Reversible Watermarking on Stereo Image Sequences

John N. Ellinas

Abstract—In this paper, a new reversible watermarking method is presented that reduces the size of a stereoscopic image sequence while keeping its content visible. The proposed technique embeds the residuals of the right frames to the corresponding frames of the left sequence, halving the total capacity. The residual frames may result in after a disparity compensated procedure between the two video streams or by a joint motion and disparity compensation. The residuals are usually lossy compressed before embedding because of the limited embedding capacity of the left frames. The watermarked frames are visible at a high quality and at any instant the stereoscopic video may be recovered by an inverse process. In fact, the left frames may be exactly recovered whereas the right ones are slightly distorted as the residuals are not embedded intact. The employed embedding method reorders the left frame into an array of consecutive pixel pairs and embeds a number of bits according to their intensity difference. In this way, it hides a number of bits in intensity smooth areas and most of the data in textured areas where resulting distortions are less visible. The experimental evaluation demonstrates that the proposed scheme is quite effective.

Keywords—Stereoscopic video, Reversible watermarking, Disparity compensation, Joint compensation

I. Introduction

STEREOSCOPIC vision has an extensive range of applications, such as 3-D television, 3-D video applications, robot vision, virtual machines, medical surgery and so on. Typically the transmission or the storage of a stereo image sequence requires twice as much data volume as a monoscopic video system. Coding, in a stereoscopic video system, is based on the intra-frame, inter-frame and inter-sequence redundancy [1], [2].

However, once the stereo sequence has been compressed, its content is no more visible. Someone should first decompress the stereo image sequence in order to visualize its content, which is time consuming and not desirable in some cases, as in a stereo video data library. Therefore, there are applications where the capacity of a stereo video must be reduced without losing the ability to identify its content.

This paper focuses on the problem of reducing the size of a stereoscopic image sequence while keeping its content visible. Our proposal may halve the capacity of a stereo video using reversible watermarking. Given a stereoscopic image sequence, the proposed method embeds into the left or reference frames the information needed to recover the right or target frames employing reversible watermarking techniques. Therefore, apart from having one image sequence for storage or transmission, its content is still visible as no compression is used. At any instant, the extraction of the embedded data may provide the initial target frames and the stereo video to be recovered. In reversible watermarking, the reference frames are exactly recovered after the extraction of the embedded data at the detector. The stereoscopic frames may be exactly recovered if the data hiding capacity is high enough and the embedding algorithm is efficient to embed the entire required information needed to reconstruct the target frames. Usually, the embedding bit-rate of the reference frames is a limiting factor for the exact reconstruction of the target frames. In that case, the embedded information is subjected to lossy compression and therefore the stereoscopic video at the detector is distorted.

Stereo image embedding by reversible watermarking has been proposed in [3], where the difference of the two images is embedded into the reference image employing an integer transform. However, the difference of the two images contains a lot of energy and therefore more bits are designated for embedding. As the embedding capacity of the reference image is limited, the compression of the difference is unavoidable and is performed with a low quality factor. Our proposal, exploiting the intra-sequence as well as the inter-sequence redundancy, adopts the use of the motion and disparity compensated difference which is widely employed in stereoscopic video compression [1], [2]. Two corresponding frames of the stereoscopic image sequence are subjected to motion and disparity compensation resulting in a residual frame and a disparity vector field. The energy of the residual frame is much lower than difference and consequently it is compressed with a higher quality factor providing better target frame recovery and consequently better stereoscopic sense.

In our proposal, the employed embedding algorithm is the Pixel Value Differencing (PVD), where the reference or cover frames are partitioned into blocks of two consecutive non-overlapped pixels in a scan order [4]. The number of bits embedded in each pair of pixels depends on their gray value difference implying that a small number of bits are placed in smooth intensity areas whereas the majority are placed in textured areas.

This paper is organized as follows: In section 2, an overview of reversible watermarking is presented. Section 3 introduces the proposed scheme for the reversible watermarking of a stereo image sequence, while in section 4 the performance of the proposed scheme is examined. Finally, section 5 concludes this work.

