ELSEVIER

Signal Processing: *Image Communication* 15 (2000) 705–727

SIGNAL PROCESSING: IMAGE COMMUNICATION

www.elsevier.nl/locate/image

A wavelet multiresolution compression technique for 3D stereoscopic image sequence based on mixed-resolution psychophysical experiments☆

Po-Rong Chang*, Meng-Jer Wu

Department of Communication Engineering, National Chiao-Tung University, Hsin-Chu, Taiwan

Received 18 August 1997

Abstract

Three-dimensional visual communications, made through the use of stereoscopic images, are able to achieve total display realism. In order to create a 3D system with two images (left and right) that should be transmitted simultaneously, the large amount of information contained in the images should be reduced, and an efficient coding appropriate for stereoscopic images is developed. Stereoscopic image sequence compression involves the exploitation of the spatial redundancy between the left and right image frames to achieve the compression ratio higher than that are by the independent compression of the two frames. In order to further achieve the higher compression ratio, in this paper, we employ the mixed-resolution psychophysical experiments to the stereo image compression. The experiments have shown that a stereo image pair with one high-resolution image and one lower-resolution image is sufficient to provide good stereoscopic depth perception. Thus, one image sequence is compressed independent of the other sequence using the motion compensation, while the other sequence is estimated at a lower resolution from this stream using the low-resolution disparity compensation. To implement the mixed-resolution coding, a wavelet multiresolution framework has been adopted to facilitate such an estimation of motion and disparity vectors at different resolutions. Experimental results indicate that the compression ratio for a typical stereoscopic image sequence is about 90, without any significant loss in the perceived 3D stereoscopic image quality.

Keywords: 3D stereoscopic image; Wavelet resolution coding

1. Introduction

Stereoscopic image display is a simple and compact means of portraying depth information on a two-dimensional screen. The binocular parallax or disparity between two images of the same scene, shot from two nearby points-of-view, contains information about the relative depths of the objects in the scene. Two pictures acquired in this manner form a stereo pair. This relative depth can be deduced by humans, when each eye is presented with its corresponding image. In other words, the sense of stereovision can be simulated by acquiring these two pictures of the same scene with the disparity and by presenting the left picture to the left eye and the right picture to the right eye. Thus, stereoscopic

* Corresponding author. Fax: 886-35-710116.

E-mail address: prchang@cc.nctu.edu.tw (P.R. Chang)

☆ This work was supported in part by the National Science Council, R.O.C., under contract NSC-88-2213-z-009-128.

0923-5965/00/$ - see front matter
PII: S0923-5965(99)00016-8

image transmission (storage) requires twice the conventional monocular transmission (storage) bandwidth. However, several schemes [2,5,8,14,16,17] have been developed to exploit the disparity relation to achieve compression ratios higher than that are obtained by the independent compression of the two pictures.

In order to further increase the compression ratio, in this paper, we apply the mixed-resolution coding technique [11,14,18] to incorporate with the disparity-compensated stereoscopic image compression. The mixed-resolution coding is a perceptually justified technique that is achieved by presenting one eye with a low-resolution picture and the other eye with one high-resolution picture. Psychophysical experiments [11,14,18] have shown that a stereo image pair with one high-resolution image and one low-resolution image provides the almost same stereoscopic depth to that of a stereo image with two high-resolution images. By combining both the mixed-resolution coding and disparity-compensated techniques, one reference (left) high-resolution image sequence can be compressed by motion-compensated scheme [19,20] independent of the other (right) image sequence. By performing low-resolution disparity-compensated technique, the disparity is able to predict the low-resolution right image sequence from the left image sequence at a lower resolution using the disparity relation. The low-resolution images are obtained using the well-known novel wavelet decomposition [1,10,20]. Another advantage of the wavelet decomposition is that it is very suitable for image compression [1,20]. After the wavelet decomposition, an image is divided into several layers with different importance. Subimages at different layers correspond to different resolution and different frequency ranges, which match the frequency-selected properties of human visual system. Antonini et al. [1] proposed a multiresolution-codebook-based vector quantization (VQ) technique for the wavelet transform, where each subcodebook corresponds to a wavelet subimage at its corresponding resolution level. In Section 3, the VQ-based wavelet multiresolution compression technique is proposed to code the left image of a still stereo image. However, for stereo image sequence, we propose an interframe hybrid DPCM/DWT/VQ scheme to code its left image sequence, where DWT denotes the discrete wavelet transform. The DPCM produces a number of prediction error subimages for the motion-compensated wavelet-decomposed subimages in order to improve the reconstruction image quality. These prediction error subimages are vector quantized using a multiresolution codebook. For low-resolution right image sequence, an inter-view hybrid DPCM/DWT/VQ scheme is applied to the low-resolution right sequence. The DPCM generates a number of prediction error subimages for the disparity-compensated subimages. Similarly, these prediction error subimages are also vector quantized. Since the estimation of both the motion vector and disparity is the computational burden of the joint motion and disparity compensated technique, we apply the variable block-size multiresolution block matching method [17,20] to reduce their computational complexity. These estimated motion vector and disparity are then DPCM coded, and all quantities are entropy-coded prior to the transmission.

2. Stereoscopic image compression using mixed-resolution coding techniques

2.1. Theory of stereovision

The sense of stereovision is normally simulated by viewing a true, three-dimensional scene. It is possible to stimulate the sense of stereovision artificially by acquiring two pictures of the same scene from separated positions, and by presenting the left picture to the left eye and the right picture to the right eye. Two pictures acquired in this manner form a stereopair. One of the most important ideas in the study of stereopairs is that of disparity. Fig. 1 illustrates the concept of disparity. Given a point A in the left picture, its matching point B in the right picture does not in general lie directly underneath A. The vector connecting B to A has been called the disparity, the stereo disparity, the binocular disparity, and the binocular parallax of the point pair (A, B). The disparity $\boldsymbol{d}$ associated with the point pair (A, B) consists of two components: a horizontal component d_h and a vertical component d_v. Depending on the camera geometry being

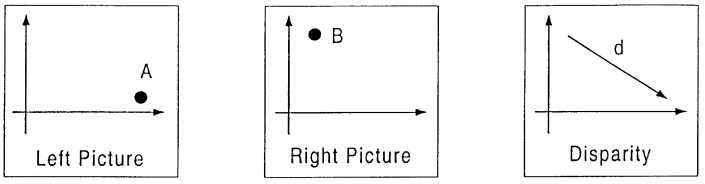


Fig. 1. Stereo disparity: A and B are matching points in the stereopair, and $\boldsymbol{d}$ is the disparity vector.

used, each component of the disparity can be either positive or negative. When negative disparity occurs, the scene being viewed will appear to be floating in the space between viewer's eyes and monitor. This type of imagery cannot be produced without the aid of stereoscopic devices such as shutter glasses. For another case of parallel axes camera geometry, the vertical component of disparity is always zero and the horizontal component of the disparity is always positive. This implies that the parallel axes geometry processes a simple mathematical relationship between the disparity of a point pair and the distance to the object it represents. In general, the disparity vector $\boldsymbol{d}$ can be used to predict one image of a stereopair from the other. For example, given a luminance level of left picture at a position $\boldsymbol{p}$, $I_L(\boldsymbol{p})$, the luminance level of its corresponding right picture can be calculated as

$$I_R(\boldsymbol{p}) = I_L(\boldsymbol{p} + \boldsymbol{d}), \tag{1}$$

where $\boldsymbol{d}$ denotes the disparity vector whose direction from left to right.

2.2. Mixed-resolution coding for stereopair data compression using wavelet multiresolution techniques

Mixed-resolution coding is a perceptually justified technique for compressing stereopairs. The compression is achieved by presenting one eye with a low-resolution picture and the other eye with a high-resolution picture. Psychophysical experiments [11,14,18] have shown that a stereo image with one high-resolution image and one lower resolution image is sufficient to provide the almost same stereoscopic depth as that of a stereo image with two high-resolution images. Thus, the eye/brain can easily fuse such stereopairs and perceive depth in them. In summary, the concept of mixed-resolution technique can be symbolically represented by the following equations:

$$\begin{aligned}&\text{(Stereo image)}\\ &\quad\approx \text{(High resolution left image)}\\ &\qquad + \text{(Low resolution right image)},\end{aligned} \tag{2}$$

$$\begin{aligned}&\text{(LR right image)}\\ &\quad= \text{(LR left image)}\\ &\qquad + \text{(Disparity between both LR images)},\end{aligned} \tag{3}$$

where LR denotes the low-resolution image.

From the above discussion, the mixed-resolution coding is able to significantly reduce the bit-rate required to transmit a stereo image with two high-resolution images. To implement the mixed-resolution coding, one of the novel techniques is based on the well-known wavelet multiresolution signal representation [1,10,20]. In the remainder of this section, we would give a brief review on the wavelet multiresolution technique.

Wavelets are functions generated from one single function ψ by dilations and translations

$$\psi_{a,b}(x) = |a|^{1/2}\psi\left(\frac{x-b}{a}\right), \quad a \in R^+, b \in R, \tag{4}$$

where the shift parameter "b" gives the position of wavelet, whereas the dilation parameter "a" governs its frequency.

The mother wavelet ψ has to satisfy the admissibility condition

$$\int_{-\infty}^{\infty} |\hat{\Psi}(\omega)|^2 \, |\omega|^{-1} \, \mathrm{d}\omega < \infty, \tag{5}$$

where $\hat{\Psi}$ is the Fourier transform of ψ.

For $a \gg 1$, the wavelet $\psi_{a,b}$ is a very highly concentrated "shrunken" version of the mother wavelet, with frequency content mostly in the high-frequency range. Conversely, for $a \ll 1$, $\psi_{a,b}$ is very much spread out and has mostly low frequencies.

Grossman and Morlet [7] showed that any square integrable function $f(x) \in L_2(R)$ can be represented in terms of a set of wavelet basis functions that cover all the scales. Such a representation writes f as an integral over a and b of $\psi_{a,b}$ with appropriate weighting coefficients [4]. Practically, one prefers to write f as a discrete superposition since sum is preferred rather than integral. Thus, one introduces a discretization, $a = 2^m$, $b = n2^m$, with $m, n \in Z$. The wavelet becomes therefore

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n), \quad (m,n) \in Z^2, \tag{6}$$

where m is a scaling parameter and n is a shift parameter.

The basic idea of the wavelet transform is to represent any arbitrary function f as a superposition of wavelets,

$$f(x) = \sum_{m,n} Wf(m,n)\psi_{m,n}(x), \tag{7}$$

where the wavelet transform $Wf(m,n)$ is defined as in terms of an inner product

$$\begin{aligned} Wf(m,n) &= \langle f(x), \psi_{m,n}(x) \rangle \\ &= \int_{-\infty}^{\infty} f(x)\psi_{m,n}(x)\,\mathrm{d}x. \end{aligned} \tag{8}$$

$Wf(m,n)$ yields to the detail signal of $f(x)$ at the resolution 2^m. It is characterized by the set of such inner products

$$W_{2^m}f = (\langle f(x), \psi_{m,n}(x) \rangle)_{n \in Z} = \{Wf(m,n); n \in Z\}, \tag{9}$$

where $W_{2^m}f$ is called the discrete detail signal or wavelet at the resolution 2^m. It contains the difference of information between the approximation of $f(x)$ at the resolution 2^{m+1} and 2^m.

