Pattern Recognition, Vol. 21, No. 6, pp. 639–650, 1988.
Printed in Great Britain.

0031-3203/88 $3.00 + .00
Pergamon Press plc
Pattern Recognition Society

FILTER-BASED MODELS FOR PATTERN CLASSIFICATION

TERRY M. CAELLI*†, WALTER F. BISCHOF‡ and ZHI-QIANG LIU‡

* Department of Psychology, Queens University, Kingston, Ontario, Canada K7L 3N6 and ‡ Centre for Machine Intelligence and Robotics, The University of Alberta, Edmonton, AB T6G 2E9, Canada

(*Received* 29 *April* 1987; *in revised form* 9 *February* 1988)

Abstract—In this paper we consider a technique for pattern classification based upon the development of prototypes which capture the distinguishing features ("disjunctive prototypes") of each pattern class and, via cross-correlation with incoming test images, enable efficient pattern classification. We evaluate such a classification procedure with prototypes based on the images *per se* (direct code), Gabor scheme (multiple fixed filter representation) and an edge (scale space-based) coding scheme. Our analyses, and comparisons with human pattern classification performance, indicate that the edge-only disjunctive prototypes provide the most discriminating classification performance and are the more representative of human behaviour.

Pattern classification Filters Scale space Edges Disjunctive prototypes
Cross correlation

1. INTRODUCTION

Pattern recognition or classification methods, which have evolved over the past 20 years, usually involve three processes: (a) pattern encoding or *feature selection*, (b) supervised learning where *prototypical descriptions* of pattern classes are produced within the chosen feature space, and (c) *classification* of new input samples in terms of their proximity to class prototypes.[1] A variety of *explicit* features have been evaluted over the decades—varying from perceptron-type feature analyzers [edge, bar detectors; e.g. Ref. (2)] to components of orthogonal transforms such as Fourier, Hadamard and Walsh, and global features (moment parameters, etc.).[3] Such mappings of patterns from the $n \times m$ dimensional input image space to feature space have proved to be relatively inefficient, since the number of features (feature space dimensions) required to retain the *uniqueness* of input patterns is large given the overhead of the transformations involved. Further, most of these feature extraction techniques are limited to specific types of patterns and do not allow for pattern matching invariant to position, size and orientation of the input.

With recent developments in computational vision and the availability of fast parallel pixel processors, a number of new approaches to feature selection have emerged which pose new ways of developing prototypes and classifying images invariant to specific transformations. In this paper we concentrate upon what we refer to as *implicit* descriptions of patterns based upon three different types of image representations of recent interest to human and computer vision research. All three representations are based on attempts to arrive at pattern prototypes *defined as images* rather than centroids in feature spaces. In this sense the *implicit features* are extracted by filters and the prototypes themselves are viewed as adaptive filters which are designed to optimize the classification processes.

These three different representations are as follows. (1) The "direct scheme" (DS) where it is assumed that the appropriate feature space is the $n \times m$ dimensional vector space of the original image. As will be seen, in this method prototypes are generated as combinations of appropriately aligned samples and classification is accomplished through the development of disjunctive prototypes generated by comparing across classes. (2) A "Gabor scheme" (GS) is discussed and illustrated. Here the input samples are decomposed via sets of (fixed) orientation/size specific (Gabor) filters. Such filtered versions are aligned across samples and classification is accomplished via how images match each prototype with respect to the similarities between their associated filter responses. Though such filters are proposed to represent the types of decompositions which occur within the primary projection area of the vertebrate visual cortex,[4,5] it is clear that this scheme, in its linear form, can perform no better than the DS scheme, due to the linearity of the matching (correlation) process as enacted piece-wise via each filter.

It is relatively obvious that efficient pattern classification techniques should be based upon feature encoders which differentiate between classes as much as possible while still capturing the common features within given class samples. Techniques based directly upon pixels or filtered images have the disadvantage of not suppressing regions of common contrast between classes. Indeed, by using image

† To whom correspondence should be addressed.

encoding techniques which emphasize the regions of contrast change, as in edge-only images, it is possible to further differentiate between pattern classes. The problem remains, however, how to obtain a reasonable definition of image edges. In our third implicit feature representation scheme, the "edge scheme" (ES), we have used two-dimensional "scale space"[6] to extract an edge mapping of the input signals. Here scale space refers to the location of zero (or level) crossings in an image as a function of many different isotropic bandpass filters (or "scales") defined by $\nabla^2 G$ kernels (as before). We (see Discussion) will also generalize this scheme to what we call "extended scale space" (ESS) where images are described by edges and the luminance gradient at each zero-crossing contour.

As already mentioned, all three methods retain the image as the "feature space". However, only "critical regions" of this space are selected for correlational consideration as a function of the signals being processed and the registration method used. Further, such regions are presumed to be more relevant to classification if they index *structural* relationships or properties of the imaged object and are stable under noise perturbations. In this sense we propose a model for pattern recognition which is not dependent on prior known feature lists or transformation methods. Rather, we supply a set of computationally efficient *algorithms* which adaptively generate "critical features" for classification as outputs of generic classes of filters and comparison methods.

2. COMPUTATIONAL PROCEDURES

Figures 1–4 define, in general terms, the three processes of feature extraction, prototype formation and pattern classification as viewed from the three different encoding models. That is, we assume a supervised learning environment where the recognition system has a data base of samples for each class and is able to construct "prototypes" via the processes

