Natural Computing **3:** 293–321, 2004.

Early cognitive vision: Using Gestalt-laws for task-dependent, active image-processing

FLORENTIN WÖRGÖTTER[1,*], NORBERT KRÜGER[1,2], NICOLAS PUGEAULT[1], DIRK CALOW[3], MARKUS LAPPE[3], KARL PAUWELS[4], MARC VAN HULLE[4], SOVIRA TAN[5] and ALAN JOHNSTON[5]

[1]*Department of Psychology, University of Stirling, Stirling FK9 4LA, U.K (*Author for correspondence, e-mail: worgott@cn.stir.ac.uk);* [2]*Computer Science, Aalborg University, Esbjerg, Denmark;* [3]*Institute of Psychology, University Münster, Germany;* [4]*Laboratorium voor Neuro- en Psychofysiologie, Departement Neurowetenschappen en Psychiatrie, KU Leuven, Belgium;* [5]*Department of Psychology, University College London, U.K.*

Abstract. The goal of this review is to discuss different strategies employed by the visual system to limit data-flow and to focus data processing. These strategies can be hard-wired, like the eccentricity-dependent visual resolution or they can be dynamically changing like mechanisms of visual attention. We will ask to what degree such strategies are also useful in a computer vision context. Specifically we will discuss, how to adapt them to technical systems where the substrate for the computations is vastly different from that in the brain. It will become clear that most algorithmic principles, which are employed by natural visual systems, need to be reformulated to better fit to modern computer architectures. In addition, we will try to show that it is possible to employ multiple strategies in parallel to arrive at a flexible and robust computer vision system based on recurrent feedback loops and using information derived from the statistics of natural images.

Key words: mid-levelvision, multi-modal processing, feature integration

1. Introduction

Serious attempts to create machine vision systems have now a history of about 30–40 years. In spite of this extended period of research and development, we find ourselves in a position that these systems are still exceedingly limited in their performance. Machine vision problems have essentially been addressed using two different approaches, the engineering and the biological approach, which were, however also often mixed. The undisputed power of our own visual system very strongly

suggests that machine vision systems should also follow a "neuronal approach", if only we would know, how the brain does it ... As a consequence machine vision systems are still wedged in-between the technological restrictions imposed by current cameras and computers and the limited knowledge about biological visual processing. Only more recently a better understanding of some fundamental neural processing principles has been reached, especially about the dynamic interactions that take place between different areas in the visual pathway.

This review article will discuss briefly a wide variety of different active or passive, overt or covert mechanisms known from the visual system and used, to some degree, in modern machine vision systems to improve image analysis. This is meant to provide an overview across the different strategies employed by the visual system to arrive at a flexible and highly functional mode of operation. In view of the different structure of technological versus biological systems, we will ask the question to what degree this kind of natural computation should be copied by modern machine vision system. Should such systems really include all these aspects or is there a chance to arrive at functional abstractions which allow reducing the effort while leading to the same good results? Furthermore we will discuss some more novel ideas about active, task-dependent vision adopting the view-point that efficient vision systems can only be built when they close the perception–action loop (at least by internal processing, if not by active behaviour). The second part of this article will present results obtained in the context of a large European project group where we are building a multi-modal, task-focused image processing system based on abstractions of mechanisms which are found in the vertebrate visual system.

2. Mechanisms to control visual input flow

Only rather recently it has become clear that vision must be regarded as an active process, where the vision system (the observer) decides what he/she "wants" to see leading to enhanced resolutions there and limiting the data-flow elsewhere. At first this seems paradox, because our eyes are taking in everything with which they are faced, or don't they? Thus, this notion needs some explanation. Obviously, we perform eye- and head-movements and thereby actively guide our visual perception to some degree. This overt behaviour has rather early also been built into machine vision systems (Aloimonos et al., 1987); only more recently,

however, it was accompanied by the notion that there must be brain-intrinsic, covert mechanisms, too, which support active perception.

Marr's seminal book on early vision (Marr, 1982) still treats vision as a bottom up filter-process. Camera pixels are combined into higher level entities, for example edges. This way different levels of representation are obtained, for example a "primal sketch" which makes properties of the 2-D image explicit, a – in Marr's terminology – 2½-D sketch which is a viewer centred representation of the visual scene and finally a 3-D reconstruction of the scene in world coordinates. About at the same time, however, it became clear that advanced vision systems also need built-in knowledge (memory) without which higher level processes such as object recognition cannot take place. After all, how would the concept of objects emerge from the simple geometrical entities that Marr's analysis provides? Early attempts towards object recognition tried to solve this problem within rather restricted environments, for example limiting the number and geometrical configurations of the objects in the scene (for a critique of this approach see Brooks (1991)). While this may work, soon it became clear that our visual system operates in a more general way: We are able to recognise an object (e.g., a cup) regardless of how it is placed, we can generalise easily subsuming all kinds of containers into the concept of a "cup", and we can recognise a vast number of objects on a breakfast table and not only the cup. All this knowledge cannot be built into machine vision systems in a naive top-down manner, it is just too much and too diverse. This problem is also known as the "bias-variance" dilemma (Geman et al., 1995). If you build too much bias (Knowledge) into a system you are reducing the variance that it can express, which means the degrees of freedom that it can show when confronted with a novel situation are reduced.

The combinatorial explosion problem which arises when trying to cover each and every aspect of the diversity of "the world" also raises two more issues, for which experimental support shall be provided below: (1) It is inconceivable that our vision system analyses each part of the scene with the same accuracy. (2) It seems highly unlikely that our vision system "jumps" from any kind of early visual representation directly to the stage of object recognition. Instead, one should assume that there are intermediate scene analysis steps interspersed in-between.

We will discuss experimental support for this next. Let us, however, first point out that these two notions have recently been combined realizing that the human visual system seems to actively make probabilistic guesses about sub-structures in the scene and that these guesses are guided by the currently existing task that the observer performs

(Rao and Ballard, 1999; Körding and Wolpert, 2004). Thus, there are not only intermediate visual representation levels existing, but these levels can be actively influenced by the observer which leads to a changed functional resolution of the analysis at different locations in the visual field (on top of the existing visiotopic maps).

First, we will briefly review some of the older known facts about space- and context-dependent changes in visual resolution as well as overt (behaviourally expressed) active vision, before we embark on the more complex aspect of modern ideas about covert active vision. The second part will also try to show how Gestalt laws emerge in a natural way from recursive, multi-modal image processing (Prodöhl et al., 2003; Spillmann and Ehrenstein, 2004).

2.1. *Structuring the visual input – Cortical maps*

It has been know since around 1977 that visual topography and several visual features are represented in an orderly fashion in the cortex (for reviews see Erwin et al., 1995; Swindale, 1996; Chapman, 2004). In a first approximation, the radial coordinates of the visual hemifield that projects to a hemisphere are transformed by the complex logarithmic function onto the cortical surface in area V1 (Schwartz, 1977, 1980; but see Johnston, 1986 for a critical review of this proposal and an alternative structural model). As a consequence, the central visual field around the fovea is strongly magnified at the cortical surface while the periphery is underrepresented. This transformation leads to the effect that concentric circles (i.e. on the retina) are mapped onto equally spaced vertical lines while radial lines will map onto horizontal lines (on the cortical surface). In addition, the properties of such a complex logarithmic mapping lead to scaling for objects that increase in image size in proportion to visual eccentricity and rotational invariance. More complex topographical representations have also been observed and modelled in higher cortical areas (Mallot, 1985). Since these topographical representations are almost everywhere continuous (at least in the lower visual areas), they allow utilising neighbourhood relationships for processing. This is an obvious advantage, because the structure of our world is such that neighbouring entities will with a greater likelihood belong together than those that are not adjacent to each other (Gestalt-law). This will also lead to higher accuracy in the processing when utilising interpolation processes between such adjacent ''pixels''. In addition, it has been found in the lower cortical areas that responses

from the two eyes are essentially represented in different, interleaved slabs of the cortex (ocular dominance maps, Chapman, 2004). Cortical cells in V1 respond preferentially to oriented stimuli and also orientation preference is represented in a map. If one performs two-dimensional Fourier analysis of such orientation maps, one will for many species, obtain a (distorted) annulus shaped spectrum (Niebur and Wörgötter, 1994). This points to the fact that orientation preference is repeated along the cortical surface on average with a constant frequency along all directions (hence isotropically) and that orientation maps are homogeneous and "look the same" at every location in V1 (Niebur and Wörgötter, 1994). Cells with enhanced colour preference are interspersed in these orientation maps and form blob-like clustered groups (Wong-Riley, 1979; Livingstone and Hubel, 1984; Ts'o and Gilbert, 1988) in V1. Indications exist that also visual disparity is represented in a map-like structure in V2 (Hubel and Livingstone, 1987; Ts'o et al., 2001). Other features, like velocity or direction preference, do not follow a map-like arrangement, but clustering is observed as well. In V2 maps follow a thick-stripe, thin-stripe, inter-stripe structure for colour and orientation, essentially also almost everywhere preserving the neighbourhood property (Livingstone and Hubel, 1984; Peterhans and von der Heydt, 1993; Shipp and Zeki, 2002a, b). We defer the reader to the literature for details about these maps and their models (Erwin et al., 1995, Swindale, 1996; Chapman, 2004). Here we will focus on the question, how such structures can facilitate image processing instead.

