Proceedings of the 2003 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Las Vegas, Nevada · October 2003

Judging Distance by Motion-based Visually Mediated Odometry

Igor Katsman*, Alfred Bruckstein*, Robert J. Holt†, Ehud Rivlin*
*Department of Computer Science, Technion — IIT 32000, Haifa, Israel
†Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey, USA
{igork, ehudr, freddy}@cs.technion.ac.il, rjh@research.bell-labs.com

Abstract— **Inspired by the abilities of both the praying mantis and the pigeon to judge distance by use of motion-based visually mediated odometry, we create miniature models for depth estimation that are similar to the head movements of these animals. We develop mathematical models of the praying mantis and pigeon visual behavior and describe our implementations and experimental environment. We investigate structure from motion problems when images are taken from a camera whose focal point is translating according to each of the biological models. This motion in the first case is reminiscent of a praying mantis peering its head left and right, apparently to obtain depth perception, hence the moniker "mantis head camera." In the second case this motion is reminiscent of a pigeon bobbing its head back and forth, also apparently to obtain depth perception, hence the moniker "pigeon head camera." We present the performance of the mantis head camera and pigeon head camera models and provide experimental results of the algorithms. We provide the comparison of the definitiveness of the results obtained by both models. The precision of our mathematical model and its implementation is consistent with the experimental facts obtained from various biological experiments.**

I. INTRODUCTION

The study of vision guided abilities in animals has become significant not only for biologists, but also for scientists working in robotics and computer vision who are using unique functional principles learned from the study of animals to develop mathematical models, and then to build an intelligent robot utilizing these principles for better performance in certain tasks.

This study examines, experimentally evaluates and compares visually mediated motion based depth determination of two species of animals, namely the Praying Mantis and the Pigeon. Praying Mantis relates to the insect group of animals, while Pigeon relates to the birds. In this paper we focus on robot vision for depth estimation purposes. With the continuously growing development of autonomous robots, many groups of researchers (both engineers and biologists) have conducted studies in different directions of biologically inspired robotics vision. Generally, work in this domain can be classified as top-down ([1]) and bottom-up (e.g. [2]). In the top-down approach, a certain task, such as path planning ([3], etc.) or visually mediated odometry ([4], etc.), looks for inspiration in a biological model, whereas in the bottom-up approach, a certain biological behavior, such as visually mediated navigation ([5], etc.) or visually mediated flight control ([6], etc.) is directly modeled with real robots. Our work belongs to the bottom-up approach.

The work presented here can be thought of as an attempt to model visual information acquisition and processing behaviors in the animal kingdom. We investigated the Praying Mantis and the Pigeon. A number of studies of formal behavioral models (such as the schema-theoretic model) of the praying mantis have been presented [7], [8]. In these studies, several visually mediated activities or behaviors of the praying mantis such as prey acquisition, predator avoidance, mating and Chantlitaxia were formulated in detail. Each of the above behaviors could be implemented by a set of visually based functions, one of which is investigated here.

Fig. 1. Mantis Head versus Mantis Head Camera

To compensate for an extremely limited ability of movements of their eyes, birds' heads are able to move significantly. More specifically, some birds develop frequent forward-backward head-movements at the rate of a few cycles per second. This head bobbing seems to play a significant role in Pigeons' vision, which has been extensively investigated (e.g. [9], [10]).

The Mantis and Pigeon provide excellent opportunities for the study of distance estimation by means of self-generated retinal image motion. The biological model could then form the basis for a biologically relevant mathematical model that would take all of the experimental findings into account and that could be of assistance for research in computer and robotic vision. Thus the eye of a mobile robot could make peering-like or bobbing-like movements, like a Mantis or Pigeon, to estimate distances in order to avoid objects among other purposes. The precision of our mathematical model and its implementation is consistent with the experimental facts obtained from biological experiments.

Depth estimation from motion is a subject of several works in computer vision [11], [12], [13]. For example, authors in [14] handle reconstruction of 3D geometrical primitives from controlled motion, authors in [15] analyze

0-7803-7860-1/03/$17.00 © 2003 IEEE

Fig. 2. Pigeon Head versus Pigeon Head Camera

the influence of intrinsic and extrinsic camera parameters errors on 3D depth distortion, and authors in [16] present a scheme for getting selectable quality 3D structure reconstruction. Here we present a biologically motivated simplified new model for motion based depth estimation and its robotics implementation.

