

Fast Quadtree Based Normalized Cross Correlation Method for Fractal Video Compression using FFT

R. E. Chaudhari[†] and S. B. Dhok*

Abstract – In order to achieve fast computational speed with good visual quality of output video, we propose a frequency domain based new fractal video compression scheme. Normalized cross correlation is used to find the structural self similar domain block for the input range block. To increase the searching speed, cross correlation is implemented in the frequency domain using FFT with one computational operation for all the domain blocks instead of individual block wise calculations. The encoding time is further minimized by applying rotation and reflection DFT properties to the IFFT of zero padded range blocks. The energy of overlap small size domain blocks is pre-computed for the entire reference frame and retaining the energies of the overlapped search window portion of previous adjacent block. Quadtree decompositions are obtained by using domain block motion compensated prediction error as a threshold to control the further partitions of the block. It provides a better level of adaption to the scene contents than fixed block size approach. The result shows that, on average, the proposed method can raise the encoding speed by 48.8 % and 90 % higher than NHEXS and CPM/NCIM algorithms respectively. The compression ratio and PSNR of the proposed method is increased by 15.41 and 0.89 dB higher than that of NHEXS on average. For low bit rate videos, the proposed algorithm achieve the high compression ratio above 120 with more than 31 dB PSNR.

Keywords: Fractal video coding, Quadtree, FFT, Motion estimation, SSIM, Normalized cross correlation

1. Introduction

Video compression using fractal transform is a new approach to video coding. Jacquin [1] proposed the first algorithm for image compression using block wise fractal coding. It reduces the affine redundancy existing in various parts of images by using a self - similarity feature which can give high compression ratio and fast decompression [2, 3]. Lin [4] described a frequency domain based image coding method for best matched solution. Video information is characterized by its large storage requirement and it is also difficult to transmit raw video data efficiently. Fractal video compression seems a favorable method to achieve high compression ratios [5]. But fractal codec is not popular because of its computational complexity and time-consuming search for the best-matched block in a pool of domain blocks. In order to overcome the above difficulties a fast quad-tree based normalized cross correlation using FFT is proposed for fractal video compression to improve the encoding speed and to achieve optimal performance.

Cube-based [3, 6, 7] and frame-based [8, 9] fractal

compression methods are used more frequently for video compression. In cube-based compression a video is divided into groups of frames, each of which in turn is partitioned into three dimensional (3D) domain and range blocks. It has high computing complexity and low compression ratio. In the frame based compression, each frame is encoded using the previous frame as a domain pool. The current frame is related to the previous frame which introduces and spreads the error between frames. It may be used to obtain a high compression ratio. Wang proposed a fixed block size hybrid compression algorithm [10] and an adaptive partition instead of fixed-size partition [11], which merges the advantages of a cube-based and frame-based fractal compression method. Another hybrid coder scheme which combines neighborhood vector quantization with fractal coding to compress the video as a 3D volume [12]. For low bit rate videos, fractal approach [13] is used in those areas where the residual error between two consecutive frames is exist.

Circular prediction mapping (CPM) and non-contractive inter-frame mapping (NCIM) is proposed by Kim [14], to combine the fractal sequence coder with well-known motion estimation/motion compensation algorithm that exploits the high temporal correlations between the frames. In CPM/NCIM, domain block is of the same size as the range block and each range block is motion compensated by a domain block in the previous frame. The main difference between CPM and NCIM is that CPM should be

[†] Corresponding Author: Dept. of Electronic Engineering (center for VLSI and Nanotechnology), Visvesvaraya National Institute of Technology, Nagpur, India. (rec77@rediffmail.com)

* Dept. of Electronic Engineering (center for VLSI and Nanotechnology), Visvesvaraya National Institute of Technology, Nagpur, India. (sanjaydhok@gmail.com)

Received: December 18, 2014; Accepted: October 20, 2015

contractive and iterative for decoding process while NCIM is non-contractive and non iterative since the decoding depends on already decoded frames. A new object based CPM/NCIM with quadtree partition is introduced by zhu, [15] in which alpha map is used to segment the image into objects and each object is encoded independently.

In this article, a fast normalized cross correlation algorithm for fractal video coding is proposed. It is based on the combination of two concepts, i.e. quadtree decomposition and the normalized cross correlation (NCC) applied in the frequency domain using FFT. Quadtree partitioning scheme provides a good level of adaption to picture contents. The optimal block size decreases the motion compensated prediction error without increasing the large computations. Cross correlation is evaluated in faster manner by using FFT and the correlated coefficients of all domain blocks are calculated in a single computation. In addition to this, substantial reduction in computation for eight isometry transformations is achieved by applying the rotation and reflection DFT properties on the IFFT of zero padded range blocks. Motion estimation speed can be further increased by two ways, pre-computing the energy of reference frame according to the smallest block size and other by retaining the energies of the overlapped search area of previous block because two nearby range blocks shared more than 50 % of blocks energy of each other. This method will improve the subjective quality of the video as well as the coding efficiency. The structural similarity (SSIM) index [16] distortion measure is more consistent with human perception to measure the video quality as compared to Peak Signal to Noise Ratio (PSNR).

Most of the fast algorithms reduce the computational complexity at the cost of the accuracy of motion estimation. The full search algorithm obtains the optimal result of exhaustive searching for the best matching block. In this paper, we try to achieve better performances of video coder in terms of three parameters, i.e. speed, compression ratio and output quality.

