TOWARDS A PERCEPTUAL QUALITY METRIC FOR COLOR STEREO IMAGES

Rafik BENSALMA, Mohamed-Chaker LARABI, Senior member, IEEE

XLIM-SIC Laboratory, University of Poitiers, France

ABSTRACT
In this paper, we propose a quality metric for color stereo images. The concept of our metric is inspired by the behavior of simple and complex cells located in the primary visual cortex. These cells are responsible for merging left and right retinal images. To replicate the task performed by these cells, we adopted an approach based on spatial-frequency transform with the processing of selective orientations. From that, a model that calculates the binocular energy contained in the left and right retinal images has been proposed. The amplitude variation of the binocular energy defines the quality criterion of the reconstructed depth within the Human Visual System (HVS). Finally, from the experimental results, the used criterion seems to be correlated to human judgment obtained by psychophysical tests.

1. INTRODUCTION
Binocular vision can be defined as the subjugation of left and right eyes when analyzing the same region of the scene. The depth perception of a given scene can be considered as a purely psychological process that allows to merge left and right retinal images. The quality of the reconstructed image in the brain is very correlated with the quality of the retinal images serving for the reconstruction. From there, the assessment of the artificial stereoscopic vision becomes a very important challenge in the framework of multimedia application dealing with 3D stereo images. However, to date, very few works have dealt with this problem and there is an important need in the imaging community.

Among the few works in literature, Campisi et al.[1] proposed a basic metric that calculates a weighted average of the quality scores separately between original couple of images and impaired couple of images. Unfortunately, the adopted process cancels the stereoscopic aspect of the assessment. Benoit et al.[2] tried to improve the approach proposed in [1] by using a 2D metric separately on left and right images of the original and the impaired couples. Then, they compare the average of scores with the one obtained on the disparity map using the same 2D metric. This metric has a number of drawbacks related to the exclusion of the stereoscopic aspect by using a 2D metric and the lack of correlation between the perceived depth and the disparity from the observer point of view. Awawedeh et al.[3] proposed a metric using a Pseudo Cepstrum filtering based on the DCT transform. Gorley et al.[4] proposed a metric based on the Stereo Band Limited Contrast (SBLC) calculated using the ratio between average contrast and average luminance for every matched pixel. However, the last two metrics do not deal with the image geometry which is a key aspect in the stereoscopic vision. Also, the color aspect is not taken into account in all the papers found to date. So, they can be used only on a marginal basis i.e. separately on the red, green and blue components.

In this paper, we propose a color metric, essentially based on the behavior of the human visual system (HVS). To achieve this, we focused on the simple and complex cells properties that are the "center" of the binocular fusion of the left and right retinal images. The most efficient way to mathematically model these cells is to use multi-resolution tools. In our case, we adopted the Complex Wavelet Transform (CWT) and the bandelet transform for the definition of the binocular energy model (BEM). This aspect will be detailed in the following sections. The quality scoring of our metric is based the binocular energy comparison computed using BEM.

The paper is organized as follows: Section 2 is a state of the art of simple and complex cells modeling used for the development of the binocular energy models. In section 3, we describe the proposed approach. After a brief description of the bandelet transform, section 5 is dedicated to the construction of our binocular energy model. The following section is dedicated to experimental analysis on the described metric. This paper ends with some conclusions and opens some future directions.

2. STATE-OF-THE-ART OF BINOCULAR ENERGY MODELS
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 [5] 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. 1. According to Campbell et al.[6], 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 cover the OFF region.

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.[7] proposed a complete model to compute the binocular energy. This work has inspired ours for the definition of the BME 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.

3. PROPOSED APPROACH
This section describes the global architecture of our metrics that can be summarized by Fig. 2. Our scheme is based on the behavior of the simple and the complex cells, situated in the human visual cortex. To model these cells and formalize their function, a set of pre-processings is performed on the pairs of stereo images. The
first stage of the process is represented by the CIE $L^*a^*b^*$ color transform, in order to separate the luminance information from the chrominance. In the second stage, we apply the complex wavelet transform (CWT) on the luminance component and the discrete wavelet transform (DWT) on the two color components. These transforms give a complex representation of the luminance, represented by a real part and imaginary part obtained by the CWT, and chrominance represented by a real part represented by the DWT of the component $a^*$ and an imaginary part represented by the DWT of the component $b^*$. The real parts (luminance and chrominance) represent the responses obtained with simple cells of type (ON/OFF) and the imaginary parts (luminance and chroma) represent the responses obtained by simple cells of type (OFF/ON). To define the size and the orientation of every simple cell, one bandelet transform is performed on the coefficients of the CWT and DWT. Every complex cell (luminance or chrominance) takes as input two simple cells, one of type (ON/OFF) and the other of type (OFF/ON) of the same size and the same orientation but with a phase-shift of $\pi/2$ (cf. Section 2). The matching of the complex cells is performed with a binocular energy model; this model is inspired by the works described in the state of the art. This model is applied to the original couple of images and the impaired one. The difference of the binocular energy between both pairs represents the score of our metric described below.

