Whereas visual crowding does not greatly affect the detection of the presence of simple visual features, it heavily inhibits combining them into recognizable objects. Still, crowding effects have rarely been directly related to general pattern recognition mechanisms. In this study, pattern recognition mechanisms in visual periphery were probed using a single crowding feature. Observers had to identify the orientation of a rotated T presented briefly in a peripheral location. Adjacent to the target, a single bar was presented. The bar was either horizontal or vertical and located in a random direction from the target. It appears that such a crowding bar has very strong and regular effects on the identification of the target orientation. The observer’s responses are determined by approximate relative positions of basic visual features; exact image-based similarity to the target is not important. A version of the “standard model” of object recognition with second-order features explains the main regularities of the data.

Introduction

Object recognition is one of the main functions of vision. At present, we have some general understanding of the computations necessary for this task and how they might be implemented in the human brain. The widely accepted “standard model” of biological pattern recognition (Hubel & Wiesel, 1965; Fukushima, 1980; Riesenhuber & Poggio, 1999) is essentially a set of hierarchically organized feature detectors, tuned to increasingly complex features and with increasing extent of spatial pooling across the levels. Still, there are many unanswered questions, interesting for researchers of both biological and computer vision (e.g., Mutch & Lowe, 2008; Jarrett, Kavukcuoglu, Ranzato, & LeCun, 2009; DiCarlo, Zoccolan, & Rust, 2012). We only have vague ideas about the actual number of levels, set of features at each level, and rules of combining lower-level features into higher-level ones. There are different opinions on the role of interactions within the levels and top-down connections, and it has been argued that the feed-forward model may be fundamentally wrong (e.g., Mumford, 1992; Rao & Ballard, 1999).

Visual crowding is a deterioration of object recognition in the visual periphery caused by other objects nearby in the visual field (e.g., Bouma, 1970). Usually, crowding does not affect the detection of the presence of simple visual features much, but seems to heavily inhibit both their relative positions and combining them into recognizable objects (Levi, Hariharan, & Klein, 2002; Pelli, Palomares, & Majaj, 2004). Therefore, crowding is closely related to object recognition, and the crowding paradigm might help to understand pattern recognition in human vision (Levi, 2008; Pelli & Tillman, 2008).

Still, this possibility has rarely been the main motive of crowding studies. Usually, these studies attempt to reveal new regularities of crowding, test theories of crowding, or build better models of this phenomenon. Quantitative models of crowding frequently use simple feature discrimination as the observer’s task (e.g., Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; van den Berg, Roerdink, & Cornelissen, 2010). It is not clear whether the proposed mechanisms are relevant for object recognition. Of course, many studies of crowding provide some useful information on object recognition. However, the supposedly central problem of combining features into objects deserves more special attention.

There are several recent studies on the integration of small line segments (or Gabor elements) into longer contours in cluttered images (May & Hess, 2007; Chakravarthi & Pelli, 2011). Whereas this task assumes a kind of feature integration too, it is obviously different from combining features in typical object recognition. Although the stimuli of these studies (“snakes” and “ladders”) may be related to certain objects, the main issue in these studies is the saliency of two kinds of contours, not recognition of objects. Also, the integration of color, orientation, and spatial frequency studied by Pöder and Wagemans (2007) can...
hardly be regarded as true object recognition that should combine features and their relative positions. An article by Dakin, Cass, Greenwood, and Bex (2010) seems to be a better example of combining object recognition and crowding research. In that study, subjects had to identify the orientation of a rotated T in the presence of a nearby irrelevant T-like object. The authors analyzed the distributions of incorrect answers as dependent on the features of the flanking object. They found that the nature of interactions was mostly determined by the configuration formed by the target and flanker, irrespective of its absolute orientation; i.e., the observed effects were predominantly object-centered rather than viewer-centered. They observed twice as many 90° target rotations among incorrect responses. There were many more errors with flankers above or below (end flankers) compared to left or right (side flankers) of an upright T target, and the end flankers induced errors that resembled the flanker more often. These regularities were accounted for by probabilistic weighted averaging of feature positions within the objects. However, the main idea of their study was to see how much the supposed low-level feature interactions could explain crowding effects with these simple objects, and their model was not intended to simulate generic object recognition.

