Integrating iconic and structured matching

R. B. Fisher and A. MacKirdy

Department of Artificial Intelligence, Edinburgh University
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK
rbf@dai.ed.ac.uk
http://www.dai.ed.ac.uk/daidb/staff/Robert_Fisher.html

Abstract. Several investigations [11, 16, 19–21] have recently been undertaken into object recognition based on matching image intensity neighborhoods rather than geometric matching of features extracted from the images. These projects have used small subwindows or complete image regions and matching has been based on the similarity of extracted descriptors to previously stored descriptors. One characteristic common to these approaches is the representation of objects as a whole, rather than as a structured ensemble. This paper describes an extension to these approaches wherein a set of related features recognized at an earlier iteration also contribute to the complete object recognition. The paper describes an iconic, or image-based, matching approach that incorporates an element of geometric matching and shows that use of the subfeatures improves matching efficiency, position accuracy and completeness.

1 Introduction

Symbolic matching algorithms have been popular and well-explored (*e.g.* [9]). They depend for their success on a combinatorial search process to establish feature correspondence and thus have an "all-or-nothing" behavior. To improve reliability and efficiency the use of a subcomponent hierarchy in matching algorithms has also been common in the symbolic domain, for both machine vision (*e.g.* [17, 4, 6, 7]) and biological vision (*e.g.* [13, 3]). In contrast, image template matching has been used in restricted domains for many years (*e.g.* [1], pg 65). The template matching approach has problems with rotation and scale invariance, and has requires much image computation. With the use of the log-polar representation [18, 22], the invariance problems can be overcome and the recent great increase in computational power of standard processors has reduced the computation time. As a consequence of these two factors, several investigations [11, 16, 19–21] have recently been undertaken into object recognition based on matching images (or some non-symbolic representation of them) directly rather than geometric matching of symbolic features extracted from the images.

We have been investigating [10, 11] the capabilities of iconic, or image-based, approaches to object recognition (described in some detail in Section 2) as have other research groups. Rao and Ballard [16] used a number of filters, derivatives of gaussians at several different scales, to build an n-dimensional feature vector. The feature vectors are input into a simple neural network which associates each vector with one of a number of objects. Their system is able to distinguish between a large number of objects under varying pose, by learning a set of poses. However, their system is unable to distinguish between objects with similar global frequency responses and the global filtering approach does not represent the spatial distribution of features necessary for distinguishing subtle appearance differences. Schiele and Crowley [19] have matched 2D histograms of pairs of image properties (mainly gradient-based) and achieved good matching results using a χ^2 metric, but their approach also ignores the global organization of the image features. Seibert and Waxman [20] used an ART network to match feature vectors extracted from log-polar processed images. Their features were interest points extracted from binary images of single isolated objects. 2D image-based recognition was linked into a 3D aspect and multiple competing identity object recognition network. Siebert and Eising [21] used the log-polar architecture with a difference-of gaussians receptive field and their matching scheme used templates applied directly to the log-polar image. An alternative approach [2] uses log-log sampling in the fourier domain but it is not considered here as it confounds the structural information exploited in our approach.

There has been much work on property-based image indexing from databases (*e.g.* [14]) but most of it has not used image geometry other than [8], which uses graph-like models of human and animal limb relationships. Mundy *et al* [15] have compared an iconic and a projective invariant object recognition system on a small database of simple parts and found that the iconic approach had a higher false-positive rate and lower false-negative rates. In part this was due to the iconic system not having a verification stage. The computation rate was also affected as the model base grew. On the other hand, their invariant approach was limited by the ability to model only simple shapes, but did cope better with illumination problems. One of the questions addressed in our research is how the associated model evidence affects the false-positive rate.

Another approach has been to represent families of similar objects by a weighted set of eigenvectors, with the recognition mechanism comparing the projection weights of a sample image to the eigenvectors in a database (*e.g.* [24]). In this case, object geometry is implicit in the representation, so the technique investigated here is inappropriate. Tistarelli [23] investigated the combination of the active space-variant sensor (as used here) with the eigenvector approach, and concluded that the accuracy of recognition can be much improved while simultaneously reducing the database size.

This paper describes an iconic matching approach that incorporates a simplified geometric model and shows that use of the subfeature matching promotes matching efficiency, position accuracy and correctness. More details can be found in [12].