The rest of the paper is structured as follows. Cross correlation related and other fractal video coder are compared in section 2. The basic fractal block coding for the image is described in Section 3. Normalized cross correlation for motion estimation is presented in section 4. The proposed fast fractal video coding including quadtree partitions and new fast fractal prediction algorithm is explained in section 5. The experimental results and comparative study of the proposed algorithm with existing algorithms are presented in section 6. The conclusion is outlined in section 7.

2. Related Work

FFT based fractal image coding was proposed by Hannes [17] to speed up the encoding computations. The collage error between the range and domain image is

measured using five different inner product operations. Each inner product implementation uses the FFT based cross correlation operation. In [17], the quantized gray level transformation (s and o) parameters are calculated for each domain block to determine the collage error. The mean subtracted normalized cross correlation using FFT is presented in [18] to find similarity between range and domain block. The computation of energies of mean subtracted overlapped domain blocks is computationally intense. Fractal image coding point of view, single computation of domain image is required for all the range blocks. However, in frame based fractal video coding search area of all range blocks are different and overlapped with other search areas.

Fractal video coding using a new cross-hexagon search (NHEXS) algorithm is proposed [19] for higher motion estimation speed for searching stationary and quasi-stationary blocks. Initially, it uses two cross shaped search patterns and then large/small hexagon-shaped Patterns in the subsequent steps. NHEXS employs halfway stop technique and modified partial distortion criterion (MPDC) to minimize encoding time. Video sequences are also encoded by region-based approach [20] using NHEXS. The regions can be defined according to the previously computed segmentation map and are encoded independently of each other. Object based stereo video compression using the combinations of shape-adaptive DCT and fractals is introduced in [21]. Mobile videos are compressed using fractal, [22] in which genetic algorithm and particle swarm optimization techniques are used to improve the quality of video and speed up factor respectively. For low bit rate videos, intercube correlation search in the spatial and spatial-temporal directions is presented in [23] to increase the coding performance. Motion and non-motion wavelet subtrees of each inter frame are coded separately using a fractal variable tree with set partitioning algorithm [24, 25], which is suitable for low bit rate videos.

Based on the idea of [18] fractal image coding, fractal video compression without mean subtracted NCC using quadtree partition is proposed in this paper. The best matched domain block having high NCC value may have large average gray level difference. This difference is reduced to zero or very small value by selecting proper fractal encoding parameters. Fractal based video compression algorithm operating in the frequency domain, pre computing and re-using the energies of adjacent overlap search areas to speed the encoding process are the main features of our work.

3. The Fractal Block Coding

Fractal image coding is based on the theory of the partitioned iterated function system (PIFS) [1]. It consists of a set of contractive transformations, when this transformation is applied iteratively to an arbitrary image it

will converge to an approximation of the original image. Images are stored as a collection of transformations will result in image compression.

The original image of size $M_1 \times M_1$ is initially partitioned into non-overlapping range blocks (R_i) of each size $N \times N$ $\{i=1, 2 \dots M_1^2/N^2\}$. Similarly, the same image is partitioned into overlapping domain blocks (D_i) of each size $2N \times 2N$ as a domain pool with 'k' pixel shift in horizontal and vertical direction $\{j=1, 2 \dots (M_1-2N+1)^2\}$. For each range block, locate the best matching domain block from the domain pool and then apply contractive mapping which minimizes the mean squared error (MSE) between range and contractive domain block. A range - domain mapping consists of three operations sequentially on each domain block of size $2N \times 2N$ such as, 1) Spatial contraction of the domain block by down sampling or averaging the four neighboring pixels of disjoint group forming a block of size $N \times N$. 2) Take 8 geometrical transformations of each block which includes 4 rotations with 90 degrees and 4 reflections. 3) For each geometrical block perform contractive affine transformation on the grayscale values to minimize the MSE. The error between range and domain block is given by Eq. (1) and scaling factor 's' and brightness factor 'o' of an affine transformation are given by Eq. (2) and (3) respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^n (s \cdot d_i + o - r_i)^2 \tag{1}$$

$$s = \frac{n \sum_{i=1}^n r_i d_i - (\sum_{i=1}^n r_i)(\sum_{i=1}^n d_i)}{n \sum_{i=1}^n d_i^2 - (\sum_{i=1}^n d_i)^2} \tag{2}$$

$$o = \frac{1}{n} \left[\sum_{i=1}^n r_i - s \sum_{i=1}^n d_i \right] \tag{3}$$

Where n is the number of pixels in the block and r_i, d_i are the pixel values of the range block and contractive domain block. For each range block, the parameters which need to be stored as a fractal encoded data are the coordinates of domain block along with s, o and geometric transformation index. Coefficients of s should be in the range of -1 to 1 to make sure that the transformation is contractive.

4. Normalized Cross Correlation for Motion Estimation

In fractal video coding, finding the best-matched domain block with proper affine transformation is achievable if the blocks are structurally similar. Applying the NCC as the similarity measure to motion estimation leads to more uniform residuals between the range and best domain block. It can improve the coding efficiency and subjective visual quality in video compression. Hence, the structural

similarity (SSIM) is a more appropriate for quality measurement of the human visual system. NCC operates on the macroblocks (MB) of the two consecutive frames of a moving sequence [26-29]. Motion estimation is done by taking the MB (R) from the current frame and this MB is shifted pixel by pixel across the search window (W) of the reference frame. To find the motion vector of each MB using the NCC is defined as

$$C_{R,W}(x,y) = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R(i,j) \cdot W(x+i,y+j)}{\sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} W^2(x+i,y+j)} \cdot \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} R^2(i,j)}} \tag{4}$$

