Depth-based color stereo images retrieval using joint multivariate statistical models

E. Ghodhbani, Mounir Kaaniche, A. Benazza-Benyahia

Depth-based color stereo images retrieval using joint multivariate statistical models

E. Ghodhbani[a,*], Mounir Kaaniche[b], Amel Benazza-Benyahia[a]

[a] *University of Carthage, SUP'COM, LR11TIC01, COSIM Lab., 2083, El Ghazala, Tunisia.*
[b] *Institut Galilée, L2TI, Université Paris 13, Sorbonne Paris Cité, France.*

Abstract

The growing interest in using the three dimensional information in various application fields has led to the generation of huge color stereo image databases. As a result, it becomes necessary to design efficient content-based image retrieval system well adapted to the indexing of such large databases. To this end, we propose in this paper different statistical-based retrieval approaches where the associated estimated model parameters are considered as a feature vector in the indexing process. More precisely, the Gaussian copula based multivariate Generalized Gaussian model will be used to capture the different correlations existing in color stereo images. While the first strategy aims at exploiting the cross-view as well as the cross-color channel redundancies, the second one resorts to a more general joint statistical model exploiting the correlation between the texture and depth information. Experimental results, performed on various datasets, confirm the benefits that can be drawn from the proposed approaches.

Keywords: Content-based retrieval, color stereo images, depth map, wavelet transform, multivariate modeling, Gaussian copula, Generalized Gaussian distribution.

---

*Corresponding author
Email address: `emna.ghodbani@supcom.tn` (E. Ghodhbani)

1. Introduction

One of the most useful 3D acquisition technologies is the stereoscopic vision system which consists in generating two images, called left and right views, by recording two slightly different angles of the same scene. The main advantage of these images is their ability to provide the 3D information (called also depth information) of the perceived scene. For this reason, such data has been extensively used in various application fields such as obstacle detection for autonomous vehicle navigation [1] and laparoscopic surgery planning in medicine [2]. Moreover, the color information plays a crucial role in the binocular vision. Indeed, in [3, 4], the authors proved that, compared to the use of luminance information only, chromatic features revealed an interesting gain in the stereoscopic correspondence process. Thus, such gain results in improving the disparity estimation step, and so, yielding a more accurate 3D reconstruction of the scene. The benefits of exploiting color information in enhancing the amount of binocular perceived depth data is also shown in [5].

The growing interest in color stereo imaging has led to the generation of huge stereo image (SI) databases. In this context, the first challenge is to facilitate the access to the database images given a query image presented by the user. To this end, Content-Based Image Retrieval (CBIR) systems have been extensively employed, and are mainly composed of two stages. The first one is the extraction of relevant visual features characterizing the color, texture and/or shape contained in both query and database images. At the second stage, the database images whose features are closest to those of the query one, according to a predefined similarity measure, are identified. It is worth pointing out that very few works have recently been reported in the literature for color SI retrieval. The first one, developed by Feng *et al.* [6], consists in extracting the MPEG-7 edge histograms from the left image. Then, a refinement of the resulting image candidates is performed using a re-ranking procedure based on the disparity cues. The major drawback of this approach is that it favors only one view while ignoring the other one. In [7], a retrieval method devoted to high-resolution optical satellite SI is proposed. It aims at comparing features extracted from digital surface models and ortho-images. In [8], an object-based stereo image retrieval method is designed for color SI. Note that unlike the aforementioned works, the objective is to find in the database images containing objects similar to the query objects. Thus, to handle such partial queries, a prior segmentation of the views is performed. Then, salient features (typically, the MPEG-7 color layout descriptor, the pyramid histogram of visual words and local binary pattern) are computed.

However, it should be noted that most of the reported works were interested in retrieving either color images (i.e. mono-view) or gray level stereo images. More precisely, it has been shown that, among these works, statistical-based approaches lead to high retrieval performance. The main idea behind them consists in resorting to a parametric modeling of the distributions of the coefficients resulting from an image decomposition such as the Wavelet Transform (WT), and then taking the fitted parameters as features. Such approach presents two main advantages: the accuracy of the parametric model and, the availability of metrics such as the Kullback-Leibler divergence allowing a fast computation of the similarity measure between the distribution parameters. For instance, regarding the mono-view color images, multivariate statistical models with different margins, such as the Gamma, Gaussian, Laplacian and Generalized Gaussian (GG) distributions, have been used to capture the correlations existing between three color channels [9, 10]. More sophisticated models based on copula have also been used in [11, 12]. In the gray level SI context, Chaker *et al.* resorted to a Bivariate GG distribution model of the wavelet subbands of both views combined with an univariate GG model of the wavelet coefficients of the disparity map [13]. Recently, Karine *et al.* have captured the cross-view dependencies through a Gaussian copula-based multivariate model [14] and, have shown that such modeling approach outperforms the BGG-based one [13].

In this paper, we propose to extend such statistical-based approaches to the context of color SI retrieval. Indeed, while the existing works devoted to color mono-view images (resp. gray level SI) could be easily generalized to the context of color SI by applying them independently to each view of the stereo pairs (resp. to each channel of the color space), our main contributions aim at designing new retrieval strategies exploiting *simultaneously* the cross-view redundancies as well as the color channel dependencies. This is first achieved through copula-based multivariate approaches for modeling the wavelet coefficients of the color SI. Moreover, in addition to the stereo pairs, the depth information is exploited in two ways. In the first one, it is modeled using an univariate model and its resulting distribution parameters are combined with those modeling the color SI wavelet coefficients. However, in the second one, we propose to resort to a more general joint statistical model to capture the dependencies existing between the texture (i.e two color views) and depth information.

The remainder of this paper is organized as follows. In Section 2, an overview of the univariate and bivariate modeling based SI retrieval approaches is given. Then, the proposed multivariate-based SI retrieval approaches are described in Section 3. Finally, experimental results, carried out on different natural color stereo image databases, are shown and discussed in Section 4, and some conclusions are drawn in Section 5.

2. Related works

2.1. Wavelet-based stereo image retrieval methods

As mentioned in Section 1, wavelet-based image retrieval methods have attracted a considerable attention over the last years. In this context, a WT is often applied to the database images, resulting in one approximation subband and J detail subbands for each image [15]. The generated detail subbands represent the image edges at different orientations and scales, and their statistical properties are very often exploited in the indexing process. For instance, the detail coefficients w_j, for each subband j with $j \in \{1, \dots, J\}$, are often viewed as realizations of a zero-mean continuous random variable whose probability density function f is approximated by a specific distribution. While Laplacian and Gamma distributions have already been used in the literature [16, 17], the GG distribution [18, 19] has been extensively employed for modeling the wavelet coefficients. Let us recall that the probability density function of the GG distribution is given by:

$$\forall w_j \in \mathbb{R}, \quad f_{\mathrm{GG}}(w_j; \mathbf{p}_j) = \frac{b_j}{2a_j \Gamma(1/b_j)} \exp(-(\frac{\mid w_j \mid}{a_j})^{b_j}), \quad \text{with} \quad \mathbf{p}_j = (a_j, b_j) \tag{1}$$

where $\Gamma(z) \triangleq \int_0^{+\infty} t^{z-1} e^{-t} dt$ and a_j and b_j are two positive reals, called the scale and shape parameters, that can be estimated using the maximum likelihood method [18, 20].

Then, the estimated parameters $\mathbf{p}_j = (a_j, b_j)$ of the different detail subbands are merged to construct the final feature vector of each image. Finally, during the retrieval procedure, the similarity between a database image $I^{(\mathrm{db})}$ and a query one $I^{(\mathrm{q})}$ is measured through the computation of an appropriate metric. While several metrics have been reported in the literature [21], the Kullback-Leibler Divergence (KLD) is an appealing tool to assess the similarity between two probability density functions [22]. It is widely used in content-based image retrieval systems since it has a closed form for a great number of model distributions. By considering two GG probability density functions of a given query and database image subbands, characterized respectively by $\mathbf{p}_j^{(\mathrm{q})}$ and $\mathbf{p}_j^{(\mathrm{db})}$, the KLD is expressed as:

$$\begin{aligned} \widetilde{\mathcal{D}}_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{q})} \parallel \mathbf{p}_j^{(\mathrm{db})}) &= KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{q})} \parallel \mathbf{p}_j^{(\mathrm{db})}) \\ &= \log\Big(\frac{b_j^{(\mathrm{db})} a_j^{(\mathrm{q})} \Gamma(1/b_j^{(\mathrm{q})})}{b_j^{(\mathrm{q})} a_j^{(\mathrm{db})} \Gamma(1/b_j^{(\mathrm{db})})}\Big) - \frac{1}{b_j^{(\mathrm{db})}} \\ &\quad + (\frac{a_j^{(\mathrm{db})}}{a_j^{(\mathrm{q})}})^{b_j^{(\mathrm{q})}} \frac{\Gamma((b_j^{(\mathrm{q})}+1)/b_j^{(\mathrm{db})})}{\Gamma(1/b_j^{(\mathrm{db})})}, \end{aligned} \tag{2}$$

where $\mathbf{p}_j^{(\mathrm{q})} = (a_j^{(\mathrm{q})}, b_j^{(\mathrm{q})})$ and $\mathbf{p}_j^{(\mathrm{db})} = (a_j^{(\mathrm{db})}, b_j^{(\mathrm{db})})$ represent respectively the estimated model parameters of the j-th subband of the query image $I^{(\mathrm{q})}$ and the database one $I^{(\mathrm{db})}$.

