Improved Line/Edge Detection and Visual Reconstruction

João Rodrigues
Escola Superior de Tecnologia

J.M.H. du Buf
Vision Laboratory - FCT

University of Algarve - Faro, Portugal
{jrodrig,dubuf}@ualg.pt

Abstract

Lines and edges provide important information for object categorization and recognition. In addition, one brightness model is based on a symbolic interpretation of the cortical multi-scale line/edge representation. In this paper we present an improved scheme for line/edge extraction from simple and complex cells and we illustrate the multi-scale representation. This representation can be used for visual reconstruction, but also for non-photorealistic rendering. Together with keypoints and a new model of disparity estimation, a 3D wireframe representation of e.g. faces can be obtained in the future.

Keywords

Visual cortex; line and edge detection; multi-scale; visual reconstruction; non-photorealistic rendering.

1. INTRODUCTION

Computer graphics, for example (non-)photorealistic rendering (NPR), is closely related to visual perception. Trying to understand perception requires insight into low-level processes in the eyes and visual cortex. In addition, artistic rendering in the style of a certain painter requires insight into the painter's techniques (brushes, strokes, color palettes) and high-level cognitive processes [Zeki00, Livingstone00]. The latter relate to image interpretation and certain effects, for example illusions like brightness and color induction, i.e. the opposite effects of simultaneous contrast (a high surround brightness pushes the center towards a lower level) and assimilation (a high surround may also pull up the center brightness), see [du Buf95].

Because such opposite center-surround effects are not yet well understood, there is a need to develop e.g. brightness models from which we can learn. Such models must be based on state-of-the-art models of visual representations in the primary cortex: area V1 and beyond.

In this paper we present an improved model for line and edge detection, suitable for modeling brightness and NPR, concentrating on the multi-scale representation, curvature and visual reconstruction. This model is based on simple and complex cells in V1, and complements existing models for end-stopped cells [Heitger92], keypoints [Wütz00, Rodrigues04b], saliency for focus-of-attention [Deco04, Rodrigues05a], inhibition [Grigorescu03, Petkov93] and disparity [Fleet91, Rodrigues04a]. Together, these models provide input for object and face detection [Rodrigues05a, Rodrigues05b] if embedded into an architecture with ventral and dorsal data streams: the where and what systems [Deco04, Rensink00].

2. CELL MODELS AND NCRF INHIBITION

Gabor quadrature filters provide a model of cortical simple cells [Lee96]. In the spatial domain (x, y) they consist of a real cosine and an imaginary sine, both with a Gaussian envelope. A receptive field (RF) is denoted by (see e.g. [Grigorescu03]):

$$G_{\lambda,\sigma,\theta,\varphi}(x,y) = \exp\left(-\frac{\tilde{x}^2 + \gamma^2\tilde{y}^2}{2\sigma^2}\right)\cos\left(2\pi\frac{\tilde{x}}{\lambda} + \varphi\right), \qquad (1)$$

$\tilde{x} = x\cos\theta + y\sin\theta$; $\tilde{y} = x\cos\theta - y\sin\theta$, where the aspect ratio $\gamma = 0.5$ and σ determines the size of the RF. The spatial frequency is $1/\lambda$, λ being the wavelength. For the bandwidth σ/λ we use 0.56, which yields a half-response width of one octave. The angle θ determines the orientation (we use 8 orientations), and φ the phase symmetry (0 or $\pi/2$). We can apply a linear scaling between $f_{\min}$ and $f_{\max}$ with a few discrete scales or hundreds of contiguous scales.

The responses of even (Fig. 1(A)) and odd (Fig. 1(B)) simple cells, which correspond to the real and imaginary parts of a Gabor filter, are obtained by convolving the input image with the RF, and are denoted by $R_{s,i}^E$ and $R_{s,i}^O$, s being the scale, i the orientation $\left(\theta_i = i\pi/(N_\theta-1)\right)$ and N_θ the number of orientations. The responses of complex cells (Fig. 1(C)) are modelled by the modulus

$$C_{s,i}(x,y) = \left[\left\{R_{s,i}^E(x,y)\right\}^2 + \left\{R_{s,i}^O(x,y)\right\}^2\right]^{\frac{1}{2}}. \qquad (2)$$