where $x \in \{0,1,2, \dots, 2V\}$ and $y \in \{0,1,2, \dots, 2V\}$

This calculation is repeated for each pixel location across the entire search window. $C_{R,W}(x,y)$ represents the NCC matrix between the current MB and reference search window. If the window area is $\pm V$ pixels in the both directions from the same position of R , then the size of the $C_{R,W}(x,y)$ will be $(2V+1, 2V+1)$. The coordinates x_m and y_m of the maximum correlation value in the NCC matrix, $C(x_m, y_m) = \max(C_{R,W}(x,y))$ can be used as a motion vector of the R . In Eq. (4), numerator part is the cross correlation of two blocks and the denominator represents a multiplication of square root of energies of two MB's. NCC method is more robust under uniform illumination changes and less sensitive to noise, but computationally intense and time consuming. For identical regions, it may give a high NCC value for the best match, but with large average gray level difference.

5. New Fractal Video Coding

All blocks based motion estimation methods attempts to find the block with smallest MSE as a matching criterion. Similarly, in fractal video compression the block with smallest MSE after performing geometrical and contractive affine transformation on each domain block is taken as a best matched block. It can be interpreted as a kind of motion compensation technique due to unavailability of error frame. Video sequence exhibits a high temporal correlation so range-domain mapping becomes more effective if the size of range and domain block is same even though the size of domain block is always greater than range block size in fractal image coding [8].

Intra frame coding makes a great impact on decoded video quality and compression ratio. If inter frame motion vector prediction is based on good quality reference frame (previous frame) then prediction error will be low. But for subsequent frames this error monotonically increases due to cumulative process. So after some interval of frames this problem can be resolved by inserting intra frame. To get a better quality reference frame at lower compression ratio,

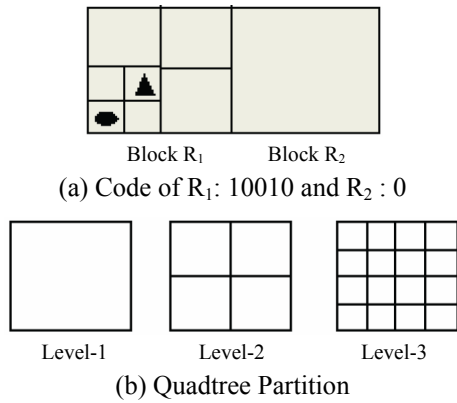


Fig. 1. (a) Code of two range blocks, (b) 3 Level quadtree partition

intra frame is compressed by using 8×8 DCT transformation / quantization techniques [30].

5.1 Quadtree partition

Quadtree partition [31] makes a large impact on the calculation of compression ratio and encoding speed of video compression. Here, 3 levels of quadtree partitioning are employed with block sizes from 16×16 to 4×4 pixels. The current frame is initially partitioned into bigger size (16×16 pixels) range blocks, i.e. level-1 and then each block is partitioned as shown in Fig. 1 (b). Partition criterion is based on the prediction MSE error between range and domain block as per Eq. (1). If the MSE of the bigger size range block is smaller than a predefined threshold, then save the fractal code and process the next block. Otherwise, that block is further partition into 4 quadrants i.e. level-2. The above procedure is repeated in each quadrant until the MSE becomes smaller than the threshold or 4×4 pixel block (level-3) is created. The algorithm is processed sequentially and iteratively on a block basis starting from top left to bottom right of each frame. Each subsequent partition of 16×16 block size is represented by code as shown in Fig. 1 (a). In general, it takes on an average 1 bit or 5 bits per 16×16 block size. Quadtree partitioning information is encoded separately using the following procedure to substantially save the number of bits required to denote quadtree level.

If the matched block size is 16×16 then it is denoted by index ‘0’ and the next range block is encoded. Otherwise assign index ‘1’ and further partition the block and go to the next step.

If the matched block size is 8×8 then it is denoted by index ‘0’, otherwise it is denoted by index ‘1’ and this step is repeated for the next 3 blocks of size 8×8 . Here, the block indexing ‘1’ represents that block is further partitioned into 4 quadrants as shown in Fig. 1(a) i.e. process all four child blocks (at level 3) one-by-one if the parent block (at level-2) index is 1.

5.2 Fast fourier transform based computations

To calculate the DFT, Cooley and Tukey has invented a Fast Fourier Transform (FFT) algorithm [18, 32]. The most commonly used FFT algorithm operates on the principle of divide and conquers approach. When the algorithm is dividing the transform into two sequences of size $N/2$ recursively in each sub sequences then it is called as radix-2 FFT. If the input signal $x(n)$ consist of N number of samples then the DFT is given as

$$X(k) = \sum_{n=0}^{N-1} x(n).W^{nk} \tag{5}$$

where $W^{nk} = e^{-\frac{j2\pi.nk}{N}}$ and $k=0,1,2, \dots, N-1$

The direct computation of equation (b) requires on the order of N^2 complex arithmetic operation. While FFT offers a major reduction in computational complexity to the order of $[(N/2) \times \log_2 N]$. Due to incredibly reduction in the complexity, FFT is more efficient than DFT to produce the same result. Using the separability property of DFT, 2D FFT can be computed by performing 1D FFT on row wise followed by column wise transformation. FFT algorithm is also used to determine the convolution or correlation of two sequences with significant reduction in complexity.

