Multiview Image Compression:
Future Challenges and Today's Solutions

N.Sgouros, M.Sangriotis, D.Maroulis

Dept. of Informatics and Telecommunications, University of Athens
Panepistimiopolis, Ilissia, Athens 15784, Greece
E-mail: {nsg, sathin, sagri, dmarou, optel}@di.uoa.gr

Abstract

Large data volumes that are produced during multiview image generation have to be efficiently compressed in order to be stored or transmitted. Two main classes of encoders use transform coding techniques either by utilizing spatial prediction methods or by using higher degree transforms. Our work summarizes the various types of multiview image sets and the corresponding coding techniques and, provides a useful comparison of the compression efficiency of these two classes of multiview image encoders over a variety of test images. Representation quality and application specific requirements are taken into account in order to decide in favour of the encoder to be used.

Keywords: Multiview image coding, Transform coding, Higher order transformations

1. Introduction

One of the dominant standards in today's stereoscopic viewing methods is the use of stereoscopic image pairs that are appropriately projected to the viewer's eyes and are called stereopairs. As new technologies evolve in the field, a series of multiview stereoscopic displays appear, that use larger image sets which are captured from slightly different viewpoints. Due to this fact the procedure produces more realistic three dimensional representations in regard to the classic two view approach. Most of the multiview techniques incorporate all necessary optical components in the display device, providing an unconstrained three dimensional experience within a viewing zone.

It is evident that these image sets contain intra-image as well as inter-image redundancy that when properly exploited can reduce the total amount of image data that have to be stored or transmitted. To this end the large amount of data that result from capturing two or more views of a scene in order to create its stereoscopic representation or produce a stereoscopic video sequence, has to be efficiently compressed prior to storage or transmission.

There are a number of techniques developed to day that deal with the classic problem of efficiently coding a stereopair [1] as well as a set of multiview images [2-4]. All these techniques are based on the fact that the image set represents the same scene from slightly different aspects. Most of the proposed methods [2-3] are generally based on block matching algorithms, which trace corresponding points between a stereopair and coding the resulting differences. Recent work in multiview image compression is summarized in [3] although no specific compression ratios are noted for quantitative analysis purposes.

Two of the core techniques that can be used to provide efficient coding of stereoscopic image pairs are based on the energy compaction properties of the Discrete Cosine Transform (DCT) that is widely used in image compression, and the exploitation of inter-image correlation using a disparity estimation scheme. The properties of the Discrete Cosine Transform (DCT) make it one of the most valuable tools in the field of signal and image coding [5]. Some of the most famous compression standards like JPEG [6] are based on the energy compaction properties of the DCT. In addition, an extension of the disparity estimation scheme is extensively used for compressing video sequences, in the form of motion vectors that predict similarities between consecutive frames.

Due to the three dimensional nature of the problem higher order transforms can be also used in order to combine the merits of the transform techniques while simultaneously solve the disparity estimation problem. Moreover time efficiency of a three dimensional compression scheme should be considered as it is required when real time constraints are imposed, i.e. real time three dimensional video transmission.

In this work we summarize the properties of a multiview disparity encoder (MDE) and a three dimensional coding technique based on the three dimensional DCT (3D-DCT). These two methods can be considered as today’s solutions to the future problem of three dimensional image compression, as they are characterized by their efficiency and robustness in a wide variety of different setups.

Our work is organized as follows. Firstly, we provide a brief description of the dominant three dimensional acquisition methods used in order to produce a multiview image data set. The volume structure that has to be constructed in both cases by proper arrangement of the multiview image set is described in section 3. In section 4 we present an overview of the MDE and the 3D-DCT encoders. Due to their novel characteristics, the 3D-DCT quantization unit and the coefficient scanning strategies are further described in section 5. In Section 6 we describe the additional entropy coding stages in each case. Finally, section 7 contains the limitations of each method imposed by theoretical and technological restrictions along with a discussion on the next steps in three dimensional image coding.

2. Image Acquisition

The first step prior to encoding a multiview image set is to find a way to rearrange the data in a proper way in order to improve the efficiency of the compression scheme. This procedure usually depends on the methodology used to capture a 3D scene. In general, arbitrary configurations of camera setups can be used in the acquisition stage. This happens because of the different 3D display devices used today [3,4,7].

Two of the basic capturing methodologies use camera setups which consist of multiple cameras arranged in a convergent or parallel axes topology as depicted in Fig.1. These two topologies can be also realized with a single camera performing a rotational-translational or just a translational movement, as marked with the arrow in Fig.1a and Fig.1b respectively.