4. 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.

$$\text{Image} = \{\text{Re}[L], \text{Im}[L], \text{Re}[C], \text{Im}[C]\} = \{\text{Re}[\text{CWT}(L^*)], \text{Im}[\text{CWT}(L^*)], \text{DWT}[C(a^*)], \text{DWT}[C(b^*)]\}$$

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 [8] 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 3 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.

5. BME FORMALIZATION

Starting from the definition given above, the model that we propose to calculate the binocular energy is based on the model proposed by Ohzawa [7] and the one proposed by Fleet [9]. 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 ($Re[L(x)]^2$) and the imaginary part of the CWT ($Im[L(x)]^2$). Dyadic square pairs of the chromatic component belong respectively to the real part represented by the DWT ($Re[L(x)]^2$), applied to the component $a^*$ and the imaginary part represented by the DWT ($Im[L(x)]^2$), 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. 1)), Complex-valued response in left and right eyes, are expressed by their amplitude and orientation of the complex function $L(x) = \rho_i(x) \exp(\phi_i(x))$. where:

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

$\rho_i(x)$ is the monocular amplitude of the complex function and $\phi_i(x)$ is the complex phase of the complex function.

$$\phi_i(x) = \arctan(Im[L(x)] / Re[L(x)]) \quad (3)$$

Table 1 summarizes all the parameters described above by giving the appropriate definition. After all the preprocessing steps come 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 [9].

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

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 $L(x) = \rho_i(x) \exp(\phi_i(x))$ and $R(x) = \rho_i(x) \exp(\phi_i(x))$ by their respective definition, we obtain the following equation:

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

$E(x)$ 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 interocular phase difference using a Taylor series of $\phi(x) = \phi(x - d)(Eq. 6)$:

$$\Delta \phi(x, d) = \phi_i(x) - \phi_i(x - d) = \phi_i(x) - \phi_i(x - d) = d \phi_i(x) + O(d^2) \quad (6)$$

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, $\cos(d \phi_i(x))$.

$$\Delta \phi_i(x, d) = \phi_i(x) - \phi_i(x) = \phi_i(x) - \phi_i(x) = \phi_i(x) + O(d^2) \quad (7)$$

In [7], authors 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:

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

$\Delta \psi$ 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:

$$\Delta \phi_i(x, d, \Delta \psi) = \phi_i(x) - \phi_i(x) - \Delta \psi = d \phi_i(x) - \Delta \psi \quad (9)$$

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

$$E(x, d, \Delta \psi) = \rho_i^2(x) + \rho_i^2(x) + 2\rho_i(x)\rho_i(x)\cos(\Delta \phi_i(x) - \Delta \psi) \quad (10)$$

The metric proposed in this paper is obtained thank to the binocular energy of the both pairs of dyadic squares. This latter is compared to the binocular energy computed with the same pairs of dyadic squares, after the two images are impaired. The pairs of dyadic squares have the same behavior of two binocular simple cells, which are matched. The complex cell, which compute the binocular energy of these binocular simple cells, has a different response in the event that the retinal images are degraded or not. Thus we get the equation that computes the score between the two pairs of original and impaired images. It is expressed as follows:

$$S = \sum_{i=1}^{N} \frac{T_{dyadic\ square}^2}{T_{image}^2} \left| E'(x) - E(x) \right| \quad (11)$$

where $T_{dyadic\ square}$ is the size of the dyadic square pairs, $T_{image}$ is the size of the stereo images. $E'(x)$ is the binocular energy of both pairs of the original dyadic squares. $N$ is the number of the pairs of the matched dyadic squares.