The goal of the present study is to obtain a better understanding of the main problem of object recognition—combining features into objects. I chose a minimal stimulus for that purpose: an object composed of two “features” and one “free-floating feature” that can be mistakenly integrated with those from the object.

My experiment was similar to that from Dakin et al. (2010). A target (tumbling T) was presented together with a single bar (either vertical or horizontal) in a random position around the target. Observers identified the orientation of the target. Response distributions as dependent on the position and orientation of the flanker were analyzed and modeled. I found that such a simple crowding object has very strong and regular effects on target identification, effects which may signify important constraints on object recognition mechanisms.

**Methods**

The stimuli were presented on a 15-in. CRT monitor with a resolution of 1024 × 768 pixels. The objects were black on the grey background (50 cd/m²).

The target (rotating T) was composed of two bars (vertical and horizontal, width 2 pixels, height 8 pixels). The crowding bar had the same size. The target was presented in one of the four orientations (0°, 90°, 180°, and 270°). The crowding bar could be either vertical or horizontal and was presented in a fixed distance of 12 pixels from the target center in a random direction (0°–360°) from it. Examples of stimuli are given in Figure 1.

The stimulus was presented unpredictably either left or right of the fixation point (which was permanently visible) at an eccentricity of 160 pixels (measured from the center of the target). The viewing distance was approximately 50 cm. Thus, the size of the target was about 0.3° × 0.3°, eccentricity was 6°, and center-to-center distance between the target and the crowding bar was 0.45° of visual angle.

Trials were initiated by the observer, clicking the “next trial” button. After a short delay (350 ms), a stimulus was exposed for 60 ms, either left or right of the fixation point. The observer had to identify the orientation of the target and insert his/her choice by clicking the corresponding icon on the screen. The program informed the observer whether the response was right or wrong.

Three observers took part in the experiment: the author and two other people who were naive about the background of this study. Each of them ran 1,000 trials. The data are response distributions as dependent on orientation and position of the crowding bar. For the main part of analysis, the data were pooled across absolute orientations of the target. Thus, responses are expressed relative to target orientation: correct (target orientation), 90° clockwise, 90° counterclockwise, and inverted. Similarly, orientation of the crowding bar was
classified as either “vertical” or “horizontal” relative to the upright orientation of the target. Positions of the bar were also measured relative to the upright oriented target and were grouped into 16 bins of 22.5°.

Several models were tried in order to understand possible mechanisms of the observed regularities. The first simple (similarity choice) model was implemented in MS Excel and fit using the Excel Solver add-in and maximum likelihood criterion. The remaining models were built as simulations in Matlab, and the fminsearch function was used to search for optimal values of parameters (corresponding to the minimum of likelihood ratio statistic G). The predictions of experimental data given a particular combination of parameter values were generated by simulation of 2,000–10,000 trials per stimulus configuration.

Results

The average proportion correct was around 0.5 (0.52, 0.48, and 0.60 for observers EP, RV, and VT, respectively). Observers had some individual preferences for particular response alternatives, but these response biases were relatively small (the probabilities for all response categories for all observers were in the range of 0.2–0.32, very close to unbiased 0.25).

Usually, crowding effects are stronger when flankers are located in the radial as compared to tangential direction from the target in the retinal coordinates (e.g., Toet & Levi, 1992). For the display used here, performance should be worse when a flanker is left or right as compared to above or below the target. A small effect in this direction was observed (proportion correct 0.50 for radial and 0.58 for tangential flankers). However, a more detailed analysis (Figure 2) revealed that the (strong) radial-tangential difference in the expected direction was present for horizontally oriented flankers only; when the flanker was a vertically oriented bar, the effect was actually opposite. Thus, the largest difference here is configuration-based and orientation-invariant: Crowding is stronger when a flanker is oriented radially compared to tangentially in the target-centered coordinates (proportion correct 0.43 and 0.64, respectively). This result might be explained by a conjecture that center-to-center distance is not a perfect measure of crowding zone with these stimuli, but edge-to-edge (or edge-to-center) distance plays a role as well. However, we will see some more specific interpretations in the Modeling section. (I could not analyze the other usual retinotopic anisotropy—difference between “inner” and “outer” flankers—because the position of stimuli (left or right) from the fixation point was not registered in this experiment).

