Fractal Depth Map Sequence Coding Algorithm with Motion-vector-field-based Motion Estimation

Shiping Zhu[1] and Dongyu Zhao[1]

[1] Department of Measurement Control and Information Technology, School of Instrumentation Science and Optoelectronics Engineering, Beihang University, Beijing 100191, China
[e-mail: spzhu@163.com, dongyuzhao@msn.cn]
*Corresponding author: Shiping Zhu

Abstract

Three-dimensional video coding is one of the main challenges restricting the widespread applications of 3D video and free viewpoint video. In this paper, a novel fractal coding algorithm with motion-vector-field-based motion estimation for depth map sequence is proposed. We firstly add pre-search restriction to rule the improper domain blocks out of the matching search process so that the number of blocks involved in the search process can be restricted to a smaller size. Some improvements for motion estimation including initial search point prediction, threshold transition condition and early termination condition are made based on the feature of fractal coding. The motion-vector-field-based adaptive hexagon search algorithm on the basis of center-biased distribution characteristics of depth motion vector is proposed to accelerate the search. Experimental results show that the proposed algorithm can reach optimum levels of quality and save the coding time. The PSNR of synthesized view is increased by 0.56 dB with 36.97% bit rate decrease on average compared with H.264 Full Search. And the depth encoding time is saved by up to 66.47%. Moreover, the proposed fractal depth map sequence codec outperforms the recent alternative codecs by improving the H.264/AVC, especially in much bitrate saving and encoding time reduction.

Keywords: Fractal coding, depth map coding, 3D video, motion estimation, Hexagon Search.

1. Introduction

With the development of multimedia technology, the traditional two-dimensional video has been unable to meet people's needs, while the three-dimensional (3D) immersive and interactive video gives the users an unprecedented wonderful experience owing to its unique depth of field effect. 3D video is composed of texture videos and the corresponding depth map sequences. Based on the principle of binocular parallax, the users can obtain a three-dimensional image, which enhances the senses of visual reality and verisimilitude. Compared with monocular video, the original data of 3D video is much huge. The transmission bandwidth is limited, which limits the widespread applications of 3D video, and therefore, it is necessary to study high efficiency 3D video compression technology.

The current 3D video compression methods mainly contain multi-view video coding [1] and depth map-based compression [2]. The multi-view video uses multiple cameras to shoot the same scene from different angles. Multi-view video is characterized by high complexity of coding and large amount of data. The multi-view video coding needs to consider the redundancy in the same viewpoint as well as that between adjacent viewpoints. Depth map represents the information about the distance from the scene to the camera imaging plane, which quantizes the actual depth value to [0, 255] so as to obtain the grayscale map representing the depth information. In the Multi-view Video plus Depth (MVD) [3] structure, each color image contains its corresponding depth map. This paper is to study the coding of depth map sequence.

The depth map can be represented by 3D-mesh, which can also be seen as a two-dimensional image so that the depth map compression can be treated as two-dimensional image compression and the existing compression standard can be applicable. In [4], an early SKIP-mode decision scheme and a selective prediction-mode decision scheme have been developed based on the relationship between coding modes of collocated macroblocks on texture and depth frames. Depth map sequence coding aims to maximize the perceived visual quality of the synthesized virtual view instead of the depth map sequence itself, so that a structural similarity-based synthesized view distortion (SS-SVD) model to relate perceptual distortion in coded depth map and synthesized view is proposed in [5]. It obtains better rate distortion performance and perceptual quality of synthesized views than JM reference software. In [6], an efficient intra prediction algorithm for smooth regions in depth coding is proposed by considering one single prediction direction instead of multiple prediction directions. The original context-based adaptive binary arithmetic coding (CABAC) was originally designed for lossy texture coding, and it cannot provide the best coding performance for lossless depth map coding. In [7], an enhanced CABAC coding mechanism for lossless depth map coding based on the statistics of residual data is proposed.

All the above depth map sequence encoding methods are based on H.264/AVC standard [8]. Its basic theory is to form a prediction of the macroblock based on previously-coded data, either from the current frame (intra prediction) or from other frames that have already been coded and transmitted (inter prediction) and then residual is transformed, quantified and entropy coded. Different from it, fractal image coding method is a novel technique in the image compression field using partial self-similarity of image [9]. Its theoretical bases are the Iterated Function System and Collage theorem [10], and it is recognized as one of the most promising new generation image coding technologies at present, which is characterized by novel concepts, high compression ratio and great potential, but with high coding complexity.

The depth map sequence shows a strong similarity between adjacent frames. Therefore, good compression performance can be obtained by use of fractal theory.

To obtain high efficiency encoding performance, and to decrease the complexity of depth map sequence codec as well, we propose a novel fractal depth map sequence coding algorithm with motion-vector-field-based motion estimation. This paper is organized as follows: The theory of fractal coding is summarized in section 2. The proposed fractal depth map sequence coding algorithm is presented in section 3. The experimental results are presented in section 4. And finally the conclusions are outlined in section 5.

2. Background of Fractal Compression

Jacquin described the first practical fractal image compression algorithm based on a Partitioned Iterated Function System (PIFS). Divide the image I to be coded into two sizes of blocks, with the size of B×B pixels being range blocks called R_i, allowing non-overlapping and covering the entire image; and with the size of 2B×2B pixels being domain blocks called D_i, allowing overlapping. All the domain blocks form a domain pool. The image I can be represented by

$$I = \{R_i\}_{0 \leq i \leq N} \tag{1}$$

where N is the total number of range blocks.

Each range block R_i shall be approximated by a certain domain block $D_{m(i)}$ with the transformation $\omega_i : D_{m(i)} \rightarrow R_i$, which is composed of spatial contraction transformation γ_i and grayscale modification transformation λ_i.

$$\omega_i = \lambda_i \circ \gamma_i \tag{2}$$

ω_i is often a contracted affine transformation, its general form is:

$$\omega_i(D_{m(i)}) = \lambda_i(\gamma_i(D_{m(i)})) = s_i \cdot t_k(\gamma_i(D_{m(i)})) + o_i \tag{3}$$

γ_i uses the average value of four adjacent pixels to get B×B pixel block $\hat{D}_{m(i)}$. The average value of four adjacent pixels is:

$$\hat{d}_{h,v} = \frac{1}{4}(d_{2h,2v} + d_{2h+1,2v} + d_{2h,2v+1} + d_{2h+1,2v+1}) \tag{4}$$

where $d_{h,v}$ and $\hat{d}_{h,v}$ represent pixel values of $D_{m(i)}$ and $\hat{D}_{m(i)}$ at the pixel position (h, v),

respectively.

