Hans Du Buf, Joao Rodrigues

Image Morphology: From Perception to Rendering

Abstract

A complete image ontology can be obtained by formalising a top-down meta-language which must address all possibilities, from global message and composition to objects and local surface properties. In computer vision, where one general goal is image understanding, one starts with a bunch of pixels. The latter is a typical example of bottom-up processing, from pixels to objects to layout and gist. Both top-down and bottom-up approaches are possible, but can these be unified? As it turns out, the answer is yes, because our visual system does it all the time. This follows from our progress in developing models of the visual system, and using the models in re-creating an input image in the form of a painting.

1 Introduction

A trained painter is able to look at a scene and almost instantaneously take decisions concerning composition (spatial and semantic relations between objects), abstraction (which objects to paint and the level of detail) and techniques (colour palette, brushes and stroke types). Painting can be done very fast (wet-in-wet) when mixing colours on the canvas, or in different sessions (wet-on-dry) for applying new layers. Most painters who apply traditional styles will work from background to foreground, even with the possibility to start the background with dark colours and finish by high-lighting important regions using bright colours (clair-obscur or chiaroscuro). If it were possible to take a look into the brain of painters and unravel all the processes that are going on, we could develop a sound theory. We wrote “could” instead of “can” because of complications that can be expected: every painter has developed an own style, and it is likely that a specific style is related to a specific way the “input image” has been or is being analysed.

Unfortunately, we cannot take a look into van Gogh’s head and we do not know the exact landscapes that he saw. We can only analyse the paintings that he produced. Functional magnetic-resonance imaging (fMRI), which is a relatively new technology for analysing activities in brain areas, is not yet mature enough to be applied systematically. Besides, current fMRI technology lacks the resolution to analyse brain activity down to the cell level, i.e., only bigger regions and pathways between regions can be obtained. For the moment there are different and complementary solutions: (a) study composition and abstraction using methods employed in empirical aesthetics, (b) study specific visual effects such as colour and brightness using psychophysics, and (c) study available data concerning cells, layers and pathways using neurophysiology, hoping that basic processes in the brains of humans and other primates are the same or at least similar.

Here we will concentrate on visual perception and the visual cortex, without going too much into detail. One of the goals of the Vision Laboratory is to develop models of the visual cortex for explaining brightness effects and illusions, now also object categorisation and recognition. A new development is to apply low-level processing to non-photorealistic rendering (NPR), i.e., painterly rendering using discrete brush strokes. This combines two developments: a standard observer and a standard painter, with a user interface that allows to select e.g., brush and stroke types for influencing the painting process and therefore the style of the painting.

Below we first present a general description of the visual system and specific processes, including layers, pathways and cells, in the cortex. Then we illustrate how the cortical image representation can be used for NPR. We conclude with a Discussion in which we return to image ontology.

2 The visual system

The goal of our visual system, but in combination with the other senses, is to recognise objects, to establish a spatial layout of our environment, and to prepare for actions, for example looking at a computer monitor and keyboard when typing a text. All this is done automatically and very fast. In addition, the image that we perceive looks perfect for those without deficiencies—except for vision scientists familiar with illusions. However, how all this is done is still a mystery. Despite the tremendous progress in research during the past decades, there still remain many open questions although our view of basic processes has become clearer. A few aspects are the following:

2.1 The retina

The projected image on the retina is pre-processed there: rods and cones, the basic photoreceptors, are connected by horizontal cells with excitative and inhibitory synapses, a first indication for spatial (or spatio-temporal) filtering. They are also connected to bipolar cells which connect to amacrine and ganglion cells. Already 12 types of bipolar cells have been identified, with at least 4 types of ON and OFF cone-connected cells. Cones play a role in daylight colour vision whereas rods are for black-white vision when the light level is low. ON and OFF refer to light increments and decrements on a background, for example white and black spots or bars on a grey background. Amacrine cells are inhibitory interneurons of ganglion cells, and as many as 50 morphological types exist. At least 10–15 types of retinal ganglion cells have been identified. These code ON and OFF signals for spatial, temporal, brightness and colour processing, and their outputs, the axons, connect to the lateral geniculate nucleus (LGN) and other brain areas (the LGN is a relay station between the retina and the visual cortex, input area V1; see below). For further details we refer to Wässle (2004).

Most important here is that receptive fields of ON and OFF retinal ganglion cells can be seen as isotropic spatial bandpass filters, i.e., without a preferred orientation and therefore with a circularly-symmetric point spread function, often modelled by means of a “Mexican hat” function with a positive centre and a negative surround. Such filters only respond to transitions like dark-bright edges, and responses in homogeneous regions are zero or very small. The size of the receptive fields is a function of the retinal eccentricity: the fields are small in the centre (fovea) and they are increasingly bigger towards the periphery. According to another theory (!), big fields exist over the entire retina, medium fields inside a circular region around the fovea, and the smallest fields are only found in the centre of the fovea. Related to the field size is the notion of scale representation: at the point that we fixate fine-scale information is available, for example for resolving printed characters of a text we are reading, whereas the surround is blurred because only medium-and coarse-scale information is available there. The notion of scale analysis or scale representation will become clearer in Section 3.

Also important is the fact that one very specific type of retinal ganglion cell is not connected, directly nor indirectly, to rods and cones (Berson 2003); their own dendrites act as photoreceptors, they have very big receptive fields, and they connect to central brain areas for controlling the circadian clock (day-night rhythm) and, via a feedback loop, the eye’s iris (pupil size). These special cells also connect to at least the ventral area of the LGN (LGNv); hence, in principle they can play a role in brightness perception, for obtaining a global background brightness on which lines and edges etc. are projected. This is still speculative and far from trivial, but we need to keep in mind that (a) pure bandpass filters, both retinal ganglion cells and cortical simple cells (see below), cannot convey a global (lowpass) background brightness level, (b) colour information is related to brightness and processed in the cytochrome-oxidase (CO) blobs embedded in the cortical hypercolumns, colour being more related to homogeneous image (object) regions instead of to lines and edges extracted on the basis of simple cells etc. in the hypercolumns and not in the CO blobs, (c) colour constancy, an effect that leads to the same perception of object colours when the colour of the light source (illumination spectrum) changes, is intrinsically related to brightness, i.e., in a more global sense rather than object edges etc., and (d) very fine dot patterns, for example a random pattern composed of tiny black dots on a white kitchen table, are difficult to code with normal retinal ganglion cells or cortical simple cells (Zucker

& Hummel, 1986; Allman & Zucker, 1990). Colour and dot-pattern processing suggest that there are more “pathways” from the retina to the visual cortex, although the availability of a cone-sampled image in the cortex is speculative (blindsight, the ability of a blind person to sense the presence of a light source or even a moving object, points at pathways that do not lead, at least directly, to area V1 in the cortex). Most of these aspects are subject to research. An amazing fact is that, in each eye, the information of 125 million rods and cones is coded by means of about one million retinal ganglion cells. The compression rate of 0.8% is impossible to achieve by current image and video compression standards like JPEG and MPEG if image quality may not deteriorate.