Figure 3 depicts the response distributions relative to target orientation, as dependent on flanker position and orientation (relative to the upright target orientation), pooled over absolute orientations and across three observers. Individual results were qualitatively similar (presented as Supplemental materials). It is clear that the crowding bar has a very strong and regular effect on the identification of target orientation. In some conditions, performance is almost perfect but drops to chance level in others. There are certain regions where the particular wrong answers dominate over the correct ones. It is not difficult to see that wrong answers are evoked by a global stimulus configuration that resembles the correspondingly oriented target (see examples in Figure 4). Also, one may notice that only rough relative positions of features matter; exact metrical relations are not important. Frequently, the crowding bar, although distant from the target bars, appears to be combined with one of these.

There were slightly more ±90° as compared to 180° target rotations among incorrect answers for all three observers, but the effect (ratio of two types of errors was about 1.25) was much less than the twofold difference observed by Dakin et al. (2010). The other regularity reported by Dakin et al. (2010) was actually in the opposite direction in this study: Performance was worse with flankers left or right compared to above or below an upright T target (proportion correct 0.42 vs. 0.65). This effect can be attributed mainly to the “horizontal” side flanker (first example in Figure 4), which reduced the proportion correct to 0.23, below chance level.

There is a complementary way to look at the present data that reveals some additional information. Instead of presenting response distributions
relative to the target orientation across different flanking conditions, we can take a particular response alternative (e.g., upright T) and plot the probability of choosing this response as dependent on the orientation of the target and the position (and orientation) of the flanker. An example of this format is given in Figure 5 (full data set is presented in Supplemental materials).

Figure 3. Results of the experiment. Distributions of responses relative to target orientation (corr = correct, inv = inverted, acw = rotated 90° counterclockwise, cw = rotated 90° clockwise) as dependent on orientation and position of the flanking bar relative to the target. For the “vertical” flanker, the target-flanker configurations corresponding to some points on x axis are shown. Data are averaged across three observers.

These graphs show that the position of the flanker affects the probability of a particular response quite uniformly, regardless of the target orientation. Especially, the curves for all “incorrect” (not matching a given response category) target orientations virtually overlap. For the matching target orientation, probability is of course higher but follows a similar curve. These results indicate that target orientation and flanker position have mostly independent effects.

**Modeling**

In this part, I present several ways of modeling used for better understanding the mechanisms behind
the data. First, the data are approximated by a behavioral similarity choice model that reveals a surprisingly simple underlying structure. Second, simulation models based on different theoretical ideas and of different complexity are tried in order to reveal possible roles of various computational mechanisms.

**Similarity choice model**

The observation of mostly independent effects of the target orientation and the position of flanker (Figure 5) suggests that the data can be reproduced by a relatively simple similarity choice model (Luce, 1963; Estes, 1982), which allows separation of two independent effects on choice probabilities. This model assumes that the tendency for a particular response $r_j$ is determined by the similarity $\eta_{ij}$ of the presented stimulus $s_i$ with the correct stimulus for that response $s_j$, and the bias $\beta_j$ of producing that response independently of the stimulus.

Probability of response $r_j$ for stimulus $s_i$ is calculated as follows:

$$p(r_j|s_i) = \frac{\eta_{ij} \beta_j}{\sum_k \eta_{ik} \beta_k}, \quad 0 \leq \eta_{ij} \leq 1, \quad \eta_{ii} = 1,$$

$$\eta_{ij} = \eta_{ji}, \quad 0 \leq \beta_j \leq 1, \quad \sum_j \beta_j = 1. \quad (1)$$

Thus, I assume that all the distributions of responses in this experiment can be reproduced by the similarities between differently oriented Ts and the biases induced by the flanker in different angular positions. Also, I assume that the model is orientation-invariant (independent of absolute orientation) and symmetrical relative to the upright orientation of the target. Under these assumptions, the effects of the horizontal and vertical flanker are represented by the same model; we just need to rotate the display (or model) by $90^\circ$ (I used the data pooled across the flanker orientations to fit the model).

