Noname manuscript No.
(will be inserted by the editor)

Quasi-Parallax for Nearly Parallel Frontal Eyes

—a possible role of binocular overlap during rapid locomotion

Loong-Fah Cheong · Zhi Gao

the date of receipt and acceptance should be inserted later

Received: date / Accepted: date

Abstract In this paper, we explore how a visual system equipped with a pair of frontally-placed eyes/cameras can rapidly estimate egomotion and depths for the task of locomotion by exploiting the eye topography. We eschew the traditional approach of motion-stereo integration, as finding stereo correspondence is a computationally expensive operation. Instead, we propose a quasi-parallax scheme by pairing appropriate visual rays together, thus obviating the need for stereo correspondence and yet being able to leverage on the redundant information present in the binocular overlap. Our model covers realistic visual systems where the two eyes might deviate from the strictly frontal-parallel configuration, and yet the results show that the advantages of the parallax-based approach are retained. In particular, it offers better disambiguation of translation and rotation over conventional two-frame structure from motion approaches, despite not having views covering diametrically opposing directions like that of spherical eyes or laterally-placed eyes. The rapid processing that such scheme entails seems to offer a more realizable and useful alternative for depth recovery during locomotion.

Keywords: Structure from Motion, Quasi-parallax, Binocular Vision

The support of the research grant NRF2007IDM-IDM002-069 from MDA of Singapore is gratefully acknowledged.

L.-F. Cheong
Electrical and Computer Engineering Department
National University of Singapore
4 Engineering Drive 3, Singapore 117576
E-mail: eleclf@nus.edu.sg

Z. Gao
Interactive and Digital Media Institute
National University of Singapore
No.21, Heng Mui Keng Terrace, Singapore 119613
E-mail: idmgz@nus.edu.sg

1 Introduction

The subject of this paper is to re-examine visual system with binocular overlap in the visual field of two eyes with a different perspective. By different perspective, we mean that the binocular overlap is not necessarily leveraged in terms of stereo matching and stereoscopic depth recovery, within which computational questions had usually been posed in the computer vision community. That is, vision may indeed be diplopic even if the visual fields of the two eyes have overlap. The motivations for such re-examination can be traced to at least three aspects.

1.1 The challenge of using stereopsis for locomotion

The rich development of feature descriptors such as SIFT (Lowe, 2004) and modern optimization techniques such as graph cut have resulted in significant increase in the performance of stereo systems (Szeliski et al., 2008). It seems reasonable to say that we have unlocked the secret of creating a successful stereo system. However, all of these successes presuppose the availability of abundant computational resources. Unless we are prepared to identify ourselves with the brute-force power of modern computer, we must re-examine the premises of computational stereo.

If we trace the origins of AI movements such as active and purposive vision, one of the dominant themes of these movements in the 90's is that vision is for serving action like locomotion in the world. Detailed, general purpose scene reconstruction might contain too much information and is too slow. Such view is also echoed by Tsotsos' argument on complexity (Tsotsos, 1998). All these arguments still ring true today. At the risk of sweeping generalization, modern computational approaches have little hope of achieving the swift flight

of say a bird in structurally complex environment such as woodland. Indeed, who can say that these modern approaches have not resulted in a form more unrealizable than any conceived in the 90's? Clearly, we are not saying that stereopsis has no role to play, but we are suggesting that it is perhaps more for attending to slower tasks and for recovering shape in the immediate space around the body (e.g. for visually guided manipulation of items held in the hands/jaws or bills), rather than for distance perception of far objects during locomotion.

1.2 An instructive look at the natural world

Stereopsis has been investigated in only a narrow range of species that share the characteristic of a relatively wide frontal binocular field produced by eyes that are typically widely spaced and forward-facing, with parallel axes and conjugate movements. However, these are significant features of the visual system in only a small proportion of extant vertebrates, most notably primates, and may be a highly specialized rather than general arrangement in the animal world.

The example of the avian world is instructive. Except for the case of owl, the presence of global stereopsis in other bird species is based upon conjecture. McFadden (1994) found some behavioral evidence of local stereopsis in pigeons; however, these results did not support the presence of global stereopsis. Short-toed Snake-eagles are diurnal predators that are described in handbooks as having forward facing eyes and a wide binocular field. However, appearances are deceptive. The functional binocular field of these birds is vertically long and relatively narrow (maximum width 20°, see Fig.1). This situation is not unique and has been found in other species including ostrich, heron and owl, and suggests that many bird species do not make full use of the potentially available binocular field. For many other birds, the frontal binocular fields are less than 10° wide and even as narrow as 5° but they are sufficient for the control of flight and landing at relatively high velocities and in structurally complex (e.g., woodland) habitats.

Martin (2007, 2009) argued that stereopsis is too slow, especially for the purpose of locomotion. Davies and Green (1994)have pointed out that stereopsis involves considerable neural processing and is too slow to control the estimation of distance and depth when a bird is landing upon a perch. McFadden (1993) pointed out that pigeons have depth perception, and are sensitive to disparities of about 1 arc min (compared to 4 s in humans), but it is doubtful that this ability is used in foraging.

Thus, for the case of birds at least, rather than try to find evidence of binocular fusion and stereopsis, Martin (2007; 2009) argued that it might be more parsimonious to consider what the function of binocularity could be if birds viewed objects diplopically within their binocular fields. Do two eyes retrieving information from almost identical flow-fields provide more information than one eye? Does it add anything beyond mere redundancy? Or does the visual system gain anything for such overlapping arrangement? The scheme put forth in our paper provides one possible way to leverage such arrangement.

Even for the case of primates, where one of the most conspicuous visual specializations is the large area of overlap between the fields of vision of the two eyes and the evident existence of stereopsis, stereo processing might still be too slow for fast locomotion (e.g. when primates execute swift arboreal movements among the canopy). The concern during fast locomotion is also likely to be more about obstacles that are further away, since the clearance away from these obstacles needs to be greater at higher speed of movements. Stereopsis is useful only for depth perception of the immediate space surrounding the body, typically for tasks such as object manipulations. As such, one might ask: if one conjectures that there is no fusion of the disparate right-eyed and left-eyed views of the scene during fast motion, how can the readily available redundance be made use of? Is it just mere greater robustness from redundance?

We hasten to add that we own to no better motive for studying an alternative scheme for leveraging such redundance other than simple curiosity about its possibility, without necessarily drawing general conclusions about the existence of such a scheme in the natural world.

1.3 Dynamic depth cues during locomotion

At this point, it may be asked: what about motion cues? The primary depth cue during locomotion is indeed motion cue. Yet due to the difficulty in structure from motion (SFM), coupled with the demand for real-time processing, so far only simple mechanisms have been proposed for such tasks in biomimetic works for locomotion, e.g. the rather minimalistic visual system consisting of elementary motion detectors and dealing with translation only (Franceschini et al., 1992). In the field of robotics, many authors have made use of the "optic flow balance" hypothesis in designing visually guided wheeled vehicles (Coombs and Roberts, 1993; Duchon and Warren, 1994; Santos-Victor et al., 1995; Dev et al., 1997; Weber et al., 1997; Carelli et al., 2002; Argyros et al., 2004; Hrabar et al., 2004; Humbert et al., 2007), or aerial vehicles (Corke et al., 2004; Griffiths et al., 2006), and simulating flying agents (Neumann and Bulthoff, 2001; Muratet et al., 2005) and hovercraft (Humbert et al., 2005). The "optic flow balance" hypothesis has been tested mainly in corridors

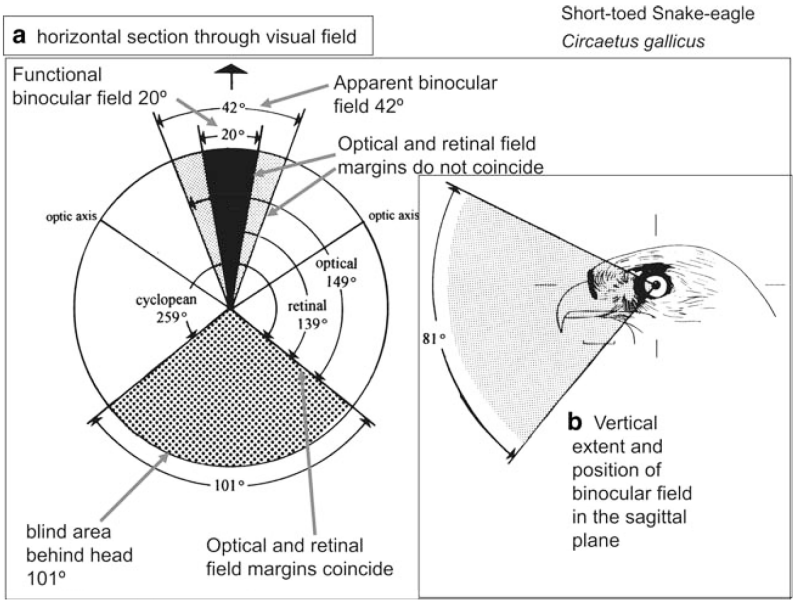


Fig. 1: Visual field of Short-toed Snake-Eagle. (from Martin (2007), courtesy of Graham R. Martin).

and urban canyons. Despite the success of the "optic flow balance" hypothesis in robotics, new behavioural experiments have shown that honeybees actually do not necessarily centre when traversing a corridor (Ruffier et al., 2007; Serres et al., 2007). They may follow one of the two walls at a certain distance. Serres et al. (2008) designed a flying agent that can shift from 'wall-following behavior' to 'centring behavior'.

Difficulties remain for negotiation of more challenging corridors including L-junctions or T-junctions, not to mention in complex scenes such as a forested environment. No one seemed too anxious to grapple with these more challenging situations using richer SFM cues. Yet, as we will show in this paper, the processing of these richer SFM cues can be made real-time by exploiting the binocular constraint afforded by the two eyes.

2 Literature Review

There has been a long history in SFM research that exploits stereo. In the earlier works of motion-stereo integration, no matter it is the mere juxtaposition of the results from independent processing of the motion and stereo information (Ayache and Faugeras, 1989; Grosso et al., 1989; Kriegman et al., 1989), where the final estimates of structure were based on some combination of the outputs of these separate processes (coupled loosely together sensus (Clark and Yuille, 1994)), or the tightly coupled approach where the processing of one type of visual information may depend on the presence of another (Balasubramanyam and Snyder, 1991; Li and Duncan, 1993; Shi et al., 1994; Waxman and Duncan, 1993; Zhang and Negahdaripour, 2008), the all but universal assumption is that the overlap in the visual field is used for computing binocular disparity. This assumption remains true in the later approaches with the advent of more sophisticated techniques such as PDE (Strecha and Gool, 2002), variational approach (Huguet and Devernay, 2007; Pons et al., 2007; Williams et al., 2005)and factorization (Ho and Chung, 2000).

Then there is another class of related works where the multiple cameras that are in simultaneous motion may not have overlap in their field of view and thus stereopsis is not possible. The general camera model (GCM) and the generalized essential matrix put forth (Pless, 2004; Kim et al., 2010) have the advantage of generality (in that it admits any arrangement of the cameras). So too in the works of Neumann (2004) except that the input to each camera is processed independently and the output of each camera is only integrated at the final stage with those of others. As mentioned, these works do not assume any binocular overlap. However, in many biological vision systems, even for those animals in which the eyes are laterally placed, there exists some degree of overlap in the visual field in the frontal direction, and it is only by some effort of imagination that one can conceive of nature not using it in some form, even though, for various reasons discussed in the preceding section, the leveraging may not be in the form of stereopsis.

