A Hybrid Compression Method for Integral Images Using Discrete Wavelet Transform and Discrete Cosine Transform

E. Elharar, Adrian Stern, Ofer Hadar, *Member, IEEE*, and Bahram Javidi, *Fellow, IEEE*

***Abstract*—Integral imaging (II) is a promising three-dimensional (3-D) imaging technique that uses an array of diffractive or refractive optical elements to record the 3-D information on a conventional digital sensor. With II, the object information is recorded in the form of an array of subimages, each representing a slightly different perspective of the object In order to obtain high-quality 3-D images, digital sensors with a large number of pixels are required. Consequently, high-quality II involves recording and processing large amounts of data. In this paper, we present a compression method developed for the particular characteristics of the digitally recorded integral image. The compression algorithm is based on a hybrid technique implementing a four-dimensional transform combining the discrete wavelet transform and the discrete cosine transform. The proposed algorithm outperforms the baseline JPEG compression scheme applied to II and a previous compression method developed for II based on MPEG II.**

***Index Terms*—Integral imaging, three-dimensional (3-D) image compression, 3-D imaging.**

I. Introduction

A THREE-DIMENSIONAL (3-D) sensing display technique based on integral imaging (II) produces true 3-D images with full parallax and continuous viewing points. Basically it enables us to capture a 3-D image by two-dimensional (2-D) elemental images obtained by a lenslet [1]. In the pickup process of II, typically a lenslet array or a pinhole array are used to capture rays of light emanating from a 3-D object from different directions while creating small elemental images, each with its own perspective of the object. An example of an elemental image array is shown in Fig. 1(top). These elemental images are recorded in a 2-D light-sensitive device such as a charge coupled device (CCD). The II image is composed of many element images, with their number corresponding to that of the number of lenslets. To reconstruct the 3-D image from the recorded 2-D elemental images, typically a spatial light modulator such as a liquid crystal display (LCD) together with another lenslet array are used to produce rays in the opposite direction to the captured ones, as illustrated in Fig. 1(bottom).

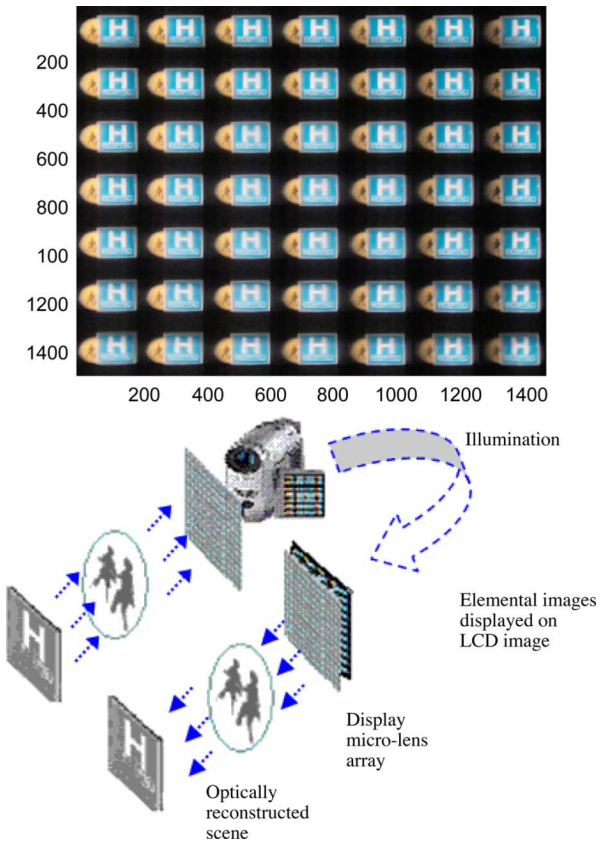

Fig. 1. (Top) Elemental images. (Bottom) Pickup and reconstruction of a 3-D scene, using the II scheme.

Representing a captured 3-D image with high resolution by 2-D elemental images requires a very large number of pixels. In practical applications, the recorded images need to be stored and transmitted, which involve considerable storage capacity and large transmission bandwidth, thus promoting the need for compression of II images [1].

In general, conventional 2-D images are characterized by high spatial correlation between neighboring pixels and therefore contain redundant information. In addition to correlation between adjacent pixels, II images exhibit high correlation between adjacent elemental images, as clearly seen in Fig. 1(top). The straightforward approach to compress integral images is by applying conventional compression methods like JPEG and JPEG 2000 based on discrete cosine transforms (DCTs)

Manuscript received June 17, 2006; revised October 30, 2006.