λ_i introduces grayscale adjustment parameter s_i, grayscale offset parameter o_i, and isometric transformation t_k in order to get a better approximation of R_i by further modifying the grayscale of $D_{m(i)}$. The general form of λ_i is:

$$\lambda_i \begin{bmatrix} x \\ y \\ z(x,y) \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & s_i \end{bmatrix} \begin{bmatrix} x \\ y \\ z(x,y) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ o_i \end{bmatrix} \tag{5}$$

where a, b, c and d determine 8 kinds of pixel rearrangements (4 kinds of rotation and 4 kinds of flip); s_i and o_i represent brightness adjustment and brightness offset respectively.

The fractal coding is aimed to look for the best matching block $D_{m(i)}$ and block transformation ω_i for each R_i, so that $R_i' = \omega_i(D_{m(i)})$ and R_i are as close as possible in the given distortion metric, namely

$$d(R_i, \omega_i(D_{m(i)})) = \min\{d(R_i, \omega(D))\} \tag{6}$$

The global transformation of the image onto a close approximation of itself is then constructed from a combination of these local block transformations. **Fig. 1** shows some mappings from domain to range blocks.

Domain blocks   Range blocks
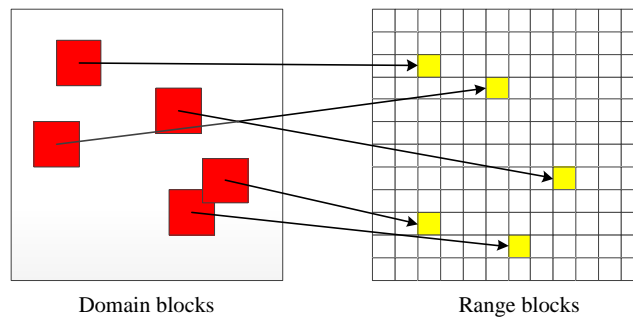
Fig. 1. Mappings from domain to range blocks

3. The Proposed Fractal Depth Map Sequence Coding Scheme

This paper proposes a fractal depth map sequence coding scheme, as shown in **Fig. 2**. Actually, it is a hybrid codec with intra prediction and inter fractal. When a depth map F_n is input to the encoder, it will be processed with range block as the basic unit, and the encoding mode will be selected between intra prediction and inter fractal. The intra prediction is to use the previously decoded adjacent pixels within the same frame to predict the current block based on spatial correlation. Compared with the previous encoders, the inter fractal adds pre-search restriction

and improves motion estimation. A motion-vector-field-based adaptive hexagon search method is designed. Locate the best matching block of current range block from the reference frame according to the motion vector (x, y), and then use the best matching block iteration to produce the currently predicted block *PRED* according to the recorded scale factor *s* and offset factor *o*. The iterative formula is as follows:

$$r_i = s \times d_i + o \tag{7}$$

where r_i represents pixel value of the predicted block, and d_i represents pixel value of the corresponding matching block.

In the figure, the reference frame represents the reconstructed image F'_{n-1} of the previous depth map. A residual block D_n is generated by subtracting the predicted block *PRED* from the current range block. It is transformed and quantized as a coefficient set *X*, and these coefficients are reordered for entropy coding. Because the reconstruction process requires iteration of fractal parameters to produce the reference frame, entropy coding coefficients of fractal parameters need to be written into the code stream. The entropy coding coefficients and side information (prediction mode, quantization step size, etc.) form the compressed bit stream.
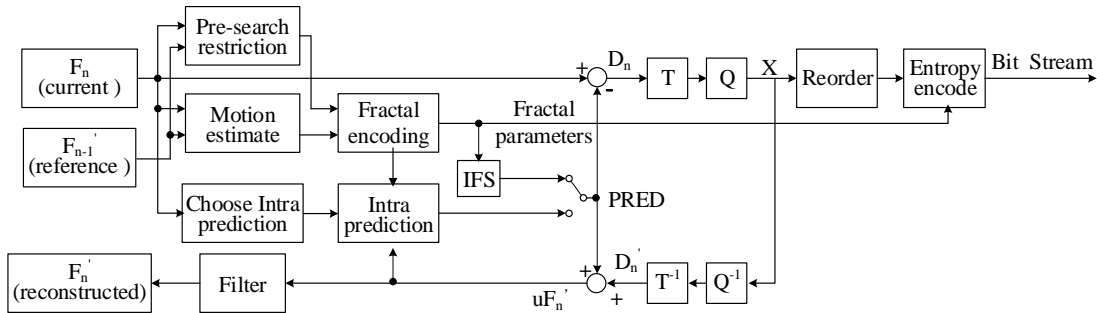

Fig. 2. The proposed fractal depth map sequence encoder

3.1 Pre-search Restriction

In the process of fractal coding, the most important step is to search in the domain pool to find the best matching block for each range block in order to minimize the matching error mean square error (MSE).

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (s \cdot d_i + o - r_i)^2 \tag{8}$$

where N represents the number of pixels in the current block.

It can be seen that, the fractal coding process is very time-consuming. Because the best matching block is required to be searched for every range block in the codebook *A*, while in theory, the number of candidate domain blocks is usually very large. For example, for each range block in the original image, if exhaustive search is used, the search time is $O(|A|)$

(linearity depends on |A|), where |A| represents the number of blocks in A. If |A| can be reduced, the fractal encoding time must be decreased. Some efforts have been made to reduce the number of blocks need to search. For example, by using a nearest-neighbor search techniques [11] or a local search operation [12] or searching in the area centered on last best matching domain block [13] or classification of image blocks [14].