In a multiresolution analysis, one really has two functions: the mother wavelet ψ and a scaling function ϕ. Let V_m denote the vector space spanned by $\phi_{m,n}$, which are generated by the dilation and translation of the scaling function ϕ:

$$\phi_{m,n}(x) = 2^{-m/2}\phi(2^{-m}x - n). \tag{10}$$

The vector space V_m can be interpreted as the set of all possible approximations at the resolution 2^m of functions in $L^2(R)$. These spaces V_m describe successive approximation spaces, $\cdots V_2 \subseteq V_1 \subseteq V_0 \subseteq V_{-1} \subseteq V_{-2} \cdots$, each with resolution 2^m. For each m, the $\psi_{m,n}$ span a space O_m which is exactly the orthogonal complement in V_{m-1} of V_m; the discrete detail signal $W_{2^m}f$, therefore, describes the information lost when going from an approximation of f with resolution 2^{m-1} to the coarser approximation with resolution 2^m.

Therefore, a discrete wavelet transform at resolution depth M decomposes a signal $f(x)$ into a set of signal segments at different scales:

$$f(x) \to \{S_{2^M}f, W_{2^M}f, W_{2^{M-1}}f, \ldots, W_{2^1}f\}. \tag{11}$$

Let $\{S_{2^m}f : m = 0, 1, \ldots, M\}$ represent a set of approximations of $f(x)$ at resolution levels $\{m = 0, 1, \ldots, M\}$. $S_{2^m}f$ is defined as

$$S_{2^m}f = (\langle f(x), \phi_{m,n}(x) \rangle)_{n \in Z}, \tag{12}$$

where $Sf(m,n) = \langle f(x), \phi_{m,n}(x) \rangle$, $m, n \in Z$.

Obviously, S_{2^M} is the approximation of $f(x)$ at the lowest resolution level M and $S_{2^0}f = f$ is the original signal. Therefore, $\{S_{2^m}f : m = 0, 1, \ldots, M\}$ form a multiresolution pyramid where resolution increases as the layer decreases. All layers in the pyramid structure are highly correlated since the higher layers of the pyramid constitute a subset of the lower layers, i.e., $S_{2^m}f \in S_{2^{m-1}}f$. In order to remove the interlayer redundancies, the pyramid $\{S_{2^M}, W_{2^m} : m = 0, 1, \ldots, M\}$ is defined as

$$W_{2^m}f = \mathrm{Diff}(S_{2^{m-1}}f, S_{2^m}f), \quad m = 0, 1, \ldots, M, \tag{13}$$

$$S_{2^m}f = \mathrm{Add}(W_{2^{m+1}}f, S_{2^{m+1}}f), \quad m = 0, 1, \ldots, M-1, \tag{14}$$

which carries the information from layer m to $m-1$ (finer approximation) or conveys the lost information from layer $m-1$ to m (coarser approximation). The Diff( ) and Add( ) operations in (13) and (14) correspond to the expansion of the scaling function and wavelet functions, respectively [20].

The extension of the 1D wavelet transform to 2D is straightforward. A separable wavelet transform is the one whose 2D scaling function $\phi(x,y)$ can be

written as

$$\phi(x, y) = \phi(x)\phi(y). \tag{15}$$

Mallat [10] showed that the 2D wavelet at a given resolution 2^m can be completely represented by three separable orthogonal 2D wavelet basis functions in $L^2(R^2)$:

$$\psi^1_{mnl}(x, y) = \phi_{mn}(x)\psi_{ml}(y), \tag{16}$$

$$\psi^2_{mnl}(x, y) = \psi_{mn}(x)\phi_{ml}(y), \tag{17}$$

$$\psi^3_{mnl}(x, y) = \psi_{mn}(x)\psi_{nl}(y), \tag{18}$$

where n and l are shift parameters for x- and y-direction, respectively. Therefore, a 2D dyadic wavelet transform of image $f(x, y)$ between the scale 2^1 and 2^M can be represented as a sequence of subimages:

$$\{S_{2^M}f, [W^j_{2^M}f]_{j=1,2,3}, \ldots, [W^j_{2^1}f]_{j=1,2,3}\}, \tag{19}$$

where $S_{2^m}f$ is the approximation of image $f(x, y)$ at the lowest resolution 2^m and the three detail subimages at resolution 2^m are defined by

$$W^j_{2^m}f = (\langle f(x, y), \psi^j_{mnl}(x, y)\rangle)_{(n,l)\in Z^2}, \quad 1 \leqslant j \leqslant 3. \tag{20}$$

The 2D separable wavelet decomposition can be implemented first in columns and then in rows independently. Fig. 2 shows the data structure of wavelet decomposition. $g(n) = (-1)^n h(-n+1)$ and $h(n) = 2^{1/2}\int\phi(x-n)\phi(2x)\,dx$ denote the 1D high-pass and low-pass filters, respectively. $\tilde{h}(n)$ and $\tilde{g}(n)$ are the conjugate filters of $h(n)$ and $g(n)$, respectively. The filter pair H and G related by $h(n)$ and $g(n)$ correspond to the expansion of the scaling and wavelet functions, respectively. This decomposition provides subimages corresponding to different resolution levels and orientations shown in Fig. 3 with resolution depth 2 ($M = 2$). The wavelet reconstruction scheme of the image is illustrated in Fig. 4. Zhang and Zafar [20] have shown that the decomposed image forms a pyramid structure up to M layers with three detail subimages in each layer and one lowest resolution subimage on the top. The pyramid structure of the 2D wavelet decomposition with resolution depth 3 is depicted in Fig. 5,

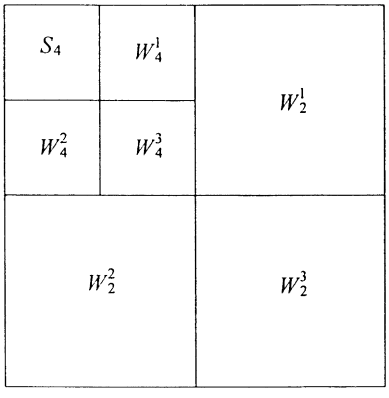


Fig. 3. Frequency band distribution of wavelet decomposition.

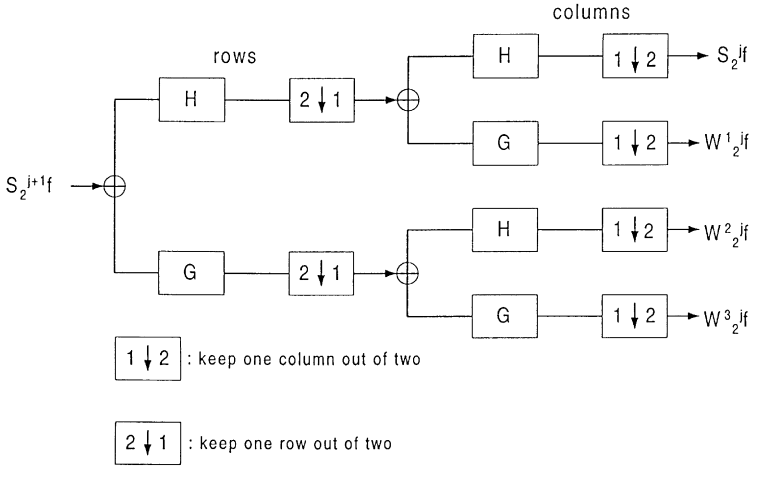


Fig. 2. Wavelet decomposition of an image $S_{2^{j+1}}f$ into $S_{2^j}f$, $W^1_{2^j}f$, $W^2_{2^j}f$ and $W^3_{2^j}f$.

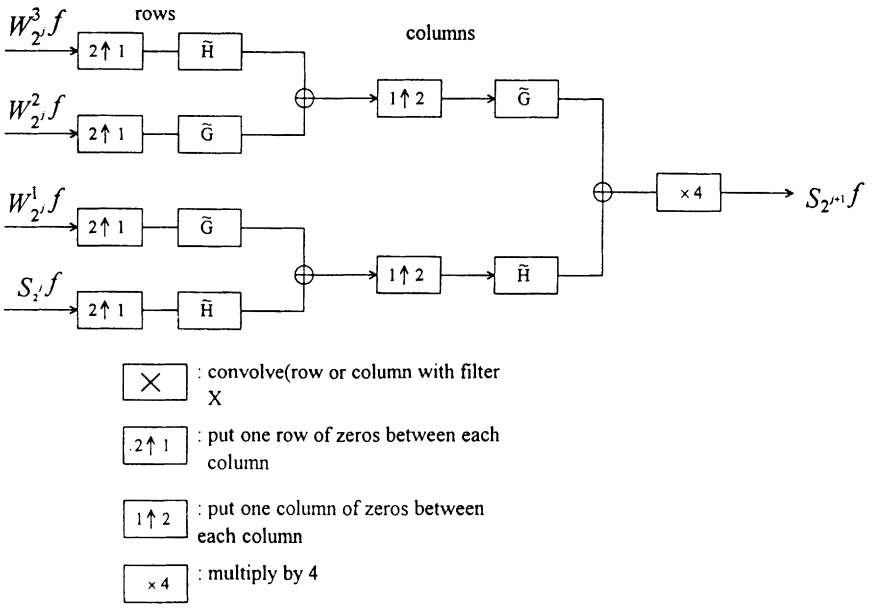


Fig. 4. Wavelet reconstruction of an image $S_{2^{j+1}}f$ from $S_{2^j}f$, $W_{2^j}^1 f$, $W_{2^j}^2 f$ and $W_{2^j}^3 f$.

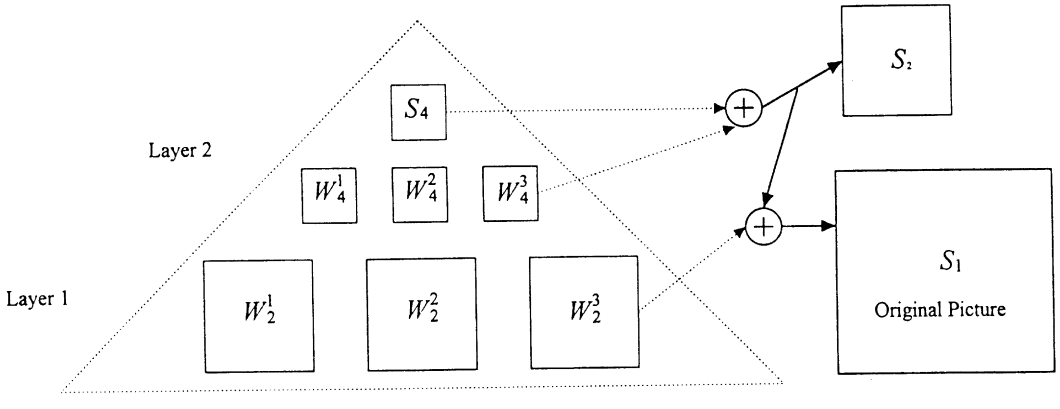


Fig. 5. The pyramid structure of wavelet decomposition and reconstruction.

which consists of a total of seven subimages $\{S_4, W_4^1, W_4^2, W_4^3, W_2^1, W_2^2, W_2^3\}$. The resolution decreases by a factor of 4 (2 in the horizontal direction and 2 in the vertical resolution) with each layer increased. After the wavelet decomposition, an image is divided into several layers with different importance. Subimages at different layers correspond to different resolutions and different frequency ranges, which match the frequency-selected properties of human visual system. It is well-known that human viewers are more sensitive to lower frequency than higher frequency image components. Additionally, energies after wavelet decomposition become highly nonuniform. The higher the layer is, the higher the energy becomes. For example, over 80% of the energy is concentrated in the subimage S_4.