Finally, there are those works which exploit the organizational possibility offered by the eye topography. In particular, by pairing visual rays from different eyes (or cameras), useful information such as heading direction can be obtained by parsimonious visual pro-

cessing (Hu and Cheong, 2009). At the heart of such approach of pairing appropriate visual rays is the idea of parallax. The traditional formulation of parallax is based on the fact that the difference in velocity between two points that are nearby in the image but at different depths is nearly independent of rotation (Hildreth, 1992; Longuet-Higgins and Pradzny, 1980; Rieger and Lawton, 1985). Canceling rotation is advantageous as it enhances translation pickup; the residual function for the heading direction has a deeper minimum than the one based on general motion recovery, leading to a minimization that is more robust to the bas-relief ambiguity and image noise in general. However, the twin hard problems of determining pairs of image features along depth boundaries and measuring their image velocities (given the interference of the boundary) plagued such approaches from its earliest beginnings. To avoid these problems, Tomasi and Shi (1993) measured the differential changes in the angles between the projection rays of pairs of point features. Given spherical field of view, Lim and Barnes (2008) measured the difference in flow for visual rays that are in the opposite direction on the image sphere. Hu and Cheong (2009) extends it to the case of a pair of laterally-placed eyes or compound eyes, where again, visual rays from the opposite directions are paired together, and the difference in their optic flows is computed. Although the resulting quantity contains a weak residual term induced by the rotation of the head, the rotation is largely removed, hence the name quasi-parallax being coined for the difference term.

In this paper, we propose a quasi-parallax approach for visual systems equipped with a pair of eyes/cameras with overlapping visual field in the frontal direction. The simplest of these systems is that of a frontally-placed pair of eyes/cameras with parallel optical axes pointing straight ahead. We also extend the formulation to the case where the optical axes may not be parallel, and this covers the important case when the eyes of an animal are divergent (in many animals, the bony sockets of the eyes are somewhat outward-pointing). As opposed to most works on such systems with overlapping visual field cited in the preceding paragraph, the crucial difference in our system is that no binocular disparities are computed, thus obviating the need for stereo correspondence and making the method particularly useful for real-time locomotion. At a more fundamental level, we show through our work that the binocular arrangement of eyes admits another possibility for exploitation, that vision may indeed be diplopic and yet we can gain important information from this arrangement. As both this work and (Hu and Cheong, 2009) are based on the notion of quasi-parallax but each with different eye configurations, it is of interest to examine whether the eye configuration will impact on how well the parallax-based methods can resolve the bas-relief ambiguity. As we will show later in the experimental section, all the advantages of parallax-based methods shown for spherical eyes or lateral eyes covering diametrically opposing directions are retained for the case of frontal eyes examined in this paper. In particular, given quasi-parallax measurements with equal quality, the performance of the frontal eyes is on par with similar parallax-based systems with spherical eyes or lateral eyes, despite not having views covering diametrically opposing directions. The crucial factor in a parallax-based scheme is the quality of the parallax measurements, not the field of view per se.

3 Basic Model for Frontally-placed Camera Pair

We start with the simplest model (Fig. 2), that of two frontally placed cameras, with their image planes coplanar and the optical axes parallel. The two cameras are mounted rigidly on a platform, each displaced an equal distance b from the platform origin O_p. The world coordinate system (WCS) is placed at the platform origin and its axes align with the axes of camera coordinate system (CCS). We will use the subscripts l and r to represent the entities associated with the left and the right cameras respectively.

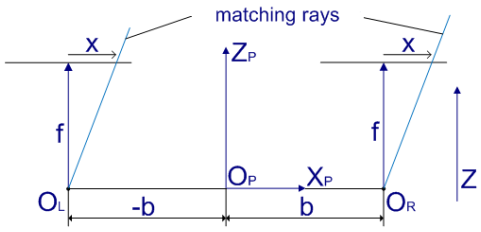


Fig. 2: Top view of a frontally-placed camera pair.

3.1 Motion and Flow Representation

With respect to the WCS, let the platform move with a translation $\boldsymbol{v}=(U,V,W)^T$ and a rotation $\boldsymbol{\omega}=(\alpha,\beta,\gamma)^T$. This induces the following right and left camera motions $(\boldsymbol{v}_r,\boldsymbol{\omega}_r)$ and $(\boldsymbol{v}_l,\boldsymbol{\omega}_l)$ in their own reference frames:

$$\begin{aligned}\boldsymbol{v}_r &= (U, V+b\gamma, W-b\beta)^T, \boldsymbol{\omega}_r = (\alpha,\beta,\gamma)^T \\ \boldsymbol{v}_l &= (U, V-b\gamma, W+b\beta)^T, \boldsymbol{\omega}_l = (\alpha,\beta,\gamma)^T\end{aligned} \tag{1}$$

whose terms differ by a sign whenever the variable b appears. For brevity of subsequent presentation, we introduce $b_r = b$ and $b_l = -b$ for the right and the left camera respectively. Thus, for the i^{th} camera, $i = r, l$, with

perspective projection model and known focal length f, the following equations relate the 3D camera motions and the optical flow (u_i,v_i) at the image point (x,y) arising from a 3D scene point with depth Z_i (Longuet-Higgins and Pradzny, 1980):

$$\begin{aligned} u_i &= \frac{(W - b_i\beta)x - fU}{Z_i} + \frac{\alpha xy}{f} - \beta(f + \frac{x^2}{f}) + \gamma y \\ &= \frac{1}{Z_i}u_i^{tr} + u_i^{rot} \\ v_i &= \frac{(W - b_i\beta)y - f(V + b_i\gamma)}{Z_i} - \frac{\beta xy}{f} + \alpha(f + \frac{y^2}{f}) - \gamma x \\ &= \frac{1}{Z_i}v_i^{tr} + v_i^{rot} \end{aligned} \tag{2}$$

where $\frac{1}{Z_i}(u_i^{tr},v_i^{tr})$ and (u_i^{rot},v_i^{rot}) are the components of the flow due to the translation and rotation respectively. Note that the translational flow contains terms that depend on the platform rotation parameters β and γ, because these platform rotations induce translations in the cameras. Eliminating the depth Z_i from the respective pair of equations gives us the differential epipolar constraint of the individual camera:

$$u_i v_i^{tr} - v_i u_i^{tr} = u_i^{rot} v_i^{tr} - v_i^{rot} u_i^{tr} \qquad i = r, l \tag{3}$$

whose bilinear nature has been noted by various authors (Heeger and Jepson, 1992; Ma et al., 2000; MacLean, 1999; Vieville and Faugeras, 1995).

3.2 Quasi-parallax

Conventional two-frame SFM works that do not leverage on the structural constraint afforded by the two cameras usually suffer from the bas-relief ambiguity, especially under small field-of-view (FOV). Here we make use of the structural constraint by collecting from the two cameras projection rays that are parallel to each other. We call the points associated with such a pair of matching rays as matching point (see Fig. 2). For the case of simple setup addressed in this section, these are simply points with the same image coordinates. These points have the desirable property that their rotational flows are the same. Taking inspiration from the classical parallax idea proposed by Rieger and Lawton (Rieger and Lawton, 1985), we subtract the two equations in (3) from one another. This removes many of the rotational terms, thus enhancing translation pickup and alleviating the bas-relief ambiguity:

$$\begin{aligned} &u_r v_r^{tr} - u_l v_l^{tr} - v_r u_r^{tr} + v_l u_l^{tr} \\ &\quad = u^{rot}(v_r^{tr} - v_l^{tr}) - v^{rot}(u_r^{tr} - u_l^{tr}) \end{aligned} \tag{4}$$

Since the rotational flows at the matching pairs are equal, we have omitted the subscript of the rotational flows u^{rot} and v^{rot}. Writing out the translational and rotational flows in full, we obtain:

$$\begin{aligned} &Uf(v_r - v_l) - Vf(u_r - u_l) + W(yu_r - yu_l - xv_r + xv_l) \\ &\quad = b\Big(2f(x\alpha\beta + y\beta^2 - y\gamma^2) - 2xy\alpha\gamma + (2f^2 - 2y^2)\beta\gamma \\ &\qquad + (y\beta + f\gamma)(u_r + u_l) - x\beta(v_r + v_l)\Big) \end{aligned} \tag{5}$$

Clearly, when b=0, we obtain perfect parallax. The RHS of equation (5) vanishes and we can solve the translation directly by linear least squares. In the general case of $b \neq 0$, the RHS is non-zero: this corresponds to the translational flow component still containing induced terms caused by the platform rotation. Thus, we term the resulting flow difference between matching points as quasi-parallax. Since our approach is similar in spirit to the parallax approach, it enjoys similar numerical advantages with regards to the bas-relief ambiguity. Yet it circumvents the limitations of the parallax approach mentioned in the preceding section because there is no need to restrict ourselves to flow pairs near depth boundaries.

Collecting all the N equations from the entire set of matching points, we can write the system of equations in the following form:

$$\mathbf{A}\boldsymbol{x}_1 = b(\mathbf{B}\boldsymbol{x}_2) \tag{6}$$

here $\boldsymbol{x}_1$=$(U,V,W)^T$, $\boldsymbol{x}_2$=$(\alpha\beta, \beta^2, \alpha\gamma, \beta\gamma, \gamma^2, \beta, \gamma)^T$, and the j^{th} row of $\mathbf{A}$ and $\mathbf{B}$ are respectively as below (omitting the subscript j on the RHS for brevity):

$$\begin{aligned} \boldsymbol{a}_j &= (fv_r - fv_l, fu_l - fu_r, yu_r - yu_l - xv_r + xv_l) \\ \boldsymbol{b}_j &= (2fx, 2fy, -2xy, 2f^2 - 2y^2, -2fy, \\ &\qquad y(u_r + u_l) - x(v_r + v_l), f(u_r + fu_l)) \end{aligned} \tag{7}$$

The term $b\mathbf{B}\boldsymbol{x}_2$ on the RHS in equation (6) can be regarded as the residue arising from the quasi-parallax and its value is typically very small due to the small baseline value b and that most of the terms are second order in the rotational parameters.

3.3 Solving the Motion Parameters

As the RHS in equation (6) are negligibly small, we propose a two-stage scheme to solve for the translation and rotation separately.

Step 1: estimate initial translation. We ignore the RHS of (6) and solve the resulting homogeneous system $\mathbf{A}\boldsymbol{x}_1$=$\mathbf{0}$. The initial translation $\hat{\boldsymbol{v}}$ is recovered up to a scalar unknown.

Step 2: estimate initial rotation. Given the current translation estimate $\hat{\boldsymbol{v}}$, we can now use the epipolar constraint equation (3) to recover the rotation parameters. Equation (6) is not suitable for rotation recovery because the rotational terms are largely removed here. Writing out in full, we have: for $i = r, l$ representing the right and the left cameras respectively. Substituting the initial translation estimation $\hat{\boldsymbol{v}}$=$(sU, sV, sW)^T$ into (8) and collecting all the equations from both the right and the left cameras, we obtain a system of equations in the form of:

$$\begin{aligned}&\boldsymbol{M}\cdot(\alpha,\beta,\gamma,\frac{b}{s}\beta,\frac{b}{s}\gamma,\frac{b}{s}\alpha\beta,\frac{b}{s}\alpha\gamma,\frac{b}{s}\beta^2,\frac{b}{s}\beta\gamma,\frac{b}{s}\gamma^2)^T\\&=\boldsymbol{M}\cdot\boldsymbol{\theta}=\boldsymbol{d}\end{aligned}\tag{9}$$

The unknown vector $\boldsymbol{\theta}$ contains both first order and higher order rotational terms. Considering the typical rotation values, the contribution of the higher order terms can be ignored in the initial estimation step. Thus:

Step 2.1: we simplify equation (9) to obtain:

$$\boldsymbol{M}_1\cdot(\alpha,\beta,\gamma,\frac{b}{s}\beta,\frac{b}{s}\gamma)^T=\boldsymbol{M}_1\cdot\boldsymbol{\theta}_1=\boldsymbol{d}\tag{10}$$

where $\mathbf{M_1}$ is a $N\times 5$ matrix comprising of the first five columns from $\mathbf{M}$. We can now solve the linear system and the first three components of the solution vector $\boldsymbol{\theta}_1$ correspond to the initial rotation estimate $\hat{\boldsymbol{\omega}}_0$=$(\hat{\alpha}_0,\hat{\beta}_0,\hat{\gamma}_0)^T$.

