3B2v8.06a/w (Dec 5 2003).51c +model NEUCOM : 10272 Prod.Type:FTP pp.1–4(col.fig.:NIL) ED:M.S.Usha PAGN:sree SCAN:thankavel

ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Neurocomputing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Computational harmonic analysis for human and robotic vision systems

Jacek Turski*

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

Dedicated to the memory of Professor Elias Deeba

Abstract

This paper studies the relevance of the conformal camera to computational vision with a particular focus on stereopsis. First we review projective Fourier analysis of the conformal camera and point to its unique attributes for modeling physiological aspects of perception. Then we design the *head–eye–visual cortex* integrated system with each eye modeled by the conformal camera. It provides a biologically realistic computational approach to the process of stereoscopic depth perception.

Keywords: The conformal camera; Projective Fourier transform; Retinotopic mapping; Binocular vision; Disparity; Stereopsis

1. Introduction

In biologically motivated machine vision systems, such as the active vision system in [8], the camera head with foveated sensor architecture (silicon retina) produces a digital image similar to the cortical image of the retinotopic mapping—the mapping that provides the initial stage in the process by which the brain processes visual information. More precisely, the nonuniform distribution of photoreceptors in the retina with the highest density at the fovea, the highly structured retinotopic arrangement of axons along the visual pathway, and a constant packing density of neurons in the visual cortex produce topographic images in the primary visual cortex (V1), with a significant magnification of the foveal region. The brain sends this information further to higher cortical areas (V2, V3/V3a, MT, ...) and in particular extracts visual cues about the 3D world from 2D images, such as the depth cues from binocular disparity and monocular motion parallax, and constructs a vivid percept of depth.

A complex logarithm provides approximations of a local topography of primate foveal and parafoveal regions [1]. To take advantage of the logarithmic space-variant imaging, the exponential chirp transform, a modified Mellin–Fourier transform, was constructed [2]. It retains the translation invariance, but also complicates the control of aliasing. More fundamentally, in active vision systems, the data model representing visual inputs should be well adapted to image projective transformations produced by different perspectives between objects and the camera, as well as to the output from the silicon retina sensors resembling the cortical topographic image. To this end, we constructed the conformal camera, and, based on it, we developed projective Fourier analysis furnishing the data model for image representation well adapted to image perspective transformations of planar objects [10]. Later, the discrete projective Fourier transform was obtained and numerical aspects, including aliasing, were discussed [11,12]. It was shown there that projective Fourier analysis is also well adapted to the retinotopic mapping of the brain visual pathway and provides the computational framework well suited for developing image processing tools explicitly designed for silicon retina sensors. We design here a model based on projective Fourier analysis that computationally integrates the head, eyes and visual cortex and expect that it will motivate building such a camera head as a part of an integrated multiple sensory system [3].

2. Projective Fourier analysis of the conformal camera

Projective Fourier analysis has been constructed from geometric harmonic analysis of the conformal camera shown in Fig. 1(A). For a comprehensive discussion we

*Tel.: 001 713 221 8401; fax; 001 713 221 8086.
E-mail address: TURSKI@DT.UH.EDU.

0925-2312/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.neucom.2005.12.091

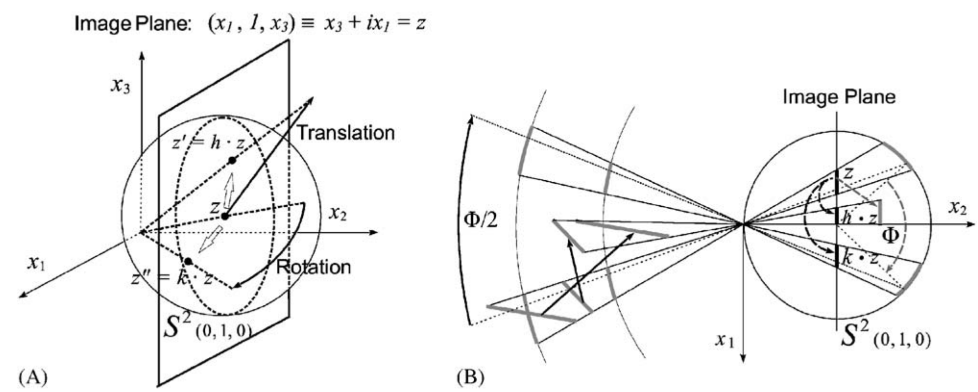

Fig. 1. (A) Image projective transformations described in the text are generated by iterations of "h" and "k" transformations. (B) Projective degrees of freedom reduced in the conformal camera (only $x_3 = 0$ section is shown).