For the implementation of mixed-resolution coding, the low-resolution right image can be obtained by performing the wavelet decomposition.

However, in order to further achieve the higher compression ratio, the next section will present a disparity-compensated multiresolution coding scheme which is able to compress the stereo image by aid of low-resolution disparity estimation.

3. Disparity-compensated wavelet multiresolution coding for still stereo image compression

This section presents a disparity-compensated wavelet multiresolution coding scheme for still stereo image in order to achieve the higher compression ratio. The left image is compressed independently of the right image using a combination of discrete wavelet transform (DWT) and vector quantization (VQ) [9,12,15]. Since the disparity can be used to predict the right image of a stereopair from the left image, only disparity has to be transmitted, together with the reconstruction (prediction) error or residual image for the disparity-compensated right image in order to improve the reconstruction image quality. Instead of transmitting the right image directly, this section proposes a new inter-view hybrid DPCM/DWT/SQ scheme to obtain both the disparity and the prediction error (residual) image. Furthermore, by employing the mixed-resolution psychophysical experiments, the disparity is estimated from both the left/right images at a low resolution, and the prediction error image is also obtained using the inter-view DPCM scheme from the low-resolution disparity-compensated right image.

3.1. Intraframe hybrid DWT/VQ scheme for still left images

Fig. 6 shows the basic architecture of the intraframe hybrid DWT/VQ scheme applied to still image compression for $M = 2$. Wavelet transform is made to organize the information of the original image into several resolutions. The image is split into $3 \times M$ detail subimages of wavelet coefficients (resolution level 1 to M) and a subimage at the lowest resolution level M. Since statistics of the subimage at the lowest resolution is similar to the statistics of the original image, a DPCM technique is used to encode it. Only the detail subimages or wavelet coefficients are vector quantized. Antonini et al. [1] have shown that statistics of detail or wavelet coefficient's subimages were modelized by the generalized Gaussian. They use the generalized Gaussian probability density to design the vector quantizer for each detail subimages.

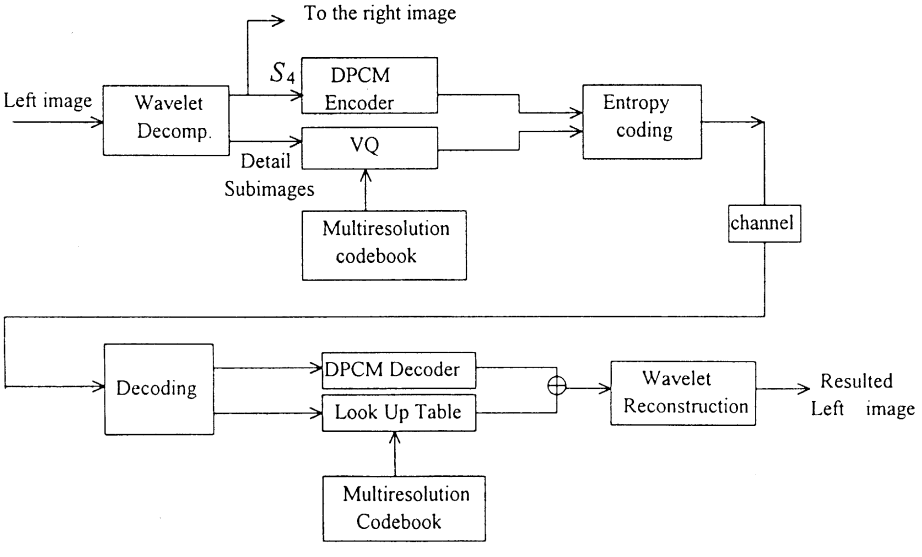


Fig. 6. Interframe hybrid DWT/VQ scheme for still left images.

Vector quantization has proven to be a powerful tool for digital image compression [9,12,15]. The principle involves encoding a sequence of samples (vector) rather than encoding each sample individually. Encoding is performed by approximating the sequence to be coded by a vector belonging to a catalogue of shapes, usually known as a codebook.

Let $\boldsymbol{x}$ be an n-component source vector with joint probability function (pdf) $f_X(\boldsymbol{x}) = f_X(x_1, x_2, \ldots, x_n)$. A vector quantizer VQ of dimension n and size L is defined as a function that maps an arbitrary vector $\boldsymbol{x} \in R^n$ into one of L output or reproduction vectors $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_L$ called codewords belonging to R^n. Thus, we have the chart

$$\boldsymbol{VQ}: R^n \to Y, \tag{21}$$

where $Y = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_L\}$ is the set of reproduction vectors called the codebook and $\boldsymbol{VQ}(\boldsymbol{x}) = \boldsymbol{y}_i$ if $\boldsymbol{x} \in C_i$ where C_i is a region corresponding to $\boldsymbol{y}_i$.

The vector quantizer is completely specified by listing the L codewords $\boldsymbol{y}_i$ and their corresponding nonoverlapping partitions C_i $(i = 1, 2, \ldots, L)$ of R^n called Voronoi regions. A Voronini region is defined by the equation

$$C_i = \{\boldsymbol{x} \in R^n \,|\, \|\boldsymbol{x} - \boldsymbol{y}_i\| \leqslant \|\boldsymbol{x} - \boldsymbol{y}_j\| i \neq j\} \tag{22}$$

and represents the subset of vectors of R^n, which are well matched by the codeword $\boldsymbol{y}_i$ of the codebook. $\|\ \|$ denotes the L^2 norm. The total distortion per dimension of this quantizer is then given by

$$d_n = \frac{1}{n} E\{\|\boldsymbol{x} - \boldsymbol{VQ}(\boldsymbol{x})\|^2\}$$

$$= \frac{1}{n} \sum_{i=1}^{L} \int_{\boldsymbol{x} \in C_i} \|\boldsymbol{x} - \boldsymbol{y}_i\|^2 f_X(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{23}$$

For the purposes of transmission or storage, a binary word c_i of length b_i bits called the index of the codeword is assigned to each reproduction vector $\boldsymbol{y}_i$. Thus, vector quantization can also be seen as a combination of two functions: an encoder that views the input vector $\boldsymbol{x}$ and generates the index of the reproduction vector specified by $\boldsymbol{VQ}(\boldsymbol{x})$ and a decoder uses this index to generate the reproduction vector $\boldsymbol{y}_i$ via the same codebook as the coder. The average binary word length is given by the formula

$$H(\boldsymbol{y}) = -\sum_{i=1}^{L} p_i(\boldsymbol{y}_i) \log_2 p_i(\boldsymbol{y}_i) \quad \text{bits/vector}, \tag{24}$$

which is the so-called entropy measure of the codebook that specifies the minimum bit-rate necessary to achieve a distortion d_n.

The codebook is created and optimized using the well-known Linde–Buzo–Gray (LBG) [9] classification algorithm with a mean square error criterion (23). This algorithm is designed to perform a classification based on a training set which comprises of vectors belonging to different images; it converges iteratively toward a locally optimal codebook. In addition, a globally optimal codebook can be obtained using the simulated annealing techniques [15]. However, global codebook design has drawbacks in that it results in edge smoothing (loss of resolution) and is not easy to take account the properties of human visual system. To tackle this difficulty, a multiresolution codebook [1] has been proposed to preserve edges. The multiresolution codebook contains a number of subcodebooks for wavelet coefficients or detail subimages from resolution level 1 to M. Each subcodebook has been designed for the detail subimages at its corresponding resolution level. The subcodebook has a low distortion level and contains few words, which clearly facilitate the search for the best code vector; the coding computational load is reduced.

In the multiresolution codebook, each detail subimage was associated with one subcodebook that was generated by LGB, and all the reproduction vectors for this subimage were quantized based on this subcodebook. Hence, before applying the Huffman entropy coding, binary code words of equal length were used to represent the quantizer output. Antonini et al. [1] have proposed an optimal bit allocation method, which takes into consideration the fact that the human eye is not equally sensitive to subimages at all spatial frequencies. According to the contrast sensitivity data collected by Campbell and Robson [3], a controlled degree of noise shaping across the subimage $W_2^{j_m}$ is defined as follows:

$$B_{m,j} = r^m \log(\sigma_{m,d}^{2\beta_{m,j}}), \quad 1 \leqslant m \leqslant M,\ 1 \leqslant j \leqslant 3, \tag{25}$$

where $\sigma_{m,j}$ is the standard deviation corresponding to $W_{2^m}^j$ and the values of r and $\beta_{m,j}$ are chosen experimentally in order to match human vision. To apply noise shaping across the VQ subimages, a total distortion of the image for a total target bit-rate R_T is given by

$$D_T(R_T) = \frac{1}{2^{2M}} D_{\mathrm{SQ+DPCM}}^M(R_{\mathrm{SQ+DPCM}}^M) + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{j=1}^{3} D_{m,j}(R_{m,j}) \times B_{m,j}, \quad (26)$$

where $D_{\mathrm{SQ+DPCM}}^M(R_{\mathrm{SQ+DPCM}}^M)$ corresponds to the DPCM MSE distortion with a scalar quantizer of $R_{\mathrm{SQ+DPCM}}^M$ bits per pixel and $D_{m,j}(R_{m,j})$ denotes the MSE distortion in the VQ coding of the subimage $W_{2^m}^j$ for $R_{m,j}$ bits per pixel. Note that $D_{\mathrm{SQ+DPCM}}^M = G_{\mathrm{p}}^{-1} D_{\mathrm{SQ}}^M$ where G_{p} denotes the prediction gain and D_{SQ}^M denotes the PCM MSE distortion. Therefore, the problem of finding an optimal bit assignment for each subimage vector quantizer is formulated as

$$\min_{R_{m,j}} D_T(R_T) \quad (27)$$

$$\text{subject to } R_T = \frac{1}{2^{2M}} R_{\mathrm{SQ+DPCM}}^M + \sum_{m=1}^{M} \frac{1}{2^{2m}} \sum_{j=1}^{3} R_{m,j}. \quad (28)$$

This minimization problem can be solved using Lagrangian multipliers. The optimal bit allocation for $M = 2$ is carried out according to the suggestion of Antonini et al. [1]. The resulted bit assignment is illustrated in Fig. 7. Subimage W_2^3 (diagonal orientation) is discarded. Subimages W_2^1, W_2^2 and W_4^3 are coded using 256-vector subcodebooks (codeword size 4×4) resulting in a 0.5 b/pixel rate, while subimages W_4^1 and W_4^2 are coded at a 2-b/pixel rate using 256-vector subcodebooks (codeword size 2×2). Finally, the lowest resolution is DPCM coded at 8 b/pixel.