2.2. *Limiting the visual input*

2.2.1. *Eye- and head-movements*

The above discussed topographical representation implicitly leads to the situation that only the central part of the visual field (fovea) is analysed with high accuracy. As a consequence, it requires substantial effort if we want to see what happens in the periphery without moving our eyes. However, even when concentrating on the periphery we will not be able to distinguish finer details. Such an architecture leads to the advantage that less neuronal machinery is required in the periphery as compared to the fovea, effectively limiting the visual input. Accordingly receptive fields are narrow in the fovea and much wider outside (Zeki, 1978; Adams and Horton, 2003).

Head- and eye-movements have been "invented" to solve the problem of reduced peripheral resolution and soon the same mechanisms have been introduced into artificial systems as well. Concerning premotor aspects, similar mechanisms might actually underlie covert and overt shifts of attention and gaze (respectively) (Rizzolatti et al., 1987; Hamker, 2003). The neuronal (and technical) control mechanisms required for an overt goal directed gaze, however, are complex and constitute their own area of research. This article will not cover any of these aspects because we would like to focus on covert mechanisms instead.

2.2.2. *Attention*

The most notable covert mechanism is visual attention which leads to a substantial reduction of the amount of input that needs to be processed by dealing only with the momentarily most important visual information.

An operational definition of attention can be given with: Exposed to a number of stimuli, that are equal in their physical appearance, both animals and humans can respond to certain stimuli while neglecting others without having to move head or eyes. This internal, covert spatial focus is the basic operation of selective attention. From a computational point of view, one can distinguish bottom-up versus top-down mechanisms of visual attention. Bottom-up (pre-) attention is data driven and involuntary. Any kind of salient stimulus will automatically attract our attention. Bottom-up mechanisms account for pre-attentive effects like the pop-out phenomenon. These effects are transient, because the saliency of a new object decays with time, and they are fast, allowing for quick reactions. Speed is assured by low-level mechanisms and points to the involvement of early visual processing levels like thalamus and primary visual cortex.

By contrast, top-down attention acts on a different time scale of up to several seconds and it represents effects of voluntary attention. This task-driven form of attention to a specific region of interest must involve high-level regions, i.e. those responsible for cognitive functions.

2.2.3. *Psychological and computational models of selective attention*

Psychophysical reaction time studies provided early evidence for the existence of covert visual attention (see Treisman, 1969; Posner et al., 1982; Wolfe, 1998, for reviews). Subjects were asked to find targets in a visual display and when a valid cue was given before the target presentation, reaction times were significantly reduced.

The results of several years of research led to the formulation of many different hypotheses as to how these search experiments could be explained, e.g. Treisman's *feature integration theory* (Treisman and Gelade, 1980) or Julesz's *texton theory* (Julesz, 1981). Many of the proposed models share the spot- or searchlight paradigm (Broadbend, 1965; Neisser, 1967; Crick, 1984) as the same basic idea: It has two stages. First, as soon as a new stimulus is presented, the whole visual field is processed in parallel during the pre-attentive mode. When this mode does not suffice, e.g. when the task is complex (conjunctive search), a different, second strategy is needed. Then, only a limited area is highlighted and analysed in detail, whereas the rest is processed with less priority. In a serial process, the whole field is scanned. Since not all search experiments could be explained by the purely bottom-up approach of the searchlight hypothesis, top-down components were added to the model and different more complex models were devised (Desimone and Duncan, 1995; Wolfe, 1998). Over the years, many hybrid models have been developed which integrate serial, parallel, bottom-up and top-down processing (e.g. Grossberg et al., 1994).

The computational models of selective attention are mainly concerned with the problem of how a focus of attention can be selected and how its information can be routed through the network, treated, for example by the so-called *selection and routing models*. A second class of model deals with the question of how to implement selective attention with neurons (*tagging models*): Those neurons under the focus of attention will have to change their (temporal) firing characteristic (Niebur et al., 1993).

2.2.4. *Neuronal basis for selective attention*

One of the first cellular studies of selective attention was made by Wurtz et al. (1982). Experiments were carried out in awake primates while recording neurons from the superior colliculus (SC), striate cortex (V1) and posterior parietal cortex (PP). The basic finding was that cells in V1 and in SC responded with a higher firing rate when the animal oriented to an attracting spot with a saccade, while there was no change in response when the animal maintained fixation at a central spot and only shifted its attention covertly. PP neurons, in contrast, also show an activity enhancement due to covert shifts of attention. Moran and Desimone extended the first approach by showing that also neurons from the inferior temporal cortex (IT) and from V4 behave differently during an attention task (Moran and Desimone, 1985). In their experimental setup, two objects were placed inside the receptive field of an IT

neuron, one being an effective stimulus for that neuron and the other ineffective. If attention was focused on the ineffective stimulus, the activity of the neuron decreased, while it increased if attention was on the effective stimulus. It seems as if the receptive field shrinks around the attended object. Similar attentional modulations have been found in other areas and with other techniques (Büchel and Friston, 1997; Connor et al., 1997; McAdams and Maunsell, 1999; Treue and Martinez Trujillo, 1999; Maunsell and McAdams, 2001).

2.2.5. *Arousal*

Arousal is a second mechanism which probably also leads to a restriction, albeit more globally, of the information flow into the visual system. For example, the degree of synchronisation of visual cortical responses as reflected in the EEG can be influenced in a longer lasting way by electrical stimulation of the brain stem (Munk et al., 1996) experimentally inducing an aroused state of the animal.

The state of arousal is normally reflected in the frequency content of the EEG. During drowsiness α-waves (approx. 8–13 Hz), interspersed with Θ-waves (4–7 Hz), prevail, while deep sleep is characterised by a so-called synchronised EEG mainly containing δ-waves (approx. 0.5–4 Hz). During alert wakefulness mainly β-waves (approx. 13–30 Hz) are observed (non-synchronised EEG). Spontaneous state-transitions occur even in the anesthetised preparation. These spontaneous transitions are strongly correlated with dramatically changed response characteristics of the cortical afferents – the thalamic relay cells (Funke and Eysel, 1992). During synchronised EEG (''drowsiness, sleep'') thalamic cells are hyperpolarised (Dossi et al., 1992) and respond in the so-called ''burst-mode'': spontaneous activity is low and responses to stimulation are dominated by brief high-frequency bursts (for a review see Steriade, 1991). Intriguingly, the temporal behaviour of cortical cells upstream of the thalamus was much less affected by EEG state changes (Ikeda and Wright, 1974). The spatial structure of the receptive fields, however, was; and we observed that cortical receptive fields decreased in size when EEG switched from the synchronised to the non-synchronised EEG state (Wörgötter et al., 1998). This effect can be attributed to a changing effective connectivity of the thalamocortical synapses during different EEG states. Part of these effects seems to be mediated by the thalamo-cortico-thalamic loop, because the EEG dependency of cortical response is removed when chronically eliminating this feedback loop (Wörgötter et al., 2002; Eyding et al., 2003).

