

Texture Analysis via Unsupervised and Supervised Learning¹

H. Greenspan and R. Goodman
Department of Electrical Engineering
116-81
California Institute of Technology
Pasadena CA 91125

R. Chellappa
Signal and Image Processing Institute
Department of Electrical Engineering - Systems
University of Southern California
Los Angeles, CA 90089

Abstract

A framework for texture analysis is proposed based on combined unsupervised and supervised learning. The textured input is represented in the frequency - orientation space via a Gabor - wavelet pyramidal decomposition. This representation has optimal localization properties in space and in frequency, and is biologically motivated. In an unsupervised learning phase a neural network vector - quantization scheme is used for the quantization of the feature - vector attributes and a projection onto a reduced dimension clustered map for initial segmentation. A supervised stage follows in which labeling of the textured map is achieved using a rule based system. This information theoretic technique is utilized to find the most informative correlations between the attributes and the texture class specification. Ultimately, a minimal representation for a library of textures is learned in a training mode, following which the classification of new test images is achieved. This combined neural network - rule based approach is suggested as a general framework for pattern classification. Simulation results for the texture classification are given.

1 Introduction

Visual texture is one of the most fundamental properties of a visible surface. It participates as one of the major modalities which help us in the understanding of our visual environment. The different textures in an image are usually very apparent to a human observer, but automatic description or recognition of these patterns is yet an unsolved and complex problem.

In this work we suggest a general framework for classification, incorporating unsupervised and supervised learning, via a neural network (NN) model and a rule - based system [1,2] respectively. This is presented in Fig. 1. The proposed framework, consisting of three major stages, is applied to the texture analysis task. The first stage performs feature extraction and transforms the image space into an array of N - dimensional feature vectors, each vector corresponding to a local window in the original image. The second stage analyzes the N -dimensional feature space and finds clusters in it, thus quantizing the continuous input features. The quantized feature vectors are the input to the third stage of analysis in which categorization of the input into one of the prelearned texture classes is achieved.

There is much evidence supporting the use of Gabor filters in the feature - extraction phase [3,4,5]. The features extracted are in the orientation and frequency domains. An open issue is the decision regarding the appropriate number of frequencies and orientations required for the representation of the textures. We suggest a learning mechanism that is able to define a minimal set of primitives which convey sufficient information about the visual input for its differentiation and labeling. In the proposed approach we reduce the feature space both in the unsupervised and supervised stages of analysis. In the unsupervised stage a NN vector-quantization procedure is used based on Kohonen's topology-preserving mapping [6]. A set of informative features are extracted in the supervised stage as congruency rules between attributes using an information theoretic measure. This learned set can now act as a classification set for test images.

¹This work is funded in part by DARPA under the grant AFOSR-90-0199 and in part by the Army Research Office under the contract DAAL03-89-K-0126. Part of this work was done at Jet Propulsion Laboratory. The advice and software support of the image-analysis group there, especially that of Charlie Anderson, is greatly appreciated.

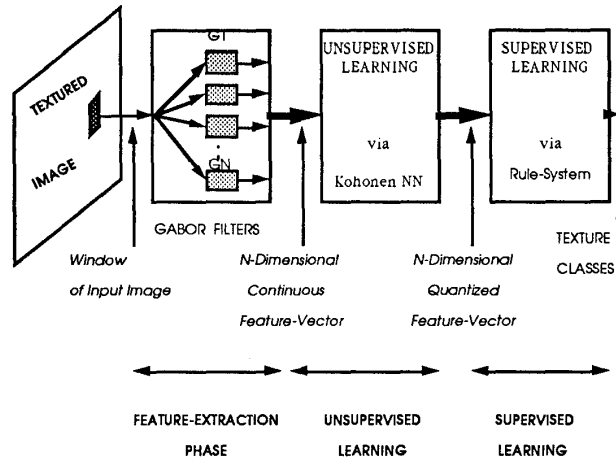


FIG. 1. - Texture Classification Architecture

2 Feature Extraction Stage

The initial stage for a classification system is the feature extraction phase through which the attributes of the input domain are extracted and presented towards further processing. The chosen attributes are to form a representation of the input domain, which encompasses information for any desired future task, in a compact and non - redundant form.