Thus, by adding the resulting KLD terms across the different detail subbands, the global similarity measure $\bar{\mathcal{D}}_{\mathrm{GG}}$ between the query and database images is obtained as follows:

$$\bar{\mathcal{D}}_{\mathrm{GG}}(I^{(\mathrm{q})}, I^{(\mathrm{db})}) = \sum_{j=1}^{J} \widetilde{\mathcal{D}}_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{q})} \parallel \mathbf{p}_j^{(\mathrm{db})}). \tag{3}$$

When dealing with gray level SI retrieval, a straightforward solution may consist in applying the wavelet-based image retrieval approach described above to each view of the stereo pairs. More precisely, at the retrieval stage, the query SI $(I^{(\mathrm{l,q})}, I^{(\mathrm{r,q})})$ is compared to any database SI $(I^{(\mathrm{l,db})}, I^{(\mathrm{r,db})})$ by evaluating the sum of the global KLDs (i.e $\bar{\mathcal{D}}_{\mathrm{GG}}$ given by Eq. (3)) of the left and right images:

$$\check{\mathcal{D}}(I^{(\mathrm{l,q})}, I^{(\mathrm{r,q})}, I^{(\mathrm{l,db})}, I^{(\mathrm{r,db})}) = \bar{\mathcal{D}}_{\mathrm{GG}}(I^{(\mathrm{r,q})}, I^{(\mathrm{r,db})}) + \bar{\mathcal{D}}_{\mathrm{GG}}(I^{(\mathrm{l,q})}, I^{(\mathrm{l,db})}). \tag{4}$$

Despite its simplicity, this approach has the shortcoming of ignoring the cross-view dependencies existing between the left and right images. This fact has motivated the design of more efficient solutions for gray level SI retrieval [13, 14]. For instance, in addition to the statistical features extracted from the left and right images, the first strategy proposed in [13] consists in extracting similar statistical features from the disparity (or depth) map u. Thus, the closeness between the query and database SI scenes is assessed through the sum of the global KLDs related to the two views as well as the disparity maps $u^{(\mathrm{q})}$ and $u^{(\mathrm{db})}$:

$$\mathcal{D}_{\mathrm{GG}} = \bar{\mathcal{D}}_{\mathrm{GG}}(I^{(\mathrm{r,q})}, I^{(\mathrm{r,db})}) + \bar{\mathcal{D}}_{\mathrm{GG}}(I^{(\mathrm{l,q})}, I^{(\mathrm{l,db})}) + \bar{\mathcal{D}}_{\mathrm{GG}}(u^{(\mathrm{q})}, u^{(\mathrm{db})}). \tag{5}$$

While this strategy exploits explicitly the cross-view dependencies through the use of the depth information, the second strategy developed in [13] as well as the method proposed in [14] aim to exploit implicitly these inter-view redundancies by resorting to bivariate statistical modeling approaches.

2.2. Bivariate statistical modeling-based approaches

2.2.1. Bivariate Generalized Gaussian model

The second strategy described in [13] involves a bivariate parametric model of the *joint* distribution of the two views to reflect the cross-view redundancies. To this end, a Bivariate Generalized Gaussian (BGG) distribution has been employed. Indeed, let us denote by $\mathbf{w}_j = (w_j^{(\mathrm{l})}, w_j^{(\mathrm{r})})^\top$ the bivariate vector composed of the wavelet coefficients of the left subband $w_j^{(\mathrm{l})}$ and those of the right one $w_j^{(\mathrm{r})}$. The latter can be viewed as the realization of a zero-mean random vector whose probability density function is given by:

$$\forall \mathbf{w}_j \in \mathbb{R}^2, f_{\mathrm{BGG}}(\mathbf{w}) = \frac{2}{\pi\Gamma(1+\frac{1}{b_j})2^{1+\frac{1}{b_j}}} \mid \mathbf{S}_j \mid^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{w}^T\mathbf{S}_j^{-1}\mathbf{w})^{b_j}\right), \tag{6}$$

where $b_j > 0$ is the shape parameter and $\mathbf{S}_j$ represents the scaling matrix of size 2×2, which can be estimated using the maximum likelihood criterion [23]. With this BGG model, it is important to note that a closed form expression of the KLD exists and can be found in [13], which will allow to measure easily the similarity between the query and database SI.

2.2.2. Copula-based bivariate modeling

The recent approach developed in [14] is based on the copula tool which presents the advantage of exploiting the dependencies between many random variables independently of their marginal distributions. Note that the basic concepts behind copula theory can be found in [24]. Such modeling approach is completely defined once the appropriate marginal distributions and the copula are fixed. Among the several copula families proposed in the literature, the authors in [14] retain the Gaussian copula, which has been extensively employed due to the following reasons. First, it is a good fit to the statistics of the wavelet coefficient subbands allowing an accurate capture of both marginal and joint distributions [11, 12]. Moreover, its related hyperparameters can be easily estimated using maximum likelihood technique and its associated KLD has a closed form. In [14], Karine *et al.* have considered the Gaussian copula with GG as well as Gamma marginal distributions, and the resulting marginal parameters are combined with the copula ones to construct the feature vector for the indexing step.

Based on the aforementioned recent works, we first propose in the following to extend them, and then, design novel Gaussian copula-based multivariate modeling approaches adapted to the context of color stereo images.

3. Proposed depth-based retrieval approaches through bivariate and multivariate models

3.1. Motivation

Let us first assume that RGB cameras have been used to acquire the color stereo images. Thus, the resulting images will be denoted by $I^{(\mathrm{c},\mathrm{v})}$ with $\mathrm{v} \in \{\mathrm{l},\mathrm{r}\}$ represents either the left or the right view, and $\mathrm{c} \in \{\mathrm{R},\mathrm{G},\mathrm{B}\}$ represents the red, green and blue color components. When dealing with color stereo images retrieval, a straightforward solution would consist in applying simple univariate statistical approaches to each color component of each view, as well as to the depth map u. More precisely, by considering again the GG distribution for modeling the wavelet coefficient subbands of both color views and the depth maps, the scale and shape parameters are used to construct the following feature vector:

$$
\begin{aligned}
&\forall j \in \{1,\ldots,J\}, \qquad \mathbf{V}_{\mathrm{UGG},j} = \left(\mathbf{p}_j^{(\mathrm{c},\mathrm{v})}, \mathbf{p}_j^{(\mathrm{u})}\right),\\
&\text{with } \mathrm{v} \in \{\mathrm{l},\mathrm{r}\}, \quad \text{and} \quad \mathrm{c} \in \{\mathrm{R},\mathrm{G},\mathrm{B}\}.
\end{aligned} \tag{7}
$$

Once the feature vectors are extracted from the query and database color SI as well as their associated depth maps, which will be denoted respectively by $\left(\mathbf{p}_j^{(\mathrm{c},\mathrm{v},\mathrm{q})}, \mathbf{p}_j^{(\mathrm{u},\mathrm{q})}\right)$, $\left(\mathbf{p}_j^{(\mathrm{c},\mathrm{v},\mathrm{db})}, \mathbf{p}_j^{(\mathrm{u},\mathrm{db})}\right)$, the sum of the KLDs over all the subbands of the color channels of the left and right views and those of the depth maps is computed as a similarity measure:

$$
\begin{aligned}
\mathcal{D}_{\mathrm{UGG}} = &\sum_{j=1}^{J} \sum_{\mathrm{v}\in\{\mathrm{l},\mathrm{r}\}} \sum_{\mathrm{c}\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{c},\mathrm{v},\mathrm{q})} || \mathbf{p}_j^{(\mathrm{c},\mathrm{v},\mathrm{db})})\\
&+ KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{u},\mathrm{q})} || \mathbf{p}_j^{(\mathrm{u},\mathrm{db})}).
\end{aligned} \tag{8}
$$

However, such approach is not so efficient since the statistical modeling step is performed in an independent way while the stereo images as well as their color components present strong correlations. For this reason, we propose to capture these dependencies by resorting to bivariate modeling approaches as it will be addressed in what follows.

3.2. Gaussian copula-based bivariate modeling approach

Inspired by the previous work of Karine *et al.* [14] for gray level SI retrieval, we propose first to extend it by resorting to Gaussian Copula based Bivariate Generalized Gaussian model (GC-BGG) to capture the cross-view dependencies for each color channel. Thus, as it can be seen in Fig. 1, each color channel of the SI is considered independently from the other ones by defining the following vector:

$$\forall \ \mathrm{c} \in \{\mathrm{R,G,B}\}, \qquad \mathbf{w}_j^{(\mathrm{c})} = \Big(w_j^{(\mathrm{c,l})}, w_j^{(\mathrm{c,r})}\Big)^\top. \tag{9}$$

By assuming that each vector $\mathbf{w}_j^{(\mathrm{c})}$ is the realization of a stochastic vector, the GC-BGG probability density function is given by:

$$\begin{aligned}
&\forall \ \mathrm{c} \in \{\mathrm{R,G,B}\}, \quad \forall \ \mathbf{w}_j^{(\mathrm{c})} = \Big(w_j^{(\mathrm{c,l})}, w_j^{(\mathrm{c,r})}\Big)^\top \in \mathbb{R}^2, \\
&\qquad f_{\text{GC-BGG}}(\mathbf{w}_j^{(\mathrm{c})}) = |\boldsymbol{\Sigma}_j^{(\mathrm{c})}|^{-1/2} \exp\Big(- \frac{(\widehat{\mathbf{w}}_j^{(\mathrm{c})})^\top((\boldsymbol{\Sigma}_j^{(\mathrm{c})})^{-1} - \mathbf{I})\widehat{\mathbf{w}}_j^{(\mathrm{c})}}{2}\Big) \\
&\qquad\qquad \times f_{\mathrm{GG}}(w_j^{(\mathrm{c,l})}; \mathbf{p}_j^{(\mathrm{c,l})}) f_{\mathrm{GG}}(w_j^{(\mathrm{c,r})}; \mathbf{p}_j^{(\mathrm{c,r})}),
\end{aligned} \tag{10}$$

where $\widehat{\mathbf{w}}_j^{(\mathrm{c})} = \Big(\phi^{-1}(w_j^{(\mathrm{c,l})}), \phi^{-1}(w_j^{(\mathrm{c,r})})\Big)^\top$ with ϕ^{-1} is the inverse cumulative distribution function of the normal distribution $\mathcal{N}(0,1)$, $\boldsymbol{\Sigma}_j^{(\mathrm{c})}$ is the covariance matrix of the vector $\widehat{\mathbf{w}}_j^{(\mathrm{c})}$ (i.e with size 2×2), and $\mathbf{p}_j^{(\mathrm{c,v})}$ is the parameters vector of the GG margins used to model the wavelet coefficients of the color components of both views. The latter parameters (i.e $\boldsymbol{\Sigma}_j^{(\mathrm{c})}$ and $\mathbf{p}_j^{(\mathrm{c,v})}$) can be estimated using the method described in [24].

