Corner detection in color images through a multiscale combination of end-stopped cortical cells

R.P. Würtz[a,*], T. Lourens[b]

[a]*Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany*
[b]*Symbiotic Intelligence Group, Japan Science and Technology Corporation, Shibuya-ku, Tokyo 150-0001, Japan*

Abstract

We assess the corner-detection capabilities of a model for end-stopped cells in the visual cortex (F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, O. Kübler, Simulation of neural contour mechanisms: from simple to end-stopped cells, Vision Research 32(5) (1992) 963–981). The responses of one of these cells alone cannot account for the percept of a corner. This shortcoming can be greatly alleviated by a combination over several scales. The resulting corner detection method can deal with high frequency texture, low contrast, and rounded corners and is competitive in comparison with other corner detectors. Starting from known cortical cell types we hypothesize a color-sensitive equivalent of simple cells. This allows to extend corner detection to color-sensitive channels. The combination of grey-scale and color corner-detection yields a biologically plausible model of corner perception and may also be of interest for computer vision applications. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Corner detection; End-stopped cells; Neural network; Multiscale methods; Color vision

1. Introduction

Corners are important image features because of their robustness with respect to changes in perspective or small distortions. As they are zero-dimensional, they do not pose an aperture-problem, and thus allow efficient matching for recognition purposes [1]. It is known that removing the corners from images can impede human recognition performance drastically, while removing much of the edge information does not [2].

Robustness and speed of corner matching can be improved by adding specific feature attributes to the corners [3]. This makes computationally "expensive" detection algorithms attractive, because they usually yield a variety of information beside the corner location. Useful attributes are *subtended angle*, *sharpness*, *size* (i.e. the range of scales where the corner exists), and the distribution of grey-values or colors in the neighborhood of the corner.

For clarity of the subject we define a corner as follows. A *junction* is an image point where one or more line or edge segments end. A *corner* is a junction where one of the angles

between line segments is larger than 180°. Thus, junctions of the V-, I-, and ↓-type are corners, while such of Y-, T-, K-, or X-type are not.

Our method for detecting corners will yield position, sharpness, size and color and grey-scale distribution (contrast). The subtended angle can be determined a posteriori by following the line segments that constitute the corner.

Our interest in corner detection is twofold. First, we want to find robust detection algorithms for technical systems, which can support model matching algorithms [1] and second, we want to provide a computational model for human vision. For the latter, an answer to the question will be proposed of how the percept of a corner might be linked to the responses of end-stopped cells.

The approach we are following is the one of a feedforward neural network, a common view for early visual processing. We are aware of the fact that not every perceived corner could be accounted for by this model. For example, corners of two different objects located close to each other will disturb the receptive fields that should give rise to corner detection. A feasible solution to this is a top-down component that finds the corners relying on the experience that the recognized objects ought to have corners there. Such processes go beyond the model we are presenting here and are mentioned only to make the limitations of the proposed model explicit.

* Corresponding author. Tel. + 49-234-322-7994; fax: + 49-234-3214-210.
E-mail address: Rolf.Wuertz@neuroinformatik.ruhr-uni-bochum.de (R.P. Würtz).

2. Processing in the visual cortex

For the early stages of vision, good functional models are available. The usual view here is hierarchical processing: Retinal cells feed their output to the lateral geniculate nucleus, whose output can be nicely modeled by convolution with a difference of Gaussians, or Laplacian of Gaussians. Their responses are combined to form orientation selective simple cells, which in turn feed their output to complex cells, which constitute the input to end-stopped cells. The basic operations are only summation of outputs and simple nonlinearities, so this could be implemented by a feedforward neural network. See Ref. [4] for an explicit model of those connections. It must be noted, however, that in spite of the good functional models this view provides for static images, it is far from clear (and currently under debate) that the anatomy works that simply.

2.1. Gabor functions as a model of simple and complex cells

In the visual system, orientation selective edge and line detectors form the first stage in corner detection. These detectors are called simple cells and are found in the primary visual cortex [5]. Various mathematical functions can be used to model responses of those cells, one of them, which we will adapt here, are two dimensional Gabor functions [6–8]. The responses of even (line-enhancing) and odd (edge-enhancing) simple cells are elegantly modeled as the real and imaginary parts of a Gabor function. In order to keep the formulae short we will use the following abbreviations throughout:

$$c = \cos\theta, \quad s = \sin\theta, \quad c_i = \cos\theta_i, \quad s_i = \sin\theta_i, \qquad (1)$$

where θ and θ_i denote the preferred orientation of the cells in question. The formula for the simple cells then becomes:

$$\hat{\mathscr{G}}_{\sigma,\theta}(x,y) = \exp\left(i\frac{\pi}{\sqrt{2}\sigma}(xc+ys)\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \qquad (2)$$

where σ is related to the width of the receptive field. We are also using the *radius r* of the receptive field, assuming that the receptive field ends where the value drops below 10^{-9}. This yields the relation $r = 4.24\sigma$.

For our simulations, the Gabor functions were slightly modified such that their integral vanishes and their one-norm (the integral over the absolute value) is independent of σ. The first is a known property of simple cells, and the second makes the responses at different scales easy to compare. The activities of a whole two-dimensional layer of simple cells are then modeled by convolution of the input image with $\mathscr{G}_{\sigma,\theta}$.

The next step in the computational hierarchy are *complex cells*, which can be modeled by the modulus of the Gabor function, i.e. each of these cells receives input from an odd and an even simple cell and the nonlinearities involved are squares and square roots. A layer of them is modeled by

$$\mathscr{C}_{\sigma,\theta}(x,y) = |I * \mathscr{G}_{\sigma,\theta}|, \qquad (3)$$

where I denotes the input image.

For computation it suffices to use orientations $\theta \in [0°, 180°)$ since $\mathscr{C}_{\sigma,\theta} = \mathscr{C}_{\sigma,\theta+180°}$. The $\mathscr{C}$-*operator* localizes intensity discontinuities but its output does not carry information about their type and polarity.