The literature presents, for the filtering stage, extreme cases such as close to two hundred filters acting in parallel and in a multi-level scheme [3], to the selection of one filter per texture based on its Fourier spectrum analysis [4]. A systematic way for the extraction of initial features is needed. In this work, we use the Gabor pyramid, or the Gabor wavelet decomposition to define an initial finite set of filters. Gabor functions are complex sinusoidal gratings modulated by 2-D Gaussian functions in the space domain, and shifted Gaussians in the frequency domain. Equation 1. gives the general form of the 2-D Gabor function and its Fourier transform:

$$g(x, y; u_0, v_0) = e^{-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i[u_0x + v_0y]} ; G(u, v) = e^{-2\pi^2(\sigma_x^2(u-u_0)^2 + \sigma_y^2(v-v_0)^2)} \quad (1)$$

where (σ_x, σ_y) define the widths of the Gaussian in the spatial domain and (u_0, v_0) is the frequency of the complex sinusoid. The 2-D Gabor filters form a complete but non-orthogonal basis which can be used for image encoding into multiple spatial frequency and orientation channels. The Gabor filters are appropriate for textural analysis as they have tunable orientation and radial frequency bandwidths, tunable center frequencies, and optimally achieve joint resolution in space and spatial frequency.

We implement a pyramidal approach in the filtering stage, reminiscent of the Laplacian Pyramid [7]. Results from vision research indicate that the central frequencies of localized frequency operators have octave relationships [8]. Coding an image with logarithmic scaling along the frequency axis results in spatial narrowing of the filters with increasing frequency-numbers, since the product of the effective spatial and spectral widths remains constant in the Gabor scheme. The filters, logarithmically transformed, have a similar shape but vary in their effective widths by multiples of two. It has been shown [8] that the completeness of the image representation is preserved in this scheme. In our simulations a computationally efficient scheme involves a pyramidal representation of the image which is convolved with fixed spatial support oriented Gabor filters. Five scales are used with 4 orientations per scale (0,90,45,-45 degrees) to produce a 20-Dimensional feature vector as the output of the feature extraction stage.

The pyramidal approach allows for a hierarchical, multiscale framework for the image analysis. This is a desirable property as it enables the identification of features at various scales of the image and thus is attractive for scale-invariant pattern recognition. A more rigorous analysis of the pyramidal approach, its extracted features and its application to scale-invariant recognition will be the topic of a future paper.

3 Vector Quantization via Unsupervised Learning

The unsupervised learning phase can be viewed as a preprocessing stage for achieving yet another, more compact representation, of the filtered input. The goal is to quantize the continuous valued features which are the result of the initial filtering stage. The need for discretization becomes evident when trying to learn associations between attributes in a statistically-based framework, such as a rule-based system. Moreover, in an extended framework, the network can reduce the dimension of the feature domain. This shift in representation is in accordance with biological based models.

The output of the filtering stage consists of N ($=20$), continuous valued feature maps. Thus, each local area of the input image is represented via an N -dimensional feature vector. An array of such N -dimensional vectors, viewed across the input image, is the input to the learning stage. We wish to detect characteristic behavior across the N -dimensional feature space, for the family of textures to be learned. By projecting an input set of samples onto the N -dimensional space, we search for clusters to be related to corresponding code-vectors, and later on, recognized as possible texture classes. A NN vector quantization procedure, based on Kohonen's model [6] is utilized for this stage.

The Kohonen mapping is an adaptive process in which the parameters (weights) of each unit in the network converges to a value such that every unit becomes matched to a particular domain of the input signals, in a regular order. Each unit thus represents a cluster in the input domain.

The vector-quantization learning algorithm defines a mapping from an N -dimensional input vector, X , to an M -dimensional output vector Y . A set of code-vectors M_i ($i = 1..M$) are defined together with the M L_2 -norms $D_i(X) = \|X - M_i\|$, and a mapping $X \rightarrow Y_c$ is defined such that $Y = Y_c$ minimizes $D_i(X)$ with respect to i . The discrete index c is the output, or the codeword representing the continuous-valued input.

In the presented simulations, each dimension, out of the N -dimensional attribute vector, is individually clustered. All samples are thus projected onto each axis of the space and one-dimensional clusters are found. The output of this preprocessing stage is an N -dimensional quantized vector of attributes which is the result of concatenating the discrete valued codewords of the individual dimensions. Each dimension can be seen to contribute a probabilistic differentiation onto the different classes via the clusters found. As some of the dimensions are more representative than others, it is the goal of the supervised stage to find the most informative dimensions for the desired task (with the higher differentiation capability) and to label the combined clustered domain.

