Compression of IP Images for Autostereoscopic 3D Imaging Applications

N.P.Sgouros[1], A.G.Andreou[1], M.S.Sangriotis[1], P.G.Papageorgas[2], D.M.Maroulis[1], N.G.Theofanous[1]
1. Dep. of Informatics and Telecommunications, University of Athens, Greece
2. Dep. of Electronics, Technological Education Institute of Piraeus, Greece
E-mail: {nsg,andyhot,sagri}@di.uoa.gr,ppapag@cc.uoa.gr,{dmarou,optel}@di.uoa.gr

Abstract

An Autostereoscopic 3D viewing system that operates on the principles of Integral Photography-IP provides a unique sense of depth, full parallax and multi-view functionality. The inherent redundancy of these images results into great amounts of data that should be efficiently coded for transmission or storage operations. In this communication a method for efficient coding of such images is presented, targeting to 3D imaging but video applications. The method is based on common techniques broadly used in image compression and properly adjusted in order to take advantage of the spatial redundancies of IP images. The generality and flexibility of the proposed approach along with the stability for a wide range of bit rates constitutes the basic characteristics of the technique. The proposed technique can be easily realized in software or hardware for computer based or standalone applications.

1. Introduction

Over the past few years the rapid increase in processing power and graphic card acceleration, combined with improvements in high fidelity optical systems, revived the interest for 3D applications. Many promising technologies evolved, ranging from polarizing glasses, mostly used at the early stages of 3D cinema, to most sophisticated techniques like eye shuttering glasses [1] and more recently autostereoscopic display devices [2].

Autostereoscopic display devices provide 3D stereoscopic view without the need of additional glasses, reducing eye fatigue while most of them allow multiple viewers which all experience the 3D effect.

A special category of autostereoscopic displays, functions on the principles of integral photography (IP) first introduced by Lippman[3] back in 1908.

Such display devices usually consist of a high resolution LCD equipped with an appropriate parallax barrier system. Most recently lenticular or microlens based autostereoscopic monitors are manufactured.

The basic functionality of such displays is shown in Figure 1.

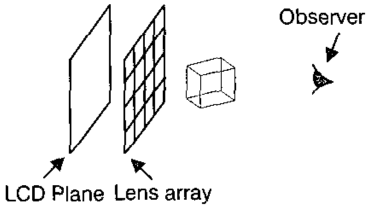


Figure 1. Concept of 3D IP display system

Specifically such autostereoscopic display systems can be of great use in medical[4], educational and entertainment[5] applications.

One of the main concerns in developing such applications is the necessity to cope with high resolution images that result in high bandwidth and storage requirements for the capturing and reproduction of 3D objects. Consequently a high-efficiency compression scheme of the associated data is crucial for the successful implementation of 3D applications.

The compression technique described takes advantage of redundancies present in IP images by efficiently estimating neighboring blocks of an image that correspond to slight different aspects of a scene. This approach aims in the maximization of the fidelity of an image for a given bit rate.

To the knowledge of the authors these are the first compression results presented for microlens-encoded IP images that exhibit multidirectional parallax. The versatility of the size of each image element introduced by the lens pitch to LCD pitch ratio enforces the need of algorithms that estimate the high cross correlation areas between adjacent image blocks. The generality of such an algorithm and the improvement in performance for a wide range of bit rates makes it an intriguing candidate for all kinds of IP imaging systems. Such applications include not only microlens systems but also lenticular based representations.

2. Microlens image essentials

Microlens-encoded images (also called by many IP field experts as aspectograms [6] due to the fact that each microlens represents a different aspect of a scene), can be created by means of photography (Integral Photography-IP) or artificially constructed (Computer Generated IP-CGIP) from an initial three dimensional object or scene.

A system that uses a microlens array as a mean of capturing or display of a 3D object, provides horizontal as well as vertical parallax, in contrast with a lenticular system that exhibits only unidirectional parallax.

A shortcoming of the multidirectional parallax is the limited resolution of such a system which is expected to be resolved by the advances in lens and LCD manufacturing technologies.