2.2. End-stopped cells

Cells that respond only to those edges or bars that *terminate* within their receptive field have first been found by Hubel and Wiesel [9]. Two types are distinguished: the *single end-stopped cells* respond well to line-ends; the *double end-stopped cells* respond best to very short line-segments or small spots, circular objects or blobs. A combination of output responses of complex cells is used as input for the end-stopped cells. We model them according to Heitger et al. [10], but our notation is slightly different. The first attempt at constructing an end-stopped operator is an approximation of the first derivative of the $\mathscr{C}$-operator in the direction orthogonal to its preferred one:

$$\hat{\mathscr{S}}_{\sigma,\theta}(x,y) = \mathscr{C}_{\sigma,\theta}(x+d\sigma s, y-d\sigma c)$$
$$-\mathscr{C}_{\sigma,\theta}(x-d\sigma s, y+d\sigma c). \qquad (4)$$

The operator which models the receptive field functionality of *double* end-stopped cells approximates the second derivative of the $\mathscr{C}$-operator:

$$\hat{\mathscr{D}}_{\sigma,\theta}(x,y) = \mathscr{C}_{\sigma,\theta}(x,y) - \frac{1}{2}\mathscr{C}_{\sigma,\theta}(x+2d\sigma s, y-2d\sigma c)$$
$$-\frac{1}{2}\mathscr{C}_{\sigma,\theta}(x-2d\sigma s, y+2d\sigma c). \qquad (5)$$

$\theta \in [0, 2\pi)$ is the orientation and d a positive constant, which represents the distance between the excitatory and inhibitory region of the end-stopped cells relative to σ. Its value is of great importance for the functioning of the model. If it is too large, the line segment (or edge) will be eliminated in the middle and two shorter line segments will remain instead of the end points. When d is chosen too small a corner will be detected at the wrong position. We have found that the choice of $d = 1.79$ constitutes a reasonable compromise.

In order to adjust the parameters of the model we determined the *correct locations* of the corners manually in the test image shown in Fig. 1. Every response of the modeled end-stopped cells other than at these corners is a *false response* and must be eliminated by further machinery. Following Heitger et al. [10] this is done by a tangential and a radial inhibiting operator, which are defined as
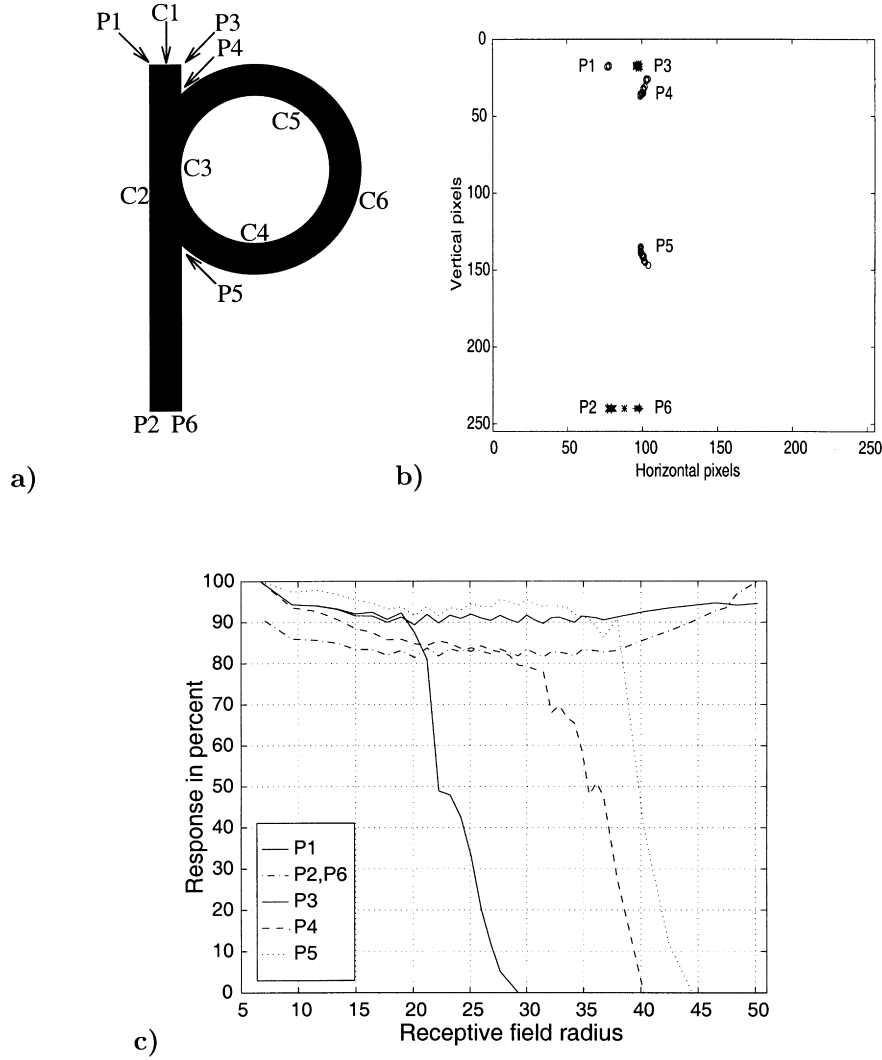
Fig. 1. (a) The character P used as input stimulus with manually marked corners; (b) the drift of these corners under variation of the receptive field radius *r*; and (c) the strength of these corners with variation of *r*.

follows:

$$\mathscr{I}_\sigma^t(x,y) = \sum_{i=0}^{2N-1} [-w_t \mathscr{C}_{\sigma,\theta_{i\backslash N}}(x,y)$$

$$+ \mathscr{C}_{\sigma,\theta_{i\backslash N}}(x + d\sigma c_i, y + d\sigma s_i)]^{\geq 0} \quad (6)$$

$$\mathscr{I}_\sigma^r(x,y) = \sum_{i=0}^{2N-1} \Bigg[[\mathscr{C}_{\sigma,\theta_{i\backslash N}}(x,y)$$

$$- w_r \mathscr{C}_{\sigma,\theta_{(i+N/2)\backslash N}}\left(x + \frac{1}{2}d\sigma c_i, y + \frac{1}{2}d\sigma s_i\right) \Bigg]^{\geq 0}$$

$$(7)$$

In these definitions, the function $[f]^{\geq 0}$ is defined to be equal to 0 where its argument is negative and equal to *f* elsewhere. This is a nonlinearity quite commonly used in neural networks, and the notation follows [10]. *N* is the

number of orientation samples in 180°, which must be even for technical reasons. $\theta_i = (i/N)180°$. The symbol "$a\backslash b$" stands for the remainder of *a* under division by *b*. The weighting factors w_t and w_r are positive constants, which must be adapted properly in order to suppress all false responses. The values $w_t = 1$ and $w_r = 4$ have been found empirically, and work well for many input images and all values of *N* and σ.