2.3. *Context sensitive receptive fields*

So far we have discussed rather global mechanisms for the control of information flow in the visual system: Attention and Arousal. There are, however much more local mechanisms existing which operate at the level of single receptive fields, which can also be interpreted in the sense of information flow control. Of particular relevance for visual information processing are those studies that showed how cortical cells change their responses in a context dependent way.

2.3.1. *Direction selectivity*

Experiments performed in cats by the groups of Hammond and Orban (Hammond and McKay, 1975, 1977, 1981; Gulyas et al., 1987, 1990; Orban et al., 1987, 1988) showed that the perception of relative motion and some motion after-effects are influenced by context dependent receptive field effects. It was found that responses to relative motion are amplified which could underlie the similar enhancement effect observed at the perceptual level. In addition, it was observed that large field, unidirectional background motion leads to adaptation of those cells, which are selective for this particular direction.[1]

2.3.2. *Orientation selectivity*

A different group of experiments demonstrated that the orientation tuning of cortical cells is also affected by the stimulation context and a wide variety of effects were observed (Gilbert and Wiesel, 1990; Knierim and v.Essen, 1992; Lamme, 1995; Sillito et al., 1995; Sillito and Jones, 1996; Zipser et al., 1996; Das and Gilbert, 1999). Of particular relevance for a guided information processing could be two effects: (1) Presenting a stimulus with preferred orientation in the receptive field centre together with many orthogonally oriented surround stimuli will enhance the response while similar orientations in the surround will suppress it (Knierim and v.Essen, 1992; Jones et al., in press). (2) Cells also adapt to stimulus orientation in a way that is similar to motion adaptation (see ''waterfall effect'', above). After prolonged presentation of a sinusoidal grating, the orientation preference of the cells will shift away from the orientation of the grating stimulus (Dragoi et al., 2000).

Above, we had briefly discussed pop-out phenomena in the context of visual attention. Here we argue that both findings could contribute to

the perception of ''orientation pop-out''. This is the effect that in a field of similarly oriented lines any group of lines that has a significantly different orientation will immediately ''pop-out'', making serial search unnecessary (Lamme, 1995; Kastner et al., 1997; Nothdurft et al., 1999).

2.3.3. *Receptive field size*

Also the phenomenon of perceptual filling-in could have a direct neuronal correlate at the level of receptive fields in striate cortex. To demonstrate this, a grey area was centred on the receptive field of a cell within a surrounding pattern of moving lines or twinkling dots. After prolonged presentation of this stimulus it was shown that receptive fields inside the grey area are significantly expanded (Pettet and Gilbert, 1992; Volchan and Gilbert, 1994; Gilbert, 1998). The robustness of this effect, however, was questioned by other groups (DeAngelis et al., 1995; Chapman and Stone, 1996) which may have been a consequence of a different interpretation of the data, though (Chapman and Stone, 1996).

The observations described in the previous section have demonstrated that significant spatial influences arise from regions that are distant from the classical receptive field. All these effects alter the response of the neurons essentially in a way that any stimulus which contains a degree of novelty is passed on while predictable situations are less strongly valued and which can in some instances be interpreted in terms of visual perception.

2.4. *Summary of the older findings*

The above sections have discussed six aspects that guide and limit the input data-flow into the visual system. (1) A higher visual resolution exists at the fovea as compared to the periphery. (2) Map-like structures exist in the visual areas which possibly connect computationally relevant features more efficiently with each other (the simplest example being neighbourhood relationships). (3) Eye- and head-movements are performed to direct the gaze to the location of interest. (4) Visual attention leads (very likely) to a reduced data flow at uninteresting image locations, while those that draw our attention are being analysed with greater detail. (5) Mechanisms of arousal lead to enhanced visual

processing capabilities when alert, while they may be reduced during drowsiness. (6) Visual receptive fields alter their spatial and temporal properties in response to the context within which a stimulus is presented.

All these mechanisms are used to enhance and optimise the computational power of our visual system. The first group (1 and 2) represent essentially hard-wired properties laid down in the anatomy of the visual system. The third aspect reflects clearly an overt behavioural strategy. Aspects 4 and 5 represent more covert mechanisms to control the input flow. Here we would think arousal related effects are not very important for machine vision systems which do not need to rest. The last aspect (6), where we have discussed that receptive fields are context sensitive, represents a truly covert mechanism, which must be rooted in the dynamical structure of the different networks involved. The moment-to-moment changing effective connectivity (Aertsen et al., 1989) between neurons, which is determined by the sum of all their inputs, controls these effects.

In the next sections we would like to discuss two questions: (1) How should visual information be represented at higher processing stages? Thus, asking what higher level receptive fields should look like to facilitate processing. And (2) How a given task will influence visual processing? (Figure 1).

3. Intermediate levels of visual representation – recurrent processing in vision

In the introduction we had stated that neither purely data-driven, feedforward nor purely knowledge-driven top-down processing mechanisms were successful for solving the machine vision problem. Data-driven mechanisms do not lead to any ''understanding'' of the objects analysed in a scene. For knowledge-driven mechanisms a different problem exists: The knowledge of the system has usually been built into it by its designer. Hence, it's the designer's knowledge and not the system's. The system's data-processing structures are vastly different from that of its designer, its input signals are not the same and the way it calculates is also different. Thus, there is always a necessarily existing mismatch between the way knowledge is represented in the system as compared to its designer. Furthermore – and even more problematic – the designer can never foresee all relevant events for the system such that imposed knowledge will have to be incomplete or sometimes irrelevant. Dennett

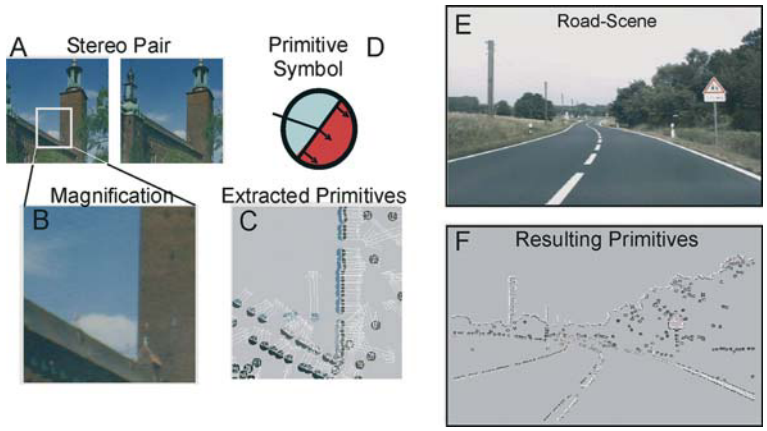

Figure 1. The concept of visual primitives. (A) A stereo pair from an image sequence. (B) Magnification of an image part. (C) Primitives extracted from this image part. (D) The symbol used to present a primitive. Encoded are colour (red versus blue), orientation (line in the middle), Optic flow (arrows), The long line points to the corresponding stereo match in the other primitive-image, which is omitted here. In (C) the same symbols are used. (E) A scene recorded while driving (left image from a stereo pair). (F) Resulting primitives at low magnification.

(1984) has called this the ''frame-problem'': The system's frame of reference for its analysis of a visual scene is necessarily different from the designer's frame of reference. This was the central problem that traditional AI (artificial intelligence) had faced. The computer vision community, however, has not given this much of a consideration so far. This has resulted in many rather naive approaches towards advanced scene analysis which were exceedingly limited in their performance. As a consequence, this has led to a rather critical perception of *any* approach towards scene ''understanding''.

So the question arises: What is missing? Why are humans so good at scene understanding, while general purpose machine vision systems still fail miserably?

