

Research Article

Robotic Vision with the Conformal Camera: Modeling Perisaccadic Perception

Jacek Turski

Department of Computer and Mathematical Sciences, University of Houston-Downtown, One Main Street, Houston, TX 77002, USA

Correspondence should be addressed to Jacek Turski, turskij@uhd.edu

Received 21 September 2009; Revised 17 January 2010; Accepted 8 February 2010

Academic Editor: Noriyasu Homma

Copyright © 2010 Jacek Turski. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Humans make about 3 saccades per second at the eyeball's speed of 700 deg/sec to reposition the high-acuity fovea on the targets of interest to build up understanding of a scene. The brain's visuosaccadic circuitry uses the oculomotor command of each impending saccade to shift receptive fields (RFs) to cortical locations before the eyes take them there, giving a continuous and stable view of the world. We have developed a model for image representation based on projective Fourier transform (PFT) intended for robotic vision, which may efficiently process visual information during the motion of a camera with silicon retina that resembles saccadic eye movements. Here, the related neuroscience background is presented, effectiveness of the conformal camera's non-Euclidean geometry in intermediate-level vision is discussed, and the algorithmic steps in modeling perisaccadic perception with PFT are proposed. Our modeling utilizes basic properties of PFT. First, PFT is computable by FFT in complex logarithmic coordinates that also approximate the retinotopy. Second, the shift of RFs in retinotopic (logarithmic) coordinates is modeled by the shift property of discrete Fourier transform. The perisaccadic mislocalization observed by human subjects in laboratory experiments is the consequence of the fact that RFs' shifts are in logarithmic coordinates.

1. Introduction

In this article, we demonstrate that a mathematical data model we have developed for image representation intended for biologically-mediated machine vision systems [1–6] may be useful to process visual information during the motion of a camera with silicon retina that resembles saccadic eye movements. Our data model is based on the projective Fourier analysis that we have constructed in the framework of representation theory of the Lie group $SL(2, \mathbb{C})$ by restricting the group representations to the image plane of the conformal camera—the camera with image projective transformations given by the action of $PSL(2, \mathbb{C})$ [4, 5]. The analysis provides an efficient image representation that is well adapted to (a) the projective transformations of retinal images and (b) the retinotopic mappings of the brain's oculomotor and visual pathways. This latter assertion stems from the fact that the projective Fourier transform (PFT) is computable by a fast Fourier transform (FFT) algorithm in coordinates given by a complex logarithm that transforms PFT into the standard Fourier integral, while simultaneously

approximating the local retinotopy [7]. Consequently, PFT of the conformal camera integrates the head, eyes, and retinotopic mapping of the visual pathways into a single computational binocular system [6]. As already suggested in [8], this integrated system may efficiently model visuomotor processes during saccadic eye movements that reposition the high-acuity fovea—the retinal region of a central angle subtending a US quarter an arm length from the eyes—on the targets of interest to build up understanding of a scene.

Humans make about three saccades per second at the eyeball's maximum speed of 700 deg/sec, producing about 200,000 saccades per day [9]. As we are not aware of these fast moving retinal images, the brain, under normal circumstances, suppresses visual sensitivity during saccadic eye movements and compensates for these incessant interruptions. This visual stability is maintained by the brain's widespread neural network [10]. Converging evidence from psychophysics, functional neuroimaging, and primate neurophysiology indicates that the most attractive neural basis that underlies visual stability is the mechanism causing visual cells in various visual and visuomotor cortical areas to

respond to stimuli that will fall in their receptive fields (RFs) before the eyes move them there, commonly referred to as the shifting RFs mechanism [11–15]. The identification of the visuosaccadic pathways (see references in [10]) 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 [16] for a review), to shift transiently the RFs of stimuli. This shift of RFs, starting 50 ms before a saccade onset and ending 50 ms after the saccade landing, is hypothesized to update (or remap) the retinotopic maps in the anticipation of each upcoming saccade. In fact, in a recent experiment [17], when human subjects shifted fixation to the clock, their reported time was earlier than the actual time on the clock by about 40 ms.

Although interruptions caused by saccades remain unnoticed in our daily life, in laboratory experiments, it becomes possible to probe the unexpected consequences of the saccadic eye movement. Specifically, laboratory experiments in lit environments have shown that briefly flashed probes around the saccade’s onset are perceived as compressed toward the saccadic target [18–21], while, in total darkness, the probes’ localizations are characterized by a uniform shift in the direction of the saccade [22–24]. The experimental studies [25–27] investigating the influence of the saccade’s parameters on perisaccadic mislocalization showed that perisaccadic visual compression and the unidirectional shift are probably governed by different neural processes. Although the perisaccadic shift can be mainly explained by delays in the processing of visual information [24], the mechanism of perisaccadic compression, commonly related to the neural processes of the RF shift [28], remains relatively elusive.

In this article, we argue that the conformal camera’s complex projective geometry and the related harmonic analysis (projective Fourier analysis) may be useful in perisaccadic perception. In particular, the image representation in terms of PFT may efficiently model the RFs shift that remaps cortical retinotopy in the anticipation of each saccade and the related phenomenon of perisaccadic perceptual space compression. During fixations the brain acquires visual information, resolving inconsistencies of the brief compression resulting from remapping. The computational significance of this remapping, when incorporated into neural engineering design of a foveate visual system, stems from the fact that it may integrate visual information from an object across saccades, eliminating the need for starting visual information processing anew three times per second at each fixation and speeding up a costly process of visual information acquisition [29]. This transfer of object features across saccadic eye movements [30–33] that is believed to maintain visual stability of transsaccadic perception is not considered here because, at present not much is known about the whole process [13].

This paper is organized as follows. We outline the neural processes of the visuosaccadic system involved in the preparation, execution, and control of the saccadic eye movement in Section 2 and later continue in Section 6. In Sections 3 and 4, we lay out the background that explains the mathematical tools we use in modeling human visual system.

To this end, in Section 3, we introduce the conformal camera, discuss its conformal geometry, and evaluate the effectiveness of this geometry in the early- and intermediate-level vision computational aspects of natural scene understanding. Then, in Section 4, we show that the conformal camera possesses its own harmonic analysis—projective Fourier analysis—which provides image representation given in terms of the discrete PFT (DPFT) fast computable by FFT in coordinates given by a complex logarithm. Section 5 of this article deals with the implementation of the DPFT in retinocortical image representation that efficiently integrates the head, eyes, and retinotopic maps into one computational system. We also mention hardware setups that could be supported by DPFT-based software and compare conformal camera-based modeling to other approaches in foveate vision. Using this integrated visual system, in Section 6, we model perisaccadic perception, including the perisaccadic compression observed in psychophysical laboratory experiments. Finally, in Section 7, we discuss and compare our model with other numerical approaches to perisaccadic perception and discuss the directions in advancing our modeling. The paper is summarized in the last section.

2. The Visuosaccadic System

One of the most important functions of any nervous system is sensing the external environment and responding in a way that maximizes immediate survival chances. For this reason, the perception and action have evolved in mammals by supporting each other’s functions. This functional link between visual perception and oculomotor action is well demonstrated in primates when they execute the eye-scanning movements (saccades) to overcome the eye’s acuity limitation in building up the scene understanding.

In fact, humans can only see clearly the central part of the visual field of a 2 deg central angle. This region is projected onto the central fovea, where its image is sampled by the hexagonal mosaic of photoreceptors consisting mainly of cone cells, the color-selective type of photoreceptors for a sharp daylight vision. The visual acuity decreases rapidly away from the fovea because the distance between cones increases with eccentricity as they are outnumbered by rod cells, photoreceptors for a low-acuity black-and-white night vision. Moreover, there are a gradual loss of hexagonal regularity of the photoreceptor mosaic and an increased convergence of the photoreceptors on the ganglion cells whose axons carry visual information from the eye to the brain. For example, at 2.5 deg radius, which corresponds to the most visually useful region of the retina, acuity drops 50%. In Figure 1, (b) shows a progressively blurred image from (a), simulating the progressive loss of acuity with eccentricity.

With three saccades per second, the saccadic eye movement is the most common bodily movement. The eyes remain relatively still between consecutive saccades for about 180–320 ms, depending on the task performed. During this time period, the image is processed by the retinal circuitry and sent, mainly, to the visual cortex (starting with the

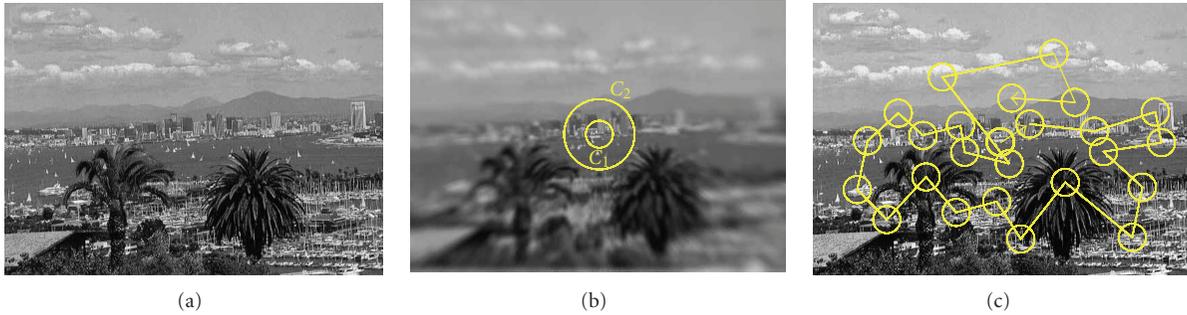


FIGURE 1: (a) San Diego skyline and harbor. (b) Progressively blurred image of (a) simulating the progressive loss of retinal acuity with eccentricity. The circle C_1 encloses the part of the scene projected onto the high-acuity fovea of a 2 deg diameter. The circle C_2 encloses the part projected onto the visually useful foveal region of a 5 deg diameter. (c) A scanning path the eyes may take to build up the scene understanding. Adapted from [34].

primary visual cortex, or V1, and reaching higher cortical areas, including the cognitive areas), with a minor part sent to the oculomotor midbrain areas. During the saccadic eye movement, the visual sensitivity is markedly reduced, although some modulations of low spatial frequencies (contrast and brightness) are wellpreserved or even enhanced [35]. This phenomenon is known as saccadic suppression. The sequence of saccades, fixations, and, often, smooth-pursuit eye movements for tracking a slowly moving small object in the scene, is called the scanpath and was first studied in [36]. In Figure 1(c) we depict the scanpath that eyes might actually take to build up understanding of the scene.