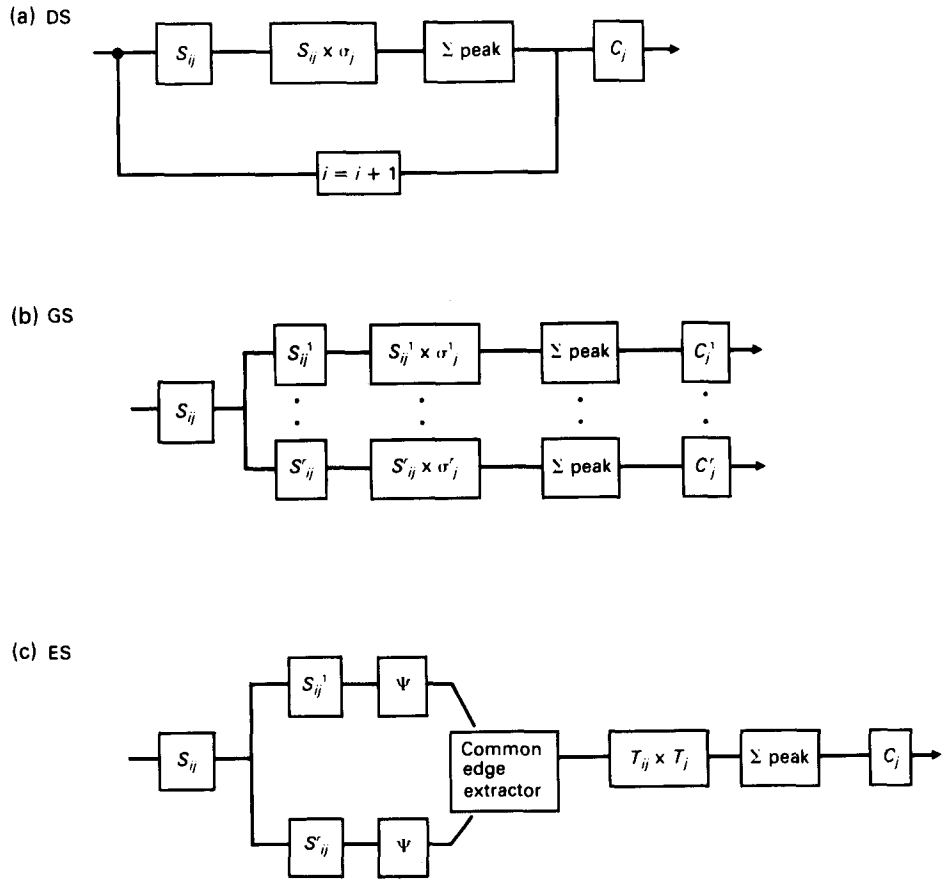


Fig. 1. Three prototype learning schemes (stage I) as determined by their representations of the input samples (S_{ij}: sample i of class j) are shown here. (a) Direct scheme (DS)—class prototypes (C_j) are formed via the direct sum of samples aligned by a normalized cross-correlation $\otimes$ criterion; σ_j refers to the reference sample for class j. (b) Gabor scheme (GS)—as above, except that a prototype image (C_j^k is formed for each filter (k). S_{ij}^k refers to the kth filtered version of sample i of class j, while σ_j^k refers to the kth filtered version of the reference sample for class j. (c) Stability scheme (ES)—as in (a) except "stable edge" versions (T_{ij}) are produced for each sample in terms of the accumulated evidence for edges over various scales (see text) at each position. Here Ψ refers to the non-linear process of registering the zero (or level) crossings of the filter outputs; T_j refers to the reference sample. In all three cases Σ peak refers to the accumulation of samples via shifts corresponding to the peak of the sample-to-reference cross-correlation.

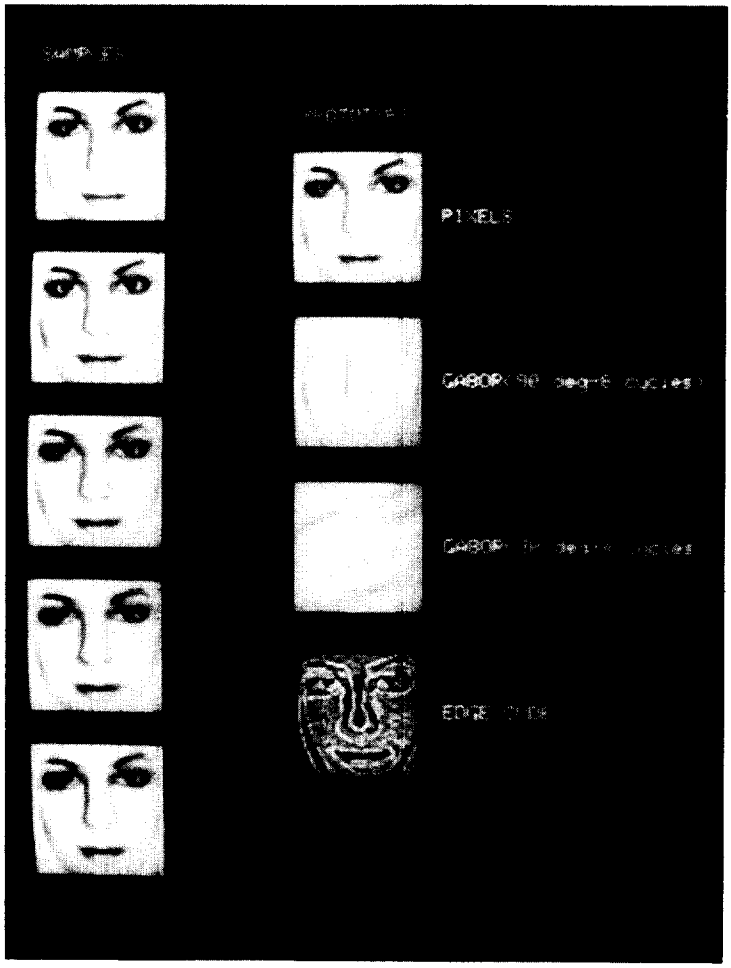


Fig. 2. Examples of prototypes emerging from the three encoding processes for five input samples for a given class. Here the resultant prototypes are derived by the alignment and summation procedures defined in Fig. 1 and examples of pixel-by-pixel (direct alignment), Gabor (vertical and 30° oriented filters), and stability analysis (edge encoding) are shown.

defined in Fig. 1. Here the prototypes are analogous to centroids of sample classes in feature space and, in this case, correspond to one image, or a set of images, per class. Each of these processes will be discussed in the following sections. However, it should be noted that the prototype formation and classification processes vary as a function of the feature encoding system. We also assume, for the purposes of illustration, that the sample patterns are already matched for orientation and size though the system can be extended to automatically accommodate for such differences (see Section 3 for details).