3.2. Inter-view hybrid DPCM/DWT/SQ scheme

The aim of disparity estimation is the matching of corresponding picture elements in the simultaneous two-dimensional (2D) pictures of the same 3D scene, viewed under different perspective angles. Two of those pictures may be the left and right

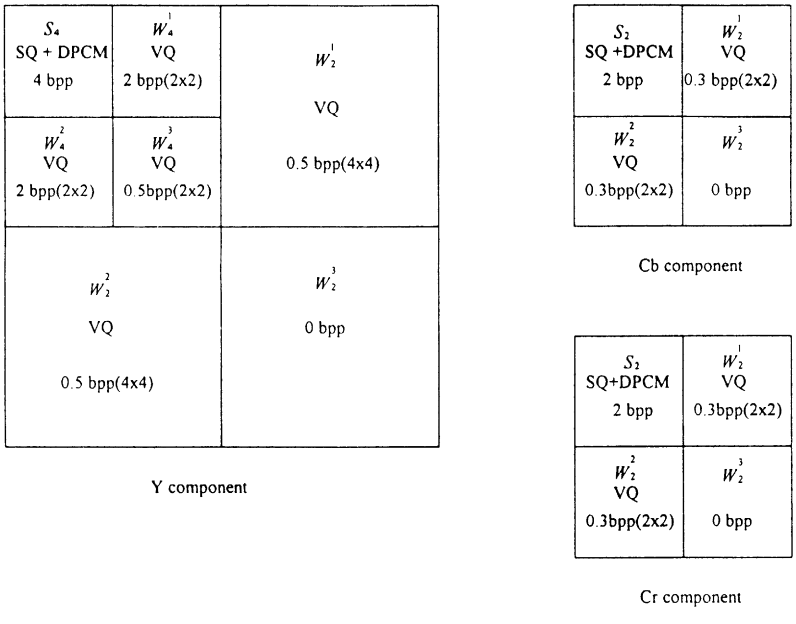


Fig. 7. Bit-rate allocation for subimages in the intraframe hybrid DWT/VQ scheme for still color left images.

views of a stereopair. A number of block-based disparity-compensated methods have been proposed for coding stereopairs [17,19,20]. With block methods, it is assumed that the disparity between left image and right image is constant within a small two-dimensional block B_d of pels. Therefore, the disparity $\boldsymbol{d}$ can be estimated by minimizing the L_1 norm of disparity prediction error such as

$$\mathrm{DPE}(\boldsymbol{d}) = \sum_{z \in B_d} |I_{\mathrm{L}}(z) - I_{\mathrm{R}}(z - \boldsymbol{d})|. \tag{29}$$

Consider a block of $P_d \times Q_d$ pels centered around pel z_0 in the left image (reference image). Assume that a maximum horizontal and vertical disparity are p_d pels and q_d pels, respectively. Thus, the search region in the right image would be an area containing $(P_d + 2p_d)(Q_d + 2q_d)$ pels. A simplified version of the criterion of (29) is given by

$$\mathrm{DPE}(z_0, x, y) = \frac{1}{P_d Q_d} \sum_{|p| \leqslant P_d/2} \sum_{|q| \leqslant Q_d/2} |I_{\mathrm{L}}(z_1 + p, z_2 + q) - I_{\mathrm{R}}(z_1 + p + x, z_2 + q + y)|, \tag{30}$$

where $-p_d \leqslant x \leqslant p_d$, $-q_d \leqslant y \leqslant q_d$, and z_1 and z_2 are the x- and y-coordinates of z_0, respectively.

The minimization of the disparity prediction error of (30) is performed using any one of the four promising methods used in motion estimation [13]. There are (i) full search, (ii) 2D-logarithmic search, (iii) three-step search, and (iv) modified conjugate direction. According to the mixed-resolution psychophysical experiments, the block-based disparity estimation is performed at low resolution. The computational complexity of the low-resolution disparity estimation becomes smaller due to the smaller search area at the lowest resolution. At the receiver, the low-resolution right subimage is estimated using the disparity from the low-resolution left subimage. A full-size reconstruction is obtained by upsampling a factor of 4 and reconstructing with the synthesis low pass filter. However, in order to improve the reconstruction quality of low-resolution right image, disparity has to be transmitted, together with the reconstruction error. In this section, we present an interview hybrid DPCM/DWT/SQ scheme (shown in Fig. 8) to both improve the reconstruction image quality and achieve the higher compression ratio.

An original right image S is first decomposed into a number of subimages. $\{S_{2^M}, W_{2^m}^j;\ m = 1, \ldots, M, j = 1, 2, 3\}$. Only the lowest resolution

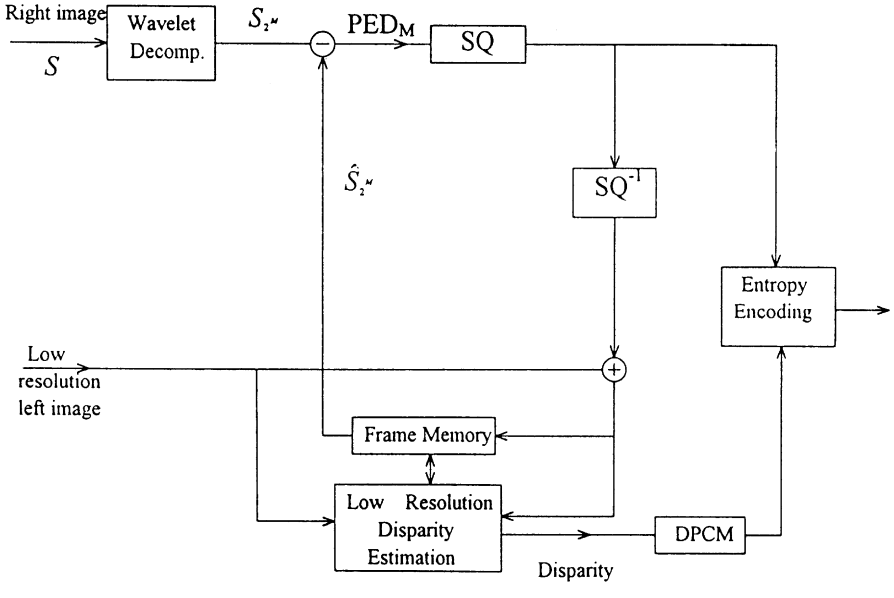


Fig. 8. Inter-view hybrid DPCM/DWT/SQ scheme.

subimage S_{2^M} is considered in our system. After using the block-based disparity compensation scheme, the disparity-compensated right subimage $\hat{S}_{2^M}$ is estimated from its corresponding low-resolution left images using the low-resolution disparity. Then the lowest resolution prediction error subimages PED_M for the disparity-compensated subimage $\hat{S}_{2^M}$ is formed using the disparity compensation. Since all the pixels in PED_M are uncorrelated, it is suggested that the lowest resolution reconstruction error subimage PED_M is coded using an optimal Lloyd Max scalar quantizer instead of DPCM. Furthermore, in order to increase the compression ratio, a DCT-based compression with a DPCM coding for its DC term and a scalar quantizer for its AC terms can be applied to PED_M. However, Zhang and Zafar [20] have shown that DCT-based compression may introduce the undesired blocking artifacts. Thus, in the paper, we consider only the scalar quantizer for PED_M. Finally, disparity vectors are DPCM-coded, and all the quantities are entropy-coded prior to the transmission according to a variable length code table which is quite similar to Table B.10 for motion vectors in MPEG-2 standard [6].

4. Joint motion/disparity-compensated wavelet multiresolution coding for stereo image sequence compression

Stereo image sequence processing requires estimation of the displacements created by the motion of objects and, also, by the disparity between two views of the 3D scene projected on the two images. The estimation of motion and disparity displacements may be performed separately. Motion estimation problem is similar to that of disparity estimation. The associated motion prediction error for the left (reference) image sequence is given by

$$\mathrm{MPE}(\boldsymbol{v}) = \sum_{\boldsymbol{z}\in B_m} |I_{\mathrm{L},k}(\boldsymbol{z}) - I_{\mathrm{L},k-1}(\boldsymbol{z}-\boldsymbol{v})|, \tag{31}$$

where $I_{\mathrm{L},k}(\boldsymbol{z})$ denotes the video frame k at $\boldsymbol{z} = (x, y)$ for left image sequence, $\boldsymbol{v}$ represents the motion vector, and B_m is the 2D block for motion estimation of size $P_m \times Q_m$ pels. Assume that the maximum horizontal and vertical displacements are p_m and q_m, respectively. Thus, the search area in the previous frame is $(P_m + 2p_m)(Q_m + 2q_m)$. Similarly, the simplified version of (34) can be written as

$$\mathrm{MPE}(\boldsymbol{z}_0, x, y) = \frac{1}{P_m Q_m} \sum_{|p|\leqslant P_m/2} \sum_{|q|\leqslant Q_m/2} |I_{\mathrm{L},k}(z_1 + p, z_2 + q) - I_{\mathrm{L},k-1}(z_1 + p + x, z_2 + q + y)|, \tag{32}$$

where $-p_m \leqslant x \leqslant p_m$, $-q_m \leqslant y \leqslant q_m$ and $\boldsymbol{z}_0 = (z_1, z_2)$.

Similarly, the motion vector can be found by applying any one of the promising search methods [13] to the minimization of (32). In order to further reduce both the computational complexity and searching time of above four methods, an efficient hierarchical block matching algorithm [17,20] has been applied to both motion and disparity estimation, in which agreement of large block is first attained and block size is subsequently and progressively decreased. One approach to hierarchical block matching uses multiple resolution versions of the image, and variable block-size at each level of the pyramid [17,20]. In a multiresolution motion estimation (MRME) scheme, the motion vector field is first calculated for the lowest resolution subimage, which sits on the top of the pyramid [17,20]. Motion vectors at lower layers, of the pyramid are refined using the motion information obtained at higher layers and again propagated to the next pyramid level until the high-resolution level is reached. The motivation for using the MRME approach is the inherent structure of the wavelet representation. MRME schemes significantly reduce the searching and matching time and provide a smooth motion vector field.

A video frame is decomposed up to the two levels. A total of seven subimages are obtained with three subimages at the first level, and four on the top level including subimage S_4 with the lowest frequency band. It is well known that human vision is more perceptible to errors in low frequencies than those incurred in higher bands to be selective in spatial orientation and position, e.g., errors in smooth areas are more disturbing to a viewer than

those near edges. The subimage S_4 contains a large percent of the total energy though it is only 1/16th of the original video frame size. Additionally, errors in higher layer subimages will be propagated and expanded to all subsequent lower layer subimages. To tackle this difficulty, Zhang and Zafar [20] have proposed a variable block-size MRME scheme to take all these factors into considerations.

Assume that b_M is the size of the block used at the lowest resolution or highest layer M. The size of the block varies with the resolution level; at level m it is $b_M \times 2^{M-m}$; then at level $m+1$ the block becomes $b_M \times 2^{M-m-1}$ in pixels of the level. If $\boldsymbol{v}_{i,2^{m+1}}(x,y)$ represents the motion vector centered at (x,y) for the subimage $W^i_{2^{m+1}}$ belonging to the $(m+1)$th layer, the initial estimation $\boldsymbol{v}^{(0)}_{i,2^m}$ of the motion vector for the same block at the mth level of the pyramid will be

$$\boldsymbol{v}^{(0)}_{i,2^m}(x,y) = 2\boldsymbol{v}_{i,2^{m+1}}(x,y), \quad i = 1,2,3. \tag{33}$$

At the mth pyramid level, a full search is used with a small search area about the displaced position $\boldsymbol{z}_0 + \boldsymbol{v}^{(0)}_{i,2^m}(x,y)$. In other words, the motion vectors at level m are given by

$$\begin{aligned}\boldsymbol{v}_{i,2^m}(x,y) &= 2\boldsymbol{v}_{i,2^{m+1}}(x,y) + (\Delta x, \Delta y)\\ &= 2^{M-m}\boldsymbol{v}_{0,2^M}(x,y) + (\Delta x, \Delta y),\end{aligned} \tag{34}$$

where $\boldsymbol{v}_{0,2^M}(x,y)$ denotes the motion vector for the subimage S_{2^M} and $(\Delta x, \Delta y)$ is the incremental motion vector found by a full search with reduced search area.

