Image and Vision Computing 32 (2014) 424–436

Contents lists available at ScienceDirect

Image and Vision Computing

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

Improving texture categorization with biologically-inspired filtering ☆

Ngoc-Son Vu [a,*], Thanh Phuong Nguyen [b], Christophe Garcia [c]

[a] ETIS — ENSEA, UCP, CNRS, Cergy, France
[b] U2IS / ENSTA - Paristech, France
[c] Université de Lyon, CNRS, INSA-Lyon, LIRIS, France

ARTICLE INFO

Article history:
Received 14 October 2013
Received in revised form 10 February 2014
Accepted 2 April 2014
Available online 13 April 2014

Keywords:
Texture classification
Retina filtering
DoG
Rotation invariant preprocessing
Completed LBP
LBC
WLD
SIFT

ABSTRACT

Within the domain of texture classification, a lot of effort has been spent on local descriptors, leading to many powerful algorithms. However, preprocessing techniques have received much less attention despite their important potential for improving the overall classification performance. We address this question by proposing a novel, simple, yet very powerful biologically-inspired filtering (BF) which simulates the performance of human retina. In the proposed approach, given a texture image, after applying a difference of Gaussian (DoG) filter to detect the edges, we first split the filtered image into two maps alongside the sides of its edges. The feature extraction step is then carried out on the two maps instead of the input image. Our algorithm has several advantages such as simplicity, robustness to illumination and noise, and discriminative power. Experimental results on three large texture databases show that with an extremely low computational cost, the proposed method improves significantly the performance of many texture classification systems, notably in noisy environments.

The source codes of the proposed algorithm can be downloaded from https://sites.google.com/site/nsonvu/code.

1. Introduction

Texture classification is a fundamental issue in computer vision and image processing, playing a significant role in many applications such as medical image analysis, remote sensing, object recognition, document analysis, environment modeling, content-based image retrieval and many more. As the demand of such applications increases, texture classification has received considerable attention over the last decades and numerous novel methods have been proposed [1–14].

The texture classification problem is *typically divided into the two subproblems of representation and classification* [15], and to improve the overall quality of texture classification, researchers often focus on improving one of (or both) those steps. It is generally agreed that texture features play a very important role, and the last decade has seen numerous powerful descriptors being proposed such as modified SIFT (scale invariant feature transform), intensity domain SPIN images [3], MR8 [6], the rotation invariant basic image features (BIF) [10], (sorted) random projections over small patches [14], local binary pattern (LBP) [2] and its variants [7,8,13,16]. Also, different similarity measures such as χ^2 statistic [2,6], Bhattacharyya distance [10], and Earth Mover's Distance [3] are often used in conjunction with nearest neighbor classifiers [2] or non-linear (kernel-based) support vector machines (SVMs) [17]. Undoubtedly, an efficient preprocessing which enhances the robustness and discriminative power of texture features is an important factor towards enhancing such texture classification systems. However, to the best of our knowledge, there does not exist any "sufficiently efficient" preprocessing methods which can significantly improve texture features. In other words, preprocessing seems to be neglected for texture classification. And this paper aims at catching up on this topic by proposing a novel efficient preprocessing technique for improving texture classification performance.

2. Related work

Most of earlier work on texture analysis focused on the development of filter banks and on characterizing the statistical distributions of their responses. Davis [18] exploited polarograms and generalized co-occurrence matrices to obtain rotation invariant statistical features. Duvernoy [19] proposed Fourier descriptors to extract texture feature on the spectrum domain. Cohen et al. [20] characterized texture as Gaussian Markov random fields and used the maximum likelihood to estimate rotation angles. Presented more recently for texture

☆ This paper has been recommended for acceptance by Sinisa Todorovic.

* Corresponding author.
E-mail address: son.vu@ensea.fr (N.-S. Vu).

http://dx.doi.org/10.1016/j.imavis.2014.04.006
0262-8856/

description are Gabor filters [21], the filter bank of Leung and Lalik [22], the MR8 [4], the filter bank of Crosier and Griffin [10], the morphological approaches of Hanbury et al. [23] and Aptoula and Lefevre [24,25], and so on.

