# Disparity based stereo image retrieval through univariate and bivariate models

A. Chaker, Mounir Kaaniche, A. Benazza-Benyahia

# Disparity based stereo image retrieval through univariate and bivariate models

A. Chaker[1], M. Kaaniche[2], A. Benazza-Benyahia[1]

[1] *COSIM Lab., SUP'COM, Carthage Univ.,*
*Cité Technologique des Communications,*
*Tunisia*
*amani.chaker@supcom.tn*
*benazza.amel@supcom.rnu.tn*
[2] *Institut Galilée, L2TI,*
*Université Paris 13,*
*Sorbonne Paris Cité, France*
*mounir.kaaniche@univ-paris13.fr*

**Abstract**

The widespread use of stereovision in various application fields has led to the constitution of very huge stereo image databases. Therefore, the design of effective content based image retrieval system devoted to stereo pairs becomes an issue of importance. To this end, we propose in this paper two retrieval methods which combine the visual contents of the stereo images with their corresponding disparity information. After modeling the distribution of their associated wavelet coefficients by the generalized Gaussian statistical model, the resulting distribution parameters are selected as salient features. While the two views are processed separately through a univariate modeling in the first method, the second one exploits the correlation between the views by resorting to a bivariate modeling. Experimental results show the benefits which can be drawn from the proposed retrieval approaches.

*Preprint submitted to Signal Processing: Image Communication* *December 16, 2014*

## 1. Introduction

Recent developments of stereoscopic display technologies have accelerated the usage of Stereo Images (SI) in various application fields such as 3DTV, telepresence in videoconferences and stereo geographical information systems. Stereoscopic image display offers a simple way of presenting the depth information in a real world scene. Indeed, the disparity information which corresponds to the displacement that exists between the corresponding pixels of the left and right images, allows to provide the 3D-depth information of the scene. As a result, very huge stereoscopic image databases are continuously generated. For example, a single view of a scene acquired by the IKONOS satellite corresponds to 360 MB every 3 or 4 days. Hence, there is a strong need for both managing and storing these large amounts of stereo data [1].

Conventional Content-Based Image retrieval (CBIR) systems allow a convenient and efficient data access by organizing images based only on their visual contents [2]. On the other hand, in order to reduce the memory requirements, images are saved in a compressed format. In this respect, Wavelet Transforms (WT) have been found to be an efficient tool to provide very compact representations of still images and, they have been adopted in most of the recent image compression algorithms [3]. Therefore, it seems interesting to design a CBIR system operating in the WT domain. Thus, the objective will consist in defining relevant signatures from the resulting wavelet coefficients. For this purpose, different wavelet-based image retrieval approaches have been proposed [4, 5, 6, 7]. For instance, in [5], the energy of the subbands are combined with the color autocorrelogram. In [6], the authors use a B-spline wavelet transform and fractal signature. In [8], an edge histogram descriptor is computed to gather the information of dominant edge orientations. Besides, another method, called wavelet correlogram, based on a fusion of multiresolution image decomposition and color correlation histogram has been introduced in [7]. However, it can be pointed out that most popular techniques resort to a statistical modeling of the distribution of the wavelet coefficients [9, 10, 11]. To this end, several models such as the generalized Gaussian distribution [9], the Gamma distribution [12, 13], and the Weibull one [14] have been used. It is worth pointing out that the effectiveness of these techniques have been studied in the case of mono-view images, including still and multicomponent images and video sequences [15]. However, to the best of our knowledge, there is only one research work developed in the context of SI [16]. More precisely, the reported method consists of two steps: a conventional CBIR system is applied to *only* one view (for example the left one). Then, the obtained results are refined by comparing the histograms of the estimated disparity maps. However, such retrieval method presents a drawback as the visual contents of the right image are not directly exploited. To alleviate this shortcoming, we investigate in this paper different techniques to improve the efficiency of a content-based stereo images retrieval system. Our major contribution is to exploit the dependencies between the two views thanks to the disparity information. More precisely, in order to extract relevant features allowing an accurate SI retrieval, we propose to use *simultaneously* the visual contents of the left and right images as well as their related disparity fields. To this end, two retrieval strategies are addressed. In the first one, the subbands of the left view, the right one and, the disparity

map are modeled by a generalized Gaussian distribution [17]. The resulting distribution parameters, considered as features of the SI pair, are combined at the retrieval stage. While the two views are modeled *independently* by using a univariate statistical model in the first strategy, the second one aims to exploit the high statistical dependencies between the two views. After defining the appropriate vectors of wavelet coefficients, extracted from the right and left (or compensated left) subbands at the same scale and orientation, we propose to resort to a *bivariate* modeling in order to capture the cross-view dependencies. At this level, it is important to note that the joint modeling of multivalued wavelet coefficients has already been investigated in different applications involving only mono-view images such as denoising [18, 19, 20] or retrieval of still and multicomponent images [11, 21, 22, 23], but there is no reported work related to the context of SI.
The remainder of this paper is organized as follows. In Section 2, we first give a brief description of conventional CBIR operating in the WT domain. Then, the straightforward extension of this system to the context of SI is discussed. In Sections 3 and 4, we describe the proposed disparity-based SI retrieval approaches based on univarite and bivariate statistical modeling, respectively. Finally, the performance of the proposed approaches is illustrated in Section 5 and some conclusions are drawn in Section 6.

## 2. Conventional wavelet-based CBIR system

### *2.1. Wavelet-based representation*

The Lifting Scheme (LS) is a flexible tool for computing the discrete WT [24]. LS was found to be a very effective structure for encoding still and stereo images [25], and it has been retained in the JPEG2000 image compression standard [3]. A generic LS is performed in three steps namely *split*, *predict* and *update*. At the first step, the input 1D signal $a_j(k)$ is divided into two subsets composed respectively of even $a_j(2k)$ and odd samples $a_j(2k+1)$. Then, thanks to the local correlation, the samples of one subset (say the odd ones) are predicted from the neighboring even samples. Thus, the prediction error, referred to as detail signal, is computed as follows:

$$d_{j+1}(k) = a_j(2k+1) - \mathbf{p}_j^\top \mathbf{a}_j(k) \tag{1}$$

where $\mathbf{p}_j$ is the prediction weighting vector and $\mathbf{a}_j(k)$ is a reference vector containing some even samples used in the predict step. Finally, the update step produces the approximation signal $a_{j+1}(k)$ by smoothing the even samples using the detail coefficients:

$$a_{j+1}(k) = a_j(2k) + \mathbf{u}_j^\top \mathbf{d}_{j+1}(k) \tag{2}$$

where $\mathbf{u}_j$ is the update weighting vector and $\mathbf{d}_{j+1}(k)$ is a reference vector containing the detail coefficients used in the update step. Note that, the compactness ability of a lifting scheme is related to the choice of the prediction and update weights. The extension of this 1D structure to 2D signals is straightforward: the lifting steps are generally performed along the lines then the columns (or inversely) of the image in a separable manner leading to an approximation subband and three detail subbands oriented horizontally, vertically and diagonally. This procedure is again repeated on the approximation sub-images, over $J$ resolution levels, leading to $(3J+1)$ subbands.

