Centering Behavior Using Peripheral Vision

David Coombs
NIST, Robot Systems Div. *
Building 220, Room B-124
Gaithersburg, MD 20899
coombs@cme.nist.gov

Karen Roberts‡
Center for Neural Systems
Boston University
Boston, MA 02215
kroberts@cns.bu.edu

Abstract
The ability to control egomotion using low resolution peripheral vision is crucial for enabling a small high resolution fovea to attend to features that require detailed examination. The bee-bot demonstrates the ability to use low resolution motion vision over large fields of view to steer between obstacles. The system uses the maximum flow observed in left and right peripheral visual fields to indicate obstacle proximity. Each peripheral field constitutes one-third of a wide-angle lens. The left and right proximities are compared to steer through the gap. Negative feedback control of steering is able to tolerate inaccuracies in the signal estimation. This interpretation of the flow is based on the assumption that the camera is translating along the gaze vector. This condition is maintained under egomotion by active gaze stabilization. Head rotation is countered by eye rotation, and gaze is returned to the heading by rapid camera movements when necessary. The low cost of such basic navigation competence can free additional resources for attending to the environment.

Keywords: active vision, real-time robot systems, motion vision, mobility

1 Introduction
Low-resolution peripheral vision has not been widely applied to low-level vehicle mobility competence. In contrast, it is possible for humans to safely travel a hallway while reading a book or paper (which occupies our scant high-resolution visual sensing) looking up only occasionally to investigate potential threats identified at low-resolution. In fact it must be assumed that much visual support of mobility can be achieved with very low resolution vision. The acuity of human vision drops from 60 cycles/degree at the fovea to 3 cycles/degree at 40° eccentricity [7]. This is roughly equivalent to having at least 120 pixels/degree at the fovea and 6 pixels/degree at 40° eccentricity. (To sharpen the comparison, consider that a “wide” angle camera with a field of view of only 115° (3mm lens) will have fewer than 4.5 pixels/degree in a 512 x 512 image!) Cutting et al. [5] estimate that a human jogging at 2m/s must estimate her instantaneous heading or within about 4° of its true direction (in order to have sufficient warning to avoid an obstacle). Running at 10m/s requires accuracy of about 1°. Interestingly, the same analysis suggests that accuracy of 1° is also sufficient for ordinary automobile driving, downhill skiing, and landing an airplane. It may be that the richest visual information for this task is found in the visual periphery, and yet acuity is low in the periphery. Similarly, Nelson and Aloimonos [11] note the utility of panoramic fields of view for locating the foci of expansion and contraction (FOE and FOC), which is useful for estimating heading. These observations suggest it may be profitable to consider the use of wider fields of view and lower resolution than are often studied in motion understanding for mobility.

The present work was undertaken to provide rudimentary visual mobility capabilities to support goal-directed activities. It represents a preliminary version of a robot system that is being developed to investigate issues of representation for effective operation in dynamic environments. This early system is intended to be a simple “auto-pilot” the robot control system can employ to follow a hall or traverse an uncluttered room with minimal computational effort in the course of accomplishing a task. For instance, the robot might explore its environs to build a map, and the robot would later use the map and its basic mobility capaci-
of cameras facing to the sides of a translating robot can compare the distances to objects on each side of the robot. For instance, larger flows to the left than the right indicate the robot is nearer the left side of the gap. (In fact it is not strictly necessary to use side-looking cameras, but the peripheral fields must be approximately symmetric about the heading.) The difference of the flows reflects the centering error, which is all that is needed to servo the robot's heading toward the middle of the gap. Thus the steering behavior is realized in real time by using measures of optical flow that are easily implemented on image processing hardware: normal flow can be estimated with spatiotemporal gradients in local operations. (A related but more sophisticated system uses flow field divergence to qualitatively detect obstacles without explicitly estimating their range [12].)