2.2 The LGN

The traditional view of the LGN is a passive relay station between the retina and V1, the cortical input layer that connects to higher areas V2, V4 etc. The more recent view is that the LGN plays an active role in visual attention: perhaps only 10% of its input stems from the retina and all other input it receives by means of feedback loops from inferior-temporal (IT) and prefrontal (PF) cortex, where short-term memory is thought to reside, via V4, V2 and V1. This implies that the magno and parvo subsystems, also called the ‘what’ and ‘where’ systems or pathways in ventral and dorsal areas throughout the visual cortex, already exist at LGN level: LGNv and LGNd (Kastner *et al.* 2006). The names ‘what’ and ‘where’ stem from the functionality of the system in testing hypotheses in the interpretation of the coded input information, i.e., what there is (object categorisation and recognition) and where it is (Focus-of-Attention and eye fixations). However, it should be stressed that the LGN is not involved in object recognition. Feedback from the visual cortex only modulates information passing through the LGN.

2.3 The visual cortex

The ‘what’ and ‘where’ pathways lead to V1 and via V2 and V4 to higher areas IT and PP (posterior-parietal). In the computational model by Deco and Rolls (2004), information in the ventral ‘what’ system propagates, bottom-up, from V1 via V2 and V4 to IT cortex. The dorsal ‘where’ system connects V1 and V2 through MT (medial-temporal) to PP. Both systems are controlled, top-down, by attention and short-term memory with object representations in PF cortex, i.e., a ‘what’ component from PF46v to IT and a ‘where’ component from PF46d to PP. Deco and Rolls showed that the bottom-up (visual input code) and top-down (expected object and position) data streams are necessary for obtaining size, rotation and translation invariance in object detection and recognition: object templates in memory are thought to represent a few canonical object views, probably normalised (if we close our eyes and imagine a few objects like a cup, a bottle, a cat and a house, one after the other, they all have more or less the same size). Invariance is obtained by dynamic routing in V2 and V4 etc., such that cells at higher levels (a) have bigger receptive fields until they cover the entire visual field, (b) perform more complex tasks, for example a face detector at a high level can combine outputs of eye and mouth detectors at a lower level, the eye and mouth detectors combining feature detectors at yet lower levels, and (c) can control attention and adapt/optimise local detection processes at the lower levels. Although Deco and Rolls (2004) explored attention and invariance, they did not apply any functional feature extractions, i.e., they only used simple cells in V1 instead of line, edge, keypoint and grating cells (see Section 3, which focuses on processing in area V1). A nice example of feature extraction is the multi-scale keypoint representation in V1 and beyond for face detection: the use of keypoints (singularities like line and edge crossings and end points) for detecting eyes etc. until a face is detected, see Rodrigues & du Buf (2006c). Such a hierarchical architecture can explain the well-known Thatcher illusion: the vertically mirrored picture with normal mouth and eye regions looks fine but when it is rotated it looks terrible. Explanation: mouth and eye detectors have no problem with the friendly facial expression and a face detector groups outputs of mouth and eye detectors; the mouth can be above or beneath the eyes, for the face detector this is the same when it only groups outputs of the other detectors.

2.4 Information propagation

Although we can detect and recognise objects very fast, almost instantaneously as it seems, processing in the different cortical areas and the information propagation, both bottom-up and top-down, take time. When seeing an image for a split second, we are able to extract the gist and detect specific objects. What happens is that the flashed image enters the system and, after the computer screen goes blank again, the information propagates through the different levels (the same occurs between fixations, during saccadic eye movements when the image is not stable and the input is inhibited). Typically, objects are recognised within 150–200 ms, and first category-specific activation of PF cortex starts after about 100 ms (Bar 2004). In addition, instead of all information propagating at the same time, or in parallel, it is known that coarse-scale information propagates faster than fine-scale information to IT cortex (Bar *et al.* 2006). This suggests that object segregation, categorisation and recognition are sequential but probably overlapping processes: the system starts with coarse scales for a first test to select possible object templates, then employs medium scales in order to refine the categorisation, until finest scales are available for final confirmation of the recognition result. For another view of the cortical architecture we refer to Rensink (2000). Rensink explains the fact that the “bandwidth” of the visual system is limited: only one object can be attended at any time, although the presence of multiple objects must be stored in what he calls layout and gist subsystems. He also explains that our brain does not need to store a complete map of our entire environment; the (normally) stable environment we are looking at can be seen as external memory. Indeed, when we close our eyes we are very poor in naming colours and other aspects of objects

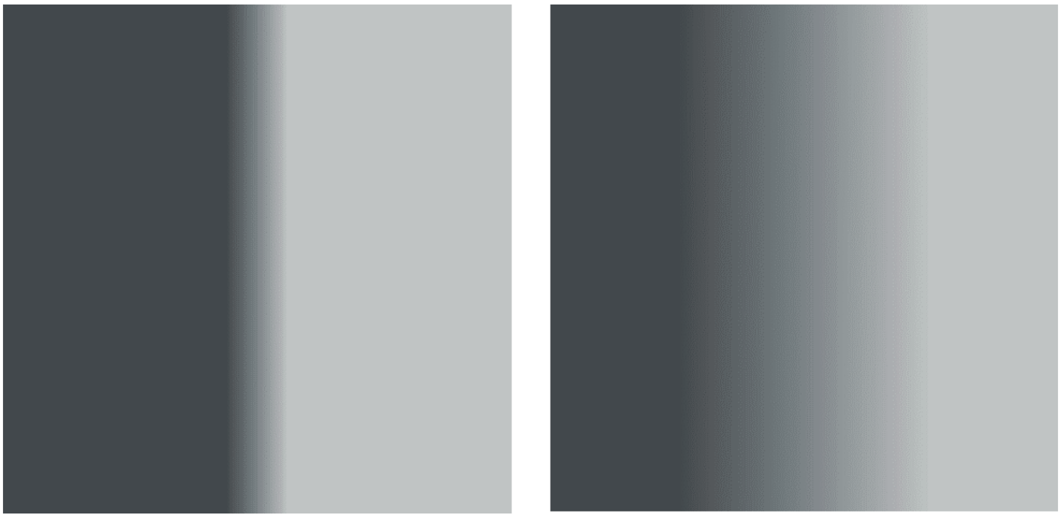

Figure 1: Both narrow and broad but linear transitions between dark and bright regions lead to the perception of Mach bands, a dark band to the left and a bright one to the right in both images. This illusion is explained in the text and in Figure 4

that are on the table in front of us. In vision science a related effect is called change blindness: when looking sequentially at two images of a house, the one with a chimney to the left and the other with the same chimney but moved to the right, only few people will notice the difference. Apparently, the house looks normal (gist), the position of the chimney is irrelevant (layout), and the system can spend its limited "bandwidth" on more important tasks, until we are told to look for differences and we start screening consciously different parts of the two images.