The motivation for this research comes from the intuition that *if* I am looking at a feature and have some moderate evidence that it is a given model (*e.g.* "it might be an eye") *and* I have seen other nearby related features (*e.g.* another eye and a nose in the correct relative orientation and placement) *then* this feature is more likely to be the hypothesized object. The accuracy will increased by the response from the correct orientation and placement of the other features with respect to the current feature. The better the predicted and actual associated labels match, the more accurate the current match is likely to be.

2 Overview of the iconic recognition process

We have used a foveated (R, θ) log-polar coordinate system [22] for retino-centric coordinates, with 20 bands, each containing 48 sectors. The receptive fields (*i.e.* the area of the (i, j) image from which they take input) of each pixel in the (R, θ) representation increase (logarithmically by 1.2) as R grows larger in order to cover the entire foveated area. The receptive fields in the innermost bands take their input from only one or a few pixels, averaging the value. This gives high resolution around the foveation point. Receptive fields in the outermost bands average large numbers of pixels, giving lower resolution. Receptive fields overlap by about 33% to avoid gaps, leading to some blurring. The polar representation is attractive because it maps rotation and scaling into translation, and this feature is used in the matching algorithm described below to deliver scale and rotation invariance.

The main representations are:

1. **The World** - a large static (r, g, b) image (here 512^2) within which the iconic matcher saccades and extracts smaller (here 128^2) foveated views.
2. **The Image Stack** - Foveating the world image maps part of the raw (r, g, b) image to (R, θ) space, to form the first part of the image stack.
   The feature extraction process extracts 42 log-polar images registered with the current foveated image [10]. The images consist of 3 scales (extracted from the $\frac{1}{2}$, 1 and 2 size images) of 14 feature types: the red/green/blue intensity component images, two on- and off-center-surround features, four radial and orthogonal on- and off- bars, four orientations of edges and an unoriented corner measure. Each receptive field in each of the feature planes gives a measure of the strength of the given property at the corresponding spatial image location. The feature images are extracted by applying a small neighborhood operator at each location in the foveated image.
3. **The Model Base** - a set of models that may be matched to the current image stack. The iconic portion of each model has the same format and contents as the image stack. Each feature plane has a weight associated with it, indicating how useful the feature is in identifying this object. The structured portion of each model may have a list of associated models (*e.g.* an eye may link to a likely nearby nose position). This list has of the form { (*model_type, relative_position, relative_orientation, relative_scale, importance_weight*) }. These links can also be used to form an iconic geometric model (described in Section 3). Models are created by a learning process using pictures that are representative of the class. A model is normally registered on a feature that will attract the attention system [11]. Models are learned at three scales (50%, 100% and 200%) because not all features will be visible at all scales.

4. **The Interest Map** - The interest map [11] is an image structure registered with the world. Its contents record a value representing the interestingness of a given point in the scene. Interestingness values increase as center-surround and corner image features are found and as models are identified (as these predict locations of likely associated models). Interestingness values decrease at parts of the image that have been explored. Details of the calculation of the interest map is given in [11].

   The scene is explored in a saccade-like process by selecting the current highest interest point as the next location to foveate.

The matching process uses a modified multi-variate cross-correlation function:

$$V_f = f(\sum_{k=1}^{42} \rho_k w_k + w_0) \tag{1}$$

where ρ_k is the single channel match score (using the standard statistical cross-correlation between 2 feature images at each given rotation and scale offset) and w_k are weights which are learned by a perceptron learning algorithm. These weights reflect the relative importance of the feature correlation scores in determining object identity. The bias w_0 reflects the *a priori* probability of data belonging to this class (defaults to 0). $f(x)$ is the sigmoid function $1/(1+e^{-x})$. The matching process compares the stack of 42 iconic feature images to entries in a model database. In order to achieve rotation and scale invariance, the log-polar images are shifted in a convolution-like process. One shift direction is equivalent to a scale change, the other is equivalent to a rotation (well known properties of the log-polar transform). The shift process, for example, allows alignment of a rotated model with a database entry, thus improving correspondence. To achieve translation invariance, the matcher outer loop has a saccade-like process that shifts foveation to the next highest point in the attention map. As models are created by foveation at high interest points, using these points to direct foveation increases the chances of aligning a model image with the corresponding image in a test scene. Thus, the highest model score (greater than a "recognition threshold") is the recognized model, the current saccade position is the object's position and the rotation and scale at which the best match occurs is the estimated model orientation and size.

While this is an unusual approach to object recognition, its advantages are: the primal-sketch-like features provide an element of illumination invariance, the correlation matching allows a graceful degradation of correspondence and thus an element of generalization and the log-polar representation allows rotation and scale invariance. Its disadvantages include: an unconventional, idiosyncratic and generally unexplored architecture, moderate computation time per foveation position (a second per saccade and model), and somewhat heuristic feature extraction processes.

