Studying and Modeling the Binocular Energy for Stereoscopic Images

Rafik Bensalma and Mohamed-Chaker Larabi
XLIM lab, SIC dept, Université de Poitiers, France

Abstract

Stereovision is a research field attracting people from various fields such as psychology, physiology, mathematics and recently computer vision. Physiological and psychological studies allowed to understand, in a significant way, the behavior of the visual cortex. Plenty of these results have been modeled but never implemented in imaging applications for various reasons. However, this step is very important in order to take advantage of the advances in the aforementioned fields. This paper tries to formulate analytically the binocular behavior of the HVS by applying the binocular fusion made by the complex cells to merge the retinal information provided by the simple cells. This allows to study the evolution of the binocular energy with regards to different impairments on one or both of the stereo pairs. Results demonstrate and important correlation between the binocular energy and the quality of the 3D reconstruction.

Introduction

The human visual system can analyze our complex environment represented in a spatial-frequency space of four dimensions (x,y,z,t). The visual system has several capabilities related to physiological or psychological aspect of vision. Among these capabilities, binocular vision or stereoscopic vision has led, historically, the interest of many scientists in different fields. Several studies and experiments in the field of physiology and psychology of vision have been conducted to understand how does the human visual system work, and explain the factors involved in vision. Vision in general and binocular vision in a particular is considered as an important research field for neurobiologists and psychologists, and now researchers from all fields such as artificial intelligence, mathematics, computer and computer vision.

Figure 1. *Recomposition of the left and right retinal images into 3D.*

Binocular vision can be defined as the combination of the left and right retinal images for the analysis of the same region of the scene (figure 1). The depth perception of a scene may be considered as a purely psychological process that merges the left and right images of the retina. The quality of the reconstructed image in the brain, is highly correlated to the quality of retinal images. This concept is very subjective, it requires taking into account many parameters such as color, contrast, depth, . . .

Figure 2 shows the various steps taken by the retinal information before binocular fusion. The information leaves the retina by retinal ganglion cells. These cells respond to stimulation of a small circular area of the retina forming the receptive field*(RF)*(figure 2). These receptive fields are classified into three types shown in Figure 2 by their antagonist colors. This information is then conveyed to the geniculate receptive field having similar properties than those of the ganglion cells. After a series of processing, the retinal information is transmitted to the visual cortex. Neurons in the primary visual cortex $V1$ are classified into three main classes: simple cells, complex cells and hypercomplex cells. These cells are responsible of the binocular fusion (See Figure 2).

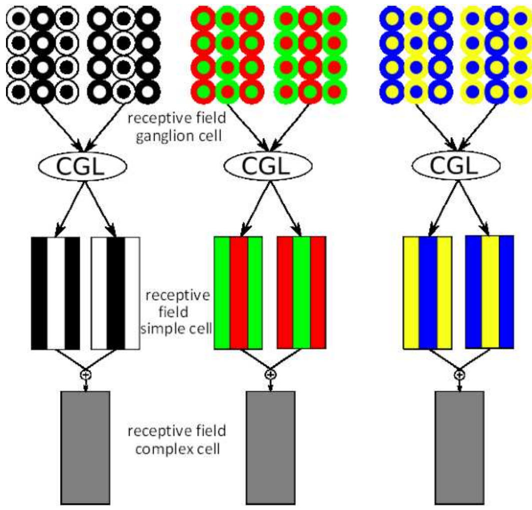


Figure 2. *Retinal information flow.*

In this paper we propose an analytical model for calculating the energy generated by a binocular pair of stereoscopic images. This work is inspired from the field of physiology and psychology of vision. The aim is to understand and model the energy generated by the HVS when viewing a stereo-pair. This will help to determine the artifacts playing an important role in the reconstruction process.

