

Harmonic Analysis for Cognitive Vision: Perisaccadic Perception

Jacek Turski

Dept. of Computer and Mathematical Sciences, Univ. of Houston Downtown
One Main Street, Houston, TX, USA 77002

ABSTRACT

The data model for image representation in terms of projective Fourier transform (PFT) is well adapted to both image perspective transformations and the retinotopic mappings of the brain visual pathways. Here we model first aspects of the human visual process in which the understanding of a scene is built up in a sequence of fixations for visual information acquisition followed by fast saccadic eye movements that reposition the fovea on the next target. We make about three saccades per second with an eyeball's maximum speed of 700 deg/sec. The visual sensitivity is markedly reduced during saccadic eye movements such that, three times per second, there are instant large changes in the retinal images without almost any information consciously carried across fixations. Inverse Projective Fourier transform is computable by FFT in coordinates given by a complex logarithm that also approximates the retinotopy. Thus, it gives the cortical image representation, and a simple translation in log coordinates brings the presaccadic scene into the postsaccadic reference frame, eliminating the need for starting processing anew three times per second at each fixation.

Keywords: The conformal camera, projective Fourier transform, retinotopy, saccades, perisaccadic mislocalization

1. INTRODUCTION

Building on projective Fourier analysis of the conformal camera [13,14,15], a mathematical model numerically integrating the head, eyes and visual cortex into one binocular system was proposed in [16]. Here, by extending that binocular system, we model the very first dynamic aspects of the human visual perception processes in which the understanding of a scene is built up in a sequence of fixations of salient parts that stabilize the retinal image on the fovea for detailed analysis and fast saccadic eye movements that reposition the fovea on the next target of interest. This sequence of fixations and saccades, called the scanpath, is the most basic feature of the foveate vision and has evolved to overcome the limitation imposed by the fact that the high-acuity fovea is receiving less than 2% of the field of vision.

In this article we discuss how projective Fourier analysis could provide efficient computational framework for processing information along the scanpath. Firstly, because the inverse projective Fourier transform (IPFT) can be efficiently computed by a fast Fourier transform (FFT) in complex logarithmic coordinates that also approximate the retinotopy of the brain visual and oculomotor pathways [10,12], IPFT output resembles the cortical as well as the collicular image representations. Secondly, a simple translation in logarithmic coordinates remaps the presaccadic scene into the postsaccadic reference frame [17], eliminating the need for starting visual information processing anew three times per second at each fixation. It may also build up perceptual continuity across fixations in the scanpath. Moreover, the property, $\ln z + \ln w = \ln zw$, which makes fast computing of PFTs with FFT possible, causes the transient perceptual space compression around saccade's incoming target, referred to as perisaccadic mislocalization.

2. PROJECTIVE FOURIER TRANSFORM FROM THE CONFORMAL CAMERA

Projective Fourier analysis has been constructed from geometric harmonic analysis of the conformal camera shown in Fig. 1(a). In this camera, the image transformations are iterations of the h and k transformations shown in Fig. 1(a). It results in the group $SL(2, C)$ of 2×2 complex matrices of determinant 1, acting by linear-fractional mappings on the

image plane with complex coordinates via: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{dz+c}{bz+a}$; $z = x_3 + ix_1$. Thus, if $f(z)$ is the image intensity function, its

projective transformations are given by $f(g^{-1} \cdot z)$, $g \in SL(2, C)$. The conformal camera reduces the projective degrees of freedom to the minimal set of projections; see Fig. 1(b). For a detailed account we refer to [14].

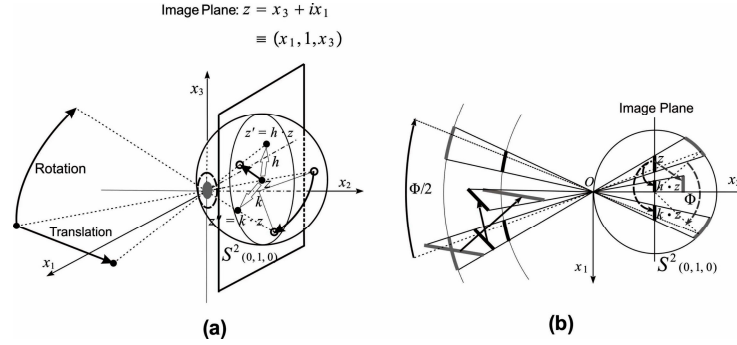


Fig. 1. (a) Image projective transformations in the conformal camera are generated by iterations of h and k transformations covering, respectively, translations and rotations of planar objects (see (b)) in the scene. (b) This 2D $x_3 = 0$ section of the conformal camera further explains how image projective transformations are generated and how the projective degrees of freedom are reduced in this monocular camera—there is one image projective transformation in the conformal camera corresponding to different planar objects translations and rotations in the 3D world.

