J. Vis. Commun. Image R. 33 (2015) 85–93

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A new image retrieval model based on monogenic signal representation ☆

Zhiyong Zeng *, Liwei Song, Qicai Zheng, Yanling Chi

Faculty of Software, Fujian Normal University, Fuzhou 350108, China

ARTICLE INFO

Article history:
Received 4 June 2014
Accepted 27 August 2015
Available online 14 September 2015

Keywords:
Monogenic signal representation
Local binary pattern
Directional texture descriptor
Content-based image retrieval
Block-based Fisher linear discriminant
Feature fusion
Amplitude
Orientation
Phase

ABSTRACT

This paper proposes a novel image retrieval model based on monogenic signal representation. An original image is decomposed into three complementary components: amplitude, orientation and phase by monogenic signal representation. The monogenic variation in each local region and monogenic feature in each pixel are encoded, and then the statistical features of the local features encoded are calculated. In order to overcome the problem of high feature dimensionality, the local statistical features extracted from the complementary monogenic components are projected by block-based fisher discriminant analysis, which not only reduces the dimensionality of the features extracted, but also enhances its discriminative power. Finally, these features reduced are fused for effective image retrieval. Experimental results show that our scheme can effectively describe an image, and obviously improve the average retrieval precision.

1. Introduction

With the development of digital media devices and internet, a large amount of data of image and video are produced, it is critical that effective index and retrieval are developed for large image database system. Images are manually annotated using keyword for traditional image retrieval method; however, it is difficult to describe diversity and fuzziness of image through manual annotation. Content-based image retrieval (CBIR) was proposed by some researchers, which is the key issues in image processing and computer vision community. In CBIR technology research, the representation and extraction is a very important basic work. Color, texture, shape and space relation are the most frequently-used features in CBIR.

Textures are perceived by human visual system, particularly, on the aspects of orientation and scale of texture patterns. Texture is a very important visual feature that reflects innate surface properties of an object and relationship to the surrounding environment. Many objects in an image can be recognized solely by their texture without any other information. Conventional textures extraction method used include gray level co-occurrence matrix [1], Tamura texture descriptor [2], Gaussian mixed model [3,4], Markov random field [5], wavelet transform [6], DCT transform [7] and Gabor filtering [8] etc. Most texture description method in transform domain such as Gabor transform descriptor only uses magnitude acquired Gabor transform, little reports use phase information of Gabor transform, few reports use simultaneously magnitude, orientation and phase information of Gabor transformation [9,10].

In recent years, local binary pattern (LBP) attained significant achievements in texture analysis and face recognition [11,12], and many improved algorithms were proposed [13]. Heikkila proposed central symmetry LBP to reduce the dimensionality of LBP features so that interest regions in an image were described by LBP [14]. To overcome the sensibility of intensity caused by image variety, Zhang presented multiscale Gabor magnitude feature map, then he used LBP operator to encode LBP magnitude feature map to extract local statistical feature [15]. Afterward, Zhang used multiscale Gabor phase instead of Gabor magnitude to encode the global and local variety of real part and imaginary part of Gabor filtering coefficients [16]. Xie utilized XOR operator to encode local change of Gabor phase and fused Gabor magnitude and phase into as a whole, this method obtained satisfactory result of image recognition [17]. Zhou exploited the fine statistical characteristics of Haar feature and he proposed Haar local binary pattern (HLBP) [18]. Trefny proposed directional local binary descriptor based on LBP [19]. Wu presented local edge binary pattern and improved orientation texture feature using the definition of local edge and

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: zzyong@fjnu.edu.cn (Z. Zeng).

http://dx.doi.org/10.1016/j.jvcir.2015.08.014

orientation [20]. Su extracted global and local feature of Discrete Fourier Transform and Gabor Wavelet Transform respectively, the global and local features were combined in both serial and parallel manner, which not only greatly increased the system accuracy but also improve system speed [21]. Owing to the relative simplicity, low compute complex and fusing structural and statistical feature of texture, meanwhile, with rotate and scale invariant prominent advantage, this method is widely used in the field of image matching, pedestrian, detection and tracing of car object, living things and medical image analysis [22–24].

In order to further improve image retrieval performance, many feature combination methods and relevant feedback techniques have been proposed recently [25–31]. Liu proposed a novel image feature detecting and describing method called color difference histogram (CDH). CDH encodes the perceptually uniform color difference between two points under different backgrounds with regard to color and edge orientation via feature vector in a similar manner to the human visual system [25]. Wang proposed a novel image retrieval scheme based on structure elements' descriptor (SED) [26]. SED effectively represents the spatial correlation of color and texture by extracting the color and texture features of an image. Gao et al. present hypergraph learning approach to determine the relevance estimation, and simultaneously utilize both visual word and tags to estimate the relevance of user [27]. Lee et al. propose a novel image retrieval scheme that extracts the image feature by fusing Advanced Speed-Up Robust Feature (ASURF) and Dominant Color Descriptor (DCD); the system can run in real-time on iPhone and find a natural color image for mobile image retrieval [28]. Kafai et al. describe Discrete Cosine Transform (DCT) hashing for creating index structures for face descriptors, this method can efficiently reduce the cost of the linear search and improve retrieval efficiency and accuracy [29]. Yang et al. propose a semi-supervised Local Regression and Global Alignment (LRGA) algorithm for data ranking and a semi-supervised long-term Relevance Feedback (RF) algorithm for using data distribution and the history RF information, then they integrate the two algorithms into multimedia content analysis and retrieval framework [30]. Spyromitros-Xioufis et al. use the framework of VLAD and Product Quantization to develop an enhanced framework for large-scale image retrieval, the system significantly improves the performance for image retrieval [31].