Step 2.2: refine the estimate $\hat{\boldsymbol{\omega}}_0$. Substituting the current value of $\hat{\boldsymbol{\omega}}_0$ into the terms containing $\frac{b}{s}$ in (9) and rearranging, we obtain:

$$\boldsymbol{M}_2\cdot(\alpha,\beta,\gamma,\frac{b}{s})^T=\boldsymbol{d}\tag{11}$$

We solve the above equation for an updated rotation estimate. The newly obtained estimate is substituted back into (9) to generate an updated (11) which is solved again for a more refined solution. This process is repeated until the solution converges and a stable rotation estimate $\hat{\boldsymbol{\omega}}$=$(\hat{\alpha},\hat{\beta},\hat{\gamma})^T$ is obtained. Essentially we can make do with a simple linearization because the second order effect is small. Numerical tests conducted under a range of motion-scene configurations and baseline values reported in the experimental section show that the estimate usually converges to a stable solution after two or three iterations.

Step 3: refine the motion estimation.

Step 3.1: We substitute the current rotation estimate $(\hat{\alpha},\hat{\beta},\hat{\gamma})^T$ back into $\boldsymbol{x}_2$ of (6) and form a new equation:

$$(\boldsymbol{A},-\boldsymbol{B}\boldsymbol{x}_2)\cdot\left(\boldsymbol{x_1}^T,b\right)^T=\tilde{\boldsymbol{A}}\cdot\tilde{\boldsymbol{x}}_1=\mathbf{0}\tag{12}$$

We solve the homogeneous system (12) for an updated translation estimate $\hat{\boldsymbol{v}}$. Obviously, if b is known, the absolute value of the translation can be determined. Otherwise, only the translation direction can be determined.

Step 3.2: Given this updated translation estimate $\hat{\boldsymbol{v}}$, we use the scheme in step 2 to obtain a more updated rotation estimate $\hat{\boldsymbol{\omega}}$.

Step 3.3: If the current motion estimate differs from that of the previous iteration by less than 0.1%, stop. Otherwise, repeat steps 3.1 and 3.2 until the solution is stable.

4 Extensions to the Basic Arrangement

In the natural world, one is met with eyes that are neither purely frontal nor lateral. For instance, even a predatory bird such as the Short-toed Snake-eagle does not have completely frontal eyes (see Fig. 1) (Martin, 2007). Even a visual system possessing frontally-placed eyes may destroy this simple arrangement via eye movements such as sideway gaze and convergence. While it is possible to perform image rectification to restore the parallel-axes stereo geometry and then use the basic model solution, it is very much against our philosophy of computational parsimony, because feature correspondence would be needed to establish the rectification transformation. Thus, if the quasi-parallax solution is to be a useful strategy for locomotion at all, we must seek extensions to the basic solution provided above.

4.1 Quasi-parallax of Sideway Configuration

From the quasi-parallax framework introduced in the preceding section, it is an easy step to extend it to the case of the two eyes gazing sideway.

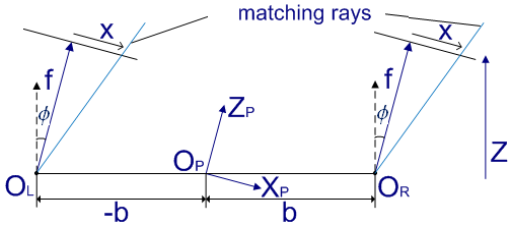


Fig. 3: Top view of the sideway configuration.

$$\Big((f^2+y^2)U - fWx - Vxy\Big)\alpha + \Big((f^2+x^2)V - fWy - Uxy\Big)\beta + \Big((x^2+y^2)W - fVy - fUx\Big)\gamma$$
$$+ b_i\Big((uy-vx)\beta + (\beta^2-\gamma^2)fy + fx\alpha\beta + fu\gamma - xy\alpha\gamma + (f^2-y^2)\beta\gamma\Big) = fUv - fuV - vWx + uWy \tag{8}$$

Fig. 3 shows the top view of a sideway configuration where the two cameras have been both rotated by the same angle ϕ around their Y axes and now gaze in a sideway direction. We call ϕ the sideway gaze angle.

If we let the new world coordinate system be O_p-$X_pY_pZ_p$, and if the platform motion expressed in this coordinate system is $\boldsymbol{v}=(U,V,W)^T$ and $\boldsymbol{\omega}=(\alpha,\beta,\gamma)^T$, then the 3D camera motions expressed in the respective camera coordinate systems are, for $i=r,l$:

$$\begin{aligned}\boldsymbol{v}_i &= (U-\beta b_i\sin\phi, V+\gamma b_i\cos\phi+\alpha b_i\sin\phi,\\ &\qquad W-\beta b_i\cos\phi)^T\\ \boldsymbol{\omega}_i &= (\alpha,\beta,\gamma)^T\end{aligned} \tag{13}$$

Carrying out analogous operation as before, we obtain the counterpart of equations (5) and (8), that is, the quasi-parallax equation (14) and the two epipolar constraint equations (15) ($i=r,l$).

While there are more terms in these equations, the nature of the equations are essentially the same as those of (5) and (8), since the rotational flows at the matching points are still the same. Thus, a scheme very similar to that in Sect. 3.3 can be used to solve for the motion parameters and the unknown ϕ. Readers can refer to Appendix for more details.

4.2 Quasi-parallax of Convergent/Divergent Configuration

The two eyes or cameras may not be parallel to each other, either because they converge to fixate on some object at some finite distance, or because the optical axes are divergent even in the relaxed state. Without loss of generality, we represent such convergent or divergent configuration as that shown in Fig. 4, where the right and the left cameras are rotated by an angle of $\theta_r=\theta$ and $\theta_l=-\theta$ around their Y axes respectively and the WCS O_p-$X_pY_pZ_p$ is as that in Section 3.1. We define θ as the convergence angle.

Given motion $\boldsymbol{v}=(U,V,W)^T$, $\boldsymbol{\omega}=(\alpha,\beta,\gamma)^T$ of the platform, the individual camera motions expressed in their own camera coordinate systems are, for $i=r,l$, as follows:

$$\begin{aligned}\boldsymbol{v}_i &= \Big(U\cos\theta_i - (W-b_i\beta)\sin\theta_i, V+b_i\gamma,\\ &\qquad (W-b_i\beta)\cos\theta_i + U\sin\theta_i\Big)^T\\ \boldsymbol{\omega}_i &= (\alpha\cos\theta_i - \gamma\sin\theta_i, \beta, \gamma\cos\theta_i+\alpha\sin\theta_i)^T\end{aligned} \tag{16}$$

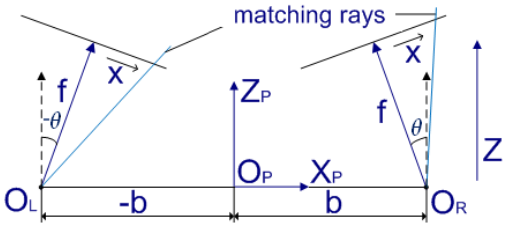


Fig. 4: Top view of the convergent configuration.

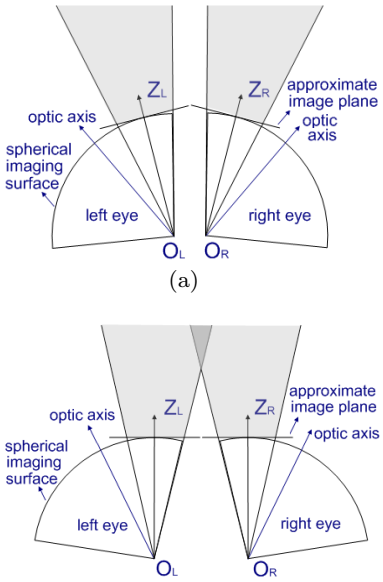


Fig. 5: Eyes with divergent optic axes. (a) Z-axes placed near the frontal direction so that the angle of divergence θ is small. (b) θ can be made to approach zero if the visual field of each eye extends sufficiently into the opposite hemisphere.

Clearly, the two camera rotations $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_l$ are different. This poses a problem for the rotation cancellation step in our quasi-parallax formulation. It seems clear that one must accept compromise if we were to retain the virtue of simplicity in our formulation. Here we assume that the angle θ is small enough, so that most of the rotational flow can be canceled by the subtraction operation carried out at corresponding matching points with the same (x,y) coordinates. This assumption on θ may not be as restrictive as it seems for the following two reasons. Firstly, any convergent eye movements during high-speed locomotion are likely to be for ob-

$$
\begin{aligned}
&Uf(v_r - v_l) + Vf(-u_r + u_l) + W\Big(y(u_r - u_l) - x(v_r - v_l)\Big) \\
&\quad = b\cos\phi(2fx\alpha\beta + 2fy\beta^2 - 2xy\alpha\gamma + 2f^2\beta\gamma - 2y^2\beta\gamma - 2fy\gamma^2 + y\beta u_r + f\gamma u_r + y\beta u_l + f\gamma u_l - x\beta v_r - x\beta v_l) \\
&\qquad - b\sin\phi(2xy\alpha^2 - 2x^2\alpha\beta + 2y^2\alpha\beta - 2xy\beta^2 + 2fy\alpha\gamma - 2fx\beta\gamma - f\alpha u_r - f\alpha u_l - f\beta v_r - f\beta v_l)
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
&(fWx - f^2U + Vxy - Uy^2)\alpha + (fWy - f^2V - Vx^2 + Uxy)\beta + (fUx - Wx^2 + fVy - Wy^2)\gamma \\
&\quad + b_i\cos\phi\Big((vx - uy)\beta - fu\gamma - fx\alpha\beta - fy(\beta^2 - \gamma^2) + xy\alpha\gamma - (f^2 - y^2)\beta\gamma\Big) \\
&\quad - b_i\sin\phi(fu\alpha - xy\alpha^2 + fv\beta + (x^2 - y^2)\alpha\beta + xy\beta^2 - fy\alpha\gamma + fx\beta\gamma) = fuV - fUv + vWx - uWy
\end{aligned} \tag{15}
$$

jects at some distances away, and thus the convergence angle will be small. Secondly, even for significantly divergent eye configuration like that of the Short-toed Snake-eagle illustrated in Fig. 1, all is not lost. If the eyes are sufficiently spherical, we are free to position the Z-axes of the two cameras appropriately such that for the regions of the two eyes facing the front (the regions shaded in grey in Fig. 5(a)), it is equivalent to a slightly divergent binocular configuration with small angle of divergence θ. Note that if the visual field of each eye extends sufficiently into the opposite hemisphere (as in Fig. 5(b)), we can even position the Z-axes of the two cameras such that it approaches the simple parallel configuration. Given this small θ assumption, we make the following simplifications: $\cos\theta \approx 1$, $\cos^2\theta \approx 1$, and $\sin^2\theta \approx 0$. Carrying out the same operations as before, we obtain the quasi-parallax equation as (17) and the epipolar constraint equations (for $i = r, l$) as (18).