2.1. *Feature encoding and initial prototype formation processes*

In the DS scheme it is assumed that the feature space is defined by the image itself (of dimension $n \times m$) and that the feature "strengths" are determined by the input pixel intensities. Under this model initial estimates of the class prototypes are formed from the sum of individual samples aligned with respect to the peak of their normalized cross-correlation with a reference sample σ_j. That is, the prototype image for class j, C_j, is determined by:

$$C_j(x, y) = \frac{1}{m_j} \sum_{i=1}^{m_j} S_{ij}(x - \alpha_i, y - \beta_i) \tag{1}$$

where S_{ij} is the ith sample of class j, (α_i, β_i) corresponds to the position of the peak value of $S_{ij} \otimes \sigma_j$, and

$$\mathrm{O} \leq S_{ij} \otimes \sigma_j = \frac{\int_\beta \int_\alpha S_{ij}(x + \alpha, y + \beta)\, \sigma_j(\alpha, \beta) d\alpha d\beta}{\left[\int_\beta \int_\alpha S_{ij}^2 (x + \alpha, y + \beta) d\alpha d\beta \int_\beta \int_\alpha \sigma_j^2(\alpha, \beta) d\alpha d\beta \right]^{1/2}} \leq 1.0, \tag{2}$$

with equality, if and only if $\sigma_j = \lambda S_{ij}$. m_j refers to the number of samples in class j. It is well known [see Ref. (7)] that such normalized cross-correlations align or match patterns according to their "shape" similarity compared to the direct cross-correlator [numerator in equation (2)] which can be contaminated by local luminance offsets unrelated to shape. Such a process is, by definition, shift invariant (Fig. 1a).

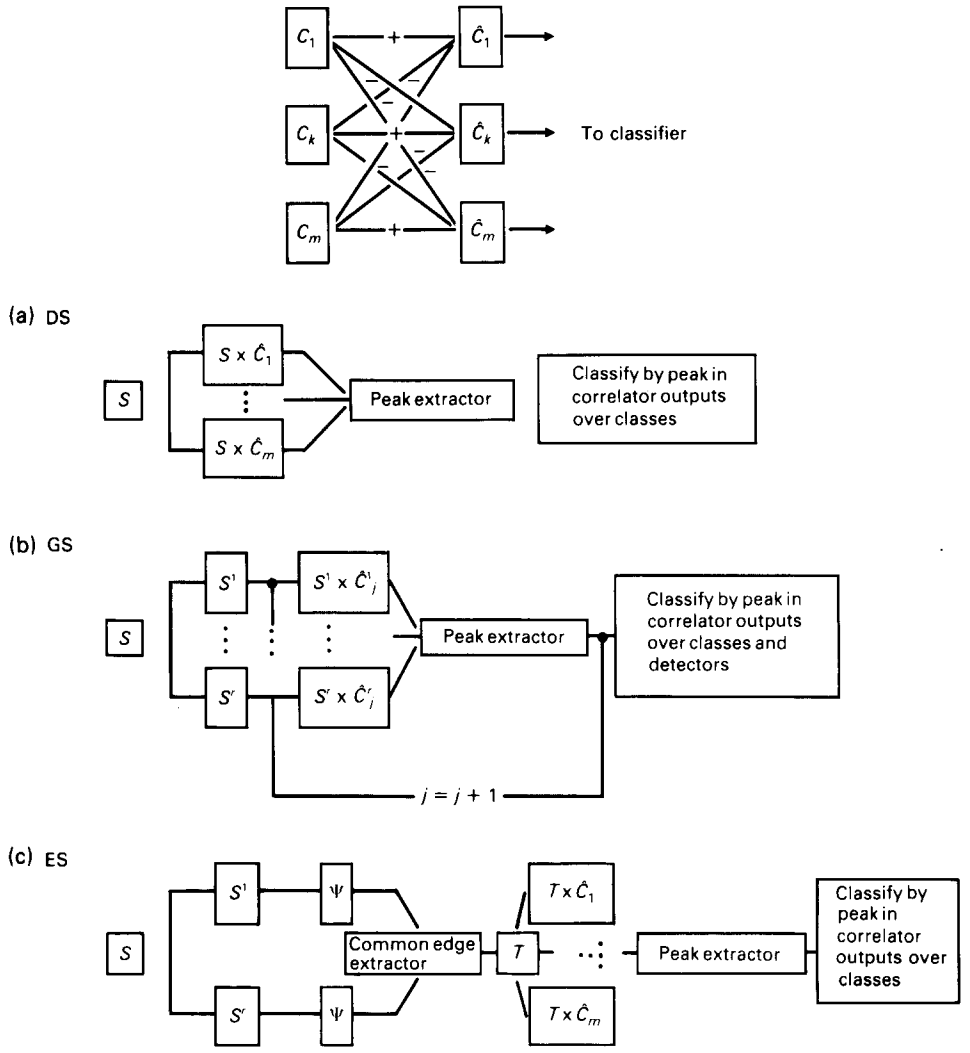


Fig. 3. Stages II and III of the recognition system are given here. (a) Stage II — involves an orthogonalizing "disjunctive" process where new prototypes are formed which emphasize the differentiating features of each class. (b) Stage III — classification: (i) direct scheme — new input samples (S) are classified according to which prototype produces the greatest peak in the associated normalized cross-correlation image(x); (ii) Gabor scheme — as in (a) except that peak outputs over the r filters are combined; (iii) stability scheme — as in (a) except the samples are converted to "edge-only" images as in Fig. 1c.

The GS scheme (Fig. 1b) differs from the DS scheme insofar as each class sample (S_{ij}) is decomposed by a set of r filters of specified spatial, spectral resolutions and bandwidths. From these images r different prototypes are formed for each class via the alignment process described by (1) and (2) above. Each "Gabor filter" is determined by a two-dimensional form of even elementary Gabor signals:[4]

$$g(x, y) = e^{-[(x-x_0)^2 + (y-y_0)^2]/\alpha^2}.\cos[2\pi(u_0 x + v_0 y)]. \quad (3)$$

Here the isotropic spatial aperture is determined by α (the space constant), positioned at (x_0, y_0) of center spatial frequency (u_0, v_0), having a band width of $1/\alpha$ in picture cycles.

The ES scheme illustrated in Fig. 1c differs from the above (independent) model insofar as outputs of each filtering/thresholding operation are combined and then aligned to form only one prototype per class: a "conjunctive" process. The purpose of this is to capture features common to a set of filters, in particular, edge features such as zero-crossings. Termed "spatial stability" analysis,[8] this process is aimed at capturing more salient image edges by determining the degrees to which the same zero-crossing positions of band-pass filters [in particular, $\nabla^2 G(\underset{\sim}{X}, \sigma)$ filters] occur over a range of filters (or "scale") values.

