Pattern Recognition Letters 51 (2015) 16–22

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Compact color–texture description for texture classification☆

Fahad Shahbaz Khan[a,*], Rao Muhammad Anwer[b], Joost van de Weijer[c], Michael Felsberg[a], Jorma Laaksonen[b]

[a] Computer Vision Laboratory, Linköping University, Sweden
[b] Department of Information and Computer Science, Aalto University School of Science, Finland
[c] Computer Vision Center, CS Dept. Universitat Autonoma de Barcelona, Spain

ARTICLE INFO

Article history:
Received 27 April 2014
Available online 12 August 2014

Keywords:
Texture features
Color features
Texture classification
Image classification

ABSTRACT

Describing textures is a challenging problem in computer vision and pattern recognition. The classification problem involves assigning a category label to the texture class it belongs to. Several factors such as variations in scale, illumination and viewpoint make the problem of texture description extremely challenging. A variety of histogram based texture representations exists in literature. However, combining multiple texture descriptors and assessing their complementarity is still an open research problem. In this paper, we first show that combining multiple local texture descriptors significantly improves the recognition performance compared to using a single best method alone. This gain in performance is achieved at the cost of high-dimensional final image representation. To counter this problem, we propose to use an information-theoretic compression technique to obtain a compact texture description without any significant loss in accuracy. In addition, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the problem of texture classification. Experiments are performed on four challenging texture datasets namely, KTH-TIPS-2a, KTH-TIPS-2b, FMD and Texture-10. The experiments clearly demonstrate that our proposed compact multi-texture approach outperforms the single best texture method alone. In all cases, discriminative color names outperforms other color features for texture classification. Finally, we show that combining discriminative color names with compact texture representation outperforms state-of-the-art methods by 7.8%, 4.3% and 5.0% on KTH-TIPS-2a, KTH-TIPS-2b and Texture-10 datasets respectively.

1. Introduction

Classifying textures is a difficult problem in computer vision and pattern recognition. The task is to associate a class label to its respective texture category. In recent years, a variety of texture description approaches have been proposed [30,10,20,5,9,52,12,41]. These approaches can be divided into two categories, namely sparse and dense representations. The sparse representation works by detecting feature points either based on interest point or dense sampling strategy. Feature description is then performed on these sampling points [20,49]. The second strategy, dense representations, involves extracting local features for each pixel in an image [30,10,5]. In this paper, we investigate the problem of texture classification using dense local texture representations.

A variety of texture description approaches exist in literature [30,10,20,5,9,52]. One of the most successful approaches is that of Local Binary Patterns (LBP) [30] based image representations. Other than texture classification, LBP have been successfully employed to solve other vision problems as well, such as object detection [48], face recognition [1] and pedestrian detection [42]. LBP describes the neighbourhood of a pixel by its binary derivatives which are used to form a short code to describe the pixel neighbourhood. A variety of LBP variants have been proposed [10,47,45]. Combining multiple texture features, such as variants of LBP features, is still an open research problem. The work of Guo et al. [9] proposes a learning framework to combine variants of LBP features for texture classification. Tan and Triggs [38] propose to combine Gabor wavelets and LBP features for the problem of face recognition. In this paper, we propose to use a heterogeneous feature set by combining multiple texture description methods.

Combining multiple texture description methods have an inherent problem of high-dimensional final image representations. Recently, Elfiky et al. [7] proposed to use a divisive information theoretic clustering (DITC) method [6] to counter the problem of high-dimensionality of bag-of-words based spatial pyramid

☆ This paper has been recommended for acceptance by Y. Liu.
* Corresponding author.

http://dx.doi.org/10.1016/j.patrec.2014.07.020
0167-8655/

representations. The DITC compression was shown to reduce the dimensionality of image representations without any significant loss in accuracy. Similar to the work of Elfiky et al. [7], we propose to use the DITC approach to compress the high-dimensional multi-texture representation. However, different to the work of Elfiky et al. [7], here we investigate compressing a multi-texture histogram to obtain a single heterogeneous texture representation.

Generally, state-of-the-art texture descriptors operate on grey level images thereby ignoring the color information. Color in combination with shape features has been shown to yield excellent results for object recognition [32,18,19], object detection [16] and action recognition [15]. Color description is a challenging problem due to significant variations in color caused by changes in illumination, shadows and highlights. Recent works have shown that an explicit color representation improves the performance for object recognition [18,19], object detection [16] and action recognition [15]. In this paper, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the task of texture classification.

