Autonomous Robots 18, 21–42, 2005

Head Movements for Depth Perception: Praying Mantis versus Pigeon

ALFRED BRUCKSTEIN
Department of Computer Science, Technion, IIT 32000, Haifa, Israel

ROBERT J. HOLT*
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey, USA

IGOR KATSMAN AND EHUD RIVLIN
Department of Computer Science, Technion, IIT 32000, Haifa, Israel
igork@cs.technion.ac.il

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 and error analysis 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.

Keywords: depth estimation, range estimation, depth from motion, motion based visually mediated odometry, motion parallax

1. 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.

During certain types of head movement, objects within the retinal image change their position both relative to one another and to the observer. This can be used for the determination of object distances. For example, when the head moves from side to side, objects within the field of view appear to change position relative to one another, with nearby objects exhibiting a greater apparent displacement than more remote objects. Note that this phenomenon, known as motion parallax, can be observed in head movements of various animals. Exploring animals' motor output triggered by motion stimuli, biologists conduct behavioural-analytical studies (Kral and Devetak, 1999; Lambrinos et al.,

*Present address: Department of Mathematics and Computer Science, Queensborough, City University of New York, Bayside, New York 11364, USA.

2000; Land, 1999), to obtain the evidence that animals use such a mechanism for distance perception. Animals' behavioural reaction, ideally suitable for such studies, is a targeted jump. Owing to its ballistic nature, correction of the required motor output and control of the amplitude and velocity of the movement during the jump are virtually impossible. Therefore, exact distance measurement must take place prior to the jump. This is mainly important for long jumps, accompanied by both large energy consumption and high risk of injury. A biological behavioural-analytical investigation, studying distances and other visual attributes during jumps, was conducted for several animal species, while the unproved hypotheses were formulated for all others. This provides the opportunity to conduct robotics and computer vision experimental studies to verify the validity of some of the hypotheses on the one hand, and build intelligent robots, utilizing learned principles for better performance in certain tasks on the other hand.

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. For each group of animals, the discussion is first put into the context of why motion factors may be important for depth perception. A comparison of the apparent movements of objects is also conducted to determine the relative distances, while the absolute distance to the object is determined via the velocity of the head motion and the motion of the object's image across the retina of the eye.

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 (e.g. Weber et al., 1998; Lambrinos et al., 2000; Collett and Rees, 1997) and bottom-up (e.g. Möller and Lambrinos, 1997; Möller et al., 1998, 2000). In the top-down approach a certain task, such as path planning (Chameron et al., 1999; Collett, 1996; Collett and Rees, 1997) or visually mediated odometry (Srinivasan et al., 1997, Argyros (Srinivasan et al., 1997; Argyros et al., 2001; Collett and Rees, 1997; Nicholson et al., 1999; etc.) looks for inspiration in a biological model, whereas in the bottom-up approach, a certain biological behavior, such as visually mediated navigation (Srinivasan et al., 1996; Iida, 2001; Möller and Lambrinos, 1997; Möller et al., 2000) or visually mediated flight control (Srinivasan et al., 1988, 1996, 1997; Iida, 2001; etc.) is directly modeled with real robots. Chaumette et al. (1996) handle reconstruction of various 3D geometrical primitives from controlled motion, Fah and Xiang (2001) analyse the influence of the errors of the camera intrinsic and extrinsic parameters to 3D depth distortion, and Dalmia and Trivedi (1996) present the scheme for selectable quality 3D structure reconstruction. Lewis and Nelson (1998) tried to model the peering behavior for mid-range depth perception using a standard normal-type robot, rolling left and right. Their implementation was too rough to obtain high precision of depth estimation since there are several sources of noise factors in the model that affect the final results. The authors came to this conclusion in their discussion and analysis of the precision of their results. Our work belongs to the bottom-up approach.

Insects' eyes are immobile and have fixed-focus optics (Kral and Devetak, 1999), which makes them unable to use the degree of dynamic convergence of the lines of sight to estimate the distance from the objects they are viewing. Refractive power required to focus the image of an object on the retina cannot be used by insects for this purpose either. In some animals, including humans, distance information can be obtained from binocular cues, i.e. by making use of the fact that two eyes, due to their slightly different viewpoints, receive different information. However, the eyes of insects are located significantly closer together than human eyes, and thus have a lower spatial acuity (Kirschfeld, 1976). As a result, insects are capable of estimating the distance to an object by means of binocular cues with a relatively low precision (Köck et al., 1993). Such a distance is reasonably restricted by the closest objects, located within an interval ranging from a few millimetres to a few centimetres. It follows from this that at greater distances, cues other than binocular ones may play a role in distance measurement. The hypothesis that motion cues could be involved has been indicated by various studies such as (Srinivasan et al., 1999).

Kral and Poteser (1997) indicate that the praying mantis accompanies each sideways movement of the body with a compensatory counter-rotation of the head about the yaw axis, so that the head is always directed straight forward and thus remains oriented toward the edge of the object. The mechanism controlling the linearity of the head movement is not clear yet, although there is some evidence of co-ordination between the visual system and the mechanical position and movement receptors, particularly those located in the neck (Poteser et al., 1998).

The praying mantis provides excellent opportunities for the study of depth estimation by means of self-generated retinal image motion and motion parallax. The biological experimental facts could then form the basis for a biologically relevant mathematical model that would take the experimental findings into account, which could be of assistance for research in computer and robot vision. Thus the eye of a mobile robot could peer from side to side like a praying mantis, to estimate depth in order to avoid objects and for other purposes.

The work presented here can be thought of as an attempt to model visual information acquisition and processing behaviors in the animal kingdom. Let's first concentrate on the Praying Mantis. A number of studies of formal behavioral models (such as the schema-theoretic model) of the praying mantis have been presented (Ali and Arkin, 1998; Arkin et al., 2000). 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. Several experimental biologic studies researching the visual abilities and behaviors of the Praying Mantis were presented in Kamon and Rivlin (1997), Poteser and Kral (1995), Poteser et al. (1998), and Yamawaki (2000a, 2000b). The precision of our mathematical model and its implementation is consistent with the experimental facts obtained from various biological experiments.

It is well known that the eyes of most non-predatory birds are located in a fairly lateral position (Martin and Katzir, 1995, 1999). Such a position provides an extensive panoramic field of view, whereas the small area of binocular overlap leads to either limited or non-existent binocular vision. However, the precise pecking behaviour of birds such as pigeons, which have a binocular overlap of up to 30 degrees, suggests that binocular cues may be involved in depth judgement (Martinoya et al., 1988). Nevertheless, some authors exclude binocular disparity as a possible cue (McFadden, 1993, 1994). On the other hand, to compensate for an extremely limited ability of movement of the birds' eyes, birds' heads are able to move significantly. More specifically, some birds develop a frequent forward-backward head-movement at the rate of a few cycles per second. This "head-bobbing" seems play a significant role in vision and is unlikely to represent any type of social or mating behaviour.

For pigeons, the visual role of head-bobbing has been extensively investigated (e.g. Dunlap and Mowrer, 1930; Frost, 1978; Davies and Green, 1988, 1991; Troje and Frost, 2000). During walking, the head movement consists of two alternating phases: a thrust phase and a hold phase. In the thrust phase, the head is quickly thrust forward. While in the hold phase, the head remains in a completely fixed position in space, in terms of both horizontal (along the roll axis), and vertical (along the yaw axis) translations, and rotation about the pitch and yaw axes (Troje and Frost, 2000). It is evident that the hold phase is under visual control. The head-bobbing seems to be an optokinetic reaction, stabilising the retinal image during the hold phase. Its effect is comparable to eye saccades in humans. As a result, image processing might be applicable here, especially for the detection of the moving objects (Frost, 1978; Troje and Frost, 2000).

In their study, Davies and Green (1990) observed pigeons exhibiting head-bobbing during landing, whereas they did not reveal this feature during steady flight or takeoff. In particular, during landing, in addition to ensuring image stabilisation, head-bobbing may also provide depth information via motion cues (Troje and Frost, 2000). On the other hand, a flying bird has a problem of having no direct access to information concerning its own speed over the ground. Proprioceptors can provide information concerning the velocity relative to the surrounding air. The movement of the air itself, caused by wind or convection currents, would constitute another velocity component. It is supposed that differential motion parallax information could be used to find a solution to this problem (Troje and Frost, 2000).

The pigeon provides 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 bobbing-like movements back and forth, like a pigeon, to estimate distances in order to avoid objects among other purposes. The precision of our mathematical model and its implementation is

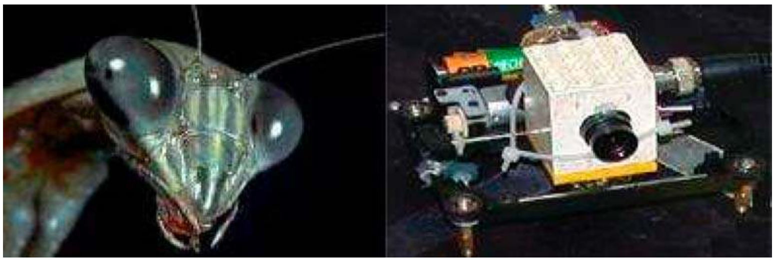

Figure 1. Mantis Head versus Mantis Head camera.

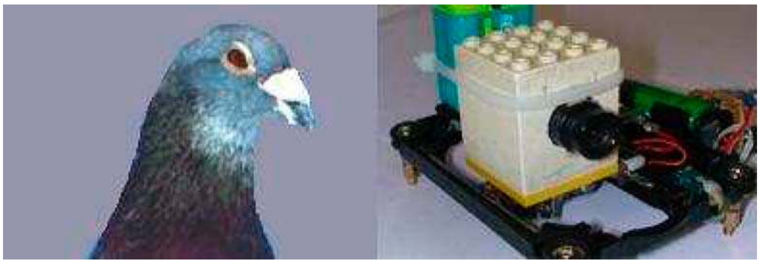

Figure 2. Pigeon Head versus Pigeon Head camera.

consistent with the experimental facts obtained from biological experiments.

