Max-Planck-Institut
für biologische Kybernetik
Arbeitsgruppe Bülthoff

Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 30 May 1996

Scene Recognition Workshop Tübingen July 3-5, 1996

Organized by: Michael J. Tarr and Heinrich H. Bülthoff

Abstract

In the past few years the question of how humans recognize and analyze natural, complex scenes has received renewed attention, resulting in new techniques and theories. The purpose of this workshop is to bring together researchers interested in these questions from both Europe and the United States to discuss latest results and ideas for the future.

Supported by the Max-Planck Gesellschaft and a TransCoop grant from DAAK to MJT and HHB

1 Introduction to the Scene Perception Workshop

Heinrich Bülthoff
Max-Planck Institute for Biological Cybernetics
hhb@mpik-tueb.mpg.de
http://www.mpik-tueb.mpg.de/bu.html

In the past several years we have made significant progress on how man, monkeys and machines recognize three-dimensional objects, provided that these objects are presented in isolation or can be easily segmented from the background. However, we have very little understanding of how one can recognize objects in a more ecological valid context, i.e. in a realistic scene. While the bottom-up segmentation of objects from a complicated natural scene makes the recognition process more difficult it is quite plausible that the context information provided by the scene can help a top-down recognition process. In order to integrate the context information into a novel recognition framework we have to understand first how we integrate the patch-wise information when we change our gaze position 3-4 times per second. The question of how we integrate information across saccades is the main topic of most talks of this workshop and the major tools to study it are the recently introduced "flicker" paradigm of Rensink, O'Regan and Clark (1995) and high speed eye tracking techniques for TSI experiments introduced by McConkie.

State of the art computer graphics and virtual reality technology open up new opportunities for studying scene perception in an even more ecological context by allowing us to interact with the scene ("perception for action").

In this workshop we would like to review our current understanding of scene perception and discuss new ideas about how we can make use of this new technology to get a better understanding of scene perception.

2 Attention and the perception of dynamic scenes

Ronald A. Rensink
Cambridge Basic Research Laboratory
rensink@pathfinder.cbr.com
http://pathfinder.cbr.com/people/rensink/rensink.html

When looking at a dynamic environment, our impression as observers is that we simultaneously see all the changes that are taking place. It will be shown that this impression is an illusion, and that humans instead have a severely limited ability to detect change. It will be argued that attention is required to perceive changes in a scene, and that the limited ability to detect change is a direct consequence of the limited capacity of the attentional mechanisms involved.

Previous studies (e.g., McConkie et al. 1996) showed that it is difficult to detect changes in scenes when these changes are made during saccades. Although this effect can be attributed to saccade-specific mechanisms, the blurring of the retinal image during the saccade also masks the transient motion signals that normally accompany an image change. Since transients play a large role in the drawing of attention, it may be that the failure to detect change is not due to saccade-specific mechanisms, but is simply due to a failure to correctly allocate attention.

To investigate this, an original image A was repeatedly alternated with a modified image A', with brief blank fields interposed. The resulting "flicker" created a global transient that swamped the local motion signals caused by the image change, preventing attention from being drawn to the location of the change. When this was done, a dramatic effect was found: even when the change was substantial and was made repeatedly, subjects had great difficulty seeing what it was. Changes were easily identified when a verbal cue was given, showing that visibility was not the limiting factor. Rather, it appears that this effect is due to a true failure to detect change, similar to that found with saccade-contingent display changes. Indeed, it is argued here that the same (purely attentional) mechanisms are the cause of "change blindness" under both these conditions.

Results also showed that the effect is robust to several kinds of variation in the parameters of the blank fields. The size of the effect is roughly the same for black, white, and gray blanks, and is also roughly the same for blank durations of 80 ms and 160 ms. Increasing the duration of the blanks to 320 ms caused a deterioration in the ability to detect change, possibly because of the decay of an underlying iconic memory. This suggests that scenes may be represented by a spatiotopic schematic map containing relatively unstructured elements, with attention selecting some subset of these elements and entering them into a more durable memory (possibly visual short-term memory) that allows comparisons to be made.

If this view is correct, the perception of change under flicker conditions acts as a "litmus test" that indicates when attention is being allocated to the relevant object or region. In keeping with this interpretation, it was found that identification was easy for changes in those objects mentioned in brief verbal descriptions of each scene, and much more difficult for objects that were never mentioned. Evidently, in the absence of low-level control, attention is attracted to various parts of a scene on the basis of high-level "interest". As such, the careful

mapping out of "attentional scans" may provide a useful new way to study scene perception.

