Journal of Visual Communication and Image Representation 50 (2018) 27–39

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

A novel statistical model for content-based stereo image retrieval in the complex wavelet domain☆

Ayoub Karine[a], Ahmed Drissi El Maliani[b], Mohammed El Hassouni[a,c,*]

[a] LRIT - CNRST URAC 29, Rabat IT Center, Faculty of Sciences, Mohammed V University in Rabat, Morocco
[b] L.I.M. Faculty of Sciences Dhar el Mahraz, USMBA, Fez, Morocco
[c] LRIT - CNRST URAC 29, Rabat IT Center, FLSH, Mohammed V University in Rabat, Morocco

ARTICLE INFO

Keywords:
Content-based stereo image retrieval
Gaussian copula
Non-Gaussian distribution
Jeffrey divergence
Complex wavelet transform
Feature extraction

ABSTRACT

This paper presents a new stereo image (SI) retrieval method based on a statistical model of complex wavelet coefficients subbands. In this context, a Gaussian copula-based multivariate model is used to capture the dependence between complex wavelet coefficients of both left and right images, and a non-Gaussian univariate model is used to characterize the statistical behavior of the disparity map. Thanks to its flexibility, the copula tool allows us to choose several marginal densities while keeping the multivariate properties. Features are extracted by estimating parameters for both multivariate and univariate models. Finally, a weighted Jeffrey divergence (JD) is used as a similarity measurement between the underlying models. Experimental results on a stereo image database demonstrate the performance of the proposed method in terms of the retrieval rates as well as the computational time.

1. Introduction

Stereo image (SI) is of great importance for various realistic applications such as video game systems, robot vision, 3DTV, military surveillance and telepresence in videoconferences [1,2]. It is the reason why more scientists gave special attention to this subject. It is also the main focus of this research paper.

The principle of stereoscopic imaging systems is inspired from the mechanism of the human brain in which it consists in generating two views called left image and right image, taken from different points of view of the same scene. Basing on this pair images, a disparity map can be estimated to describe the depth information of the scene [3]. However, in some applications such as 3D-TV, the disparity map is extracted from monoview image in order to convert 2D scene to the 3D one [4].

Recently, with the development of acquisition systems and display technologies, a large collection of stereo images is being generated. To simplify the management of these images, it is necessary to develop powerful and effective methods to describe visual information. For instance, content-based image retrieval (CBIR) deals with retrieving the relevant images, similar to a user query image. Generally, this system runs on two steps. The first one is feature extraction (FE) which consists of computing features (signatures) from each image in the database. The second is devoted to the calculation of the distance between the query and each image in the database which is called similarity measurement (SM).

The extraction of good features to characterize different textures is a fundamental step of various image processing applications and it has been subject of an intensive study. It is done by applying different transformations like wavelet transforms that produce multi-scale/multi-resolution image representations. In the literature, several methods of wavelet decomposition are developed, the most used in CBIR system are: discrete wavelet transform (DWT) [5], steerable pyramid (SP) [6], lifting scheme (LS) [7], Gabor [8], Pyramidal Dual Tree Directional Filter Bank (PDTDFB) [9], curvelet [10], contourlet [11] and dual-tree complex wavelet transform (DT-CWT) [12]. Moreover, several works proved the efficiency of using statistical distribution modeling of wavelet coefficients in a subband [13–15]. This way of representation is characterized by a few and pertinent sets of parameters and it is widespread in texture retrieval frameworks.

Recent studies verified that histograms of wavelet detail coefficients have a non-Gaussian behavior with heavy tails. In the context of using the statistical framework in the wavelet-based texture retrieval problem in the case of mono-view image, several univariate models have been used such as the generalized Gaussian distribution (GGD) [13], the generalized Gamma distribution (GΓD) [15], the Weibull distribution [14] and the Gamma distribution [14]. Recently, authors underlined

☆ This paper has been recommended for acceptance by Zicheng Liu.
* Corresponding author at: LRIT - CNRST URAC 29, Rabat IT Center, Faculty of Sciences, Mohammed V University in Rabat, Morocco.
E-mail address: mohamed.elhassouni@gmail.com (M. El Hassouni).

https://doi.org/10.1016/j.jvcir.2017.11.006
Received 1 March 2016; Received in revised form 21 August 2017; Accepted 5 November 2017
Available online 10 November 2017
1047-3203/

the importance of exploiting the dependencies between different wavelet subbands. In consequence, multivariate statistical models have been studied. Bombrun et al. modeled the wavelet coefficient by using a SIRV (Spherically Invariant Random Vectors) model [16]. Verdoolaege et al. used a multivariate generalized Gaussian model (MGGD) [17] in the wavelet domain. EL Maliani et al. employed a multivariate generalized Gamma distribution (MGΓD) for modeling the dependence between color components [18]. Other approaches lie in the direction of using the multivariate wavelet modeling using copula theory [19]. Sakji-Nsibi et al. [20,21] employed different families of copula for multicomponent image indexing in the wavelet transform domain. Bombrun et al. [16] proposed a multivariate Gamma model using Gaussian copula (GC-MGam) for texture classification. In their pioneering work, Kwitt et al. compared a Student *t* and a Gaussian copulas with Weibull margins (GC-MWbl). They concluded that the Gaussian copula is the best one for texture retrieval [22]. Based on this work, Lasmar et al. [23] introduced a motivation of using a dependence inside each subband and studied the exploitation of a GGD as a marginal density function of Gaussian copula (GC-MGG).