Although they are the simplest of bodily movements, the eyes' saccades are controlled by a widespread neural network that involves nearly every level of the brain. In Figure 2 we show the diagram of well-established pathways of the primate visuosaccadic system. They includes most prominently, the superior colliculus (SC) of the midbrain for representing possible saccade targets, and the parietal eye field (PEF) and the frontal eye field (FEF) in the parietal and frontal lobes of the neocortex (which obtains inputs from many visual cortical areas) for assisting the SC in the control of the involuntary (PEF) and voluntary (FEF) saccades. They also project to the simple neural circuits in the brainstem reticular formation (pons) in the midbrain that ensure the saccade's outstanding speed and precision. The course of events in the visuosaccadic system, which is based on [10], is outlined in the caption of Figure 2.

Although many of the neural processes involved in saccade generation and control are amenable to precise quantitative studies [37], some neural processes of the visuosaccadic system remain virtually unknown. The saccadic suppression, the fact that we do not see moving retinal images, is barely understood. There is accumulating evidence that viewers integrate information poorly across saccades during tasks such as reading, visual search, and scene perception [38]. This means that, three times per second at each fixation, there are instant large changes in the retinal images without almost any information consciously carried between fixations. Furthermore, because the next saccade target selection for the voluntary saccades takes place in

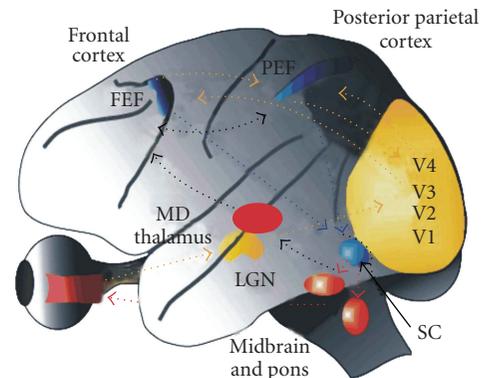


FIGURE 2: The visuosaccadic system. The course of events is the following. (1) *Fixations* (~ 300 ms). When eyes complete a saccade, visual information is acquired and sent via LGN to V1 and higher cortical areas including FEF and PEF (yellow paths). The next saccade's target is selected in FEF/PEF and its position is calculated in SC (blue paths) and the oculomotor command (OC) is prepared (Pons). However, before the saccade is executed, a copy of OC is sent via MD of Thalamus to FEF/PEF (black paths) to shift the receptive fields of the current fixation landmarks to the receptive fields of their future positions centered at the impending saccade's target. (2) *Saccades* (~ 30 ms). The OC executes the eyes' saccade with the motor neurons (red paths). The shift of receptive fields (starting ~ 50 ms before the saccade onset and ending ~ 50 ms after the saccade landing) gives access to the upcoming saccade's target before eyes arrival, integrating objects' features between fixations and maintaining stability of perception.

the higher cortical areas involving cognitive processes [39], the time needed for the oculomotor system to plan and execute the saccadic eye movement could take as long as 150 ms. Therefore, it is critical that visual information is efficiently acquired during each fixation without repeating much of the whole process since it would require too much computational resources. However, visual constancy, the fact that we are not aware of any discontinuity in the scene perception when executing the scanpath, is not perfect. About 50 ms before the onset of the saccade, during saccadic movement (~ 30 ms), and about 50 ms after the saccade,

the salient stimuli are not perceived in veridical locations. In particular, a transient compression around the saccade target, called perisaccadic mislocalization, is observed in lit laboratory experiments; see Section 1. We continue this discussion in Section 6 where we present the algorithm for modeling some of the neural processes underlying perisaccadic perception.

3. The Conformal Camera, Geometry, and Perception

We model the human eyes' imaging functions with the conformal camera; the name of which will be explained later. In the conformal camera, shown in Figure 3, the retina is represented by the image plane $x_2 = 1$ with complex coordinates $x_3 + ix_1$, on which a 3D scene is projected under the mapping

$$j(x_1, x_2, x_3) = \frac{(x_3 + ix_1)}{x_2}. \quad (1)$$

The implicit assumption $x_2 \neq 0$ will be removed later.

The image projective transformations are generated by the two basic transformations h and k shown in Figure 3. Both transformations have the form of linear-fractional transformations

$$z \mapsto g \cdot z = \frac{\alpha z + \beta}{\gamma z + \delta} \quad (2)$$

with $\alpha\delta - \gamma\beta = 1$. Therefore, all finite iterations of the mappings h and k give the group $\mathbf{SL}(2, \mathbb{C})$ acting on the image plane of the conformal camera by linear-fractional, or Möbius, transformations

$$\mathbf{SL}(2, \mathbb{C}) \ni \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{dz + c}{bz + a}, \quad z = x_3 + ix_1 \equiv (x_1, 1, x_3). \quad (3)$$

Because $\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ have the same action, we need to identify matrices in $\mathbf{SL}(2, \mathbb{C})$ that differ in sign. The result is the quotient group $\mathbf{PSL}(2, \mathbb{C}) = \mathbf{SL}(2, \mathbb{C}) / \{\pm Id\}$, where Id is the identity matrix, and the action (3) establishes a group isomorphism between linear-fractional, or Möbius, mappings and $\mathbf{PSL}(2, \mathbb{C})$. Thus,

$$\mathbf{PSL}(2, \mathbb{C}) \ni g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto f(g^{-1} \cdot z) = f\left(\frac{az - c}{-bz + d}\right) \quad (4)$$

gives the image projective transformations of the intensity function $f(z)$.

3.1. Geometry of the Conformal Camera. In the homogeneous coordinate framework of projective geometry [40], the conformal camera is embedded into the complex plane

$$\mathbb{C}^2 = \left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \mid z_1 = x_2 + iy, z_2 = x_3 + ix_1 \right\}. \quad (5)$$

In this embedding, the ‘‘slopes’’ ξ of the complex lines $z_2 = \xi z_1$ can be numerically identified with the points on the extended image plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ where ∞ corresponds to the line $z_1 = 0$. In fact, if $x_2 \neq 0$ and $y = 0$, the slope ξ corresponds to the point $x_3 + ix_1$ at which the ray (line) in \mathbb{R}^3 that passes through the origin is intersecting the image plane of the conformal camera. Then, the standard action of the group $\mathbf{SL}(2, \mathbb{C})$, on nonzero column vectors \mathbb{C}^2 ,

$$\begin{pmatrix} z'_1 \\ z'_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{pmatrix} \quad (6)$$

implies that the slope $\xi = z_2/z_1$ is mapped to the slope

$$\xi' = \frac{z'_2}{z'_1} = \frac{c + d\xi}{a + b\xi} \quad (7)$$

agreeing with the mappings in (3). However, the action (3) must be extended to include the line $z_1 = 0$ of ‘‘slope’’ ∞ as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \infty = \frac{d}{b}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \left(-\frac{a}{b}\right) = \infty. \quad (8)$$

The stereographic projection $\sigma = j|_{S^2_{(0,1,0)}}$ (with j in (1)) establishes isomorphism $\widehat{\mathbb{C}} \cong S^2_{(0,1,0)}$ and $\sigma(0,0,0) = \infty$ gives a concrete meaning to the point ∞ such that it can be treated as any other point. The set $\widehat{\mathbb{C}}$ is referred to as the Riemann sphere and the group $\mathbf{PSL}(2, \mathbb{C})$ acting on $\widehat{\mathbb{C}}$ consists of the bijective meromorphic mappings of $\widehat{\mathbb{C}}$ [41]. Thus, it is the group of holomorphic automorphisms of the Riemann sphere that preserve the intrinsic geometry imposed by complex structure, known as Möbius geometry [42] or inversive geometry [43].

The mappings in (4) are conformal, that is, they preserve the oriented angles of two tangent vectors intersecting at a given point [41]. Because of this property, the camera is called ‘‘conformal’’. Although the conformal part of an image projective transformation can be removed with almost no computational cost, leaving only a perspective transformation of the image (see [4, 5]), the conformality provides an advantage in imaging because the conformal mappings rotate and dilate the image's infinitesimal neighborhoods, and, therefore, locally preserve the image pixels.

The image plane of the conformal camera does not admit a distance that is invariant under image projective (linear-fractional or Möbius) transformations. Therefore, geometry of the conformal camera does not possess a Riemannian metric; for instance, there is no curvature measure. It is customary in complex projective (Möbius or inversive) geometry to consider a line as a circle passing through the point ∞ . Then, the fundamental property of this geometry can be expressed as follows: linear-fractional mappings map circles to circles [41]. Thus, circles can play the role of geodesics.

3.2. The Conformal Camera and the Intermediate-Level Vision. As discussed before, circles play a crucial role in the

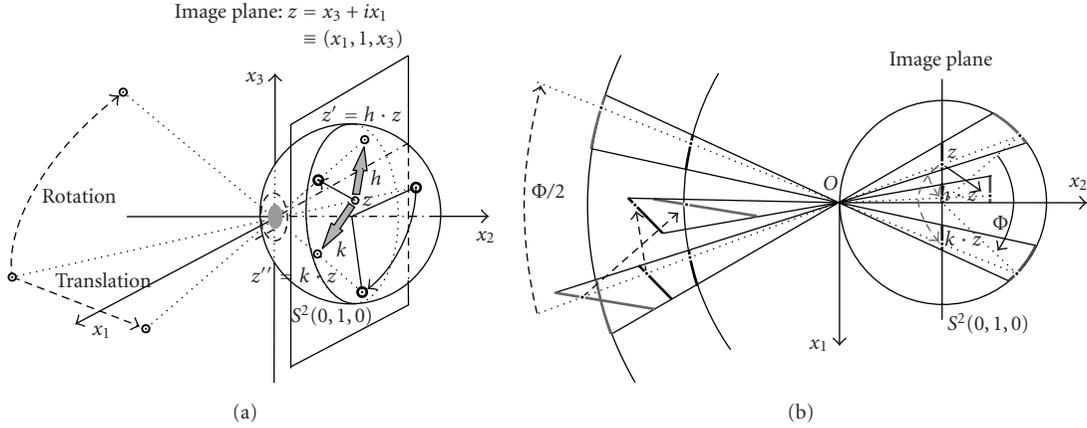


FIGURE 3: The conformal camera. (a) Image projective transformations are generated by iterations of transformations covering translations “ h ” and rotations “ k ” of planar objects in the scene. (b) The 2D section of the conformal camera further explains how image projective transformations are generated and how the projective degrees of freedom are reduced in the camera; one image projective transformation in the conformal camera corresponds to different planar objects translations and rotations in the 3D world.

conformal camera geometry and it should be reflected in psychological and computational aspects of natural scene understanding whether this camera is relevant to modeling primate visual perception.