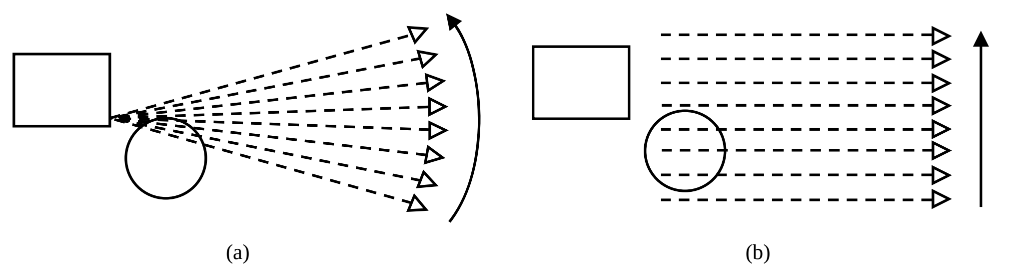

Fig. 1. Camera setup with (a) Convergent and (b) parallel axes.

While many other capturing setups and corresponding display devices are proposed over the years as discussed earlier, i.e. using 2-D camera arrays or 2-D lens arrays, we confined our work in multiview image sets generated in the ways described in Fig.1 as these techniques dominate current trends in multiview image generation.

However most of the techniques devised for multiview datasets can be easily adapted in order to address different setups.

3. Data Structure Formation

A data structure has to be defined for each encoder by proper formation of the acquired image set. In the case of the MDE encoder the multiview image set is rearranged based on the direction of the camera movement as depicted in Fig. 2a. The corresponding rearrangement for the case of the 3-DCT encoder is realized by creating a data volume that is actually a parallelepiped formed by placing consecutive images in an order that depicts the translational or translational-rotational movement of the camera as depicted in Fig 2b.

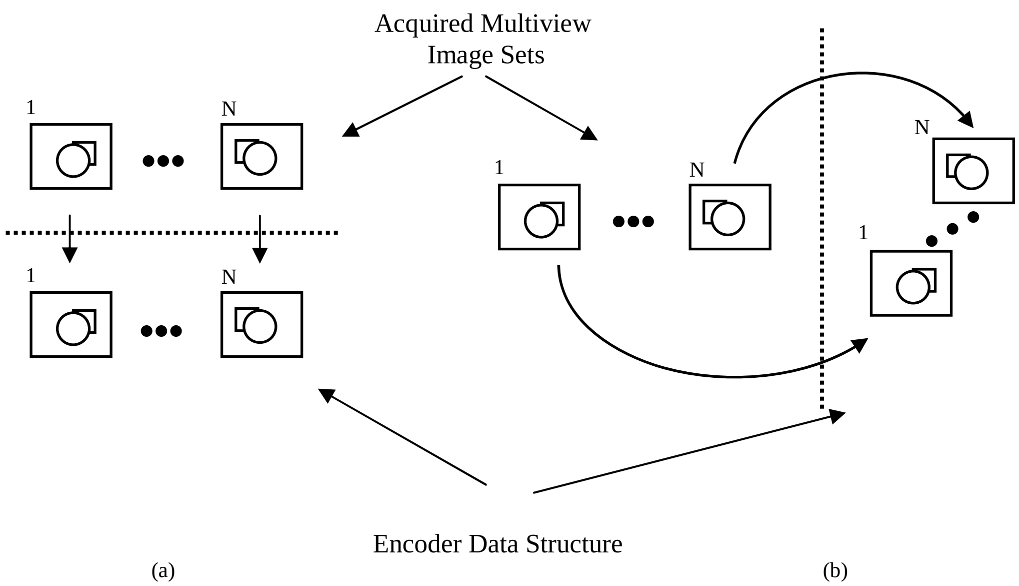

Fig. 2. Ordering of a multiview image set to form an image data volume (a) for the MDE and (b) for the 3D-DCT encoder

In Fig. 2 a multitude of N images is considered in each case. However in both cases a limited amount of consecutive images is considered as the fundamental data unit. In the case that the MDE encoder is used, a standard strategy of forming a group of pictures (GOP) is followed in order to increase the efficiency of the encoder. Equivalently when using a 3D-DCT approach a volume of pictures (VOP) is used as the fundamental data unit. The choice of the size of GOP and VOP was done upon the simple assumption that most of the existent multiview setups use 8 distinct views for each multiview set combining compression efficiency with low computational complexity.