Obviously, there is at the moment no clear answer to this. However, during the last years many neuroscientists have suggested that the solution to these problems may lie in the recurrent and parallel signal processing properties of the visual pathway. Several parallel processing streams exist in the visual system, e.g., the ''what'' and the ''where'' pathway, where predominantly form and location/movement are processed (for a recent review see Ungerleider and Pasternak, 2004). Receptive fields in these pathways are getting increasingly more complex

at higher level of hierarchy. This, however, is not only the result of feed-forward afferent wiring; much more important are the influences of feedback within an area and from higher areas to lower ones. Through this feedback a highly dynamic recurrent processing architecture emerges that leads to the self-emergence of advanced features. This has been nicely demonstrated by the studies of Somers et al. (1995) and Suarez et al. (1995) which have shown that local excitatory feedback can lead to the self-emergence of sharp orientation tuning or to the generation of direction selectivity. Less, however, is known about the influence of higher order feedback mechanisms, for example about the feedback between cortical areas. It has been suggested that such mechanisms might be used to refine the afferent data flow and to restrict it. Here, Rao and Ballard (1999) have suggested that predominantly unpredictable events are parsed to the higher areas, while predicted events will be processed with a lower priority.

3.1. *Multi-modal processing of stereo information*

The above discussed neuronal mechanisms cannot be directly used in a computer vision context and more abstract representations are needed to achieve similar ends. Recently we have proposed (Krüger et al., 2004) to use a specific type of multi-modal, local representation, called a visual primitive to facilitate computations. Primitives are making use of low-level feature extraction which yield orientation, edge-information (Felsberg and Sommer, 2001), colour, optic-flow (Nagel and Enkelmann, 1986), and stereo disparity information (Krüger et al., 2002). In addition, these primitives also carry information about edgeness and junctionness in for of confidences (Felsberg and Krüger, 2003; Krüger and Felsberg, 2003). The low-level features are computed pixel-wise. However, the Primitives represent this information in a condensed way (97% compression). A sparse and meaningful representation is created where the number of primitives is much smaller than the number of original pixels. In a very abstract sense, these primitives could, thus, be associated to hypercolumns in the visual cortex (Krüger et al., 2004). This offers the advantage that the next image processing stages have to be performed only on a much reduced number of inputs that however have a higher semantic meaning (data condensation). In the next stages, these primitives can be used to perform grouping and motion analysis (Krüger and Wörgötter, in press). In order to perform grouping, we rely on some statistically significant properties of images, most importantly

on the fact that collinear line segments prevail in a scene (Krüger and Wörgötter, 2002).

Thus, if line segments are collinear, it seems more likely that they belong to the same object. When relying only on collinearity, the confidence in such a 'guess' is still rather low. However, one can take also other features into account; for example, colour, disparity and flow and perform multi-modal line processing. Hence, if a pair of collinear lines shares the same colour, disparity, and flow, then it is much more likely that it belongs to the same object (Figure 2B). In a similar way, we can improve stereo estimation by assuming that only that particular stereo pair is correct that at the same time also shares other features. Such multi-modal stereo processing leads to a first sorting out of false stereo matches (Figure 2C and D; Krüger et al., 2002; Pugeault and Krüger, 2003).

Normally, however, we are still left with many false stereo matches and many wrong line combinations after having performed these two steps of multi-modal processing. But now we can go one step further and combine multi-modal line information with multi-modal stereo information: If I have found a set of possible stereo matches for one given line in the left (or right) image, then that particular match which also belongs to a collinear pair is highly likely to be the correct one. Figure 3 demonstrates the idea behind this and shows how false matches can be eliminated by employing this mechanism to a real scene. Of course this argument is symmetrical and we can also sort out wrong line-matches on the grounds of stereo information (Pugeault et al., 2004).

In the next step, also motion information is integrated into the common image analysis scheme (Figure 4). Here we rely on the rigid body motion (RBM) principle: If the motion parameters in a visual scene are known, and if its objects are rigid, then it is at least in principle possible to predict the development of the scene ad infinitum (see, e.g. Krüger and Wörgötter, in press). While this principle sounds simple, nevertheless, several requirements have to be fulfilled (Figure 5). To calculate the RBM, it is necessary to track a number of corresponding image entities for every moving object in a reliable way. Thus, noise, occlusions, and the ever prevalent correspondence problem interfere with this requirement. In addition, RBM requires calculations in the 3D real-world domain. Thus, 3D coordinates have first to be reconstructed from the (cleaned-up) stereo pairs before the RBM can be determined. The goal of using the RBM principle is to predict how a scene would develop. Hence, we must then calculate the 3D coordinates for the next stereoscopic camera frame pair and we must also be able to re-project

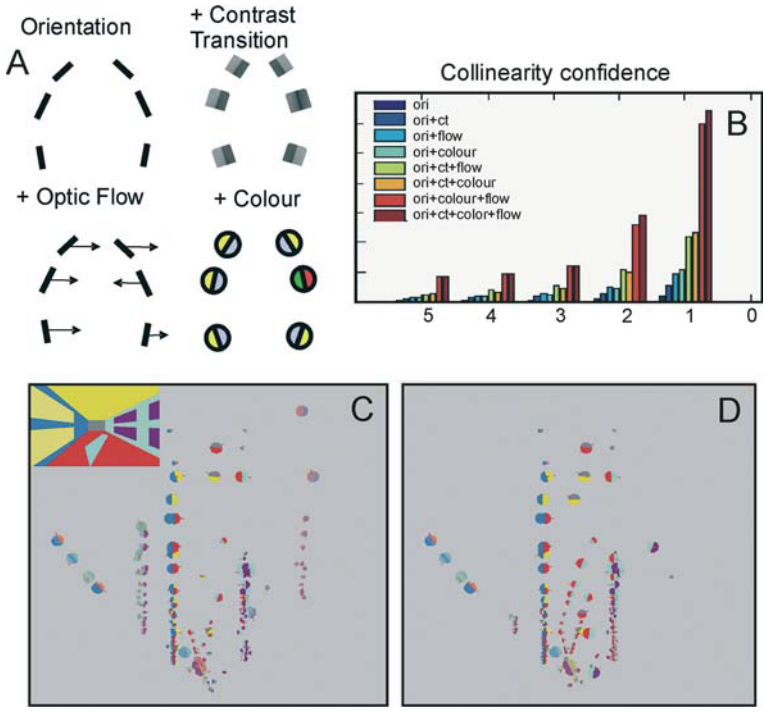

Figure 2. Influence of multi-modal processing on the confidence of collinearity estimation (A,B) and of stereo matching (C,D). (A) Schematic of the processing principle. Collinearity of orientation is by itself not sufficient to guarantee the existence of an uninterrupted line. Only in the left part of each panel a line can be inferred with some confidence from the existence of multi-modal matches of different image features. (B) The increase in the confidence in a collinear line pair can be statistically measured. For details see Krüger and Wörgötter (2002). For a given line segment the numbers at the abscissa refer to the distance from this line segment. *A priori* collinearity becomes less likely at greater distances. Different bar show, how confident one can be, in a found collinear pair. Confidence is lowest when only the orientations match (which is the minimal condition for a pair to be collinear). When all features match, confidence is highest. (C,D) Stereo matches for an artificial stereo pair; one image of which is shown in the inset. Large panels represent a reconstructed a top view (X–Z axis view). In (C) only orientation was used to determine a stereo match and many wrong matches exist. In (D) stereo matches where calculated using all other multi-modal image features (except flow, because these are still images).

the 3D coordinates onto the 2D image domain. Only this way we can finally compare the re-projected coordinates with the ones obtained from the next frame-set. All these steps require solving several tedious technical problems before such a scheme will robustly work.

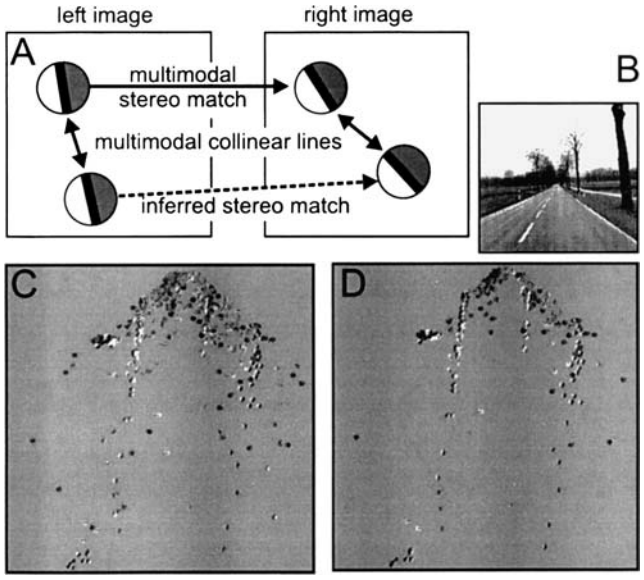

Figure 3. Finding stereo matches based on lines. (A) Processing principle. In the left and right image two highly probable collinear line segments have been found, but only for one (top) a good multi-modal stereo match exists. This may be due to the perspective distortion which may have led to a bad matching situation for the bottom pair. However, the strong confidence that both pairs belong to lines leads to the assumption that also the bottom stereo match must be correct. (B) One image from a driving scene. (C) Multi-modal stereo matches (top view) found without taking line-correspondences into account and (D) with line-correspondences. The number of wrong matches is substantially reduced in (D).