E. Elharar and A. Stern are with the Electro Optical Engineering Department, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: elharare@bgu.ac.il; stern@bgu.ac.il).

O. Hadar is with the Communication Systems Engineering Department, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: hadar@bgu.ac.il).

B. Javidi is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-1157 USA (e-mail: bahram@engr.uconn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JDT.2007.900915

or discrete wavelet transforms (DWTs), respectively, while ignoring the redundancy or correlation between adjacent elemental images. In this study, we exploit the cross-correlation between adjacent elemental images to increase the compression ratio and improve the reconstruction quality. As can be seen in Fig. 1(top), integral images exhibit correlation in four dimensions. Consider, for example, the central pixel in each elemental image and let (i, j) denote the indices of the column and rows of the elemental image in the array. This pixel exhibits 2-D correlation to neighboring pixels within the elemental image. The same pixel is also correlated to the central pixels in adjacent elemental images in a plane defined by the indexes of the elemental images (i, j), providing additional 2-D correlation. In order to exploit this 4-D correlation, we apply here a hybrid scheme of compression combining 2-D DCT and 2-D DWT.

The hybrid compression algorithm developed in this study is described in Section II. In Section III, the performance of this technique is compared with standard JPEG compression and with a compression technique developed for II presented in [2]. The compression technique presented in [2] rearranges the elemental images to form a sequence of elemental images. Then, the standard MPEG-II compression technique is applied to this sequence to exploit redundancy between elemental images. It is found that the compression technique developed here outperforms the other two compression methods.

II. 4-D Hybrid DWT-DCT Coding Scheme

In this paper, a 4-D DWT-DCT compression algorithm is proposed. The proposed hybrid algorithm is based on pixel transformation methods, DWT and DCT, a unique packet partition stage based on the high cross-correlation characteristic between adjacent elemental images, followed by quantization and entropy coding. The block diagram of the encoder and decoder of the proposed hybrid compression scheme is shown in Fig. 2. The decoder reconstructs the image by inverting the steps of the encoder.

The DCT and DWT transformations were integrated together in a hybrid architecture in a way which enables us to exploit the local characteristics within each elemental image and the redundancy due to the correlation between adjacent elemental images. In the first step, a DWT is applied to each elemental image. The DWT was successfully applied to digital holograms of 3-D objects [3]. Wavelet transform enables us to decorrelate the spatial correlation of pixels without the limitation of the block-based DCT. Hence, undesirable blocking artifacts in the reconstructed image, typical to DCT-based compression techniques (see, for instance, [4] and [5]), are avoided. In the next step, called *packet partition*, DWT coefficients belonging to different elemental images are grouped into blocks which are DCT'ed. The DCT stage excels in concentrating most of the signal in the lower spatial frequencies. For a typical 8×8 block, from the input coding, most of the spatial frequencies have zero or near-zero amplitude and can be approximated and then encoded efficiently.

The first algorithmic step in the hybrid algorithm (Fig. 2) divides the II image $X[r, s]$ into rectangular tiles on a regular grid with their number corresponding to that of the elemental image $x_{ij}[n, m]$. Each tile consists of one elemental image. A 2-D DWT is performed on each tile $x_{i,j}[n, m]$, with respect to n and m.

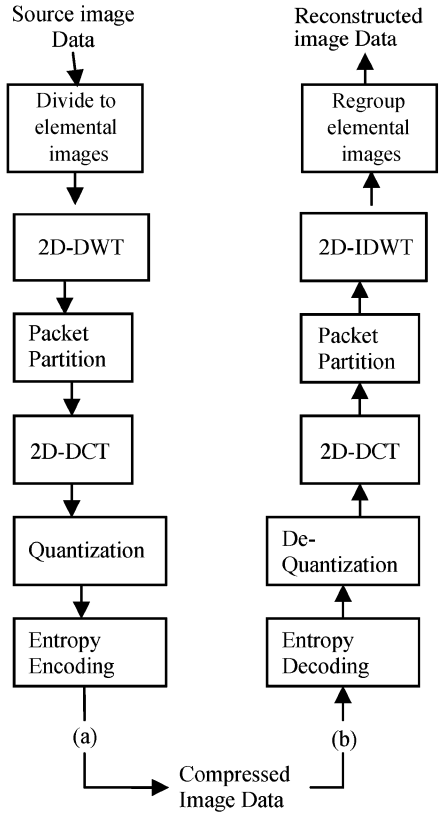

Fig. 2. Block diagram of the hybrid compression scheme.

