# USING BINOCULAR ENERGY MODELING FOR STEREOSCOPIC COLOR IMAGE CODING

*Rafik BENSALMA, Chaker LARABI*

XLIM lab., SIC dept., University of Poitiers
Poitiers, FRANCE
Email: {bensalma,larabi}@sic.univ-poitiers.fr

## ABSTRACT

In this paper, we propose a coder for stereoscopic color images based on the binocular properties of the human visual system (HVS). In the preprocessing stage we modeled some properties of the simple and complex cells. These cells characterized by their orientation and amplitude are responsible for binocular fusion; they take as input a set of signals representing the two retinal images and give as output binocular signals. To model this process we have used mathematical functions that have the same characteristics as the simple and complex cells such as wavelet and bandelet. After the matching process, we obtain a residual image, a disparity map and the reference image; these allow to predict the target image. The residual image contains the matching error. The one obtained with our approach contains a very few amount of data generating low bitrate. The experimentation stage showed that our coder gives better results than the two famous coders coming from literature.

## 1. INTRODUCTION

The stereoscopy can be defined as the association of two eyes in the visual analysis of the same region of the scene. it improves significantly this analysis but it comes along, in return, with an increase of the information to be treated and to be stored. The main advantage to have two shots of the same scene is that each of them takes a slightly different view. The space between eyes generates binocular disparities (difference between the retinal images of the left and the right eyes) which are exploited, via the stereoscopic vision, to reconstruct the third dimension from flat retinal images. The matching is a crucial stage, for all stereo applications (stereo coding, stereo quality assessment, ...) because the quality of the 3D reconstruction depends on it.

Stereo coding aims to reduce the size of the couple of images by exploiting the redundancy between them. The most significant research addressing this problem has been carried out during the last decade. Among the first proposed approaches, we find the one introduced by *Dinstein et al.* [1] based on the deletion theory in the human visual system (HVS). The method was subjectively estimated by a panel of observers asked to compare stereo compressed images with original images and then rate them on a discrete scale. *Aydinoglu et al.* [2] proposed an approach based on regions matching. Three kinds of regions are considered: the occulted regions, the redundant regions and the contours. A matching is performed between regions of the same type which is not efficient for the disparity map computation. *Tzovara et al.* [3] proposed a solution based on the object contours. This is by supposing that all pixels in the estimated region have the same disparity. The coordinates of estimated regions must be transmitted with the estimated disparity. This increases the size of transmitted information. Techniques based on the object contours can be generally classified into two distinct categories: shape-oriented or blocks-oriented (blocks can be of variable or fixed size).

*Woo et al.* [4] and *Magnor et al.* [5] proposed approaches based on a hierarchical representation of images. The disadvantage lies in the fact that the disparity is computed for all the scales that do not reduce effectively the size of the compressed images. *Boulgouris et al* [6] proposed an approach in closed circuit where the residual image with its disparity are calculated by using the reference image locally decoded because such an image is available for the decoding stage. For coding residual images, *Frajka et al.* [7] proposed that the occluded and the non occluded regions are coded separately, and for this reason they used a technique containing variable blocks. *Kim et al.* [8] and *Woo et al.* [9] proposed an approach based on blocks of the same size with a directional search of the best match which does not allow to find the best one, but reduces the computation time. *Pagliari et al.* [10] and *Moellenhoff et al.* [11] proposed approaches based on *DCT* transform which do not exploit the spatial distribution of the information and consequently the stereo properties are not exploited. *Aydinoglu et al.* [12] proposed an approach based on a subspace projection that allows the adaptation to the local transform. This approach is not efficient on textured images.

The most important work using wavelet representation has been introduced by *Ellinas et al.* [13]. These latter used the wavelet decomposition for coding residual images. The originality of both approaches is the separated processing of the occulted regions and the rest of regions represented on the residual image. *Nath et al.* [14], *Ellinas et al.* [13] and *Moran et al.* [15] proposed wavelet-based approaches. These approaches present a set of common drawbacks relative to the calculation of the disparity map. Indeed, this map is calculated on each scale of the decomposition which do not reduce efficiently the size of the compressed images.

