M.V. Srinivasan, R. Hengstenberg,
H. A. Mallot, and S. Venkatesh

Active Vision

Traditionally, the design of machines and robots that "see" has been approached from "first principles" — primarily physics, mathematics and geometry — with relatively little consideration as to how the problems that are tackled might be solved by natural visual systems. While the first-principles approach is a perfectly valid one, the solutions that it produces often tend to be computationally complex, sensitive to noise, and sometimes too general-purpose to be really effective.

On the other hand, it is patently clear that biological vision is rapid, reliable and robust. One reason for the impressive performance of natural vision may be that — unlike many automated vision systems — it is tailored to a specific purpose, or set of purposes, that best serve the animal's needs. Another salient feature of natural vision is that it is "acquisitive". That is, animals (humans included) rarely sit and "take in" a scene in a purely passive manner: they continually move and interact with the environment in order to glean information about it. This strategy considerably simplifies the task of analysing the scene.

The aim of the working group on "Active Vision" was to review (if possible, in a book) some of the distinctive principles of natural vision and to explore ways of taking advantage of these tricks and "short cuts" for the design of robust algorithms for machine vision. The group consisted of (in alphabetical order) *Dr. Roland Hengstenberg,* a neurobiologist from the Max-Planck-Institut für biologische Kybernetik, Tübingen, *Dr. Hanspeter Mallot,* a theoretical biologist from the Max-Planck-Institut für biologische Kybernetik, Tübingen, *Dr. Mandyam Srinivasan,* an animal behaviourist from the Australian National University, Canberra and *Dr. Svetha Venkatesh,* a computer scientist from Curtin University, Perth. During the latter half of our stay we were joined by *Dr. Erhardt Barth,* an applied mathematician who joined the Kolleg as an Ansgar Rumler Stipendiat (financed by the Freundeskreis des Wissenschaftskollegs) and assisted us with computer simulations.

The group began by holding weekly meetings where we informed each other about our work, background and research interests. These meetings were interspersed with additional get-togethers where we had in-depth discussions about topics of potential interest that should be considered for inclusion in the book. Over the first month it became clear that there were a number of relevant topics within each of our individual research areas. Moreover, although books have already been published on the topic of "Active Vision", all of them so far have been a collection of chapters written by different authors and put together by an editor (or editors). They have almost always been the outgrowth of symposia or workshops on the topic. This is not necessarily the ideal recipe for a book: the resulting product may not be fully comprehensive (given the difficulty of finding an author to cover each topic of interest) or fully coherent (given to the multiplicity of authors). In most circumstances, the best that an editor can do (apart from heavy-handed editing which quickly loses friends and potential collaborators) is to write an introduction that attempts to draw a unifying thread *(einen roten Faden)* through the various chapters. Speaking from some of our own earlier experience, this is an unenviable task that is not always completely successful. Therefore, we decided that there was indeed a need for a coherent book on the topic (coherent, at least from our point of view!) written by a small number of closely collaborating authors. Accordingly, we formulated a rough outline for the book, and decided that each of us would write a set of chapters covering his or her speciality in relation to the field. However, as time progressed and the process of reading and writing evolved, we began to discover that we may have taken on a rather Herculean task. We realised that there is a veritable plethora of examples of "Active Vision". Indeed, almost every aspect of vision — animate or inanimate — can be considered to be "active" in one way or another. Thus, writing a comprehensive book on "Active Vision" would entail writing a comprehensive book on Vision itself! This point was brought home to us rather forcefully at the Kolleg-sponsored workshop that we organised on the theme of *Active Vision in Animals and Machines* (details below). There, in one of the sessions, we presented a working outline of our book and invited comments and constructive criticism from the participants. The response was loud and clear: "Your topic is too general and too diffuse! Forget about trying to cover the whole field, you can't please everybody. Focus your attention instead on a specific audience interested in a specific aspect of `active vision'. It was suggested to us that the thing that engineers would most like to read about would be a "bug book for engineers": a compendium of facts about insect sensory mechanisms that may suggest novel engineering applications. After the workshop, we discussed the matter amongst ourselves at length and decided that the advice was indeed valuable. We drew up a revised outline — not necessarily for a "bug book for engineers" — but for a book that would restrict itself to the principles of vision in insects, together with actual and potential applications in the fields of machine vision and robotics.