Depth estimation from motion is a subject of several works in computer vision (Dellaert et al., 2000; Sandini and Tistarelli, 1990; Zheng et al., 2000). For example, authors in Chaumette et al. (1996) handle reconstruction of 3D geometrical primitives from controlled motion, authors in Fah and Xiang (2001) analyze the influence of intrinsic and extrinsic camera parameters errors on 3D depth distortion, and authors in Dalmia and Trivedi (1996) 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.

Time-ordered image sequences are popular tools that allow the estimation of projected two-dimensional image motion as either instantaneous image velocities or discrete image displacements. These are usually called the optic flow or image velocity fields, respectively. If the optic flow is a sufficiently accurate approximation to a two-dimensional image motion, it may be used to determine the three-dimensional motion and structure of an object, up to a common scale factor. Optic flow has many other applications, including object segmentation, collision detection and avoidance, determination of the focus of expansion, motion compensated encoding, and measurement of stereo disparity. Barron et al. (1994) contains an extensive comparison of several flow generation algorithms.

In this paper we provide computation evaluation of head-motion-based depth estimation in several animals (Introduction-Section 1); 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 and error analysis of the algorithms (Praying Mantis Head Camera Model, Pigeon Head Camera Model—Sections 2 and 3); we also compare the performance of the presented models and discuss the advantages and disadvantages of the compound movements and their variations (Comparison of the two models and some variations—Section 4); and we discuss the results and also propose potential usage of the systems in a mobile robot environment (Section 5).

2. The Praying Mantis Head Camera

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

There are several sets of controlled motion parameters that allow the calculation of the depth from self-generated head movements. Some of them are based on retinal velocity measurements while others deal with motion amplitude of the objects in the retinal image. To determine absolute distances via motion parallax generated by head movements, the following mechanisms are suggested: (1) the velocity of the head movement is kept constant, so that the distance to the target object is inversely proportional to the image velocity; (2) the amplitude of the head movement is kept constant, while velocity of the head movement varies, thus allowing computation of the distance to the target object relying on the relationship between the head amplitude and the amplitude of the image motion; (3) the amplitude of the head movement is adjusted in such a way that the amplitude of the image motion is kept constant, as close as possible to the threshold; a linear increase in the amplitude of head movement with distance is predefined, with the result that the object distance is determined on the basis of the amplitude of head movement only; (4) the velocity of the head movement is adjusted in such a manner that the velocity of the image motion approaches a constant, as close as possible to the threshold; as a result, a linear increase of the velocity of head movement with distance is predefined. It can be concluded from the nature of the head movements that the simplest principles among the above (1 and 2) based on constant velocity or amplitude of the head movement are the feasible methods to utilize (Kral and Devetak, 1999; Lambrinos et al., 2000; Land, 1999). For the mantids, it is very likely that the distance information is provided by image velocity, described in the first case above (Kral, 1998, 1999; Poteser et al., 1998; Srinivasan et al., 1999). Our current experimental configuration utilizes the velocity measurements, which is similar to the method utilized by the Praying Mantis.[1]

2.1. 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.

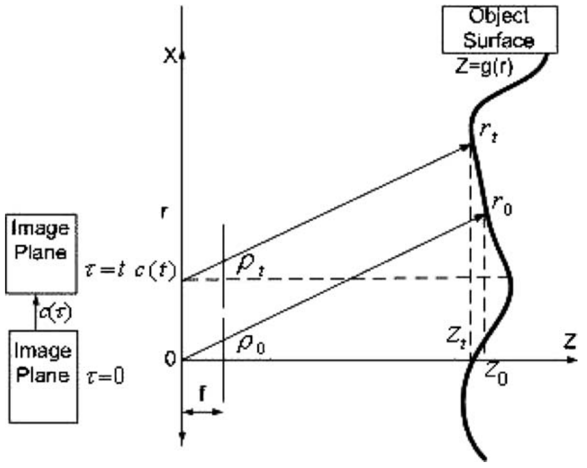

Figure 3. Mantis head camera model. The surface, whose cross section is given by $z = g(r)$, is viewed by a camera with focal point moving along the X-axis. ρ denotes the displacement along the X-axis from the CCD center on the image plane at which a feature is projected ($\rho_0 = \rho_t = \rho$), and $r_0(\rho)$ 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 ($r_0 \neq r_t$).

We start from the following relationship, where f is the focal length of the camera:

$$\frac{\rho}{f} = \frac{r}{z}.$$

When the camera is in its initial position ($\tau = 0$),

$$\frac{\rho}{f} = \frac{r_0}{z_0}.$$

For $\tau = t$, when the camera is displaced along the X-axis according to the function $c(t)$ we have

$$\rho = f\frac{r_t - c(t)}{z_t} = f\frac{r_t - c(t)}{g(r_t)},$$

in the same coordinate system. Whence,

$$\frac{1}{f}\rho g(r_t) = r_t - c(t),$$

or

$$r_t = \frac{1}{f}\rho g(r_t) + c(t). \tag{2.1}$$

In the most general case we define the inverse function as

$$r_t = h_c(\rho). \tag{2.2}$$

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:

$$\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}} = \frac{\frac{\partial F(r_t)}{\partial \rho}}{\frac{\partial F(r_t)}{\partial t}} = \frac{F'(h_c(\rho))\frac{\partial h_c}{\partial \rho}}{F'(h_c(\rho))\frac{\partial h_c}{\partial t}} = \frac{\frac{\partial h_c}{\partial \rho}}{\frac{\partial h_c}{\partial t}}. \tag{2.3}$$

To evaluate (2.3) we combine (2.1) and (2.2) 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

$$\frac{\partial h_c(\rho)}{\partial \rho} = \frac{1}{f}\left[\rho g'(h_c(\rho))\frac{\partial h_c(\rho)}{\partial \rho} + g(h_c(\rho))\right],$$

and

$$\frac{\partial h_c(\rho)}{\partial \rho} = \frac{\frac{1}{f}g(r_t)}{1 - \frac{1}{f}\rho g'(r_t)}; \tag{2.4}$$

$$\frac{\partial h_c(\rho)}{\partial t} = \frac{1}{f}\rho g'(h_c(\rho))\frac{\partial h_c(\rho)}{\partial t} + \frac{dc}{dt},$$

and

$$\frac{\partial h_c(\rho)}{\partial t} = \frac{\frac{dc}{dt}}{1 - \frac{1}{f}\rho g'(r_t)}. \tag{2.5}$$

From (2.3), (2.4), and (2.5) it follows that:

$$\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}} = \frac{\frac{\partial h_c}{\partial \rho}}{\frac{\partial h_c}{\partial t}} = \frac{\frac{\frac{1}{f}g(r_t)}{1-\frac{1}{f}\rho g'(r_t)}}{\frac{\frac{dc}{dt}}{1-\frac{1}{f}\rho g'(r_t)}} = \frac{\frac{1}{f}g(r_t)}{\frac{dc}{dt}}, \tag{2.6}$$

$$g(r_t) = z_t = f\frac{dc}{dt}\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}}. \tag{2.7}$$

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.

2.2. 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.

The video output signal of the camera is connected to the miniature wireless video RF transmitter, which broadcasts a video signal remotely, thus enabling autonomic usage of the device on the mobile robot. Both the camera and transmitter are operated from a single 9 V battery. The total size of the platform with the camera and transmitter is [10 cm × 5 cm × 2 cm].

The video signal is then received by an RF video receiver that is connected to the PCI frame grabber located inside a Dual Pentium III workstation, which performs the image processing of all the incoming frames. In addition, based on the incoming image analysis, the workstation could send action commands back to the remote robot, supporting the peering platform.

Varying the target distance and peering velocity parameters, performance of the system was measured. Targets were placed at various distances in front of the camera: 5, 6, 7, 8, 9, and 10 cm. Peering

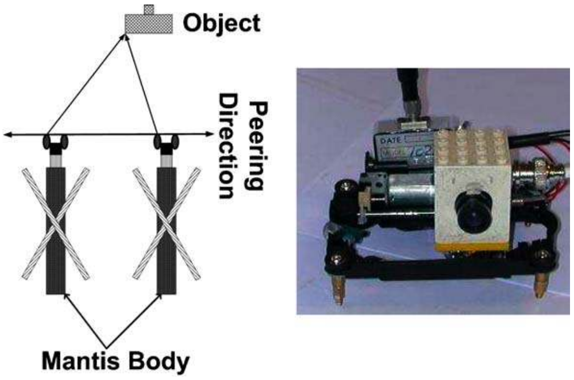

Figure 4. Scheme of peering behavior of praying Mantis and the implementation of the miniature Mantis Head camera platform, which utilizes peering behavior for distance estimation.

velocities of 1.5 cm/sec and 2 cm/sec were used. [For Mantis religiosa individuals 50 to 70 mm in size, peering amplitudes are approximately 2 to 10 mm and peering velocities approximately 6 to 18 mm s^{-1}.]

2.3. Optic Flow and Experimental Results

Optic flow is a visual displacement flow field that can be used to explain changes in an image sequence. The underlying assumption used to obtain an equation is that the gray level is constant along the visual trajectories. In other words, the partial derivative of the gray level $I(x, y)$ along the optic flow $V = (v_1, v_2) = (\frac{dx}{dt}, \frac{dy}{dt})$ is zero:

$$\frac{\partial I_t}{\partial x}\frac{dx}{dt} + \frac{\partial I_t}{\partial y}\frac{dy}{dt} + \frac{\partial I_t}{\partial t} = 0. \tag{2.8}$$

This equation alone is not sufficient to determine a unique vector field, since at any location we have only a single scalar constraint with which to find a two-dimensional vector (v_1, v_2), which constitutes an ill-posed problem. Since in our case the camera moves along the X-axis (with constant speed dc/dt), the component of the velocity dy/dt along the Y-axis is zero, so we can reduce Eq. (2.8) to the following: $\frac{\partial I_t}{\partial x}\frac{dx}{dt} + \frac{\partial I_t}{\partial t} = 0$, i.e.

$$\frac{dx}{dt} = -\frac{\frac{\partial I_t}{\partial t}}{\frac{\partial I_t}{\partial x}}. \tag{2.9}$$

Using (2.9) one can rewrite Eq. (2.7) as (denoting $\rho = x$):

$$g(r) = z = -f\frac{\frac{dc}{dt}}{\frac{dx}{dt}} = -f\frac{sV_0}{v_1}. \tag{2.10}$$

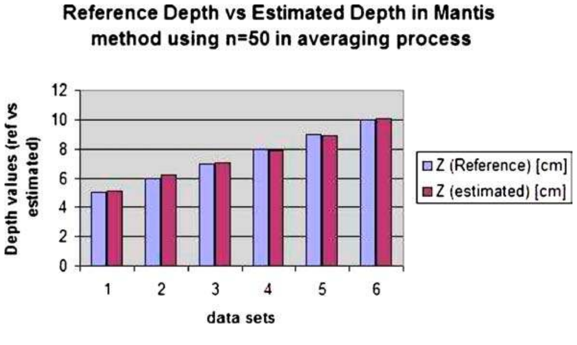

Figure 5. Reference distance versus averaged estimated distance.