Neurophysiological experiments demonstrate that the retina filters impinged images extracting local contrast spatially and temporally. For instance, center surround cells at the retinal processing stage are triggered by local spatial changes in intensity referred to as edges or contours. This filtering is enhanced in the primary visual cortex, the first cortical area receiving the retinal output. This area itself is a case study in dense packing of overlapping visual submodalities: motion, orientation, frequency (color), and oculomotor dominance (depth). In psychological tests, humans easily detect a significant change in spatial intensity (low-level vision), and effortlessly and unambiguously group this usually fragmented visual information (contours of occluded objects, for example), into coherent, global shapes (intermediate-level vision). Considering its computational complexity, this grouping is one of the difficult problems that primate visual system has to solve [44].

The Gestalt phenomenology and quantitative psychological measurements established the rules, summarized in the ideas of good continuation [45, 46] and association field [47], that determine interactions between fragmented edges such that they extend along continuous contours joining the edges in the way they will normally be grouped together to faithfully convey a scene meaning. Evidence accumulated in psychological and physiological studies suggests that the human visual system utilizes a local grouping process (association field) with two very simple rules: collinearity (receptive fields aligned along a line) and cocircularity (receptive fields aligned along a circle with the preferred orientation orthogonal to the tangents of the circle [48]) with underlying scale invariant statistics for both geometric arrangements in natural scenes. These rules were confirmed in [49, 50] by statistical analysis of natural scenes. Two basic intermediate-level descriptors that the brain employs

in grouping elements into global objects are the medial axis transformation [51], or symmetry structure [52, 53], and the curvature extrema [54, 55]. In fact, the medial axis, which the visual system extracts as a skeletal (intermediate-level) representation of objects [56], can be defined as a set of the centers of maximal circles inscribed inside the contour. The curvatures at the corresponding points of a contour are given by the inverse radii of the circles.

This discussion shows that the conformal camera should effectively model the eye’s imaging functions related to lower- and intermediate-level visions of natural scenes.

4. Projective Fourier Analysis

The projective Fourier analysis has been constructed by restricting geometric Fourier analysis of $\mathbf{SL}(2, \mathbb{C})$ —a direction in the representation theory of the semisimple Lie groups [57]—to the image plane of the conformal camera [5, Section 5.1]. The resulting projective Fourier transform (PFT) of a given image intensity function $f(z) \in L^2(\mathbb{C})$ is the following:

$$\hat{f}(s, k) = \frac{i}{2} \int f(z) |z|^{-is-1} \left(\frac{z}{|z|} \right)^{-k} dz d\bar{z}, \quad (9)$$

where $(s, k) \in \mathbb{R} \times \mathbb{Z}$, and, if $z = x_3 + ix_1$, then $(i/2) dz d\bar{z} = dx_3 dx_1$ is the Euclidean measure on the image plane.

In log-polar coordinates (u, θ) given by $\ln r e^{i\theta} = \ln r + i\theta = u + i\theta$, $\hat{f}(k, s)$ takes on the form of the standard Fourier integral

$$\hat{f}(s, k) = \int \int f(e^{u+i\theta}) e^u e^{-i(us+\theta k)} du d\theta. \quad (10)$$

Inverting it, we obtain the representation of the image intensity function in the (u, θ) -coordinates:

$$e^u f(u, \theta) = \frac{1}{(2\pi)^2} \sum_{k=-\infty}^{\infty} \int \hat{f}(s, k) e^{i(us+\theta k)} ds, \quad (11)$$

where $f(u, \theta) = f(e^{u+i\theta})$. We stress that although $f(e^{u+i\theta})$ and $f(u, \theta)$ are numerically equal, they are given on different spaces.

4.1. Discrete Projective Fourier Transform. In spite of the logarithmic singularity of log-polar coordinates, PFT of any function f integrable on $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ is finite:

$$\begin{aligned} |\hat{f}(s, k)| &\leq \int_0^{2\pi} \int_{-\infty}^{u_1} f(e^{u+i\theta}) e^u du d\theta \\ &= \int_0^{2\pi} \int_0^{r_1} f(re^{i\theta}) dr d\theta < \infty. \end{aligned} \quad (12)$$

This observation is crucial in constructing the discrete PFT as follows. By removing a disk $|z| \leq r_a$, we can regularize f such that the support of $f(u, \theta)$ is contained within $(\ln r_a, \ln r_b) \times [0, 2\pi)$ and approximate the integral in (10) by a double Riemann sum

$$\hat{f}\left(\frac{2\pi m}{T}, n\right) \approx \frac{2\pi T}{NM} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{uk} f(e^{uk} e^{i\theta_l}) e^{-2\pi i(mk/M + nl/N)} \quad (13)$$

with $M \times N$ partition points

$$\begin{aligned} (u_k, \theta_l) &= \left(\ln r_a + \frac{kT}{M}, \frac{2\pi l}{N} \right), \quad 0 \leq k \leq M-1, \\ &0 \leq l \leq N-1, \quad T = \ln(r_b/r_a). \end{aligned} \quad (14)$$

Then, introducing

$$f_{k,l} = \left(\frac{2\pi T}{MN} \right) f(e^{uk} e^{i\theta_l}), \quad f_{k,l} = \left(\frac{2\pi T}{MN} \right) f(u_k, \theta_l) \quad (15)$$

and defining $\hat{f}_{m,n}$ by

$$\hat{f}_{m,n} = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f_{k,l} e^{uk} e^{-i2\pi mk/M} e^{-i2\pi nl/N}, \quad (16)$$

we obtain

$$f_{k,l} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \hat{f}_{m,n} e^{-uk} e^{i2\pi mk/M} e^{i2\pi nl/N}. \quad (17)$$

Both expressions (16) and (17) can be computed efficiently by FFT algorithms.

On introducing complex coordinates $z_{k,l} = e^{uk} e^{i\theta_l}$ into (16) and (17), these expressions are referred to as the discrete projective Fourier transform (DPFT) and its inverse, respectively [4, 5]. When ‘‘pixels’’ locations $e^{uk} e^{i\theta_l}$ are transformed by the conformal camera’s action of $g \in \mathbf{PSL}(2, \mathbb{C})$ so that the function undergoes projective transformation $f(g^{-1} \cdot e^{uk} e^{i\theta_l}) = f(e^{u'_k} e^{i\theta'_l})$, its representation in (17) is given in terms of u'_k and θ'_l (instead of u_k , and θ_l) but with $\hat{f}_{m,n}$ unchanged. We refer to the representation transformations as projectively adapted characteristics of the projective Fourier transforms [4, 5]. These projective transformations are not given explicitly here since they are not used in this work.

5. DPFT in Computational Vision

We discussed before the relevance of the conformal camera to the intermediate-level vision task of grouping image elements into individual objects in natural scenes. Here we discuss the relevance of the data model of image representation based on DPFT to image processing in biologically mediated machine vision systems.

5.1. Modeling the Retinotopy with DPFT. The mapping $w = \ln(z \pm a) - \ln a$, where $a > 0$ removes logarithmic singularity and $\pm a$ indicate, for different signs, the left or right brain hemisphere, is the accepted approximation of the retinotopic structure of primate visual cortical areas and the midbrain SC [7, 58]. However, the DPFT that provides the data model for image representation can be efficiently computed by FFT only in log-polar coordinates given by the complex logarithm $w = \ln z$. This mapping has distinctive rotational and zoom symmetries

$$\ln(e^{i\theta} z) = \ln z + i\theta, \quad \ln(\rho z) = \ln z + \ln \rho \quad (18)$$

important in image identification (rotations and dilations in the retinal space corresponding to translations in the cortical (log-polar) space). Thus, we see that the Schwartz model of the retina comes with drastic consequences by destroying the rotation and zoom symmetries.

The following facts support our modeling of retinotopy with DPFT. First, for small $|z| \ll a$, $\ln(z \pm a) - \ln a$ is approximately linear while, for large $|z| \gg a$, it is dominated by $\ln z$. Secondly, to construct discrete sampling for DPFT, the image is regularized by removing a disc representing the fovea, which is possible because PFT in log-polar coordinates does not have a singularity at the origin; see (12). Thirdly, there is accumulated evidence pointing to the fact that the fovea and periphery have different functional roles in vision [59, 60] and very likely involve different image processing underlying principles. Finally, by the split theory of hemispherical image representation, the foveal region has a discontinuity along the vertical meridian, with each half processed in a different brain hemisphere [61].

We conclude this discussion with the following remark. Both models discussed above, as well as all other similar models, are, in fact, fovea-less models [62]. However, because the fovea is explicitly removed in our model, we plan to complement it in the future by including the foveal image representation.

5.2. Image Sampling for DPFT. The DPFT approximation was obtained using the rectangular sampling grid (u_k, θ_l) in (14), corresponding, under the mapping

$$u_k + i\theta_l \mapsto z_{k,l} = e^{u_k + i\theta_l} = r_k e^{i\theta_l}, \quad (19)$$

to a nonuniform sampling grid with equal sectors

$$\alpha = \theta_{l+1} - \theta_l = \frac{2\pi}{N}, \quad l = 0, 1, \dots, N-1 \quad (20)$$

but with exponentially increasing radii

$$\begin{aligned}\rho_k &= r_{k+1} - r_k = e^{u_{k+1}} - e^{u_k} \\ &= e^{u_k} (e^\delta - 1) = r_k (e^\delta - 1), \quad k = 0, 1, \dots, M-1,\end{aligned}\quad (21)$$

where $u_{k+1} - u_k$ is the spacing $\delta = T/M$ and $r_0 = r_a$, where r_a is the radius of the disc that has been removed to regularize logarithmic singularity. This sampling interface is shown in Figure 4.

Let us assume that we have been given a picture of the size $A \times B$ in pixel units, which is displayed with K dots per unit length (dpl). Then, the physical dimensions, in the chosen unit of length, of the pixel and the picture are $1/K \times 1/K$ and $A/K \times B/K$, respectively. Also, we assume that the retinal coordinates' origin (fixation) is the picture's center.