We use the following scheme to solve for the motion parameters and the convergent angle θ.

Step 1: estimate initial translation and $\sin\theta$. We ignore the RHS of (17) as before, and also further omit the terms containing rotation parameters in the LHS as they are all coupled with $\sin\theta$ and thus constitute second order effects. As a result, we obtain:

$$
\begin{aligned}
&U\Big(fv_r - fv_l - (v_l x + v_r x - u_l y - u_r y)sin\theta\Big) \\
&\quad + V(fu_l - fu_r) + W\Big(v_l x - v_r x - u_l y + u_r y \\
&\quad - (fv_l + fv_r)sin\theta\Big) = 0
\end{aligned} \tag{19}
$$

Gathering all such equations, we solve the resulting homogeneous system $\boldsymbol{A}_4\boldsymbol{y}=\boldsymbol{0}$ for the initial direction of translation and the initial value of $\sin\theta$. Here $\boldsymbol{y}=(sU, sU\sin\theta, sV, sW, sW\sin\theta)^T$ is treated as a vector of independent unknowns, and s is the unknown scale factor. The translation can only be solved up to the scale factor s, but the value of $\sin\theta$ can be obtained as the average of the ratio of the first two and the last two components of $\boldsymbol{y}$.

Step 2: estimate initial rotation. Given $\boldsymbol{y}$, we can now solve equations (18) for the rotation parameters. Dropping second order rotational terms from (18), we obtain equation (20).

Gathering all such equations and solving the resulting linear system, we can obtain the solution vector $\boldsymbol{r}_{init}=(\alpha, \beta, \gamma, b\beta/s, b\gamma/s)^T$. The dependency between the components of $\boldsymbol{r}_{init}$ is ignored and the initial rotation estimate is simply obtained as the first three components of $\boldsymbol{r}_{init}$. The newly obtained estimate is substituted back into equations (18) to take into account the second order rotational terms. The updated (18) is solved again for a more refined rotation estimate. This process is repeated until the solution converges.

Step 3: refine motion estimate.

Step 3.1: Substituting the current rotation estimate into (17), we obtain an improved solution for translation and $\sin\theta$ by solving a new homogeneous system $\boldsymbol{A}_5\boldsymbol{y}_1=0$. Here, we need to specify the unknown $\boldsymbol{y}_1=(sU, sU\sin\theta, sV, sV\sin\theta, sW, sW\sin\theta, sb, sb\sin\theta)^T$.

Step 3.2: Given $\boldsymbol{y}_1$, we repeat step 2 to refine the rotation estimate.

Step 3.3: If the current motion estimate differs from that of the previous iteration by less than 0.1%, stop. Otherwise, repeat steps 3.1 and 3.2 until the solution is stable.

4.3 Implementation Details

We first use the variational method of (Bruhn et al., 2005) to obtain 100% dense flow field. Clearly, not all flow estimates have the same reliability. Following (Bruhn et al., 2005), an energy-based confidence measure c_{energy} is used to assess the relative reliability of the flow estimate (u, v) at every image location. If E_i is the energy functional that penalizes deviations from model assumptions (such as brightness constancy and smoothness) at pixel i, we define the confidence measure c_{energy} to be inversely proportional to E_i:

$$
c_{energy} = \frac{1}{E_i + \epsilon^2} \tag{21}
$$

$$U\Big(fv_r - fv_l + (u_l y + u_r y - v_l x - v_r x + 4fx\alpha + 2f^2\gamma - 2x^2\gamma + 2fy\beta)\sin\theta\Big) + V\Big(fu_l - fu_r + (2fy\alpha - 2xy\gamma)\sin\theta\Big)$$
$$+ W\Big((v_l - v_r)x - (u_l - u_r)y - (fv_l + fv_r + 2xy\beta + 2f^2\alpha - 2x^2\alpha - 4fx\gamma)\sin\theta\Big) \quad (17)$$
$$= b\Big((u_l + u_r)(f\gamma + y\beta) - (v_l + v_r)x\beta + (2f^2 - 2y^2)\beta\gamma + 2fy(\beta^2 - \gamma^2) - 2xy\alpha\gamma + 2fx\alpha\beta + (v_l - v_r)f\beta\sin\theta\Big)$$

$$\alpha\Big(Vxy - f^2U + fWx - Uy^2 + (fVy + f^2W + 2fUx - Wx^2)sin\theta_i\Big) + \beta\Big(Uxy - f^2V - Vx^2 + fWy + (fUy$$
$$- Wxy)sin\theta_i\Big) + \gamma\Big(fVy + (fUx - Wx^2 - Wy^2) + (f^2U - 2fWx - Ux^2 - Vxy)sin\theta_i\Big) + b_i\Big((vx - uy)\beta - fu\gamma$$
$$- fx\alpha\beta + xy\alpha\gamma - fy\beta^2 - (f^2 - y^2)\beta\gamma + fy\gamma^2 + (fv\beta + xy\beta^2 + fy\alpha\gamma - xy\gamma^2 + x^2\alpha\beta - f^2\alpha\beta + 2fx\beta\gamma)sin\theta_i\Big) \quad (18)$$
$$= fuV + vWx - fUv - uWy + (fvW + Uvx - Uuy)sin\theta_i$$

$$\alpha\Big(Vxy - f^2U + fWx - Uy^2 + (fVy + f^2W + 2fUx - Wx^2)\sin\theta_i\Big) + \beta\Big(Uxy - f^2V - Vx^2 + fWy + (fUy$$
$$- Wxy)\sin\theta_i\Big) + \gamma\Big(fVy + fUx - Wx^2 - Wy^2 + (f^2U - 2fWx - Ux^2 - Vxy)sin\theta_i\Big) + b_i(v_r x\beta - u_r y\beta - fu_r\gamma) \quad (20)$$
$$= fu_rV - fUv_r + v_rWx - u_rWy + (fv_rW + v_rxU - u_ryU)\sin\theta_i$$

Here ϵ serves as a small regularisation parameter that prevents the denominator from becoming singular. Equipped with the above confidence measure, we divide the flow field of each camera into 20 confidence levels. In our experiment with 600×480 images, we do a first cut by selecting only matching points with confidence level at level 16 and above (that is, the top 25%). The confidence level of a matching pair is taken to be the lower of the two confidence levels of the two flow measurements.

Besides the confidence criterion, we also need to make sure that the flow difference at matching points is not too small with respect to the noise expected. For this purpose, we add a further selection measure c_{mag}:

$$c_{mag} = \frac{\sqrt{(u_r - u_l)^2 + (v_r - v_l)^2}}{max(|(u_r, v_r)|, |(u_l, v_l)|)}$$

We use the following thresholding scheme to trim the number of matching points to the top 150 pairs:

$$c_{mag} > \tau(i) \quad (22)$$

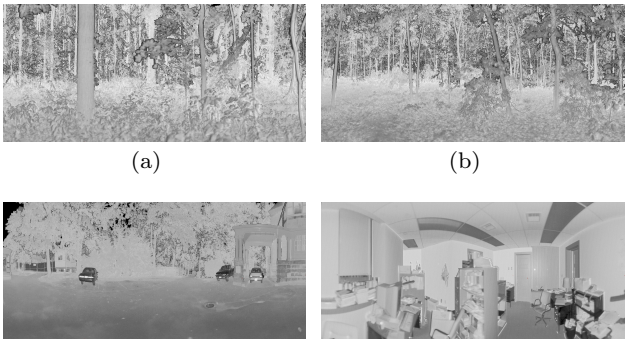
where the threshold $\tau(i)$ depends on how much confidence we have in the quasi-parallax measurements, as indicated by the confidence level i. In our experiment, we use a simple linear relationship $\tau(i) = -0.05i + 1.15$ (that is, τ varies linearly between 0.35 to 0.15 for i ranging from 16 to 20).

5 Experiments on Synthetic Data

We first carry out experiments on synthetic but realistic data by using the Brown range image database (Lee and Huang, 2000) which contains many static natural scenes. Fig. 6 shows some typical scenes in the database: forest, outdoor and indoor. Unless otherwise stated, we use scene (a) in the simulations that ensue. The average scene depth of this scene is about 7 m. We endow the scene with 3D motions, and project the points and their flows onto each camera's image plane. As the relative motion between the cameras and the scene are known, the result of our method can be evaluated by comparing with the ground truth. Note that while the direction and magnitude of the rotation can be estimated, only the direction of the translation is recovered, with its magnitude only recoverable when b is known. The camera pair is in the ideal frontal parallel configuration, unless otherwise noted, as from Section 5.7 onwards. Through a series of experiments, we clarify the effects of various factors such as motion-scene configuration on the performance of our quasi-parallax method. Its performance is also compared against that of the so-called gold standard bundle adjustment method.

5.1 Different Motion Configurations

We present different 3D motions with varying parameter: **translation-to-rotation ratio** ϵ (ϵ=0.1, 0.2, 1, 5, 10) so as to investigate the effect of motion configurations. The value of ϵ is computed as the ratio of the total magnitude of the translational flow and that of the rotational flow from all available points. We set other parameters as follows: the focal length is 6 mm, FOV is 50°, b=0.2 m, and the image dimension is 600×600 pixels (5.6 × 5.6 mm in metric unit).

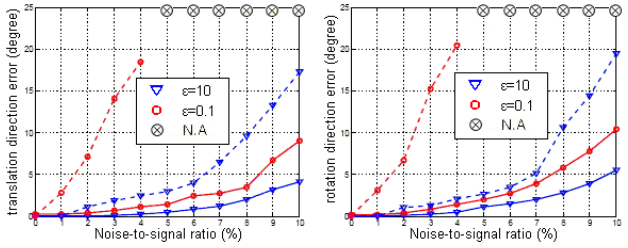


(a) (b) (c) (d)

Fig. 6: Range images of typical scenes. (a) Forest 1, (b) Forest 2, (c) Outdoor and (d) Indoor scenes.

We compare the recovered motion parameters ($\boldsymbol{\upsilon}_{re}$, $\boldsymbol{\omega}_{re}$) with the ground truth ($\boldsymbol{\upsilon}_{gt}$, $\boldsymbol{\omega}_{gt}$). The results are shown in Table 1. The errors of the translation $\boldsymbol{\upsilon}$ and the rotation $\boldsymbol{\omega}$ are defined as the angles between $\boldsymbol{\upsilon}_{re}$ and $\boldsymbol{\upsilon}_{gt}$, and between $\boldsymbol{\omega}_{re}$ and $\boldsymbol{\omega}_{gt}$ respectively. The error in the magnitude is defined as $|\frac{||\boldsymbol{\upsilon}_{re}||}{||\boldsymbol{\upsilon}_{gt}||} - 1|$ for translation and $|\frac{||\boldsymbol{\omega}_{re}||}{||\boldsymbol{\omega}_{gt}||} - 1|$ for rotation.

Without noise in the flow input, the results in Table 1 look fairly good. Yet, even without noise, it can be seen that with decreasing ε, the accuracy of the system deteriorates. At $\varepsilon = 0.1$, the translation recovery incurs significant errors of 20% or more in both direction and magnitude. It could be that with rotation-dominant motion (small ε), the quality of the quasi-parallax degrades with the weak translation. This is especially a problem since we did not impose any selection criterion on the quasi-parallax measurements in this experiment.



(a) (b)

Fig. 7: Estimation errors as a function of noise level, and over a range of ε. Dotted curves represent the case where all quasi-parallax measurements are used, solid curves represent the case that only the top 150 quasi-parallax measurements are retained. The symbol N.A. represents that the solution does not converge.