Figure 6 shows some results obtained with artificial and real scenes. This data shows that only primitives which belong to the edges of an object are confirmed (white) across multiple frames, while all others will eventually be sorted out (black).

In summary: In this section, we have shown how to improve image information by combining stereo-disparity analysis with other visual modalities like colour, orientation and contrast transition information. Finally, we have also added motion information in order to more reliably extract 3D information from visual scenes.

In the next section, we will focus on optic flow analysis and ask how we can analyse the different flow patterns in a scene which arise from ego-motion as well as from the motion of individual objects. This adds on to the use of RBM by providing us with a flow-based image segmentation in parallel to the above described multi-modal processing

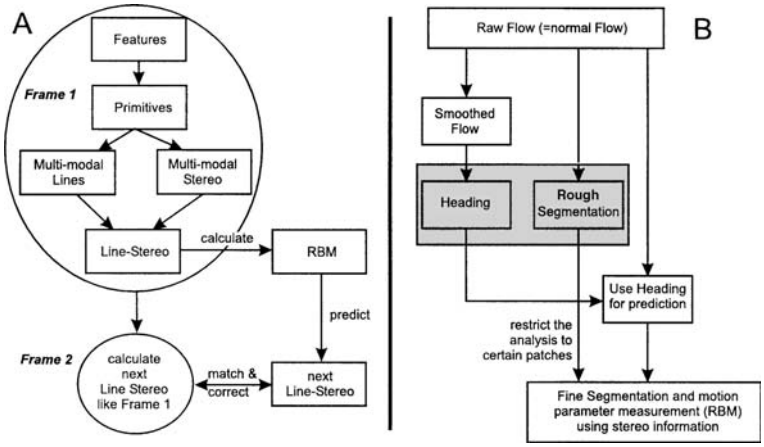

Figure 4. Multi-modal processing of stereo and motion. (A) The big ellipse shows the processing steps for stereo as described in Section 3.1 above. Combining this with rigid body motion estimation, as described below, we can predict the development of the different stereo pairs in the next camera frames and compare the prediction with the actually obtained results. (B) Processing of optic flow. For explanation see Section 3.2.

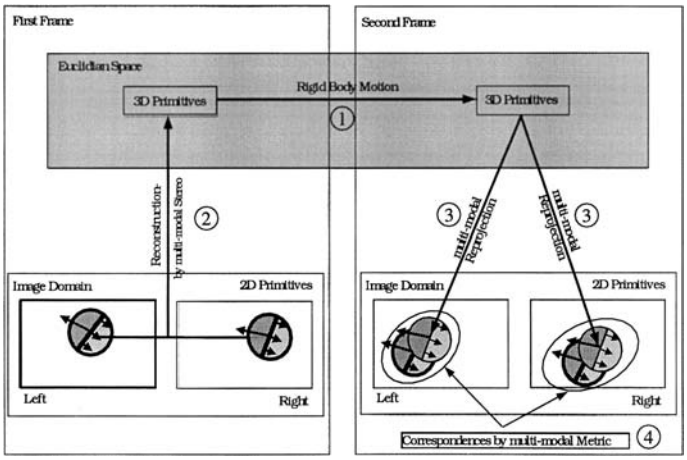

Figure 5. Including rigid body motion in the processing scheme for stereo. 2D information must be transferred to Euclidian space to calculate the RBM (2). This allows predicting the next set of 3D coordinates (1) which are back-projected into the 2D image space (3) and then compared with the actually obtained results from direct stereo calculations (4).

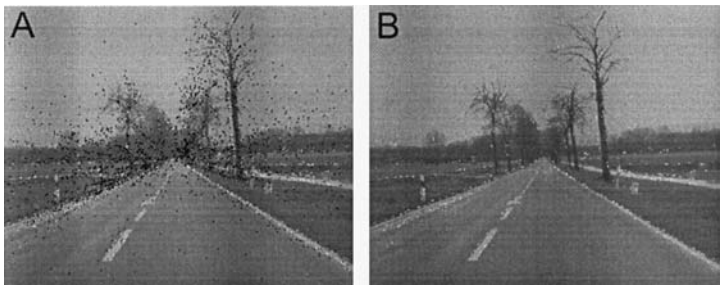

Figure 6. Results from including rigid body motion in the stereo processing calculated accumulating the information over 10 frame pairs. (A) Uncorrected matches, all matches after 10 frame pairs are shown and it is evident that mainly the black, wrong matches dominate. (B) Corrected matches superimposed onto the scene. The sky was artificially darkened to be able to show the confirmed matches in white.

steps. Why is this useful? Three aspects stand out: (a) As can be seen from the examples in Figures 3 and 6, the above described multi-modal processing is still not fully error-free and cross-checking with the image segmentation structure will further improve on this. (b) One result of the flow analysis described next is accurate information about the heading direction of the observer. This in itself is valuable for navigation purposes and (c) along the same lines: Quickly obtained information about the general structure (segments) of all moving objects, even if this information is coarse, is very valuable for obstacle avoidance and other related tasks.

3.2. *The processing of optic flow*

Optic flow analysis is mainly haunted by the so-called aperture problem: If an edge is viewed through a small aperture then only the ''normal'' motion-component, which is the one orthogonal to the orientation of the edge, will be resolved. This leads to the effect that all algorithms used to analyse flow will at edges yield normal flow as a result. Only at corners can true optic flow be measured. In addition, flow cannot be determined at untextured, smooth surfaces. Thus, flow-field maps always consist of many false estimates with only very few correct flow vectors and it is a very hard problem to extract meaningful information from such a map. Since observers are almost never entirely motionless themselves, it is, however, fair to say that most of the time a flow-field map will be dominated by the ego-motion flow pattern. Thus, to achieve improved flow analysis we have adopted a two-step strategy (Figure 4):

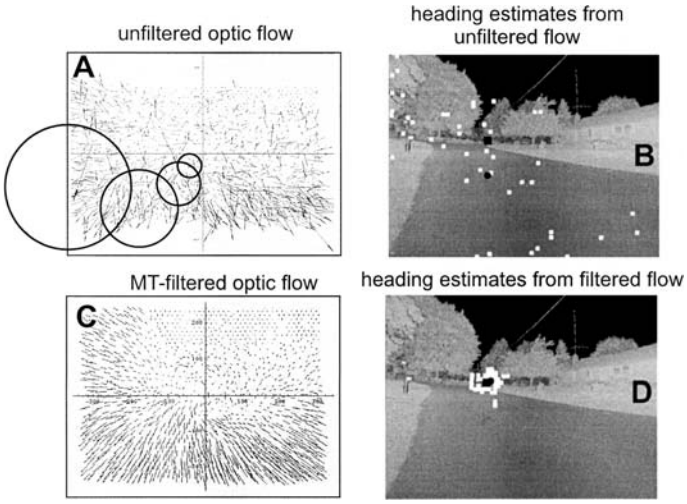

Figure 7. Processing of optic flow. (A) Flow field obtained with the Nagel algorithm. (B) Range image of an outdoor scene taken from the Brown Range Image data base (Huang et al., 2000). These images contain for all pixels the complete 3D information from laser range finder measurements. By geometrical extrapolation a movie of 10 frames has been created from this image which simulates a certain heading. True heading direction is given by the black square. Little white squares are estimates of the heading by analysing 150 randomly selected flow vectors from the flow field in (A). The black disc represent the average of these heading estimates. Clearly the individual estimates are widely scattered and their average is also not in correspondence with the true heading direction. (C) Flow field after MT filtering. The size of the used filters is indicated by the circles superimposed onto (A). The filtered flow field is much smoother. (D) Same as (B) but now the heading estimates are calculated from the MT-filtered flow field. Individual estimates are good and the average closely matches the true heading direction.