Geometric harmonic analysis of the group $SL(2, C)$, when restricted to the image plane of the conformal camera, gives projective Fourier transform (PFT) $\hat{f}(s, k)$; $s \in R, k \in Z$, of an image $f(z)$ (Section 5.1 in [15]). In log-polar coordinates (u, θ) given by the complex logarithm, $\ln(re^{i\theta}) = \ln r + i\theta = u + i\theta$, this PFT becomes the standard Fourier integral. Its IPFT decomposes $\hat{f}(u, \theta)$ ($= f(e^u e^{i\theta})$) in terms of $\hat{f}(s, k)$. Since log-polar coordinates approximate locally the retinotopy, $\hat{f}(u, \theta)$ resembles the collicular or cortical projection of the retinal image $f(re^{i\theta})$.

The uniform $M \times N$ sampling grid (u_k, θ_l) , corresponding to a constant packing density of neurons in the cortex, gives the image (uniform) samples $\hat{f}_{k,l} = \hat{f}(u_k, \theta_l)$ in cortical log-polar coordinates and the non-uniform samples $f_{k,l} = f(r_k e^{i\theta_l}) = f(e^{u_k} e^{i\theta_l})$ in the image plane of the conformal camera that conforms to the density distribution of ganglion cells on the retina. Since PFT of an integrable image is finite, the sampling can be regularized with respect to the logarithmic singularity by removing a disc centered at the origin representing the fovea. The resulting discrete projective Fourier transform (DPFT) and its inverse, both computable by FFT, are given by

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

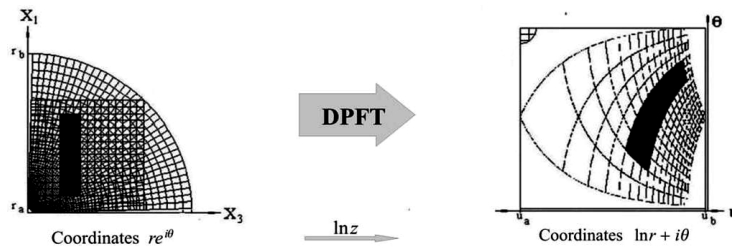


Fig. 2. Log-polar sampling (the distance between circles changes exponentially) of a bar pattern on the retina is shown on the left. The bar pattern in the cortical space rendered by the inverse DPFT computed with FFT is shown on the right. The cortical uniform sampling grid is shown only in the upper left corner.

In summary, the image representation in terms of DPFT can be computed by FFT only in cortical log-polar coordinates that approximate a local retinotopy, Fig. 2. This representation is fovea-less, as are all similar representations [18].

3. INTEGRATED BIONOCULAR VISION SYSTEM

Projective Fourier analysis provides an efficient computational framework that integrates the head, the eyes modeled by conformal cameras and visual cortex into one quantitative binocular system [16]; see Fig. 3 for a short description.