refer to [12]. In the conformal camera, the image projective transformations are given by the Möbius group $\mathbf{PSL}(2,\mathbb{C}) = \mathbf{SL}(2,\mathbb{C})/\{\pm\mathbf{Id}\}$ acting by linear-fractional mappings on the image plane with complex coordinates via.

$$\pm\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{dz+c}{bz+a}, \quad z = x_3 + ix_1 \equiv (x_1, 1, x_3).$$

Thus, we must take the factor group $\mathbf{SL}(2,\mathbb{C})/\{\pm\mathbf{Id}\}$ consisting of 2×2 complex matrices of determinant 1 where those matrices that differ in sign are identified. The conformal camera reduces the projective degrees of freedom to the minimal set of projections—there is one image projective transformation in the conformal camera shown in Fig. 1(B) corresponding to different objects translated and rotated in the 3D world. Since this camera is a monocular system, rotations in the 3D world are generated by projecting an object into a sphere with the center at the focal point of the camera and rotating the sphere.

In geometric harmonic analysis of the group $\mathbf{SL}(2,\mathbb{C})$, an image is decomposed in terms of 1D unitary representations of the Borel subgroup of $\mathbf{SL}(2,\mathbb{C})$, with the coefficients of the representation given by *projective Fourier transform*. This transform in log-polar coordinates becomes the standard Fourier transform; it can be computed efficiently by a 2D fast Fourier transform (FFT). Using this decomposition, we could render digital image perspective transformations of a pattern (planar object) by computing only one discrete projective Fourier transform of the original pattern.

The mapping $w = k\ \ln(z + a)$ is an accepted approximation of the topographic structure of primate V1 [1]. The parameter a removes singularity but also destroys a nice behavior of the logarithm, namely $\ln(vz) = \ln v + \ln z$, which is important in computing projective Fourier transform with FFT. Remarkably, the newest study [4] supports a line singularity in the fovea: according to the split theory, which provides a greater understanding of vision cognitive processes than the bilateral theory of overlapping projections, the ganglion cells in the fovea are divided (cones in the fovea are individually connected to ganglion cells) such that the corresponding half of their axons travel with the nasal axons of each eye to join the temporal axons of the other eye. We choose $w = \ln z$ to model the packing density of the ganglion cells away from the foveola, which defines log-polar coordinates $(\ln r, \theta) \equiv \ln r + \mathrm{i}\theta = \ln(r\mathrm{e}^{\mathrm{i}\theta})$ in the cortex. This choice implies that both rotations and scale changes of a pattern correspond (although for the monocular vision) to cortical translations via.: $\ln(\rho\mathrm{e}^{\mathrm{i}\alpha} r\mathrm{e}^{\mathrm{i}\theta})$ corresponds to $(\ln r + \ln \rho, \theta + \alpha)$, which provides a significant computational advantage. Now, the feasibility of the data model of image representation based on projective Fourier analysis follows from the fact that, in log-polar coordinates $(\ln r, \theta)$, the projective Fourier transform takes on the standard Fourier integral form and can be computed by FFT. It can be done in the following steps [11]: an analog image is digitized ($d \times d$ pixels in Fig. 2(A)), its discrete projective Fourier transform (DPFT) is formed and expressed in log-polar coordinates which is finally computed by FFT. The output—the cortical image of the bar pattern—is shown in Fig. 2(B). Without DPFT the digital image must be resampled with nonuniform log-polar sampling geometry—pixels with the radial size changing logarithmically and a constant angular size—and expressed in the log-polar coordinate plane with uniform sampling geometry ($\delta_1 \times \delta_2$ pixels of the cortical image), a rather costly procedure for real images.

3. Integrated binocular vision system

The idea of how percepts are represented in the brain is still not well understood. For example, a striking feature of human stereopsis is that depth perception depends on relative disparity (i.e., the difference in retinal disparities of two features in the visual field), but encoding of relative disparity has only been recently investigated [5]. Thus, linking known physiological details with computational modeling and engineering designs are vital not only to the emerging field of neural engineering but can also be useful in interpreting relevant neurophysiological data [7]. In this section we present, based on the computational framework