More formally, a sample image S_{ij} is initially decomposed by a set of band-pass $\nabla^2 G(\underset{\sim}{X}, \beta)$ filters to result in outputs $S_{ij}^k(\underset{\sim}{X})$ by:

$$S_{ij}^k(\underset{\sim}{X}) = \nabla^2 G(\underset{\sim}{X}:\beta_k) * S_{ij}(\underset{\sim}{X}), \text{ for } \underset{\sim}{X} = (x, y) \quad (4)$$

$$= \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right](\beta_k\sqrt{2\pi})^{-1}$$

$$\exp\{-r^2/2\beta_k^2\} * S_{ij}(\underset{\sim}{X}),$$

* denoting convolution. These operations produce the

r images S_{ij}^1–S_{ij}^r shown in Fig. 1c. From these outputs we extract the zero-crossings to result in a broad sample of the image "scale-space" (see Ref. (6) for more details). That is, the transformation process Ψ shown in Fig. 1c corresponds to the selection of sample filter points in the $S_{ij}^k(\underset{\sim}{X})$ images where:

$$S_{ij}^k(\underset{\sim}{X}) = 0, \text{ and } \nabla S_{ij}^k(\underset{\sim}{X}) \neq 0, \text{ for } \nabla \equiv \frac{\partial}{\partial x} + \frac{\partial}{\partial y}. \quad (5)$$

We then define the "spatial stability" image as that corresponding to the number of zero-crossings which occur at a given pixel over a range of β_k values. We use this latter term since highly stable image edge points must correspond to zero-crossings over a large number of scales [β in equation (4)]—consistent with the fact that this *local* region must have a broad (local) Fourier power spectrum indicative of high contrast and steep gradient—the ingredients for distinct edge information. Secondly, this definition of *stable* edges minimizes the occurrence of edges in white noise alone. This follows from the observation that, for sets of isotropic (even) filters $\{f_1 \ldots f_r\}$ the cross correlation between their outputs, to white noise inputs, is determined by:

$$C(\underset{\sim}{X}) = \sum_{\underset{\sim}{u}} F_k(\underset{\sim}{u}) F_l(\underset{\sim}{u}) \cos(\underset{\sim}{u}.x), \quad (6)$$

where F_k and F_l correspond to the Fourier power spectra of filters f_k and f_l ($\underset{\sim}{u}$ refers to spectral co-ordinates: spatial frequencies). Consequently the correlations between pixels is specifically dependent on the inner product or spectral separation between the filter power spectra. However, as the scales (β) of the $\nabla^2 G(\underset{\sim}{X}, \beta)$ operators differ their correlation decreases, so decreasing the probability of consistent zero-crossing points (see Ref. (8) for more details). If the image is uncorrelated, as in white noise images, then the likelihood of common zero-crossings must decrease as the filters become more spectrally separated. Conversely, the method used here to define stable edges is one way of depicting correlations over scales within the image. Such encoding characteristics of this scheme are not inherent in the direct and Gabor encoding schemes.

A prototype for each class is then formed by summing each T_{ij} with respect to the relative position of each sample's normalized cross-correlation peak [equation (2)], to result in the class j prototype:

$$C_j(\underset{\sim}{X}) = \sum_{i=1}^{m_j} T_{ij}(x - \underset{\sim}{\alpha}_i), \quad (7)$$

for m_j samples in class j. Again, normalization is necessary in order to decrease false alignments based upon higher edge densities uncorrelated with the reference edge information, according to (2).

Figure 2 shows examples of the prototypes generated by the three different encoding models though only the vertical and 30° orientated Gabor filter prototypes are shown. The direct scheme prototype was defined from (1) and (2) to result in an image representative of the (aligned) "average" image from the samples.

In the Gabor scheme we have assumed that each Gabor signal [see equation (3)] is defined in every image pixel and so we could enact the decomposition by filtering in the Fourier domain. Here equation (3) assumes the spectral form:

$$G(\underset{\sim}{u}) = e^{-\alpha^{2(\underset{\sim}{u} - \underset{\sim}{u}0)^2}} \quad \text{for } \underset{\sim}{u} \equiv (u, v) \text{ and } \underset{\sim}{u}_0 = (u_0, v_0). \quad (8)$$

We have employed 18 such filters of three different centre frequencies and six different orientations: equally spaced in 30° increments from centre-to-centre. For an input 128 × 128 pixel format the centre frequencies and bandwidths were $\{8 \pm 4,\ 16 \pm 8,\ 32 \pm 16\}$ picture cycles, for each orientation component. Such filters are consistent with orientation detector profiles recorded by electrophysiologists in the primary projection area of the vertebrate visual cortex.[4,5]

For the edge scheme (Fig. 1c) we have implemented equations (4), (5) and (7) with respect to 17 different β values ranging from 1 to 16 in $\frac{1}{4}$ octave steps to produce the prototype shown in Fig. 2, where edge strength is denoted by intensity.

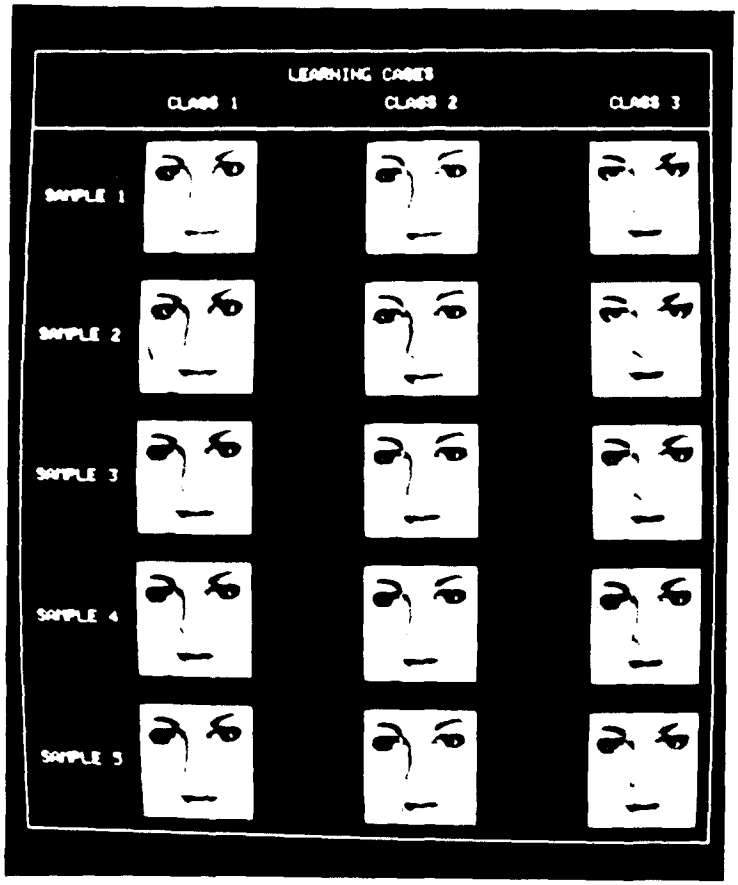
The five samples shown in Fig. 2 were generated by taking linear combinations of three basic faces with weights shown in column 1 of Fig. 4b. That is, the five samples corresponding to class 1 (Fig. 4a, b) were constructed by the choice of five different weights of the three basic faces. For example, sample 1 of class 1 consisted of 0.6 of face 1, 0.3 of face 2 and 0.1 of face 3. Clearly class 1 was biased towards a prototype resembling face 1 (of the three basic shapes) relative to the other two. These basic faces are shown in the top row of Fig. 6a (see weights in Fig. 6b). As will be seen, we have employed such small differences within and between pattern classes to challenge both models and human observers.