The major drawback of the previous works is the lack of taking into account the binocular process allowing to generate the 3D image in the HVS. All the coders tries to minimize the residual information, obtained by the subtraction of the similar regions in both images; in that case we can have two similar images allowing a bad depth reconstruction but giving a good PSNR. The second drawback is that the matching of the images, in these coders, is realized without consideration of the behavior of the HVS.

In this work, we propose a stereo coder that takes into account the binocular process performed by the HVS. It con-

 ISSN 2076-1465 

sists in modeling the behavior of the simple and complex cells whose are the main actors, of the matching operation in the primary visual cortex. From this, we formalize a model for the calculation of the binocular energy. This latter is exploited then in addition to stereoscopic constraints to find the best matching. So this is the main innovative part and which offers a better comprehension of the HVS 3D perception.

The rest of this paper is organized as follows: After the description of the proposed approach in section 2, a brief description of the bandelet transform used to model the simple and complex cells behavior is given in section 3. Section 4 is dedicated to the proposed binocular energy model (BME). The experimentation is described in section 5 and this paper ends with some conclusion and future directions.

## 2. PROPOSED APPROACH

In this section, we describe our coder illustrated in Figure 1. The first step in our coder is to represent images in CIE $L^*a^*b^*$ color space; that choice will be justified later. Then, some transformations are applied on the stereoscopic images to be able to model the properties of simple and complex cells. These latter are characterized by their size, their orientation and their phase.

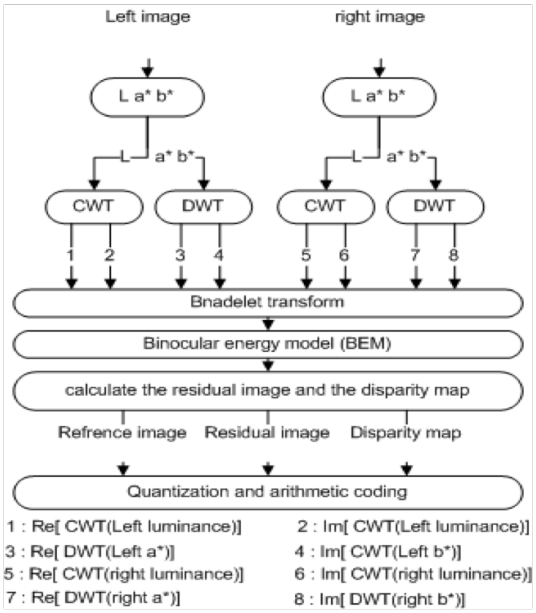


**Fig. 1**. Proposed approach.

The simple cells work in pairs. As shown in Figure 3, a cell complex takes as input two simple cells. The simple cells belonging to the same pair has a shift phase equal to $\pi/2$. From this description, we modeled simple cells, taking into account some of their characteristics. A complex wavelet transform (CWT) is applied to the luminance component (L) of the right and left image. To obtain the real and imaginary parts of the luminance component we use two filters with a shift phase equal to $\pi/2$. For the chromatic component a discrete wavelet transform (DWT) is applied to the component ($a^*$) and also to the component ($b^*$), of the two images. In the CIE $L^*a^*b^*$ color space, the two color components are orthogonal. Thus the first component ($a^*$) represents the real part of the color information and the second represents the imaginary part ($b^*$). This explains why it has been chosen for our approach. The real part of the luminance (Re [CWT(L)]) and chrominance (Re [DWT(a)]) have a shift phase equal to $\pi/2$ with their imaginary parts (Im [CWT(L)]) and (Im [DWT(a)]).