According to (2.10), 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 other words, in the case of a visual field comprised of stationary objects, retinal image motion and motion parallax initiated by the observer can be used to determine the absolute and relative distance of objects. In the case of the praying mantis, translatory side-to-side movements of the head in a horizontal plane are performed to determine the jump distance to stationary objects. The speed of the retinal image motion is the relevant parameter for determining the distance to the object; thus, by computing the above optic flow one can estimate the distance to the objects.

There are two main approaches to computing optic flow: Token matching or correlation (extracting features from each frame and matching them from frame to frame) and Gradient techniques (relating optic flow to spatial and temporal image derivatives). Gradient-based methods only work when the motion is "small" and the derivative can be reliably computed. Note that for "large" motion one can employ multi-resolution methods. Tracking algorithms can compute motion when the motion is "large" by using correlation based methods or feature tracking. In our model we used the token matching approach.

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 v_1 in Eq. (2.10) 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 constant frame rate of 30 Hz, and the average v_1 was computed as $v_i = \frac{1}{n}\sum_{i=1}^{n} v_{1i}$, which greatly improves the accuracy of the estimation algorithm.

2.4. Error Analysis

In what follows we discuss the sources of inaccuracy of the described algorithm. Then a scheme for accuracy evaluation of the estimated depth will be proposed.

The algorithm presented here requires flow calculation; therefore the approximation accuracy of v_1 in expression (2.10) has a critical role for the depth estimation accuracy. The approximate precision of

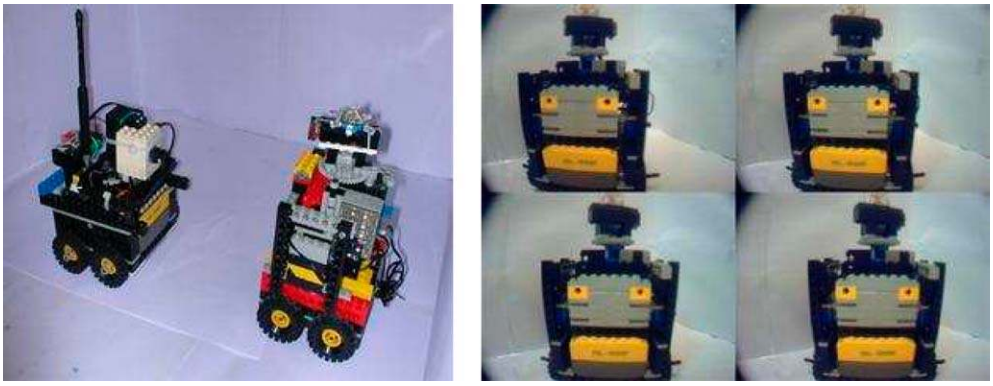

Figure 6. Tiny lego robot utilizes miniature Mantis head camera. Sequence order: top left, top right, bottom left, bottom right.

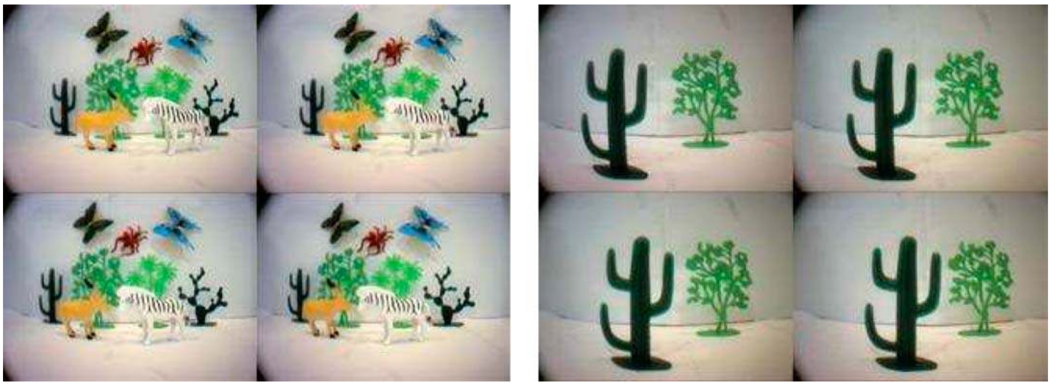

Figure 7. Two subsets (of 4 samples each) captured during left to right peer of the Mantis head camera platform (nature scenarios). Sequence order: top left, top right, bottom left, bottom right.

v_1 is mainly determined by noise and quantization. The influence of these two factors is investigated below.

The deviation in the camera motion speed is another factor, which also has a significant impact on the algorithm accuracy. The algorithm assumes a precise mechanical setup, i.e. constant camera motion speed. However, this assumption was difficult to satisfy absolutely. The influence of this factor on the algorithm accuracy is also investigated.

Let $\hat{v}$ and $\hat{u}$ be the approximated retinal images and observer velocities, respectively. Then the absolute error in the depth estimation ΔZ can be calculated using the following equation, obtained from (2.10). (For brevity, we introduce $k = -fs$ in all equations below):

$$g(r) = Z = k\frac{u}{v}, \qquad (2.11)$$

where $u = \frac{dc}{dt}$, $v = \frac{dx}{dt}$.

$$\Delta Z = \hat{Z} - Z = k\left(\frac{\hat{u}}{\hat{v}} - \frac{u}{v}\right). \qquad (2.12)$$

Let δ_u, δ_v be the relative errors in the approximation

$$\begin{bmatrix} \delta_u \\ \delta_v \end{bmatrix} = \begin{bmatrix} \frac{\hat{u}-u}{u} \\ \frac{\hat{v}-v}{v} \end{bmatrix}. \qquad (2.13)$$

Then,

$$\hat{u} = u\delta_u + u = (1+\delta_u)u,$$
$$\hat{v} = v\delta_v + v = (1+\delta_v)v.$$

Using the above notation and (2.12) we obtain:

$$\Delta Z = \hat{Z} - Z = k\left[\frac{(1+\delta_u)u}{(1+\delta_v)v} - \frac{u}{v}\right],$$

which can be simplified to

$$\Delta Z = k\frac{u}{v}\left(\frac{1+\delta_u}{1+\delta_v} - 1\right) = k\frac{u}{v}\frac{(\delta_u - \delta_v)}{(1+\delta_v)}. \quad (2.14)$$

Now, letting the upper bound of $|\delta_v|$ be M, and denoting B to be the following constant

$$B = \frac{1}{1-M}, \quad (2.15)$$

(2.14) implies that

$$\begin{aligned} \Delta Z &\approx Z(\delta_u - \delta_v) \quad \text{and} \\ |\Delta Z| &\leq |Z|(|\delta_u| + |\delta_v|)B. \end{aligned} \quad (2.16)$$

To estimate the expected depth calculation error, we insert the standard deviation σ_u and σ_v into the error expressions δ_u and δ_v:

$$|\delta_u| = \frac{\sigma_u}{|u|} \quad \text{and} \quad |\delta_v| = \frac{\sigma_v}{|v|}.$$

The relative error σ_Z of the estimated depth can be evaluated using the relations above and (2.16):

$$\begin{aligned} \delta_Z &= \frac{\hat{Z} - Z}{Z} = \frac{\Delta Z}{Z}, \\ |\delta_Z| &\leq \left(\frac{\sigma_u}{|u|} + \frac{\sigma_v}{|v|}\right)B \approx \frac{\sigma_u}{|u|} + \frac{\sigma_v}{|v|}. \end{aligned} \quad (2.17)$$

(2.17) gives a rough evaluation of the accuracy of the estimated depth in (2.10). Sampling n times, during each peer of the camera, and computing the average v_1 according to $v_1 = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} \overset{\leftrightarrow}{v}_{1i}$, where $\overset{\leftrightarrow}{v}_{1i}$ is sorted v_{1n}, greatly improves the accuracy of the estimated depth.

In this part we will investigate the expected accuracy of the proposed depth estimation method. This accuracy evaluation is based on the algorithm implementation scheme described before.

At first, we will concentrate on the imaging factor $\hat{v}$ of the algorithm's expected inaccuracy. The inaccuracy of this term is caused by several factors: the image function $I(\rho, t)$ is given on a discrete grid and not on a continuous one, and in addition, the values of the function $I(\rho, t)$ are quantized and noisy.

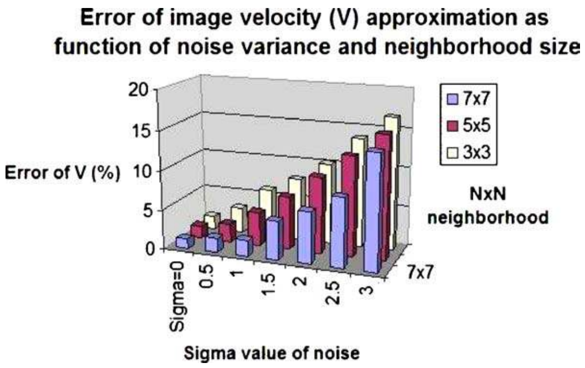

Figure 8. The simulation-based errors of the depth estimation as a function of the noise variance σ^2 and neighborhood size used for the approximation of $\hat{v}$. The first column $\sigma = 0$ shows the error values caused by quantization only.

We use synthetic image sequences for the estimation of the quantization and noise influence. The generated images were degraded with a Gaussian additive noise with mean $\mu = 0$ and variance σ^2, with the noise values at two different pixels being independent. The $\hat{v}$ approximations of the degraded data were calculated and compared to the ideal $\hat{v}$ values (without noise and quantization). The mean square value of the error was calculated for the different values of noise variance σ^2 and neighborhood size. The received values of the approximation errors of $\hat{v}$ are shown in Fig. 8. This table demonstrates the approximation accuracy as a function of noise variance σ^2 and neighborhood size used for $\hat{v}$ approximations. The error values shown in the columns corresponding to $\sigma^2 = 0$ are approximation errors caused by quantization only.

