



Sparse representation with multi-manifold analysis for texture classification from few training images[☆]



Xiangping Sun^{a,b,*}, Jin Wang^{a,b}, Mary F.H. She^b, Lingxue Kong^b

^a Center for Intelligent Systems Research, Deakin University, Waurn Ponds, Victoria 3216, Australia

^b Institute for Frontier Materials, Deakin University, Waurn Ponds, Victoria 3216, Australia

ARTICLE INFO

Article history:

Received 24 July 2013

Received in revised form 17 May 2014

Accepted 5 July 2014

Available online 11 July 2014

Keywords:

Texture classification

Sparse representation

Manifold learning

Multi-manifold analysis

Few training image

ABSTRACT

Texture classification is one of the most important tasks in computer vision field and it has been extensively investigated in the last several decades. Previous texture classification methods mainly used the template matching based methods such as Support Vector Machine and k-Nearest-Neighbour for classification. Given enough training images the state-of-the-art texture classification methods could achieve very high classification accuracies on some benchmark databases. However, when the number of training images is limited, which usually happens in real-world applications because of the high cost of obtaining labelled data, the classification accuracies of those state-of-the-art methods would deteriorate due to the overfitting effect. In this paper we aim to develop a novel framework that could correctly classify textural images with only a small number of training images. By taking into account the repetition and sparsity property of textures we propose a sparse representation based multi-manifold analysis framework for texture classification from few training images. A set of new training samples are generated from each training image by a scale and spatial pyramid, and then the training samples belonging to each class are modelled by a manifold based on sparse representation. We learn a dictionary of sparse representation and a projection matrix for each class and classify the test images based on the projected reconstruction errors. The framework provides a more compact model than the template matching based texture classification methods, and mitigates the overfitting effect. Experimental results show that the proposed method could achieve reasonably high generalization capability even with as few as 3 training images, and significantly outperforms the state-of-the-art texture classification approaches on three benchmark datasets.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Among all properties of an object, such as colour, shape, and motion, texture is one of the most significant characteristics which human vision and machine vision systems utilize in interpreting scenes and performing object identification. As an active field in computer vision, texture classification has been widely applied in many areas including medical image analysis [1–3], remote sensing [4], material characterization [5], and content-based image retrieval (CBIR) [6]. Due to the existence of the large intra-class variation (Fig. 1) and inter-class similarity in textural images robust texture classification is very challenging.

A general texture classification framework comprises two steps: the textural feature extraction, and classification. Previous works mainly focused on designing the robust feature extraction methods and using classifiers such as Support Vector Machine (SVM) and k-Nearest-Neighbour (kNN) for supervised classification [8,9]. Some representative

feature extraction methods include the Local Binary Patterns (LBP) [10], texon-based approaches [11,12,7,13], filter bank based methods [14–17] and bag-of-keypoints [18,8]. The texon-based approaches represent a textural image as a histogram of textons, where the textons are computed by utilizing a clustering algorithm such as k-means to cluster the local feature vectors extracted from the training images. There are numerous ways to extract the local feature vectors from images in the texon-based approaches. In [11,12,7], a bank of filters is utilized and all the filter responses on each pixel are concatenated as a local feature vector. By carefully selecting the filters, the filter bank based feature could be invariant to image translation and rotation. Another popular kind of local features is image patches [13], which utilize the raw pixel values in a fixed-size image patch around each pixel as a local feature vector, and it was proven to outperform the filter bank based feature on some datasets. Recently, Liu and Fieguth [19] applied random projection on the image patch to get a random feature based on the compressive sensing theorem, which not only reduced the dimension of the feature, but also achieved comparative results with the patch-based method [13]. The filter bank based methods filter an image using a bank of well designed filters, and extract features from each filtered image. All the features are then concatenated together as the image feature. Different from the texon-based approaches using

[☆] This paper has been recommended for acceptance by Matti Pietikainen

* Corresponding author. Tel.: +61 4 20747345.

E-mail addresses: xiangping.sun@gmail.com (X. Sun), jay.wangjin@gmail.com (J. Wang), mary.she@deakin.edu.au (M.F.H. She), lingxue.kong@deakin.edu.au (L. Kong).

