Modeling Active Vision During Smooth Pursuit of a Robotic Eye

Jacek Turski; University of Houston-Downtown; Houston, Texas, U.S.A.

Abstract

The computational framework, based on the conformal camera, is developed for processing visual information during smooth pursuit movements of a robotic eye. During smooth pursuit, the image of the tracked object remains nearly stationary while the image of a stationary background sweeps across the image plane of the camera. The background's image transformation derived in the second-order approximation enable the anticipation of the perceptual outcome of the camera pursuit. This can be used to support the correct visual information of the moving object in front of the stationary background, These results complement the author's previous study on the predictive image processing for visual stability during the conformal camera movements resembling primate's saccadic eye rotations. The visual information processing algorithms that can support visual stability during smooth pursuit and saccadic movements of an anthropomorphic robotic camera are needed for an autonomous robot efficient interactions with the real world, in real time.

Introduction

In this article we develop a computational framework for biologically-mediated processing of visual information during foveated-sensor camera rotations resembling primate's smooth pursuit eye movements (SPEM). The presented algorithms can support perceptual stability for autonomous robotic systems needed for their efficient interaction with dynamic environments.

Primate Perception During Eye Movements

Whenever primates direct the gaze to attend a scene detail, their highest acuity is confined to the foveal region subtending the visual angle of 2 degrees. To overcome the acuity limitation of foveated vision, humans and other primates explore the scene by making about 3 to 5 saccades per second with the eyeballs speed of up to 900 deg/s to fixate the high-acuity fovea on the salient and behaviorally relevant parts of the scene. In addition, primates are able to execute smooth pursuit eye movements that keep the fovea focused on a slowly moving object (up to 100 deg/s), but they also employ a combination of smooth pursuit and saccades to track an object moving unpredictably or moving faster than 30 deg/s).

Visual neurons in the primate brain have spatial receptive fields (RFs) that encode the position of an object in gaze-centered coordinates, that is, in respect to the frame centered on the fovea in retinotopic maps. Because of incessant eye movements, the retinotopic representation of this position information is constantly changing. For this reason, very sophisticated neural processes have evolved to maintain a temporally continuous, stable perception of the world.

The identification of visuosaccadic pathways [19] supports the idea that the brain uses a copy of the oculomotor command of the impending saccade, referred to as efference copy or corollary discharge (see a review in [3], to shift transiently the RFs of stimuli to their future location before the eyes saccade takes them there. This shift remaps the retinotopic maps from the presaccadic frame to the postsaccadic frame in the anticipation of each upcoming saccade, giving access to the visual information at the saccadic target before the saccade is executed. It is believed that this predictive remapping mechanism contribute to visual stability [4, 11].

Smooth pursuit eye movements (SPEMs) by stabilizing the tracked objects image on the fovea, superimpose additional motion on the retinal images of the stationary background and other moving objects, leading to the possible distortion of the perceived speed and direction of motion. Moreover, in natural viewing corrective saccades often accompany smooth pursuit eye movement [2, 17]. Whenever eyes are pursuing a moving object in front of a stationary background, the object's image is stabilized on the fovea and the background image sweeps across the retina in opposite direction. Nevertheless, we generally perceive the object as moving and the background as stationary, despite opposite retinal information. This indicates that the stable perception that we enjoy has to be maintained at all time during SPEM. Consequently, in addition to the extraretinal information such as eye intended movement (efference copy), the retinal motion information should also contribute to processing visual information during smooth pursuit [7, 9, 13].

Modeling Visual Information During Robotic Eye Movements

Visual information acquisition sensors, with software and/or hardware-implemented image processing used in anthropomorphic cameras, can be classified into two broad categories. The foveated-sensor, frame-based, video acquisition systems [1, 27] and asynchronous spiking, frame-less, contrast-variation acquisition systems [10, 16]. Frame-less based vision systems have a huge advantage over standard frame-based vision systems due to very high dynamic range and temporal resolution, but very low computational load and internal latency [14].

The frame-based cameras with foveated sensor architecture have been implemented in numerous applications in the context of machine vision, but they have not been widely used in commercial applications due to the lack of efficient tools for processing and analyzing visual information [28]. On the other hand, so far only a limited range of applications has been demonstrated for frame-less vision systems mainly because the visual information processing and analyzing tools are not available [5].

The conformal camera framework [23, 24] has been recently used in [25, 26] to model some of the front-end processes underlying stability of perception for a fovated-sensor camera head mounted on the moving platform that replicates human saccadic eye movements [6, 8]. In this study, the conformal camera's geometric framework is applied to develop anticipatory processing

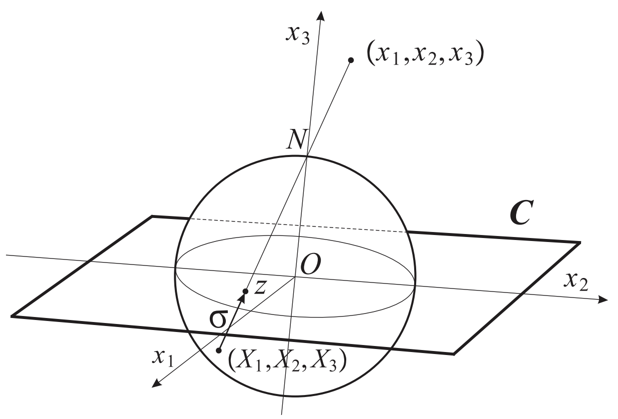

Figure 1. *The conformal camera. The points of space are centrally projected into the sphere, representing the retina, and into the image plane. The sphere and the center of projection N represent the eyeball and the pupil. The image representation suitable for computational processing is given by stereographic projection σ from the sphere (the retina) to the image plane.*

of visual information than can support perceptual stability during SPEM. The algorithms for image projective transformations are given in the initial image plane of the conformal camera during its smooth pursuit movement.

