Contribution to Improvement of Compression Algorithms for Stereoscopic Image and Video

John N. Ellinas[1]
(iellinas@di.uoa.gr)

1 Abstract

The demand for 3D imaging has been increasing because stereopsis provides realism to 2D images. The price for this added realism is the doubling of data and thus the limited bandwidth of existing channels and the shortage of the storing media become the main bottleneck. To achieve more effective coding for a pair of stereo images or stereo video, various encoding schemes have been proposed. In particular, three novel encoding methods have been developed for stereoscopic images and two for stereoscopic video. All of them employ the Discrete Wavelet Transfrom (DWT) and a morphological compression algorithm that applies on the transform coefficients. The proposed algorithms were experimentally evaluated on various synthetic and real images and videos, where they have shown beneficial performance against other state-of the-art coders.

2 Introduction

Stereoscopic vision is based on the projection of an object on two slightly displaced image planes and has an extensive range of applications, such as 3-D television, 3-D video applications, robot vision, virtual machines, medical surgery and so on. Two pictures of the same scene taken from two nearby points form a stereo pair and contain sufficient information for rendering the captured scene depth. Typically the transmission or the storage of a 3-D system requires twice as much data volume as a 2-D system. Nevertheless, in a stereoscopic system a more efficient coding scheme may be developed if the inter-image redundancy is exploited.

In general, there are two methods of implementing either motion or disparity estimation in monoscopic or stereoscopic compression applications. The first method, based on intensity processing, handles this estimation by the block-matching algorithm (BMA) [1]. The target frame is divided into blocks of fixed size (FBS) and is matched under a matching criterion, which may be the Mean Square Error (MSE) or the Mean Absolute Difference (MAD), which minimizes a cost function. Several compression algorithms have been developed that use block matching or alternative implementations, including hierarchical disparity estimation [2], multiresolution block matching [3], block matching with geometric transform [4], DWT with morphological coefficient-to-coefficient stereo coding [5] and so on. The second method for either motion or disparity estimation,

[1] M. S. Sangriotis

based on object handling, firstly defines or derives the features of the participating objects in the processed frame and then estimates the temporal or disparity field between corresponding frames [6]. Apart from the above-mentioned methods, several others have been proposed that try to improve the performance or combine their characteristics such as the segmentation based coding [7], stereo image projection [8], post-compensation residual coding [9], overlapped block disparity compensation [10], a hybrid scheme between block and object based technique [11] and so on. Of the two methods described above, the former is the most commonly used because of its simplicity. However, it fails to code efficiently homogeneous regions. The frame segmentation into blocks of variable sizes that incorporate homogeneous motion and/or disparity characteristics may improve the coding efficiency by allocating fewer bits to larger regions. Such a method is presented in [12] and employs rectangular segmentation of the disparity field in a multiresolution framework, in order to improve coding efficiency of the Right image in a stereo pair.

Several schemes have been proposed for the compression of stereoscopic image sequences. In [13] the Left image stream is MPEG-type encoded and the Right frames are estimated from their corresponding Left frames, using hierarchical multiresolution disparity estimation. In [6] and [14] the proposed 3-D motion estimation methods are integrated in a stereoscopic inter-frame coding scheme. Another study proposes the successive exploitation of disparity and motion redundancy [7]. In [15] an object-based coding algorithm is proposed relying on modelling of objects with 3-D wire-frames. Another scheme proposes the MPEG-type encoding of the Left stream and the joint motion and disparity estimation of the Right frames with blocks of fixed size [16].

