Interactions between self-motion and depth perception in the processing of optic flow

V. Cornilleau-Pérès and C.C.A.M. Gielen

Moving and acting in a 3D environment requires the perception of its 3D structure. Vision is known to play a crucial role in the control of self-motion, particularly through the changes in the retinal image subsequent to movements of the observer. Reciprocally, signals related to self-motion can also influence our visual perception of 3D space. These interactions between 3D visual perception and self-motion, as demonstrated behaviourally, are now better understood thanks to the development of computational models for processing moving images. They also bear a particular interest in the context of the recent intensive exploration of the inferior parietal lobe (IPL) by neurophysiologists. The IPL is now firmly established as one site of interaction between 3D visual perception and motor control. The parallel between behaviour and neurophysiology leads to a set of crucial, yet unanswered, questions.


GIBSON SET the basis for modern studies of visual perception by describing the visual input as an optic flow, rather than a succession of static images. When an animal moves, optic flow carries information related to the motion of the animal, as well as to the 3D layout of the environment and the movement of objects. Generally, these two functional roles (proprioceptive and exteroceptive) have been approached separately in the past. However, several studies, most of which are quite recent, have explored the complexity and neural substrates of the interactions between self-motion and 3D shape.

Acknowledgements
The author thanks his colleagues in the laboratory for helpful discussions and critical reading of this manuscript. The work reported here was funded by grants from the CNRS, INSERM, AFM, RMHE, International Human Frontier Science Program and Rhône-Poulenc Rorer.
The 3D structure of the environment and the movement of the observer interact in the production of optic flow. For a translation of the 3D image by the observer, the velocity of the retinal image varies with the distance of the object points (Fig. 1). This variation is termed motion parallax. In computer vision, recovering depth information from monocular motion parallax has proved to be a difficult task\textsuperscript{1,3}, whereas it is readily achieved by human observers in many situations\textsuperscript{1,5}. Motion parallax is therefore a depth cue for the human visual system and, in this sense, presents similarities with binocular disparity. It also presents different properties that are worth mentioning. First, it provides depth information only up to a certain scale factor (a distant object that is moving quickly can yield the same optic flow as a nearby object that is moving slowly). Second, rotations of an object around an axis that passes through the eye displaces the retinal image globally, without providing any depth information (similarly, eye rotations provide no depth information). Finally, when the relative movement between the observer and the visual scene is not restricted to such a rotation, the 3D structure of 3D objects can, in principle, be extracted from optic flow\textsuperscript{2}.

Influence of 3D structure on self-motion

Studies of both the perception and the control of self-motion reflect the theoretical interplay between depth and motion. The perception of self-motion

Self-motion perception depends not only on the motion of the observer within a visual scene, but also on the 3D structure of this scene. This is shown in experiments onvection (the sensation of self-motion caused by visual motion) and heading (the direction of self-motion). It can be illusory, as experienced in a stationary train, when the train alongside starts moving away. Vection depends on various parameters such as the velocity and area of visual stimulation, and also the perceived depth within the visual scene. For example, if several patches within the visual field have different movements, the patch which is apparently the most distant is perceived as stationary in space and determines vection\textsuperscript{6,8}.

In 'heading' experiments the movement of a subject in a visual environment is simulated on a display screen. The subject has to report the direction of self-motion (or 'heading') within this environment. High accuracies (about one degree), are usually achieved in such a task, whereas it is a difficult task\textsuperscript{5,3} to achieve high accuracies in visual field experiments. That absolute distances can also be perceived as threshold effects; for large distances, larger postural sways are required to provide the minimum visual motion necessary to detect a change in the position of the observer (relative to his environment), and drive a postural readjustment. Recently, Schöner\textsuperscript{20} proposed a more elaborated model, taking into account the dynamics of the postural-control system, and the coupling between visual motion and head movements.

As a consequence of this coupling, vision induces postural responses with amplitudes that decrease as D increases. The model then predicts that the stability of postural responses also decreases, in the sense of a larger variability in its phase delay relative to the visual stimulation. This prediction was tested by exposing observers to planar surfaces oscillating along a sagittal axis, which showed that the temporal locking of postural sway and visual expansion effectively improves as 3D decreases\textsuperscript{18,19}. Hence, it is not only the visual expansion that drives the postural response, but also the temporal coupling between self-motion and visual signals. That absolute distances can also be perceived from this coupling is suggested by perceptual studies (see below), although the demonstration that has been performed is for lateral, rather than antero-posterior, movements in humans.