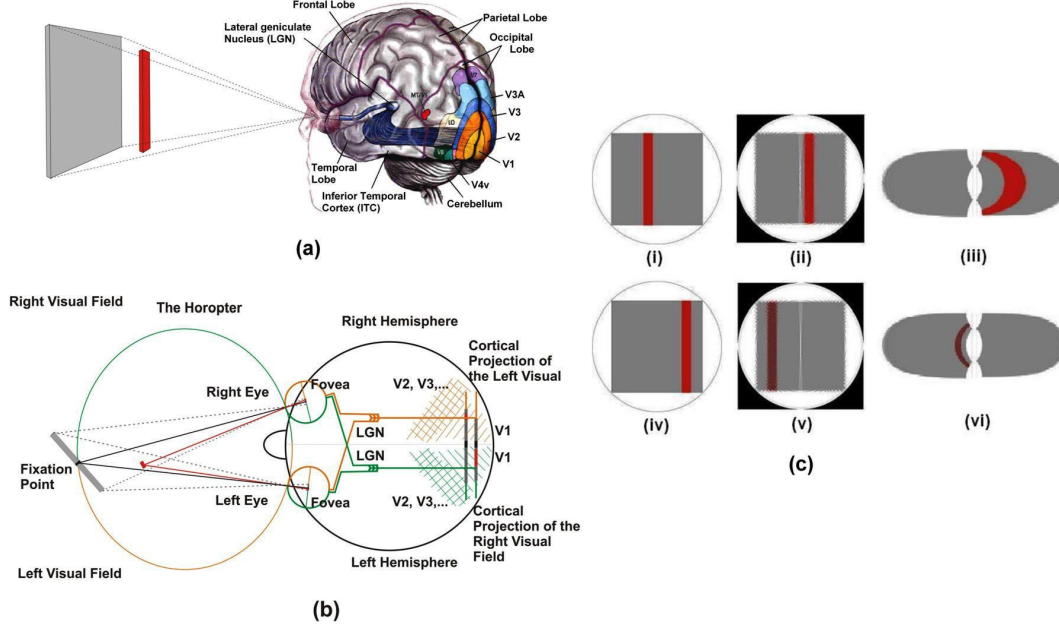


Fig. 3. (a) The 3D scene consisting of a gray square with a red bar located in the front of it. (b) Head-eyes-visual cortex integrated system with eyes modeled by conformal cameras and the scene from (a), as seen from above. The horopter of the system resembles the empirical horopter [16]. (c) (i) and (iv) show the scene seen by the right eye and the left eye, respectively. The retinal projections (ii) and (v), sampled by ganglion cells, are split along vertical meridians according to the split theory [7]. The retino-cortical mappings (iii) and (vi) are simulated in Matlab [1] and the cut-and-paste transformations are used to account for the global retinotopy.

According to the split theory [7], which provides a better understanding of vision cognitive processes than the bilateral theory of overlapping hemispherical cortical projections, the ganglion cells in the fovea are divided along the vertical meridian (cf., Fig. 3 (c) in (ii) and (v)) such that, for each eye, the nasal half of their axons cross over to the other side of the brain to join the temporal axons of the other eye along their pathway to the corresponding hemisphere, Fig. 3 (b).

4. FOVEATE VISION: THE SCANPATH

Whenever we open eyes in daylight, we execute, involuntarily or not, about three saccades per second, with an eyeball's maximum speed of about 700 deg/sec, to reposition the fovea for detailed analysis of salient objects of a scene in order to build up a 3D understanding. Although saccades are the simplest of bodily movements, they are controlled by very sophisticated neural processes of nearly every level of the brain. Their neuronal circuits include the superior colliculus (SC) of the midbrain for representing possible saccade targets, the parietal eye field (PEF) and frontal eye field (FEF) in the parietal and frontal lobes of the neocortex (which obtain inputs from many visual cortical areas) for assisting the SC in the control of the involuntary (PEF) and voluntary (FEF) saccades, cf., Fig. 4. They also project to the simple neuronal circuits, in the brainstem reticular formation in the midbrain, that ensure the saccade's outstanding speed and precision. Remarkably, many of the neural processes involved in saccade generation and control are amenable to precise quantitative studies such that even questions regarding the operation of the whole structure can be addressed [4]. This not only carries immense clinical significance [3], but also forms an essential preliminary stage in building our understanding of human vision—the knowledge that will eventually be transferred to the emerging field of neural engineering. Nevertheless, some of the neural processes of the visuosaccadic system remain virtually unknown. In what follows, we propose to model some of these neural processes with our DPFT.

4.1 Visual information along the scanpath

The visual sensitivity is markedly reduced during saccadic movements as we do not see moving images on the retinas. This largely unknown neural process is called saccadic suppression. Thus about three times per second, there are instant large changes in the retinal images without almost any information consciously carried across fixations. Furthermore, because the next target selection for the voluntary saccades takes place in the higher cortical areas involving cognitive processes [4], 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 period of about 300 ms, but without repeating much of this whole process at each fixation since it would require too many computational resources. Visual constancy, i.e., the fact that we are not aware of any discontinuity in the scene perception when executing the scanpath, is not perfect, however. Before the onset of the saccade, perceptual space is compressed around the saccade's incoming target [6,9], a phenomenon called perisaccadic mislocalization, which is outlined in Fig. 4.

