Pattern Recognition 46 (2013) 434–448

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

Object recognition using Gabor co-occurrence similarity

Jian Zou [a,b,*], Chuan-Cai Liu [b], Yue Zhang [a,b], Gui-Fu Lu [a,b]

[a] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b] School of Mathematics and Physics, Anhui Polytechnic University, Wuhu 241000, China

ARTICLE INFO

Article history:
Received 14 September 2011
Received in revised form
18 May 2012
Accepted 26 June 2012
Available online 4 July 2012

Key words:
Object recognition
Gabor magnitude
Co-occurrence matrix
Multinomial manifold

ABSTRACT

We present an object recognition approach using co-occurrence similarities of Gabor magnitude textures in this paper. A novel image descriptor, multichannel Gabor magnitude co-occurrence matrices (MGMCMs), is designed to characterize Gabor textures for object representation and similarity matching. The descriptor is a generalization of multichannel color co-occurrence matrices (MCMs), which focus on using robust and discriminative magnitude textures in filtered images. Our approach starts from Gabor wavelet transformation of each object image. An exploratory learning algorithm is proposed for learning channel-adaptive magnitude truncation parameters and level parameters. This allows us to design the magnitude quantization that can reduce overall biased and peaked levels of resulting feature distributions in each channel, to avoid over-sparse co-occurrence distributions on average. The direction-based grouping is adopted for computational complexity reduction of MGMCMs extraction under a specific neighborhood mode on the grouped rescaled magnitude images of per object image. When each MGMCM is treated as a probability distribution lying on a multinomial manifold, we represent per object image as a point on a product multinomial manifold. Using multinomial geometry and metric extension technique, we construct the *p*-order Minkowski co-occurrence information distance for similarity matching between the albums of Gabor magnitude textures. The feasibility and effectiveness of the approach is validated by the experimental results on the Yale and FERET face databases, PolyU palmprint database, COIL-20 object database and Zurich buildings database.

1. Introduction

Object recognition is studying and classifying an unknown object into one of the set of predefined classes. Oftentimes, it is assumed that the object being observed has been detected or there is a single object in the image. Object recognition is currently one of the most actively researched areas of computer vision, image processing and analysis. The main challenge in object recognition arises from the varying factors, such as shape, scaling, rotation, distortion, illumination and poses etc., and a successful recognition system should be robust to such changes. According to the way image data is represented, objects can be described by different cues thereby deriving model-based, shape-based, and appearance-based recognition methods in mainstream. Based on the applied features, these methods can be sub-divided into two main classes, i.e., local methods and global methods [1,2].

Texture describes visual information that is related to local spatial variation of the pixel intensities (or local filter responses) in image (or filtered image) subregions. It is reasonable to assume that the intensity (or filter response) variation of different objects is different. Therefore, one can help object recognition by means of characterizing texture [3,4]. Texture analysis using 2D Gabor filters falls into the category of frequency-based approaches, which are based on the premise that texture is an image pattern containing a repetitive structure that can be effectively characterized in a frequency domain. The Gabor filters have been found to yield local responses robust to many variations in the imaging conditions, including translations, deformations, and background changes [7,8]. Gabor features constructed from post-processed Gabor filter responses have been successfully used in various important computer vision tasks. The local multivariate and global high-dimensional feature forms bring lots of challenges to common learning or recognition systems, when taking the complete set of Gabor filtered images of each associated image as a whole. This situation causes local methods to be very popular though many of them are either time-consuming or manual annotation-involved, while global methods need to be supported by certain compressing or sampling technique [10,11]. The redundant information existing among different Gabor features makes magnitude features to be widely used due to the easy-to-use univariate positive value space and discriminative texture information [7–11]. Psychological research

* Corresponding author at: School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China.
Tel.: +86 18605695628; fax: +86 84315751.
E-mail address: qingwenzou@yahoo.com.cn (J. Zou).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.06.018

on human texture perception finds that two homogeneous textures are discriminative if they produce dissimilar marginal distributions [12]. This fact indicates the feasibility to recognize object by matching multichannel magnitude distributions similarities under specific texture representation. The existing problems, however, obstruct the application prospect in this way. On the parameterized consideration, Gamma models used to be adopted for modeling Gabor magnitude textures. But, it seems impracticable to use accurate intrinsic metric for model-based match, because of complex geometry structure of Gamma family and high computational complexity in computing information distance on Gamma manifold [13,14]. Similar problem also appears in the case where each channel magnitude set is modeled as a sampling realization of a probability density function lying on a nonparametric statistical manifold [15]. On the non-parameterized consideration, the multi-dimensional histograms are widely applied to multichannel data or feature sets. But, the loss of spatial information, over-sparseness and high computational cost degrade the performance of the descriptors in similarity matching [9].