In the remainder of this paper we develop mathematical models of the biologically motivated visual-motor systems for depth estimation, describe an implementation of the system and experimental environment, present and discuss the performance of the systems and provide experimental results of the algorithms in Sections II and III; we compare the performance of the presented models in Section IV; and discuss the results in Section V.

II. THE PRAYING MANTIS HEAD CAMERA

In this section we describe the Mathematical Model, Experimental Environment and Experimental Results of the Praying Mantis experimental evaluation.

A. The Model

Figure 3 illustrates the process. The camera moves left and right (pure translation) along the X-axis according to the function $c = c(t)$, where we set $c = c(0)$. Typically this motion is with constant speed (and changing direction at the edges of the platform) such as $c(\tau) = sV_0\tau$, where s is 1 or -1 depending on the peering direction.

We start from the following relationship, where f is the focal length of the camera: $\rho/f = r/z$. When the camera is in its initial position ($\tau = 0$), $\rho/f = r_0/z_0$. For $\tau = t$, when the camera is displaced along the X-axis according to the function $c(t)$, we have $\rho = f[r_t - c(t)]/z_t = f[r_t - c(t)]/g(r_t)$ in the same coordinate system. Hence $(1/f)\rho g(r_t) = r_t - c(t)$ or

$$r_t = \frac{1}{f}\rho g(r_t) + c(t) \ . \tag{1}$$

In the most general case we define the inverse function as

$$r_t = h_c(\rho) \ . \tag{2}$$

The image might be regarded as a function of r, which itself is a function of time and ρ, say $I = F(r_t) = F(h_c(\rho)) = F(h_c(t)(\rho))$.

Useful information can be obtained by observing the ratio of the derivatives of I with respect to ρ and t:

$$\frac{\partial I/\partial\rho}{\partial I/\partial t} = \frac{\partial F(r_t)/\partial\rho}{\partial F(r_t)/\partial t} = \frac{F'(h_c(\rho))(\partial h_c/\partial\rho)}{F'(h_c(\rho))(\partial h_c/\partial t)} = \frac{\partial h_c/\partial\rho}{\partial h_c/\partial t} . \tag{3}$$

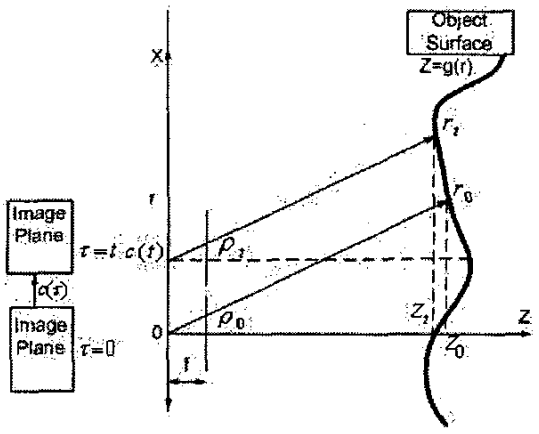

Fig. 3. Mantis head camera model. The surface, whose cross section is given by $z = g(r)$, is viewed by a camera with focal point moving along the X-axis. ρ denotes the displacement along the X-axis from the CCD center on the image plane at which a feature is projected, and $r_0(\rho)$ and $r_t(\rho)$ are the displacements where points observed at the displacement ρ on the image plane are located on the surface of the object.

To evaluate (3) we combine (1) and (2) to obtain

$$h_c(\rho) = r_t = \frac{1}{f}\rho g(r_t) + c(t) = \frac{1}{f}\rho g(h_c(\rho)) + c(t) \ .$$

Differentiation with respect to ρ and t yields

$$\frac{\partial I/\partial\rho}{\partial I/\partial t} = \frac{\partial h_c/\partial\rho}{\partial h_c/\partial t} = \frac{g(r_t)/f}{dc/dt} \tag{4}$$

and

$$g(r_t) = z_t = f\frac{dc}{dt}\frac{\partial I/\partial\rho}{\partial I/\partial t} \ . \tag{5}$$

In this expression $c = c(t)$ and dc/dt are given, while $\partial I/\partial\rho$ and $\partial I/\partial t$ are determined by observation.