The final corner operators on a single scale in a single direction then are:

$$\mathscr{S}_{\sigma,\theta} = [[\hat{\mathscr{S}}_{\sigma,\theta}]^{\geq 0} - g(\mathscr{I}_\sigma^t + \mathscr{I}_\sigma^r)]^{\geq 0}, \quad (8)$$

$$\mathscr{D}_{\sigma,\theta} = [[\hat{\mathscr{D}}_{\sigma,\theta}]^{\geq 0} - g(\mathscr{I}_\sigma^t + \mathscr{I}_\sigma^r)]^{\geq 0}, \quad (9)$$

where *g* is a gain factor, which must be chosen appropriately. A reasonable value, found empirically, is $g = 2.0$.
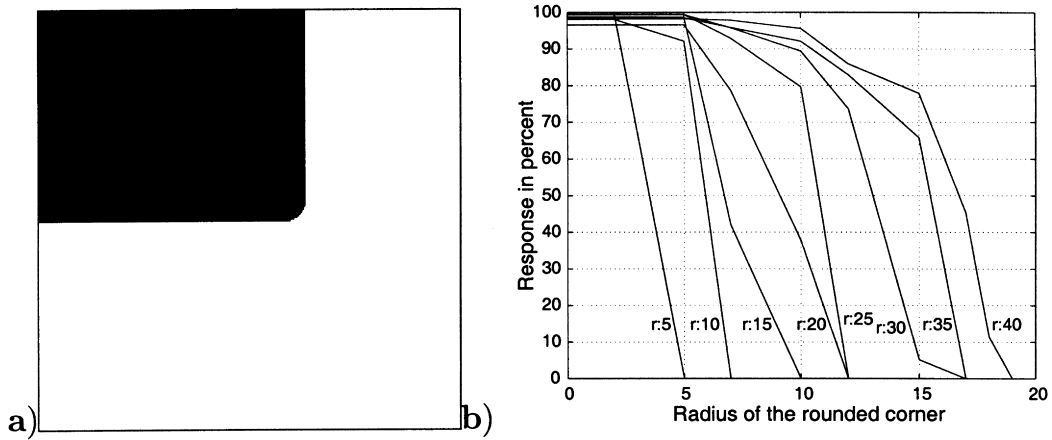
Fig. 2. (a) Synthetic image with a rounded corner; (b) shows the dependence of the response of the $\mathscr{E}_\sigma$ operator on the radius (curvature) of the corner for different values of the receptive field radius $r = 4.24\sigma$.

3. Scale-dependence of end-stopped cells

If an image contains a sharp convex corner, end-stopped cells will react on all scales. In real-world images the corners may not be ideal, and therefore the variation of the response over different scales is important. If the location of a corner should shift in an unpredictable way with a scale change a useful combination would be much harder to find than if it is relatively stable. We have scrutinized that relation using simple, artificial stimuli. The upper half of Fig. 2 shows the expected dependence of the response on the corner radius. The good results achieved by the proposed multiscale combination indicates that natural corners behave similarly. These results cannot hold in full generality because examples can be constructed where the corner position actually *must* shift when moving across scales. For example, with a large rounded corner located close to a small sharp one detection will switch from the sharp one to the large one at some point when scale is increased gradually.

The deviations of the strongest response from the correct locations we call *drift*, with "correct" meaning correct by visual inspection. Without explicit knowledge of the correct location the movement of the response under scale variation yields an estimate of the drift. For simplicity, we only look at the maximal response of the $\mathscr{S}_\theta$ and $\mathscr{D}_\theta$ operators over all orientations. This means that, for the moment being, we ignore information about the orientations that make up the corner and the distinction of corners and line-ends. For applications, this useful information can, of course, be retained. The resulting operator is:

$$\mathscr{E}_\sigma = \max_{i=0,\dots,2N-1} (\max(\mathscr{S}_{\sigma,\theta_i}, \mathscr{D}_{\sigma,\theta_i})). \tag{10}$$

In order to extract single pixels as corner points we begin by thresholding the results of $\mathscr{E}_\sigma$ at each scale with a threshold T (Fig. 3) This yields little patches, which we call *potential corner features* (PCF). These patches are thinned out by

taking local maxima, i.e. a point (x,y) becomes a *marked corner* (MC) if it is a PCF and its value of $\mathscr{E}_\sigma$ is larger than at the eight neighboring pixels. This means that each PCF must contain at least one MC, but several are possible.

We used artificial images of corners with subtended angles varying from $3°$ to $150°$ (corners with larger subtended angles remain undetected). Fig. 4 shows that corner detection is reliable and practically independent of σ. Also, the value of the $\mathscr{E}_\sigma$ operator depends very little on the angle in the range between $20°$ and $120°$. Thus, the threshold T can be chosen independently of σ, and we have used $T = 5$ throughout. The drift varied in the range of ± 3 pixels without any clear relation to σ. This is enough to prohibit the simplest possible combination by a Boolean operator on the presence of a corner at a given pixel on the single scales derived from the local maxima of $\mathscr{E}_\sigma$.

The results of the corner operator $\mathscr{E}_\sigma$ on a image of a real-world scene are shown in Fig. 3. That image poses a difficult test-case for corner detection, it has a background with many high-frequency features, and some corners are rounded, vague, or with low contrast. If we take different scales separately as illustrated in Fig. 3, we obtain many PCF areas in the background for small scales. The middle scales still show over 50% falsely detected PCF areas and fail to detect the corner between the marker and the pencil. The coarsest scale, which is the most reliable one in terms of ruling out texture still shows false responses at the round part of the scissors.

The use of a small scale produces small PCF areas, which is ideal for corner localization. On the other hand, most of these PCF areas do not contain a real corner but small details of the background. With increasing scale the size of the PCF area increases, too, and the number of PCF areas caused by the texture decreases. Fig. 3e shows areas at all corner-points but also a considerable amount of undesired PCF areas. In the largest scale the areas are large and some corners are not covered by a PCF area but the undesired ones are gone. Beside the tradeoff between localization
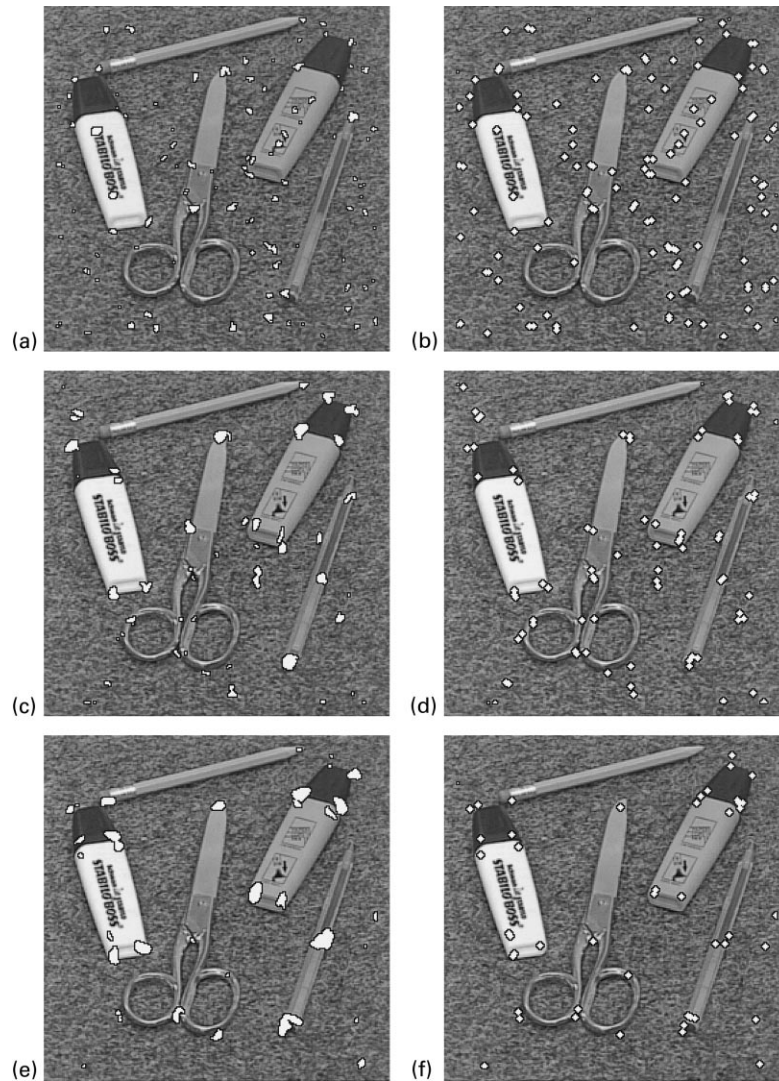
Fig. 3. The left column shows potential corner features (PCF) using the receptive field radii $r = 15$, $r = 25$, and $r = 35$, the right column the corresponding marked corners (MC). The corner marks are larger than one pixel for visualization.