For the indexing step, the estimated parameters resulting from modeling each stereo pair channel of the database will represent the following texture feature vector:

$$\begin{aligned}
&\forall \mathrm{c} \in \{\mathrm{R,G,B}\}, \quad \forall j \in \{1, \dots, J\}, \\
&\mathbf{V}^{(\mathrm{c})}_{\text{GC-BGG},j} = \Big(\mathbf{p}_j^{(\mathrm{c,v})}, \boldsymbol{\Sigma}_j^{(\mathrm{c})}\Big)_{\mathrm{v}\in\{\mathrm{l,r}\}} = \Big(\mathbf{p}_j^{(\mathrm{c,l})}, \mathbf{p}_j^{(\mathrm{c,r})}, \boldsymbol{\Sigma}_j^{(\mathrm{c})}\Big).
\end{aligned} \tag{11}$$

Then, the feature vectors associated to the query and database SI, denoted by $(\mathbf{p}_j^{(\mathrm{c,v,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,q})})$ and $(\mathbf{p}_j^{(\mathrm{c,v,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,db})})$, are compared using the KLD. It is important to note that the latter has a closed form for such GC-BGG, and is given by [24]:

$$\widetilde{\mathcal{D}}^{(\mathrm{c})}_{\mathrm{GC\text{-}BGG,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,db})}) = \sum_{\mathrm{v}\in\{\mathrm{l,r}\}} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{c,v,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})})$$
$$+ \frac{1}{2}\left(\mathrm{tr}((|\boldsymbol{\Sigma}_j^{(\mathrm{c,q})}|)^{-1}|\boldsymbol{\Sigma}_j^{(\mathrm{c,db})}|) + \log\frac{|\boldsymbol{\Sigma}_j^{(\mathrm{c,q})}|}{|\boldsymbol{\Sigma}_j^{(\mathrm{c,db})}|} - 2\right). \quad (12)$$

Therefore, the global distance $\mathcal{D}_{\mathrm{GC\text{-}BGG}}$ between a query color SI and a database one is computed through summing the above similarity measure $\widetilde{\mathcal{D}}_{\mathrm{GC\text{-}BGG}}$ over the three spectral channels and across the different wavelet subbands.

$$\mathcal{D}_{\mathrm{GC\text{-}BGG}} = \sum_{j=1}^{J} \sum_{\mathrm{c}\in\{\mathrm{R,G,B}\}} \widetilde{\mathcal{D}}^{(\mathrm{c})}_{\mathrm{GC\text{-}BGG,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{c,db})}). \quad (13)$$

3.3. Gaussian copula-based multivariate modeling approaches

3.3.1. Accounting for cross-channel dependencies

In the previous approach, the cross-view redundancies are exploited by processing the three color channels separately. A dual approach would consist in exploiting the cross-channel dependencies while processing the left and right views independently. This approach is illustrated in Fig. 2. More precisely, for each view of the stereo pair, a 3-dimensional vector is defined as follows:

$$\forall\ \mathrm{v} \in \{\mathrm{l,r}\}, \qquad \mathbf{w}_j^{(\mathrm{v})} = \left(w_j^{(\mathrm{c,v})}\right)^{\top}_{\mathrm{c}\in\{\mathrm{R,G,B}\}} = \left(w_j^{(\mathrm{R,v})}, w_j^{(\mathrm{G,v})}, w_j^{(\mathrm{B,v})}\right)^{\top}. \quad (14)$$

Then, for each view, the wavelet coefficients of the three color channels are modeled by a 3-dimensional Gaussian Copula-based Multivariate Generalized Gaussian distribution, denoted in the following by GC-MGG-3. Thus, by assuming that $\mathbf{w}_j^{(\mathrm{v})}$ is a realization of a random vector, the GC-MGG-3 probability density function is given by:

$$
\begin{aligned}
&\forall\ \mathrm{v} \in \{\mathrm{l,r}\}, \quad \forall\ \mathbf{w}_j^{(\mathrm{v})} = \Big(w_j^{(\mathrm{R,v})}, w_j^{(\mathrm{G,v})}, w_j^{(\mathrm{B,v})}\Big)^\top \in \mathbb{R}^3,\\
&\qquad f_{\text{GC-MGG-3}}(\mathbf{w}_j^{(\mathrm{v})}) = |\mathbf{\Sigma}_j^{(\mathrm{v})}|^{-1/2} \exp\Big(- \frac{(\widehat{\mathbf{w}}_j^{(\mathrm{v})})^\top((\mathbf{\Sigma}_j^{(\mathrm{v})})^{-1} - \mathbf{I})\widehat{\mathbf{w}}_j^{(\mathrm{v})}}{2}\Big)\\
&\qquad\qquad \times f_{\mathrm{GG}}(w_j^{(\mathrm{R,v})}; \mathbf{p}_j^{(\mathrm{R,v})}) f_{\mathrm{GG}}(w_j^{(\mathrm{G,v})}; \mathbf{p}_j^{(\mathrm{G,v})}) f_{\mathrm{GG}}(w_j^{(\mathrm{B,v})}; \mathbf{p}_j^{(\mathrm{B,v})}),
\end{aligned} \tag{15}
$$

where $\widehat{\mathbf{w}}_j^{(\mathrm{v})} = \Big(\phi^{-1}(w_j^{(\mathrm{R,v})}), \phi^{-1}(w_j^{(\mathrm{G,v})}), \phi^{-1}(w_j^{(\mathrm{B,v})})\Big)^\top$, $\mathbf{\Sigma}_j^{(\mathrm{v})}$ is the covariance matrix of the vector $\widehat{\mathbf{w}}_j^{(\mathrm{v})}$ (i.e with size 3×3), and $\mathbf{p}_j^{(\mathrm{c,v})}$ is the parameters vector defined previously. The latter parameters (i.e $\mathbf{\Sigma}_j^{(\mathrm{v})}$ and $\mathbf{p}_j^{(\mathrm{c,v})}$) can be again estimated used the method described in [24].

The estimated hyperparameters resulting from modeling the cross-channel dependencies for each view constitute the following texture feature vector:

$$
\begin{aligned}
&\forall\ \mathrm{v} \in \{\mathrm{l,r}\}, \quad \forall\ j \in \{1, \ldots, J\},\\
&\mathbf{V}_{\text{GC-MGG-3},j}^{(\mathrm{v})} = \Big(\mathbf{p}_j^{(\mathrm{c,v})}, \mathbf{\Sigma}_j^{(\mathrm{v})}\Big)_{\mathrm{c} \in \{\mathrm{R,G,B}\}}.
\end{aligned} \tag{16}
$$

Then, for each subband view of the query and database SI, the associated feature vectors $(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{v,q})})$ and $(\mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{v,db})})$ are compared using again the KLD [24]:

$$
\begin{aligned}
&\widetilde{\mathcal{D}}_{\text{GC-MGG-3,j}}^{(\mathrm{v})}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{v,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{v,db})}) =\\
&\qquad \sum_{\mathrm{c} \in \{\mathrm{R,G,B}\}} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{c,v,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})})\\
&\qquad + \frac{1}{2}\left(\mathrm{tr}((|\mathbf{\Sigma}_j^{(\mathrm{v,q})}|)^{-1}|\mathbf{\Sigma}_j^{(\mathrm{v,db})}|) + \log\frac{|\mathbf{\Sigma}_j^{(\mathrm{v,q})}|}{|\mathbf{\Sigma}_j^{(\mathrm{v,db})}|} - 3\right).
\end{aligned} \tag{17}
$$

By computing the above similarity measure across all the detail subbands, and applying it independently on each view, the resulting global distance between a query and database SI yields:

$$
\mathcal{D}_{\text{GC-MGG-3}} = \sum_{j=1}^{J} \sum_{\mathrm{v} \in \{\mathrm{l,r}\}} \widetilde{\mathcal{D}}_{\text{GC-MGG-3,j}}^{(\mathrm{v})}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{v,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{v,db})}). \tag{18}
$$

3.3.2. Accounting for cross-view and channel dependencies

After separately exploiting the cross-view dependencies and the cross-channel ones, we propose now to combine these two kinds of dependency in one multivariate statistical model. Thus, in this new modeling approach, shown in Fig. 3, we will consider a 6-dimensional vector $\mathbf{w}_j^{(\mathrm{r,l})}$ which gathers the wavelet coefficient subbands of the three channels of the left and right views:

$$\begin{aligned}\mathbf{w}_j^{(\mathrm{r,l})} &= \Big(w_j^{(\mathrm{c,v})}\Big)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}} \\ &= \Big(w_j^{(\mathrm{R,l})}, w_j^{(\mathrm{G,l})}, w_j^{(\mathrm{B,l})}, w_j^{(\mathrm{R,r})}, w_j^{(\mathrm{G,r})}, w_j^{(\mathrm{B,r})}\Big)^{\top}. \end{aligned} \tag{19}$$