5.3 Frame based fractal video coding

Current frame (f_{t+1}) is initially partitioned into non overlapped range blocks of size 16×16 pixels. In fractal coding if the number of domain blocks is more, than the probability to get best matched block with contractive affine transformation will be more. In this paper search area is selected as ± 8 pixels in both horizontal and vertical direction from the same location of the range block. The search window is defined on the previously decoded frame (f_t) i.e. reference frame. The flow diagram of the proposed fractal video coder is shown in Fig. 2.

In video sequences, most of the blocks can be observed as stationary or non-motion block. The NCC matching process may find the wrong matched domain block if the range block is homogeneous. To avoid this problem, it is verified whether the range block is stationery or homogeneous in the beginning itself and then these blocks are coded separately. The early termination of the search process, in case the range block is homogeneous or zero/non-zero motion with zero luminance factors is employed. It significantly reduced the computational cost for determining the motion vector as possible. Otherwise, fractal parameters with MSE error E_2 for each block are determined from the fast fractal prediction algorithm. If the isometry transformation index is no change, i.e. $t=0$ then that block is again checked for exact self-similarity and MSE (E_1) is calculated. Final fractal information depends

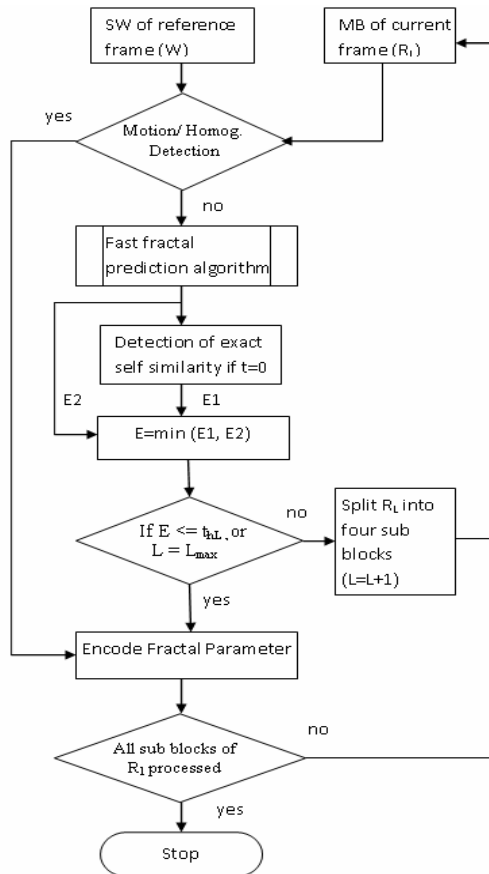


Fig. 2. Flow diagram of the proposed fractal video coder

on the minimum of two MSE errors E1, E2. If the range block cannot find a good matching domain block within the given MSE error threshold, then it is divided into four quadrant blocks and each sub blocks are encoded separately. Since the depth of quadtree is 3, so the two error thresholds are set i.e. t_{h1} and t_{h2} . The algorithm proceeds iteratively from level 1 to level2 and from level2 to level3. It terminates when either the given threshold condition is satisfied for all the sub blocks or smallest block size has been reached.

For each and every range block, all the fractal parameters including the quadtree indexing need to be encoded. The encoded information can be used for storage or transmission of sequences. We encode the position or motion vector of domain block with respect to range block using 5-bit codeword. 5-bits and 7-bits are used for quantized value of s and o parameters respectively to give good quality output. Finally, use 3-bit codeword for all the isometry transformations.

5.4 Fast fractal prediction algorithm

Range block ' R_L ' (size $N \times N$) and the search window ' W_L ' (size $(N+16) \times (N+16)$) are the two inputs for fast fractal prediction algorithm. Computational complexity of NCC can be reduced by implementing the numerator of

Eq. (1) in the frequency domain using FFT. But, the primary requirement of cross correlation in frequency domain is that two input blocks must be of equal size. The range block size is increased by padding zeros on the left and down the side of the block and makes it equal to search window size. Cross correlation operation is the complex conjugate multiplication of two frequency domain components (range block and search window). This complex conjugate term additionally increases the complexity. It is minimized by taking the inverse FFT instead of forward FFT of one of the input. If the input signal is real then its conjugate of the FFT is equal to IFFT of same input with constant multiplying term. Without using this constant also final required NCC result will not differ. Numerator of Eq. (4) using 2D-FFT (FFT) and the same equation as a NCC matrix is given as Eq. (6).

$$C_{R_L W_L} = \frac{IFFT[FFT(W_L), IFFT(R_L)]}{\|R_L\| \cdot \|W_L\|} \quad (6)$$

Where $\|\cdot\|$ is a 2-norm, L is quadtree index, if $L=1$ then $N=16$, if $L=2$ then $N=8$ and if $L=3$ then $N=4$.

$$C(x_m, y_m) = \max\left(\left|\operatorname{Re}\{C_{R_L, W_L}\}\right|\right) \quad (7)$$

Form Eq. (6) $\|R_L\| = \sqrt{\sum_{i,j=0}^{N-1} R_L^2(i, j)}$ and

$$\begin{aligned} \|W_L(x, y)\| &= \sqrt{\sum_{i,j=0}^{(N/4)-1} g_L(x + (4 \times i), y + (4 \times j))} \\ &= \sqrt{\sum_{i,j=0}^{N-1} W_L^2(x + i, y + j)} \end{aligned}$$

where $g_L \subseteq g$ and $x, y = 0, 1, 2, \dots, 16$

$$g(m, n) = \sum_{i,j=0}^3 f_{i-1}^2(m+i, n+j),$$

Where $m = \{0, 1, 2, \dots, M_2-4\}$ and $n = \{0, 1, 2, \dots, N_2-4\}$, f_{i-1} - The previous reconstructed frame of size $M_2 \times N_2$.