We now add isotropic, Gaussian noise to the 2D motion field, with the standard deviation of the noise amplitude ranging from 0% to 10% of the length of the individual 2D motion vector. One hundred separate runs are carried out for each noise level, and the mean values of the direction errors are plotted in Fig. 7 as dotted curves. We only plot the direction errors since the trend is similar for the magnitude errors (and in the magnitude case, the error remains bounded by 16% for the range of conditions tested). Note that for some noise level, no stable solution can be obtained (indicated as N.A. in Fig. 7). This is not surprising since in this simulation, all quasi-parallax measurements are used, without considering their reliability.

We now improve the quality of the quasi-parallax by selecting the top 150 pairs of matching points ranked according to the magnitude of the flow difference (here the confidence level of the flow does not come in since we are not dealing with real images). The improved results are plotted in Fig. 7 as the solid curves. From these results, we conclude that: (1) selecting good quasi-parallax plays an important role in reducing the negative impact caused by noise in the optic flow input; (2) our quasi-parallax scheme can handle a wide spectrum of motions ranging from translation-dominant motion (ϵ=10) to rotation-dominant motion (ϵ=0.1); and (3) the accuracy of the motion estimates improves with increasing the translation-to-rotation ratio ε.

5.2 Different b

The distance b is half of the baseline which is one of the most important parameters of any stereo configuration. It also plays an important role in our quasi-parallax framework, as it appears as the multiplier in the RHS of equation (6). It determines the magnitude of the RHS and thus might affect the convergence of our iterative algorithm which is initialized by ignoring the RHS. On the other hand, it is also intuitively clear that large b improves the quality of the quasi-parallax, because it is more likely to yield matching points with significant depth differences.

In this group of experiments, we vary the offset b from 0.01m to 0.5m, with ϵ=1 and other parameters the same as before. Fig. 8 shows the errors in the motion estimation with different offsets and different noise levels. Generally, there is an increase in performance with larger b, with the improvement leveling off when the offset exceeds a certain threshold. The first conclusion is that under the range of operating conditions tested, the effect caused by initially ignoring the RHS of equation (6), even under large b, is negligible. As for the decreasing errors with b in Fig. 8(a), (b), and (c), the phenomenon can be explained by Fig. 8(d),which expresses the relationship between the number of good matching

Table 1: Motion recovery for different ϵ.

	Motion parameter		Error of motion estimation			
ε	$\boldsymbol{v}_{gt}$(cm/s)	$\boldsymbol{\omega}_{gt}$($\times$0.001rad/s)	$\boldsymbol{v}$ direction	$\boldsymbol{v}$ magnitude	$\boldsymbol{\omega}$ direction	$\boldsymbol{\omega}$ magnitude
10	(3,3,11)	(0.5,0.5,0.1)	0.0006	0.0004	0.0002	0.0000
5	(2,2,8)	(0.5,0.5,0.1)	0.0031	0.0010	0.0009	0.0001
1	(1,1,5)	(0.5,0.5,0.1)	0.0079	0.0013	0.0011	0.0006
0.2	(1,1,3)	(1,2,0.23)	0.0571	0.0821	0.0083	0.0039
0.1	(1,1,2)	(2,4,0.58)	0.2310	0.2021	0.0322	0.0407

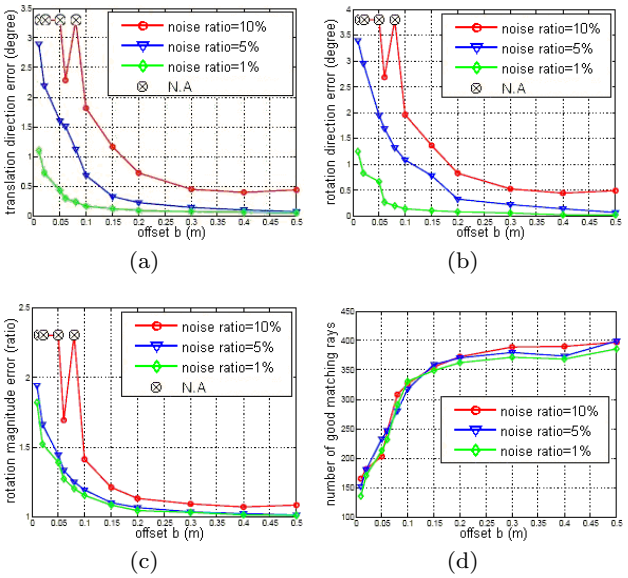


Fig. 8: Estimation errors as a function of offset b. (a) Translation direction. (b) Rotation direction. (c) Rotation magnitude. (d) Number of good matching rays defined as those with $c_{mag} > 0.2$.

rays (defined as those with $c_{mag} > 0.2$) and b. Intuitively, increasing baseline is conducive to forming good parallax because it is more likely to have large depth differences in the matching points. Having a larger pool of good parallax measurements to choose the top 150 matching points in turn improves the quality of the input to the algorithm. However, once the baseline is large enough, the depth difference in the matching pair is no longer correlated to the baseline, thus explaining the plateau in the plots. Thus, in the design of a quasi-parallax-based system, we should seek the point beyond which increase in baseline does not lead to further improvement in error performance. For our case, where the scene depth is at least 10m away, the threshold seems to be $b = 0.2$m from the plots in Fig. 8. For nearer scene depths such that the depth changes relative to the average scene depth are larger, the threshold will be smaller. We repeat the experiments with the same forest scene but with the scene content placed much closer at an average depth of 3.5m and at 2m. At these settings, which seem closer to the conditions under which some animals navigate in enclosed forests, we obtain a threshold value of $b = 0.05$m and $b = 0.03$m respectively. These values of b seem to be in keeping with the eye separation distances found in some mammals.

5.3 Different Scenes

The effect caused by different scenes is similar to that caused by b in the preceding subsection, as different scene types with different degree of roughness has a direct impact on the amount of good quasi-parallax measurements. Here, we perform experiment on typical scenes individually, including the two forest scenes, the outdoor and the indoor scenes shown in Fig. 6. The conditions for this group of experiments are as follows: ϵ=1, FOV = 50°, the noise level (noise-to-signal ratio) = 5%, and the offset b ranging from 0.05m to 0.5m.

Fig. 9 shows that the performance on the forest scene is better than that on the indoor scene given the same offset b. Such observation can be explained from the fact that the indoor scene contains much more planar areas than the forest scene, and nearly zero parallax is generated over much of the image when the offset b is small, which does not lend to numerical stability. Fig. 9(d) corroborates our explanation.

5.4 Different FOV

In addition to baseline, the FOV is another important parameter of a vision system, especially so for the function of motion estimation. We let the FOV range from 10° to 60°, fixing ϵ=1 and with other parameters remaining unchanged. Clearly, by virtue of the fact that a smaller FOV is viewing a smaller part of the scene, there will be smaller amount of depth changes; thus the quality of the quasi-parallax measurements will be affected, resulting in the deterioration of performance, as shown in Fig. 10(a) and (b). However, this is not our main point of interest here. What we are keen to

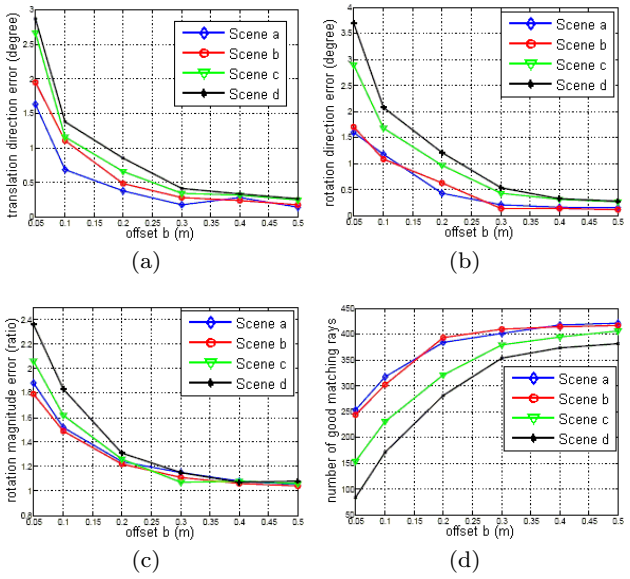


Fig. 9: Estimation errors as a function of offset b and different scenes (a: forest 1, b: forest 2, c: outdoor, d: indoor). (a) Translation direction. (b) Rotation direction. (c) Rotation magnitude. (d) Number of good matching rays defined as those with $c_{mag} > 0.2$.

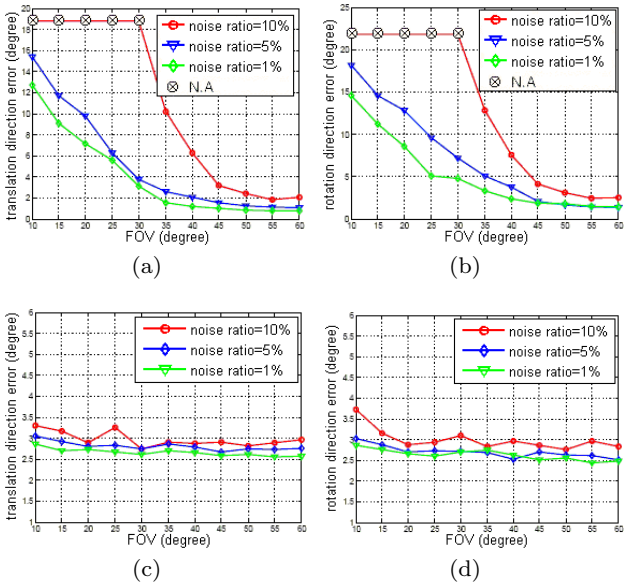


Fig. 10: Estimation errors as a function of the FOV. For (a) and (b), the quality of the quasi-parallax input varies (naturally) with the FOV; for (c) and (d), we control c_{mag}, the quality of the quasi-parallax input, to be the same despite changes in the FOV.

get at is that, given the *same* quality in the quasi-parallax measurements, do we expect the algorithm's performance to vary with the change in the FOV per se? For this purpose, we control the quality of quasi-parallax input—measured by c_{mag}, the magnitude of the flow difference of the matching points—to be the same, despite changes in the FOV.

As can be seen from Fig. 10(c) and (d), the errors in the motion estimates are relatively independent of the FOV. Thus the deciding factor for the accuracy of the quasi-parallax based method is the amount of depth difference, not the FOV per se.

5.5 Comparison Against Bundle Adjustment (BA)

In this section, we compare our method against the solution obtained by the so-called gold standard BA algorithm. The purpose of the comparison is not intended to establish the superiority of our method over BA or otherwise; in any case, the BA method is usually applied to scenarios where the differential techniques cannot be applied, and more importantly, it usually serves to refine the initial estimates from other algorithms. The purpose of the following comparison is rather to shed some light on the issue of bas-relief ambiguity which usually plagues two-frame SFM. While the superiority of the parallax scheme has been demonstrated for spherical FOV and for a pair of laterally placed eyes covering diametrically opposite viewing sphere (Hu and Cheong, 2009; Lim and Barnes, 2008), do we expect the parallax-based approach to exhibit the same disambiguation of the bas-relief problem for the frontally-placed eyes covering a small part of the viewing sphere, especially in comparison to the conventional two-frame SFM approached represented by BA?

We applied the extended BA algorithm proposed in (Hu and Cheong, 2009), with appropriate modifications taking into account the obvious geometry difference in the eye configuration. The outline of the extended BA are as follows: (1) use linear subspace method to obtain an initial estimate of each camera motion separately; (2) from the camera motion estimates, obtain an initial estimate of the global platform motion; (3) bundle adjust the global platform motion by minimizing the difference of the actual flow and the back-projected flow generated from the platform motion.