Although many efforts have been carried out along this direction, the supremacy of filter bank-based descriptors for texture analysis has been challenged by several authors [2,6] who showed that it is possible to discriminate between textures using the intensities or differences of pixel within small-scale neighborhoods. They demonstrated that despite the global structure of the textures, very good discrimination could be achieved by exploiting the distributions of such pixel neighborhoods. Two particularly important works along these lines are the VZ-Joint classifier [6] and the LBP method [2]. The simple, elegant and efficient local texture descriptor LBP may be the preferable choice over VZ-Joint classifier since LBP uses a pre-defined texton dictionary and does not need to use nearest neighbor to obtain the texton labels, a time consuming step. Due to its impressive computational efficiency and good texture discriminative property, the LBP descriptor [2] has gained considerable attention since its publication, and has already been used in many other applications, including image retrieval, dynamic texture recognition, face image analysis, motion analysis, and outdoor scene analysis [26–28]. Despite its great success, the conventional LBP operator comes with disadvantages and limitations, such as small spatial support region, loss of local textural information, and rotation and noise sensitivities. To overcome those, a lot of effort has been done. To recover from the loss of information, local image contrast was introduced by Ojala et al. [2] as a complementary measure, and better performance has been reported. By a completed LBP model, Guo et al. [8] included both the magnitudes of local differences and the pixel intensity itself, and claimed better performance. In terms of locality, [7] proposed to extract global features from the Gabor filter responses as a complementary descriptor. Dominant LBP (DLBP) also presented in [7] relies on dominant patterns which were experimentally chosen from all rotation invariant patterns. Regarding noise robustness, Ojala et al. [2] introduced the concept of uniform and rotation invariant patterns (LBP^{riu2}), while Tan and Triggs [29] proposed local ternary patterns (LTP). Liu et al. [16] have recently generalized LBP with two different and complementary types of features which are extracted from local patches, based on pixel intensities and differences. In [13], a LBP variant, the Local Binary Count (LBC), is proposed, in which a pixel is encoded by the number of neighbors whose intensities are larger than that of the considered pixel. Also, presented in [28] are several LBP variants for image and video description, that is, the LBP histogram Fourier (LBP-HF) features, and the LBPs from three orthogonal planes (LBP-TOP) features. In [9], Chen et al. proposed WLD (Weber Local Descriptor) method based on the fact that human perception of a pattern depends not only on the change of a stimulus but also on the original intensity of the stimulus.

An alternative method to improve the strength of texture descriptor is to perform efficient preprocessing. For example, in face recognition, Vu and Caplier [30,31] applied the LBP operators upon three edge distribution maps across different directions, and reported state-of-the-art performance. However, to the best of our knowledge, with regard to feature extraction in texture classification, no such efficient preprocessing method exists (in [7], the DLBP features and Gabor filters are extracted separately). This is the motivation for the algorithm presented in this paper.

Neuroscience has made lots of progress in understanding the visual system and how images are transmitted to the brain. It is believed that the difference of Gaussians (DoG) filter simulates how the human retina processes the images observed and extracts theirs details. We propose to somehow mimic the same strategy to generate richer and more robust information from the image before carrying out the feature extraction step.

The rest of the paper is structured as follows. Section 3 details the proposed approach. Experimental results are presented in Section 4 and conclusions are finally drawn in Section 5.

3. Proposed method

The main objective of the proposed method is to enhance robustness and discriminative power of texture classification systems at the level of preprocessing. We propose to use a simple yet efficient DoG filter which simulates the performance of human retina. Given an input texture image, we first use a DoG filter to detect its edges and then split the filtered image into two "maps" alongside two sides of the detected edges (the term "edges" used here refer to the positions where there are changes in intensity). Feature extraction algorithms, e.g., the LBP encoding method, are then carried out on those resultant maps to obtain the final texture representation.

This section first briefly describes the human retina, in particular the bipolar cells by which our algorithm is inspired. We then detail the proposed method and discuss its properties.

3.1. Model of retinal processing

The retina lies at the back of the eye. Basically, it is made of three layers: the photoreceptors layer with cones and rods; the outer plexiform layer (OPL) with horizontal, bipolar and amacrine cells; and the inner plexiform layer (IPL) with ganglion cells.

Photoreceptors: rods have the ability to see at night, under conditions of very low illumination whereas cones have the ability to deal with bright signals. Photoreceptor layer plays therefore the role of a light adaptation filter.

Outer plexiform layer (*OPL*): the photoreceptor performs a low pass filter. Horizontal cells perform a second low pass filter. In OPL, bipolar cells calculate the difference between photoreceptor and horizontal cell responses. Then, bipolar cells act as a band-pass filter: they remove high frequency noise and low frequency illumination. Typically, to model the processes of OPL, two Gaussian low pass filters corresponding to the effects of photoreceptors and horizontal cells are used. Thus, bipolar cells act like a Difference of Gaussians (DoG) filter.

Inner plexiform layer (*IPL*): IPL works similarly to OPL but it performs on the temporal information rather than on the spatial one as in OPL.

In the literature, different algorithms inspired by the human retina have been proposed for different applications [32,33]. The two first layers of the retina, photoreceptor and OPL, have been modeled and successfully used for face recognition under difficult lighting conditions [32]. In [33], Benoit and Caplier modeled all the three layers for moving contours enhancement. Our algorithm is inspired by the performance of the bipolar cells.

3.2. Details of the proposed method

In fact, there are two types of bipolar cells, called ON and OFF. The ON bipolar cells take into account the difference of photoreceptor and horizontal cell responses, whereas the OFF bipolar cells compute the difference of horizontal and photoreceptor cells. More precisely, if we apply a DoG filter on an image for simulating the bipolar cells, a "map" with positive and negative values will be obtained. Within this resultant map, the positive values and the absolute of the negatives values correspond respectively to the responses of the ON and OFF bipolar cells.

Also, the DoG filter is often used to approximate a LoG filter (Laplacian of Gaussian) due to its low computational cost [34]. It calculates the second spatial derivative of an image. In areas where the image has a constant intensity, the filter response will be zero. Wherever an intensity change occurs, the filter will give a positive response on the darker side and a negative response on the lighter side. At a reasonably sharp edge between two regions of uniform but different intensities, the filter response will be: (1) zero at a