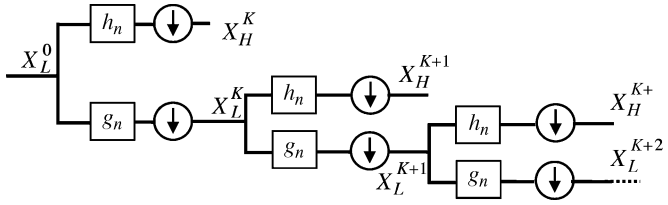

Fig. 3. 1-D DWT decomposition. At each stage K, the input signal is filtered through high-pass and low-pass filters h_n and g_n, respectively. The filtered outputs are then down sampled with a factor of 2.

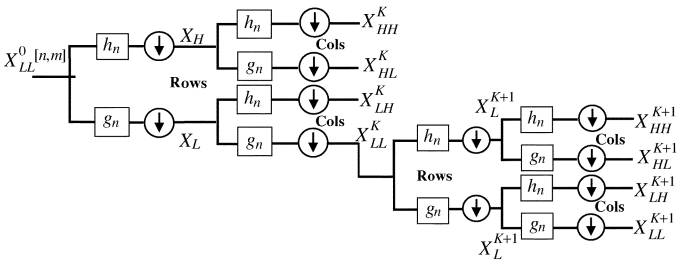

Fig. 4. 2-D DWT decomposition.

The wavelet transform decomposes the signal into a band of energy which is sampled at different rates [6]–[8]. These rates are determined to maximally preserve the information of the signal while minimizing the sampling rates or the resolution of each subband [6], [7]. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low-pass and high-pass filters, respectively. After each filtering step, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has a highest frequency

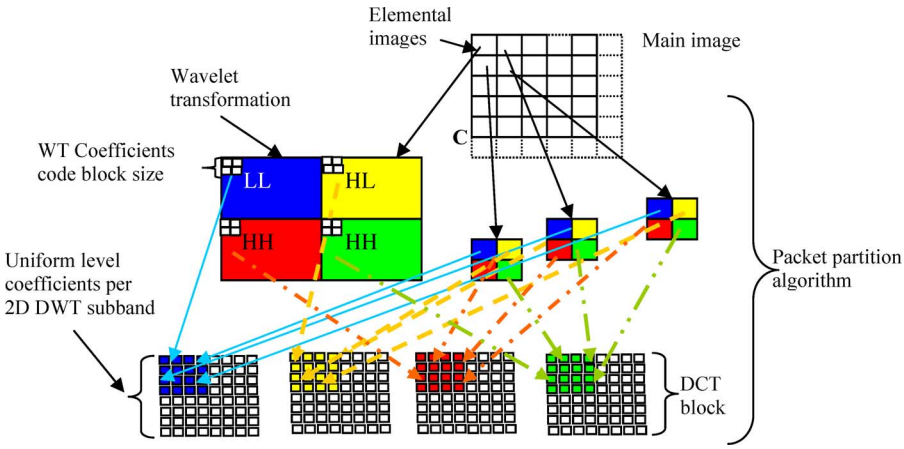

Fig. 5. "Packet partition" scheme.

of $\pi/2$ radians instead of π. The signal can therefore be subsampled by 2 simply by discarding every other sample. The generic form for a one-dimensional (1-D) DWT is illustrated in Fig. 3 and briefly described below.

For the sequence of low- and high-frequency coefficients of the decomposition layer K, we use the symbols X_L^K and X_H^K, respectively, as follows:

$$X_L^{K+1}[n] = \sum_{l=0}^{N_L-1} g[l] \times X_L^K[2n-l] \tag{1}$$

$$X_H^{K+1}[n] = \sum_{l=0}^{N_H-1} h[l] \times X_L^K[2n-1-l] \tag{2}$$

where N_H and N_L are the number of taps of the high-pass (h) and low-pass (g) filters, respectively, and K is the decomposition level. A 2-D DWT decomposition scheme is illustrated in Fig. 4. For each level, the input signal is filtered along the rows, and the resulting signal is filtered along the columns. In this way, the 2-D decomposition of an input signal $x_{\mathrm{i,j}}[n,m]$ represented as $X_{LL}^K[n,m]$, with n columns and m rows, is given by