The conditions of the comparison are as follows: b=0.2 m, ϵ=1 and 0.1, the noise level in the flow is 5%, and the FOV ranging from 10° to 60°. The top 150 pairs of matching points are used by our quasi-parallax (QP) algorithm as before, whereas 150 feature points are selected randomly for each camera in the BA algorithm. In Fig. 11, the motion estimation results of the BA algorithm and our QP method are compared. It is clear that QP outperforms BA significantly under small FOV; however, given sufficient FOV (e.g. FOV>45°),

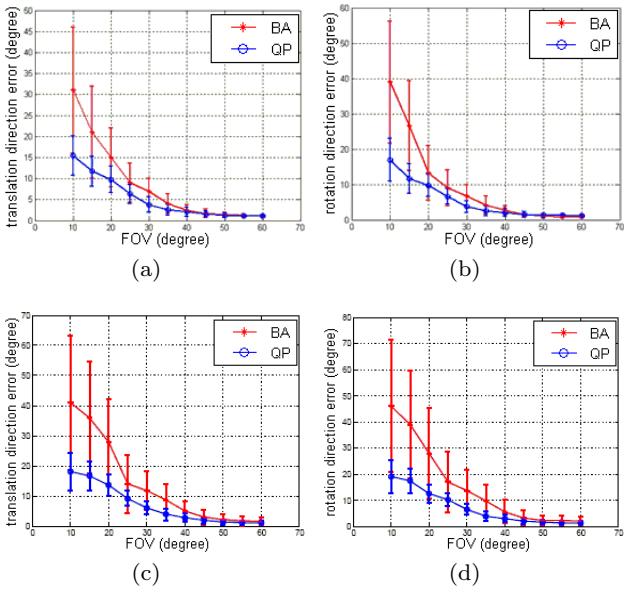


Fig. 11: Comparison of quasi-parallax (QP) and Bundle adjustment (BA) methods under different FOV. For (a) and (b), ϵ=1; for (c) and (d), ϵ=0.1.

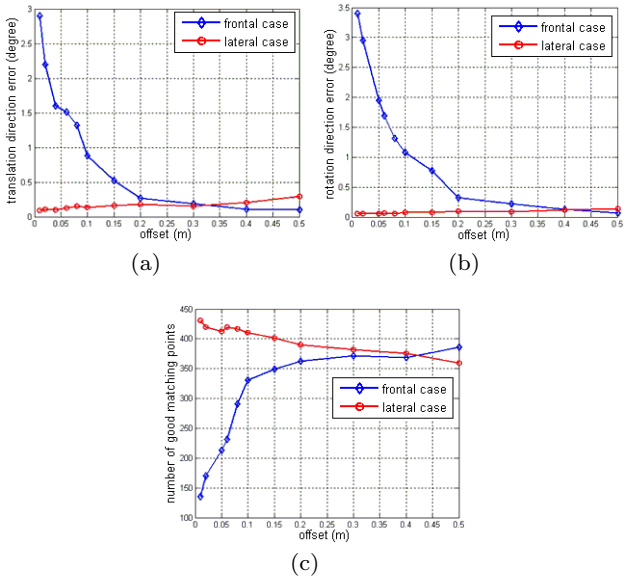


Fig. 12: Estimation errors of the frontal and lateral configurations as a function of offset. (a) Translation direction. (b) Rotation direction. (c) Number of good matching rays defined as those with $c_{mag} > 0.2$.

their performances are comparable since the bas-relief ambiguity is no longer a problem. In view of earlier results such as (Hu and Cheong, 2009) for lateral eyes and the current set of results for frontal eyes, we can conclude that the parallax-based method is generally more effective in removing the bas-relief ambiguity, and this superiority is independent of having eyes covering diametrically opposite viewing sphere.

5.6 Frontal vs Lateral Configuration

We now directly pit the performance of the frontal eyes versus the lateral eyes in resolving the bas-relief ambiguity using the quasi-parallax approach. Again, the noise level is 5% of the flow. One hundred separate runs are done for every condition tested and the mean value is reported. The same forest scene is used as before. The image content seen by the lateral eyes can be generated from the same forest scene without any problem, since the Brown range data is captured from a sensor encompassing a horizontal FOV of almost 270°. We use the same number of matching points (150 pairs) for both configurations.

We use the same parameters as in section 5.2 but with the offset b changing from 0.01m to 0.5m. Fig. 12 shows the results. As was shown in section 5.2, in the frontal case, larger b results in matching points with larger depth differences, leading to better results until the improvement in performance plateaus off at $b \approx$ 0.2m. In the lateral case, the performance is basically independent of the offset value b, as the two cameras are always viewing very different parts of the scene, irrespective of the value of b. Fig. 12(c) plots the relationship between the number of good matching pairs and b, which also corroborates the preceding conjecture. It also reveals that when the number of matching points reaches comparable level in both eye configurations, there is then no significant difference in the performance.

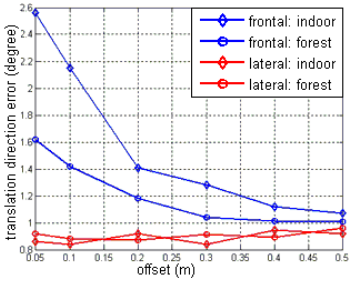


Fig. 13: Estimation results of frontal and lateral configurations with different scenes.

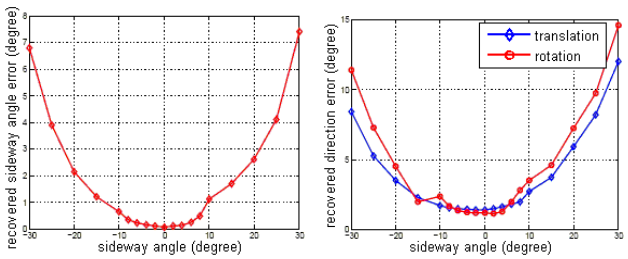
When we repeat the experiments with different scene type such as an indoor scene, we obtain very similar results (Fig. 13). The performance does not depend on the configuration of the eyes, but more on the number of good matching points available. In the case of indoor scene, due to the largely planar surfaces seen in

such scenes, the number of good matching points in the frontal eyes may never reach that of the lateral eyes, at least for the range of practical baselines tested.

From the above, we can conclude that there is no inherent advantage of having a lateral eye over having a frontal eye configuration, as far as resolving the bas-relief ambiguity is concerned, save for the fact that the lateral eye configuration is always more likely to yield good parallax measurements.

5.7 Estimation Results under Sideway Gaze

The sideway gaze configuration contains an additional unknown ϕ which might introduce new numerical instability. Here, we perform simulation on the forest scene as ϕ varies from -30° to 30°, with other parameters being: FOV=50°, b=0.2 m, ϵ=1, and the noise level is 5%.



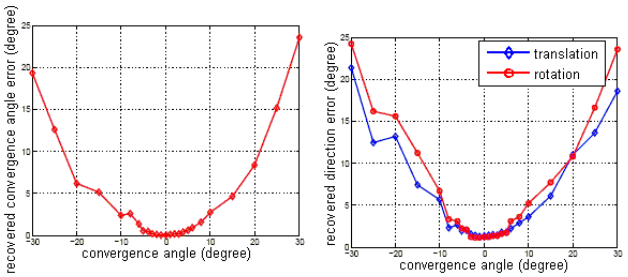
(a) (b)

Fig. 14: Estimation results as a function of the sideway angle.

Fig. 14(a) shows the mean error of the recovered ϕ value and Fig. 14(b) depicts the mean error of the recovered motion parameters over 100 trials. Clearly, the quasi-parallax framework remains effective when the sideway angle is within the range of 30°. The slight asymmetry of the result is due to the fact that with different ϕ, the cameras are observing different parts of the scene.

5.8 Estimation Results under Convergent Configuration

The convergent configuration not only introduces an additional unknown θ, it also results in an incomplete cancellation of the rotational components in the quasi-parallax measurements. We perform simulation on the convergent configuration with θ varying from -30° to 30°, and other parameters same as those in the preceding section.



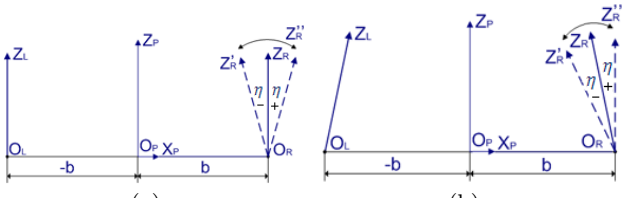
(a) (b)

Fig. 15: Estimation result as a function of the convergence angle.

Fig. 15(a) shows the mean error of the recovered convergence angle θ and Fig. 15(b) depicts the mean error of the recovered motion directions over 100 trials. It seems that the quasi-parallax method remains effective, especially if the total angle between the two Z-axis 2θ is less than 30°, under which the motion direction errors remain less than about 10°. Even as this total angle 2θ reaches towards the value of 60°, the error performance degrades gracefully with errors less than 25°. Thus, our quasi-parallax method remains a viable scheme for rapidly estimating heading directions during locomotion, even if the eyes are moderately convergent due to fixation, or divergent because the bony orbits are divergent.

5.9 Sensitivity Analysis

In practice, there might be imperfection in the camera postures such that they deviate from the canonical configurations modeled in the preceding sections. In this section, we conduct two experiments to test the sensitivity of our recovery methods when such modeling errors are present. Referring to Fig. 16, we perturb the right camera by an angle η so that the binocular setup deviates from the perfectly parallel case (Fig. 16(a)) and from the symmetrical convergent case (Fig. 16(b)).

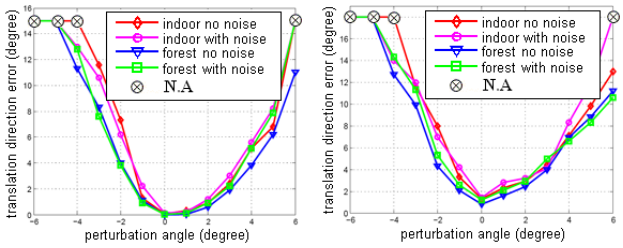


(a) (b)

Fig. 16: The right camera perturbed by an angle η from its (a) frontal, and (b) convergent configurations.

The other parameters were as follows: FOV=50°, b=0.2 m and ϵ=1. Both indoor and forest scenes were investigated. We also tested the case of perfect flow field and when 5% flow noise is added. Fig. 17(a) and Fig. 17(b) show the mean errors of the recovered translation direction for the frontal and convergent configurations respectively. On the whole, both the algorithms for the respective configurations are similarly affected by the perturbation; they fail to converge when the perturbation is more than 4°. Note that in Fig. 17, the estimation result for the case of positive η is better than that of negative η; this is nothing but the previously discussed fact that a stronger divergence resulting from positive η gives rise to better quasi-parallax than the case of negative η, as the observed scene content between the two cameras is more different.

From the preceding finding, it is evident that if significant eye convergence can take place during locomotion, there is a need to extend the basic model in Section 3, and instead use the extensions discussed in Section 4. While the convergent model is itself also sensitive to deviation from symmetry, we expect such asymmetry to be much less prevalent and its range of deviation to be small. Thus, under most cases, the extended models presented in this paper should be adequate for most practical situations.



(a) (b)

Fig. 17: Estimation errors of the translation direction as a function of the perturbation angle η, for (a) frontal, and (b) convergent configurations. For each case, both indoor and forest scenes, as well as noisy (5% standard deviation) and noiseless flows are investigated.