3 From objects to scenes: Speculations on similarities and differences

Michael J. Tarr & Vlada Aginsky
Brown University
Michael_Tarr, vlada_aginsky@brown.edu
http://www.cog.brown.edu/brochure/people/mjt/tarr.html

The Theory. While the study of object perception has a long and storied history, the study of scene perception has been relatively neglected. In part this is because researchers have generally assumed that the perception of a scene is essentially a summation of the perception of its components. Recent results, however, indicate that this is not the case. In particular, observers are apparently unable to retain a complete representation of a scene across disruptions such as visual saccades or flicker. Not surprisingly, some parts of the scene are more salient in visual memory in that changes in such parts are not easily detected across disruptions. Remarkably, even the presence or absence of an entire object or part of an object may be misremembered. First, we consider the implications of such findings for how scenes may be represented and how these representations may differ from those used to represent objects. Second, we consider how observers integrate object information embedded within scenes into more complex representations, specifically asking what object features and relations contribute to scene perception? Our assumption is that both object and background features influence the initial representation of a scene, but that they are differentially encoded in terms of their specificity within the representation. Several experiments are reported that investigate this issue by examining the degree to which cueing particular features influences detection of changes within a scene. As summarized below, our results suggest that foreground objects are automatically integrated into the structure of scenes, but that background information is less salient within the representation, and as such, is subject to increased saliency through cueing. These results are interpreted in the context of a theory in which object, category, and scene representations are formed through statistical associations between local features and their relations.

The Experiments. We used the "flicker" paradigm introduced by Rensink, O'Regan, and Clark (1996) in which one element of a scene continuously alternates between its original appearance and a noticeably changed appearance. A brief blank field disrupts the scene at the moment that the change occurs and presumably delocalizes transients produced by the change which may otherwise draw the perceiver's attention. The crucial assumption of this technique is that visual properties that are more salient within the representation of the scene are preserved across such transients and, therefore, are easier to detect when changed.

Four different experiments were conducted. The first was a replication of the original flicker study, the second a control in which we measured the detection of changes without the presence of flicker, and the third and fourth investigated the impact of cueing the type of change (color, location, or presence) on the speed of detection. Our assumption was that cueing would differentially enhance detection of changes in features that were typically less salient in visual memory. Across these latter two experiments two different cueing manipulations were used. In the first scenes were blocked according to the type of change. In the second, scenes containing each type of change were randomly intermixed, but a cue informing participants as to the type of change was provided prior to each trial. Several results stand out. First, regardless of the type of change, changes were rapidly detected when there was no flicker. Second, we replicated Rensink et al.'s finding that changes in foreground objects are detected far more rapidly than are changes in background information. Third, cueing made no difference in the detection of changes in foreground objects. Fourth, cueing facilitated the detection of color changes in background information to a greater extent than it did location or presence changes.

The Conclusions. These results suggest that attention is initially directed towards elements of the scene that perceivers consider to be informative or interesting, e.g., stable foreground objects. Moreover, it is likely that such elements are represented with the greatest salience in visual memory. There is apparently less salience in the representation of background information, and based on the cueing advantage, even less salience in the representation of color information in the background. Planned follow-up experiments will use computer graphics psychophysics to create both synthetic familiar and nonsense scenes.

4 Transsaccadic object representations

Karl Verfaillie & Peter De Graef
University of Leuven
Karl.Verfaillie, Peter.DeGraef@psy.kuleuven.ac.be
http://www.psy.kuleuven.ac.be/ karl/objdb.html

In everyday scene exploration, the eye typically

saccades within or between objects which underlines the ecological importance of object processing but also indicates that integration across saccade-based discontinuities in the proximal stimulus is a pervasive and multi-faceted challenge for object perception. In a series of experiments, we investigate whether and how object position, orientation, and identity are represented across saccades. In the experiments on the representation of position and orientation, we measure the detectability of saccade-contingent changes in a point-light walker. The representation of object identity is studied in a search task that requires exploration of the visual scene in order to perform a series of object decisions. These issues are examined as a function of the object's transsaccadic status (target, source, or bystander) and as a function of the integration interval (within one fixation-saccade cycle, across multiple cycles). Implications for theories of stored visual knowledge are discussed.