Contributions: We first show that combining multiple texture description methods significantly improves the performance compared to using the single best texture method alone. We further propose to use information theoretic compression approach to compress high-dimensional multi-texture features into a compact heterogeneous texture representation. Finally, we provide a comprehensive evaluation of color features, popular in object recognition, for the task of texture classification. This paper extends our earlier work [17] for texture classification that only evaluated the contribution of color for texture recognition. Beyond the work in [17], we here investigate the problem of combining multiple local texture descriptors for robust texture description. We perform extensive experiments on four challenging texture datasets namely, KTH-TIPS-2a, KTH-TIPS-2b, FMD and Texture-10.

The results of our experiments clearly demonstrate that combining multi-texture descriptors significantly improves the performance compared to the single best method alone. We further show that multi-texture representations can be compressed efficiently without any significant loss in accuracy. Finally, our comprehensive evaluation of color features suggest that discriminative color names outperforms other color descriptors for texture recognition. By combining the best color descriptor with our compact heterogenous texture representation provides state-of-the-art results on three of the four texture datasets.

The paper is organized as follows. In Section 3 we investigate the problem of combining multiple texture descriptors. A comprehensive evaluation of pure color descriptors for texture description is provided in Section 4. In Section 5 we provide experimental results. Section 6 finishes with concluding remarks.

2. Related work

A variety of texture description approaches have been proposed in recent years [30,10,20,5,9,52,47,22]. Varma and Zisserman [41] propose a statistical approach for texture modeling using the joint probability distribution of filter responses. A multiresolution approach based on local binary patterns (LBP) is proposed by Ojala et al. [30] for gray-scale and rotation invariant texture classification. The LBP is one of the most successful approaches for texture classification with several variants existing in literature [10,47,45]. Chen et al. [5] propose a method based on Weber's law consisting of two components namely differential excitation and orientation. An image is represented by concatenating the two components in a single representation. ul Hussain and Triggs [12] introduce an approach that uses lookup-table based vector quantization for texture description. A set of low and mid-level perceptually inspired image features are proposed by Sharan et al. [35] for texture classification.

Combining multiple texture representations for robust classification [9,24,38,11] is an interesting problem. The work of Tan and Triggs [38] combines Gabor wavelets and LBP for the problem of face recognition. Ylioinas et al. [46] combine contrast information together with local binary patterns for improved gender classification. A combination of HOG, LBP and Gabor features is used by Li et al. [23] for gender classification. To counter the dimensionality of the proposed image representation, Partial Least Squares (PLS) is used to learn a low-dimensional representation. Hong et al. [11] propose a numerical variant of LBP which is efficient and rotation invariant. The method is combined with other cues by a covariance matrix. Guo et al. [9] propose a learning framework to fuse a variety of LBP variants such as conventional LBP, rotation invariant patterns, local patterns with anisotropic structure, completed local binary patterns and local ternary patterns. Similar to Guo et al. [9], we investigate the problem of combining multiple texture description approaches. However, instead of only combining LBP variants [9], we here investigate fusing multiple texture descriptors to obtain a single heterogeneous texture representation.

A variety of color description approaches have been proposed in the field of object and scene recognition [8,3,43,32,18,19]. Bosch et al. [3] propose to compute SIFT descriptors directly on HSV channels for image classification. A comprehensive evaluation of color descriptors is performed by Sande et al. [32]. It has been shown that using an explicit color descriptor significantly improves the performance for object recognition [18,19], object detection [16], texture recognition [17] and action recognition [15]. In this work, we perform a comprehensive evaluation of pure color descriptors, popular in object recognition, for the problem of texture classification.

3. Combining multiple texture descriptors

Here we present our framework of combining multiple texture features and obtaining a compact heterogeneous texture representation. We combine five texture descriptors namely, completed local binary patterns [10], WLD descriptor [5], binary Gabor pattern [51], local phase quantization descriptor [31] and binarized statistical features [14]. We start by providing a brief overview of the five texture descriptors used in this work.

Completed local binary patterns [10]: The completed local binary patterns (CLBP) extends the conventional LBP operator by incorporating local difference sign-magnitude transform information (LDSMT).[1] The LDSMT further consists of two components, namely the difference sign and difference magnitude encoded by a binary code. Likewise the conventional LBP, a region is also represented by its center pixel encoded by a binary code after global thresholding. The final image representation is obtained by concatenating the three binary code maps to form a single histogram.

WLD descriptor [5]: The WLD descriptor is inspired by Weber's Law and encodes both differential excitations and orientations at locations. The first component, differential excitation, captures the ratio between the intensity difference of a pixel with its neighbors and the intensity of the current pixel. The second component captures the gradient orientation of the current pixel.

Binary Gabor patterns [51]: The binary Gabor patterns (BGP) is a rotation invariant texture descriptor. Unlike MR8 filters [41], BGP uses pre-defined rotation invariant binary patterns and does not require a pre-training phase to learn a texton dictionary. Unlike LBP, where each sign is binary coded from the difference of two

[1] We experimented with different variants of LBP and found CLBP to provide superior performance.