In our scheme, before the matching search process for a range block, we add a pre-search restriction to rule the improper domain blocks out of the matching search process based on the variance matching condition of each block so that the number of blocks involved in the matching search process can be restricted to a smaller size, thus to greatly accelerate the coding. The derivation process of the pre-search restriction condition is detailed as follows.

In the fractal coding process, by setting an appropriate scale factor s and offset factor o may allow the value of D_i after affine transformation has a minimum squared distance of the value of R_i. When the partial differentiations of s and o are 0, the minimum matching error MSE can be obtained, in which the corresponding s and o are shown in Formulas (9) and (10).

$$s = \frac{[N\sum_{i=1}^{N} r_i d_i - \sum_{i=1}^{N} r_i \sum_{i=1}^{N} d_i]}{[N\sum_{i=1}^{N} d_i^2 - (\sum_{i=1}^{N} d_i)^2]} \tag{9}$$

$$o = \frac{1}{N}[\sum_{i=1}^{N} r_i - s\sum_{i=1}^{N} d_i] \tag{10}$$

Therefore, substitute the Equations (9) and (10) again into the Formula (8) to obtain the following MSE form different from the basic one:

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

$$= \frac{1}{N}\sum_{i=1}^{N}(s\cdot d_i + \frac{1}{N}[\sum_{i=1}^{N} r_i - s\sum_{i=1}^{N} d_i] - r_i)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}((d_i - \frac{\sum_{i=1}^{N} d_i}{N})\cdot\frac{[N\sum_{i=1}^{N} r_i d_i - \sum_{i=1}^{N} r_i \sum_{i=1}^{N} d_i]}{[N\sum_{i=1}^{N} d_i^2 - (\sum_{i=1}^{N} d_i)^2]} + \frac{\sum_{i=1}^{N} r_i}{N} - r_i)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}((d_i - \bar{d})\cdot\frac{[\sum_{i=1}^{N} r_i d_i - N\bar{r}\bar{d}]}{[\sum_{i=1}^{N} d_i^2 - N\bar{d}^2]} + \bar{r} - r_i)^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} ((d_i - \bar{d}) \cdot \frac{\sum_{i=1}^{N}(r_i - \bar{r})(d_i - \bar{d})}{\left\| d_i - \bar{d} \right\|^2} + \bar{r} - r_i)^2$$

$$= \frac{1}{N} \left\| r_i - \bar{r} \right\|^2 \sum_{i=1}^{N} (\frac{(d_i - \bar{d})}{\left\| d_i - \bar{d} \right\|} \cdot \frac{\sum_{i=1}^{N}(r_i - \bar{r})(d_i - \bar{d})}{\left\| r_i - \bar{r} \right\| \left\| d_i - \bar{d} \right\|} - \frac{r_i - \bar{r}}{\left\| r_i - \bar{r} \right\|})^2 \qquad (11)$$

Make $\hat{d} = \frac{(d_i - \bar{d})}{\left\| d_i - \bar{d} \right\|}$, $\hat{r} = \frac{r_i - \bar{r}}{\left\| r_i - \bar{r} \right\|}$, and it's obvious that $\left\| \hat{d} \right\|^2 = 1$, $\left\| \hat{r} \right\|^2 = 1$, Then

Formula (11) can be further derived as follows：

$$MSE = \frac{1}{N} \left\| r_i - \bar{r} \right\|^2 \sum_{i=1}^{N} (\hat{d} \cdot \sum_{i=1}^{N} \hat{r}\hat{d} - \hat{r})^2$$

$$= \frac{1}{N} \left\| r_i - \bar{r} \right\|^2 \left(1 - (\sum_{i=1}^{N} \hat{r}\hat{d})^2 \right) \qquad (12)$$

Let

$$m = 1 - (\sum_{i=1}^{N} \hat{r}\hat{d})^2 \qquad (13)$$

For each range block, $\frac{1}{N} \left\| r_i - \bar{r} \right\|^2$ is already known. In order to obtain the minimum matching error MSE, the value of m should be as smaller as possible, which is the start point of our pre-search restriction scheme.

In our pre-search restriction scheme, for each range block in the matching process, if $Th < m < 1$, then we need to search the next block in the codebook to find the minimum matching error MSE; otherwise it indicates that the current MSE has been small enough without requirement of the next step's comparison and the result can be directly saved, then proceed to search for the next range block. Th is a threshold experimentally determined in advance and it's set to 0.9 in our scheme.

The pre-search restriction algorithm is outlined in the pseudo-code as follows.

Pre-search Restriction Algorithm:

- Prepare R_i , $i = 1, 2, ..., N_R$

- Prepare D_j , $j = 1, 2, ..., N_D$

- Extend D_j by spatial contraction transformation and 8 kinds of affine transformation to

form a codebook

$$\Omega = \{\Omega_l = t_k D_j, k = 1,...,8, j = 1,...,N_D, l = 1,...,8N_D\}$$

● Loop Range Block: For every range block R_i, $i = 1, 2,..., N_R$, do

{

　▪ Loop Domian Search: For every domain block in Ω, do

{

　Compute *m* using formula (13)

　· if *0.9<m<1* and *l<8N_D*

　　go on to next domain block

　· else

　　compute and record scale factor snd offset factor

　　go on to next range block

}

}

3.2 Improvements of Motion Estimation

Currently, the fractal coding speed is very low, of which the most important reason is that the search for the best matching block is time-consuming, and to improve the coding speed, it is necessary to improve search skills. Unsymmetrical-Cross Multi-Hexagonal Search (UMHexagonS [15]) is a fast motion estimation algorithm, which has been applied in the JM reference software of H.264/AVC. It greatly outperforms full search in computation with up to 90% computational reduction while still maintaining good rate-distortion performance. However it's proposed based on the multiple prediction modes and multiple reference frames of H.264, which is significantly different from the fractal depth map-sequence coding method in this paper. Therefore, the following improvements have been made.

