M3AV workshop on Neural Networks
in natural and artificial vision
Chamrousse (France) Jan. 29-30 1991
pp. 37-42 (ESPRIT Programme)

Gabor space and texture segmentation

J.M.H. du Buf and J. Bigün
Signal Processing Laboratory
Swiss Federal Institute of Technology
EPFL-Ecublens, CH-1015 Lausanne, Switzerland

Abstract

Considering the Gabor spectral decomposition as a model of the image representation in the striate cortex, we can try to exploit this representation for texture segmentation purposes. However, the dimensionality of this representation requires a compaction which makes a segmentation feasable in practice.

Here we use a Gabor filter set with 5 frequency and 6 orientation bands, resulting in 30 complex output images from which the phase information is discarded - as a result, we have a 4D local power spectrum with dimensions x, y, f, and o. This power spectrum is compacted by a) applying a Gaussian model with 5 free parameters, or b) computing only very few central moments. These reduced representations allow for the application of unsupervised segmentation algorithms.

The experimental results obtained are quite good. However, there seems to exist a basic problem which is related to the sensitivity of model parameters to the mixture of two local power spectra across a texture boundary. This leads to separate regions at the boundaries quite often.

Introduction

This contribution is not concerned with the exact spectral decomposition applied. A Gabor spectral decomposition is only one of many possibilities, including wavelets, prolate spheroidal functions, or Hermite polynomials. All these schemes allow to model the image representation in the cortical hypercolumns, that is a frequency and orientation selective decomposition. The Gabor representation applied here is based on a filter set with polar separable Gaussian transfer functions, see Fig. 1. We apply 5 frequency bands and 6 orientation bands. If the image size is 256x256 pixels, this leads to a representation with size 256x256x5x6x2. The factor of two is due to the fact that the filters are complex. Since the role of the local phase spectrum is rather obscure until now, we compute the local power spectrum only (size 256x256x5x6).

The real problem addressed here is how to exploit all this information. This is not an easy problem. One might apply prior knowledge of course, which leads to what is called a *supervised* segmentation. Knowing the frequency content of the textures in an image may lead to a straightforward filter selection, thereby reducing the image representation by discarding the irrelevant filters. An alternative and still supervised segmentation can

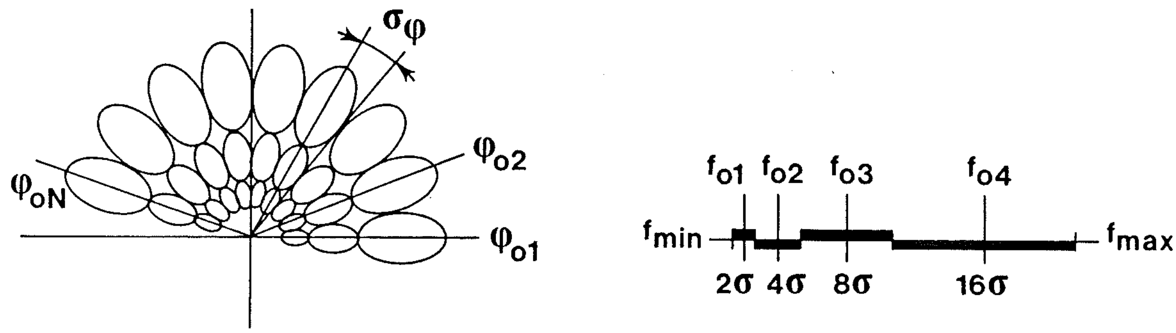

Figure 1: Filter configuration in the spatial spectral domain. In practice we use 5 frequency bands and 6 orientations.

be achieved if we know what textures we have to deal with, but applying the complete filter set. This requires the training of a classifier, like a minimum distance one. Figure 2 shows a test image (left) with texture patches taken from aerial images, together with the result of a minimum distance classification. As can be seen, the centers of all patches have been classified correctly, but the varying filter results at the texture boundaries lead to many subregions. This effect is caused by utilizing the Euclidian distance from the computed class centroids in Gabor space.