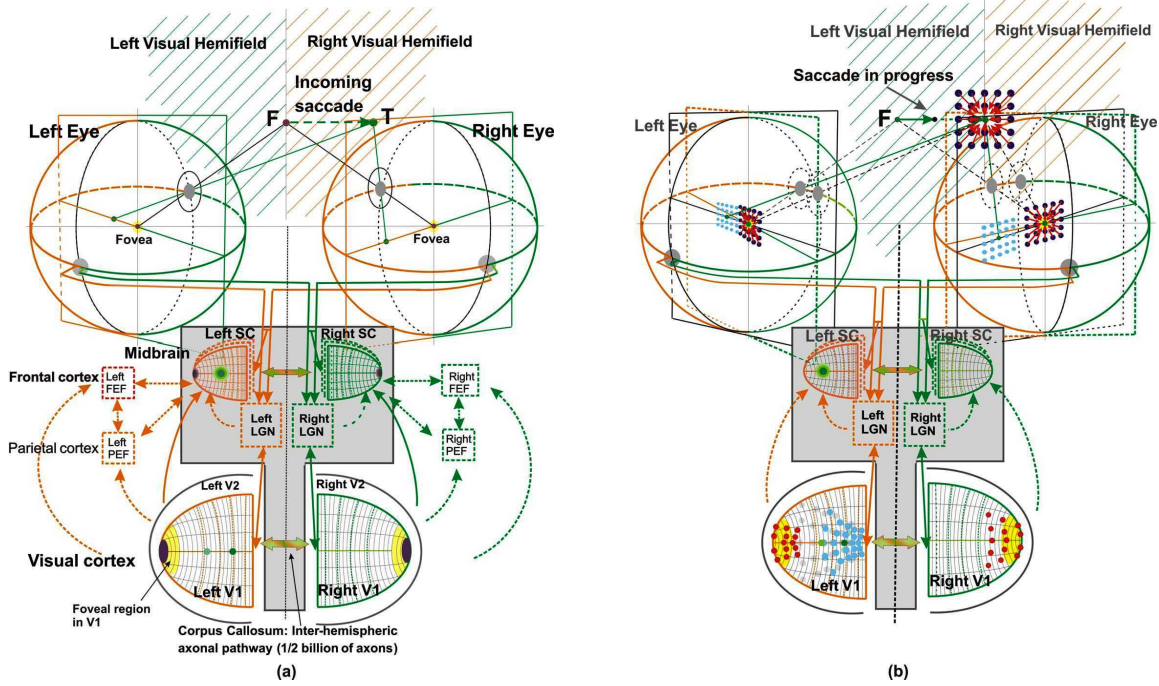


Fig. 4. (a) Eyes are held fixated at F by the foveal neuronal population (dark) in SC. The population at the retinotopic image of T (green) of SC calculates the position of the target T of an incoming saccade. SC also codes the motor command for the execution of the saccade. (b) Before the onset of the saccade, during, and after the saccade, the visual sensitivity is reduced and flashed dots around T are not perceived in veridical locations. Instead, the copy of the motor command (efference copy) is sent to translate the cortical image (light blue dots in V1) of flashed dots to remap it into a target-centered frame (red dots in V1). This internal remapping results in the illusory compression of flashed dots, shown by red arrows, called perisaccadic mislocalization. The compression is shown around the incoming target T even though the eyes fixation is moving from F to T. The cortical area of neural correlates of remapping is uncertain; the only requirement is that it is retinotopically organized. Although it could be PEF/FEF, here this area is represented by V1.

4.2 DPFT in modeling the scanpath vision

During the fixation of about 300 ms, the image is sampled by ganglion cells $f_{k,l} = f(r_k e^{i\theta_l})$ and its 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 gives cortical image representation $\hat{f}_{k,l} = \hat{f}(u_k, \theta_l) = f(e^{u_k} e^{i\theta_l})$ where disparity-sensitive cells contribute to the building 3D understanding of the scene. In the same fixation period, the next saccade's target is selected and position with respect to fovea is calculated and converted into the motor command to move the eyes. About 50 ms before the onset of saccade, during the saccade (about 30 ms) and about 50 ms after the saccade, the visual sensitivity is reduced. During this time interval, neural processes, using a copy of the eyes motor command, or afference copy [11], transiently shift the cortical image. In our modeling, this shift is obtained using the shift property of Fourier transform as applied to (l_2) :