This model has 15 independent parameters (two similarities and 13 biases) and approximates 192 independent response probabilities ($G = 342, R^2 = 0.92$). (When recalculating the predictions into the format used in Figure 3, the fit is even better [$R^2 = 0.96$]. That fit is not surprising, given 15 free parameters.) The estimated target similarities and biases, as dependent on the position of the flanker, are given in Figure 6. The modeling demonstrates that the assumption about the independence of the effects of the target and flanker is really a good approximation. Also, it shows that a radially oriented flanking bar has a much stronger biasing effect than does a tangentially oriented one.

**Computational models with two features and positional noise**

It is reasonable to suppose that positional uncertainty plays some role in perception of peripherally presented stimuli (e.g., Hess & Hayes, 1994; Michel & Geisler, 2011). Also, positional noise was an important component in the model used by Dakin et al. (2010) for similar data. Therefore, I start with a very simple model which assumes that the positions of simple features (horizontal and vertical bars) are registered on a retinotopic map with additive 2D Gaussian noise. I suppose that the observer has no access to absolute positions and can use relative positions of features only. Then, the most likely target orientation can be determined by a comparison of the noisy relative positions in a display with the four template vectors corresponding to the possible target orientations (Figure 7). For my simple three-bar stimulus, there were two relevant feature pairs in each display, and eight comparisons per display were needed. Euclidean distance was used as the measure of difference, and the target orientation corresponding to the best match (minimal difference) was chosen. (This model is a
simplification of the ideal observer for the same task. An ideal observer might use an additional piece of information that, without the noise, the flanking bar is always presented at the fixed radius from the target center.

Trying to approximate the experimental data, the sole free parameter, standard deviation of positional noise $\sigma$ was varied. The best fit (according to log likelihood statistic $G$) was obtained with $\sigma = 3.0$. Although the model correctly predicts some observed regularities (Figure 8), the overall fit is very poor ($G = 1050, R^2 = 0.70$).

There is a straightforward way to improve this simple model. Suppose that the observer is uncertain about the size of the target or cannot apply the optimal (small) templates in the visual periphery. Therefore, it is possible that templates with somewhat larger offsets of horizontal and vertical features are used. I varied this offset (length of the template vectors) $l$ together with standard deviation of positional noise $\sigma$ and found that
the best fit could be obtained with \( l = 9 \) pixels, more than twice that corresponding to the actual target. This model with enlarged templates approximates the data much better as compared with the one with “ideal” templates (see Figure 8). Still, this model fails to reproduce some important details of the data. For example, it predicts that the vertical flanker above and horizontal flanker below the target should have exactly identical effects. Obviously, this prediction is wrong. A better model should handle simple features differently, dependent on their possible role within a candidate object. Making this distinction needs something more than just registration of these two features.

The simple models with positional noise of features predict a strong predominance of \( \pm 90^\circ \) compared to \( 180^\circ \) target rotations among incorrect responses, which was observed in Dakin et al. (2010) study but not in the present experiment. Therefore, the low-level positional uncertainty of features is likely not the main limitation of performance in this study.

**A pattern recognition model with second-order features**

This model follows the general ideas of the “standard model” for pattern recognition in biological vision (e.g., Riesenhuber & Poggio, 1999). It assumes a set of local feature detectors, pooling the outputs of these detectors over some second-level receptive fields and combining the results into second-order features. Then, a classification mechanism is applied to these features.

Although the stimuli used here were built from only two “features” (horizontal and vertical bars), a real visual system could obviously detect many more features in these images (e.g., low-pass blobs, line terminations, curved edges). I tried not to increase the complexity too much and used three simple features in my model—horizontal and vertical bars plus an un-oriented, low-pass blob corresponding to the target center.

The model assumes that the simple features (a blob at the target centroid and the horizontal and vertical bars, both belonging to the target and the crowding bar) are registered on a retinotopic map. Further, I assume that the positions of the oriented features relative to the hypothetical object center are estimated using the relative activity of neural units with large receptive fields positioned with some spatial offset from the target center. This operation can be viewed as a computation of second-order features (e.g., presence of a horizontal bar above the object center). This kind of complex features has been proposed by some neurobiological studies (Connor, Preddie, Gallant, & Van Essen, 1997; Pasupathy & Connor, 2001). The main points of this model are illustrated in Figure 9.