Fig. 9 shows an example of the proposed variable block-size MRME scheme. First, the motion vector $\boldsymbol{v}_{0,4}$ for the highest layer subimage S_4 are estimated by full search with block size (b_4) of 2×2. These motion vectors are then scaled appropriately to be used as initial estimates for motion estimation in higher resolution subimages. An alternative approach to find the incremental motion vector at level m is to use 2^{M-m} times the motion vector for level M as an initial estimation of a full search with a relatively small search area.

Similarly, Tzovaras et al. [17] have shown that the above variable block-size multiresolution block matching techniques are also valid for disparity estimation in order to reduce the amount of processing time. This particular disparity estimation

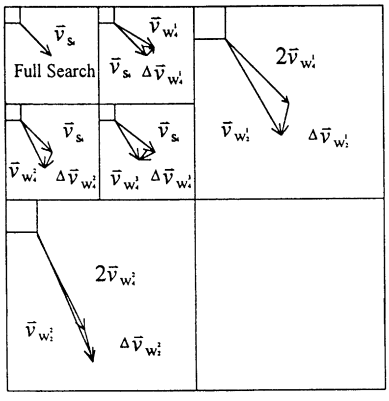


Fig. 9. Variable block-size multiresolution motion estimation (MRME) of the left image sequence.

scheme is called the variable block-size multiresolution disparity estimation (MRDE). The main difference between MRME and MRDE is that the disparity prediction error of MRDE is given by

$$\begin{aligned}&\mathrm{DPE}(\boldsymbol{z}_0, x, y) =\\ &\frac{1}{P_d Q_d}\sum_{|p|\le P_d/2}\sum_{|q|\le Q_d/2}|I_{\mathrm{L},k}(z_1+p, z_2+q)\\ &\quad - I_{\mathrm{R},k}(z_1+p+x, z_2+q+y)|,\end{aligned} \tag{35}$$

where $I_{\mathrm{L},k}(\boldsymbol{z})$ and $I_{\mathrm{R},k}(\boldsymbol{z})$ are the kth left and right video frames at position $\boldsymbol{z}$. Note that the MRDE is performed for both the left and right frames at the same time instant.

Similarly, according to the mixed-resolution psychophysical experiments, this section presents a low-resolution MRDE scheme for disparity estimation and compensation, in order to improve the computational efficiency of estimation. The low-resolution MRDE scheme illustrated in Fig. 10 requires a two-level wavelet decomposition. The first-level wavelet decomposition is used to obtain both the lowest resolution left and right video frames. The second-level wavelet decomposition is conducted to perform the MRDE at the lowest resolution. Fig. 11 shows the block flow diagram of the joint motion/disparity-compensated wavelet multiresolution coding for stereo image sequence. However, the quality of the reconstructed stereo image sequence may be degraded using only the

motion and disparity vectors for some cases. In order to improve the reconstruction image quality, the motion/disparity vectors have to be transmitted together with prediction errors. The next two sections will propose two novel schemes to achieve the goal.

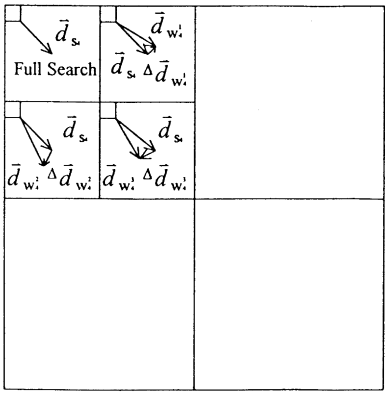


Fig. 10. Variable block-size multiresolution disparity estimation (MRDE) of the low resolution right/left image sequences.

4.1. Interframe hybrid DPCM/DWT/VQ scheme for left image sequence using a variable block-size MRME scheme

The left image stream is compressed using the hybrid DPCM/DWT/VQ scheme illustrated in Fig. 12 independent of the right image stream. Using wavelet decomposition, an original image S is first decomposed into subimages $\{S_{2^M}, W^j_{2^m}; m = 1, \ldots, M, j = 1, 2, 3\}$. After using the block-based motion compensation, the prediction error subimages, $\{\mathrm{PEM}_M, \mathrm{PEM}^j_m; m = 1, \ldots, M, j = 1, 2, 3\}$ for the motion-compensated subimages $\{\hat{S}_{2^M}, \hat{W}^j_{2^m}; m = 1, \ldots, M, j = 1, 2, 3\}$ are obtained. The lowest resolution prediction error subimage PEM_M is PCM-coded using an optimum scalar quantizer instead of DPCM since the pixels in PEM_M are uncorrelated. However, the prediction error

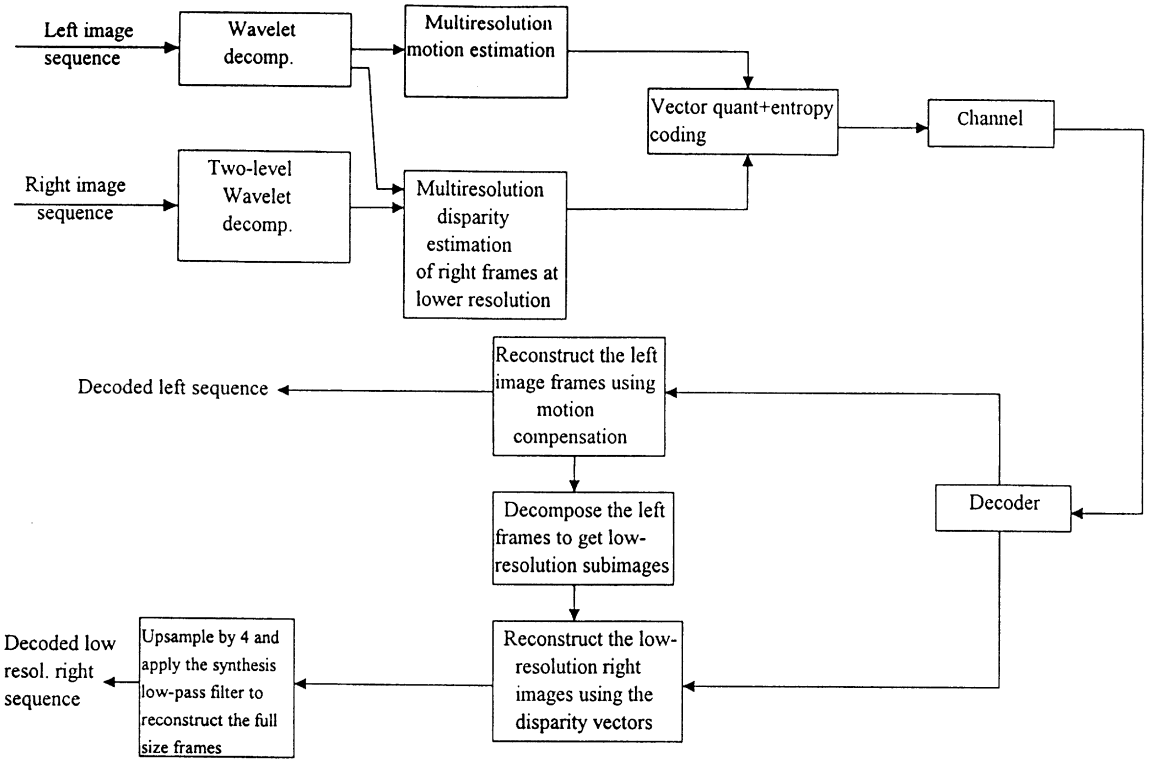


Fig. 11. Encode-decoder structure for the joint motion/disparity-compensated stereoscopic image sequence compression scheme.

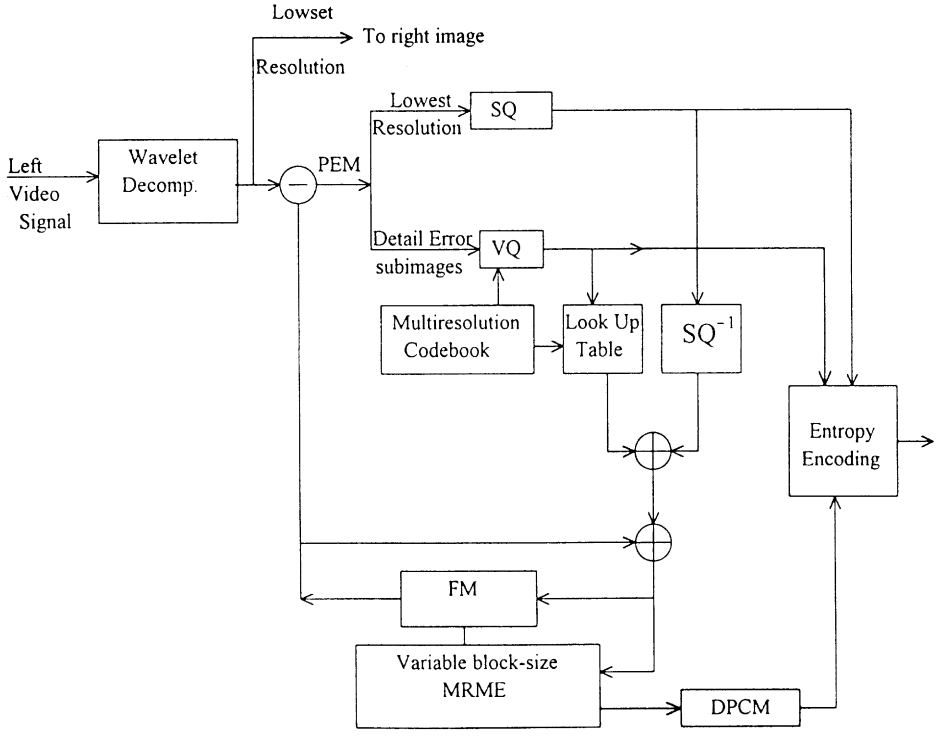


Fig. 12. Interframe hybrid DPCM/DWT/VQ scheme for the left image sequence using a variable block-size MRME.

subimages PEM_m^j are vector quantized using a multiresolution codebook, where each error image has its corresponding subcodebook. Fig. 13 shows the bit assignment for PEM_M and PEM_m^j, $j = 1, 2, 3$; $1 \leqslant m \leqslant M$ and $M = 2$. The motion vectors are DPCM-coded, and all quantities are then entropy-coded prior to the transmission. Furthermore, the variable block-size MRME scheme is applied to perform the motion estimation in order to improve the computational efficiency of the proposed interframe compression scheme. For color video frames, the motion vectors obtained from the luminance component of the color video frames are used as the motion vectors for both the Cr and Cb chrominance components. Since the luminance component contains more than 60% of the total energy of the original image and Cr- and Cb-chrominance components have less than 20% of the total energy, the size of Cr- or Cb-component is 1/4th of the original video frame size. Thus, each chrominance component is decomposed into four subimages. Fig. 13(b,c) shows the bit assignment for each chrominance component.