6 Experiments on Real Scenes

In our experiment on real data, we mounted a frontally parallel pair of synchronized cameras on a mobile platform, with the offset b being about 0.1m. We use two Dragonfly cameras from Point Grey Research with 50° FOV each. The frame rate is 30 frames per second and the image size is 640×480 pixels. Since no ground truth is available to evaluate our ego-motion estimation result directly, we compare our results against those obtained by inputting the four images involved to the Bundler software (Snavely et al., 2008) which is based on BA. We also color-code the dense depth maps that are reconstructed based on the motion parameters estimated by QP and by Bundler respectively, from which we can make some observation about the accuracy of the estimated motion parameters. Various indoor and outdoor scenes are tested and the results are shown below.

The first column of Fig. 18 is the original image of the scene; the second column depicts the confidence levels of the flow (the red, green and blue pixels represent the top three confidence levels respectively), and the third column depicts the 100% dense depth map reconstructed from QP, displayed as chroma-depth images with warm colors representing near depths and vice versa. In order to compensate for the effect of noise in real images, we regard those pixels with negative depths or very large positive depths as incorrect, and instead fill in the depth values from neighboring areas using a procedure similar to image quilting used in texture synthesis (Alexei and William, 2001). From visual inspection, it can be seen that the recovered depth is in good qualitative agreement with the actual scene. In particular, scene (a) is a quite textureless case and yet we can recover a reasonably good dense depth map.

We tabulate in Table 2 the difference in motion estimates recovered by the QP approach and the Bundler approach. As can be seen, the difference is small. Roughly speaking, indoor scenes (scenes a and b) exhibit greater differences in the recovered motion estimates. It is not clear, however, which approach yields a better results under this kind of scenes. The QP approach suffers from a lack of high-quality parallax measurement in this kind of scenes with many planar structures, whereas for the Bundler approach, the effective FOV is small due to the lack of distinctive features. To shed some light on this issue, for each approach, we reconstruct depths from the estimated motion parameters using the flow equations. It is known that the recovered depths near the estimated focus of expansion (FOE) are very sensitive to the accuracy of the motion estimation (Cheong et al., 1998). In particular, the reconstructed depths in the image region between the true and the estimated FOE are likely to have negative values. Hence we can use the amount of negative depths as a rough gauge for the accuracy of the motion estimates. In Table 2, we list in the last two columns the number of negative depths recovered from QP and Bundler respectively. In the case of scene (a), the negative depth region is clearly larger for the Bundler approach, indicating that its motion recovery contains a larger error. For other scenes, we obtain comparable statistics from both approaches. Thus, at 50° FOV, there is generally no difference in

Table 2: Difference in the motion parameters and the number of pixels with negative recovered depth recovered by QP and by Bundler .

Scene	Translation direction(°)	Rotation direction(°)	Rotation magnitude(%)	#negative depth(QP)	#negative depth(Bundler)
a	2.193	1.228	4.1	43	62
b	3.427	1.371	4.7	42	47
c	1.432	0.826	2.0	38	34
d	1.493	0.846	2.2	41	36
e	1.329	0.788	1.7	37	43
f	1.745	0.963	2.6	41	38

performance between the two approaches, corroborating the results obtained in Fig. 11.

Next, we compare the performance of QP and Bundler under the difficult scenario of small FOV (25°) and scenes depicting man-made environment. Here, the difference in performance between the two approaches is much more pronounced. From Fig. 19, we can clearly see the difference in extent of the negative depth regions (the black regions overlaid on the chroma depth maps)[1]. Thus, we can conclude that QP is better able to resolve the bas-relief ambiguity compared to conventional BA-based approach.

Lastly, we report on the amount of computations incurred by the two algorithms. Computation times are reported for a Dual-Core 2.5 GHz Intel processor executing C++ codes. Excluding the preprocessing steps (flow estimation for QP and SIFT feature detection for Bundler) and the depth reconstruction step, the average processing times on 640×480 images are 1.300 sec for QP and 6.713 sec for Bundler. If we run the algorithms on larger images (2400×1800), the corresponding processing times are 2.337 sec for QP and 11.597 sec for Bundler.

7 Conclusions

It is a commonplace that binocular overlap in the two eyes is utilized by the visual system as a form of stereoscopic depth cues. Yet empirical observation of the natural world gives us no warrant for supposing that stereopsis exists widely in vertebrates, still less birds. Given this lack of empirical evidence for stereopsis, and if we think it unlikely that nature will ignore the binocular overlap, it should then be possible for the binocular overlap to be exploited in some other way. In this paper, we showed that the arrangement of two frontally placed eyes — whose optical axes may or may not be parallel — can be leveraged for quasi-parallax instead of binocular disparities and we have demonstrated its feasibility over both synthetic and real data. Indeed, quasi-parallax, with the better disambiguation of translation and rotation over two frames, and the rapid processing that it entails, seems to us a more realizable and useful alternative during locomotion. It is more realizable because it involves mere matching of visual rays that are (approximately) parallel in directions, without the heavy optimization needed for solving the correspondence problem, and simple linear algebra for solving the 3D motion parameters. It is useful because it resolves the traditional difficulties associated with using parallax in a single image. It is particularly useful for locomotion because it provides a reasonably accurate translation and rotation estimates within a reasonable range of convergence angle θ and sideway gaze angle ϕ, and doing so in a real-time manner without deliberate processing. We feel that such a solution is not less useful because it exploits a particular eye topography and its general validity cannot be established for all θ and ϕ. To crave for a generally valid solution for all tasks may be a deep intellectual need but to allow such a need to dictate the design of a visual system that can move about in the real world is a symptom of an equally deep scientific fallacy.

[1] The geometry of the negative depth areas has been algebraically characterized by the Cremona transformation in terms of the errors in the estimated camera motion parameters (Cheong and Ng, 1999).

8 Appendix

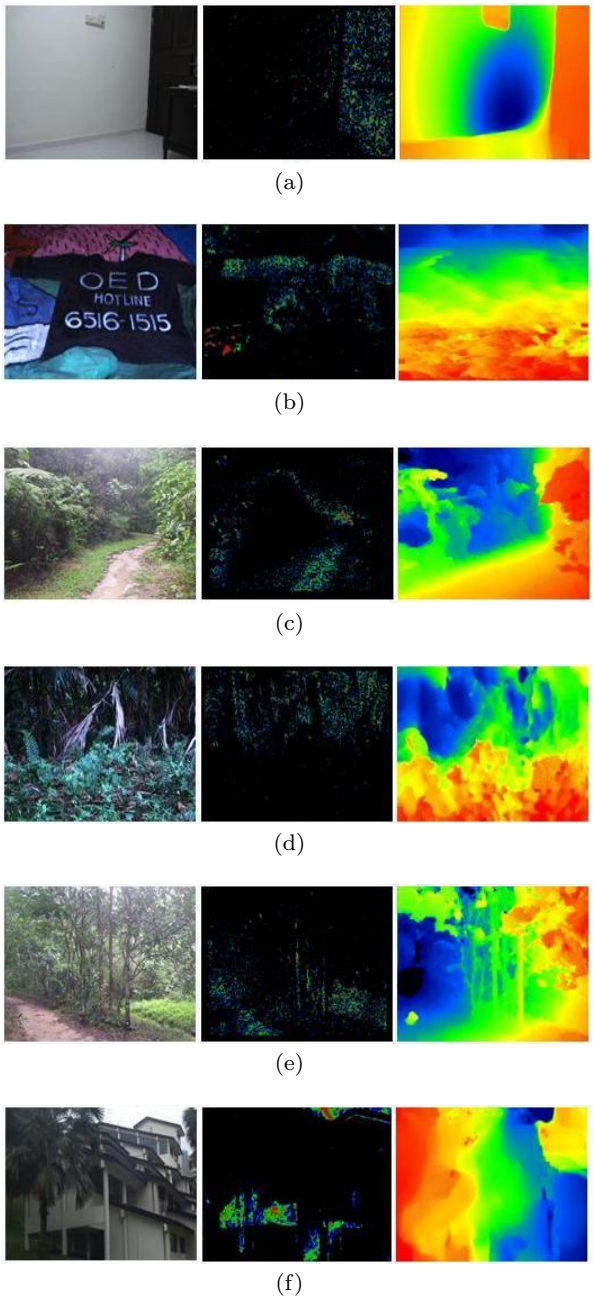
Solving the Motion Parameters for Sideway Configuration

To solve the motion parameters and the sideway angle ϕ, we use the following scheme which is highly similar to that of Section 3.3. First, we note that by collecting all equations (14) from N matching points, we can write the system of equations in the following form:

$$\boldsymbol{A}_3\boldsymbol{x}_3 = b\cos\phi(\boldsymbol{B}_3\boldsymbol{x}_4) + b\sin\phi(\boldsymbol{B}_4\boldsymbol{x}_5) \tag{23}$$

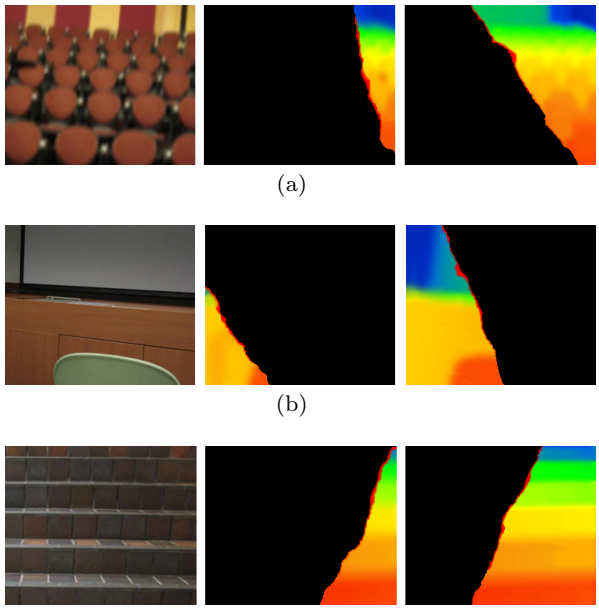
Here, $\boldsymbol{x}_3{=}(U,V,W)^T$.

Step 1: We ignore the RHS of (23) at first, and obtain the initial translation estimate $\hat{\boldsymbol{v}}$ up to a scalar



(a) (b) (c) (d) (e) (f)

Fig. 18: Results of six real image sequences with moderate FOV (50°). For each sequence, the left image depicts the scene, the middle those matching points with the top three confidence levels, and the right the dense depth map reconstructed from QP.



(a) (b) (c)

Fig. 19: Depth reconstruction results under small FOV of 25°. The second column is the result from Bundler and the third column from QP. The black pixel represents point whose recovered depth is negative.

factor s by solving the homogeneous system $\boldsymbol{A}_3\boldsymbol{x}_3{=}\boldsymbol{0}$. **Step 2:** Given $\hat{\boldsymbol{v}}$, we now turn to equation (15) to solve for the rotation parameters and the sideway angle ϕ. Substituting the initial translation estimation $\hat{\boldsymbol{v}}{=}(sU, sV, sW)^T$ into (15) and ignoring the higher order terms in the rotational parameters such as $\alpha\beta$ and $\beta\gamma$, we gather all measurements to obtain:

$$\begin{aligned}&\boldsymbol{M}_4 \cdot (\alpha, \beta, \gamma, \frac{b\cos\phi}{s}\beta, \frac{b\cos\phi}{s}\gamma, \frac{b\sin\phi}{s}\alpha, \frac{b\sin\phi}{s}\beta)^T \\ &\quad = \boldsymbol{M}_4 \cdot \boldsymbol{\phi}_1 = \boldsymbol{d}_1\end{aligned} \tag{24}$$

We can solve the above linear system, obtaining an initial estimate for the rotation $\hat{\boldsymbol{\omega}}_0{=}(\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0)^T$ in the first three components of $\boldsymbol{\phi}_1$.