These wavelet subbands are modeled using a 6-dimensional Gaussian copula-based MGG, which will be designated in what follows by GC-MGG-6. Its related probability density function is expressed as:

$$\begin{aligned}&\forall\ \mathbf{w}_j^{(\mathrm{r,l})} = \Big(w_j^{(\mathrm{c,v})}\Big)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}} \in \mathbb{R}^6, \\ &\qquad f_{\text{GC-MGG-6}}(\mathbf{w}_j^{(\mathrm{r,l})}) = |\mathbf{\Sigma}_j^{(\mathrm{r,l})}|^{-1/2} \exp\Big(- \frac{(\widehat{\mathbf{w}}_j^{(\mathrm{r,l})})^{\top}((\mathbf{\Sigma}_j^{(\mathrm{r,l})})^{-1} - \mathbf{I})\widehat{\mathbf{w}}_j^{(\mathrm{r,l})}}{2}\Big) \\ &\qquad\qquad \times \prod_{\mathrm{v}\in\{\mathrm{l,r}\}} \prod_{c\in\{\mathrm{R,G,B}\}} f_{\mathrm{GG}}(w_j^{(\mathrm{c,v})}; \mathbf{p}_j^{(\mathrm{c,v})}),\end{aligned} \tag{20}$$

where $\widehat{\mathbf{w}}_j^{(\mathrm{r,l})} = \Big(\phi^{-1}(w_j^{(\mathrm{c,v})})\Big)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}}$, $\mathbf{\Sigma}_j^{(\mathrm{r,l})}$ is the covariance matrix of the vector $\widehat{\mathbf{w}}_j^{(\mathrm{r,l})}$ (i.e with size 6×6), and $\mathbf{p}_j^{(\mathrm{c,v})}$ is the parameters vector of the GG margins. The latter parameters (i.e $\mathbf{\Sigma}_j^{(\mathrm{r,l})}$ and $\mathbf{p}_j^{(\mathrm{c,v})}$) can be again estimated using the method described in [24].

During the indexing step, the resulting estimated hyperparameters set is used as texture feature vector:

$$\forall j \in \{1,\ldots,J\}, \qquad \mathbf{V}^{(\mathrm{r,l})}_{\text{GC-MGG-6},j} = \Big(\mathbf{p}_j^{(\mathrm{c,v})}, \mathbf{\Sigma}_j^{(\mathrm{r,l})}\Big)_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}}. \tag{21}$$

Then, the similarity between the feature vectors of the query and database SI wavelet subbands, $(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,q})})$ and $(\mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,db})})$, is computed as follows:

$$\begin{aligned}
&\widetilde{\mathcal{D}}^{(\mathrm{r,l})}_{\text{GC-MGG-6,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,q})}||\mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,db})}) = \\
&\quad \sum_{\mathrm{v}\in\{\mathrm{l,r}\}} \sum_{\mathrm{c}\in\{\mathrm{R,G,B}\}} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{c,v,q})}||\mathbf{p}_j^{(\mathrm{c,v,db})}) \\
&\quad + \frac{1}{2}\left(\mathrm{tr}((|\mathbf{\Sigma}_j^{(\mathrm{r,l,q})}|)^{-1}|\mathbf{\Sigma}_j^{(\mathrm{r,l,db})}|) + \log\frac{|\mathbf{\Sigma}_j^{(\mathrm{r,l,q})}|}{|\mathbf{\Sigma}_j^{(\mathrm{r,l,db})}|} - 6\right).
\end{aligned} \tag{22}$$

Finally, by adding the above measure over all the wavelet subbands, the global distance used to compare a given query and database SI is obtained:

$$\mathcal{D}_{\text{GC-MGG-6}} = \sum_{j=1}^{J} \widetilde{\mathcal{D}}^{(\mathrm{r,l})}_{\text{GC-MGG-6,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,q})}||\mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{\Sigma}_j^{(\mathrm{r,l,db})}). \tag{23}$$

3.4. Improved depth modeling based retrieval approaches

In the previous described approaches, only texture features have been extracted from the left and right color images to compare the query and database SI. However, in order to improve the indexing process, it would be interesting to exploit another relevant feature of stereo data which is the 3D information. This can be achieved by using the depth maps of the SI database. To this end, and after performing a wavelet transform on these maps, we propose to resort to the following two strategies.

3.4.1. Independent univariate depth modeling approach

In the first one, an independent univariate modeling of the depth maps will be performed. Indeed, as shown in [13, 14], the wavelet coefficients of the depth maps can be modeled using the GG distribution. Thus, in addition to the texture feature vector, the depth one is constructed by taking the shape and distribution parameters $\mathbf{p}_j^{(\mathrm{u})} = (a_j^{(\mathrm{u})}, b_j^{(\mathrm{u})})$. Then, for each stereo pair, the KLD between their associated query $u^{(\mathrm{q})}$ and database $u^{(\mathrm{db})}$ depth maps, whose features are denoted by $\mathbf{p}_j^{(\mathrm{u,q})}$ and $\mathbf{p}_j^{(\mathrm{u,db})}$, can be computed as follows:

$$\mathcal{D}(u^{(\mathrm{q})}||u^{(\mathrm{db})}) = \sum_{j=1}^{J} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{u,q})}||\mathbf{p}_j^{(\mathrm{u,db})}). \tag{24}$$

Finally, in order to compare a given query SI and the database one, the above depth similarity measure (i.e. Eq. (24)) will be added to the previous image similarity measures resulting from the Gaussian copula based bivariate and multivariate modeling of both views (i.e. the measures given by Eqs. (13), (18) and (23)).

3.4.2. Joint texture and depth information modeling

The depth wavelet coefficients may share some similarities with those of the color components in both views. For instance, recent studies have already shown the existence of such correlation and exploited it in the context of stereo image quality assessment [25, 26]. In order to confirm such dependencies between the color stereo images and their corresponding depth maps, we resort to a Chi-plot [27], which is similar to the scatterplot but with more explicit information. Fig. 4 displays an example of Chi-plot for each color channel and the depth map of a given database SI. The deviation of the points from the tolerance band indicates the dependence behavior between this data. For this reason, we have proposed to investigate a more general joint statistical modeling framework to simultaneously exploit the cross-view/channel and depth map redundancies. To this end, the detail wavelet subbands of all these inputs are considered to define the following 7-dimensional vector:

$$\begin{aligned}
\mathbf{w}_j^{(\mathrm{r,l,u})} &= \left(w_j^{(\mathrm{c,v})}, w_j^{(\mathrm{u})}\right)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}} \\
&= \left(w_j^{(\mathrm{R,l})}, w_j^{(\mathrm{G,l})}, w_j^{(\mathrm{B,l})}, w_j^{(\mathrm{R,r})}, w_j^{(\mathrm{G,r})}, w_j^{(\mathrm{B,r})}, w_j^{(\mathrm{u})}\right)^{\top}. \qquad (25)
\end{aligned}$$

Then, the wavelet coefficients of this vector are modeled using a 7-dimensional Gaussian copula based MGG, which will be denoted by GC-MGG-7, whose probability density function is given by:

$$
\begin{aligned}
&\forall\ \mathbf{w}_j^{(\mathrm{r,l,u})} = \left(w_j^{(\mathrm{c,v,u})}\right)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}} \in \mathbb{R}^7,\\
&f_{\text{GC-MGG-7}}(\mathbf{w}_j^{(\mathrm{r,l,u})}) = |\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u})}|^{-1/2} \exp\left(-\frac{(\widehat{\mathbf{w}}_j^{(\mathrm{r,l,u})})^{\top}((\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u})})^{-1}-\mathbf{I})\widehat{\mathbf{w}}_j^{(\mathrm{r,l,u})}}{2}\right)\\
&\qquad\qquad \times f_{\mathrm{GG}}(w_j^{(\mathrm{u})};\mathbf{p}_j^{(\mathrm{u})}) \prod_{\mathrm{v}\in\{\mathrm{l,r}\}} \prod_{\mathrm{c}\in\{\mathrm{R,G,B}\}} f_{\mathrm{GG}}(w_j^{(\mathrm{c,v})};\mathbf{p}_j^{(\mathrm{c,v})}), \quad (26)
\end{aligned}
$$

where $\widehat{\mathbf{w}}_j^{(\mathrm{r,l,u})} = \left(\phi^{-1}(w_j^{(\mathrm{c,v})}), \phi^{-1}(w_j^{(\mathrm{u})})\right)^{\top}_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}}$, $\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u})}$ is the covariance matrix of the vector $\widehat{\mathbf{w}}_j^{(\mathrm{r,l,u})}$ (i.e with size 7×7), and $\mathbf{p}_j^{(\mathrm{c,v})}$ and $\mathbf{p}_j^{(\mathrm{u})}$ are the parameter vectors of the GG margins.

