Pattern Recognition Letters 51 (2015) 57–62

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Scale invariant texture representation based on frequency decomposition and gradient orientation ☆

Jun Zhang, Jimin Liang*, Chunhui Zhang, Heng Zhao

School of Life Science and Technology, Xidian University, Xian, Shaanxi 710071, China

ARTICLE INFO

Article history:
Received 18 January 2014
Available online 9 October 2014

Keywords:
Scale invariance
Gradient orientation
Frequency decomposition
Wedge filters
Texture classification

ABSTRACT

This paper proposes an effective scale invariant texture representation based on frequency decomposition and gradient orientation. First, the image intensities are decomposed into different orientations by using wedge filters in the frequency domain, and the N-nary coding method is adopted for the vector quantization. Second, the scale invariant gradient orientation is generated by selecting the most stable value of the gradient orientation with different Gaussian scales. Finally, the 2D joint distribution of the two types of local descriptors is used as the representation. The performance was evaluated on texture classification using a nearest neighbor classifier. Simple but not ordinary, our method achieves state of the art classification performance on the KTH-TIPS dataset under the traditional experimental design. Moreover, the main experiments were conducted on the KTH-TIPS and KTH-TIPS2-b datasets with the experimental designs of scale invariance validation. Compared with the methods of basic image features (BIFs) and local energy pattern (LEP), the proposed representation achieves superior performance with much lower dimension of representation.

1. Introduction

Texture representation with scale invariance is a challenging and long lasting problem for texture analysis. The scale invariance is an important property for classifying the textures imaged from different distances or with different resolutions. The approaches based on keypoint detection have achieved a very impressive performance on scale analysis [1–4]. In this paper, we focus on another branch of approaches based on the statistical properties of local structure. Of the statistical texture representations, in order to achieve the scale invariance, the methods could be summarized into three categories.

First, the scale level of the texture is estimated and the specific scale is selected to generate representation. For example, Li et al. adopted the Laplacian of Gaussian filters to select an optimal scale and construct scale invariant local binary pattern (LBP) descriptor [5].

Second, the multi-scale or multi-resolution representation is applied and the dissimilarity between the training model and testing sample is calculated by shifting along different scales or resolutions [6,7]. Crosier and Griffin [6] used the basic image features (BIFs) with scale shifting and Zhang et al. [7] used local energy pattern (LEP) with resolution shifting distance measurements to achieve the scale invariance respectively.

Third, the scale invariant measures are calculated to represent the texture [8,9]. Some existing methods used the scale invariant transforms such as the Fourier–Mellin transform, log polar Fourier transform and wavelet transform to calculate the scale invariant measures [10].

For the first category, the representation relies too much on the estimation step. It might meet the situation that if the estimation is not accurate, the representation will vary widely. For the second category, the representation itself is still not scale invariant. Moreover, the multi-scale or multi-resolution leads to a very high dimension of the representation and needs more computation for the calculation of dissimilarity with the shifting method. For the third category, the scale invariant measurements or local descriptors are specially designed for extracting scale-independent values. The original representation is scale invariant and it does not need any other steps such as scale estimation or scale shifting. However, it is not easy for a few scale invariant measurements or a pure scale invariant feature to describe the local structure very distinctively, and it often leads to a not very competitive classification performance compared with the other two categories.

In this paper, we propose a simple, yet very effective scale invariant local descriptor for texture classification. Our method aims to extract scale invariant measurements belonging to the third category above. We propose to use the 2D joint distribution of two new scale invariant descriptors to represent the texture. Specifically, the first descriptor is based on the gray intensity decomposition in the frequency

☆ This paper has been recommended for acceptance by Dr. G. Moser.

* Corresponding author.
E-mail address: jimleung@mail.xidian.edu.cn, jiminliang@gmail.com (J. Liang).

http://dx.doi.org/10.1016/j.patrec.2014.08.002

domain and the second descriptor is based on the scale invariant gradient orientation. Since the two proposed descriptors achieve scale invariance from different information, the two descriptors are very complementary for the representation. Consequently, the proposed method achieves state of the art classification performance on the KTH-TIPs dataset under a traditional experimental design. Moreover, superior classification performance is achieved in the experiments for the validation of scale invariance compared with the two representative scale invariant approaches of BIFs [6] and LEP [7].

2. Algorithm

The proposed algorithm combines two different sorts of scale invariant descriptors. The first local descriptor is based on the frequency decomposition by using scale invariant wedge filters, and the second one is based on the scale invariant gradient orientation. In order to obtain the 2D joint distribution, the two types of local descriptors need to be quantized respectively. The first descriptor is a vector and we adopt the method of N-nary coding [7] for the vector quantization. The second descriptor is a scalar and it is easy to be quantized into different states.

2.1. Frequency decomposition

2.1.1. Generation of the local descriptor

As Varma and Zisserman pointed out, the distribution of gray intensities cannot distinguish textures with different scales. Thus, they thought that the distribution of derivatives or compact patches of intensities could avoid such a drawback [11]. However, the 'negative property' is positive for our method. Our objective is to find the scale invariant descriptor. Obviously, the distribution of gray intensities is a scale invariant representation. Unfortunately, using the distribution of intensities directly could not obtain satisfactory performance due to the lack of a good description of the local structure and poor distinguishing ability. Therefore, we seek a kind of orientation decomposition which could describe the local structure. The wedge filters are well-known scale invariant filters and have been widely used for image processing [12,13]. The wedge filters are shown in Fig. 1.