First, we determine the ego-motion flow component (''heading'') and ''subtract'' the resulting flow-field from the originally measured one. Then, we can find the other independently moving objects (IMOs) by analysing the residual error. Thus, this process amounts to scene segmentation into different moving objects.

To robustly extract ego-motion, we have invented a technique which utilises a neuronal property of the cells in the middle temporal area (MT) of monkey brain. In this area, big flow-sensitive receptive fields exist (Albright and Desimone, 1987). These fields scale with visual eccentricity, thus, fields that represent larger eccentricities are larger than those close to the fovea. As a consequence, it has been concluded that these fields average the flow to increase the robustness of its representation (Lappe, 1996). To this purpose, we have implemented

similar receptive field filters. The filtering technique improves the stability of the flow representation by averaging flow vectors within local neighbourhoods to stabilise the motion signal. Based on the properties of area MT this filtering method decreases noise by averaging flow vectors over image areas, which increase in size proportional to the eccentricity from the centre of the field of view (Figure 7A and C).

While averaging over large areas is more favourable for noise reduction and smoothing, averaging over small areas saves information such as local speed and velocity or local motion parallax. The spatial integration over peripherally increasing image areas is a compromise between both goals and well adjusted to the typical structure of the flow field elicited by self-motion. Small areas surrounding the centre of the view field contain sets of vectors with large deviations in the local flow direction, whereas the flow field in the periphery is more homogeneous allowing spatial averaging over a large scale without loosing information. The filter procedure results in an improved representation of the flow that is especially well suited for self-motion estimation (Figure 7A and C). The next stage of flow analysis, the heading estimation stage, acts on this representation. It is modelled after the next area of flow processing in the visual system, area MST (Lappe, 1998).

The MT-like filtering model was tested with optical flow fields obtained from image sequences with an optical flow algorithm. These tests involved both simulated camera motion through a natural scene with ground truth (Calow et al., 2004) and real motion sequences recorded in a moving car. To estimate performance improvements with respect to unfiltered optic flow, we randomly selected 150 flow vectors for a single run of the heading estimation and computed mean and standard deviation over several runs (Figure 7B and D). How the heading is actually determined will be discussed below. Here, Figure 7B and D shows first that MT-like filtering strongly improves the results. After MT-like filtering the width of the distribution of estimated headings dropped from 25° to 6°, the average error from 12° to 4°. These results demonstrate that MT-like filtering is a reasonable strategy to decrease noise in optical flow fields and to improve heading detection. The method works well on optical flow fields based on natural scenes affected by strong noise and the aperture problem.

The actual heading can be performed in different ways but we rely on an algorithm with which we can calculate heading as well as extract the motion patterns of other objects at the same time. In realistic scenes, the motion field generated by a moving observer is highly complex. This is primarily due to the interactions between translational and rotational

motion, the depth layout of the environment and the presence of IMOs. Nonetheless, often not more than a few entities are responsible for this motion field. The largest amount of structure is due to the observer's egomotion as discussed above, hence determining the latter is an important step towards the discovery of IMOs. The structure of Figure 8B shows how we extract such motion gestalts from optical flow fields. The algorithm consists of three main components which operate simultaneously. First, a novel method for the extraction of all egomotion parameters from optic flow fields has been developed. Using fixed-point iterations, the method robustly and efficiently deals with the nonlinear interaction between translation and rotation parameters. Thus, complex combinations of translation and rotation can also be reliably extracted. A second component corrects the intrinsic bias in the translation estimate that originates from the error norm used by most instantaneous-time egomotion algorithms. Finally, IMOs are considered as outliers and techniques from robust statistics are applied to make the egomotion estimation insensitive to their presence. Once the observer's motion parameters are extracted, IMOs are identified by evaluating the residual errors of all flow vectors against the egomotion model. Note that this approach can also cope with non-rigid IMOs.

The algorithm is demonstrated on a synthetic data set (McCane et al., 1998) for which the true optic flow is known. Three frames from

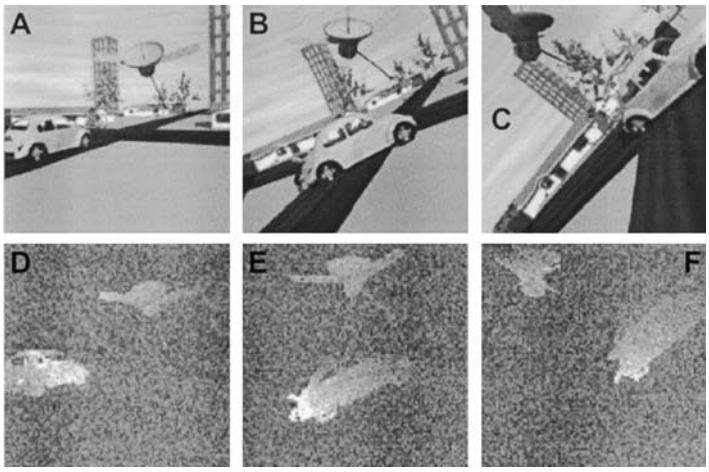

Figure 8. Processing of independently moving objects (IMOs). (A–C) Image sequence and (D–F) residual errors for frames 1, 15 and 30 of a complex motion sequence taken from McCane et al. (1998). The residual error can be used to extract the different IMOs in this sequence.

the sequence are shown in the top row of Figure 8. To account for the errors induced by optical flow algorithms, Gaussian noise vectors with a standard deviation of 10% of the average flow vector size were added to these flow fields. Both scene structure and motion patterns are complicated in this sequence. The observer motion consists of a combination of translation and rotation. There are two IMOs present, the car and the satellite, and they too undergo complex combinations of translation and rotation, which differ from the observer's. The segmentation results are shown in the bottom row of Figure 8. A lighter colour means a larger residual errors or a larger discrepancy between the flow vector and the model parameters. It is clear that the IMOs pop out quite strongly while the flow field generated by the scene structure is ignored.

3.3. *Summary of the employed multi-modal image analysis steps*

In the previous sections, we have described how to combine different scene analysis steps into an integrated image analysis system. The motivation for this was that each individual step is still error-ridden such that only a combination will yield reliable results. Specifically, we have shown that collinear line segments can be better trusted when not only the orientation of the lines match but also other features like colour or their associated optic flow vectors. The same is true for stereo pairs. Once such collinear lines are found, stereo analysis can then be further improved by relying on line-matches instead of point matches. Rigid body motion can be used to predict the development of the scene (more specifically of the found stereo-matches) and optic flow analysis can be used to extract heading information and to find all independently moving objects.

4. Covert active vision – Task dependent image processing

In Section 1, we have argued that active vision should be "task dependent". For example, the centre of interest should be different when performing a heading task (centre of interest focused on the heading direction) as compared to an obstacle avoidance task where the centre of interest should be always on the nearest obstacle. Very little is known about how the brain would implement such mechanisms, but we are able to provide the first basic results and concepts for a computer vision system. Here we rely mainly on a task-dependent dynamic remapping of the visual information: Our system is designed such that the magnifi-

cation of the image changes relative to the centre of interest. This part of the complete image analysis system is still not fully integrated with the rest but Figure 9 shows the basic principle for a task.

In this example, the task is to spot an approaching object (a car) in a rear view mirror image. Obviously the image of a car is small when it is far away, while it becomes bigger close by. This example is motivated by the demand for a driver assistant system which eliminates the blind-angle problem when a human monitors his/her rear view mirror while driving. Thus, it is important to spot an approaching car relatively early and to be able to track it in a reliable way until it is getting dangerously close (when a warning signal should be elicited). Part A of Figure 9 shows two original images, part B the remapped ones.

For other tasks, different mapping functions are possible. For example, the extracted heading information (see Section 3.2 above) can be used to magnify the image most at the focus of the heading (data not shown).