From Fig. 8-[graph series] it follows that for a camera characterized by 1% noise (~2.0 grey level per pixel), which is typical for real cameras, the accuracy of the depth estimation is about 2%.

The inaccuracy in the observer velocities $\hat{u}$ is due to the speed of the camera not being constant. This value is determined by counting the number of frames between the border frames (frames with no motion detected). There could be an error of 2 frames from a total of 66 frames in this procedure. Using this calibration procedure a number of times and applying simple averaging on the noisy results, the estimated error of the observer is about 1%. From the above calculations it follows that one can expect a depth estimation error of about 3% from the actual camera setup. This expectation is consistent with the errors in the

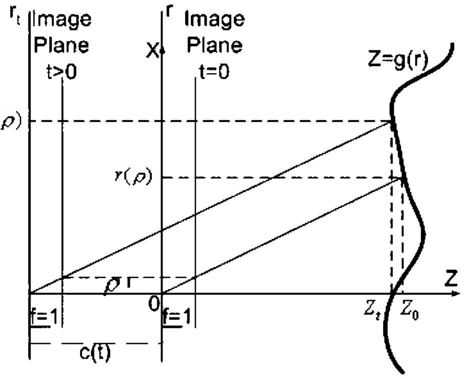

Figure 9. The pigeon head camera model. The surface, whose cross section is given by $z = g(r)$, is viewed by a camera whose focal point moves along its optical axis Z. ρ is the height on the image plane at which a feature is projected, and r and r_t are the heights at which the point observed at height ρ on the image plane are located on the object ($r_0 \neq r_t$). 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 $t > 0$.

depth estimation values obtained from the algorithm's execution.

3. The Pigeon Head Camera

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

3.1. The Model

Figure 9 illustrates what is going on. 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.

We start from the following relationship, where f is the focal length of the camera:

$$\frac{\rho}{f} = \frac{r}{z}.$$

When the camera is in its initial position $\tau = 0$,

$$\rho = f\frac{r_0}{z_0}.$$

At time $\tau = t$, when the camera is displaced along its optical axis according to the function $c(t)$ we have

$$\rho = f\frac{r_t}{z_t + c(t)},$$

in the same coordinate system, whence

$$r_t = \frac{1}{f}\rho(z_t + c(t)) = \frac{1}{f}\rho(g(r_t) + c(t)). \tag{3.1}$$

In the most general case we define the inverse function as:

$$r_t = h_c(\rho). \tag{3.2}$$

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,

$$\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}} = \frac{\frac{\partial F(r_t)}{\partial \rho}}{\frac{\partial F(r_t)}{\partial t}} = \frac{F'(h_c(\rho))\frac{\partial h_c}{\partial \rho}}{F'(h_c(\rho))\frac{\partial h_c}{\partial t}} = \frac{\frac{\partial h_c}{\partial \rho}}{\frac{\partial h_c}{\partial t}}. \tag{3.3}$$

To evaluate (3.3) we combine (3.1) and (3.2) to obtain

$$h_c(\rho) - \frac{1}{f}\rho g(h_c(\rho)) = \frac{1}{f}\rho c.$$

Differentiating with respect to ρ and t yields

$$\frac{\partial h_c(\rho)}{\partial \rho} - \frac{1}{f}\rho g'(h_c(\rho))\frac{\partial h_c(\rho)}{\partial \rho} - \frac{1}{f}g(h_c(\rho)) = \frac{1}{f}c,$$
$$\frac{\partial h_c(\rho)}{\partial \rho}\left[1 - \frac{1}{f}\rho g'(r_t)\right] - \frac{1}{f}g(r_t) = \frac{1}{f}c,$$

and

$$\frac{\partial h_c(\rho)}{\partial \rho} = \frac{\frac{1}{f}[c + g(r_t)]}{1 - \frac{1}{f}\rho g'(r_t)}; \tag{3.4}$$
$$\frac{\partial h_c(\rho)}{\partial t} - \frac{1}{f}\rho g'(h_c(\rho))\frac{\partial h_c(\rho)}{\partial t} = \frac{1}{f}\rho\frac{dc}{dt},$$
$$\frac{\partial h_c(\rho)}{\partial t}\left[1 - \frac{1}{f}\rho g'(r_t)\right] = \frac{1}{f}\rho\frac{dc}{dt},$$

and

$$\frac{\partial h_c(\rho)}{\partial t} = \frac{\frac{1}{f}\rho\frac{dc}{dt}}{1 - \frac{1}{f}\rho g'(r_t)}. \tag{3.5}$$

From (3.3), (3.4), and (3.5) we obtain

$$\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}}=\frac{\frac{\partial h_c}{\partial \rho}}{\frac{\partial h_c}{\partial t}} = \frac{\frac{\frac{1}{f}[c+g(r_t)]}{1-\frac{1}{f}\rho g'(r_t)}}{\frac{\frac{1}{f}\rho\frac{dc}{dt}}{1-\frac{1}{f}\rho g'(r_t)}} = \frac{c+g(r_t)}{\rho\frac{dc}{dt}}, \tag{3.6}$$

$$g(r_t)=z_t = \rho\frac{dc}{dt}\frac{\frac{\partial h_c}{\partial \rho}}{\frac{\partial h_c}{\partial t}} - c = \rho\frac{dc}{dt}\frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}} - c. \tag{3.7}$$

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.

3.2. *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.

The video output signal of the camera is connected to the miniature wireless video RF transmitter, which remotely broadcasts a video signal enabling autonomic usage of the device on the mobile robot. Both the camera and transmitter are operated from a single 9 V battery. The total size of the platform with camera and transmitter is [10 cm × 5 cm × 2 cm].

The video signal then received by the RF video receiver that is connected to the PCI frame grabber which is located inside a Dual Pentium III workstation, which performs the image processing of all the incoming frames. Optionally the workstation could, based on the incoming image analysis, send back action commands to the remote robot, which carries the bobbing platform on board. Varying the target distance and peering velocity parameters, the performance of the system was measured. Targets were placed at various distances in front of the camera: 5, 6, 7, 8, 9, and 10 cm. Peering velocities of 1.5 cm/sec and 2 cm/sec were used.

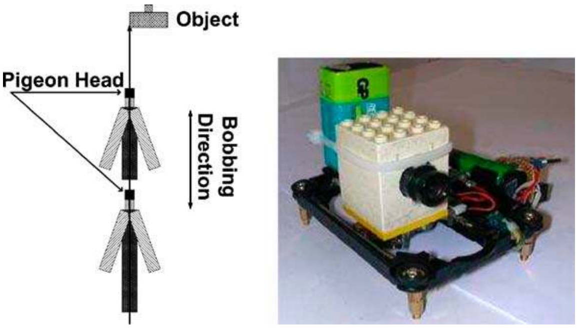

Figure 10. Scheme of Bobbing behavior of Pigeon and the implementation of the Miniature Pigeon Head camera platform, which utilizes bobbing behavior for distance estimation.

3.3. *Optic Flow and Experimental Results*

The situation regarding optic flow is the same as that in Mantis Head Camera section. That is, the partial derivative of the gray level $I(x, y)$ along the optic flow $V = (v_1, v_2) = (\frac{dx}{dt}, \frac{dy}{dt})$ is zero:

$$\frac{\partial I_t}{\partial x}\frac{dx}{dt} + \frac{\partial I_t}{\partial y}\frac{dy}{dt} + \frac{\partial I_t}{\partial t} = 0. \tag{3.8}$$

By computing the above optic flow we can estimate the distance to the objects by using Eq. (3.7).

As noted in Section 2.3, 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.

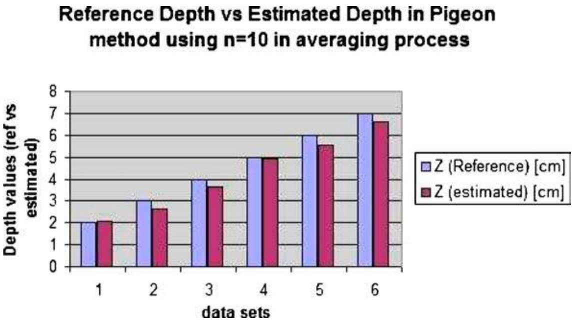

Figure 11. Reference depth versus averaged estimated depth.

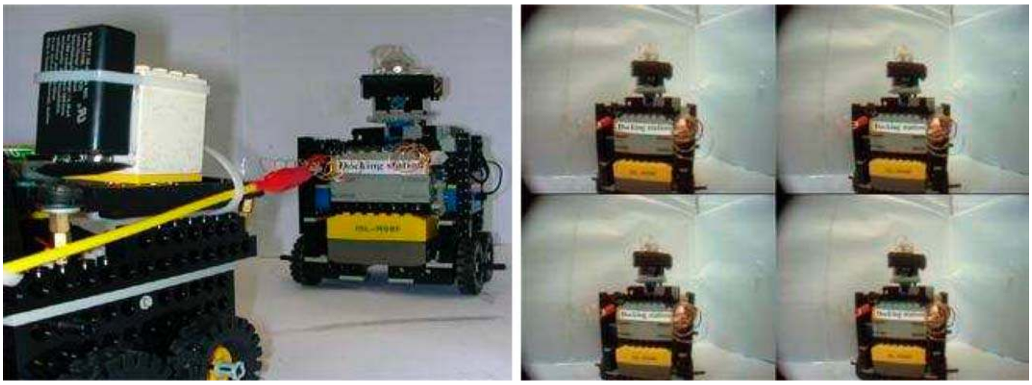

Figure 12. Tiny Lego Robot utilizes miniature Pigeon Head camera. Sequence order: top left, top right, bottom left, bottom right.

