ROBOT MOTION VISION
Part I: Theory

M. A. Taalebinezhaad

MIT Artificial Intelligence Laboratory
Cambridge, MA 02139, USA

Abstract A direct method called *fixation* is introduced for solving the general motion vision problem, arbitrary motion relative to an arbitrary environment. This method results in a linear constraint equation which explicitly expresses the rotational velocity in terms of the translational velocity. The combination of this constraint equation with the *Brightness-Change Constraint Equation* solves the general motion vision problem. Avoiding *correspondence* and *optical flow* has been the motivation behind this direct method which uses the image brightness information such as temporal and spatial brightness gradients directly. In contrast to previous direct methods, the fixation method does not put any severe restrictions on the motion or the environment. Moreover, the fixation method neither requires *tracked images* as its input nor uses *tracking* for obtaining *fixated images*. Instead, it introduces a *pixel shifting process* to construct *fixated images* for any arbitrary *fixation point*. This is done entirely in software without any use of camera motion for *tracking*.

Key Words Active Vision, Computer Vision, Feature Correspondence, Fixation, Motion Vision, Optical Flow, Pixel Shifting, Robot Vision

چکیده دراین مقاله، روش مستقیمی بنام «تثبیت» (Fixation) ارائه شده است که مسئله «دید حرکتی» (Motion Vision)، یعنی پیدا کردن حرکت و شکل، با استفاده از یک سری تصاویر الکترونیکی را در حالت کلی حل می نماید. دستاورد این روش معادله ای است که بطور مشخص حرکت دورانی را بصورت یک تابع خطی از حرکت خطی بیان می دارد. این روش، اطلاعات مربوط به «میزان درخشندگی» (Brightness) تصویر مانند مشتق روشنائی در زمان و فضا را بطور مستقیم مورد استفاده قرار داده ولی از بکارگیری «ارتباط نقاط ممیزه» (Feature Correspondence) یا «جریان نوری» (Optical Flow) کاملاً اجتناب می ورزد. برخلاف روشهای مستقیم قبلی، روش «تثبیت» هیچ محدودیت خاصی روی حرکت اجسام و یا شکل اشیاء نمی گذارد. بعلاوه، این روش نه نیاز به «تصاویر تثبیت شده» (Fixated Images) دارد و نه از روش «پیگیری» (Tracking) برای دستیابی به تصاویر تثبیت شده استفاده می کند. درعوض، روش سریع و مطمئنی بنام «جابجائی سلول تصویری» (Pixel Shifting Process) برای هر نقطه تثبیت دلخواه ارائه می شود. روش «جابجائی سلول تصویری» بطور کامل با نرم افزار حرکت دادن دوربین انجام می پذیرد.

INTRODUCTION

In motion vision, the goal is to recover, from time varying images, the relative motion between an observer and an environment as well as the structure of objects in the environment. A survey of previous literatures on machine vision is given by Barron [1]. Some of the current issues in image flow theory and motion vision are discussed by Waxman & Wohn [2], and Aloimonos & Shulman [3]. Much of the earlier work on recovering motion has been based on either establishing correspondences between the images of prominent features (points, lines, contours, and so on) in an image sequence, *correspondence*, (for example, Prazdny [4], Ullman [5,6], Longuet-Higgins [7], and Aloimonos & Basu [8] or establishing the velocity of points over the whole image, commonly referred to as the *optical flow* (for example, Ballard & Kimball [9], Bruss & Horn [10], and Adiv [11].

In general, identifying features here means determining gray-level corners. For images of smooth objects, it is difficult to find good features or corners. Furthermore, the *correspondence* problem has to be sloved,that is, feature points from consecutive frames have to be matched.

The computation of the local flow field exploits a constraint equation between the local brightness changes and the two components of the *optical flow*. This only gives the components of flow in the direction of the brightness gradient. To compute the full flow field, one needs additional constraints such as the heuristic assumption that the flow field is locally smooth [12,13]. This leads to an estimated optical flow field that is not the same as the true motion field.

Both solving the *correspondence* problem, and computing *optical flow* reliably, have proven to be rather difficult and computationally expensive. This has motivated the investigation of *direct methods* which use the image brightness information directly to recover the motion and shape.

Recently, direct motion vision methods have used the *Brightness-Change Constraint Equation* (BCCE) for solving the motion vision problem in special cases such as *Known Depth* [12], *Pure Translation* or *Known Rotation* [14], *Pure Rotation* [14], and *Planar World* [15].

In this work, a direct method called *fixation* is introduced which results in a constraint equation that explicitly expresses the rotational velocity as a linear function of the translational velocity. The combination of this *fixation constraint equation* and the BCCE offers a solution to the motion vision problem of arbitrary motion relative to an arbitrary rigid environment. That is, it recovers the shape, rotational velocity and translational velocity without putting any severe restrictions on the motion or the shape.

The *fixation method* is not only different from the previous tracking methods, but is also general. For example, Aloimons & Tsakiris [16] propose a method for tracking a target of known shape; Bandopadhay et al. [17] use optical flow and feature correspondence for tracking the principal point in order to find the motion in a special case (no rotation along the optical axis) without considering noise; and Sandini & Tistarelli [18] use an optical flow based tracking method for finding the depth in a special case (no rotation along the optical axis). Also, Thompson [19] introduces an optical flow method for recovering the motion in the special case where the rotational velocity along the optical axis is zero. His method requires a sequence of tracked images at the principal point but he acknowledges that the actual implementation of such tracking requirement in engineering system is not possible yet.

In contrast to these tracking methods, the *fixation method* dose not require tracked images as its input. Instead, it introduces a *pixel shifting process* for constructing the fixated images at any arbitrary *fixation point*. This is done entirely in software without any use of camera motion for *tracking*.