4.2. Inter-view hybrid DPCM/DWT/VQ scheme for low-resolution right image sequence using a variable block-size MRDE technique

All the low-resolution right video frames can be estimated from their corresponding left video frames using the low-resolution variable block-size MRDE procedure described above. In order to improve the image quality of the reconstructed right video frames, both the disparity and reconstruction (prediction) errors should be transmitted. We use the similar architecture of inter-view hybrid DPCM/DWT/SQ scheme proposed in Section 3.2 to implement this concept. Fig. 14 shows its basic structure. The main difference between Figs. 8 and 14 is that Fig. 14 contains a two-level wavelet decomposition and a variable block-size multiresolution disparity estimation. Both the

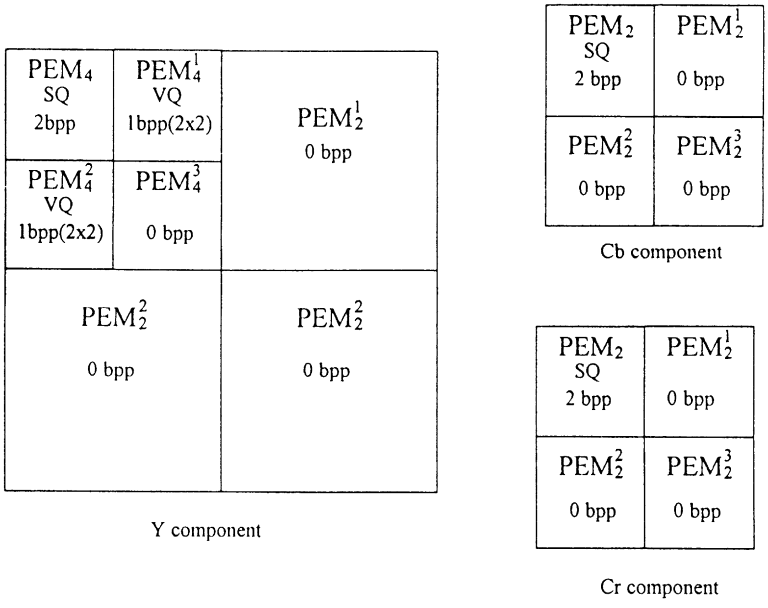


Fig. 13. Bit-rate allocation for prediction error subimages, PEM in the interframe hybrid DPCM/DWT/VQ scheme using MRME.

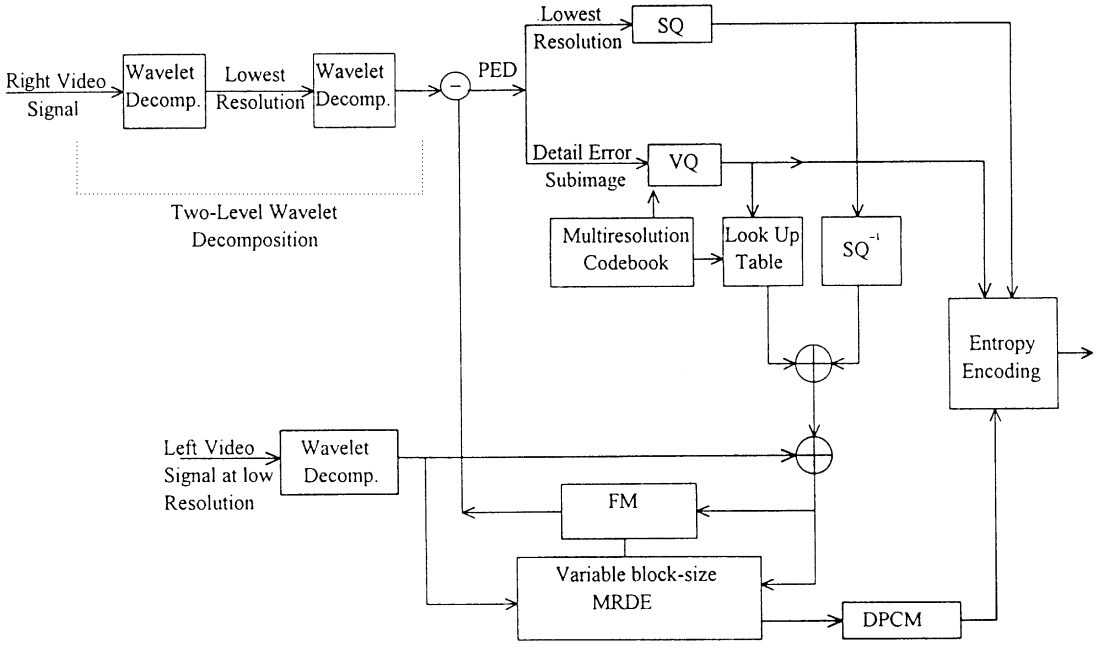


Fig. 14. Inter-view hybrid DPCM/DWT/VQ scheme for low-resolution right and left image sequences using MRDE.

lowest resolution left and right video frames are obtained by the first-level wavelet decomposition. The resulting low-resolution left and right video frames are again decomposed into the second-level lowest resolution subimage and the wavelets in different scales and resolutions. The second-level prediction error subimages $\{\mathrm{PED}_M, \mathrm{PED}_m^j;\ m = 1, \ldots, M,\ j = 1, 2, 3\}$ for low-resolution disparity-compensated subimages are formed using the low-resolution variable block-size multiresolution disparity compensation (MRDE) scheme. Like the motion-compensated interframe compression scheme, the PED_M is coded using an optimum scalar quantizer, and the subimages PED_m^j are coded using multiresolution vector quantizer. Fig. 15(a) shows the bit assignment for PED_M and PED_M^j. For color stereo video frames, the disparity vectors obtained from their luminance components are used as the disparity vectors for their Cr- and Cb-chrominance components. Fig. 15(b,c) shows the bit assignment for the Cr- and Cb-chrominance components, respectively. Certainly, the proposed interview DPCM/DWT/VQ scheme with variable block-size MRDE can be directly applied to still stereo image instead of interview hybrid DPCM/DWT/SQ scheme in order to improve its computational efficiency.

5. Simulation results

To examine the performance of the proposed compression methods, a still stereoscopic color test image "Achoo" with both the left and right pictures illustrated in Fig. 16 is considered in our system. Each picture is a 640×480 color image and eight bits for each of R-, G- and B-color components. Since the RGB color components are strongly correlated, the encoding of a color image in the RGB domain is not very efficient in the image compression. By performing a transformation of the RGB signals to the Y, Cr, Cb domain, it produces nearly decorrelated components and most of the signal energy is contained within the luminance (Y) signal. For simplicity, both the left and right pictures are decomposed to their lowest resolution subimages and detail subimages using Daubechie's W_6 wavelet [4]. According to the mixed-resolution psychophysical experiments, the low-resolution disparity can be obtained from the luminance

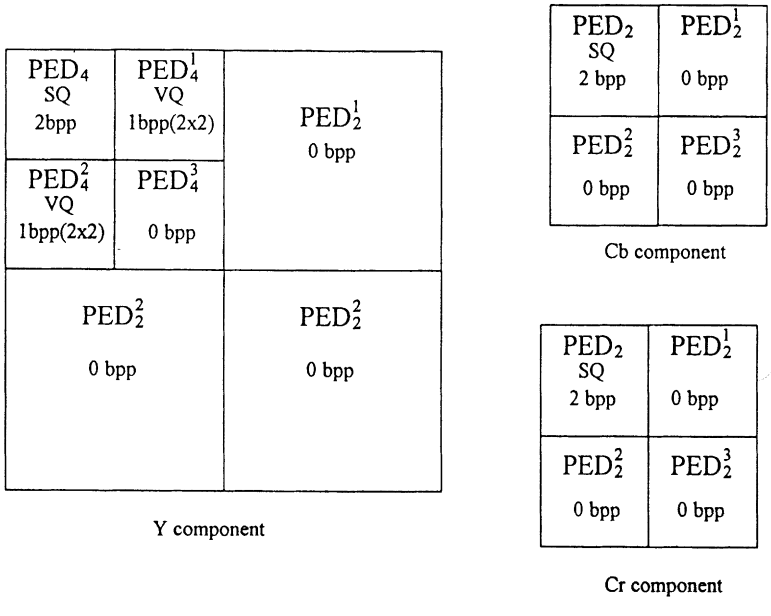


Fig. 15. Bit-rate allocation for prediction error subimages, PED, in the inter-view hybrid DPCM/DWT/VQ scheme using MRDE.

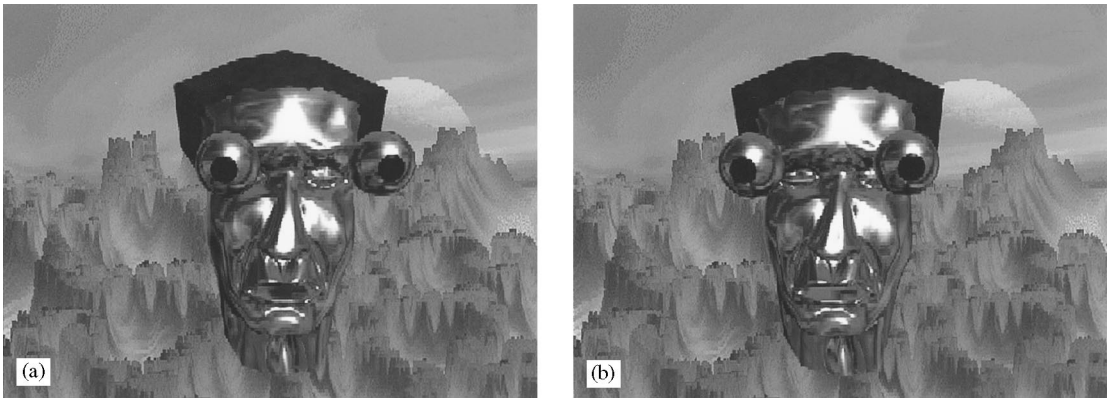


Fig. 16. Still color stereoscopic image "Achoo". (a) Left image; (b) right image.

Fig. 17. Wavelet decomposition of the right picture of Achoo.

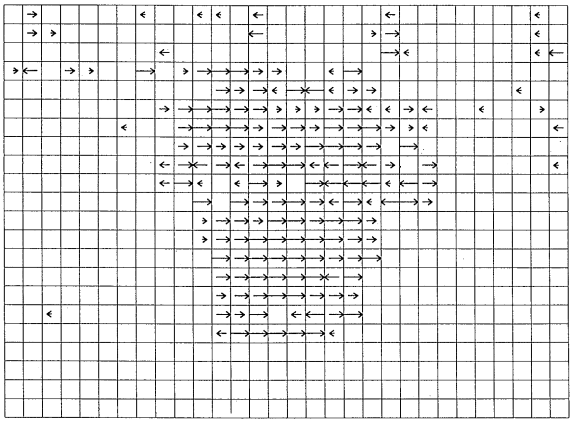


Fig. 18. Low-resolution disparity field for still stereoscopic image "Achoo".