I suppose that the receptive field profiles are circular 2D Gaussian. The size of them should be consistent with the size of crowding zones at a given eccentricity. I chose the diameter (4\( \sigma \)) equal to 0.5 of the target eccentricity and offset of the center from the supposed target center 2\( \sigma \). I used the half-wave rectified difference between the signals of two opponent units as a signal of a second-order feature. Gaussian noise was added to these signals. With two instances of the same simple feature within a receptive field, a max rule was applied—the feature with the highest response (after spatial weighting and noise addition) was chosen. Different second-order features provide evidence for different target orientations. The signals representing horizontal and vertical features were combined additively. I used two weighting factors in order to approximate the data: a differential weight of (1) the feature signals from the target relative to the signals from the flanker \( w_t \), and (2) the signals from radial relative to tangential oriented features, in respect to each target alternative (i.e., the vertical relative to horizontal stroke of an upright oriented T) \( w_r \).

Mathematically, the second-order features were calculated as follows:

\[
f_j = \max_i \left\{ \max \left[ w_p(i) \cdot (g_{i,j}g_{0,j} - g'_{i,j}g_{0,j}), 0 \right] + N \right\},
\]

where \( g_{i,j} \) and \( g'_{i,j} \) are the responses of the two units underlying second-order feature \( j \) with spatially oppo-
nent pooling fields (e.g., -x and +x in Figure 9B) to a simple feature signal \( i \), \( g_{i,j} \) and \( g'_{i,j} \) are the responses of the same pooling units to the feature signal corresponding to the target center, \( N \) is independent Gaussian noise, and \( w_p \) is positional weight (\( w_p = w_t \) for target features, and \( w_p = 1 \) for flanker features).

\[
g_{i,j} = \exp \left[ -\frac{(x_i - \bar{x}_j)^2 + (y_i - \bar{y}_j)^2}{2\sigma^2} \right],
\]

where \( x_i \) and \( y_i \) are coordinates of simple feature \( i \) in a display, \( \bar{x}_j \) and \( \bar{y}_j \) are coordinates of the center, and \( \sigma \) is the size (standard deviation) of the integration field \( j \) for calculating relative position signals.

The response alternative with the maximum support was selected in each trial:

\[
r = \arg \max_k \left( \sum_j w_f(j,k)f_j \right),
\]

where \( w_f \) is feature weighting factor (\( w_f = w_r \) for radial feature, \( w_f = 1 \) for tangential feature, and \( w_f = 0 \) for features not consistent with given response alternative \( k \)) and \( f_j \) is signal of second-order feature \( j \).

Three parameters \( w_r \), \( w_f \), and \( \sigma_N \) (standard deviation of noise \( N \)) were adjusted to fit the experimental results. The model fits the data much better than the previous models (\( R^2 = 0.95 \)) and predicts the main qualitative regularities well (Figure 10). The fit parameters are given in Table 1.

Parameter \( w_f \) can be interpreted as a measure of the selectivity of simple feature signals from the target location relative to those from the flanker location. The simple features to be combined in a given trial are selected by maximum rule from the noisy feature

![Figure 10](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933685/)

Figure 10. Approximation of experimental results by the object recognition model (left) and probabilistic positional averaging model (right). Symbols represent experimental data (d) and lines are predictions of the model (m). Responses are relative to target orientation: corr = correct, inv = inverted, acw = rotated 90° counterclockwise, cw = rotated 90° clockwise.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Observer</th>
<th>EP</th>
<th>VT</th>
<th>RV</th>
<th>Three observers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise, ( \sigma_N )</td>
<td></td>
<td>0.023</td>
<td>0.019</td>
<td>0.019</td>
<td>0.020</td>
</tr>
<tr>
<td>Target (re flanker) weight, ( w_t )</td>
<td></td>
<td>3.0</td>
<td>3.4</td>
<td>2.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Radial (re tangential) feature weight, ( w_r )</td>
<td></td>
<td>2.0</td>
<td>2.2</td>
<td>2.6</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Table 1. The fit parameters of the pattern recognition model for individual observers and for the pooled data.
signals. The spatial weight parameter \( w_r \) creates some preference (higher probability) for the features from the target location. I suppose that its value is determined by proximity to the candidate target center, which corresponds to the position of the blob-like feature. Selective amplification of the target (or inhibition of the flanker) signals can be a combined effect of both top-down attention and bottom-up saliency (e.g., Pöder, 2006).