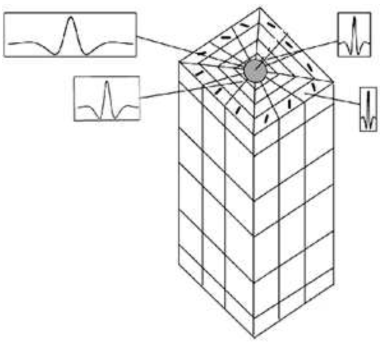
The paper is organized as follows: A state of the art about the physiology of vision is given in section 2. Section 3 is devoted to the description of the model based on simple and complex cells. Results are presented in section 4 and the paper ends with some conclusions and future works.

Modeling the binocular fusion

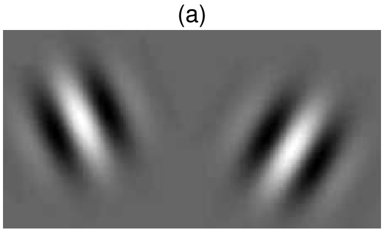
In this work, we are interested in the merging step of the retinal information. Several studies have tried to understand and model the simple and complex cells, responsible for binocular fusion. *Barlow et al.*[1] were the first to show the existence of different cells in the primary cortex. These cells respond preferentially when a lag is introduced between the bars of a network of luminance, and they reveal the existence of disparity detectors. The structure of receptive fields of cells was first defined by *Hubel and Wiesel* [2] in the 1950s. These receptive fields are considered as linear spatial filters [2] and [3]. These filters are characterized by their elongated shape composed of two distinct regions antagonists ON and OFF, inherited from the receptive fields of ganglion cells 2, respectively activated and inhibited. Many physiological experiments have shown that these cells can be modeled as linear filters, whose impulse response was measured at various points in the visual cortex. *DeAngelis et al.* [4] showed that these impulse responses can be approximated by wavelets of *Gabor* [5], constructed with a gaussian window $g(x_1,x_2)$ multiplied by sinusoids.

$$\psi^k(x_1,x_2) = g(x_1,x_2)\exp[-i(x_1 cos\alpha_k + x_2 sin\alpha_k)] \qquad (1)$$

Peyre et al. [6] [7] have linked this work on the bandelets with hypercomplex cells. As shown in this figure 3 (taken from [6]), simple cells are connected to complex cells, whose are connected to hypercomplex cells; represented by the cylinder. The closer we get to the center of the cylinder, smaller the size of single cells will be. Bandelet transform offers a similar decomposition. Thus, the dyadic squares can model many of the characteristics of simple cells, as size, amplitude, phase and orientation.



(a)



(b)

Figure 3. *Organization of simple cells in visual cortex [8]. (a) Hypercomplex cell representation and (b) simple cell receptive field.*

Complex cells do not exhibit the same characteristics as simple cells, except that they inherit from their size. Unlike simple cells that are sensitive to a given orientation and position of light in their receptive field, complex cells provide optimal responses of the position of some stimulus in the receptive field and orientation. The receptive fields of complex cells are not opposing regions (ON / OFF) which implies a loss of sensitivity to monocular phase, hence the insensitivity to the position of the stimulus.

Several models have been proposed to model the behavior of these complex cells *Fleet et al.* [9] et *Ohzawa et al.* [10]. Like simple cells, there are two types of complex cells, monocular complex cells and binocular complex cells. The receptive fields of monocular complex cells receive signals from two monocular simple cells of the same side of the retina, to calculate the monocular signals corresponding to the occluded regions. The receptive fields of binocular complex cells receive signals from two simple cells to generate a binocular signal (see figure 3). The two simple cells are grouped in pairs with phase quadrature (The arrangement of ON and OFF regions has an offset of $\pi/2$.