2.2. *Prototype disjunction: orthogonalization of filters*

As outlined in the Introduction, these representation schemes for pattern classification are based upon the notion of *implicit* description of discriminating class features in terms of filter operations on images. That is, class prototypes formed in the supervised learning stage (as in Figs 1 and 2), are viewed as images which are to be matched, by cross-correlation, with new inputs for classification. The "implicitness" refers to the fact that the critical features are contained within the image: the features are located at various positions without using an explicit feature list.

However, if classification is to be accomplished via such a matching processs, then the degree of class differentiation is determined by the orthogonality of the prototypes (being adaptive filters). Here two prototypes are *orthogonal* via the usual zero scalar product criterion:

(a)



(b)

	Class 1	Class 2	Class 3
Sample 1	.6, .3, .1	.3, .6, .1	.3, .1, .6
Sample 2	.7, .2, .1	.2, .7, .1	.2, .1, .7
Sample 3	.5, .25, .25	.25, .5, .25	.25, .25, .5
Sample 4	.4, .3, .3	.3, .4, .3	.3, .3, .4
Sample 5	.5, .3, .2	.3, .5, .2	.3, .2, .5

Fig. 4. Samples (difficult cases) used in the training sequence for a three-class classification task. They were generated via linear combinations of three faces having coefficients defined in (b).

$$C_i \perp C_j \Leftrightarrow \int_{\underset{\sim}{x}} C_i(\underset{\sim}{x}) C_j(\underset{\sim}{x}) d\underset{\sim}{x} = 0. \quad (9)$$

We can further define *absolute orthogonality* (A) in terms of the two prototypes having shift invariant zero-scalar product, or:

$$\left\{ \forall \underset{\sim}{\alpha} \int_{x} C_i(\underset{\sim}{x}) C_j(\underset{\sim}{x} + \underset{\sim}{\alpha}) d\underset{\sim}{x} = C_{ij}(\underset{\sim}{\alpha}) \equiv \underset{\sim}{0} \Leftrightarrow C_i \neq C_j. \right. \quad (10)$$

By the convolution theorem we know that equation (9) holds if, and only if, C_i and C_j have Fourier spectra which are perfectly disjoint. That is:

$$\forall \underset{\sim}{u}, A_i(\underset{\sim}{u}) . A_j(\underset{\sim}{u}) = 0, \quad (11)$$

where A_i corresponds to the Fourier amplitude (or power) spectrum of C_i.

The prototype formation process, so far defined, does not guarantee orthogonality (in either sense) of the filters insofar as they only attempt to represent common features within the class samples. One simple way of "orthogonalizing" the prototypes is to compute the class filter by a "disjunctive" process such as:

$$\hat{C}_j(\underset{\sim}{x}) = \max \left\{ 0, C_j(\underset{\sim}{x}) - \frac{1}{n-1} \sum_{\substack{k=1 \\ k \neq j}}^{n} C_k(\underset{\sim}{x} - \underset{\sim}{\delta_{jk}}) \right\}, \quad (12)$$

where δ_{jk} is the shift corresponding to the peak of the (normalized) cross-correlation between prototypes j and k, as defined by equation (2) (see Fig. 3a). Such disjunctive operations, which emphasize prototype differences, constitute a simple method of introducing a degree of orthogonality into the class filters, via (9) and (10) while still preserving the important image features. This process would result in a greater difference between the peaks in the cross-correlations between prototypes and samples from each class.

2.3. *Classification stage: pattern recognition*

This final classification stage is illustrated in Fig. 3b where the class membership (pattern recognition) is determined by the class prototype with the highest peak cross-correlation. In the case of multiple (independent) decompositions, as in the Gabor scheme (GS), classification is made according to a norm defined over the associated peak cross-correlation vector. Further, the likelihood of class membership (or relative strength of membership) would be determined from the comparison of cross-correlation peaks over the classes. Here, again, normalization via equation (2), is required. More formally, the class membership likelihood vector L_j for an input sample S is defined by:

$$L(j) = \left\{ \left[\sum_{k=1}^{r} [\max_{\sim} \{ S \otimes \hat{C}_j^k \}]^2 \right]^{1/2} \right\}, \quad (13)$$

for n classes and r filters per class (for DS and ES $r = 1$). $\hat{C}_j^k$ corresponds to the kth disjunctive filter (prototype) for class j, $\otimes$ to normalized cross-correlation.[2]

Before describing some test examples, an important point should be noted about the GS process. In its linear formulation it is clear that the GS scheme can differentiate *no better* than the DS scheme since the Gabor filters cannot capture any more features than those contained in the original image, given that we use cross-correlation (matched filtering) as the comparison process. For this reason we have not *directly* included this scheme in our simulations. Rather, we have included some discussion of the scheme since it is of interest to biological spatial vision (see Introduction). Further, as will be seen in Section 3, this scheme has interesting properties for matching invariant to rotation, shift and size differences between samples and prototypes (see Zetsche and Caelli[13]).

Consequently we have run the DS and ES feature extraction schemes over the prototype and classification procedures with a number of face types with varying within- and between-class differences. Figure 4 shows the most difficult case involving three classes whose samples are quite similar. Such faces were generated from linear combinations of three basic faces. Subtle mouth, nose and eye differences between classes were produced by the different weighting coefficients shown in Fig. 4b. Figure 5 shows prototypes and disjunctive prototypes formed for the DS and ES schemes as determined by equations (3) and (11).