Above we did not address other issues like motion and disparity. In the next section we will focus on feature extractions in V1, by means of specialised cells. But some general questions remain: if things are quite complicated, with still many gaps in our knowledge, how is the image created that we perceive? Where in our brain is it created? Well, nobody knows exactly, but researchers who are developing, e.g., computational brightness models should have an idea. If we require that a brightness model should at least be able to predict Mach bands, the bright and dark bands that are seen at ramp edges (see Fig. 1), the number of published models is surprisingly small (Pessoa 1996).

If, in addition, we require that a model that can predict Mach bands should also be able to predict most of all known brightness illusions like brightness induction, with the two opposite effects of simultaneous brightness contrast and assimilation (see Fig. 2), the number of models is even smaller. Our own model was first tested on 1D patterns (du Buf 1994; du Buf & Fischer 1995), but a 2D version has already been tested and will soon be submitted for publication. It is based on a specific philosophy that answers the two questions posed above.

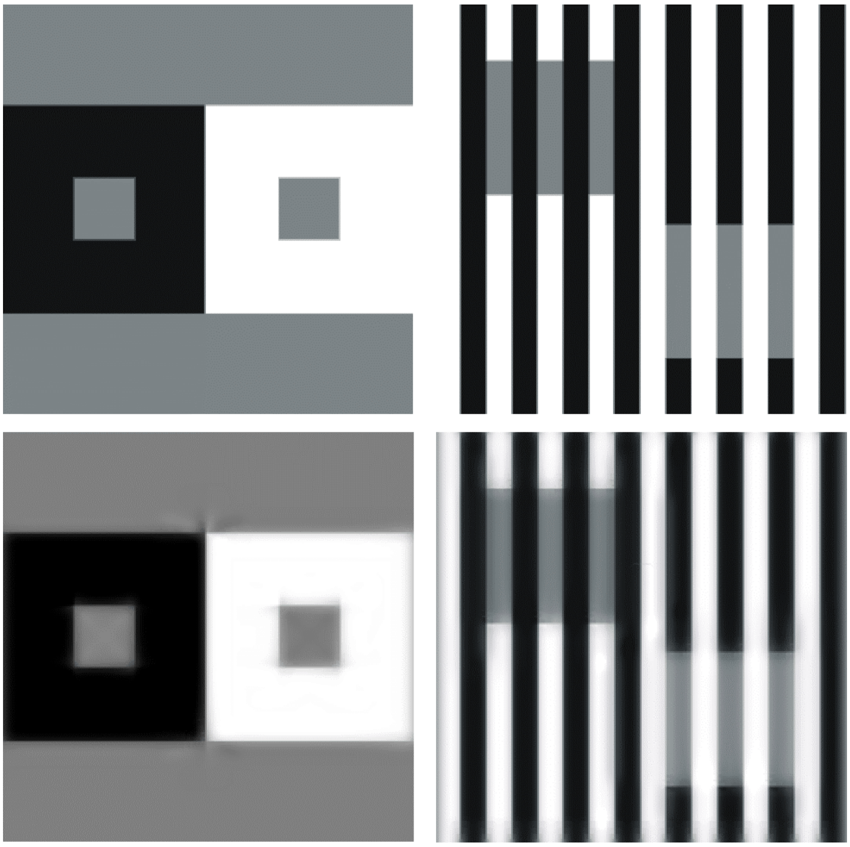

Figure 2: Top: two examples of brightness induction, simultaneous brightness contrast (left) and assimilation (right). In both images the grey squares and the bars are of the same intensity physically, but there is a big difference in our brightness interpretation; Bottom: model predictions show correct effects

3 Feature extractions in V1 and brightness perception

V1 is the input layer of the visual cortex in both left and right hemispheres of the brain. It is organised in so-called cortical hypercolumns, with neighbouring left-right regions which receive input—via the optic chiasm and one of the two LGNs—from the left and right eyes, with small “islands,” the CO blobs. In the hypercolumns there are simple, complex and end-stopped cells. Simple and complex cells are thought to serve line and edge extraction, whereas end-stopped cells respond to singularities (line/edge crossings, vertices, end points). There are many cells tuned to different scales, i.e., with receptive fields that range from very small to very big. If we penetrate the surface of the cortex perpendicularly, we find cells tuned to different orientations. Many cells are also disparity-tuned, which indicates that stereo processing starts in V1, if not already in the LGN. It is likely that stereo processing involves simple cells with non-zero phase characteristics (Ohzawa *et al.* 1997; Read & Cumming 2006).

V1 is composed of at least nine major layers, but the processing in those layers is not yet well understood. For nice overviews see Hubel (1995) and Schmolesky (2000).[1] Apart from simple, complex and end-stopped cells there also are bar and grating cells. These are specialised for extracting aperiodic bars and periodic gratings. In contrast to simple and complex cells, which can be seen as linear filters because they respond to all patterns, bar and grating cells are highly nonlinear: a bar cell does not respond to bright or dark bars in a periodic grating and a grating cell does not respond to isolated bars; see du Buf (2006) for a computational model of these cells and texture coding. There also are cells that respond to illusory contours, e.g., gaps in edges, for example caused by occluding objects like tree branches in front of other branches (von der Heydt *et al.* 1992; Heitger *et al.* 1998). Without doubt, there remain cells with other specific functions that will be discovered in the near future.

The tuning of cells to different frequencies (scales), orientations and disparities, together with the existence of e.g., bar cells, points at a multi-scale image representation: lines, edges, keypoints, gratings etc. It is even possible that disparity is attributed to extracted lines and edges, i.e., in principle it is possible to construct a 3D “wireframe” model of objects, like the solid models used in computer graphics, but this is still speculative. However, it is likely that there are at least three (interconnected) data streams within the ‘what’ and ‘where’ streams:

(1) The multi-scale line/edge representation serves object segregation, categorisation and recognition, with coarse-to-fine-scale processing, the latter also being applied to disparity in order to solve the correspondence problem. We may assume that this stream is responsible for line/edge-related brightness perception (see below).

(2) The multi-scale keypoint representation serves Focus-of-Attention (FoA), a process that directs our eyes—and mental attention—to points with a certain complexity: it does not make much sense to fixate points in homogeneous image regions where there are no structures to be analysed. In combination with motion and other cues, like colour contrast, this stream could be the basic cornerstone of the ‘where’ stream (Itti & Koch 2001; Rodrigues & du Buf 2006c).

(3) Colour and texture are surface properties of objects, normally in homogeneous regions but also with global modulations like shading due to light sources (shape-from-shading) and/or the shape of 3D objects (shape-from-texture). This shape information complements disparity information. Since lines and edges are 1D transitions (1D singularities; keypoints are 2D singularities) without colour, colour is supposed to be “sampled” and represented in the CO blobs (but see below!).

[1] cf. http://webvision.med.utah.edu/VisualCortex.html .

