Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition New York, June 1993, pp 638-639.

MIT Media Lab Perceptual Computing TR # 215

Real-time Recognition with the entire Brodatz Texture Database

Rosalind W. Picard
MIT Media Lab
Cambridge, MA 02139
picard@media.mit.edu

Tanweer Kabir
CSIRO; PO Box 76
Epping, NSW 2121
tkabir@rp.csiro.au

Fang Liu
MIT Media Lab
Cambridge, MA 02139
fliu@media.mit.edu

Abstract

The Brodatz Album has become the de facto standard for evaluating texture algorithms, with hundreds of studies having been applied to small sets of its images. This paper compares two powerful recognition algorithms, principal components analysis and multiscale autoregressive models, by evaluating them on a 999-image database derived from the entire Brodatz Album. The variety of homogeneous and non-homogeneous images studied is thus nearly an order of magnitude larger than has been compared before, giving one snapshot of the "state of the art" in real-time texture recognition.

1 Introduction

Image recognition applications are shifting rapidly from traditional areas of target recognition and satellite imagery to new areas in multi-media image/video analysis and retrieval of visual information. Many of the old applications can be typified by having a small number of "classes" of patterns, e.g., wheat, grass, water, and a large availability of training samples of each. In contrast, many new applications have a huge number of classes and a small set of representative samples, e.g., searching through video for a particular shot. This paper examines and compares two powerful recognition methods in a new multimedia search environment where real-time discrimination between over a hundred competing classes is important.

2 Brodatz texture database

The "Brodatz texture database" is derived from the Brodatz Album [1]. It was formed by cropping nine 128×128 subimages from the centers of 111[1] different original 8-bit 512×512 images received on tape from the Georgia Institute of Technology. Thus the database consists of 999 different 128×128 8-bit images, which can be considered to represent 111 different "classes" of data. Consequently it has a relatively large number of classes, and a small number of examples for each class.

Most texture studies on classification, discrimination, and segmentation have been run on small subsets of test data from the Brodatz Album, typically four to sixteen images at once. Moreover, the tested images usually exhibit strong homogeneity within each class as well as visual and semantic dissimilarity between classes. Often they are chosen to all be "microtextures". This study differs in that it includes approximately an order of magnitude greater variety, including many inhomogeneous and large-scale patterns.

The commonplace restriction of texture studies to homogeneous microtextures is an artificial and potentially misleading scenario. Results may obtain 90-100% accuracy on small sets of such data, but these results do not typically extend to real scenes, which usually also contain non-homogeneous and non-textured regions. By including the non-homogeneous Brodatz images in the database, a much more difficult and realistic scenario is obtained. We have found *subsets* of the Brodatz database which reach 100% classification accuracy with no false alarms [2], but performance drops significantly when considering the whole database.

For a very inhomogeneous Brodatz image, humans might not classify its 9 database images as being from the same class. A true "semantic classification" would, for example, be expected to identify that all lizard skin images are similar even if they came from different original Brodatz images. This study does not explicitly consider semantic criteria; subsequently, its notion of "similar images" may not match that of humans.

Note additionally that the Brodatz Album has limited variety in pattern scale, rotation, contrast, and perspective. Developing methods to handle these transformations is essential for recognition in real scenes, but cannot be addressed with the present Brodatz data unless it is altered. Nonetheless, the current database is significantly more diverse than has been considered in prior texture analysis studies. Consequently, it provides an important benchmark for evaluating progress in texture recognition.

3 Two recognition methods

Principal component analysis is an important tool in pattern recognition, and has been successfully applied to small sets of textures [3] as well as many other patterns for representation, recognition and discrimination. The key difference in the method used here is that it has been made shift-invariant which greatly improves its performance on textures. The shift-invariance is a consequence of discarding the phase of the patterns in a method similar to that used for face recognition by Akamatsu, et. al. [4]. The training and evaluation for this method, its implementation in the image database environment, and its use in ranking "how easy" each Brodatz image is to classify are described in [2]. The features produced for each database image consist of 99 projection coefficients onto a set of eigenfunctions. The eigenfunctions are from the pooled covariance matrix of

[1] The Brodatz Album actually has 112 different textures, of which D45 was missing from the tape.