A lot of efforts have been made for developing efficient compression algorithms for stereo images and image sequences. The embedment of the multiview profile (MVP) to the MPEG-2 coding standard is towards that direction [17]. Also, the recent emergence of the highly efficient advanced video codec H.264/AVC [18] has provoked the expectation of an equally efficient stereo codec. In parallel, JPEG-2000 and MPEG-4 shows a trend to replace DCT-based techniques with DWT-based schemes. Following these trends, in the present work, three methods of coding sterescopic images and two methods of coding stereoscopic image sequences have been emerged. Their operation is based on the DWT and the morphological compression algorithm, MRWD (Morphological Representation of Wavelet Data), [19]. The aim of the proposed coders is to tackle the excessive bandwidth requirements of the typical stereoscopic applications, as 3DTV or 3D video entertainment, and to keep a fair trade-off between quality and bandwidth. The use of a robust and high efficient wavelet-based morphological coder serves the need of bandwidth reduction in conjunction with more effective motion-disparity compensation methods. This coder presents excellent compression efficiency, low complexity, fast execution and embedded bitstreams. The proposed motion-disparity compensation schemes exploit the high degree of correlation between the same scene content, achieving: more effective coding, allocation of fewer bits to larger homogeneous regions and less annoying artifacts. Also, subband coding methods exploit the non-uniform distribution of energy across the different frequency bands. Since the entire frame is filtered and subsampled to obtain the subbands, these methods do not suffer from blocking artifacts that are common in block-based transform coding methods. The inherent advantages of the wavelet transform are the creation of almost decorrelated coefficients, energy compaction and variable resolution. The proposed algorithms for stereoscopic image coding are the following:

- **Stereo image compression using wavelet coefficients morphology, [5]:** The proposed coder decomposes both images of the stereo pair by DWT and then employs MRWD encoder. Firstly, the morphological coder encodes the reference image. Then, the morphological coder starts encoding the target image and during this process it produces the residual target image, which is the difference of the subband coefficients of the original target image with the predicted ones after a pixel-to-pixel matching with the reference image. Alternatively, the coder completes the formation of the clusters in the subbands, which are groups of significant coefficients, and then produces the residual target image after a cluster-to-cluster matching. Finally, the reference image, the residual target image and the disparity vectors are entropy coded and transmitted. The encoding algorithm exploits the statistical properties of the wavelet coefficients to form clusters and uses a morphological operator to efficiently encode the significant ones together with their positions. The disparity field in this paper, which consists of the disparity vector (DV) and the disparity compensated difference (DCD), is estimated simultaneously with the morphological operations on the subbands in order to form a single framework. The pixel-to-pixel matching allows the estimation of a dense disparity field when a better quality image is required, whereas the cluster-to- cluster matching allows the estimation of a sparse disparity field when low complexity is required.
- **Stereo image coder based on quad-tree analysis and morphological representation of wavelet coefficients, [20]:** The novelty of the proposed work is the combination of a robust still image coder with a disparity compensation procedure in the spatial domain based on variable size block-matching. The proposed coders are based on the DWT decomposition in combination with the MRWD encoding unit, which is a robust algorithm yielding very good lossy compression. The proposed disparity compensation is based on the segmentation of one view given the other and achieves a coding representation that is commensurate with the local disparity detail. Typical stereoscopic images contain areas of almost constant disparity such as background or objects of large size. The disparity estimation schemes based on blocks of fixed size divide these areas into small blocks creating more disparity vectors than those are actually needed. To overcome this drawback, a disparity estimation based on a quad-tree segmentation of the Right image is proposed. The proposed quad-tree decomposition (QTD) is performed employing a rate-distortion splitting criterion such that the total

cost aftrer disparity compensation, i.e. distortion of the residual image and bit-rate of the disparity vectors, to be minimum.

- **Stereo Image Coder Based on MRF Analysis for Disparity Estimation and Morphological Encoding, [21]:** A robust still image encoder and a disparity compensation process, which is based on the Markov Random Field (MRF) model are the novelties of the proposed coder. MRF model takes into account the contextual constraints by considering that the disparity field is smooth except near object boundaries. The probabilistic aspect of the MRF analysis is converted to energy distribution through its equivalence to Gibbs distribution (GRF). The usual statistical criterion for optimality is the maximum a posteriori probability (MAP) that provides the MAP-MRF framework. According to this model, occlusion field is initially separated into three regions by setting two threshold levels. The blocks of the intermediate region, which is called uncertain, are finally characterized as occluded or non-occluded. This reduces the number of regions needed for the MAP search procedure, which is normally implemented in the entire occlusion field, making the algorithm simpler and faster. Also, mean absolute error (MAE) is selected instead of mean square error (MSE), in order to render our algorithm less sensitive to noise.

