Learning an Efficient Texture Model by Supervised Nonlinear Dimensionality Reduction Methods

Elnaz Barshan, Mina Behravan, and Zohreh Azimifar

School of Electrical and Computer Engineering,
Shiraz University, Shiraz, Iran
barshan@cse.shirazu.ac.ir

Abstract. This work investigates the problem of texture recognition under varying lighting and viewing conditions. One of the most successful approaches for handling this problem is to focus on textons, describing local properties of textures. Leung and Malik [1] introduced the framework of this approach which was followed by other researchers who tried to address its limitations such as high dimensionality of textons and feature histograms as well as poor classification of a single image under known conditions.

In this paper, we overcome the above-mentioned drawbacks by use of recently introduced supervised nonlinear dimensionality reduction methods. These methods provide us with an embedding which describes data instances from the same classes more closely to each other while separating data from different classes as much as possible. Here, we take advantage of the superiority of modified methods such as "Colored Maximum Variance Unfolding" as one of the most efficient heuristics for supervised dimensionality reduction.

The CUReT (Columbia-Utrecht Reflectance and Texture) database is used for evaluation of the proposed method. Experimental results indicate that the algorithm we have put forward intelligibly outperforms the existing methods. In addition, we show that intrinsic dimensionality of data is much less than the number of measurements available for each item. In this manner, we can practically analyze high dimensional data and get the benefits of data visualization.

Keywords: Texture Recognition, Texton, Dimensionality Reduction.

1 Introduction

Texture is a fundamental characteristic of natural materials and has the capacity to provide important information about scene interpretation. Consequently, texture analysis plays an important role both in computer vision and in pattern recognition. Over the past decades, a significant body of literature has been devoted to texture recognition based on mainly over-simplified datasets. Recently, more and more attention has been paid to the problem of analyzing textures achieved in different illumination and viewing directions. As Figure 1 shows, recognizing textures with such variations generally causes much trouble. Leung and Malik [1] were amongst the first to comprehensively study such variations. They proposed 3-D textons which are cluster centers of a number of predefined filter responses (textons) over a stack of images with different viewpoint and lighting conditions. The basic idea here is to build a universal vocabulary from these

E. Bayro-Corrochano and J.-O. Eklundh (Eds.): CIARP 2009, LNCS 5856, pp. 209–216, 2009.

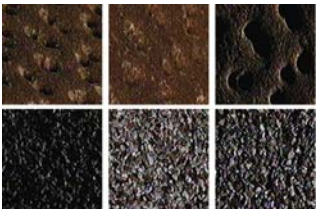


Fig. 1. Changing viewpoint and illumination can have a dramatic impact on the appearance of a texture image. Each row shows texture images of the same class under different viewpoint and lighting conditions.

textons describing generic local features of texture surfaces. Given a training texture class, the histogram of its 3-D textons forms the model corresponding to that texture. In the training stage, the authors acquired a model for each material using stacked images of different albeit a priori known conditions. This model, however, requires the test images to be in the same order as in the training. Leung and Malik also developed an algorithm for classifying a single image under known conditions. Yet, this method does not classify a single image as efficient as the case for the multiple images. Later, Varma and Zisserman [2] presented an algorithm based on Leung and Malik's framework, without requiring any a prior knowledge of the imaging conditions. In Varma and Zisserman's method, textons are obtained from multiple unregistered images of a particular texture class using K-means clustering. A representing model for each class is brought out using the texture library which is, actually, a collection of textons from different texture classes.

For the purpose of achieving a faithful representation of various textures, a finite set of textons (i.e., texton library) closely representing all possible local structures, ought to be obtained. Hence, the cardinality of our texton library must be considerably large. Nevertheless, this may, by itself, cause high-dimensional models. To address this issue, Cula and Dana [3] employed the method of Principal Component Analysis (PCA) and compressed the feature histogram space into a low-dimensional one. Applying PCA as a method for unsupervised linear dimensionality reduction causes a number of limitations to be discussed later on.

