Projective Fourier Analysis and Perisaccadic Perception

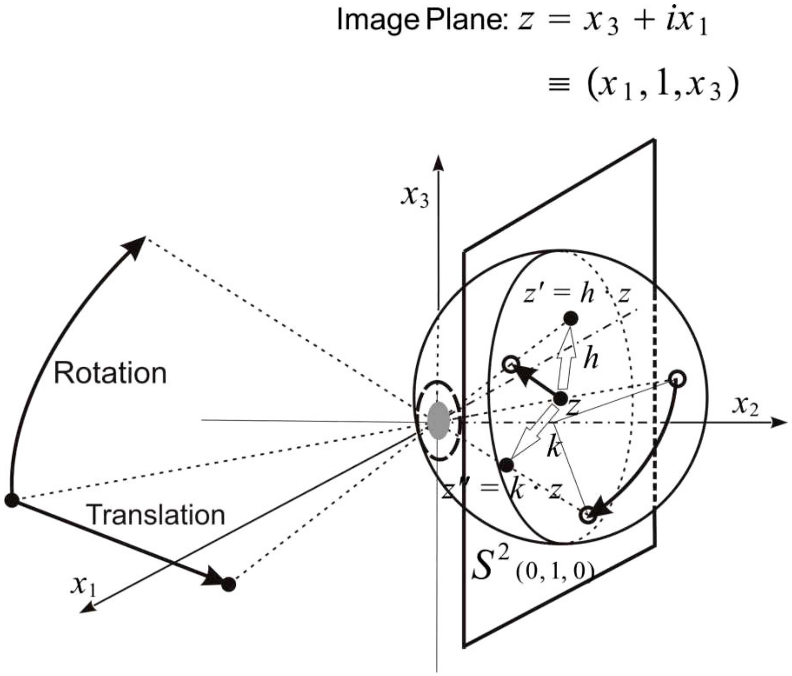
Jacek Turski
University of Houston-Downtown
turskij@uhd.edu
http://cms.dt.uh.edu/Faculty/TurskiJ/

OSA Vision Meeting, Rochester, NY, October 24-25, 2008

ABSTRACT

In this work, using **projective Fourier transform** we have developed [1,2], we model first aspects of the human visual process in which the understanding of a scene is built up in a sequence of attentional visual information acquisition followed by a fast saccadic eye movement that repositions the fovea on the next target. This sequence, called the **scanpath**, is the most basic feature of the **foveate vision**. We make three saccades per second with the maximum eyeball's speed of 700 deg/sec. The visual sensitivity is reduced during saccadic movement as we do not see moving images on the retinas. Therefore, three times per second, there are instant large changes in the retinal images without any information consciously carried between images. Inverse projective Fourier transform (IPFT) is fast computable by in complex logarithmic coordinates that also *approximates* the *retinotopy*. The output from IPFT resembles the cortical image and a simple translation in logarithmic coordinates brings the presaccadic scene into the postsaccadic reference frame [5]. It eliminates the need for starting processing anew three times per second at each fixation, but it introduces **perisaccadic mislocalization** [4,5]. It may also build up perceptual continuity across fixations in the scanpath.

The Conformal Camera – The Eye Model



$$k(\psi, \phi, \psi') = \pm \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}$$

$$k \cdot z = \frac{\overline{\alpha}\, z - \overline{\beta}}{\beta\, z + \alpha}$$

$$h(b_1, b_2, b_3) = \pm \begin{pmatrix} a & 0 \\ b & a^{-1} \end{pmatrix}$$

$$h \cdot z = \frac{a^{-1} z + b}{a}$$

$$\mathbf{SL}(2, \mathbb{C}) \ni \pm \begin{pmatrix} a\ b \\ c\ d \end{pmatrix} \cdot z = \frac{dz + c}{bz + a}$$