1.　Initial search point prediction

Use the following three methods for initial search point prediction:

(1)　Spatial median prediction: According to the spatial correlation, take the median of motion vectors of the left, up and right-up adjacent blocks in the current frame as the predicted motion vector, as shown in **Fig. 3**;
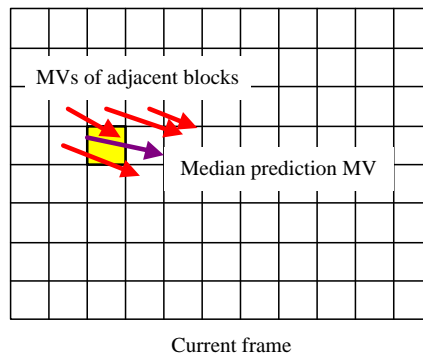

Fig. 3. Spatial median prediction

(2)   Origin prediction: Take $(0,0)$ as the motion vector value;

(3)   Adjacent reference frame prediction: According to the time correlation, get current motion vector from MV in the same position of the previous reference frame.

Then calculate the MSE using Formula (12) for each candidate, compare and choose the minimum MSE. The point with the minimum MSE is named as MMD (minimum matching distortion), which is used as the initial search point.

2.   Threshold transition condition at the asymmetrical cross template search

The criteria of the similarity between range block and transformed domain block in fractal coding uses MSE as shown in Formula (12). We use different thresholds based on different sizes of blocks; after asymmetrical cross template search is completed, select the best matching point as a new start point for the subsequent template matching.

3.   Early termination condition

Based on the characteristics of fractal coding algorithm in this paper, the early termination circumstances include two aspects:

Firstly, in the multi-level non-uniform hexagonal grid integer pixel motion search process, in addition to the early termination conditions of the algorithm itself, in order to reduce the complexity of search, if the optimal matching point is at the center of the hexagon, the search can be stopped;

Secondly, the inter fractal coding algorithm in this paper uses the hierarchical tree-structure partition as shown in **Fig. 4**. Search using mode 1 at first, if the threshold condition is satisfied, then terminate the current block coding and proceed to the next block; otherwise, use mode 2 to divide the current range block into two smaller sub-blocks and search for each sub-block. This partition goes on to mode 3 and mode 4 until the threshold condition is satisfied. When the range block has been partitioned to 4 equal sub-blocks as shown in mode 4, each sub-block allows further partition recursively until it reaches the allowed minimum size.
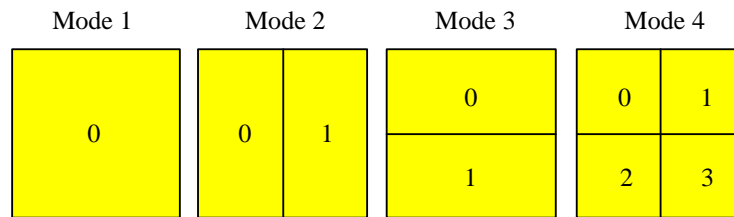

Fig. 4. Hierarchical tree-structure partition

3.3 Motion-vector-field-based Adaptive Hexagon Search Algorithm

In the actual video sequences, most of the depth motion vectors are enclosed in the central area [16]. We use full search algorithm in fractal depth map sequence codec and make some statistical analysis about motion vector. As shown in **Fig. 5(a)** and **Fig. 5(b)**, about 81.80% of motion vectors are found located in the central 5×5 area, i.e. $p = \pm 2$ pixels, using the search window $W = \pm 7$. The large cross pattern (A+B+C) accounts for 74.71% of the motion vector distribution, and the small cross pattern (A+B) accounts for 68.98% of the motion vector distribution. A lot of experimental results show that more than 80% of the blocks can be regarded as stationary or quasi-stationary blocks, suggesting that motion vector is of center-biased distribution characteristics [16]. Most of the best matching points of inter fractal encoding are located in a small scope around the initial point.

(a) General view   (b) Detail statistics
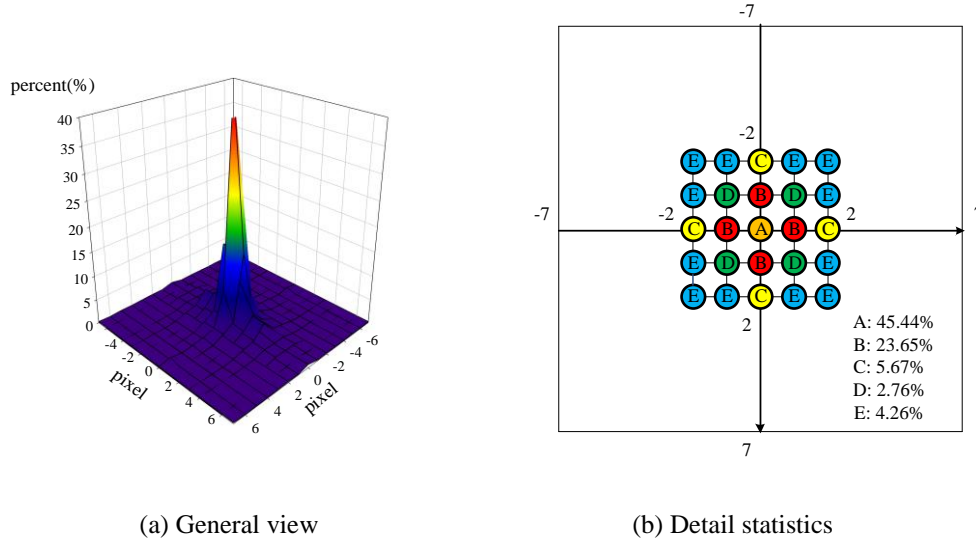
Fig. 5. Average distributive probability of depth motion vector for the search window $W = \pm 7$