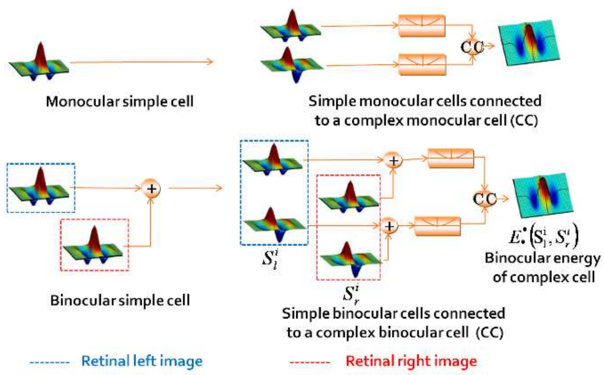


Figure 4. *Connections between simple and complex cells.*

Proposed analytical modeling

Modeling of simple cells

The first step consist to convert the stereo images in L^*a^*b* color space, which represents the color antagonism of the HVS. Simple cells are working in pairs that have similar characteristics with a phase shift equal to $\pi/2$. So we apply a complex wavelet transform CWT [11] to the different components of stereoscopic images. The wavelets used in CWT are directional wavelets (V,D,H) (See figure) to calculate the real and imaginary part of the transform. The two wavelet functions used for calculating the real and imaginary part of the CWT has a quadrature phase. In the same time, the wavelet transform allows decomposition of the image into perceptual channels, represented by the different scales of the transform [12].

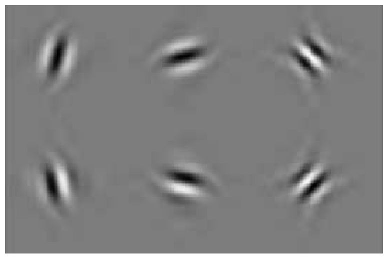


Figure 5. *Directional wavelets.*

The second step, i.e. modeling simple cells, is to apply a bandelet transform on the coefficients obtained by using the CWT. The geometry obtained on the real part of the CWT is replicated on the imaginary part. This provides pairs of dyadic squares with the same characteristics and a phase quadrature, as the simple cells pairs.

Modeling complex cells

Complex cells are the those responsible for the binocular fusion. A complex cell takes as input a pair of simple cell. Thus, after modeling the simple cells comes the matching stage of the retinal image pairs . For this, the dyadic squares pair of one image are matched with those of the second image by calculating the binocular energy produced by them (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 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. 2) is calculated as described in *Fleet et al.*[9].

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

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 (Eq. 3):

$$E(x) = \rho_l^2(x) + \rho_r^2(x) + 2\rho_l^2(x)\rho_r^2(x)cos(\Delta\phi(x)) \tag{3}$$

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 interocular phase difference using a Taylor series of $\phi_l(x-d)$(Eq. 4):

$$\Delta\phi_l(x,d) = \phi_l(x) - \phi_r(x) = \phi_l(x) - \phi_l(x-d) = d\phi'_{l/r} + O[d^2] \tag{4}$$

Combining equation 4 with equation 3 provide a useful characterization of the binocular energy as described by equation (Eq. 5). As the disparity is increased slightly above zero, the binocular energy response decreases as the cosine of disparity times instantaneous frequency, $cos(d\phi'_{l/r})$.

$$\Delta\phi_l(x,d) = \phi_l(x) - \phi_r(x) = \phi_l(x) - \phi_l(x-d) = d\phi'_{l/r} + O[d^2] \tag{5}$$

Ohzawa et al.[10], showed that if the simple cells have not the same orientation, the disparity between them is useless. Fleet[9] defined this relation in the following way:

$$R(x) = \exp(i\Delta\psi)L(x-d) = \rho_l(x-d)\exp(\phi_l(x-d) + \Delta\psi) \tag{6}$$

$\Delta\psi$ 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:

$$\Delta\phi_l(x,d,\Delta\psi) = \phi_l(x) - \phi_r(x) - \Delta\psi = d\phi'_{l/r} - \Delta\psi \tag{7}$$

Finally, the binocular energy(Eq. 8), computed by the complex cell for the both pairs of dyadic squares, is equal to:

$$E(x,d,\Delta\psi) = \rho_l^2(x) + \rho_r^2(x) + 2\rho_l^2(x)\rho_r^2(x)cos(d\phi'_l - \Delta\psi) \tag{8}$$

Experimental results and discussion

In this section we present some experimental results in order to validate the proposed approach and to study the influence of the binocular energy. It is performed in two steps :

- Show matching results obtained by the proposed model using the pair of stereoscopic images (Doll) from the used image database.
- Study the evolution of the binocular energy according to different types of artifacts such as: JPEG, JPEG 2000 compression and noise.