A block diagram of the ideas behind this work is shown in Figure 1. We start with a brief review of the BCCE in section 1. Then in section 2, it is shown that by choosing an arbitrary *fixation point* in the image plane, $\mathbf{r}_o$, and knowing the component of rotational velocity along the position vector of the fixation point $\omega_{\mathbf{R}o}$, we can obtain a *Fixation Constraint Equation* (FCE) which explicitly expresses the *rotational velocity* ω as a linear function of the *translational velocity* $\mathbf{t}$. Section 3 shows how the FCE can be combined with the BCCE and applied to *fixated images* in order to find $\mathbf{t}$, ω and depth map Z in the general case. Recovering $\omega_{\mathbf{R}o}$, used in the FCC, and finding the components of *fixation velocity*, u_o and v_o, necessary for constructing a fixated 2nd image, are discussed in section 4. In order to apply the FCE, a sequence of two fixated images is needed. Initial lst image is used directly and section 5 shows how a fixated 2nd image is constructed from the initial 2nd image. Partial implementation of the fixation method on real images, Taalebinezhaad [20], supports some

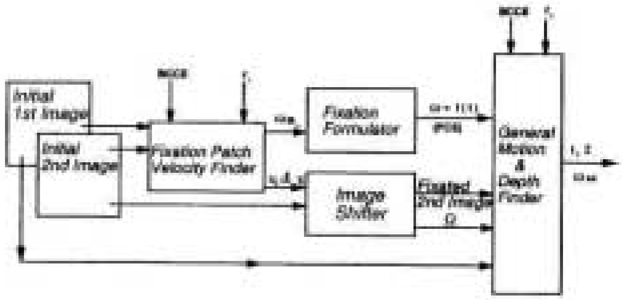

Figure 1: The fixation method modules.

of the presented algorithms in this work.

THE BRIGHTNESS CHANGE CONSTRAINT EQUATION

Using a viewer-centered coordinate system which is adopted from Longuet-Higgins & Prazdny [21] is very common in direct motion vision. Figure 2 depicts the coordinate system under consideration.

In such a coordinate system, a world point

$$\mathbf{R} = (XYZ)^T \tag{1}$$

is imaged at

$$\mathbf{r} = (xy1)^T. \tag{2}$$

That, is the image plane has the equation $Z = 1$ or in other words the *focal length* f is 1. The origin is at the projection center and the Z-axis runs along the optical axis. The X- and Y- axes are parallel to the x- and y-axes of the image plane. Image coordinates are measured relative to the *principal point*, the point $(0 \;\; 0 \;\; 1)^T$ where the optical axis pierces the image plane. The position vectors $\mathbf{r}$ and $\mathbf{R}$ are related by the perspective projection equation

$$\mathbf{r} = (x\; y\; 1\,)^T = \left(\frac{X}{Z}\frac{Y}{Z}\frac{Z}{Z}\right)^T = \frac{\mathbf{R}}{\mathbf{R}\,.\,\hat{\mathbf{z}}} \tag{3}$$

where $\hat{\mathbf{z}}$ denotes the unit vector along the Z-axis and $\mathbf{R}.\,\hat{\mathbf{z}} = Z$.

When the observer moves with instantaneous translational velocity $\mathbf{t} = (U\; V\; W\,)^T$ and instantaneous rotational velocity $\omega = (A \;\; B \;\; C\;)^T$ relative to an environment, then the time derivative of the position vector of a point in the environment, $\mathbf{R}$, relative to the observer can be written as

$$\mathbf{R}_t = -\mathbf{t} - \omega \times \mathbf{R}. \tag{4}$$

The motion of the world point $\mathbf{R}$ results in the motion of its corresponding image point $\mathbf{r}$. It can be shown [15] that the motion field in the image plane is obtained by differentiating Equation 3 with respect to time as

$$\mathbf{r}_t = \frac{d}{dt}\left(\frac{\mathbf{R}}{\mathbf{R}\,.\,\hat{\mathbf{z}}}\right) = \frac{\hat{\mathbf{z}} \times (\mathbf{R}_t \times \mathbf{r})}{\mathbf{R}\,.\,\hat{\mathbf{z}}}. \tag{5}$$

Substituting for $\mathbf{R}$, $\mathbf{r}$ and $\mathbf{R}_t$ from Equations 1, 2 and 4 into Equation 5 gives [21, 10]

$$\mathbf{r}_t = \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \frac{-U+xW}{Z} + Axy - B\,(x^2+1) + Cy \\ \frac{-V+yW}{Z} - Bxy + A\,(y^2+1) - Cx \\ 0 \end{pmatrix}. \tag{6}$$

This result is just the *parallax equations of photogrammetry* that occur in the incremental adjustment of relative orientation [22, 23]. It shows how, given the environment motion, the motion field can be

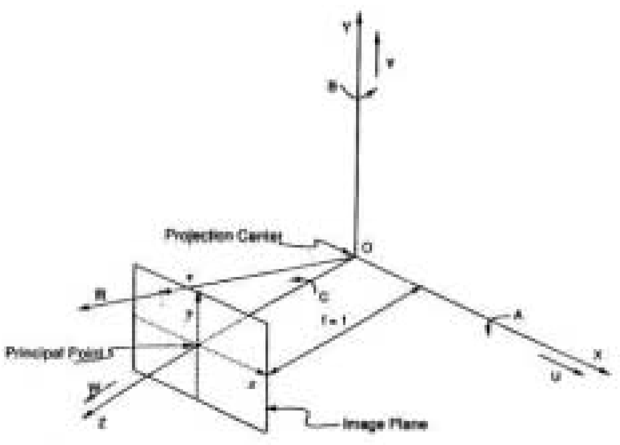

Figure 2: The viewer-centered coordinate system. The translational velocity of the camera is $\mathbf{t} = (U\; V\; W\;)^T$, and its rotational velocity is $\omega = (A\; B\; C\,)^T$.

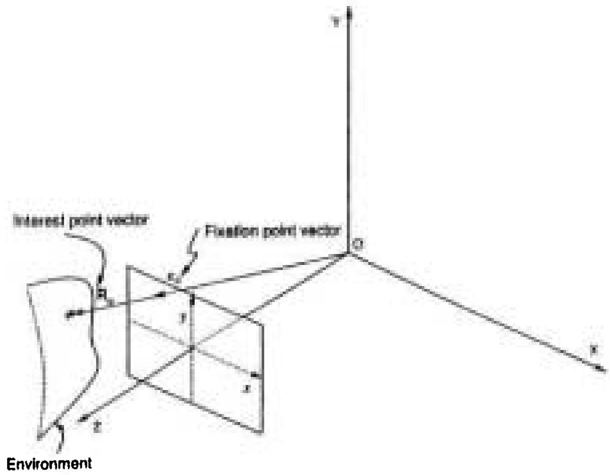

Figure 3: In fixation method, the image of the interest point, is kept stationary in the image plane despite the relative motion between the camera and the environment.