The Minimum Degree of Freedom

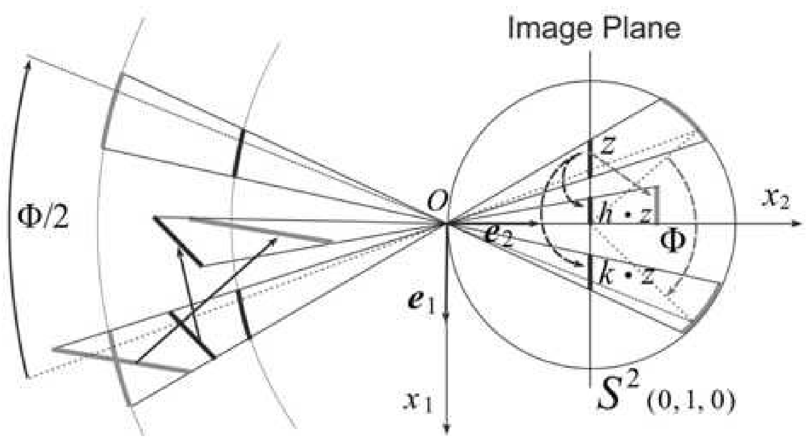


Image Projective Transformations

$$f(z) \rightarrow f(g^{-1} \cdot z); \quad g \in \mathbf{SL}(2, \mathbb{C})$$

Harmonic analysis on $\mathbf{SL}(2, \mathsf{C})$ gives **projective Fourier transform**

$$\widehat{f}(s,k) = \frac{i}{2}\int f(z)|z|^{-is-1}\left(\frac{z}{|z|}\right)^{-k} dz d\overline{z}; \quad k \in \mathsf{Z}, \ s \in \mathsf{R}$$

where if $z = x + iy$ then $\frac{i}{2} dz d\overline{z} = dxdy$

In **log-polar coordinates** (u, θ) where $u + i\theta = \ln r + i\theta = \ln(re^{i\theta})$

it becomes the standard Fourier transform

$$\widehat{f}(s,k) = \int\int f(e^{u+i\theta})e^{u}e^{-i(us+\theta k)}\, dud\theta$$

and can be inverted

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

However, $f(e^{u+i\theta}) = \mathrm{f}(u,\theta)$ are defined in **different** domains.

Uniform sampling grid $(\mathbf{u}_k, \theta_l)$ gives (uniform) cortical samples $\mathrm{f}_{k,l}$

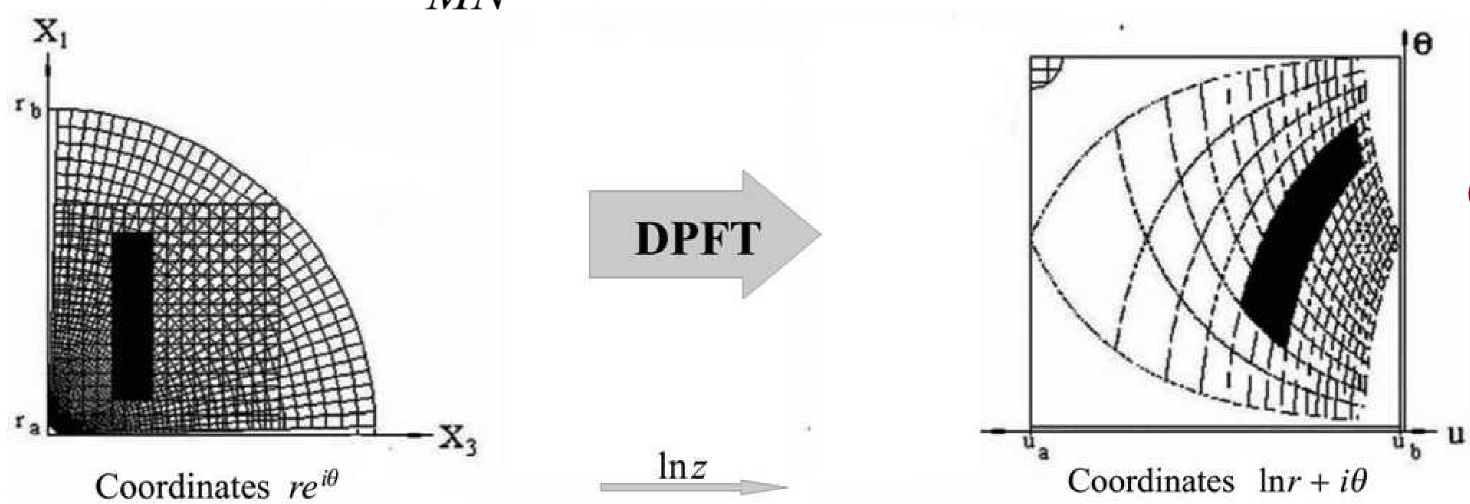
$$\mathrm{f}_{k,l} = \mathrm{f}(u_k, \theta_l) = f(e^{u_k} e^{i\theta_l}) = f(r_k e^{i\theta_l}) = f_{k,l}$$

and (non-uniform) retinal samples $f_{k,l}$. **DPFTs** can be computed by **FFT**