The paper is organized as follows. In Section 2, we discuss imaging with the conformal camera. Section 3 develops the approximations of the image projective transformations during a sequence of a small horizontal rotations of the conformal camera. In Section 4, we use the first-order approximation of the image projective transformations to model horizontal SPEM.

The Conformal Camera

The conformal camera was introduced in [21, 22]. We demonstrated in these references that the conformal camera possesses many features that are particularly useful in computationally efficient modeling the eye's imaging functions. In particular, the underlying geometric analysis of the conformal camera integrates the scene projected onto the rotated eye's retina with the correspondingly changing retinotopic maps into a single numerical system.

The conformal camera consists of the unit sphere S^2 and an image plane $\mathbf{C}$ through the sphere center O, see Fig. 1. The end point N of the sphere diagonal, which is orthogonal to the image plane, is the center of the projection of the spatial points into the plane.

The camera's orientation is described by the positively-oriented orthonormal frame $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, such that $\mathbf{e}_3 = \overrightarrow{ON}$. The frame is attached at the camera's center and resulting spatial coordinates are denoted by (x_1, x_2, x_3). The image plane $x_3 = 0$ is parametrized with complex coordinates $x_1 + ix_2$.

A spatial point (x_1, x_2, x_3), projected along the ray through N to the point z of the image plane, defines the mapping

$$j_N(x_1, x_2, x_3) = z = \frac{x_1 + ix_2}{1 - x_3}. \tag{1}$$

Taking restriction $\sigma = j_N|_{S^2}$, we obtain the stereographic projection

$$\sigma : S^2 \to \widehat{\mathbf{C}} = \mathbf{C} \cup \{\infty\} \tag{2}$$

given by

$$\begin{aligned} \sigma(X_1, X_2, X_3) &= z = \frac{X_1 + iX_2}{1 - X_3}, \\ \sigma(0,0,1) &= \infty. \end{aligned} \tag{3}$$

The extra point ∞ representing $\sigma(N)$ is appended to $\mathbf{C}$ such that stereographic projection is one-to-one and onto and, therefore, identifies the extended image plane with the sphere. $\widehat{\mathbf{C}}$ is known as the Riemann sphere [12].

Two fundamental properties of stereographic projection are particularly useful in imaging with the conformal camera. First, σ maps circles in the sphere that do not pass through N to circles in the image plane, while any circle containing N is mapped to a line. It is customary in the Riemann sphere to regard lines as circles passing through ∞. Second, σ is conformal, that is, it preserves the oriented angles.

Since the pixels are preserved by stereographic projection σ between the sphere (the retina) and the image plane, we conclude that the retinal illuminance is preserved as well.

Image Projective Transformations

The basic image transformations, the h- and k-transformations, are shown in Fig. 2. In the h-transformation, the image of a planar object (e.g., a planar surface of an object) is translated out of the image plane by the vector $\mathbf{b} = \langle b_1, b_2, b_3 \rangle$, as a result of the object undergoing the translation in the scene by the vector $\mathbf{v}$, and projected back into the image plane by j_N. This results in the following mapping:

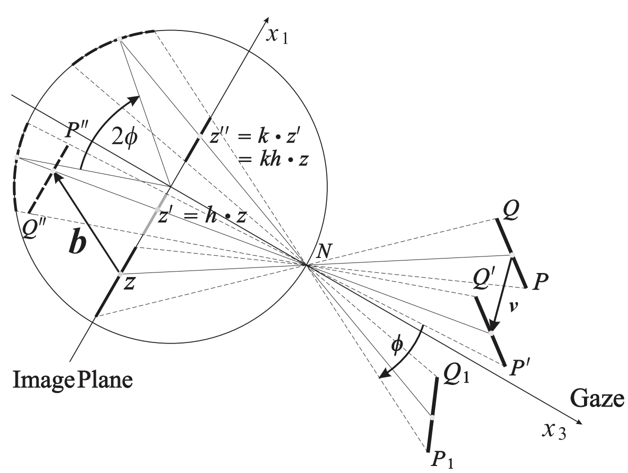

Figure 2. *The image projective transformation hk in the conformal camera resulting from the horozontal motion of the object PQ that undergoes translation by $\mathbf{v}$ and the rotation by ϕ.*

$$z' = h(b_1, b_2, b_3) \cdot z = \frac{z + b_1 + ib_2}{1 - b_3} = \frac{\delta z + \gamma\delta}{1/\delta}, \tag{4}$$

where $\delta = (1 - b_3)^{-1/2}$ and $\gamma = b_1 + ib_2$. From here on, we use the following action

$$z \longmapsto \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot z = \frac{az + b}{cz + d}, \quad ad - bc = 1, \tag{5}$$

and note that

$$h(b_1,b_2,b_3)\cdot z = \begin{bmatrix} \delta & \gamma\delta \\ 0 & 1/\delta \end{bmatrix} \cdot z.$$

The h-transformation is followed by the k-transformation that results from the rotation of the object by (ψ,ϕ,ψ') relative to the center of projection N. That is, the image is first projected into the sphere S^2 by σ^{-1}, then it is rotated with the sphere by the Euler angles $(\psi,2\phi,\psi')$ (here, ψ is the rotation angle about the x_3-axis, followed by the rotation angle 2ϕ about the x_2-axis and the rotation angle ψ' about the x_3-axis) and projected back to the image plane, giving

$$z'' = k(\psi,2\phi,\psi')\cdot z' = \begin{bmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{bmatrix} \cdot z' = \frac{\alpha z' + \beta}{-\overline{\beta} z' + \overline{\alpha}} \qquad (6)$$

where

$$\alpha = e^{-i(\psi+\psi')/2}\cos\phi, \quad \beta = -e^{-i(\psi-\psi')/2}\sin\phi. \qquad (7)$$

The composition of the h- and k-transformations gives the image g-transformation of an object undergoing rigid motions in the scene

$$z'' = g\cdot z = k(\psi,2\phi,\psi')h(b_1,b_2,b_3)\cdot z. \qquad (8)$$

The Group of Image Projective Transformations

In tis section we briefly describe the action of the group of projective transformations of the image intensity function and refere to [22, 23] for detailed discussion.