B. The Experimental Environment

A miniature video camera was mounted on a specially designed micro-translation platform, which provides precise periodic side-to-side peering movements of the camera with constant speed. When an electromotor of the platform is activated, the camera translates in the direction that is parallel to the image plane. This behavior simulates the peering behavior of the praying mantis.

C. Experimental Results

In our case the camera moves along the X-axis (with constant speed dc/dt), so the component of the velocity dy/dt along the Y-axis is zero, and we can reduce the basic flow equation to the following:

$$\frac{\partial I_t}{\partial x}\frac{dx}{dt} + \frac{\partial I_t}{\partial t} = 0 \ ,$$

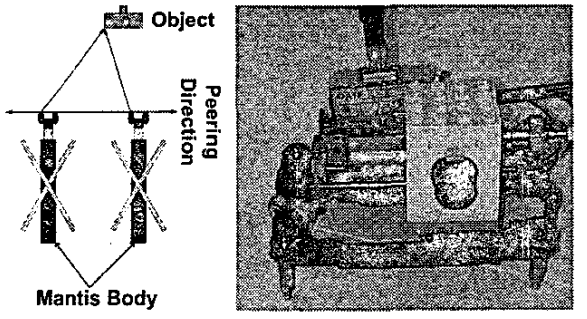

Fig. 4. Scheme of Peering Behavior of Praying Mantis and the implementation of the Miniature Mantis Head Camera Platform, which utilizes peering behavior for distance estimation.

i.e.

$$\frac{dx}{dt} = -\frac{\partial I_t/\partial t}{\partial I_t/\partial x} \ . \tag{6}$$

Using (6) one can rewrite (5) as (denoting $\rho = x$, $sV_0 = dc/dt$, and $v_1 = dx/dt$)

$$g(r) = z = -f\frac{dc/dt}{dx/dt} = -f\frac{sV_0}{v_1} \ . \tag{7}$$

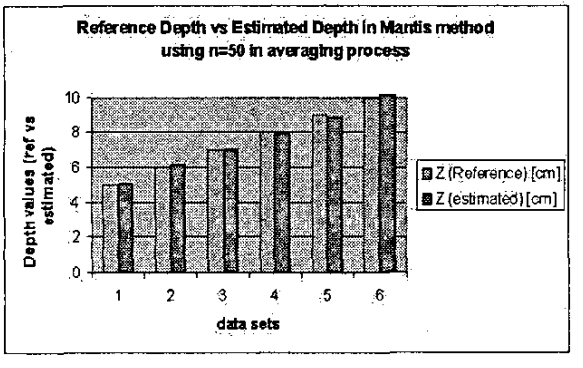

Fig. 5. Reference distance versus averaged estimated distance

According to (7), when the observer moves with speed V_0, the retinal images of objects close to the eye (smaller z) are displaced more quickly (bigger v_1) than those of more distant objects (bigger z).

In our experiments the target object was placed at various known distances in front of a constantly peering camera. The distance to the object was estimated by computing equation (7) via the token matching (fast feature tracking) technique. The experimentally estimated distances were compared to their true values and the accuracy of the estimations was calculated. For each peer of the camera the object was sampled $n = 50$ times with a constant frame rate of 30 Hz, and the average v_1 was computed as $(1/n)\sum_{i=1}^{n} v_{1i}$, which greatly improves the accuracy of the estimation algorithm.

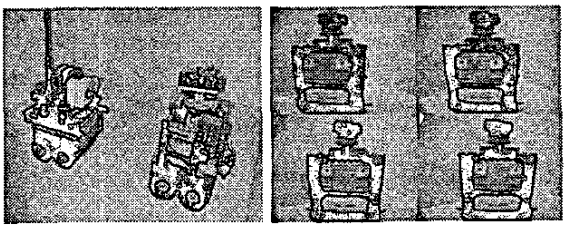

Fig. 6. Tiny Lego Robot utilizes Miniature Mantis Head Camera. Sequence order: top left, top right, bottom left, bottom right.

III. THE PIGEON HEAD CAMERA

In this section we describe the Mathematical Model, Experimental Environment and Experimental Results of the Pigeon experimental evaluation.