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

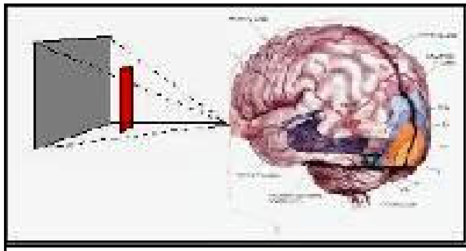
$$\mathrm{f}_{k,l} = \frac{1}{MN} \sum^{M-1} \sum^{N-1} \hat{f}_{m,n} e^{-u_k} e^{i2\pi nl/N} e^{i2\pi mk/M}$$

Retinal Image

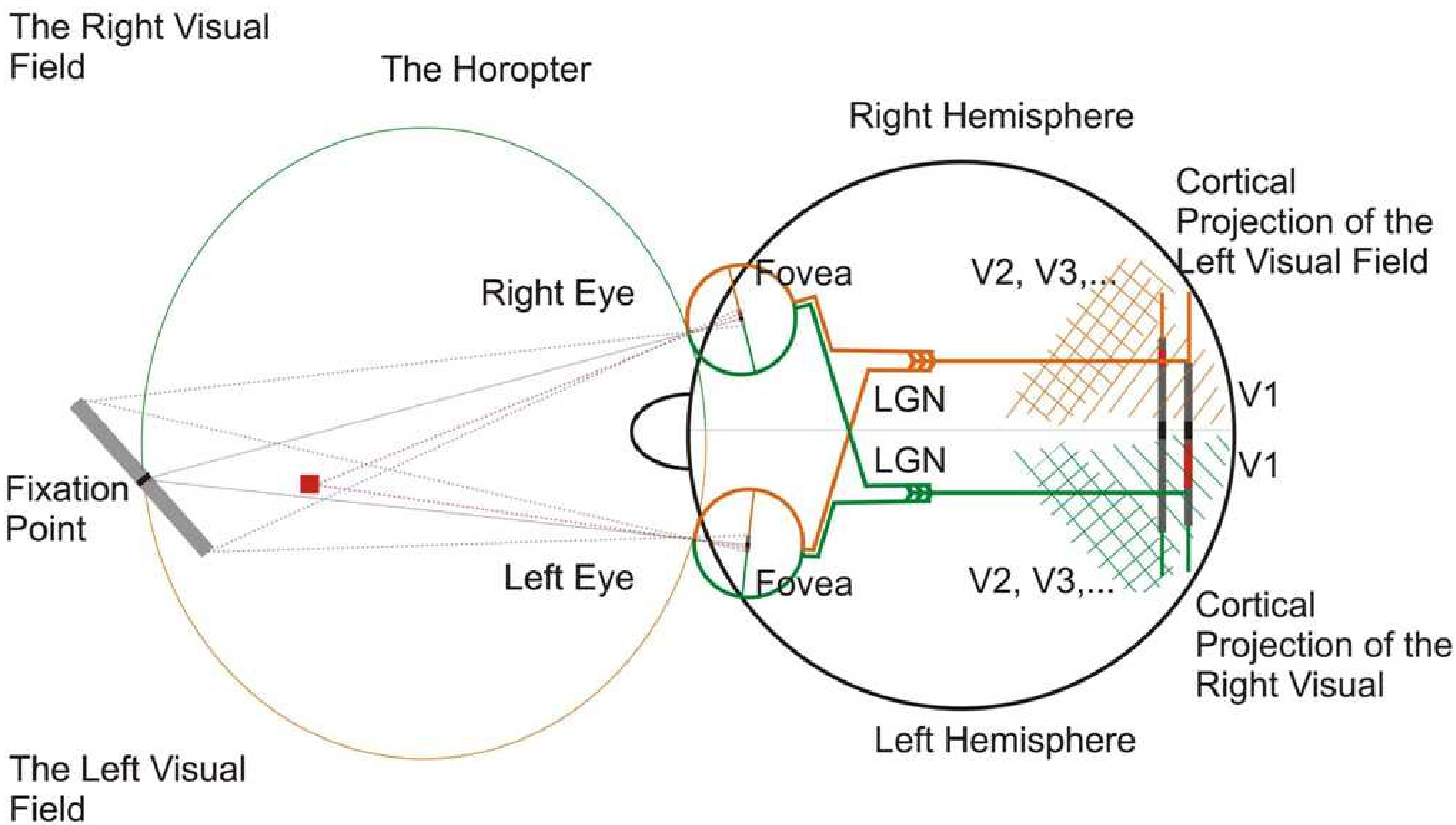


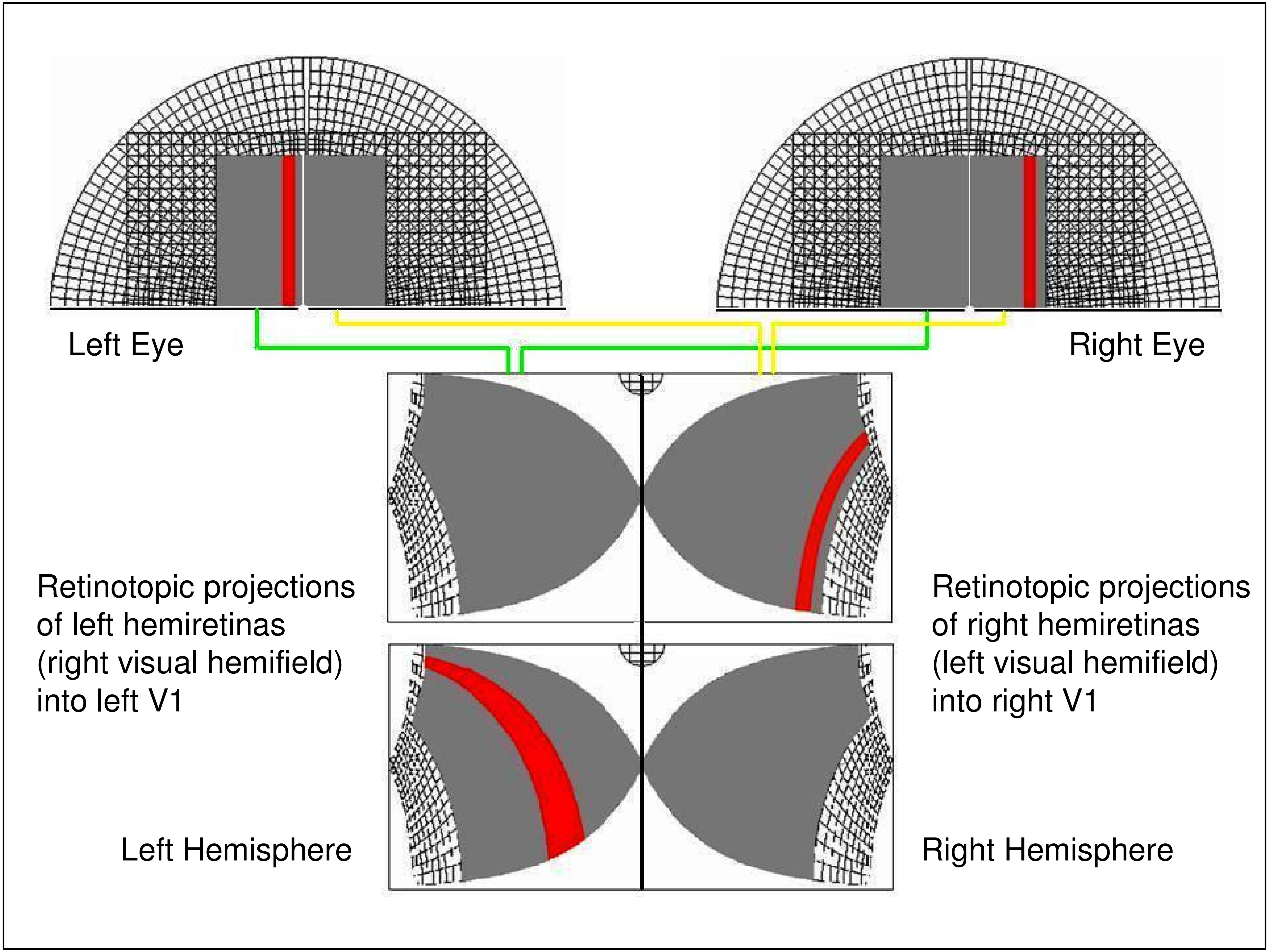
Cortical Image

Retinotopy

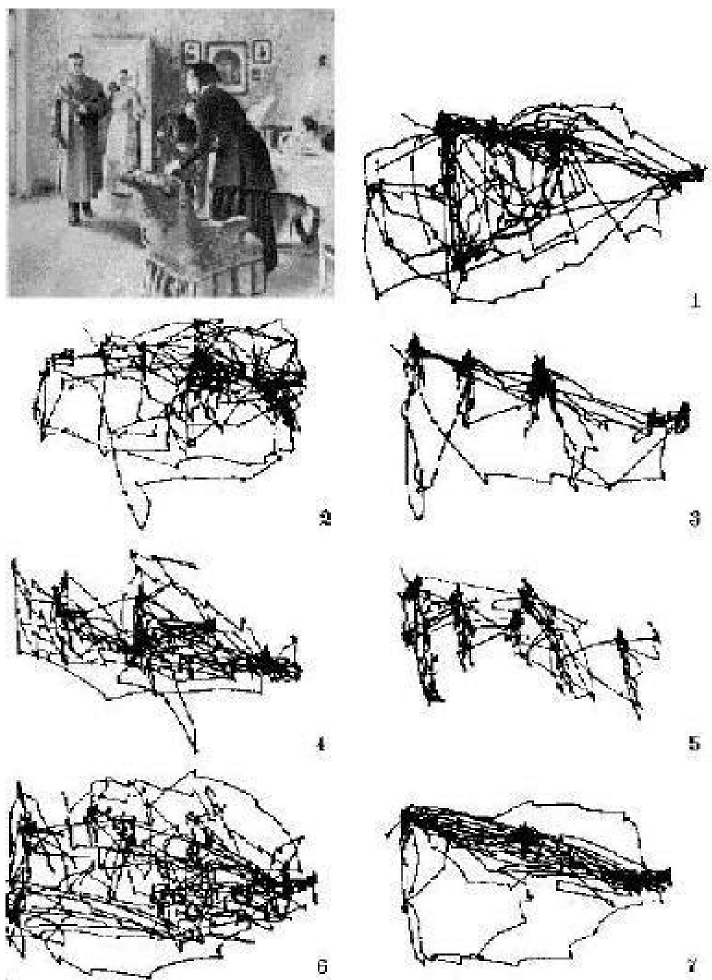


Binocular System with Conformal Cameras




Left Eye
Right Eye
Retinotopic projections of left hemiretinas (right visual hemifield) into left V1
Retinotopic projections of right hemiretinas (left visual hemifield) into right V1
Left Hemisphere
Right Hemisphere

Foveate Vision: The Scanpath



A.Yarbus, Plenum Press, New York, 1967.