The architecture does allow exploration of the questions addressed in this paper: **is it possible to integrate geometric models within an iconic**

image matching paradigm? If so, does it provide any benefit in terms of speed, spatial accuracy or recognition completeness?

3 Integrating substructures into matching

To extend the architecture summarized in Section 2, five data structures or processes needed to be developed:

1. Structured model representation,
2. Subcomponent evidence recording,
3. Subcomponent evidence location,
4. Extended matching function, and
5. Interest map update.

These points are discussed in the following subsections.

3.1 Structured model representation

The models need to be extended to include other models associated with the current model. The associated models include subcomponents (such as an eye as being a subcomponent of a face) as well as more generally associated objects (such as a keyboard and a monitor).

Each augmented model becomes:

model type
42 feature (R, θ) feature planes
42 feature evidence combination weights ω_k plus ω_0
a set of N associated models:
$\mathcal{A} = \{\ \boldsymbol{a}_i\ \} = \{\ (\ \text{associated model type}_i,$
relative position $\boldsymbol{t}_i$, relative orientation ψ_i and scale σ_i,
relative importance $\gamma_i\ )\ \}$

Each related model is normally also a proper full model, containing its own 42 feature planes, as well as its own associated models, which may or may not refer back to the initial model.

3.2 Subcomponent evidence recording

The Stable Feature Frame (SFF) [5] represents the system's visual memory. It is registered on the world rather than the gaze location and incrementally records a stable, non-retinocentric view of the world. It contains defoveated (r, g, b) data obtained during the system's visual exploration, plus a list $\{\boldsymbol{s}_j\} = \{(M_j, \boldsymbol{t}_j, \psi_j, \sigma_j, V_j)\}$ of recognized image structures (*i.e.* model instances) M_j, their image locations $\boldsymbol{t}_j$, estimated orientation ψ_j, estimated scale σ_j and matching scores V_j.

3.3 Subcomponent evidence location

When matching at the current foveation position $\boldsymbol{t}_f$, initially only the 42 feature planes are involved. The matching algorithm computes a feature-based match at $\boldsymbol{t}_f$ with score V_f (using Eqn (1)) and an estimated model scale σ_f and rotation ψ_f. For each associated subcomponent i with relative position $\boldsymbol{t}_i$, relative orientation ψ_i and relative scale σ_i, one can predict where in the scene the associated models are likely to be found:

$$\boldsymbol{t}_{pi} = \boldsymbol{t}_f + \sigma_f \mathcal{R}(\psi_f)\boldsymbol{t}_i \tag{2}$$

their expected scale:

$$\sigma_{pi} = \sigma_i \sigma_f$$

and their expected orientation:

$$\psi_{pi} = \psi_f + \psi_i$$

The SFF can then be searched for model instances of the correct type that are within a search window of the predicted position, and that have a scale and orientation that are within a tolerance of the predicted position, scale and orientation.

3.4 Extended match evaluation function

The match evaluation score originally used only the match score V_f provided by correlating the 42 feature planes and combining their match scores (see Eqn (1)). With the associated model matching scores, the overall matching function has been extended to be:

$$\alpha V_f + (1-\alpha)V_a$$

where V_a is the associated model evidence and $\alpha = 0.7$ (chosen arbitrarily).

The associated model evidence V_a is given by:

$$V_a = \sum_{\boldsymbol{a}_i \in \mathcal{A}} \gamma_i \ max_{\boldsymbol{s}_j \in SFF} \ \{ \ h(pred(\boldsymbol{a}_i), \boldsymbol{s}_j) \ \}$$

where $pred(\boldsymbol{x})$ is the predicted properties of the observed model, as given by the formulas in Section 3.3, γ_i is the relative importance of each associated model, $\boldsymbol{a}_i$ is an associated model and $\boldsymbol{s}_j$ is a previously found model instance recorded in the stable feature frame (SFF). Thus, the more associated models that are successfully found, the larger is the combined evidence score V_a.

The associated model goodness evaluation $h()$ is given by:

$$h((..., \boldsymbol{t}_{pi}, ...), (..., \boldsymbol{t}_j, ...)) = M_j e^{-\beta d(\boldsymbol{t}_{pi}, \boldsymbol{t}_j)}$$

where M_j is the associated model's match evaluation score, $d()$ is a position dissimilarity metric and β is a scaling factor.