Parameter \( w_r \) is a measure of the relative “importance” of the two oriented features for the identification task. It indicates that the radial (relative to the object center) bar (i.e., vertical bar of an upright T) is weighted more heavily than the tangential one (horizontal bar of an upright T). It is possible that the radial feature is more reliable because it extends somewhat farther from the object center. However, a more detailed analysis is needed for the true explanation.

In my simplified model, I used four integration fields at the fixed positions around the target object. However, this model performs equally well when a trial-by-trial jitter (up to the sigma of the integration field profile) is added to the position of integration fields. (In that case, the noise parameter must be reduced because of additional noise introduced by the spatial jitter). For full position invariance, a larger set of overlapping integration fields is necessary.

**Probabilistic anisotropic positional averaging**

This model was proposed by Dakin et al. (2010) in order to explain the results of their (quite similar) experiment. The model assumes that the positions of the target features are, with some probability, distorted by similarly oriented flanker features. When distorted, the position of the target feature is replaced by a weighted average of the positions of target and flanker features. Both the probability and the extent of distortion (when it occurs) depend on the distance between the features according to a 2D Gaussian profile (interference zone). Also, a noise \( \sigma_{\text{pos}} \) was added to the positions of features.

I tried to apply these mechanisms to the stimuli of my present experiment. Because there was only one flanking feature, only the position of the target feature with the same orientation could be distorted. I found that in order to approximate my data, both horizontal and vertical coordinates of the respective target feature should be distorted and that the parameter corresponding to the maximum weight of flanker (\( w_{\text{average}} \)) should be set to its limit, 1. With optimal parameters (maximum probability of distortion \( w_{\text{prob}} = 0.8 \), noise \( \sigma_{\text{pos}} = 0.31W \), dimensions of interference zone \( 1.2W \times 2.4W \), where \( W \) is width of the target), this model fit my data significantly better than the simple two-feature model but much worse than my three-feature object recognition model (see Figure 10). Interestingly, the spatial parameters (in units of target size) were very similar to those reported by Dakin et al. (2010), although the target size and eccentricity were very different in these studies.

The accuracy of fit measures for the reported models are given in Table 2. The differences are obvious without any detailed statistical tests (note that a difference in AIC about 100 means that one model is about \( 10^{21} \) times more likely than the other). The simulations suggest that the ideas about second-order features and encoding positions of features relative to an object center are likely important for the explanation of the present data.

### Discussion

The present study used a novel method of probing pattern recognition mechanisms with an additional single “feature” in different positions relative to the target object. It was found that such a minimal crowding object has very strong and regular effects on the identification of the target. Interestingly, the effects of the flanking feature were mostly independent of target orientations, indicating that spatial interactions between particular features of the target and flanker are not important in these conditions.

The results are broadly consistent with the main logic of the “standard model” of visual object recognition that combines feature detection and spatial pooling. In that model, spatial pooling is a necessary component of object recognition invariant to spatial transformations. Crowding, however, has been frequently explained by a pooling of visual features over inappropriately large areas in the periphery. There have
been very few attempts to relate these two views of feature pooling (Pelli & Tillman, 2008; Isik, Leibo, Mutch, Lee, & Poggio, 2011). The present study shows how the necessary pooling within object recognition mechanism may become “inappropriate” in the visual periphery.

I found that simple models that used two oriented features could not explain the experimental results well. Adding a third feature—unoriented blob corresponding to the target center—helped to improve the fit considerably. Although computational models frequently use the set of simple features comprising oriented bar and edge detectors (or Gabor filters), neurobiological studies have reported more diverse optimal stimuli for low-level visual neurons, including unoriented blobs, single lines, and multi-period gratings (e.g., Ringach, 2004). Perhaps all these features play some role. Combining nonoriented blob-like features with bars and edges may be a more efficient way to encode and recognize real objects in biological vision.

