Video Compression Scheme using DEMD based Texture Synthesis

Mithilesh Kumar Jha, Brejesh Lall, Sumantra Dutta Roy
Department of Electrical Engineering, Indian Institute of Technology, Delhi
Jham73@gmail.com, brejesh@ee.iitd.ac.in, sumantra@ee.iitd.ac.in

Abstract— **In this paper we present a video coding scheme based on texture synthesis through Directional Empirical Mode Decomposition (DEMD). In this scheme P and B-frames of the video sequence are decomposed and parametrically coded with the help of DEMD algorithm, while I-frames are coded with the help of H.264. All P and B frames are decomposed into Intrinsic Mode Function (IMF) image and its residue. Only the first level IMF image for P and B frames are coded. At decoder stage subsequent IMF images are synthesized with the help of correlation search. Wavelet decomposition is performed over residual image and energy level at the HH band is used as a decision criterion for number of decomposition to be performed for optimum synthesis. The experimental result demonstrates the effectiveness of the algorithm in multi-resolution parametric modeling of image data which can be efficiently coded to achieve significant compression with acceptable quality. This scheme also enables to perform scalable coding of IMF parameters to achieve higher compression with perceptual quality.**

Keywords-Video Coding; Texture Synthesis; Directional Empirical Mode Decomposition (DEMD); Feature Value; Correlation Search; Intrinsic Mode Functions (IMF)

I. INTRODUCTION

Current day video compression schemes utilize prediction, motion compensation and transform based approaches to remove the inter-pixel redundancy. These schemes however do not exploit the perceptual redundancy across frames in the appearance of objects and textures present to attain higher compression rates. In recent times video compression has seen a lot of work focused on achieving higher compression with perceptually acceptable quality by means of texture synthesis. This is done because MSE reconstruction of texture is not required but a sample from the same random process is sufficient to maintain visual fidelity. In this paper we present a texture synthesis based architecture which enables to parse and encode the data separately for I, P and B frames in a video sequence and thus provides a framework to achieve very high compression with perceptual quality.

In this paper we propose a video compression scheme based on traditional video coding for I-frames followed by DEMD based decomposition and synthesis of texture pattern with P and B frames in a video sequence. Here both encoder and decoder structures have been proposed. At encoder: I-frames are encoded with the help of H.264. Texture images with P and B frames are decomposed based on DEMD criterion and only first level decomposed image is encoded with parametric approach. At decoder: I-frames are first decoded through the traditional H-264 approach and decomposed in to L level of IMF Images. The criterion for selection of decomposition level L is also suggested in the scope of this paper. With the help of L-level of IMF image from I-frame and first level IMF image from P and B-frames, all other IMF images for P and B-frames are synthesized through pattern correlation search based on feature criterion. Composing all levels of decomposed images a synthesized image for P and B-frames are generated.

The rest of the paper is organized as follows: section II, provides a literature review of recent work. Section III and IV describes the texture synthesis scheme in detail. The experiment and results are discussed in section 5 followed by a conclusion in section 6.

II. LITERATURE REVIEW

In the context of related work, C. Zhu, X. Sun, F. Wu and H. Li proposed a video coding scheme in which textural and structural regions are selectively removed in the encoder and reconstructed at the decoder by spatio-temporal texture synthesis and edge based in-painting [1-2]. To keep temporal consistency, they consider different types of motion in region removal and restoration for both textural and structural regions. Besides some parameter which can effectively guide restoration are also extracted and coded. In the proposed architecture the author apply their scheme to only B-frames, where as in our scheme processing are done for all P and B frames and thus providing the potential for higher compression.

Another texture based video coding approach is suggested by P. Ndjiki-Nya, B. Makai, A. Smolic, H.Schwaz and T. Wiegand [3-4]. In their scheme the video is divided into texture with non-important subjective detail and remainder and an improved coding scheme is proposed using a texture analyzer and synthesizer. The texture analyzer identifies the regions with texture on the basis of MPEG-7 descriptors and generates coarse mask and side information required by texture synthesizer at the decoder side. The texture synthesizer replaces identified texture with synthetic texture using warping based parametric method. The warp model takes care of rigid motion of texture while another part of synthesizer enables local motion activity in texture.