*2.2. Feature extraction and similarity measure*

In this subsection, we only focus on mono-view images. The key step in a wavelet-based CBIR system consists of extracting salient features from the wavelet coefficients of the images. As aforementioned, a statistical framework could be adopted to model the wavelet coefficients of the different subbands. For instance, the Generalized Gaussian (GG) distribution has been extensively used [26]. Thus, in a given subband $j$, the wavelet coefficients are modeled by a GG distribution whose probability density function (pdf) is defined by:

$$\forall \xi \in \mathbb{R} \quad f_j(\xi) = \frac{\beta_j}{2\alpha_j \Gamma(1/\beta_j)} e^{-(|\xi|/\alpha_j)^{\beta_j}} \tag{3}$$

where $\Gamma(z) \triangleq \int_0^{+\infty} t^{z-1} e^{-t} dt$ represents the Gamma function, $\alpha_j$ and $\beta_j$ are respectively the scale and shape parameters. The latter two parameters can be estimated by using the Maximum Likelihood technique [9]. Following the modeling step, the feature vector of each image of the database is composed of the distribution parameters of all the detail subbands $(\alpha_j, \beta_j)_{1 \le j \le 3J}$.
Finally, for the different subbands $j$ with $j \in \{1, \ldots, 3J\}$, an appropriate metric should be defined in order to measure the similarity between the pdf $f_j^{\mathrm{db}}$ of an image in the database $I^{\mathrm{db}}$ and the pdf $f_j^{\mathrm{q}}$ of the query image $I^{\mathrm{q}}$. Very often, the Kullback-Leibler Divergence (KLD) is retained [9, 23, 27, 28]. In the case of the GG distribution, the KLD, denoted by $\widetilde{D}_{\mathrm{GG}}$, is expressed as:

$$\widetilde{D}_{\mathrm{GG}}(\alpha_j^{\mathrm{db}}, \beta_j^{\mathrm{db}} \parallel \alpha_j^{\mathrm{q}}, \beta_j^{\mathrm{q}}) \triangleq \log\left(\frac{\beta_j^{\mathrm{db}} \alpha_j^{\mathrm{q}} \Gamma(1/\beta_j^{\mathrm{q}})}{\beta_j^{\mathrm{q}} \alpha_j^{\mathrm{db}} \Gamma(1/\beta_j^{\mathrm{db}})}\right) - \frac{1}{\beta_j^{\mathrm{db}}} + \left(\frac{\alpha_j^{\mathrm{db}}}{\alpha_j^{\mathrm{q}}}\right)^{\beta_j^{\mathrm{q}}} \frac{\Gamma((\beta_j^{\mathrm{q}}+1)/\beta_j^{\mathrm{db}})}{\Gamma(1/\beta_j^{\mathrm{db}})} \tag{4}$$

where $(\alpha_j^{\mathrm{db}}, \beta_j^{\mathrm{db}})$ and $(\alpha_j^{\mathrm{q}}, \beta_j^{\mathrm{q}})$ represent the distribution parameters of $f_j^{\mathrm{db}}$ and $f_j^{\mathrm{q}}$, respectively. Thus, the resulting similarity measure $D_{GG}$ between the two images $I^{\mathrm{db}}$ and $I^{\mathrm{q}}$ is deduced as follows:

$$D_{\mathrm{GG}}(I^{\mathrm{db}}, I^{\mathrm{q}}) = \sum_{j=1}^{3J} \widetilde{D}_{\mathrm{GG}}(\alpha_j^{\mathrm{db}}, \beta_j^{\mathrm{db}} \parallel \alpha_j^{\mathrm{q}}, \beta_j^{\mathrm{q}}). \tag{5}$$

*2.3. A straightforward extension to SI*

Let us now proceed to the retrieval problem in the case of database composed of stereo images. A straightforward solution consists in *separately* applying the aforementioned conventional CBIR system to each view. More precisely, the retrieval procedure aims at comparing the left and right images of the query stereo pair $(I^{(\mathrm{l,q})}, I^{(\mathrm{r,q})})$ to those of the database $(I^{(\mathrm{l,db})}, I^{(\mathrm{r,db})})$. Thus, after extracting their corresponding feature vectors $(\alpha_j^{(\mathrm{l,q})}, \beta_j^{(\mathrm{l,q})})_{1\le j\le 3J}$, $(\alpha_j^{(\mathrm{r,q})}, \beta_j^{(\mathrm{r,q})})_{1\le j\le 3J}$, $(\alpha_j^{(\mathrm{l,db})}, \beta_j^{(\mathrm{l,db})})_{1\le j\le 3J}$ and $(\alpha_j^{(\mathrm{r,db})}, \beta_j^{(\mathrm{r,db})})_{1\le j\le 3J}$, the similarity criterion $D_{\mathrm{GG}}^{(\mathrm{r,l})}$ can be simply obtained by computing the KL divergences defined on the right and left images:

$$D_{\mathrm{GG}}^{(\mathrm{r,l})} = D_{\mathrm{GG}}(I^{(\mathrm{r,db})}, I^{(\mathrm{r,q})}) + D_{\mathrm{GG}}(I^{(\mathrm{l,db})}, I^{(\mathrm{l,q})}). \tag{6}$$

## 3. Proposed disparity-based retrieval approaches through univariate model

*3.1. Motivation*

The aforementioned intuitive strategy for indexing a SI pair is not so efficient since only the left and right images are used during the comparison process. Indeed, an important feature of the stereoscopic system, which corresponds to the estimated disparity map, has not been taken into account. This kind of information is inversely proportional to the depth of the objects in the scene [29] and, hence, it is expected to provide salient features. From this point of view, a more efficient retrieval method could be designed by incorporating the disparity information in the feature vector. In what follows, and before describing our disparity-based retrieval approaches, we present the disparity estimation issue.

### *3.2. Disparity estimation*

The principle of the disparity estimation is to find for a given pixel in the right image the best corresponding one in the left image. When the stereo images are rectified, the disparity field is limited to a horizontal component that will be denoted by $u$. Several methods have been proposed to solve the stereo matching problem [30, 31]. For instance, global methods have been extensively employed. They consist in minimizing a global energy function over the entire image based on some specific algorithms, such as the graph-cuts [30] and variational methods [31]. While most of the existing methods operate in the spatial domain, some wavelet-based disparity estimation methods have been recently attracted much attention thanks to the hierarchical and scale-space localization properties of the wavelets [32, 33].

With the ultimate goal of generating a *dense* and *smooth* disparity map while preserving the depth discontinuities, we have resorted to the estimation method presented recently in [34] to compute the disparity maps of stereo image database. We should note that, among the developed disparity estimation techniques, it is important to employ an efficient method that guarantees the smoothness property of the produced disparity map. Indeed, such property allows us to interpret this map as an image, and therefore can undergo a wavelet decomposition in order to be efficiently exploited in the retrieval process of the stereo images.

### 3.3. Disparity-based retrieval strategies

Once the disparity maps are available, two approaches could be considered in order to exploit the disparity information.

