Recognition Model with Extension Fields

Peter Kalocsai
University of Southern California
Los Angeles, CA 90089

Abstract

A recognition model which defines a measure of shape similarity on the direct output of multiscale and multiorientation Gabor filters does not manifest qualitative aspects of human object recognition of contour-deleted images in that: a) it recognizes recoverable and nonrecoverable contour-deleted images equally well whereas humans recognize recoverable images much better, b) it distinguishes complementary feature-deleted images whereas humans do not. Adding some of the known connectivity pattern of the primary visual cortex to the model in the form of *extension fields* (connections between collinear and curvilinear units) among filters increased the overall recognition performance of the model and: a) boosted the recognition rate of the recoverable images far more than the nonrecoverable ones, b) increased the similarity of complementary feature-deleted images, but not part-deleted ones, more closely corresponding to human psychophysical results. Interestingly, performance was approximately equivalent for narrow (±15°) and broad (±90°) extension fields.

1 Introduction

A task that both biological and artificial vision systems have to solve is to recover boundaries of objects from many times imperfect, noisy input. The Gestalt grouping principles of co-curvilinearity, proximity, constancy of curvature can help recovering meaningful information under these circumstances. There is considerable evidence from neuroscience [2] and psychophysics [3] that these grouping principles are built into the mammalian visual system in the form of connectivity patterns among processing units. There is both anatomical and physiological evidence that cells with approximately collinear orientation are interconnected mainly by excitatory connections [4]. Psychophysical results seem to suggest a broader field of connections between not only collinear units, but also curvilinear ones [3]. For either the narrow or the broad fields, the excitatory connections reveal smoothly decreasing strength with increasing distance and curvature differences [3,11]. There is also evidence for facilitation (increase in sensitivity for detecting Gabor patches) when local and global orientations are 90 degrees offset (the virtual line connecting two segments is perpendicular to their orientation) which is not modeled here [3,11]. The smoothly decaying excitatory field around an oriented segment is referred to as an *extension field* in this paper. The terms 'association field' or 'stochastic completion field' can be found in the literature to refer to similar constructs. These terms are generally applied to fields considered to manifest broad tuning. The term 'horizontal connections' has been employed to refer to the more narrowly tuned excitatory connections documented for neural units. To compare the effects of both narrow (collinear) and broad (collinear and curvilinear) connectivity patterns among processing units we decided to

implement two versions of the extension field: a narrow and a broad one. In the absence of precise neurophysiological data for the strength of connections between collinear and curvilinear units we choose the algorithmic definition of narrow and broad extension fields to be an excitatory gradient +/-15 and +/-90 degrees respectively centered on an oriented segment.

The goal of the present study was to investigate the consequences of adding extension fields to a recognition model that computed shape similarity based on representations of V1 hypercolumn activity. Specifically, we studied whether the extension fields would increase the resemblance of the recognition performance of the model to that shown by humans.

1.1 Brief comparison with previous work

Several previous computer vision models have used extension field type algorithms to guide the grouping process [5,6,7,10,12,14]. The main contribution in the present effort is the implementation of such a scheme on a biologically plausible multiscale and multiorientation filter representation, roughly similar to that of a lattice of V1 hypercolumns. This representation allowed a measure of shape similarity based on the combined activity produced by both the input image and the grouping process (although this does not necessarily mean that grouping results in activity that is indistinguishable from that produced by the original image). The previously cited efforts did not result in a measure of shape similarity.

Other differences distinguishing the present effort from prior ones was that the latter studies used only one scale as opposed to our multiscale approach. Since our test images were line-drawings, only one scale size--chosen to be the width (frequency) of the lines--could have very well been used, but a multiscale representation better resembles the sampling properties of biological vision systems. Many of the studies in the grouping literature [10,12] used an iterative relaxation algorithm as opposed to the more biologically plausible one-pass operation which was implemented here. An additional feature of the current study is that it directly compared the recognition performance of a grouping model to that of humans on a large number of test images, which is relatively rare in the literature. In the following we will describe two experiments on object recognition and compare human data to the performance of our baseline model.