Non-classical receptive field (NCRF) inhibition (Fig. 1(D)) can be used to suppress information in textured regions [Grigorescu03]. There are two types: (a) ani-

sotropic, in which only responses obtained for the same preferred RF orientation contribute to the suppression, and (b) isotropic, in which all responses over all orientations equally contribute to the suppression.
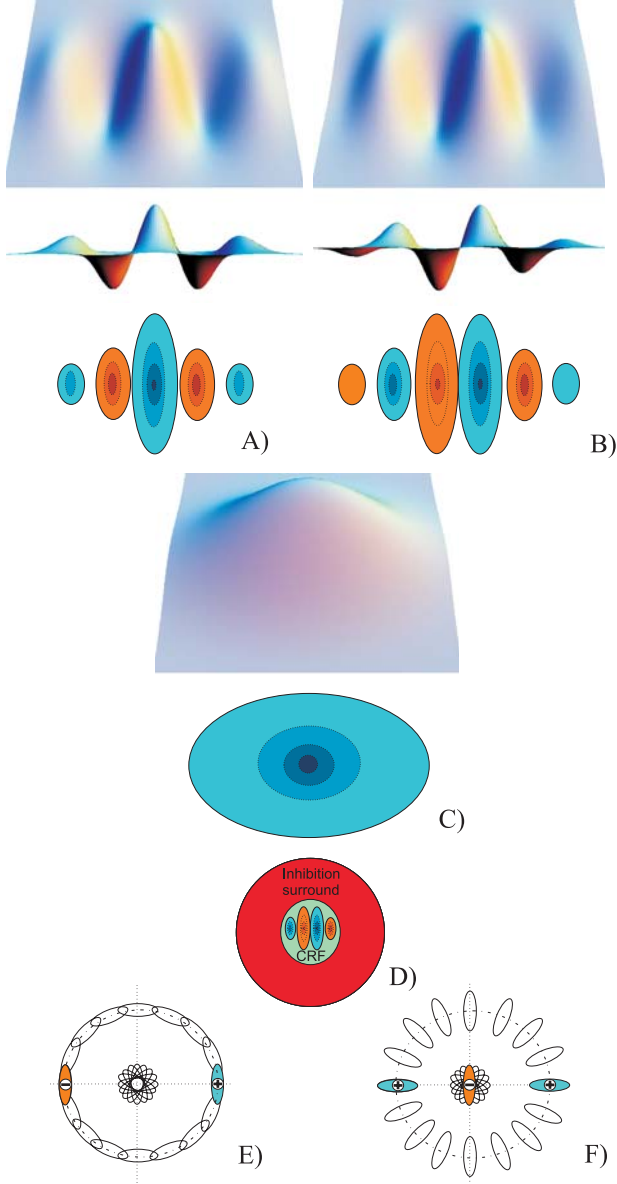
Figure 1: 3D and 2D cell representations (see text).

The anisotropic non-classical receptive field inhibition (A-NCRF) model is computed by an inhibition term $t_{s,\sigma,i}^A$ for each orientation i,

$$t_{s,\sigma,i}^A = C_{s,i} * w_\sigma , \qquad (3)$$

as a convolution of the complex cell response $C_{s,i}$ with the weighting function w_σ, with $\|.\|_1$ being the L_1 norm and $[.]^+$ denotes the suppression of negative values:

$$w_\sigma(x,y) = \left[DoG_\sigma(x,y)\right]^+ \Big/ \left\|\left[DoG_\sigma\right]^+\right\|_1 , \qquad (4)$$

$$DoG_\sigma(x,y) = \frac{1}{2\pi(4\sigma)^2}\exp\left(-\frac{x^2+y^2}{2(4\sigma)^2}\right) - \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \qquad (5)$$

The operator $b_{s,\sigma,i}^A$ corresponds to the inhibition of $C_{s,i}$, i.e. $b_{s,\sigma,i}^A = \left[C_{s,i} - \alpha t_{s,\sigma,i}^A\right]^+$, with α controlling the strength of the inhibition.