The central disc of radius r_0 represents the fovea with a uniform distribution of grid points, the number of the foveal pixels N_f given by $\pi r_0^2 = N_f/K^2$. This means that the fovea cannot increase the resolution, which is related to the distance of the picture from the eye. The number of sectors is obtained from the condition $2\pi(r_0 + r_1)/2 \approx N(1/K)$, where $N = \lfloor 2\pi r_0 K + \pi \rfloor$. Here $\lfloor a \rfloor$ is the closest integer to a . To get the number of rings M , we assume that $\rho_0 = r_0(e^\delta - 1) = 1/K$ and $r_b = r_M = r_0 e^{M\delta}$. We can take either $r_b = (1/K) \min(A, B)/2$ or $r_b = (1/K) \sqrt{A^2 + B^2}/2$. Thus, $\delta = \ln[(1 + 1/r_0 K)]$ and $M = (1/\delta) \ln(r_b/r_0)$.

Example 1. We let $A \times B = 512 \times 512$ and $K = 4$ per mm, so the physical dimensions in mm are 128×128 and $r_b = 128\sqrt{2}/2 = 90.5$. Furthermore, we let $N_f = 296$, so $r_0 = 2.427$ and $N = 64$. Finally, $\delta = \ln(10.7084/9.7084) \approx 0.09804$ and $(1/0.09804) \ln(90.5/2.427) \approx M = 37$. The sampling grid consists of points in polar coordinates: $(r_k + \rho_{k+1}/2, \theta_l + \pi/64) = (2.552e^{k \cdot 0.09804}, (2l + 1)\pi/64)$ $k = 0, 1, \dots, 36, l = 0, 1, \dots, 63$.

5.3. Imaging with the Conformal Camera. In the example from the previous section, the number of pixels in the original image is 262,144, whereas both foveal and peripheral representations of the image contain only 2,664 pixels. Thus, it results in about 100 times less pixels than in the original image. However, this reduction in the number of pixels comes at a price: not only does the small central region have the resolution required for a clear vision, it also has to be removed to regularize the logarithmic singularity. Therefore, the conformal camera with the DPFT-based image processing in the present state of development can support only the peripheral imaging functions of the visual system.

The most basic and frequent eye's imaging functions are connected with an incessant saccadic eye movement (about 200,000 saccades per day). The neural mechanisms of the RFs shifts and perisaccadic mislocalization, hypothesized to be involved in maintaining visual stability, are mainly supported by the peripheral visual processing. We use DPFT to model these phenomena in Section 6. The DPFT-based image representation could support the following hardware

setup. A set of samples $\{f_{k,l}\}_{M \times N}$ of an image f , where $f_{k,l} = (2\pi T/MN) f(e^{u_k} e^{i\theta_l})$, is obtained from a camera with anthropomorphic visual sensors (silicon retina) [63] or from an "exp-polar" scanner with the sampling geometry similar to the distribution density of the retinal ganglion cells. The DPFT is applied to $\{f_{k,l}\}_{N \times M}$ according to (16), and $\hat{f}_{k,l}$ is efficiently computed with FFT. Next, the IDPFT of $\hat{f}_{k,l}$, given in (17), is again computed with FFT. However this output from IDPFT renders the retinotopic image $f_{k,l} = (2\pi T/MN) f(u_k, \theta_l)$ (numerically equal to $f_{k,l}$) of the retinal samples in the cortical log-polar coordinates.

When the eyes remain fixed, motion of objects is perceived by the successive stimulation of adjacent retinal loci. These image transformations are modeled in the conformal camera by the corresponding covariant transformations of the image representation in terms of DPFT; see the end of Section 4.1. These transformations are not important in modeling perisaccadic perception and are not dealt with in this work. However, they will become important in modeling smooth-pursuit eye movements, which we plan to undertake in the near future.

5.4. Other Approaches to Foveate Vision. Of the numerical approaches to foveate (also called space-variant) vision, involving, for example, Fourier-Mellin transform or log-polar Hough transform, the most closely related to our work are results reported by Schwartz' group at Boston University. We note that the approximation of the retinotopy by a complex logarithm was first proposed by Eric Schwartz in 1977. This group introduced the fast exponential chirp transform (FECT) [64] in their attempt to develop numerical algorithms for space-variant image processing. Both FECT and its inverse were obtained by the change of variables in the spatial and frequency domains applied to the standard Fourier integrals. The discrete FECT was introduced somehow ad hoc and some basic components of Fourier analysis, such as underlying geometry or Plancherel measure, were not considered. In comparison, projective Fourier transform (PFT) provides an efficient image representation well adapted to projective transformations produced in the conformal camera by the group $\text{SL}(2, \mathbb{C})$ acting on the image plane by linear-fractional mappings. Significantly, PFT can be obtained by restricting geometric Fourier analysis of the Lie group $\text{SL}(2, \mathbb{C})$ to the image plane of the conformal camera. Thus, the conformal camera comes with its own harmonic analysis. Moreover, PFT is computable by FFT in log-polar coordinates given by a complex logarithm that approximates the retinotopy. It implies that PFT can integrate the head, eyes, and visual cortex into a single computational system. This aspect is discussed, with special attention to perisaccadic perception, in the remaining part of the paper. Another advantage of PFT is the complex (conformal) geometric analysis underlying the conformal camera. We demonstrated in Section 3.2 the relevance of this geometry to the intermediate-level vision problem of grouping local contours into individual objects of natural scenes.

The other approaches to space-variant vision use the geometric transformations, mainly based on a complex

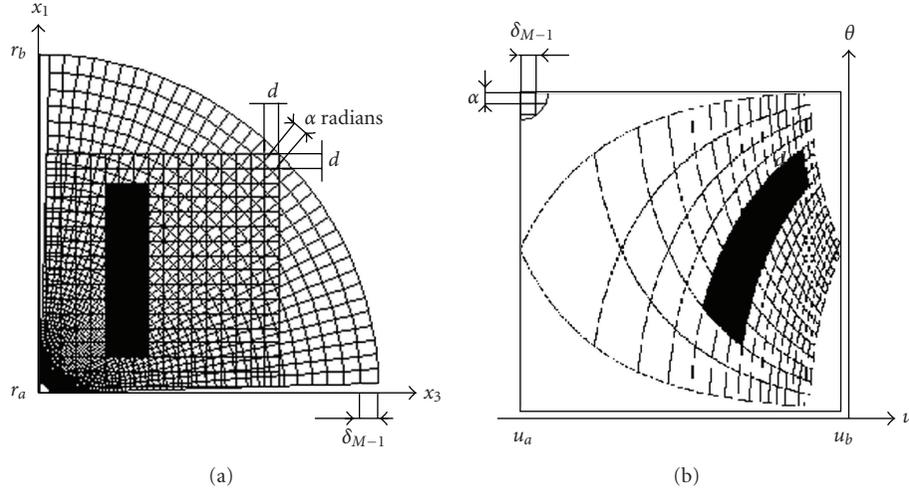


FIGURE 4: The retinocortical sampling interface. (a) The exp-polar sampling (the distance between circles displayed in the first quadrant changes exponentially) of a bar pattern. (b) The bar pattern in the cortical coordinates rendered by the inverse DPFT computed with FFT. The cortical uniform sampling grid, obtained by applying the complex logarithm to the exp-polar grid in (a), is shown in the upper left corner.

logarithmic function between the nonuniform (retinal) sampling grid and the uniform (cortical) grid for the purpose of developing computer programs for problems in robotic vision. We give only a few examples of such problems: tracking [65], navigation [66], detection of salient regions [67], and disparity estimation [68]. However, in contrast to our projectively covariant image processing carried out with FFT, they share high computational costs in the geometric transformation process for dynamic scenes.

6. Perisaccadic Perception with DPFT

A sequence of fast saccadic eye movements is necessary to process the details of a scene by fixating the fovea successively on the targets of interest. Given the frequency of three saccades per second and limited computational resources, it is critical that visual information is efficiently acquired without starting anew much of this acquisition process at each fixation. This is critically important in robotic designs based on the foveate vision architecture (silicon retina), and in this section we propose the front-end algorithmic steps in addressing this problem.

The model of perisaccadic perception presented in this section is based on the theory in [29] that states (as is most classically assumed) that an efference copy generated by SC, a copy of the oculomotor command to rotate the eyes in order to execute the saccade, is used to briefly shift the flashed probes' RFs in FEF/PEF toward the cortical fovea. Because the shift occurs in logarithmic coordinates approximating retinotopy, the model can also explain observed in laboratory experiments perisaccadic compression shown in Figure 5.

We recall the time course of events (Figure 2) that we are going to model. During the eyes' fixation, lasting, on average, 300 ms, the retinal image is sampled by ganglion cells and sent to cortical areas, including higher areas in the parietal and the frontal lobes. In particular, the next

saccade's target is selected in PEF/FEF areas and its position computed in subcortical SC area. About 50 ms before the onset of the saccade, during the saccade (~ 30 ms), and about 50 ms after the saccade, the visual sensitivity is reduced, and probes flashed around the impending saccade's target are not perceived in veridical locations; see Figure 5. Instead, a copy of the oculomotor command (efference copy) is used to translate the receptive fields of flashes recorded in the fovea-centered frame of the current fixation, remapping them into a target-centered frame. This internal remapping results in the illusory compression of flashes about the target. The cortical locations of the neural correlates of remapping are uncertain; it is required that these areas are retinotopically organized. These areas, of which most likely include PEF/FEF and V4 (and to a progressively lesser degree V3, V2, and V1), can be represented here by one retinotopic area [69].

6.1. The Model. The modeling steps are the following.

Step 1 (see Figure 6). The eye initially fixated at **F** is making the saccade to the target located at **T**. The four probes flashed around the upcoming saccade's target at **T** are projected into the retina and sampled by the photoreceptor/ganglion cells to give the set of samples $f_{m,n} = (2\pi T/MN)f(e^{u_m}e^{i\theta_n})$. Next, DPFT $\hat{f}_{k,l}$ is computed by FFT in log-polar coordinates (u_k, θ_l) , where $u_k = \ln r_k$. The inverse DPFT, computed again by FFT (the gray arrow), renders the image representation $f_{m,n} = (2\pi T/MN)f(u_m, \theta_n)$ in Cartesian log-polar coordinates—the four dots in (u, θ) -coordinates. The fovea, which is shown in yellow in Figure 6, is not included in log-polar coordinates, for these coordinates approximate only the extrafoveal part of the retina. For simplicity, we can take the radius of the fovea to be 1 so that the u -coordinate starts at 0.