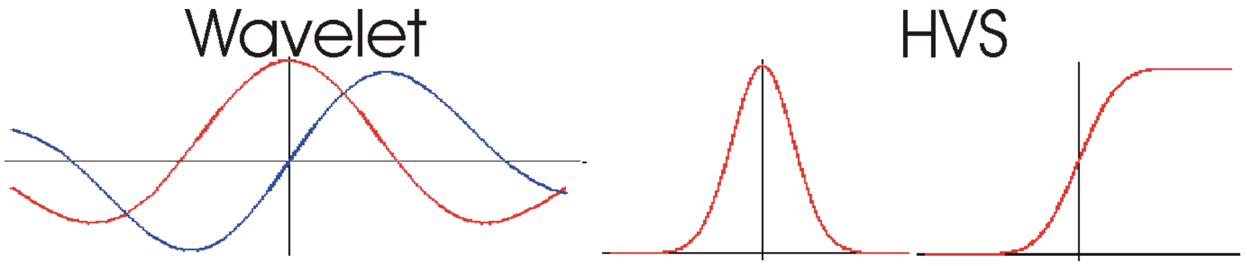

Figure 3: 1D cross sections of Gabor wavelets with sine and cosine components (left) and line and edge symbolic representations (right). A Gaussian window can truncate the error function at the far right.

This is an over-simplification of course, because FoA in textured regions can direct attention for scrutinising detail, i.e., a conscious action that may complement an unconscious process like automatic texture segregation, and global modulations (shape-from-X) can invoke different analyses. It is therefore important to stay focused on the main themes: basic processing serves (a) object structure, (b) surface structure, and (c) scene structure. Coming back to brightness processing, our model was conceived from three rather simple—not trivial—observations that are not so easy to explain to non-specialists:

(1) Simple cells are often modelled by complex Gabor (wavelet) functions, or quadrature filters with a real cosine and an imaginary sine component, both with a Gaussian envelope (see Fig. 3 (left), and du Buf (1993)). Such filters have a bandpass characteristic: the integral over the sine component is zero and the integral over the cosine component is very small or residual. Wavelets are also being used in image coding: the use of a complete set of bandpass filters tuned to all frequencies and orientations, plus one isotropic lowpass filter, which sum up to an allpass filter (a linear filter that passes all frequency components), allows to reconstruct the input image. Therefore, in principle the brain could use the same strategy: sum the activities of all simple cells plus one "lowpass channel," for example from the special retinal ganglion cells with photoreceptive dendritic fields, if available in the CO blobs, into a retinotopic projection map in some neural layer. However, this leads to a paradox: it would be necessary to construct "yet another observer" of this map in our brain. Therefore, we assume that brightness is related to the multi-scale line/edge representation, which is necessary for object recognition.

(2) Basic line and edge detection involves simple cells in phase quadrature: positive and negative lines and edges (1D cross sections) can be detected and classified by combining detectors of zero-crossings and extrema (positive or negative) of the sine and cosine components, in combination with (positive) extrema of activities of complex cells. Our previous (van Deemter & du Buf 2000) and recent (Rodrigues & du Buf 2006a) models are based on simple and complex cells and are multi-scale, since many spatial patterns cannot be described using only one or few scales. However, there is one complication: at ramp edges, where a linear ramp meets a plateau, for example in trapezoidal bars or gratings

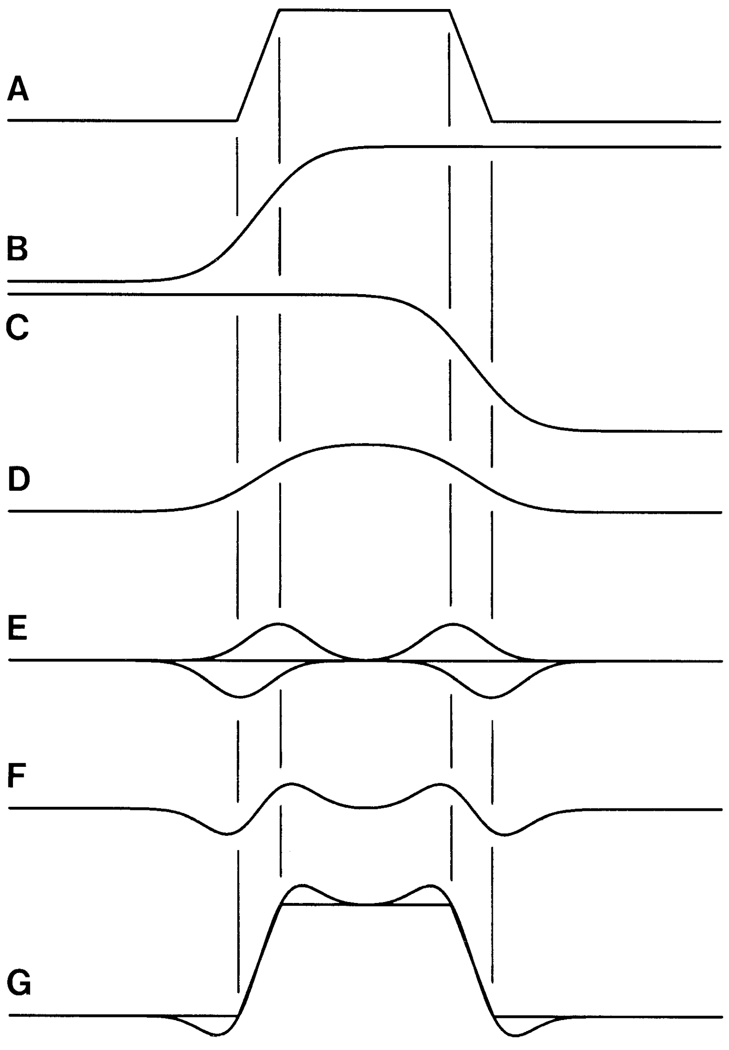

Figure 4: Mach bands at a trapezoidal luminance bar (A) can be explained by the multi-scale line/edge representation. At a very coarse scale a wide bar is detected (not shown here). At medium scales the two edges are represented by scaled error functions (B,C) which, when summed, also form a wide bar (D). At fine scales the four ramp edges are represented by positive and negative lines (E), which when summed (F) and combined with signal D create the typical overshoots (G).

(Fig. 4), the system will detect positive and negative lines. Responses of filters in quadrature do not allow distinguishing between lines and ramp edges, which explains Mach bands at ramp edges (du Buf 1994).

(3) The implicit, multi-scale line and edge representation must provide information for brightness construction by means of an interpretation. In other words, instead of a reconstruction the system builds a virtual impression on the basis of a learned interpretation of responding line and edge cells, perhaps much like a trained neural network. We “simply” assume that a responding line cell (at a certain position, tuned to a scale and orientation) is interpreted as having a Gaussian cross-profile there, with a certain amplitude (the response of the complex cell) and width (the scale of the underlying simple and complex

cells). The same way responding edge cells are interpreted, but with a bipolar (positive-negative) cross-profile and modelled by a Gaussian-windowed errorfunction (see also Fig. 10 in du Buf (1994)). Figure 4 illustrates the process in the case of a trapezoidal bar: the entire bar is represented by a very broad vertical line at coarse scales, by positive and negative edges at the ramps at medium scales, and positive and negative lines at the ramp edges at fine scales. If all is combined, the detected lines at fine scales cause Mach bands.