$$X_H^{K+1}[\mathrm{row},m] = \sum_{l=0}^{N_L-1} g[l] \times X_{LL}^K[\mathrm{row},2n-l] \tag{3}$$

$$X_H^{K+1}[\mathrm{row},m] = \sum_{l=0}^{N_H-1} h[l] \times X_{LL}^K[\mathrm{row},2m-1-l] \tag{4}$$

$$X_{LL}^{K+1}[n,\mathrm{col}] = \sum_{l=0}^{N_L-1} g[l] \times X_L^K[2n-l,\mathrm{col}] \tag{5}$$

$$X_{LH}^{K+1}[n,\mathrm{col}] = \sum_{l=0}^{N_H-1} h[l] \times X_L^{K+1}[2n-1-l,\mathrm{col}] \tag{6}$$

$$X_{HL}^{K+1}[n,\mathrm{col}] = \sum_{l=0}^{N_L-1} g[l] \times X_H^K[2n-l,\mathrm{col}] \tag{7}$$

$$X_{HH}^{K+1}[n,\mathrm{col}] = \sum_{l=0}^{N_H-1} h[l] \times X_H^K[2n-1-l,\mathrm{col}] \tag{8}$$

where $l = \{0,1,\ldots,L-1\}$, $\mathrm{row} = \{0,1,\ldots,N/2^K-1\}$, $m = \{0,1,\ldots,M/2^{K+1}-1\}$, $\mathrm{col} = \{0,1,\ldots,M/2^{K+1}-1\}$, $n = \{0,1,\ldots,N/2^{K+1}-1\}$, and $X_{LL}^0[n,m] = X_{i,j}[n,m]$.

The following stage in our compression algorithm is the *packet partition* stage. With *packet partition*, we implement a split-merge algorithm with which wavelet coefficients X_{LL}^K, X_{LH}^K, X_{HL}^K, and X_{HH}^K are rearranged to form a set of similar level input data. The process of *packet partition* is illustrated in Fig. 5. Blocks of DWT coefficients of each elemental image are grouped together with the respective blocks of DWT coefficients from adjacent elemental images. The obtained packets, consisting of similar groups of DWT coefficients originating from neighbor elemental images (Fig. 5), are then 2–D-DCT'ed in the next stage. Thus, in the packet partition process, the inter-elemental images information is gathered, and then it is decorrelated using DCT. The number of elemental images involved in a single 2-D DCT computation N_{EI} depends on the choice of the DCT block size N_{DCT} and the size of the wavelet coefficient block is N_{WB}:

$$N_{\mathrm{EI}} = \frac{N_{\mathrm{DCT}}}{N_{\mathrm{WB}}}. \tag{9}$$

For example, in the packet partition illustrated in Fig. 5, $N_{\mathrm{DCT}} = 8 \times 8$, the DWT coefficients blocks are of size 2×2 so that $\mathrm{N_{WB}} = 4$ and $N_{\mathrm{EI}} = 16$.

In our experiments, described in the following section, we used DCT blocks of size 8×8. We also find empirically for the set of images we considered that best results are obtained with $N_{\mathrm{EI}} = 4$. However, since N_{EI} indirectly represents the correlation length between elemental images, the optimal $N_{\mathrm{EI}} = 4$ value depends on the II setup (mainly on the lenslet array pitch and object location).

In DCT coding, each component of the image is typically subdivided into blocks of 8×8 pixels. A 2-D DCT is applied to each block of the input data to obtain an 8×8 array of DCT coefficients. If $Y[n,m]$ represents the wavelet coefficient value in the packet partition block, then the DCT is computed for each block of the coefficient data as follows:

$$y[n,m] = \frac{c[u]c[v]}{4} \sum_{m=0}^{7} \sum_{n=0}^{7} Y[n,m] \cos\frac{(2n+1)u\pi}{16} \times \cos\frac{(2m+1)v\pi}{16} \tag{10}$$

where

$$0 \le u; \quad v \le 7; \quad c[u] = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}}, & u = 0 \\ 1, & 1 \le u, v \le 7 \end{array} \right\}.$$

The original wavelet coefficient code block samples can be recovered from the DCT coefficients by applying inverse DCT as follows:

$$Y[m,n] = \sum_{u=o}^{7}\sum_{v=0}^{7} \frac{c[u]c[v]}{4} y[u,v] \cos\frac{(2m+1)u\pi}{4} \times \cos\frac{(2n+1)v\pi}{16} \tag{11}$$

where $0 \leq m$ and $n \leq 7$.

The magnitude of DCT coefficients exhibits a pattern of their occurrences in the rearranged wavelet coefficient array after the packet partition stage. The DCT coefficients corresponding to the lowest basis function are usually large in magnitude, and they are also deemed to be perceptually most significant. These features are exploited in various methods of quantization and symbol coding. At the quantization step, each DCT coefficient $x[m,n]$, where $0 \leq m$ and $n \leq 7$ are mapped into one of a finite number of levels determined by the desired compression factor. This is achieved by dividing each block (point-by-point) by an 8×8 quantization matrix $\mathbf{q}$ and rounding the result as follows:

$$\mathbf{q}_y[m,n] = \left[\frac{y[m,n]}{q[m,n]}\right]_{\text{round}}. \tag{12}$$

As the quantization makes the coding lossy, it provides the major contribution in this step of the compression. The following stage consists of entropy coding in a similar manner as performed in JPEG [8]. The symbols defined for dc and ac coefficients are entropy coded by using Huffman coding. Huffman coding is a method of variable-length coding in which shorter code words are assigned to the more frequently occurring symbols in order to achieve the shortest description possible. For color images, the technique is implemented on YUV components separately.

III. Experimental Results

Here, we present the experimental results for II compression using the proposed hybrid technique. For demonstration purposes, we compressed the image by the proposed hybrid algorithm and by conventional JPEG at different compression depths. We evaluate the efficiency of compression by evaluating the peak-to-peak signal to noise ratio (PSNR) defined as

$$\text{PSNR}(I_o, I_u) = 10\log_{10}\left(\frac{P^2}{\text{MSE}(I_o, I_u)}\right) \tag{13}$$

where P is the maximum value in one pixel, I_O is the original image, and I_u is the image obtained after de-compression.

The compression ratio is defined as

$$r = \frac{\text{Original image size}}{\text{Compressed image size}}. \tag{14}$$

Fig. 6 compares reconstructions of an image compressed with standard JPEG and with the proposed hybrid technique. The image compressed shown in Fig. 1(top) is 1456×1456 pixels, and the number of elemental images is $7 \times 7 = 49$. The image was compressed with both compression techniques to have the same size at the ratio of 1:100. In Fig. 6, it can be seen that a better contrast is obtained with the hybrid compression technique. From the enlarged images in Fig. 6(c) and (d), it is evident

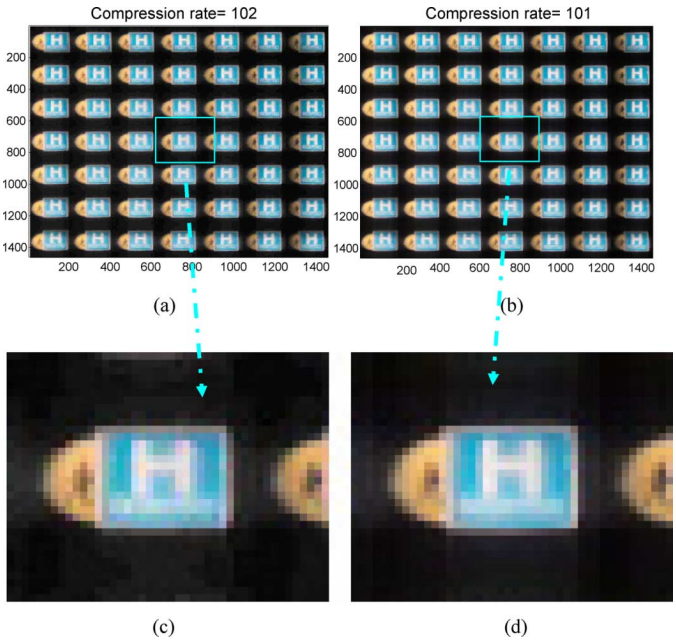

Fig. 6. Decompressed integral images of (a) JPEG compressed image and (b) using Hybrid compression technique. (c), (d) Respective enlargements of the upper row images.