For the indexing step, the estimated hyperparameters set resulting from this joint modeling of texture and depth data are used to build the feature vector:

$$
\forall j \in \{1,\dots,J\}, \qquad \mathbf{v}^{(\mathrm{r,l,u})}_{\mathrm{GC-MGG-7},j} = \left(\mathbf{p}_j^{(\mathrm{c,v})}, \mathbf{p}_j^{(\mathrm{u})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u})}\right)_{\substack{\mathrm{v}\in\{\mathrm{l,r}\}\\ \mathrm{c}\in\{\mathrm{R,G,B}\}}}. \quad (27)
$$

The associated similarity measure, used to compare the two feature vectors $(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{p}_j^{(\mathrm{u,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,q})})$ and $(\mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{p}_j^{(\mathrm{u,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,db})})$ of the query and database SI, is given by:

$$
\begin{aligned}
&\widetilde{\mathcal{D}}^{(\mathrm{r,l,u})}_{\text{GC-MGG-7,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{p}_j^{(\mathrm{u,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{p}_j^{(\mathrm{u,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,db})}) =\\
&\quad \sum_{\mathrm{v}\in\{\mathrm{l,r}\}} \sum_{\mathrm{c}\in\{\mathrm{R,G,B}\}} KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{c,v,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}) + KLD_{\mathrm{GG}}(\mathbf{p}_j^{(\mathrm{u,q})} || \mathbf{p}_j^{(\mathrm{u,db})})\\
&\quad + \frac{1}{2}\left(\mathrm{tr}((|\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,q})}|)^{-1} |\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,db})}|) + \log \frac{|\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,q})}|}{|\boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,db})}|} - 7\right). \quad (28)
\end{aligned}
$$

Finally, the resulting measure is obtained by adding the above measure over all the detail wavelet subbands:

$$
\mathcal{D}_{\text{GC-MGG-7}} = \sum_{j=1}^{J} \widetilde{\mathcal{D}}^{(\mathrm{r,l,u})}_{\text{GC-MGG-7,j}}(\mathbf{p}_j^{(\mathrm{c,v,q})}, \mathbf{p}_j^{(\mathrm{u,q})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,q})} || \mathbf{p}_j^{(\mathrm{c,v,db})}, \mathbf{p}_j^{(\mathrm{u,db})}, \boldsymbol{\Sigma}_j^{(\mathrm{r,l,u,db})}). \quad (29)
$$

4. Experimental results

4.1. Color stereo image datasets

To evaluate the performance of the proposed retrieval methods, our simulations have been carried out on the following datasets:

1. The standard Tsukuba SI dataset [28]. Due to the lack of the public availability of huge color SI databases, we propose to use three different collections of this dataset[1], which have been generated under these illumination conditions:

   - **Fluorescent**: The considered default lightening condition, with perfect appearance for all the objects and typical contrast between light and shadow.
   - **Flashlight**: The scene is only lit with a flashlight attached to the stereo camera. The light condition is low in most of the scene except the area where the flashlight is pointing.
   - **Lamps**: This illumination is the most darkened one as the scene is lit only by the subtle moonlight that comes from the window. As a result, the majority of the scene is dark and contains a wide set of under exposed objects.

   Note that each dataset collection is composed of $1,800$ color stereo scenes, of size 640×480, with their associated ground truth depth maps. This database has been employed in various stereo vision applications such as stereo matching, depth super-resolution and ego-motion estimation [29, 30]. In this work, we adapted it to be exploited and used for retrieval purpose by assigning the stereo images sharing similar visual contents and texture to the same class. Thus, 17 classes are obtained where the number of images per class ranges from 80 to 150. Fig. 6 illustrates some class samples of the Fluorescent dataset, and Fig. 7 shows one sample obtained

[1] http://www.cvlab.cs.tsukuba.ac.jp/dataset/tsukubastereo.php

under the three illumination conditions.

2. Another dataset selected from the large **FlyingThings3D**[2] database [31]. The latter is composed of 1,000 color stereo pairs as well as their associated ground truth depth maps, of size 960×540, and contains 100 classes with 10 samples per class. Similarly to Fig. 6, Fig. 8 illustrates some class examples of this database.

4.2. Comparison methods

To show the relevance of our proposed approaches compared to the existing ones, we will consider the following methods:

- Re-ranking [6]: This method consists in extracting the MPEG-7 edge histogram from the left image. Then, a refinement of the resulting image candidates is performed using a re-ranking procedure based on the depth features. To this end, histograms of the depth maps are considered as relevant features and the diffusion distance is used to measure the similarity between these histograms.

- GC-BGG-LR-D [14]: This very recent state-of-the-art method aims at using a Gaussian copula-based bivariate GG model to extract texture features from a grayscale stereo pair [14]. Moreover, depth features are extracted based on the GG modeling of the depth wavelet coefficients. It is important to note here that this method has been retained in our comparison since it has been recently shown in [14] that it outperforms other state-of-the-art stereo image retrieval methods such as the object-based stereo image retrieval algorithm [8] and the bivariate GG modeling-based approach [13].

- M-Laplacian-LR-D [10]: It is a multivariate modeling approach with Laplacian distribution which has been found to be an efficient method for color

[2] https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html

monoview image retrieval. In our context, it is separately applied to each view of the stereo pair. Furthermore, and similarly to the previous retrieval approach, note that depth features are also extracted through GG modeling of the depth wavelet coefficients.

- UGG-LR: This method consists in applying a Univariate GG (UGG) modeling to the wavelet coefficients of each color channel of the left and right views. Let us recall that this straightforward approach is described in Section 3.1

- UGG-LR-D: While only the texture information is used in the previous approach, this one propose to exploit the depth information through its UGG modeling.

- GC-BGG-LR-D-ext: It corresponds to the proposed *extension* of the Gaussian copula-based bivariate model with GG margins [14] exploiting the cross-view dependencies for each color channel, as described in Section 3.2. The UGG modeling of the depth information is also used in the indexing step.

- GC-MGG-3-LR-D: It is the first version of the proposed 3-dimensional Gaussian copula-based MGG model exploiting only the cross-channel dependencies while assuming that the two views are independents, as explained in Section 3.3.1. The UGG modeling of the depth information is also used in the indexing step.

- GC-MGG-6-LR-D: It corresponds to the second version of the proposed 6-dimensional Gaussian copula-based MGG model exploiting both the cross-view and channel dependencies, as presented in Section 3.3.2. The UGG modeling of the depth information is also used in the indexing step.

- GC-MGG-7-LRD: It designates the third version of the proposed 7-dimensional Gaussian copula-based MGG model exploiting simultaneously the cross-view/channel and depth dependencies, as addressed in Section 3.4.2.

Regarding the wavelet modeling-based retrieval approaches, it should be noted that the 8 order Daubechies wavelet transform is used for the multiscale decomposition of the color SI as well as the depth maps. Moreover, three decomposition levels were considered yielding one approximation and $J = 9$ detail subbands for each transformed input.

4.3. Performance evaluation metrics

Several objective criteria were defined to evaluate the performance of the retrieval procedure. The most widespread metrics are the precision PR *versus* recall RC ratios and the the Average Retrieval Rate (ARR):

- The precision $PR = N^r/N$ is the ratio between the number of relevant images in the returned ones N^r and the number of returned images N, whereas the recall $RC = N^r/N^t$ is the ratio between N^r and the number of relevant images in the database N^t. These two metrics are used to plot PR-RC curve in order to illustrate the exhaustive retrieval performance of the algorithm.
- The ARR is the mean percentage of relevant retrieved images over the whole database.

Note that a retrieved image is considered as relevant if it belongs to the same class of the query one.

4.4. Results and discussion

First, the performance of the different aforementioned approaches are compared in terms of precision-recall. Figures 9, 10, 11 and 12 illustrate the PR-RC curves for the fluorescent, flashlight, lamps and FlyingThings3D datasets. Several interpretations could be derived from the obtained results.

Indeed, the re-ranking approach, performed in the spatial domain, has the lowest performance for the different datasets, while the wavelet-based state-of-the-art approaches (GC-BGG-LR-D and M-Laplacian-LR-D) lead to better results. Moreover, for the Fluorescent, Lamps and FlyingThings3D datasets, one can observe that the two latter methods outperform the UGG-LR-D approach. This shows the interest of exploiting the inter-view or inter-channel dependencies in the retrieval process. It should be noted here that the UGG-LR-D approach leads to a substantial gain compared to the UGG-LR method for all the databases. This result confirms the benefits of taking advantages from both texture and depth information in stereo image retrieval. For this reason, the next proposed Gaussian copula-based bivariate and multivariate modeling approaches have been directly shown using the depth features combined with the texture ones. Thus, it can be noticed that the GC-BGG-LR-D method outperforms the independent univariate modeling approach (UGG-LR-D), which corroborates again the interest of the color cross-view correlation in enhancing the indexing outcomes. An additional gain is achieved using the GC-MGG-3-LR-D approach. This result shows that the color dependencies in each view of the stereo pair are higher than the cross-view ones. Now, by exploiting both the cross-color channel and cross-view dependencies through the GC-MGG-6-LR-D approach, the PR-RC performance is further improved. Finally, by resorting to a joint modeling approach of the texture and depth information, the last proposed approach GC-MGG-7-LRD yields the best retrieval performance for the four dataset collections.

If we focus now on the performance of the proposed retrieval approaches for the Tsukuba dataset collections, one can observe that their PR-RC curves are slightly impacted by the illumination changes and the obtained results are quite similar. Therefore, it is important to note another advantage of our proposed retrieval approaches, which is their robustness with respect to the illumination variations.

In addition to the PR-RC, the performance of the different methods is also evaluated in terms of average retrieval rates (ARR). While Table 1 provides the ARR results for the state-of-the-art methods, Tables 2, 3, 4 and 5 illustrate the results for the fluorescent, flashlight, lamps and FlyingThings3D dataset collections, respectively. Again, it can be seen that the proposed Gaussian copula-based multivariate modeling approaches lead to the best results. Moreover, Tables 2, 3, 4 and 5 show the retrieval performance for different resolution levels of the multiscale decomposition. While a gain in ARR can be observed by increasing the number of resolution levels from 1 to 2, a very small gain is achieved by considering 3 levels. This allows us to deduce that the retrieval performance becomes more stable from this level and it would be enough to set it to 3 (i.e. $J = 9$).

Finally, we propose to compare the computational complexity of the different Gaussian copula-based bivariate and multivariate models. To this end, we present in Table 6 the length of the feature vectors associated to the different statistical models as well as the execution time required to compare the features of a given query and candidate stereo images. Note that the execution time is obtained using a computer with an Intel Core i7 processor (2.6 GHz) and a Matlab implementation. While the smallest computational complexity is obtained with the GC-BGG-LR-D approach [14] since it is performed using only the luminance information, it can be noticed that the proposed approaches lead to a slight increase of the computational complexity due to the joint modeling of all color channels of both views and depth maps.

