Vision Based Autonomous Robot Navigation: Motion Segmentation[1]

Michael R. Blackburn and Hoa G. Nguyen
Naval Command, Control and Ocean Surveillance Center
Research, Development, Test & Evaluation Division
San Diego, CA, USA 92152-7383

95RO023

Abstract

The ability to acquire and respond appropriately to targets or obstacles, moving or stationary, while underway, is critical for all unmanned mobile robot applications. This is achieved by most animate systems, but has proven difficult for artificial systems. We propose that efficient and extensible solutions to the target acquisition, discrimination and maintenance problem may be found when the machine sensor-effector control algorithms emulate the mechanisms employed by biological systems. In nature, visual motion provides the basis for these functions. Because visual motion can be due either to target motion or to platform motion, a method of motion segmentation must be found. We present a solution to this problem that emulates natural strategies, and describe its implementation in an autonomous visually controlled mobile robot.

The Problem of Motion Segmentation

Motion segmentation is used here as the process of identifying the 3-D spatial location of a unique source of motion from the optic flows created by one or more independently moving objects in the visual field. Motion segmentation is a greater problem than identifying and localizing a source of motion because the observer himself may also be moving in a complex visual space causing all other objects, whether stationary or moving, to contribute to the optic flow though induced motion.

Consider a person driving his car down a busy street. In the course of driving, the driver uses saccades to select targets upon which to fix his gaze, and pursuit eye movements to hold his gaze upon his selected target. These targets may be other moving cars, pedestrians, billboards or other objects resting on the ground. While the relative target motion is minimized by the fixation mechanisms, images of the objects in the foreground and background relative to the target will continue to contribute to the optic flow on his retina during target fixation due to the motion of the car as well as to the pursuit eye movements. This flow can be non-uniform in direction as well as speed, depending upon the distance to the objects and their position with respect to the target and the relative direction of gaze with respect to the direction of travel of the car. Still, new targets are quickly noticed in the driver's peripheral vision, especially if they are moving independently of their surround. The mechanism that draws attention to these new targets is the focus of our interest in motion segmentation.

Studies of eye movements of drivers have shown that unusual motion is a strong attractor of attention (Thomas, 1969). For example, a car that takes off from the parked position is irresistible, so is a blinking light, like a turn indicator, or a bouncing ball that appears across the traffic. While the driver most often notices objects that are themselves moving, motion in itself is inadequate to

[1] Proceedings for the Dedicated Conference on Robotics, Motion, and Machine Vision in the Automotive Industries. 28th ISATA, 18-22 September 1995, Stuttgart, Germany, 353-360.

guarantee attention. On a busy street many objects are in independent motion and yet are easily ignored. The common factor in the events that attract the attention of automobile drivers apparently is the unpredictability of the motion. When a motion parameter changes, it is temporarily unpredictable.

Consider now the mechanisms that may be employed to identify and attend to the location of independently moving objects. One such mechanism may null the common elements of motion, thus favoring the unique. Center-surround interactions, similar to those that provide pattern enhancement (Kuffler and Nicholls, 1977), could be used to compare the motion of small objects with the motion of the background. Such mechanisms have been described for the Hawk Moth (Collett, 1971), for the optic tectum of birds (Frost et al., 1990), for the tectum of cats (Sterling and Wickelgren, 1969), for the cat visual cortex (Jones, 1970), and for the primate medial temporal cortex (Allman, et al., 1985; Tanaka, et al., 1986; Tanaka, et al., 1993). Essentially, the center-surround mechanism suppresses the response to motion in the center of a receptive field if the motion in the surround of the receptive field is in the same direction. If the motion of the surround is in the opposite direction to the motion of the center, the response to the motion in the center can be enhanced. The center-surround mechanism could be speed independent as nearby objects generate optic flows with greater speeds than do distant objects during auto translation but are not perceived as uniquely moving.

The center-surround mechanism accomplishes a form of spatial prediction; the surround predicts the motion of the center. If both the surround and the center changed direction of motion together, this change would not be noticed at the output. Spatially consistent motion could be expected during induced motion, but regions of inhomogeneous contrast, such as a spot on the plaster wall, could dominate attention even during induced motion for lack of an adequately moving surround. That it does not do so suggests either that the surround involves the entire visual field, that there is in addition a mechanism of reafferent inhibition, or that the consistent motion of the spot contributes to its predictability and insignificance. Of course, all three possibilities could participate.