Although the LBP in texture recognition has achieved great success, most of the LBP features are encoded to the image gray level, and the amplitude, orientation and phase in transform domain are rarely utilized. Practically, human visual system is reasonable sensitive to the information of contour profile, direction and phase, it is necessary for us to pay more attention to the effective use of this information. Meanwhile, when LBP is encoded, only the gray-value changes between the central pixel and its neighbor pixels in a local region is utilized, and the gray-value variance changes between the central pixel and its neighbor pixels in a local region is omitted, this means texture directions are ignored, which leads to low accuracy of image recognition. In addition, many local texture extraction methods don't take into account the information of central pixel, at the same time, some methods of feature reduction will decrease the performance of image classification. Finally, different LBP texture analyzing methods mainly are aimed at texture database which is not too complex, it is a challenge that the underlying mechanism of LBP would be further studied for more complicated natural image database.

In order to solve the drawbacks of single basic LBP mode difficult to distinguish natural texture local changes, in this paper, we propose a texture descriptor called directional monogenic binary pattern (DMBP) by fusing the amplitude, orientation, phase and the central pixel information of monogenic signal, not only the frequency-value changes of monogenic signal but also its variance changes between the central pixel and its central-symmetric neighbor pixels in a local region are considered together in the new descriptor, this method is simple and can improve effective expression of complicated natural texture feature.

The content of the paper is organized as follows. Section 1 introduces briefly related works of texture analysis method. Section 2 presents the monogenic signal representation. Section 3 describes in detail DMBP feature extraction and representation model. Section 4 presents the calculation of DMBP histogram and the reduced dimensionality method of DMBP histogram. Section 5 is the dissimilarity measurement. Section 6 describes the experimental results, and Section 7 concludes the paper.

2. Construction of monogenic signal of an image

Monogenic signal is proposed by Felsberg and Sommer in 2001 [32], and used in many applications such as face recognition [33]. Monogenic signal is an important isotropic generation of analytic signal from one dimension to two dimensions. It is an effective tool to analyze intrinsic one dimension 2-D signal (such as lines and edges) in a rotation invariant manner. The monogenic signal idea is that it implements the decomposition of identity of signal. For an image $f(x,y)$, its monogenic signal is defined as:

$$f_{MS}(x,y) = f(x,y) - (i,j)f_R(x,y) \tag{1}$$

where,

$$f_R(\bar{x}) = \frac{\bar{x}}{2\pi|\bar{x}|^3} * f(\bar{x}) \tag{2}$$

refers to Riesz transform, (i,j) is two imaginary axes of incomplete quaternion, $\bar{x} = [xy]^T \in R^2$. Here, bandpass filters is log-Gabor filters, its transfer function is:

$$H_{LG}(u,v) = \exp\left(-\frac{\left(\log\left(\sqrt{u^2+v^2}/\omega_0\right)\right)^2}{2(\log(k/\omega_0))^2}\right) \tag{3}$$

where, ω_0 refers to central frequency of filters, k is scale factor. and its transfer function in Fourier domain is:

$$(H_1(u), H_2(u)) = \left(-i\frac{u}{|u|}, -i\frac{v}{|u|}\right), \quad u = (u,v) \in R^2 \tag{4}$$

Therefore, for an image $f(x,y)$, its monogenic representation can be considered as an incomplete quaternion, which is defined as the combination of $f(x,y)$ and its Riesz transform:

$$f_{MS}(x,y) = (f_{\lg}(x,y), f_{\lg}(x,y) * h_1(x,y), f_{\lg}(x,y) * h_2(x,y)) = (h, h_1, h_2) \tag{5}$$

In Eq. (5), h is the real part of the monogenic signal, and h_1 and h_2 are the two imagery parts. Based on the real and imagery parts, the original image $f(x,y)$ could be orthogonally decomposed into three components, local amplitude A, local orientation θ and local phase ϕ:

$$A = \sqrt{h^2 + h_1^2 + h_2^2} \tag{6}$$

$$\theta = a\tan(h_2/h_1), \quad \theta \in [0 \quad \pi] \tag{7}$$

$$\phi = -sign(h_1)a\tan 2\left(\sqrt{h_1^2 + h_2^2}/h\right), \quad \phi \in [0 \quad 2\pi] \tag{8}$$

where,

$$h = f(x,y) * F^{-1}(G(\omega)) = f_{\lg}(x,y) \tag{9}$$

$$h_d = F^{-1}\left(\left(\sqrt{-1}\cdot\omega_d/\sqrt{\omega_1^2 + \omega_2^2}\right)H\right), \quad d \in \{1,2\} \tag{10}$$