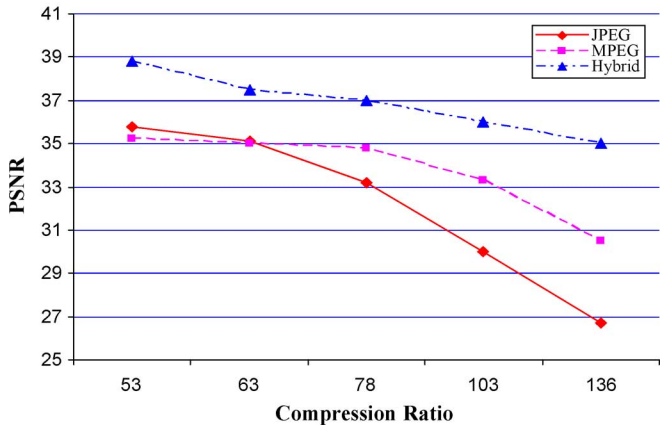

Fig. 7. PSNR (dB) versus compression ratio for Hybrid, JPEG and MPEG-based compression schemes.

that the edges are sharper and the background is less noisy with the hybrid compression technique.

Fig. 7 shows a comparison between the PSNR obtained with our hybrid method, the regular JPEG method, and the MPEG-based compression scheme proposed in [2]. With the compression method in [2], the integral image is transformed into a sequence of elemental images which are compressed with a standard MPEG-II algorithm. It can be seen that our method outperforms the regular JPEG and MPEG-based algorithms with at least 2–4 dB for compression ratios up to $r = 70$. The difference is much greater at high compression ratios. In fact, it can be seen that the hybrid scheme maintains an acceptable PSNR (higher than 35 dB) even at a compression ratio as high as $r = 136$, where the JPEG image quality is quite poor ($\text{PSNR} = 27$ dB).

The MPEG-based technique presented in [2] provides better results than standard JPEG does at high compression ratios, as it takes to account the similarity between adjacent elemental

images. It can be seen in Fig. 7 that the DCT-DWT hybrid compression technique provides higher PSNR values of about 2–5 dB than the MPEG-based technique. This is attributed to the fact that the MPEG-based technique exploits the correlation between the elemental images along one dimension only (the direction of the elemental image sequence), whereas the DCT-DWT hybrid compression exploits 2-D correlation between elemental images.

IV. Conclusion

A new compression scheme has been developed for 3-D images captured using an II technique. The proposed compression scheme is based on a hybrid technique implementing a 4-D transform combining DWT and DCT while exploiting both the correlation within elemental images and the cross correlation between adjacent elemental images. The image quality obtained with the presented hybrid technique is compared with that obtained by using JPEG and MPEG-based compression schemes at the same compression ratio. The hybrid compression scheme has shown significantly better results compared with the other two compression schemes.

References

[1] A. Stern and B. Javidi, "Three dimensional sensing, visualization, and processing using integral imaging," *Proc. IEEE*, vol. 94, Special Issue on 3-D Technologies for Imaging and Display, no. 3, pp. 591–607, Mar. 2006.

[2] S. Yeom, A. Stern, and B. Javidi, "Compression of 3-D color integral images," *Opt. Express*, vol. 12, pp. 1632–1642, Apr. 2004.

[3] A. Shortt, T. J. Naughton, and B. Javidi, "Compression of digital holograms of three-dimensional objects using wavelets," *Opt. Express*, vol. 14, pp. 2625–2630, Apr. 2006.

[4] Y. Lee, H. C. Kim, and H. Park, "Blocking effect reduction of JPEG image by Signal adaptive filtering," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 229–234, Feb. 1998.

[5] G. Lakhani, "Improved equations for JPEG's blocking artifacts reduction approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 6, pp. 930–934, Dec. 1997.

[6] B. E. Usevitch, "A tutorial on modern lossy compression wavelet image compression: Foundation of JPEG 2000," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 22–35, Sep. 2001.

[7] M. W. Marcellin1, M. J. Gormish2, A. Bilgin1, and M. P. Boliek, "An overview of JPEG-2000," in *Proc. IEEE Data Compression Conf.*, Mar. 2000, pp. 523–541.

[8] Z. Xiong and K. Ramchandran, "Wavelet image compression," in *Handbook of Image and Video Processing*, A. Bovik, Ed., 2nd ed. New York: Academic, 2005, ch. 4–5.

E. Elharar, photograph and biography not available at the time of publication.

Adrian Stern received the B.Sc., M.Sc. (*cum laude*), and Ph.D. degrees from Ben Gurion University of the Negev, Beer-Sheva, Israel, in 1988, 1997, and 2003, respectively, all in electrical and computer engineering.