A second example concerns convergence eye movements. As a target moves along a sagittal axis, the accuracy of convergence eye movements is highly improved by the presence of a visual scene surrounding
motion parallax provides only information about relative depth, and has to be combined with other sensory signals to yield absolute depth perception. Self-motion, in this sense, abolishes the ambiguity surrounding motion parallax, as proprioceptive signals relative to the motion velocity can be used to scale the motion parallax. Thus, animals such as the locust or the gerbil produce spontaneous head movements to judge absolute distances. Similarly, humans are able to report absolute distance from active head movements (see also J.N. Park, PhD thesis, p. 69, University of Kansas, 1964). However, when a target of limited size is presented alone rather than in a structured environment, errors in the estimation of its distance can be very large. The head movements are then accompanied by apparent 3D movements of the object.

Using random patterns of dots displayed on an oscilloscope screen, Rogers and Graham compared the perceived absolute distance within a surface (that is, the absolute distance in depth separating two surface points) during comparable head and object movements (Fig. 2). The relative motion between dots could be yoked to the head movements, simulating the presence of a stationary surface corrugated in depth. In another condition, the dot movements simulated a corrugated surface moving in front of the stationary subject. Perceived depth was larger, and in better agreement with the actual depth during head movements than during object movements, suggesting a co-operation between self-motion signals and optic flow for the perception of absolute 3D distances between object points.

Similarly, the sign of relative depth is ambiguously defined by motion parallax for objects covering a small viewing angle (for example, <15°), and apparent reversals of depth order frequently occur when a passive observer views corrugated surfaces in motion. Here again, during self-motion, proprioceptive information co-operates with optic flow in the sense that the concavity or convexity of the surface is always unambiguously perceived in this case.

The notion that the detection of relative depth from optic flow could also be improved during self-motion, as compared with object motion, is suggested by a study on reports of surface slant (J.N. Park, op. cit.), and by the theoretical approach to the 3D analysis of visual motion. We tested this hypothesis by measuring the ability to discriminate between spherical and planar surfaces under three conditions (Fig. 3): (1) during head translations; (2) during object translations; and (3) during object rotations around a frontoparallel axis. Across these conditions, the apparent relative movement between the image points (that is, motion parallax) was strictly identical. Hence the potential visual information about surface structure remained unchanged. The factors that varied across the conditions were the origin of the movement (self-motion or object motion), and the global retinal motion, which is inversely correlated with the quality of the retinal-image stabilization (RIS). Retinal-image stabilization was best achieved during object rotation, since the central image point was stationary on the screen, and had an averaged image velocity that was roughly zero. During object translation, the stimulus reached fast velocities, and pursuit eye movements failed to stabilize the image motion accurately. During head translations, however, oculomotor reflexes of vestibular origin are known to improve the

Influence of self-motion on perceived depth within a visual scene

Reciprocally, self-motion is a crucial source of information for depth perception. For example, monocular
gain of pursuit eye movements. Indeed, whether measured with computer-generated surfaces, or with real objects, the subjects' performance co-varied with the quality of the RIS, being always optimal for object rotation, intermediate for head translation, and worst for object translation. Hence RIS is probably a prominent factor influencing the perception of depth from motion, whereas the movement source (object or self-motion) is only secondary in our experimental situation. This conclusion is supported by results of Nakayama showing that small amounts of retinal-image motion can severely affect the visual detection of velocity gradients.

The RIS interpretation, however, was challenged in a complementary experiment where small-field stimuli (with a diameter of 8° of the visual angle) were replaced by large-field stimuli (with a diameter of 90° of the visual angle). In the case using large-field stimuli we failed to find any difference of performance in the three conditions above. An interpretation based on retinal motion might still hold, however, for two reasons. First the gain of smooth pursuit is known to increase with the size of the stimulus; thus, for object translation, the RIS improves as the field size increases. Second, whenever there exists a small phase lag (a few degrees) between eye and object position, foveal vision is not always effective for small-field stimuli, whereas large-field stimuli keep stimulating the foveal and parafoveal retina. Future investigations will clarify whether these factors fully explain our results. An alternative explanation might be that the processing of optic flow for large-field stimuli (usually involved during self-motion) could be mediated by different processes, as compared to the motion of small-field stimuli, which is usually related to object movements.

To summarize, a functional complementarity between optic-flow and self-motion signals exists, in the sense that these signals can specify information that is ambiguously defined by motion parallax (such as absolute depth and depth order). However, the processing of relative depth from optic flow does not seem to be improved by self-motion, but rather appears to depend on the stabilization of the retinal image, at least in conditions using small-field stimuli. Under conditions using large-field stimuli, the stability of visual performance under the different test conditions questions the existence of individual processes dedicated to the analysis of optic flow with large-field stimuli.