In this paper, we focus on the problem of high dimensionality of texture models and, furthermore, introduce an efficient algorithm for classifying a *single* texture image under *unknown* imaging conditions. Here, our attempt is to shed light on a new approach to overcome this difficulty. In this viewpoint of ours, a richer space is sought after that can reflect modes of variability which are of particular interest. As a result, we propose to project data onto an ideal space peculiar to our problem. Not only is the new space thus gained supposed to be of low dimension, but also it has to provide us with a better representation of data, i.e. to be more discriminative for the classification algorithm. In other words, we aim at transforming the ill-posedness of texture classification into a better-posed problem. To find this transformation, we take the benefits of recently introduced supervised nonlinear dimensionality reduction methods.

The rest of this paper is organized as follows: Section 2 provides an overview on texton-based texture representation. Next, we briefly describe one of the most efficient

heuristics for reducing the dimensionality of nonlinear data. In section 3, we introduce our new method which represents a model of enough capability for classifying a single image under unknown imaging conditions. Experimental results of the proposed algorithm are presented in section 4 followed by "Conclusion" in section 5.

2 Background Review

2.1 Texton-Based Texture Representation

A texture image is constructed based on certain structures, such as spot-like features of various sizes, edges with different orientations and scales, bar-like characteristics and so on. It was reported that local structure of a texture can be closely represented by its responses to an appropriate filter bank [4,5,6].

Different filter banks focus on different constructive structures. Accordingly, Leung and Malik [1] introduced an appropriate LM filter bank which was later employed by a number of other researchers. The LM set is a multi-scale, multi-resolution filter bank that has a combination of edge, bar and spot filters. It consists of the first and the second derivatives of Gaussian (at six orientations and three scales), eight Laplacian of Gaussian (LOG) filters and four Gaussian filters, a total of 48 filters. In this study we used this filter bank.

One of the fundamental properties of textures is pattern repetition, which means that filter responses to only a small portion of texture image are sufficient to describe its structure. This small set of prototype response vectors of one image was called 2-D textons by Leung and Malik. They also proposed 3-D textons definition; this definition is based on the idea that the vectors obtained from concatenating filter responses of different images of the same class will encode the appearance of dominant features in all of the images. They used 3-D textons to represent a framework for recognizing textures under different imaging conditions. Since the inspiration of our work comes from the Leung and Malik's algorithm [1], let us briefly review this method.

The Leung and Malik's algorithm uses 3-D textons from all the texture classes to compute a universal vocabulary. To construct such a desirable vocabulary, the K-means clustering algorithm is applied to the data from each class individually. The class centers are, then, merged together to produce a dictionary. This dictionary should be pruned in order to produce a more efficient, faithful and least redundant second version. After constructing the vocabulary, different images from each class are passed through the filter bank and stored in a large vector, which is then assigned to the nearest texton labels from the said dictionary. The histogram of texton frequencies is computed in such a manner as to obtain one model per class. Textons and texture models are learnt from training images. Once this is done, classification of a test image is done by computing a model from images with different imaging conditions, as in the training stage. The algorithm selects the class for which chi-square distance between the sample histogram and the model histogram could be minimized. Readers interested in other aspects of the original algorithm are referred to Leung and Malik's original paper [1]. Despite the fact that Leung and Malik's algorithm has numerous advantages, it has its own limitations discussed by several authors from different aspects. These disadvantages were to be

addressed by the very authors. In section 3, we discuss shortcomings of this algorithm and introduce a new approach based on dimensionality reduction methods.

2.2 Dimensionality Reduction

The problem of dimensionality reduction and manifold learning has recently attracted much attention on the part of many researchers. Manifold learning is a method to retrieve low dimensional global coordinates that faithfully represent the embedded manifold in the high dimensional observation space.