and robustness, this experiment shows that for images like the one used it is not sufficient to choose one appropriate scale for corner detection, but several scales must be taken into account. It should be noted that the exclusion of the background texture is motivated by the assumption that corners are located at the end of significant lines. Clearly, line following is a different problem, and *some* corners can only be detected by a high-level perceptual decision. However, we argue that the combination of different scales yields a better low-level corner detector than any combination of filtering and thresholding on a single scale. Fig. 3 illustrates that there is no single scale at which all corners are detected. For example, the corner at the top of the ball-point at the right in the image is covered by a PCF area in the small and middle scales. The corners at the top of the left marker appear in the middle and large scales only, this makes it clear that multiple scales

are not only necessary to obtain all corners but also to eliminate the PCF areas caused by the texture.

The use of multiple scales also has the advantage that additional information about the corner can be extracted, e.g. the corner can be classified as responding to small, middle, or large scales, or a combination of them. Once we have combined the operators on multiple scales, this *signature* becomes an additional attribute of the corners, which may reduce the complexity of corner matching [3]. Analysis of the robustness of this property will be subject to future research.

4. Combination of a range of scales

Our analysis of the end-stopped operators on simple synthetic as well as real-world images has shown that the extension of the PCF areas is roughly proportional to the
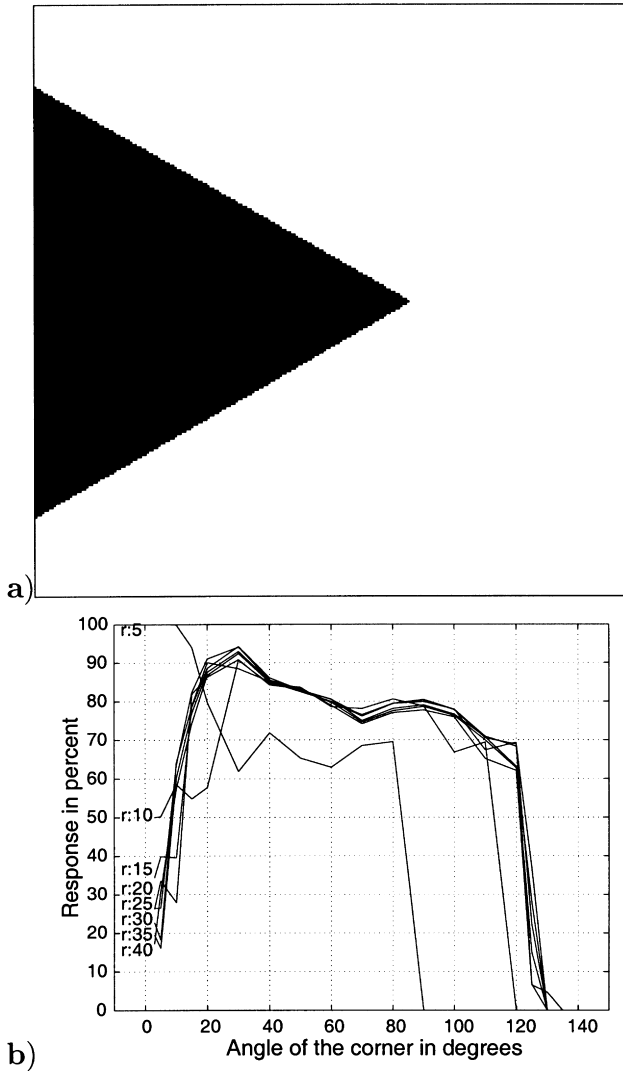
a)

b)

Fig. 4. The relative output of the $\mathscr{E}$-operator at corners with different angles. The $\mathscr{E}$-operator with receptive field radius $r = 5$ responds best to a corner with an angle of three degrees, this response is normalized to 100 and all others are relative to this response. The lines labeled from $r = 5$ to $r = 40$ represent the different receptive field radii of the modeled simple cells. The receptive field radius r is equal to 4.24σ.

4. The responses of the $\mathscr{E}_\sigma$-operator at large scales yield large spots as potential corner feature areas. These spots cover the proper locations.
5. The $\mathscr{E}_\sigma$-operators at small scales locate corners very well but are very sensitive to local changes (high frequency noise) and therefore less reliable.

After testing several combination methods containing nonlinearities such as, e.g., powers and various sampling schemes for σ we found that the simple *average* of $\mathscr{E}_\sigma$ over an equidistantly sampled range of scales is sufficient:

$$\mathscr{E}_{\mathrm{avg}}(x, y) = \frac{1}{\mathscr{S}} \sum_{i=0}^{S-1} \mathscr{E}_{\sigma_0 + i\Delta\sigma}(x, y), \tag{11}$$

This is yet another application of a feedforward neural network, the nonlinearity is introduced by the same thresholding that gives rise to the PCFs on a single scale.

So the final corner operator is achieved by:

1. evaluation of $\mathscr{E}_{\mathrm{avg}}$ on the image;
2. thresholding to obtain the PCF areas; and
3. thinning of the PCF areas to obtain single corners.

The quality of this detector is shown in Fig. 5. The parameter values for these results have been $N = 8$, $\sigma_{\min} = 1.18$, $\sigma_{\max} = 9.43$, and $S = 15$.