This preprocessing step is realized to separate the responses obtained with cells belonging to the same pair (Fig. 3). Both cells belonging to the same pair have the same size in the same orientation the same position and the same amplitude with a shift phase equal to $\pi/2$. The next step is to define the size and position of each pair of simple cells. For this we applied the bandelet transform on the wavelet coefficients obtained with the CWT and the DWT. Figure (fig 2) shows the obtained result by applying the same geometry on the real and imaginary parts of luminance and chrominance components of each image. After this last stage, the left and the right images are represented with a set of dyadic squares characterized by their orientation phase, size and position. A dyadic square belonging to the real part of the luminance (Re [CWT(L)]) with the dyadic square that have the same position in the imaginary part of the luminance (Im [CWT(L)]) represent one pair of dyadic squares (the same for chrominance) (Fig. 2).

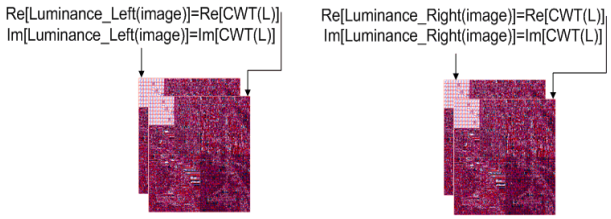


**Fig. 2**. Wavelet and Bandelet transform of both retinal images.

Once the simple cells modeled, we model the complex cells to match dyadic squares pairs of the couple of images (See Section 4). The matching process generates a disparity map and a residual image. The residual image undergoes adaptive quantization according to different scales of the CWT and DWT decomposition. The quantized image is encoded with the disparity map using arithmetic coder. The quality of the reconstructed target image depends on the quality of compressed residual image. In the decoding process, the reverse of coding operation is performed. It uses the reference image with the residual image and the disparity map to reconstruct the target image.

## 3. BANDELET TRANSFORM

In the previous section, we mentioned the spatial-frequency transform that we use in our metric scheme. As shown in the following figure, a CWT is applied to the luminance component [$L^*$] of the left and right retinal images. The filters used to compute the real and the imaginary parts presents a shift-phase equal to $\pi/2$. DWT is applied to the chromatic component of the both images ($a^*$ and $b^*$), knowing that these components are orthogonal. This preprocessing step allows a complex writing of the luminance and chrominance components as described by equation 1.

$$\begin{aligned} Image &= \{Re[L], Im[L], Re[C], Im[C]\} \\ &= \{Re[CWT(L^*)], Im[CWT(L^*)], DWT[C(a^*)], DWT[C(b^*)]\} \\ &L: luminance, C: Chrominance \end{aligned} \quad (1)$$

To ease the comprehension of the approach, a brief review of the Bandelet transform is given in the following paragraphs. The reader can refer to [16] for a full detailed description of the Bandelet transform.

The bandelets are defined as anisotropic wavelets that are warped along the geometric flow, which is a vector field indicating the local direction of the regularity along edges. The dictionary of bandelet frames is constructed using a dyadic square segmentation and parameterized geometric flows. The ability to exploit image geometry makes its approximation optimal for representing the images. For image surfaces, the geometry is not a collection of discontinuities, but rather areas of high curvature. The Bandelet transform recasts these areas of high curvature into an optimal estimation of regularity direction. Figure 2 shows an example of bandelets along the geometric flow in the direction of edges. In real applications, the geometry is in the direction of the edge. The support of the wavelets is deformed along the geometric flows in order to exploit the edge regularity.

## 4. BINOCULAR ENERGY MODELS (BEM)

### 4.1 State-of-the-art

Modeling the binocular energy created by the simple and complex cells is an important step to be included in computer vision applications dedicated to stereoscopic images. Several works exist in literature and we propose, in this section, to describe the most important ones related to our work. Hubel and wisel [17] defined two types of binocular cells, namely the simple cells and the complex cells, qualifying the degree of complexity in the internal structure of a receiver. The complex cells in their model are built by the association of a number of simple cells as described in Fig. 3. According to Campbell et al.[18], the receiving fields of this cells are described as a linear filter constituted by different regions of type "ON" (activated) and "OFF" (inhibited). The optimal activation of these cells is made by a grating of luminance so that the white bar covers all the ON region while the black one covers the OFF region.