The possibility of integration over the entire visual field is reduced, however, by the absence of long range connections in either the subcortical or cortical visual areas. Yet, the size of the receptive fields of cortical elements is known to increase in sequential visual processing areas leading away from the primary visual cortex. Feedback from later motion processing areas, such as the dorsal region of the medial superior temporal cortex (Tanaka et al., 1993), could provide global motion information to the cortical region or regions responsible for motion segmentation.

The difficulties in summarizing the optic flow globally for an active vision system are great. In primates the bilateral organization of the nervous system allocates half of the visual field to one side of the brain and the remaining half of the visual field to the other side of the brain. Thus integration of visual information in one half of the brain generally does not cover the entire visual field, but is limited to the contralateral hemifield. Furthermore, motion analysis takes place in the cortex on a projected field that has undergone a log-polar mapping from the retinal surface (Tootell et al., 1982). Global patterns of motion on the retinal surface are quite different in the cortex following this log-polar mapping. The log-polar transformation simplifies the motion patterns due to translation and rotation on the Z perceptual axis, resulting in parallel flows, but complicates the motion patterns of translations on the X and Y perceptual axes. Thus on the occasions when an observer is looking at a target that is not in line with his direction of travel, the patterns of optic flow on the cortex of other non target objects will differ in direction between the upper and lower visual field quadrant projections.

The direction and magnitude of the motion components of fixed non-target objects will depend upon the direction of travel of the observer relative to the target and of the location of the other objects relative to the target and observer. Objects located on either side of the target on the distance dimension will move in opposite directions. Objects proximate to the observer will move faster than objects located beyond the target. Objects located on the same side of the target both on the distance dimension and on the lateral dimension will be consistent in direction but not in magnitude. Therefore, only under a few conditions is the integrated motion of the global motion field descriptive of common motion.

One such condition is when the observer is moving in the direction of the target upon which he has fixed his vision. Then the motion of new non-rigid moving objects in the peripheral visual fields will have expansion and rotation components that clearly differ from the background. The center/surround mechanisms described above can enhance these differences. When, however, the observer is moving lateral to the target but maintaining a visual fix upon it, the determination of unique motion is more complicated. Even within a quadrant, the background motion can be inconsistent depending upon the distance of the fixed objects relative to the target. However, this motion parallax can provide powerful depth cues, and when once established could be used to mediate the prediction of background motion for center-surround comparisons. We cannot rule out the possibility that global measures of field motion participate in motion segmentation, but the computational difficulties involved militate against it. We also suspect that nature selects the simplest solutions to any computational problem.

Reafferent suppression of anticipated optic flow upon the execution of eye, head, or body movements (with or without feedback from either the vestibular or proprioceptive systems) would require considerable predictive accuracy of the resulting motion patterns given any of a great number of movements. This is also unlikely because it would require in addition some information about the source of the visual contrast. Proprioceptive and vestibular input do contribute to image stabilization by reflexes that move the eyes in a direction opposite to the motion of the head. The calculations required to do this are much simpler than calculations required to predict the optic flow resulting from a movement of the eyes or head. The stabilization is effective only for the immediate region of visual fixation anyway. Motion still is generated in peripheral vision with a stabilized eye during head movements. Thus the remaining likely mechanism for suppression is temporal consistency. The temporal consistency of a local motion vector is easily calculated, requiring only the persistence of the previous state. Then, if the trajectory of a moving spot, locally determined, is either consistent with its immediate surround or with itself over time, it may be ignored. The evidence suggests that to command attention, the motion of an object must be both spatially and temporally unique. Therefore, the vision system could be primarily calculating the second derivative of a spatio-temporal function (acceleration).

For the purpose of prediction, change in a motion vector is not easily assessed within a receptive field. This is because the motion of an object takes the object across receptive fields in a short period of time. When the motion has been detected and assessed, the target has moved on. However, within certain spatial and temporal limits, the activity resulting from directed motion can persist and influence, through lateral projections, the activity of the output elements along a direction determined by the previous activity. In this way a projection of inhibition could be used to suppress predictable motion. Such a projection is consistent with the network organization that subserves basic motion analysis. That is, no new organization need be invented to assess the change in the pattern of motion, the process only needs to be repeated, like taking the derivative of velocity to assess acceleration. The final output of the motion segmentation process then may represent a region of visual space that contains new (unpredicted) motion that differs from its surround.