In the conformal camera, the image projective transformations are defined as the composition of finite iterations of the h- and k-transformations.

Then, by the representation of h- and k-transformations as the matrices acting on the points of the image plane, see (5), we obtain that the image projective transformations are given by the mappings

$$g\cdot z = \frac{az+b}{cz+d}, \quad g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{SL}(2,\mathbf{C}), \qquad (9)$$

which are extended to the Riemann sphere $\widehat{\mathbf{C}} = \mathbf{C}\cup\{\infty\}$ as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\cdot\infty = a/c, \begin{bmatrix} a & b \\ c & d \end{bmatrix}\cdot(-d/c) = \infty, \quad \text{if } c \neq 0,$$
$$\text{and} \quad \begin{bmatrix} a & b \\ 0 & 1/a \end{bmatrix}\cdot\infty = \infty.$$

In (9)

$$\mathbf{SL}(2,\mathbf{C}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : ad - bc = 1 \right\},$$

is the group of 2×2 matrices of complex numbers with determinant one.

For the given intensity function $f(z)$, its image projective transformations are given by

$$f(z) \mapsto f_g(z) = f(g^{-1}\cdot z) = f\left(\frac{dz-b}{-cz+a}\right),$$
$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{SL}(2,\mathbf{C}). \qquad (10)$$

Imaging with the Conformal Camera

The basic image transformations in the conformal camera can only be defined for the planar objects in the scene, cf. [23]. To justify this requirement, we note that only the most basic features are extracted from the impinged visual information on the retina before they are sent to the brain areas for processing. Thus, the initial image of the centrally projected scene is comprised of numerous brightness and color spots from many different locations in space, without explicit information about the perceptual organization of the scene [18]. What is initially perceived is a small number of objects' surfaces that are segmented from the background and from each other [15]. The object's 3D attributes are acquired when 2D projections on the retina are processed downstream the visual pathway by neuronal populations extracting the monocular information (texture gradients, relative size, linear and aerial perspectives, shadows and motion parallax) and, whenever disparity is also available (two eyes seeing the scene), the binocular information.

Active Imaging: the Vector $\mathbf{b}$

We demonstrate in Fig. 3 that when the line of sight of the camera is rotated, the image transformation of a planar stationary object is given by the kh-transformation,

$$\begin{aligned} z' &= k(\psi,-2\phi,\psi')h(b_1,b_2,b_3)\cdot z \\ &= \begin{bmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{bmatrix}\begin{bmatrix} \delta & \gamma\delta \\ 0 & 1/\delta \end{bmatrix}\cdot z \\ &= \frac{\alpha\delta^2 z + \alpha\gamma\delta^2 + \beta}{-\overline{\beta}\delta^2 z - \overline{\beta}\gamma\delta^2 + \overline{\alpha}}, \end{aligned} \qquad (11)$$

where $\delta = (1-b_3)^{-1/2}$ (with $b_3 < 1$), $\gamma = b_1 + ib_2$ and α and β are given in (7).

In fact, Fig. 3 explains that when the camera rotates by ϕ about the x_2-axis, the center of projection undergoes translation by $\mathbf{t} = \overrightarrow{NN_1}$. Then, the object both translates by $-\mathbf{t}$ and rotates by $-\phi$ with respect to the center of projection, the point N representing the eye's pupil.

In the image transformation (11), the eye's rotation angles can be assumed known, as they are used to program the eye movements. However, the vector $\mathbf{b}$ is the model's internal parameter which has to be derived in terms of the eye's intended gaze change and object's geometry and location. In this section, the vector $\mathbf{b}$ is obtained in a full generality.

Referring to Fig. 3, the coordinates of the object's endpoints $P(p_3,p_1)$ and $Q(q_3,q_1)$ are

$$p_1 = s\tan\varphi, \quad p_3 = s+1$$

where s is the distance to the fronto-parallel plane containing the endpoint P and

$$q_1 = s\tan\varphi + w\sin\alpha, \quad q_3 = s+1+w\cos\alpha.$$

Because under the gaze rotation ϕ, the center of projection N is translated by $\mathbf{t}$, PQ is translated by the vector $-\mathbf{t}$, where

$$\mathbf{t} = \langle t_3, t_1\rangle = \langle \cos\phi - 1, \sin\phi\rangle \qquad (12)$$