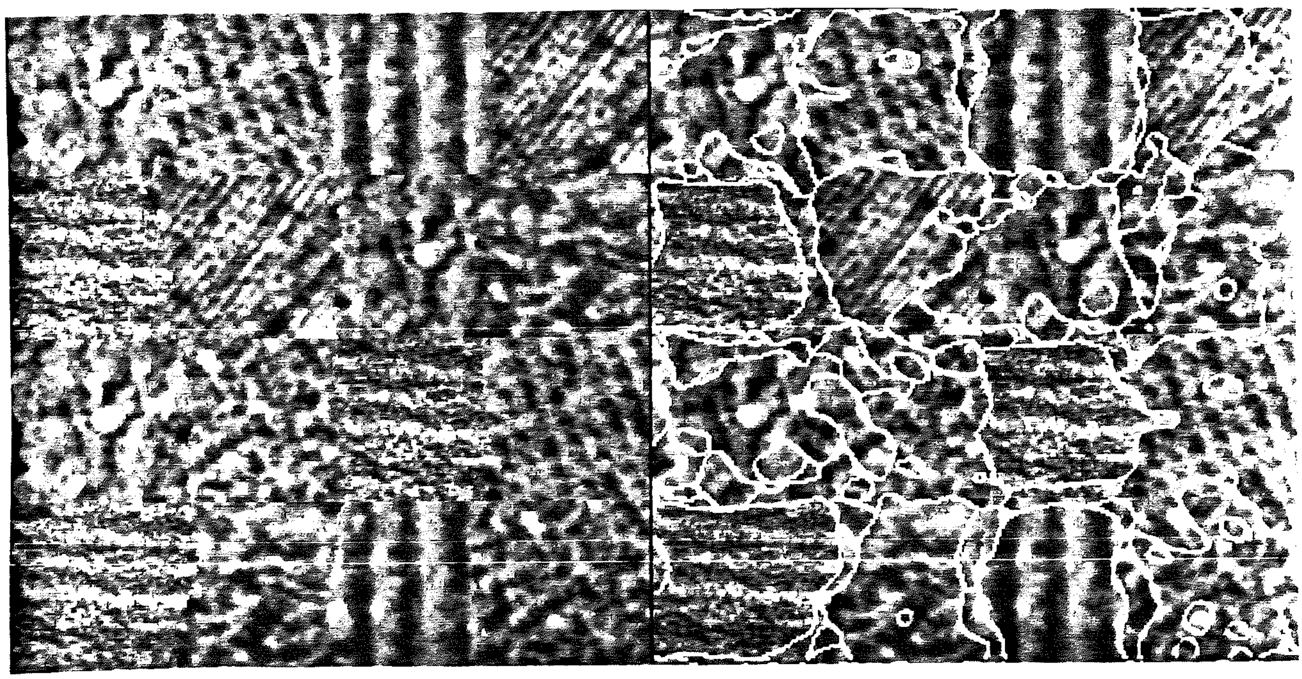

Figure 2: Test image with 7 different textures in 16 patches (left), and a minimum distance classification (right).

In order to circumvent this problem we have to apply a segmentation algorithm to the Gabor representation. Such an algorithm should be able to extract the homogeneous regions on the basis of local texture feature homogeneity, detecting important seed regions and growing these without leaving too many small subregions (although this heavily depends on the homogeneity of the textures - the test images considered here are extremely difficult to segment). In this study we apply a quadtree based segmentation algorithm. In this algorithm a centroid clustering is performed at a fixed level in the tree. The clusters found are applied to classify the textures at that tree level, and the boundaries are projected down, utilizing a boundary refinement procedure at each step going down. This method is rather *unsupervised* and therefore applicable in cases that we don't have prior information about an image. However, this method cannot directly be applied to the Gabor local power spectrum, because a clustering in a 30 dimensional feature space is impossible (it can be done of course but it will give nonsense). The solutions to this dimensionality problem explored here are based on a dimensionality reduction, that is reducing the number of texture features from 30 to 3 or 5, say, which allows for applying the unsupervised segmentation algorithm to the reduced feature set. This process can be seen as a coding of the local power spectrum by means of very few parameters, see Fig. 3. Two codes have been tested and will be compared below: modelling the local power spectrum using a Gaussian with 5 parameters, and computing central moments of the local power spectrum, in which case often 3 uncorrelated moments can be selected.