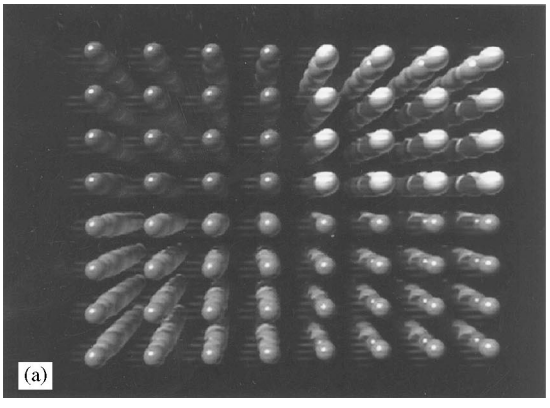
components of these two resulted lowest resolution pictures. However, in order to improve the computation efficiency of disparity estimation, a variable block-size MRDE scheme is applied to these two lowest resolution pictures. However, it requires the second-level wavelet decomposition for both the lowest resolution pictures. Fig. 17 shows the image decomposition of the right picture after performing a two-level wavelet decomposition. The resulting low-resolution disparity field is illustrated in Fig. 18. For the left picture, its lowest resolution subimage is DPCM-encoded, and its three detail subimages are vector quantized according to the bit assignment of Fig. 7. On the other hand, an inter-view hybrid DPCM/DWT/VQ scheme using MRDE is applied to the right picture instead of the original inter-view hybrid DPCM/DWT/SQ scheme. The proposed compression techniques are also valid for Cr- and Cb-chrominance components except for that their disparities are obtained from Y-component. Finally, all quantities are entropy-coded.

Throughout this paper, there are two measures used to evaluate the compression performance, i.e., compression ratio and peak signal-to-noise ratios (PSNR) for Y-, Cr- and Cb-components. The compression ratio is defined by

$$\mathrm{CR} = \frac{(\text{the number of bits in the original image})}{(\text{the number of bits in the compressed image})}. \tag{36}$$

Fig. 19. Left image reconstruction.

Fig. 20. Low-resolution right image reconstruction with up-sampling.



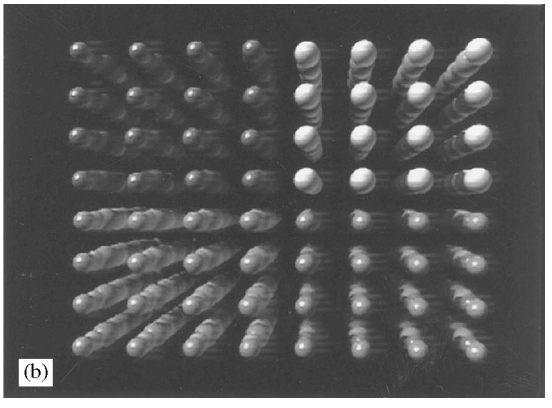


Fig. 21. The 8th stereoscopic image frame of Spiral Ball. (a) Left image frame; (b) right image frame.

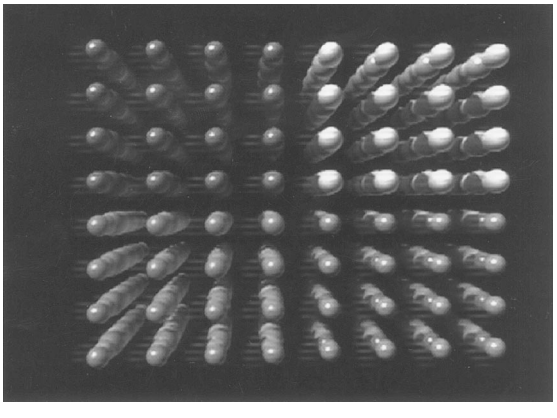


Fig. 22. Left image frame reconstruction.

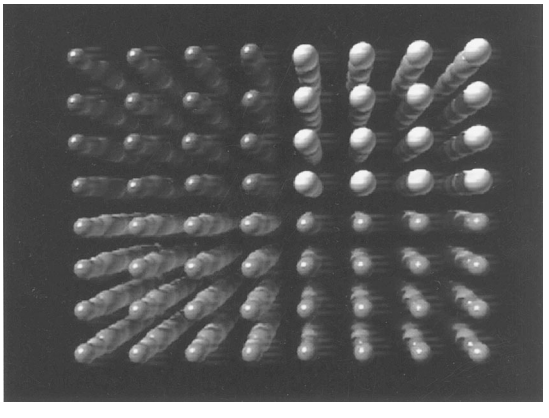


Fig. 23. Low-resolution right image frame reconstruction.

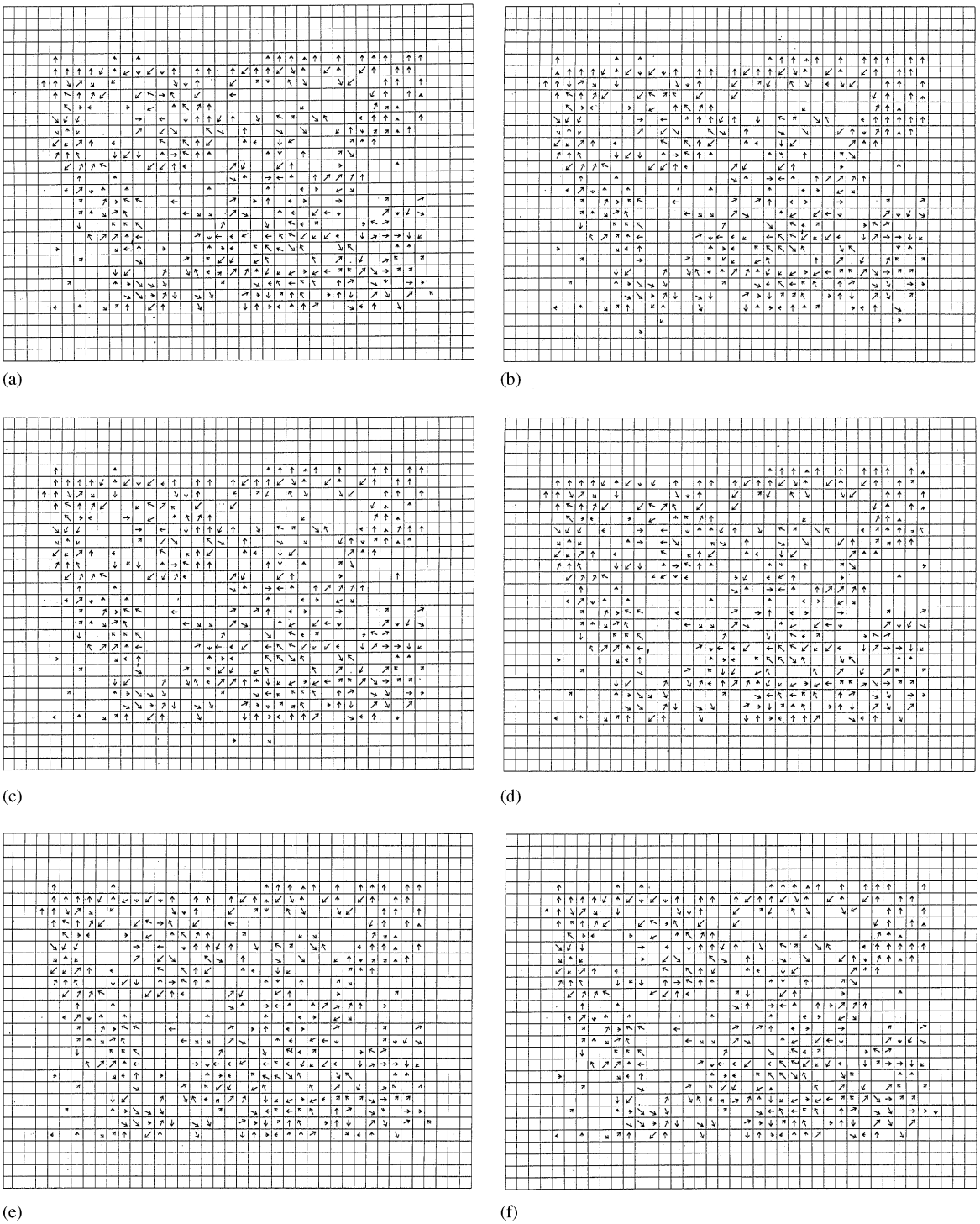


Fig. 24. Motion vector field obtained from the 8th and 9th left image frames. (a) Lowest resolution subimages S_4's; (b) wavelet subimages W_4^1's; (c) wavelet subimages W_4^2's; (d) wavelet subimages W_4^3's; (e) wavelet subimages W_2^1's; (f) wavelet subimages W_2^2's.

The peak signal-to-noise ratio in decibels (dB) is computed as

$$\mathrm{PSNR}_X = 20\log_{10}\frac{I_{X,\mathrm{pp}}}{\mathrm{RMSE}_X}, \quad X = \mathrm{Y}, \mathrm{Cr}, \mathrm{Cb}, \quad (37)$$

where $I_{X,\mathrm{pp}}$ denotes the peak-to-peak value for the X-component of input color image, and RMSE_X is the root mean-squared reconstruction error between the X-components of the input and reconstructed color images, where X denotes each of Y-, Cr-, Cb-components. It should be noted that PSNR_X for the low-resolution right picture are obtained at the low-resolution level.

By performing our still stereoscopic color image compression technique, the compression ratio for the color stereo image Achoo is found to be 42.3, where the total number of bits in the original image and in the compressed image are 1.0368×10^6 and 2.451×10^4 bytes, respectively. Fig. 19 shows the reconstructed left image. Its signal-to-noise ratios are $\mathrm{PSNR}_{\mathrm{Y,L}} = 38.52\,\mathrm{dB}$, $\mathrm{PSNR}_{\mathrm{Cr,L}} = 34.02\,\mathrm{dB}$ and $\mathrm{PSNR}_{\mathrm{Cb,L}} = 33.44\,\mathrm{dB}$. Fig. 20 shows a full-size reconstructed right image which is obtained by upsampling the resulting low-resolution right image. Its signal-to-noise ratios in the low-resolution manner are $\mathrm{PSNR}_{\mathrm{Y,R}} = 39.72\,\mathrm{dB}$, $\mathrm{PSNR}_{\mathrm{Cr,R}} = 38.80\,\mathrm{dB}$ and $\mathrm{PSNR}_{\mathrm{Cb,R}} = 38.22\,\mathrm{dB}$. However, its