Human texture discrimination in terms of texture statistical properties has been investigated in past few decades. A basic conclusion is that the textures in gray-level images are discriminated spontaneously only if they differ in second order moments. Classical co-occurrence matrix methods aim to characterize image texture by use of spatial information of original pixel features [15,16,40]. The most popular second-order statistical features for texture analysis are derived from gray level co-occurrence matrix (GLCM), originally addressed by Haralick et al. [15]. But it is only suitable for single band image. Rosenfeld et al. [16] generalized GLCM to multiband co-occurrence matrix (MBCM), the most appropriate descriptor for multispectral texture in multiband image. Unfortunately, MBCM is impracticable for image with dozens of spectral bands, because the computational complexity of MBCM extraction and dissimilarity quantization exponentially grows with the number of spectral bands and the number of quantization levels of spectral components [16,17]. Palm [18] firstly simplified joint co-occurrence probability model of color texture by introducing integrative color co-occurrence matrices (CCMs) and treating them as the generalization of GLCM. Similar techniques were exploited by Arvis et al. and Muselet and Macaire [19,20]. According to the mode of considered spectral band pairs, the CCMs can be categorized into single-channel CCMs (SCMs) and multi-channel CCMs (MCMs). Most of above approaches aim to use the vectorized normalized Haralick features. But, the inherent shortages hinder them from getting the better performance in some more challenging tasks. Then the actual problem is the ignorance of the best combination of existing classical features and the incomplete utilization of the statistical information of associated co-occurrence matrices. Moreover, the co-occurrence information extracted from raw pixel features tends to be impacted by some imaging condition changes. Therefore, the traditional co-occurrence matrix methods tend to underperform in the recognition tasks where these changes are involved.

Comparative studies showed that when applied to texture classification, the approach based on multidimensional co-occurrence matrices has certain advantages to wavelet-based opponents [21,23,40]. This inspires us to develop a more competitive co-occurrence matrix method using robust Gabor magnitude features rather than raw pixel features for recognition tasks. Because the magnitude distributions of filtered images are biased, original magnitude quantization tends to obtain over-sparse co-occurrence distribution that is not conducive to significant similarity matching [20]. For this reason, we rescale each magnitude image by a transformation determined by channel-adaptive truncation and level parameters. Each pair of channel-adaptive parameters is the optimal solution of proposed learning algorithm that minimizes the cumulative sum of absolute kurtosis and skewness of a collection of channel magnitude sets. Taking the complexity into account, we choose the direction-based grouping scheme to limit multichannel co-occurrence information extraction in the grouped rescaled magnitude images. Three interpolation-free neighborhood modes are proposed for possible manner to count co-occurring feature pairs in a local rhombic region of a center pixel on a pair of rescaled magnitude images. Thereby the model of MGMCMs is defined as a generalization of MCMs, and a novel object representation is introduced in the pattern of MGMCMs-characterizing Gabor magnitude textures. We treat each MGMCM as a probability distribution lying on a multinomial manifold (i.e. simplex) [21]. Thus, an object image can be further represented as a point on a product multinomial manifold. For algorithmic implementation, the geodesic distance metric of each factor multinomial manifold is extended to its closure. The p-order Minkowski co-occurrence information metric is then built by the extended factor distances for match purpose.

The main contributions of our method include: (1) a novel image descriptor MGMCMs is designed to characterize multichannel magnitude texture for object representation and similarity match; (2) an exploratory learning algorithm is designed for learning channel-adaptive parameters for robust and discriminative co-occurrence information extraction; (3) with multinomial geometry and metric extension technique, the p-order Minkowski co-occurrence information metric is built for magnitude co-occurrence similarity matching using complete information of MGMCMs.

The rest of this paper is organized as follows. Section 2 gives a brief review of Gabor magnitude and the models of CCMs. Section 3 details Gabor magnitude co-occurrence probability model. Experiments and analysis are presented in Section 4, and Section 5 gives some conclusions and future work.

2. A brief review of Gabor magnitude and CCMs

2.1. Gabor filter and Gabor magnitude

In the spatial domain, a 2D Gabor filter (kernel, wavelet or function) is a Gaussian kernel function modulated by a sinusoidal plane wave [9,10]:

$$G_{u,v}(z;u,v,\sigma,k_{u,v}) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\|k_{u,v}\|^2\|z\|^2/2\sigma^2}\left[e^{-ik_{u,v}z} - e^{-\sigma^2/2}\right], \tag{1}$$

where $z=(x,y)$ represents a 2-dimensional input point. The parameters u and v define the orientation and scale of the Gabor kernel. $\|\cdot\|$ denotes the norm operator, and σ refers to the standard deviation of the Gaussian window in the kernel. An attractive mathematical property of Gabor filters is that they minimize the joint uncertainty in space and frequency [5,28]. In general, the filter bank associated with U directions and V scales are employed to characterize image brightness appearance. The wave vector $k_{u,v}$ can be defined as $k_{u,v} = k_v e^{i\phi_u}$, where $k_v = k_{max}/f^v$, $\phi_u = \pi u/U$ and k_{max} is the maximum frequency, while f^v is the spatial frequency between kernels in the frequency domain. The filter bank is denoted as

$$\{G_{u,v}(\cdot) : (u,v) \in \{0,1,\ldots,U-1\} \times \{0,1,\ldots,V-1\}\}, \tag{2}$$

The acquisition of Gabor magnitude features of an input image involves convoluting the image with each filter of a given Gabor filter bank under the fixed convolution mask in the size of $a\times b$. The magnitude at the point $z_0=(x_0,y_0)$ obtained by convoluting