II. Reversible Watermarking Overview

Reversible watermarking not only embeds the watermark into an image but also can recover the original image from

Manuscript received November 28, 2008.

J. N. Ellinas is with the Department of Electronic Computer Systems, Technological Educational Institute of Piraeus, 250 P. Ralli & Thivon, 12244 Egaleo,Greece (phone:+30-210-5381454; e-mail: jellin@teipir.gr).

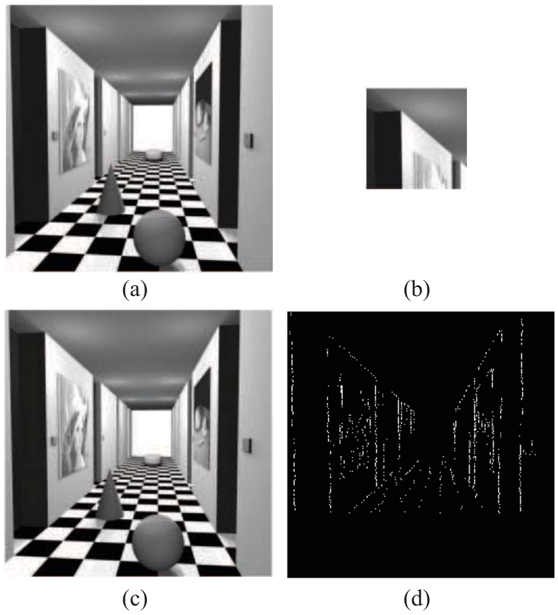

Fig. 1. PVD embedding algorithm for "Room" stereo pair. (a) Reference image; (b) Segment of the target image; (c) Reference image embedded with the information of the segment; (d) Difference between initial and embedded frame.

the received copy. Apart from robustness and imperceptibility, which are the desired features of a watermarking algorithm, reversible watermarking should also provide data embedding efficiency and data retrieving capability. There are several reversible watermarking schemes applying data compression in order to reduce the size of the embedded data [5], [6], [7], [8]. Some other methods are using difference expansion to embed information [9], [10], [11], [12] or histogram bin shifting [13], [14], [15]. The robustness, imperceptibility and the embedding capacity are the challenges of reversible watermarking algorithms, which all the above mentioned research works are dealing with. The first type of algorithms is more elaborate and fragile to distortions, as any loss of the compressed data may cause severe degradation of the embedded data. Most schemes of the second type are pixel or block-based, have simple implementation but are also weak as far robustness is concerned. The third group of algorithms is block-based, more resistible to some attacks but with lower embedding capacity.

The algorithm employed in the present work is the PVD, which combines data embedding efficiency and simplicity of implementation at both the encoder and detector sides. Belonging to the second group of algorithms, it presents weak robustness but this is not a significant requirement for applications like storing and retrieving stereoscopic image sequences in a data base. PVD hides data in gray-scale images imperceptibly based on the differences of the gray values in the two-pixel blocks of the cover image. A small difference value indicates that the pixels belong to a smooth area whereas a large difference indicates that they lie near edges. According to the human visual system, where distortions near edges are less noticeable than those in smooth areas, the pixels near edges can hide more secret bits than those in smooth areas. Thus, different amounts of data can be embedded in different blocks according to the degree of smoothness in the cover image. Fig. 1 illustrates the effect of the PVD embedding algorithm on the two images of the stereo pair "Room". A segment of the right image has been embedded into the reference image without having a noticeable effect. The difference of the initial image with the embedded one indicates where the information has been mostly embedded.

In reversible watermarking, given a stereoscopic pair of frames, the left or reference is considered as cover frame and the information needed to recover the right or target frame should be prepared in such a way so that to fulfill the prerequisites of the method, like embedding capacity and imperceptibility. Let f_1 and f_2 be the cover frame and the one that is to be embedded respectively. The residual of the two frames is

$$r = f_1 - f_2 \tag{1}$$

Once the residual is recovered, the two frames are readily available at the detector side. If the residual is embedded into the cover frame without any loss, both frames may be exactly recovered at the detector. If this is not feasible, the residual is subjected to lossy compression. In that case, the cover frame is exactly recovered while the other frame is recovered with distortion. This type of residual still contains a lot of energy, since it does not take into account the inter-image redundancy. The residual that turns up a disparity compensated procedure is a more rational approach. Fig. 2 shows the difference between the two images of the "Room" stereo pair, the residual after disparity compensation and the resulting disparity vectors. It is obvious that the energy of the second case is much lower and therefore less effort is needed to embed it into the cover image. Obviously, the disparity vectors must be embedded without any loss, since they ensure the accurate reconstruction of the target image at the detector. Smoothness of the disparity field may result in a small contribution to the increment of the embedding rate but at the cost of computing complexity.

III. Proposed Reversible Watermarking Scheme

The encoder of the proposed scheme has two modes of operation that resemble to the stereoscopic video compression modes [1], [2]. Both modes treat the stereo video as two image sequences with a GOP in the form IBBPBB. In the first mode, which is called *compatible*, the residuals result in after the disparity compensation between the corresponding left and right frames of the stereoscopic image sequence as Fig. 3 illustrates. The residual frame r is estimated according to Eq. (1)

$$r = \widetilde{f}_L - f_R \tag{2}$$

where $\widetilde{f}_L$ is the disparity compensated reference frame and f_R is the target frame.

This mode of operation is straight forward but it does not take into account the inter-frame redundancy of the right sequence. The second mode of operation, which is called *joint*, is illustrated in Fig. 4. In this case, the residuals result in after the motion and disparity compensation between the

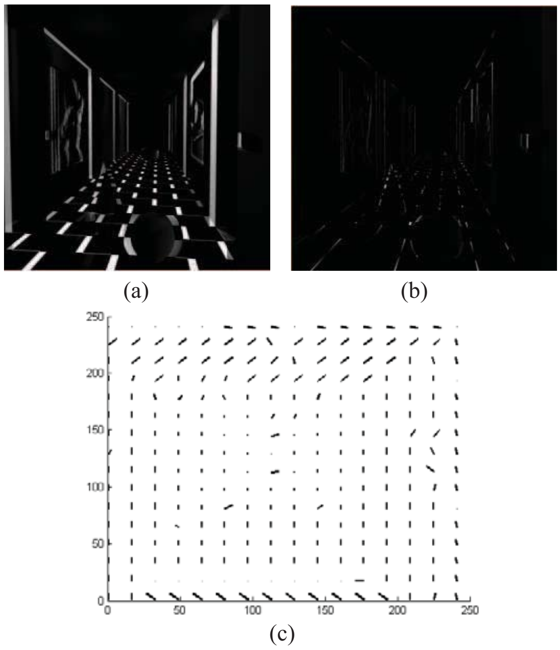

Fig. 2. Residuals of the stereo pair "Room". (a) Difference; (b) Disparity Compensated Difference; (c) Disparity Vectors.

corresponding left and right frames of the stereoscopic video. The I-frames are processed in the same way with the first mode of operation because they are reference frames in the video stream. The residuals of the P-frames result in after interpolating the motion forward prediction of the target frame with the disparity prediction of the corresponding left frame. Using Eq. (2), the residuals resulting from this consideration are expressed as

$$r_P = (w_p \times \widetilde{f}_I^R + w_d \times \widetilde{f}_P^L) - f_P^R \qquad (3)$$

where f_P^R is the current P-frame, f_I^R is the immediate preceding I-frame of the right sequence, f_P^L is the corresponding left P-frame and w_p, w_d are the weighting factors for the forward motion and disparity prediction correspondingly. The weighting factors are considered to have an equal role for this prediction and they are set to $w_p = 0.5$ and $w_d = 0.5$.

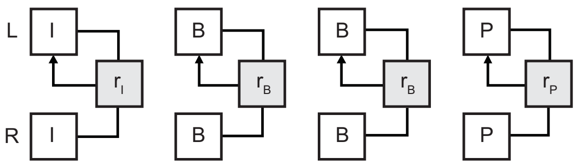

Fig. 3. Compatible mode of operation.

The residuals of the B-frames result in after interpolating the motion bidirectional prediction of the target frame with the disparity prediction of the corresponding left frame. Using Eq. (2), the residuals resulting from this consideration are expressed as

$$r_B = (w_f \times \widetilde{f}_I^R + w_b \times \widetilde{f}_P^R + w_d \times \widetilde{f}_B^L) - f_B^R \qquad (4)$$

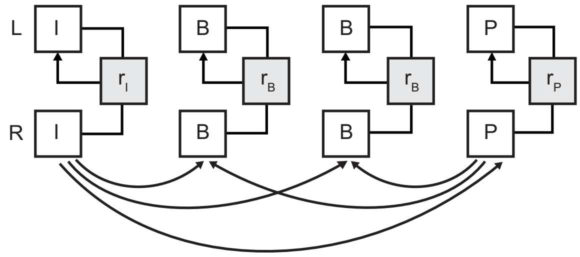

Fig. 4. Joint mode of operation.

where f_B^R is the current B-frame, f_I^R is the immediate preceding I or P-frame of the right sequence, f_P^R is the following P or I-frame of the right sequence, f_B^L is the corresponding left B-frame and w_f, w_b, w_d are the weighting factors for the bidirectional motion and disparity prediction correspondingly. The motion weighting factors are considered to have an equal role with the disparity weighting factor for this prediction and they are set to $w_f + w_b = 0.5$ and $w_d = 0.5$, because $w_f + w_b + w_d = 1$. The motion weighting factors for the prediction of the first B-frame is arbitrarily set to 70% for forward prediction and 30% for backward prediction. These factors are reversed for the prediction of the second B-frame. Therefore, the motion weighting factors for the first B-frame are set to $w_f = 0.35$ and $w_b = 0.15$ and their values are reversed for the second B-frame. The selection of the weighting factors is arbitrary and their tuning for every group of frames could provide better results at the cost of computational complexity. However this choice suffices for the purposes of this paper and is made in order to improve the equal weight case which is clearly sub-optimal.

For gray-scale images represented in the intensity range [0, 255], the residual pixel values lie in the range [-255, 255]. Finally, the residual is transformed in the normal range with the following normalization procedure.

$$\overline{r} = \lceil \frac{255 + r}{2} \rceil \qquad (5)$$

The ceiling function is not reversible and therefore the recovered values may have an inherent error of maximum one gray-level, which is negligible. Disparity compensation is implemented on 16×16 pixel blocks using exhaustive search over an area of 7 pixels. Thus, their values lie in the range of -7 to +7 or in the equivalent range of 0 to 15 and they are further losslessly compressed. The residual frames are JPEG compressed for the reduction of their size and are embedded into the corresponding cover frame of the left sequence, along with the corresponding disparity vector fields. The embedding unit employs the PVD algorithm for hiding the residuals along with their vector fields. The embedding capacity of each cover frame is pre-estimated and the compression quality factor of the residuals is adjusted, so that they can be incorporated in the cover frame.

The watermarked frame contains the information needed to recover the target frame. At the decoder, the compressed normalized residuals are first extracted along with their respective disparity fields. Thus, the initial cover frames are recovered and the target frames are further recovered from them using the disparity vectors and the decompressed residuals. The recovery of the target P-frames is performed by applying Eq. (6) which is derived from Eq. (3) and (5).

$$f_P = 255 + (w_p \times \widetilde{f}_I^R + w_d \times \widetilde{f}_P^L) - 2 \times \overline{r}_P \qquad (6)$$

The recovery of the target B-frames is performed by applying Eq. (7) which is derived from Eq. (4) and (5).

$$f_B = 255 + (w_f \times \widetilde{f}_I^R + w_b \times \widetilde{f}_P^R + w_d \times \widetilde{f}_B^L) - 2 \times \overline{r}_B \quad (7)$$

The cover frames are exactly recovered without any distortion, as watermarking is reversible. The recovered P or B-frames are distorted due to the compression of the residuals, which is at that stage unavoidable to match the embedding bit-rate of the cover frames. Obviously, the embedding bit-rate ability of a cover frame is image dependent. For images with textured regions, the embedding capacity is increased providing high quality factors and conversely for smooth images, the embedding capacity is decreased providing low quality factors.