These optical flows arise from both the structure of the environment and the vehicle's motion. When the camera translates through the environment, the optical flow reflects the scene structure. However, when the camera rotates, the flow is smooth and does not depend on the scene structure. Therefore the camera must be rotationally stabilized so that the optical flow will describe the scene structure even when the robot is rotating.

In order to stabilize gaze, the camera must be rotationally stabilized. Note that at any time, gaze can be stabilized on only a single point in the world. Nevertheless, points near the fixation point will be nearly stabilized as well. One of the easiest ways to begin stabilizing gaze is to simply rotate the camera with velocity \( \omega_g = -\omega_h \), the opposite of the head's angular velocity. This stabilizes gaze on the horizon (i.e., on a point of infinite distance along the gaze direction). Visual signals alone indicate the true degree of stabilization, and they may be used in various ways to hold gaze [4, 15, 13, 16]. Although a complete gaze holding system's success ultimately relies on vision, most visual processing is too slow to respond to high frequency disturbances, so nature relies on the fast vestibular and otolith organs (which are essentially rotational and translational accelerometers) to sense head motions. Open loop stabilization systems compensate for high frequency components of head motions, and visual following (e.g., the optokinetic system) corrects low frequency visual drifts [9]. Open loop stabilization suffices for this application.

2 Peripheral Optical Flow

The robot needs a robust measure of optical flow obtainable in real time in order to gauge the relative clearance to each side of its path. Robustness at least requires that results be consistent and that failures not be catastrophic. Ideally, the method could handle large flows. However, it would still be difficult to estimate the small translational component well enough.
to permit relative range estimation, because it is hard to measure with enough accuracy the large flows that are primarily due to camera rotation. The requirement of real-timeliness limits the techniques at least to non-iterative methods unless there is a guarantee of convergence in a small number of iterations, and promising results are appearing in this area [2, 3, 8].

This work takes the expedient approach of estimating normal optical flow at full spatial resolution. The single forward-looking camera is equipped with a wide-angle (115°) 3mm lens. The magnitude of flows that can reasonably be estimated lies somewhere between 0.25 and 1.5 pixels per inter-image interval, Δt. Δt was chosen to be 6/60s = 100ms to keep flows in the measurable range for tight clearance (e.g., 15cm) at low speeds (e.g., 10cm/s). The robot’s speed is adjusted to keep the flows within the measurable range as clearance changes.

The simplest case is opposed side-looking cameras. Further assume that the cameras are translating in a plane. Due to the aperture problem, normal flow only reflects the projection of optical flow onto the local image intensity gradient. In order to compare normal flows from the side-looking cameras, only flows with complementary phases in each image can be compared meaningfully. (See Figure 2.) For instance, flows in direction θ in the right image should be compared with flows in direction π - θ in the left.

The choice of θ affects the relationship between the normal and true optical flows. Let φ be the true direction of flows in the right image that arise from the camera translation. If φ can be chosen to be near φ, then the normal flow will nearly reflect the true optical flow. However, it is also desirable that there be considerable areas of large image intensity gradient in directions θ and π - θ in the right and left images, respectively (e.g., edges oriented perpendicular to θ and π - θ). A single measure (the maximum) is taken of the flows in each peripheral visual field. For the task of comparing clearance to obstacles to each side of the path, the maximum flow indicates the distance to the nearest objects in the respective peripheral fields. The optical flow processing in each field is implemented in two parts: first, normal flow is estimated, then the maximum flow is identified by examining a histogram of the flows. On PIPE (pipelined image-processing engine) [10], normal flow is estimated using local neighborhood operations. The intensity images are blurred by Gaussian kernels, and the normal flow is estimated by a gradient method. A histogram is made for each peripheral field of all flows in locations with large horizontal intensity gradients (i.e., vertical edges). The histogram is transferred to the host workstation, where it is searched for the largest flow.