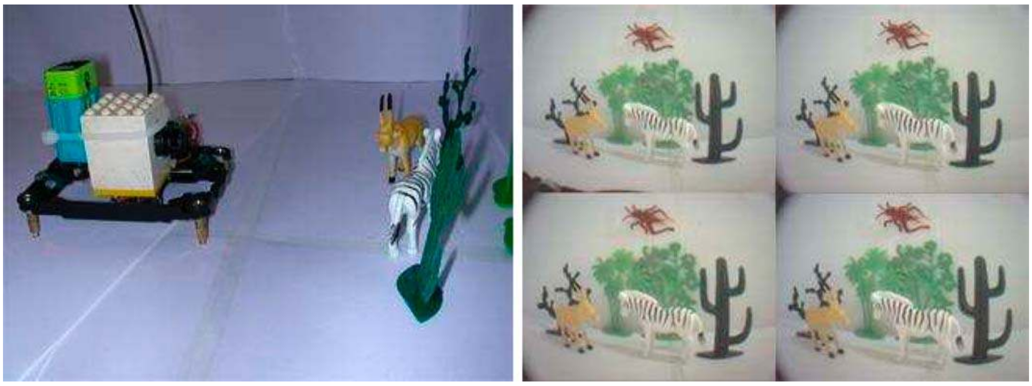

Figure 13. The subsets (of 4 samples) captured during forward bobbing of the Pigeon Head camera platform (nature scenarios). Sequence order: top left, top right, bottom left, bottom right.

3.4. *Error Analysis*

In what follows we discuss the sources of inaccuracy of the described algorithm. Then a scheme for accuracy evaluation of the estimated depth will be proposed. The algorithm presented here requires flow calculation; therefore the approximation accuracy of v in expression (3.9) has a critical role in the depth estimation accuracy. The approximate precision of v is mainly determined by noise and quantization. The influence of these two factors is investigated below. Another important factor for the depth estimation accuracy is ρ, which is the distance from the image center to the feature point. Practically, the image center point is not known, and its coordinates are determined by using a center of expansion algorithm by Lenz and Tsai (1988).

The deviation in the camera motion speed is another factor, which also has a significant impact on the algorithm accuracy. The algorithm assumes a precise mechanical setup, i.e. constant camera motion speed and precise zero value of c at time $t = 0$. However, these assumptions were difficult to satisfy absolutely. The influence of this factor on the algorithm accuracy is also investigated.

Let $\hat{v}$ and $\hat{u}$ be the approximated retinal images and observer velocities, respectively (inaccuracy in $\hat{v}$ is caused by error in feature tracking due to feature movements and changes in their size, while inaccuracy in $\hat{u}$ is due to the speed of the camera not being constant), $\hat{\rho}$ be the approximated value of the feature point coordinate from the image center, and $\hat{c}$ be the approximated value of the camera displacement from its initial position along its optical axis at the given time. Then the absolute error in the depth estimation ΔZ can be calculated using the following equation, obtained from (3.7):

$$g(r_t) = z_t = \rho \frac{dc}{dt} \frac{\frac{\partial I}{\partial \rho}}{\frac{\partial I}{\partial t}} - c = \rho \frac{u}{v} - c, \qquad (3.9)$$

where $u = \frac{dc}{dt}$, $v = \frac{\frac{\partial I}{\partial t}}{\frac{\partial I}{\partial \rho}}$. This leads to

$$\Delta Z = \hat{Z} - Z = \left(\hat{\rho}\frac{\hat{u}}{\hat{v}} - \hat{c}\right) - \left(\rho\frac{u}{v} - c\right). \quad (3.10)$$

Let δ_u, δ_v, δ_ρ, δ_c be ther relative errors in the approximation (we assume all are small values):

$$\begin{bmatrix} \delta_u \\ \delta_v \\ \delta_\rho \\ \delta_c \end{bmatrix} \triangleq \begin{bmatrix} \frac{\hat{u}-u}{u} \\ \frac{\hat{v}-v}{v} \\ \frac{\hat{\rho}-\rho}{\rho} \\ \frac{\hat{c}-c}{c} \end{bmatrix}. \quad (3.11)$$

Then,

$$\begin{aligned} \hat{u} &= u\delta_u = (1+\delta_u)u \\ \hat{v} &= v\delta_v = (1+\delta_v)v \\ \hat{\rho} &= \rho\delta_\rho = (1+\delta_\rho)\rho \\ \hat{c} &= c\delta_c = (1+\delta_c)c. \end{aligned}$$

Using the above notation and (3.10) we obtain

$$\Delta Z = \rho(1+\delta_\rho)\frac{u(1+\delta_u)}{v(1+\delta_v)} - c(1+\delta_c) - \rho\frac{u}{v} + c,$$

which can be simplified to:

$$\begin{aligned} \Delta Z &= \rho\frac{u}{v}\left[(1+\delta_\rho)\frac{1+\delta_u}{1+\delta_v} - 1\right] - c\delta_c, \quad (3.12) \\ \Delta Z &= \rho\frac{u}{v}\frac{(\delta_\rho + \delta_u - \delta_v + \delta_\rho\delta_u)}{1+\delta_v} - c\delta_c. \quad (3.13) \end{aligned}$$

If we assume δ_u, δ_v, δ_ρ, δ_c are all small compared to 1, then letting the expression $A \lesssim B$ denote "A is less than or equal to a quantity approximately equal to B", we have

$$\begin{aligned} |\Delta Z| &\lesssim \rho\frac{|u|}{|v|}(|\delta_\rho| + |\delta_u| + |\delta_v|) + c|\delta_c| \\ &= \left(\rho\frac{|u|}{|v|} - c\right)(|\delta_\rho| + |\delta_u| + |\delta_v|) \\ &\quad + c(|\delta_\rho| + |\delta_u| + |\delta_v| + |\delta_c|). \quad (3.14) \end{aligned}$$

To estimate the expected depth calculation error, we insert the standard deviations σ_u, σ_v, σ_ρ and σ_c into the error expressions δ_u, δ_v, δ_ρ and δ_c: $\delta_u \approx \frac{\sigma_u}{|u|}$, $\delta_v \approx \frac{\sigma_u}{|u|}$, $\delta_\rho \approx \frac{\sigma_\rho}{|\rho|}$ and $\delta_c \approx \frac{\sigma_c}{|c|}$. The relative error δ_Z of the estimated depth can be evaluated using the relations above and (3.14):

$$\begin{aligned} |\delta_Z| &= \frac{|\hat{Z} - Z|}{Z} = \frac{|\Delta Z|}{Z} \\ &\le \frac{\left(\rho\frac{|u|}{|v|} - c\right)(|\delta_\rho| + |\delta_u| + |\delta_v|) + c(|\delta_\rho| + |\delta_u| + |\delta_v| + |\delta_c|)}{\rho\frac{|u|}{|v|} - c}. \end{aligned}$$

Also, using the condition that the estimated depth is at least twice the bobbing path length of the platform ($\rho\, u/v \ge 2c$) we obtain

$$\begin{aligned} |\delta_Z| &\le (|\delta_\rho| + |\delta_u| + |\delta_v|) + (|\delta_\rho| + |\delta_u| + |\delta_v| + |\delta_c|) \\ &\le 2(|\delta_\rho| + |\delta_u| + |\delta_v|) + |\delta_c|, \quad (3.15) \\ |\delta_Z| &\lesssim 2\left(\frac{\sigma_u}{|u|} + \frac{\sigma_v}{|v|} + \frac{\sigma_\rho}{\rho}\right) + \frac{\sigma_c}{c}. \end{aligned}$$

(3.15) gives a rough evaluation of the accuracy of the estimated depth.

Sampling n times during each bobbing movement of the camera, and computing the average z according to $z = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} \overset{\leftrightarrow}{z_i}$ improves the accuracy of the estimated depth. Here, $\overset{\leftrightarrow}{z_n}$ is a sequence of the sorted values of z_n, and only $(n - 2k)$—the less noisy median values—are used in computing the average.

In this part we will investigate the expected accuracy of the proposed depth estimation method. This accuracy evaluation is based on the algorithm implementation scheme described before.

At first, we will concentrate on the imaging factor $\hat{v}$ of the algorithm's expected inaccuracy. The inaccuracy of this factor is caused by several factors. First, the image function $I(\rho, t)$ is given on a discrete grid and not on a continuous one. Furthermore, the values of function $I(\rho, t)$ are quantized and noisy.

We use synthetic image sequences for the estimation of the quantization and noise influence. The generated images were degraded with a Gaussian additive noise with mean $\mu = 0$ and variance σ^2, with the noise values at two different pixels being independent. The $\hat{v}$ approximations of the degraded data were calculated and compared to the ideal $\hat{v}$ values (without noise and quantization). The mean square value of the

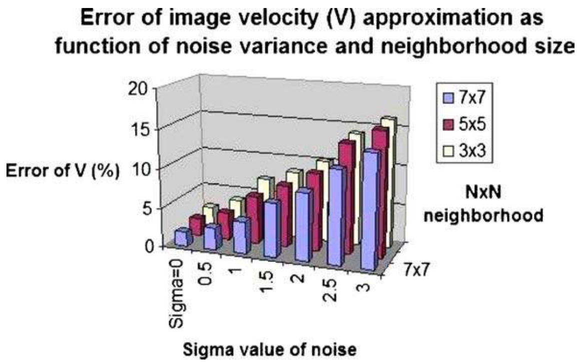

Figure 14. The simulation-based errors of the depth estimation as a function of the noise variance σ^2 and neighborhood size used for the approximation of $\hat{v}$. The first column $\sigma = 0$ shows the error values caused by quantization only.

error was calculated for the different values of noise variance σ^2 and neighborhood size. The received values of the approximation errors of $\hat{v}$ are shown in Fig. 14. This table illustrates the approximation accuracy as a function of noise variance σ^2 and neighborhood size used for the $\hat{v}$ approximations. The error values shown in the columns corresponding to $\sigma^2 = 0$ are approximation errors caused by quantization only.

From Fig. 14-[graph series] it follows that for a camera characterized by a 1% noise (~2.0 grey level per pixel), which is typical for real cameras, the accuracy of the depth estimation is about 4%.