Coordinates of the endpoints $P'(p_3',p_1')$ and $Q'(q_3',q_1')$ of the translated line segment $P'Q'$, are the following

$$p_1' = s\tan\varphi - \sin\phi, \quad p_3' = s+2-\cos\phi$$

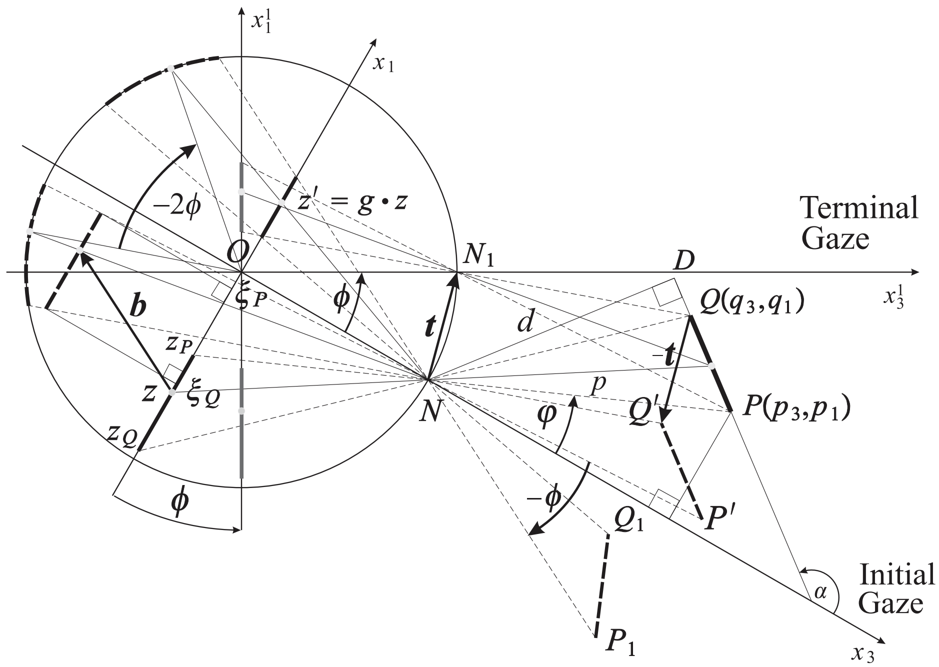

Figure 3. *When the camera gaze is rotated by ϕ, the image projective transformation in the initial coordinate system is given by the g-transformation where $g = hk$ is the result of the composition of the relative object movements (translation by $-\mathbf{t}$ and rotation by $-\phi$ in the scene.*

and

$$\begin{aligned} q_1' &= s\tan\varphi + w\sin\alpha - \sin\phi, \\ q_3' &= s + 2 + w\cos\alpha - \cos\phi. \end{aligned}$$

The components b_1 and b_3 of the vector $\mathbf{b}$ are derived from the relation,

$$b_1 = \xi_P - z_P = \xi_Q - z_Q \tag{13}$$

and the proportions obtained from similar triangles $\triangle Nz_PO \sim \triangle NPp_3$, $\triangle Nz_QO \sim \triangle NQq_3$, and similar triangles that can be easily identified from the last two of the following proportions

$$\begin{aligned} \frac{-z_P}{1} &= \frac{p_1}{p_3 - 1}, \\ \frac{-z_Q}{1} &= \frac{q_1}{q_3 - 1} = \frac{s\tan\varphi + \sin\alpha - \sin\phi}{s + 1 - \cos\phi}, \\ \frac{\xi_P}{1 - b_3} &= -\frac{p_1 - t_1}{p_3 - 1 - t_3} = -\frac{s\tan\varphi - \sin\phi}{s + 1 - \cos\phi}, \end{aligned}$$

and

$$\frac{-\xi_Q}{1 - b_3} = \frac{q_1 - t_1}{q_3 - 1 - t_3} = \frac{s\tan\varphi + w\sin\alpha - \sin\phi}{s + 1 + w\cos\alpha - \cos\phi}.$$

To this end, using the proportions in $\xi_P - z_P = \xi_Q - z_Q$, we first derive

$$\begin{aligned} & - \quad \frac{(1 - b_3)(s\tan\varphi + s\sin\phi)}{s + 1 - \cos\phi} + \tan\varphi \\ &= -\frac{(1 - b_3)(s\tan\varphi + w\sin\alpha - \sin\phi)}{s + 1 + w\cos\alpha - \cos\phi} \\ &+ \quad \frac{s\tan\varphi + w\sin\alpha}{s + w\cos\alpha} \end{aligned}$$

and then solve for b_3, to obtain

$$b_3 = 1 - \frac{\left(1 + \frac{1 - \cos\phi}{w\cos\alpha + s}\right)(s + 1 - \cos\phi)}{s + \frac{\tan\alpha(1 - \cos\phi)}{\tan\alpha - \tan\varphi} + \frac{\sin\phi}{\tan\alpha - \tan\varphi}} \tag{14}$$

Similarly, from $b_1 = \xi_P - z_P$, we have

$$b_1 = -\frac{\left(1 + \frac{1 - \cos\phi}{w\cos\alpha + s}\right)(s\tan\varphi - \sin\phi)}{s + \frac{\tan\alpha(1 - \cos\phi)}{\tan\alpha - \tan\varphi} + \frac{\sin\phi}{\tan\alpha - \tan\varphi}} + \tan\varphi, \tag{15}$$

where we substituted the expression for b_3.

Sequence of Horizontal Gaze Rotations

Whenever eyes are pursuing a moving object, both eye movements and object movements induce an image motion on the retina. Therefore, eye movements must be compensated to allow a clear and stable perception of our surroundings. Given the complexity of the smooth pursuit system, the cortical processing of different coordinate systems (retinal information, information about object movement in space, and information about eye movement relative to the head) and the visual-to-motor transformations necessary to generate these precise eye movements remain largely unclear [20]. Hence, our goal here is to layout the geometric and computational framework for visual information processing during SPEM without assuming the neural mechanisms maintaining visual stability. In particular, we study the background image projective transformations that occur when the conformal camera's gaze is horizontally rotated.