1.2 Human experimental results and comparison of performance with a baseline model

In a psychophysical experiment [1] equal amount of contour was deleted from line drawings in such a way that the parts were either *recoverable* or *nonrecoverable* as illustrated in Figure 1. Subjects were able to recognize recoverable versions, but not nonrecoverable ones. A model [9] based on the direct output of a number of columns of multiscale and multiorientation Gabor filters (each column is roughly analogous to the simple cells in a V1 hypercolumn) was tested on the same images. The model recognized the nonrecoverable images as well as the recoverable ones, a result that does not correspond to human data (see results later).

In another task, subjects named briefly presented contour-deleted images [2]. For each image, two sets of complementary pairs were created by deleting every other

vertex and edge from each simple part in the first set (feature-deleted) and by deleting approximately half the components from each image in the second set (part-deleted) (Figure 2). If the members of the complementary feature-deleted pair or the part-deleted pair were superimposed they would provide an intact image without any overlap in contour.

Members of a complementary feature-deleted image pair (Figure 2. left) were equivalent to each other for human subjects as tested with the priming paradigm [2], but not for the model since the similarity of members of a pair was markedly lower than similarity of one of the images from the pair to itself. Part-deleted complementary images (Figure 2. right) were not equivalent neither to humans nor to the model.

Figure 1. Examples of intact (left), recoverable (middle) and nonrecoverable (right) test images [1].

2 Additions to the baseline model

The extension field is essentially a probability directional vector-field describing the contribution of a single unit-length edge element to its neighborhood in terms of direction and strength [6]. In other words, it describes the preferred direction and the probability of ex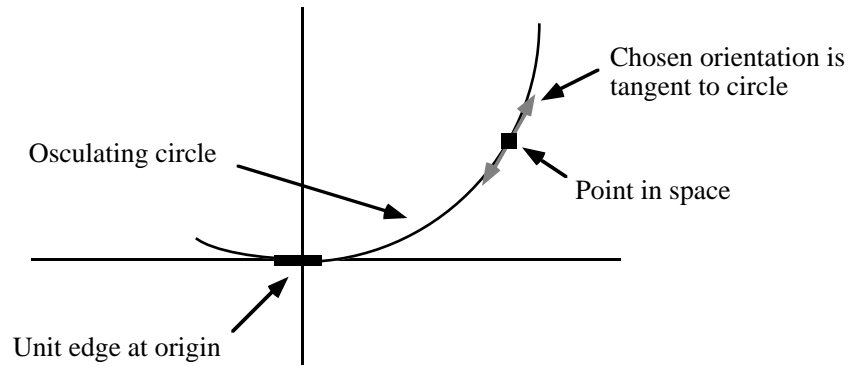istence of every point in space to share a curve with the original segment. The field direction at a given point in space is chosen to be tangent to the oscul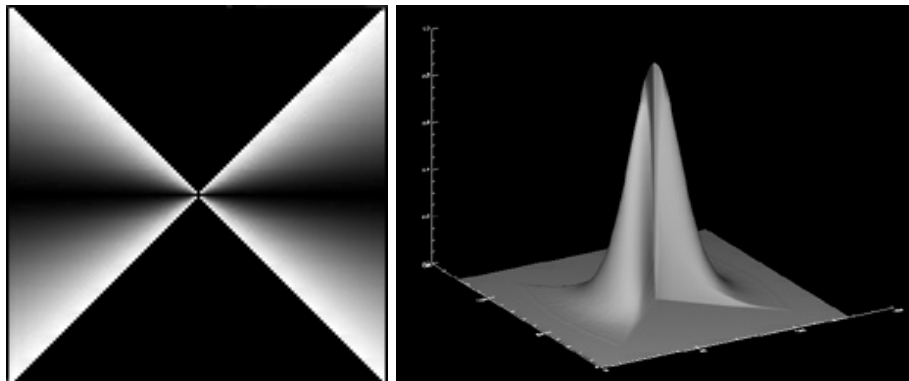ating circle passing through the edge element and that point, while its strength is proportional to the radius of that circle (Figure 3.). Also, the strength decays with distance from the origin (the edge segment). The decay of extension field strength is set to be Gaussian for both the proximity and curvature constraints:

Figure 2. Examples of feature-deleted complementary image pairs (left) and part-deleted complementary image pairs (right). Each member contains approximately half the parts of the object.

$$\overline{EF}(x,\rho) = e^{-Ax^2} e^{-B\rho^2} \qquad\qquad (1)$$

where x is the distance along the circular arc and ρ is the curvature of the given arc. Recently, Williams and Jacobs [13] described a very similar type of prior probability distribution of boundary completion based on computing the probability that a particle following a random walk will pass through a given position and orientation on a path joining two edge segments.