On the other hand, for the similarity measurement step, several metrics have been used, e.g., the Kullback-Leibler divergence (KLD) and its symmetric version Jeffery divergence (JD) [24], the normalized Euclidean distance [25], the geodesic distance [26] and the Cauchy-Schwarz divergence [27,28]. In the copula framework, Sakji-Nsibi et al. [20] employed the normalized Euclidean distance. Same authors [21] approximated the KLD between copula-based models by using a Monte-Carlo (MC) approach. Kwitt et al. [22] measured the similarity between copula-based models by employing a likelihood similarity. The lack of a closed-form analytic KLD expression for copulas has been addressed in the work of [23], where authors derived a closed forms of both KLD and JD. Latterly, [29,30] also proposed a closed form of JD and KLD in the case of rotation-invariant texture retrieval.

For stereo image retrieval, the literature is very limited. Feng et al. [31] presented a strategy for retrieving stereo image and video. The idea behind this work is to use the disparity map in order to refine the retrieval results provided from the left image. For this end, the MPEG-7 edge histogram is used to extract the feature from the left image and the diffusion distance is adopted to calculate the similarity measurement between the histograms of the disparity maps. Recently, Peng et al. [32] developed a generic framework for content-based HrosSI (high-resolution optical satellite stereo images retrieval). The proposed approach has two steps. First, regional fractal dimension (RFD) and normalized regional fractal cooccurrence matrix (NRFCM) are calculated. After that, the similarity match between descriptors is performed. In [33], Xu et al. proposed an object-based stereo image retrieval framework (OBSIR) for color stereo images based on a salient object segmentation and visual feature descriptors which are the pyramid histogram of visual words (PHOW) and local binary pattern (LBP). In [34], the stereo image and disparity map information were studied in the wavelet domain. Then, a feature vector was constructed by the statistical model parameters of wavelet coefficients resulting from LS decomposition. For this end, two strategies were investigated: univariate configuration by using a GGD and multivariate configuration by using bivariate generalized Gaussian (BGG) distribution. Finally, the distance between the query and each candidate image was computed by using a KLD. This approach leads to interesting results thanks to the pertinent characterization achieved by BGG in order to model dependence between left and right stereo pairs.

In this paper, we develop a new content-based stereo image retrieval methods closest in methodology to the work by Chaker et al. [34]. Our goal is to propose a method that provide a good trade-off between retrieval rate and runtime comparing to competing methods. Actually, even if that the Bivariate Generalized Gaussian model offers a considerable improvement over univariate based stereo image retrieval approaches, it presents some limitations. First, the marginal modeling and dependence structure are jointly considered. Nevertheless, it is known that in some cases the dependence information is not uniform with the margins, and has to be modeled separately. Furthermore, even the margins might have different distributions. Second, the BGG based approach could suffer from the complexity of its similarity measurement due to a hyper-geometric function of the KLD. To overcome these shortcoming, our challenge is to find (i) a flexible wavelet decomposition which can produce more and pertinent informations and (ii) a statistical model which is able to capture both marginal distribution of wavelet subbands and dependence structure among them. The related hyper-parameters can be easily estimated in a reasonable runtime and have a closed form of similarity measurement between their PDFs. For this end, our contribution is based on multivariate modeling of both left and right images in the complex wavelet domain by using the copula theory and DT-CWT. Precisely, we use a Gaussian copula-based multivariate model which provides a way of constructing a multivariate distribution while taking into account the information of margins and the dependence between wavelet coefficients for both left and right images. We also consider the related disparity field of stereo pair in retrieval scheme, using different non-Gaussian distributions. For similarity measure, closed forms of the JD between statistical models are exploited. The retrieval experiments follows the standard protocol adopted for CBIR systems assessment.

The remainder of this paper is organized as follows. We address in Section 2 the general statistical framework of the stereo image retrieval. After that, the proposed stereo image retrieval approach is described in Section 3. Then, in Section 4, we present the experimental results which is followed by conclusions and future works.

2. Content-based stereo image retrieval framework in a statistical context

The statistical behavior of stereoscopic images and disparity map in the wavelet domain have underlined firstly by Yang et al. [35,36]. By definition, the architecture of CBIR systems is composed by two major tasks which are feature extraction and similarity measurement. This scheme can be extended to a stereo image retrieval in a statistical framework as illustrated in Fig. 1. It consists on three parts: the data transformation, the statistical modeling and the similarity measurement. The first two parts produce a feature of each stereo image and the related disparity map, while the third part aims to calculate similarity between them.

2.1. Feature extraction

The stereo pair and disparity map can be seen as three monoview images. Each one of them is transformed separately in a wavelet domain. This way of transformation captures details of an image at different scales and orientations and are extensively used in the texture retrieval. Mathematically, the bivariate vector $V_{i,j}^{J}$ of a stereo pair i in an orientation j and a scale J of wavelet transform is expressed as follows:

$$V_{i,j}^{J} = \begin{pmatrix} \underset{i_{l_{1,j}}}{\overset{J}{\mathrm{W}}} & \cdots & \underset{i_{l_{n^J,j}}}{\overset{J}{\mathrm{W}}} \\ \underset{i_{r_{1,j}}}{\overset{J}{\mathrm{W}}} & \cdots & \underset{i_{r_{n^J,j}}}{\overset{J}{\mathrm{W}}} \end{pmatrix} \qquad (1)$$

where:

- $\underset{i_{l_{1,j}}}{\overset{J}{\mathrm{W}}},\ldots,\underset{i_{l_{n^J,j}}}{\overset{J}{\mathrm{W}}}$ and $\underset{i_{r_{1,j}}}{\overset{J}{\mathrm{W}}},\ldots,\underset{i_{r_{n^J,j}}}{\overset{J}{\mathrm{W}}}$ are respectively, the wavelet coefficient of left and right images.
- n^J is the size of the subband and i is the index of the stereo pair.

In addition, the univariate vector $U_{i,j}^{J}$ of disparity map i is considered as a random variable in an orientation j and a scale J of wavelet transform as follows: