Signal Processing 103 (2014) 331–347

Contents lists available at ScienceDirect

Signal Processing

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

Sparse representation of image and video using easy path wavelet transform

Syed Akbar Raza Naqvi *, Imran Touqir, Adil Masood Siddiqui

Department of Electrical Engineering, Military College of Signals, National University of Sciences & Technology, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 9 June 2013
Received in revised form
13 January 2014
Accepted 19 January 2014
Available online 4 February 2014

Keywords:
Sparse representation of image and video
Path vectors
Geometric wavelet transform
Adaptive and non-adaptive wavelet transform

ABSTRACT

This paper utilizes a wavelet based image compression technique, Easy Path Wavelet Transform (EPWT), for sparse representation of video. The algorithm has been extended for several aspects including Rigorous EPWT for three dimensional data compression and Relaxed EPWT for cost minimization. The proposed algorithm when applied on sampled frames of a video segment operates along permutations of frame indices exploiting the local correlations efficiently. These arrays of rearranged frame indices, in the case of redundant frames, are then uniquely stored while taking the repetition into consideration. The favorite direction criterion in the encoding process has also been modified to achieve enhanced sparsity for video segments while improving the existing quantitative results for images. The extended EPWT is effective and experimental results are comparable with existing state of the art techniques.

1. Introduction

The primary goal of all image and video coding techniques is to obtain a sparse representation of their significant features with minimum data loss. The conventional tensor product wavelet transform does not operate in the diagonal direction effectively therefore it is inadequate in the compression of images containing geometric features. This inability of wavelet transforms to adapt to the geometric features of images and videos leads to the research in areas of adaptive and non-adaptive image representation schemes. Instead of selecting a fixed frame as in classical wavelet transform one can adapt the function systems depending on the local frame structures. Based on this very notion, adaptive wavelet systems that include Wedgelets [1], Bandelets [2], Grouplets [3], Tetrolets [4], Generalized Tree Based Wavelet Transforms (GTBWTs) [5] and Easy Path Wavelet Transform (EPWT) [6,19] were developed. Similarly non-adaptive image compression techniques that comprise of Shearlets [7], Contourlets [8], Curvelets [9] and Directionlets [10] continued to be employed for effective image compression.

In EPWT, the process of evaluating and encoding a path vector resembles the coding techniques proposed for chain codes [11] with the major difference in the adaptability of the encoding symbols with respect to the varying directional features. The proposed methodology, in this paper, introduces an improved adaptive encoding technique for video segments while also improving its performance on images, with the primary aim of reducing the entropy of the respective significant features. In order to improve the efficiency and sparseness of the algorithm, modifications were made to the "favorite direction criterion" of the original algorithm in Ref. [6], resulting in reduced processing cost. In order to eliminate temporal redundancy multi-scale motion compensation [12] is performed that compresses the inter-frame difference by utilizing wavelet transformation. Non-Uniform Directional

* Corresponding author.
E-mail addresses: akbar.msee16@students.mcs.edu.pk (S.A.R. Naqvi), imrantqr@mcs.edu.pk (I. Touqir), dradil@mcs.edu.pk (A.M. Siddiqui).

http://dx.doi.org/10.1016/j.sigpro.2014.01.024
0165-1684

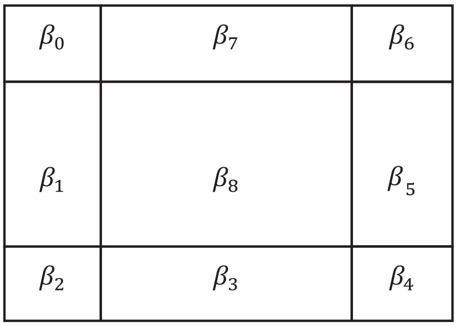

Fig. 1. Common neighborhood blocks of a frame.

Frequency Decomposition [13] for video representation designs non-uniform directional filter banks to make the geometrical transform sparse. This scalable video coding technique is equipped with dual multiresolution transform. The proposed EPWT, for video segments, avoids temporal redundancy by evaluating the inter-frame difference and applying the algorithm uniquely on redundant frames while storing the frame repetition. This in turn reduces the complexity while increasing the storage capacity. Adaptive three-dimensional wavelet transform based on lifting [14,15], permits the exploitation of the inter-frame redundancy, yielding PSNR results that are competitive with other video compression techniques. Other adaptive techniques that include adaptive lossless video compression using an Integer Wavelet Transform (IWT) [16], lossless video sequence compression [17] and adaptive vector quantization based video compression scheme [18] involve the removal of spatial, temporal and spectral redundancies by backward prediction systems and vector quantization.