Now, we will concentrate on the non-imaging factors $\hat{u}$, $\hat{\rho}$, $\hat{c}$ of the algorithm's expected inaccuracy. The inaccuracy of these terms is caused by several factors:

- The inaccuracy in the observer velocities $\hat{u}$ is due to the speed of the camera not being constant. This value is determined by counting the number of frames between the border frames (frames with no motion detected). There could be an error of 2 frames from a total of 66 frames in this procedure, which is about 3% error.
- The inaccuracy in the feature point coordinates from the image center $\hat{\rho}$ is bounded by about 1.25% at the used range of values of coordinates. This is one pixel error from a minimum range of 80 pixels.
- The inaccuracy in the camera displacement from its initial position along its optical axis at the given time is bounded by about 2.8%. This is about 1 millimeter of possible initial displacement out of a total range of 37 millimeters.

From the above calculations it follows that one cannot expect a depth estimation error less than 11% from the actual camera setup. This expectation is consistent with the errors in the depth estimation values obtained from the algorithm's execution.

4. Comparison of the Two Models and Some Variations

4.1. *Comparison of Mantis and Pigeon Methods*

The precision of the Mantis Head model seems to be much higher than that of Pigeon Head model. The mathematical explanation of this difference is provided next. The precision of the Mantis Head model obtained from raw experiments is on the order of single digit percents, while that of the Pigeon Head model is around ten percent. This difference is caused by the differences in mechanical and algorithmic parts in the compared methods which are discussed later on in terms of particular error factors contributing to the total error. As obtained in Sections 2.4 and 3.4 the total error in both of the methods can be roughly estimated by the following expressions:

$$|\delta_Z| \lesssim \frac{\sigma_u}{|u|} + \frac{\sigma_v}{|v|} \quad \text{—for the Mantis Head model.}$$

$$|\delta_Z| \lesssim \frac{\sigma_u}{|u|} + \frac{\sigma_v}{|v|} + \frac{\sigma_\rho}{|\rho|} + \frac{\sigma_c}{|c|} \quad \text{—for the Pigeon Head model.}$$

It is easy to see that in the second expression (for the Pigeon) there are two factors that are missing in the first expression (for the Mantis). This of course

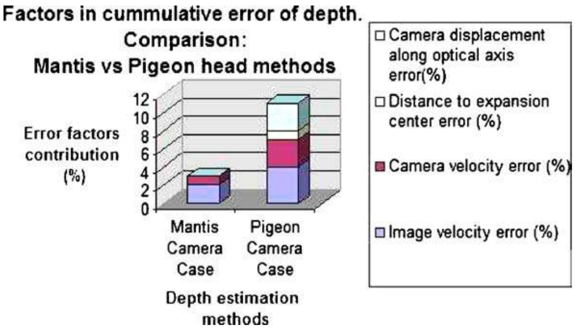

Figure 15. Cumulative error of depth.

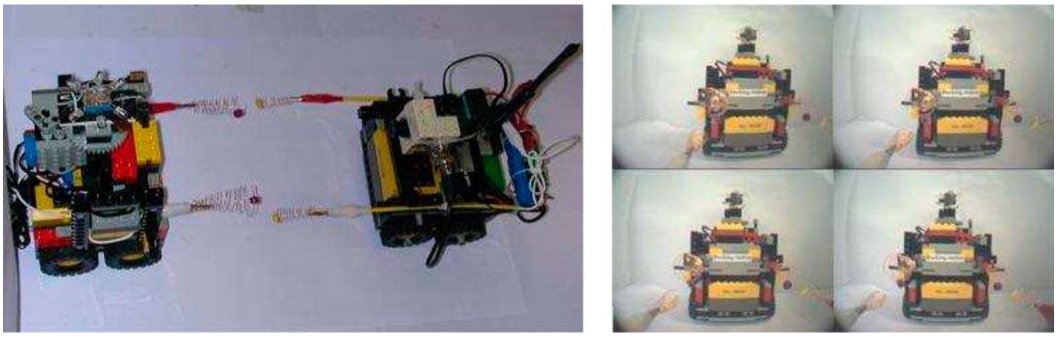

Figure 16. Tiny Lego Robot performs docking to base station utilizing peering method. Sequence order: top left, top right, bottom left, bottom right.

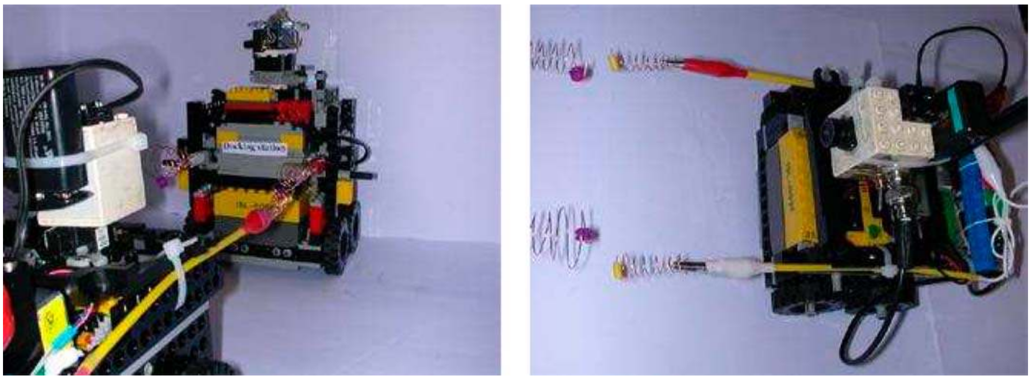

Figure 17. Tiny Lego Robot performs docking to base station utilizing peering method—another view.

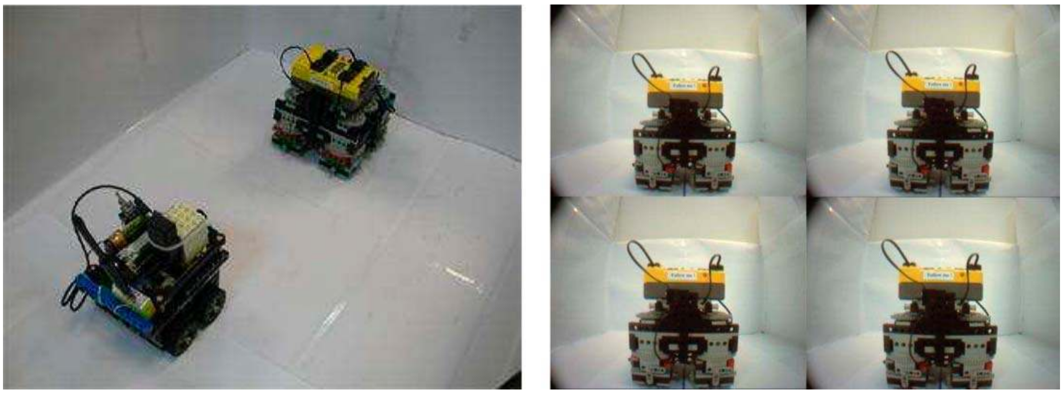

Figure 18. Tiny Lego Robots perform convoy maneuver utilizing bobbing method. Sequence order: top left, top right, bottom left, bottom right.

means that assuming similar values of the coexisting factors, the second expression is a bigger total error. Actually, as discussed in Sections 2 and 3, the specific factors in the Pigeon model are also slightly bigger that those also appearing in the Mantis model. The following graph compares the total (cumulative) error in both methods and the relative contribution of various factors that influence the total error.

4.2. *When to Use What* (*or Advantages and Disadvantages of the Proposed Methods*)

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. Praying Mantis uses a head-peering method during the hunting for static prey from ambush locations, while Pigeon uses its method during walking, performing head-bobbing with hold and thrust phases.

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 required. This implies that the selection of usage of the depth estimation method should be task dependent.

Here we may follow the purposive vision approach (Aloimonos, 1993, "Active perception") that does not consider vision in isolation, but as part of a complex system, which interacts in a specific way with the world. Accordingly, the visual categories a visual system uses, and consequently the algorithms it needs to develop or learn in order to derive them, depend totally on two things:

(a) the characteristics of the system itself (its physiology, its mobility—is it flying, crawling, walking, etc.—and its computational capacity); and
(b) the tasks it needs to accomplish.

In what follows we discuss various experiments that 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, for precise docking of the robot to its base station, the peering method should be used. This way, the robot could slowly approach the station, then pause, measure the distance to the station and, based on this measurement, perform the next approaching step. The closer the robot is to its base station, the finer the approaching steps should be performed, taking into account the possible inaccuracy of the depth estimation method.

Another example, illustrated in Fig. 18, is performing simple robot convoy of a few robots. Here the bobbing method might be used. This way, when the leading robot pauses, the following robots could roughly estimate the distance to the leader, approach to some extent and pause in turn, signaling to the next following robot to start its part of the maneuver. When the last robot in the chain has finished its part, it could signal to the leader to continue to the next step of the chain maneuver. Here the rough estimate of the distance will be enough to successfully perform this task.

In the example illustrated in Fig. 19, some robot locates the nearest robot among the others. Here only ordinal depth estimation will be enough to successfully perform this task.

Another improvement that could be made based on the previous analysis is to use a hybrid (or mix) of the methods. For example, if the peering were to be done in a diagonal direction in the XY plane, the two-dimensional information could be used to obtain higher precision of the depth estimation. This hybrid could perform better in a multi-directional pattern environment, which improves the robustness of the model. The possible variations of this are discussed later on.

4.3. *Using Compound Movements*

Some insects, such as Empusid Mantid shown in Fig. 20, demonstrate peering movements that are not simply sideways translational (Kral and Devetak, 1999). Field observations indicate that when climbing among the branches of shrubs and jumping from one branch to another, the insects use these complex head movements to estimate the distance to the nearest and most readily grasped object or landing target. The complexity of the peering movements might be directly dependent on the complex structure of the surroundings. Changing the peering axis, distance information concerning the objects in a variety of directions could be obtained without the need of turning.