The outline of the book is now as follows:

Tentative title:

SEEING WITH SIX LEGS: INSECT VISION, AND APPLICATIONS TO COMPUTER VISION AND ROBOTICS

1. General Introduction

This section outlines why it is useful to study insects, and how this can help us design better machines.

2. Eye Structures and Optics

This section describes the compound eyes of insects, how they extract and represent panoramic (360 degree) vision, and how information on intensity, colour and polarisation is represented by the photoreceptors in the compound eye. A variety of specialisations are also described, such as the "mirror" compound eyes with special reflecting structures that have been emulated in X-ray telescopes, the dorsal eye region, or "love spot", in male insects that is used to detect and chase females, ocelli that help stabilise flight attitude by detecting shifts of the visual horizon, the eyes of certain spiders and marine organisms that have evolved special scanning mechanisms to recognise objects through the spatio-temporal intensity signatures that they generate, and so on.

3. Reflexive Vision

This section describes a variety of reflexive behaviours. Some examples are the optomotor response, which stabilises roll, yaw and pitch and helps the flying insect maintain a straight course, the centering response, which enables flying insects to negotiate narrow gaps safely, visual control of flight speed and landing, chasing and tracking behaviours, and active camouflage of self-motion. The principles and models of motion detection underlying these behaviours are discussed, as are the underlying neural mechanisms. New simulations incorporating these principles for course control and landing have been carried out and are described. We also describe computer-vision algorithms and robots that have been developed in a number of different laboratories around the world, based on these principles.

4. Acquisitive Vision

Insects are more than just reflexive creatures: they actively acquire information about their surroundings. Unlike us, insects are ill-equipped to extract stereoscopic depth cues from the environment. This is because the eyes of most insects are so close together that they capture virtually identical images, making it difficult to compute depth in the conventional way. Instead, most insects use cues based on image motion to infer depth: when an insect flies in a straight line, the images of nearby objects appear to move faster than those of remote objects. Thus, the range of an object or surface is estimated in terms of the speed of its image on the eye. Grasshoppers, for example, peer (rock the head from side to side) before they jump on to a nearby target. The range of the target is inferred in terms of the motion of its image: the closer the target, the larger the motion. Flying insects use a similar principle to gauge the distances to various objects. Objects are distinguished from their backgrounds in terms of the relative motion between their images. Wasps leaving their nest for the first time or bees leaving a newly-discovered food source perform stereotyped, arcing flights around the object of interest. The structure of these flights appears specially adapted to extract the three-dimensional layout of the environment in the vicinity of the nest or food source. We describe these behaviours, as well as a variety of computational algorithms and robots that use these principles for navigating in unknown environments. We also describe the perception of depth by the praying mantis, so far the only insect in which binocular depth vision has been unequivocally demonstrated.

5. Navigation

The book ends with a brief chapter on visual navigation in insects. Among the topics discussed are the celestial compass and its use by ants and bees, dead-reckoning, odometry, landmark-based navigation, and search strategies. Computational models and robots exploring some of these strategies are also described.

At the time of writing this report, the text for the book is ca. 80% complete. However, it requires polishing. Many illustrations, presently in the form of rough sketches, have to be redrawn to publication quality. Given that progress on the book is likely to be slower once we are all back at our home institutions (due to other demands on our time), we anticipate that it will be another six months before a final version of the manuscript is available.

Workshop on Active Vision in Animals and Machines

With the help and financial support of the Wissenschaftskolleg, the working group on "Active Vision" organised a workshop on the theme of "Active Vision in Animals and Machines" which was held at the Wissenschaftskolleg on 22-24 March, 1997. The object of the workshop was to bring together biologists, studying human and animal vision, with engineers and computer scientists, working on computer vision and robotics. The aim was to explore the state of the art in this relatively new area of interdisciplinary research, and to examine whether recent progress in either field — biology or engineering — could stimulate new concepts or approaches in the other. In order to facilitate useful, in-depth discussion, the number of participants was deliberately kept relatively small. We are grateful to the Wissenschaftskolleg for financing this workshop and helping organise it. Special thanks go to *Andrea Friedrich* and *Katharina Wiedemann* for helping make the workshop such a successful and enjoyable event for all of the participants.