In the first one, the disparity is *implicitly* taken into account by computing the *compensated* left image in the wavelet domain $I_j^{(\mathrm{c})}$ from the multiresolution representation of the left image $I_j^{(\mathrm{l})}$ as follows:

$$I_j^{(\mathrm{c})}(x,y) = I_j^{(\mathrm{l})}(x + u_j(x,y), y) \tag{7}$$

where the disparity $u_j$ is obtained by sampling and dividing by $2^j$ the initial disparity field $u$:

$$u_j(x,y) = \frac{1}{2^j} u(2^j x, 2^j y). \tag{8}$$

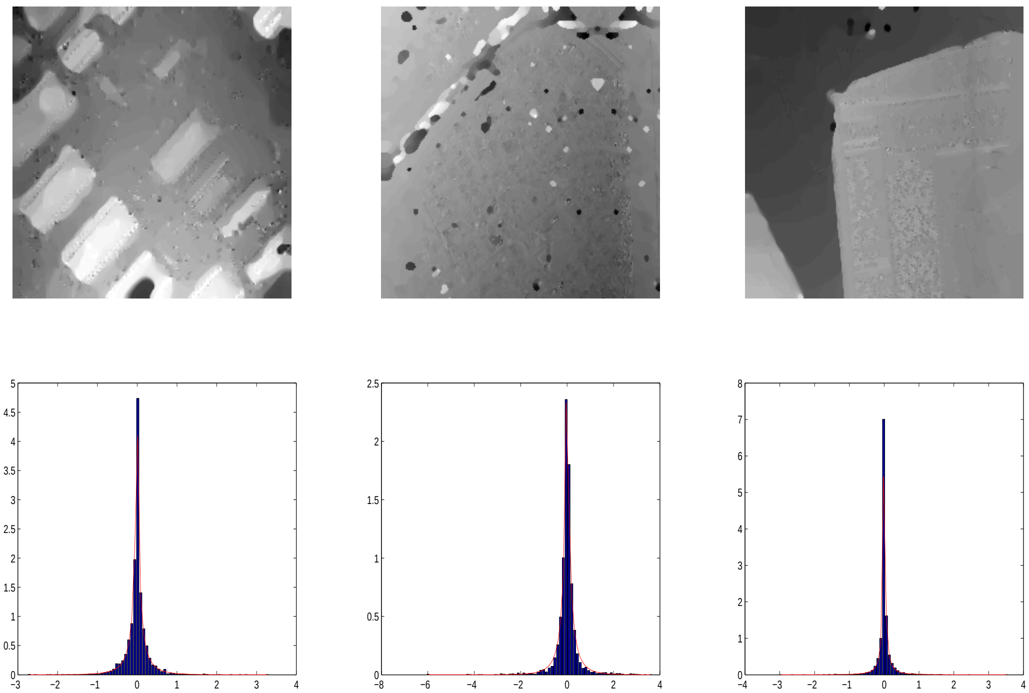
Then, the GG distribution parameters of the different detail subbands of the right image $(\alpha_j^{(\mathrm{r})}, \beta_j^{(\mathrm{r})})_{1 \le j \le 3J}$ and the compensated left one $(\alpha_j^{(\mathrm{c})}, \beta_j^{(\mathrm{c})})_{1 \le j \le 3J}$ are extracted. Finally, the retrieval procedure for a given query stereo pair $(I^{(\mathrm{l,q})}, I^{(\mathrm{r,q})})$ aims at finding the best stereo pairs $(I^{(\mathrm{l,db})}, I^{(\mathrm{r,db})})$ that minimize the KL divergences, $D_{\mathrm{GG}}^{(\mathrm{r,c})}$, defined on the right image and the compensated left one:

$$D_{\mathrm{GG}}^{(\mathrm{r,c})} = D_{\mathrm{GG}}(I^{(\mathrm{r,db})}, I^{(\mathrm{r,q})}) + D_{\mathrm{GG}}(I^{(\mathrm{c,db})}, I^{(\mathrm{c,q})}) \tag{9}$$

where $I^{(\mathrm{c,q})}$ and $I^{(\mathrm{c,db})}$ represent respectively the compensated left images of the query and candidate stereo pairs.

Unlike the first approach, the second one aims to exploit *explicitly* the disparity information by extracting a relevant signature from the disparity map and, combining it with the features defined previously on the SI pair. To this end, since a smooth disparity map is produced while preserving the depth discontinuities, we propose to apply a wavelet transform to the estimated disparity field. After performing an intensive experiments on a large data set of the estimated disparity maps, we have noticed that their wavelet coefficients can also be successfully modeled by a GG distribution. Indeed, to objectively assess the appropriateness of the GG model, we have applied the Kolmogorov-Smirnov (KS) goodness-of-fit test [35, 36]. Note that the KS test is based on comparing the cumulative distribution functions. As an example, by taking three disparity maps and considering their horizontal detail coefficients, Fig. 1 shows the histograms of these coefficients (in blue) and the fitted GG distributions (in red) as well as their resulting KS measures. By performing this test on all the disparity maps of the data set, the average of the KS values is equal to 0.1 which is very small. This confirms that the GG distribution is well-suited for modeling the disparity maps.

Based on these observations, we select the distribution parameters of all the resulting detail subbands $(\alpha_j^{(\mathrm{u})}, \beta_j^{(\mathrm{u})})_{1 \leq j \leq 3J}$ to characterize the disparity map $u$. Therefore, at the retrieval step, the candidate stereo pairs of the database $(I^{(\mathrm{r,db})}, I^{(\mathrm{l,db})})$ that are similar to the query one $(I^{(\mathrm{r,q})}, I^{(\mathrm{l,q})})$ are determined by comparing the right and left images as well as their associated disparity maps $u^{(\mathrm{db})}$ and $u^{(\mathrm{q})}$. More precisely, we propose to define the similarity



$\beta^{(u)} = 0.44, \alpha^{(u)} = 0.015$ KS=0.04 | $\beta^{(u)} = 0.52, \alpha^{(u)} = 0.048$ KS=0.02 | $\beta^{(u)} = 0.27, \alpha^{(u)} = 0.0002$ KS=0.09

Figure 1: Modeling the distribution of the horizontal detail subband at the second resolution level for three examples of disparity maps using the GG model.

measure as follows:

$$
\begin{aligned}
D_{\mathrm{GG}}^{(\mathrm{r,l,u})} &= a \cdot D_{\mathrm{GG}}(I^{(\mathrm{r,db})}, I^{(\mathrm{r,q})}) + b \cdot D_{\mathrm{GG}}(I^{(\mathrm{l,db})}, I^{(\mathrm{l,q})}) + c \cdot D_{\mathrm{GG}}(u^{(\mathrm{db})}, u^{(\mathrm{q})}) \\
&= \textstyle\sum_{j=1}^{3J} \Big( a \cdot \widetilde{D}_{\mathrm{GG}}(\alpha_j^{(\mathrm{r,db})}, \beta_j^{(\mathrm{r,db})} \parallel \alpha_j^{(\mathrm{r,q})}, \beta_j^{(\mathrm{r,q})}) \\
&\qquad + b \cdot \widetilde{D}_{\mathrm{GG}}(\alpha_j^{(\mathrm{l,db})}, \beta_j^{(\mathrm{l,db})} \parallel \alpha_j^{(\mathrm{l,q})}, \beta_j^{(\mathrm{l,q})}) \\
&\qquad + c \cdot \widetilde{D}_{\mathrm{GG}}(\alpha_j^{(\mathrm{u,db})}, \beta_j^{(\mathrm{u,db})} \parallel \alpha_j^{(\mathrm{u,q})}, \beta_j^{(\mathrm{u,q})}) \Big)
\end{aligned}
\tag{10}
$$

where $a$, $b$ and $c$ are three positive weights.

To conclude this part, we should note that this first category of SI retrieval approaches are based on univariate model since, up to now, the right and left images are modeled separately without taking into account the cross-view dependencies.