Derivation of General Fixation Constraint Equation

For a sequence of two fixated images, at fixation point we should have

$$\mathbf{r}_{of} = 0 \tag{11}$$

where $\mathbf{r}_{of}$ is the time derivative of the fixation point vector and similar to Equation 5 it can be written as

$$\mathbf{r}_{of} = \frac{\hat{\mathbf{z}} \times (\mathbf{R}_{of} \times \mathbf{r}_o)}{\mathbf{R}_o . \hat{\mathbf{z}}}. \tag{12}$$

$\mathbf{R}_{of}$ is the time derivative of the interest point vector. Combination of Equations 11 and 12 shows that for fixation we need to have

$$\hat{\mathbf{z}} \times (\mathbf{R}_{of} \times \mathbf{r}_o) = 0. \tag{13}$$

In other words, we want to find out when $\mathbf{R}_{of} \times \mathbf{r}_o$ is zero or parallel to $\hat{\mathbf{z}}$. For $\mathbf{R}_{of} \times \mathbf{r}_o$ to be parallel to $\hat{\mathbf{z}}$, we should have $\mathbf{r}_o$, perpendicular to $\hat{\mathbf{z}}$ which is not possible with a finite field of view, hence only $\mathbf{R}_{of} \times \mathbf{r}_o = 0$ applies. Conclusively, considering that $\mathbf{R}_o$ and $\mathbf{r}_o$ have the same direction, Equation 13 is simplified as

$$\mathbf{R}_{of} \times \mathbf{R}_o = 0 \tag{14}$$

Now substituting for $\mathbf{R}_{of} = -\mathbf{t} - \omega \times \mathbf{R}_o$, Equation 4, into Equation 14 gives

$$(\omega \times \mathbf{R}_o) \times \mathbf{R}_o + \mathbf{t} \times \mathbf{R}_o = 0. \tag{15}$$

Expansion of Equation 15 by using $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{c}.\mathbf{a})\, \mathbf{b} - (\mathbf{c}.\mathbf{b})\, \mathbf{a}$ results in

$$(\mathbf{R}_o . \omega)\, \mathbf{R}_o - (\mathbf{R}_o . \mathbf{R}_o)\, \omega + \mathbf{t} \times \mathbf{R}_o = 0. \tag{16}$$

As long as the translational velocity $\mathbf{t}$ is neither zero nor parallel to the interest point vector $\mathbf{R}_o$, then any vector, including ω, can be expressed in terms of the triad of vectors $\mathbf{R}_o$, $\mathbf{t} \times \mathbf{R}_o$ and $\mathbf{t}$. So we can write ω in its general form as

$$\omega = \alpha \mathbf{R}_o + \beta (\mathbf{t} \times \mathbf{R}_o) + \gamma \mathbf{t} \tag{17}$$

where α, β and γ are parameters to be determined. Later in this section we will consider the special cases where $\mathbf{t}$ is zero or parallel to $\mathbf{R}_o$ by defining ω based on another triad of vectors.

Substituting for ω from Equation 17 into Equation 16 yields

$$[1 - \beta (\mathbf{R}_o . \mathbf{R}_o)]\, (\mathbf{t} \times \mathbf{R}_o) + \gamma\, (\mathbf{R}_o . \mathbf{t})\, \mathbf{R}_o - \gamma\, (\mathbf{R}_o . \mathbf{R}_o)\mathbf{t} = 0. \tag{18}$$

Now, we should find the parameters β and γ such that Equation 18 holds without putting any restrictions on either $\mathbf{R}_o$ or $\mathbf{t}$. We start by finding the dot product of Equation 18 by $\mathbf{t} \times \mathbf{R}_o$ which results in

$$[1 - \beta (\mathbf{R}_o . \mathbf{R}_o)]\, \| \mathbf{t} \times \mathbf{R}_o \|^2 = 0. \tag{19}$$

Equation 19 will hold without restricting either $\mathbf{R}_o$ or $\mathbf{t}$ if

$$\beta = \frac{1}{\|\mathbf{R}_o\|^2}. \tag{20}$$

Another possibility for satisfying Equation 19 is to have $\|\mathbf{t} \times \mathbf{R}_o\| = 0$ which implies that either $\mathbf{t}$ or $\mathbf{R}_o$ is zero, or $\mathbf{t}$ is parallel to $\mathbf{R}_o$. But $\mathbf{R}_o$ cannot be zero and we have also assumed that here $\mathbf{t}$ is neither zero nor parallel to $\mathbf{R}_o$. As a result, $\|\mathbf{t} \times \mathbf{R}_o\|$ can not be zero.

Similarly the dot product of Equation 18 by $\mathbf{t}$ yields

$$\gamma(\mathbf{R}_o \cdot \mathbf{t})(\mathbf{R}_o \cdot \mathbf{t}) - \gamma(\mathbf{R}_o \cdot \mathbf{R}_o)(\mathbf{t} \cdot \mathbf{t}) = 0. \quad (21)$$

Knowing that $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{c} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{b} \cdot \mathbf{c})(\mathbf{d} \cdot \mathbf{a})$, Equation 21 can be simplified as

$$\gamma \|\mathbf{t} \times \mathbf{R}_o\|^2 = 0. \quad (22)$$

We discussed that $\|\mathbf{t} \times \mathbf{R}_o\|$ cannot be zero here hence Equation 22 is satisfied only if γ is zero

$$\gamma = 0. \quad (23)$$

Substituting for β from Equation (20) and γ from Equation 23 into Equation 17 gives

$$\omega = \alpha \mathbf{R}_o + \frac{1}{\|\mathbf{R}_o\|^2}(\mathbf{t} \times \mathbf{R}_o) \quad (24)$$

where α is still unknown. This means that the component of the rotational velocity along $\mathbf{R}_o$ cannot be determined by the *fixation formulation*. Physically this makes sense because the rotational velocity along $\mathbf{R}_o$, denoted by ω_{R_o}, dose not move the fixation point. This observation leads us to find $\omega_{\mathbf{R}_o}$ in a separate step before using the fixation formulation results. Derivation of ω_{R_o}, will be shown in section 4.