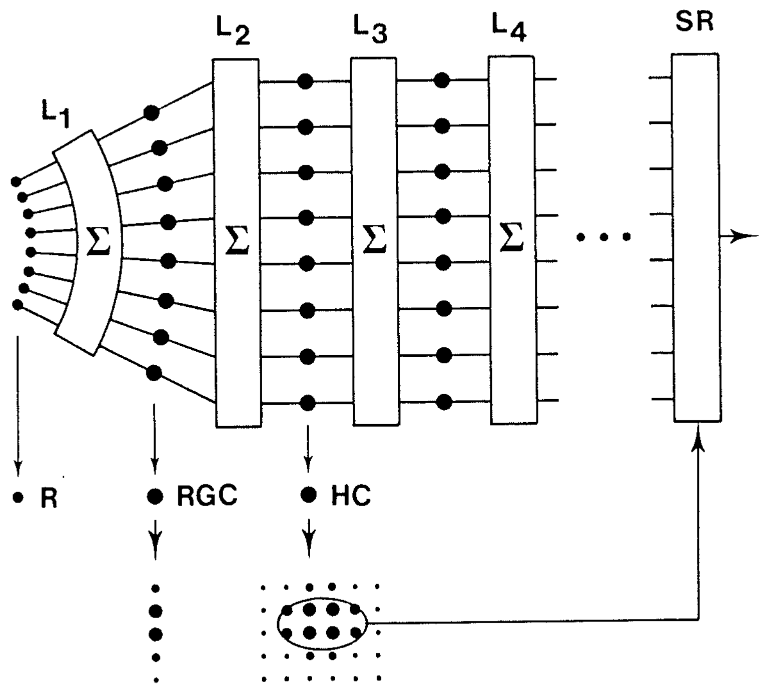

Figure 3: Schematic representation of the first functional layers $L_{1,2,...}$ in the visual system, mainly characterized by linear weighted summation ($\sum$). A hypercolumn (HC) can be seen as a 2D array of filter outputs, organised in frequency and orientation. The information in this array is modelled, the model parameters being forwarded to a semantic representation layer (SR).

The Gaussian model

The local power spectrum $P(x, y, f, o)$ can be seen as a 4D array. At each position we have a subarray $P'(f, o)$, giving the local spectrum in the frequency f and orientation o. This subarray mimicks a cortical hypercolumn (Fig. 3). We now see this array as a 2D rectangular array. Hence, instead of using the f_x and f_y filter coordinates (Fig. 1) we use a $\log(f_r) - \varphi$ coordinate system. The filter results are thus represented in an equidistant 2D lattice. The local power spectrum can be seen as values on a square grid, spanning a surface. This surface can be modelled by applying a 2D Gaussian with only 5 free parameters: an amplitude, two positions (one in the frequency, the other in the orientation), and two standard deviations. The parameters can be computed by means of a least-squares fit. An analysis has shown that the local power spectrum is indeed unimodal for most (real) textures. However, the local power spectrum is periodic in the orientation (filters with orientations φ_{o1} and φ_{oN} in Fig. 1 are neighbours). Hence, the power spectrum array is in fact cylindrical, and the Gaussian should be centered around the maximum in the array. The 5 computed model parameters can be used as texture features and the unsupervised segmentation algorithm can be directly applied to these. The results for two test images are shown in Fig. 4. As can be seen, some textures have not been discriminated and there are still many subregions. However, many boundaries and (sub)regions are perceptually significant. Recall that these test images are extremely difficult to segment, we applied many different texture feature extraction methods to these images, giving very disappointing results even if feature subsets are selected by visual inspection (supervised!), whereas the results shown in Fig. 4 were obtained in a completely unsupervised way.