Linearization of the Vector b

Introducing power series $\sin\phi = \phi - \phi^3/6 + ..$ and $\cos\phi = 1 - \phi^2/2 + ..$ into expressions for b_3 and b_1 given in (14) and (15), where '..' indicates the corresponding higher order terms in ϕ, we get for (14),

$$b_3 = 1 - \frac{1 + \left(\frac{1}{s} + \frac{1}{w\cos\alpha + s}\right)\frac{\phi^2}{2} + ..}{1 + \frac{1}{s(\tan\alpha - \tan\varphi)}\phi + ..},$$

which is expanded in powers of ϕ as follows,

$$\begin{aligned} b_3 &= \frac{1}{s(\tan\alpha - \tan\varphi)}\phi - \left[\frac{1}{w\cos\alpha + s}\right. \\ &- \left.\frac{\tan\varphi}{s(\tan\alpha - \tan\varphi)} - \frac{2}{s^2(\tan\alpha - \tan\varphi)^2}\right]\frac{\phi^2}{2} + .. \end{aligned}$$

Similarly, for (15), we have

$$b_1 = -\frac{\tan\varphi - \frac{1}{s}\phi + \frac{\tan\varphi}{w\cos\alpha + s}\frac{\phi^2}{2} + ..}{1 + \frac{1}{s(\tan\alpha - \tan\varphi)}\phi + \frac{\tan\alpha}{s(\tan\alpha - \tan\varphi)}\frac{\phi^2}{2} + ..} + \tan\varphi,$$

which is expanded as follows,

$$\begin{aligned} b_1 &= \frac{\tan\alpha}{s(\tan\alpha - \tan\varphi)}\phi - \left[\frac{\tan\varphi}{w\cos\alpha + s}\right. \\ &- \left.\frac{s\tan\varphi\tan\alpha - 2}{s^2(\tan\alpha - \tan\varphi)} - \frac{2\tan\varphi}{s^2(\tan\alpha - \tan\varphi)^2}\right]\frac{\phi^2}{2} + ... \end{aligned}$$

Then, substituting

$$s(\tan\alpha - \tan\varphi) = \frac{s\sin(\alpha - \varphi)}{\cos\alpha\cos\varphi} = \pm\frac{d}{\cos\alpha}$$

where d is the distance to the plane containing the planar object,

$$d = \frac{s|\sin(\alpha - \varphi)|}{\cos\varphi}, \tag{16}$$

which can be easily obtained from the right triangle $\triangle NDP$ in Fig. 3, we arrive at the second order approximations

$$\begin{aligned} b_3 &\approx \pm\frac{\cos\alpha}{d}\phi \\ &- \left[\frac{1}{w\cos\alpha + s} \pm \frac{\sin\alpha}{d} + \frac{2\cos^2\alpha}{d^2} + \frac{1}{s}\right]\frac{\phi^2}{2} \end{aligned} \tag{17}$$

$$\begin{aligned} b_1 &\approx \pm\frac{\sin\alpha}{d}\phi + \left[-\frac{\tan\varphi}{w\cos\alpha + s}\right. \\ &\pm \left.\frac{s\tan\varphi\sin\alpha - 2\cos\alpha}{sd} - \frac{2\tan\varphi\cos\alpha^2}{d^2}\right]\frac{\phi^2}{2} \end{aligned} \tag{18}$$

that make the set of parameters especially convenient for error analysis. In the above expressions, the upper sign is for $\alpha > \varphi$ while the lower sign is for $\alpha < \varphi$.

Thus, the approximation of vector **b** to the first order in ϕ is the following

$$\mathbf{b} = \pm\frac{\phi}{d}\langle\cos\alpha, \sin\alpha\rangle. \tag{19}$$

where α gives the orientation of the object PQ relative to the initial coordinate system and d is the distance (16) to the plane containing the (planar) object.

From the second order terms in (17) and (18), we see that the approximation breaks down when $\alpha = \varphi$ or $\alpha = \varphi \pm \pi$, and when $\varphi = \pm\pi/2$. From (16) and $s = p\cos\varphi$, where p is the distance to P satisfying $p \neq 0$, we see that $s = 0$ if and only if $\varphi = \pm\pi/2$, and $d = 0$ if and only if $\alpha - \varphi = 0, \pm\pi$. Under the first condition ($s = 0$), some points do not project to the image plane (that is, they project to ∞), and under the second condition ($d = 0$), a planar two-dimensional object has a one-dimensional projection on the image plane. These conditions are the same as the ones that prevent the proper functioning of the primate vision system.

Maybe the most significant result is the fact that in the first order approximation, the vector **b** direction (relative to the initial coordinate system) is given by the angle α (**b** is parallel to $\overrightarrow{PQ}$), while its length depends linearly on ϕ (or $-\phi$) and inversely on d, the distance to the plane containing the planar object.

Thus, since the values of α and d fix the plane containing the planar object, the leading term approximation of **b** does not depend on the object size w or where on the plane it is located. We conclude that the approximation (19) applies directly to 2D planar objects (or planar surfaces of 3D objects) as it does not depend on the shape of these objects.

Sequence of Small Gaze Rotations

We consider a sequence of horizontal rotations ($\phi_m, m = 1, 2, 3, ...$), where ϕ_m is the angle rotating gaze $m-1$ to gaze m with the 0th gaze being the initial gaze. The image transformations are given in the initial coordinate system.