It would be worthwhile to compare the average operator to the actual mechanism used in human vision. However, these mechanisms are not known well enough to allow a comparison. The assumption underlying our model is that PCF-areas are detected by feedforward neural network acting on the output of complex cells. From that viewpoint, the average operator in Eq. (11) followed by thresholding is only the simplest possible network, which we have verified to produce reasonable results on real-world images. It should be noted that even the neuronal wiring of complex cells is subject of a current debate in neuroscience. The articles of Alonso and Martinez [11] and Chance et al. [12] present conflicting evidence in the same journal only half a year apart.

5. Extension to color channels

The corner detection described so far as well as most corner detectors from the literature act on grey-level images. In this section we describe a biologically motivated extension to color channels. *Opponent* color-sensitive cells are found at the first levels of processing after the photoreceptors. Some of them are orientation selective and have an elongated area which is *excited* by one color, and one or two flanks which are *inhibited* by the opposite color with color pairs blue–yellow, red–green, or vice versa [13]. In area V4, color and orientation selective cells have been found [14].

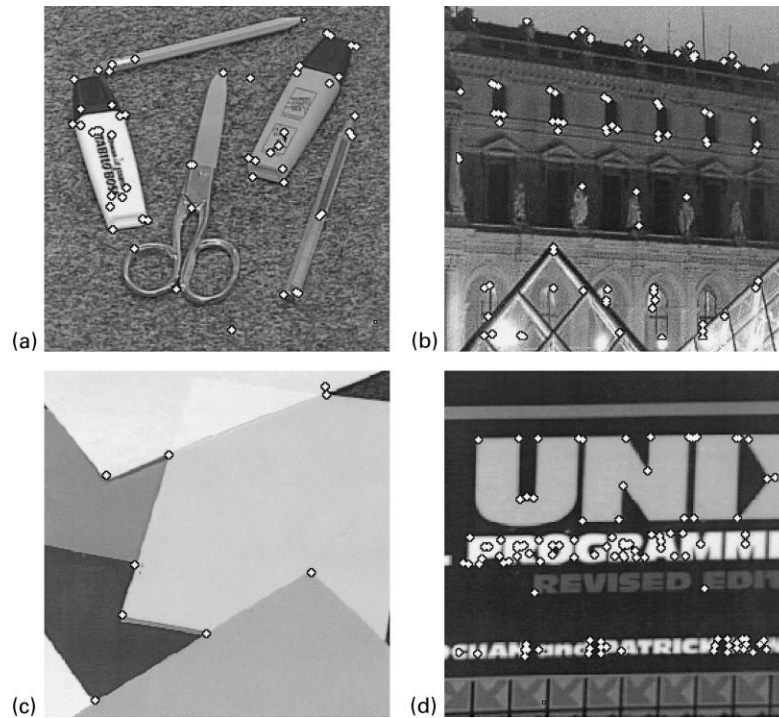In the parvo-cellular layer of the visual system of the

receptive field radius r (or the parameter σ). If the responses exist over a variety of scales, the drift is usually smaller than the width of those areas. Thus, the operation applied for combining can remain local, there is no need to integrate different locations. In detail, we have made the following observations:

1. The outputs of the $\mathscr{E}_\sigma$-operators at a corner or line-end are (almost) identical at all different receptive field sizes.
2. Usually if a corner is detected with a receptive field radius r, it is detected also in the "neighboring" receptive field radii $r \pm r'$ (In the simulations $r' = 2.5$ or 5 are used).
3. The $\mathscr{E}_\sigma$-operators respond well to sharp corners at all different scales but they do not respond to rounded corners, except for large σ.

Fig. 5. Corners detected by the detector $\mathscr{E}_{\text{avg}}$, which combines end-stopped responses from a range of scales, in a variety of natural images.

macaque monkey three different types of cells are found [15]. Two types are called color opponent cells because they always have an excitatory and an inhibitory color. The first type of these opponent cells is called the *center-surround color-opponent cell*. This type is divided into two groups, one with red-green color opponency and another one with blue-yellow opponency. The second type of opponent cells is the *center-only color-opponent cells*, and make up about ten percent of the four upper geniculate (parvo-cellular) layers. Cells of the third type, about 15% of the cells in the parvo-cellular layers, have center-surround receptive fields and do not show a color preference. Their receptive field size is about the same as for the first type of cells, whereas the receptive fields of the second type are more like those of the achromatic center-surround cells in the magnocellular pathway, which are about four times larger than the first and third type [16].

The opponent-color theory uses the above three types of cells and combines the outputs from red, green, and blue (R,G,B) cones into an achromatic channel (black and white) and two chromatic channels (red–green) and (blue–yellow). The fourth type of color-opponent cells are the *double-opponent cells*, first discovered in the goldfish [17,18]. Such cells are also found in the cortex of the monkey but not in the lateral geniculate nucleus [15]. They have excitatory and inhibitory regions for each of the two colors they are sensitive to. Double opponent cells are found in the cortical upper-layer cells in so-called *blobs*. Within the blobs of the maca-que, Livingstone and Hubel found three types of double opponent cells, known as red–green, yellow–blue, and

black–white double opponent cells. The most common double-opponent cell has red–green color opponency [13,34]. It seems likely that cells in the blobs project to cells around the blobs, a few of which are specific for both color and orientation [13,19]. In all species that have been studied, double opponent cells are found one stage higher than opponent cells and are probably formed by convergence of the opponent color cells [19].

In this paper we use color images in the (r, g, b)-format as provided by a color video camera for input and ignore all subtleties about variations in spectral sensitivity of the real photoreceptors. That is, the grey-scale image I from the first sections is replaced by the triple (I^r, I^g, I^b). The first transformation is calculating the remaining channels yellow and grey-scale (achromatic):