This model provides a completely new way for image (re)construction, not like coding based on wavelets or simple cells. An additional observation is that there is a lot of neural noise in the system and we do not know whether there exist simple and complex cells etc. at all retinotopic positions and tuned to all scales and all orientations (representation noise and completeness). Stained maps of hypercolumns and dendritic/axonal fields of most if not all cells look rather random (Hubel 1995). Nevertheless, the image that we perceive looks rather stable and complete. It is very simple to simulate what happens when we suppress information, both in the brightness model as described above and in wavelet coding, the latter being modelled by considering the summation of responses of simple cells. For example, we can suppress one entire scale channel, or 50% of all information by a random selection. Figure 5 shows what happens: the result is a very graceful degradation in the case of the brightness model, but a very disturbing rippling in the case of wavelet coding. This rippling in image coding requires sophisticated post-processing to reduce the effect, see for example Ye et al. (2004).

In the meantime the two questions at the end of Section 2.4 have been addressed: (a) The image that we perceive is a virtual construction by a symbolic line and edge interpretation, i.e., it is not a re-construction with no need for “yet another observer” in our brain who must analyse the reconstructed image for object recognition etc. In fact, object recognition and brightness perception have been combined into a single process: indeed, our simulations showed that object categorisation and recognition can be obtained by using different multi-scale image representations, i.e., either line/edge maps with event positions and types, or by the unimodal line and bimodal edge representations (Rodrigues & du Buf 2006a,b). (b) There is no precise region in our brain where the image that we perceive is created. Our model is limited to feature extractions in V1 and beyond, but this information must propagate to higher brain regions, eventually leading to consciousness, at the least being aware of our position in our actual environment. In other words, we may say that our perceived image, and therefore also at least part of our consciousness, are constructed by the entire brain, perhaps with an emphasis on the visual cortex. This is a holistic view, but it should be mentioned that the local-global discussion about consciousness might be a hornets’ nest (Koch 2004; Bauer 2004).

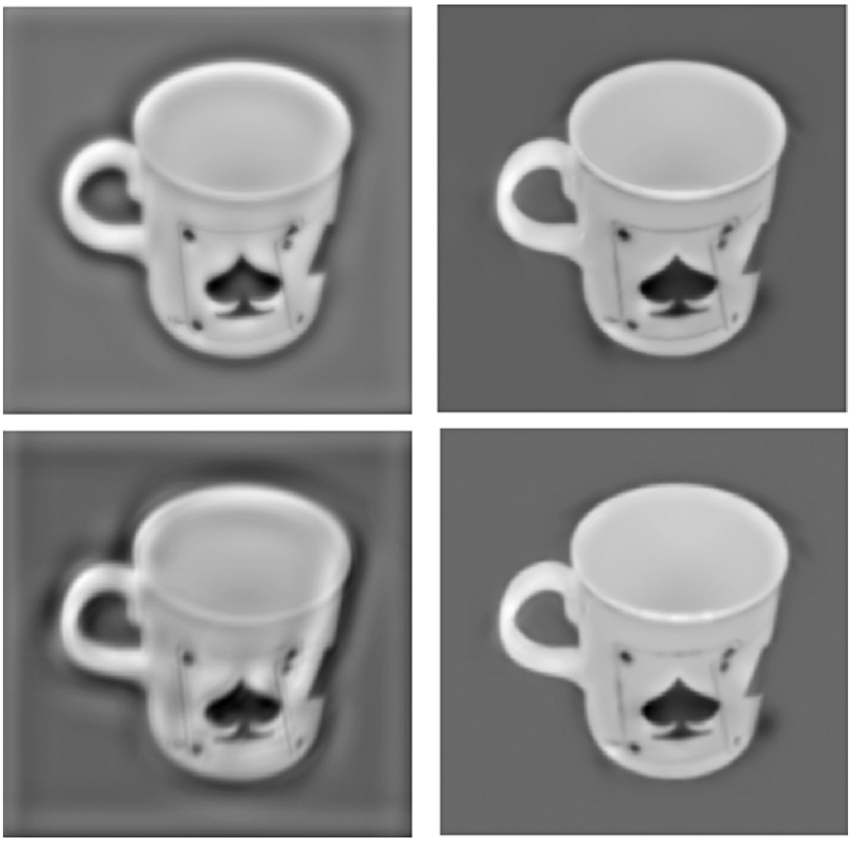

Figure 5: Image coding based on wavelets (left) and the brightness model (right). In both cases a limited number of scales has been used (top), which leads to severe rippling in the case of coding. If from all information only half is randomly selected, the coding result further deteriorates (bottom-left) but not the brightness result (bottom-right)

Above we wrote that colour is represented in the CO blobs in V1, possibly in the form of sampled values that represent homogeneous object regions. However, recently it was found that many colour cells in V1 are orientation tuned (Friedman *et al.* 2003). This probably means that such oriented edge (contour) cells also contribute to colour perception and not only to achromatic brightness as exploited in our brightness model. In addition, contour processing may play an important role in colour constancy, with different weights of near and far (local and global) contour components in the normalisation process, in addition to near and far colour samples; for a computational model see for example Rizzi *et al.* (2003). It should also be added that part of all neural connections may be more static and a result of evolution, i.e., brightness as an ecological interpretation of learned patterns in natural images (Yang & Purves 2004). All such complications, including long- and short-term adaptation effects and input-output amplitude nonlinearities, which have not even been mentioned until here, make us realise that we are far away from a unified framework.

The same can be said about object categorisation and recognition. Change blindness, the fact that we do not notice things at positions where we are not looking, points at an interpretational filling-in process. Even the filling in of the blind spots in the retinas, where the two optical nerves leave the eyes and there are no photoreceptors, is not noticed under normal viewing conditions. The latter effect could at least be explained by the fact that input from the other eye might be used there, but not change blindness. If we do not perceive a specific object, we do not perceive that object's brightness and colour. In such a case our brain may be guessing what the most obvious solution might be, probably on the basis of prior experience with similar images.

4 Painterly rendering

It is relatively straightforward to develop a painterly-rendering scheme on the basis of our brightness model, i.e., human vision, as is the case in similar approaches using algorithms from computer vision (Gooch *et al.* 2002; Kovács & Szirányi 2004; Shiraishi & Yamaguchi 2000). In our case, the scale of simple and complex cells is translated into the width of discrete brush strokes: single strokes in the case of detected lines and two parallel strokes in the case of detected edges, simulating coarse-to-fine painting using increasingly smaller brushes. Detected line and edge positions are stored in coordinate lists and these can be processed, for example smoothed, broken up into smaller lists, and/or linearised. For each coordinate list the stroke(s) is (are) rendered by means of triangle lists and texture mapping, for which colours are picked in the input image: one colour at the centre of line strokes and two colours at the centres of edge strokes. Texture mapping allows to simulate real brush strokes, composed of random selections of heads, bodies and tails of digitised strokes that were painted with a flat brush and, e.g., oil paint.