This motion pattern motivated us to research the option of diagonal peering. Diagonal peering is the combination of horizontal and vertical peering motions. It could be used both to improve the precision of the depth estimation since the retinal velocity information is available from the X and Y directions,

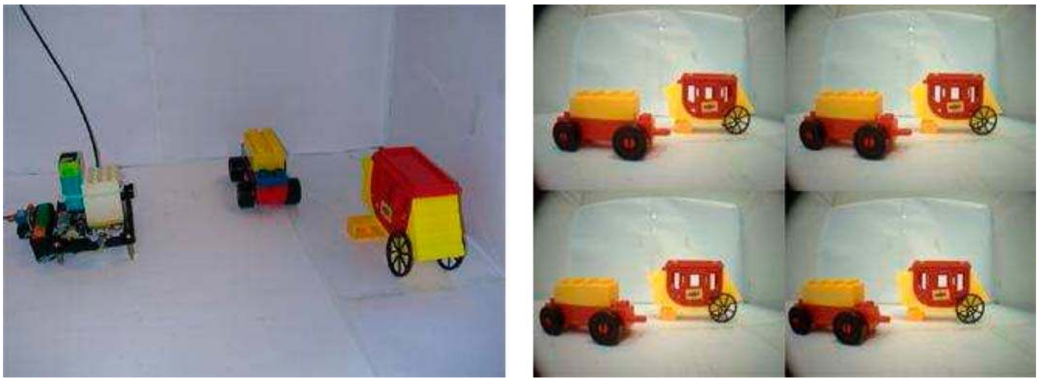

Figure 19. Tiny Lego Robot performs relative distance estimation by utilizing peering method. Sequence order: top left, top right, bottom left, bottom right.

Figure 20. The Mantis-Empusa.

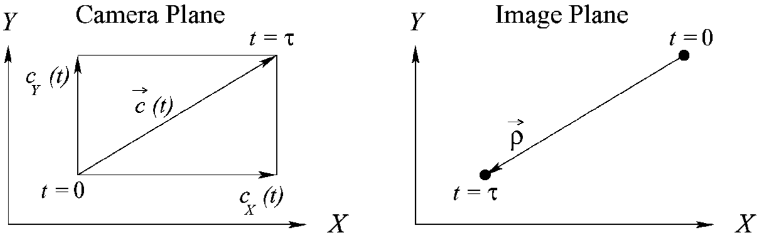

Figure 21. Diagonal Mantis head camera model. The surface, whose cross section is given by Z, is viewed by a camera with focal point moving along a diagonal in the XY-plane. $\vec{\rho}$ denotes the displacement along that diagonal on the image of the projection of the feature point located on the surface of the object to the image plane.

and to optimize the performance of the algorithm in surroundings consistent with those of vertically and horizontally directed patterns. Ideally, the parameters of such motion combinations should be configurable. In some cases when the surrounding patterns are co-directed, just the X or the Y motion direction should be used. When the directions of the surrounding patterns are multiple and mixed, both the X and Y motion

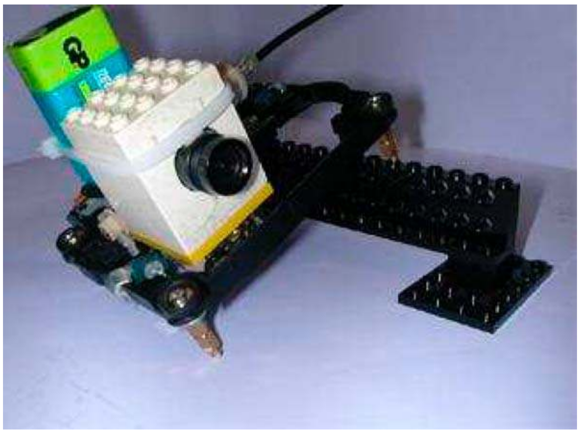

Figure 22. Scheme of the implementation of the miniature Mantis Head camera platform, which utilizes the diagonal peering behavior for distance estimation.

directions should be used. The amount of motion in each direction should be also derived from the analysis of the surrounding patterns for best performance and set dynamically.

In what follows we analyze the pure diagonal motion along the X-Y direction. The camera moves right and up and then left and down (pure translation) along the X-Y diagonal direction according to the function $\vec{c} = \vec{c}(t)$. We use notation similar to that of Section 2. Typically this motion is with constant speed (and changing direction at the edges of the platform) such as $\vec{c}(\tau) = s\vec{V}_0\tau$, where s is 1 or -1 depending on the peering direction.

We apply (2.7) separately to the X and Y directions of the compound motion.

$$z_{tX} = f\frac{dc_X}{dt}\frac{\frac{\partial I_X}{\partial \rho_X}}{\frac{\partial I_X}{\partial t}}, \quad z_{tY} = f\frac{dc_Y}{dt}\frac{\frac{\partial I_Y}{\partial \rho_Y}}{\frac{\partial I_Y}{\partial t}},$$

where indices X and Y indicate the components of the vectors projected onto the X and Y axes. In this expression, $c_K = c_K(t)$ and dc_K/dt are given, while $\partial I_K/\partial\rho_K$ and $\partial I_K/\partial\rho_K$ are determined by observation for $K = X, Y$.

A miniature video camera was mounted on a specially designed micro-translation platform, which provides precise periodic diagonal peering movements of the camera in the XY plane 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 compound two-dimensional peering behavior of the praying mantis.

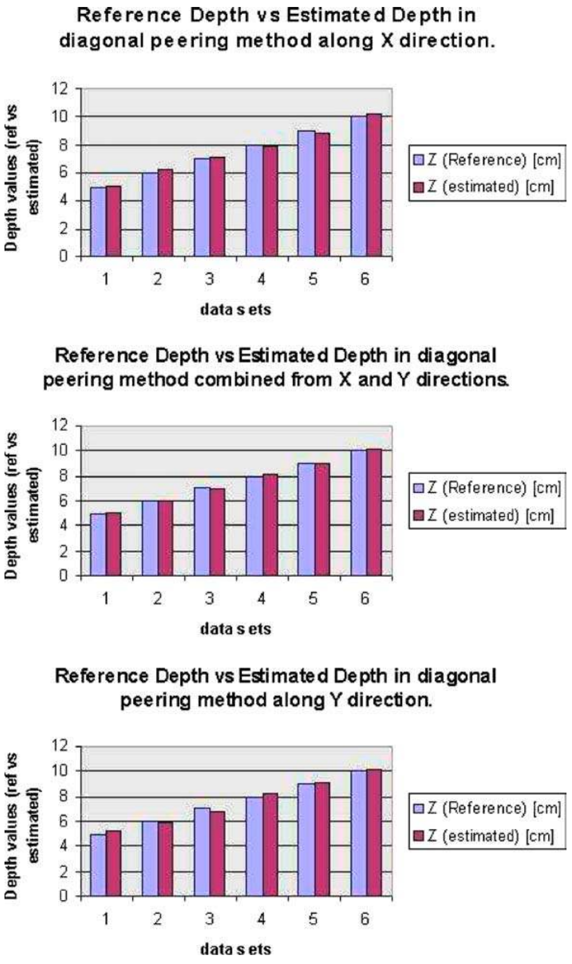

Figure 23. Reference distance versus averaged estimated distance in X (topmost) and Y (middle) directions separately and their combination (at the bottom).

In a similar way as was described in Section 2.3 we obtained two sets of the depth estimation results (one each for the X and Y directions).

The resulting depth value Z can be derived from Z_X and Z_Y using the expression $Z = \alpha Z_X + (1 - \alpha)Z_Y$, where α is the weight coefficient of the depth components and is determined by the overall pattern gradient direction in the picture. If all patterns are horizontal then $\alpha = 0$, and if they are all vertical then $\alpha = 1$. In the case where the angle of the peering diagonal is dynamically configurable, it should be set perpendicular to the dominant orientation of the gradient of the pattern in order to optimize the precision of the depth estimation.

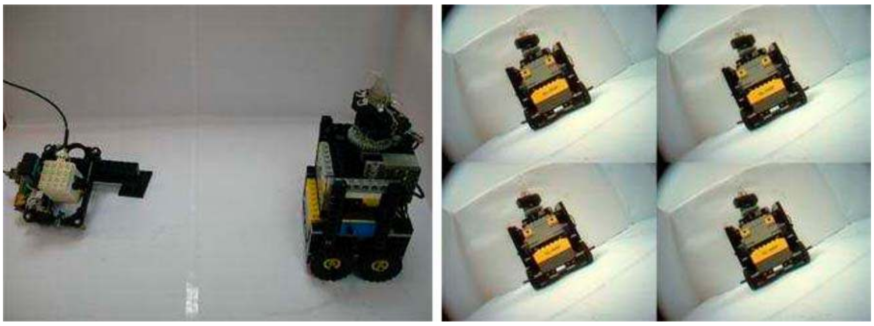

Figure 24. Miniature Mantis Head Camera performing diagonal peering movements for depth estimation. Sequence order: top left, top right, bottom left, bottom right.

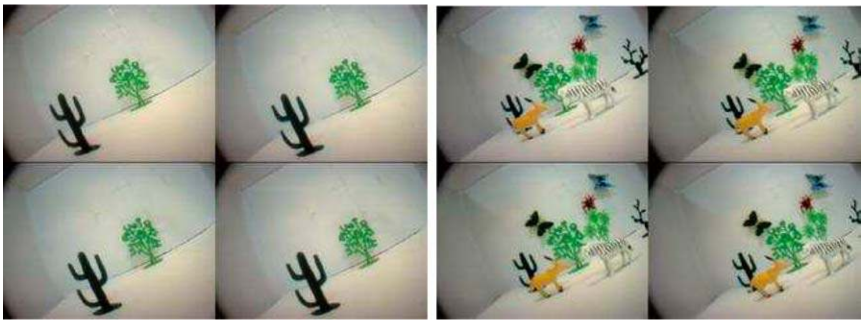

Figure 25. The subsets (of 4 samples) captured during diagonal peering movements of the Mantis Head camera platform (Nature scenarios). Sequence order: top left, top right, bottom left, bottom right.

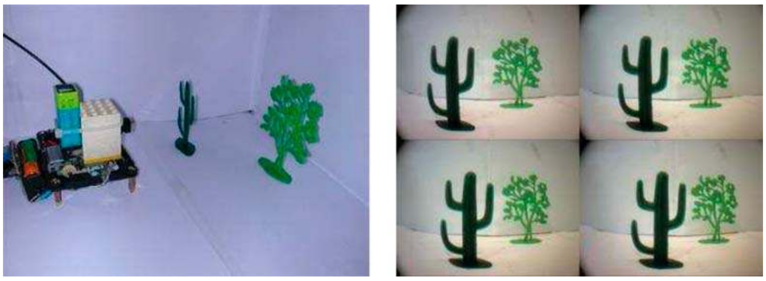

Figure 26. The subsets (of 4 samples) captured during peering movements of the Mantis Head camera platform for ordinal depth estimation (Nature scenarios).