Camera systems particularly in industrial or outdoor (e.g. while driving) environments are preferably mounted in a rigid way and fixation movements are too demanding with respect to the mechanical robustness of such systems. As described in Section 2.1, our visual system performs magnification at the fovea which is usually aligned with the centre of interest by means of fixational eye-movements. However, because of the reasons given above such a foveation strategy cannot usually be applied in technical systems. The task-dependent remapping suggested here can solve this problem. Without having to move the cameras, the magnification will be highest at the most relevant image locations.

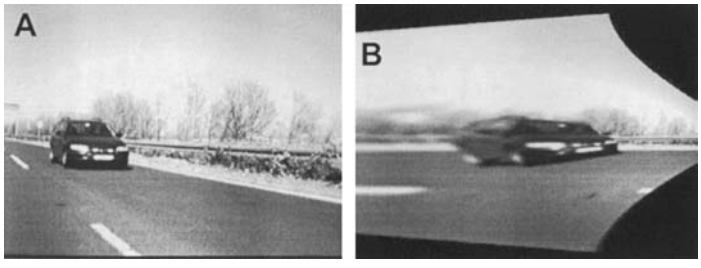

Figure 9. Task dependent image remapping. (A) Original image from a rear view mirror. (B) Remapped image. Regardless of distance the size of the car in the image is always roughly the same, which makes detection much easier.

5. Conclusions

The goal of this article was twofold: (1) To summarise the multiple processing strategies employed by the vertebrate visual system as far as they are known to date and (2) to demonstrate that it is possible to adapt some of these strategies to a computer vision system. The results shown in the second part of this paper have demonstrated that multi-modal image processing has the potential to improve scene analysis to a large degree. In particular, one must realize that it does not make sense to try to create a carbon-copy of the existing natural visual systems. Instead, it is necessary to arrive at abstractions which are better suited for implementation in a technical system. It is, however, obvious that there is still a large amount of work to be done until such systems will be robust and fast enough for any real-time application.

Acknowledgements

This paper describes concepts and results from the ECOVISION project, funded by the European Commission, which we gratefully acknowledge.

Notes

[1] This effect could be the basis of the so-called "waterfall" motion after-effect (Mather and Verstraten, 1998): i.e. when concentrating for some time on a wide motion field with constant velocity and direction (like falling water in a waterfall) the observer will perceive motion in the opposite direction as soon as he looks away. An imbalance between the spontaneous firing of the adapted downward selective cells, versus the non-adapted upward selective cells could be the source of this percept.

References

Adams D and Horton J (2003) A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. Journal of Neuroscience 23: 3771–3789

Aertsen AM, Gerstein GL, Habib MK and Palm G (1989) Dynamics of neuronal firing correlation: modulation of "effective connectivity". Journal of Neurophysiology 61(5): 900–917

Albright TD and Desimone R (1987) Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Experimental Brain Research 65: 582–592

Aloimonos Y, Weiss I and Bandopadhay A (1987) Active vision. International Journal of Computer Vision 1: 333–356

Broadbend D (1965) Information processing in the nervous system. Science 150: 457–462

Brooks R (1991) Intelligence without reason. International Joint Conference on Artificial Intelligence pp. 569–595

Büchel C and Friston K (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fmri. Cerebral Cortex 7: 768–778

Calow D, Krüger N, Lappe M and Wörgötter F (2004) Space variant filtering of optic flow for robust three dimensional motion estimation. In: Engineering in Intelligent Systems – EIS 2004

Chapman B (2004) The development of eye-specific segregation in the retinogenicolostriate pathway. In: Chalupa LM and Werner JS (eds) The Visual Neurosciences, Vol. 1(8), pp. 108–116. Cambridge, MA, USA, MIT Press

Chapman B and Stone L (1996) Turning a blind eye to cortical receptive fields. Neuron 16: 9–12

Connor C, DC P, Gallant J and van Essen D (1997) Spatial attention effects in macaque area v4. Journal of Neuroscience 17: 3201–3214

Crick F (1984) Function of the thalamic reticular complex: The searchlight hypothesis. Proceedings of the National Academy of Sciences of the USA 81: 4586–4590

Das A and Gilbert C (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature 399: 655–661

DeAngelis G, Anzai A, Ohzawa I and Freeman R (1995) Receptive field structure in the visual cortex: does selective stimulation induce plasticity? In: Proceedings of the National Academy of Science, pp. 9682–9686, USA

Dennett DC (1984) Cognitive wheels: The frame problem of AI. In: Hookway C (ed) Minds, Machines and Evolution, pp. 129–151. Cambridge University Press

Desimone R and Duncan J (1995) Neural mechanisms of selective visual attention. Annual Review of Neuroscience 18: 193–222

Dossi R, Nunez C and Steriade M (1992) Electrophysiology of a slow (0.5–4 Hz) intrinsic oscillation of cat thalamocortical neurones in vivo. Journal of Physiology 447: 215–234

Dragoi V, Sharma J and Sur M (2000) Adaptation-induced plasticity of orientation tuning in adult visual cortex. Neuron 28(1): 287–298

Erwin E, Obermayer K and Schulten K (1995) Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. Neural Computation 7: 425–468

Eyding D, Macklis JD, Neubacher U, Funke K and Wörgötter F (2003) Selective elimination of corticogeniculate feedback abolishes the electroencephalogram dependence of primary visual cortical receptive fields and reduces their spatial specificity. Journal of Neuroscience 23(18): 7021–7033

Felsberg M and Krüger N (2003) A probablistic definition of intrinsic dimensionality for images. Pattern Recognition, 24th DAGM Symposium

Felsberg M and Sommer G (2001) The monogenic signal. IEEE Transactions on Signal Processing 49(12): 3136–3144

Funke K and Eysel U (1992) Eeg-dependent modulation of response dynamics of cat "dLGN" relay cells and the contribution of corticogeniculate feedback. Brain Research 573: 217–227

Geman S, Bienenstock E and Doursat R (1995) Neural networks and the bias/variance dilemma. Neural Computation 4: 1–58

Gilbert C (1998) Adult cortical dynamics. Physiology Reviews 78: 467–485

Gilbert C and Wiesel T (1990) The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. Vision Research 30: 1689–1701

Grossberg S, Mingolla E and Ross W (1994) A neural theory of visual search: Interactions of boundary, surface, spatial and object representations. Psychological Review 101: 470–489

Gulyas B, Orban G, Duysens J and Maes H (1987) The suppressive influence of moving textured backgrounds on responses of cat striate neurons to moving bars. Journal of Neurophysiology 57: 1767–1791

Gulyas B, Spileers W and Orban G (1990) Modulation by a moving texture of cat area 18 neuron responses to moving bars. Journal of Neurophysiology 63: 404–423

Hamker FH (2003) The reentry hypothesis: linking eye movements to visual perception. Jounal of Vision 11: 808–816

Hammond P and McKay D (1975) Differential responses of cat visual cortical cells to textured stimuli. Experimental Brain Research 22: 427–430

Hammond P and McKay D (1977) Differential responsiveness of imple and complex cells in cat striate cortex to visual texture. Experimental Brain Research 30: 275–296

Hammond P and McKay D (1981) Modulatory influences of moving textured backgrounds on responsiveness of simple cells in feline striate cortex. Journal of Physiology 319: 431–442

Huang J, Lee AB and Mumford D (2000) Statistics of range images. *CVPR*

Hubel DH and Livingstone MS (1987) Segregation of form, color, and stereopsis in primate area 18. Journal of Neuroscience 7(11): 3378–3415

Ikeda H and Wright MJ (1974) Sensitivity of neutones in visual cortex (area 17) under different levels of anaesthesia. Experimental Brain Research 20: 471–484

Johnston, A (1986) A spatial property of the retino-cortical mapping. Spatial vision 1: 319–331

Julesz B (1981) Textons, the elements of texture perception, and their interactions. Nature 290: 91–97

Kastner S, Nothdurft H and Pigarev I (1997) Neuronal correlates of pop-out in cat striate cortex. Vision Research 37: 371–376

Knierim J and v. Essen D (1992) Neuronal responses to static texture patterns in area v1 of the alert macaque monkey. Journal of Neurophysiology 67: 961–980

Körding KP and Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427: 244–247

Krüger N and Felsberg M (2003) A continuous formulation of intrinsic dimension. Proceedings of the British Machine Vision Conference

Krüger N and Wörgötter F (2002) Multi modal estimation of collinearity and parallelism in natural image sequences. Network: Computation in Neural Systems 13: 553–576

Krüger N and Wörgötter F (2004) Statistical and deterministic regularities: Utilization of motion and grouping in biological and artificial visual systems. Advances in Imaging and Electron Physics, in press

Krüger N, Felsberg M, Gebken C and Pörksen M (2002) An explicit and compact coding of geometric and structural information applied to stereo processing. Proceedings of the workshop 'Vision, Modeling and Visualization 2002'

Krüger N, Lappe M and Wörgötter F (2004) Biologically motivated multi-modal processing of visual primitives. The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour 1(5): in press.