4. Overview

The MDE class of encoders functions on the principles of the MPEG [5] encoding scheme. As in an MPEG encoder a search area is defined in order to find a best match between a frame to be encoded and a reference frame. However there are two basic differences due to the nature of the problem. The time series of the frames are substituted by a spatial sequence of frames and the search area can be restricted, based on the intrinsic characteristics of the acquisition device. A simplified block diagram of a typical MDE encoder is depicted in Fig. 3.

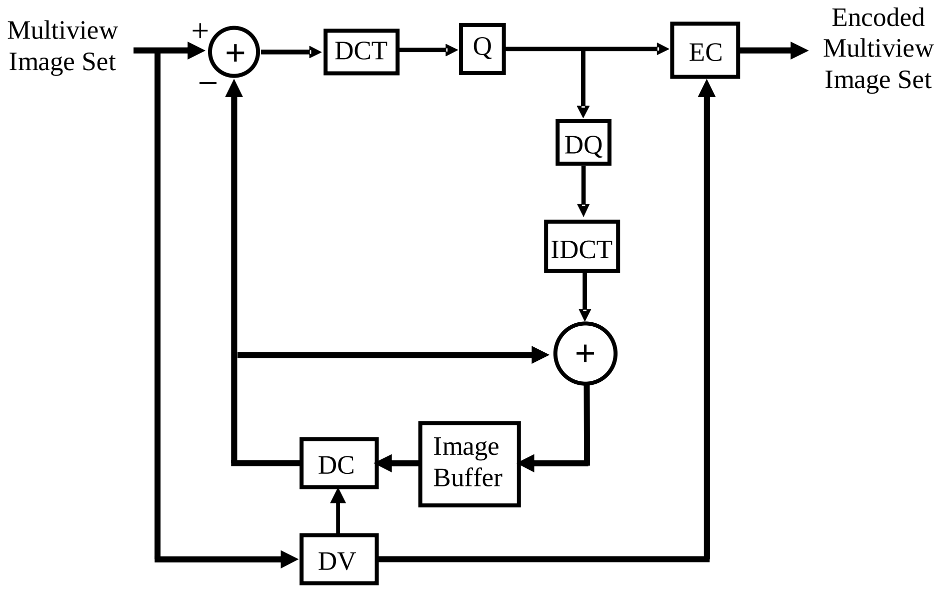

Fig. 3. Block diagram of an MDE encoder

In detail the encoder comprises a DCT and an Inverse DCT (IDCT) transform units, a quantizer (Q) and an inverse quantizer (DQ). In addition an Image buffer is used to store images that will be used in the disparity compensation (DC) process. Disparity compensation is performed by deriving a set of Disparity Vectors (DV) between the reference and the predicted –disparity compensated– image. Finally the image data along with the corresponding DV are encoded using a variable length entropy coder (EC)

In what follows we describe the 3D-DCT encoder as a case study of the class of 3D transform techniques. The developed method is based on an extension of the classic DCT compression scheme in three dimensions (3D-DCT). There are many techniques that use 3D-DCT schemes in order to perform motion estimation and compress video sequences [8] including multiview image sets [3] and other types of stereoscopic data [9].

The realized 3D-DCT scheme is applied on multiview image data sets that are accordingly transformed in one or more VOPs'. The VOP data are properly transformed and quantized using appropriate quantization volumes. On the final stage of the procedure the resultant coefficients are rearranged using a case introduced scan order technique and further compressed with the use of an entropy coder.

The encoder comprises of a 3D-DCT unit, the quantizer (Q) and the entropy coder (EC). An additional unit is added to this standard setup for determining the quantization volume values and the scan order of the 3D-DCT coefficients based on the standard deviation of the coefficients (SDU). The encoder layout is depicted in Fig. 4.

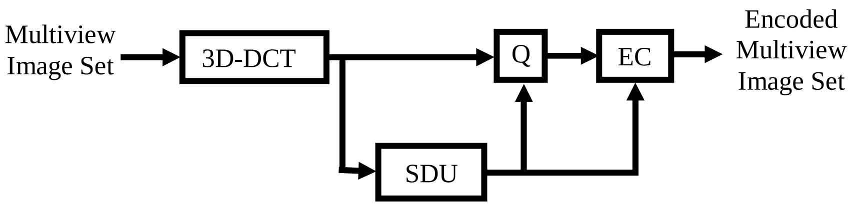

Fig. 4. Block diagram of an MDE encoder

A global standard deviation strategy is used to locate the dominant 3D-DCT coefficients directivity, determine the quantization volume values and the scan order of the quantized coefficients, aiming to augment the efficiency of the standard 3D-DCT scheme. In this way we practically incorporate into the algorithm global characteristics of the VOP, which is information on the disparity between consecutive images.