Figure 6a shows a set of new samples for classification via the prototypes shown in Fig. 5. They were also generated by linear combinations of the same three basic faces used to form the prototypes with linear combination coefficients as shown in Fig. 6b. The performance of DS, ES and their direct and disjunctive forms are shown in Figs 7a and b. Though a variety of criteria can be used to evaluate the performance of such pattern classification schemes, we have compared these predictions with human performance by training observers to 100% learning criterion and then presenting the new patterns a large number of times to determine the likelihood of classification into one of the three classes.

In more detail, three observers, with normal acuity, were presented with the training set (Fig. 4a) until they could be perfectly classified. In this learning phase a given experimental trial consisted of presenting each training sample for 200 ms on an Electrohome 12" monitor, at a visual angle of 2°, space average luminance of 18cd/m^2 and 0.8 contrast $(l_{max} - l_{min})/(l_{max} + l_{min})$, where l_{max} and l_{min} correspond to maximum and minimum luminances of the image. Immediately after the offset of the sample they were informed as to the class to which the sample belonged via the display (for 1 second) of the roman numerals I, II or III. All samples were randomized over trials. After this sequence they were tested under identical display conditions, except that they were required to respond as to the class of the sample. This procedure continued till 100% correct classification performance occurred. The test phase was similar to the latter (probe) component of the learning phase though involving new patterns each of which was displayed 50 times over trials per observer to obtain a classification

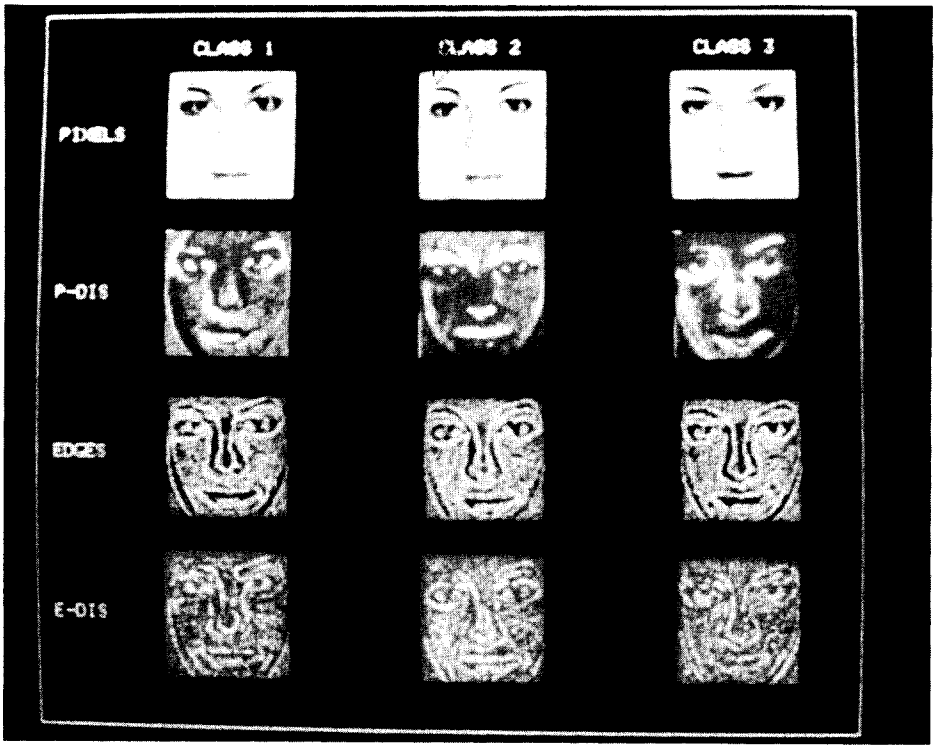


Fig. 5. Initial (rows 1, 3) and disjunctive (rows 2, 4) prototypes formed for the samples shown in Fig. 4 by the alignment process with respect to the direct (rows 1, 2) and edge (rows 3, 4) encoding schemes.

preference or choice frequency score. As shown in Fig. 6, these new faces were generated from the same basic images though using different weighting coefficients (see Fig. 6b).

Since observers showed little variation in responses (specifically, a mean standard error of estimate, MSE, of 6.7 over all examples), we have combined their data in Fig. 7c to illustrate, more clearly, performance in comparison to the models (Figs 7a and b). What, of course, is interesting is the consistency between class membership determined by the disjunctive edge prototypes and human performance, relative to that of the direct (and so, Gabor) scheme. Indeed, observers consistently reported that they accomplished the task by looking for facial features which distinguished one class from another, usually in the eye, nose and mouth region. Since the edge disjunctive prototypes both capture the salient features within a class and the discriminating features between classes in an *implicit* image form, it is not surprising that they do reflect the types of internal representations employed by observers in classification tasks involving such precision. We have correlated all the predicted disjunctive prototype matching values (Fig. 7b) with observed behaviour to result in a Pearson's correlation coefficient of $r = 0.89$, so explaining 79% of the preference results, a correlation clearly significantly greater than zero ($P < .01$). However, since the other models (Figs 7a and 7b (dotted lines)) did not show any systematic relationship to observer responses we have omitted further statistical comparisons.

3. EVALUATION AND EXTENSIONS

In the above procedures we have used the normalized cross-correlator[2] as a criterion for the development of *implicit* feature descriptions of class prototypes, their differentiating (disjunctive) features, and pattern classification. We have found that the disjunctive edge scheme proved to be the most discriminating type of prototype for classification. This follows from the observations that edge codes emphasize the distinguishing pattern features of the class and the disjunctive process emphasizes the unique features within a class—relative to others. Further, it is this code which seems to best parallel human performance.