Optic-flow analysis by visual cortical neurones

Within the cortical visual system, one of probably two major pathways progresses dorsally from V1 to the inferior parietal lobe (IPL). In this dorsal pathway, visual information flows mainly from V1 to area MT (directly or through other prestripate visual areas), then to the IPL, where areas MST and 7a seem to play a prominent role in the processing of visual information about motion and spatial location. By contrast, the ventral pathway that originates from V1 to reach the inferotemporal cortex (IT) is principally involved in processing visual information involving colour, form and pattern (Fig. 4).

According to Gibson's view, 3D shape is not recovered from motion through an inverse 3D computation, but through the coding of spatio-temporal transformations related to surface layout and relative movement between the observer and his environment. Computational studies have proposed two types of such transformations corresponding to the first and second spatial derivatives of the retinal velocity. Indeed, 3D parameters of structure and motion, such as surface orientation and curvature, can be derived from the extraction of these image transformations. Hence, many authors attempted to demonstrate a specific sensitivity to these derivatives among neurones of the dorsal pathway. Particular attention was paid to the sensitivity of neurones in areas MT and MST to first-order derivatives of optic flow, such as expansion or rotation. Although neurones in area MT respond vigorously to such motion patterns, they do not present a specific

---

Fig. 3. Comparison of the perception of structure from self-motion and object motion. The surface presented to the subject consists of a set of dots which are displayed on a computer screen. (A) In the condition 'self-motion', the subject actively translates his head in front of the display screen. His head translations are measured and transmitted to the computer. The dot motion simulates a stationary surface that can be either spherical or planar. (B) In the condition 'object translation', the head translations recorded in the condition 'self-motion' are used to translate rigidly the simulated surface in front of the stationary subject. (C) In the condition 'object rotation', the recorded head translations are combined with rotations around the subject's eye. The resulting surface movement is a rotation of the surface around one of its tangent axes which lies in a frontoparallel plane. The relative motion of the dots on the screen is exactly as in B, but their overall motion on the display screen is minimal.

---

V. Cornilleau-Pérès and C. Gielen – Processing of optic flow
sensitivity to them. By contrast, a large percentage of
neurones in area MST respond selectively to one or
several first-order changes of the retinal image. These neurones are mostly located in the dorsal part
of area MST, which is also called area MSTd.

Neurones in area MT present another type of sensi-
tivity to spatial variations of image velocity, because their responses to stimulation within the receptive field can be strongly modified by a movement outside this receptive field. This centre-surround antagonism can be described as a tuning to second-order variations of image velocity across the visual field. However, it remains to be explored whether this tuning requires the presence of velocity discontinuities (as has been proposed up to now) or whether it could also detect smooth second-order variations of velocity across the border of the receptive field.

Finally, this sensitivity of areas MT and MST to higher-order characteristics of visual motion is also combined with a sensitivity to binocular disparity, sometimes in a complex way. For example, in area MT, the interactions between patterns moving in coincident directions (one in the preferred direction, the other in the null direction of the cell) depend on their disparity; the neuronal response is inhibited by the second pattern only if the two motion signals have similar disparities. Since area MT belongs to the cortical pathway that feeds into subcortical structures involved in the OKN generation, such a mechanism could be responsible for the sensitivity of the OKN to binocular disparity described by Howard and his colleagues.

Sensorimotor interactions and optic-flow processing in the brain

The IPL is not only involved in visuospatial percep-
tion but also in the integration of multisensory informa-
tion for the purpose of motor control. This was already suggested in the last century, as classical deficits following IPL lesions include visuospatial neglect, disorientation and failure of oculomotor control. More recently, our knowledge of the functional role of the IPL has strongly benefitted from electrophysiological data obtained from alert primates.

Within the dorsal pathway, the first stage at which visual information is combined with extra-retinal signals related to eye movements seems to be at area MST. Prior to area MST, area MT is involved in the generation of pursuit eye movements but does not seem to receive extra-retinal inputs, since suppression of the visual stimulation during pursuit abolishes the neuronal activity. By contrast, responses in area MST can be maintained, suggesting a convergence of visual and extra-retinal inputs at this area. Furthermore, area 7a in this pathway contains many neurones that respond to visual stimulation, but with a strong dependence on the behaviour of the animal, for example, regarding gaze direction. Finally, the distance of fixation can also modulate the activity of a population of neurones in area 7a (and probably area MST), which led Sakata and colleagues to conclude that extra-retinal and visual information relative to the 3D location of the target in space seem to converge on these neurones.

The influence of extra-retinal inputs related to head movements is also revealed by recordings in areas 7a (Ref. 55) and MST (Ref. 56). In particular, in the lateral ventral part of area MST (MSTd) the activity of a population of visual-tracking neurones is correlated with visual stimulation, ocular pursuit and head movements. Since the preferred directions for these three types of stimulation are generally in good agreement, Thier and Erickson proposed that these neurones encode the direction of the target in space.