The 5B layer of the visual cortex is a good candidate for the final output of the motion segmentation process for it projects to the superior colliculus (tectum in amphibians), which integrates activity to determine the next target location. As motion information is filtered through cortical layers or between processing stations on its way out to layer 5B, its uniqueness could be assessed by the above spatial and temporal criteria.

Implementation of the Algorithm

The motion segmentation algorithm was built on previous models of biological vision (Blackburn et al., 1987; Blackburn and Nguyen, 1990, 1994). First in our implementation of motion segmentation, the activity in the surround of each motion detector (*Y6B*) is integrated. We set the outer radius of the surround at approximately 32% of the lateral dimension of the surface of the processing layer, while the inner radius of the surround, which is exempt from integration, was set at about 8% of the surface dimension, so these regions of integration are actually quite large. As the resolution of the processing layer is increased the relative dimensions of the regions of integration can be reduced.

$$surround_{cd,i,j} = \Sigma_{a,b}\ Y6B_{cd,i+a,j+b},$$

where the subscript *cd* refers to one the four cardinal directions on the log-polar transformed computational plane of motion sensitivity that we currently analyze (down *d*, up *u*, right *r*, and left, *l*), subscripts *i* and *j* index the location of the direction detector on the computational plane, and *a* and *b* range from the dimensions of the inner radius to the dimensions of the outer radius.

Second, the input to the unique motion detector (*Y5B*) from the current motion detector is suppressed by residual activity of like motion detectors that reside geometrically in its trail (*Y6Bt*). For example, the output of motion detectors that represented motion to the right in a local region of the computation plane, would be inhibited by the short term history of rightward motion in detectors located immediately to its left on the plane. In this way, attention to continuous motion is soon suppressed. At the same time, both surround inhibition of contemporary motion and feedforward inhibition of the motion history combine to enhance and focus the output of unique or unusual motion.

$$Y5B_{d,i,j} = (Y6B_{d,i,j} - Y6Bt_{d,i-k,j})*surround_{u,i,j} / surround_{d,i,j},$$

$$Y5B_{u,i,j} = (Y6B_{u,i,j} - Y6Bt_{u,i+k,j})*surround_{d,i,j} / surround_{u,i,j},$$

$$Y5B_{r,i,j} = (Y6B_{r,i,j} - Y6Btr_{r,i,j-k})*surround_{l,i,j} / surround_{r,i,j},$$

$$Y5B_{l,i,j} = (Y6B_{l,i,j} - Y6Bt_{l,i,j+k})*surround_{r,i,j} / surround_{l,i,j}.$$

Positive activity in the output layer (*Y5B*) is then sent to the tectum where it competes in the selection of new targets.

Application to an Autonomous Mobile Robot

The objective of our application was to develop an autonomous mobile robot capable of visual target detection, tracking, and trailing. Specifically, the robot is tasked with following a human walking through an office complex. For a demonstration of autonomy, all sensor-effector loops had to be completed on the robot, without external assistance in the form of target designation or environmental modeling. In order to be applicable to unstructured environments the robot had to accomplish this task without the aid of any explicit a priori knowledge of the floor plan, or the aid

of any special codings or markings in the environment, including any special treatment of the target. Vision was the only means by which the robot was permitted to gain information about the external environment. Further, only visual motion information was used. Because of the visual complexity of the office complex, the robot visual system had to segment the unique motion of the walking human from the induce motion of the background in order to accomplish acquisition and maintain the target during tracking and trailing.