Where $\|R_L\|$ represents the square root of energy of range block and $\|W_L\|$ is the 2-norm matrix (size 17×17) of search window according to the size of range block. Start pixel location of g_L and W_L are same from g and f_{i-1} respectively. Each x and y coordinates location, $\|W_L(x, y)\|$ or $\|W_L\|$ represent the energy of corresponding overlapped domain block of size $N \times N$. C_{R_L, W_L} is the NCC matrix with size 17×17 and $\operatorname{Re}\{ \cdot \}$ represent the real part of the NCC array block. The coordinates (x_m, y_m) of Eq. (7) denote the position of highly correlated block and can be used as an estimation of motion vectors. MSE will be minimized by selecting the highly correlated domain block. This

correlated domain block is used to match with range block using the scaling and brightness parameters at different isometry transformations as per Eq. (1). If the scaling and brightness parameter values are outside the predefined ranges, then select the next correlated block and repeat the procedure till they satisfies the conditions $|s| \leq 1.0$ and $|o| \leq 255$, so that fractal parameter can be quantized effectively. Fig. 3 shows the flow diagram of fast fractal prediction algorithm which returns all the fractal parameters along with the MSE error of the particular range block.

This proposed algorithm has the difficulty of applying an isometry transformation to the individual domain blocks since it operates on entire domain image (search window) and uses the frequency domain. Hence the isometry transformations are proposed for the range block. IFFT of original range block and FFT of search window are of equal size due to padding zero of range block. IFFT of the isometry transformed range block is expressed in terms of IFFT of the original range block using the rotation and reflection properties of 2D IDFT [32]. This helps to minimize the IFFT calculations and increase the searching speed of the algorithm.

A further number of computations to compute the energy

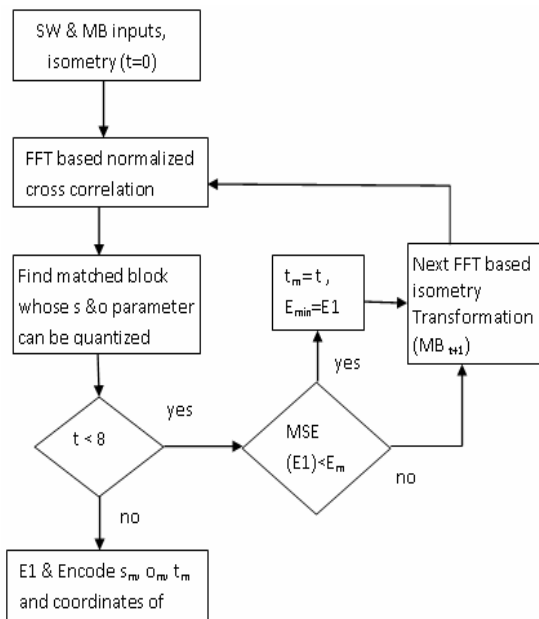


Fig. 3. Flow diagram of the fast fractal prediction algorithm

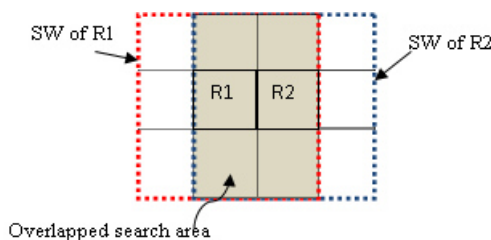


Fig. 4. Overlapped search area of two adjacent range blocks

of the search window are significantly reduced by using pre-computed and reusing the energy of overlapping search area. The energy $g(m,n)$ of the reference frame f_{t-1} is calculated and stored according to the small block size and acts as a look-up table. As the range block size changed from level 1 to level 2 or 3, energy of search window gL is selected from the look-up table and computed with respective larger block size with minimum computations. Two adjacent range blocks (R1 and R2) share their more than 50 % search areas as shown in Fig. 4. If the two adjacent blocks are of the same size, then reusing the energy of first block overlapped search regions for the next block will be beneficial to avoid unnecessary computations.

6. Experimental Results

The proposed NCC based fast fractal video compression algorithm is simulated and tested on 7 real video sequences ('Foreman', 'Carphone', 'News', 'Paris', 'Tennis', 'Bus', 'Highway'). The tested videos are gray-level image sequences [33] with the size of 352×288 pixels. Range blocks are formed according to quadtree partition criteria with minimum and maximum block sizes of 4×4 pixels and 16×16 pixels respectively. A search area on the reference frame is ± 8 pixels in both vertical and horizontal directions from the same position as of range block on the target frame. Along with proposed method, CPM/NCIM and NHEXS algorithms are also implemented to compare their performance. The prerequisite information like quadtree partition with block size, two thresholds and the encoded bit length for fractal parameters are the same for all the implemented algorithms. In CPM/NCIM method first 3 frames are set for CPM and the remaining frames are using NCIM. All the three methods including proposed method are implemented in MATLAB 7.14 and simulated on PC (Intel Core i5-2400 CPU, 3.10GHz, 3.16GB RAM).