Most dimensionality reduction methods are unsupervised. That is to say, they do not respect the label or the real-valued target covariate. Therefore, it is not possible to guide the algorithm towards those modes of variability that are of particular interest. For example, where possible, by using labels of a subset of the data according to the kind of variability that one is interested in, the algorithm can be guided to reflect this kind of variability.

Amongst the proposed supervised nonlinear dimensionality reduction methods, "Colored Maximum Variance Unfolding" (CMVU) [7] is of much interest and capability.This method is built upon "Maximum Variance Unfolding" (MVU) method [8]. By integrating two sources of information, data and side information, CMVU is able to find an embedding which: 1) preserves the local distances between neighboring observations, and 2) maximally aligns with the second source of information (side information). Theoretically speaking, CMVU constructs a kernel matrix $\mathbf{K}$ for the dimension-reduced data X which has the capacity to keep the local distance structure of the original data Z unchanged, so that X maximally depends on the side information Y as described by its kernel matrix $\mathbf{L}$. This method is formulated by the following optimization problem:

Maximize tr HKHL subject to:
$1.K \succeq 0$
$2.K_{ii} + K_{jj} - 2K_{ij} = d_{ij}$ for all (i,j) with $\eta_{ij} = 1$

where $K, L \in \mathbb{R}^{m*m}$ are the kernel matrices for the data and the labels, respectively, $H_{ij} = \delta_{ij} - m^{-1}$ centers the data and the labels in the feature space, and binary parameter η_{ij} denotes whether inputs z_i and z_j are k-nearest neighbors or not. The objective function is an empirical estimate of "Hilbert-Schmidt Independence Criterion" (HSIC) that measures the dependency between data and side information [9]. This optimization problem is an instance of semi-definite programming (SDP). From the solution of SDP in the kernel matrix K, output points X_i could be derived using singular value decomposition. Figure 2 illustrates embedding of 2007 USPS digits produced by CMVU and PCA, respectively.

3 Methodology

In this section, we discuss different texton-based texture representation methods. Then, we present our new method to address all accompanying drawbacks, and will show its superiority compared to other methods each focusing on a specific limitation.

As stated in the previous section, issues associated with the use of 3D textons to classify 3D texture images are:

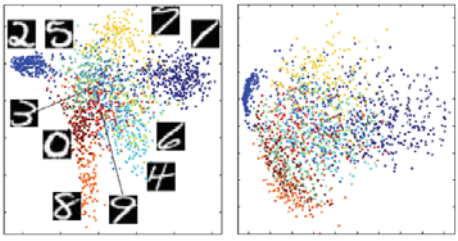


Fig. 2. Embedding of 2007 USPS digits produced by CMVU and PCA, respectively [7]

- increased dimensionality of feature space to be clustered in the later stages,
- increased time complexity of the iterative procedure to classify a single image which causes the convergence problem,
- necessity of a set of ordered texture images captured under known imaging conditions, and
- introduction of only one comprehensive model per class whereas it is unlikely that a single model can fully account for the various appearance of real-world surfaces.

By use of 2-D textons, none of the above problems would appear. 2-D textons are cluster centers of filter responses over a single image (and not over a stack of images) captured at different conditions. The problem here is how we should represent different instances from the same class as being inter-related while preserving the between-class distances. One solution is to select the models which best represent their texture classes. Cula and Dana [3] proposed a model selection algorithm in a low dimensional space. They fitted a manifold to low dimensional representation of models specifically generated for each class and removed the models which least affected the manifold shape. Their algorithm, notwithstanding, introduces some drawbacks. For projecting models into a low dimensional space, they utilized PCA which is an unsupervised linear dimensionality reduction method. The PCA works well if the most important modes of data variability are linear. But in this study, the variability of models cannot be expressed linearly and this causes poor performance of PCA. The second problem stems from the fact that two different distance measures are used in constructing the manifold path in the training stage and selecting the closest model in the classification stage. In other words, when constructing the manifold path, at each step the closest point in terms of imaging angles is chosen, while in classification phase, the closest surface class is selected in terms of distance between models feature vectors. Another significant issue is that this algorithm ignores inter-class variation between textures since the models for a texture are selected without considering the other texture classes.