5. 3D-DCT Coefficient Quantization and Scan Order

Quantization is one of the most crucial factors in a coding procedure. A quantization unit generally returns the quantized coefficients values $F_{quantized}(u,v,w)$ by performing the operation defined by Eq. 14

$$F_{quantized}(u,v,w) = \left\lfloor \frac{F(u,v,w)}{Q(u,v,w)} \right\rfloor \tag{1}$$

where $F(u,v,w)$ is the 3D-DCT coefficient volume and $Q(u,v,w)$ is the quantizer value at position u,v,w. The main idea of the quantization procedure is to attenuate coefficients that are believed not to introduce significant error. In this manner the number of zero valued coefficients is increased leading to higher compression ratios.

The quantizer unit that is used in this work is a modified version of the spatio-temporal quantization volume presented in [8]. In detail the quantizer volume values are determined by Eq.2

$$Q(u,v,w) = qf \cdot (a \cdot u^p + b \cdot v^p + c \cdot w^p) \qquad (2)$$

where $Q(u,v,w)$ is the quantization value at position u,v,w. Other parameters that adjust the shape of this quantizer volume are qf which determines the quality level of the encoded data and p that provides an exponential increase in the quantization values. In addition a,b,c enhance this increase in the quantization values in specific directions. These three parameters are determined upon the values of the standard deviation in each coefficient position in order to maintain significant coefficients.

This is based on the fact that in general the variance of a DCT coefficient over an image is analogous to the energy content of that particular coefficient [10]. Extending this fact to the three dimensional case, the transform coefficients derived are subjected to the standard deviation unit (SDU) where the total standard deviation for each coefficient is calculated. Standard deviation is finally used to determine the quantization values and the scan order of the coefficients as previously discussed. A simplified approach for determining the dominant coefficients is realised but this is enough for elaborating the concept. Due to a priori knowledge of certain image characteristics we expect the strong coefficients to occupy certain spectral planes [8] in the 3D-DCT frequency domain.

In total there are three types of scan order of the transform coefficients used in many applications that implement a 3-DCT. The first type of scan that is used is a classic 3D Zig-Zag scan as defined in [11]. However this type of scan is used when the data have increased homogeneity and little or no translational movements are noted. In the case where translational and rotational movements are present two alternative scan orders are applied in order to provide longer runs of zero coefficients and increase the efficiency of the proposed algorithm.

The first of the scans is utilized when the standard deviation coefficients has large values along the horizontal axis of the quantized coefficients cube, as produced by Eq. 1. This case for a VOP comprised of 8 images is depicted in Fig. 5a. On the contrary, when the standard deviation has high values along the vertical direction of the cube the scan order depicted in Fig. 5b produces optimal results.

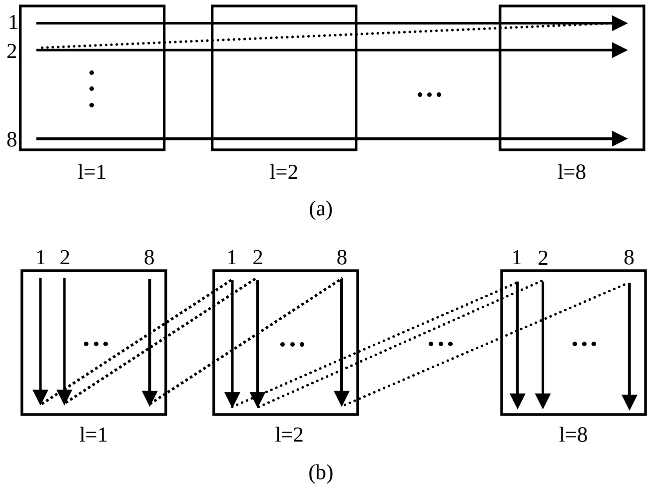

Fig. 5. (a) Horizontal and (b)Vertical scan.

6. Entropy Coding

The entropy coder (EC) further encodes the data by performing a run length coding of the ordered coefficients, followed by Huffman entropy coding. In the case of the MDE encoder the EC also encodes the set of disparity vectors, along with the coefficient values in order to create the encoded multiview image set using a standard MPEG approach. The EC scheme for the 3D-DCT architecture follows the strategy of the JPEG approach, where the DC coefficient is coded separately using a DPCM scheme. In addition the EC encodes the quantization parameters and scan order of the coefficients as they are depended on the statistical characteristics of each multiview image set.