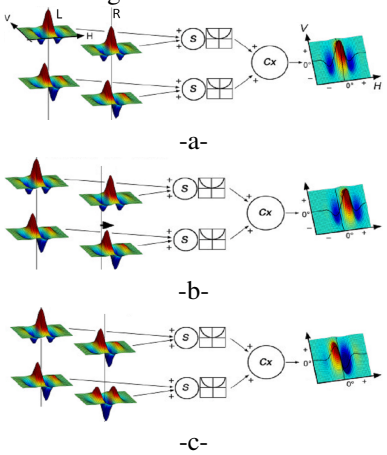


**Fig. 3**. Illustration of the receiving fields of simple and complex cells. (a) identical for both eyes, (b) a shift position, (c) a phase shift (In the model of binocular energy, the neurone of energy $Cx$ represents the complex cell)

The complex cells are not like the simple cells in the sense that they have different receiving fields. These latter have positive and negative responses for the simple cells while the responses are positive or zero for complex cells. In 1990, Ohzawa et al.[19] proposed a complete model to compute the binocular energy. This work has inspired ours for the definition of the BEM used in our metric. The function of the simple and complex cells can be mathematically described by the orientation adaptive wavelets (Gabor wavelet, curvelet, bandelet, ...). The ON and OFF regions, of these cells, correspond respectively to peaks and hollows of these functions.

### 4.2 Proposed binocular energy model

Starting from the definition given above, the model that we propose to calculate the binocular energy is based on the model proposed by Ohzawa [19] and the one proposed by Fleet [20]. Bandelet transform, applied on the wavelet coefficients of luminance and chrominance components, allows to define the image geometry. This latter is defined by a set of dyadic squares (the same geometry is applied to the real and imaginary parts of the luminance and chrominance). Each dyadic square is characterized by its size and orientation. Dyadic squares obtained with CWT applied to the luminance are arranged in pairs, similar to the dyadic squares obtained with the DWT, applied to both chrominance components. Dyadic squares of a given pair belong to the real part of the CWT ($Re[L(x)]^L$) and the imaginary part of the CWT ($Im[L(x)]^L$). Dyadic square pairs of the chromatic component belong respectively to the real part represented by the DWT ($Re[L(x)]^C$), applied to the component $a^*$ and the imaginary part represented by the DWT ($Im[L(x)]^C$), applied to the component $b^*$. Dyadic squares of a pair have given the same orientation and same size with a shift-phase equal to $\pi/2$. $L(x)$ and $R(x)$ (responses of two simple cells (Fig. 3)), Complex-valued response in left and right eyes, are expressed by their amplitude and orientation of the complex function($L(x) = \rho_l(x)\exp(\phi_l(x))$). where:

$$\rho_l^2(x) = |L(x)|^2 = Re[L(x)]^2 + Im[L(x)]^2 \quad (2)$$

$\rho_l(x)$ is the monocular amplitude of the complex function and $\phi_l(x)$(Eq. 3) is the monocular phase of the complex function.

$$\phi_l(x) = arg|L(x)| = arctan(Im[L(x)]/RE[L(x)]) \quad (3)$$

**Table 1**. parameters table.

| parameters | Definitions |
|---|---|
| X | Retinal position |
| L(x), R(x) | Complex-valued response in left and right eyes, at position x |
| $Re[L(x)]^L$ | Luminance real part of left monocular response(dyadic square) |
| $Im[L(x)]^L$ | Luminance imaginary part of left monocular response |
| $Re[L(x)]^C$ | Color real part of left monocular response |
| $Im[L(x)]^C$ | Color imaginary part of left monocular response |
| $\rho_{l/r}(x)$ | Monocular (left eye) amplitude signal |
| $\phi_{l/r}(x)$ | Monocular (left eye) phase signal |
| $\phi'_{l/r}(x)$ | Left-eye instantaneous frequency at position x |
| $\Delta\psi$ | simple cell phase shift |
| $d$ | Stimulus disparity |
| $\Delta\phi(x)$ | phase difference |
| $E(x)$ | Binocular energy response at retinal position x |
| $E(x,d)$ | Response of binocular energy neuron with simple cell position shift |
| $E(x,\Delta\psi)$ | Response of binocular energy neuron with simple cell phase shift |
| $E(x,d,\Delta\psi)$ | Response of binocular hybrid energy neuron with position shift $d$ and phase shift $\Delta\psi$ |