Having discussed the above issue, we propose to analyze this problem from another viewpoint: *reducing the dimensionality of model histograms to their intrinsic dimensionality*. By mapping the models to a very low dimensional space, the complexity of the classification decreases and model selection can take the benefits of data visualization. It is important to note that the basic modes of variability of our data are nonlinear. Therefore, the dimensionality reduction method should be capable of unfolding the manifold on which the nonlinear dataset is lying. On the other hand, we are searching

for a space in which models from the same classes stay more closely while models from different classes remain as much discriminated as possible.

Here we take the advantages of CMVU, which is one of the most efficient heuristics for supervised dimensionality reduction, as discussed in § 2. This method generates brilliant results for training data, e.g., it is empirically observed that the most significant modes of the variability of a dataset with dimensionality of 1200 can be presented in a space of as low as five dimensions. It confirms our reasoning of selecting the CMVU to visualize the train data. This method, however, faces some complications in projecting the test data. Desired embedding for training data could be computed with respect to its labels. Because of the fact that at the testing time the second source of information (the labels) is not available, this method does not provide us with an embedding of testing data to the space in which the training data is embedded. Herein, we choose to project the testing data based on the fundamental idea of "Locally Linear Embedding" (LLE) [10]. The projection procedure for testing data S is as follows:

Alg. 1. The projection procedure for testing data

Input: training data matrix in the original space, **Z**, projected training data matrix, **X**, testing data matrix in the original space, **S**, and the number of testing data, m

Onput: Projected testing data matrix, **P**

1: **for all** $i \in \{1 \dots m\}$

2: $N = \{z_j \in Z | \eta_{ij} = 1\}$

3: $W = \text{argmin}_W E(W) = |s_i - \Sigma_j w_{ij} N_j|$

4: $p_i = \Sigma_j W_{ij} x_j$

5:**end for**

The projection of testing data using this LLE-like method causes some negligible differences under the circumstances of the presence of labels being computed using CMVU.

4 Experimental Results

We perform all experiments on the CUReT dataset [11]. This dataset provides a starting point in empirical studies of texture surfaces under different viewing and illumination directions. This database contains 61 different textures, each observed with over 200 combinations of viewpoint and illumination conditions.

In order to construct our texton library, 40 unregistered images with different imaging conditions from 20 texture classes are employed. We use the same texture set as the one examined by Cula and Dana [3]. We extract 2-D textons from each image, and apply K-means algorithm to the texture classes individually in order to obtain 60 centers from each different materials. These 1200 centers are used as initial points for the final clustering step, which produce an efficient and least redundant dictionary from 1200 textons.

To justify the effectiveness of our approach, we have performed three sets of experiments. 10 arbitrary texture images from each texture class are selected in all three of experiments. Thence, the total number of test images is 200. In the first experiment,

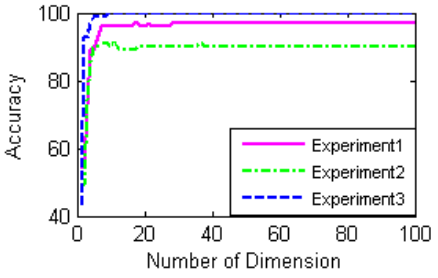
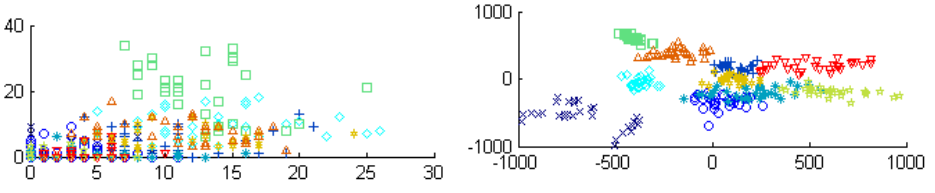


Fig. 3. Classification rate on CURet dataset for different projection dimensions. Three sets of experiments have been performed. In experiment (1) exactly the same images involved in constructing the vocabulary have been used. Experiment (2) is a bit more complex, in the sense that testing image conditions differ from those used in constructing the vocabulary. In Experiment (3) two disjoint sets of texture classes are used in library construction and the texture recognition, separately.