IV. Experimental Results

The proposed coder for reversible watermarking was tested on two stereoscopic image sequences, namely “crowd”and “book-sale”. The temporal length of the tested sequences is 28 frames. The size of each frame is 320×240 pixels and the type of sequence is IBBPBB.

Initially, the compatible mode of operation is evaluated where the residual frame of each pair of frames is embedded to the corresponding frame of the left sequence together with the disparity vectors. Table I presents the average characteristics of the proposed reversible watermarking process. The embedding capacity of each cover frame (EC) depends on the frame content and the effectiveness of the used PVD data hiding algorithm. PVD embedding algorithm may be improved as far as embedding capacity is concerned at the expense of slightly increased computational overhead [16]. The pixel value differencing is applied to edged areas, while a least-significant-bit (LSB) replacement method is employed on smooth areas. The embedding capacity increases on average by 20% for the same objective quality. Therefore, target frames may be compressed with better quality factors implying better quality of their reconstructed counterparts and improvement of the stereoscopic sense. The quality factor (qf) for the compression of the residuals is selected so that the resulting capacity together with the data resulting from the lossless compression of the disparity vector fields to be slightly less than the corresponding capacity of the cover frame. The disparity vector field consists of 600 vectors, expressed in 4,800 bits. This binary string is reduced by 40% in average employing lossless compression with a winRAR encoder. The embedded bits (EB) are less than the capacity of the cover frame so that to let enough room for overhead information (length of the stream, dimensions of the frames etc) or for an authentication key. The average objective quality of the watermarked frames (PSNRw) is well over 35 dB which is the lower limit for imperceptibility. Also, the average objective quality of the recovered right frames (PSNRr) is quite satisfactory. In a reversible watermarking scheme, it is evident that the quality of the recovered frames depends on the embedding bit-rate of the cover frames.

TABLE I
Average characteristics of the compatible mode

Video	qf	PSNRw (dB)	PSNRr (dB)	EC (bits)	EB (bits)
crowd	89	40.51	34.79	117,700	115,096
book-sale	90	40.65	36.07	116,380	109,910

Next, the joint mode of operation is evaluated where the residual frame, which is embedded with the disparity vectors to the corresponding frame of the left sequence, is estimated by an interpolative scheme which takes into account both the motion prediction of the current right frame and the disparity prediction of the corresponding left frame. Table II presents the average characteristics of the proposed reversible watermarking process for the same quality factor of the previous mode. It is evident that the objective quality is better and the embedded bits are much less than in the previous mode of operation because of the more accurate prediction of the target frames by the proposed interpolative scheme. Although the embedded disparity vector fields are two for P-frames and three for B-frames, there is an overage in the embedding capacity that permits a better quality factor for the compression of some residuals. This will result an increase in the objective quality of the recovered right frames and consequently better stereoscopic sense.

TABLE II
Average characteristics of the joint mode

Video	qf	PSNRw (dB)	PSNRr (dB)	EC (bits)	EB (bits)
crowd	89	40.89	35.60	117,700	98,229
book-sale	90	40.94	36.64	116,380	97,104

Figure 5 illustrates the relationship between the embedding capacity of the cover frames and the embedded bits for the two modes of operation in the stereoscopic video of “crowd”. It can be seen that the embedded bits for I-frames are identical for the two modes of operation as the residuals result in after disparity compensation. Also, the embedded bits in B-frames are less than those in P-frames showing that the residual B-frames are more accurately predicted.

