Pattern Recognition 48 (2015) 447–457

Contents lists available at ScienceDirect

Pattern Recognition

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

Effective texture classification by texton encoding induced statistical features

Jin Xie [a,b], Lei Zhang [a,*], Jane You [a], Simon Shiu [a]

[a] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] Department of Electrical and computer Engineering, New York University in Abu Dhabi, Abu Dhabi, UAE

ARTICLE INFO

Article history:
Received 7 April 2014
Received in revised form
16 July 2014
Accepted 13 August 2014
Available online 24 August 2014

Keywords:
Texture classification
Texton learning
Texture feature extraction
Sparse representation

ABSTRACT

Effective and efficient texture feature extraction and classification is an important problem in image understanding and recognition. Recently, texton learning based texture classification approaches have been widely studied, where the textons are usually learned via K-means clustering or sparse coding methods. However, the K-means clustering is too coarse to characterize the complex feature space of textures, while sparse texton learning/encoding is time-consuming due to the l_0-norm or l_1-norm minimization. Moreover, these methods mostly compute the texton histogram as the statistical features for classification, which may not be effective enough. This paper presents an effective and efficient texton learning and encoding scheme for texture classification. First, a regularized least square based texton learning method is developed to learn the dictionary of textons class by class. Second, a fast two-step l_2-norm texton encoding method is proposed to code the input texture feature over the concatenated dictionary of all classes. Third, two types of histogram features are defined and computed from the texton encoding outputs: coding coefficients and coding residuals. Finally, the two histogram features are combined for classification via a nearest subspace classifier. Experimental results on the CUReT, KTH_TIPS and UIUC datasets demonstrated that the proposed method is very promising, especially when the number of available training samples is limited.

1. Introduction

Texture feature extraction and classification is an important problem in many image processing, computer vision and pattern recognition tasks, including medical imaging, remote sensing, material identification and image classification, etc. In the past decades, a variety of texture classification methods have been proposed [1–5, 7,8,10,12,13,11,29,32,30,31,33–36]. According to [28], these methods can be categorized based on the following five criteria. (1) Image sampling—The sampling can be either dense or sparse, depending on whether or not the image pixels are probed for feature extraction. (2) Feature type—including image filter responses and image patches. (3) Feature space partition—There are two strategies to partition the feature space: priori strategy and posteriori strategy. The difference between the two strategies lies in that the partitioning rule is defined independently of the features in the former, while the rule is learned from the features in the later. (4) Feature labelling—Each feature is labeled in the partitioned feature space via hard labeling or soft labeling. The hard labeling method assigns each feature to one single partition of the feature space, while the soft labeling method assigns each feature to multiple partitions. (5) Image representation—including histogram and signature, e.g., representation by the clustering center and the size of the center.

As a simple and efficient statistical descriptor, the local binary pattern (LBP) histogram [1] has been successfully used for rotation invariant texture classification. Some variants [2–4,29,32,34–36] of LBP have been proposed to improve the accuracy and robustness of LBP operator. It is pointed out in [28] that LBP and its variants can be viewed as bag of feature descriptors with dense image sampling, image patch feature, a priori partitioning of the feature space, hard label assignment and histogram based image representation.

Apart from LBP based methods, another popular approach to texture classification is texton learning based methods. The methods [11,12] share similar attributes: sparse image sampling; image filter response feature; a posteriori partitioning of the feature space; hard label assignment; signature based image representation. In [11], Lazebnik et al. detected the invariant regions in the texture image for texton learning. A combination of spin image (SPIN) and rotation invariant feature transform (RIFT) descriptors is then used to build the texture descriptor on these detected regions. Textons are learned from these descriptors by the K-means clustering method and the

* Corresponding author.
E-mail addresses: csjxie@gmail.com (J. Xie),
cslzhang@comp.polyu.edu.hk (L. Zhang), csyjia@comp.polyu.edu.hk (J. You),
csckshiu@comp.polyu.edu.hk (S. Shiu).

http://dx.doi.org/10.1016/j.patcog.2014.08.014

texton histogram is built as the statistical feature. Based on Lazebnik et al.'s work, Zhang et al. [12] combined three kinds of descriptors, SPIN, RIFT and scale invariant feature transform (SIFT) [19], to learn textons and employed a kernel SVM classifier for classification.

The texton learning based methods [7–10,20,13] can be categorized with the following attributes: dense image sampling; image filter response and patch features; a posterior partitioning of the feature space; hard label assignment; and histogram based image representation. In [7], for each class, the training images are registered and filtered to produce a set of 48-dimensional response vectors. Then the feature vectors are clustered by using the K-means clustering method and the texton histogram is used as the texture model for classification. Cula and Dana [8] extended Leung and Malik's algorithm [7] to 2D texton learning for un-calibrated and single texture image classification. Varma and Zisserman [22,9] modeled textures as their distributions over a set of textons, which are learned from their responses to the MR8 filter bank. In [23,10], textons are clustered directly from the patches of original images instead of their MR8 filter responses, which can lead to slightly better results. In order to deal with large scale and viewpoint changes of texture images, Varma and Garg [20] extracted the local fractal dimension and length feature from the MR8 filter responses to learn textons for classification. Liu et al. [13] proposed to couple random projection [27] with texton learning for texture classification. Sorted pixels and pixel differences in a patch are projected into a low dimensional space with random projection, and then textons are learned with the K-means clustering method in the compressed domain.