Image-plane position and in-depth orientation

Karl Verfaillie

Using a transsaccadic integration paradigm, in which participants had to detect saccade-contingent changes in a moving point-light walker, Verfaillie et al. (1994) observed that the global image-plane position of the walker was not maintained accurately across saccades, whereas saccade-contingent changes in the walker's in-depth orientation were readily noticed. This suggests that transsaccadic object representations are position invariant but orientation dependent. We will give an overview of three follow-up studies.

First, as far as the object's position is concerned, we examine whether and under what conditions transsaccadic memory for the object's position improves when viewers can code the point-light walker's position in an allocentric rather than an egocentric reference frame, i.e., when the object's position can be coded relative to neighboring landmark objects rather than relative to the viewer only.

Second, as far as the object's orientation is concerned, we explore two alternative explanations for the conclusion that the object's in-depth orientation can be carried across saccades. First, by manipulating the presaccadic orientation, we show that our earlier finding of accurate detection of saccade-contingent rotations was not due to the fact that the presaccadic orientation was always a (canonical) 3/4 view. Second, when the walker is rotated in depth, the relative positions of point-lights in the image change. Therefore, maybe what is maintained across saccades is the relative position of lights rather than a walker in a particular orientation. An experiment with inverted walkers provides evidence against this hypothesis.

The third series of experiments generalizes the earlier basic findings from the case of saccades within the same object to the case of saccades between different objects. Whereas in Verfaillie et al.'s (1994) experiments, subjects made a saccade within the same point-light walker, observers now saccade from one point-light walker to another, and we probe transsaccadic memory for position and orientation of saccade source and saccade target.

Absolute and episodic object identity

Peter De Graef

When objects are encountered in the context of real-world scenes, strictly data-driven accounts of object recognition appear to be incomplete. Relative size, position, and semantic plausibility of the object in its context have been claimed to affect the speed with which the object can be recognized. We have opted to study these context effects in a search task that requires exploration of the visual scene in order to perform a series of object decisions (De Graef et al., 1990).

Under these conditions, context effects proved to be a dynamic phenomenon based on various mechanisms each with their own spatial and temporal restrictions. Based on published and unpublished data from our own work we will distinguish between a) immediate context effects, operational during the first glance at a scene, b) delayed context effects, which only develop as scene exploration progresses, and c) inter-object priming effects, which can occur throughout the course of scene exploration.

While immediate effects are largely intra-fixational, delayed and priming effects involve a transsaccadic component: Information needs to be integrated over at least one and often multiple fixation-saccade cycles. To document this, we will discuss four experiments. In the first two, we looked at the impact of prime and background information on the peripheral and foveal processing of a target object. In both experiments, we elicited prime-target fixation sequences and orthogonally manipulated prime-target and target- background relatedness. In addition, prime-target distance and the length of the interval between prime and target fixations were taken into account to delineate spatial and temporal properties of the observed context effects. In a third experiment, we used intrasaccadic changes of a target object to determine whether semantic object information was integrated transsaccadically and to what extent this was modulated by the object's context. Finally, in a fourth experiment we are looking at the position-specificity of transsaccadic integration of object in-

formation by using intrasaccadic changes of object position while controlling for attentional strategies that may have compromised earlier studies of position-specificity. While this study has been carried out with objects out-of context, we hope to be able to also report some data on an in-context version of the experiment.

5 The influence of scene context on object perception

John M. Henderson & Andrew Hollingworth
Michigan State University
john@eyelab.psy.msu.edu
Phillip A. Weeks, Jr.
AT&T Bell Laboratories

How is the encoding of perceptual information concerning an object affected by the scene within which that object appears? Two general classes of theories concerning the influence of scene context on object perception will be considered, both of which are capable of accounting for the current evidence. Data from an eyetracking experiment and several tachistoscopic presentation paradigms will be presented in an attempt to offer new data with which to adjudicate between these classes of theories.

In order to illustrate the distinction between the alternative theoretical views, the following view of object identification will be adopted: It will be assumed that object identification requires the generation of object primitives (e.g., surface properties, edges), the generation of an episodic description of the object (e.g., a structural description of some type) from those primitives, and the matching of the description to long-term memory representations. This general outline is consistent with a number of current computational theories of object identification, despite large differences in the specifics of those theories. The first general class of theory instantiates the assumption that the episodic description generated for an object is modulated by the consistency of that object with the currently active scene representation, such that construction of the description is facilitated for consistent objects and/or inhibited for inconsistent objects. This type of signal modulation hypothesis can easily be conceptualized in terms of a connectionist architecture similar to the McClelland and Rumelhart interactive activation model of word recognition, with scenes corresponding to words, objects corresponding to letters, and primitives corresponding to features. While the representation of the spatial relationships between objects and scenes may be more complex than that between letters and words, the overall analogy is relatively straightforward.