(a) Original Space (b)CMVU Space

Fig. 4. The first two dimensions of CUReT dataset in the original space and the space produced by CMVU, respectively. Dot shapes are used to denote textures from different classes.

exactly the same images involved in constructing the vocabulary have been used. The second experiment is a bit more complex, in the sense that testing image conditions differ from those used in constructing the vocabulary. In the last experiment, the most complex one, two disjoint sets of texture classes are used in library construction and texture recognition, separately. Figure 3 shows the percentage of correctly classified test images as a function of dimensions used to represent projected models by CMVU. This figure clearly shows that up to a certain level the accuracy increases with dimensionality and converges to a fixed point with very low variability. Additionally, this Figure shows better results for experiment 3, which is the consequence of selecting more discriminative classes than the other sets chosen for constructing the library. In Figure 4 the first two dimensions of data in the original space is shown as well as its projection in/on to the new space using CMVU. Obviously enough, CMVU introduces a clear data separation with an excellent visualization.

5 Conclusions

This paper introduces the idea of supervised nonlinear dimensionality reduction to alleviate the difficulties associated with texture recognition. Although we were not the first

to address the high dimensionality of texture models, the contribution of this work is its efficient mapping of data nonlinearity, i.e., we have shown how to represent the data intrinsic information while magnifying the descriptive properties of the original feature space. Besides, we proposed a LLE-like approach to cope with shortcoming of CMVU in projecting the test data when carrying no side information. This paper presents a new framework to efficiently visualize a hugely dimensioned data in a very low dimension yet rich space.

This study can be extended in different directions: 1) orientation and scale invariant features may be extracted using techniques such as gradient histograms, 2) advanced classifiers and clustering algorithm can be investigated, and 3) the data visualization techniques may also be employed in selecting the most discriminative texture models.

References

1. Leung, T.K., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1), 29–44 (2001)
2. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. International Journal of Computer Vision 62(1)
3. Cula, O.G., Dana, K.J.: 3d texture recognition using bidirectional feature histograms. International Journal of Computer Vision 59(1), 33–60 (2004)
4. Clark, M., Bovik, A.C., Geisler, W.S.: Multichannel texture analysis using localized spatial filters. IEEE Trans. Pattern Anal. Mach. Intell. 12(1), 55–73 (1990)
5. Randen, T., Husoy, J.H.: Filtering for texture classification: A comparative study. IEEE Trans. Pattern Anal. Mach. Intell. 21(4), 291–310 (1999)
6. Prabhakar, S., Jain, A.K., Hong, L.: A multichannel approach to fingerprint classification. IEEE Trans. Pattern Anal. Mach. Intell. 21(4), 348–359 (1999)
7. Smola, A.J., Borgwardt, K.M., Song, L., Gretton, A.: Colored maximum variance unfolding. In: NIPS (2007)
8. Weinberger, K.Q., Saul, L.K.: An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In: AAAI (2006)
9. Bousquet, O., Smola, A.J., Gretton, A., Scholköpf, B.: Measuring statistical dependence with hilbert-schmidt norms. In: ALT, pp. 63–77 (2005)
10. Saul, L.K., Roweis, S.T.: Think globally, fit locally: Unsupervised learning of low dimensional manifold. Journal of Machine Learning Research 14, 119–155 (2003)
11. Nayar, S.K., Koenderink, J.J., Dana, K.J., Ginneken, B.v.: Reflectance and texture of real-world surfaces. ACM Trans. Graph. 18(1), 1–34 (1999)