The position dissimilarity function $d()$ is evaluated by:

$$d(\boldsymbol{t}_{pi}, \boldsymbol{t}_j) = \frac{|| \boldsymbol{t}_{pi} - \boldsymbol{t}_j ||}{|| \frac{1}{2}(\boldsymbol{t}_{pi} + \boldsymbol{t}_j) - \boldsymbol{t}_f ||}$$

which has a small value if the distance between the predicted $\boldsymbol{t}_{pi}$ and observed $\boldsymbol{t}_i$ model positions is small relative to the distance of the matched models from the foveation point $\boldsymbol{t}_f$. This relative distance is important because the accuracy of position location declines as models are located distant from the foveation point (in part due to the averaging effects of the log-polar representation). An orientation and scale dissimilarity metric could also be incorporated into $h()$.

3.5 Interest map update

The original interest map was updated by computing an interest function at each of the 3 scales on the 14 features. Activity from high-interest features was combined with activity from opponent color features and then defoveated into the existing interest map. Details of the interest map calculation are given in [11].

If a model is successfully recognized, this means that there may be other nearby associated models. Section 3.3 discussed how the position of these associated models was predicted and how the previously recognized instances of the models were located. However, not all of the associated models may have been found so far. Therefore, it makes sense to look for these other models in the predicted locations ($\boldsymbol{t}_{pi}$) that had no successful match. The appropriate mechanism for causing this search is to increase the level of activation in the interest map at the predicted locations. The uncertainty in match position and scale grows as the predicted position becomes more distant from the foveation point. Therefore, the interest map is updated according to this function:

$$Imap(\boldsymbol{x}) += \alpha e^{-\tau \delta(\boldsymbol{x})}$$

where $\boldsymbol{x}$ are points near the predicted position $\boldsymbol{t}_{pi}$, $\boldsymbol{t}_f$ is the foveation point, $\alpha = 250$, $\tau = 100$ and

$$\delta(\boldsymbol{x}) = \frac{|| \boldsymbol{t}_{pi} - \boldsymbol{x} ||^2}{|| \boldsymbol{t}_{pi} - \boldsymbol{t}_f ||^2}$$

The interest map is updated only if no associated model has been previously found at the predicted point $\boldsymbol{t}_{pi}$. Determining whether a model has been found there uses the mechanism discussed in Section 3.4.

4 Experiments

The claim of the paper is that use of the subcomponent evidence improves the speed, positional accuracy and completeness of recognition. We demonstrate this by running the iconic recognition system with and without the subcomponent evidence process enabled. The recognition system is iconic, so only objects transformed by translation, scaling and rotation about the optical axis are appropriate. The experiments used a set of images containing frontal views of faces as the experimental scenes, with eyes, nose and mouths as the associated models. Faces have standard substructure and face images are commonly obtainable. Note, we are making no claims about this system as a face recognizer, or identifier system.

The model database contained (for these comparison experiments) 4 models: eye, nose, mouth and full face. When the subcomponent mechanism was not used, the interest map was not updated with the predicted associated model positions and the SFF was not searched for associated model matches. All other components of the system were identical including the model base. The time per saccade was virtually identical, with or without the subcomponent process.

All experiments started with a foveation at the center of the image (which was always on the face). The stopping criteria were when either all features were found, or the system had completed 20 saccades.

Table 2 summarizes the measurements taken from the experiments without and with subcomponent evidence, using 7 test images. Table 1 (left) describes the contents of the five feature columns of Table 2, and Table 1 (right) describes the contents of the rightmost summary column of the data tables. Italicized entries in Table 2 denote incorrect recognitions. The feature column boxes record on which saccade the feature was found, how many pixels error there were between the estimated feature position and what we thought was the correct registration point (correct matches only), the match score for the correct recognition and the number of false recognitions of that feature. (As eyes were indistinguishable, false recognitions of eyes are listed only for the left eye.) The rightmost column lists the number of saccades needed until the last feature was found, the total number of correct features found, the average position error in pixels of the correctly found features, and the total number of mismatched features.

The results for image C6 need a special explanation: in this case the initial foveation point was very close to the registration point for the face model, so the face was immediately recognized. This results in a large inhibition region (to prevent re-saccading back to already recognized features) suppressing the recognition of other nearby features. Consequently, it was hard to recognize any other nearby features. This is why the right eye and nose were never found and why it took a long time to saccade to the mouth.