1) Free examination 3) give ages of the people

6) remember position of people and objects

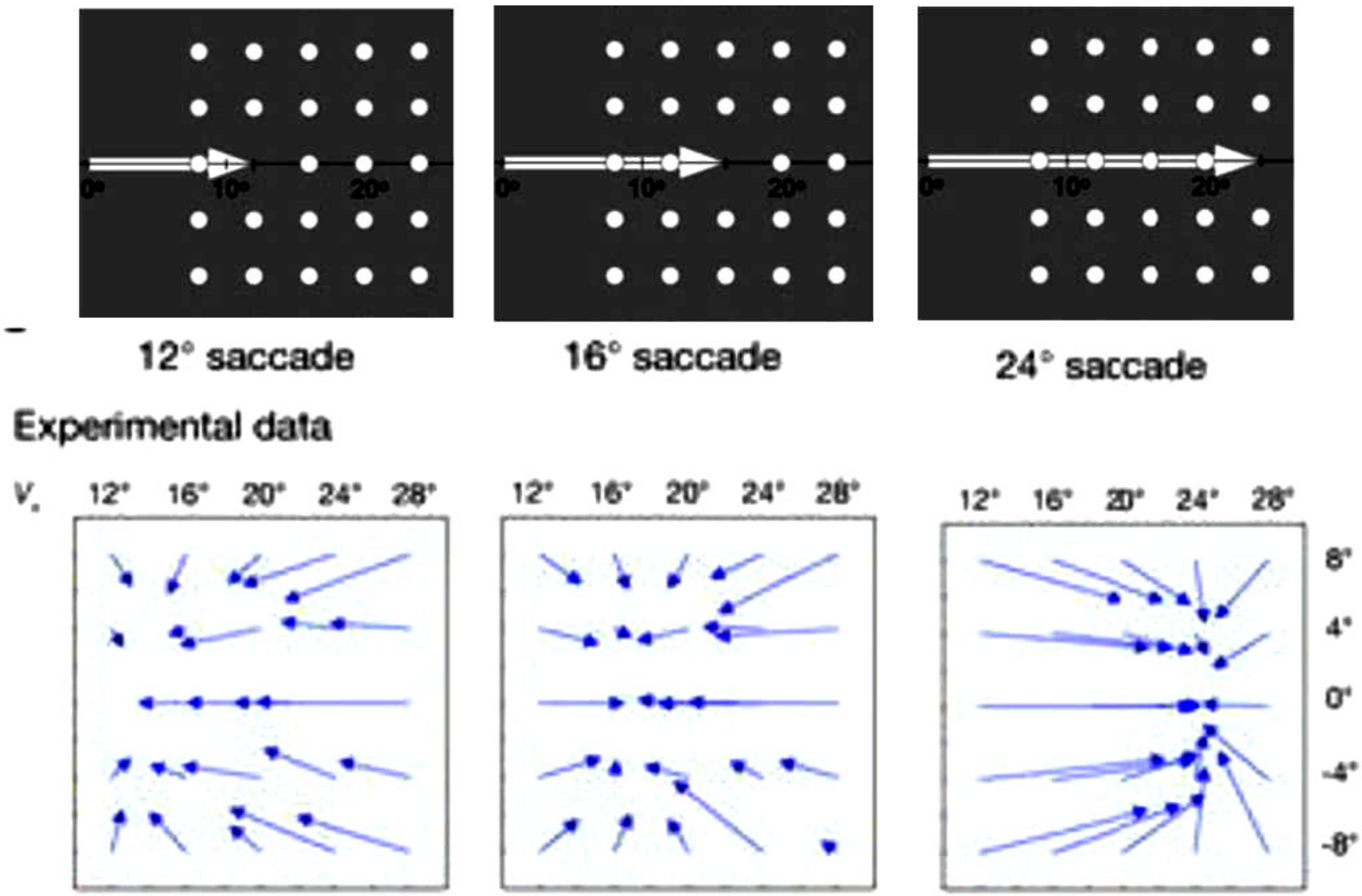
The understanding of a scene is built up in a sequence of visual information acquisition when a target is attended (***fixation***) followed by a fast eyes movement that repositions the fovea on the next target of interest (***saccade***). **It is *called the scanpath***

We make about ***3 saccades per second*** with the eyeball speed up to ***700 deg/sec*** !

Visual Information along the Scanpath

1. ***Saccadic Suppression*** Visual sensitivity is reduced during saccades as we do not see fast moving images when our eyeballs rotate up to 700 deg/sec
2. ***Visual Constancy*** 3 times per second there are huge instant changes in the image projected on the retina, but we are not aware of discontinuities in the scene perception
3. ***Perisaccadic Mislocalization*** However, this visual constancy is not perfect: *Before* the onset of a saccade the perceptual space is *transiently compressed* around the saccade's intended target

Note: All three phenomena are not well understood in visual neuroscience.



Adapted from:

Hamker FH, Zirnsak M, Calow D, Lappe M (2008) The peri-saccadic perception of objects and space. PLoS Comput Biol 4(2): e31.

PFT in Modeling the Scanpath Vision

1. **Fixation (~300 ms)** An image is sampled by retina photoreceptors

$$f(r_i e^{i\theta_l}) = f(e^{u_k} e^{i\theta_l}) = f_{k,l}$$

and its DPFT $\widehat{f}_{m,n}$ is computed by FFT in log-polar coordinates.

The IDPFT, computed again by FFT, gives the image representation

$$\mathrm{f}_{k,l} = \mathrm{f}(u_k, \theta_l) = f(e^{u_k} e^{i\theta_l})$$

in **cortical coordinates** (rectangular **log-polar coordinates** attached at the fovea center) where disparity-sensitive cells contribute to 3-D understanding of the scene.