The isotropic NCRF (I-NCRF) model is obtained by computing the inhibition term $t_{s,\sigma}^I$ which does not depend on orientation i. For this we construct the maximum response map of the complex cells $\widetilde{C}_s = \max\{C_{s,i}\}$, with $i = 0,...N_\theta - 1$. The isotropic inhibition term $t_{s,\sigma}^I$ is computed as a convolution of the maximum response map $\widetilde{C}_s$ with the weighting function w_σ, and the isotropic operator is $b_{s,\sigma,i}^I = \left[C_s - \alpha t_{s,\sigma}^A\right]^+$.

3. LINE/EDGE DETECTION METHOD

Van Deemter and du Buf [Deemter96] presented a scheme for line and edge detection based on the responses of simple cells. A positive line is detected where R^E shows a local maximum in the orthogonal filter orientation and R^O shows a zero crossing. In the case of an edge the even and odd responses must be swapped. This gives 4 possibilities for positive and negative events: local maxima/minima plus zero crossings. Rodrigues and du Buf [Rodrigues04b] combined the responses of simple and complex cells, i.e. simple cells serve to detect positions and event types, whereas complex cells are used to increase the confidence. Since the use of Gabor modulus (complex cells) implies some loss of precision at vertices [du Buf93], increased precision was obtained by considering multiple scales.

The two algorithms described above work reasonably well but there are still a few problems: (a) either one scale is used or only a very few scales for increasing confidence, (b) some parameters must be optimized for specific input images or even as a function of scale, (c) detection precision can be improved, and (d) detection continuity at highly curved lines/edges must be guaranteed.

We present an improved algorithm with no free parameters, truly multi-scale, and with new solutions for problems (c) and (d). With respect to precision, simple and complex cells respond beyond line and edge ends, for example beyond the corners of a rectangle. In addition, at line or edge crossings, detection leads to continuity of the dominant events and gaps in the sub-dominant events. These gaps must be reduced in order to reconstruct continuity. Both problems can be solved by introducing new inhibition schemes, like the radial and tangential ones used in the case of keypoint operators [Rodrigues04b].

Here we use lateral (L) and cross-orientation (C) inhibition, see Fig. 1(E,F), defined as:

$$I_{s,i}^L(x,y) =$$
$$\left[C_{s,i}\left(x+d\dot{C}_{s,i}, y+d\dot{S}_{s,i}\right)-C_{s,i}\left(x-d\dot{C}_{s,i}, y-d\dot{S}_{s,i}\right)\right]^+ +$$
$$\left[C_{s,i}\left(x-d\dot{C}_{s,i}, y-d\dot{S}_{s,i}\right)-C_{s,i}\left(x+d\dot{C}_{s,i}, y+d\dot{S}_{s,i}\right)\right]^+ ;$$
(6)

$$I_{s,i}^E(x,y) = \left[C_{s,(i+N_\theta/2)}\left(x+2d\dot{C}_{s,i}, y+2d\dot{S}_{s,i}\right)- \right.$$
$$\left. 2.C_{s,i}(x,y)+C_{s,(i+N_\theta/2)}\left(x-2d\dot{C}_{s,i}, y-2d\dot{S}_{s,i}\right)\right]^+ ,$$
(7)

where $(i+N_\theta/2)\perp i$, with $\dot{C}_{s,i}=\cos\theta_{s,i}$, $\dot{S}_{s,i}=\sin\theta_{s,i}$ and $d=0.6s$. We can apply the inhibition to the complex cell responses, with β controls the strength of the inhibition (normally we use $\beta=1.0$),

$$\hat{C}_{s,i} = \left[C_{s,i}(x,y) - \beta\left(I_{s,i}^L(x,y) + I_{s,i}^E(x,y)\right)\right]^+ .$$
(8)

Line/edge (event) detection is achieved by constructing a few cell layers on top of simple and complex cells.

The first layer serves to select active regions and dominant orientations. At each position responses of complex cells are summed

$$\hat{C}_s = \sum_{i=0}^{N_\theta-1} \hat{C}_{s,i} .$$
(9)

and only at positions with $\hat{C}_s$ not equal to zero an output cell is activated. At active output cells, the dominant orientation is selected by gating one complex cell on the basis of non-maximum suppression of $\hat{C}_{s,i}$. The gating is confirmed or corrected by an excitation/inhibition process of dominant orientations in a local neighborhood.