This paper is the extension of the Easy Path Wavelet Transform proposed by Plonka [6] from image compression towards the sparse representation of video segments. The EPWT technique is preferred in this paper over existing wavelet based techniques because of its superior reconstruction results, less computational complexity and its unique ability to adapt to the directional features of the data under consideration. The algorithm proposed in Ref. [6] is improved with respect to sparse path vector evaluation in this paper and implemented on sampled frames such that the application of the improved EPWT results in an overall compression of the entire video segment. The paper is structured in the following manner. Section 2 discusses the important notations and terminologies used in this paper. Section 3 elaborates the Rigorous EPWT and its application on three-dimensional data. Section 4 emphasizes upon the modification of the Rigorous EPWT to the Relaxed EPWT such that the cost of storage is minimized resulting in economical storage of path vectors evaluated at all decomposition levels of EPWT. Section 5 provides the simulation results of both lossy and lossless reconstructions using EPWT algorithm. Comparisons of EPWT with existing adaptive wavelet transform techniques are also performed proving its efficiency and effectiveness for sparse representation of both 2D and 3D data. Finally, Section 6 deals with the summary of the entire research work proposed in this paper along with the possible future work.

2. Important terminologies

To facilitate the understanding of the Easy Path Wavelet Transform, some notations and symbols are required. Let M and N be positive integers and denote the dimensions of each frame of a video segment, such that $MN = 2^L$ where L represents the wavelet decomposition level. The input three-dimensional function is denoted by $v(m,n,k) : m,n,k \in \mathbb{N}, \Lambda k$ represents frame number in the video segment. This data is initially decomposed to its constituent frames from$k = 1$ to $k = k_{max}$. Where k_{max} denote the maximum number of frames. Each frame, k, now relates to a typical image represented by $f_k = (f_k(m,n))_{m=0,n=0}^{M-1,N-1}$. Let the index sets be represented as$I = \{(m,n) : m = 0,\dots,M-1,\ n = 0,\dots,N-1\}$. In order to have a one-dimensional representation of the original frame indices bijective mapping is used such that $S : I \to \{0,\dots,MN-1\}$. If $(m_1,n_1) \in I$, then the neighborhood, $\tilde{n}$, of an index $(m,n) \in I$ is defined by

$$\tilde{n}(m,n) = \{|m-m_1| \le 1, |n-n_1| \le 1 : m \ne m_1, n \ne n_1\} \quad (2.1)$$

If 'i' is the current index of the path vector and β_t be the block number where the subscript 't' ranges from $\{0,1,\dots,8\}$ then any frame can be divided into nine different blocks as indicated in Fig. 1 and denoted by the generalized neighborhood equations as follows:

$$\tilde{n}(i) = \begin{cases} \beta_0 \to \{(i+M),(i+M+1),(i+1)\} \\ \beta_1 \to \left\{ \begin{matrix} (i+M),(i+M+1),(i+1), \\ (i-1),\ (i+M-1) \end{matrix} \right\} \\ \beta_2 \to \{(i+M),(i-1),(i+M-1)\} \\ \beta_3 \to \left\{ \begin{matrix} (i+M),(i-M),(i-M-1), \\ (i-1),\ (i+M-1) \end{matrix} \right\} \\ \beta_4 \to \{(i+M),(i-M-1),(i-1)\} \\ \beta_5 \to \left\{ \begin{matrix} (i+1),(i-M+1),(i-M), \\ (i-M-1),\ (i-1) \end{matrix} \right\} \\ \beta_6 \to \{(i+1),(i-M+1),(i-M)\} \\ \beta_7 \to \left\{ \begin{matrix} (i+M),(i+M+1),(i+1), \\ (i-M+1),(i-M) \end{matrix} \right\} \\ \beta_8 \to \left\{ \begin{matrix} (i+M),(i+M+1),(i+1), \\ (i-M),\ (i-M+1), \\ (i-M-1),(i-1) \end{matrix} \right\} \end{cases} \quad (2.2)$$

Disjoint partitions E of $S(I)$ are used, i.e. $E = \{S_1, S_2, S_3, \dots, S_r\}$ where $S_u \cap S_w = \varnothing$ for $u \ne w$ and $\cup_{w=1}^{r} S_w = S(I)$. Two subsets S_u and S_w from E are neighbors i.e. $S_w \in \tilde{n}(S_u)$ if there exists two indices $i \in S_w$ and such that $i \in \tilde{n}(i_1)$. Path vectors through index subsets of $S(i)$ are evaluated and then a one-dimensional wavelet transform is applied along these path vectors. This complete path vector will constitute L number of path vectors. The entire path is composed of path vectors of all the frames at each level of decomposition i.e. $(p_k^1, p_k^2, p_k^3, \dots, p_k^L)_{k=1}^{k=k_{max}}$. EPWT algorithm is applied on the constituting frames of the input video after which the compression of each frame path vector is performed.