Lamme V (1995) The neurophysiology of figure-ground segregation in primary visual cortex. Journal of Neuroscience 15: 1605–1615

Lappe M (1996) Functional consequences of an integration of motion and stereopsis in area MT of monkey extrastriate visual cortex. Neural Computation 8: 1449–1461

Lappe M (1998) A model of the combination of optic flow and extraretinal eye movement signals in primate extrastriate visual cortex. Neural Networks 11: 397–414

Livingstone MS and Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. Journal of Neuroscience 4(1): 309–356

Mallot HA (1985) An overall description of retinotopic mapping in the cat's visual cortex areas 17, 18, and 19. Biological Cybernetics, 52: 42–51

Marr D (1982) Vision. W. H. Freeman and Company, New York

Mather G and Verstraten F (1998). The Motion Aftereffect: A Modern Perspective. MIT Press, Cambridge

Maunsell JHR and McAdams CJ (2001) Effects of attention on the responsiveness and selectivity of individual neurons in visual cerebral cortex. In: Braun J, Koch C and Davis JL (ed), Visual Attention and Cortical Circuits, pp. 103–119. Cambridge, MA, USA, MIT Press

McAdams C and Maunsell J (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. Journal of Neuroscience 19: 431–441

McCane B, Galvin B and Novins K (1998) On the evaluation of optical flow algorithms. In: Fifth International Conference on Control, Automation, Robotics &Vision, Vol. 1, pp. 1563–1567, Singapore

Moran J and Desimone R (1985) Selective attention gates visual processing in extrastriate cortex. Science 229: 782–784

Munk M, Roelfsema P, Engel A and Singer W (1996) Role of reticular activation in the modulation of intracortical synchronization. Science 272: 271–274

Nagel HH and Enkelmann W (1986) An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 8: 565–593

Neisser U (1967) Cognitive Psychology. Appleton-Century-Crofts, New York

Niebur E and Wörgötter F (1994) Design principles of columnar organization in visual cortex. Neural Computation 6: 602–614

Niebur E, Koch C and Rosin C (1993) An oscillation-based model for the neuronal basis of attention. Vision Research 33(18): 2789–2802

Nothdurft H, Gallant J and v.Essen D (1999) Response modulation by texture surround in primate area v1: correlates of "popout" under anesthesia. Visual Neuroscience, 16: 15–34

Orban G, Gulyas B and Vogels R (1987) Influence of a moving textured background on direction selectivity of cat striate neurons. Journal of Neurophysiology 57: 1792–1812

Orban G, Gulyas B and Spileers W (1988) Influence of moving textured backgrounds on responses of cat area 18 cells to moving bars. Progress in Brain Research, 75: 137–145

Peterhans E and von der Heydt R (1993) Functional organization of area V2 in the alert macaque. European Journal of Neuroscience, 5(5): 509–524

Pettet M and Gilbert C (1992) Dynamic changes in receptive-field size in cat primary visual cortex. In: Proceedings of the National Academic of Sciences, pp. 8366–8370. USA

Posner M, Cohen Y and Rafal R (1982) Neural systems control of spatial orienting. Philosophical Transactions of the Royal Society of London B 298: 187–198

Prodöhl C, Würtz RP and von der Malsburg C (2003) Learning the gestalt rule of collinearity from object motion. Neural Computation 15(8): 1865–1896

Pugeault N and Krüger N (2003) Multi–modal matching applied to stereo. Proceedings of the BMVC 2003

Pugeault N, Wörgötter F and Krüger N (2004) A non-local stereo similarity based on collinear groups. Fourth International ICSC Symposium on Engineering of Intelligent Systems

Rao RPN and Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2: 79–87

Rizzolatti G, Riggio L, Dascola I and Umiltá C (1987) Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. Neuropsychologica 25: 31–40

Schwartz EL (1977) Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. Biological Cybernetics 25: 181–194

Schwartz EL (1980) Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. Vision Research 20: 645–669

Shipp S and Zeki S (2002a) The functional organization of area V2, I: specialization across stripes and layers. Visual Neuroscience 19(2): 187–210

Shipp S and Zeki S (2002b) The functional organization of area V2, II: the impact of stripes on visual topography. Visual Neuroscience 19(2): 211–231

Sillito A and Jones H (1996) Context-dependent interactions and visual processing in v1. Journal of Physiology (Paris) 90: 205–209

Sillito A, Grieve K, Jones H, Cudeiro J and Davis J (1995) Visual cortical mechanisms detecting focal orientation discontinuities. Nature 378: 492–496

Somers DC, Nelson SB and Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. Journal of Neuroscience 15(8): 5448–5465

Spillmann L and Ehrenstein WH (2004) Gestalt factors in the visual neurosciences. In: Chalupa LM and Werner JS (ed) The Visual Neurosciences, Vol. 2(106), pp. 1573–1589. Cambridge, MA, USA, MIT Press

Steriade M (1991) Cerebral Cortex, Vol. 9, chapter Alertness, quiet sleep, dreaming, pp. 279–357. Kluwer Academic/Plenum Publishers

Suarez H, Koch C and Douglas R (1995) Modeling direction selectivity of simple cells in striate visual cortex within the framework of the canonical microcircuit. Journal of Neuroscience 15(10): 6700–6719

Swindale NV (1996) The development of topography in the visual cortex: A review of models. Network: Computation in Neural Systems 7: 161–247

Treisman A (1969) Strategies and models of selective attention. Psychological Review 76: 282–299

Treisman A and Gelade G (1980) A feature-integration theory of attention. Cognitive Psychology 12: 97–136

Treue S and Martinez Trujillo J (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399: 575–579

Ts'o DY and Gilbert CD (1988) The organization of chromatic and spatial interactions in the primate striate cortex. Journal of Neuroscience 8(5): 1712–1727

Ts'o DY, Roe AW and Gilbert CD (2001) A hierarchy of the functional organization for color, form and disparity in primate visual area V2. Vision Research 41(10–11): 1333–1349

Ungerleider LG and Pasternak T (2004) Ventral and dorsal cortical processing streams. In: Chalupa LM and Werner JS (ed) The Visual Neurosciences, Vol. 1(34), pp. 541–562. Cambridge, MA, USA, MIT Press

Volchan E and Gilbert C (1994) Interocular transfer of receptive field expansion in cat visual cortex. Vision Research 35: 1–6

Wolfe J (1998) Attention, chapter Visual Search, pp. 13–74. Psychology Press Ltd

Wong-Riley M (1979) Changes in the visual system of monocularly sutured or enucleated cats demonstrable with cytochrome oxidase histochemistry. Brain Research 171(1): 11–28

Wörgötter F, Suder K, Zhao Y, Kerscher N, Eysel U and Funke K (1998) State-dependent receptive-field restructuring in the visual cortex. Nature 396: 165–168

Wörgötter F, Eyding D, Macklis JD and Funke K (2002) The influence of the corticothalamic projection on responses in thalamus and cortex. Philosophical Transactions of the Royal Society of London B Biological Science 357(1428): 1823–1834

Wurtz R, Goldberg M and Robinson D (1982) Brain mechanisms of visual attention. Scientific American 246: 124–135

Zeki SM (1978) Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. Journal of Physiology 277: 273–290

Zipser K, Lamme V and Schiller P (1996) Contextual modulation in primary visual cortex. Journal of Neuroscience 15: 7376–7389