## 4. Proposed disparity-based retrieval approaches through bivariate model

### *4.1. Strategy*

In the previous section, the SI retrieval techniques have been developed through univariate statistical approach by modeling the right image independently of the left one. Since the left and right views correspond to the same scene, and so have similar contents, the wavelet coefficients of both images could present strong statistical dependencies. For example, such dependencies between the two images have already been successfully exploited for SI compression purposes [25, 37]. Therefore, a suitable statistical model should be employed to capture the dependencies across the wavelet coefficients of the left and right images.

### *4.2. Bivariate generalized Gaussian model*

First, we should recall that recent works on color image retrieval [11, 21, 22, 23] have shown that the distribution of the wavelet coefficients could be modeled by some specific multivariate models. The latter outperform the conventional univariate approach in terms of accuracy since they account for both the spatial and the cross-channel dependencies. Motivated by these reasons, we propose in this work to resort to a bivariate model to further exploit the dependencies between the wavelet coefficients of the left and right images. Generally, the choice of the appropriate model should fulfill the following constraints:

- The retained model should reflect accurately the sparsity of the wavelet coefficients of each view and also the cross-view dependencies.
- The model structure should enable a straightforward estimation of the parameters from the coefficients in each subband.
- The bivariate model should allow an easy computation of a meaningful similarity measure. For example, knowing that the KLD has been widely used in mono-view image retrieval [9, 23, 27, 28], it would be interesting to select a bivariate model from which a closed form expression of the KLD could be derived in order to facilitate the retrieval stage by avoiding Monte-Carlo estimation procedures.

Based on the previous points, we find that using the well-known Bivariate Generalized Gaussian (BGG) distribution could be an appropriate way for characterizing the dependencies between the wavelet coefficients of the left and right images.

To introduce this model, let us denote by $\mathbf{w}_j$ the bivariate vector composed of the wavelet coefficients of two correlated components, for each subband $j$. We assume that the set of coefficient vectors $\mathbf{w}_j$ in each subband constitutes an independent identical distributed sample of a random vector $\mathbf{W}_j$. Under the hypothesis of a zero-mean vector, the generic expression of the probability density function $f_{\mathbf{W}_j}$ of the BGG distribution is given by:

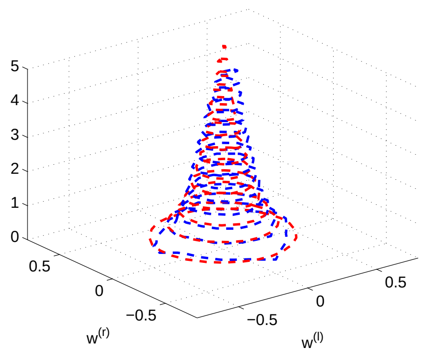
$$\forall \mathbf{w} \in \mathbb{R}^2,\ f_{\mathbf{W}_j}(\mathbf{w}; \beta_j, \mathbf{\Sigma}_j) = \frac{2}{\pi\Gamma(1+\frac{1}{\beta_j})2^{1+\frac{1}{\beta_j}}} \mid \mathbf{\Sigma}_j \mid^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{w}^\intercal \mathbf{\Sigma}_j^{-1} \mathbf{w})^{\beta_j}\right). \tag{11}$$

where $\beta_j > 0$ denotes the shape parameter and, $\mathbf{\Sigma}_j$ is a symmetric positive-definite matrix of size $2 \times 2$ (the scaling matrix). Note that these parameters can be estimated using the moment method [38] or the maximum likelihood criterion [22].

In order to validate the appropriateness of the BGG model, we have conducted the multivariate Kolmogorov-Smirnov (KS) test [36] on the stereo images database. Indeed, by taking the horizontal detail subbands of the left and right images of four stereo pairs, Fig. 2 illustrates the empirical bivariate histograms (in blue) fitted with the BGG distribution (in red), and provides their associated KS measures. By repeating the same test on the whole set of stereo images in the database, an average KS value of about 0.09 is obtained. These results corroborate also that stereo wavelet subbands can be well modeled by a BGG distribution.

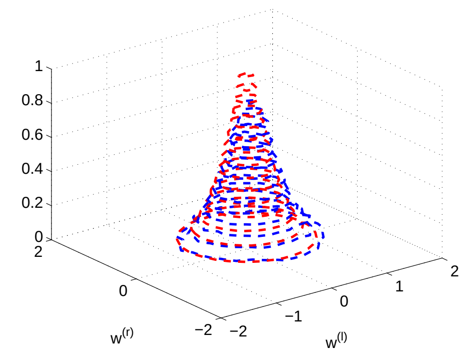
It is important to note that a closed form expression of the KLD is available for such BGG model. Indeed, for two BGG distributions with parameters $(\beta_j^{\mathrm{q}}, \mathbf{\Sigma}_j^{\mathrm{q}})$ and $(\beta_j^{\mathrm{db}}, \mathbf{\Sigma}_j^{\mathrm{db}})$, the KLD is given by [22, 39]:

$$\begin{aligned}
\widetilde{D}_{\mathrm{BGG}}(\beta_j^{\mathrm{db}}, \mathbf{\Sigma}_j^{\mathrm{db}} \| \beta_j^{\mathrm{q}}, \mathbf{\Sigma}_j^{\mathrm{q}}) = \ln & \left[ \frac{\Gamma\left(\frac{1}{\beta_j^{\mathrm{q}}}\right)}{\Gamma\left(\frac{1}{\beta_j^{\mathrm{db}}}\right)} 2^{\left(\frac{1}{\beta_j^{\mathrm{q}}} - \frac{1}{\beta_j^{\mathrm{db}}}\right)} \left(\frac{|\mathbf{\Sigma}_j^{\mathrm{q}}|}{|\mathbf{\Sigma}_j^{\mathrm{db}}|}\right)^{\frac{1}{2}} \frac{\beta_j^{\mathrm{db}}}{\beta_j^{\mathrm{q}}} \right] - \frac{1}{\beta_j^{\mathrm{db}}} \\
+ & \left[ 2^{\left(\frac{\beta_j^{\mathrm{q}}}{\beta_j^{\mathrm{db}}} - 1\right)} \frac{\Gamma\left(\frac{\beta_j^{\mathrm{q}} + 1}{\beta_j^{\mathrm{db}}}\right)}{\Gamma\left(\frac{1}{\beta_j^{\mathrm{db}}}\right)} \times \left(\frac{\mu_1 + \mu_2}{2}\right)^{\beta_j^{\mathrm{q}}} \right. \\
\times & \left. {}_2\mathrm{F}_1\left(\frac{1 - \beta_j^{\mathrm{q}}}{2}, \frac{-\beta_j^{\mathrm{q}}}{2}; 1; \left(\frac{\mu_1 - \mu_2}{\mu_1 + \mu_2}\right)^2\right) \right]
\end{aligned} \tag{12}$$



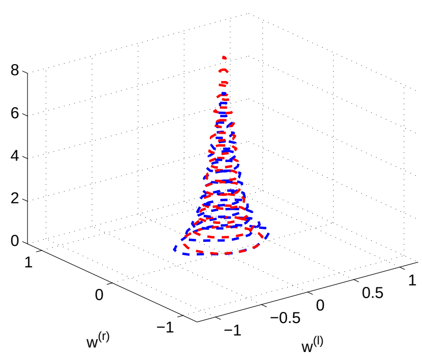

$\beta_1 = 0.44$, $\mathbf{\Sigma}_1 = \begin{pmatrix} 0.0029 & 0.0001 \\ 0.0001 & 0.0029 \end{pmatrix}$