Figure 6 shows the worst watermarked frame and the worst recovered target frame for the joint mode of the stereoscopic image sequences “crowd”and “book-sale”respectively. Their objective quality shows that the proposed methods provide imperceptibility and a very good final stereoscopic sense. Finally, the joint scheme outperforms compatible because of the more accurate prediction of the P and B-frames, which implies residuals with less energy that are subjected to lossy compression with a better quality factor.

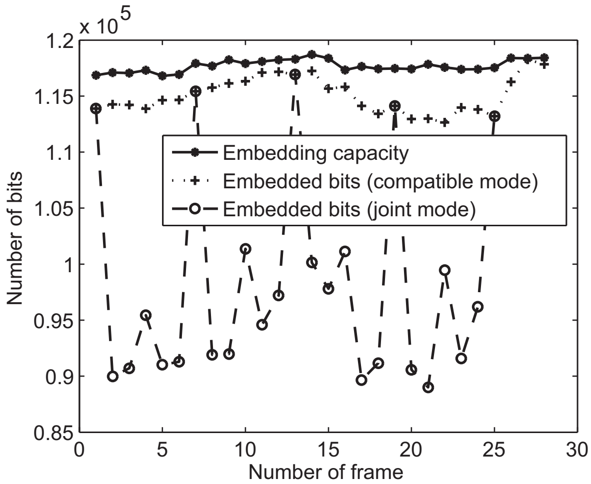

Fig. 5. Embedding capacity of cover frames and embedded bits for the two modes of operation of the stereo video "crowd".

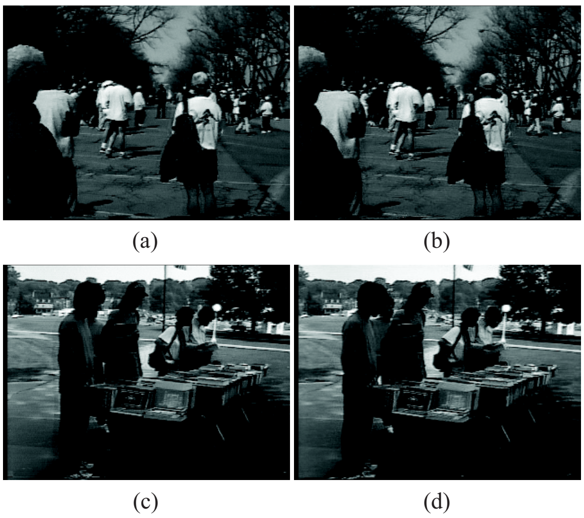

Fig. 6. Worst frames for the joint mode of the stereoscopic image sequences "crowd"and "book-sale", (a) and (c) watermarked frames, (b) and (d) recovered target frames.

V. CONCLUSION

In this work, a reversible watermarking scheme for halving the capacity of a stereoscopic image sequence is presented and its performance is evaluated. Two modes of operation are proposed that manipulate the residual estimation in a different way. The compatible method embeds, for each pair of frames, the disparity compensated difference and the corresponding disparity vectors into the reference frame. The joint method embeds, for each pair of frames, the motion and disparity estimated residual and the corresponding disparity vectors into the reference frame. The motion and disparity prediction of the P and B-frames follows an interpolative scheme used in stereo video compression. The embedding method used is the PVD algorithm where the cover frame is partitioned into blocks of two consecutive non-overlapped pixels where the secret bits are embedded according to their gray value difference. Because the embedding bit-rate of the cover image, which depends on the image content and the embedding algorithm, is not high enough, the residual is JPEG compressed whereas the disparity vectors are losslessly compressed. At the decoder's side, the watermarked frames are subjected to an inverse process where the initial left frames are exactly recovered while the right frames are recovered with distortions as the residuals were lossy compressed. The compression quality factor may be tuned so that to compensate the quality of the recovered right frame with the cover frame embedding bit rate. The joint mode of operation outperforms compatible because of the more accurate prediction of P and B-frames that results in lower energy residuals.

REFERENCES

[1] J. N. Ellinas, M. S. Sangriotis, "Stereo Video Coding Based on Interpolated Motion and Disparity Estimation", Proc. EURASIP 3rd Int. Conf. on Image and Signal Processing and Analysis, Rome, Italy, pp. 301-306, Sept. 2003.