I used second-order features that encode the presence of simple visual features in some approximate position relative to candidate object center. This idea is supported by several neurobiological studies (Connor et al., 1997; Pasupathy & Connor, 2001; Freiwald, Tsao, & Livingstone, 2009). The absence of direct combination of oriented features also seems to be consistent with independent effects of these features, as found using the choice model in this study. Theoretically, using a kind of “reference” feature may help to avoid a combinatorial explosion when all features can be combined with all others. There are computer vision algorithms (e.g., Lowe, 2004) that reduce the total amount of computation by calculating complex features only around some salient points (e.g., local contrast energy maximum).

Many crowding studies have measured a single interference zone around the target object. This study, as well as Dakin et al. (2010), has revealed some structure within this zone and specified subfields for pooling different visual features. While Dakin et al. explained their results by low-level automatic feature interaction, the results of the present study are more consistent with combining of features within a pattern recognition mechanism. Of course, both mechanisms may exist, but available data do not allow telling them clearly apart.

Although the results of this study and Dakin et al. (2010) are not directly comparable, some differences (different proportions of 90° and 180° errors, different effects of side vs. end flankers) may need an explanation. At present, I have no good one. In the light of my modeling, the two-bar flankers should behave differently, because they form their own “object center” besides the target, but the exact predictions need further assumptions. However, in addition to different flankers, our experiments were different in regard to spatial uncertainty, and using visual masking. Also, in the present study, the size of stimuli was about six times, and the target-flanker distance (in eccentricity units) about two times smaller, exposure duration was four times shorter, and consequently, proportion correct was 0.53 compared to 0.82 in Dakin et al. study. Some of these factors could produce the observed differences in response distributions too.

I did not use positional noise in my best model. The simple models with positional noise of features were not well consistent with my experimental results. Of course, a large amount of intrinsic positional uncertainty can be observed in the visual periphery (e.g., Michel & Geisler, 2011). However, this uncertainty does not distort seriously the relative positions of features in a simple pattern recognition task (e.g., Levi, Klein, & Sharma, 1999). These findings seem to support the idea about special mechanisms for the estimation of relative positions of visual features.

There is some resemblance between the present and Petrov and Popple (2007) studies. These authors, too, used stimuli composed of three simple “features” and analyzed the distributions of observers’ responses. However, in their experiment, all features were equally relevant, a condition that needs a somewhat different model. Qualitatively, their conclusion about the important role of preattentive feature contrast is very similar to my assumption about bottom-up amplification of feature signals at the locations of salient features.

The results of the present study do not contradict the general idea that pattern recognition in visual periphery is based on “textural” statistics calculated over relatively large receptive fields (Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011). Actually, second-order features of my model are equivalent to certain correlation statistics. However, the present study indicates that a relatively small number of optimally selected statistics might be sufficient to accomplish simple visual tasks. Note that the present model used only eight second-order features as compared with about 700 statistics used in the popular Portilla and Simoncelli (2000) texture model. Also, we should consider a pooling according to the maximum rule besides summation-based statistics within receptive fields.

The present results do not exclude the possibility that some alternative models can fit my data as well or even better. Also, the proposed model is highly simplified and several details are probably wrong. Still, the main assumptions (computation of second-order features, encoding the positions of features relative to candidate objects) seem to be necessary for the explanation of the present data.
Conclusions

This study shows that a simple psychophysical crowding experiment can reveal interesting aspects of computations carried out within the biological visual system. The results are consistent with the main ideas behind the “standard model” of object recognition and support some more specific assumptions about the set of elementary features and principles which guide the formation of higher-order features.

Keywords: crowding, object recognition, peripheral vision, visual features, relative positions, spatial pooling

Acknowledgments

This study was supported by Estonian Ministry of Education and Research, projects SF0180027s12 and IUT20-40. I thank Valdar Tammik for help in running experiments and Preeti Verghese, Jaan Aru, and three anonymous reviewers for their useful comments and suggestions.

Commercial relationships: none.
Corresponding author: Endel Põder.
Email: endel.poder@ut.ee.
Address: Institute of Psychology, University of Tartu, Tartu, Estonia.

References