Then we refine the rotation estimate by reinstating the higher order terms in (15) (by using $\hat{\boldsymbol{\omega}}_0$) and solving the resulting linear system of equations. This process is repeated until a stable rotation estimate $\hat{\boldsymbol{\omega}}{=}(\hat{\alpha}, \hat{\beta}, \hat{\gamma})^T$ is obtained. Numerical tests in the experimental section again show that the estimate always converges to a stable solution within five iterations.

Step 3: We substitute the current rotation estimate $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})^T$ back into the RHS of (23) and form a new system of equations:

$$\begin{aligned}&(\boldsymbol{A}_3, -\boldsymbol{B}_3\boldsymbol{x}_4, -\boldsymbol{B}_4\boldsymbol{x}_5) \cdot (\boldsymbol{x}_3, b\cos\phi, b\sin\phi)^T \\ &\qquad = \tilde{\boldsymbol{A}}_3 \cdot \tilde{\boldsymbol{x}}_3 = \boldsymbol{0}\end{aligned} \tag{25}$$

We solve the homogeneous system (25) for an updated translation estimate $\hat{\boldsymbol{v}}$. The value of ϕ is also

recovered from the relationship between the estimate for $b\cos\phi$ and $b\sin\phi$. Given the current translation estimate $\hat{\boldsymbol{v}}$, we repeat the procedure in step 2 to obtain an updated rotation estimate $\hat{\boldsymbol{\omega}}$. If the current motion estimate differs from that of the previous iteration by less than 0.1%, stop. Otherwise, repeat step 3 until the solution is stable.

References

Alexei A. E. and William T. F. 2001. Image quilting for texture synthesis and transfer. In *SIGGRAPH'01 Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 341-348).

Ayache, N. and Faugeras, D. 1989. Maintaining representations of the environment of a mobile robot. *IEEE Robotics and Automation Magazine*, 5(5):804-819.

Argyros, A. A., Tsakiris, D. P. and Groyer, C. 2004. Biomimetic Centering Behavior - Mobile Robots with Panoramic Sensors. *IEEE Robotics and Automation Magazine*, 11(4):21-30.

Balasubramanyam, P. and Snyder, M. A. 1991. The p-field: A computational model for binocular motion processing. In *CVPR* (pp. 115-120).

Bruhn, A., Weickert, J., and Schnorr, C. 2005. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211-231.

Carelli, R., Soria, C., Nasisi, O. and Freire, E. O. 2002. Stable AGV corridor navigation with fused vision-based control signals. In *Proceedings of the 28th Annual Conference of the IEEE Industrial Electronics Society–IECON 02*, 3:2433-2438.

Cheong, L.-F., Fermuller, C., and Aloimonos, Y. 1998. Effects of errors in the viewing geometry on shape estimation. *Computer Vision and Image Understanding*, 71(3):356-372.

Cheong, L-F. and Ng, K. 1999. Geometry of Distorted Visual Space and Cremona Transformation. *International Journal of Computer Vision*, 32(2):195-212.

Clark, J. and Yuille, A. 1994. *Data fusion for sensory information processing*. Dordrecht: Kluwer Academic.

Coombs, D. and Roberts, K. 1993. Centering behavior using peripheral vision. In *CVPR* (pp. 440-451).

Corke, P. I., Hrabar, S. E., Peterson, R., Rus, D., Saripalli, S. and Sukhatme, G. S. 2004. Autonomous deployment and repair of a sensor network using an unmanned aerial vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*, 3602-3609.

Davies, M. N. O. and Green, P. R. 1994. Multiple sources of depth information: an ecological approach. In: Davies M. N. O. and Green P. R(Eds.), *In Perception and Motor control in Birds: an ecological approach*, Berlin Springer, pp. 339-356.

Dev, A., Ben, K., and Groen, F. 1997. Navigation of a mobile robot on the temporal development of the optic flow. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, 2:558-563.

Duchon, A. P. and Warren, W. H. 1994. Robot Navigation from a Gibsonian Viewpoint. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 3:2272-2277.

Franceschini, N., Pichon, J. M., and Blanes, C. 1992. From insect vision to robot vision. *Philosophical Transaction of the Royal Society of London B: Biological Sciences*, 337:283-294.

Griffiths, S., Saunders, J., Curtis, A., Barber, B., McLain, T. and Beard, R. 2006. Maximizing miniature aerial vehicles: Obstacle and terrain avoidance for MAVs. *IEEE Robotics and Automation Magazine*, 13(3):34-43.

Grosso, N., Sandini, G. and Tistarelli, M. 1989. 3-D Object Reconstruction Using Stereo and Motion. *IEEE Trans. Systems, Man and Cybernetics*, 19(6):1465-1476.

Heeger, D. J. and Jepson, A. D. 1992. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95-117.

Hildreth, E. C. 1992. Recovering heading for visually-guided navigation. *Vision Research*, 32(6):1177-1192.

Ho, P. and Chung, R. 2000. Stereo-Motion with Stereo and Motion in Complement. *IEEE Trans. on PAMI*, 22(2):215-220.

Hrabar, S. E., Corke, I., Sukhatme, G. S., Usher, K. and Roberts, J. M. 2005. Combined optic flow and stereo-based navigation of urban canyons for a UAV. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 302-309).

Hu, C. and Cheong, L.-F. 2009. Linear Quasi-Parallax SfM Using Laterally-Placed Eyes. *International Journal of Computer Vision*, 84(1):21-39.

Huguet, F. and Devernay, F. 2007. A variational method for Scene Flow Estimation from Stereo Sequences. In *ICCV* (pp. 1-7).

Humbert J. S., Hyslop AM, and Chinn M. 2007. Experimental validation of wide-field integration methods for autonomous navigation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 2144-2149).

Humbert J. S., Murray RM, and Dickinson MH. 2005. Sensorimotor convergence in visual navigation and flight control systems. In *Proceedings of 16th IFAC World Congress*, Prague.

Kim, J., Li, H., and Richard, H. 2010. Motion estimation for nonoverlapping multicamera rigs: linear alge-

braic and L-infinity geometric solutions. *IEEE Trans. on PAMI*, 32(6):1044-1059.

Kriegman, D. J., Triendl, E. and Binford, T. O. 1989. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. Robotics and Automation*, 5(6):792-803.

Lee A. B. and Huang J. 2000. Brown range image database. http://www.dam.brown.edu/ptg/brid/index.html.

Li, L. and Duncan, J. H. 1993. 3-d translational motion and structure from binocular image flows. *IEEE Trans. on PAMI*, 15(7):657-667.

Lim, J. and Barnes, N. 2008. Directions of egomotion from antipodal points. In *CVPR* (pp. 1-8)..

Longuet-Higgins, H. C. and Pradzny, K. 1980. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London, Series B*, 208:385-397.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91-110.

Ma, Y., Kosecka, J. and Sastry, S. 2000. Linear differential algorithm for motion recovery: a geometric approach. *International Journal of Computer Vision*, 36(1):71-89.

MacLean, W. J. 1999. Removal of translation bias when using subspace methods. In *ICCV* (pp. 753-758).

Martin, G. R. 2007. Visual fields and their functions in birds. *Journal of Ornithology*, 148(2):547-562.

Martin, G. R. 2009. What is binocular vision for? A birds' eye view. *Journal of Vision*, 9(11):14, 1-19.

McFadden, S. A. 1993. Constructing the three-dimensional image. In: H. Philip Zeigler and Hans-Joachim Bischof(Eds.), *Vision, brain and behavior in birds*, Cambridge MA, MIT Press, pp. 47-61.

McFadden, S. A. 1994. Binocular depth perception. In: Davies M. N. O. and Green P. R(Eds.), *In Perception and motor control in birds: an ecological approach*, Berlin Springer, pp. 54-73.

Muratet, L., Doncieux, S., Briere, Y. and Meyer, J. A. 2005. A contribution to vision-based autonomous helicopter flight in urban environments. *Robotics and Autonomous Systems*, 50(4):195-209.

Neumann, J. 2004. Compound eye sensor for 3D egomotion estimation. In *Proceedings of IEEE International Conference on Intelligent Robots and Automation*, (pp. 3712-3717).

Neumann, T. R. and Bulthoff, H. H. 2001. Insect inspired visual control of translatory flight. In *Advances in Artificial Life, Proceedings of ECAL*, Berlin Springer, pp. 627-636.

Pless, R. 2004. Camera cluster in motion: Motion estimation for generalized camera designs. *IEEE Robotics and Automation Magazine*, 11(4):39-44.

Pons, J., Keriven, R. and Faugeras, O. 2007. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision*, 72(2):179-193.

Rieger, J. H. and Lawton, D. T. 1985. Processing differential image motion. *Journal of the Optical Society of America A*, 2(2):354-359.

Ruffier, F., Serres, J., Masson, G. P. and Franceschini, N. 2007. A bee in the corridor: regulating the optic flow on one side. In *Proceedings of the 7th meeting of the German neuroscience society—31st Gottingen neurobiology conference*, Gottingen Germany, abstract T14-7B.

Santos-Victor, J., Sandini, G., Curotto, F., and Garibaldi, S. 1995. Divergent stereo in autonomous navigation: From bees to robots. *International Journal of Computer Vision*, 14(2):159-177.

Serres, J., Dray, D., Ruffier, F. and Franceschini, N. 2008. A vision-based autopilot for a miniature air vehicle: joint speed control and lateral obstacle avoidance. *Autonomous Robots*, 25:103-122.

Serres, J., Ruffier, F., Masson, G. P. and Franceschini, N. 2007. A bee in the corridor: centring or wall-following? In *Proceedings of the 7th meeting of the German neuroscience society–31st Gottingen neurobiology conference*, Gottingen Germany, Abstract T14-8B.

Shi, Y., Shu, C. and Pan, J. 1994. Unified optical flow field approach to motion analysis from a sequence of stereo images. *Pattern Recognition*, 12:1577-1590.

Snavely, N., Steven M. Seitz and Szeliski, R. 2008. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision*, 80(2):189-210.

Strecha, C. and Gool, L. V. 2002. Motion-stereo integration for depth estimation. In *ECCV* (pp. 495-497).

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. 2008. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Trans. on PAMI*, 30(6):1068-1080.

Tomasi, C. and Jianbo Shi. 1993. Direction of heading from image deformations. In *CVPR* (pp. 422-427).

Tsotsos, J. K. 1988. A 'complexity level' analysis of immediate vision. *International Journal of Computer Vision*, 1(4):303-320.

Vieville, T. and Faugeras, O. D. 1995. Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. In *ICCV* (pp. 750-756).

Waxman, A. M. and Duncan, J. H. 1986. Binocular image flow: steps toward stereo-motion fusion. *IEEE Trans. on PAMI*, 8:715-729.

Weber, K., Venkatesh, S., and Srinivarsan, M. V. 1997. Insect inspired behaviours for the autonomous control of mobile robots. In: Srinivasan M. V. and Venkatesh S.(Eds.), *From Living Eyes to Seeing Ma-*

chines, Oxford University Press, pp. 226-248.

Williams, O., Isard, M. and MacCormick, J. 2005. Estimating disparity and occlusions in stereo video sequences. In *CVPR* (pp. 250-257).

Zhang, H. and Negahdaripour, S. 2008. Epiflow: a paradigm for tracking stereo correspondences. *Computer Vision and Image Understanding*, 111(3):307-328.