A. The Model

Figure 7 illustrates the setup. The camera moves back and forth (pure translation) along the optic axis according to the function $c = c(t)$, where we set $c(0) = 0$. Typically this motion is periodic and with constant speed (and changing direction at the edges of the cart platform) such as $c(\tau) = sV_0\tau$, where s is 1 or -1 depending on the bobbing direction.

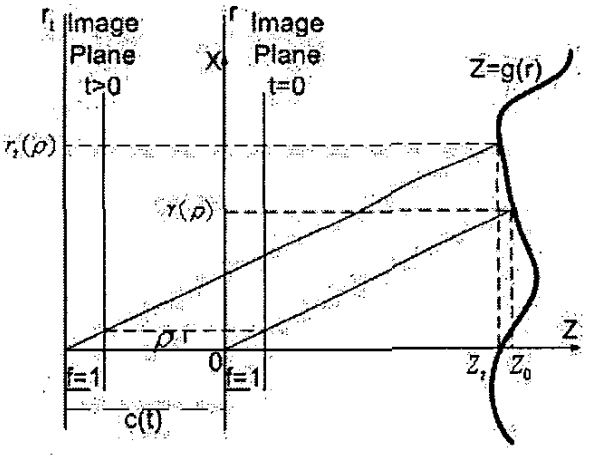

Fig. 7. The pigeon head camera model. The surface, whose cross section is given by $z = g(r)$, is viewed by a camera whose focal point moves along its optical axis Z. ρ is the height on the image plane at which a feature is projected, and r and r_t are the heights at which the point observed at height ρ on the image plane are located on the object. The camera moves in time along the optical axis according to the function $c(t)$. ρ is the same pixel coordinate on the image at times $t = 0$ and $t > 0$. r and r_t are two object points projected to the same pixel coordinate ρ at times $t = 0$ and $t > 0$.

We start from the following relationship, where f is the focal length of the camera: $\rho/f = r/z$. When the camera is in its initial position $\tau = 0$, $\rho = fr_0/z_0$. At time $\tau = t$, when the camera is displaced along its optical axis

according to the function $c(t)$, we have $\rho = fr_t/[z_t + c(t)]$ in the same coordinate system, whence

$$r_t = \frac{1}{f}\rho[z_t + c(t)] = \frac{1}{f}\rho[g(r_t) + c(t)] \ . \quad (8)$$

In the most general case we define the inverse function as

$$r_t = h_c(\rho) \ . \quad (9)$$

The image may be regarded as a function of r which itself is a function of time and of ρ, say $I = F(r_t) = F(h_c(\rho)) = F(h_c(t)(\rho))$. We claim that we can obtain useful information by observing the ratio of the derivatives of I with respect to ρ and t. To wit,

$$\frac{\partial I/\partial \rho}{\partial I/\partial t} = \frac{\partial F(r_t)/\partial \rho}{\partial F(r_t)/\partial t} = \frac{F'(h_c(\rho))(\partial h_c/\partial \rho)}{F'(h_c(\rho))(\partial h_c/\partial t)} = \frac{\partial h_c/\partial \rho}{\partial h_c/\partial t}. \quad (10)$$

To evaluate (10) we combine (8) and (9) to obtain $h_c(\rho) - \rho g(h_c(\rho))/f = \rho c/f$. Differentiating with respect to ρ and t yields

$$\frac{\partial I/\partial \rho}{\partial I/\partial t} = \frac{\partial h_c/\partial \rho}{\partial h_c/\partial t} = \frac{c + g(r_t)}{\rho(dc/dt)} \quad (11)$$

and

$$g(r_t) = z_t = \rho\frac{dc}{dt}\frac{\partial h_c/\partial \rho}{\partial h_c/\partial t} - c = \rho\frac{dc}{dt}\frac{\partial I/\partial \rho}{\partial I/\partial t} - c \ . \quad (12)$$

In this expression $c = c(t)$ and dc/dt are given, while $\partial I/\partial \rho$ and $\partial I/\partial t$ are determined by observation.

B. The Experimental Environment

A miniature video camera was mounted on a specially designed micro-translation platform, which provides precise periodic back-and-forth bobbing movements of the camera with constant speed. Thus as an electromotor of the platform is activated, the camera translates along its optic axis. This actually simulates the bobbing behavior of the walking pigeon.