A second class of theory is one in which an activated scene representation exerts its influence not by directly modulating the generation of a perceptual description, but instead by affecting processes that come later in the processing sequence. For example, the consistency of an identified object and the activated scene representation might affect a memory consolidation process of the sort investigated by Potter and Intraub. A second possibility that will be pursued in this talk is that the activated scene representation affects the goodness-of-fit criterion that is used to determine whether a match exists between an episodic object description and a stored object description. In this latter criterion modulation hypotheses, an active scene representation does not directly influence the generation of the episodic description for the objects in a scene. Instead, an active scene representation affects object processing by modifying the goodness-of-fit criterion: Consistent objects require less perceptual evidence for an entry-level match and so will be detected faster. The criterion modulation hypothesis predicts an indirect effect on the construction of episodic descriptions, but in the opposite direction from that predicted by an interactive activation model: To the extent that a match between the constructed description and the stored description takes longer to achieve when an object is inconsistent with the scene than when it is consistent (due to a lowered criterion in the latter case), more time should be taken in constructing the object description in the former case, and so more detailed perceptual representations should be formed.

In order to investigate these competing theoretical perspectives, we have conducted a series of experiments using line drawings of complex real-world scenes as stimuli. Scenes were paired so that an object that was consistent in each scene could be swapped across scenes, creating inconsistent context conditions. In an initial eyetracking study, inconsistent objects were fixated more often and for more time than were consistent objects, replicating past research. However, several other findings that have been reported in the literature did not replicate, including a failure to find earlier fixation on inconsistent objects.

In a second series of experiments, we sought to develop a paradigm in which it was possible to investigate the nature of the perceptual description that is formed for an object as a function of the consistency of the object with the scene in which it appears, and further that would minimize as much as possible the influences of the activation of object categories, the generation of object names, post-perceptual guessing, and task-specific response strategies. The paradigm we developed is a simple same-different task. In this

paradigm, the participant is presented with a picture of a real-world scene for some controlled period of time, followed by a pattern mask, followed by a re-presentation of the original scene. The participant's task is to determine whether any of the objects has changed across the two presentations of the scene. In this task, a signal modulation perspective predicts facilitated detection of a change to an object when that object is consistent with the scene context, while a criterion modulation perspective predicts facilitated detection when the object is inconsistent with the scene. Second, we also used two types of object manipulations so that we could examine the influence of scene context on the episodic encoding of two types of perceptual information. In the object deletion condition, we removed an object across scene presentations to examine the effect of scene context on the encoding of information about object presence. The decision concerning object deletion can presumably be made either at the level of the episodic description or the activated entry-level concept. In the left-right orientation reversal, we changed an object's left-right orientation across scene presentations in order to examine the effect of scene context on the encoding of information about object orientation. Left-right orientation is thought not to be encoded as part of an entry-level concept, and so should more directly reflect the quality of the perceptual description constructed.

Finally, we have used the flicker paradigm developed by Rensink, O'Regan, and Clark to examine object encoding. In this paradigm the display is alternated between presentation of a scene and of a blank or masked interval. Each alternation of the scene is either the same or is changed in some way. We displayed our scene stimuli in this paradigm with either a 250 or 500 ms scene presentation and an 80 ms intervening mask duration, with changes of either deletion/addition or orientation. The main finding across both the change detection and flicker paradigms is that changes are better detected when an object is inconsistent rather than consistent with the scene in which it appears. These results support the criterion modulation class of theory.

6 Object blanking reveals properties of transsaccadic memory

Werner X. Schneider & Heiner Deubel
Ludwig-Maximilians-University, Munich
wxs@mip.paed.uni-muenchen.de, kdeub@mpipf-muenchen.mpg.de

Changing the location of an object (e.g. the saccade target) during a saccade can hardly be seen when the change is less than 20%of the saccade size. This result seemed to indicate that transsaccadic memory of stimulus location is relatively inaccurate. We have recently demonstrated, however, that such an intrasaccadic displacement of an object is reported with high accuracy when the object is momentarily blanked after the saccade (Deubel, Schneider, Bridgeman, 1996, Vision Research). This means that blanking an object after the saccade makes precise transsaccadic location information available. In a number of experiments, further properties of this postsaccadic gap effect were investigated.