After all the preprocessing steps comes the stage of matching of the retinal pairs of images. For this, the dyadic

squares pair of one image are matched with another pair of the second image by calculating the binocular energy produced by these two pairs of dyadic squares (which represents the response of two simple cells). The cell responsible of the information fusion, in the human visual system, is the complex cell. The binocular complex cell takes as input two responses from two simple cells (two pairs of dyadic squares belonging respectively to the left and right retinal images). If the complex cell is of type monocular, it will take as input a response of a simple cell (a pair of dyadic squares). In the case of a binocular complex cell, the binocular energy (Eq. 4) is calculated as described in [20].

$$E(x) = |L(x)+R(x)|^2 = (Re[L(x)]+Re[L(x)])^2 \\ (Im[L(x)]+Im[R(x)])^2 \quad (4)$$

The two pairs of matched dyadic squares, belonging respectively to the right image $R(x)$ and the left image $L(x)$ must have the same orientation and the same size.
When we replace $L(x) = \rho_l(x)\exp(\phi_l(x))$and $R(x) = \rho_r(x)\exp(\phi_r(x))$ by their respective definition, we obtain the following equation :

$$E(x) = \rho_l^2(x)+\rho_r^2(x)+2\rho_l(x)\rho_r(x)cos(\Delta\phi(x)) \quad (5)$$

$E(x)$ is the energy of the response obtained by the binocular complex cell. When the both pairs of dyadic squares have not a same position, the right monocular response $R(x)$ is a shifted version of the left monocular responses $L(x)$, i.e. $R(x) = L(x-d)$. Similarly, when the phase signal is not the same between the pairs of dyadic squares $\phi(x) = \phi(x-d)$. From this, we can express the inter-ocular phase difference using a Taylor series of $\phi_l(x-d)$(Eq. 6):

$$\Delta\phi_l(x,d) = \phi_l(x)-\phi_r(x) = \phi_l(x)-\phi_l(x-d) = d\phi'_{l/r}+O[d^2] \quad (6)$$

Combining equation 6 with equation 5 gives us a useful characterization of a binocular energy as described by equation 7. As the disparity is increased slightly above zero, the binocular energy response decreases as the cosine of disparity times instantaneous frequency, $cos(d\phi'_{l/r})$.

$$\Delta\phi_l(x,d) = \phi_l(x)-\phi_r(x) = \phi_l(x)-\phi_l(x-d) = d\phi'_{l/r}+O[d^2] \quad (7)$$

In [19], authors showed that if the simple cells have not the same orientation, the disparity between them is useless. Fleet [20] defined this relation in the following way:

$$R(x) = \exp(i\Delta\psi)L(x-d) = \rho_l(x-d)\exp(\phi_l(x-d)+\Delta\psi) \quad (8)$$

$\Delta\psi$ denotes a phase shift between the couple of simple cells. So, the binocular energy of the left and the right pairs of dyadic squares are then related. The phase difference has now the form:

$$\Delta\phi_l(x,d,\Delta\psi) = \phi_l(x)-\phi_r(x)-\Delta\psi = d\phi'_{l/r}-\Delta\psi \quad (9)$$

Finally, the binocular energy(Eq. 7), computed by the complex cell for the both pairs of dyadic squares, is equal to:

$$E(x,d,\Delta\psi) = \rho_l^2(x)+\rho_r^2(x)+2\rho_l(x)\rho_r(x)cos(d\phi'_l-\Delta\psi) \quad (10)$$