For the experiments, we used an image database created and provided by the university of Toyama. Some examples from this database are given in figure 6.

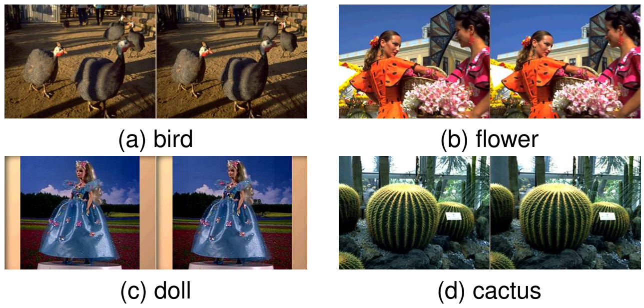
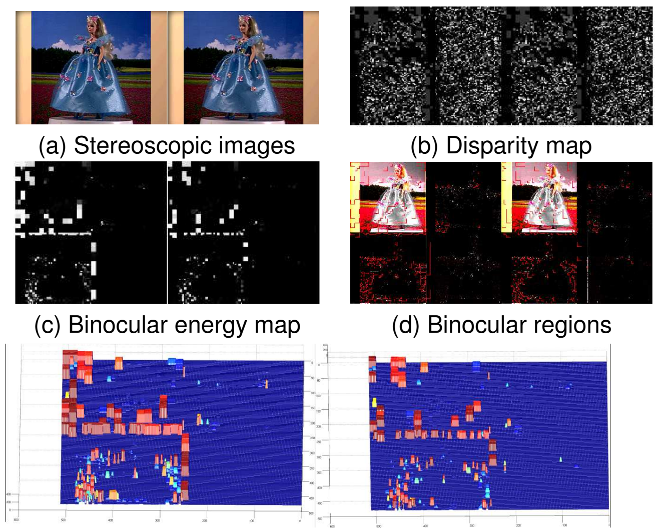


Figure 6. *Some stereoscopic images of the database used.*

Generation of the disparity and binocular energy map

For these experiments, we focused first on the generation of the disparity and the binocular energy maps. We took, the "doll" image to illustrate the different aspects. Figure 7-b shows the disparity maps obtained from figure 7-a. Figure 7-c gives the binocular energy map and Figure 7-d highlights the binocular regions used to compute the binocular energy as described in our approach. We can notice that the binocular energy depends on the spectral information contained in the different levels of the decomposition and is greater high frequency subbands.



(e) Binocular energy map

Figure 7. *Matching of the stereo image-pairs. Doll is used for the illustration.*

In order to study the evolution of the binocular energy, we used 2 compression standards (JPEG and JPEG 2000) and a random noise. Several artifacts ca be generated depending on the bitrate as blockiness, bluriness, ringing ... or on the strength of the noise. Both symmetric and asymmetric configurations have been adopted. This means that for each bitrate (resp. strentgh) of the left image, we applied different bitrates (resp. strenghts) for the right image. This generates several configurations to be tested.

JPEG

Figure 8 gathers for 6 images from the database (Cattle, Flower, Goat, Gate, Doll, Woman), the results of binocular energy for JPEG compression. Each plotted point on the figure represent the results for a couple of bitrates (one for the left image and the other for the right). We can notice that there is a strong relation between the binocular energy and the bitrate (i.e. quality). Also the binocular energy has not a completely linear behavior because the blocking effect is not perceived in the same way in $2D$ and $3D$ vision. In $3D$, the blocking effect causes false depths and at the same time reduces the quality of the original depth.