4 Supervised Learning via a Rule-Based System

The goal of the supervised stage is to utilize the existing information in the feature maps for higher level analysis, such as input labeling and classification. In particular we need to learn a classifier which maps the output features of the Kohonen stage to the texture class labels. Any classification scheme could be used. However, we utilize a rule - based information theoretic approach which is an extension of a first order Bayesian classifier, because of its ability to output probability estimates for the output classes [1]. The classifier defines correlations between input features and output classes as probabilistic rules of the form: If $Y = y$ then $X = x$ with prob. P . A data driven supervised learning approach utilizes an information theoretic measure to learn the most informative links or rules between features and class labels [2]. The classifier then uses these links to provide an estimate of the probability of a given output class being true. When presented with a new input evidence vector, a set of rules R can be considered to "fire". The classifier estimates the posterior probability of each class given the rules that fire in the form $\log(p(x)/R)$, and the largest estimate is chosen as the class label decision. In addition, the actual class probability could be used for further higher level processing.

5 Simulation Results

The performance of the presented framework is illustrated on several images. The model was tested on both synthetic and natural textures, taken from the standardized texture sets of textons [9] and the Brodatz library of natural textures [10] , respectively.

An example of 2 and 3 class texton images is presented in Fig. 2. Shown are the mosaic textures, a subset of Gabor filters and the corresponding filtered maps (a). The classification map, in discretized form, and an error map, are shown in (b). A correct classification rate of 95 - 99% was obtained. These are representative results for the texton analysis. Fig. 3. presents 2 and 3 natural-texture labeling results. Shown (starting from the left) are the mosaic input image, its labeled map, followed by an error map and a probability estimation map. The gray levels of the label maps correspond to prelearned texture classes. In the error maps, the white regions indicate errors, where as in the probability maps - the darker areas indicate areas of least confidence. In the 2 class case, a classification rate of 90 - 98% was achieved. It is noteworthy that structured textures (such as Canvas and Herring) and stochastic textures (such as Sand and Bubble) are handled in a similar framework of analysis. Looking at the label map (second column) and the corresponding error map (third) it is clear that structured textures with strong orientation components are easier to differentiate than stochastic unoriented textures. Even so, a classification result of 90% was obtained for the Bubble/Sand case which is a difficult texture mosaic (even to our own eyes). It is our aim to arrive at probabilistic mappings of the input image, rather than the presented hard-decision labeling. Initial results of the probability estimates in the classification stage are presented in column 4. In Canvas/Herring and Bubble/Sand cases we see that the few areas with labeling errors are places of very low confidence rating. Based on these probability estimates we feel that many of the current mislabeled areas can be removed and an even higher classification result may be achieved. In the 3 texture mosaic of Canvas Rock and Sand an 83% classification rate is achieved. It is apparent (from the label and error maps) that local areas in the Canvas (bright squares) convey similar information as the Rock texture. Thus, the 2 labels are intermixed. Again, the probability estimation map indicates the uncertainty in the decisions made at the boundary and at the locations of error. The performance can be enhanced by taking into account this information. A similar observation can be made in the case of Canvas and Sand in Fig. 4. The right label-map was achieved based on the learned library of the 3 textures above; the Sand is recognized and labeled accurately while the Canvas background is intermixed with Rock- labeling. For comparison, the left label map is obtained based on a library of sand and canvas textures only.

6 Discussion

The proposed learning scheme achieves a high percentage classification rate on both artificial and natural textures. Several points are worth noting. First, the examples presented pose very difficult classification tasks. The decision making process is a local one (as can be seen in the labeling of the 3-texture case). A global interpretation can be incorporated via a smoothing operation. We suggest the use of the probability maps as a guide to such a goal. This is currently under investigation. Second, the segmentation of the image is achieved via the labeling process. Third, a method for transition from continuous to discrete representation is suggested with an application to a combined neural-network and rule based learning schemes. The framework has been applied to both structural and unstructural (stochastic) textures. This is a much desired generality as most schemes suggested in the literature are applicable to only one of these categories.

The presented learning framework starts with a complete representation of the input image in the frequency and orientation domains and then combines unsupervised and supervised learning. An information based approach utilizes the most informative correlations between the input features for the recognition process.