In the second layer, event type and position is determined on the basis of active output cells (1st layer) and gated simple and complex cells. A first cell complex checks simple cells $R_{s,i}^E$ and $R_{s,i}^O$ for a local maximum (or minimum by rectification) using a dendritic field size of $\pm\lambda/4$, λ being the wavelength of the simple cells (Gabor filter).

The active output cell is inhibited if there is no maximum or minimum. A second cell complex does exactly the same on the basis of complex cells. A third cell complex gates four types of zero-crossing cells on the basis of simple cells, again on $\pm\lambda/4$. If there is no zero-crossing, the output cell is inhibited. If there is a zero-crossing, the active cell at the position of the zero-crossing cell determines event position and the active zero-crossing cell determines event type.

In the third layer, the small loss of accuracy due to the use of complex cells in the second layer is compensated. This is done by correcting local event continuity, considering the information available in the second layer, but by using excitation of output cells by means of grouping cells that combine simple and complex cells tuned to the same and adjacent orientations, see Fig. 2. The latter

process is an extension of linear grouping [Deemter96] and a simplification of using banana wavelets [Krüger97]. In the same layer event type is corrected in small neighborhoods, restoring type continuity, because the cell responses may be distorted by interference effects when two events are very close [du Buf93].
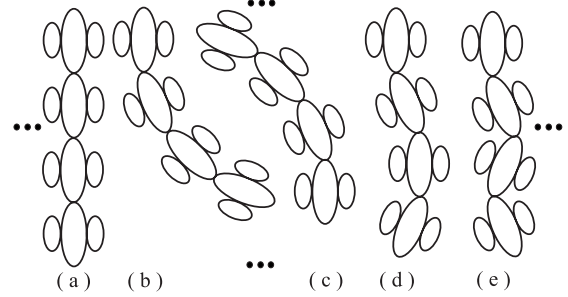
Figure 2: Curvature continuity.

Figure 3 shows four input images together with detection results (positive and negative lines and edges are coded by different gray levels). Detection accuracy is very good and there are many small events due to low-contrast textures (Fig. 3 does not show event amplitudes!) and the fact that there is no threshold value in the detection scheme.

Figure 4 shows results obtained by standard edge-only (!) detection algorithms from computer vision in the case of the orange image with, from left to right: Canny, Nalwa, Bergholm and Iverson [Heath97] (for more results see *http://marathon.csee.usf.edu/edge/edge_detection.html*).

Most events in textured regions can be suppressed by NCRF inhibition. Figure 5 shows results obtained by I-NCRF at the finest filter scale. For more results obtained with NCRF we refer to [Grigorescu03], but we note that they developed contour (edge) detection algorithms, whereas we can distinguish between edges and lines with positive and negative polarities. This is necessary for visual reconstruction; see below.

4. MULTI-SCALE REPRESENTATION

Here we focus on the multi-scale representation. Although NCRF inhibition can be applied at each scale, we will not do this for two reasons: (a) we want to illustrate line/edge behavior in scale space for applications like face detection and recognition, also visual reconstruction, and (b) in many cases a coarser scale, i.e. increased RF size, will automatically eliminate texture detail.

For illustrating scale space we can create an almost continuous, linear scaling with 320 scales between $4\leq\lambda\leq40$, but here we will present only a few scales in order to show complications. Figure 6 shows events detected at five scales in the case of ideal, solid square and star objects. At fine scales (at left) the edges of the square are detected, as are most parts of the star, but not at the very tips of the star. This illustrates an important difference between normal image processing and developing cortical models. The latter must be able to construct brightness maps, and at the tips of the star, where two edges converge, there are very fine lines. The same