The participants in the workshop were (in alphabetical order): *Ruzena Bajcsy* (University of Pennsylvania, Philadelphia), *Dana Ballard* (University of Rochester), *Martin Banks* (University of California, Berkeley), *Gary Bernard* (Boeing Aeroplane Company, Seattle), *Heinrich Bülthoff* (Max-Planck-Institut für biologische Kybernetik, Tübingen), *Jan-Olof Eklundh* (Royal Institute of Technology, KTH, Stockholm), *Nicolas Franceschini* (CNRS, Marseilles), *John Frisby* (University of Sheffield), *Roland Hengstenberg* (Wissenschaftskolleg zu Berlin and Max-Planck-Institut für biologische Kybernetik, Tübingen), *Hanspeter Mallot* (Wissenschaftskolleg zu Berlin and Max-Planck-Institut für biologische Kybernetik, Tübingen), *Ingo Rentschler* (University of Munich), *Samuel Rossel* (University of Freiburg), *Giulio Sandini* (University of Genoa), *Mandyam Srinivasan* (Wissenschaftskolleg zu Berlin and Australian National University, Canberra), *Svetha Venkatesh* (Wissenschaftskolleg zu Berlin and Curtin University, Perth) and *William Warren* (Brown University, Providence).

In toto, the participants covered a wide range of sub-fields of research, ranging from insect vision through human psychophysics to robotics. The workshop was organised in the form of five serial sessions, each covering a fairly broad topic. These sessions, and a (very) brief summary of each participant's contribution are given below (with apologies to the individual participants for any inadvertent misrepresentations!).

Session 1: Foveation and its Advantages

The foveal region of the human retina, with its higher photoreceptor density, conveys to the brain more spatial detail about the environment than do other retinal regions. *Ingo Rentschler* discussed the implications of the so-called "cortical magnification factor", its variation across the visual field, and the extent to which various aspects of human visual performance are influenced by this factor. *Svetha Venkatesh* highlighted the advantages of incorporating active, foveate vision into machine vision. She illustrated this in the context of two specific tasks. One task was the determination of whether or not an object is capable of "containment", e.g. whether an object can be used as a cup to hold tea. If a camera moves over a cup-like object whilst fixating a point on the upper rim, the emergence of new features (such as the lower edge of the bottom) will reveal that the vessel has containment capability. In a second example, she showed how a fovea with a log-polar mapping property can simplify the process of detecting salient features in an object's image and moving successively from one feature to the next, or tracking a given feature in time as the parent object moves. *Jan-Olaf Eklundh* described a novel strategy, using gaze control, accommodation and movement detection, by which a computer vision system could separately track three people moving at three different speeds and depths in a room, whilst the vision system itself was in motion as part of a moving robot.

Session 2: Vision for Mobility

When a fly makes a banked turn, it holds its head horizontal relative to the external world. *Roland Hengstenberg* described the visual and mechanosensory processes by which flies stabilise their heads. He also described the functional organisation of the large-field motion-detecting neurons in the brain of the fly that serve to detect and compensate for disturbances in roll, pitch and yaw, and showed that these neurons are "matched filters" tuned to detect rotation about, or translation along, specific axes. *Nicolas Franceschini* described an autonomously navigating robot, endowed with fly-like vision, which successfully negotiates a cluttered environment by using cues derived from image motion to detect potential obstacles and avoid collisions with them. An interesting feature of this machine is the control of speed based on a "radius of vision". This strategy ensures that the speed of the robot is always safely matched to the proximity of the nearest potential obstacle. *Mandyam Srinivasan* summarised the ways in which flying insects use image-motion cues to perform a variety of manoeuvres and visual tasks, such as negotiating narrow gaps, regulating flight speed, executing smooth landings, distinguishing objects from backgrounds, and estimating distance flown. He also described autonomously navigating robots which incorporated some of these principles. *Giulio Sandini* presented a broad discussion of new challenges in the fields of neuroscience and artificial system design, and illustrated them with specific issues such as the design of artificial gaze control systems. He explained that log-polar sensors, for example, are excellent for control of binocular vergence because they place high acuity only in the region where it is needed. Evidence was also presented to suggest that compensatory eye movements, driven by the vestibular apparatus, may be more sophisticated than previously believed: this control system seems to correct for the range of the viewed object, and also for the fact that the two eyes need to rotate by different amounts when the object is at a finite range. *Ruzena Bajcsy* discussed the problems of dynamic vision in relation to the control of autonomously mobile, visually guided robots. She presented a scheme for tracking a target in an environment cluttered with obstacles, which incorporated attractor and repellor fields, nonlinear control to facilitate smooth transitions between different modes of operation, and which contained impressively few "free" (adjustable) parameters. *Martin Banks* discussed the use of retinal and extra-retinal information in the subjective evaluation of self-motion and scene layout. Data from his laboratory revealed that the perceived path of self-motion depends upon the velocity field in the retina and upon a variety of extra-retinal signals that inform the nervous system about the rotation of the eyes, head, and body. Estimates of the slant, tilt and curvature of a surface are based on three calculations: horizontal disparities corrected by felt eye position, horizontal disparities corrected by vertical disparities, and monocular perspective cues. *William Warren* elucidated the dynamics of perception and action, viewing the two in combination in a closed loop. Babies on a baby-bouncer quickly learn to adjust leg stiffness and pushing frequency to maximise the amplitude of the bounce (hence the term *bouncing baby boy?!)*. The visual control of locomotion was explored in subjects walking on a treadmill whilst viewing a dynamic visual display. Whilst walking through a doorway, for example, the evolution of the walking trajectory is controlled by pattern-based cues, as well as the location of the focus of expansion.