With regard to the three qualities claimed, the evidence shows that:

speed - in all images except C6 (which searched a long time to find the 3rd model at the edge of the inhibition region), the number of saccades to recognize all features was less in the subcomponent evidence case.

position - For the 17 features recognized by both systems, the average subcomponent case position error was smaller (13.5 *versus* 16.7 pixels).

completeness - more features were correctly found (30 *versus* 17 out of 35 possible), and fewer incorrect features were found (12 *versus* 17).

The results on other images (results not shown here) have slight variations, but the same general properties still hold. Thus, we claim that the use of subcomponent evidence increases the speed, positional accuracy and completeness of recognition.

A side effect of the subcomponent evidence mechanism is that recognition scores of the correctly found features are often reduced. This arises because the recognition score now requires associated model evidence in order to obtain top scores. As initial feature recognitions will *not* have many previously found associated features, their recognition scores will be lower. Also, features recognized with inaccurate positions reduce the match scores in proportion to the position error.

Figure 1 shows the saccade path on the FCE5 image (left) without and (right) with subcomponent evidence. The search is clearly much more focussed with subcomponent evidence.

Saccade found	*Match score*		*Max. saccades*	*Total correct features found*
Offset error	*False instances*		*Average offset*	*Total mismatches*

Table 1. Key to the entries in Table 2.

Fig. 1. Saccades on FCE5 without (left) and with (right) use of subcomponent evidence.

a)

Image	left eye	right eye	nose	mouth	face	Summary
c1	4 1.00	*1 1.00*	*5 1.00*	*3 1.00*	7 1.00	7 2
	16 2		1	1	22 0	19 4
c3	1 1.00	*7 1.00*	8 1.00	4 1.00	3 1.00	8 4
	14 1		12 1	11 0	26 0	15.8 2
c4	10 1.00	2 1.00	1 1.00	*8 1.00*	*12 0.99*	12 3
	16 1	19	13 0	1	1	16 3
c5	10 1.00	12 1.00	1 1.00			12 3
	2 1	28	0 0			10 1
c6	5 1.00			2 0.86	1 0.98	5 3
	32 0			34 1	27 0	31 1
fce5	*9 1.0*	9 1.0	*3 1.0*		*5 0.79*	20 1
	1	9	2		2	9 5
bebie1	1 1.00		*18 1.00*		*2 0.94*	18 1
	3 0		1		1	3 2

b)

Image	left eye	right eye	nose	mouth	face	Summary
c1	3 0.71	4 0.76	6 0.82	5 0.79	7 0.91	7 5
	9 2	14	9 0	7 0	10 0	7.8 2
c3	1 0.70	2 0.76	3 0.76	4 0.83	5 0.89	5 5
	14 0	14	16 0	12 0	5 0	15.6 0
c4	4 0.71	9 0.81	10 0.82	7 0.77		10 4
	19 0	25	20 1	23 0		17.2 1
c5	3 0.79	2 0.72	1 0.70	4 0.82	5 0.74	5 5
	7 1	17	7 0	5 0	9 0	9 1
c6	2 0.83			18 0.80	1 0.68	18 3
	25 0			9 3	27 0	20.3 3
fce5	13 0.83	12 0.76	14 0.81	11 0.75	15 0.92	15 5
	4 2	5	4 2	6	2	4 4
bebie1	1 0.70	6 0.80		5 0.75	9 0.86	9 4
	3 1	10		22 0	8 0	8.6 1

Table 2. Results of testing (a) *without* and (b) *with* full subcomponent evidence.

Fig. 2. Test images for reported experiment (c1, c3, c4, c5, c6, fce5, bebie1)

5 Discussion

The claim made earlier in the paper is that the use of the simplified geometrical model and the associated subcomponent recognition processes improves the recognition process in several ways:

– **recognition speed** - in most cases, the number of saccades needed to recognize the features in the scene was reduced.
– **position accuracy** - the average error in the estimated position of the feature was reduced.
– **recognition correctness** - more features were correctly found and fewer incorrect features were found.

The experimental evidence supports this claim.

Sometimes, when a model has been incorrectly recognized (*e.g.* recognizing a squint eye with the mouth model), then prediction can lead the process to search several non-feature positions before returning to true feature positions. Adding a cumulative evidence process to the attention mechanism could help reduce the effect of this.

At the moment, the models do not distinguish between left and right eyes. This reduces matching accuracy but means that a single recognized eye will predict two possible positions for the second eye and the recognition will also expect to find evidence from 2 positions. The multiple prediction is reasonable, but the model representation that we use should be extended to allow mutually exclusive alternatives.