It should be noted that this implementation does not attempt to compute range for objects in the image. To do so would require compensating for the eccentricity from the heading at each flow estimate. In particular, flows are smaller in a forward- or oblique-looking camera than in a side-looking camera, since the focal plane only observes the projection of the motion vector onto its plane. I.e., an object near the focus of expansion (FOE) or contraction (FOC) will generate less optical flow than an object at the same range to the side of the robot. This implies that it is desirable to keep the peripheral visual fields symmetric about the heading. Otherwise, objects to the side to which the robot is turning will appear farther away than objects of the same range on the other side. In other words, if the robot heads to the right, the right hand field becomes more forward-looking, and the left hand field approaches a side-looking orientation. Further, wide angle lenses distort flows as a function of their retinotopic eccentricity. For these reasons, the control system is designed to keep the gaze nearly aligned with the heading.
3 Control

Like its visual system, the robot's control system is simple. The robot's goal is merely to move forward without collision. The motion coordinator commands either centering or deflection behaviors. The human operator requests deflection (when collision is imminent, for instance), and the coordinator sends commands to deflect the robot's path; otherwise the coordinator sends commands for the centering behavior.

The system of the robot provides an estimate (in the form of left- and right-hand flows, $F_L$ and $F_R$) of the robot's proximity to the sides of the "gap" seen by its camera. The change in heading, $\Delta \Phi$, is proportional to the centering error, $\Phi_e$.

$$\Phi_e = |F_R| - |F_L|$$

$$\Delta \Phi = K_\Phi \Phi_e$$

Thus, when $|F_R| < |F_L|$, indicating that the nearest object is to the left, the robot veers to the right. The controller achieves smooth turning behavior by controlling the angular velocity of the robot. An open-loop proportional gain is used with the estimated interval time, $T$, to set the drive velocity.

$$\phi = K_\phi \Delta \Phi / T$$

The control gain is less than unity to ensure stability.

The nonlinear gaze control is a nystagmus, a repetitive eye motion of slow phase rotations punctuated by quick phase rapid returns. The visual system of the robot provides an estimate (in the form of left- and right-hand flows, $F_L$ and $F_R$) of the robot's proximity to the sides of the "gap" seen by its camera. The change in heading, $\Delta \Phi$, is proportional to the centering error, $\Phi_e$. If the maximum flow magnitude is small and decreased if it is large. In practice, the assumption of forward camera translation in a static environment is quite often violated. For instance, the movement of a human can easily overshadow the flows due to camera translation. For stability, when the flows are inconsistent with this assumption, the robot slows down and maintains the current heading.

4 Results

The PIPE program histograms horizontal normal flows in $16/60s = 0.27s$. The sample interval $T \approx 0.7s$. No attempt has been made to optimize execution speed (e.g., by interleaving communications and computations). Figure 4 plots the robot's path in a trial in the lab. The robot began in the lower left corner of the lab $(125, 125)$ and began traveling up. After about 12 minutes, the trial ended with the robot in the same corner of the lab, although the dead-reckoned estimate of robot location is just through the lab wall. The dead-reckoning worsened throughout the course of the trial, with a pronounced negative bias in $Y$. (The robot did not drive through the table or the wall.) The robot drove back and forth across the lab and sometimes negotiated the turns into the left and right sides of the lab. In the trial, the operator signaled the robot when collision seemed imminent. The robot stopped, turned, and resumed operation.

Gaze is stabilized fairly effectively in the slow phase. Disruptions occur during the quick phase, when flows reflect the quick camera rotation rather than forward translation. This may result merely in the robot slowing down, but often the disturbance causes the robot to slow to a stop. The steering control performs poorly when gaze deviates more than about $10^\circ$ from the robot heading, when the sideways translation of the camera produces significant enough flows to corrupt the flows due to forward translation.

The peripheral vision has difficulty, predictably, seeing objects that lack high contrast vertical edges. The robot sometimes tries to clip the corners of the table. Interestingly, the robot's response to a moving object is fairly sensible. Typically a moving object produces motion that is inconsistent with the forward camera translation assumption, and the robot stops. If the object moves in the direction consistent with robot egomotion, the robot will turn away. Although it was not intended, this behavior causes the robot to steer around its tether where it dangles, swinging back and forth at times, from the ceiling. In a bustling room, this behavior paralyzes the robot. It is, however, an error in the conservative direction.