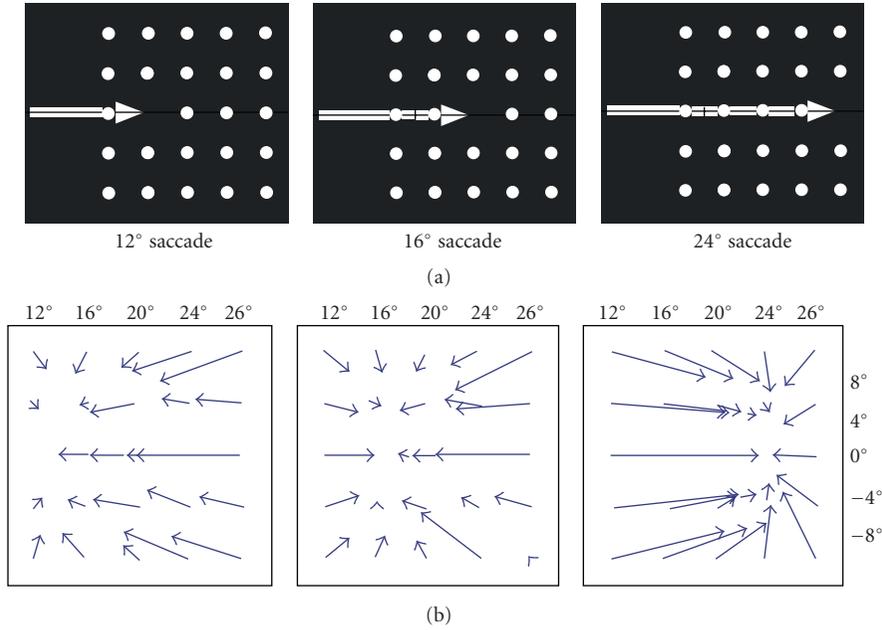


FIGURE 5: Experimental data (see [20]) of the absolute mislocalization (lower row) referenced to the true position of a flashed dot randomly chosen from an array of 24 dots and three different saccade amplitudes (upper row) are shown.

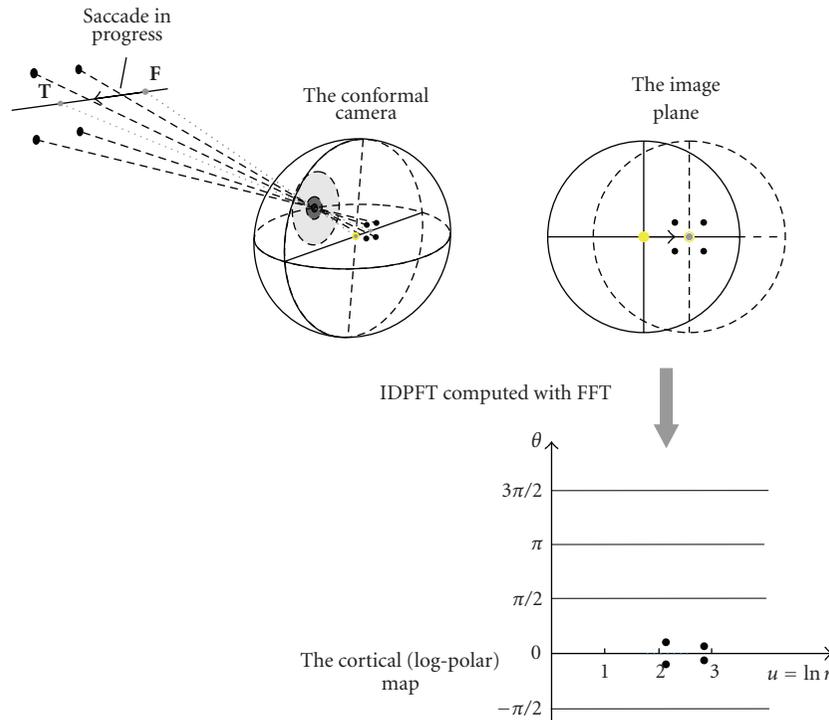


FIGURE 6: The local retinotopic mapping of four probes flashed around the saccade target at T.

Step 2 (see Figure 7). The log-polar image is multiplied by two characteristic set functions, $\chi_{A_i}(u_m, \theta_n)$, $i = 1$ or 2 ; the domain of each is shown in Figure 7 in a different color, the blue-enclosed region for A_1 and the green-enclosed for A_2 . We obtain two images $f_i(u_m, \theta_n) = \chi_{A_i}(u_m, \theta_n)f(u_m, \theta_n)$, $i = 1, 2$ representing cortical half-images into which the

image would be divided by the retinal vertical meridian after the fovea landed on the target at T. We recall that the characteristic function of a set A is defined by the following condition: $\chi_A(x)$ takes on 1 if $x \in A$ and 0 if $x \notin A$. The blue-enclosed image is reflected both in the vertical ($u = c_1$ line where c_1 is the midpoint of the projection of A_1 into

the u -axis) and in the horizontal ($\theta = 0$ line) axes and translated (blue arrow) while the green-enclosed image is only translated in the u -direction (green arrow).

These transformations are shown on the left of the gray arrow in Figure 7 while the results of these transformations (red dots) are shown on its right. The translation in the (u, θ) -coordinates (blue arrow) is obtained by the shift of the IDPFT

$$f_{m+h, n-j} = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{i2\pi hk/M} e^{-i2\pi jl/N} \hat{f}_{k,l} e^{-(u_k+j\delta)} e^{i2\pi mk/M} e^{i2\pi nl/N}, \quad (22)$$

(here $i = \sqrt{-1}$) which can be computed by FFT. The formula in (22) is the standard shift property of Fourier transform; the cortical image $f_{m,n}$ is translated by $-h$ pixels in the u -coordinate and by j pixels in the θ -direction, as the blue arrow shows in Figure 7. (Equivalently, the coordinate system is translated by h pixels in the u -coordinate and by $-j$ pixels in the θ -direction.) In (22), the inverse discrete Fourier transform is applied to $e^{i2\pi hk/M} e^{-i2\pi jl/N} \hat{f}_{k,l} e^{-(u_k+j\delta)}$ where $\delta = T/M$, $T = \ln R$, is the spacing in the u -coordinate (see (21) in Section 5.2) and $\hat{f}_{k,l} e^{-(u_k+j\delta)}$ is the original Fourier transform with the normalizing area factor in log-polar coordinates.

Further, the image reflection about the vertical axis of the A_1 region can be done with two consecutive transformations, each computable with FFT. These transformations consist of the reflection $f_1(u_m, \theta_n) \rightarrow f_1(-u_m, -\theta_n)$ followed by the translation $f_1(-u_m, -\theta_n) \rightarrow f_1(2c_1 - u_m, -\theta_n)$. We note that the reflection can be obtained by applying Fourier transform twice to the original image, which follows directly from the Fourier transform definition. The red dots represent peripheral receptive fields shifted to the frame centered at the upcoming saccade target.

Step 3 (see Figure 8). This perisaccadic compression is obtained by decoding the cortical image representation to the visual field representation:

$$\begin{aligned} f_{m-h, n+j} &= \left(\frac{2\pi T}{MN} \right) f(u_m + h\delta, \theta_n - j\gamma) \\ &= \left(\frac{2\pi T}{MN} \right) f(e^{u_m+h(\delta)} e^{i(\theta_n-j\gamma)}) \\ &= \left(\frac{2\pi T}{MN} \right) f(e^{h\delta} r_m e^{i(\theta_n-j\gamma)}), \end{aligned} \quad (23)$$

where $i = \sqrt{-1}$ and $\delta = T/M$, $\gamma = 2\pi/N$. We see that, under the shift of the coordinate system (u_m, θ_n) by $(h\delta, -j\gamma)$, the original position of a dor at $r_m e^{i\theta_n}$ is transformed to $e^{-h\delta} r_m e^{i(\theta_n-j\gamma)}$, resulting in the compression shown in the scene in Figure 8 with red dots referenced by red arrows to the original positions of flashed probes (black dots). Although this step is not supported by FFT, a commonly used look-up table [65] could efficiently decode the cortical image

see the discussion in Section 7.1. Where and how the brain accomplished this step is the greatest mystery of primate perception.

In the modeling steps we presented above, the cortical translation shown by the green arrow in Figure 7 gives only compression of the two corresponding probes in the scene because $j = 0$. However, in the translation shown in Figure 7 by the blue arrow, $j \neq 0$ and the corresponding image undergoes the compression and rotation, both needed to have the fovea at the center of the four red dots in the scene shown in Figure 8, when the saccade lands at **T**. Because of this rotation, we need two reflections to have the original parity of the image.

Although we do not show here quantitative results of the modeling steps, the qualitative results can be seen if we translate the cortical image of the bar in Figure 4(b), say in the u -coordinate to the left by a number of pixels (pixels are shown in the left upper corner), and trace out mentally its retinal copy using the $d \times d$ square lattice in (a) and its log-polar image in (b). We can see that the bar in (a) will be compressed with respect to the origin of (x_3, x_1) -coordinates. We note that, as mentioned in Step 3, we cannot apply FFT to render this compressed retinal image.

The model presented in this section complements the theory proposed in [29]. Experimental results suggest a very tight time course followed by perisaccadic compression with its duration of about 130 ms and with the maximum mislocalization immediately before the saccade. In the model we propose here, this saccadic dynamics can be easily accounted for: the distance of the shift $(-h, j)$ (in terms of cortical pixels) can be taken as a function of time. Another aspect of perisaccadic compression that is accounted for in our modeling is the fact that not all of RFs undergo shift during eyes saccadic motion. In Step 2, the translations are applied to selected (salient) retinotopic areas. These two aspects of perisaccadic perception are not supported in a natural way by the model presented in [29].

6.2. On Modeling Global Retinotopy. The global retinotopy reflects the anatomical fact that the axons in the optic nerve leaving each eye split along the retinal vertical meridian when the axons originating from the nasal half of the retina cross at the optic chiasm to the contralateral brain's hemisphere and join the temporal half, which remains on the eye's side of the brain. This splitting and crossing reorganize the local retinotopy (log-polar mapping) such that the left brain hemisphere receives the right visual field projection and the right brain hemisphere receives the left visual field projection. According to the split theory [61, 70], which provides a greater understanding of vision cognitive processes than the bilateral theory of overlapping projections, there is a sharp foveal split along the vertical meridian of hemispherical cortical projections. Both hemispheric projections are connected by a massive corpus callosal bridge of about 200 M of neuronal fibers [71].