Similarly, A. Dumitras and B. G. Haskell [5] propose a video coding framework in which texture is removed from selected regions of the original frames. The resulting frames with the texture removed and the parameter of the removed texture are than encoded. At the decoder, the boundaries of the regions without texture are identified and new texture which is synthesized using the decoded texture parameters is mapped onto these regions.

Recently A. Khandelia, S.Gorecha, B. Lall, S. Chaudhury and M. Mathur [6] have suggested Video Compression scheme based on AR based texture synthesis. In this scheme each macro block is characterized either as an edge block or non-edge block containing texture. The non-edge blocks are coded by modeling them as an auto regressive process. AR model is applied in spatio-temporal domain to ensure both spatial as well as temporal consistency. Edge blocks are coded using standard H.264/AVC. This approach has however seen some limitation for structural texture synthesis. In our approach suggested here can be used with high quality for variety of structural textures.

Our scheme is different from all the above approach because we do not send parameters considering all source data for reconstructing the textures, instead we decompose the image based on DEMD approach and send only first level decomposed image for P and B frames. All other levels of P and B frames are synthesized based on pixel based feature correlation search. Multi resolution approach ensures that exact patches are found at each resolution with-in the decomposed image. Also this scheme can be seamlessly integrated with the existing H.264 compression scheme and can be used as a preprocessor block for better compression.

III. DIRECTIONAL EMPIRICAL MODE DECOMPOSITION OVERVIEW

Empirical Mode Decomposition is first proposed by N. E. Huang, Z. Shen and S. R. Long [7]. Its main idea is to decompose a given signal into a limited set of frequency components, called Intrinsic Mode Functions (IMF). Later Z. Liu, H. Wang and S. Peng in [8-10] proposed Directional Empirical Mode Decomposition (DEMD), which takes image direction into account in the decomposition and extracts three feature values for each pixel. DEMD is significantly different from the classical multi-scale structure. Firstly the distance between extrema is introduced to make the local scale. Therefore the DEMD decomposition is self adaptive and completely data driven. Secondly, an iteration method is adopted to extract each component. For these reasons the DEMD is locally self-adaptive and a unique advantage in extracting the content for visual perception.

The EMD decomposition of the original signal *x*(*t*) can be denoted by:

$$x(t) \; = \; \sum_{i=1}^{N} \mathrm{imf}_i(t) + r_N(t) \qquad (1)$$

Where $\mathrm{imf}_i(t)$ is the set of IMF components and the monotonic function $r_N(t)$ is the residue. These definitions can be extended to 2-D IMF and DEMD as follows: Given a signal, $u(x, y)$, $\theta \in \left[0, \left(\pi/2\right)\right]$ and c ∈ R, the bi-dimension IMF for associated θ should satisfy following conditions:

$$\begin{aligned} v_{1,c}^{\theta}(x) &= u(x, [\tan\theta]x + c) && 0 \le \theta < \tfrac{\pi}{2} \\ v_{2,c}^{\theta}(x) &= \begin{cases} u(.,x) & \theta = 0 \\ u\left(x, \left[\tan\left(\theta + \frac{\pi}{2}\right)\right] x + c\right) & 0 < \theta < \frac{\pi}{2} \end{cases} && (2) \end{aligned}$$

Where $v_{1,c}^{\theta}(x)$ and $v_{2,c}^{\theta}(x)$ satisfy both of the IMF conditions.

DEMD of a 2-D image f(x, y) for angle θ is defined as the following decomposition:

$$f(x, y) \; = \; \sum_{i=1}^{N} \mathrm{imf}_i^{\theta}(x, y) + r_N^{\theta}(x, y) \qquad (3)$$

Where $\mathrm{imf}_i^{\theta}(x, y)$ is a bi-dimensional IMF on θ, and there exists at least a monotonic one-dimensional sample for $r_N^{\theta}(x, y)$on $v_{1,c}^{\theta}(x)$ and $v_{2,c}^{\theta}(x)$.

Similar to one dimensional IMF, “sifting algorithm” is used to obtain bi-dimensional algorithm as described by Y. Zhang, Z. Sun, and W. Li [11]. Thus for any given image f(x, y), we can decompose it in the form (3), which use the instant frequency of the decomposed image as the signature and the algorithm has an excellent local self adaptability. Fig. 1 shows a texture image, it’s first two IMF images and the residue obtained by DEMD respectively.

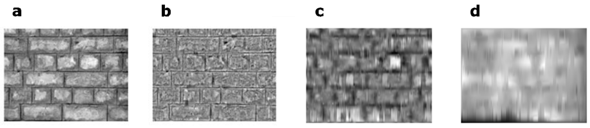

Fig. 1: (a) original image, (b) first IMF image, (c) Second IMF image, (d) Residue Image

IV. APPLYING DEMD IN TEXTURE SYNTHESIS

We propose DEMD for texture synthesis with the idea that each texture is having some inherent direction which can be exploited effectively while decomposing the frame into the IMF images and its residue. We propose to extract the first level IMF image for P and B frames at encoder stage and code them parametrically. I-frames are coded with the help of standard block based video coding approach like H.264. The detailed implementation process is as follows:

A. Encoding Flow (Fig. 2)

- Given a video sequence S with,

  S = {FI, FP, . . . ,FB}, where Fi, i = I, P, B represents the I, P and B-frames in the sequence. Decompose these frame sequence into its IMF images and residue as demonstrated in Fig. 3

- Extract all the extremas for first level IMF images against P and B frames and encode them as follows:

If, total number of maxima is N_{max}, total number of minima is N_{min}, maximum variation in maxima is V_{max}, maximum variation in minima is V_{min} than the total number of bits required to code the (N_{max}-1) maximas and (N_{min}-1) minimas are

$$\{Nmax-1\}*log2Vmax+ \{Nmin-1\}*log2Vmin \qquad (4)$$

Also the differences between two locations of extremas are coded. If the difference is D, the required bits to code the same is log2(D).

- Encode I-frame with the help of standard block based compression scheme like H.264

B. Synthesis Flow (Fig. 4)

- Decompose I frame into L IMF levels. The number of IMF level, L is decided based on the energy content with the residue image. The energy in the HH band of the residue image is computed from wavelet decomposition and checked against the threshold. If the energy content is greater than the threshold, further IMF decomposition and synthesis is continued. In our experiment threshold is fixed to a pre-defined value
- There are three feature values extracted for each pixels using DEMD approach. These include instant frequencies and envelope. These feature values are used during correlation search to calculate and minimize the visual deviation. The following formula is used to calculate the deviation error.

$$\boldsymbol{d(V_{in(imf_i)}, V_{out(imf_i)})} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\boldsymbol{V^{j}_{in(imf_i)}} - \boldsymbol{V^{j}_{out(imf_i)}})^2} \qquad (5)$$

Where $V_{in(imf_i)}$ refers to a particular boundary zone of the ith IMF image and $V_{out(imf_i)}$ refers to the boundary zone of the current position in the corresponding synthesized ith IMF image. N denotes the number of pixels in the boundary zones.

- We use correlation search to ensure the inherent correlations among all levels of IMF images. This implies that the search range in each level of IMF image is determined by the corresponding patch in the upper levels of the IMF images. Once all the IMF levels are synthesized the final synthesized texture can be obtained by rotating inversely the synthesis result in the inherent direction of the original texture. Fig. 2 shows the pictorial representation of the IMF based synthesis approach.

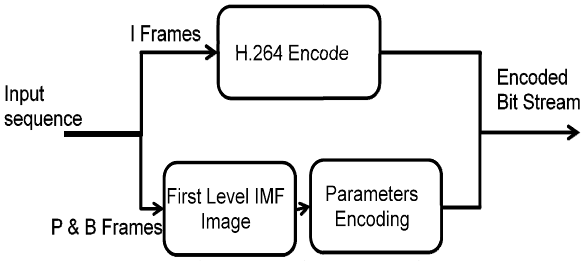

Fig. 2: Encoder Framework

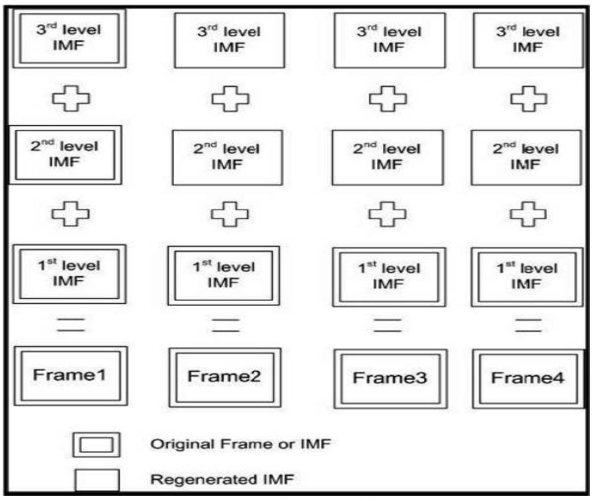

Fig.3 : IMF decomposition structure

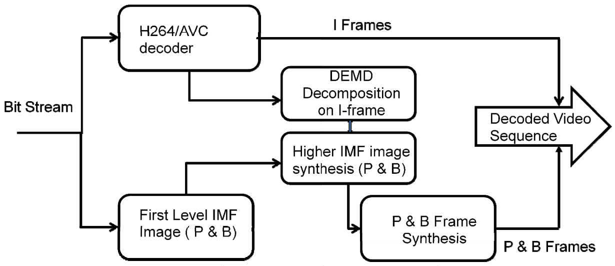

Fig.4: Decoder Framework

V. EXPERIMENTS AND RESULTS

In our experiments four video sequences are selected. All the experiments are carried out by MATLAB code, running on a desktop PC with P4 CPU and 1 GB of main memory and windows XP as an operating system. The results presented in Table I, do not include any quantization and entropy coding of the IMF parameters (maxima and minima) hence the compression ratio is symbolic and can be further improved significantly. Fig.6 and Fig.7 demonstrates the synthesis results for brick video and escalator video sequences. The scalability of the algorithm is also proved by sending higher level of IMF with fewer maxima and minima to achieve more compression on the cost of quality. Further to that it is experimentally proved that the compression can be further improved by applying a probabilistic approach to send only those parameters which have maximum information and discard others as shown in Fig. 5.

TABLE I. EXPERIMENTAL RESULTS AND COMPRESSION DATA

Sequence Name	No. Of Maxima	No. Of Minima	Encoded size (In Bytes)	Raw Frame size (In Bytes)	Compression Ratio (In %)	PSNR (In db)
Brick Video (128 x 128)	3961	3658	11429	16384	69.75	24.08
Circular Hole Video (128 x 128)	3003	2883	9940	16384	60.66	60.57
Escalator Video (128x128)	2956	2955	10344	16384	63.13	30.83
Floor Box Video (128x128)	2448	2476	8001	16384	48.83	24.07

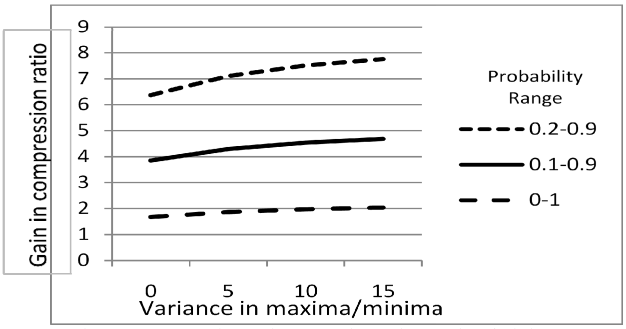

Fig. 5: Compression gain vs. variance in maxima/minima

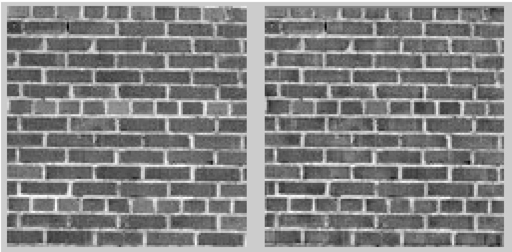

Fig. 6. Original (Left) and reconstructed (Right) Frame for brick video with 1st level IMF for P & B frames

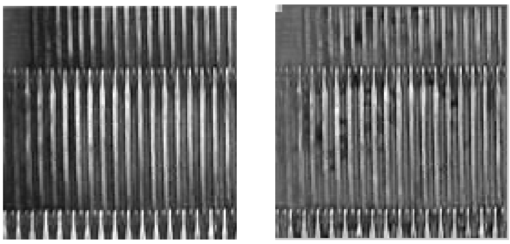

Fig. 7. Original (Left) and reconstructed (Right) Frame for escalator video with 1st level IMF for P & B frames

VI. CONCLUSIONS

In this paper a DEMD based video compression scheme is proposed. In this method only first level IMF Image is parametrically coded for P and B-frames in a given video sequence, while the I-frames are proposed to be coded with the help of standard H-264 video coding scheme. The synthesis approach is based on feature correlation search at different IMF level in the nearest reconstructed frame. The number of IMF level decomposition is computed based on the residue energy with the HH band of the wavelet decomposition. The scheme can be further enhanced with more optimized parameter coding approach to receive even better compression. Since this approach is based on parametric coding compression vs quality trade off can be easily done based on the application requirements. Also the comparative study of the results against the standard video codec like H.264 is targeted as future work.

REFERENCES

[1] C. Zhu, X. Sun, F. Wu and H. Li, “Video Coding with Spatio-Temporal Texture Synthesis and edge-based inpainting,” IEEE International Conference on Multimedia and Expo (ICME), 2008.

[2] C. Zhu, X. Sun, F. Wu, and H. Li, "Video Coding with Spatio-Temporal Texture Synthesis," IEEE International Conference on Multimedia and Expo (ICME), 2007.

[3] P. Ndjiki-Nya, B. Makai, A. Smolic, H.Schwaz and T. Wiegand, “Improved H.264/AVC Coding using Texture Analysis and Synthesis,” Proc. of ICIP, IEEE International Conference on Image Processing, 2003.

[4] P. Ndjiki-Nya, C. Stuber and T. Wiegand, “Texture Synthesis method for generic Video Sequence,” in Proc. of IEEE International on Image Processing, 2007.

[5] A. Dumitras, B. G. Haskell, “An Encoder–Decoder Texture Replacement Method With Application to Content-Based Movie Coding”, IEEE Transactions on Circuits and systems for Video Technology, Vol. 14, No. 6, June 2004.

[6] A. Khandelia, S.Gorecha, B. Lall, S. Chaudhury and M. Mathur, “ Parametric Video Compression Scheme using AR based Texture Synthesis,” in Indian Conference on Computer Vision, Graphics and Image Processing, 2008.

[7] N. E. Huang, Z. Shen, S. R. Long, “The Empirical Mode Decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”. Proceedings of Royal Society of London (Series A) 1998;454(1):903–95.

[8] Z. Liu, H. Wang, S. Peng, “Texture classification through Directional Empirical Mode Decomposition”. in: Proceedings of the 17th International Conference on Pattern Recognition (ICPR), 2004.

[9] Z. Liu, H. Wang, S. Peng. “Texture segmentation using Directional Empirical Mode Decomposition”. in: 2004 International Conference on Image Processing (ICIP), 2004, p. 279–82.

[10] Z. Liu, S. Peng. “Directional Empirical Mode Decomposition and its application to texture segment”. Science in China Series E: Information Sciences 2005;35(2):113–23.

[11] Y. Zhang, Z. Sun, and W. Li, "Texture synthesis based on Directional Empirical Mode Decomposition," Computer and Graphics, 2008, pp. 175-186.