References

1. DH Ballard and CM Brown. Computer Vision, Prentice-Hall, New Jersey, 1982.
2. J Ben-Arie and Z Wang. Pictorial recognition using affine-invariant spectral signatures. Proc. Int Conf on Comp. Vis. and Pat. Rec., pp 35–39, San Jose, Puerto Rico, 1997.
3. I Biederman. Recognition-by-components: A theory of human image understanding. Psychological Review. vol 4, pp 115-147, 1987.
4. RA Brooks. Symbolic reasoning among 3D models and 2-D images. Artificial Intelligence Journal, vol 17, pp 285–348, 1981.
5. JA Feldman. Four frames suffice: a provisional model of vision and space. Behavioral Brain Sciences Vol 8, pp 265-313, 1985.
6. RB Fisher. From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis. John Wiley, UK, 1989.

7. RB Fisher. Hierarchical Matching Beats The Non-Wildcard and Interpretation Tree Model Matching Algorithms. Proc. 1993 British Machine Vision Association Conf., pp 589–598, Surrey, 1993.
8. D. A. Forsyth and M. M. Fleck. Identifying nude pictures. In IEEE Workshop on the Applic. of Comp. Vis. pp 103–108, 1996.
9. W. E. L. Grimson. Object Recognition By Computer: The Role of Geometric Constraints. MIT Press, 1990.
10. TD Grove. Attention directed iconic object matching. M.Sc. dissertation, Dept. of Artificial Intelligence, University of Edinburgh, 1995.
11. TD Grove, RB Fisher. Attention in Iconic Object Matching. Proc. British Machine Vision Conference BMVC96, Edinburgh, pp 293–302, September 1996.
12. A MacKirdy. Full Subcomponent Evidence and Further Parallelism in an Iconic Object Recognition System. BSc Honours Dissertation, Dept. of Artificial Intelligence, Univ. of Edinburgh, 1997.
13. D Marr. Vision. W.H. Freeman and Company, 1980.
14. K Messer, J Kittler, M Kraaijveld. Selecting features for neural networks to aid an iconic search through an image database. Proc. IEE 6th Int. Conf. on Image Proc and Its Applic, pp 428-432, 1997.
15. J Mundy, A Liu, N Pillow, A Zisserman, S Abdallah, S Utcke, S Nayar and C Rothwell. An experimental comparison of appearance and geometric model based recognition. in Proc. Int Workshop on Object Representations in Comp. Vis, 1996. In assoc. with ECCV 1996.
16. RPN Rao and DH Ballard. Object indexing using an iconic sparse distributed memory. Technical Report TR 559, Computer Science Dept., U. Rochester, 1995.
17. LG Roberts. Machine Perception of Three-Dimensional Solids, in Tippett, J. T. (ed.), Optical and Electro-Optical Information Processing, MIT Press, Ch. 9, Cambridge, Massachusetts, pp 159-197, 1965.
18. G Sandini and M Tistarelli. Vision and Space-Variant Sensing. In (ed. H. Wechsler), Neural Networks for Perception, Vol 1: Human and Machine Perception, Academic Press, Ch II.9, pp 398–425, 1992.
19. B Schiele and JL Crowley. Object Recognition Using Multidimensional Receptive Field Histograms. Proc. 1996 Eur. Conf. on Comp. Vision, Vol 2, pp 610–619, 1996.
20. M Seibert and AM Waxman. Learning and Recognizing 3D Objects from Multiple Views in a Neural System. in (ed. H. Wechsler), Neural Networks for Perception, Vol 1: Human and Machine Perception, Academic Press, Ch II.12, pp 426–444, 1992.
21. JP Siebert and I Eising. Scale-space recognition based on the retino-cortical transform. Proc. IEE Conf on Image Processing and its Applications, Edinburgh, 1995.
22. J van der Spiegel, G Kreider, C Claeys, I Debusschere, G Sandini, P Dario, F Fantini, P Belluti, and G Soncini. A foveated retina-like sensor using CCD technology. In C. Mead and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*, Ch 8, pp 189–212. Kluwer Academic Publishers, Boston, 1989.
23. M Tistarelli. Active/Space-Variant Object Recognition. Image and Vision Computing, Vol. 13, No 3, pp 215-226, 1995.
24. MA Turk and AP Pentland. Eigenfaces for Recognition. J. Cognitive Neuroscience, vol 3 no 1, pp 71–86, 1991.