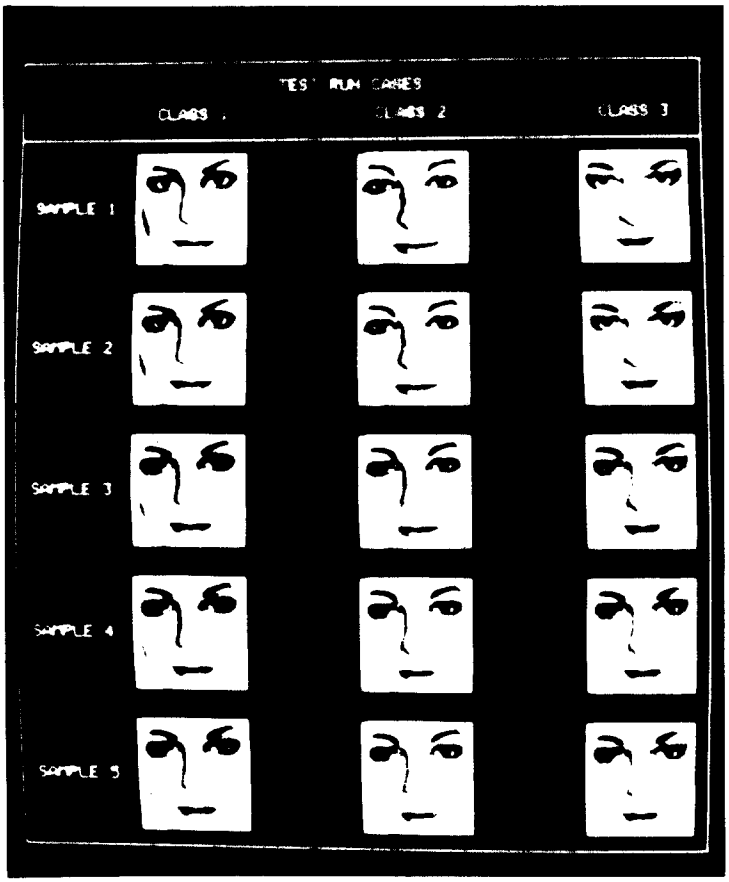
Such procedures differ from more traditional pattern classification techniques insofar as the following.

(i) The underlying feature space consists of the image itself and critical features are depicted by various filtering and thresholding processes which are adaptive to within- and between-class comparisons.

(ii) Similarity is defined by normalized cross-correlation, and the optimal pattern classifier is determined by a set of adaptive filters determined by within-class conjunction and between-class disjunction operations. The disjunctive technique endeavors to orthogonalize the filters (prototypes) as much as possible without detracting from the shape characteristics involved.

(iii) The classifier is analogous to the Least Squares Minimum Distance Classifier[1] insofar as the classification vector's dimension is equal to the number of

(a)



(b)

	Class 1	Class 2	Class 3
Sample 1	1.0, 0.0 0.0	0.0, 1.0 0.0	0.0, 0.0 1.0
Sample 2	.75, .15, .1	.1, .75, .15	.15, .1, .75
Sample 3	.55, .3, .15	.3, .55, .15	.15, .3, .55
Sample 4	.45, .35, .2	.2, .45, .35	.35, .2, .45
Sample 5	.45, .25, .3	.25, .45, .3	.3, .25, .45

Fig. 6. (a) Examples of new sample images used to evaluate the different classification methods. Here the prior allocation of samples to classes was done according to their linear weightings with respect to the basic three faces used to generate all images. Linear weighting coefficients are shown in (b).

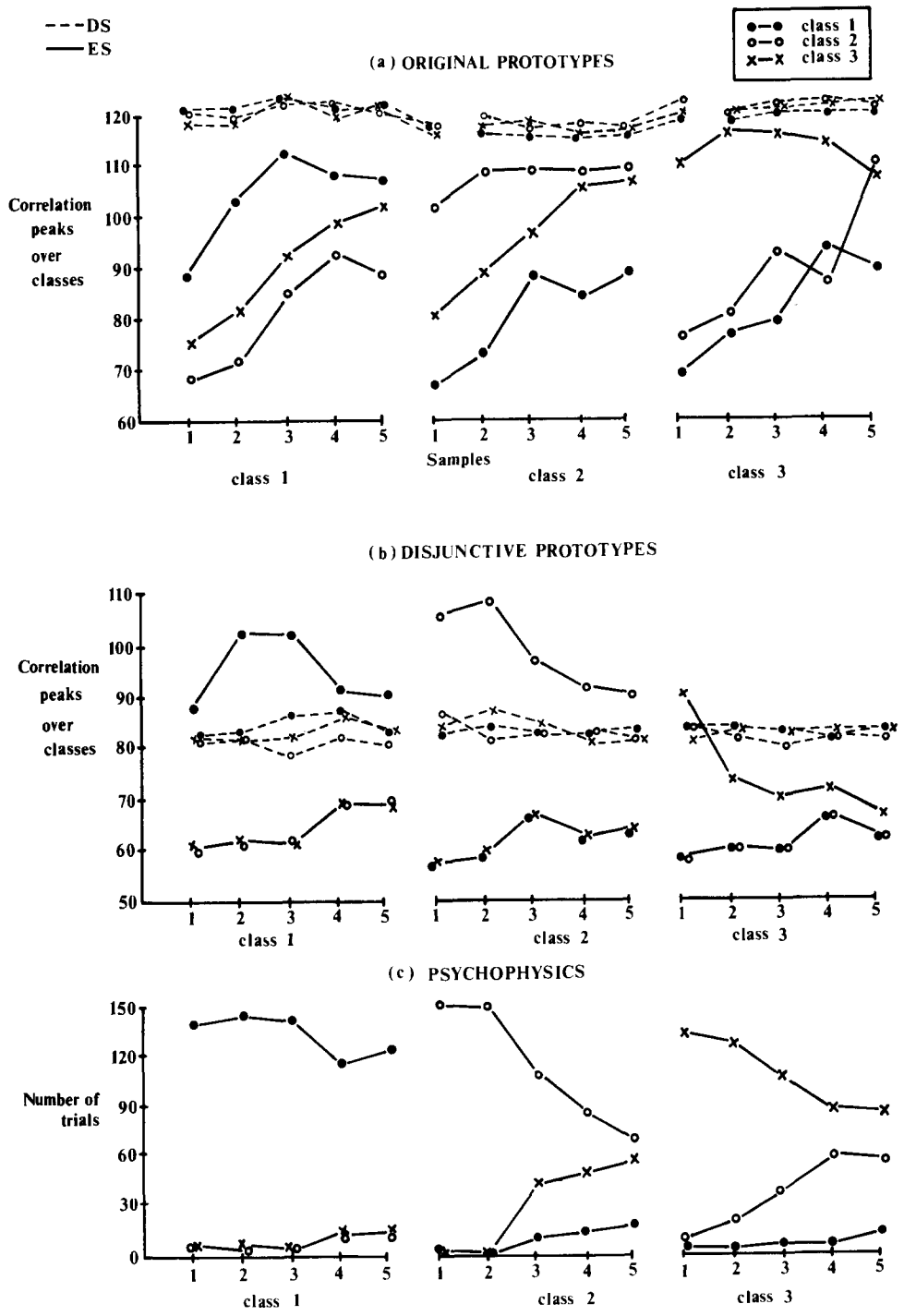


Fig. 7. Performance of the direct (DS: - - -) and edge (ES:——) encoding schemes using no disjunction (a) and disjunctive (b) prototypes. The curves refer to the peaks in the normalized cross-correlation for each of the three class prototypes used on each sample. (c) shows results from the psychophysical experiment where the class classification frequencies are shown per sample. Here incorrect classifications correspond to the curves which do not match the class number. In all cases the abscissas are numbered from 1 to 5 according to their class sample label (see Fig. 6).

classes and the disjunctive method attempts to maximally differentiate the cross-correlations over classes.