$$I^y = \frac{1}{2}(I^r + I^g), \tag{12}$$

$$I^a = \frac{1}{3}(I^r + I^g + I^b). \tag{13}$$

Then the color images are transformed into the achromatic (upper index a), the red–green and the blue–yellow channel. For simplicity of the model we assume that the M- and P-pathway have similar properties and that the relative outputs of the cones is similar to the (r, g, b) triple. In the P-pathway *double opponent color sensitive simple cells* can be obtained in a similar way as the simple cells. Their receptive
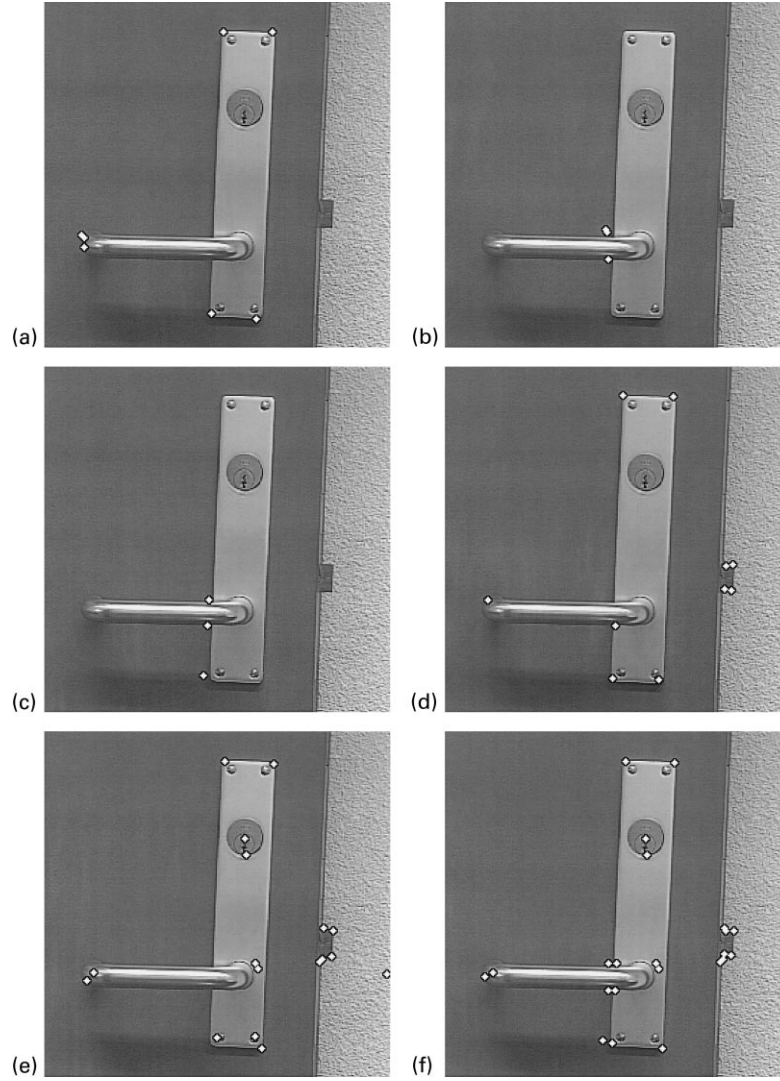
Fig. 6. Corners detected in the color image of a blue door with a shiny knob: (a)–(d) show the results in the (red, green), (green, red), (blue, yellow), and (yellow, blue) channels, respectively; (e) the results in the achromatic channel; and (f) corners detected by combining all five channels with the $\mathscr{E}_{\text{all}}$-operator.

field properties are modeled by:

$$\mathscr{G}_{\sigma,\theta}^{e,i} = I^e * G_{\sigma,\theta} - I^i * G_{\sigma,\theta}, \tag{14}$$

where the variables (e, i) stand for one of the four pairs (r, g), (g, r), (y, b), and (b, y).

Proceeding in the same way as in Section 2 we obtain the following operator for the double opponent color sensitive complex cell (analogous to Eq. (3)):

$$\mathscr{C}_{\sigma,\theta}^{e,i}(x,y) = |\mathscr{G}_{\sigma,\theta}^{e,i}| \tag{15}$$

Consequently, the following double opponent color operators are constructed: $\mathscr{S}_{\sigma,\theta}^{e,i}$, $\mathscr{D}_{\sigma,\theta}^{e,i}$, $\mathscr{E}_{\sigma}^{e,i}$, $\mathscr{E}_{\text{avg}}^{e,i}$, which are the color-sensitive equivalents of the end-stopped cells and corner detectors, respectively.

The percept of a corner is not confined to a single channel, rather corners from any channel are accepted. To model this, we finally combine corner detectors from all channels by:

$$\mathscr{E}_{\text{all}}(x,y) = \mathscr{E}_{\text{avg}}^{a}(x,y) + \mathscr{E}_{\text{avg}}^{\text{r,g}}(x,y) + \mathscr{E}_{\text{avg}}^{\text{g,r}}(x,y)$$
$$+ \mathscr{E}_{\text{avg}}^{\text{b,y}}(x,y) + \mathscr{E}_{\text{avg}}^{\text{y,b}}(x,y). \tag{16}$$

In Fig. 6 the detected corners are illustrated by using the MC-operator, for better visualization we illustrated only the corners in color. The door is blue, the handle is shiny, and the background has a grayish texture. Fig. 6e shows that the combination of all five channels is necessary to find all corners (Fig. 6f). Combining corners at multiple scales can cause a local maximum in a channel to disappear in the combination of channels and vice versa a local maximum can appear due to summing the responses of different channels which are not in one of the single channels.

The use of color is not primarily intended to improve technical corner detection. Of course, it is possible to construct images that have corners only in the color channels but not in
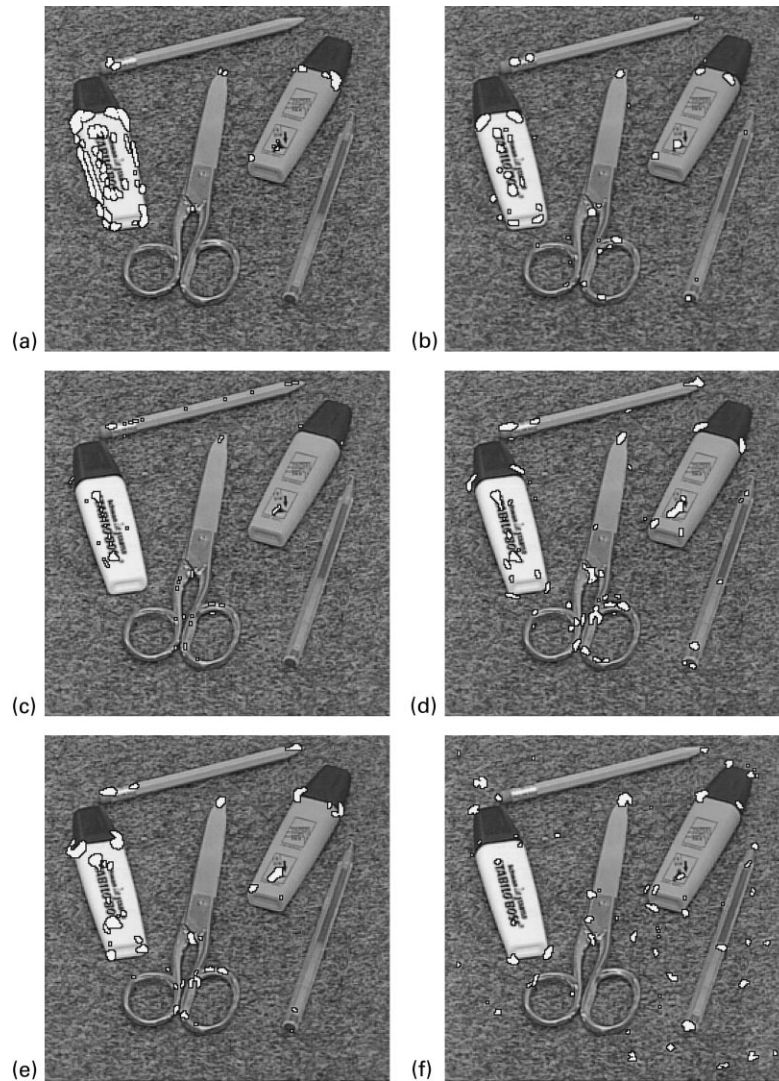
Fig. 7. Thresholded results of the outputs at a single scale of different operators: (a)–(e) the PCF areas obtained with the Diff-operator, DET-operator, Kitchen–Rosenfeld operator, Plessey feature point operator, and Plessey–operator, respectively. The outputs of the operators are thresholded at 100, 2, 3, 15, and 0.4. The parameters used are: $\mu = 1$, and $\sigma = 3$ for the Gaussian kernel; and (f) The result of our operator at $r = 25$, thresholded at $T = 5$.