The proposed algorithms for stereoscopic video coding are the following:

- **Stereo Video Coding Based on Interpolated Motion and Disparity Estimation, [16]:** The proposed stereo image sequence compression scheme, which is called IMDE (Interpolated Motion and Disparity Estimation) and belongs to intensity methods, uses the MPEG standard for coding the left or main image stream and an interpolative scheme, which is applied on the corresponding motion and disparity frames, for predicting and coding the P and B type of frames of the right or auxiliary image stream. The estimated residual frames of both channels are coded using a DWT followed by the MRWD encoder. Furthermore, the performance of the proposed interpolative scheme is investigated with respect to the joint estimation of motion and disparity vectors as well as the optimized choice of the weighting factors of the predicted frames that participate in this scheme.
- **Stereo video coding based on quad-tree decomposition of B-P frames by motion and disparity interpolation, [22]:** This new stereo image sequence compression scheme is called Enhanced Interpolated Motion and Disparity Estimation (EIMDE) and belongs to intensity processing methods. The Left image sequence is MPEG-like encoded, whereas P and B frames of the Right image sequence are predicted by a joint motion-disparity interpolative scheme and segmented into variable size macroblocks. The processed frame is initially segmented into blocks of homogeneous intensity by its quad-tree decomposition with an intensity difference threshold criterion. These blocks may probably belong to the same object or the background of an image and may present homogeneous motion or disparity characteristics. Then, quad-tree decomposition follows with a simplified Rate-Distortion

criterion that permits splitting if there is a Rate-Distortion benefit. Furthermore, the performance of the proposed interpolative scheme is enhanced by using a suitable search method for the estimation of the best joint motion and disparity vectors and by optimizing the weighting factors of the participating frames.

3 Overview of the proposed methods

This section quotes an overview of the developed algorithms for coding stereoscopic images or stereoscopic sequences.

3.1 Stereo Image Compression Using Wavelet Coefficients Morphology

The proposed scheme estimates the disparity vector field and the disparity compensated difference with a pixel-to-pixel matching method, which was embedded into the morphological coder. The morphological coder was constructed to produce four partitions of coefficients in every subband, in order to obtain the best performance. In a second experiment, the disparity vector and the disparity compensated difference were estimated by the cluster-to-cluster matching method. This approach operates after the cluster formation in the subbands. The produced clusters of the target image are corresponded to their relatives of the reference image by shifting them, so that the minimum MAE is obtained. Consequently the DV field is estimated for the whole cluster and the DCD field is estimated by subtracting the corresponding coefficients between the two clusters. This method degrades the objective quality but is less computational expensive. Hence, the two examined methods may be used together in order to transmit some clusters with better quality than others. The proposed method shows a very good performance in the whole examined range. The pixel-to-pixel disparity estimation, although it is simple and produces dense disparity fields, it provides comparable entropy rates with respect to the popular block-matching methods.

3.2 Stereo image coder based on quad-tree analysis and morphological representation of wavelet coefficients

In this method, the Right image is segmented into blocks of variable size according to a quad-tree splitting procedure. Right image is initially segmented into blocks of homogeneous intensity using quad-tree decomposition with an intensity difference threshold, resulting into blocks with minimum permissible size of 8×8. These blocks may probably belong to the same object or the background and may present homogeneous disparity characteristics. Then, second quad-tree decomposition with a simplified rate-distortion criterion follows, permitting the splitting of an already existing block to four children blocks only if there is a rate-distortion benefit from this splitting. Figure 1 shows the segmentation of the target image with an intensity splitting criterion of 0.6.