Session 3: Binocular Vision and Stereopsis

Samuel Rossel described binocular vision in the praying mantis, so far the only insect known to use stereo mechanisms to evaluate depth so that it may properly aim its strike at an appetising fly (Rossel's own discovery, some 17 years ago). He presented evidence that even mantids solve the tricky "correspondence" problem, and organise predictive strikes at moving targets. While the stereoscopic mechanism is impressively precise with regard to the computation of depth, it does not use a "photographic" representation of the images captured by the two eyes. Rather, depth computation is based on the matching of a few key, global features of the objects, so that fusion can occur even if the images in the two eyes are considerably different. *John Frisby* described strategies for "active stereopsis" in human vision. Vergence-dependent torsion of the eyes, for example, facilitates the maintenance of metric stereo constancy and is well adapted to viewing horizontal surfaces below eye level or nearby overhanging objects above eye level.

Session 4: Scene Analysis

Heinrich Bülthoff presented convincing evidence to indicate that the visual system does not use internal models to represent inanimate objects, faces or scenes: rather, it uses a set of views, and interpolates between them in order to recognise objects in unfamiliar views. This is true even in recognising dynamic objects. *Hanspeter Mallot* continued this theme by marshalling evidence for view-based representations in navigation. Human subjects exploring a virtual ("Hexatown") maze can learn to navigate through the maze simply by associating specific views with specific motor actions, and without invoking a "cognitive map". He presented a graph-theoretic representation of the maze problem that can be used to model spatial memory at a number of different levels of complexity. *Dana Ballard* presented a case for viewing the human cortex as a predictive Kalman filter which used current and past observations to make a prediction of what to expect next. Information was transmitted to the brain only when the prediction did not agree with what was actually observed or experienced. Such a model accounts for a number of complex, psychophysically observed after-effects.

Session 5: What Exactly is "Active Vision", and Where is it Going?

In this discussion session we attempted to arrive at a clear definition of the term "Active Vision". It became evident that the term is rather fuzzy and has been used (and abused) in a variety of different contexts. It was difficult to arrive at a precise definition, since almost every aspect of vision could be considered to be "active" in some sense. For example, a grasshopper "peers" (sways its head from side to side) to estimate the range to an object in terms of the speed of the object's image on the retina. While one might declare this to be an example of `Active Vision" without any hesitation, it is noteworthy that the same information on depth can be obtained by using two stationary eyes (binocular vision) and measuring the apparent motion of the target in the two images. Is binocular vision then also "Active Vision"? And can vision be termed "active" simply because the camera or eye moves? The general consensus was that it is difficult to define "Active Vision" precisely, and that the term itself has probably outlived its useful life.

In this session we also presented a working outline of our book-in-progress to the participants, and invited comments and criticisms. This proved to be a very useful exercise. A very pertinent question raised by the audience was "To whom are you addressing the book? Engineers, biologists, or both?" Critical questions like this helped us focus our efforts, and sharpened our ideas about what belonged in the book and what didn't. The result, we hope, is a book which — though now directed at a smaller, more specialised audience — will prove to be of interest to that audience.

We thank the Wissenschaftskolleg for bringing us together in Berlin and providing us with a wonderful opportunity to work on this project in a creative, stimulating, environment, free from everyday concerns. We would never have undertaken this project on our own, without the encouragement, support and hospitality of this unique institution.