Central moments

Central moments are wellknown in mechanics and statistics, so we can skip the definition here. These moments can be directly computed from the local power spectrum, but different configurations can be selected. Moments can be computed in the $\log(f_r) - \varphi$ lattice as used with the Gaussian model, but gave inferior results. Alternatively, moments can be computed taking the actual filter positions (f_x, f_y), assuming a Hermitian spectrum (making the power spectrum symmetric with respect to the origin), or we can apply a double angle representation, which solves the 180 degree orientation ambiguity. However, the best results were obtained by applying exactly the same representation as shown in Fig. 1, that is asymmetric and in (f_x, f_y) coordinates. From the moments computed we take only three which are least correlated (this is still supervised, a profound study is performed now). Anyhow, the results are shown in Fig. 5. For the left image all 7 texture classes have been detected, but there are many elongated misclassifications at the boundaries. This effect is very prominent in the right image, where only 6 out of 7 textures have been identified.

Discussion

The results shown in Figs. 4 and 5 are encouraging, but refinements of the methods are required. Although most textures have been detected and the quality of the boundaries

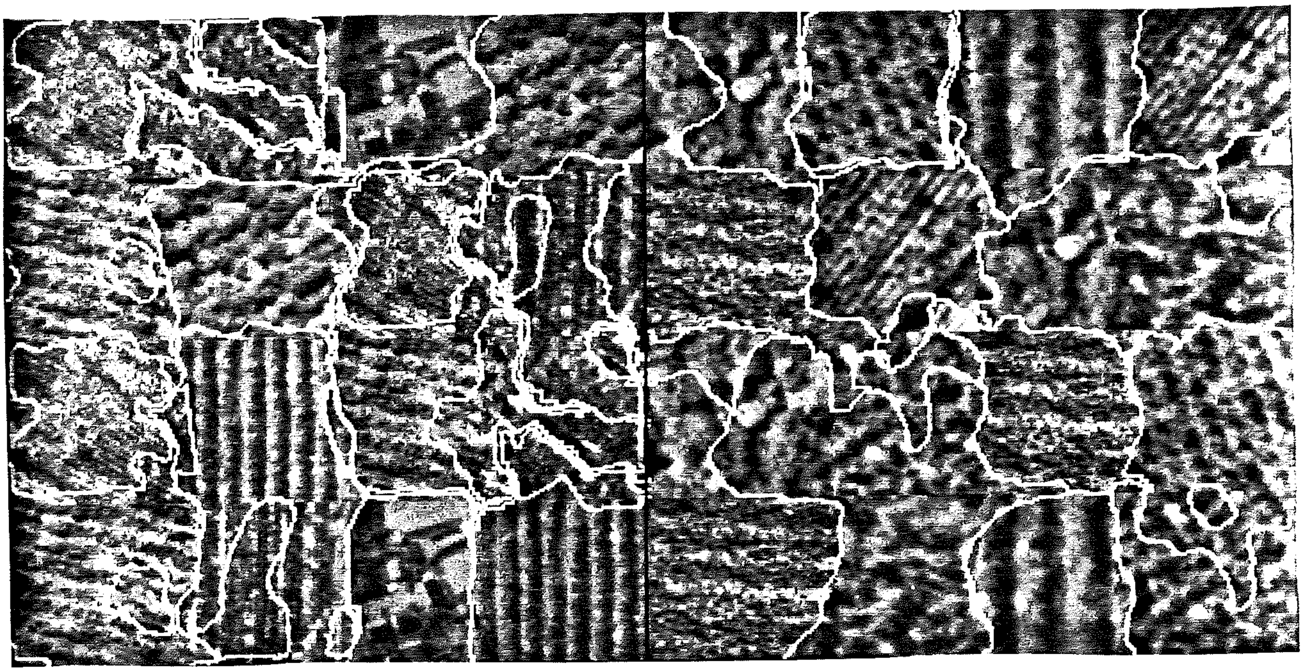

Figure 4: Two test images segmented by applying a Gaussian model to fit the shape of the local power spectrum.