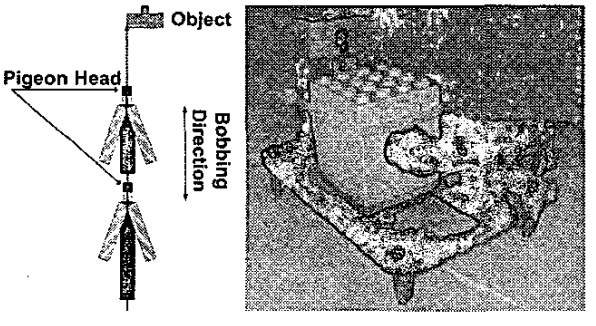

Fig. 8. Scheme of Bobbing Behavior of Pigeon and the implementation of the Miniature Pigeon Head Camera Platform, which utilizes bobbing behavior for distance estimation.

C. Experimental Results

The situation regarding optic flow is the same as that in Mantis Head Camera section. By computing the optic flow we can estimate the distance to the objects by using equation (12). As noted in Section II-C, the two prevalent approaches to computing optic flow are token matching or correlation, and gradient techniques. Here we use a fast feature-tracking scheme to calculate the optic flow. In our experiments, the target object was placed at various known distances in front of the constantly bobbing camera. The distance to the object was estimated by computing flow through the use of token matching (fast feature tracking) techniques. The experimentally estimated distances were compared to their known values and the accuracy of the estimations was calculated. In each bobbing of the camera the object was sampled $n = 10$ times with constant frame rate 30 Hz.

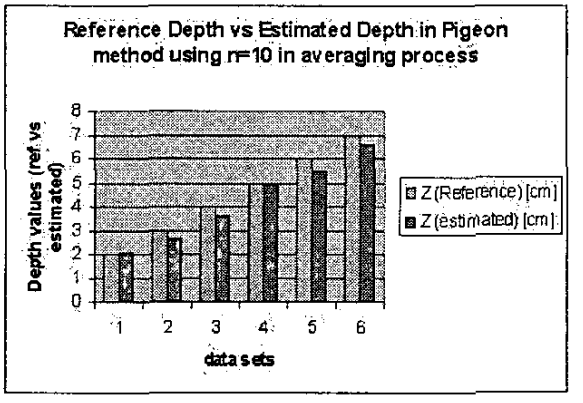

Fig. 9. Reference depth versus averaged estimated depth

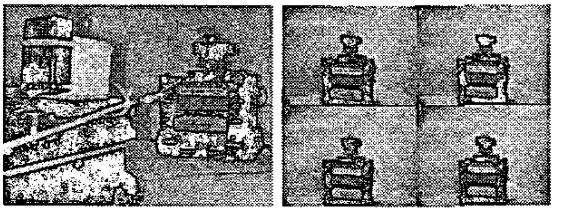

Fig. 10. Tiny Lego Robot utilizes Miniature Pigeon Head Camera. Sequence order: top left, top right, bottom left, bottom right.

IV. When to Use What

While the Mantis model is more precise than the Pigeon model, its usage is possible only from a static location while the Pigeon model can be used during motion. Thus the Mantis Head model should be used when the precise estimation of depth from the static position of a mobile agent is required, while the Pigeon Head model should be used when the rough estimation of depth from a moving or

non-moving mobile agent is needed. This implies that the selection of usage of the depth estimation method should be task dependent [17].

In what follows we mention briefly some experiments that were conducted to demonstrate the usage of the above methods for different tasks (such as robot convoy, robot docking, etc). Each robot could be equipped with one of several vision configurations such as peering camera, bobbing camera, both peering and bobbing cameras, or combinations of these — e.g. a camera that could be rotated 90 degrees and used for peering or bobbing, dependent on task.

For example, we used the peering method to perform precise docking of the robot to its base station. Finer steps are taken as the robot gets closer to its target.

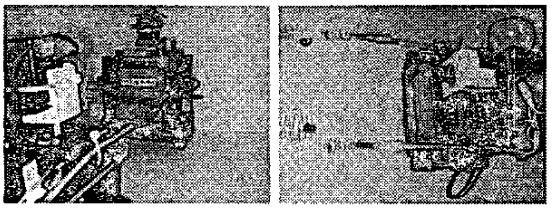

Fig. 11. Tiny Lego Robot performs docking to base station utilizing peering method-another view.