Although crucial for synthesizing 3D representation from the binocular disparities in the pair of 2D retinal images, we cannot fully address the global retinotopy here because

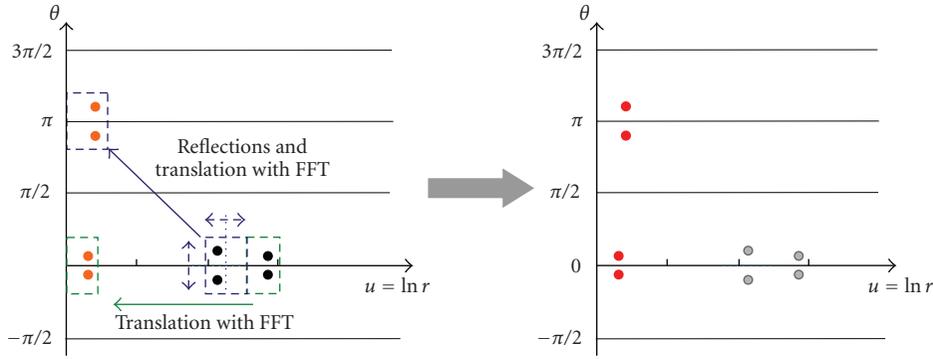


FIGURE 7: Modeling shifts of RFs of flashed probes. See text for detailed description.

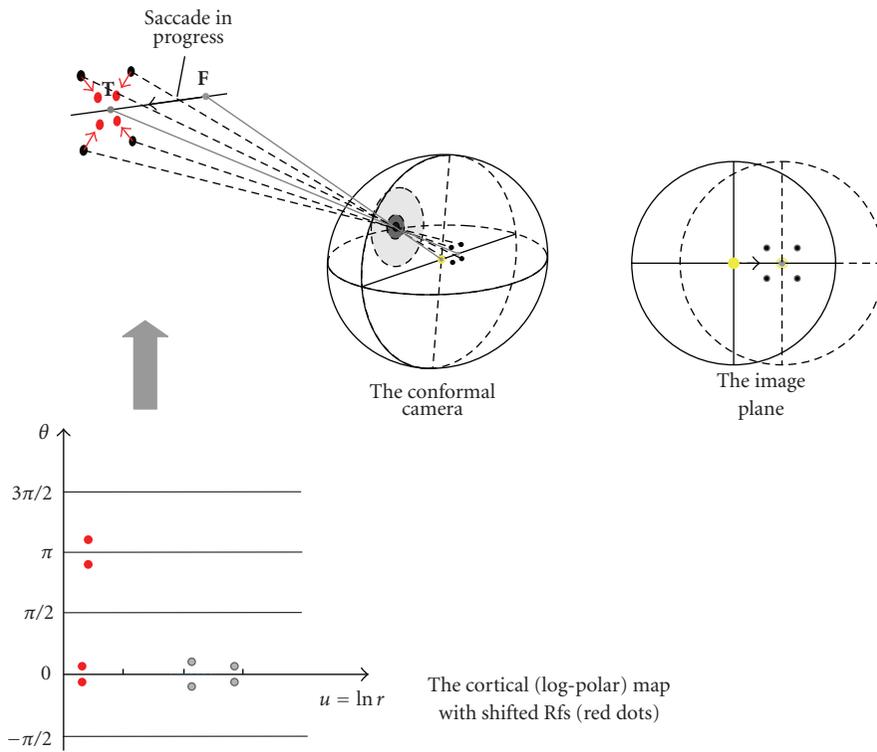


FIGURE 8: Perisaccadic compression of flashed probes perceived briefly during the saccadic eye movement. Note that this perception does not involve the visual sensory system.

the foveal image representation is not included in our modeling. However, the fovea-less global retinotopy can be easily modeled with DPFT by two reflections (cf., Step 2 in our modeling) both computable with FFT. Figure 9(b) shows the result of the reflection about the line $u = 0$ of the peripheral region given by $\pi/2 \leq \theta \leq 3\pi/2$. It is followed in 9(c) with the result of the reflection about $\theta = \pi/2$.

At this point, we can only graphically show what we expect to obtain when the foveal image representation complements the peripheral (log-polar) image representation we have developed in terms of projective Fourier transform. To this end, Figure 10 shows the peripheral region (gray) and the central foveal region (yellow). These two regions are connected by the transitional region (shaded with gray lines).

The green curve in Figure 10 shows the cortical projection of a straight line making an angle of $\pi/4$ with the retinal horizontal meridian and passing through the center of the fovea.

7. Discussion

Our model, which is based on PFT, uses the approximation of the retinotopy given by the complex logarithmic mapping $z \mapsto \ln z$ with $|z| \geq 1$, where 1 represents, with the appropriate normalization, the radius of the foveal region removed in our modeling. This mapping transforms PFT into the standard Fourier integral that is computable by FFT, providing the benchmark of efficiency not available in any

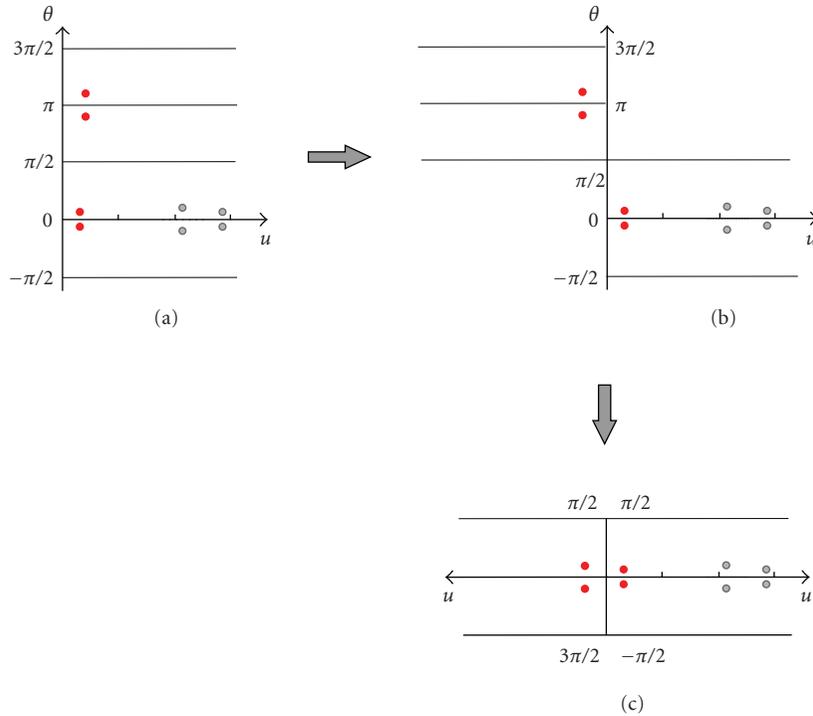


FIGURE 9: Two consecutive reflections that can be computed with FFT account for the global retinotopy without the foveal region.

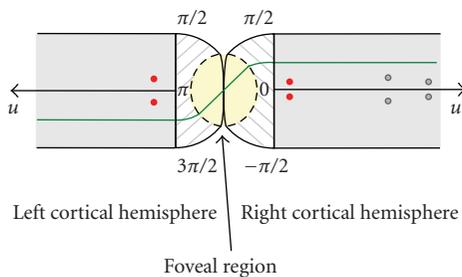


FIGURE 10: Schematic depiction of the global retinotopy with the peripheral region (gray) and the foveal region (yellow). Shaded in gray lines is the perifoveal region, which connects the other two regions. The red dots show the shifted neuronal activity in the receptive fields from the veridical position (gray dots) resulting in the illusory space compression.

other computational modeling of perisaccadic perception. Although the foveal system is indispensable to primate vision, it is rather less important to the proper functioning of the visuosaccadic system. In fact, neuronal cells in FEF/PEF, the higher cortical areas implicated in the RF shift mechanism that obtain retinotopic projections from the occipital cortex, have larger RFs and primarily code stimuli spatial locations rather than stimuli features [72]. We should note that, although we use the complex logarithmic approximation for all retinotopically organized brain areas, this approximation is well established only for the first visual cortical areas and SC of the midbrain [7, 58]. This is justified if we realize that most algorithmic principles employed by

natural visual systems need to be reformulated to better fit modern computational algorithms (FFT in our case).

The results obtained in the simulations in [29], where the Schwartz' retinotopic mapping $z \mapsto \ln(z + a)$ was used to approximate the cortical magnification factor, showed the unidirectional shift component of mislocalization superimposed on perisaccadic compression. However, it was noticed there that this component did not scale with the compression according to the experimental data. The unidirectional shift is absent in the model presented here because, in our approximation of retinotopy, the parameter a is zero. Since the unidirectional perisaccadic shift has a different neural origin (as it is primarily caused by delays in neuronal signals) than perisaccadic compression (caused by remapping), it should not be accounted for by just this parameter.

Further, in both our model and the model in [29], perceptual compression is attributed to a translation of the origin of the logarithmic coordinate system, which results in a linear relation between perceived and actual probes' positions. Thus, the nonlinearity observed in [19, 21] is not accounted for in our modeling. However, asymmetry and nonlinearity present in experimental data could have casual origin resulting from multiple sources; we mention here three: (1) an asymmetric distribution of photoreceptor/ganglion cell density [73], (2) the average preferred fixation area located from the point of the highest cone density a distance of about half of the central fovea's radius ($70 \mu\text{m}$) [74], and finally, (3) fluctuations of the cortical surface curvature (and therefore the lengths of the geodesics) across hemispheres [75]. Given the incomplete understanding of the neural processes underlying perisaccadic perception,

it is impossible to distinguish between these “accidental” causes and the real neural mechanisms captured in modeling.

7.1. Relations to Other Models and the Current Research. Two computational theories of transsaccadic vision that have been proposed in visual neuroscience are related to our modeling, both with similar functional explanation of perisaccadic mislocalization by the cortical magnification factor. The first theory [29], which motivated our research, was discussed and compared with our model in the previous sections. In summary, our modeling can be seen as complementing the approach proposed in [29] by providing efficient image representation suitable for processing visual information during saccadic eye movement, and in particular, the classically assumed process of active remapping compensating for receptive fields displacements.