is very good in quite some cases, there are often elongated regions between two textures. This can be seen in all results, but best in the right part of Fig. 5. This effect is similar to the misclassifications which result from applying a minimum distance classifier (Fig. 2). However, this effect is here caused by the sensitivity of certain parameters to the mixture of two local power spectra at a boundary between two textures. This problem is probably a fundamental one, and inherently connected to the modelling of the local power spectrum. More work will be necessary to obtain more insight into this phenomenon and to improve the results. Since we are only working on the algorithmic level until now, practical schemes of neural net like structures have to be elaborated.

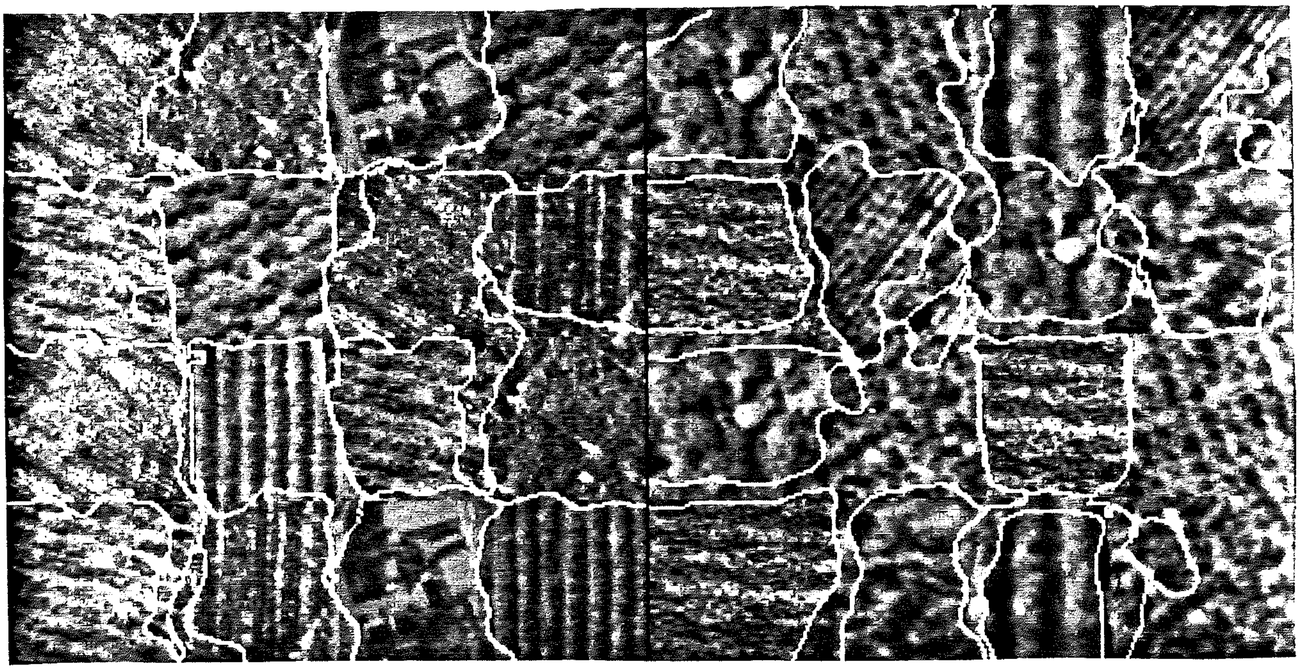

Figure 5: Two test images segmented by computing central moments of the local power spectrum.