effect occurs at coarser scales, until entire triangles are detected as lines and even pairs of opposite triangles (right image). In the case of the square, we show medium-scale results with little cross-orientation inhibition (edges will continue at the corners). Lines will be detected at the diagonals, which will vanish at very coarse scales (also in case of the star object).
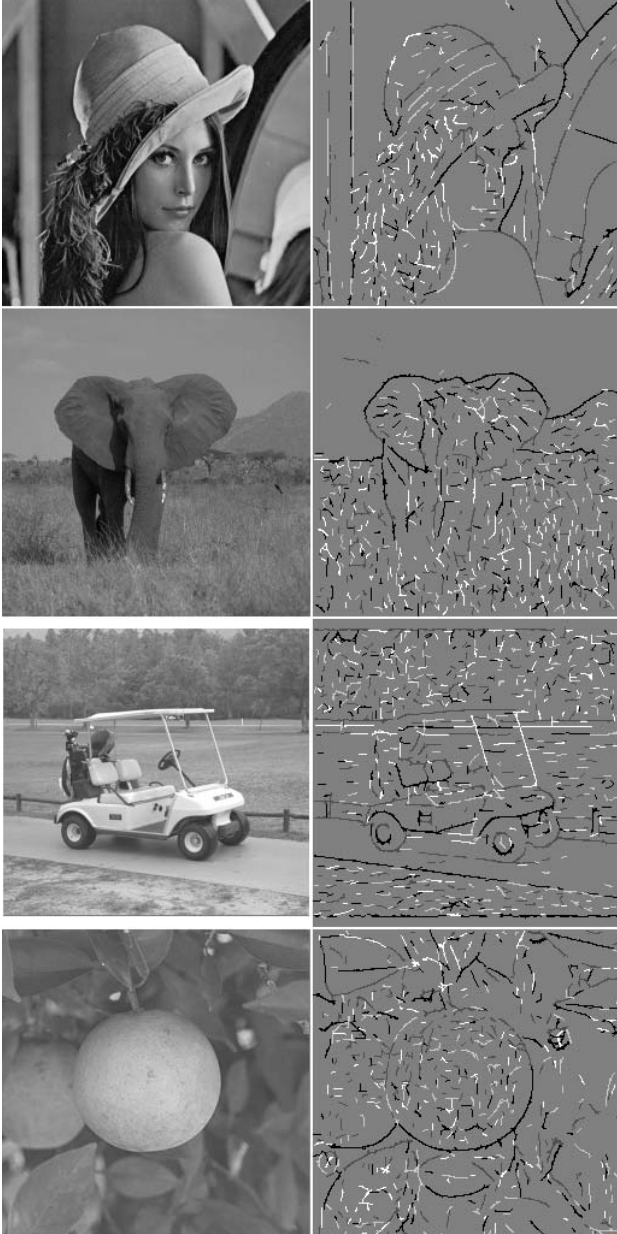

**Figure 3: Fine-scale line/edge detection (see text).**

Figure 7 shows the Fiona image with events detected at six scales, with decreasing amount of detail at coarser scales, and the result of a face detection algorithm that only exploits keypoint scale space [Rodrigues05b]. In the future, keypoints and lines/edges will be combined in order to also recognize faces in a database.

Figure 8 shows image reconstructions. These are not based on a complete set of (bandpass) wavelets complemented by one lowpass filter, a technique common in image coding, but on a virtual interpretation of lines and