Nowadays different geometries of lens placement patterns and lens types are used with full fill factor arrays providing better results. The placement of the lenses on an array can be hexagonal or square while the lens type used is usually hexagonal or square, satisfying in this manner the full fill factor property.

3. Aspectogram Compression

Basic compression schemes that apply in still images and video, achieve significant reduction in the amount of two dimensional data that must be stored or transmitted by exploiting the inter-pixel and inter-frame correlation respectively.

Regarding still images most of these techniques like JPEG [7] involve a transform coding stage of the data, followed by quantization and entropy coding.

In the case of video sequence coding, apart from the previous stages for intra-frame coding, a motion estimation algorithm predicts subsequent frames, a technique vastly used in the MPEG video coding algorithm [8].

In the case of aspectograms as in Fig. 2 the inherent redundancy that is present in the IP image due to the high correlation of neighboring areas, as it can be noticed in figures 3a, 3b remains unexploited by the aforementioned procedures.

Many alternatives have being proposed by different groups in order to take advantage of the intrinsic redundancies in IP images.

These alternative schemes include hybrid DCT/DPCM encoding [10], 3D-DCT [11] and recently an adaptive 3D-DCT technique [12] to improve coding efficiency. Until now all simulation results of the previous algorithms are presented for lenticular sheet images with matching dimensions for the lenticules and the transform block used.

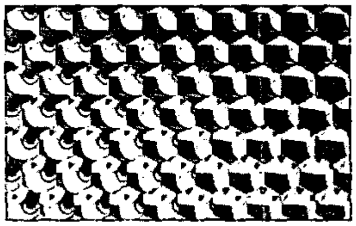
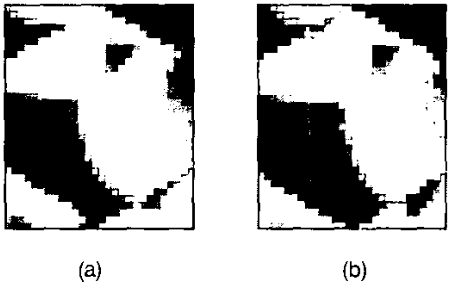


Figure 2. Hexagonal geometry and placement of lenses for a specific IP image[9]



(a) (b)

Figure 3a,b. Adjacent sub-images exhibiting horizontal disparity

The results with the use of the above techniques are promising since all achieve good results compared to the Baseline JPEG algorithm.

4. Compression by disparity compensation

Our coding technique aims in exploiting the extra redundancy present, by considering that neighboring areas represent a scene by slightly different aspects.

This disparity between adjacent lens encoded sub-images can be modeled as the disparity in the standard left-right stereo image pairs.

More specifically a segmentation of the image is performed in areas of dimensions close to the aspects dimensions in order to maximize the cross correlation in a sequence of sub-images and provide flexibility and feasibility like the present 2D image coding architectures.

As shown in Fig. 4 an image can be approximated as a sequence of rectangular sub-images that all exhibit a certain degree of disparity which can be characterized as a spatial motion.

In the developed method a set of consecutive sub-images is realized, as shown in Fig. 5, that forms a fixed size coding sequence of the original IP image.

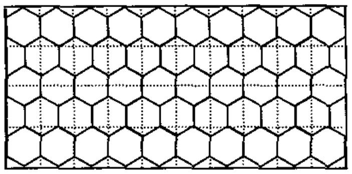


Figure 4. Rectangular lattice for segmentation of a hexagonal geometry image

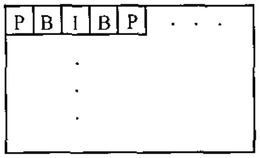


Figure 5. Coding sequence

5. Coding of I, P and B sub-images

At first an Intra sub-image (I) is encoded in a JPEG like manner by using a DCT transform, quantization and then efficient coding to optimize performance.

Specifically an I sub-image is segmented to 8X8 pixel blocks and a 2D DCT is applied on each block followed by quantization as described in the JPEG standard.

The coefficients are zigzag scanned and finally coded with a combination of run-length and Huffman procedures.

