AN ES BASED EFFECIENT MOTION ESTIMATION TECHNIQUE FOR 3D INTEGRAL VIDEO COMPRESSION

S. Adedoyin[1], W.A.C. Fernando[1], A.Aggoun[2], W.A.R.J.Weerakkody[1]

[1]Centre for Communications Systems Research, University of Surrey, Guildford, GU2 7XH, UK
[2]School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, United Kingdom

sadedoyin@hotmail.com; w.fernando@surrey.ac.uk; amar.aggoun@brunel.ac.uk; rajithawee@yahoo.com

ABSTRACT

In this paper we propose a novel approach to use both motion and disparity information to compress 3D integral video sequences. The integral video sequence is decomposed into 8 viewpoint video sequences and a block search is performed to jointly exploit the motion and disparity redundancies to maximize the compression. An Evolutionary Strategy (ES) based search algorithm is used to reduce the complexity. Experimental results show that an ES based strategy can reduce the motion estimation complexity by 95%.

Index terms - *3D video, Evolutionary Strategy, 3D integral imaging*

1. INTRODUCTION

Emulating human 3D vision, which relies on processing 'left eye' and 'right eye' images, has been a dream for many years. Current stereo-spectacles, cameras, and displays have relied on creating two images focusing on two perspective viewpoints (the parallax effect) leading to an impression of depth. Such systems require users to simultaneously focus on the screen plane whilst converging their eyes to a different point in space, which causes eyestrain and fatigue, and therefore have not found wide applications.

Furthermore multiview stereoscopic systems tend to suffer from the effect of 'card boarding' and 'flipping'. So ideally we require true autostereoscopic 3D visualisation systems, exhibiting full parallax and continuous view points which allow accommodation and convergence to function in unison. Holographic techniques demonstrating this feature, are being researched by various groups to produce full colour realistic spatial images. However, such systems are inherently slow and costly, their disadvantages including: the need for coherent radiation, poor colour rendering, specialised environmental conditions, and the huge data content needed for processing.

Integral imaging is a technique that is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution by using unique optical components [1]. It is akin to holography in that 3D information recorded on a 2D medium can be replayed as a full 3D optical model, however, in contrast to holography, coherent light sources are not required. This conveniently allows more conventional live capture and display procedures to be adopted. A 3D integral image is represented entirely by a planar intensity distribution, which may be recorded on to a photographic film for later electronic scanning and processing or directly recorded as an intensity distribution using a CCD with a standard camera lens.

TV based on 3D integral imaging video technology, that requires only one camera, will be attractive to service providers because it will seamlessly provide the added value of 3D realism. 3D integral imaging encoded video can be designed to be scalable with 2D video and can be encoded efficiently so as to economically provide attractive high value services over high value systems. Thus, the development of a 3D integral imaging TV system will also demonstrate how added-value broadband services of this type can be delivered, providing benefit to designers of these type of services in the future. However, lots of work needs to be done in this area especially in the compression of 3D integral imaging video. In this paper we propose an ES based motion estimation algorithm for 3D integral imaging video.

The rest of the paper is organized as follows. Section 2 comprises of the proposed 3D video encoder with ES-Based ME, section 3 contains the results and section 4 concludes this paper.

2. PROPOSED 3D VIDEO ENCODER WITH ES-BASED ME

Evolutionary computation (EC) theories were developed originally from observing natural evolution of life form. Because of this, the terminology surrounding the field of EC is full of analogies with natural evolutionary process. It was particularly from Darwin's theories [2] that the best techniques regarding the optimization, modelling and the control of unknown processes were developed. EC has long been exploited in the video coding field. A very well known form of EC called Genetic Algorithm (GA) was used to perform image registration as part of a larger Digital Subtraction Angiography (DAS) system [3][4]. Subsequently, GA search algorithm has been applied for

motion estimation [5][6][7][8]. Hardware implementation of Four-Step genetic search algorithm was proposed in [5].

Similarly, the Evolutionary Strategies (ES) were developed to solve technical optimization problems in video coding field. Thus, the motion and disparity estimation has been carried out using a $(1+\lambda)$ rudimentary ES for stereoscopic video sequences, which includes calculation of P- and B-frames, weighted prediction, joint motion disparity estimation [10][11]. In this paper, we apply ES to estimate the motion and inter-view disparity vectors in lenticular video coding.

2.1 Coding structure

Integral imaging involves many microlenses in one recording. Consequently a number of perspective 2-D images, as many as there are lenses in the lens array, are obtained in a single capture process. These images are called "elemental images" according to the convention in the literature. In this paper, the images are recorded using a lenticular lens sheet (1D cylindrical microlens array). This results in vertically running bands to present in the planar intensity distribution captured by the 3D integral imaging camera [1].