KS=0.05



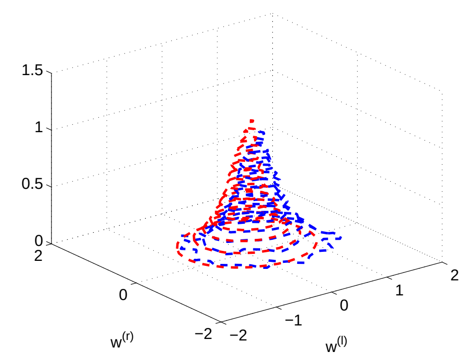

$\beta_1 = 0.61$, $\mathbf{\Sigma}_1 = \begin{pmatrix} 0.0574 & 0.0002 \\ 0.0002 & 0.0548 \end{pmatrix}$

KS=0.02




$\beta_1 = 0.43$, $\mathbf{\Sigma}_1 = \begin{pmatrix} 0.0014 & -0.000 \\ -0.000 & 0.0012 \end{pmatrix}$

KS= 0.04




$\beta_1 = 0.48$, $\mathbf{\Sigma}_1 = \begin{pmatrix} 0.0255 & -0.0006 \\ -0.0006 & 0.0194 \end{pmatrix}$

KS=0.03

Figure 2: Empirical bivariate histogram of the horizontal wavelet coefficients of the left and right images (denoted here by $w^{(\mathrm{l})}$ and $w^{(\mathrm{r})}$)(in blue) fitted with a BGG density (in red) for four different stereo images as well as their resulting KS measures.

where $\mu_1$ and $\mu_2$ are the inverse of the eigenvalues of $(\mathbf{\Sigma}_j^{\mathrm{db}})^{-1}\mathbf{\Sigma}_j^{\mathrm{q}}$ and ${}_2\mathrm{F}_1$ represents the Gauss hypergeometric function [40].

*4.3. An improved disparity-based retrieval strategies*

Now, we will describe three retrieval strategies based on the selected BGG distribution.

Intuitively, in the first one, the bivariate vector, defined by $\mathbf{w}_j^{(\mathrm{r,l})} = \left(w_j^{(\mathrm{r})}, w_j^{(\mathrm{l})}\right)^\top$ is composed of the wavelet coefficients in the right and left images for each subband $j$. Let us denote by $\left(\boldsymbol{\Sigma}_j^{(\mathrm{r,l})}, \beta_j^{(\mathrm{r,l})}\right)$ the distribution parameters of vector $\mathbf{w}_j^{(\mathrm{r,l})}$. Thus, in the indexing step, the comparison between the stereo pair in the database, characterized by its feature vector $\left(\beta_j^{(\mathrm{r,l,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,db})}\right)_{1\leq j\leq 3J}$ and the query one parameterized by the feature $\left(\beta_j^{(\mathrm{r,l,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,q})}\right)_{1\leq j\leq 3J}$ is achieved by computing the global KLD:

$$D_{\mathrm{BGG}}^{(\mathrm{r,l})} = \sum_{j=1}^{3J} \widetilde{D}_{\mathrm{BGG}}(\beta_j^{(\mathrm{r,l,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,db})} \parallel \beta_j^{(\mathrm{r,l,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,q})}). \tag{13}$$

Since the highly similar pixels of the left and right images are located at different spatial positions identified by the disparity information, it would be more interesting to focus on the right image $I^{(\mathrm{r})}$ and the compensated left one $I^{(\mathrm{c})}$ by using their wavelet coefficients to build the bivariate vector $\mathbf{w}_j^{(\mathrm{r,c})} = \left(w_j^{(\mathrm{r})}, w_j^{(\mathrm{c})}\right)^\top$, for each subband $j$. Thus, in this second retrieval strategy, the feature vectors deduced from the right and compensated left images of the query stereo data $\left(\beta_j^{(\mathrm{r,c,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,c,q})}\right)_{1\leq j\leq 3J}$ will be compared to those of the database stereo pair $\left(\beta_j^{(\mathrm{r,c,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,c,db})}\right)_{1\leq j\leq 3J}$ by computing the following measure:

$$D_{\mathrm{BGG}}^{(\mathrm{r,c})} = \sum_{j=1}^{3J} \widetilde{D}_{\mathrm{BGG}}(\beta_j^{(\mathrm{r,c,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,c,db})} \parallel \beta_j^{(\mathrm{r,c,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,c,q})}). \tag{14}$$

Although it is clear that stereo images contain nearly similar contents since they correspond to the same scene, there are some areas in one image that are absent in the other one, referred to as occluded areas. This occlusion effect is well known in stereovision problems and is due to the different viewpoints of the cameras and the presence of discontinuities in the scene. Generally, increasing the dependencies between the two components of the bivariate vector leads to an efficient retrieval procedure. For this reason, we propose in the third strategy to improve the previous one by taking into account the effect of the occlusion. Note that the occluded areas are mainly located at the boundaries of the SI. Let us denote by $w_j^{(\mathrm{r_{ocl}})}$ and $w_j^{(\mathrm{c_{ocl}})}$ the wavelet coefficients of the right and compensated left images resulting from the removal of the occluded regions. These coefficients will constitute the components of the bivariate vector $\mathbf{w}_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}})} = \left(w_j^{(\mathrm{r_{ocl}})}, w_j^{(\mathrm{c_{ocl}})}\right)^\top$. After estimating their associated model parameters and building the feature vectors for the query stereo pair $\left(\beta_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{q})}, \mathbf{\Sigma}_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{q})}\right)_{1\le j\le 3J}$ and the candidate one $\left(\beta_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{db})}, \mathbf{\Sigma}_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{db})}\right)_{1\le j\le 3J}$, the global KLD is therefore obtained:

$$D_{\mathrm{BGG}}^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}})} = \sum_{j=1}^{3J} \widetilde{D}_{\mathrm{BGG}}(\beta_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{db})}, \mathbf{\Sigma}_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{db})} \parallel \beta_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{q})}, \mathbf{\Sigma}_j^{(\mathrm{r_{ocl}},\mathrm{c_{ocl}},\mathrm{q})}). \tag{15}$$

Finally, as it was performed with disparity-based retrieval approaches through univariate model, these three strategies should further incorporate the disparity information into their feature vector. In other words, in addition to the BGG distribution parameters of the right and left (or compensated left) images, it would be interesting to consider also the GG distribution parameters of the disparity maps $u^{(\mathrm{db})}$ and $u^{(\mathrm{q})}$ respectively associated to the database and query SI. Consequently, during the comparison process between a query and a candidate stereo pair, we should add the measure $D_{\mathrm{GG}}(u^{(\mathrm{db})}, u^{(\mathrm{q})})$ to each of the KL divergences given by Eq. (13), (14) and (15).

## 5. Experimental results

### *5.1. Experimental setup*

Since there are no SI databases publicly available to evaluate the performance of SI retrieval methods, we have built a database which can be downloaded from [1]. This database is composed of real SI pairs of size $248 \times 248$ taken from various sources. The images correspond to a variety of contents including natural scenes [2], [3] and [4], man-made objects available at the Middlebury stereo vision website[5], and SPOT5 scenes. According to their texture, these stereo images have been divided into 17 classes, with 40 pairs per class, such as wooded area, tree, bushes, mountains, urban area. Note that SI of the same class are taken from the same scene. An example of some right images of different classes in the database is shown in Fig. 3.

