Visual Pattern Recognition in Humans

I. Evidence for Adaptive Filtering*

T. Caelli, I. Rentschler, and W. Scheidler
Institut für Medizinische Psychologie der Universität, Goethestrasse 31/1, D-8000 München 2, Federal Republic of Germany

Abstract. We have investigated how observers learn to classify compound Gabor signals as a function of their differentiating frequency components. Performance appears to be consistent with decision processes based upon the least squares minimum distance classifier (LSMDC) operating over a cartesian feature space consisting of the real (even) and imaginary (odd) components of the signals. The LSMDC model assumes observers form prototype signals, or adaptive filters, for each signal class in the learning phase, and classify as a function of their degree of match to each prototype. The underlying matching process can be modelled in terms of cross-correlation between prototype images and the input sample.

1 Introduction

In recent years it has become relatively clear that the internal encoding of image information cannot generally be represented by the basic dimensions (spatial frequencies) and values (amplitude and phase) of the image Fourier transform (e.g., Julesz and Caelli 1979; Badcock 1984; Rentschler and Treutwein 1985). An apparent exception to this is the known sensitivity of the visual system to parameters of Gabor signals which are simply and directly related to their Fourier components (Watson et al. 1983; Daugman 1984). However, we (Caelli and Moraglia 1987) have recently shown how for such simple signals, these Fourier domain parameters (amplitude and phase values for each component frequency) are directly related to luminance modulations as measured by the normalized cross-correlation function between signal pairs. Such results question the directness of the proposed visual encoding of images by Fourier domain descriptors and further suggest that, after all, it is the characteristics of the luminance profiles which are directly encoded by the human visual system.

Since such Gabor signals have Fourier parameter states which directly relate to the perceived modulation of the luminance waveform we have chosen them to study how human observers accomplish pattern recognition or pattern classification in comparison to techniques used in computer pattern recognition. We assume that the input signals are sampled $N \times N$-pixel digital images. Independently of whether these signals are considered in the space domain or as the outcome of an orthogonal transform (e.g., the discrete Fourier Transform) they constitute vectors in a signal space of $N \times N$ dimensions (see Ahmed and Rao 1975, Chap. 10). Consistent with the generally accepted view of perception as a series of processes of data compression and labelling of image structures by the extraction of the salient features, or correlations, successful pattern recognition depends upon the reduction of this dimensionality by capitalizing on such correlations present in the image. The resulting signal space of lower dimensionality is called the feature space and its characteristics are to be determined with the sole purpose of minimizing the corresponding increase in pattern classification errors. From this it is clear that the critical issue for pattern recognition, be it technical or biological, is not the type of representation of the $N \times N$-dimensional input signals, but the assumptions made concerning feature selection and classification.

The signals considered in the present study are compound Gabor signals as used by Caelli et al. (1986) and Rentschler et al. (1987) in their recent studies on the perception of grey-scale textures. Such signals have
intensity profiles defined by
\[ g(x, y) = e^{-\frac{r^2}{2^2}} \left[ \beta_1 \cos(2\pi f x) + \beta_2 \cos(2\pi 3f x + \phi) \right], \]
where \( \alpha \) determines the space constant of the aperture, \((\beta_1, \beta_2)\) correspond to the amplitudes of the first and third harmonic frequencies \((f, 3f)\), while \( \phi \) denotes the phase angle of the third harmonic. An isotropic Gaussian aperture was used where
\[ r = \sqrt{x^2 + y^2}, \]
the signal being centered at a given origin \((0, 0)\). In generating the manifold of image patterns used we have restricted variations of the signals to the \((\beta_2, \phi)\) parameters only, and for the purpose of developing an interpretable (metric) feature space, we have used the equivalent cartesian real (even) and imaginary (odd) feature definitions:
\[ z_1 = \beta_2 \cos \phi \]
and
\[ z_2 = \beta_2 \sin \phi. \]
Given that these signals parameters adequately capture the perceptually salient properties of the luminance profile, in the following experiments we were concerned with evaluating whether human observers form internal prototype and classify signals according to their proximity to each class prototype. Proximity is usually determined in either the feature space – as in the Minimum Distance Classifier (MDC), or in the decision space – as in the Least Squares Minimum Distance Classifier (LSMDC, see Ahmed and Rao 1975, Chap. 10).

Both classification strategies involve the positioning of class samples in a feature space where each feature dimension corresponds to a given perceptually encoded signal property. Prior training (usually in the form of supervised learning) is used to construct prototypes in two ways. One, for the MDC case, a class prototype is determined as the centroid of the sample cluster. That is, we define pattern class \( C_i \) as set of \( n \)-dimensional feature vectors
\[ C_i: \{Z_{i1}, ..., Z_{in}\}, \]
where the class prototype is defined by
\[ Z_i = \frac{1}{k} \sum_{j=1}^{k} Z_{ij} \]
for \( k \) samples in \( C_i \). Classification of any new input pattern is then made according to which class prototype is closest in feature space.

The second LSMDC strategy does not assume that the centroid is the best estimate of the class prototype – particularly when the distribution of samples in feature space lacks symmetry and compactness. In this case samples are mapped into a decision space wherein the patterns belonging to each class \((C_i)\) are made to cluster about prototype locations \( V_i \) \((i = 1, ..., m; m = \text{number of classes})\) usually corresponding to vertices of a \( m \)-dimensional hypercube. The LSMDC technique consequently maps samples from a \( n \)-dimensional feature space into a \( m \)-dimensional decision space while minimizing the within-variance of the samples about the class vertex in decision space. That is, the LSMDC seeks a matrix \( A \) which maps \( \{Z_{ij}\} \) into \( V_i \) so that the mean-square error in the mapping is minimized. For \( L_{ij} \) being the mapping of \( Z_{ij} \), then the error vector is
\[ e_j = L_{ij} - V_i = AZ_{ij} - V_i, \]
and so the total mean-square error is
\[ \varepsilon = \frac{1}{N_i} \sum_{j=1}^{N_i} ||e_j||^2, \]
where \( N \) corresponds to the number of samples in class \( i \). That is, we wish to minimize \((5)\) with respect to \( A \). For convenience we use the “augmented feature space”
\[ \tilde{Z} = \{Z_1, Z_2, ..., Z_{nt} - 1\} = \{Z', -1\} \]
which results in (see Ahmed and Rao 1975, Chap. 10 for details)
\[ A = S_{YZ} S_{ZV}^{-1}, \]
where
\[ S_{YZ} = \sum_{i=1}^{m} \sum_{j=1}^{N_i} \frac{P_i}{N_i} (V_i Z_{ij}), \]
\[ S_{ZZ} = \sum_{i=1}^{m} \sum_{j=1}^{N_i} \frac{P_i}{N_i} (Z_{ij} Z_{ij}'). \]
\( p_i \) is the a priori probability of class \( i \), and \( Z_{ij}' \) the transpose of \( Z_{ij} \). Each prototype is then determined by the sample \( Z_i \) which is mapped to the vertex \( V_i \) in decision space. That is, the prototype \( Z_i \) for class \( C_i \) is defined by
\[ A \times Z_i = V_i, \]
where \( A \) is determined from the sample covariances \([6]; \text{see Ahmed and Rao (1975)}\). The centroid in feature space, used in the MDC formulation, is similar to this prototype formation rule when the samples are symmetrically distributed about the centroid in feature. Otherwise this latter definition of the prototype is more appropriate since it is derived as the best fitting sample in the least squares sense [see (4) and (5)].

Again, class membership is determined by the proximity of samples to vertices (prototypes) in decision space, using the matrix \( A \) as estimated from the