From now on, we change to the following notation: the coordinate system rotated with the camera by the angle ϕ_m is denoted by (x_3^m, x_1^m). We choose both ϕ_m and the scene parameters, such that the corresponding image h-transformation is well approximated by using the linear term of the vector $\mathbf{b}_m$,

$$\mathbf{b}_m = \frac{\phi_m}{d_{m-1}}\langle\cos\beta_{m-1}, \sin\beta_{m-1}\rangle. \tag{20}$$

Here, for $m = 1$, $d_0 = d$ and $\beta_0 = \alpha$, so that

$$\mathbf{b}_1 = \frac{\phi_1}{d}\langle\cos\alpha, \sin\alpha\rangle. \tag{21}$$

We need to find β_m and d_m for $m \geq 2$. To this end, when the conformal camera undergoes the second gaze rotation ϕ_2, the angle β_2 of $\mathbf{b}_2$ is $\alpha - \phi_1$.

Thus, for the mth gaze rotation, the angle β_{m-1} of $\mathbf{b}_m$ is the following

$$\beta_{m-1} = \alpha - \sum_{k=1}^{m-1}\phi_k, \quad m \geq 2. \tag{22}$$

Finally, using the results derived in Appendix, we have the following formulas for d_m,

$$\begin{aligned} d_{m-1} &= \pm d + 2\sin\left[\frac{1}{2}\sum_{k=1}^{m-1}\phi_k\right]\cos\left[\alpha - \frac{1}{2}\sum_{k=1}^{m-1}\phi_k\right] \\ &= \pm d + \sin\alpha - \sin\beta_{m-1}, \quad m \geq 2 \end{aligned} \tag{23}$$

with the upper sign holding for $\alpha > \varphi$ and the lower sign holding for $\alpha < \varphi$. The distance d is given in (16).

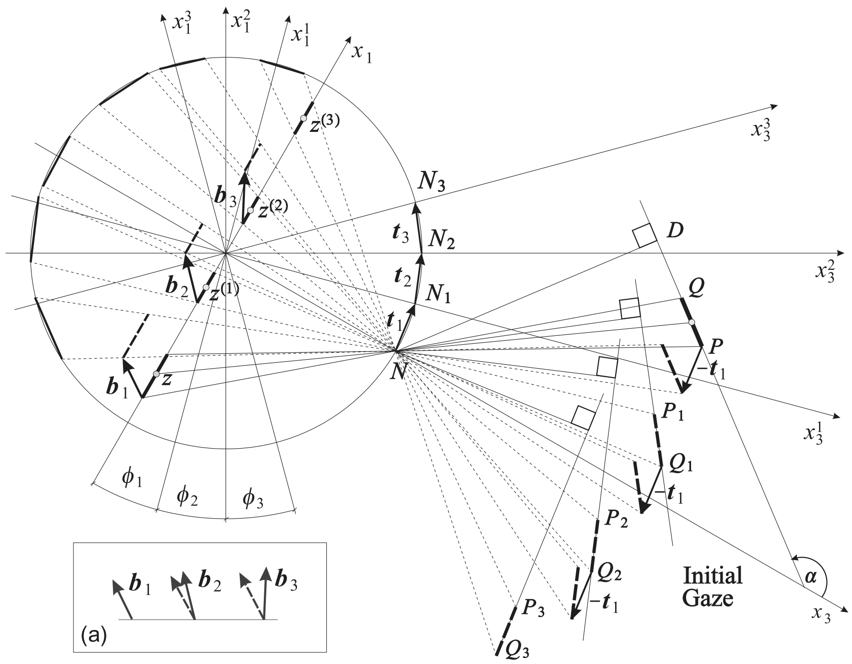

Figure 4. *The image projective transformations during the sequence of the gaze rotations ϕ_1, ϕ_2 and ϕ_3. The relative motions of the object PQ used in the construction of the image transformations, as explained in Fig. 3, are also shown. The panel (a) confirms the results from the previous section on the approximation of the vector $\mathbf{b}$ in (20) and (22): the direction of $\mathbf{b}_2$ is the rotated direction of $\mathbf{b}_1$ by ϕ_1 and the direction of $\mathbf{b}_3$ is the rotated direction of $\mathbf{b}_1$ by $\phi_1+\phi_2$*

Conclusions: Modeling Horizontal SPEM

The horizontal SPEM with angular speed $\omega(t)$, $t \geq a$ is approximated for a time step Δt by a sequence of discrete rotations $(\phi_m, m = 1,2,3,...)$ where $\phi_m = \omega_m \Delta t$ is the rotation from gaze $m-1$ to gaze m with the average angular speed ω_m in time interval $[t_{m-1}, t_m]$, $t_m = a + m\Delta t$. For $\omega(t)$ we choose Δt such that for each ϕ_m, the approximation (20) holds.

Let the intensity function $f(z)$ denote the image of the object. Using the results from the previous section and Fig. 4, the image projective transformation of the rotation from gaze $m-1$ to gaze m, in the initial coordinate system, is the following

$$f\left(z^{(m-1)}\right) \mapsto f\left(z^{(m)}\right) = f\left(g_m^{-1} \cdot z^{(m-1)}\right), \tag{24}$$

where

$$\begin{aligned} g_m &= k(0,-2\phi_m,0)\, h(b_m \sin\beta_{m-1}, 0, b_m \cos\beta_{m-1}) \\ &= \begin{bmatrix} \cos\phi_m & \sin\phi_m \\ -\sin\phi_m & \cos\phi_m \end{bmatrix} \begin{bmatrix} \delta_m & \gamma_m \delta_m \\ 0 & 1/\delta_m \end{bmatrix} \end{aligned} \tag{25}$$

and hence,

$$\begin{aligned} g_m^{-1} &= h^{-1}(b_m \sin\beta_{m-1}, 0, b_m \cos\beta_{m-1}) k^{-1}(0,-2\phi_m,0) \\ &= \begin{bmatrix} 1/\delta_m & -\gamma_m \delta_m \\ 0 & \delta_m \end{bmatrix} \begin{bmatrix} \cos\phi_m & -\sin\phi_m \\ \sin\phi_m & \cos\phi_m \end{bmatrix} \end{aligned} \tag{26}$$

Here $b_m = \phi_m / d_{m-1}$ and β_{m-1} and d_{m-1} are given in (22) and (23), respectively. Further, $\delta_m = (1 - b_m \cos\beta_{m-1})^{-1/2}$ and $\gamma_m = b_m \sin\beta_{m-1}$. In implementation, the product of matrices in (26) has to be linearized in ϕ_m.