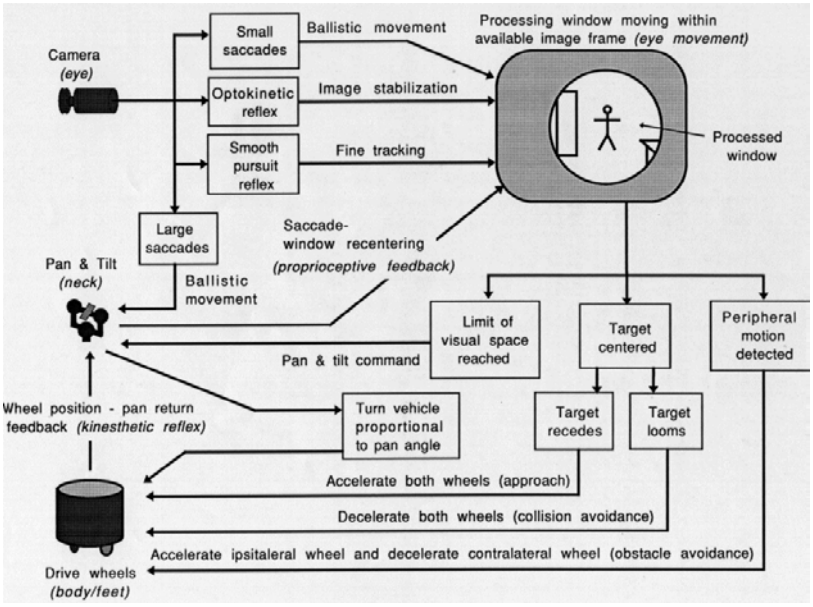

Fig 1. Visual-motor functions and relationships

Fig 2. The autonomous visually guided robot trailing a walking human in a cluttered room.

We fitted a mobile robot with video camera, pan and tilt mechanism, on-board computer and biologically based visual-motor control algorithms. The above motion segmentation algorithms were added to the robot control algorithms developed in earlier work (Blackburn and Nguyen, 1994).

Figure 1 diagrams the various visual-motor functions of the robot. Because the robot visual system is essentially reactive, the behavior of the animate target determines the behavior of the robot. Targets are detected by a model of the vertebrate optic tectum, using a biased cooperative mechanism between hemifields. The algorithm determines the center of mass of potential targets from the unique motion information from the *Y5B* layer, and directs motors controlling the pan and tilt mechanism to bring the center of the receptor field upon the target. A smooth pursuit reflex, which takes input from motion detectors in the foveal region (using the *Y6B* layer activity), attempts to keep the fovea centered on the acquired moving target. An optokinetic reflex, which responds to full field motion, stabilizes the eye when the robot body is in motion. Reorientation of the robot to trail an acquired target is accomplished by basing commands to the robot drive motors on the camera pan angle, requiring the robot to drive in the direction of the gaze. This process is analogous to the targeting motion of the eyes, head and body in biological systems. Trailing is accomplished by triggering forward thrust of the robot when the predominant motion of a centered target is toward the center of the visual field (contracting motion field). Collision is avoided by decreasing forward thrust when the target motion is away from the center (expanding). Obstacle avoidance is achieved by decreasing thrust on the side of the robot opposite to the peripheral motion away from the center of the visual field. The obstacle avoidance reflex, which is transitory, assumes

precedence over the pursuit reflex, allowing the robot to skirt around obstacles in pursuit of a target. The target motion can then drive the robot forward, but to the robot, forward is temporarily away from nearby obstacles. The camera windowing and pan and tilt mechanisms meanwhile continue to attempt to maintain central fixation on the target. For a discussion of the biological visual processes from which we derived our algorithms, see Blackburn and Nguyen (1994).

We used a Transitions Research Corporation (TRC) Labmate Mobile Robot Base. A single CCD video camera with a 90 degree field of view, mounted on a pan and tilt mechanism built in-house, provided monocular input to the vision processing hardware. Camera position was taken from shaft encoders located on the pan and tilt axles. Wheel motion information was obtained from encoders located on both left and right drive motor axles. Vision processing hardware included an Imaging Technologies OFG Frame Grabber coupled to a Hyperspeed Technology coprocessor board with two i860 microprocessors. The vision processing hardware cards were hosted on an 80486 PC computer located in the robot housing. The PC provided I/O to the Labmate and pan and tilt controllers. The Hyperspeed board received video data directly from the OFG board at frame rate over an ITI vision bus. One i860 processor was dedicated to subsampling the input frame and making decisions about the required motor responses, while the other i860 processor integrated the visual input into receptive fields and performed motion analysis. Pan, tilt and drive motor commands were sent to the 80486 for integration and execution. Actual processing rate with the algorithms described herein was approximately 8 frames per second. Frame rates of 20 frames per second could be achieved if all outputs used for process monitoring were disabled.

Testing was performed in a large partitioned room with an open work area of 32 by 18 feet. Three walls of this work area contained windows, doors and office furniture. An example of target acquisition and pursuit is shown in the photograph of Figure 2. From a resting position the robot turned and moved forward in pursuit of a human walking into its visual space. Using only motion information the robot was able to detect targets while either stationary or in transit. While in transit, the segmented motion provided information on the target's behavior adequate to maintain tracking and trailing. Because the system segments and attends to the dominant unique motion, target maintenance can fail whenever the target motion does not differ significantly from the background. Because we have not yet implemented motion parallax based depth perception, the motion sheer due to the induced motion of objects residing at depths which differ from the target and its background is probably a major source of confounding unique motion. In addition, common motion at computational boundaries, where both surround inhibition and directionally specific inhibition can be weak, is particularly difficult to eliminate.