In a first series, a saccade target was presented together with a second visual object (a "distractor"). One of both objects was displaced during the saccade, and one of both then reappeared only after a temporal blanking. Subjects consistently perceived the blanked object as jumping and the later appearing object as stationary, indicating that transsaccadic location correspondence seems to be computed on the basis of that object that appears immediately after the saccade.

In a second experimental series, subjects had to judge the spatial position of the presaccadic distractor with respect to a postsaccadic indicator. The data show that the judgement is largely determined by postsaccadic target position: when the target is displaced, the presaccadic distractor is perceived in a displaced position. We conclude from these findings that objects available immediately after the saccade serve as a reference for transsaccadic spatial correspondence.

Finally, we tested whether the blanking manipulation also improves the perception of intrasaccadic changes such as size, luminance, orientation, color, and shape. The task required again subjects to saccade to a peripheral target. Triggered by the saccade, one attribute of the target was changed (e.g., the size of the target) and subjects had to report this intrasaccadic change. The results show that for "dorsal" attributes (size, orientation), i.e. attributes used for spatial-motor actions like grasping, the gap manipulation improved the perception of the intrasaccadic changes. Ventral attributes (color, form), i.e. attributes used by the object recognition system, did not profit from the postsaccadic gap.

The findings will be discussed in relationship to the questions of what kind of temporary object representations are implemented across the saccade in transsaccadic memory and how they are updated.

7 A computational perspective on scene understanding

Alan L. Yuille
Smith-Kettlewell Institute
yuille@skivs.ski.org http://www.ski.org/ALYuille_lab/

Scene understanding poses formidable representational and computational problems. Current systems only work well in restricted domains or when used to give quick, crude scene classification as part of interactive systems for database retrieval. This talk speculates on how our current work on object recognition might be generalized to scene understanding using approaches based on twenty questions.

8 Flexible scene categorizations in a scale space

Philippe G. Schyns & Aude Oliva
University of Glasgow
philippe, aude@psy.gla.ac.uk
http://tornade.ere.umontreal.ca/ hebertpa/prof/schynsp.htm

Efficient categorizations of complex visual stimuli require effective encodings of their distinctive properties. In the object recognition literature, scene categorization is often pictured as the ultimate result of a progressive reconstruction of the input scene from simple local measurements. Boundary edges, surface markers and other low-level visual cues are serially integrated into successive layers of representations of increasing complexity, the last of which derives the identity of a scene from the identity of a few objects. For example, in Figure 1, combinations of fine-grained edge descriptors and other local cues suggest the presence of cars, road panels, highway lamps and other objects which typically compose a highway scene. Precise scene categorization often requires that the identification of component objects from such fine-grained measurements precedes the identification of the scene.

However, there is data challenging this exclusive "object-before-scene" recognition. Complex visual displays composed of many partially hidden objects are often recognized quickly, in a single glance–in fact, as fast as a single component object (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982; Potter, 1976; Schyns & Oliva, 1994). This suggests that categorization processes could sometimes directly extract global representations of the input scene; representations allowing "express," but comparatively less precise classifications of the input. To illustrate the different routes to scene categorization, squint or blink while looking at Figure 1, another scene should appear (if this demonstration does not work, step back from the picture until you perceive a city).

Figure 1: Two scenes presented simultaneously at two scales.

Figure 1 simultaneously presents visual cognition with two scenes, each associated with a different spatial scale (Schyns & Oliva, 1994). Although it is possible to identify the background city scene from the spatial layout of its major "blobby" components, it is virtually impossible to reliably identify each isolated blob as a building (a single blob can potentially correspond to many objects). This illustrates that coarse scene properties, or a scene-before-object strategy, could provide an alternate route, an express-way to recognition (see Henderson, 1992). In sum, Figure 1 illustrates that the information associated to the two scales composing the picture can elicit two independent classifications.

We will discuss how processes of scene categorization use the information associated with different perceptual spatial scales. The psychophysics of scale perception would suggest that recognition should use coarse, blobby information before fine scale edges because Low spatial Frequencies (LF) are perceptually available before High spatial Frequencies (HF). Although possible, this interpretation does not take into account the nature of the task the recognition system must solve. If different spatial scales transmit different information about the input, an identical scene might be preferentially encoded at the scale optimizing information for the considered task. For example, while precise categorizations could progressively reconstruct the input from local fine-grained measurements (e.g., boundary edges), express routes could encode the same stimulus at a cruder resolution; a resolution highlighting the global scene structure. Our experiments will show how task constraints (the requirement of locating diagnostic information) can affect the encoding of scenes in a space of spatial scales.