The second theory [18, 20] explains perisaccadic compression by spatial attention being directed to the target of a planned saccade. The authors proposed an elaborate computational modeling that assumes that the flashed stimuli RFs in cortical areas dynamically change position toward the saccade target RF as the result of the gain of feedback of the retinotopically organized activity hill of the saccade target in the oculomotor SC layer. This attention directed to the target increases spatial discrimination at the saccade target location before the saccade onset. The perceived spatial distortion of stimuli is the result of the cortical magnification factor of the visuo-cortical mapping (or retinotopy of the visuomotor pathways) when the position of each stimulus is decoded from activity of the neural ensemble. Thus, in this theory, different neural processes to those proposed in [29] are assumed: a local and transient change in the gain control around the saccade target in retinotopic-organized stimulus position represented by a hill of neural activity is inducing perisaccadic distortion of the perceived stimulus location. However, because circuitry underlying receptive field remapping is widespread and not well understood, it cannot be easily decided whether saccadic remapping is the cause or consequence of saccadic compression [29].

What really sets apart our modeling from other models is the fact that the computational efficiency is built into the modeling process, as all algorithmic steps (except the last one) involve computations with FFT. This is especially important because the incessant occurrence of saccades and the time needed for the oculomotor system to plan and execute each saccade require that visual information is efficiently processed during each fixation period without repeating, afresh, the whole process at each fixation [29].

All models proposed so far capture only the initial, front-end stage of remapping for a particularly simple scene of flashed probes and, though they explain the perisaccadic mislocalization phenomenon, they leave out the crucial modeling step of the integration of the objects’ features (pattern, color, ...) across saccades [30–33] that achieves stability of perception [14, 76]. Many issues must be understood better before this crucial modeling step is achieved. We give two examples. During a scene viewing, a salient map of the landmarks and behaviorally significant objects of the scene is

created and the RF shift updates the retinotopy of only this saliency map [13]. Although it is still unclear what a saliency map should be when viewing complex natural scenes, it points to the possibility of working with a sparse visual information data when performing Step 3 in the model outlined in Section 6.1. Thus, a lookup table approach [65] could be efficient enough for this step even when viewing a complex scene. Further, the time course of different stages in visual information processing in trans-saccadic perception and their influences on other cognitive processes is unclear. It is well known that scene gist recognition, when the scene is viewed for 50 ms [77], is critical in the early stage of scene perception, influencing more complex cognitive processes, such as directing our attention within a scene, facilitating object recognition, and influencing long-term memory. Only very recently [78] has it been found that peripheral vision is more useful for recognizing the gist of a scene than central vision (i.e., foveal + parafoveal vision) is, even though central vision is more efficient per pixel at processing gist.

Although the understanding of neural mechanisms involved in trans-saccadic perception is incomplete, a significant progress in understanding dynamic interaction taking place between different pathways in the visuosaccadic system has been recently made. In particular, the fundamental principles underlying perception of objects across saccades have been outlined [13]. Therefore, we should expect major advances in the near future. As a consequence, in robotic vision research, which is still wedged in-between the limited knowledge about biological visual processing and technological and software restrictions imposed by current cameras, scanners, and computers, it is becoming more important than ever to propose different, even if competing, perspectives on how to model known processes involved in trans-saccadic perception. In this article we proposed a comprehensive, biologically mediated engineering approach to model an active vision system. Our modeling, which is efficiently supporting both hard-wired eccentricity-dependent visual resolution and front-end modeling of mechanisms that may contribute to continuity and stability of trans-saccadic perception, is based on an abstract and less intuitive camera model with underlying nonmetric (conformal) Möbius geometry. However, our initial study of the smooth-pursuit eye movements, which complement fixations and saccades in the scanpath, indicates that the conformal camera with its DPFT-based image representation will also be important in processing visual information during the pursuits.

Further, the conformal camera geometry’s effectiveness in the intermediate-level vision problems and the perspective covariant projective Fourier analysis, well adapted to retinotopy, strongly suggest that DPFT-based image representation should be useful in modeling the neural processes that underlie the transfer of the objects’ features across saccades and maintain the continuity and stability of perception.

Finally, it was observed that saccades cause not only a compression of space, but also a compression of time [79]. In order to preserve visual stability during the saccadic

scanpath, receptive fields undergo a fast remapping at the time of saccades. When the speed of this remapping approaches the physical limit of neural information transfer, relativistic-like effects are psychophysically observed and may cause space-time compression [80, 81]. Curiously, this suggestion can also be accounted for in our model based on projective Fourier analysis since the group $SL(2, \mathbb{C})$ of image projective transformations in the conformal camera is the double cover of the group of Lorentz transformations of Einstein's special relativity; for a simple presentation [82, Section 1.2].

8. Summary

In this article, we presented a comprehensive framework we have developed for computational vision over the last decade, and we applied this framework to model some of the processes underlying trans-saccadic perception. We have done this by bringing, in one place, physiological and behavioral aspects of primate visual perception, the conformal camera's computational harmonic analysis, and the underlying conformal geometry. This allowed us to discuss the conformal camera's effectiveness in modeling a biologically mediated active visual system. First, the conformal camera geometry fully accounts for the basic concepts of cocircularity and scale invariance employed by the human vision system in solving the difficult intermediate-level vision problems of grouping local elements into individual objects of natural scenes. Second, the conformal camera has its own harmonic analysis—projective Fourier analysis—providing image representation and processing that is well adapted to image projective transformations and the retinotopic mapping of the brain visual and oculomotor pathways. This later assertion follows from the fact that the projective Fourier transform integrates the head, eyes (conformal cameras), and visual cortex into a single computational system. Based on this system, we presented a computational model for some neural processes of the perisaccadic perception. In particular, we modeled the presaccadic activity, which, through shifts of stimuli current receptive fields to their future postsaccadic locations, is thought to underlie the scene remapping of the current foveal frame to the frame at the upcoming saccade target. This remapping uses the motor command of the impending saccade and may help maintain stability of primate perception in spite of three saccadic eye movements per second with the eyeball's maximum speed of 700 deg/sec producing 200,000 saccades per day. Our modeling also accounted for perisaccadic mislocalization observed by human subjects in laboratory experiments. Finally, we compared our model with the other computational approaches in the modeling trans-saccadic perception and discussed further developments.

Acknowledgment

The author thanks Dr. Noriyasu Homma for helpful comments.

References

- [1] J. Turski, "Projective fourier analysis in computer vision: theory and computer simulations," in *Vision Geometry VI*, A. Y. Wu and L. J. Latecki, Eds., vol. 3168 of *Proceeding of SPIE*, pp. 124–135, 1997.
- [2] J. Turski, "Harmonic Analysis on $SL(2, \mathbb{C})$ and projectively adapted Pattern representation," *Journal of Fourier Analysis and Applications*, vol. 4, no. 1, pp. 67–91, 1998.
- [3] J. Turski, "Projective Fourier analysis for patterns," *Pattern Recognition*, vol. 33, no. 12, pp. 2033–2043, 2000.
- [4] J. Turski, "Geometric Fourier analysis of the conformal camera for active vision," *SIAM Review*, vol. 46, no. 2, pp. 230–255, 2004.
- [5] J. Turski, "Geometric Fourier analysis for computational vision," *Journal of Fourier Analysis and Applications*, vol. 11, no. 1, pp. 1–23, 2005.
- [6] J. Turski, "Computational harmonic analysis for human and robotic vision systems," *Neurocomputing*, vol. 69, no. 10–12, pp. 1277–1280, 2006.
- [7] E. L. Schwartz, "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding," *Vision Research*, vol. 20, no. 8, pp. 645–669, 1980.
- [8] J. Turski, "Harmonic analysis for cognitive vision: perisaccadic perception," in *Human Vision and Electronic Imaging XIV*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 7240 of *Proceeding of SPIE*, pp. 72401A-1–72401A-5, 2009.
- [9] P. H. Schiller, "The neural control of visual guided eye movements," in *Cognitive Neuroscience of Attention: A Developmental Perspective*, J. E. Richards, Ed., pp. 3–50, Erlbaum Associates, Lawrence, Kan, USA, 1998.
- [10] R. H. Wurtz, "Neuronal mechanisms of visual stability," *Vision Research*, vol. 48, no. 20, pp. 2070–2089, 2008.
- [11] C. L. Colby and M. E. Goldberg, "Space and attention in parietal cortex," *Annual Review of Neuroscience*, vol. 22, pp. 319–349, 1999.
- [12] J.-R. Duhamel, C. L. Colby, and M. E. Goldberg, "The updating of the representation of visual space in parietal cortex by intended eye movements," *Science*, vol. 255, no. 5040, pp. 90–92, 1992.
- [13] D. Melcher and C. L. Colby, "Trans-saccadic perception," *Trends in Cognitive Sciences*, vol. 12, no. 12, pp. 466–473, 2008.
- [14] E. P. Merriam, C. R. Genovese, and C. L. Colby, "Remapping in human visual cortex," *Journal of Neurophysiology*, vol. 97, no. 2, pp. 1738–1755, 2007.
- [15] M. A. Sommer and R. H. Wurtz, "A pathway in primate brain for internal monitoring of movements," *Science*, vol. 296, no. 5572, pp. 1480–1482, 2002.
- [16] O. J. Grüsser, "On the history of the ideas of efference copy and reafference," *Clio Medica*, vol. 33, pp. 35–55, 1995.
- [17] A. Hunt and P. Cavanagh, "Clocking saccadic remapping," *Journal of Vision*, vol. 8, p. 818, 2008.
- [18] F. H. Hamker, M. Zirnsak, D. Calow, and M. Lappe, "The peri-saccadic perception of objects and space," *PLOS Computational Biology*, vol. 4, no. 2, pp. 1–15, 2008.
- [19] M. Kaiser and M. Lappe, "Perisaccadic mislocalization orthogonal to saccade direction," *Neuron*, vol. 41, no. 2, pp. 293–300, 2004.
- [20] M. Lappe, H. Awater, and B. Krekelberg, "Postsaccadic visual references generate presaccadic compression of space," *Nature*, vol. 403, no. 6772, pp. 892–895, 2000.