As a result, the *fixation constraint equation* (FCE) is written as

$$\omega = \omega_{R_o} \widehat{\mathbf{R}}_o + \frac{1}{\|\mathbf{R}_o\|}(\mathbf{t} \times \widehat{\mathbf{R}}_o) \quad (25)$$

where $\mathbf{t}$ is the translational velocity and $\widehat{\mathbf{R}}_o = \hat{\mathbf{r}}_o$ is the unit vector along the position vector of an arbitrary fixation point, an arbitrary point in the image chosen for fixation. Equation 25 shows that after *fixation*, the rotational velocity ω can be explicitly expressed as a linear function of the translational velocity $\mathbf{t}$.

Derivation of Special Fixation Constraint Equation

When the translational velocity $\mathbf{t}$ is zero or parallel to the interest point vector $\mathbf{R}_o$, Equation 16 is simplified as

$$(\mathbf{R}_o \cdot \omega)\mathbf{R}_o - (\mathbf{R}_o \cdot \mathbf{R}_o)\omega = 0. \quad (26)$$

This time, ω is defined based on the triad consisting of vectors $\mathbf{R}_o$, $\hat{\mathbf{x}}$, and $\hat{\mathbf{x}} \times \mathbf{R}_o$ as

$$\omega = l\,\mathbf{R}_o + m(\hat{\mathbf{x}} \times \mathbf{R}_o) + n\hat{\mathbf{x}} \quad (27)$$

where l, m, and n are parameters to be determined. Here we assume that $\mathbf{R}_o$ is not parallel to $\hat{\mathbf{x}}$. This is a reasonable assumption because otherwise we should at least have a field of view of 180° to be able to choose an awkward interest point along the X-axis, which results in a fixation point at infinite distance from the principal point and near the border of an infinite image plane.

Substituting for ω from Equation 27 into Equation 26 yields

$$[m\,\mathbf{R}_o \cdot (\hat{\mathbf{x}} \times \mathbf{R}_o) + n(\mathbf{R}_o \cdot \hat{\mathbf{x}})]\,\mathbf{R}_o - m(\mathbf{R}_o \cdot \mathbf{R}_o)(\hat{\mathbf{x}} \times \mathbf{R}_o) - n(\mathbf{R}_o \cdot \mathbf{R}_o)\hat{\mathbf{x}} = 0. \quad (28)$$

The dot product of Equation 28 with $(\hat{\mathbf{x}} \times \mathbf{R}_o)$ results in

$$-m(\mathbf{R}_o \cdot \mathbf{R}_o)\|(\hat{\mathbf{x}} \times \mathbf{R}_o)\|^2 = 0. \quad (29)$$

Considering that $\mathbf{R}_o$ is not zero nor parallel to $\hat{\mathbf{x}}$, Equation 29 is satisfied only if m is zero

$$m = 0. \quad (30)$$

Substituting for m into Equation 28 and finding its dot product by $\hat{\mathbf{x}}$ results in

$$n\,(\mathbf{R}_o \cdot \hat{\mathbf{x}})\,(\mathbf{R}_o \cdot \hat{\mathbf{x}}) - n\,(\mathbf{R}_o \cdot \mathbf{R}_o)\,(\hat{\mathbf{x}} \cdot \hat{\mathbf{x}}) = 0. \tag{31}$$

Using $(\mathbf{a}\times\mathbf{b}) \cdot (\mathbf{c}\times\mathbf{d}) = (\mathbf{c.a})\,(\mathbf{b.d}) - (\mathbf{b.c})\,(\mathbf{d.a})$, Equation 31 can be written as

$$n\,\|\hat{\mathbf{x}} \times \mathbf{R}_o\|^2 = 0. \tag{32}$$

Again $\mathbf{R}_o$ cannot be either zero or parallel to a result, Equation 32 will hold for arbitrary $\mathbf{R}_o$ if $n = 0$. Substituting for n and m into Equation 27 gives

$$\omega = l\,\mathbf{R}_o \tag{33}$$

where l is still unknown. We can substitute $\omega_{Ro}\hat{\mathbf{R}}_o$ for $l\,\mathbf{R}_o$. The procedure for computing ω_{Ro} is given in section 5. Consequently, for the special cases we obtain the *special fixation constraint equation* (SFCE) as

$$\omega = \omega_{Ro}\hat{\mathbf{R}}_o \tag{34}$$

which means that when the translational velocity t is zero or parallel to $\mathbf{R}_o$ then the corresponding rotational velocity may only have a component along $\mathbf{R}_o$.

This procedure for deriving the SFCE, Equation 34, is not essentially different from what we did for deriving the FCE, Equation 25. In fact, Equation 34 is a special case of Equation 25. But we did not directly derive Equation 34 from Equation 25 because Equation 25 was derived based on the assumption that **t** is neither zero nor parallel to $\mathbf{R}_o$. As a result, for implementation it is enough to use the FCE, Equation 25, without knowing whether the present condition is the special case or not.

3. SOLVING THE GENERAL DIRECT MOTION VISION PROBLEM

At this step, we assume that a sequence of two fixated images are constructed. In other words, we have made the fixation point stationary in the image plane. This can be done first by finding the *fixation velocity*, the apparent velocity at the fixation point in the lst image, as shown in section 4. Then the *pixel shifting process* explained in section 5 can be used for constructing a new image, *fixated 2nd image*, in which the image of interest point is positioned at the same point as in the initial lst image.

We start by studying the general case where the translational velocity **t** is neither zero nor parallel to the interest point vector $\mathbf{R}_o$. The special cases of **t** will be discussed later.