The benefits of such processes are that they are implementable in current (parallel) pipe-line pixel and array processors[9] and they do not require the prior selection of features either in the image or some transform domain.

Like most other pattern recognition techniques, the proposed algorithms (Figs 1 and 3) are only shift

invariant and so the sample and test patterns must be standardized for size and orientation before the process can be implemented. A number of procedures have been recently developed to overcome this problem including the definition of features via the equivalent two-dimensional circular harmonic power spectrum amplitudes,[10,11] log-polar transforms of image edge coordinates[9] and matching in joint image log-polar spectral domains for shift, rotation and size invariance.[13]

Though we have shown support for the ES scheme in the case of classifying patterns such as faces, etc., it has one important limitation which is, in principle, solvable. Images are not uniquely represented by such *samples* of scale space slices and, if we are to keep the number of filters to a minimum, what is required is an extra determination of the contrast direction at each edge point of the image, particularly those which have high stability. For example, augmenting the scale space [as determined by equations (4) and (5)] to include both minimum (zero-crossing contour) and maximum (direction of maximum contrast change) principle curvatures seem appropriate extensions to normal scale space encoding of pattern information. Such a representation is under investigation as a logical extension of the ES scheme. By treating such image pairs as one complex image, the matched filtering processes described above could be applied to these augmented features for alignment, prototype formation and pattern classification. To this date, however, we have found few cases in *pattern* recognition where such additional calculations seem necessary given the fact that the relative position of zero-crossings at different scales already implicitly encodes contrast direction to some extent.

4. CONCLUSIONS

In this paper we have explored a class of pattern classification methods based upon image domain operations. We have developed procedures which endeavor to maximally differentiate classes by the emphasis of differentiating features and have shown these techniques to be particularly useful with edge-only representations. The algorithms are being extended to include problems of recognition-under-transformations and more details of contrast gradients. Since all computations are based upon convolution processes, it is envisaged that such an approach to pattern recognition is readily applicable to current parallel architectures.

SUMMARY

In this paper we consider how human and machine pattern learning and classification may be accomplished by cross-correlation procedures based upon one of three types of encoding processes: (1) direct *pixel-by-pixel* registration; (2) via the outputs of two-dimensional *Gabor filters*; and (3) the registration of edge information derived from pattern *scale space* properties. Further, a *disjunctive* prototype function stage is introduced to produce orthogonalized class filters (or templates) to optimize between-class discrimination. We show that the edge-only (disjunctive) encoding process performs better than the others and more closely corresponds to human performance. Finally, some extensions of these processes are discussed for matching inferred shape characteristics of objects and recognition invariant to the position, orientation and size of target patterns.

Acknowledgements—T. M. Caelli was supported by grant no. A2568 from the Natural Sciences and Engineering Research Council of Canada, and W. F. Bischof was supported by grant no. 81.166.0.84 of the Swiss National Science Foundation.

REFERENCES

1. N. Ahmed and K. Rao, *Orthogonal Transforms for Digital Signal Processing*. Springer, Berlin (1972).
2. M. Minsky and S. Pappert, *Perceptions: an introduction to Computational Geometry*. MIT Press, Cambridge, MA (1969).
3. R. Chin and C. Dyer, Model-based recognition in robot vision, *Comput. Surv.* **18**, 67–108 (1986).
4. J. Daugman, Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity, *IEEE Trans. Syst. Man. Cyber.* **SMC-13**, 882–888 (1983).
5. D. Pollen, K. Foster and J. Gaska, Phase-dependent response characteristics of visual cortical neurons, *Models of the Visual Cortex*, D. Rose and V. G. Dobson, (eds). John Wiley (1985).
6. A. P. Witkin, Scale-space filtering: a new approach to multi-scale descriptions, *Image Understanding*, S. Ullman and W. Richards, (eds), pp. 79–95. Ablex, Norwood, NJ (1984).
7. A. Rosenfeld and A. Kak, *Digital Picture Processing*. Academic Press, New York (1982).
8. W. F. Bischof and T. M. Caelli, Parsing scale-space and spatial stability analysis, *Comput. Vis., Graphics Image Process.* **42**, 192–205 (1988).
9. T. M. Caelli and S. Nagendran, Fast edge-only matching techniques for robot pattern recognition, *Comput. Vision, Graphics, Image Process.* **39**, 131–143 (1987).
10. R. Wu and H. Stark, Rotation-invariant pattern recognition using a vector reference, *Appl. Optics* **23**, 838–840 (1984).
11. N. Ferrier, Invariance coding in image processing, M.Sc. thesis, University of Alberta (1987).
12. H. Glünder, On functional concepts for the explanation of visual pattern recognition, *Human Neurobiol.* **5**, 37–47 (1986).
13. C. Zetsche and T. Caelli, Invariant pattern recognition using multiple filter image representations, *Comput. Vision, Graphics, Image Process.* (in press).

About the Author—TERRY M. CAELLI is Professor of Psychology at Queens University, Kingston, Ontario. He received his Ph.D. at the University of Newcastle, Australia.

Professor Caelli's research interests are in human and computer pattern recognition and texture segmentation.

About the Author—WALTER F. BISCHOF is Assistant Professor at the University of Alberta. He received his Ph.D. at the University of Bern, Switzerland.

Professor Bischof's research interests lie in the area of computational vision.

About the Author—ZHI-QIANG LIU is a Postdoctoral research associate at the University of Calgary, Canada. He received his Ph.D. at the University of Alberta.

Dr Liu's research interests are image processing, computational vision and signal processing.