$$\hat{f}(u_k + j\delta_M, \theta_l) = \hat{f}_{k+j,l} = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} e^{i2\pi mj/M} \hat{f}_{m,n} e^{-(u_k + j\delta_M)} e^{i2\pi mk/M} e^{i2\pi nl/N} \quad (2)$$

where δ_M is the spacing. It brings the presaccadic scene in fovea-centered coordinates into postsaccadic scene in target-centered coordinates. However, $\hat{f}(u_k + j\delta_M, \theta_l) = f(e^{u_k + j\delta_M} e^{i\theta_l}) = f(e^{j\delta_M} r_k e^{i\theta_l})$ compresses perceptual space.

5. CONCLUDING REMARKS

In this work we outlined a computational model based on DPFT of the conformal camera for the perisaccadic perception. The implementation of the model poses some challenges when computing with FFT and will be considered elsewhere.

It has been observed that saccades cause both a compression of space and a compression of time [8]. 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 may cause space-time compression [2]. Curiously, this suggestion can also be accounted for in our model based on projective Fourier analysis since the group of image projective transformations in the conformal camera, is the double cover of the group of Lorentz transformations of Einstein's special relativity.

REFERENCES

- [1] Bernardino, A., "LogPolar Wrapper and LogPolar Mapper," Matlab codes available over the Internet (2002).
- [2] Burr, D. and Morrone, C., "Time Perception: Space-Time in the Brain," *Current Biology*, 16, R171-R173 (2006).
- [3] Carpenter, R.H.S., "A Neurological Microcosm," *Advances in Clinical Neuroscience and Rehabilitation*, 4, 6-7 (2004).
- [4] Girard, B. and Berthoz, A., "From brainstem to cortex: Computational models of saccade generation circuitry," *Progress in Neurobiology*, 77, 215-251 (2005).
- [5] Glimcher, P. W., "Making choices: the neurophysiology of visual-saccadic decision making," *TRENDS in Neuroscience: Review*, 24, 654-659 (2001).
- [6] Lappe, M. Awater, H. and Krekelberg, D., "Postsaccadic visual references generate presaccadic compression of space," *Nature*, 403, 892-895 (2000).
- [7] Lavidor, M. and Walsh, V., "The nature of foveal representation," *Nature Reviews: Neuroscience*, 5, 729-735 (2004).
- [8] Morrone, M.C., Ross, J. and Burr, D., "Saccadic eye movements cause compression of time as well as space," *Nature Neuroscience*, 8, 950-954 (2005).
- [9] Ross, J., Morrone, M.C. and Burr, D., "Compression of visual space before saccades," *Nature*, 386, 698-701 (1997).
- [10] Schwartz, E., "Computational anatomy and functional architecture of striate cortex," *Vision Research*, 20, 645-669 (1980).
- [11] Sperry, R.W., "Neural basis of the spontaneous optokinetic response produced by visual inversion," *Journal of Computational Physiology and Psychology*, 43, 462-489 (1950).
- [12] Tabareau, N., Bennequin, D., Berthoz, A. and Slotine, J-J., "Geometry of the superior colliclic mapping and efficient oculomotor computation," *Biological Cybernetics*, 97, 279-292 (2007).
- [13] Turski, J., "Projective Fourier analysis for patterns," *Pattern Recognition*, 33, 2033-2043 (2000).
- [14] Turski, J., "Geometric Fourier Analysis of the Conformal Camera for Active Vision," *SIAM Review*, 46, 230-255 (2004).
- [15] Turski, J., "Geometric Fourier Analysis for Computational Vision," *Journal of Fourier Analysis and Applications*, 11, 1-23 (2005).
- [16] Turski, J., "Computational Harmonic Analysis for Human and Robotic Vision Systems," *Neurocomputing*, 69, 1277-1280 (2006).
- [17] VanRullen, R., "A simple translation in cortical log-coordinates may account for the pattern of saccadic localization errors," *Biological Cybernetics*, 91, 131-137, (2004).
- [18] Weiman, C.F.R., "Log-polar Binocular Vision System," Transition Research Corporation: NASA Phase II SBIR Final Report (1994).