5 Discussion

In summary, the centering behavior demonstrates the use of low resolution motion vision over large fields of view to steer through clear space. The system uses one
Figure 3: Description of the robot system: There will soon be a task coordinator and a path planner. For the moment, the motion coordinator has standing orders to move forward, and it accepts a signal from the human interface to deflect the robot's path. The motion coordinator issues commands for either the centering or the deflection behavior. Each behavior is elaborated by a task decomposition tree. The centering behavior steers through clear space using peripheral flows to indicate relative proximity to the sides of the robot's current path. The deflection behavior stops and turns the robot and re-centers the robot's gaze.

Dead-reckoned Robot Path

Figure 4: robot's Path: The robot's path was recorded from dead-reckoning. The path began in the lower left corner of the lab (125, 125). The dead-reckoning worsened throughout the course of the trial, with a negative bias in Y, and rotating from Y toward X. (The robot did not drive through the table or the wall.)
receptive field for each of the left and right peripheral fields of view. The largest optical flow in a receptive field indicates the proximity of the nearest obstacle. The left and right proximities are compared to steer between the obstacles. The simplicity of the approach is possible because of the wide field of view, servoed steering, and active gaze stabilization. The relatively wide field of view allows the robot to see a bit to the sides, where it will be going when it turns. Negative feedback control tolerates errors in signal estimation. Active gaze stabilization ensures that the camera is rotationally stabilized and that gaze is aligned with heading. Without these conditions, this simple visual processing would not suffice.

In general, it is not always possible to completely stabilize the camera, and residual rotational flows result. For instance, a vehicle may experience a sudden jolt from running over a cable, and the stabilization systems may not have the bandwidth to respond before the image is affected. The simplest method of coping with this problem is to ignore visual observations that are inconsistent with the assumption that gaze is stabilized. A more desirable approach would involve compensating for the disturbance to estimate the stabilized signal. Such approaches will be complicated by noise in flow estimation and distortions wide-angle lenses.

The centering behavior works fairly well when the environment is modeled reasonably by a channel or hall. However, the centering behavior has a tendency to drive the robot straight toward walls and into corners, perceiving the peripheral flows to be balanced as it does so. In a minimal "auto-pilot" mobility system, peripheral "hall" centering behavior must be joined by frontal obstacle detection and avoidance. Ultimately, navigating more complex environments that include chairs, doorways and people will require the ability to sense and comprehend, at least qualitatively, the free space and obstacles surrounding the robot. Nevertheless, there may still be a place for very simple methods. There will always be a need for both robustness and real-time performance, and the relationship between them is likely to remain a trade-off.

In contrast to walking down the hall while reading, it is very difficult (and certainly unsafe) to drive a car with only peripheral vision. The fovea is used to evaluate potential hazards (e.g., peripheral motion that could arise from an object hurtling into the path of the vehicle). Less than 3 degrees of the visual field is covered by the fovea's high acuity of 60 cycles/degree, and it appears to be sufficient to take foveal samples of the scene 3 or 4 times a second to permit survival in many environments. Clearly it is important to integrate information from these fixations. Low-resolution peripheral flow navigation and other such rudimentary capabilities can permit a system to devote more of its limited resources (e.g., time, eye movements, processing power, etc.) to attending to these crucial aspects of the world.

Acknowledgments
Ernest Kent and Martin Herman initially suggested this work and provided guidance throughout. James Albus and Daniel Raviv provided invaluable discussions. Tsai-Hong Hong proposed the normal flow algorithm and gave the benefit of her experience estimating optical flow. Marilyn Nashman and Thomas Wheasley made it easy to interface to the PIPE. Hui-Min Huang prepared the system organization and task decomposition diagrams.

References