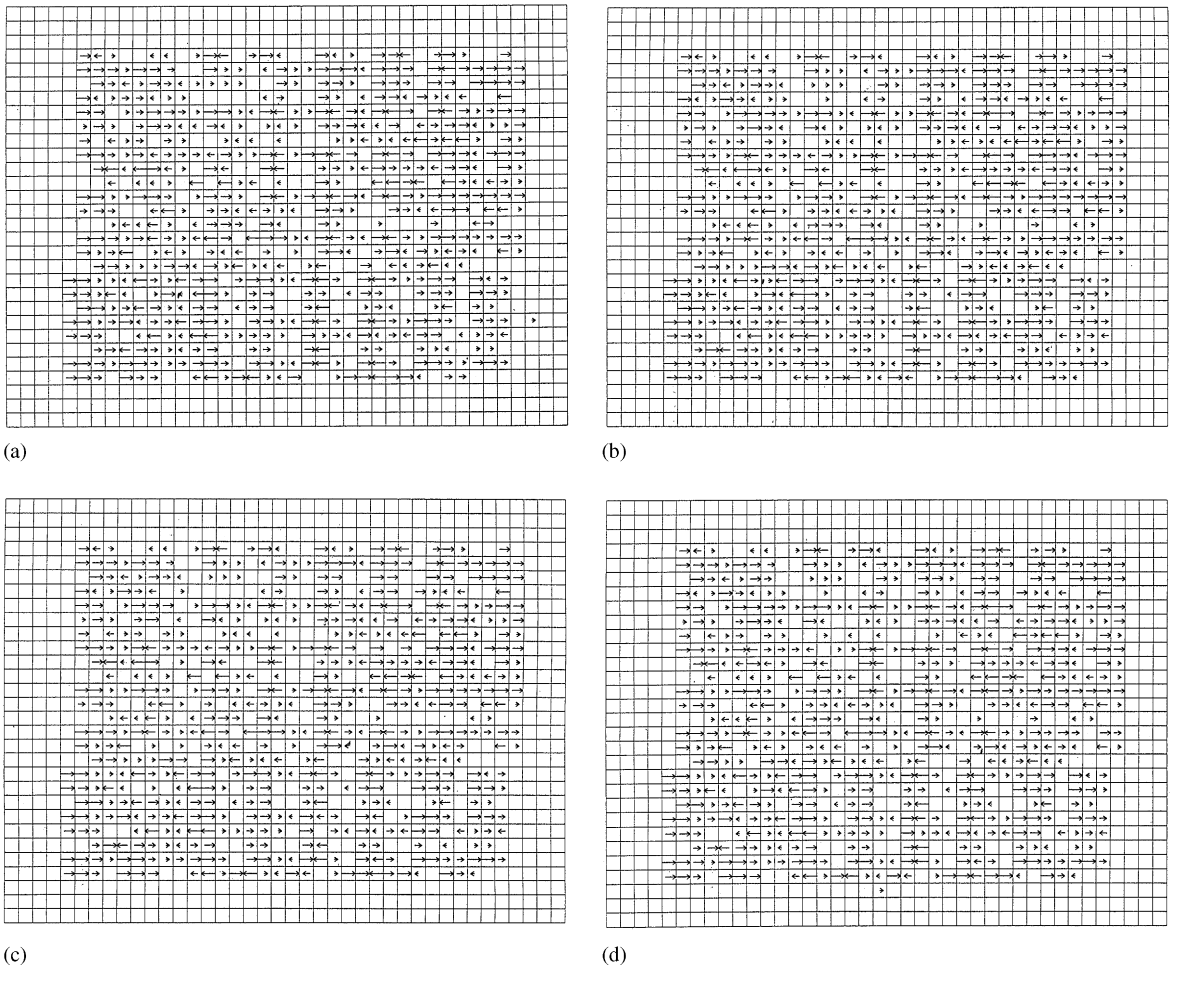


Fig. 25. Low-resolution disparity field obtained from both the 8th low-resolution left and right image frames. (a) Lowest resolution subimages S_4's; (b) wavelet subimages W_4^1's; (c) wavelet subimages W_4^2's; (d) wavelet subimages W_4^3's.

PSNRs become worst, i.e., $PSNR_{Y,R} = 34.62\,dB$, $PSNR_{Cr,R} = 32.45\,dB$ and $PSNR_{Cb,R} = 31.73\,dB$, when these PSNRs are evaluated in the same resolution level as the left images.

Next, it is of interest to conduct a similar study for a color stereoscopic image sequences. The proposed joint motion/disparity-compensated compression techniques were tested on the first 20 fields of the color (24 bits/pixel) stereo image sequence "Spiral Ball" (size 640×480), in the 3D library (delic side) of the 3D museum, constructed by Multimedia Creators Network of Pioneer Electronic Corp. The size of the block used at the lowest resolution level for both the MRME and MRDE schemes is set to 2×2. The left and right images of the 8th Spiral Ball stereo image frame are illustrated in Fig. 21. Figs. 22 and 23 show the left image reconstruction by interframe DPCM/DWT/VQ decoder and the right image reconstruction by interview DPCM/DWT/VQ decoder, respectively. Fig. 24 depicts the motion vector field obtained from both the 8th and 9th left image frames using the variable block-size MRME technique. On the other hand, the disparity field obtained from both the 8th low-resolution left and right image frames is illustrated in Fig. 25 using the variable block-size MRDE technique. In Fig. 26, 20 fields of the left color image sequence including Y-, Cr- and Cb-components are given, with the respective PSNR per field used in the proposed codec. In Fig. 27, the same results are shown from the low-resolution right image sequence. It is notable that the PSNR for each component of both the left and right image frames is greater than 33 dB. Furthermore, the compression ratio of the Spiral Ball image sequence over 20 fields is found to be 91.02, where the total numbers of bits in the original and the compressed image sequence over 20 fields are 3.6864×10^7 and 4.05×10^5 bytes (or 2.025×10^4 bytes/field), respectively.

6. Conclusion

This paper has introduced a joint motion/disparity-compensated wavelet multiresolution coding scheme based on mixed-resolution psychophysical experiments which is capable of achieving

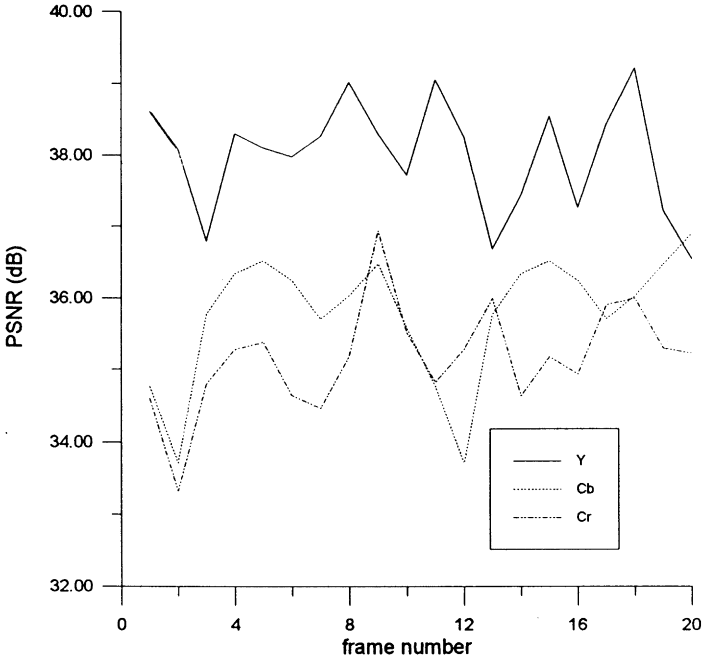


Fig. 26. PSNR for the left color image sequence.

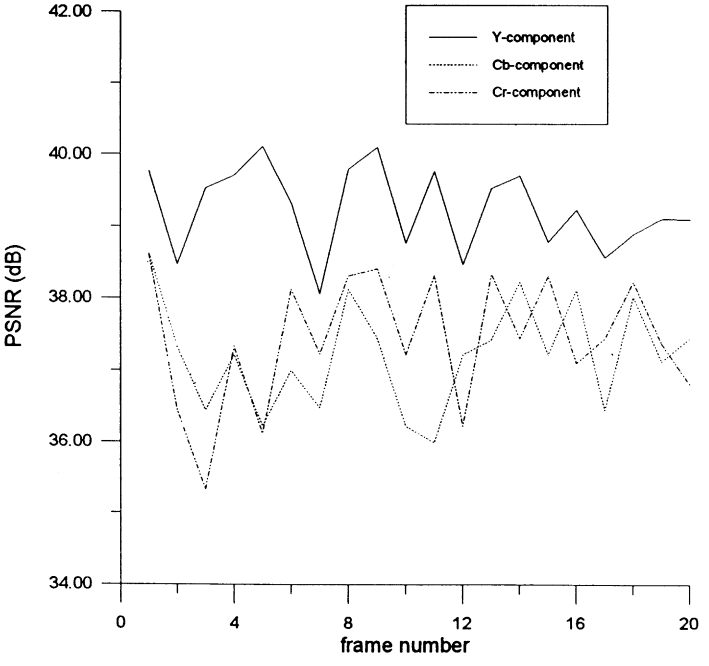


Fig. 27. PSNR for the low-resolution right color image sequence.

the high compression ratio for typical stereoscopic image sequences, without any significant loss in the perceived 3D stereo image quality. A variable block-size multiresolution block matching is applied to perform the estimation of both the motion and disparity vectors in an efficient manner. Both the interframe and interview DPCM/DWT/VQ schemes are able to produce the motion and disparity vectors together with their associated prediction error subimages, in order to both increase the compression ratio and improve its reconstruction image quality. Results show that such schemes can achieve the higher compression ratio of 91.02 for Spiral Ball stereo image sequence.

References

[1] M. Antonini et al., Image coding using wavelet transform, IEEE Trans. Image Process. 1 (2) (April 1992) 205–220.

[2] H. Aydinoglu et al., Compression of multi-view images, in: Proceedings of first IEEE International Conference Image Processing, Austin, TX, USA, November 1994, pp. 385–389.

[3] F.W. Campbell, J.G. Robson, Application of Fourier analysis to the visibility of gratings, J. Phys. 197 (1968) 551–556.

[4] I. Daubechies, Orthonormal bases of compactly supported wavelets, Commun. Pure Appl. Math. XLI (1988) 909–996.

[5] R.E.H. Franich et al., Stereo-enhanced displacement estimation by genetic block matching, SPIE Visual Commun. Image Process. 2094 (1993) 362–371.

[6] Generic Coding of Moving Pictures and Associated Audio Information: Video, ISO/IEC International Standard 13818-2, 1995.

[7] A. Grossmann, J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM J. Math. 15 (1984) 723–736.

[8] S. Lethurman et al., A multiresolution framework for Stereoscopic image sequence compression, in: Proceedings of first IEEE International Conference on Image Processing, Austin, TX, USA, November 1994, pp. 361–365.

[9] Y. Linde, A. Buzo, R.M. Gray, An algorithm for vector quantizer design, IEEE Trans. Commun. 28 (January 1980) 84–95.

[10] S. Mallat, A theory for multiresolution signal decomposition: the wavelet represen, IEEE Trans. Pattern Anal. Mach Intell. 11 (7) (July 1989) 647–693.

[11] T. Mitsuhashi, Subjective image position in Stereoscopic TV systems – consideration on comfortable stereoscopic images, SPIE 2179 (1994) 259–265.

[12] N.M. Nasrabadi, R.A. King, Image coding using vector quantization: a review, IEEE Trans. Commun. 36 (August 1988) 957–971.
[13] A. Netravali, B. Haskell, Digital Pictures – Representation and Compression, Plenum Press, New York, 1988.
[14] M.G. Perkins, Data compression of stereopairs, IEEE Trans. Commun. 40 (4) (April 1992) 684–696.
[15] K. Rose, E. Gurewitz, G.C. Fox, A deterministic encoding approach to clustering, Pattern Recognition Lett. 11 (1990) 589–594.
[16] A. Tamtaoui, C. Labit, Constrained disparity and motion estimation for 3DTV image sequence coding, Signal Processing: Image Communication 4 (1) (November 1991) 45–54.
[17] D. Tzovaras, M.G. Strintzis, H. Sahinoglou, Evaluation of multiresolution block matching techniques for motion and disparity estimation, Signal Processing: Image Communication 6 (1994) 59–67.
[18] G. Westheimer, S.P. Mckee, Stereoscopic acuity with defocused and spatially filtered retinal images, J. Opt. Soc. Amer. 70 (7) (July 1980) 772–778.
[19] S. Zafar, Y.Q. Zhang, B. Jabbari, Multiscale video representation using multiresolution motion compensation and wavelet decomposition, IEEE J. Selected Areas Commun. 11 (1) (January 1993) 24–34.
[20] Y.Q. Zhang, S. Zafar, Motion-compensated wavelet transform coding for color video compression, IEEE Trans. Circuits Systems Video Technol. 2 (3) (September 1992) 285–296.