[2] J. N. Ellinas, M. S. Sangriotis, "Stereo video coding based on quad-tree decomposition of B-P frames by motion and disparity interpolation", Proc. IEE Vision, Image and Signal Processing, vol. 152, no. 5, pp. 639-647, 2005.

[3] D. Coltuc, "On stereo embedding by reversible watermarking", Proc. Int. Symposium on Signals, Circuits and Systems (ISSCS 2007), Iasi, Romania, pp. 1-4, July 2007.

[4] Da-Chun Wu, and Wen-Hsiang Tsai, "A steganographic method for images by pixel-value differencing", Pattern Recognition Letters, Elsevier, no. 24, pp. 1613-1626, 2003.

[5] M. U. Celik, G. Sharma, A. M. Tekalp, and E. Saber, "Lossless generalized-lsb data embedding", IEEE Transactions on Image Processing, vol. 14, no. 2, pp. 253-266, 2005.

[6] J. Fridrich, M. Goljan, Q. Chen, and V. Pathak, "Lossless data embedding with file size preservation", Proc. of EI SPIE, San Jose, USA, pp. 354-365, Jan. 2004.

[7] A. Leest, M. Veen, and F. Bruekers, "Reversible image watermarking", Proc. Int. Conf. on Image Processing, Barcelona, Spain, pp. II-731-4, Sept. 2003.

[8] G. Xuan, C. Yang, Y. Zhen, Y. Q. Shi, and Z. Ni, "Reversible data hiding based on wavelet spread spectrum", Proc. IEEE 6th Workshop on Multimedia Signal Processing, Italy, pp. 211-214, Sept. 2004.

[9] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform", IEEE Transactions on Image Processing, vol. 13, no. 8, pp. 1147-1156, 2004.

[10] D. M. Thodi, and J. J. Rodriguez, "Prediction-error based reversible watermarking", Proc. Int. Conf. on Image Processing, Genova, pp. 1549-1552, Oct. 2004.

[11] J. Tian, "High capacity reversible data embedding and content authentication", IEEE Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, Hong Kong, pp. III-517-20, Apr. 2003.

[12] J. Tian, and R. O. Wells, "Reversible data-embedding with a hierarchical structure", Proc. Int. Conf. on Image Processing, Genova, pp. 3419-3422, Oct. 2004.

[13] C. C. Chang, W. L. Tai, and M. H. Lin, "A reversible data hiding scheme with modified side match vector quantization", Proc. Int. Conf. on Advanced Information Networking and Applications, Taiwan, pp. 947-952, Mar. 2005.

[14] C. D. Vleeschouwer, J. F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management", IEEE Transactions on Multimedia, vol. 5, no. 1, pp. 97-105, 2003.

[15] B. Yang, M. Schmucker, X. Niu, C. Busch, and S. Sun, "Reversible image watermarking by histogram modification for integer dct coefficients", Proc. IEEE 6th Workshop on Multimedia Signal Processing, Siena, Italy, pp. 143-146, Sept. 2004.

[16] H. C. Wu, C. S. Tsai, and M.S. Hwang, "Image steganographic scheme based on pixel-value differencing and LSB replacement methods", IEE Proc. Vision Image Signal Processing, vol. 152, no. 5, pp. 611-615, 2005.

John N. Ellinas was born in 1953. He graduated in Electrical and Electronic Engineering from the University of Sheffield, England, in 1977 and received the M.Sc. degree in Telecommunications, from the University of Sheffield, England, in 1978. He also received the Ph.D. degree from the University of Athens, Department of Informatics and Telecommunications, in 2005. Since 1983, he has been with the Department of Electronic Computer Systems, Technological Educational Institute of Piraeus in Greece, where he is an Associate Professor. His main interests are in the field of embedded computer systems and digital image processing. His research activity is focused on microprocessor systems design, stereoscopic image and video coding, image restoration and watermarking.