The output reconstructed frame (\hat{f}) quality in comparison with original frame (f) is measured by PSNR as per Eq. (8)

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{mse(f, \hat{f})} \right) \tag{8}$$

where

$$mse(f, \hat{f}) = \frac{1}{M_2 \times N_2} \sum_{i,j=0}^{M_2, N_2} (f(i, j) - \hat{f}(i, j))^2$$

Table 1 shows the comparison of average coding results of the first 60 frames of each of five video sequences. The result shows that the proposed method improves the compression ratio (CR) 2.5-9 times, PSNR 4 - 8 dB more and encoding time 16-34 times in comparison with traditional CPM/NCIM. Similarly, CR is increased by 6-28,

Table 1. Comparison of average fractal coding results of video sequences

Video sequences	Compression Ratio (CR)			PSNR (dB)			Encoding Time (Sec.)		
	Proposed	CPM/NCIM	NHEXS	Proposed	CPM/NCIM	NHEXS	Proposed	CPM/NCIM	NHEXS
Foreman	70.84	28.46	64.64	33.47	29.29	33.11	2.88	46.81	3.76
Carphone	129.27	47.86	112.50	36.19	32.18	35.58	1.80	25.54	2.27
News	356.56	38.86	328.42	38.63	30.48	38.18	0.93	33.11	1.18
Paris	44.39	11.82	24.75	30.88	24.11	28.56	4.14	105.48	13.98
Tennis	62.52	18.54	56.21	31.58	28.09	30.67	3.51	66.47	4.84

Table 2. Comparison of average compression ratio and PSNR of 15 frames

Video sequences	Compression Ratio (CR)		PSNR (dB)		SSIM
	Proposed	Region Based [20]	Proposed	Region Based [20]	
Foreman	66.00	62.53	33.42	32.31	0.9386
Paris	45.52	45.83	30.61	33.74	0.9557
Highway	265.30	82.76	37.44	35.38	0.9708
Bus	19.30	15.22	28.08	29.56	0.9097

encoding time by 1.2 -3.4 times and obtained slightly better PSNR in comparison with NHEXS algorithm. The proposed algorithm is also compared with Zhu’s region-based algorithm [20] in terms of compression ratio and PSNR of the first 15 frames of each video sequence as shown in Table 2.

In fractal video coding all the pixels of domain block are modified by applying s and o parameters. Due to this there may be large MSE between the original frame and reconstructed frame. The proposed algorithm determines highly correlated, i.e. structural similar block which improves the subjective visual quality. Wang [16] introduced SSIM index for measuring the similarity between the two images. SSIM indexing is more consistent with human visual system and more effective for measuring the quality of images than traditional method like PSNR. The overall quality of the image using SSIM includes three component measures, i.e. luminance, contrast and structure as shown in Eq. (9). It is calculated within the local window that moves pixel by pixel over the entire image and the average of SSIM is used to evaluate the overall image quality. In this paper, window of size 11×11 pixels using a Gaussian weighting function with a standard deviation of 1.5, for two local windows x and y are used. The SSIM is calculated as follows:

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) \tag{9}$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

Where μ and σ^2 are the mean and variance of the window block, σ_{xy} is the covariance between x and y , C_1 , C_2 and C_3 are constants. SSIM index value is bounded between 0 to 1, if $x = y$ the SSIM index value will be 1.

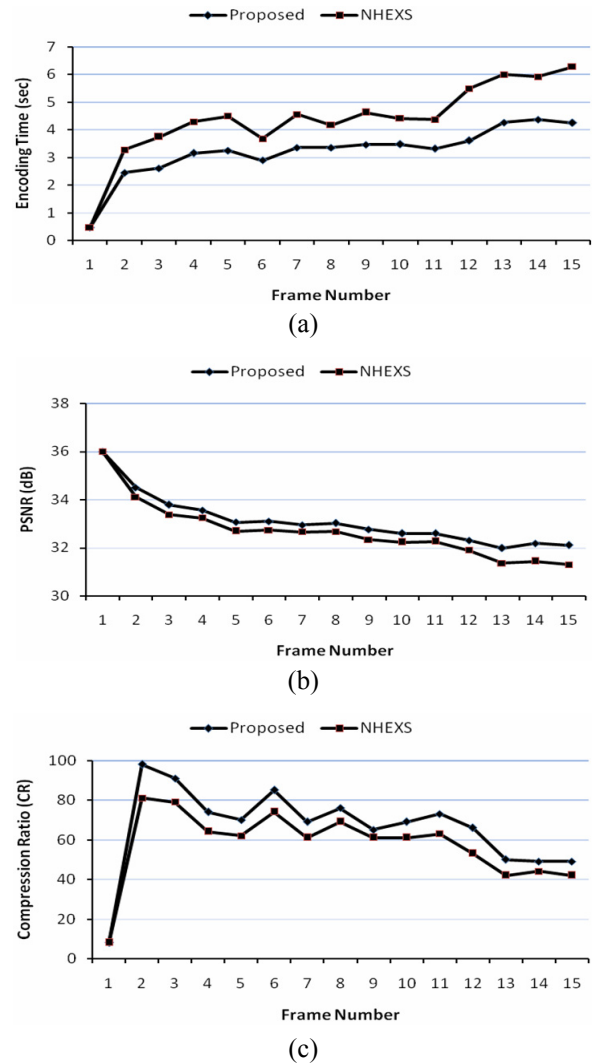


Fig. 5. frame to frame comparison of performance using NHEXS and proposed algorithm: (a) Encoding time; (b) PSNR and (c) Compression ratio

In Table 2 even though the PSNR of the proposed algorithm for ‘Paris’ video is lower than that for the region based algorithm [20] but SSIM is the second largest among all videos. In this particular video sequence entire frame is moving in the left direction from 6 to 7 pixels, so almost all the range blocks are modified structurally with s and o parameters. The best match may give a high NCC with large gray level difference, due to that MSE may increase. The average PSNR for ‘Bus’ video sequence is low, whereas