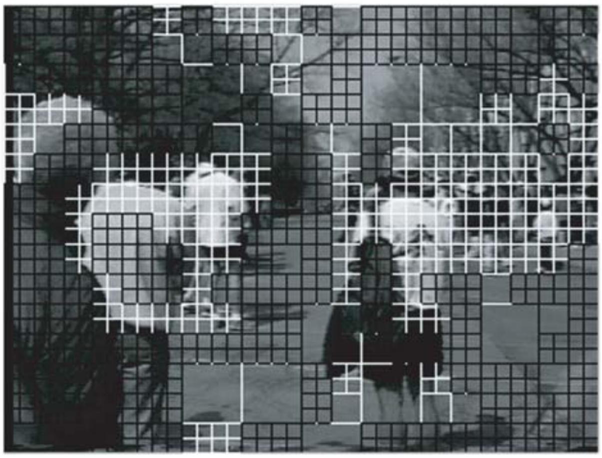

Fig. 1. QTD decomposition of a B-frame with intensity splitting criterion of 0.6.

3.3 Stereo Image Coder Based on MRF Analysis for Disparity Estimation and Morphological Encoding

The disparity field of a stereo image pair is an MRF/GRF model consisting of the disparity, D and occlusion, O , fields. The problem is to determine the disparity and occlusion fields from the observations, which are the pair of images. The configurations d and o of the disparity and occlusion fields may be estimated by:

$$(\hat{d}, \hat{o}) = \arg \min_{(d,o) \in S} [U(S^R \mid S^L, d, o) + U(d \mid o) + U(o)] \tag{1}$$

where S^L and S^R represent the reference and target images respectively. The first term of equation (1) represents the likelihood energy of the Right or target image, given the Left or reference image and the fields D and O. This term is called similarity constraint because it concerns the similarity between the two images of the stereo pair or the residual energy. The second term is the energy of the disparity field, given the occlusion field O and is called smoothness prior constraint. It concerns the smooth variation of the disparity vectors. The third term is the energy of the occlusion filed and is called occlusion constraint. It penalizes every discontinuity or occluded block. Therefore, the minimization of the total sum of these three terms will effect the coding efficiency of the residual image and the resulting disparity vectors. The initial occlusion field is formed by employing a double threshold procedure and is is separated into three regions: the non-occluded region, the occluded region and the uncertain region, where the blocks are subjected into a MAP search, in order to enrol them as occluded or non-occluded.

3.4 Stereo Video Coding Based on Interpolated Motion and Disparity Estimation

Figure 2 illustrates the proposed stereo video coding method, with the following characteristics:

- The main or reference channel is independently coded employing an MPEG-like encoder.
- I-frames of the auxiliary channel are predicted from the corresponding frames of the reference channel and the relulting residual frames are encoded together with the disparity vector field.
- P-frames of the auxiliary channel are predicted by interpolating the previous temporal reference frame of the same channel (as in a monoscopic video) and the corresponding P-frame of the reference channel.
- B-frames of the auxiliary channel are predicted by interpolating the previous temporal reference frame of the same channel, the next temporal reference frame of the same channel and the corresponding P-frame of the reference channel.

The prediction of the motion and disparity fields are performed in an iterative procedure so that the best values of the vector fields to be attained. The participation of the frames in the interpolating scheme is done with weighting constants that are estimated so that the residual frame to have minimum energy.

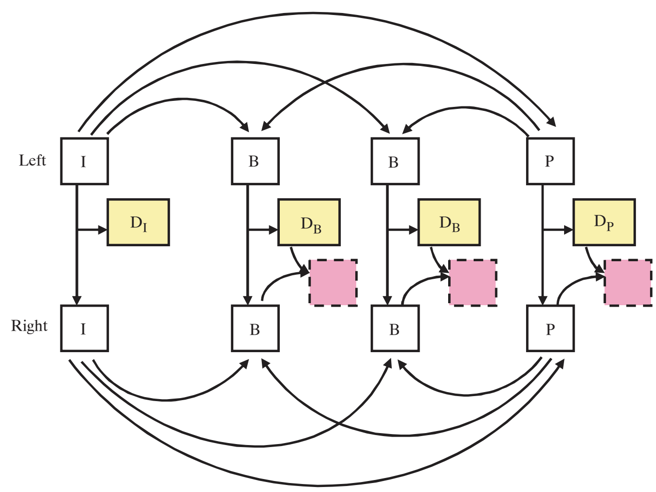

Fig. 2. *IMDE* method for stereo video encoding.