the intensity channel. Given the observation that those are readily detected by a human observer *and* the hypothesis that a bottom-up process is responsible for corner detection we would expect that similar machinery be present in the color channels.

6. Discussion

We have analyzed properties of the model for end-stopped cells proposed in Ref. [10]. In order to link the responses of these cells with the percept of a corner we have proposed to use the pointwise average of their responses over a range of scales.

Furthermore, by hypothesizing a new kind of color-sensitive cells with receptive field properties similar to simple cells we have extended corner detection to color channels.

Fig. 6 shows an example where only the combination of achromatic and color channels can reveal all corners.

We have compared the single-scale corner detection with five standard corner detection operators, namely the ones from [20–24]. On single scales their performance was worse than $\mathcal{E}_\sigma$ (Fig. 7), which may be expected given the higher computational cost of the latter. The scale problems we have discussed apply to all of these corner detectors. The same holds for the *optimal* corner detector developed in [25]. We were not aware of the existence of that detector by the time we made the experiments, and therefore it is not included in the comparison. On the other hand, the optimality of image processing operators depends on the applicability of the underlying feature model, which is violated frequently in real-world images.

After combining scales with the average operator, the Plessey feature point operator [24] scored best and yielded results comparable to our operator. All other operators were
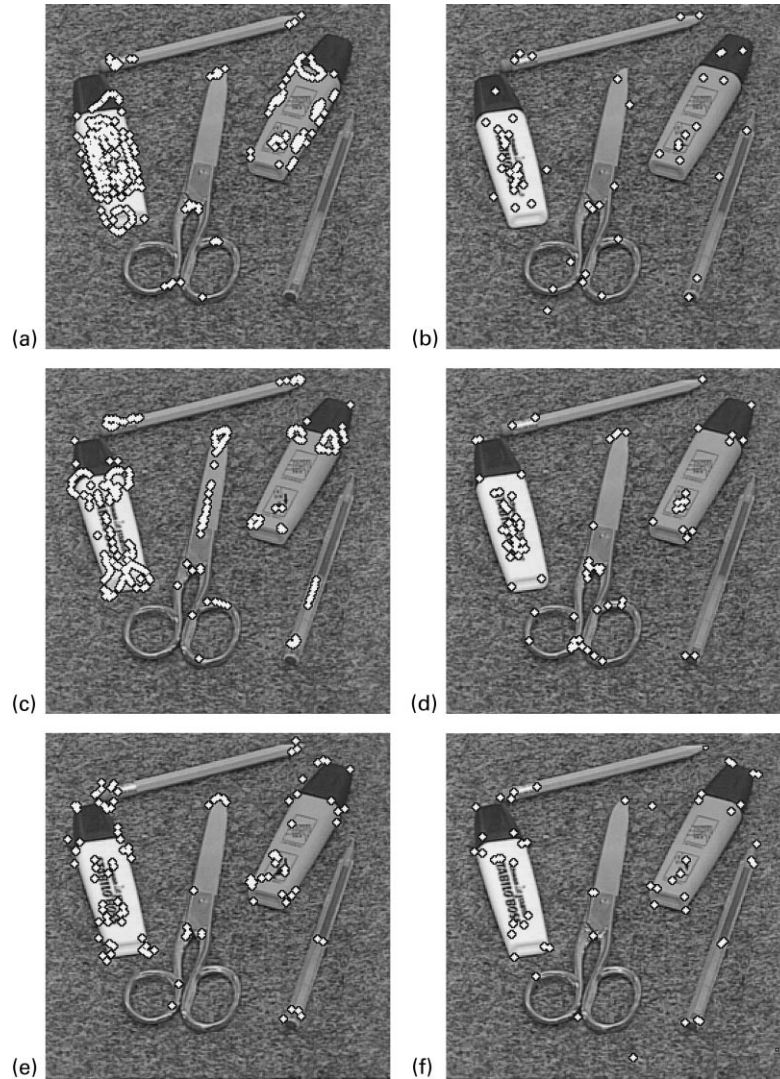
Fig. 8. Marked corners at multiple scales (combined with the avg-operator) of different standard operators: (a)–(e) the marked corners obtained with the Diff-operator, DET-operator, Kitchen–Rosenfeld operator, Plessey feature point operator, and Plessey-operator, respectively. The outputs of the operators are thresholded at 5.10^4, 200, 25, 10^6, and 0.005. The outputs of the Diff- and Kitchen–Rosenfeld operators are thresholded at the extrema and the outputs of other operators are thresholded at positive values. The parameters used are: $\mu = 1$, and linearly increasing receptive fields ranging from $r = 5$ to $r = 40$ with a step size of 2.5; and (f) the result of our operator.

weaker in the sense that they detected too many texture details or strong contrast differences rather than corners. This shows that our approach yields not only a model for corner perception but also a competitive corner detector for computer vision (see Fig. 8 for the results of the different detectors). We have tried to choose good parameter sets for all of the operators, but are not sure if we have found optimal ones in each case. The results from Fig. 8 at least suggest that other corner detectors require different methods of combination over scales than $\mathscr{E}_\sigma$.

The crucial prerequisite for applying the corner detector proposed here is a small drift, i.e. the displacement of a detected corner under variation of the detector scale must be smaller than the corresponding variation of the PCF area. For edge detectors, the equivalent of drift has been studied extensively using edge models in Ref. [26] and in Ref. [27]. A comparable

theoretical analysis to our knowledge does not exist for corner detection. On the other hand, we have deliberately avoided using a corner model, because of the known limitations of such model assumptions. Instead we have explored a "synthetic" approach, which uses known facts from biological vision, relying on the assumption that living systems have mechanisms to develop feature detectors optimally suited to the images they have to process. Nevertheless, we agree that further theoretical analysis of the proposed detector is necessary, although complicated by the form of the Gabor functions and even more by the nonlinearities throughout the processing path. Lately, an attempt has been made to understand better the nonlinearity in the formation of complex cells, i.e. the modulus of a Gabor-filtered image [28], but a complete analytical treatment of the endstopped cell model underlying our corner detector currently seems out of reach.