JPEG 2000

Unlike JPEG, JPEG 2000 does not introduce false depth but it smoothes the original edges. The impairments introduced is on amplitude, phase and orientation of Dyadic Square. Figure 9 gathers binocular energy results for 6 images from the database (Cattle, Flower, Goat, Gate, Doll, Woman) using JPEG 2000 in both symmetric and asymmetric coding. In this case we can also notice a correlation between bitrate (quality) and binocular energy. This means that the proposed model is able to transcribe the strength of a JPEG 2000 compression into a binocular energy explaining what is perceived by an observer.

Noise

The proposed model has been used on images on which an impulse noise has been applied. The aim is to study the evolution of the binocular energy function of this specific noise which is different from the impairments generated by JPEG and JPEG 2000 compression.

The results obtained with our model are given in figure 10. X-axis represents the noise variance and the Y-axis represents the binocular energy. We can draw similar conclusions than previously about the correlation between the binocular energy and the noise strength. Also, we can notice that the perception of noise in 3D is not linear.

conclusion

In this paper, we proposed an analytical model for the computation of the binocular energy between two pairs of stereoscopic images. The aim was to study the behavior of this energy with regards to different types of impairments. The results show that the proposed model is highly correlated with the impairment strength and at the same time is not linear, which fully comply with the human perception. One major direction for this work is to build a quality metric exploiting this energy. The aim is thus to predict the quality of stereoscopic pair from a 3D perceptual point of view.

References

[1] C. P. J. Barlow, HB. Blakemore, "The neural mechanism of binocular depth discrimination," *J Physiol* **193**, pp. 327–342, 1967.

[2] T. Hubel, DH. Wiesel, "Stereoscopic vision in macaque monkey. cells sensitive to binocular depth in area 18 of the macaque monkey cortex," *J Nature* **225**(41-42), 1970.

[3] G. E.-C. C. Campbell, F.W. Cooper, "The spatial selectivity of the visual cells of the cat," *J Physiol* **203**(223-235), 1969.

[4] I. F.-R. DeAngelis, G.C. Ohzawa, "Depth is encoded in the visual cortex by a specialized receptive field structure," *J Nature* **352**(156-195), 1991.

[5] D. Gabor, "Theory of communication," *J Inst Elec Engin* **93**(429-457), 1946.

[6] E. Le Pennec and S. Mallat, "Bandelet image approximation and compression," *SIAM Multiscale Modeling and Simulation* **4(3)**, p. 9921039, 2005.

[7] S. Mallat and G. Peyr, "Orthogonal bandelet bases for geometric image approximation.," *!!* **!!**, 2006.