Based on this observation, we propose our motion-vector-field-based adaptive hexagon search algorithm. Firstly, an unsymmetrical cross-shaped pattern (UCSP) is used to search points around the initial point, as shown in **Fig. 6(a)**, and then small cross-shaped pattern (SCSP, see **Fig. 6(b)**) and large hexagon-shaped pattern (LHSP) are used. In addition to the six surrounding points and the central point of the traditional LHSP, we also contains the midpoint of the two sides located above and below the search central point, as shown in **Fig. 6(c)**, so that the new LHSP contains a total of 9 search points, which could improve the search accuracy with negligible increasing complexity. Furthermore, we employ a halfway-stop technology so that the small motion vector can be searched with less points, so as to accelerate the motion estimation. The specific steps are as follows:
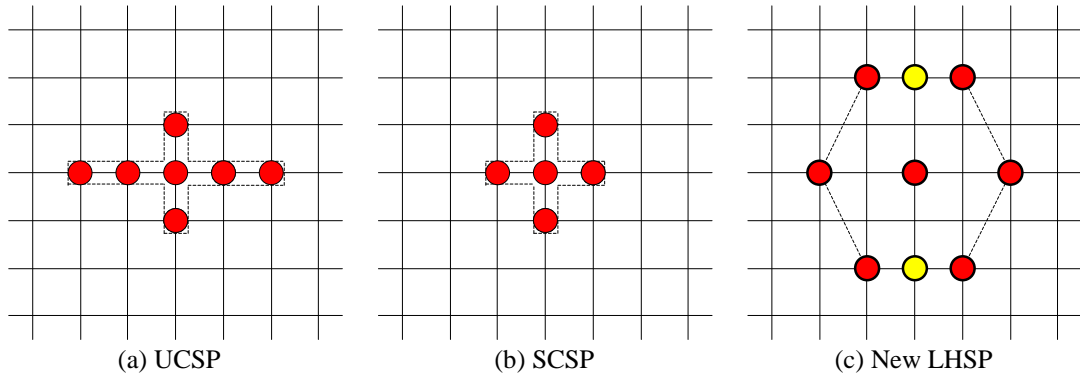
Step (i):   Search by a UCSP centering the initial point. Search stops if the MMD point occurs at the center of the UCSP, and we get the motion vector, as shown in **Fig. 7(a)**, where the best matching point is filled with red color. Otherwise, go to Step (ii).

Step (ii):   Take the MMD point in Step (i) as the center of the SCSP. Two or three new points need to be searched in this step. Search stops if the MMD point occurs at the center of the SCSP. **Fig. 7(b)** shows the case when two more points (filled with yellow color) are searched. Otherwise, go to Step (iii).
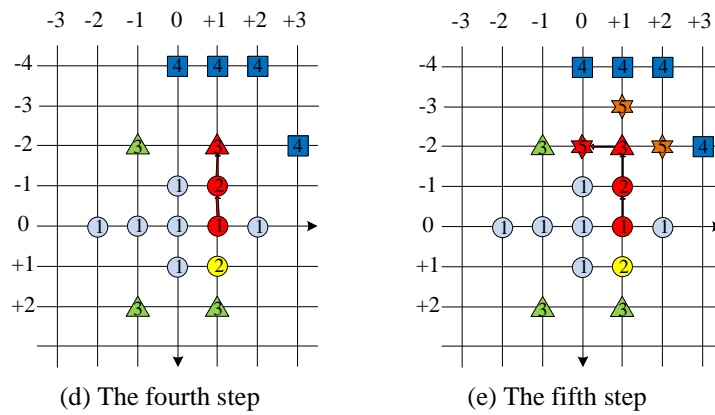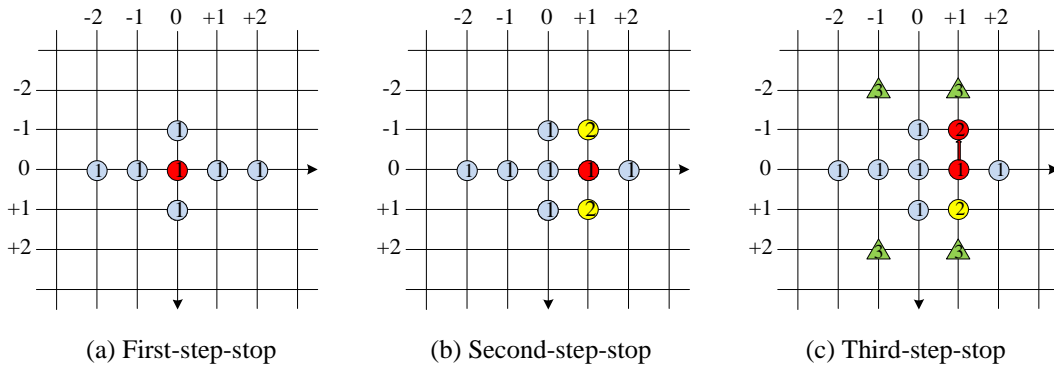
Step (iii):   Four additional points on a traditional LHSP centering the initial point in step (i) are checked. If the MMD point is not on the LHSP, meaning that the MMD point in Step (ii) is still the best matching point, the search stops, as shown in the in **Fig. 7(c)**. Otherwise, if the MMD point is on the LHSP, go to Step (iv).

Step (iv):   Search by the new LHSP centering the MMD point in Step (iii). If the new MMD point is at the center of the new LHSP, then go to Step (v), as shown in **Fig. 7(d)**. Otherwise, this step is repeated again.

Step (v):   Take the minimum MMD point in the previous step as the center, and then a SCSP is applied. Find the new MMD point, which is the final solution for the motion vector, as shown in **Fig.7 (e)**.

|(a) UCSP|(b) SCSP|(c) New LHSP|

Fig. 6. Search patterns

|(a) First-step-stop|(b) Second-step-stop|(c) Third-step-stop|

|(d) The fourth step|(e) The fifth step|

Fig. 7. The proposed motion-vector-field-based adaptive hexagon search algorithm

   The first two steps in our proposed motion estimation scheme using two cross-shaped pattern have a high probablity to find the best matching point based on center-biased distribution characteristics of motion vector. Thus the number of points need to search is heavily reduced. In the following steps, we use large hexagon-shaped pattern to avoid falling into local optimum.

