } \exists (i,j) : |\Theta^q_{i,j}| \wedge 2^n > 0 \text{ with } (i,j) \in S \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where n is the power at which the significance is tested, Γ_n is the significance function at bitplane level n, $\wedge$ is the bitwise conjunction operator, $\Theta^q_{i,j}$ is the quantized coefficient at position (i, j) of the wavelet decomposition, and S is a quadtree in X. A quadtree is a set of coefficient coordinates defined as

$$S = \bigcup_{k=l}^{\Lambda} \bigcup_{x=0}^{2^{\Lambda-l}} \bigcup_{y=0}^{2^{\Lambda-l}} \{(2^k i + x, 2^k j + y)\} \tag{2}$$