From each end of an edge segment, an extension field expanded to define a triangular area as shown in Fig. 4. The maximum orientation difference spanned by the broad extension field was ±90˚, which were at the ±45˚ boundaries of the extension field (Figure 4). Beyond those values, the Gaussians for orientation were set to zero so the broad extension fields had zero values above and below the main diagonals, as illustrated in Fig. 4. The narrow extension field is a subset of the broad extension field in that it uses the same direction and strength fields except that the excitation area is limited to ±15˚ orientation difference. The absence of grouping activity in the regions outside of the extension field merely means that additional information is needed to reconstruct curves between such pairs.

The extension fields were incorporated into the baseline model by allowing a field to operate on each of the 24 activation fields created by convolving the 24 kernels with an image. Because there were 8 orientations for the activation fields there were also 8 orientations for the extension fields. The additional excitation as provided by the extension field was distributed to the activation fields in such a manner that only the corresponding orientations of the activation fields and extension fields were convolved:

$$(EFWI)(\vec{l},\vec{x}_0) = \int EF_{\vec{l}}(\vec{x}_0 - \vec{x})WI_{\vec{l}}(\vec{x})d^2x = EF_{\vec{l}} * WI_{\vec{l}} \qquad (2)$$

Chosen orientation is tangent to circle

Osculating circle

Point in space

Unit edge at origin

Figure 3. Field direction for every point in space is chosen to be the tangent to the osculating circle passing through the edge segment and the given point.

Figure 4. The brightness coded directional map of the broad extension fields given a horizontal edge element in the middle. Within the butterfly shaped extension fields, black refers to horizontal and white to vertical orientations (left). The strength map of the extension field for locations and directions shown in the left figure. Strength declines with increasing orientation differences and distance from the edge element. There is no strength assigned above and below the diagonals (right).

where $\vec{l}$ gives the orientation of both the extension and activation fields. For the broad extension field model the activation fields not only get excitation from the extension field with the same orientation, but also from all the other orientations. For computational ease the excitation fields were divided into 8 subregions based on orientation and only the corresponding range of orientations were applied to an activation field with a given orientation. For the broad extension field model the overall excitation applied to an activation field is then given by summing up the

excitation coming from: (a) the extension field with the preferred orientation of the given activation field and (b) the excitation from all the other extension fields. In the narrow extension field model the activation field with a given orientation was only convolved with the excitation field having the same orientation.

To anticipate a point that will be made in the Discussion, the grouping activity can be distinguished from the activity produced directly by the image by keeping a reference copy of the early filter activations without any connectivity.

Figure 5 shows the activation fields created by convolving an image with the differently oriented and scaled kernels (altogether 24 kernels were used) and with the narrowly tuned extension field (again the 'boat' recoverable images is used as an example). In the visual representation the activation values of the model are normalized to integer values between 0-255 for 8-bit graphical display.
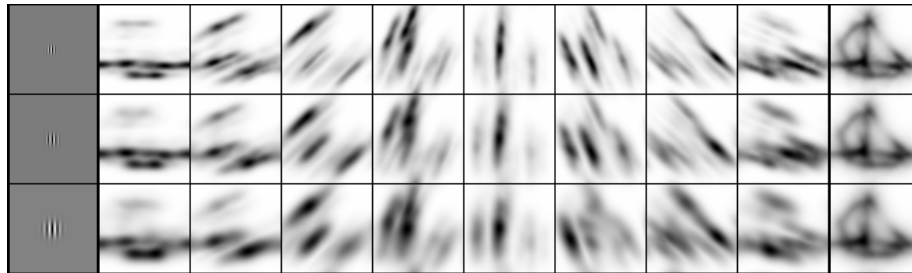
Figure 5. Activation fields produced by narrowly tuned extension fields for the recoverable 'boat' image (the model with broad extension fields gave similar results). The three rows represent the three scale sizes used in the experiment. The first column shows the 2D picture of the Gabor kernels at the three different scales. From the second to the second to last column the normalized activations of the differently oriented kernels to the recoverable 'boat' image are displayed starting with horizontal orientation and incrementing by 22.5 degrees. The last column shows the normalized cumulative activation of the three different scales at all orientations.