Current work involves further investigation into the three stages of the architecture. We hope to show that scale and rotation invariance are achievable goals using the presented framework.

References

- [1] R.M. Goodman, C. Higgins, J. Miller and P. Smyth, 'A Rule-Based Approach to Neural Network Classifiers,' *INNC 90 Paris - International Neural Network Conference*, Palais Des Congres, Paris, France, July 9-13, 1990.
- [2] R.M. Goodman and P. Smyth, 'The Induction of Probabilistic Rule Sets - the ITRULE Algorithm,' *Proceedings of the Sixth International Workshop on Machine Learning* 1989.

- [4] A. C. Bovik, M. Clark and W. S. Geisler, 'Multichannel Texture Analysis Using Localized Spatial Filters,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 55-73, 1990.
- [5] M. Porat and Y. Y. Zeevi, 'The Generalized Gabor Scheme of Image Representation in Biological and Machine Vision,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, July 1988.
- [6] T. Kohonen, 'Self Organisation and Associative Memory,' Springer-Verlag, 1984.
- [7] P.J.Burt and E. A. Adelson, 'The Laplacian Pyramid as a compact image code,' *IEEE Trans. Commun.*, Vol. COM-31, pp. 532-540, 1983.
- [8] J.G. Daugman, 'Uncertainty relation for resolution, spatial frequency, and orientation optimized by 2D visual cortical filters,' *J. Opt. Soc. Amer. (A)*, Vol. 2, no. 7, pp. 1160-1169, 1985.
- [9] B. Julesz, 'Textons, the elements of texture perception and their interactions,' *Nature*, Vol. 290, 91-97, 1981.
- [10] P. Brodatz, 'Textures', New York:Dover, 1966.

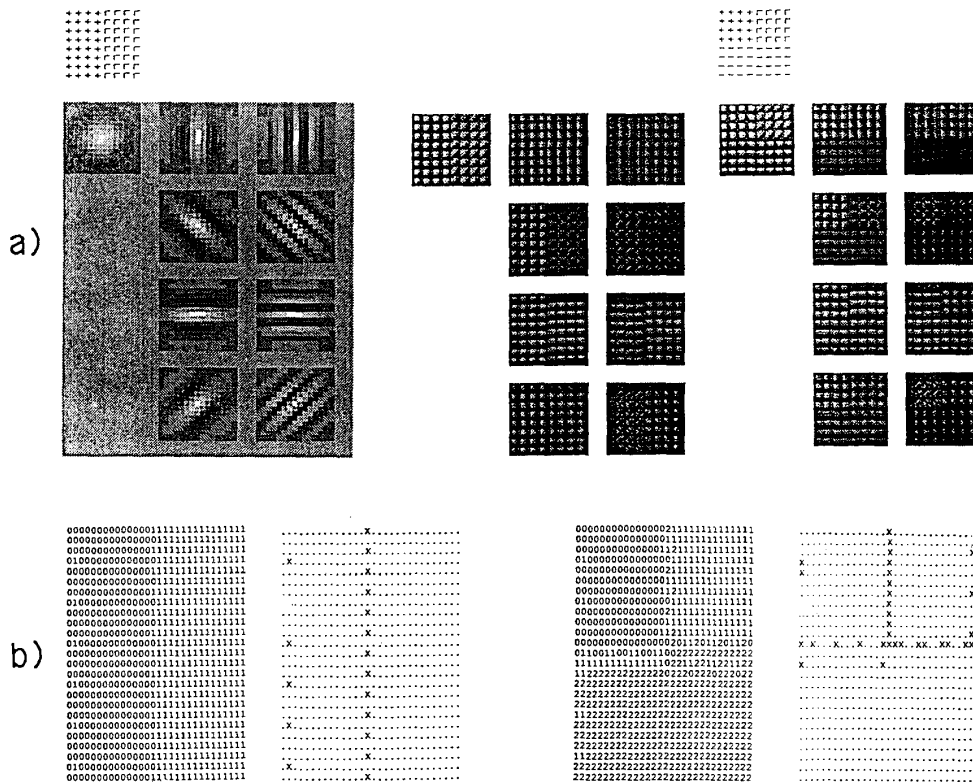


FIG. 2. - Texton classification results. a) Shown are examples of textons, a subset of Gabor filters and the corresponding filtered maps. b) Classification results; '0','1' and '2' represent prelearned texture classes + error maps where 'x' represents an error.