edges: our visual system does not reconstruct in the normal sense, i.e. there is no cortical layer in which cell activity corresponds to input luminance. An active "line cell" is interpreted as a Gaussian intensity profile with a certain orientation, amplitude and scale, the size of the profile being coupled to the scale of the underlying simple and complex cells. In the same way an active "edge cell" is interpreted, but with a bipolar, Gaussian-truncated error-function profile. As for image coding, this interpretation must be complemented with a lowpass filter, a process that happens to exist by means of retinal ganglion cells with light-sensitive dendritic fields NOT connected to rods and cones [Berson03].
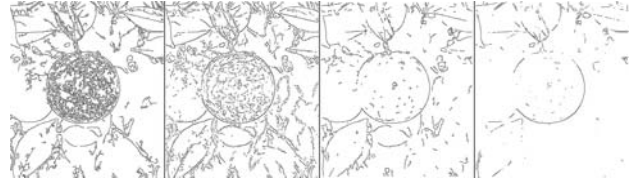
**Figure 4: Edges by Canny, Nalwa, Bergholm and Iverson algorithms.**
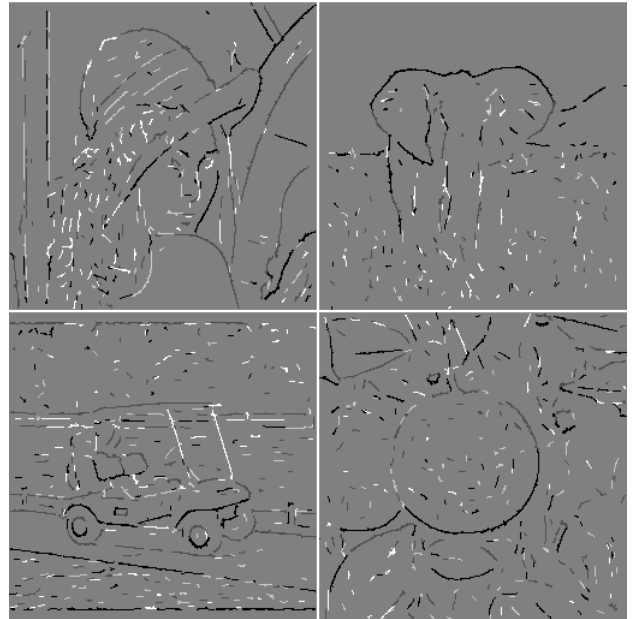
Figure 5: Line/edge detection with NCRF.

In order to show reconstructions, we must sum the lowpass (LP_σ) plus line (L_s) and edge (E_s) representations, using

$$R = \gamma.LP_\sigma + (1-\gamma).\sum_{s=1}^{N_s} W_s.(L_s + E_s), \qquad (10)$$

in which γ controls the balance between lowpass and other information, and W_s the relative weighting of the different scales. In Fig. 8 we used $\gamma = 0.5$, $W_s = 1/N_s$ and 16 equally-spaced scales $4 \le \lambda \le 20$. As can be seen, reconstructions are quite good. Better results are
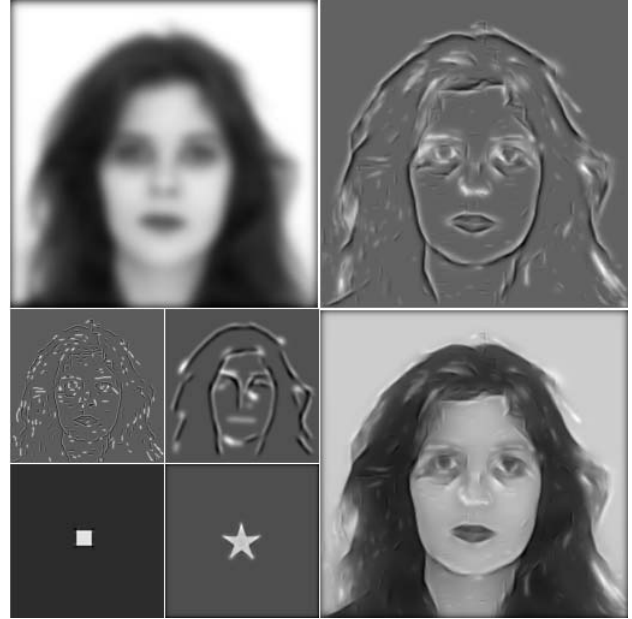
obtained when using more filter scales, but optimization requires extensive experiments with different values for γ and W_s.
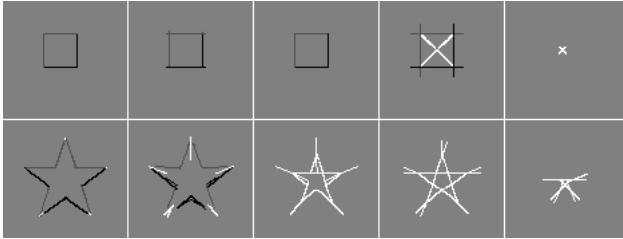
Figure 6: Multi-scale representation of a square and a star, left to right λ={4, 12, 18, 24, 40}.
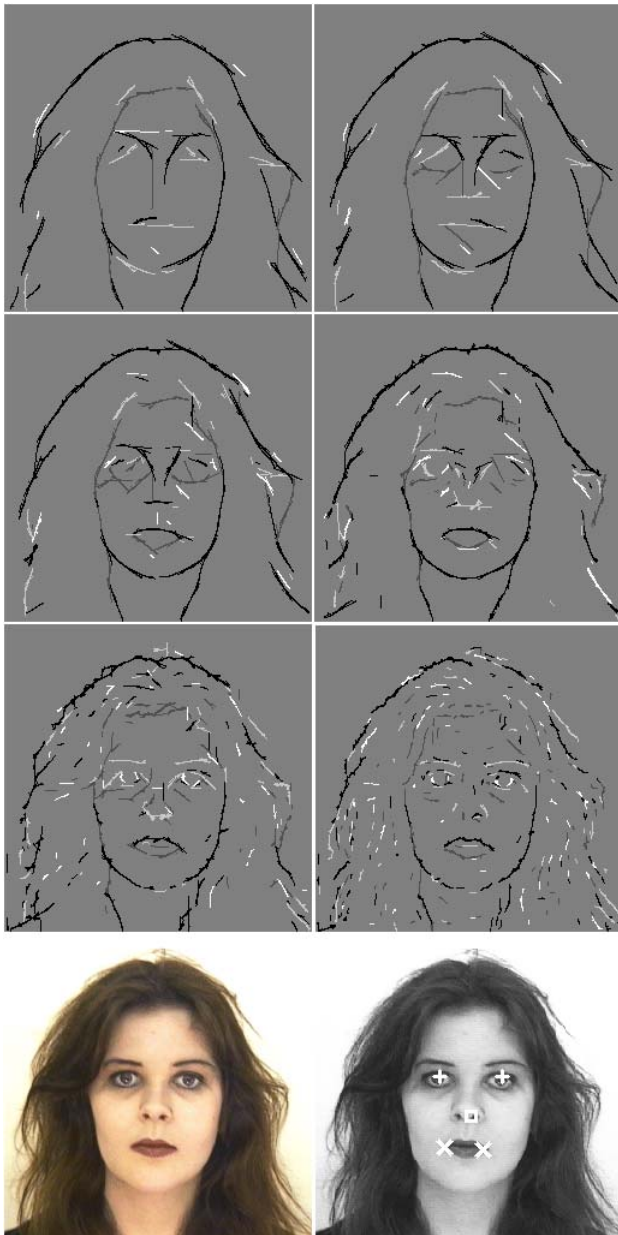
Figure 7: Multi-scale event detection on Fiona.