When the two dyadic square pairs (left and right) are matched, if the binocular energy E($x_l$, $x_r$)≈ E($x_l$, $x_l$) the error between the two dyadic squares is equal to 0, nothing is saved in the residual image. In the case of an important difference between E($x_l$, $x_r$) and E($x_l$, $x_l$) , this latter is saved n the residual image (Re(residual image)=Re[$x_l$]-Re[$x_r$]) and (Im(residual image)=Im[$x_l$]-Im[$x_r$]) to be used in the reconstruction of the target image.

## 5. EXPERIMENTAL RESULTS

In this section, the experimental evaluation of the proposed coder is reported. Two stereo image pairs were employed for the experimental evaluation (cf. fig. 4). The evaluation of the proposed method is performed by using the PSNR calculated using the following equation:

$$PSNR = 10log_{10}\frac{255^2}{(MSE_L+MSE_R)} \quad (11)$$

where $MSE_L$ and $MSE_R$ are respectively the mean square errors of the left and the right images. In this section we present results obtained with our coder. The results are presented as comparative curves between the results obtained with our coder and *Woo et al.* [4] and *Ellinas et al.* [13].

**Fig. 4**. Cones (top) and Teddy (bottom) stereo images.

The results obtained with our coder are very interesting compared to those obtained with the coder of Woo and Ellinas. This can be explained by several factors related to the conception of our coder and to the functions used in modeling. In our model we took into account the characteristics of 3D vision. In the case of 3D vision we have all artifacts unique to the 2D vision and artifacts unique to 3D vision. To assess the quality of our compressed stereoscopic images we used PSNR although it is not the best metric for that.

We can classify the 3D artifacts in two categories. In the first category are artifacts caused by the loss of depth. This case may be caused by an intense filtering of textured areas in the image. Such artifacts can be caused by the JPEG 2000 coder for example. In our coder, we use bandelets transform, the advantage of this transform is the preservation of contours using its adaptive quantization function. For each pair of dyadic squares is kept the significant wavelet coefficients (which respect the geometry of the dyadic square). This quantization method preserve the geometry of dyadic squares pair ensuring a good depth perception.

The second category of artifacts that we find in the 3D perception are the appearance of false depth artifacts caused by 2D encoder used as it is the case with the block effects caused by the JPEG coder. With the block effect we seem to see the blocks in depth especially in uniform areas where there is no geometry.

## 6. CONCLUSION

In this paper, we proposed a coder for stereoscopic color images based on properties of human visual system. Both

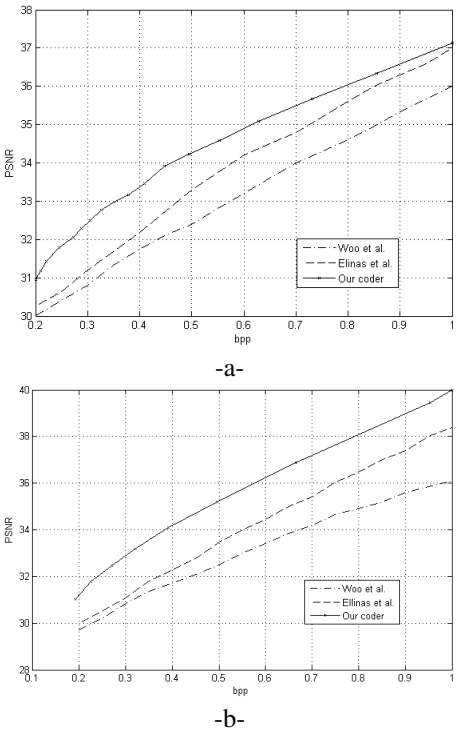


**Fig. 5**. Quality performance evaluation of the proposed coder vs. woo et al. [9] for (a) Teddy image and (b) Connes.

images undergo series of preprocessing before coding. The pretreatment phase models the operation of simple and complex cells responsible for binocular fusion. To model some properties of simple cells we applied CWT on the luminance component and a DWT on the color components of two images. A bandelets transform was then applied on the obtained wavelet coefficients. We obtained after this treatment a set of dyadic squares pair (real and imaginary) characterized by their size amplitude and orientation as the simple cells. To match these square dyadic pairs we propose a binocular energy model (BEM). After the match we get a disparity map and a residual image. These are quantified and coded with the arithmetic coder. The obtained results showed better performance than the literature coders used in this paper. The validation of this coder will continue by a specific psychophysical study.