As explained in Section 3.2, the method developed in [34] has been used to generate the disparity maps. The retrieval performance is evaluated in terms of precision $PR = \frac{R^r}{R}$ and recall $RC = \frac{R^r}{R^t}$, where $R^r$ is the number of output pairs considered as relevant, $R^t$ is the total number of relevant pairs in the database and $R$ denotes the number of returned pairs. A retrieved pair is considered as relevant if it belongs to the same category of the query pair. Note that the query images are taken from the database.

[1] `http://www-l2ti.univ-paris13.fr/~kaaniche/Download.html`

[2] http://www.mi.auckland.ac.nz/EISATS/

[3] http://vasc.ri.cmu.edu/idb/html/stereo/index.html

[4] http://vasc.ri.cmu.edu/idb/html/jisct/

[5] http://vision.middlebury.edu/stereo/

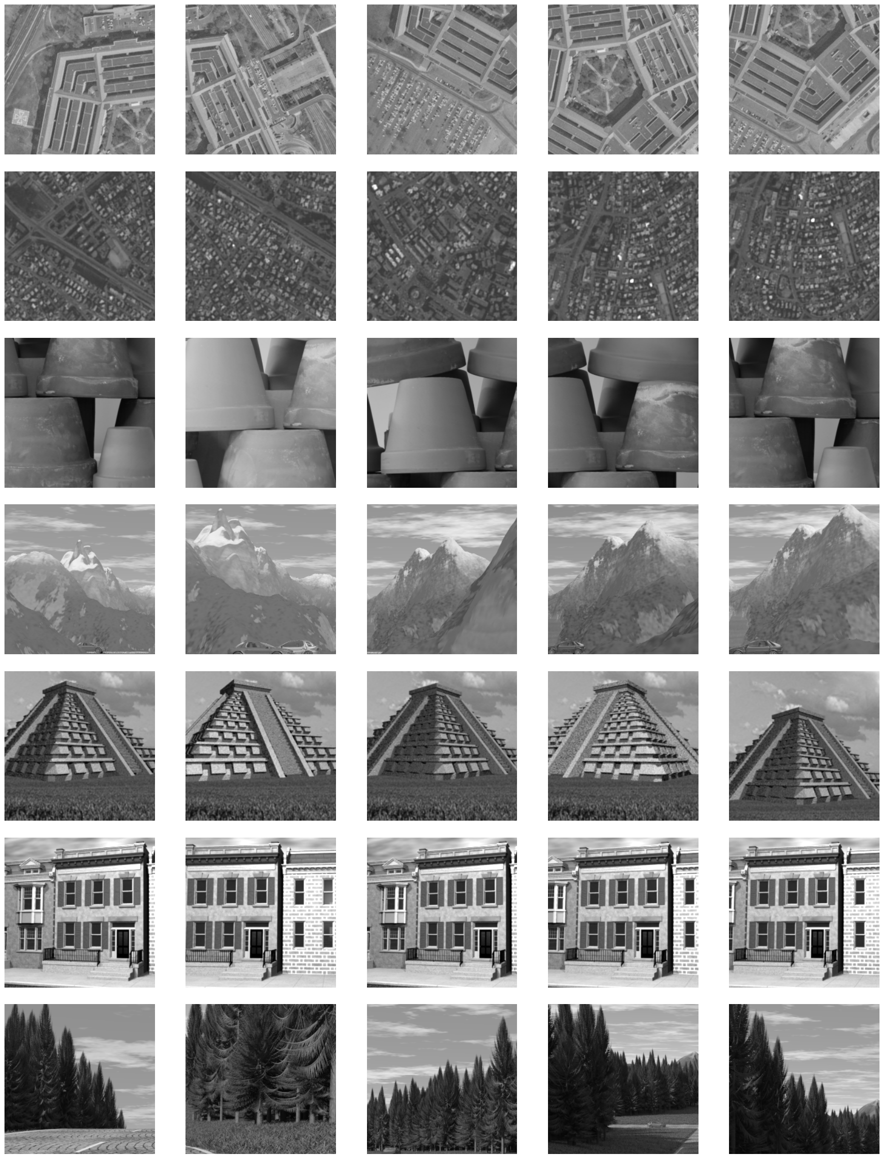


Figure 3: Some samples of right images for different classes in the database. From top to bottom: pentagon, urban area, flowerpots, mountains, pyramid, buildings and tree.

In order to show the benefits of using the disparity map in a SI retrieval system, two rounds of experiments are performed. The first one aims at illustrating the behavior of the univariate modeling-based retrieval approaches described in Section 3. The objective of the second one is to validate the interest of the bivariate modeling presented in Section 4. In what follows, we describe and discuss these experimental tests.

### *5.2. Univariate modeling-based retrieval approaches*

In this part, we are interested in evaluating the methods related to the univariate model. The first one corresponds to the straightforward approach, presented in Subsection 2.3, where the GG distribution parameters of only the right and left images are compared. This method will be designated by GG-RL. The second one takes into account the disparity information in an implicit manner by computing features of the right view and the disparity compensated left view. This method will be denoted by GG-RDCL. The third one, designated by GG-RL-GG-D, is the second version of the proposed univariate model-based method where a new feature vector is defined by incorporating *simultaneously* the visual contents of the left and right images as well as the disparity information. We have also tested for comparison the recent state-of-the-art approach [16]. Recall that its basic idea consists in using the disparity to refine the results provided by a conventional CBIR system. More precisely, the MPEG-7 edge histogram descriptor is employed for the left images and the diffusion distance is used to measure the similarity between the histograms of the disparity maps during the refinement step. Hereafter, this method will be designated by State-of-the-art [16]. Moreover, since the developed approaches operate in the wavelet transform domain, we have also proposed to apply the state-of-the-art method in the same domain. It is important to note here that designing a CBIR system operating in the wavelet domain may constitute an interesting feature in practice in the sense that the decoding procedure at the retrieval step becomes unnecessary when the images are saved in a compressed format. To this end, the first step consists of comparing the feature vector of the query left image $(\alpha_j^{(\mathrm{l,q})}, \beta_j^{(\mathrm{l,q})})_{1\leq j\leq 3J}$ to the database left images $(\alpha_j^{(\mathrm{l,db})}, \beta_j^{(\mathrm{l,db})})_{1\leq j\leq 3J}$ using the KLD as a similarity measure. Then, the disparity features $(\alpha_j^{(\mathrm{u,db})}, \beta_j^{(\mathrm{u,db})})_{1\leq j\leq 3J}$ are used in the re-ranking step which is applied to the first 10% of the returned images and the retrieval results are re-ordered. In what follows, this modified version of the state-of-the-art method, applied in the WT domain, will be designated by Mod-state-of-the-art-WT.

Fig. 4 provides the precision-recall plots of the various approaches. It indicates that using implicitly the disparity map by comparing the right image and the disparity compensated left one of the stereo pair outperforms the straightforward approach. Moreover, thanks to the second version of the proposed method where the disparity information is explicitly added to the feature vector, we achieve further improvements. Our approach becomes more performant than the other methods, and achieves a gain of about 1-12% in precision compared to the state-of-the-art method [16].

We should note here that different tests have been carried out to study the impact of the weights assigned to the disparity map and to both images on the retrieval performance of the GG-RL-GG-D approach. Since the left and right images correspond to the same scene and present very similar contents, we assume that the weights associated to these views are identical (i.e. $a = b$). Fig. 5 shows the precision-recall curves for different weight parameters.