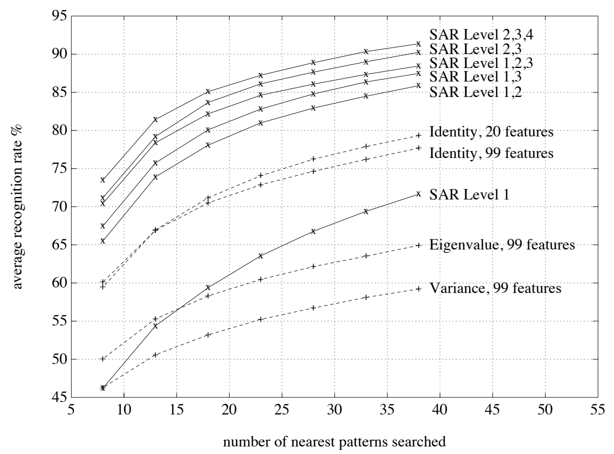

Figure 1: Performance comparison of SAR and principal components methods over entire Brodatz database.

a randomly-chosen training set of 100 images. Three weighted least-square distances are compared here on the 99 features and on a feature subset of size 20.

The second recognition method is based on a multi-scale simultaneous autoregressive (SAR) model. This is a second-order noncausal model characterized by five parameters at each resolution level. At any one scale, the model is shift-invariant in its assumption that the parameters are not spatially varying. The model used here is equivalent to that of Mao and Jain [5] with the exception of not including their rotation-invariant averaging. Parameters and parameter covariances for this model were estimated for each of the 999 images. The covariances for each pattern were used in a Mahalanobis distance during classification (found to perform significantly better than the Euclidean distance.)

For both methods the features and their weights are precomputed. Distances between features are computed in real-time on a SUN-SPARC workstation.

4 Performance analysis

Each of the nine subimages taken from an original Brodatz image are considered to be from the same class. Recognition performance for several variations of the two different methods is shown in Figure 1.

This figure was formed as follows. A search was done for each of the 999 database images. The x-axis corresponds to the number of database images closest to the test image which were searched for "matches", i.e. samples from the same class. For each image, a maximum of 8 matches could be found for a 100% recognition rate. Each y-axis value was found by averaging the recognition rates over all 999 images. For example, if 13 closest images are being searched, and if 4 matches are found for one of the 999 images, then the recognition rate for that image would be 50%. This recognition rate would be averaged with those of the other 998 images to obtain one data point in Figure 1.

These performance criteria differ from those used in traditional recognition scenarios. Here there is no need to decide a label for each image. Rather, the goal is to retrieve and display images which are close to the selected image in real-time. The environment is one of interactive image retrieval where the user can view 40 images in a display at once. As long as the images found within the viewing window contain the "correct" images, the system is successful. Moreover, a limited number of false positives can be desirable for finding similar-looking images.

The top four curves in Figure 1 correspond to recognition performance using the SAR model estimated over different resolution levels. The best results were achieved using the levels 2,3, and 4. At five parameters per level, this required estimating 15 features per 128×128 image. Below the top five SAR results are the results using features corresponding to the twenty biggest eigenvalues and a Euclidean (identity weighted) distance. It is interesting to observe that in spite of the use of decorrelated features, including more (99) actually decreases the performance. (Using less than 20 also decreased performance.) Using the principal components features and distance weights of inverse eigenvalues or inverse feature variances performed even worse. It is important to remember that although directions with greater variance may be optimal in mean-squared error for representation, they are not necessarily optimal for discrimination.

It may seem surprising that fifteen SAR features achieved greater than 90% accuracy with the "in view=success" criterion while twenty principal components achieved less than 80%. Eigenfilter-based features have been found to perform as well as co-occurrence features on small sets of Brodatz images [3]. Since co-occurrences contain more information than correlations, and since basic SAR features are estimated from correlations, it seems that significant improvement is due to the multi-scale information. The multi-scale SAR should be compared with other filter-banks such as Gabor, wavelet, and steerable pyramid.

Results here are displayed as for an "operating characteristic" indicating how the performance increases monotonically with the permitted number of false positives. In the limit as the number of allowed false positives goes to 990 all the curves reach 100%. A meaningful way to demonstrate that a texture recognition method is better than these is to show that its operating characteristic lies above the results in Figure 1.

Acknowledgments

This research was made possible in part by the sponsorship of BT, U.K. The authors would like to thank Chris Perry for his help with the experiments and Mark Shackleton for helpful comments on this paper.

References

[1] P. Brodatz. *Textures: A Photographic Album for Artists and Designers.* Dover, New York, 1966.

[2] R. W. Picard and T. Kabir. Finding similar patterns in large image databases. In *Proc. ICASSP*, pages V–161–V–164, Minneapolis, MN, 1993.

[3] M. Unser. Local linear transforms for texture measurements. *Signal Processing*, 11:61–78, 1986.

[4] S. Akamatsu, T. Sasaki, H. Fukamachi, and Y. Suenaga. A robust face identification scheme – KL expansion of an invariant feature space. In *Proc. SPIE Conf. on Intell. Robots and Comp. Vis.*, volume 1607, pages 71–84, Boston, MA, Nov. 1991.

[5] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Patt. Rec.*, 25(2):173–188, 1984.