Another example is performing simple robot convoy of a few robots. Here the bobbing method is used, and the rough estimate of the distance will be enough to successfully perform this task.

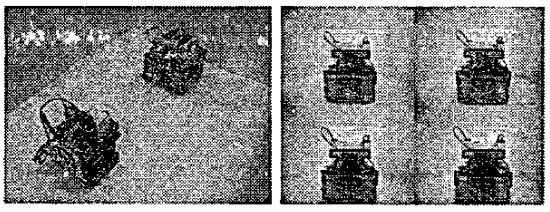

Fig. 12. Tiny Lego Robots perform convoy. maneuver utilizing bobbing method. Sequence order: top left, top right, bottom left, bottom right.

In another example some robot locates the nearest robot among the others. Here only ordinal depth estimation will be enough to successfully perform this task.

A. Ordinal Depth Estimation

From biological studies it follows that animals often use relative motion parallax for depth perception [18]. In other words, relative distances can be determined from motion parallax, if the distinction is made between the apparent motions of objects relative to one another.

In some tasks knowledge of the ordinal depth is enough for animals to perform adequately. To calculate the ordinal depth there is no need to measure the exact absolute value of the head movement velocity. This fact could be used to simplify the model implementation in several cases. Praying mantises use the so-called Chantlitaxia behavior when choosing a hunting place. They just select the nearest object, move towards it, then again select the next nearest object and move again towards it. When an appropriate location is found, the praying mantis starts hunting from it. In this Chantlitaxia behavior the estimation of ordinal depth is enough to select the nearest object in each step. Still, during the hunting process the absolute distance estimation is required.

In order to estimate which object is more distant and which one is closer (for example for Chantlitaxia purposes) the Praying Mantis could use the peering behavior and check the sign of the expression $(Z_1/Z_2)-1$, which is derived from (5). If the object with depth Z_1 is closer than the one with Z_2, the sign of the expression above is negative.

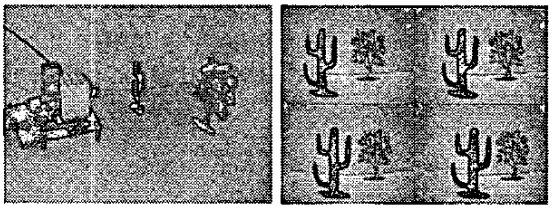

Fig. 13. The subsets (of 4 samples) captured during peering movements of the Mantis Head Camera Platform for ordinal depth estimation (Nature scenarios).

V. CONCLUSIONS

In this study biologically motivated mathematical models of depth estimation and their implementations were presented. We showed how one could recover depth using either peering behavior that is commonly used by the praying mantis, or bobbing behavior that is commonly used by the pigeon. These models are consistent with recent behavioral and anatomical evidence collected from various biologic experiments [18], [19].

The presented systems can estimate the depth of a set of objects, similarly to the ability of certain animals, which can be used by a mobile agent for learning the surrounding space, collision avoidance and navigation. The real-time performance of the models adds to its attractiveness for usage with mobile agents. The precision of the depth estimations, achieved by the models and their implementations, are consistent with the results demonstrated by animals.

As items for future work, we plan to investigate other visual routines of the mentioned animals. Particularly, we plan to use our mantis head platform mounted on miniature mobile robots in order to implement some of the visual behaviors of the praying mantis, as presented

in [7]. We also plan to implement some of the real-time indoor navigation algorithms ([20], [21], [18]), using Lego mobile robots with the mantis head platform. Using precise distance estimation by the platform, Lego robots will be available to perform accurate docking and other precision-requiring tasks, which are difficult to achieve with the standard Lego environment. As another direction for future work, we plan to study the principles of different types of self motion for precise depth estimation used by other animals, measure their sensitivity, evaluate precision and compare these principles to those used by the praying mantis and pigeon.

In this study, we have developed a mathematical model of the biologically motivated visual-motor system for distance estimation, then described an implementation of the system and experimental environment, presented and discussed the performance of the system and experimental results, and presented directions for the future work.