As distinct from the large number of cortical studies, subcortical investigations related to the interactions between depth perception and motor control are very sparse, and were usually conducted in cats, rather than in monkeys. In cats, several structures involved in gaze orientation or motor control contain cell populations that can be visually activated, and which are sensitive to binocular disparity. Based on anatomical findings, it has been hypothesized that this selectivity could be achieved through cortical inputs.

Self-motion or 3D structure?

Areas MT and MST are therefore candidates for the processing of 3D structure from motion parallax, as well as for interactions with self-motion related signals. Because the size of the receptive field is very large in area MSTd (typically 30° at a 10° eccentricity), it has been argued that this area is involved in the processing of self-motion, whereas the smaller receptive fields in area MT (typically 10° at a 10° eccentricity) make this area a better candidate for the extraction of 3D shape from motion. However, it remains unclear how the shape of the 3D object is processed from motion in area MT, since neurophysiological recordings in this area have failed to validate the computational approach proposed so far, involving first-order derivatives of velocity.

Concluding remarks

Visual perception of 3D space, as based on various cues including motion parallax, binocular disparity and the coupling between head movements and visual inputs, is involved in most aspects of self-motion perception and control. Reciprocally, translations of the eye in space are a source of depth information through retinal motion. At the interface between sensory and motor systems, with a sensitivity
to optic flow and binocular disparity, the dorsal visual pathway (involving areas MT, MST and 7a) seems a likely locus for these interactions. The recent advances in the electrophysiological exploration of this pathway are promising. A striking element of these advances is the emergence of real scientific contacts between behavioural, neurophysiological and computational approaches, and between the physiology of perception and motor control. Such contacts will be essential for our future understanding of the 3D visual perception in the moving observer.

**Selected references**

Neuronal networks for induced '40 Hz' rhythms

John G.R. Jefferys, Roger D. Traub and Miles A. Whittington

A fast, coherent EEG rhythm, called a gamma or a '40 Hz' rhythm, has been implicated both in higher brain functions, such as the 'binding' of features that are detected by sensory cortices into perceived objects, and in lower level processes, such as the phase coding of neuronal activity. Computer simulations of several parts of the brain suggest that gamma rhythms can be generated by pools of excitatory neurones, networks of inhibitory neurones, or networks of both excitatory and inhibitory neurones. The strongest experimental evidence for rhythm generators has been shown for: (1) neocortical and thalamic neurones that are intrinsic '40 Hz' oscillators, although synchrony still requires network mechanisms; and (2) hippocampal and neocortical networks of mutually inhibitory interneurones that generate collective 40 Hz rhythms when excited tonically.


FAST, GAMMA RHYTHMS have been implicated in higher cognitive function. They are also known as '40 Hz' rhythms, but actually range from 30 to 100 Hz and might vary in frequency during a response. The 20–100 Hz range we consider here overlaps with the beta band (15–30 Hz) of the EEG, but we will ignore the finer points of EEG classification. The natural history and functional roles of synchronous gamma oscillations have been reviewed recently1-3, and so will be considered only briefly.

Gamma rhythms occur in humans and other mammals following sensory stimuli, often in brief runs. 'Induced rhythms' at 50–60 Hz were first described in the olfactory bulb by Adrian4, and have since been identified in the olfactory cortex5, visual cortex6-9, auditory cortex10,11, somatosensory cortex12 and motor cortex13-15. Gamma oscillations also occur in the hippocampus16,17, where the link with external sensory stimuli is less direct, but might still exist in the form of multimodal inputs received from higher-order sensory cortices. Hippocampal gamma rhythms tend to occur during the theta band (4–12 Hz) of the EEG, which is a prominent feature of the hippocampus in vivo16,18, especially during exploration.

In humans the auditory response includes brief '40 Hz transient responses19,20, which increase when the subject pays attention, and which disappear with loss of consciousness during anaesthesia21. Repetitive auditory stimulation at ~40 Hz generates a large '40 Hz steady-state response'22. Recordings of brain magnetic activity (magnetoencephalograms or MEGs) in humans suggest that gamma rhythms can be very widespread2-3, during both waking and dream states. Other MEG measurements in humans suggest that gamma rhythms might be organized to sweep across the whole brain, perhaps providing 'temporal binding...into a single cognitive experience'24.

Neuronal firing

Single-unit recordings in vivo have revealed much about the events or features to which neurones respond. Individual neurones do not detect their preferred sensory features in isolation, but form part of neuronal networks whose emergent properties define the feature-detection properties of the cortical column. In the visual system, it used to be thought that successive hierarchies of neurones encoded progressively more-complex features of

Acknowledgements

This work was supported by Esprit Project Macons 6615. We thank Tjeerd Dijkstra, Jacques Droulez, Franck Bremmer and Werner Graf for valuable comments on the manuscript.