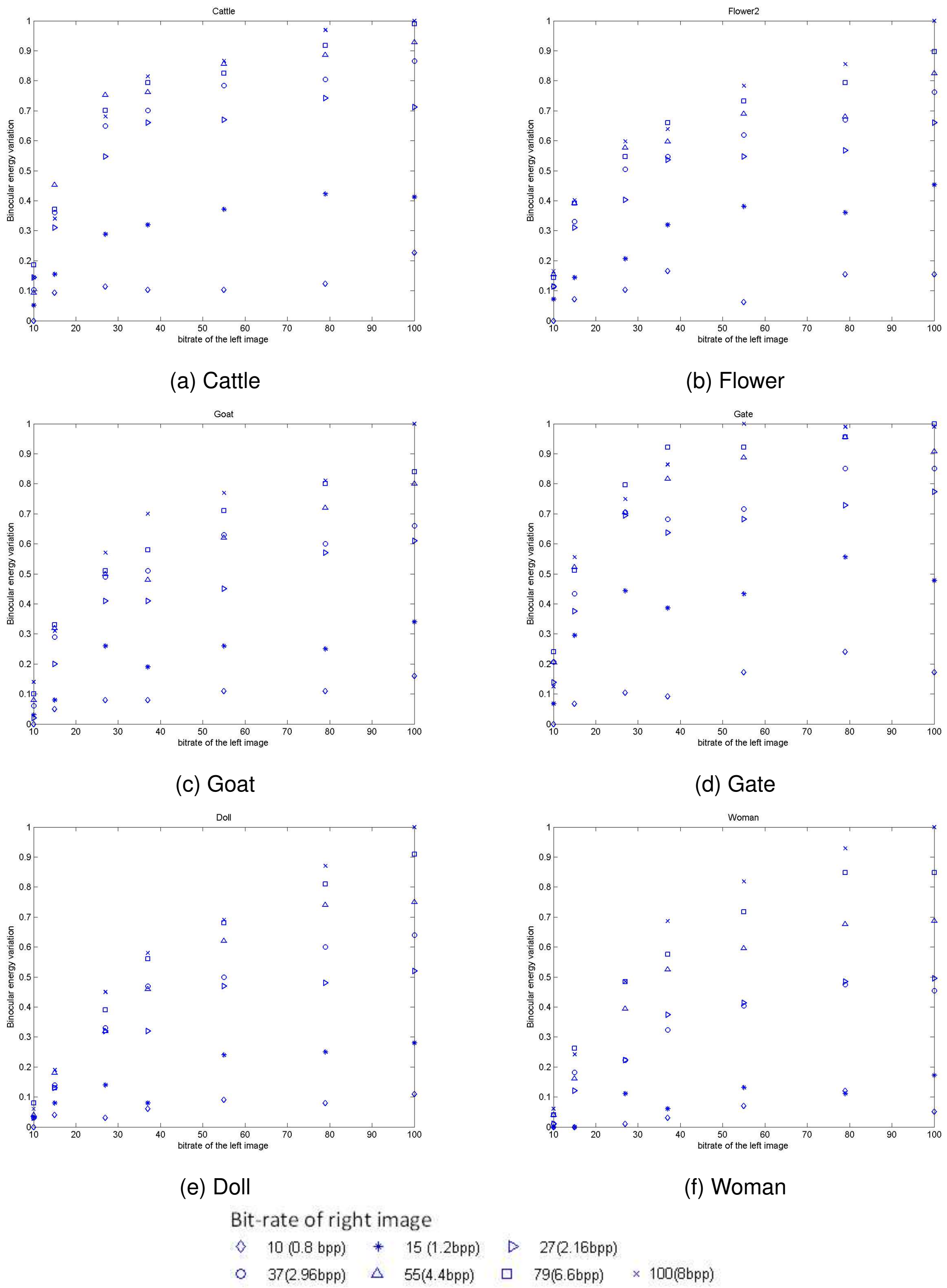
[8] G. Peyre, "Geometrie multi-chelles pour les images et les textures," *Phd thesis* **Ecole Polytechnique**, 2005.

[9] D. Fleet, H. Wagner, and D. Heeger, "Neural encoding of binocular disparity: Energy model, position shifts and phase shifts," *Vision Research* **36**(12), pp. 1839–1857, 1996.

[10] G. F.-R. Ohzawa, I. DeAngelis, "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors," *Science* **249**, pp. 1037–1041, 1990.