Substituting for ω from the fixation constraint Equation 25 into the brightness-change constraint Equation 8 gives

$$E_t + \omega_{Ro}\,\mathbf{v}.\,\hat{\mathbf{R}}_o + \frac{1}{\|\mathbf{R}_o\|}\,[\mathbf{v}.\,(\mathbf{t} \times \hat{\mathbf{R}}_o)] + \frac{1}{Z}(\mathbf{s.t}) = 0. \tag{35}$$

Knowing that $\mathbf{a}.(\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}).\,\mathbf{c}$ and doing some manipulations on Equation 35 results in

$$E_t' + [\frac{1}{Z}\mathbf{s} - \frac{1}{\|\mathbf{R}_o\|}(\mathbf{v} \times \hat{\mathbf{R}}_o)].\,\mathbf{t} = 0 \tag{36}$$

where E_t' is a notation for $E_t + \omega_{Ro}\,\mathbf{v}.\,\hat{\mathbf{R}}_o$ which is computable at any pixel. In general, Equation 36 can be solved numerically for **t** and Z using images of any size and with any field of view. For a small patch around the fixation point, *fixation patch*, Equation 36 can be simplified as

$$E_t' + (\frac{1}{Z} - \frac{1}{Z_o})\,(\mathbf{s}\,.\,\mathbf{t}) \approx 0.^{[2]} \tag{37}$$

[2]Considering that $\|\mathbf{R}_o\| = Z_o\,\|\mathbf{r}_o\|$ and $\mathbf{v} = \mathbf{r} \times \mathbf{s}$, the term $\frac{1}{\|\mathbf{R}_o\|}(\mathbf{v} \times \hat{\mathbf{R}}_o)$ from Equation 36, let's call it K, can be expanded as

$$K = \frac{1}{Z_o\|\mathbf{r}_o\|}(\mathbf{r}\times\mathbf{s})\times\frac{\mathbf{r}_o}{\|\mathbf{r}_o\|}.$$

Further expansion of K by using the relation $(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{c.a})\,\mathbf{b} - (\mathbf{c.b})\mathbf{a}$, results in →

In analogy to the pure translation case of Horn & Weldon [14], we can find the translational velocity **t**. Equation 37 shows that $1/(\frac{1}{Z}-\frac{1}{Z_o})=\frac{\mathbf{s}.\mathbf{t}}{E_t'}$. At the points that E_t' is very small, even small error in computing **t** will result in large error in $1/(\frac{1}{Z}-\frac{1}{Z_o})$ which translates into large error in the estimation of depth Z. Considering this fact, the true translational velocity **t** can be found from Equation 37 by minimizing

$$J=\iint(\frac{1}{Z}-\frac{1}{Z_o})^{-2}dxdy=\iint(\frac{\mathbf{s}.\mathbf{t}}{E_t'})^2dxdy \qquad (38)$$

with respect to **t**. In other words, we are looking for the true motion **t** which minimizes the sum of squares of $\frac{\mathbf{s}.\mathbf{t}}{E_t'}$ over the fixation patch. We also put the $\|\mathbf{t}\| = 1$ constraint on this minimization problem to avoid the trivial solution **t** = 0. This is a valid constraint on **t** because due to the *scale factor ambiguity* we can only find the direction of **t**. This constraint on **t** can be written as

$$\mathbf{t}^T\mathbf{t} = 1. \qquad (39)$$

Moreover we can rewrite J as

$$J = \mathbf{t}^T M\,\mathbf{t} \qquad (40)$$

where M is a fully computable 3×3 symmetric matrix

$$M = \iint (\frac{1}{E_t'})^2\,\mathbf{s}\mathbf{s}^T dxdy. \qquad (41)$$

Minimizing J in Equation 40 under the constraint Equation 39 is an ordinary calculus constrained minimization problem which can be solved by minimizing

$$I(\mathbf{t}, \lambda) = \mathbf{t}^T M\,\mathbf{t} + \lambda(1 - \mathbf{t}^T\mathbf{t}) \qquad (42)$$

with respect to **t** and the Lagrange multiplier λ. Then, we will obtain

$$\frac{\partial I}{\partial t}=2Mt-2\lambda\mathbf{t}=0 \qquad (43)$$

which is simplified as

$$M\,\mathbf{t} = \lambda\mathbf{t}. \qquad (44)$$

Equation 44 is an eigenvalue problem where λ is an eigenvalue of the known matrix M and **t** is the corresponding eigenvector. The eigenvalues of M are real and nonnegative because M is a positive semidefinite Hermitian matrix. Substituting for M **t** from Equation 44 into Equation 42 gives $I = \lambda$ which implies that under the given constraint, $\mathbf{t}^T M\mathbf{t}$ is minimized when the smallest of three real and nonnegative eigenvalues is used for computing the eigenvector **t**.

It is shown that the fixation method can be used for solving the motion vision problem in its general case. The translational velocity **t** is obtained from Equation 44 by using the smallest eigenvalue and computing its corresponding eigenvector. Then we can use Equation 36 for finding the depth map, depth at all image points, as

$$Z=\frac{(\mathbf{s}.\mathbf{t})}{\frac{(\mathbf{v}\times\hat{\mathbf{R}}_o).\mathbf{t}}{\|\mathbf{R}_o\|}-E_t'} \qquad (45)$$

Then, Equation 25 gives the partial rotational velocity ω

$$\omega=\omega_{R_o}\hat{\mathbf{R}}_o+\frac{1}{\|\mathbf{R}_o\|}(\mathbf{t}\times\hat{\mathbf{R}}_o) \qquad (46)$$

where $\|\mathbf{R}_o\| = Z_o\|\mathbf{r}_o\|$ and Z_o, depth at the fixation point,

---

$$K=\frac{1}{Z_o\|\mathbf{r}_o\|^2}[(\mathbf{r}_o.\mathbf{r})\mathbf{s}-(\mathbf{r}_o.\mathbf{s})\mathbf{r}].$$

It is clear at the fixation point, where $\mathbf{r} = \mathbf{r}_o$ and $\mathbf{s} = \mathbf{s}_o$, $K=\frac{1}{Z_o}\mathbf{s}_o$ and for the points near the fixation point $K\approx\frac{1}{Z_o}\mathbf{s}$

is obtained from Equation 36[3].

The total rotational velcioty of the observer relative to the environment is obtained by adding ω to the *equivalent rotational velocity* $\boldsymbol{\Omega}$ given in section 6. It can be seen that for the general case, the *fixation formulation* lets us find the shape and motion by an arbitrary choice of a fixation point $\mathbf{r}_o$.

Special Cases: t Is Zero or Parallel to $\mathbf{R}_o$

When the translational velocity **t** is zero, we showed that the partial rotational velocity ω has only a component along $\mathbf{R}_o$, Equation 34. This component is obtained from Equation 51. There are also methods for finding the total rotational velocity using the initial unfixated images [14]. In the case of $\mathbf{t} = 0$, we basically cannot obtain any estimation for the depth Z.