- [21] J. Ross, M. C. Morrone, and D. C. Burr, "Compression of visual space before saccades," *Nature*, vol. 386, no. 6625, pp. 598–601, 1997.
- [22] H. Honda, "The time courses of visual mislocalization and of extraretinal eye position signals at the time of vertical saccades," *Vision Research*, vol. 31, no. 11, pp. 1915–1921, 1991.
- [23] L. Matin and D. G. Pearce, "Visual perception of direction for stimuli flashed during voluntary saccadic eye movements," *Science*, vol. 148, no. 3676, pp. 1485–1488, 1965.
- [24] J. Schlag and M. Schlag-Rey, "Illusory localization of stimuli flashed in the dark before saccades," *Vision Research*, vol. 35, no. 16, pp. 2347–2357, 1995.
- [25] H. Awater and M. Lappe, "Mislocalization of perceived saccade target position induced by perisaccadic visual stimulation," *Journal of Neuroscience*, vol. 26, no. 1, pp. 12–20, 2006.
- [26] F. Ostendorf, C. Fischer, C. Finke, and C. J. Ploner, "Perisaccadic compression correlates with saccadic peak velocity: differential association of eye movement dynamics with perceptual mislocalization patterns," *Journal of Neuroscience*, vol. 27, no. 28, pp. 7559–7563, 2007.
- [27] S. M. C. I. Van Wetter and A. J. Van Opstal, "Experimental test of visuomotor updating models that explain perisaccadic mislocalization," *Journal of Vision*, vol. 8, no. 14, pp. 1–22, 2008.
- [28] K. Matsumiya and K. Uchikawa, "The role of presaccadic compression of visual space in spatial remapping across saccadic eye movements," *Vision Research*, vol. 43, no. 18, pp. 1969–1981, 2003.
- [29] R. VanRullen, "A simple translation in cortical log-coordinates may account for the pattern of saccadic localization errors," *Biological Cybernetics*, vol. 91, no. 3, pp. 131–137, 2004.
- [30] M. Demeyer, P. De Graef, J. Wagemans, and K. Verfaillie, "Transsaccadic identification of highly similar artificial shapes," *Journal of Vision*, vol. 9, no. 4, article 28, pp. 1–14, 2009.
- [31] D. Melcher, "Dynamic, object-based remapping of visual features in transsaccadic perception," *Journal of Vision*, vol. 9, no. 14, article 2, pp. 1–17, 2008.
- [32] D. Melcher, "Selective attention and the active remapping of object features in trans-saccadic perception," *Vision Research*, vol. 49, no. 10, pp. 1249–1255, 2009.
- [33] M. Wittenberg, F. Bremmer, and T. Wachtler, "Perceptual evidence for saccadic updating of color stimuli," *Journal of Vision*, vol. 8, no. 14, pp. 1–9, 2008.
- [34] S. Anstis, "Picturing peripheral acuity," *Perception*, vol. 27, no. 7, pp. 817–825, 1998.
- [35] D. C. Burr, M. C. Morrone, and J. Ross, "Selective suppression of the magnocellular visual pathway during saccadic eye movements," *Nature*, vol. 371, no. 6497, pp. 511–513, 1994.
- [36] Y. Yarbus, *Eye Movements and Vision*, Plenum Press, New York, NY, USA, 1967.
- [37] B. Girard and A. Berthoz, "From brainstem to cortex: computational models of saccade generation circuitry," *Progress in Neurobiology*, vol. 77, no. 4, pp. 215–251, 2005.
- [38] J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," *Nature*, vol. 434, no. 7031, pp. 387–391, 2005.
- [39] P. W. Glimcher, "Making choices: the neurophysiology of visual-saccadic decision making," *Trends in Neurosciences*, vol. 24, no. 11, pp. 654–659, 2001.
- [40] M. Berger, *Geometry I*, Springer, New York, NY, USA, 1987.
- [41] G. Jones and D. Singerman, *Complex Functions*, Cambridge University Press, Cambridge, UK, 1987.
- [42] M. Henle, *Modern Geometries. The Analytical Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 1997.
- [43] D. A. Brannan, M. F. Esplen, and J. J. Gray, *Geometry*, Cambridge University Press, Cambridge, UK, 1998.
- [44] S. Ullman, *Higher-Level Vision: Object Recognition and Visual Cognition*, MIT Press, Cambridge, Mass, USA, 1996.
- [45] K. Koffka, *Principles of Gestalt Psychology*, Harcourt & Brace, New York, NY, USA, 1935.
- [46] M. Wertheimer, *Laws of Organization in Perceptual Forms*, Harcourt, Brace & Jovanovitch, London, UK, 1938.
- [47] D. J. Field, A. Hayes, and R. F. Hess, "Contour integration by the human visual system: evidence for a local 'association field,'" *Vision Research*, vol. 33, no. 2, pp. 173–193, 1993.
- [48] P. Parent and S. W. Zucker, "Trace inference, curvature consistency, and curve detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 823–839, 1989.
- [49] C. C. Chow, D. Z. Jin, and A. Treves, "Is the world full of circles?" *Journal of Vision*, vol. 2, no. 8, pp. 571–576, 2002.
- [50] M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco, "On a common circle: natural scenes and gestalt rules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 4, pp. 1935–1940, 2001.
- [51] H. Blum, "Biological shape and visual science," *Journal of Theoretical Biology*, vol. 38, no. 2, pp. 205–287, 1973.
- [52] M. Leyton, "A theory of information structure I. General principles," *Journal of Mathematical Psychology*, vol. 30, no. 2, pp. 103–160, 1986.
- [53] M. Leyton, "A theory of information structure II: a theory of perceptual organization," *Journal of Mathematical Psychology*, vol. 30, no. 3, pp. 257–305, 1986.
- [54] F. Attneave, "Some informational aspects of visual perception," *Psychological Review*, vol. 61, no. 3, pp. 183–193, 1954.
- [55] D. D. Hoffman and W. A. Richards, "Parts of recognition," *Cognition*, vol. 18, no. 1–3, pp. 65–96, 1984.
- [56] I. Kovacs and B. Julesz, "Perceptual sensitivity maps within globally defined visual shapes," *Nature*, vol. 370, no. 6491, pp. 644–646, 1994.
- [57] A. W. Knap, *Representation Theory of Semisimple Groups: An Overview Based on Examples*, Princeton University Press, Princeton, NJ, USA, 1986.
- [58] N. Tabareau, D. Bennequin, A. Berthoz, J.-J. Slotine, and B. Girard, "Geometry of the superior colliculus mapping and efficient oculomotor computation," *Biological Cybernetics*, vol. 97, no. 4, pp. 279–292, 2007.
- [59] Y. Petrov, M. Carandini, and S. McKee, "Two distinct mechanisms of suppression in human vision," *Journal of Neuroscience*, vol. 25, no. 38, pp. 8704–8707, 2005.
- [60] J. Xing and D. J. Heeger, "Center-surround interactions in foveal and peripheral vision," *Vision Research*, vol. 40, no. 22, pp. 3065–3072, 2000.
- [61] M. Lavidor and V. Walsh, "The nature of foveal representation," *Nature Reviews Neuroscience*, vol. 5, no. 9, pp. 729–735, 2004.
- [62] C. F. R. Weiman, "Log-polar Binocular Vision System," Transition Research Corporation: NASA Phase II SBIR Final Report, 1994.
- [63] F. Berton, G. Sandini, and G. Metta, "Anthropomorphic visual sensors, in encyclopedia of sensors," C. A. Grimes, E. C. Dickey, and M. V. Pishko, Eds., vol. 10, pp. 1–16, American Scientific, New York, NY, USA, 2006.

- [64] G. Bonmassar and E. L. Schwartz, "Space-variant fourier analysis: the exponential chirp transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1080–1089, 1997.
- [65] A. Bernardino, J. Santos-Victor, and G. Sandini, "Foveated active tracking with redundant 2D motion parameters," *Robotics and Autonomous Systems*, vol. 39, no. 3-4, pp. 205–221, 2002.
- [66] G. Baratoff, C. Toepfer, and H. Neumann, "Combined space-variant maps for optical-flow-based navigation," *Biological Cybernetics*, vol. 83, no. 3, pp. 199–209, 2000.
- [67] N. Tamayo and V. J. Traver, "Entropy-based saliency computation in log-polar images," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications, (VISAPP '08)*, vol. 1, pp. 501–506, 2008.
- [68] R. Manzotti, A. Gasteratos, G. Metta, and G. Sandini, "Disparity estimation on log-polar images and vergence control," *Computer Vision and Image Understanding*, vol. 83, no. 2, pp. 97–117, 2001.
- [69] M. A. Silver and S. Kastner, "Topographic maps in human frontal and parietal cortex," *Trends in Cognitive Sciences*, vol. 13, no. 11, pp. 488–495, 2009.
- [70] C. D. Martin, G. Thierry, J.-F. Démonet, M. Roberts, and T. Nazir, "ERP evidence for the split fovea theory," *Brain Research*, vol. 1185, no. 1, pp. 212–220, 2007.
- [71] E. Zaidel and M. Iacobini, *The Parallel Brain: The Cognitive Neuroscience of the Corpus Callosum, A Bradford Book*, MIT Press, Cambridge, Mass, USA, 2003.
- [72] A. P. Saygin and M. I. Sereno, "Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex," *Cerebral Cortex*, vol. 18, no. 9, pp. 2158–2168, 2008.
- [73] V. H. Perry and A. Cowey, "The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors," *Vision Research*, vol. 25, no. 12, pp. 1795–1810, 1985.
- [74] N. M. Putnam, H. J. Hofer, N. Doble, L. Chen, J. Carroll, and D. R. Williams, "The locus of fixation and the foveal cone mosaic," *Journal of Vision*, vol. 5, no. 7, pp. 632–639, 2005.
- [75] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system," *NeuroImage*, vol. 9, no. 2, pp. 195–207, 1999.
- [76] J. Gottlieb, "From a different point of view: extrastriate cortex integrates information across saccades. Focus on "Remapping in Human Visual Cortex"" *Journal of Neurophysiology*, vol. 97, no. 2, pp. 961–962, 2007.
- [77] L. Fei-Fei, A. Iyer, C. Koch, and P. Perona, "What do we perceive in a glance of a real-world scene?" *Journal of Vision*, vol. 7, no. 1, article 10, pp. 1–29, 2007.
- [78] A. M. Larson and L. C. Loschky, "The contributions of central versus peripheral vision to scene gist recognition," *Journal of Vision*, vol. 9, no. 10, article 6, pp. 1–16, 2009.
- [79] M. C. Morrone, J. Ross, and D. Burr, "Saccadic eye movements cause compression of time as well as space," *Nature Neuroscience*, vol. 8, no. 7, pp. 950–954, 2005.
- [80] D. Burr and C. Morrone, "Time perception: space-time in the brain," *Current Biology*, vol. 16, no. 5, pp. R171–R173, 2006.
- [81] M. C. Morrone, J. Ross, and D. Burr, "Keeping vision stable: rapid updating of spatiotopic receptive fields may cause relativistic-like effect," in *Problems of Space and Time in Perception*, R. Nijhawan, Ed., Cambridge, UK, 2008.
- [82] S. Sternberg, *Group Theory and Physics*, Cambridge University Press, Cambridge, UK, 1994.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