Based on the encoded I sub-image an estimation of two predicted sub-images (P) is formed by estimating the proper disparity vectors for each P sub-image as depicted in Fig. 6. This is followed by a residual or error sub-image calculation as the difference of the original and the disparity estimated P sub-image. The residual sub-image is then coded following the rules described for the I sub-image coding.

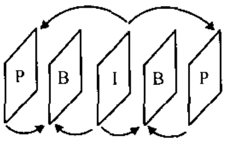


Figure 6. Disparity vector estimation of P and B subimages

For the bidirectional predicted sub-images (B) a prediction is formed by disparity vector estimations based on the adjacent set of I and P sub-images as shown in Fig. 6.

Apart from the previous scheme the encoder operates also in P-I-P sequences in order to provide an increase in performance at high bit rates. These two modes of operation are used in the encoder according to the desired quality for the reproduced image.

6. Disparity estimation technique

Theoretically in a properly formed aspectogram the disparity between adjacent aspects can be analytically calculated [13].

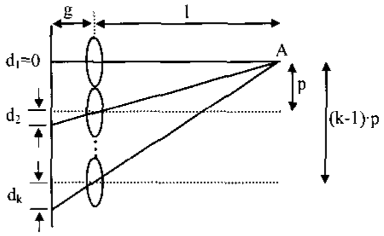


Figure 7. Geometry for disparity calculations

As it is shown in Fig. 7 for a Point A in the object space the disparity of the elemental image of that point from the k-th lens is given by the following equation,

$$d_k = (k-1)\frac{gp}{l} \qquad (1)$$

where p is the lens pitch, g is the gap between the recording plane and the lens array and l is the distance between the object plane and the lens array.

Without loss of generality we have assumed that point A lies on the primary lens axis of the first lens of the array, thus d_1=0.

By using d_2 as a metric for the disparity between lenses equation (1) can be rewritten in the form:

$$d_k = (k-1)d_2 \qquad (2)$$

Disparity estimation is currently performed by using a unidirectional full pixel search. The simplicity of this unidirectional approach is based on the previous theoretical study and the results are almost identical to the ones calculated using a classical multidirectional search over an area to estimate a disparity vector. The block size used in the disparity estimation technique is set to an 8X8 block which gave optimal results for the size of the aspects used in this study and can be dynamically adjusted to match other sub-image sizes. The best matching for a predicted block is decided by the use of a mean square error (MSE) estimator.

In our case the sub-image dimensions are chosen close to the aspect dimensions that vary slightly. This mismatch is produced by the fact that the lens pitch is not an integer multiple of the dot pitch used for the aspectogram representation, resulting in a shift in the vector coordinates during the image scan procedure.

For this reason the disparity vectors coordinates were not assumed fixed, but have been evaluated with the previously described method. All vectors were finally DPCM coded to efficiently encode the small differences occurring, followed by a Huffman encoding that reduces the extra bits needed for the disparity vectors encoding.

7. Simulation and Results

The images used as an input to the encoder described earlier in the text were designed for reproduction with a Fresnel technologies #300 hexagonal microlens array.

To evaluate the compression algorithm performance an objective quality measure was used, in order to provide comparable results to other standard image compression techniques. In particular the peak signal-to-noise ratio (PSNR) was used to evaluate the performance and the results were compared against the popular Baseline JPEG standard which is vastly used for benchmarking purposes.

As it is shown in Fig. 8 the first mode of operation of the Spatial Disparity Estimation (SDE) encoder described in the text that uses a P-B-I-B-P sequence of sub-images outperforms Baseline JPEG for the low bit rates area.