For the other special case where **t** is parallel to $\mathbf{R}_o$, we substitute for ω from Equation 34 into the BCCE Equation 8 to obtain

$$E_t' + \frac{1}{Z}(\mathbf{s}.\mathbf{t}) = 0 \qquad (47)$$

where E_t is again a notation for the computable value of $E_t + \omega_{Ro}\ \mathbf{v}.\ \widehat{\mathbf{R}}_o$. Because no approximation is involved in deriving Equation 47, an exact closed form solution exists for **t** and Z without any restriction on the field of view or the size of fixation patch. This exact solution for finding **t** and Z is the same as the solution given in the general case, starting from Equation 38, except that J is defined as $\iint Z^2\, dxdy$ for this special case.

4. COMPUTING THE FIXATION VELOCITY AND ω_{R_o}

The fixation formulation is based on the assumption that the fixation point is kept stationary in a sequence of two fixated images. We use the term *fixation velocity* to refer to the apparent velocity at the fixation point in the initial 1st image. The x and y components of the fixation velocity are represented by u_o and v_o respectively. The basic fixation requirement, a sequence of two fixated images in which $\mathbf{r}_{ot} = 0$, can be satisfied by finding u_o and v_o, and then using these components for obtaining a new image, *fixated 2nd image*. The *pixel shifting process* for obtaining the fixated 2nd image is explained in the next section.

We also saw that the component of the rotational velocity along $\mathbf{R}_o$, ω_{Ro}, cannot be obtained from the fixation formulation because this component does not move the fixation point. Here, we will introduce algorithms which can be used for finding ω_{Ro} and also the components of the fixation velocity, u_o and v_o.

The motion field velocity due to the component of the rotational velocity of the observer relative to the environment along $\mathbf{R}_o$ is given by $-(\omega_{\hat{R}_o} \times \mathbf{r}) = -\omega_{Ro}(\widehat{\mathbf{R}}_o \times \mathbf{r}) = -\frac{\omega_{Ro}}{\|\mathbf{r}_o\|}(\mathbf{r}_o \times \mathbf{r})$, where $\widehat{\mathbf{R}}_o = \hat{\mathbf{r}}_o$ is the unit vector along $\mathbf{r}_o$. Assuming that depth is approximately the same on the *fixation patch*, a small patch around the fixation point, and substituting $\mathbf{r}_o = (x_o\ y_o\ 1)^T$ and $\mathbf{r} = (x\ y\ 1)^T$ the components of the total motion field velocity due to fixation velocity and ω_{Ro},

---

[3]At the fixation point, Equation 36 is exactly expanded to an equation similar to Equation 37

$$E_t + \omega_{Ro}\ \mathbf{v}_o \,.\, \widehat{\mathbf{R}}_o + \left(\frac{1}{Z_o}\frac{1}{Z_o}\right)(\mathbf{s}_o\,.\,\mathbf{t}) = 0.$$

Theoretically, all terms of this equation vanish because E_t is zero at the fixation point, and $\mathbf{v}.\mathbf{r} = 0$ is true for all points. But at any other *i* around the fixation point, the depth Z_{oi} can be obtained from Equation 36 as

$$Z_{oi} = \frac{1}{E_{ti}'}\left[\frac{\mathbf{v}_i \times \mathbf{r}_o}{\|\mathbf{r}_o\|^2} - \mathbf{s}_i\right].\mathbf{t}$$

By averaging N of such neighboring depths, we can easily find the depth Z_o as

$$Z_o = \frac{1}{N\,\|\mathbf{r}_o\|}\mathbf{t}\,.\sum_{i=1}^{i=N}\left[\frac{\mathbf{v}_i \times \mathbf{r}_o - \|\mathbf{r}_o\|^2\,\mathbf{s}_i}{E_{ti}\|\mathbf{r}_o\| + \omega_{Ro}(\mathbf{v}_i\,.\,\mathbf{r}_o)}\right]$$

where $\mathbf{s}_i$, $\mathbf{v}_i$, and E_{ti} are computed at **N** points around the fixation point.

are given by

$$\begin{cases} x_t = u_o - \frac{\omega_{Ro}}{\|\mathbf{r}_o\|}\hat{\mathbf{x}}\,.\,(\mathbf{r}_o \times \mathbf{r}) = u_o + \overline{\omega}_{Ro}(y - y_o) \\ y_t = v_o - \frac{\omega_{Ro}}{\|\mathbf{r}_o\|}\hat{\mathbf{y}}\,.\,(\mathbf{r}_o \times \mathbf{r}) = v_o - \overline{\omega}_{Ro}(x - x_o) \end{cases} \quad (48)$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit vectors along the x and y axes and $\overline{\omega}_{Ro}$ is a notation for $\frac{\omega_{Ro}}{\|\mathbf{r}_o\|}$. Substituting for x_t and y_t from the above equations into the BCCE, Equation 7, yields

$$[u_o + \overline{\omega}_{Ro}(y - y_o)]\,E_x + [v_o - \overline{\omega}_{Ro}(x - x_o)]\,E_y + E_t = 0 \quad (49)$$

Due to noise, Equation 49 odes not necessarily hold for any point (x,y) so we try to find u_o, v_o and $\overline{\omega}_{Ro}$ by minimizing the sum of squares of errors over the fixation patch. In other words we want to minimize

$$\iint [(u_o + \overline{\omega}_{Ro}(y - y_o))\,E_x + (v_o - \overline{\omega}_{Ro}(x - x_o))\,E_y + E_t]^2\,dxdy \quad (50)$$

with respect to u_o, v_o and $\overline{\omega}_{Ro}$. This results in a system of three linear equations that can be solved for the three unknowns

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{pmatrix} u_o \\ v_o \\ \overline{\omega}_{Ro} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}. \quad (51)$$

Matrix **A** is symmetric and its elements are given by

$$\begin{cases} a_{12} = \iint E_x E_y\,dx\,dy \\ a_{13} = \iint E_x\,[E_x(y - y_o) - E_y(x - x_o)]\,dx\,dy \\ a_{23} = \iint E_y\,[E_x(y - y_o) - E_y(x - x_o)]\,dx\,dy \\ a_{11} = \iint E_x^2\,dx\,dy \\ a_{22} = \iint E_y^2\,dx\,dy \\ a_{33} = \iint [E_x(y - y_o) - E_y(x - x_o)]^2\,dx\,dy \end{cases} \quad (52)$$

and the components of vector C are as follows:

$$\begin{cases} c_1 = -\iint E_t E_x\,dx\,dy \\ c_2 = -\iint E_t E_y\,dx\,dy \\ c_3 = -\iint E_t\,[E_x(y - y_o) - E_y(x - x_o)]\,dx\,dy\,. \end{cases} \quad (53)$$

Considering that the fixation point coordinates x_o and y_o are known, then the sets of Equations 52 and 53 show that the element of matrix **A** and the components of vector **C** are fully computable.