9 The art and science of scene perception

Daniel J. Simons & Daniel T. Levin
Cornell University
djs5, dtl1@cornell.edu
http://www.psych.cornell.edu/simons/simons.html

Representing spatial relations in static displays and motion pictures

Daniel J. Simons

Over the past 5 years, research from a number of laboratories has shown that people fail to notice striking changes to objects and scenes when localized retinal information signaling a change is masked or eliminated (e.g., by eye movements or flashed blank screen). Without immediate sensory information signaling a change, the visual system must rely on representations of the scene to detect changes. Thus, greater facility at detecting particular kinds of changes to objects might indicate a more precise initial representation of that aspect of a scene. Two sorts of changes seem to be more readily detected than others: (a) changes to objects that attract attention during initial processing (i.e., changes to objects that are the center of interest), and (b) changes to the relative spatial positions or layout of objects. In this presentation, I will focus primarily on recent work with both infants and adults addressing the perception and representation of layout information. I will consider several questions concerning the nature, precision, and flexibility of layout representations. For example, how substantial a layout change is needed for detection, and to what extent are layout representations viewpoint dependent? Most of the research I present will focus on the ability to detect changes to static images that are separated by a blank ISI. I will also introduce recent work on the ability to detect changes across cuts in motion pictures.

Implicit and explicit perceptual theory in the development of motion pictures

Daniel T. Levin

Although psychologists have only recently begun to explore the process of binding different views of a scene, film makers have been using a variety of implicit theories about scene perception for at least 80 years. A key task for early film makers was to create visual narratives that were correctly apprehended by anyone with a nickel and a spare hour. Their audience often had minimal experience viewing motion pictures, and spanned most of the world's cultures. Early film makers were, therefore, required to create motion pictures that accurately tapped the core of human perception. I discuss the work of these artists and relate it to current work on scene perception using a variety of demonstrations. Dan Simons and I have also completed a number of experiments that verify and extend these artistic intuitions. In concordance with film makers' intuitions, we find that it is possible to make fairly dramatic changes across different views of a scene which participants fail to notice. These changes can extend to changing the actor present in a scene. I will go on to discuss possible constraints on the kind of change that will and will not escape notice, and relate these constraints both to the practice of continuity editing and to current research in developmental psychology.

10 Scene perception in object recognition and view-based navigation

Isabelle Bülthoff, Sabine Gillner & Guy Wallis
Max-Planck Institute for Biological Cybernetics
isa, binni, guy@mpik-tueb.mpg.de
http://www.mpik-tueb.mpg.de/people/personal/person.html

Most of the work done on object recognition, be it psychophysical or neural network based, has been concerned with the viewing of objects in isolation, phantom faces or conglomerations of simple geometries floating in space. This approach has been remarkably fruitful but can only bring us part of the way towards learning how we recognize real objects as they appear in real scenes. Due to problems of repeatability and control, conducting experiments in the real world is often impracticable. Various members of staff at the Max-Planck Institute have been seeking to take advantage of advanced computer graphics which can provide much of the detailed and compelling appearance of a natural environment whilst continuing to offer control of the experimental environment. Our interests have concerned stereoscopic depth perception, view based navigation, attention and scene based object analysis. Presentations of the current work in all of these areas will be made, including a tour of the laboratories where the work is currently underway.

Top-down influence of recognition on stereoscopic depth perception

Isabelle Bülthoff, Pawan Sinha & Heinrich Bülthoff

We have previously demonstrated that the recognition of biological motion sequences is consistent with a view-based recognition framework. We found that anomalies in the depth structure of 3D objects had an intriguing lack of influence on subject ratings of its figural goodness.

Our most recent results indicate the existence of top-down object-specific influences that suppress the perception of deviations from the expected 3D structure in a motion sequence. The absence of such an influence for novel structures (non-biological random structures) indicate that high-level expectations about an object's 3D structure can strongly influence even the relatively early processes involved in stereoscopic depth perception.