By the iteration of (24) we obtain the SPEM's sequence of image projective transformations of the object given only in terms of the initial orientation angle α and the distance d of the plane containing the object planar surface.

Explicitly, for the given initial image f of a stationary object and the given SPEM's sequence $(\phi_1, \phi_2, ..., \phi_n)$, the image undergoes the following transformation under the pursuit

$$f\left(z^{(n)}\right) = f\left(g^{-1}_{(n,n-1,...,1)} \cdot z\right)$$

where $g_{(n,n-1,...,1)} = g_n g_{n-1} ... g_2 g_1$ and

$$g^{-1}_{(n,n-1,...,1)} = g_1^{-1} g_2^{-1} ... g_{n-1}^{-1} g_n^{-1}$$

with g_m^{-1} given in (26).

We recall that each d_{m-1} in (23), which describes the vector $\mathbf{b}_m$ in (20), can be determined only by the planar object (or planar surface of the object) orientation α and the distance d to the plane containing this object. These values can be typically estimated by the primate's vision system. Therefore, our modeling of the SPEM with the conformal camera should be able to support the stability of visual informations during smooth pursuit of an anthropomorphic robotic camera that is needed for an autonomous robot efficient interaction with the real world, in real time.

Implementation in a pursuit movement by the conformal camera will be done in the future.

Appendix: The Formula For d_m

In this appendix we derive the formula for d_m that appeared in (20). To get d_m, we use Fig. 5 which shows details near projection centers N_1, N_2 and N_3 of the three first gaze rotations gaze rotations ϕ_1, ϕ_2 and ϕ_3.

To this end, using the isosceles triangle $\triangle CNN_k$ and the fact that $|CN| = 1$, we first note that

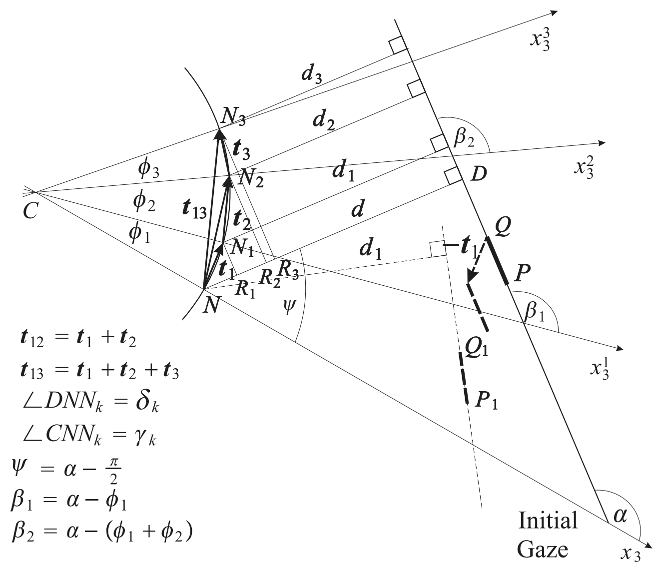

Figure 5. *Geometric details near the projection centers N and N_1 shown in Fig. 3 and the next two projection centers N_2 and N_3.*

$$
\begin{aligned}
|\mathbf{t}_{1m}| &= |\mathbf{t}_1 + \mathbf{t}_2 + \ldots + \mathbf{t}_m| \\
&= 2\sin\left[\frac{1}{2}(\phi_1 + \phi_2 + \ldots\phi_m)\right] \\
&= 2\sin\left[\frac{1}{2}\left(\sum_{k=1}^{m}\phi_k\right)\right]
\end{aligned}
$$

and

$$\gamma_m = \frac{\pi}{2} - \frac{1}{2}\sum_{k=1}^{m}\phi_k.$$

Next, at the vertex N, we can write

$$\psi + \delta_m + \gamma_m = \pi$$

which, using $\psi = \alpha - \pi/2$, can be solved for δ_m as follows

$$\delta_m = \pi - \alpha + \frac{1}{2}\sum_{k=1}^{m}\phi_k.$$

Finally, from

$$
\begin{aligned}
d - d_m &= |NR_m| = |\mathbf{t}_{1m}|\cos\delta_m \\
&= 2\sin\left[\frac{1}{2}\sum_{k=1}^{m}\phi_k\right]\cos\left[\pi - \alpha + \frac{1}{2}\sum_{k=1}^{m}\phi_k\right] \\
&= -2\sin\left[\frac{1}{2}\sum_{k=1}^{m}\phi_k\right]\cos\left[\alpha - \frac{1}{2}\sum_{k=1}^{m}\phi_k\right],
\end{aligned}
$$

which can be easily verified in Fig. 5, we obtain

$$
\begin{aligned}
d_m &= d + 2\sin\left[\frac{1}{2}\sum_{k=1}^{m}\phi_k\right]\cos\left[\alpha - \frac{1}{2}\sum_{k=1}^{m}\phi_k\right] \\
&= d + \sin\alpha - \sin\left[\alpha - \sum_{k=1}^{m}\phi_k\right].
\end{aligned}
$$

References

[1] J. A. Boluda, F. Pardo, A reconfigurable architecture for autonomous visual-navigation, Mach. Vision Appl., 13, 322 (2003).