In the special case where the fixation point is at the principal point, $x_o = y_o = 0$, elements of matrix **A** are simplified as

$$\begin{cases} a_{12} = \iint E_x E_y\,dx\,dy \\ a_{13} = \iint E_x\,(yE_x - xE_y)\,dx\,dy \\ a_{23} = \iint E_y\,(yE_x - xE_y)\,dx\,dy \\ a_{11} = \iint E_x^2\,dx\,dy \\ a_{22} = \iint E_y^2\,dx\,dy \\ a_{33} = \iint (yE_x - xE_y)^2\,dx\,dy \end{cases} \quad (54)$$

and the components of vector C are given as follows

$$\begin{cases} c_1 = -\iint E_t\,E_x\,dx\,dy \\ c_2 = -\iint E_t\,E_y\,dx\,dy \\ c_3 = -\iint E_t\,(y\,E_x - x\,E_y)\,dx\,dy\,. \end{cases} \quad (55)$$

After finding $\overline{\omega}_{Ro}$, we can easily calculate ω_{Ro} as

$$\omega_{Ro} = \overline{\omega}_{Ro}\sqrt{x_o^2 + y_o^2 + 1}. \quad (56)$$

When the fixation point is at the principal point, ω_{Ro} become equal to $\overline{\omega}_{Ro}$.

5. CONSTRUCTING A FIXATED 2ND IMAGE

The fixation method requires a sequence of two images in which the fixation point is kept stationary. Referring to Figure 1, we are given two initial images. The *initial 1st image* is used directly but we need to find a *fixated 2nd image.*

Physical rotation of the camera relative to the observer base is a hardware solution to this problem which is basically a *tracking* problem. Considering that in general the interest point has a motion relative to the observer, the fixated 2nd image cannot be obtained in one step. As a result, a feedback control loop is requied for the camera rotation system to compensate for the errors resulted from the new position of the fixation point. This tracking approach is avoided not only because of the errors involved but also because of the concern about the real time applications.

In the following, we will show how a fixated 2nd image can be constructed by applying an imaginary rotation to the initial 2nd image through software, *pixel shifting process.* It is assumed here that the fixation velocity has been already computed from Equation 51. We introduce an *equivalent rotational velocity* $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ which could result in the same fixation velocity (u_o, v_o) at the fixation point (x_o, y_o). According to Equation 6, the components of Ω must satisfy the following set of equations

$$\begin{cases} u_o = x_o y_o \Omega_x - (x_o^2 + 1)\,\Omega_y + y_o \Omega_z \\ v_o = (y_o^2 + 1)\,\Omega_x - x_o y_o \Omega_y - x_o \Omega_z . \end{cases} \tag{57}$$

Among infinite number of Ω that satisfy the system of Equations 57, we choose the only one that dose not introduce any new rotational velocity along the fixation point vector $\mathbf{r}_o$. Mathematically this means that $\Omega.\mathbf{r}_o = 0$ which results in the following constraint on the components of Ω,

$$x_o \Omega_x + y_o \Omega_y + \Omega_z = 0. \tag{58}$$

This constraint guarantees that the value of ω_{Ro} obtained by applying the system of Equation 51 to the two initial images is also valid for the fixated images. As a result, no modifications in ω_{Ro} is needed before using it in Equations 45 and 46.

Given the fixation velocity (u_o, v_o) and the fixation point coordinates x_o and y_o, the equivalent rotational velocity Ω is obtained by solving the combination of three linear equations in 57 and 58. For example in the case that the fixation point is at the principal point, $x_o = y_o = 0$, the equivalent rotational velocity becomes,

$$\Omega = (v_o, -u_o, 0). \tag{59}$$

Then, the 2nd fixated image can be constructed by applying an imaginary $-\Omega$ rotation to the initial 2nd image through software, *pixel shifting process.* Considering Equation 57, the following set of equations must be satisfied in the pixel shifting process of the initial 2nd image

$$\begin{cases} u = -xy\,\Omega_x + (x^2 + 1)\,\Omega_y - y\Omega_z \\ v = -(y^2 + 1)\Omega_x + xy\,\Omega_y + x\Omega_z . \end{cases} \tag{60}$$

Here Ω_x, Ω_y and Ω_z are known values. As a result, the *shifting vector* (u, v) can be obtained for every pixel of the initial 2nd image. The brightness at pixel (x, y) of the fixated 2nd image is obtained by finding the brightness at the corresponding point $(x - Tu, y - Tv)$ in the initial 2nd image. Where T is the time interval between two initial images. In general, a computed original point is not located at the center of a pixel in the initial 2nd image. As a result, its brightness cannot be read directly from the image file and should be computed by averaging, bilinear interpolation or bicubic interpolation of the brightnesses at its neighboring pixels.

It should be clear by now that neither we assume that the fixated images are given to us in advance nor we use tracking for obtaining the fixated images. Construction of the 2nd fixated image is based on a *pixel shifting process.* This is done entirely in software and no tracking is involved.

SUMMARY

The algorithms and formulations presented in this work show how to solve directly for the motion and shape in the general case. In contrast to previous work done in the area of motion vision, our solutions are general and does not put any severe restrictions on the motion or the environment. More importantly, the fixation method uses neither *optical flow* nor *feature correspondence*. Instead, direct image information such as temporal and spatial brightness gradients are used. This method neither requires tracked images as input nor tracking for obtaining fixated images. Instead, it introduces a *pixel shifting process* for constructing fixated images at any arbitrary *fixation point*. This process is done entirely in software without any use of tracking.