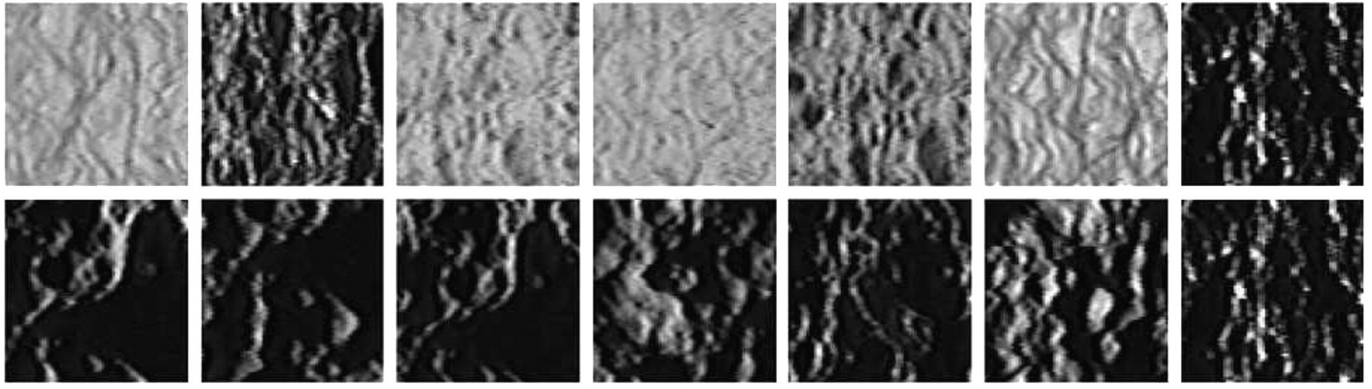


Fig. 1. Variations of the same texture in different conditions. In the first row, the images are captured in the same viewing angle but different illuminations, while in the second row the texture is imaged by changing the viewing angles with illumination fixed [7].

filter bank responses on each pixel as a local feature vector, all the features extracted by the filter bank based methods are in image-level but not pixel-level, thus they could not give a statistical characterization of textures. Therefore the filter bank based features are usually not as discriminative as those descriptors that could capture the local primitives of textures for texture classification, e.g., the texon-based approaches, though they could be designed to be invariant to many image variations such as translation, scale, and rotation. The bag-of-keypoint approaches utilize the region detectors such as Harris–Laplace (HL) detector to detect the interest areas of images, and then apply the image descriptors like Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) to extract local features from them [8]. The bag-of-keypoint features are invariant to image translation, rotation and affine transform. However, since the local features extracted from the detected keypoints of textural images are sparsely distributed, the bag-of-keypoint features are usually not discriminative enough. To tackle this problem, another method was proposed by extracting the local SIFT feature from densely sampled locations [20, 21], and it was shown in [20] that the dense feature worked better than bag-of-keypoints for scene classification. However, the dense SIFT sampling loses scale and affine invariance.

Texture classification could be regarded as a statistical learning problem, where one template is learnt from each training image (through feature extraction) and a classifier is learnt from all the templates of the training images. Ideally, if the feature extraction method is robust enough (not only invariant to different imaging conditions but also discriminative), the templates learnt from images of the same class will be close to each other and those learnt from images of different classes will be far away from each other, reaching small intra-class variation and large inter-class variation. Thus, a simple classifier with a few training images could easily distinguish test images from different classes. However, it is not always feasible to design a very discriminative feature extraction method, and also since the classifier is local in the input space (both SVM and kNN are local estimators), it requires a large number of training images to achieve a high generalization capability [22]. As shown in Section 3, the performance of the state-of-the-art texture classifiers will significantly decline when the number of training images per class decreases. Since collecting labelled image data is costly, it is common in practice that only a small number of images are available for training. Thus, it is critical to develop robust classification methods that only need a small amount of training images to achieve high generalization capability in the classification of test images (in fact not only for texture classification, most computer vision tasks have such a desire).

A few attempts have been made to solve the problem of classifying textures from a small number of training images. For instance, Drbohlav and Chantler [23] brought out a method to classify textural images captured under different illuminations from a single training image

per class. They filtered an image with a directional derivative operator to model the textural appearance under a specific illumination direction, and then utilized a filter bank to compute the image features. To compare two images under unknown illumination directions, a set of feature vectors were calculated for a complete set of illumination directions for each image. The distance between the closest pair of feature vectors of the two images is adopted as the distance between them. Targhi et al. [24] developed an approach to classify textures under unknown lighting conditions from a small number of training images by generating additional training data using a photometric stereo. However, these works only considered single variation of textures, i.e., illumination change, which were not applicable to the real world texture classification where textures are usually subject to multiple imaging condition variations, as illustrated in Fig. 1.

In this paper, we aim to develop a novel framework that only needs a few training images to classify textures with various image variations such as translation, rotation, scale, illumination and view-point change. The following three major aspects are considered in the proposed texture classification framework:

1. Since most textures are uniform and repetitive on pattern distribution, we could divide a textural image into many subimages, where each subimage represents one aspect of the texture and is regarded as a new sample. Subsequently, by using these subimages for training, more variations of the texture are incorporated which is beneficial for achieving higher generalization capability of the model.
2. It is presumed that a more compact model requires less training samples to learn a generalized representation of signals. Because textures are sparse and the sparse representation suggests a more compact model than the local estimators [22], the sparse representation is favourable to model the textural images.
3. Considering that supervised learning from a small number of training images is prone to overfitting, which results in low generalization capability in the classification of new images, it is important to consider both the discrimination and generalization of a model in the learning process. Regarding each texture as lying in a low dimensional manifold, it is expected that through a multi-manifold analysis, on the one hand the distance between different texture classes could be enlarged, thus increasing the discriminative power of the model; on the other hand the intra-class variation can be decreased, therefore mitigating the overfitting effect.

Based on the above considerations, we develop a sparse representation based multi-manifold analysis (SR-MMA) framework for texture classification from few training images. After extracting a set of image patches from each training image as the new training samples, we utilize sparse representation to model these new samples by assuming that each sample of a texture is generated from a sparse representation of a set of basis. Subsequently, we propose a supervised multi-manifold

Download English Version:

<https://daneshyari.com/en/article/10359455>

Download Persian Version:

<https://daneshyari.com/article/10359455>

[Daneshyari.com](https://daneshyari.com)