Each viewpoint video sequence represents a unique recording direction of the object scene. Hence the 3D integral video sequence can be separated into its respective distinctive viewpoint videos. Figure 1 illustrates the viewpoint extraction for a lenticular video of 4 distinctive viewpoints. Columns of pixels formed by each micro lens representing the similar view points are placed near to each other to form the viewpoint images. These viewpoint video sequences are used in the motion compensation.
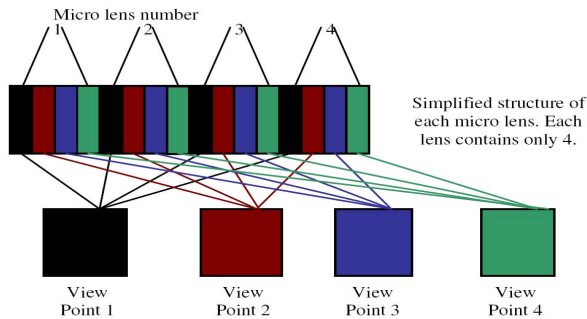

Figure 1 Viewpoint extraction

The best motion vectors for each viewpoint can be found by applying a conventional block-matching algorithm. However such a technique would be too complex and time consuming. Since viewpoint images are captured by slightly different viewing angles, there is a great deal of redundancies. Therefore, it is advantageous to find a proper

set of motion vectors only for a single viewpoint and utilize this correlation to minimize the overall coding complexity.

Compression efficiency can be maximized if disparity correlations amongst the viewpoints are also considered. Exploitation of such additional redundancies, however, increases the computational complexity further. Following this argument, we propose to motion compensate one of the middlemost viewpoints and the rest of the view points are motion and disparity compensate jointly considering the motion compensated viewpoint as the base viewpoint. Proposed structure is illustrated in figure 2. Note that the figure represents only the first five view points. After coding the base viewpoint (i.e. Viewpoint 5) with respect to the base viewpoint of the reference frame, viewpoint 3 is coded taking the corresponding viewpoint from the reference frame and the reconstructed version of the base viewpoint of the current frame as references. Subsequently, the viewpoint 4 is coded taking reconstructed versions of the viewpoint 3 and 5 of the current frame and the viewpoint 4 of the reference frame as references. This process is repeated for the other viewpoints in the sequence.
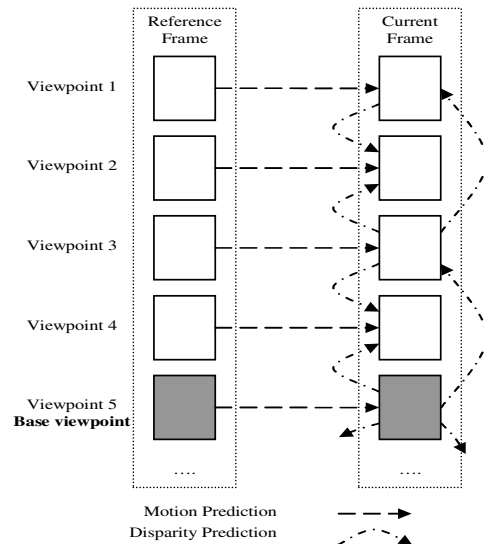

Figure 2 Proposed motion and disparity estimation technique

2.2 Proposed evolutionary strategy

ES typically uses deterministic selection in which the worst solutions are purged from the population based directly on their fitness function value. The $(\mu+\lambda)$ Evolutionary Strategy demonstrated in Figure 3 is used in this work with an increasing level of imitation of biological evolution [9], where μ means the total number of parents in

previous population, and λ stands for the number of offspring generated from mutated parents.
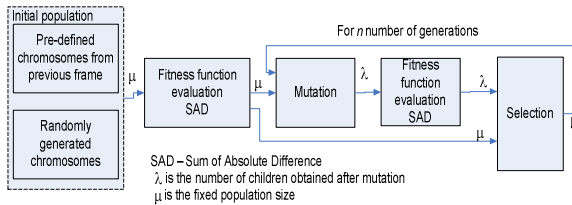

Figure 3 ($\mu+\lambda$) Evolutionary Strategy-based motion estimation algorithm

2.3 Chromosome representation

Each chromosome represents three elements of a motion vector, i.e. the data for coordinates x and y and the reference frame (in case of a motion compensation) or viewpoint (in case of a disparity compensation). Each element is described by 2 genes: object and strategy parameters, as shown in the Figure 4.