## 4. Experimental Results and Analysis

To verify the performance of the proposed method, we have made several experiments with

the Kendo, Newspaper, and Book Arrival test sequences. Experimental results obtained using our proposed fractal coding algorithm with motion-vector-field-based motion estimation are compared with those using H.264 full search, H.264 EPZS (enhanced predictive zonal search), LCMDME proposed by Gianluca et al [17], and fast MD/ME with Sobel edge detector by Zhu et al [18]. Since depth maps are not directly used to watch, but they are used to generate the synthetic views. Hence, we compare the quality between the synthetic views generated by the encoded depth maps. Synthetic views are generated using the View Synthesis Reference Software (VSRS) 3.0 provided by MPEG [19]. The views used for our experiments are listed in **Table 1**.

Table 1. Color and depth sequences used for the experiments

Sequence	Resolution	L view	R view	Synthetic view	Frame Rate
Kendo	1024×768	1	3	2	30
Newspaper	1024×768	2	4	3	30
Book Arrival	1024×768	8	10	9	16.67

To keep coincidence with the compared methods, the color videos used for the generation of the synthetic views have been encoded using our fractal codec. The RD curves are shown in **Fig. 8**, where the PSNR is evaluated between the synthetic views generated through the original depth maps and the synthetic views created through the compressed ones [17]. The bitrate is calculated by the sum of the two depth sequences' bitrate.

(a) Kendo

(b) Newspaper

(c) Book Arrival
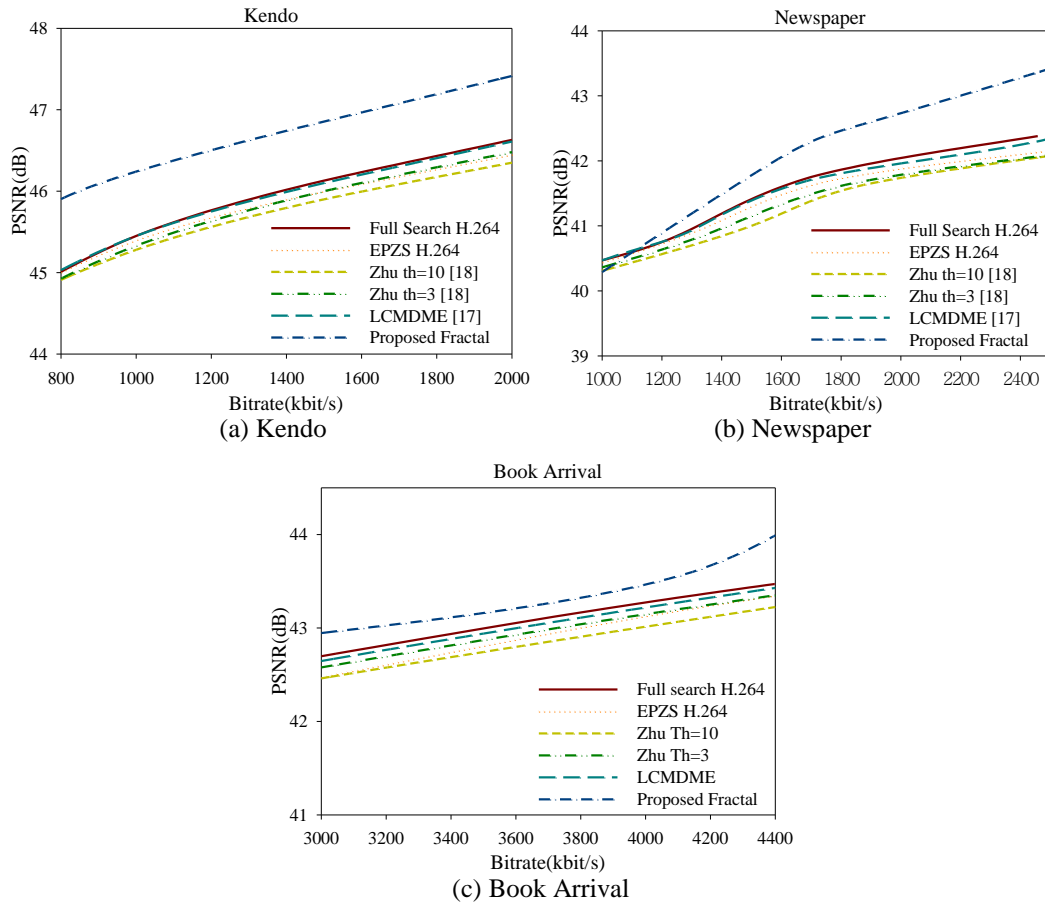
Fig. 8. RD curves for different methods

It can be seen from **Fig. 8** that our proposed fractal depth map sequence coding algorithm with motion-vector-field-based motion estimation outperforms all the other compared methods. Adjacent frames in the depth map sequence have a strong similarity, so that we can get good compression performance by inter fractal method.

The RD performance comparisons evaluated with the Bjontegaard metric [20] are summarized in **Table 2**. We note that the proposed method is consistently better than all the other compared methods, with 33.82% bit rate decrease and 0.45 dB PSNR increase on average than H.264 Full Search.

Table 2. BDBR and BDPSNR between the proposed method and the compared five methods

Scheme	Kendo		Newspaper		Book Arrival	
	BDBR(%)	BDPSNR(dB)	BDBR(%)	BDPSNR(dB)	BDBR(%)	BDPSNR(dB)
Full Search	-40.35	0.49	-33.58	0.62	-27.54	0.25
EPZS	-30.71	0.56	-43.34	0.74	-21.99	0.35
Zhu th=10	-34.25	0.66	-44.85	0.90	-37.72	0.49
Zhu th=3	-31.16	0.50	-43.72	0.70	-27.79	0.36
LCMDME	-26.15	0.48	-36.68	0.66	-18.9	0.28

Fig. 9 shows the 10th frame of the original depth map sequence Newspaper view 2 and the decoded ones by Full Search using JM18.1 [21] and the proposed method. **Fig. 10** shows the 10th frame of the synthesized Newspaper view 3 resulted from Full Search using JM18.1 and the proposed method.