Thus, it can be observed that selecting a very low or a high value of $c$ leads to

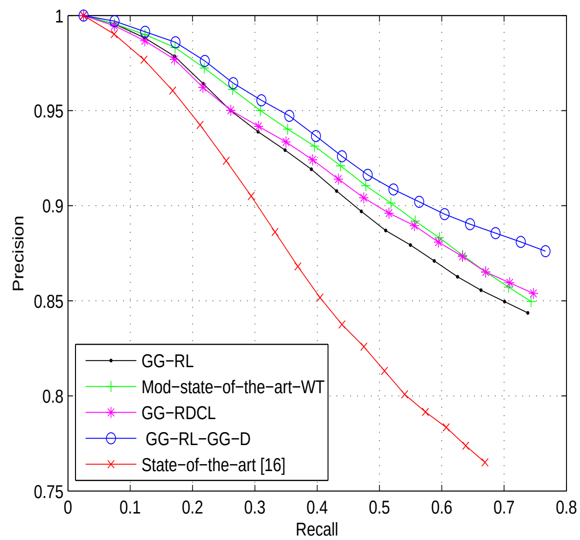



Figure 4: Retrieval performance in terms of precision and recall of the univariate approaches.

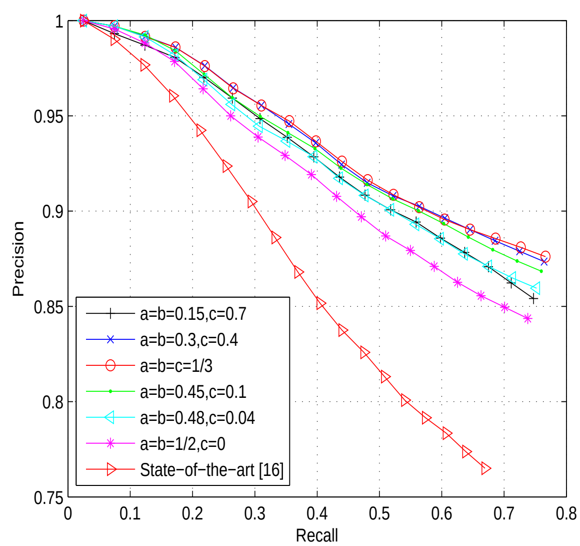



Figure 5: Impact of the weight values on the retrieval performances of the GG-RL-GG-D approach.

worse results. More generally, we conclude that weighting alike the features of the left, right and disparity images, by taking a value of $c$ around $\frac{1}{3}$, allows to achieve good retrieval performance.
Therefore, these results corroborate that disparity gives additional cues for SI retrieval when it is combined with the visual contents of the two views.

### *5.3. Bivariate modeling-based retrieval approaches*

The second series of experiments is dedicated to the study of the effect of using a bivariate statistical model to capture the dependencies between the wavelet coefficients of a stereo pair and, to the illustration of the benefits of incorporating simultaneously the disparity features and the visual ones computed through the bivariate model. To this end, we have also conducted the three following experiments related to the three retrieval strategies discussed in Subsection 4.3.
The first one, where the bivariate vector is constructed from the wavelet coefficients of the right and left images, is designated by BGG-RL. The second one, where the bivariate vector is defined by using the wavelet coefficients of the right image and the disparity compensated left one, is denoted by BGG-RDCL-1. The third approach, corresponding to the improved version of the previous one by taking into account the occlusion effect, is denoted by BGG-RDCL-2. By further adding the GG distribution parameters of the disparity map during the indexing step, these three methods will be respectively designated by BGG-RL-GG-D, BGG-RDCL-GG-D-1 and BGG-RDCL-GG-D-2. Fig. 6 depicts the precision versus recall curves for these approaches.
It can also be noticed that adding disparity improves the different bivariate modeling-based retrieval strategies. Moreover, these results show that tak-

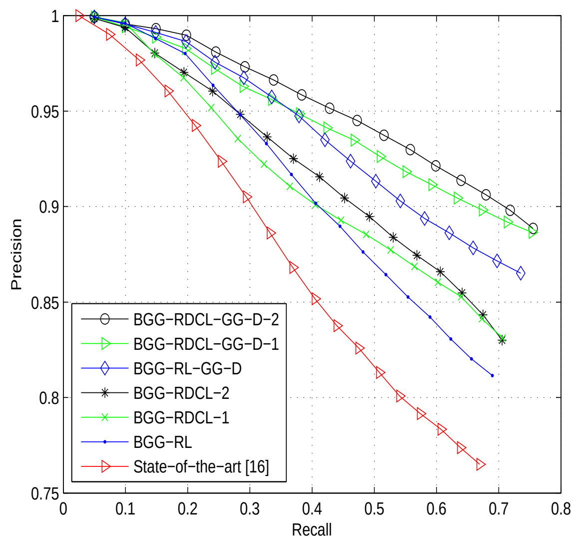



Figure 6: Retrieval performance in terms of precision and recall of the bivariate approaches.

ing into account the occlusion effect allows us to achieve the best retrieval performance.

Finally, we focus on comparing the performance of the proposed retrieval approaches based on bivariate and univariate models. It can be seen from Fig. 7 that the joint modeling of wavelet subbands BGG-RDCL-GG-D-1 achieves better retrieval performance compared to the univariate model-based approach. Further improvements are achieved when the bivariate model-based approach deals with the occlusion effect. Thus, compared to the state-of-the-art method [16], the resulting gain reaches 15% in terms of precision-recall. All these results confirm the effectiveness of incorporating the disparity information for stereo image retrieval purpose.

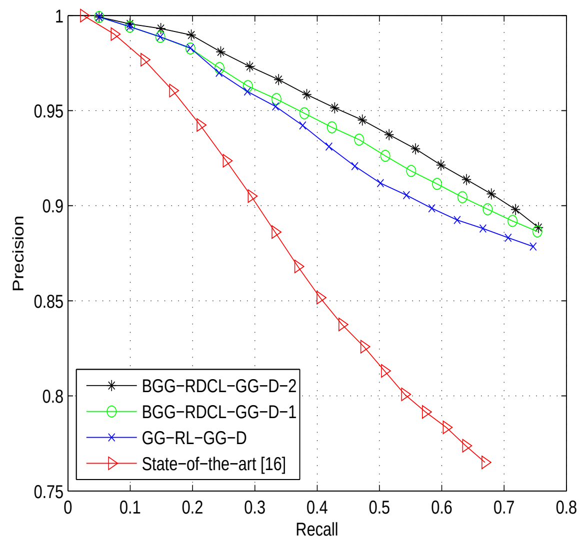



Figure 7: Retrieval performance in terms of precision and recall of the univariate and the bivariate approaches.

## 6. Conclusion

In this paper, we have addressed the problem of indexing and retrieval of stereo images in the wavelet-transform domain. Our first contribution consists in employing dense disparity maps either implicitly or explicitly during the feature extraction step. The parameters of the generalized Gaussian distribution that model the detail subbands of each view are combined with those of the disparity map to build a salient feature of the stereo pair content. Our second contribution aims at resorting to an appropriate bivariate model that accounts for the cross-view dependencies. Experimental results indicate the good performance of the bivariate approaches. In future work, it would be interesting to study the effect of quantizing the stereo images at different qualities as performed in [41].

## References


[1] K. Yoon, Y. Kim, J.-H. Park, J. Delgado, A. Yamada, F. Dufaux, R. Tous, JPSearch: New international standard providing interoperable framework for image search and sharing, Signal Processing: Image Communication 27 (7) (2012) 709–721.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, Journal of Visual Communication and Image Representation 12 (2000) 1349–1380.