[2] C. J. Erkelens, Coordination of smooth pursuit and saccades, Vision Res., 46, 163 (2006).

[3] O. J. Grüsser, On the history of the ideas of efference copy and reafference, Clio Medica 33, 35 (1995).

[4] N. J. Hall, C. L. Colby, Remapping for visual stability, Phil. Trans. R. Soc., 366, 528 (2011).

[5] H. Kim, A. Honda, R. Benosman, S-H. Ieng, A. Davison, Simultaneous Mosaicing and Tracking with an Event Camera, British Machine Vision Conf., https://wp.doc.ic.ac.uk/roboticvision/publication/simultaneous-mosaicing-and-tracking-with-an-event-camera/(2014).

[6] S. Kohlbecher, K. Bartl, S. Bardins, E. Schneider, Low-letency combined eye and head tracking system for teleoperating a robotic head in real-time, Proc. of the 2010 Symposium on Eye-Tracking Research and Applications, pg. 117. New York, NY:ACM (2010).

[7] R. J. Krauzlis, Recasting the smooth pursuit eye movement system, J Neurophysiol., 91, 591 (2004).

[8] C. Laschi, F. Patane, E Maini, L. Manfredi, G. Teti, L. Zollo, E. Guglieimelli, P. Dorio, An Anthropomorphic Robotic Head for Investigating Gaze Control, Adv. Robotics, 22, 57 (2008).

[9] S. G. Lisberger, Visual Guidance of Smooth-Pursuit Eye Movements: Sensation, Action, and What Happens in Between, Neuron, 66, 477 (2010).

[10] M. Mahowald, C. Mead, A silicon model of early visual processing, Neural Networks, 1, 91 (1988).

[11] D. Melcher, C. L. Colby, Trans-saccadic perception, Trends in Cogn. Sci., 12, 460 (2008).

[12] T. Needham, Visual Complex Analysis, Oxford Univ Press, NY, 2002.

[13] W. T. Newsome, R. H. Wurtz, H. Komatsu, Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal imputes, J Neurophysiol., 60, 604 (1988).

[14] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, T. Delbruck, Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output, Proc. IEEE, Vol. 102 pg. 1470. (2014).

[15] P. R. Roelfsema, M. Tolboom, P. S. Khayat, Different Processing Phases for Features, Figures, and Selective Attention in the Primary Visual Cortex, Neuron, 56, 785 (2007).

[16] P-F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P-Y. Burgi, S. Gyger, P. Nussbaum, A 128 $\times$ 128 pixel 120 dB dynamic range vision sensor chip for image contrast and orientation extraction, Proc. Int. Solid-St. Circ. Conf. pg. 226. (2003).

[17] D. Schoppik, S. G. Lisberger, Saccades exert spatial control of motion processing for smooth pursuit eye movements, J Neurosci., 26, 7607 (2006).

[18] R. Shapley, Early vision is early in time, Neuron, 56, 755 (2007).

[19] M. A. Sommer, R. H. Wurtz, Brain circuits for the internal monitoring of movements, Annu. Rev. Neurosci., 31, 317 (2008).

[20] J. L. Souman, I. T. Hooge, A. H. Wertheim, Frame of reference transformations in motion perception during smooth pursuit eye movements, J Comput Neurosci., 20, 61 (2006).

[21] J. Turski, Harmonic analysis on SL(2,C) and projectively adapted pattern representation, J Fourier Anal. Appl., 4, 67 (1998).

[22] J. Turski, Projective Fourier analysis for patterns, Pattern Recogn., 33, 2033 (2000).

[23] J. Turski, Geometric Fourier Analysis of the Conformal Camera for Active Vision, SIAM Rev., 46, 230 (2004).

[24] J. Turski, Geometric Fourier Analysis for Computational Vision, J Fourier Anal. Appl., 11, 1 (2005).

[25] J. Turski, Robotic Vision with the Conformal Camera: Modeling Perisaccadic Perception, J Robotics, (2010). doi:10.1155/2010/130285

[26] J. Turski, Imaging with the Conformal Camera, Proc. IPCVIPR., Vol. II, CSREA Press (2012).

[27] R. Wodnicki, G. W. Roberts, M. D. Levine, A Log-Polar Image Sensor Fabricated in a Standard 1.2- m ASIC CMOS, Proc. IEEE J Solid-St Circ., 32, pg. 1274. (2002).

[28] M. Yeasin, Optical flow in log-mapped image plane–a new approach, IEEE Trans. Pattern Anal. Mach. Intell., 23, 1 (2001).

Author Biography

Jacek Turski was awarded his Ph.D. from McGill University. After holding postdoctoral positions at the University of Manitoba and the University of Houston, he joined the University of Houston-Downtown where he is now a Full Professor in the Department of Mathematics and Statistics. Within the geometric analysis framework of the conformal camera (the Riemann sphere), he has been modeling visual information during the foveated-sensor camera rotations resembling saccadic and smooth pursuit eye movements. His latest study, which will appear in Vision Research, corrects the two century-old Vieth-Müller model of the geometric horopter. That model, which is still influencing theoretical research in binocular vision, assumes that the optical node coincides with the eye rotation center, an anatomically incorrect location. Dr. Turski's research has been supported by NSF grants and has been published in mathematics, computer science and robotics journals.