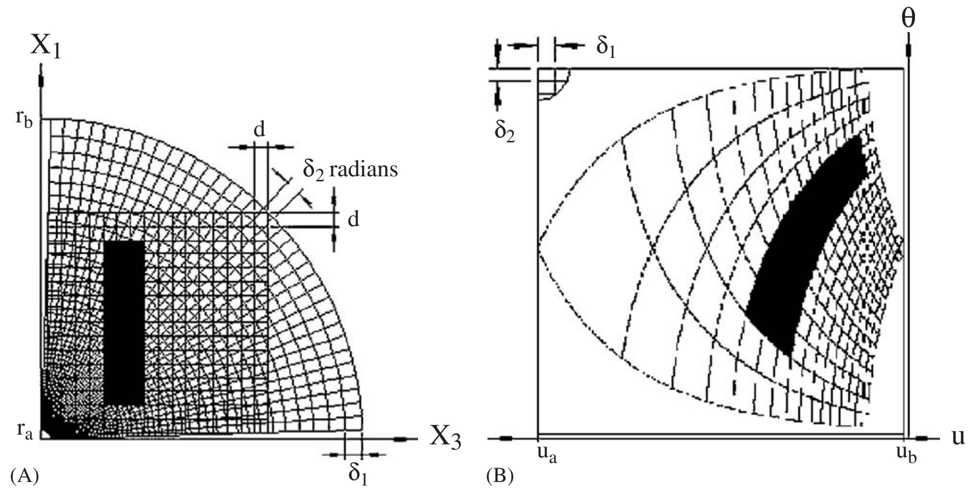

Fig. 2. (A) Log-polar sampling of a bar pattern (retinal image). (B) Inverse DPFT of the bar pattern in the log-polar coordinate plane (cortical image). Pixels are shown in the upper left corner.

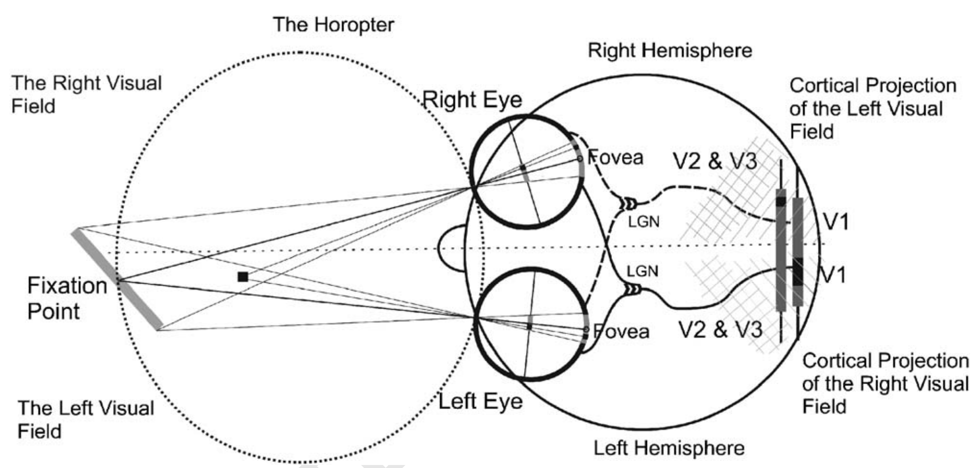

Fig. 3. Head–eye–visual cortex integrated system. The eyes are modeled by the conformal cameras.

of projective Fourier analysis, a design that *integrates* the head, eyes and visual cortex into one system. This system with eyes modeled by conformal cameras is shown in Fig. 3. The scene consists of a gray square and a black bar located in front of it, as seen from above. The fact that the scene is ''seen'' from slightly different vantage point by each camera, (A) and (C) in Fig. 4 (we ignore that images are inverted), implies that the system produces two different cortical images in V1, shown in (B) and (D) in Fig. 4. It follows from the study of iso-disparity surfaces in stereo configurations along the epipolar lines [6] that theoretical horopters of our vision system are conics that closely resemble the empirical horopters. To sum up, our binocular model *numerically integrates the head, eyes and visual cortex* and supplies the output that mimics the cortical images for which the brain's disparity-tuned neurons use simple geometry to estimate the relative depths of objects in the scene. Because disparity-tuned units, based on the response properties of real binocular cells, can effectively compute disparity maps from stereograms [7]. This bioncular system provides a computational framework for developing physiologically realistic models of stereo vision.

4. Additional remarks and conclusions

It is worth mentioning that the results we have established shed light on a recent remarkable construction. Starting in [9] with the space S of 2D shapes given by the set of all closed, simple, smooth curves in the plane, modern complex analysis shows that shapes can be classified up to an arbitrary Möbius transformation in $\mathbf{PSL}(2,\mathbb{C})$. The conformal camera has image projective transformations given by the Möbius group $\mathbf{PSL}(2,\mathbb{C})$, which implies that the classification of 2D shapes obtained in [9] is in fact both perspectively and retinotopically invariant.