(a) The 10th frame of the original depth map Newspaper view 2

(b) The 10th frame of the decoded depth map Newspaper view 2 resulted from Full Search using JM18.1

(c) The 10th frame of the decoded depth map Newspaper view 2 resulted from our proposed method

Fig. 9. the 10th frame of the original depth map-sequence Newspaper view 2 and the decoded ones

(a) The 10th frame of the synthesized Newspaper view 3 resulted from Full Search using JM18.1

(b) The 10th frame of the synthesized Newspaper view 3 resulted from our proposed method

Fig. 10. The 10th frame of the synthesized Newspaper view 3

Fig. 11 shows the encoding time for two depth views and **Table 3** shows the percentage of average depth encoding time saved by our method compared with Full Search and LCMDME.
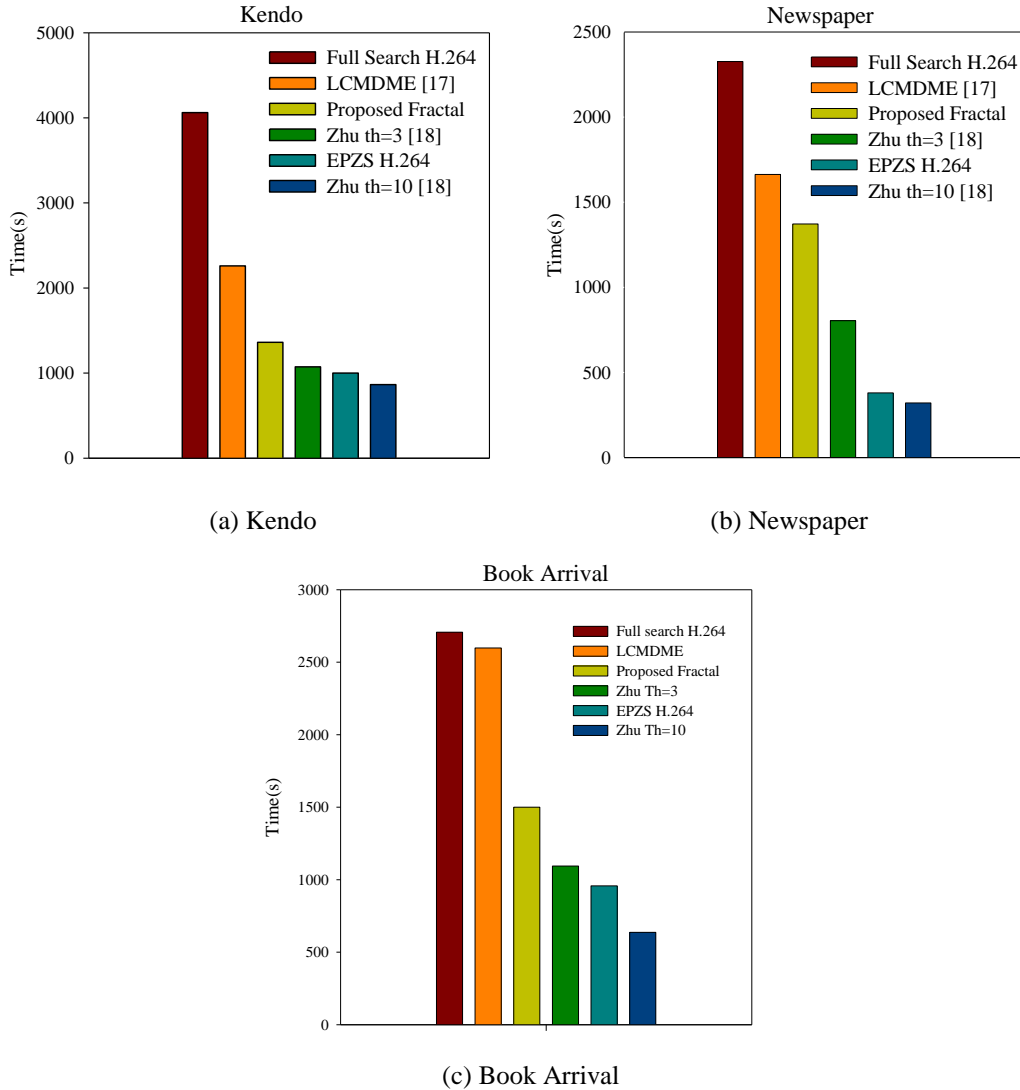

(a) Kendo

(b) Newspaper

(c) Book Arrival

Fig. 11. Comparison of depth sequences (two depth views) encoding time

Table 3 Average depth encoding time saved (%) by the proposed method compared with Full Search and LCMDME

Sequence	Full Search	LCMDME
Kendo	66.47	39.73
Newspaper	41.01	17.50
Book Arrival	44.59	42.26
Average	50.69	33.16

The encoding time comparisons show our proposed method can reduce depth compression time by up to 66.47% and 42.26% compared with H.264 Full Search and LCMDME respectively. Although the encoding time is longer for our methods than the other three fast motion estimation methods, the difference is not prominent for Kendo.

5. Conclusion

In this paper, an efficient fractal depth map-sequence coding algorithm with fast motion estimation is proposed. We firstly add pre-search restriction to rule the improper domain blocks out of the matching search process so that the number of blocks involved in the matching search process can be restricted to a smaller size. Some improvements for motion estimation including initial search point prediction, threshold transition condition and early termination condition are made based on the feature of fractal coding. The probability distribution characteristics of depth motion vectors are also analyzed and motion-vector-field-based adaptive hexagon search algorithm is proposed to accelerate the search, especially for the static block or the quasi static block.

Experimental results show that the proposed algorithm can reach optimum levels of quality and save coding time at the same time. The PSNR of synthesized view is increased by 0.44 dB with 33.82% bit rate decrease on average compared with H.264 Full Search evaluated by the Bjontegaard metric. And the depth encoding time is saved by up to 66.47%. This method makes the best of the features of fractal coding and depth motion, and achieves a great improvement with considerably good results. Also, it has built a good foundation for the further research of 3D video fractal coding. Our method doesn't consider the relation between depth map and the corresponding color video, which may be our further research direction.