The methods [24,25,30,31,33,14] have the following attributes: dense image sampling; image filter response and patch features; a posterior partitioning of the feature space; soft label assignment; and histogram based image representation. In [24,25,30,33,14], the textons are learned by the sparse representation (or sparse coding) technique. The weights of the representation coefficients are used as the soft label assignment to form the histogram for texture classification.

In the above mentioned texton learning based methods, a dictionary of textons is usually learned by the K-means clustering algorithm or the sparse coding algorithm, and the distributions of textons or the histograms of texton encoding coefficients are computed as the statistical features for classification. The K-means clustering algorithm can be viewed as a special case of sparse coding, where the coding vector has only nonzero entry (i.e., 1) to indicate the cluster label. Due to the restrictive constraint, the learned ball-like clusters may not be able to characterize accurately the intricate feature space of texture images. For the sparse coding based texton learning and encoding algorithms, although some fast sparse coding methods [15,16] have been proposed, the l_0-norm or l_1-norm minimization still makes these methods time-consuming in order to obtain good accuracy. Moreover, most of these texton learning based texture classification methods [9,10,13,14] use the histogram of texton encoding coefficients for classification, and they will become less effective when the number of training samples is insufficient.

In this paper, a novel texton learning and encoding approach is developed for effective and efficient texture classification. The main motivations and contributions of this work are as follows. (1) Considering that the texture patterns from the same class of texture images are similar, a regularized least square based texton learning method is developed to learn the texton dictionary class by class, and the whole dictionary is concatenated by all sub-dictionaries. (2) A fast two-step l_2-norm texton encoding method is proposed to efficiently code the texture feature over the dictionary. (3) Two types of texton encoding induced histogram features are computed from the coding coefficients and coding residuals, respectively, and they are combined for texture classification via the nearest subspace classifier. The proposed scheme is verified on the representative and benchmark CUReT, KTH_TIPS and UIUC datasets, showing very promising performance, especially when the number of training samples is limited.

The rest of the paper is organized as follows. Section 2 presents in detail the proposed texton learning and encoding scheme for texture classification. Section 3 performs extensive experiments and Section 3 concludes the paper.

2. Texton encoding induced statistical features for texture classification

In this section, we describe in detail the proposed texture classification approach, whose framework is illustrated in Fig. 1. There are four major components in the proposed method: texton dictionary learning, texton encoding, feature description and classification.

2.1. Texton dictionary learning

A dictionary of textons is to be learned from the training texture images so that they can be used to represent the test image. Before learning the textons, the training texture images are first converted into grey level images and normalized to have zero mean and unit standard deviation. This process reduces the image variations caused by illumination changes. The texture features (e.g., the MR8 feature [9], the patch feature [10] and the SIFT feature [19]) are then extracted from the pre-processed texture images and normalized

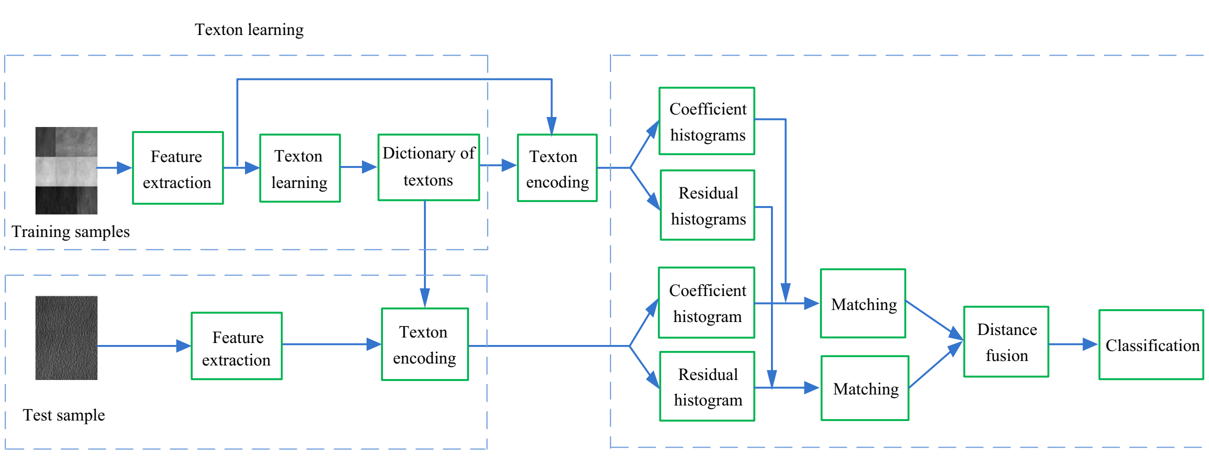

Fig. 1. Framework of the proposed texton learning and encoding based texture classification method.