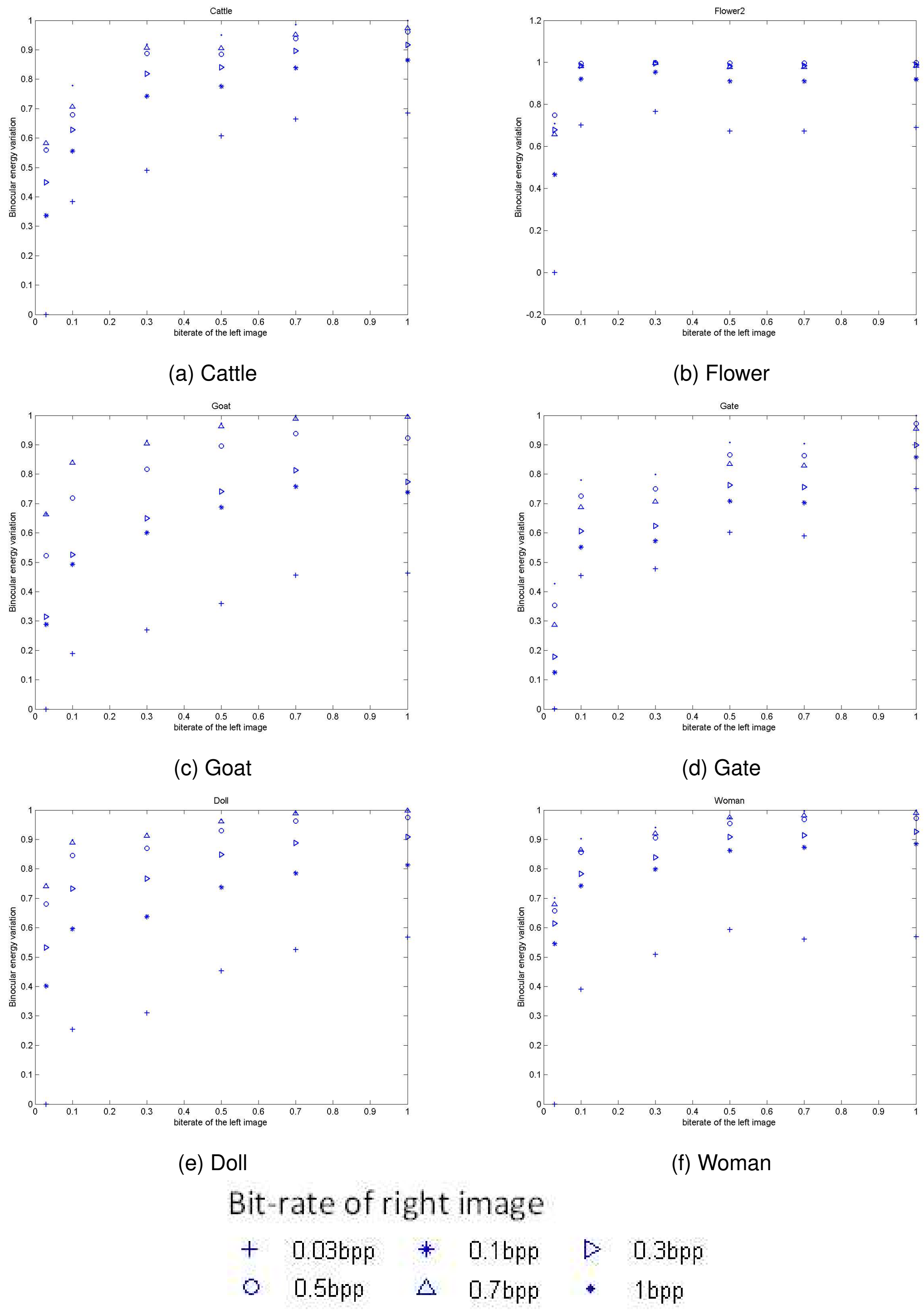
[11] R. G. Selesnick, I. W. Baraniuk and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine* (123-151), 2005.

[12] S. Mallat, "A theory for multiresolution signal decomposition : the wavelet represntation," *IEEE, PAMI* **11**(7), pp. 674–693, 1989.