Overall, all the obtained results confirm the efficiency of the proposed approaches for color stereo image retrieval.

5. Conclusion

In this work, we have taken into account the color information as well as the diverse correlations distinguishing the stereo images for their retrieval purpose. To this end, we developed various Gaussian copula-based multivariate modeling approaches, and their corresponding estimated hyperparameters set are used as relevant features in the indexing process. Experimental results, carried out on different color stereo datasets, have illustrated the good performance of the proposed approaches. Let us recall that the main reasons behind the achieved improvements are twofold. The first one is the simultaneous capture of the cross-view and channel dependencies. The second one is the joint statistical modeling of texture and depth information. In future work, these approaches could be further extended by taking into account the intra- and inter-subband correlations in each color channel of each view. Moreover, according to some obtained preliminary results showing the good performance of a deep neural network-based approach applied separately to the left and right views, it would be interesting to investigate the appropriate way(s) of integrating the depth map in the deep neural network architecture.

References

[1] N. Bernini, M. Bertozzi, L. Castangia, M. Patander, M. Sabbatelli, Real-time obstacle detection using stereo vision for autonomous ground vehicles: A survey, in: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2014, pp. 873–878.

[2] B. Sdiri, M. Kaaniche, F. A. Cheikh, A. Beghdadi, O. J. Elle, Efficient enhancement of stereo endoscopic images based on joint wavelet decomposition and binocular combination, IEEE Transactions on Medical Imaging (2018) 1–1.

[3] J. Krauskopf, J. D. Forte, Influence of chromaticity on vernier and stereo acuity, Journal of Vision 2 (9) (2002) 645–652.

[4] J. A. Martins, J. M. Rodrigues, H. du Buf, Luminance, colour, viewpoint and border enhanced disparity energy model, Public Library of Science (PLOS one) 10 (6) (2015) e0129908.

[5] H. Den Ouden, R. Van Ee, E. De Haan, Colour helps to solve the binocular matching problem, The Journal of Physiology 567 (2) (2005) 665–671.

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

[7] F. Peng, L. Wang, J. Gong, H. Wu, Development of a framework for stereo image retrieval with both height and planar features, IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing 8 (2) (2015) 800–815.

[8] X. Xu, W. Geng, R. Ju, Y. Yang, T. Ren, G. Wu, OBSIR: object-based stereo image retrieval, in: IEEE International Conference on Multimedia and Expo, 2014, pp. 1–6.

[9] A. D. El Maliani, M. El Hassouni, Y. Berthoumieu, D. Aboutajdine, Multi-model approach for multicomponent texture classification, in: International Conference on Image and Signal Processing, Springer, 2012, pp. 36–44.

[10] G. Verdoolaege, P. Scheunders, On the geometry of multivariate generalized Gaussian models, Journal of Mathematical Imaging and Vision 43 (3) (2012) 180–193.

[11] S. Sakji-Nsibi, A. Benazza-Benyahia, Copula-based statistical models for multicomponent image retrieval in the wavelet transform domain, in: 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 253–256.

[12] 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.

[13] A. Chaker, M. Kaaniche, A. Benazza-Benyahia, Disparity based stereo image retrieval through univariate and bivariate models, Signal Processing: Image Communication 31 (2015) 174–184.

[14] A. Karine, A. D. El Maliani, M. El Hassouni, A novel statistical model for content-based stereo image retrieval in the complex wavelet domain, Journal of Visual Communication and Image Representation 50 (2018) 27–39.

[15] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press, 2009.

[16] J. R. Mathiassen, A. Skavhaug, K. Bø, Texture similarity measure using kullback-leibler divergence between gamma distributions, in: European conference on computer vision, Springer, 2002, pp. 133–147.

[17] 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.

[18] 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.

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

[20] S. C. Choi, R. Wette, Maximum likelihood estimation of the parameters of the gamma distribution and their bias, Technometrics 11 (4) (1969) 683–690.

[21] S.-H. Cha, Comprehensive survey on distance/similarity, International Journal of Mathematical Models and Methods in Applied Sciences 1 (2007) 300–307.

[22] S. Kullback, R. A. Leibler, On information and sufficiency, Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[23] 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.

[24] N.-E. Lasmar, Y. Berthoumieu, Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms, IEEE Transactions on Image Processing 23 (5) (2014) 2246–2261.

[25] B. Potetz, T. S. Lee, Statistical correlations between two dimensional images and three-dimensional structures in natural scenes, Journal of the Optical Society of America 20 (7) (2003) 1292–1303.

[26] C. C. Su, L. K. Cormack, A. C. Bovik, Color and depth priors in natural images, IEEE Transactions on Image Processing 22 (6) (2013) 2259–2274.

[27] N. Fisher, P. Switzer, Chi-plots for assessing dependence, Biometrika 72 (2) (1985) 253–265.

[28] S. Martull, M. Peris, K. Fukui, Realistic CG stereo image dataset with ground truth disparity maps, in: International Conference on Pattern Recognition, Workshop TrakMark2012, 2012, pp. 40–42.

[29] G. Riegler, M. Rüther, H. Bischof, ATGV-Net: Accurate depth super-resolution, in: European Conference on Computer Vision, Springer, 2016, pp. 268–284.

[30] A. Patron-Perez, S. Lovegrove, G. Sibley, A spline-based trajectory representation for sensor fusion and rolling shutter cameras, International Journal of Computer Vision 113 (3) (2015) 208–219.

[31] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox, A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4040–4048.

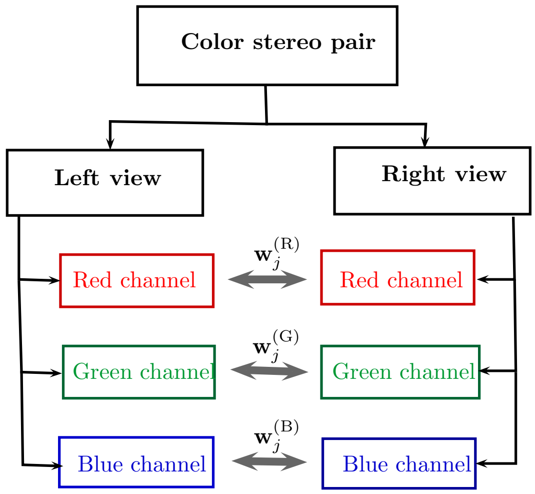

Figure 1: Exploiting cross-view dependencies for bivariate modeling approach.

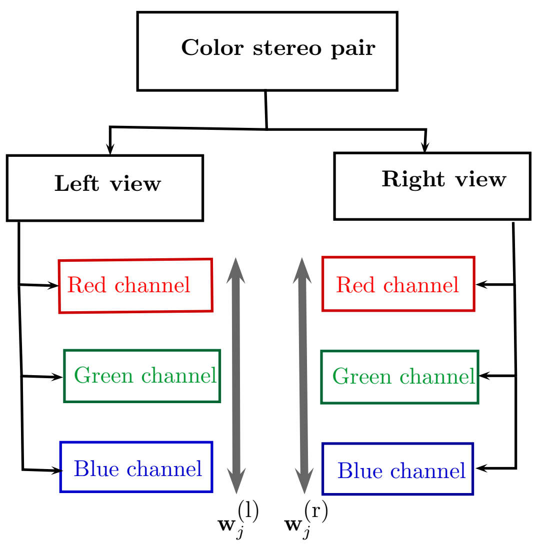

Figure 2: Exploiting cross-channel dependencies for 3-dimensional multivariate modeling approach.

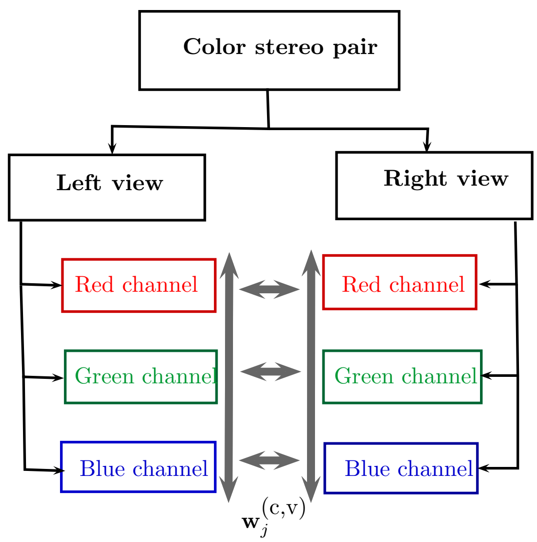

Figure 3: Exploiting cross-view and channel dependencies for 6-dimensional multivariate modeling approach.

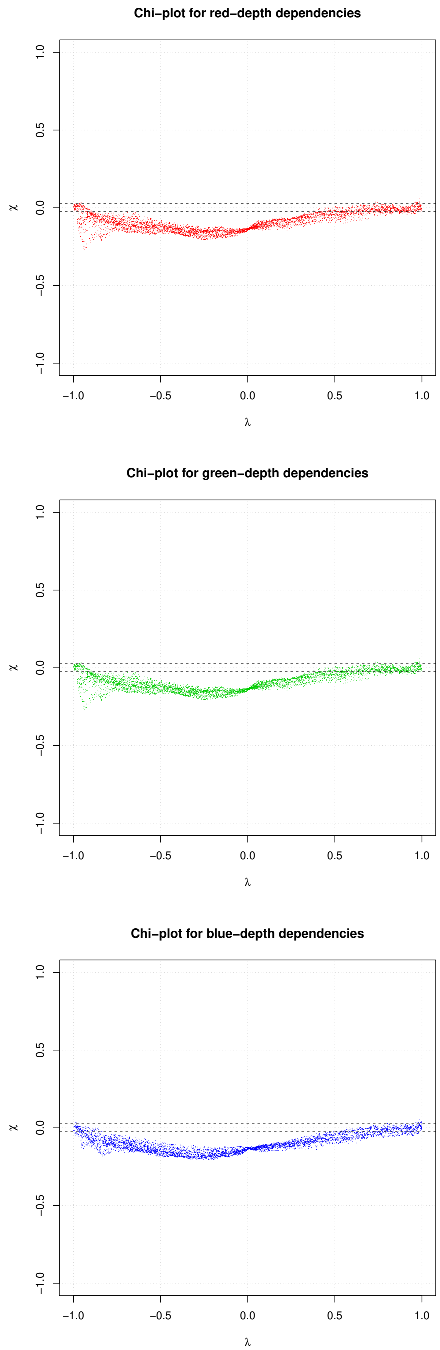

Figure 4: Chi-plots illustrating correlations between color and depth components of one view in the Fluorescent dataset.

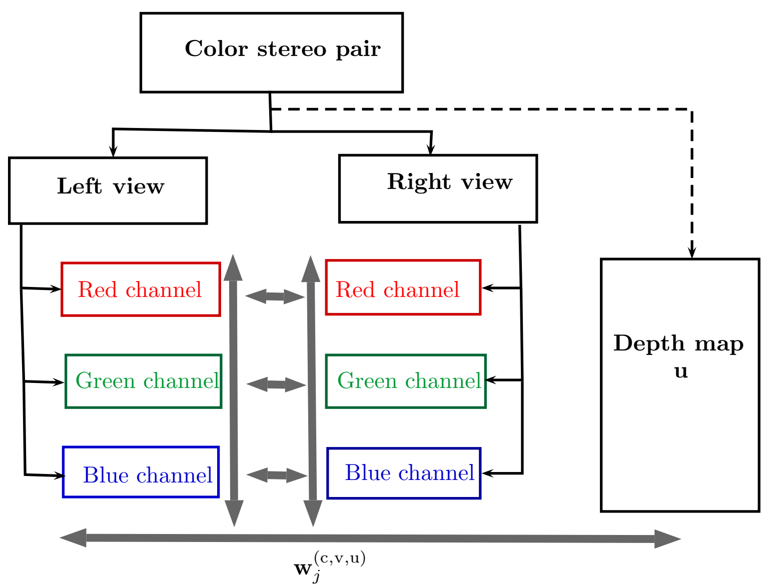

Figure 5: Exploiting cross-view/channel and depth dependencies for 7-dimensional multivariate modeling approach.

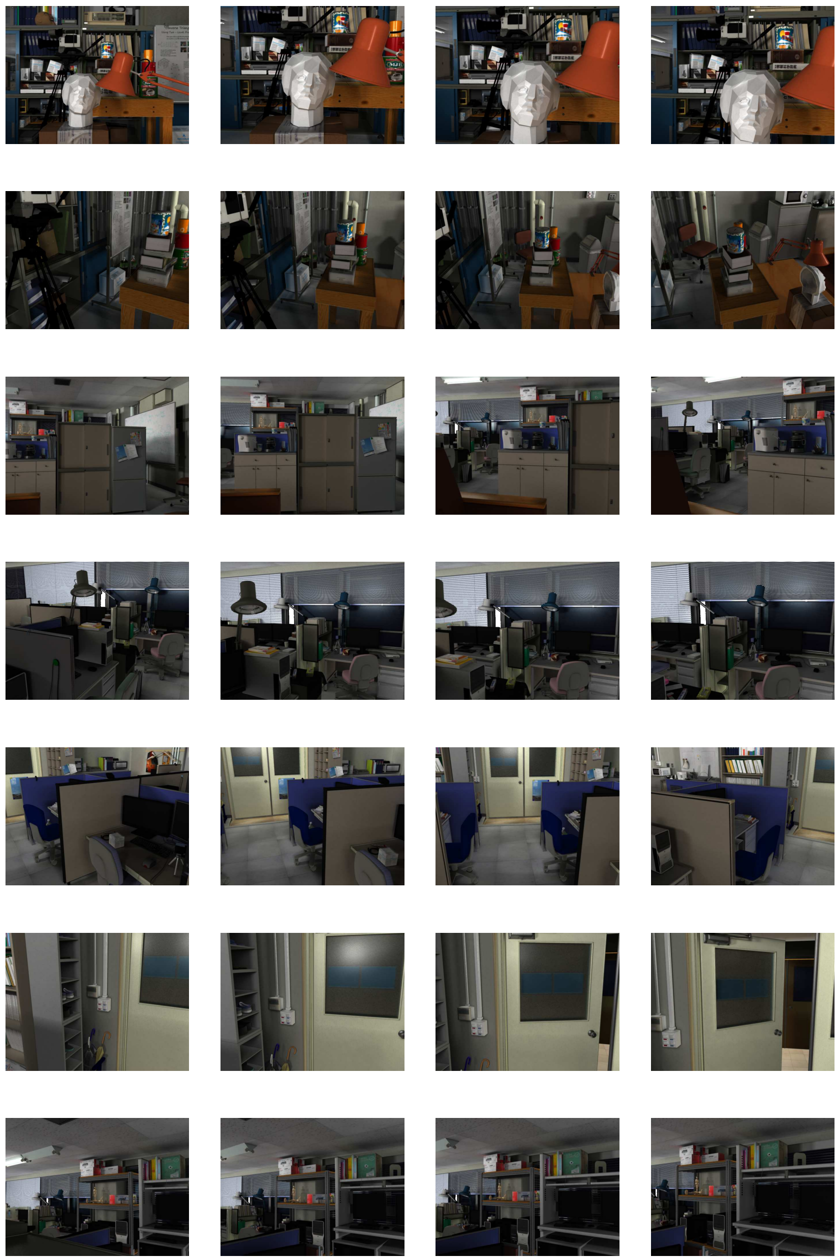

Figure 6: Some class samples of the **Fluorescent Tsukuba** database.

(a) Fluorescent. (b) Flashlight. (c) Lamps.

Figure 7: Example of the three different illuminations.

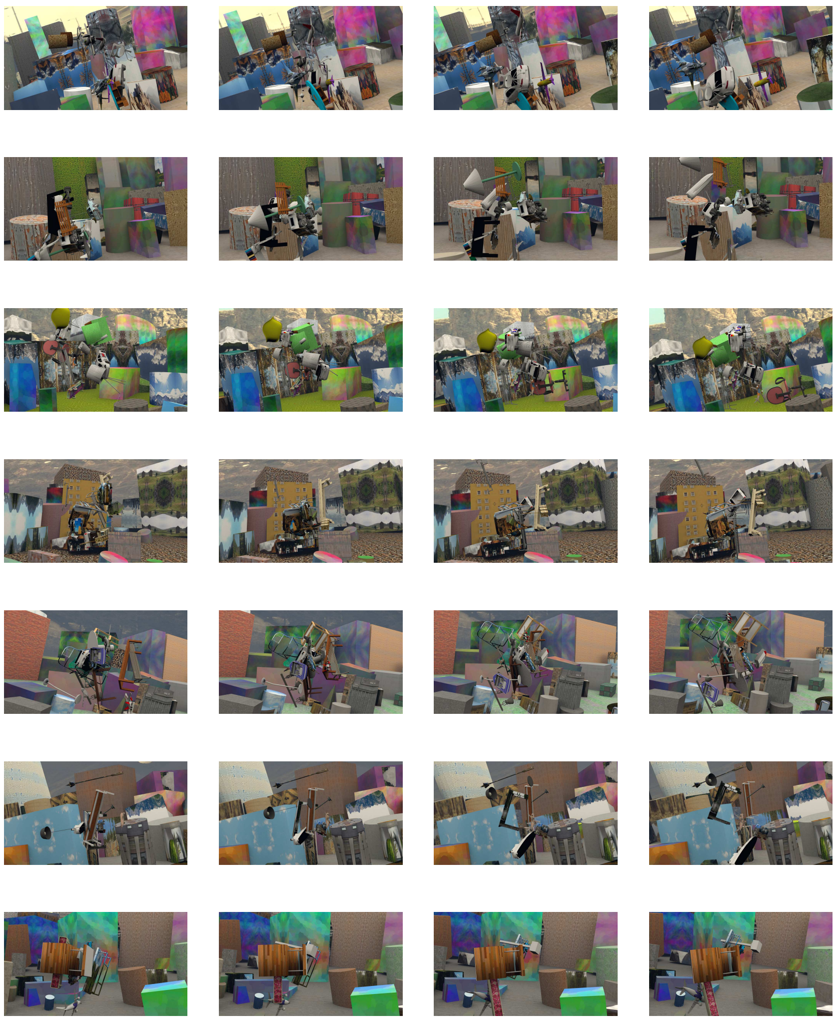

Figure 8: Some class samples of the **FlyingThings3D** database collection.

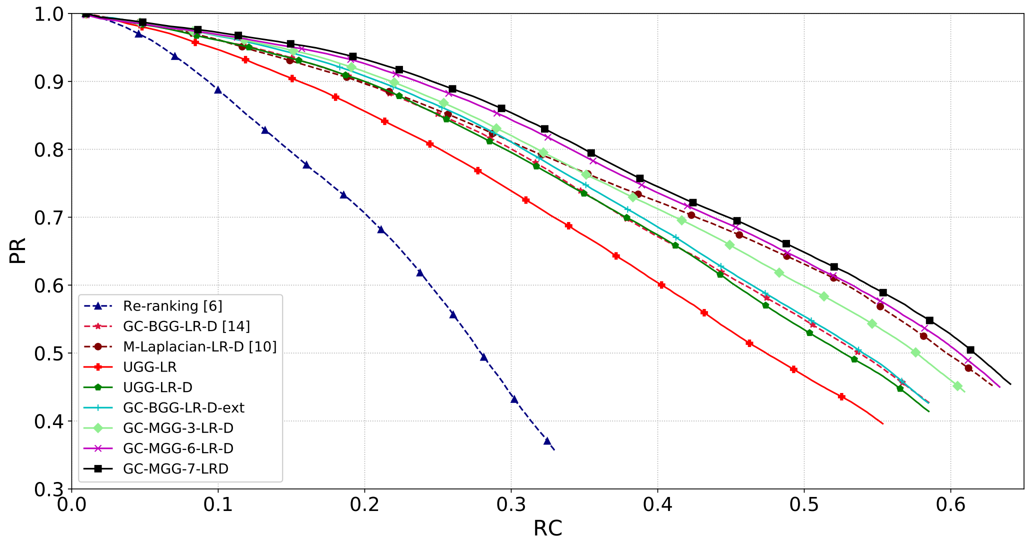

Figure 9: Precision-Recall curves of the proposed and state-of-the-art approaches for the **Fluorescent** illumination.

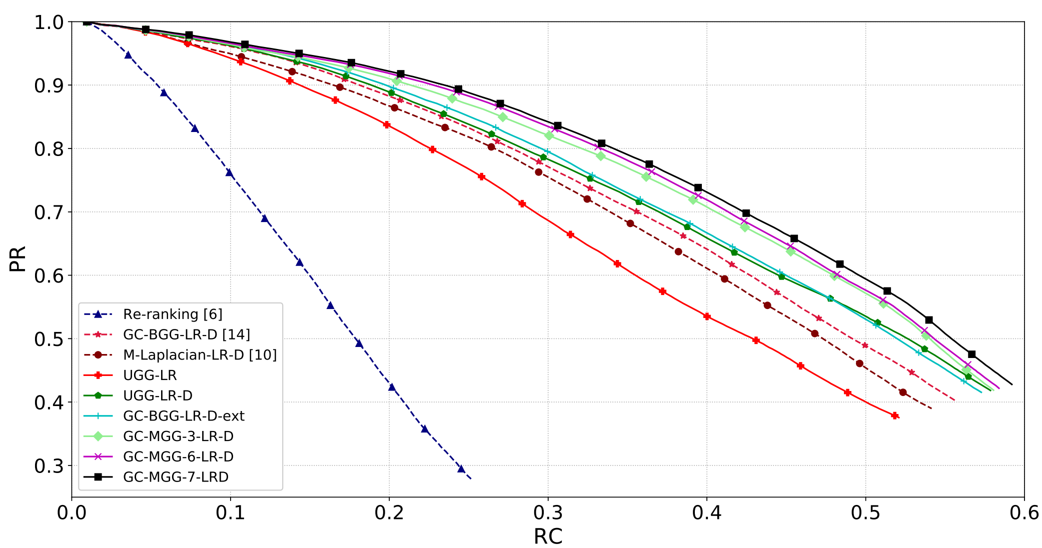

Figure 10: Precision-Recall curves of the proposed and state-of-the-art approaches for the **Flashlight** illumination.

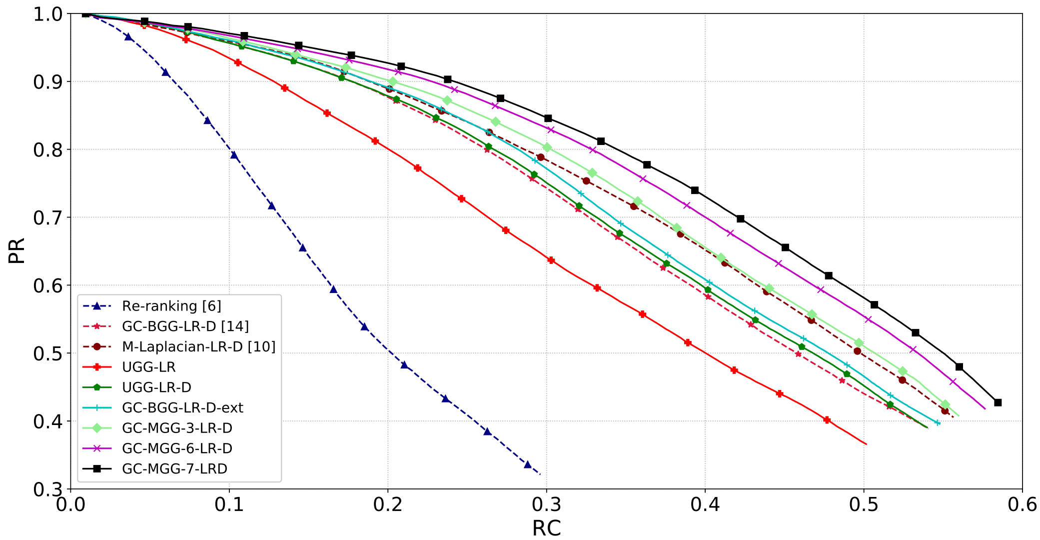

Figure 11: Precision-Recall curves of the proposed and state-of-the-art approaches for the **Lamps** illumination.

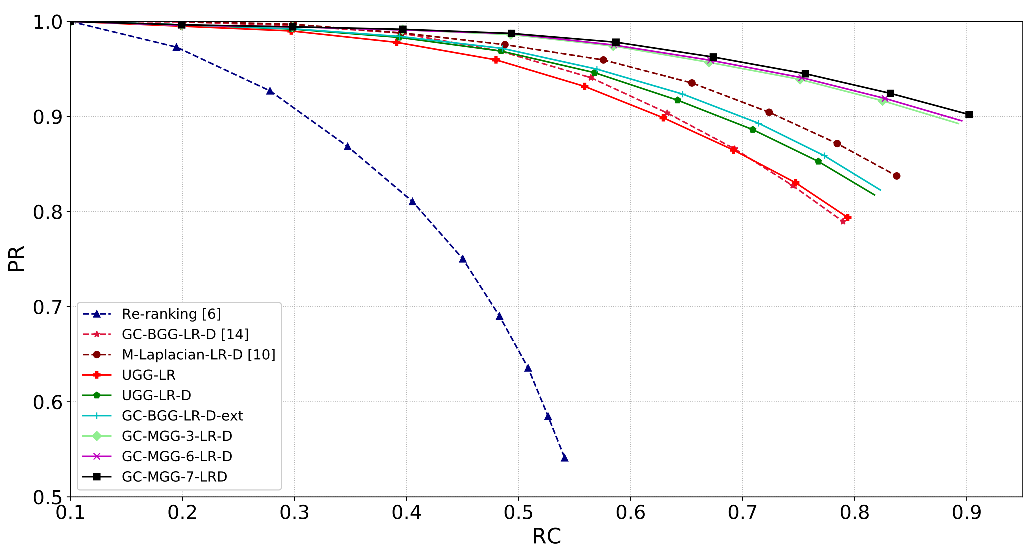

Figure 12: Precision-Recall curves of the proposed and state-of-the-art approaches for the **FlyingThings3D** database collection.

State-of-the-art methods	Fluorescent	Flashlight	Lamps	FlyingThings3D
Re-ranking [6]	56.22	52.14	60.31	64.25
GC-BGG-LR-D [14]	78.01	71.68	80.26	83.25
M-Laplacian-LR-D [10]	**80.13**	**74.23**	**80.56**	**86.18**

Table 1: ARR rates of the state-of-the-art methods.

	1 Scale	2 Scales	3 Scales
UGG-LR	75.71	77.57	78.22
UGG-LR-D	81.43	82.74	82.96
GC-BGG-LR-D-ext	84.28	84.13	84.91
GC-MGG-3-LR-D	85.29	87.35	87.52
GC-MGG-6-LR-D	87.57	88.33	89.26
GC-MGG-7-LRD	**88.64**	**89.68**	**90.00**

Table 2: ARR rates of the proposed methods for the **Fluorescent** illumination.

	1 Scale	2 Scales	3 Scales
UGG-LR	72.85	74.72	75.29
UGG-LR-D	77.43	78.00	78.82
GC-BGG-LR-D-ext	79.44	80.26	81.09
GC-MGG-3-LR-D	83.60	84.92	85.11
GC-MGG-6-LR-D	85.03	86.23	87.02
GC-MGG-7-LRD	**87.61**	**88.35**	**88.91**

Table 3: ARR rates of the proposed methods for the **Flashlight** illumination.

	1 Scale	2 Scales	3 Scales
UGG-LR	75.56	79.29	80.17
UGG-LR-D	78.71	81.46	82.39
GC-BGG-LR-D-ext	79.84	82.62	83.40
GC-MGG-3-LR-D	81.00	84.63	85.75
GC-MGG-6-LR-D	83.61	87.84	87.96
GC-MGG-7-LRD	**84.43**	**88.90**	**89.02**

Table 4: ARR rates of the proposed methods for the **Lamps** illumination.

	1 Scale	2 Scales	3 Scales
UGG-LR	80.13	82.54	83.67
UGG-LR-D	82.56	84.09	85.23
GC-BGG-LR-D-ext	83.71	85.83	85.06
GC-MGG-3-LR-D	87.16	88.49	89.28
GC-MGG-6-LR-D	88.78	89.14	90.51
GC-MGG-7-LRD	**90.11**	**92.02**	**92.83**

Table 5: ARR rates of the proposed methods for the **FlyingThings3D** dataset.

	Length of feature vector	Runtime (in seconds)
GC-BGG-LR-D [14]	36	2.19
GC-BGG-LR-D-ext	90	4.21
GC-MGG-3-LR-D	81	4.46
GC-MGG-6-LR-D	72	4.66
GC-MGG-7-LRD	72	5.19

Table 6: Computational cost of the differnet copula-based approaches.