Paratial implementation of the fixation method on real images has shown encouraging results which support some of the presented algorithms in this work [20].

Referring to Figure 1, the fixation method can be implemented in the following steps.

STEP 1: Finding the *fixation velocity* components (u_o, v_o) and the component of rotational velocity along $\mathbf{R}_o$, ω_{R_o}, by applying the system of Equation 51 to the direct image information from two *initial images*.

STEP 2: Knowing the fixation velocity components, u_o and v_o, the *fixated 2nd image* is constructed by the *pixel shifting process* explained in section 5. This is done entriely in software without any use of camera motion for tracking.

STEP 3: Using the fixation constraint Equation 25, initial 1st image, and the fixated 2nd image, the method presented in section 3 can be used for recovering the translational velocity, the partial rotational velocity, and the depth at all image points.

STEP 4: The total rotational velocity ω_{tot} is simply obtained by adding the *equivalent rotational velocity* Ω, from Equations 57 and 58, to the *partial rotational velocity* ω from Equation 46.

ACKNOWLEDGEMENT

I would like to express my gratitude to Professor Berthold K. P. Horn who patiently went through several versions of this work and provided me with his valuable ideas. Also many thanks to Professor Ellen C. Hildreth for her crucial and constructive comments, and to the anonymous reviewers of this work for their detailed and useful comments.

REFERENCES

1. J. Barron, "A Survey of Approaches for Determining Optical Flow, Environmental Layout and Egomotion," Technical Report RBCV-TR-84-5, University of Toronto, Ontario, Canada, November (1984).
2. A. M. Waxman and K. Wohn, "Image Flow Theory: A Frame Work for 3-D Inference from Time-Varying Imagery," Volume 1 of Advances in Computer Vision, chapter 3, Lawrence Erlbaum Assoc., Editor: C. Brown, (1988), 165-224.
3. J. Aloimonos and D. Shulman," Integration of Visual Modules, An Extension of the Marr Paradigm," Academic Press, Boston, (1989).
4. K. Prazdny, "Motion and Structure from Optical Flow," In Proc. of the Sixth Inter. Joint Conf. on Artificial Intelligence, Tokyo, Japan, August (1979).
5. S. Ullman, "The Interpretation of Visual Motion," MIT Press, Cambridge, MA, (1979).
6. S. Ullman, "The Effect of Similarity between Line Segments on the Correspondence Strength in Apparent Motion," Perception, 9, (1980), 617-626.
7. H. C. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene From Two Projections," *Nature*, 293: (1981), 133-135.

8. J. Aloimonos and A. Basu, "Determining the Translation of a Rigidly Moving Surface, without Correspondence," In Proceedings of *IEEE Conf.* on Computer Vision and Pattern Recognition, Miami, FL June (1986).
9. D. H. Ballard and O. A. Kimball, "Rigid Body Motion from Depth and Optical Flow," Computer Vision, Graphics, and Image Processing, 22(1), April (1983).
10. A. R. Bruss and B. K. P. Horn, "Passive Navigation," Computer Vision, Graphics, and Image Processing, 21(1): 3-20, Jan (1983).
11. G. Adiv, "Determining 3-d Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligene*, 7(4), (1985), 384-401.
12. B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, 17, (1981), 185-203.
13. E. C. Hildreth. The Measurement of Visual Motion. MIT Press, Cambridge, Mass., (1984).
14. B. K. P. Horn and E. J..Weldon Jr, "Direct Methods for Recovering Motion." *Inter. Journal of Computer Vision*, 2 (1988), 51-76.
15. S. Negahdaripour and B. K. P. Horn, "Direct Passive Navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1), Jan (1987), 168-176.
16. J. Aloimonos and D. P. Tsakiris, "On the Mathematics of Visual Tracking," Technical Report CAR-TR-390, Computer Vision Laboratory, University of Maryland, MD, Sep (1988).
17. A. Bandopadhay, B. Chandra, and D. H. Ballard, "Active Navigation: Tracking En environmental Point Considered Beneficial. In Proc. of IEEE Workshop on Motion: Representation and Analysis, Kiawash Island, May 7-9, (1986), pages 23-29.
18. G. Sandini and M. Tistarelli, "Active Tracking Strategy for Monocular Depth Inference Over Multiple Frames," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), Jan (1990), 13-27.
19. W. B. Thompson, "Structure-from Motion by Tracking Occlusion Boundaries, *Biological Cybernetics*, 62: (1989), 113-116.
20. M. A. Taalebinezhaad, "Partial Implementation of Fixation Method on Real Images. In Proc. of *IEEE Conf.* on Computer Vision and Pattern Recognition, Maui, Hawaii, June (1991).
21. H. C. Longuet-Higgins and K. Prazdny, "The Interpretation of a Moving Retinal Image. In Proceedings of the Royal Society of London B, (1980), pages 385-397.
22. B, Hallert, "Photogrammetry," McGraw Hill, New York, NY, (1960).
23. F. Moffit and E. M. Mikhail, "Photogrammetry," Harper & Row, New York, NY, 3rd edition, (1980).
24. M. A. Gennert and S. Negahdaripour, "Relaxing the Brightness Constancy Assumption in Computing Optical Flow," Technical Report MIT AI Memo 975, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, June (1987).
25. S. Negahdaripour, A. Shokrollahi and M. Gennert, "Relaxing the Brightness Constancy Assumption in Computing Optical Flow," In Proc. of Inter. conf. on Image Processing, Singapore, September (1989),
26. N. Cornelius and T. Kanade, "Adapting Optical-Flow to Measure Object Motion in Reflectance and X-Ray Image Sequences," Technical Report CMU-CS-83-119, Carnegie-Mellon University, Pittsburgh, PA, January (1983).
27. A. T. Bahil and T. LaRitz, "Why Can't Batters Keep their Eyes on the Ball," *American Scientist*, 72, (1984), 219-253.
28. A. T. Bahil and J. D. McDonald, "Model Emulates Human Smooth Pursuit System Producing Zero-Latency Target Tracking," *Biological Cybernetics*, 48, (1983), 213-222.
29. A. Bandopadhay, "A Computational Study of Rigid Motion Perception," PhD thesis, Dept. of Computer Science, Univ. of Rochester, (1986).