Navigation experiments in virtual environments: HexTown

Sabine Gillner and Hanspeter Mallot

Last year we presented a study about construction of neuronal representations from sequences of views and movement decisions in a hexagonal maze. (Dartsch & Mallot 1995). We have build up a simple driving simulator in which the participants sit in front of a computer screen and can virtually turn around or move from one place to another by pressing the appropriate button of a computer mouse.

During exploration, subjects drive off from a fixed starting point ("home") and are then asked to find their way back to this home from different starting points in the town, or to find other starting points.

We have analyzed the representation in terms of distance estimation: If subjects had to judge distances between objects in the town, they tended to answer in terms of path length in all conditions despite the fact that the exploration behavior was rather different between the conditions. In addition we have analyzed the exploration sequence in terms of association between movement decision and local view and we have found that subjects who make a larger number of errors have a tendency (up to 65 when encountering a given view for the next time. This simple association (view movement) is less frequent in good navigators.

Scene perception in real and virtual environments

Guy Wallis

The process by which we recognize and analyze scenes, composed of the fauna, flora or man-made objects of everyday life, remains largely mysterious. What evidence we do have, suggests that the instantaneous, full and detailed perception of a scene which we experience, is simply illusory and that detailed analysis of objects can only be achieved in a more piecewise, serial manner (Rensink *et al.* 1996, Currie *et al.* 1996, O'Regan *et al.* 1996, Blackmore 1995). As Rensink and Blackmore have shown in their work, astonishingly large changes can be made to the composition of individual static scenes without them being immediately obvious to the passive observer - so long as the motion associated with these changes is masked in some way, such as by a grey blank interval or motion of the entire image between changes. More than simply telling us that the detection of motion is highly influential in the analysis of scenes, it tells us that object attributes, such as colour, location, orientation etc. may initially only be encoded at a very coarse level, if at all. This fact raises interesting questions about what attributes are encoded in more detail and under what circumstances such that the almost instantaneous recognition of the overall scene can proceed despite the actual vagueness with which other attributes are encoded[1]. One manner in which we might choose to develop the ideas discussed here are by extending them to dynamic scenes in which the observer moves. This form of insensitivity to scene changes does transfer to dynamic environments, as shown in informal experiments at the Max-Planck Institute. However, the role of the observer, be he active or passive, will almost certainly also be of significance, since it will affect which elements of the scene require detailed processing. A passenger in a car may prefer to observe a house at the roadside whilst the driver will be watching for road signs or car movements for example. In the case of the passive observer real video footage can be shot and edited to create the scene changes as described in experiment I below. In the case of an active observer, however, the only practicable solution is to generate the entire world artificially so that it can be dynamically controlled, allowing the observer to interact with the world, as described in experiments II and III.

Expriment I: In order to test the transferral of insensitivity to changes within a scene from static to dynamic scenes, a series of videos were shot driving along the same stretch of road in which a number of objects were visible. In each video, two objects were altered in some way from the standard configuration. Either the orientation, colour or location of the object was changed, or it was simply removed. The film was then edited together so that short frame sequences from the standard and new configurations appeared smoothly interleaved as the observer drives towards, and ultimately past the test objects. The subjects task was to say what was changing. Scene footage was switched every 10 frames, and following the example of Rensink, motion cues were masked by translating the scene each 10 frames randomly in the plane of the monitor.

Experiment II: A new series of experiments are planned in which the observer may view the scene

[1]The impression of a scene's identity is itself influential in the speed and accuracy with which objects within the scene are recognized, as various researchers have described (Biedermann *et al.* 1982, De Graef 1996)

passively or actively (as the passenger or driver of a car). In order to see how the results from the real world transfer to the virtual environment, a copy of the scene displayed in the video footage was replicate in the virtual environment and subjects allowed to view the scenes passively as before.

Experiment III: Given that results are comparable in the two environments further experiments are planned using a virtual environment in which the observer drives along a road in traffic. This work is being conducted in conjunction with Daimler Benz in Germany and is intended to ascertain which qualities of a scene a driver does attend to.

11 Virtual reality and psychophysical experiments

Hartwig Distler and Hendrik-Jan van Veen
Max-Planck Institute for Biological Cybernetics
mad, veen@mpik-tueb.mpg.de
http://www.mpik-tueb.mpg.de/demopage.html
http://www.mpik-tueb.mpg.de/people/personal/mad/mad.html

One of the reasons for the relatively small number of experiments investigating scene recognition in the past has been the technical problems when setting up experiments. Displaying pictures in the form of videos for example, prevents the presentation of dynamical information and the participants cannot interact with the scene.