In homogeneous regions, where no lines and edges have been detected, we can prepare a background by applying strokes randomly or by influencing orientations for diagonal (or rotated) criss-crossing. In fact, we always start with painting a complete background, like most painters do, because our interface allows to select line/edge-related foreground strokes with certain brush sizes. The use of all scales and therefore brush sizes will result in a very realistic painting; when some scales are skipped the result will be more abstract. In addition, when introducing an orientation bias, i.e., for example rotating brush strokes towards horizontal, vertical and diagonal orientations, the result will become more cubistic with increasing bias.

The user interface which is being developed has very few menu lists and a structure that resembles the procedure that a painter uses: first select a surface structure (canvas or paper) and background colour, then apply a background with random or biased strokes, which can be incomplete because the user can stop the painting process at any time, for example to adjust parameters. To this end the user can set the speed of the painting process, can stop, resume or re-start the entire process or only the back-or foreground process. The interface allows to apply palette effects, for example to apply a model of colour constancy—a sort of normalisation of the dynamic ranges of the R, G and B channels—

which normally makes a painting more vivid, and/or to apply a red-orange or blue-green shift for introducing a warm or cold emotion. The interface also allows to apply a model of Focus-of-Attention based on end-stopped cells, in order to apply brush strokes only in and around regions with some complexity. Figures 6 and 7 show a few examples. For further details we refer to du Buf *et al.* (2006) and Nunes *et al.* (2006). Future research goals are to study the influence of colour shifts, not only for colour emotions, and the level of image abstraction, in simulated paintings. Such aspects are closely related to painting styles and studied in a research area called empirical aesthetics. Image and painting composition is much harder to address in terms of the visual cortex, although simple manipulations of existing paintings have been applied in some studies, see Nodine *et al.* (2003) and Locher (2003).

5 Discussion

In the Introduction we wrote that we cannot take a look into van Gogh's head and we do not know the exact landscapes that he saw. Well, after reading the subsequent sections the reader should be able to assume that we are on the way to simulate a standard observer in conjunction with a standard painter. In other words, we start being able to explore basic processes in the visual system and to combine these into an increasingly complete architecture, thereby implicitly looking into a "generic head" with the possibility to simulate specific painters in the future.

The visual system is able to construct on the basis of a brief glance a complete image/scene representation in our brain: from local syntax to objects to gist and layout of objects, including semantic interpretations and even emotions. More advanced models will therefore lead to a complete morphology, as if someone is asked to write a complete description of an image, from global aspects to local detail. Unfortunately, the development of a complete artificial visual system—or computational model—is a very long-term goal. However, the image interpretation, description, annotation etc. are expected to foster novel solutions for image and video synthesis, coding and art work for illustration purposes. The development will depend on results of ongoing and future research projects, both in visual perception and in NPR. Since even relatively simple models of the visual system require tremendous amounts of storage capacity and associated CPU times for the number crunching, new generations of more powerful computers are required. As for now, we do not know whether parallel processing in a distributed Grid environment will be beneficial because of necessary communication times, but the tremendous storage capacity that is required is no surprise: the entire brain counts 1012 (one million million) cells with 1014 to 1015 interconnections, and a significant part is devoted to vision. Today, in 2007, it is already possible to achieve 1 TFLOPS (one tera or one million million of floating point operations per second) on a normal PC using graphics boards with GPUs that are optimised for vectorised MADD (multiply-add) operations. This is not a supercomputer, but on compara-

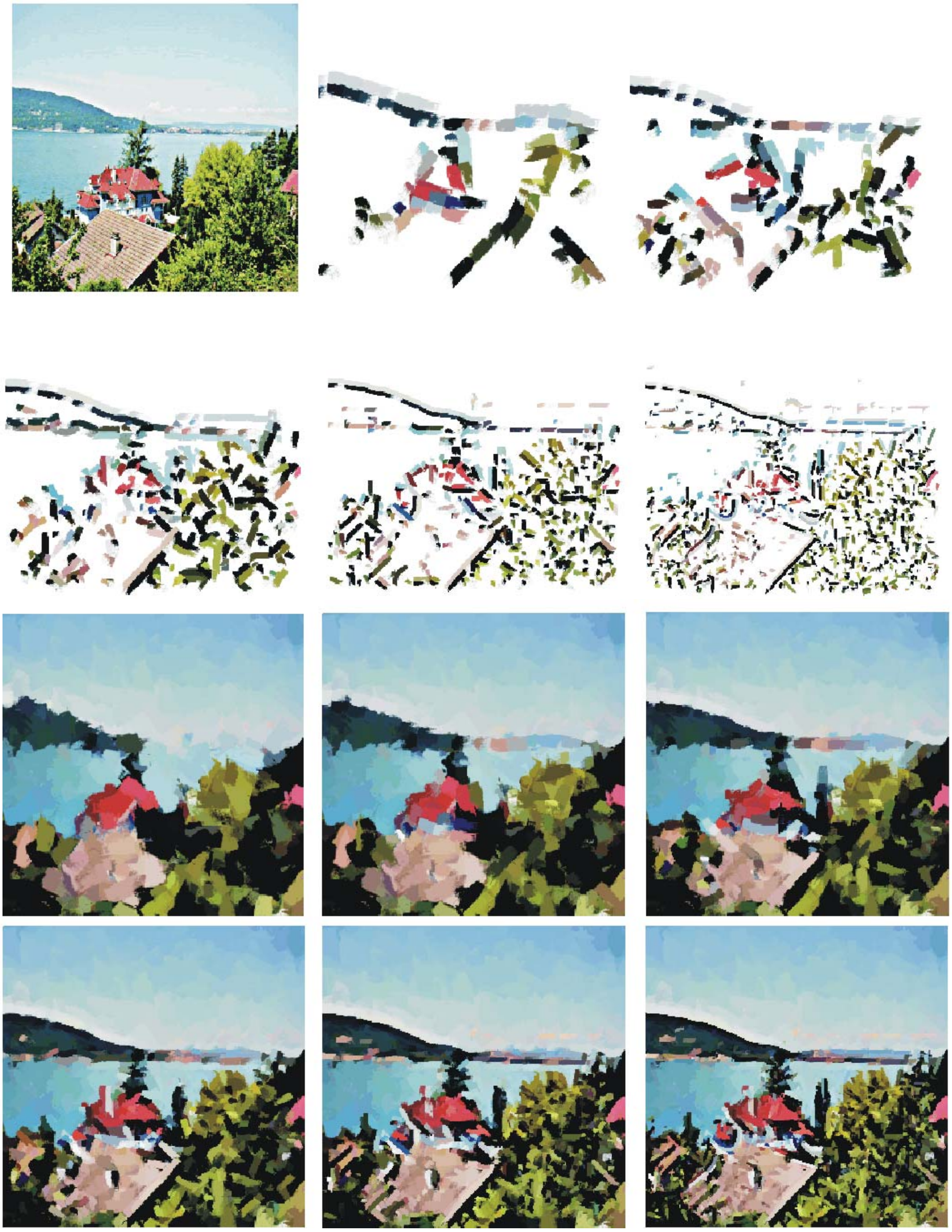

Figure 6: Rendering: the input image (top-left) is first used to paint a background with a big brush (shown on the third row at left), on which foreground strokes can be painted using increasingly smaller brushes. Not all scales need to be painted

ble systems it will soon be able to simulate the dynamics of 1012 cells at a speed which will come close to realtime, provided that enough of fast memory is available. Storage capacity being the bottleneck, future hard disks with a capacity of more than 1 TBYTE will not provide a solution because of slow access times.