Figure 6 provides a direct visual comparison of the workings of the three different model types. The top row displays three versions of the 'boat' image from the set: intact, recoverable and nonrecoverable in left, middle and right columns respectively. Below the 3 x 9 blocks of images show the cumulative activation patterns induced by the three images in the three examined models: baseline, with narrowly tuned extension fields, with broadly tuned extension fields (from top nine image to bottom nine). In each of the three nine image blocks the first row represents the cumulative activation patterns of the kernels at the highest scale and at all 8 orientations. The second row represents the cumulative response at the highest and medium scale and the last row shows the 'total' of the activation for all scales and all orientations as well (similarly to the last column of Figure 5). This visualization of model activation also shows that for the second and third block of nine images (model with narrow and broad extension fields) the activation patterns for intact and recoverable images are much more similar than for the baseline model (first block of nine images).

3 Simulations

In the recoverable-nonrecoverable experiment the similarity of 36 intact images with the recoverable and nonrecoverable versions (altogether 108 images) was calculated and compared to each other.

In the feature-deleted vs. part-deleted experiment the similarity of the feature-deleted complementary image pair was compared to the similarity of the part deleted complementary image pair for 18 images (altogether 72 used).

3.1 Result of the simulations

The results of the simulations are displayed on Figures 7 and 8. The addition of narrowly tuned extension fields between similarly oriented kernels increased the similarity of both the recoverable and nonrecoverable versions to the original intact image, although it increased the similarity of the recoverable version more. Whereas for the baseline model there was no difference between the similarity of recoverable and nonrecoverable images $t(35) = .64$, $p = .52$ the addition of narrow extension fields significantly increased the difference between the similarity of recoverable and nonrecoverable types compared with the original images $t(35) = 4.8$, $p < .001$. The addition of broad extension fields further improved similarity for recoverable images, but did not improve similarity for the nonrecoverable ones compared to the narrow extension fields. Consequently, the broad extension field model further increased the difference between the similarity of recoverable and nonrecoverable images compared with the intact versions $t(35) = 9.09$, $p < .001$.

The addition of narrow and broad extension fields significantly increased the similarity of feature-deleted complementary images pairs, but did not improve the similarity of the part-deleted pairs. The similarity of two complementary feature-deleted images was already significantly higher than of two complementary part-deleted ones $t(17) = 3.04$, $p < .01$, but the addition of horizontal connections further improved this difference $t(17) = 8.54$, $p < .001$, just as did the addition of extension fields $t(17) = 9.12$, $p < .001$. The fact that similarity did not improve for part-deleted image pairs was expected considering that there was no any global knowledge provided that could relate the two different part structures in the pair to each other. However, the significant increase of similarity for the feature-deleted pairs was not an obvious outcome of the simulation. The addition of the broad extension field did not improve similarity for feature-deleted images compared to the narrow extension fields, which might be due to the large number of man-made objects in the stimuli set with mostly straight contours.

Figure 6 (next page). The top row displays the intact, recoverable and nonrecoverable versions of an image respectively. The 3 x 9 block of images below the top row display the activation patterns of the three model types (Baseline, Narrow Extension Fields, and Broad Extension Fields) to these images. The first row in each three blocks represents the cumulative activation of the highest frequency kernels at all eight orientations to the three images. The second row in each three blocks shows the cumulative activation of the highest and medium frequency kernels at all orientations. Finally, the third row in each blocks represents the cumulative activation of all three kernel sizes at all orientations (all 24 kernels).