3.5 Stereo video coding based on quad-tree decomposition of B-P frames by motion and disparity interpolation

In a stereoscopic video, the predicted Right B-frames are estimated by interpolating the motion predicted frames from the reconstructed Right I or P frames and the disparity predicted frame from the corresponding reconstructed Left B-frame. Figure 3 shows the interpolation that must be applied to every macroblock of the frame.

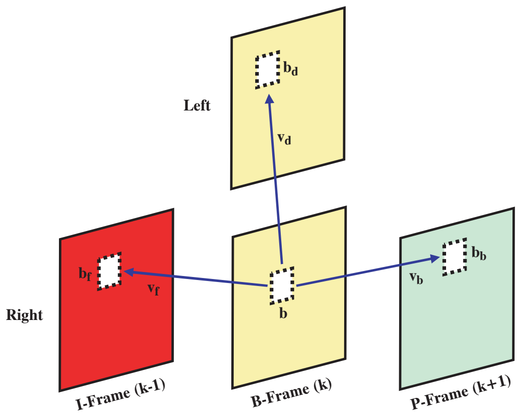

Fig. 3. Interpolative scheme for *IMDE* method.

The aim is to find the vectors that provide the best prediction of the Right frame. The independent search for the best vectors is a sub-optimal procedure. The proposed approach is to successively update each vector keeping the other two constant for a lower residual energy. It is apparent that the energy of the residual B-frame depends on the proper selection of the weighting factors between motion and disparity. For a given bit-rate, the reduction of this energy provides higher PSNR for the reproduced frame, as the resulting distortion is decreased. Basically, the relation between motion and disparity weighting factors should be adjusted for every macroblock and their choice must minimize the energy of each residual macroblock through the previously described vector optimization. Instead of estimating the weighting factors for every macroblock, which is time consuming and bit-rate expensive, a sub-optimal scheme is proposed. The motion weighting factors are initially considered to play an equal role to the disparity weighting factor. The proposed algorithm, which involves quad-tree decomposition of a P or B frame with a Rate-Distortion splitting criterion, employs the previously described interpolative scheme with the initial values of the weighting factors. These values are adjusted so that the total energy of the resulting residual frame is minimized.

4 Conclusions

In this thesis, five algorithms for compression of stereoscopic images and steroscopic video were developed and evaluated. The first one performs the disparity compensation process in the transform domain of a DWT together with the progress of a morphological compression algorithm. The second one performs the disparity compensation process by segmenting the target image into blocks of variable size. The segmentation is perfromed by a quad-tree decomposition with a two-fold splitting criteria (intensity and rate-distortion). The third one employs MRF model in order to achieve smoother disparity vector field and to manipulate the occlusion field more effectively. The fourth method encodes the auxiliary channel of a stereoscopic video by interpolating the motion and disparity field, optimizing the vectors and the weighting constants. The last method is the same with the previous one but enhances its performance by segmenting the B or P-frames into blocks of variable size.

The experimental evaluation of the proposed algorithms shows beneficial performance with respect to other state-of the-art coders. Their performance inherits the advantages of the wavelet transform, the boosted performance of the morphological compression algorithm and the effective way that disparity compensation is performed, either in the transfrom domain or with blocks of variable size or employing the MRF model.