VI. REFERENCES

[1] D. Lambrinos, R. Möller, T. Labhart, R. Pfeifer, and R. Wehner, 2000, "A mobile robot employing insect strategies for navigation," *Robotics and Autonomous Systems, special issue on Biomimetic Robots*, vol. 30, 2000, pp. 39–64.

[2] R. Möller, D. Lambrinos, T. Roggendorf, R. Pfeifer, and R. Wehner, "Insect strategies of visual homing in mobile robots," in *Biorobotics*, eds. T. Consi and B. Webb, AAAI Press, 2000.

[3] S. Chameron, G. Beugnon, B. Schatz, and T. S. Collett, "The use of path integration to guide route learning in ants," *Nature*, vol. 399, no. 6738, 1999, pp. 769–772.

[4] A. A. Argyros, C. Bekris, and S. O. Orphanoudakis "Robot homing based on panoramic vision", TR287 ICS-FORTH, 2001.

[5] F. Iida, 2001, "Goal-directed navigation of an autonomous flying robot using biologically inspired cheap vision," in *Proceedings of the 32nd ISR (International Symposium on Robotics)*, 2001, pp. 19–21.

[6] M. V. Srinivasan, M. Lehrer, W. Kirchner, S. W. Zhang, and G. A. Horridge, "How honeybees use motion cues to estimate the range and distance objects," in *Proc. IEEE SMC*, 1988, pp. 579–582, (in English).

[7] R. C. Arkin, K. S. Ali, A. Weitzenfeld, and F. Cervantes-Perez, "Behavioral models of the praying mantis as a basis for robotic behavior," *Journal of Robotics and Autonomous Systems*, vol. 32, no. 1, 2000, pp. 39–60.

[8] R. C. Arkin, "Ecological Robotics: A Schema- theoretic Approach", in *Sensing, Modelling and Planning*, eds. R. C. Bolles, H. Bunke, and H. Noltemeier, World Scientific, 1997, pp. 377–393.

[9] B. J. Frost, "The optokinetic basis of head-bobbing in the pigeon," *J. Exp. Biol.*, vol. 74, 1978, pp. 187–195.

[10] M. N. O. Davies and P. R. Green, 1991, "The adaptability of visuomotor control in the pigeon during flight," *Zool. Jahrb. Physiol.*, vol. 95, 1991, pp. 331–338.

[11] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun, "Structure from motion without correspondence," in *IEEE, CVPR00*, 2000, pp. 557–564.

[12] G. Sandini, and M. Tistarelli, "Active tracking strategy for monocular depth inference over multiple frames", *IEEE-PAMI*, vol. 12, no. 1, 1990, pp. 13–27.

[13] W. Zheng, Y. Kanatsugu, Y. Shishikui, and Y. Tanaka, "Robust depth-map estimation from image sequences with precise camera operation parameters," in *ICIP00*, 2000.

[14] F. Chaumette, S. Boukir, P. Bouthemi, and D. Juvin, "Structure from controlled motion", *IEEE-PAMI*, vol. 18, no. 5, 1996, pp. 492–504.

[15] L. Fah and T. Xiang, 2001, "Characterizing depth distortion under different generic motions," *IJCV*, 2001, pp. 199–217.

[16] L. K. Dalmia and M. Trivedi, "High speed extraction of 3D structure of selectable quality using a translating camera," *Computer Vision and Image Understanding*, vol. 64, 1996, pp. 97–110.

[17] Y. Aloimonos, "Active vision revisited," in *Active Perception*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1993.

[18] K. Kral, "Side-to-side head movements to obtain motion depth cues: a short review of research on the praying mantis," *Behavioural Processes*, vol. 43, 1998, pp. 71–77.

[19] Y. Yamawaki, "Effect of luminance, size and angular velocity on the recognition of non-locomotive prey models by the praying mantis," *Journal of Ethology*, vol. 18, no. 2, 2000, pp. 85–90.

[20] V. Lumelsky and T. Skewis, "Incorporating range sensing in the robot navigation function," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 5, 1990, pp. 1058–1069.

[21] I. Kamon, E. Rimon, and E. Rivlin, 1998, "Tangent bug: a range-Sensor-Based navigation algorithm," *International Journal of Robotic Research*, vol. 17, no. 9, 1998, pp. 934–953.