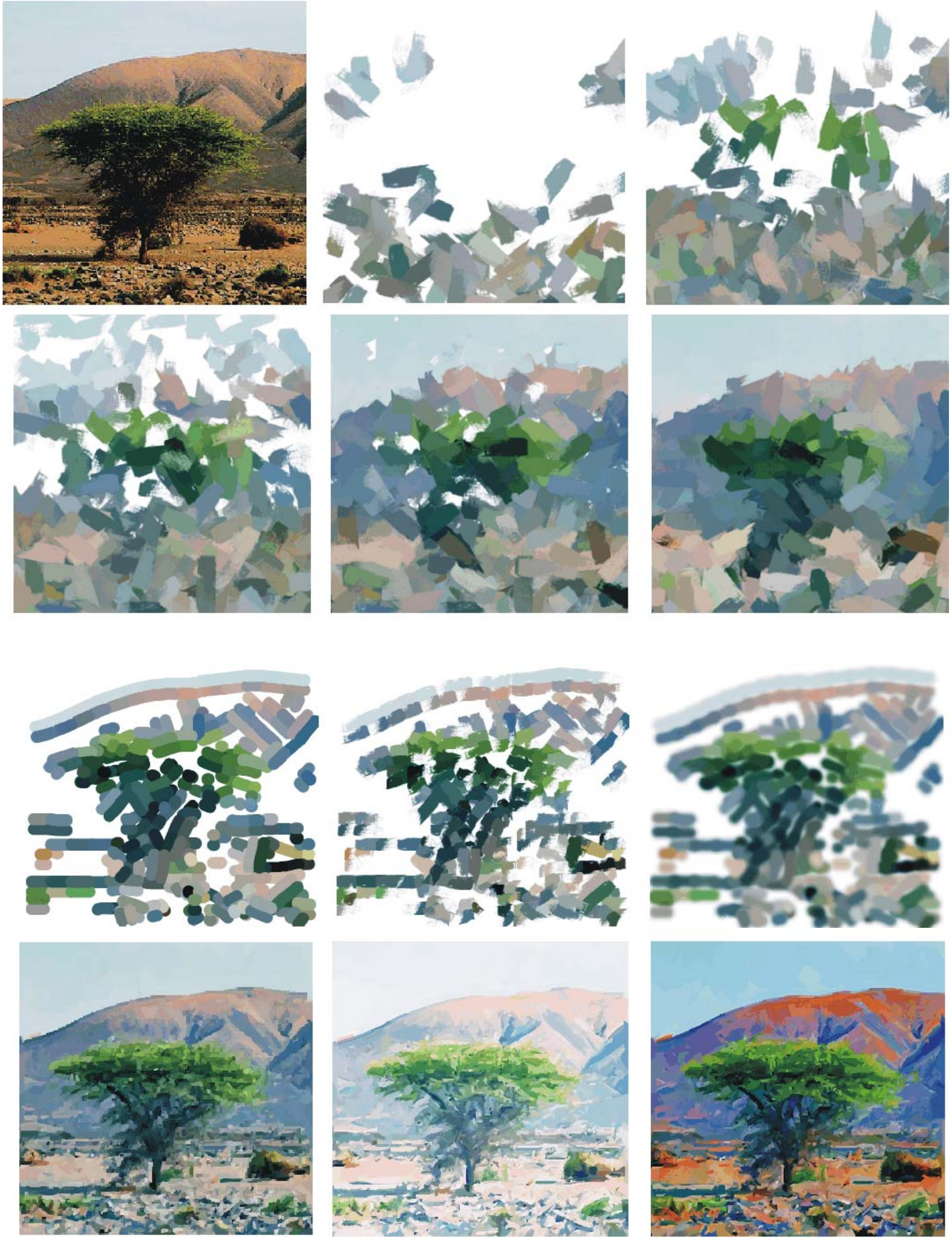

Figure 7: Top two rows: input image and the background process with random strokes of a flat brush. Third row: foreground strokes with a round brush, a flat one and spray. Bottom row: changing brightness and saturation

Although not discussed explicitly, it should be clear that our models provide a morphological image description in terms of multi-scale features on the basis of cortical cells: lines and edges for brush strokes and keypoints for Focus-of-Attention. Future extensions will cope with abstraction and composition, even with meaning or gist. All these features can be extracted by data-driven or bottom-up processes. So why could we write in the abstract that bottom-up and top-down processing can be combined?

The answer is rather straightforward: one might say that our visual system has two modes of operation. When looking at an image for a split second, long before we consciously know what objects there are, our brain already knows what the image is about. This is the fast gist and layout vision, probably implemented by feed-forward neural networks that exploit texture, colour, disparity and motion. Such features also allow for separating (segregating) entire objects, for example a tree with differently coloured and textured trunk and crown in front of a background, where trunk and crown should belong to the same object. Hence, in addition to global gist there may exist local gist, which hints at specific objects and their spatial relations (layout). This first and rapid mode of operation can be thought to "bootstrap" the second mode: select subsets of normalised templates in memory in order to scrutinise objects in the input image. The latter objects are not normalised, which implies that multi-scale line/edge and keypoint representations of input objects and normalised templates in memory must be compared. This comparison must be done sequentially, object after object, and the two feature maps must be projected such that they converge. This is the dynamic linking between neural layers at low and high levels as explored by Deco & Rolls (2004), and the fact that a big part of the visual cortex is involved in the dynamic linking limits the "bandwidth" of the system (Rensink 2000).

So, why is our visual system so fast and efficient? Because bottom-up and top-down processing are done in parallel. We do not think in terms of object edges or textures, we think in terms of gist, and gist limits the enormous amount of possible object templates in memory that must be checked. This explains why we have difficulties in recognising objects that are completely out of their normal context. In conclusion, the good news is that bottom-up and top-down image morphologies can or even must be combined. The bad news is that we are much more advanced in bottom-up processing, i.e. top-down processing is an almost completely new research area. However, in one or a few decades from now, when a lot of research effort has been put into top-down processing, this bad news will turn into good news for image morphology!