7. Discussion

In this work we presented two of the dominant trends in multiview image compression. Both methods are based on the extensively used DCT technique in image and video compression. The basic parts of the MDE technique are derived directly from the MPEG standard for video coding. However an augmentation is made based on the fact that the window size used in the matching process is smaller than the one used in the temporal case. In this way smaller windows with known search directivity provide real time performance and efficient DV coding reducing the overall bit cost. This constraint also allows the use of exhaustive search algorithms for locating the best match without increasing overall complexity of the encoder. Moreover the MDE encoder can be easily ported in hardware modules increasing the robustness of the technique.

We also realized a 3D-DCT technique with adaptive quantization based on the statistics of the multiview image set in order to evaluate the efficiency of higher order transform techniques in multiview image coding. The technique is based on the spectral characteristics of the transform coefficients in the 3D-DCT domain, in order to predict the quantizer shape parameters and scan order of the resulting coefficients. Due to the use of standard DCT modules and the separability property of this transform this technique could be proven a valuable tool for developing robust compression schemes for multiview image sets. It should be noted that the overall complexity of a 3D-DCT scheme is not prohibitive for providing real time solutions and there are many hardware implementations to accomplish this.

Simulation results prove that in low bit rates the MDE coder performs better than the 3D-DCT technique. However in high bit rates the 3D-DCT technique outperforms the MDE coder. It is to be noted that in general the 3D-DCT coder is able to provide a homogenous result over a set of multiview images while the predicted frames in the MDE coder usually introduced inhomogeneities throughout the set. These inhomogeneities deteriorated the final 3D representation as it was verified by subjective evaluation of the results. Nevertheless in low bit rates quality deterioration was not easily noticeable and subjective evaluation of the results agrees with the objective measures. Thus we conclude that the PSNR value truly depicts the efficiency of the MDE over 3D-DCT for low resolution applications.

In conclusion, the 3D-DCT coder has been proven better for high quality applications while not introducing increased complexity in regard to the MDE coder. On the contrary the MDE encoder can be used for low quality multiview image sets providing high compression of the data set.

Future work includes certain improvements in both techniques and the assessment of the results over a variety of multiview image sets and encoder architectures using wavelet transforms.

8. References

[1] J. N. Ellinas, M. S. Sangriotis, "Stereo Image Compression Using Wavelet Coefficients Morphology", *Image & Vision Computing,* Elsevier Science, vol. 22, Iss. 4, Apr. 2004, pp. 281-290.

[2] C. Kamisetty and C. V. Jawahar, "Multiview image compression using algebraic constraints," in *Proc.TENCON 2003.Conf.on Conv. Tech. for Asia-Pacific Reg.*, Bangalore, Oct. 2003, vol. 3, pp. 927-931.

[3] D. Shah and N. A. Dodgson, "Issues in multi-view autostereoscopic image compression," in *Proc. IS&T/SPIE Symp. on Stereo. Disp. and Appl. XII,* San Jose, Jan. 2001, vol. 4297, pp. 307-316.

[4] T. Palfner and E. Müller, "Coding of full-parallax multi-view images," in *Proc. IS&T/SPIE Symp. on Stereo. Disp. and Appl. XVI,* San Jose, Jan. 2005, vol. 5664, pp. 460-471.

[5] K. R. Rao and P. C. Yip, The Transform and Data Compression Handbook, Boca Raton: CRC Press LLC, ch.4, 2001.

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

[7] A. Schmidt and A. Grasnick, "Multiviewpoint autostereoscopic displays from 4D-Vision GmbH," in *Proc. IS&T/SPIE Symp. on Stereo. Disp. and Vitual Reality Syst. IX,* San Jose, Jan. 2002, vol. 4660, pp. 212-221.

[8] N. Božinović and J. Konrad, "Motion analysis in 3D DCT domain and its application to video coding," *Sig. Process.: Image Communication,* Elsevier Science, vol. 20, Iss. 6, July 2005, pp. 510-528.

[9] R. Zaharia, A. Aggoun, M. McCormick, "Adaptive 3D-DCT compression algorithm for continuous parallax 3D integral imaging," S*ignal Processing: Image communication*, Elsevier Science, 2002, vol. 17, Iss. 3,pp. 231-242

[10] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. on Image Process.*, vol. 9, No. 10, pp. 1661–1666, Oct. 2000.

[11] B. Yeo and B. Liu, "Volume rendering of DCT-based compressed 3D-scalar data," *IEEE Trans. on Vis, and Comp, Graph.*, vol. 1, No. 1, pp. 29–43, Mar. 1995.