Figure 8: Reconstruction (see text).

5. DISCUSSION

In this paper we presented an improved scheme for line and edge detection, a scheme that can be used for creating a truly multi-scale image representation suitable for object detection/recognition and visual reconstruction.

Faces can already be detected on the basis of the multi-scale keypoint representation, i.e. by grouping keypoints at eyes, nose and mouth [Rodrigues05b], but lines and edges can (must!) complement keypoint information for developing robust face recognition algorithms, for which a cortical architecture of the what and where subsystems was explored by Deco and Rolls [Deco04].

For example, eyes, nose and mouth detectors can be complemented by a detector of the quasi-oval face outline, which is quite stable over many scales (Fig. 7). More information is available of course: eyebrows and quasi-elliptical mouth and eyes.

With respect to visual reconstruction, the symbolic representation with Gaussian line profiles and Gaussian-truncated error-function edge profiles, in combination with a lowpass filter, yields a "rendering" that is suitable for extending our brightness model [du Buf95] from 1D to 2D, for example for modeling brightness illusions.

Apart from "realistic" rendering in the case of brightness perception, non-photorealistic rendering can be obtained by selecting a few scales in combination with picking a few colors of the input image [Lam05]. Figure 9 shows an example of NPR in which the typical crayon effect was added by Corel's Photo-Paint standard "impressionist" filter. In the future we plan to develop real brushes (instead of using the line and edge profiles) and use these along simulated strokes (detected line and edge positions) in order to simulate real oil paintings and watercolors in certain styles, like Van Gogh and Turner.

In the future we can develop 3D wireframe models of objects, including faces, because we already have an optimized scheme for keypoint extraction [Rodrigues05a], which link lines and edges at vertices, and an initial model for disparity (stereo) [Rodrigues04a] that can be used to attribute depth to keypoints and therefore also to lines and edges. Such extensions are very useful for analyzing automatically facial expressions and to simulate realistic speaking avatars.