(g) Legend

Figure 8. *Results of binocular energy for symmetric and asymmetric JPEG coding.*



(a) Cattle (b) Flower

(c) Goat (d) Gate

(e) Doll (f) Woman

(g) Legend

Figure 9. *Results of binocular energy for symmetric and asymmetric JPEG 2000 coding.*

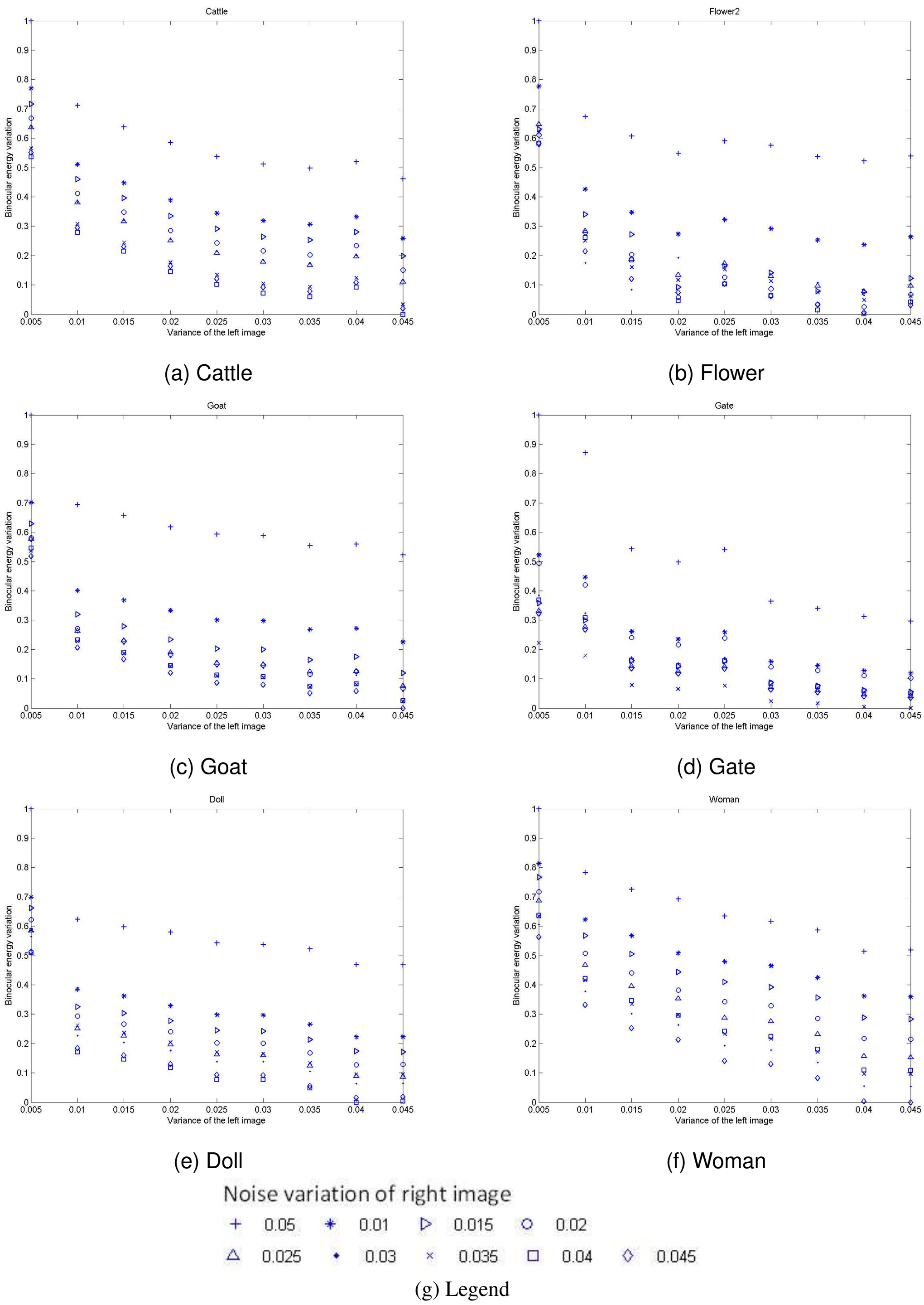


Figure 10. *Results of binocular energy for symmetric and asymmetric impulse noise.*