4.4. Ordinal Depth Estimation

From biological studies it follows that animals often use relative motion parallax for depth perception (Sobel, 1990; Kral and Poteser, 1997). 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 (2.7). If the object with depth Z_1 is closer that the one with depth Z_2, the sign of the expression above is negative.

5. 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 (Poteser and Kral, 1995; Yamawaki, 2000a, 2000b).

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 their attractiveness for usage with mobile agents.

The miniature and extremely light mechanical, optical and electronic implementations of the models were presented. These properties make it possible to install them on top of a small mobile robot or smaller vehicle and use them to obtain depth information of points of interest in the surrounding space.

The precision of the depth estimations, achieved by the models and their implementations, are consistent with both the results of error analysis and those 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 by Arkin et al. We also plan to implement some of the real-time indoor navigation algorithms (Lumelsky et al., 1990, Kamon et al., 1998, Kral et al., 1998), 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. The use of different mechanisms for depth perception based on various types of motion is widespread throughout different types of animals. Some of the mechanisms are well studied by biological experiments but some are more difficult for biological experimental study—they remain unanswered and exist as hypotheses. The robotics experiments could prove the validity of some of these hypotheses.

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, provided an error analysis of the algorithm and its high precision, which is consistent with that of the praying mantis and the pigeon, and presented directions for future work.

Acknowledgments

The authors acknowledge the useful input and motivation during the conversations with entomologist Meir Broza and biologist Gadi Katzir from Oranim University. The authors also acknowledge the support of ISL Lab, CS Department, Technion and help of Lab engineers Kuty Pekar, Eyal Gordon and Ronen Keidar during the different stages of the model implementation.

Note

[1] Several amplitude-based experiments with Scorebot robot in our Lab were conducted as well. Due to mechanical restrictions of the configuration the results were significantly less precise compared to those obtained from the custom-built miniature camera motion platform.

References

Ali, K.S. and Arkin, R.C. 1998. Implementing schema-theoretic models of animal behavior in robotic systems. In *5th International Workshop on Advanced Motion Control—AMC '98*. Coimbra, Portugal, pp. 246–254.

Aloimonos, Y. 1993. Active vision revisited. In *Active Perception*, Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Argyros, A.A., Bekris, K.E., and Orphanoudakis, S.C. 2001. Robot homing based on corner tracking in a sequence of panoramic images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, Vol. 2. Kauai, Hawaii, pp. 3–10.

Arkin, R.C., Ali, K.S., Weitzenfeld, A., and Cervantes-Perez, F. 2000. Behavioral models of the praying mantis as a basis for robotic behavior. *Journal of Robotics and Autonomous Systems*, 32(1):39–60.

Barron, J.L., Fleet, D.J., and Beauchemin, S.S. 1994. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77.

Chameron, S., Beugnon, G., Schatz, B., and Collett, T.S. 1999. The use of path integration to guide route learning in ants. *Nature*, 399:769–772.

Chaumette, F., Boukir, S., Bouthemi, P., and Juvin, D. 1996. Structure from controlled motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):492–504.

Collett, T.S. 1996. Insect navigation en route to the goal: Multiple strategies for the use of landmarks. *Journal of Experimental Biology*, 199:227–235.

Collett, T.S. and Rees, J.A. 1997. View-based navigation in hymenoptera: Multiple strategies of landmark guidance in the approach to a feeder. *J. Comp. Physiol. A*, 181:47–58.

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

Davies, M.N.O. and Green, P.R. 1988. Head-bobbing during walking, running and flying: Relative motion perception in the pigeon. *Journal of Experimental Biology*, 138:71–91.

Davies, M.N.O. and Green, P.R. 1990. Optic flow-field variables trigger landing in hawk but not in pigeons. *Naturwissenschaften*, 77:142–144.

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

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

Dunlap, K. and Mowrer, O.H. 1930. Head movements and eye functions of birds. *J. Comp. Psychol.*, 11:99–113.

Fah, L. and Xiang, T. 2001. Characterizing depth distortion under different generic motions. *International Journal of Computer Vision*, 44(3):199–217.

Frost, B.J. 1978. The optokinetic basis of head-bobbing in the pigeon. *Journal of Experimental Biology*, 74:187–195.

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

Kamon, I., Rimon, E., and Rivlin, E. 1998. Tangent bug: A range-Sensor-Based navigation algorithm. *International Journal of Robotic Research*, 17(9):934–953.

Kamon, I. and Rivlin, E. 1997. Sensor based motion planning with global proofs. *IEEE Transactions on Robotics and Automation*, 13(6):814–822.

Kirschfeld, K. 1976. The resolution of lens and compound eyes. In F. Zettler and R. Weiler (eds.), *Neural Principles in Vision*, Springer: Berlin, pp. 354–370.

Köck, A., Jakobs, A.-K., and Kral, K. 1993. Visual prey discrimination in monocular and binocular praying mantis Tenodera sinensis during postembryonic development. *J. Insect Physiol.*, 39:485–491.

Kral, K. 1998. Side-to-side head movements to obtain motion depth cues: A short review of research on the praying mantis. *Behavioural Processes*, 43:71–77.

Kral, K. 1999. Binocular vision and distance estimation. In F.R. Prete, H. Wells, P.H. Wells, and L.E. Hurd (eds.), *The Praying Mantids: Research Perspectives*, Johns Hopkins University Press: Baltimore, pp. 114–140.

Kral, K. and Devetak, D. 1999. The visual orientation strategies of Mantis religiosa and Empusa fasciata reflect differences in the structure of their visual surroundings. *J. Insect Behav.*, 12:737–752.

Kral, K. and Poteser, M. 1997. Motion parallax as a source of distance information in locusts and mantids. *J. Insect Behav.*, 10:145–163.

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

Land, M.F. 1999. Motion and vision: Why animals move their eyes. *J. Comp Physiol. A*, 185:341–352.

Lenz, R.K. and Tsai, R.Y. 1988. Techniques of calibration of the scale factor and image center for high accuracy 3-D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:713–720.

Lewis, M.A. and Nelson, M.E. 1998. Look before you Leap: Peering behaviour for depth perception. In R. Pfeifer, B. Blumberg, J.-A. Meyer, and S.W. Wilson (eds.), *From Animals to Animats 5: Proceedings of the Fifth Internatoinal Conference on Simulation of Adaptive Behavior*, MIT Press, pp. 98–103.

Lumelsky, V. and Skewis, T. 1990. Incorporating range sensing in the robot navigation function. *IEEE Transactions on Systems, Man and Cybernetics*, 20(5):1058–1069.

Martin, G.R. and Katzir, G. 1995. Visual fields in ostriches. *Nature*, 374:19–20.

Martin, G.R. and Katzir, G. 1999. Visual fields in short-toed eagles, *Circaetus gallicus* (Accipitridae), and the function of binocularity in birds. *Brain, Behavior and Evolution*, 53:55–56.

Martinoya, C., Le Houezec, J., and Bloch, S. 1988. Depth resolution in the pigeon. *J. Comp. Physiol. A*, 163:33–42.

McFadden, S.A. 1993. Constructing the three-dimensional image. In H.P. Zeigler and H.-J. Bischof (eds.), *Vision, Brain and Behavior in Birds*, MIT Press: Cambridge, MA, pp. 47–61.

McFadden, S.A. 1994. Binocular depth perception. In M.N.O. Davies and P.R.G. Green (eds.), *Perception and Motor Control in Birds, an Ecological Approach*, Springer-Verlag: Berlin, pp. 5–34.

Möller, R. and Lambrinos, D. 1997. Modelling the landmark navigation behavior of the desert ant Cataglyphis. Technical report, AI Lab, University of Zurich.

Möller, R., Lambrinos, D., Pfeifer, R., Labhart, T., and Wehner, R. 1998. Modeling ant navigation with an autonomous agent. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, MIT Press, pp. 185–194.

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

Nicholson, D.J., Judd, S.P.D., Cartwright, B.A., and Collett, T.S. 1999. Learning walks and landmark guidance in wood ants Formica Rufa. *Journal of Experimental Biology*, 202:1831–1838.

Poteser, M. and Kral, K. 1995. Visual distance discrimination between stationary targets in praying mantis: An index of the use of motion parallax. *Journal of Experimental Biology*, 198:2127–2137.

Poteser, M., Pabst, M.A., and Kral, K. 1998. Proprioceptive contribution to distance estimation by motion parallax in praying mantid. *Journal of Experimental Biology*, 201:1483–1491.

Sandini, G. and Tistarelli, M. 1990. Active tracking strategy for monocular depth inference over multiple frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:13–27.

Sobel, E.C. 1990. The locust's use of motion parallax to measure distance. *J. Comp. Physiol. A*, 167:579–588.

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

Srinivasan, M.V., Poteser, M., and Kral, K. 1999. Motion detection in insect orientation and navigation. *Vision Res.*, 39:2749–2766.

Srinivasan, M.V., Zhang, S.W., and Bidwell, N. 1997. Visually mediated odometry in honeybees. *Journal of Experimental Biology*, 200:2513–2522.

Srinivasan, M.V., Zhang, S.W., Lehrer, M., and Collett, T.S. 1996. Honeybee navigation en route to the goal: Visual flight control and odometry. *Journal of Experimental Biology*, 199:237–244.

Troje, N.F. and Frost, B.J. 2000. Head-bobbing in pigeons: How stable is the hold phase. *Journal of Experimental Biology*, 203:935–940.

Weber, K., Venkatesh, S., and Srinivasan, M.V. 1998. An insect-based approach to robotic homing. In *ICPR98*, p. CV13.

Yamawaki, Y. 2000a. Effect of luminance, size and angular velocity on the recognition of non-locomotive prey models by the praying mantis. *Journal of Ethology*, 18(2):85–90.

Yamawaki, Y. 2000b. Saccadic tracking of a light grey target in the mantis, Tenodera aridifolia. *Journal of Insect Physiology*, 46:203–210.

Zheng, W., Kanatsugu, Y., Shishikui, Y., and Tanaka, Y. 2000. Robust depth-map estimation from image sequences with precise camera operation parameters. In *ICIP00*, p. TP08.07.