Figure 9: NPR rendering of Fiona

6. ACKNOWLEDGEMENTS

7. REFERENCES

[Berson03] D. Berson. Strange vision: ganglion cells as circadian photoreceptors. Trends in Neurosciences, 26(6): 314-320, 2003.

[Deco04] G. Deco and E.T. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res., (44):621–642, 2004.

[Deemter96] J.H. van Deemter and J.M.H. du Buf. Simultaneous detection of lines and edges using compound Gabor filters. Int. J. Patt. Rec. Artif. Intell., 14(6):757–777, 1996.

[du Buf93] J.M.H. du Buf. Responses of simple cells: events, interferences, and ambiguities. Biological Cybernetics, 68:321–333, 1993.

[du Buf95] J.M.H du Buf and S. Fischer. Modeling brightness perception and syntactical image coding. Optical Eng., 34(7):1900–1911, 1995.

[Fleet91] D.J. Fleet, A.D. Jepson and M.R.M. Jenkin. Phase-based disparity measurement. CVGIP: Image Understanding, 53(2):198–210, 1991.

[Grigorescu03] C. Grigorescu, N. Petkov and M.A. Westenberg. Contour detection based on nonclassical receptive field inhibition. IEEE Trans. IP, 12(7):729–739, 2003.

[Heath97] M. Heath, et. al., A robust visual method for assessing the relative performance of edge-detection algorithms. IEEE Tr. PAMI, 19(12):1338-1359, 1997.

[Heitger92] F. Heitger et al. Simulation of neural contour mechanisms: from simple to endstopped cells. Vision Res., 32(5):963–981, 1992.

[Krüger97] N. Krüger and G. Peters. Object recognition with banana wavelets. Proc. 5th Europ. Symp. Artificial Neural Networks, 1997.

[Lam05] R. Lam, J. Rodrigues and J.M.H. du Buf. Artistic rendering of the visual cortex. Accepted for: 2nd Workshop Luso-Galaico de Artes Digitais, V. N. de Cerveira (Portugal), 27 August, 2005.

[Lee96] T.S. Lee. Image representation using 2D Gabor wavelets. IEEE Tr. PAMI, 18(10):pp. 13, 1996.

[Livingstone00] M. Livingstone. Vision and art: the biology of seeing. Abrams, New York (NY), 2000.

[Petkov93] N. Petkov, T. Lourens and P. Kruizinga. Lateral inhibition in cortical filters. Proc. Int. Conf. Dig. Sig. Proc. and Inter. Conf. on Comp. Appl. to Eng. Sys., Cyprus:122–129, July 14-16 1993.

[Rensink00] R. Rensink. The dynamic representation of scenes. Visual Cognition, 7(1-3):17–42, 2000.

[Rodrigues04a] J. Rodrigues and J.M.H. du Buf. Vision frontend with a new disparity model. Early Cogn. Vision Workshop, Isle of Skye, Scotland, 28 May - 1 June 2004.

[Rodrigues04b] J. Rodrigues and J.M.H. du Buf. Visual cortex frontend: integrating lines, edges, keypoints and disparity. Proc. Int. Conf. Image Anal. Recogn., Springer LNCS 3211(1):664–671, 2004.

[Rodrigues05a] J. Rodrigues and J.M.H. du Buf. Multi-scale cortical keypoint representation for attention and object detection. In J.S. Marques et al., editors, 2nd Iberian Conf. on Patt. Recogn. and Image Anal, Springer LNCS 3523: 255–262, 2005.

[Rodrigues05b] J. Rodrigues and J.M.H. du Buf. Multi-scale keypoints in V1 and face detection. Accepted for: 1st Int. Symp. Brain, Vision and Artif. Intell., Naples (Italy), 19-21 October, 2005.

[Würtz00] R.P. Würtz and T. Lourens. Corner detection in color images by multiscale combination of end-stopped cortical cells. Image and Vision Computing, 18(6-7):531–541, 2000.

[Zeki00] S. Zeki. Inner vision: an exploration of art and the brain. Oxford Univ. Press, 2000.