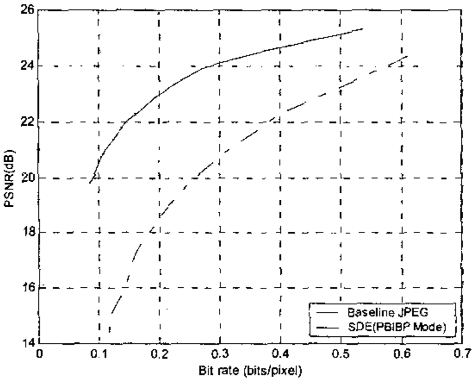


Figure 8. Results for PBIBP mode of operation

For the high bit rates area the second mode of operation of the SDE encoder is used by coding the image by using only I and P sub-images. The P-I-P sequences produced provide a stability in performance for these rates as it can be noticed in Fig. 9.

8. Conclusions

The compression technique described in this paper aims to provide a general framework to IP image compression by taking into account the spatial disparity between adjacent aspects, in the same way that MPEG takes into advantage the motion that occurs in subsequent frames.

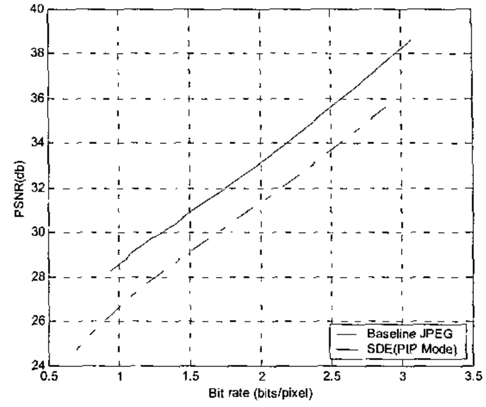


Figure 9. Results for the PIP mode of operation

The proposed technique exhibits high performance in a rate distortion sense and achieves stability for both low and high bit rates.

Due to the fact that this technique is based on the intrinsic physical properties of a CGIP or a captured IP image it can be applied in all IP variations regardless of the lens type used (lenticular or microlens) and the placement of the lenses (hexagonal or square).

The main modules of the algorithm are based on effective combinations of the JPEG coding and MPEG motion estimation techniques. The simplicity of the algorithm derives from the use of these modules improving the feasibility, robustness and simplifying the implementation of the algorithm.

Our further work includes the extension of the algorithm to encode IP video as well as certain adjustments in quantization and transform coding techniques to improve performance of the algorithm.

References

[1] *Developers Handbook*, Stereographics,1997, www.stereographics.com

[2] M.Halle, "Autostereoscopic displays and computer graphics", *Computer Graphics*, 31(2), ACM SIGGRAPH, 1997, pp. 58-62

[3] G.Lippmann,"La photographie integrale", *C. R. Acad. Sci*,146,1908,pp.446-451

[4] H.Liao,S.Nakajima et.al.,"Intra-operative Real-Time 3-D Information Display System Based on Integral Videography", MICCAI'01,LNCS 2208, 2001,pp. 392-400

[5] P.Harman, "Home based 3D entertainment-an overview",*Proc.* ICIP(1),2000,pp. 1-4

[6] Communication with Dr.S.Dudnikov.

[7] W.B. Pennebaker, J.L.Mitchell, *JPEG Still image data compression standard*,VNR,NY,1993

[8] K.R. Rao, J.J.Hwang, *Techniques & standards for image-video & audio coding,* Prentice Hall PTR,NJ,1996

[9] "Passive autostereoscopic light field displays", http://graphics.lcs.mit.edu/~aisaksen

[10] M.Forman, A.Aggoun, M.McCormick, "Compression of integral 3D TV pictures",*Conf.Pub.* IEE,vol.410, 1995, pp. 584-588

[11] M.C.Forman,A.Aggoun,M.McCormick, "A novel coding scheme for full parallax 3D-TV pictures",*Conf.Proc.*IEEE ICASSP'97,vol. 4, 1997, pp.2945-2947

[12] R.Zaharia, A. Aggoun,M.McCormick, " Adaptive 3D-DCT compression algorithm for continuous parallax 3D integral imaging",*Signal Processing: Image communication* (17), Elsevier Science,2002,pp231-242

[13] J.H. Park, S.W. Min et al. "Analysis of viewing parameters for two display methods based on integral photography", *Appl.Opt.* 40(29),Opt.Soc.America,2001,pp.5217-5232