## REFERENCES


[1] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov, and A. Henik, "On stereo image coding," in *ICPR*, 1988, pp. I: 357–359.

[2] H. Aydinoglu, F. Kossentini, Q. Jiang, and M.H. Hayes, III, "Region-based stereo image coding," in *ICIP*, ICIP, 1995, pp. II: 57–60.

[3] D. Tzovaras and M. Strintzis, "Disparity estimation using rate-distortion theory for stereo image sequence coding," *Digital Signal Processing Proceedings*, ,413-416, 1997.

[4] W.T. Woo, "Stereo image coding using hierarchical mrf model and selective overlapped block disparity compensation," in *ICIP*, 1999, pp. II:467–471.

[5] M. Magnor and B. Girod, "Hierarchical coding of light fields with disparity maps," in *ICIP*, 1999, pp. III:334–338.

[6] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Transactions on Circuits and Systems for Video Technology*, 898-903, 2002.

[7] T Frajka and K Zeger, "Residual image for stereo image compression," *ICIP*, 2002.

[8] W.H. Kim, J.Y. Ahn, and S.W. Ra, "An efficient disparity estimation algorithm for stereoscopic image compression," vol. 43, no. 2, pp. 165–172, May 1997.

[9] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *CirSysVideo*, vol. 10, no. 2, pp. 194, March 2000.

[10] C.L. Pagliari and T.J. Dennis, "Stereo disparity computation in the dct domain using genetic algorithms," in *ICIP*, 1997, pp. III: 256–259.

[11] M.S. Moellenhoff and M.W. Maier, "Dct transform coding of stereo images for multimedia applications," vol. 45, no. 1, pp. 38–43, February 1998.

[12] H. Aydinoglu and M.H. Hayes, III, "Stereo image-coding: A projection approach," *IP*, vol. 7, no. 4, pp. 506–516, April 1998.

[13] J.N. Ellinas and M.S. Sangriotis, "Stereo image compression using wavelet coefficients morphology," vol. 22, no. 4, pp. 281–290, April 2004.

[14] S.K. Nath and E. Dubois, "An improved, wavelet-based, stereoscopic image sequence codec with snr and spatial scalability," vol. 21, no. 3, pp. 181–199, March 2006.

[15] F. Moran and N. Garcia, "Comparison of wavelet-based three-dimensional model coding techniques," vol. 14, no. 7, pp. 937–949, July 2004.

[16] G. Peyre, "Geometrie multi-chelles pour les images et les textures," *Phd thesis*, vol. Ecole Polytechnique, 2005.

[17] TN. Hubel, DH. Wiesel, "Stereoscopic vision in macaque monkey. cells sensitive to binocular depth in area 18 of the macaque monkey cortex," *Nature*, vol. 225, no. 41-42, 1970.

[18] GF. Enroth-Cugell C. Campbell, FW. Cooper, "The spatial selectivity of the visual cells of the cat," *J Physiol*, vol. 203, pp. 223–235, 1969.

[19] GC. Freeman-RD. Ohzawa, I. DeAngelis, "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors," *Science*, vol. 249, pp. 1037–1041, 1990.

[20] D.J. Fleet, H. Wagner, and D.J. Heeger, "Neural encoding of binocular disparity: Energy model, position shifts and phase shifts," *Vision Research*, vol. 36, no. 12, pp. 1839–1857, 1996.