Object parameters define the actual coordinate in the image. The value of object parameter is determined from the search window size. For example, if the search window is within the range [-16, 16], then the value of object parameter can take any integer number inside of this range. The search window size depends on the maximum motion vector size. Strategy parameter determines wherever a local or global search will be carried out. The smaller the value of the strategy parameter, the more localized the search process becomes. The negative value defines the decrement of mutated gene and positive values respectively determine the increment of the mutated gene. The strategy parameter depends on the window size and can take any value up to its maximum. In order to implement the local search, we choose to set the strategy parameter to values -1 or 1.
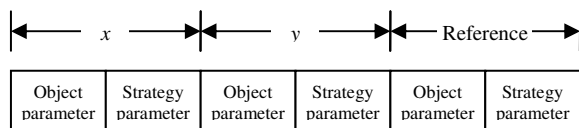

Figure 4 Chromosome representation

2.3.1 Fitness Function
The quality of the chromosome is defined by fitness function. Fitness function is calculated based on the Sum of Absolute Difference (SAD). Each chromosome in newly generated population is evaluated using fitness function.

2.3.2 Evolutionary strategy operators
In order to reduce the number of generations required to obtain the satisfactory solution, the initial population is generated from predefined and randomly generated chromosomes. The pre-defined chromosomes are determined based on the knowledge from the previously coded blocks from both adjacent viewpoints and viewpoints

from the previously encoded frame. Selection takes place only amongst the offspring's (mutated values) and parents. The size of population in the next generation is fixed. The new population is generated from the best chromosomes from the previous population that combines both parents and offspring as shown in Figure 3.

Mutation rate defines the percentage of genes to be mutated in a newly generated population. Mutation rate used in the experiments was set to 8.5%. In general the value of the strategy parameter is generated randomly from the local search increment window specified in advance. In our case, the strategy parameter value can vary within the range [-1, 1]. The new value of the object parameter (if this gene has been chosen to be mutated) is defined as following:

$$x_{op}^{new} = x_{op} + x_{sp} \qquad (1)$$

where x_{op}^{new} is the new value of mutated object gene for x coordinate, x_{op} and x_{sp} are the values of object and strategy parameters for x coordinate respectively. Similarly the parameters for y coordinate and the reference are calculated.

3. RESULTS

Proposed joint motion and disparity estimation technique is implemented in a 3D-DCT integral image codec based on the architecture described in [1] for performance evaluation. An adaptive arithmetic coder is used for the entropy coding and the quantizer step size ranged from 10 – 50. For experimental purposes a population size of 30 was used for ES. The number of generation and the mutation rate are set to 10 and 8.5% respectively based on preliminary experimental results. Peak Signal-to-Noise Ratio (PSNR) is used to measure the objective quality. Figure 5-7 show the objective quality comparison of the proposed ES based joint motion and disparity estimation technique (denoted as ES-JM&D) for the Room integral video test sequence of image size 512×512 against three reference cases namely:
(i) Motion only full search (FS-MOTION) – motion compensated prediction is used and the motion vectors are calculated using the full search algorithm.
(ii) Motion only ES search (ES-MOTION) – as above except full search is replaced with ES search.
(iii) Joint motion and disparity full search (FSJM&D) – joint motion and disparity compensated prediction is used with ES search. Figure 5 and Figure 6 depict the relative performance of the above algorithms for selected viewpoints and Figure 7 does for the entire frame. Results show that FS-JM&D has outperformed FS-MOTION by over 1 dB. This is clear evidence that the use of disparity redundancies together with the motion can greatly improve the compression efficiency. Combining ES with JM&D reduces coding complexity by approximately 95% this can be seen in Table 1. This is also

noted in comparing ES with a motion only full search, ES gives an 84% decrease in coding complexity. Combining ES with JM&D or with a motion only search results in a small reduction in image quality, as seen in Figures 5-7.