During 2002–2003, he was a Postdoctoral Fellow with the Electrical and Computer Engineering Department, University of Connecticut, Storrs. He is now with Department of Electro-Optical Engineering, Ben Gurion University of the Negev. His current research interests include sequences of images processing, image restoration, image and video compression, three-dimensional imaging, biomedical imaging, optical encryption, and nonconventional imaging.

Ofer Hadar (S'91–M'00) received the B.Sc., M.Sc. (*cum laude*), and the Ph.D. degrees from the Ben Gurion University of the Negev, Beer-Sheva, Israel, in 1990, 1992, and 1997, respectively, all in electrical and computer engineering. His Ph.D. dissertation dealt with the effects of vibrations and motion on image quality and target acquisition.

The prestigious Clore Fellowship supported his Ph.D. studies. From August 1996 to February 1997, he was with CREOL, Central Florida University, Orlando, as a Research Visiting Scientist, where he was involved with angular dependence of sampling MTF and over-sampling MTF. From October 1997 to March 1999, he was a Postdoctoral Fellow with the Department of Computer Science, Technion-Israel Institute of Technology, Haifa. Currently, he is a Senior Lecturer with the Communication Systems Engineering Department, Ben Gurion University of the Negev. His research interests include image compression, video compression, rate control in H.264, packet video, transmission of video over IP networks, video rate smoothing and multiplexing, video quality measures, and signal processing in audio and Hi Fi Systems. He is also a consultant for several hi-tech companies such as EnQuad Technologies, Ltd., in the area of MPEG-4, and Scopus in the area of video compression and transmission over satellite network.

Dr. Hadar is a member of SPIE.

Bahram Javidi (S'82–M'83–SM'96–F'98) received the B.S. degree in electrical engineering from George Washington University, Washington, D.C., and the M.S. and Ph.D. degrees in electrical engineering from the Pennsylvania State University, University Park. He is a Board of Trustees Distinguished Professor at the University of Connecticut, Storrs. He has supervised over 80 master's and doctoral graduate students, postdoctoral students, and visiting professors during his academic career. He has published over 230 technical articles in major journals. He has published over 270 conference proceedings, including over 100 invited conference papers, and 60 invited presentations. His papers have been cited over 3200 times, according to the citation index of WEB of Science. His papers have appeared in *Physics Today* and *Nature*, and his research has been cited in the *Frontiers in Engineering Newsletter,* published by the National Academy of Engineering, *IEEE Spectrum, Science, New Scientist,* and *National Science Foundation Newsletter.* He has completed eight books, including *Physics of Automatic Target Recognition* (Springer-Verlag, 2007), *Optical Imaging Sensors and Systems for Homeland Security Applications* (Springer-Verlag, 2005), *Optical and Digital Techniques For Information Security* (Springer-Verlag, 2005), *Image Recognition: Algorithms, Systems, and Applications* (Marcel Dekker, 2002), *Three Dimensional Television, Video, and Display Technologies* (Springer-Verlag, 2002), and *Smart Imaging Systems* (SPIE Press, 2001). He is currently on the Board of Editors of the Proceedings of the IEEE, the Editor in Chief of the Springer-Verlag series on Advanced Science and Technologies for Security Applications, and the Journal of Display Technologies. He has served as topical editor for Springer-Verlag, Marcel Dekker, *Optical Engineering Journal,* and the IEEE/SPIE Press Series on Imaging Science and Engineering. He has held visiting positions during his sabbatical leave at the Massachusetts Institute of Technology, United States Air Force Rome Lab at Hanscom Base, and Thomson-CSF Research Labs in Orsay, France. He is a consultant to industry in the areas of optical systems, image recognition systems, and 3-D optical imaging systems.

Dr. Javidi is Fellow of six professional societies. He was awarded the Dennis Gabor Award in Diffractive Wave Technologies by SPIE in 2005. He was the recipient of the IEEE Lasers and Electro-Optics Society Distinguished Lecturer Award twice in 2003 and 2004. He has been awarded the University of Connecticut Board Of Trustees Distinguished Professor Award, the School Of Engineering Distinguished Professor Award, University of Connecticut Alumni Association Excellence in Research Award, the Chancellor's Research Excellence Award, and the first Electrical and Computer Engineering Department Outstanding Research Award. In 1990, the National Science Foundation named him a Presidential Young Investigator.