Fig. 6. Original: (a) frame 11th; (b) frame 12th of 'Foreman video'; (c) Decoded 11th frame (SSIM=0.9313, CR=73.17); (d) Decoded 12th frame (SSIM=0.9290, CR=65.93)

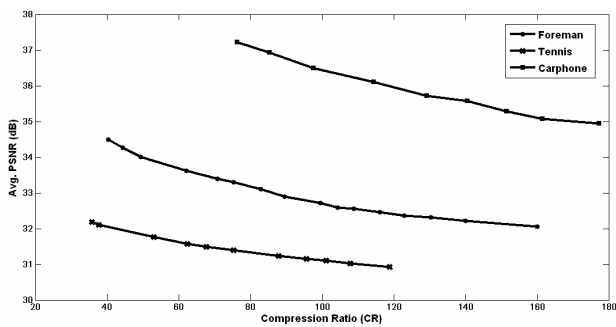


Fig. 7. CR-distortion performances of the proposed algorithm on three video sequences

the SSIM index is better. However, for the remaining videos CR and PSNR using the proposed method are better than region based [20] method. It is also observed that compression is very high for low bit rate videos with good quality of output video.

Frame by frame comparison of encoding time, PSNR and CR for the first 15 frames of 'Foreman' video are shown in Fig. 5. Performances of the proposed method are better than NHEXS algorithm with respect to all the three performance parameters. Original and decoded 11th and 12th frame of 'Foreman' video with SSIM and CR are shown in Fig. 6. The CR-distortion performances of the proposed video coder on the Foreman, Tennis and Carphone sequences are shown Fig. 7. The provided PSNR and compression ratio are averaged over all the first 60 frames. It is observed that higher compression ratio of more than 100 can be achieved by little sacrificing of the video quality by 1 to 2dB.

7. Conclusion

In this paper, a new fractal video coding method with fast normalized cross correlation is presented to increase the encoding speed and improve the coding efficiency. Three level quadtree partitions and fast NCC computation as matching criteria are used. In this approach, the sizes of range block and domain block were kept same to exploit temporal correlation in video sequences. The proposed method improves the subjective visual quality of decoded video from human perception point of view. Experimental results show that the proposed scheme can raise the encoding speed by 14.19 to 34.6 times, compression ratio by 2.49 to 9.18 times and decoded video quality is improved by 3.5 to 8.15 dB in comparison with CPM/NCIM. The speed of the proposed method is 21 % to 70 % higher than NHEXS algorithm and it also improves PSNR and compression ratio on an average by 0.89 dB and by 15.41 respectively. In this method, a high compression ratio of more than 100 can be achieved with good quality video output and significantly less encoding time.

The computation complexity of NCC limits the number of quadtree levels because at each level different size of FFT and cross correlation computations are required. The speed of the proposed algorithm will be further improved by finding the gray level transformation parameters and the denominator calculations of NCC in frequency domain. Another way, each level of quadtree partition requires different FFT computation for different sizes that can be designed using multirate processing properties.

References

- [1] A. E. Jacquin, "Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations," *IEEE Transactions on Image Processing*, vol. 1, No. 1, pp. 18-30, 1992.
- [2] Y. Fisher, *Fractal Image Compression: Theory and Application*, Ed. Springer Verlag, New York, 1995.
- [3] M. S. Lazar and L. T. Bruton, "Fractal Block Coding of Digital Video," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 4, No. 3, pp. 297-308, 1994.
- [4] Y. L. Lin and M. S. Wu, "An Edge Property-Based Neighborhood Region Search Strategy for Fractal Image Compression," *Computers & Mathematics with Applications*, vol. 62, No. 1, pp. 310-318, 2011.
- [5] C. Urmson and K. Ferens, "Video Compression through Fractal Block Coding," *IEEE Canadian Conference on Electrical and Computer Engineering*, Vol. 2, pp. 465-467, 1998.
- [6] K. U. Barthel and T. Voyer, "Three-Dimensional Fractal Video Coding," *IEEE International Conference on Image Processing*, vol. 3, pp. 260-263,