4. CONCLUSION

This paper proposes a novel technique to exploit motion and disparity redundancies in 3D integral video sequence and low complexity optimization technique. Experimental results show that ES based joint motion and disparity estimation technique achieve over 1 dB objective quality gain while maintaining up to 94% computational cost saving.
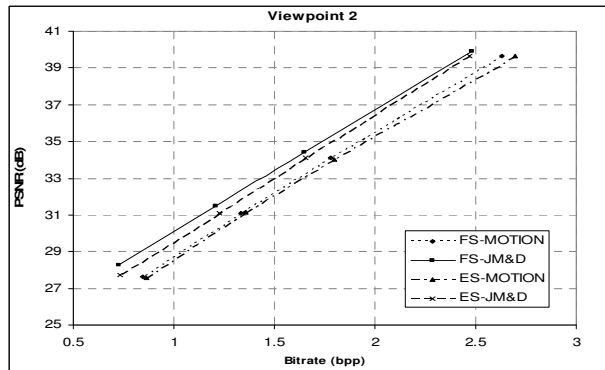
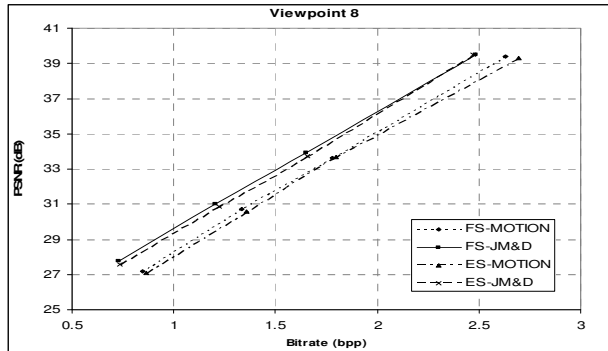Figure 5 Objective quality comparisons for the viewpoint 2

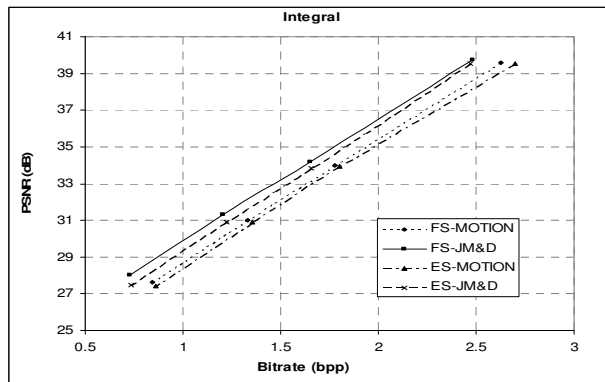Figure 6 Objective quality comparisons for the viewpoint 8

Figure 7 Objective quality comparison for the complete integral video

TABLE 1

	Number of search points		Complexity
	Full search	ES search	
Motion	1024	132	87.11%
Disparity	3072	183	94.04%

5. REFERENCES

[1] A Aggoun: 'A 3D DCT Compression Algorithm For Omnidirectional Integral Images' Proceedings of IEEE ICASSP 2006, Volume 2, 14-19 May 2006 Page(s):II-517 - II-520.

[2] Charles Darwin, "The Origin of Species: By Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life (Bantam Classic)," Bantam Classics, Reprint. 1999

[3] Fitzpatrick, J.M., Grefenstette, J.J. and Van-Gucht, D. "Image registration by genetic search," Proceedings of Southeastcon 84, Louisville, KY, 460-464, Apr 1984.

[4] Grefenstette, J.J. and Fitzpatrick, J.M. "Genetic search with approximate function evaluations," Proc. Intl. Conf. on Genetic Algorithms and their Applications, Pittsburgh, PA, 112-120, Jul. 1985.

[5] Man F. So & Angus Wu, "Hardware Implementation of Four-Step Genetic Search Algorithm", IEEE Signal Processing Society 1999 Workshop on Multimedia Signal Processing, Copenhagen, Denmark, September 1999.

[6] Xu Yuelei, Bi Duyan and Mao Baixin, " A Genetic Search Algorithm For Motion Estimation", Proceedings of 5th International Conference on Signal Processing Proceedings, Beijing, China, 2000.

[7] Guanghua Qiu, Chaohuan Hou, "A New Fast Algorithm for the Estimation of Block Motion Vectors", Proceedings of 3rd International Conference on Signal Processing, Beijing, China, 1996.

[8] Shen Li, Wei-pu Xu, Hui Wang, Nan-ning Zheng, "A Novel Fast Motion Estimation Method Based on Genetic Algorithm", Proceedings of International Conference on Image Processing, Kobe, Japan, 1999.

[9] Ingo Rechenberg, Evolutionsstrategie '94. Stuttgart: Frommann-Holzboog 1994.

[10] K. Ponudurai, W.A.C. Fernando and K.K. Loo, "Joint Motion and Disparity Estimation in Stereo Video Sequences Using Evolutionary Strategy," Proceedings of 1st Regional Conference on ICT and E-Paradigms, Colombo, Sri Lanka, 2004.

[11] K. Ponudurai, W.A.C. Fernando and K.K. Loo, "Joint Motion and Disparity Estimation in Stereo Video Sequences Using Evolutionary Strategy," Proceedings of The Seventh International Symposium on Wireless Personal Multimedia Communications, Abano Terme, Italy, September 2004.