References

Allman, J., Zucker, S.: Cytochrome oxidase and functional coding in primate striate cortex: a hypothesis. Cold Spring Harbor Symp. on Quant. Biol. 55, 1990, 979–982.

Bar, M.: Visual objects in context. *Nature Reviews: Neuroscience* 5, 2004, 619–629.

Bar, M. & Kassam, K. & Ghuman, A. & Boshyan, J. & Schmid, A. & Dale, A. & Hämäläinen, M. & Marinkovic, K. & Schacter, D. & Rosen, B. & Halgren, E.: Top-down facilitation of visual recognition. *Proc. National Academy of Science* 103 (2), 2006, 449–454.

Bauer, R.: In search of a neuronal signature of consciousness – facts, hypotheses and proposals: neuroscience and its philosophy. *Synthese* 141 (2), 2004, 233–245.

Berson, D. Strange vision: ganglion cells as circadian photoreceptors. *TRENDS in Neurosciences* 26 (6), 2003, 314–320.

Deco, G. & Rolls, E.: A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* 44, 2004, 621–642.

du Buf, J.: Responses of simple cells: events, interferences, and ambiguities. *Biol. Cybern*. 68, 1993, 321–333.

du Buf, J.: Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biol. Cybern*. 70, 1994, 449–461.

du Buf, J.: Modeling brightness perception. Chapter 2 in: van den Branden Lambrecht, C.J. (ed.): *Vision models and applications to image and video processing*. Kluwer Academic, 2001, 21–36.

du Buf, J.: Improved grating and bar cell models in cortical area V1 and texture coding. *Image and Vision Computing*, doi:10.1016/j.imavis.2006.06.005, 2006 (Vol. 25, 873-882, 2007).

du Buf, J. & Fischer, S.: Modeling brightness perception and syntactical image coding. *Optical Eng.* 34 (7), 1995, 1900–1911.

du Buf, J. & Rodrigues, J. & Nunes, S. & Almeida, D. & Brito, V. & Carvalho, J.: Painterly rendering using human vision. *VIRTUAL - Advances in Computer Graphics in Portugal*, p. 12 (http://virtual.inesc.pt/), 2006.

Friedman, H. & Zhou, H. & von der Heydt, R.: The coding of uniform colour figures in monkey visual cortex. *The Journal of Physiology* 548, 2003, 593–613.

Gooch, B. & Coombe, G. & Shirley, P.: Artistic vision: painterly rendering using computer vision techniques. Proc. ACM/SIGGRAPH-Eurographics NPAR, Annecy (France), 2002, 83–91.

Heitger, F. & von der Heydt, R. & Peterhans, E. & Rosenthaler, L. & Kubler, O.: Simulation of neural contour mechanisms: representing anomalous contours. *Image and Vision Computing* 16 (6), 1998, 407–421.

Hubel, D.: *Eye, Brain and Vision*. Scientific American Library, 1995.

Itti, L. & Koch, C.: Computational modeling of visual attention. *Nature Reviews Neuroscience* 2 (3), 2001, 194–203.

Kastner, S. & Schneider, K. & Wunderlich, K.: Beyond a relay nucleus: neuroimaging views on the human LGN. *Progress in Brain Res*. 155, 2006, 125–143.

Koch, C.: *The Quest for Consciousness. A Neurobiological Approach*. Roberts and Company Publishers, 2004.

Kovács, L. & Szirányi, T.: Painterly rendering controlled by multiscale image features. Proc. 20th Spring Conf. Comp. Graphics, Budmerice (Slovakia), 2004, 177–184.

Locher, P. E.: Experimental aesthetics: the state of the art. *Bulletin of Psychology and the Arts* 4 (2), 2003.

Nodine, C. & Locher, P. & Krupinski, E.: The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo* 26 (3), 2003, 219–227.

Nunes, S. & Almeida, D. & Brito, V. & Carvalho, J. & Rodrigues, J. & du Buf, J.: Perception-based painterly rendering: functionality and interface design. Proc. Ibero-American Symp. Comp. Graphics, Santiago de Compostela (Spain), 2006, 53–60.

Ohzawa, I. & DeAngelis, G. & Freeman, R.: Encoding of binocular disparity by complex cells in the cat's visual cortex. *J. Neurophysiol*. 18 (77), 1997, 2879–2909.

Pessoa, L.: Mach bands: how many models are possible? Recent experimental findings and modeling attemps. *Vision Res*. 36, 1996, 3205–3227.

Read, J., Cumming, B.: Does depth perception require vertical-disparity detectors? *Journal of Vision* 6 (12), 2006, 1323–1355.

Rensink, R.: The dynamic representation of scenes. *Visual Cogn.* 7 (1-3), 2000, 17–42.

Rizzi, A. & Gatta, C., Marini, D.: A new algorithm for unsupervised global and local color correction. *Pattern Recogn. Lett.* 24 (11), 2003, 1663–1677.

Rodrigues, J. & du Buf, J.: Cortical object segregation and categorization by multi-scale line and edge coding. Proc. Int. Conf. Comp. Vision Theory and Applications 2, Setúbal (Portugal), 2006a, 5–12.

Rodrigues, J. & du Buf, J.: Face recognition by cortical multi-scale line and edge representations. Proc. Int. Conf. Image Anal. Recogn., Springer LNCS Vol. 3211, 2006b, 329–340.

Rodrigues, J. & du Buf, J.: Multi-scale keypoints in V1 and beyond: object segregation, scale selection, saliency maps and face detection. BioSystems 86, 2006c, 75–90:doi:10.1016/j.bio systems. 2006.02.019.

Schmolesky, M.: The primary visual cortex. http://webvision.med.utah.edu/VisualCortex. html, 2000.

Shiraishi, M. & Yamaguchi, Y.: An algorithm for automatic painterly rendering based on local source image approximation. Proc. ACM/SIGGRAPH-Eurographics NPAR, Annecy (France), 2000, 53–58.

van Deemter, J. & du Buf, J.: Simultaneous detection of lines and edges using compound Gabor filters. *Int. J. Patt. Recogn. Artif. Intell.* 14 (6), 2000, 757–777.

von der Heydt, R. & Peterhans, E. & Dursteler, M.: Periodic-pattern-selective cells in monkey visual cortex. *J. Neurosci*. 12 (4), 1992, 1416–34.

Wässle, H.: Parallel processing in the mammalian retina. *Nature Rev. Neuroscience* 10, 2004, 747–757.

Yang, Z., Purves, D.: The statistical structure of natural light patterns determines perceived light intensity. Proc. Nat. Acad. Sci. USA 101 (23), 2004, 8745–8750.

Ye, S., Sun, Q. & Chang, E.: Edge directed filter based error concealment for wavelet-based images. Proc. IEEE Int. Conf. on Image Processing 2, 2004, 809–812.

Zucker, S. & Hummel, A.: Receptive fields and the representation of visual information. *Human Neurobiol.* 5 (2), 1986, 121–128.