The reliability of the proposed corner detection method has made it possible to use it for the construction of a symbolic representation of objects as edge graphs. We are currently testing the applicability of this data format. First promising results have been published in Ref. [1].

The methods employed here and the underlying philosophy bear some resemblance to the concept of *scale space* [29]. For a recent overview of scale space techniques for the detection of corners and other features see Ref. [30]. The major difference between (Gaussian) scale space and the use of Gabor filters is the explicit representation of local orientation in the latter. The Gabor transform derives a four dimensional representation (two spatial plus two frequency coordinates or, alternatively, scale and orientation) of the image. This representation, sometimes called *phase space*, may be viewed as an extension of the original scale space concept by the extra orientation coordinate, at least if both scales and orientations are sampled densely, which is usually avoided due to the computational effort required. What is lacking is the elegant and analytically fruitful method of moving along the scale coordinate by means of a differential equation.

A major drawback of the Gabor phase space approach is the multitude of convolution operations, which makes it expensive in terms of computing time. Analysis of a 512×512 pixel color image takes about 30 min on a Pentium I with 233 MHz and 256 MB RAM under Linux. Most of this time is spent for the convolution with 360 Gabor functions (15 scales, 8 orientations and 3 color channels). Given the broad usability of the Gabor transform (see, e.g. Refs. [31–33]) it can be expected that it will soon be available in hardware. Given such a front end technical applications of this approach will be no problem. Also, closer analysis may allow to reduce the number of Gabor scales required.

Acknowledgements

References

[1] T. Lourens, R.P. Würtz, Object recognition by matching symbolic edge graphs, in: R. Chin, T.-C. Pong (Eds.), Computer Vision—ACCV'98, Lecture Notes in Computer Science, 1352, Springer, Berlin, 1998, pp. 193–200.

[2] I. Biedermann, Recognition-by-components: a theory of human image understanding, Psychological Review 94 (2) (1987) 115–147.

[3] P.L. Rosin, Augmenting corner descriptors, Graphical Models and Images Processing 58 (3) (1996) 286–294.

[4] J.J. Atick, A.N. Redlich, Mathematical model of the simple cells in the visual cortex, Biological Cybernetics 63 (1990) 99–109.

[5] D.H. Hubel, T.N. Wiesel, Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat, Journal of Neurophysiology 28 (1965) 229–289.

[6] S. Marcelja, Mathematical description of the responses of simple cortical cells, Journal of the Optical Society of America 70 (1980) 1297–1300.

[7] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America A 2 (7) (1985) 1362–1373.

[8] J. Jones, L. Palmer, An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex, Journal of Neurophysiology 58 (1987) 1233–1258.

[9] D. Hubel, T. Wiesel, Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex, Journal of Physiology 160 (1962) 106–154.

[10] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, O. Kübler, Simulation of neural contour mechanisms: from simple to end-stopped cells, Vision Research 32 (5) (1992) 963–981.

[11] J.-M. Alonso, L.M. Martinez, Functional connectivity between simple cells and complex cells in cat striate cortex, Nature Neuroscience 1 (5) (1998) 395–403.

[12] F.S. Chance, S.B. Nelson, L. Abbott, Complex cells as cortically amplified simple cells, Nature Neuroscience 2 (3) (1999) 277–282.

[13] M.S. Livingstone, D.H. Hubel, Anatomy and physiology of a color system in the primate visual cortex, Journal of Neuroscience 4 (1984) 309–356.

[14] S. Zeki, A Vision of the Brain, Blackwell, London, 1993.

[15] D.H. Hubel, Eye, Brain and Vision, Scientific American Library, New York, 1988.

[16] R. Shapley, V.H. Perry, Cat and monkey retinal ganglion cells and their visual functional roles, Trends in Neuroscience (TINS) 9 (1986) 229–235.

[17] N.W. Daw, Colour-coded ganglion cells in the goldfish retina: extension of their receptive fields by means of new stimuli, Journal of Physiology 197 (1968) 567–592.

[18] N.W. Daw, The psychology of colour vision, Trends in Neuroscience 7 (1984) 330–335.

[19] D.H. Hubel, T.N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, Journal of Physiology 195 (1968) 215–243.

[20] P.R. Beaudet, Rotational invariant image operators, in: 4th International Conference on Pattern Recognition, Tokyo, 1978, pp. 579–583.

[21] L. Kitchen, A. Rosenfeld, Gray-level corner detection, Pattern Recognition Letters 1 (1982) 95–102.

[22] J.J. Koenderink, A.J. van Doorn, Representation of local geometry in the visual system, Biological Cybernetics 55 (1987) 367–376.

[23] C. Harris, Determination of ego-motion from matched points, in: Proceedings of the Alvey Vis. Conf., Cambridge, UK, 1987.

[24] Harris, C., Stephens, M., A combined corner and edge detector, in: Proceedings of the 4th Alvey Vis. Conf., Manchester, 1988, pp. 189–192.

[25] K. Rangarajan, M. Shah, D.V. Brackle, Optimal corner detection, Computer Vision, Graphics, and Image Processing 48 (1989) 230–245.

[26] V. Berzins, Accuracy of Laplacian edge detectors, Computer Vision, Graphics, and Image Processing 27 (1984) 195–210.

[27] D.J. Williams, M. Shah, Edge contours using multiple scales, Computer Vision, Graphics, and Image Processing 51 (1990) 256–274.

[28] I.J. Wundrich, R.P. Würtz, C. von der Malsburg, Image representation by the magnitude of the discrete Gabor wavelet transform. Submitted for publication.

[29] A. Witkin, Scale-space filtering, in: Proceedings of the 8th IJCAI, Karlsruhe, Germany, 1983, pp. 1019–1022.

[30] T. Lindeberg, Feature detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 79–116.

[31] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers 42 (3) (1993) 300–311.

[32] R.P. Würtz, Object recognition robust under translations, deformations and changes in background, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 769–775.

[33] M. Becker, E. Kefalea, E. Maël, C. von der Malsburg, M. Pagel, J. Triesch, J.C. Vorbrüggen, R.P. Würtz, S. Zadel, GripSee: A gesture-controlled robot for object perception and manipulation, Autonomous Robots 6 (2) (1999) 203–221.

[34] M. Livingstone, D. Hubel, Segregation of form, color, movement, and depth: anatomy, physiology, and perception, Science 240 (1988) 740–749.