6. EXPERIMENTAL RESULTS

We evaluate our stereo metric on the binocular image database used in [10]. The evaluation is performed by comparing the subjective quality scores obtained by our binocular metric with the subjective Mean Opinion Scores (MOS) provided by the authors of [10] for the images shown in figure 5. These images were JPEG coded with seven quality scales (QS: 10, 15, 27, 37, 55, 79 and 100). The MOS vs the objective scores obtained by our metric are shown in figure 4.
The stereoscopic vision does not provoke the same artefacts as the 2D vision in the case of both retinal images are impaired. We can define two types of artifacts: On one hand, artifacts created in homogeneous regions (no geometry) that can create a false impression of depth. On the other hand, artifacts that remove or change the geometry of the image. These artifacts can be created by an intense smoothing of the image creating a difference between the two matched information in the two images.

Our model is sensitive to this kind of transformations. The estimation of the orientation of a dyadic square is realized with a polynomial approximation. This function is sensitive to the modifications of the values of the dyadic squares. These modifications can change its orientation, phase and/or amplitude. Also, the same parameters characterizing the simple and complex cells are used in the binocular fusion. For two matched pairs of dyadic squares, belonging respectively to the left and right images, if both pairs are impaired our model can have two different behaviors. If the impairment provokes a change of orientation between pairs, these pairs are not matched \(E(x) = 0\). When they keep the same orientation, both are matched but their binocular energy is not the same, it is going to decrease because of the change of amplitude and phase.

7. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have proposed a metric with full reference for quality assessment of stereoscopic images. Unlike the stereoscopic metric that can be found in the literature, we assess the quality of 3D images as they are reconstructed in the visual cortex. This is why we focused on the processing performed on the retinal images. Thus, we proposed a model for characterizing the properties of simple and complex cells responsible of the binocular matching. The binocular energy generated by a complex cell depends on the quality of the retinal images captured by simple cells. Our metric compares the binocular energy generated by a complex cell with original images and the impaired images pairs. The average of the binocular energy obtained with all complex cells defines a quality score for the pair of impaired images. As future directions, we are refining our matching model based on additional properties (such as detection of binocular rivalry, stereoscopic constraints to optimize the matching ... ) of the lateral geniculate body.

8. REFERENCES


```
Table 2. objective assessment.

<table>
<thead>
<tr>
<th>Metric</th>
<th>CC</th>
<th>ROCC</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric [10]</td>
<td>0.96</td>
<td>0.92</td>
<td>0.069</td>
</tr>
<tr>
<td>Our metric</td>
<td>0.971</td>
<td>0.95</td>
<td>0.0142</td>
</tr>
</tbody>
</table>
```

Fig. 4. Proposed binocular metric performance, MOS vs predicted MOS (Objective scores).

Fig. 5. Example of stereo image pair.

6.1. Discussion

The stereoscopic vision does not provoke the same artefacts as the 2D vision in the case of both retinal images are impaired. We can define two types of artifacts: On one hand, artifacts created in homogeneous regions (no geometry) that can create a false impression of depth. On the other hand, artifacts that remove or change the geometry of the image. These artifacts can be created by an intense smoothing of the image creating a difference between the two matched information in the two images.

Our model is sensitive to this kind of transformations. The estimation of the orientation of a dyadic square is realized with a polynomial approximation. This function is sensitive to the modifications of the values of the dyadic squares. These modifications can change its orientation, phase and/or amplitude. Also, the same parameters characterizing the simple and complex cells are used in the binocular fusion. For two matched pairs of dyadic squares, belonging respectively to the left and right images, if both pairs are impaired our model can have two different behaviors. If the impairment provokes a change of orientation between pairs, these pairs are not matched \(E(x) = 0\). When they keep the same orientation, both are matched but their binocular energy is not the same, it is going to decrease because of the change of amplitude and phase.

7. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have proposed a metric with full reference for quality assessment of stereoscopic images. Unlike the stereoscopic metric that can be found in the literature, we assess the quality of 3D images as they are reconstructed in the visual cortex. This is why we focused on the processing performed on the retinal images. Thus, we proposed a model for characterizing the properties of simple and complex cells responsible of the binocular matching. The binocular energy generated by a complex cell depends on the quality of the retinal images captured by simple cells. Our metric compares the binocular energy generated by a complex cell with original images and the impaired images pairs. The average of the binocular energy obtained with all complex cells defines a quality score for the pair of impaired images. As future directions, we are refining our matching model based on additional properties (such as detection of binocular rivalry, stereoscopic constraints to optimize the matching ...) of the lateral geniculate body.

8. REFERENCES