Baseline Model

Narrow Extension Field Model
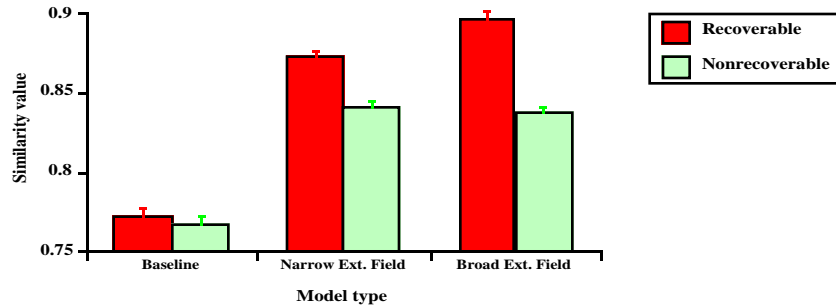
Broad Extension Field Model

Figure 7. Average similarity values for matching the original intact images with the recoverable and nonrecoverable versions in the three model types.
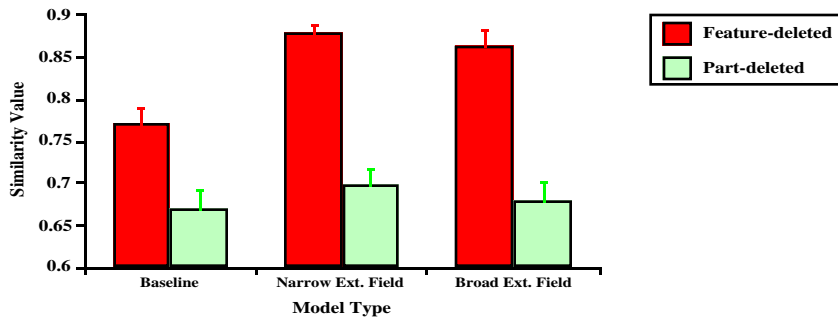

Figure 8. Average similarity values for matching complementary feature-deleted pairs and complementary part-deleted pairs in the three model types.

4 Conclusions

The addition of extension fields to a baseline model of object recognition that operates on the output of multiscale and multiorientation Gabor filters improves its overall recognition performance (at minimum for the given set of images) and brings its performance significantly and qualitatively closer to that of human object recognition.

Interestingly, adding broad extension fields to the original model did not improve its performance significantly beyond the improvement already achieved by narrow extension fields.

An obvious direction for future development of the model is to incorporate inhibition and endstopping into the connectivity pattern, both well known characteristics of biological low level vision systems. We might mention though that even without these additions the model's performance could significantly be improved just based on the addition of excitatory connections.

5 References

[1] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review,* 94, 115-147, 1987.

[2] I. Biederman and E. E. Cooper, "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition," *Cognitive Psychology,* 23, 393-419, 1991.

[3] D. J. Field, A. Hayes, and R. F. Hess, "Contour Integration by the Human Visual System: Evidence for a Local Association Field," *Vision Research,* 33(2), 173-193, 1993.

[4] C. D. Gilbert and T. N. Wiesel, "Columnar Specificity of Intrinsic Horizontal and Corticocortical connections in Cat Visual Cortex,"*The Journal of Neuroscience,* 9(7), 2432-2442, 1989.

[5] S. Grossberg and E. Mingolla, "Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations," Perception and Psychophysics 38, 141-171, 1985.

[6] G. Guy and G. Medioni "Inferring Global Perceptual Contours from Local Features," *International Journal of Computer Vision,* 20(1/2), 113-133, 1996.

[7] F. Heitger and R. von der Heydt, "A Computational Model of Neural Contour Processing: Figure-Ground Segregation and Illusory Contours," In *Proceedings of the ICCV,* 32-40, 1993.

[8] W. Konen, "ZN-Face: A system for access control using automated face recognition," In *Proceedings of the International Workshop on Automated Face- and Gesture-Recognition,* 18-23, 1995.

[9] M. Lades, J. C. Vortbrüggen, J. Buhmann J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Transactions on Computers,* 42, 300-311, 1993.

[10] P. Parent and S. W. Zucker, Trace Inference, Curvature Consistency, and Curve Detection, *IEEE Trans. PAMI,* 11(8), 823-839, 1989.

[11] U. Polat and D. Sagi, "The Architecture of Perceptual Spatial Interactions,"*Vision Research, 34(1),* 73-78, 1994.

[12] A. Sha'ashua and S. Ullman, "Structural saliency: the detection of globally salient structures using a locally connected network," In *Proceedings of the ICCV,* Tampa Fl., 321-327, 1988.

[13] L. R. Williams and D. W. Jacobs, "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience," In *Proceedings of ICCV,* 1995.