5 References

1. M. G. Perkins, *Data compression of stereopairs*, IEEE Trans. Communications, vol. 40, pp. 684-696, 1992.
2. S. Sethuraman, A. G. Jordan, M. W. Siegel, *Multiresolution based hierarchical disparity estimation for stereo image pair compression*, Proceedings of the Symposium on Application of subbands and wavelets, 1994.
3. D. Tzovaras, M. G. Strintzis, H. Sahinoglou, *Evaluation of multiresolution block matching techniques for motion and disparity estimation*, Signal Processing: Image Communications, vol. 6, pp. 59-67, 1994.
4. M. Ghanbari, et al., *Motion compensation for very low bit-rate video*, Signal Processing: Image Communications, vol. 7, pp. 567-580,1995.
5. J. N. Ellinas and M. S. Sangriotis, *Stereo image compression using wavelet coefficients morphology*, Image and Vision Computing, vol. 22, no. 4, pp. 281-290, 2004.
6. D. Tzovaras, N. Grammalidis and M. G. Strintzis, *Object based coding of stereo image sequences using joint 3D motion/disparity compensation*, IEEE Trans. CSVT, vol. 7, no. 2, pp. 312-327, 1997.
7. S. Sethuraman, M. W. Siegel and A. G. Jordan, *Segmentation based coding of stereoscopic image sequences*, Proc. SPIE EI-Digital Video Compression: Algorithms and Technologies, vol. 2668, pp. 420-429, 1996.
8. H. Aydinoglu and M. H. Hayes, *Stereo image coding: A projection approach*, IEEE Trans. IP, pp. 506-516, 1998.

9. M. S. Moellenhoff and M. W. Maier, *Transform coding of stereo image residuals*, IEEE Trans. IP, vol. 7, no. 6, pp. 804-812, 1998.
10. O. Woo, A. Ortega, *Overlapped block disparity compensation with adaptive windows for stereo image coding*, IEEE Trans. CSVT, vol. 10, pp. 194-200, 2000.
11. J. Jiang and E. A. Edirisinghe, *A hybrid scheme for low bit-rate coding of stereo images*, IEEE Trans. IP, vol. 11, no. 2, pp. 123-134, 2002.
12. S. Sethuraman, M. W. Siegel and A. G. Jordan, *A multiresolution region based segmentation scheme for stereoscopic image sequence compression*, Proc. SPIE EI-Digital Video Compression: Algorithms and Technologies, vol. 2419, pp. 265-274, 1995.
13. S. Sethuraman, M. W. Siegel and A. G. Jordan, *A multiresolution framework for stereoscopic image sequence compression*, Proc. IEEE ICIP, vol. 2, pp. 361-365, 1994.
14. N. Grammalidis, S. Malassiotis, D. Tzovaras, M. G. Strintzis, *Stereo image sequence coding based on three-dimensional motion estimation and compensation*, Signal Processing: Image Communication, vol. 7, pp. 129-145, 1995.
15. S. Malassiotis, M. G. Strintzis, *Coding of video-conference stereo image sequences using 3-D models*, Signal Processing: Image Communication, vol. 9, pp. 125-135, 1997.
16. J. N. Ellinas and M. S. Sangriotis, *Stereo video coding based on interpolated motion and disparity estimation*, EURASIP 3rd International Conference on Image and Signal Processing and Analysis, pp. 301-306, 2003.
17. MPEG-2 Video Subgroup, *Proposed Draft Amendment 3 Multi-view Profile*, ISO/IEC/JTC1/SC19/WG11 Doc. N1088, 1995.
18. T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, *Overview of the H.264/AVC video coding standard*, IEEE Trans. CSVT, vol. 13, no. 7, pp. 560-576, 2003.
19. S. D. Servetto, K. Ramchandran and M. T. Orchard, *Image coding based on a morphological representation of wavelet data*, IEEE Trans. IP, vol. 8, no. 9, pp. 1161-1173, 1999.
20. J. N. Ellinas and M. S. Sangriotis, *A novel stereo image coder based on quad-tree analysis and morphological representation of wavelet coefficients*, 1st International Conference on Information Technology and Quality, Athens, 2004.
21. J. N. Ellinas and M. S. Sangriotis, *Stereo image coder based on MRF analysis for disaprity estimation and morphological encoding*, 2^{nd} International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Sept. 2004.
22. J. N. Ellinas and M. S. Sangriotis, *Stereo video coding based on quad-tree decomposition of B-P frames by motion and disparity interpolation*, IEE Proceedings on Vision, Image & Signal Processing, accepted on 8 May 2005.