We designed here the head–eye–visual cortex system based on computational theory, paying close attention to physiological details of the brain's visual pathway. We plan

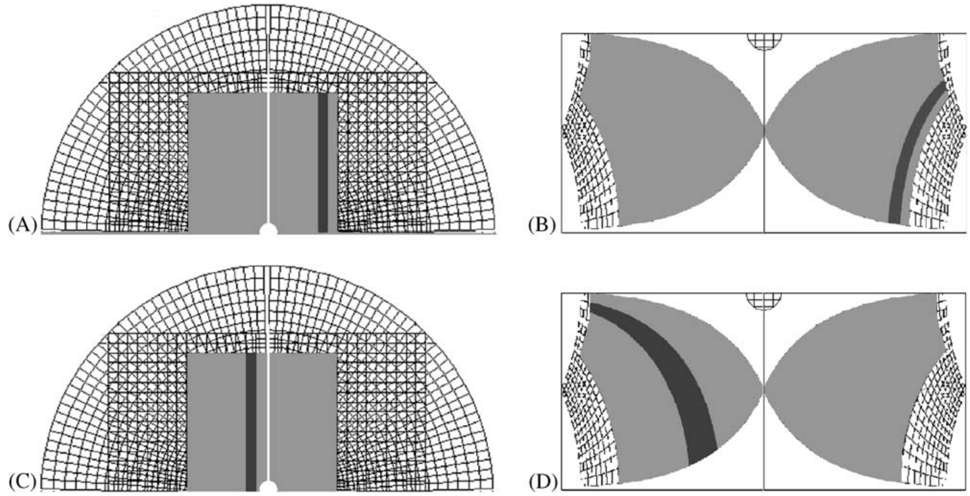

Fig. 4. 3D scene seen by the right conformal camera (A) and the left conformal camera (C). The corresponding cortical images are shown in (B) and (D).

to integrate this system further as a computational unit by developing a Lie-group formulation of its kinematics, and to simulate numerically some simple tasks of depth perception. However, to conduct quantitative experiments with tasks of depth perception, an active vision system with silicon retinas wired according to the head–eye–visual cortex system is needed. Finally, we intend to study topological characteristics of the cortical topography imposed by the partial crossing of the neuronal fibers between the brain hemispheres and a line singularity in the fovea.

Acknowledgment

The author thanks the CNS2005 Program Committee for a constructive criticism which contributed to the clarity of this paper. He also wishes to acknowledge the support from the University of Houston-Downtown to attend the CNS2005 conference.

References

[1] M. Balasubramanian, J. Polimeni, E.L. Schwartz, The V1–V2–V3 complex: quasiconformal dipole maps in primate striate and extra-striate cortex, Neural Networks 15 (2002) 1157–1163.

[2] G. Bonmassar, E.L. Schwartz, Space-variant Fourier analysis: the exponential chirp transform, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 1080–1089.

[3] R.A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, M.M. Williamson, The Cog Project: Building a Humanoid Robot, Lecture Notes in Computer Science, vol. 1562, Springer, Berlin, 1999, pp. 52–87.

[4] M. Lavidor, V. Walsh, The nature of foveal representation, Nat. Rev. Neurosci. 5 (2004) 729–735.

[5] P. Neri, H. Bridge, D.J. Heeger, Stereoscopic processing of absolute and relative disparity in human visual cortex, J. Neurophysiol. 92 (2004) 1880–1891.

[6] M. Pollefeys, S. Sinha, Iso-disparity surfaces for general stereo configurations, in: T. Pajdla, J. Matus (Eds.), Proceedings of the 8th ECCV, Part III, Lecture Notes in Computer Science, vol. 3023, Springer, Berlin, 2005, pp. 509–520.

[7] N. Qian, Binocular disparity and the perception of depth, Neuron 18 (1997) 359–368.

[8] G. Sandini, G. Metta, Retina-like sensors: motivations, technology and applications, in: T.W. Secomb, F. Barth, P. Humphrey (Eds.), Sensors and Sensing in Biology and Engineering, Springer, Berlin, 2002.

[9] E. Sharonm, D. Mumford, 2D-shape analysis using conformal mapping, in: IEEE Computer Society Conference on CVPR, vol. 2, 2004.

[10] J. Turski, Harmonic analysis on $\mathbf{SL}(2,\mathbb{C})$ and projectively adapted pattern representation, J. Fourier Anal. Appl. 4 (1998) 67–91.

[11] J. Turski, Geometric Fourier analysis of the conformal camera for active vision, SIAM Rev. 46 (2004) 230–255.

[12] J. Turski, Geometric Fourier analysis for computational vision, J. Fourier Anal. Appl. 11 (2005) 1–23.

Jacek Turski was awarded his Ph.D. from McGill University. After holding postdoctoral positions at the University of Manitoba and the University of Houston, he joined the University of Houston-Downtown where he is now a full professor in the Department of Computer and Mathematical Sciences. Recently, Turski has constructed projective Fourier analysis of the conformal camera in the framework of the representation theory of semisimple Lie groups. Based on this Fourier analysis, he is currently developing a physiologically realistic model of human and robotic vision systems.