Summary and Conclusions

A preliminary form of motion segmentation is described that allows a visual motion based mobile robot to acquire, track and trail moving targets in a visually complex environment. The motion segmentation algorithms produced to accomplish this behavior use temporal and spatial comparisons of local optic flow to highlight likely locations of unique motion.

This project demonstrates the utility of elementary biological models of visual motion processing in the autonomous control of a mobile robot. The machines that we hope to produce may be most efficiently made by following biological precedent. Natural mechanisms have been successful at all levels of complexity, and they achieved additional complexity and capability by maintaining and modulating more elementary functions. Since nature began with the production of very simple organisms, living successfully within protected ecologies, so might we begin in our design of artificial intelligence machines with very simple functions appropriate for a particular environment. When we become good at this, then like nature, we can expand our requirements and the complexity of the machines to meet them. If we are competent observers and modelers, it should

not take a billion years to reproduce human capability in a machine. The blueprint stands before and within each of us. Furthermore, once we have achieved that level of performance, we would be well positioned to produce a better intelligence than any of which we as individuals posses.

References

Allman, J., E. Miezin, and E. McGuinness. (1985). "Stimulus specific responses from beyond the classic receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons." Annual Review of Neurosciences, vol. 8, pp. 407-430.

Blackburn, M.R., H.G. Nguyen, and P.K. Kaomea. (1987). "Machine visual motion detection modeled on vertebrate retina." SPIE Proceedings, vol. 980, pp. 90-98.

Blackburn, M.R., and H.G. Nguyen (1990). "Modeling the biological mechanisms of vision: Scan paths." Mathematical and Computer Modeling, vol. 14, pp. 311-316.

Blackburn, M.R., and H.G. Nguyen (1994). "Autonomous visual control of a mobile robot." Proceedings of the 1994 Image Understanding Workshop. Monterey, CA, Nov. 13-16, 1994, pp. 781-788.

Collett, T. (1971). "Visual neurons for tracking moving targets." Nature (Lond) vol. 232, pp. 127-130, 1971

Daniel, P.M., and D. Whitterage. (1961). "The representation of the visual field on the cerebral cortex in monkeys." Journal of Physiology, vol. 159, pp. 203-21. Frost, B..J., D.R. Wylie, and Y. C. Wang. (1990). "The processing of object and self-motion in the tectofugal and accessory optic pathways of birds." Vision Research, vol. 30, pp. 1677-1688.

Jones, B.H. (1970). "Responses of single neurons in cat visual cortex to a simple and more complex stimulus." American Journal of Physiology, vol. 218, pp. 1102-1107.

Kuffler, S.W. and Nicholls, J.G. (1977) From Neuron to Brain. Sinauer Assoc., Sunderland, MA.

Sterling, P. and B.G. Wickelgren. (1986). "Visual receptive fields in the superior colliculus of the cat." Journal of Neurophysiology, vol. 32, pp. 1-15.

Tanaka, K., K. Hikosaka, H. Saito, M. Yukie, Y. Fukada, and E. Iwai. (1986). "Analysis of local and wide-field movements in the superior temporal visual areas of the Macaque monkey." Journal of Neurosciences, vol. 6, pp. 134-144.

Tanaka, K., Y. Sugita, M. Moriya, and H. Saito. (1993). "Analysis of object motion in the ventral part of the medial superior temporal area of the Macaque visual cortex." Journal of Neurophysiology, vol. 69, pp. 128-142.

Thomas, E.L. (1968) "Movements of the eye." Scientific American, vol. 219, pp. 88-95.

Tootell, R.B.H., M.S. Silverman, E. Switkes, and R.L. DeValois. (1982). "Deoxyglucose analysis of retinotopic organization in primate striate cortex." Science, vol. 218, pp. 902-904.

Acknowledgments

Supported by the Advanced Research Projects Agency and the Office of Naval Research under contract number N0001493WX2D002. Additional support for this project was provided by the Naval Command, Control and Ocean Surveillance Center, Independent Exploratory Development Program. The cooperation and assistance of CDR Bart Everett, Steve Timmer and Theresa Tran for the loan of the robot base, pan and tilt mechanism, and in hardware engineering are greatly appreciated.