Virtual environments on the other hand, offer promise to overcome these problems since they allow the experimenter to control and manipulate the presented scene. Additionally, participants can interact with the scene in realtime. Workshop attendees will have the opportunity to see our simulation environment which includes a virtual bicycle and car steering wheel interface.

Contrasting virtual environments with real environments

Since the quality of the visual display of simulation environment is still poor, the suitability of virtual environments for conducting experiments investigating scene perception has to be investigated. I have already been addressed the question of perceived velocity in virtual environments, which is affected by the spatial frequency content of the image - something which may vary widely in virtual environments.

The issue will be discussed by means of a simulation environment which includes a virtual bicycle that we have setup at the Institute in Tübingen.

Navigation experiments in virtual environments: Virtual Tübingen

Hendrik-Jan van Veen

As an extension to the experiments already conducted on navigation we are also intending to build a virtual model of the local town of Tübingen. Part of the motivation for doing this is that the model should enable us to test the transferral of learning to navigate in a real environment to a virtual one and visa versa. However, it should also enable us to ascertain which elements of the scene play a role in navigation of the winding back streets of the town, since by manipulating elements such as the appearance or location of buildings present in the virtual model, it should become clear which the key cues for navigation are.

General References

Biederman, I. 1987. Recognition by components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

Biedermann, I., Mezzanotte, R.J., & Rabinowitz, J.C. 1982. Scene perception: Detecting and judging objects undergoing relational violations.*Cognitive Psychology*, **14**, 143–177.

Blackmore, S.J., Brelstaff G., Nelson K. & Troscianko T.1995. Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, **24**, 1075–1081.

Bülthoff, H. H., Edelman, S. & Tarr, M. 1995. How are three-dimensional objects represented in the brain?, *Cerebral Cortex*, **5**, 247-260.

Bülthoff, I., Sinha, P., & Bülthoff, H. H. 1996. Top-down influence of recognition on stereoscopic depth perception, *In: Investigative Opthalmology and Visual Science*, **37**, S1125.

Currie, C., McConkie, G.W., Carlson-Radvansky, L.A., & Irwin, D.E. 1996. Maintaining visual stability across saccades: Role of the saccade target object. *Journal of Experimental Psychology: Human Performance and Psychophysics*, *In press.*

Dartsch & Mallot, *Proceedings of the 23rd Göttingen Neurobiology Conference*, **I**, 47.

De Graef, P. 1996. *Speeded object verification in real-world scenes: perceptual, decisional and attentional components.* Submitted for review.

De Graef, P., Christiaens, D., & d'Ydewalle, G. 1990. Perceptual effects of scene context on object identification. *Psychological Research*, **52**, 317-329.

Deubel, H., Schneider, W.X., & Bridgeman, B. 1996. Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research*, **36**, 985-996.

Johansson, G. 1973. Visual perception of biological motion and a model for its analysis *Perception & Psychophysics*, **14**, No. 2, 201-211.

Logothetis, N. K., Pauls, J., Bülthoff, H. H. & Poggio, T. 1994. View-dependent object recognition by monkeys, *Current Biology*, **4** 401-414.

O'Regan, J.K., Rensink, R.A., & Clark, J.J. 1996. "Mud splashes" render picture changes invisible. *Page 213 of: Investigative Opthalmology and Visual Science*. Fort Lauderdale, Florida: Lippincott-Raven: Hagerstown, MD, USA.

Poggio, T. & Edelman, S. 1990. A network that learns to recognize three-dimensional objects. Nature, 343, 263-266.

Rensink, R., O'Regan, J. K., & Clark, J. J. 1996. Visual perception of scene changes disrupted by delocalized transients. *Submitted.*

Rensink, R.A., J.K., O'Regan, & Clark, J.J. 1996. To see or not to see: The need for attention to perceive changes in scenes. *Page 213 of: Investigative Opthalmology and Visual Science*. Fort Lauderdale, Florida: Lippincott-Raven: Hagerstown, MD, USA.

Sinha, P., Bülthoff, H. H. & Bülthoff, I. 1995. View-based representations for biological motion sequences, *In: Investigative Opthalmology and Visual Science*, **36**, S417. ftp://ftp.mpik-tueb.mpg.de/pub/papers/hhb/SinhaBulBul-95-abs.ps.Z

Verfaillie, K., De Troy, A., & Van Rensbergen, J. 1994. Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 649-670.