References

[1] Anthony Vetro, Thomas Wiegand and Gary J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," in *Proc. of the IEEE,* vol. 99, no. 4, pp. 626-642, April, 2011. Article (CrossRef Link)

[2] Byung Tae Oh, Jaejoon Lee and Du-sik Park, "Depth map coding based on synthesized view distortion function," *IEEE Journal of Selected Topics in Signal Processing,* vol. 5, no. 7, pp. 1344-1352, November, 2011. Article (CrossRef Link)

[3] Qiuwen Zhang, Ping An, Yan Zhang, Liquan Shen and Zhaoyang Zhang, "Low complexity multiview video plus depth coding," *IEEE Transactions on Consumer Electronics,* vol. 57, no. 4, pp. 1857-1865, November, 2011. Article (CrossRef Link)

[4] Ling-Jiao Pan, Byung-Tak Lee and Nac-Woo Kim. "H.264-based depth map-sequence coding algorithm using mode mapping for fast 3-D video compression," *Optical Engineering,* vol. 50, no. 1, pp. 017401, January, 2011. Article (CrossRef Link)

[5] Hui-Ping Deng, Li Yu, Bin Feng and Qiong Liu, "Structural similarity-based synthesized view distortion estimation for depth map coding," *IEEE Transactions on Consumer Electronics,* vol. 58, no.4, pp. 1338-1344, November, 2012. Article (CrossRef Link)

[6] S.-H. Tsang, Y.-L. Chan and W.-C. Siu, "Efficient intra prediction algorithm for smooth regions in depth coding," *Electronics Letters,* vol. 48, no. 18, pp. 1117-1119, August, 2012. Article (CrossRef Link)

[7] Jin Heo and Yo-Sung Ho, "Improved context-based adaptive binary arithmetic coding over H.264/AVC for lossless depth map coding," *IEEE Signal Processing Letters,* vol. 17, no.10, pp. 835-838, October, 2010. Article (CrossRef Link)

[8] JVT of ISO/IEC MPEG and ITU-T VCEG, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," Doc. G050r1, March, 2003.

[9] Erjun Zhao and Dan Liu, "Fractal image compression methods: a review," in *Proc. of 3rd International Conference on Information Technology and Applications,* pp. 756-759, July 4-7, 2005. Article (CrossRef Link)

[10] Yuval Fisher, "Fractal image compression*," Fractals*, vol. 2, no. 3, pp. 347-361, September, 1994. Article (CrossRef Link)

[11] Hannes Hartenstein, Matthias Ruhl and Dietmar Saupe, "Region-based fractal image compression," *IEEE Transactions on Image Processing,* vol. 9, no. 7, pp. 1171-1184, July, 2000. Article (CrossRef Link)

[12] H. Mohamadi, S. Nodehi and M. Tayarani, "A local search operator in quantum evolutionary algorithm and its application in fractal image compression," in *Proc. of The 2nd International Conference on Computer and Automation Engineering,* pp. 710-715, February 26-28, 2010. Article (CrossRef Link)

[13] Hui Yu, Li Li , Dan Liu, Hongyu Zhai and Xiaoming Dong, "Based on quadtree fratal image compression improved algorithm for research," in *Proc. of International Conference on E-Product E-Service and E-Entertainment,* pp. 1-3, November 7-9, 2010. Article (CrossRef Link)

[14] Tamás Kovács, "A fast classification based method for fractal image encoding," *Image and Vision Computing,* vol. 26, no. 8, pp. 1129-1136, August, 2008. Article (CrossRef Link)

[15] Z. B. Chen, P. Zhou and Y. He, "Fast integer pixel and fractional pixel motion estimation for H.264/AVC," *Journal of Visual Communication and Image Representation,* vol. 17, no. 2, pp. 264-290, April, 2006. Article (CrossRef Link)

[16] Chun-Ho Cheung and Lai-Man Po, "A novel cross-diamond search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 12, no. 12, pp. 1168-1177, December, 2002. Article (CrossRef Link)

[17] Gianluca Cernigliaro, Fernando Jaureguizar, Julián Cabrera and Narciso García, "Low complexity mode decision and motion estimation for H.264/AVC based depth maps encoding in free viewpoint video," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 23, no. 5, pp. 769-783, May, 2013. Article (CrossRef Link)

[18] Bo Zhu, Gangyi Jiang, Yun Zhang, Zongju Peng and Mei Yu, "View synthesis oriented depth map coding algorithm," in *Proc. of Asia-Pacific Conference on Information Processing,* pp. 104-107, July 18-19, 2009. Article (CrossRef Link)

[19] M. Tanimoto, T. Fujii and K. Suzuki, "View synthesis algorithm in view synthesis reference software 3.0 (VSRS 3.0)," Doc. M16090, February, 2009.

[20] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," Doc. VCEG-M33, April, 2001.

[21] http://iphome.hhi.de/suehring/tml/download/old_jm/

Shiping Zhu received the B.S. and M.S. degrees from Xi'an University of Technol ogy, Xi'an, China, in 1991 and 1994, and the Ph.D. degree from Harbin Institute of T echnology, Harbin, China, in 1997. From 1997 to 1999, he was a Postdoctoral Fellow with Beihang University, Beijing, China. From 2000 to 2002, he was a Postdoctoral F ellow with the Brain and Cognition Research Center, Université Paul Sabatier, Toulo use, France. From 2002 to 2004, he was a Postdoctoral Fellow with the Department o f Computer Science and Department of Electrical and Computer Engineering, Univer sité de Sherbrooke, Sherbrooke, QC, Canada. Since 2005, he has been an associate pr ofessor with the Department of Measurement Control and Information Technology, S chool of Instrumentation Science and Optoelectronics Engineering, Beihang Universi ty, Beijing, China. His research interests include image processing and video coding, computer vision, machine vision for 3-D measurement. (E-mail: spzhu@163.com).

Dongyu Zhao received the B.S. degree in Measurement and Control Technology a nd Instrumentation from Beijing Jiaotong University, Beijing, China in 2012, and she is currently pursuing the M.S. degree in Instrumentation Science and Technology at B eihang University, Beijing, China. Her research interests include image processing and video compression. (E-mail: dongyuzhao@msn.cn).