[3] D. Taubman, M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[4] M. Kobayakawa, M. Hoshi, T. Ohmori, Robust texture image retrieval using hierarchical correlations of wavelet coefficients, in: IEEE International Conference on Pattern Recognition, 2000, pp. 3395–3400.

[5] P. Skulsujirapat, S. Aramvith, S. Siddhichai, Development of digital image retrieval technique using autocorrelogram and wavelet based texture, in: IEEE International Midwest Symposium on Circuits and Systems, 2004, pp. 273–276.

[6] Q. Wang, D. D. Feng, Texture analysis and retrieval using fractal signature and B-spline wavelet transform with second order derivative, in: IEEE International Conference on Image Processing, 2005, pp. 509–512.

[7] H. A. Moghaddam, T. T. Khajoie, A. H. Rouhi, M. S. Tarzjan, Wavelet correlogram: a new approach for image indexing and retrieval, Pattern Recognition 38 (12) (2005) 2506–2518.

[8] S. Agarwal, A. K. Verma, P. Singh, Content based image retrieval using discrete wavelet transform and edge histogram descriptor, in: IEEE International Conference in Information Systems and Computer Networks, 2013, pp. 19–23.

[9] M. N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized gaussian density and Kullback-Leibler distance, IEEE Transactions on Image Processing 11 (2) (2002) 146–158.

[10] S. Sakji-Nsibi, A. Benazza-Benyahia, Indexing of multichannelimages in the wavelet transform domain, in: IEEE International Conference onCommunication Technologies: Theory & Practice, Damascus, Syria, 2008, pp. 1–6.

[11] S. Sakji-Nsibi, A. Benazza-Benyahia, Copula-based statistical models for multicomponent image retrieval in the wavelet tranform domain, in: IEEE International Conference onImage Processing, Cairo, Egypt, 2009, pp. 253–256.

[12] J. R. Mathiassen, A. Skavhaug, K. Bø, Texture similarity measure using Kullback-Leibler divergence between gamma distributions, in: Computer Vision-ECCV 2002, Springer, 2002, pp. 133–147.

[13] S. K. Choy, C. S. Tong, Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval, IEEE Transactions on Image Processing 19 (2) (2010) 281–289.

[14] R. Kwitt, A. Uhl, Lightweight probabilistic texture retrieval, IEEE Transactions on Image Processing 19 (1) (2010) 241–253.

[15] S. Kiranyaz, M. Gabbouj, Content-based management of multimedia databases, Lambert Academic Publishing, 2012.

[16] Y. Feng, J. Ren, J. Jiang, Generic framework for content-based stereo image/video retrieval, Electronics letters 47 (2) (2011) 97–98.

[17] A. Chaker, M. Kaaniche, A. Benazza-Benyahia, Exploiting disparity information for stereo image retrieval, in: IEEE International Conference on Image Processing, Paris, France, 2014, p. 5 pages.

[18] J. Portilla, V. Strela, M. J. Wainwright, E. P. Simoncelli, Imagedenoising using scale mixtures of Gaussians in the wavelet domain, IEEE Trans. Image Process. 12 (2003) 1338–1351.

[19] A. Benazza-Benyahia, J.-C. Pesquet, Building robust wavelet estimators for multicomponent images using Stein's principle, IEEE Trans. Image Process. 14 (11) (2005) 1814–1830.

[20] P. Scheunders, S. D. Backer, Wavelet denoising of multicomponentimages using Gaussian scale mixture models and a noise-free image aspriors, IEEE Transactions on Image Processing 16 (7) (2007) 1865–1872.

[21] R. Kwitt, P. Meerwald, A. Uhl, Efficient texture image retrieval using copulas in a Bayesian framework, IEEE Transactions on Image Processing 20 (7) (2011) 2063–2077.

[22] G. Verdoolaege, P. Scheunders, Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination, International Journal of Computer Vision 95 (3) (2011) 265–286.

[23] N. Lasmar, Y. Berthoumieu, Multivariate statistical modeling for texture analysis using wavelet transforms, in: IEEE International Conference on Acoustics Speech and Signal Processing, Dallas, TX, USA, 2010, pp. 790–793.

[24] W. Sweldens, The lifting scheme: A custom-design construction of biorthogonal wavelets, Applied and computational harmonic analysis 3 (2) (1996) 186–200.

[25] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, J.-C. Pesquet, Vector lifting schemes for stereo image coding, IEEE Transactions on Image Processing 18 (11) (2009) 2463–2475.

[26] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using wavelet transform, IEEE Transactions on Image Processing 1 (2) (1992) 205–220.

[27] H. Yuan, X. P. Zhang, Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence, in:

IEEE International Conference on Multimedia and Expo, Taipe, Taiwan, 2004, pp. 1867–1870.

[28] C. Nafornita, Y. Berthoumieu, I. Nafornita, A. Isar, Kullback-Leibler distance between complex generalized Gaussian distributions, in: European Signal Processing Conference, Bucharest, Romania, 2012, pp. 1850–1854.

[29] U. R. Dhond, J. K. Aggarwal, Structure from stereo-a review, IEEE Transactions on Systems, Man and Cybernetics, 19 (6) (1989) 1489–1510.

[30] Y. Boykov, O. Veksler, R. Zabih., Fast approximate energy minimization via graph cuts., IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222–1239.

[31] S. Kosov, T. Thormählen, H.-P. Seidel, Accurate real-time disparity estimation with variational methods:, in: International Symposium on Visual Computing, Vol. 5875, Las Vegas, United States, 2009, pp. 796–807.

[32] W. L. Chan, H. Choi, R. G. Baraniuk, Multiscale image disparity estimation using the quaternion wavelet transform., in: IEEE International Conference on Image Processing, Atlanta, GA, USA, 2006, pp. 1229–1232.

[33] A. Bhatti, S. Nahvandi, Depth estimation using multi-wavelet analysis based stereo vision approach, International Journal of Wavelets, Multiresolution and Information Processing 6 (2008) 481–497.

[34] W. Miled, J.-C. Pesquet, M. Parent, A convex optimization approach for depth estimation under illumination variation, IEEE Transactions on Image Processing 18 (4) (2009) 813–830.

[35] J. Massey, J. Frank, The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association 46 (253) (1951) 68–78.

[36] A. Justel, D. Peña, R. Zamar, A multivariate Kolmogorov-Smirnov test of goodness of fit, Statistics & Probability Letters 35 (3) (1997) 251–259.

[37] W. Woo, A. Ortega, Stereo image compression based on disparity field segmentation, in: SPIE Conference on Visual Communications and Image Processing, Vol. 3024, 1997, pp. 391–402.

[38] E. Gómez, M. A. Gómez-Villegas, J. M. Marin, A survey on continuous elliptical vector distributions, Revista Matemática Complutense 36 (1) (2003) 345–361.

[39] G. Verdoolaege, Y. Rosseel, M. Lambrechts, P. Scheunders, Wavelet-based colour texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gaussian models, in: IEEE International Conference on Image Processing, 2009, pp. 265–268.

[40] M. Abramowitz, I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publishing Inc., New York, 1970.

[41] A. Chaker, M. Kaaniche, A. Benazza-Benyahia, An efficient retrieval strategy for wavelet-based quantized images, in: IEEE International

Conference on Acoustics Speech and Signal Processing, Vancouver, BC, 2013, pp. 1493–1497.