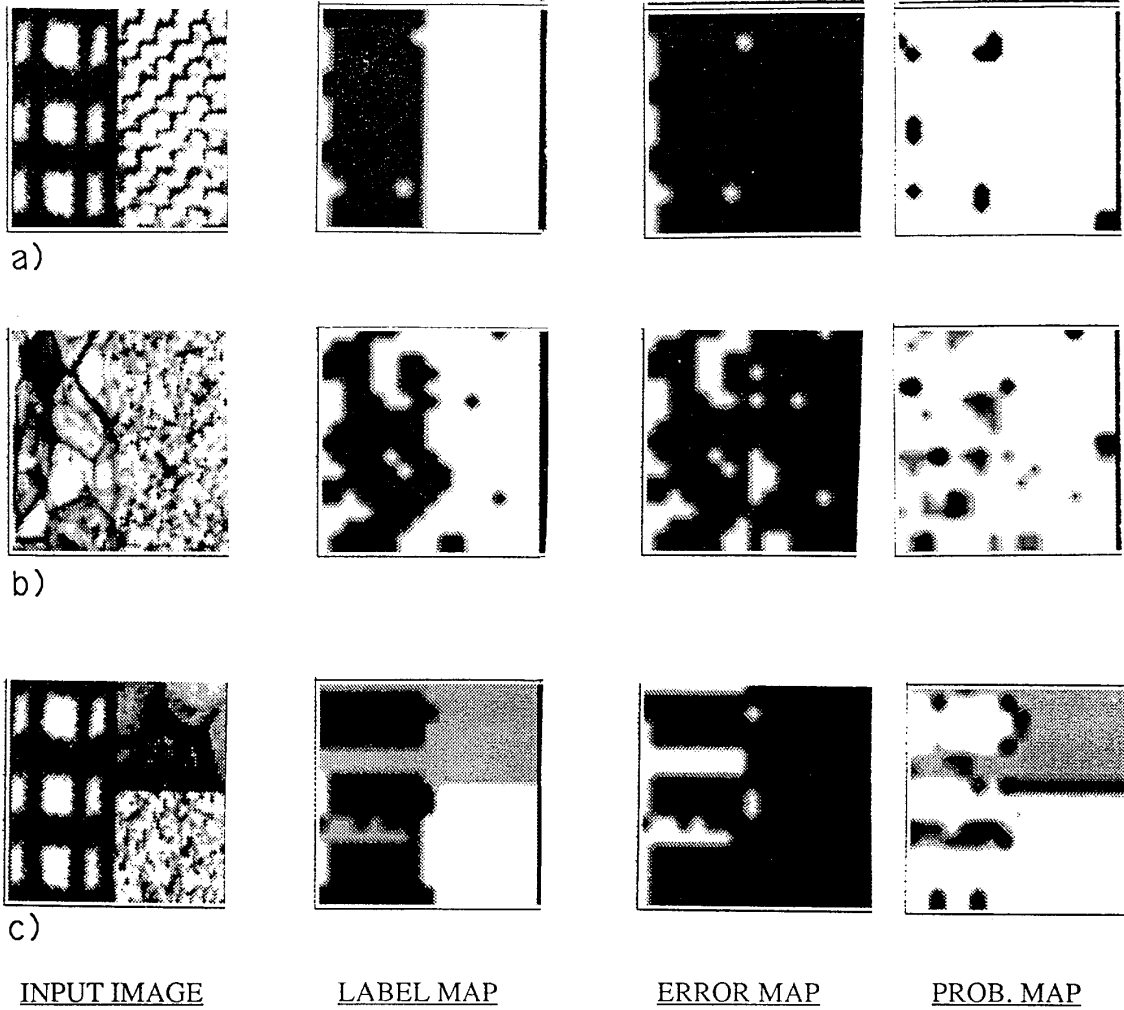


FIG. 3. - Natural texture labeling results. a) Canvas/Herring
 b) Bubble/Sand c) Canvas/Rock/Sand.
 Shown (left to right) are the input mosaic, the resultant label map,
 and the corresponding error and probability maps.

FIG. 4. - (below): Canvas/Sand. Label maps with 2 and 3 prelearned texture classes (left,right respectively).