By using the wedge filters, the image could be decomposed into several oriented frequency subimages. Then, the oriented frequency components are concatenated to describe the local structure. However, the local descriptor always performs badly in classification due to the illumination affecting the representation too much. In order to remove the effect of brightness changes and the direct current (DC) component, a Gaussian high-pass filter is used during the frequency filtering. Then, the contrast effect is weakened by the normalization of Weber's law [14]. To our knowledge, although the wedge filters are well-known and are old scale invariant filters, no one has used the filters to generate the local descriptor like we present here. The specific procedure is described as follows:

- First, the Fourier transform of image $I(x, y)$ is calculated and denoted as $\mathscr{F}(I(x, y))$.
- Second, suppose the wedge filters with different angles are $W_{\phi_j}(u, v), j \in \{1, \ldots, P\}$, where P is the number of filters with different orientations. The Gaussian high-pass filter is denoted as $H_g(u, v)$. Thus, the filter responses are

$$R(x, y, \phi_j) = \mathscr{F}^{-1}(\mathscr{F}(I(x, y)) \cdot H_g(u, v) \cdot W_{\phi_j}(u, v)). \tag{1}$$

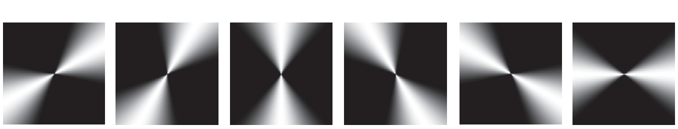

Fig. 1. Wedge filters for frequency decomposition.

The local descriptor at (x, y) is defined as

$$\mathbf{v}(x, y) = \{R(x, y, \phi_1), R(x, y, \phi_2), \ldots, R(x, y, \phi_P)\}. \tag{2}$$

- Finally, the local descriptor is normalized by Weber's law as [14]:

$$\mathbf{v}(x, y) \leftarrow \mathbf{v}(x, y)[\log(1 + \|\mathbf{v}(x, y)\|_2/0.03)]/\|\mathbf{v}(x, y)\|_2. \tag{3}$$

2.1.2. Vector quantization

The elements of the local descriptor are not in the range of $[0, 1]$ after the normalization of Weber's law. In order to use N-nary coding [7], the range should be transformed into $[0, 1]$. We use the tangent function to transform the range $[-\infty, +\infty]$ into $[-\pi/2, \pi/2]$, then linearly stretch the range into $[0, 1]$.

According to [7], all of the elements from local descriptors (regardless of category) are aggregated, and the probability distribution function (PDF) is obtained from the normalized histogram of the elements. The $N - 1$ quantization thresholds are learned according to following equation:

$$T = \left\{f^{-1}\left(\frac{1}{N}\right), f^{-1}\left(\frac{2}{N}\right), \ldots, f^{-1}\left(\frac{N-1}{N}\right)\right\}, \tag{4}$$

in which,

$$f(r) = \int_0^r p(\omega)\, d\omega, \tag{5}$$

where r denotes the possible value for the element of local descriptor and $p(\cdot)$ is the PDF of r. The elements of each local descriptor could be quantized to N states by T in the range of $[0, 1]$. Here, N states are denoted as $\{0, 1, \ldots, N - 1\}$. Thus, the quantized vector is noted as $\mathbf{v}'(x, y) = \{R'(x, y, \phi_1), R'(x, y, \phi_2), \ldots, R'(x, y, \phi_P)\}$. Thus, by N-nary coding, the label of the local descriptor is defined as:

$$\Phi(x, y) = \sum_{p=0}^{P-1} R'(x, y, \phi_p) N^p. \tag{6}$$

Therefore, $\Phi(x, y) \in \{0, 1, \ldots, N^P - 1\}$.

2.2. Scale invariant gradient orientation

Only using the distribution of the local descriptor based on the frequency decomposition is feasible, but it cannot achieve very high classification performance. Thus, we want to find another local descriptor with scale invariance based on different information. In this way, the two local descriptors should be complementary.

As we know, the gradient orientation has the property of scale invariance. For a given image $I(x, y)$, the orientation of the gradient is often defined as $arctan(I_x/I_y)$, where I_x and I_y are the first order derivatives with regards to x and y respectively. In order to preserve the signs of the derivatives, the orientation of gradient is often developed as $\theta = arctan2(I_x/I_y) + \pi$ [15], where

$$arctan\,2\left(\frac{I_x}{I_y}\right) = \begin{cases} arctan\left(\frac{I_x}{I_y}\right), & I_x > 0,\ I_y > 0 \\ arctan\left(\frac{I_x}{I_y}\right) + \pi, & I_x < 0,\ I_y > 0 \\ arctan\left(\frac{I_x}{I_y}\right) - \pi, & I_x < 0,\ I_y < 0 \\ arctan\left(\frac{I_x}{I_y}\right), & I_x > 0,\ I_y < 0. \end{cases} \tag{7}$$

Consequently, $0 \leq \theta \leq 2\pi$. For a given scale parameter α, assume the scaled image $I^1(x, y)$ of the original image I is $I^1(x, y) = I(x/\alpha, y/\alpha)$. Moreover, assume the coordinates are continuous. Hence, the first