Signal Processing: *Image Communication* 31 (2015) 174–184

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Disparity based stereo image retrieval through univariate and bivariate models

A. Chaker [a], M. Kaaniche [b,*], A. Benazza-Benyahia [a]

[a] COSIM Lab., SUP'COM, Carthage Univ., Cité Technologique des Communications, Tunisia
[b] Institut Galilée, L2TI, Université Paris 13, Sorbonne Paris Cité, France

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form
3 November 2014
Accepted 12 December 2014
Available online 24 December 2014

Keywords:
Content based image retrieval
Wavelet domain
Stereo pairs
Disparity map
Feature extraction

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.

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 [4–7]. For instance, in [5], the energy of the subbands is 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

* Corresponding author.
E-mail addresses: amani.chaker@supcom.tn (A. Chaker), mounir.kaaniche@univ-paris13.fr (M. Kaaniche), benazza.amel@supcom.rnu.tn (A. Benazza-Benyahia).

http://dx.doi.org/10.1016/j.image.2014.12.004
0923-5965/

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 [9–11]. 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 monoview 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 monoview images such as denoising [18–20] or retrieval of still and multicomponent images [11,21–23], 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 monoview 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, α_j and β_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^{db} of an image in the database I^{db} and the pdf f_j^{q} of the query image I^{q}.

Very often, the Kullback–Leibler Divergence (KLD) is retained [9,23,27,28].