- 1995.
- [7] V. D. Lima, R. S. William and H. Pedrini, "Fast Low Bit-Rate 3D Searchless Fractal Video Coding," IEEE 24th SIBGRAPI Conference on Graphics, Patterns and Images, pp. 189-196, 2011
- [8] Y. Fisher, D. Rogovin and T. P. Shen, "Fractal (self-VQ) Encoding of Video Sequences," In the Proc. of SPIE Visual Communications and Image Processing, vol. 2308, pp. 1359-1370, 1994.
- [9] C. S. Kim and S. U. Lee, "Fractal Coding of Video Sequence by Circular Prediction Mapping," In Fractals, vol. 5, pp. 75-88, 1997.
- [10] M. Wang and C. H. Lai, "A Hybrid Fractal Video Compression Method," Computers and Mathematics with Applications, vol. 50, No. 3-4, pp. 611-621, 2005.
- [11] M. Wang, R. Liu and C. H. Lai, "Adaptive Partition and Hybrid Method in Fractal Video Compression," Computers and mathematics with Applications, vol. 51, No. 11, pp. 1715-1726, 2006.
- [12] Z. Yao and R. Wilson, "Hybrid 3D Fractal Coding with Neighborhood Vector Quantization," EURASIP Journal on Advances in Signal Processing, vol. 16, pp. 2571-2579, 2004.
- [13] B. Hurtgen and P. Buttgen, "Fractal Approach to Low Rate Video Coding," in Proc. of the SPIE: Visual Communications and Image Processing '93, vol. 2094, pp. 120-131, 1993.
- [14] C. S. Kim, C. Kim, and S. U. Lee, "Fractal Coding of Video Sequence using Circular Prediction Mapping and Noncontractive Interframe Mapping," IEEE Transactions on Image Processing, vol. 7, No. 4, pp. 601-605, 1998.
- [15] S. Zhu, L. Li and Z. Wang "A Novel Fractal Monocular and Stereo Video Codec with Object-based Functionality," EURASIP Journal on Advances in Signal Processing, vol. 2012:227, pp. 1-12, 2012.
- [16] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, No. 4, pp. 600-612, 2004.
- [17] H. Hartenstein and D. Saupe, "Lossless Acceleration of Fractal Image Coding via the Fast Fourier Transform," Signal Processing: Image Communication, Vol. 16, No. 4, pp. 383-394, 2000.
- [18] S. B. Dhok, R. B. Deshmukh and A. G. Keskar, "Efficient Fractal Image Coding using Fast Fourier Transform," International Journal on Computing, vol. 1, No. 2, 2011.
- [19] K. Belloulata, S. Zhu and Z. Wang, "A Fast Fractal Video Coding Algorithm Using Cross - Hexagon Search for Block Motion Estimation," International Scholarly Research Network (ISRN) Signal Processing, pp. 1-10, 2011.
- [20] S. Zhu, Y. Hou, Z. Wang and K. Belloulata, "Fractal Video Sequences Coding with Region-Based Functionality," Applied Mathematical Modeling, vol. 36, No. 11, pp. 5633-5641, 2012.
- [21] K. Belloulata, A. Belalia, S. Zhu, "Object-Based Stereo Video Compression using Fractals and Shape-Adaptive DCT," International Journal of Electronics and Communications, Vol. 68, No. 7, pp. 687-697, 2014.
- [22] J. Wan, L. You, "A Fast Context-Based Fractal Mobile Video Compression with GA and PSO," 2nd International Conference on Teaching and Computational Science, pp. 112-115, 2014.
- [23] C. C. Wang and C. H. Hsieh, "Efficient Fractal Video Coding Algorithm using Intercube correlation Search," Optical Engineering, vol. 8, No. 39, pp. 2058-2064, 2000.
- [24] Y. Bnjmohan and S. H. Mneney, "Low Bit-rate Video Coding Using Fractal Compression of Wavelet Subtrees," IEEE 7th AFRICON Conference, vol. 1, pp. 39-44, 2004.
- [25] Y. Zhang, L. M. Po and Y. L. Yu, "Wavelet Transform Based Variable Tree Size Fractal Video Coding," IEEE International Conference on Image Processing, vol. 2, pp. 294-297, 1997.
- [26] S. D. Wei and S. H. Lai, "Fast Normalized Cross Correlation Based on Adaptive Multilevel Winner Update," Proc. of the multimedia 8th Pacific Rim conference on Advances in multimedia information processing, pp. 413-416, 2007.
- [27] R. E. Chaudhari and S. B. Dhok, "Acceleration of Fractal Video Compression using FFT," 15th International Conference on Advanced Computing Technologies (ICACT), pp. 1-4, 2013.
- [28] B. C. Song, "A Fast Normalized Cross-Correlation Based Block Matching Algorithm Using Multilevel Cauchy-Schwartz Inequality," ETRI Journal, vol. 33, No. 3, pp. 401-406, 2011.
- [29] A. A. Eskinder, R. E. Chaudhari, and S. B. Dhok, "Fast Motion Estimation for Quad-Tree Based Video Coder Using Normalized Cross-Correlation Measure," International Journal of Image processing (IJIP), vol. 7, No. 4, 2013.
- [30] Y.M. Zhou, C. Zhang and Z.K. Zhang, "An Efficient Fractal Image Coding Algorithm using Unified Feature and DCT," Chaos, Solitons & Fractals, vol. 39, No. 4, pp. 1823-1830, 2009.
- [31] G. J. Sullivan and R. L. Baker, "Efficient Quadtree Coding of Image and Video," IEEE Transactions on Image processing, vol. 3, No. 3, pp. 327-331, 1994.
- [32] A. K. Jain, Fundamentals of Digital Image Processing, PHI publications, 1989.
- [33] CIPR Sequences, <http://www.cipr.rpi.edu/resource/sequences/>



R E Chaudhari He received his B.E. degree in Electronics and Telecommunication Engineering from Government College of Engineering, Aurangabad, India, in 1999, and M. Tech, in Electronics Engineering from Visvesvaraya Regional College of Engineering, Nagpur in 2002. He is pursuing his Ph.D. degree in Electronics Engineering from Visvesvaraya National Institute of Technology, Nagpur, and working as an assistant professor at St. Francis Institute of Technology, Mumbai, India. His research interests are in the areas of image/ video and signal processing.



S B Dhok He received his B. Tech and M.Tech in Electrical Engineering from Indian Institute of Technology, Mumbai, India and Ph. D. degree in Electronics Engineering from Visvesvaraya National Institute of Technology (VNIT), Nagpur, India. He is currently an associate professor in Centre for VLSI and Nanotechnology at VNIT. His area of research includes Signal Processing, Image Processing, Data Compression, Wireless Sensor Networks and VLSI Design.