In addition, during this time **the next saccade's target is selected, its position w.r. to the fovea is computed and converted into a motor command to move the eyes.**

2. <u>Saccade (~30 ms)</u> About 50 ms ***before*** the onset of a saccade, ***during*** the saccade (~30 ms), and about 50 ms ***after*** the saccade, the **visual sensitivity is reduced** and **neural processes transiently shift the cortical image** in the cortical space and transiently **compresses** the perceptual space

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

$$\mathfrak{f}(u_k + j\delta_M, \theta_l) = f(e^{u_k + j\delta_M}e^{i\theta_l}) = f(e^{j\delta_M}r_k e^{i\theta_l})$$

(**perisaccadic mislocalization**).

It brings the presaccadic scene (fovea-centered coordinates**) into the postsaccadic frame** (target-centered coordinates) **which eliminates the need for starting visual information acquisition afresh 3 times per second at each fixation, which would be very costly to do.**

The Binocular Scanpath Vision System

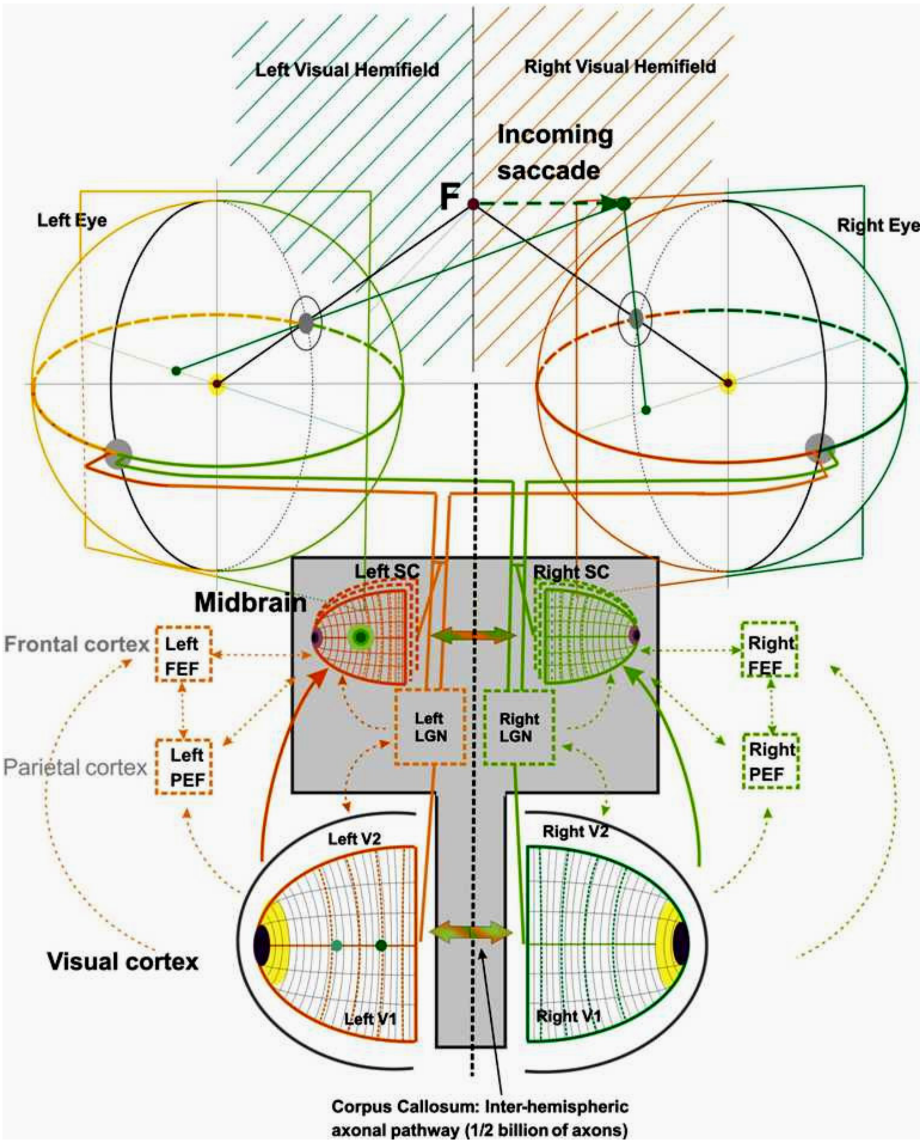
SC = Superior Colliculus

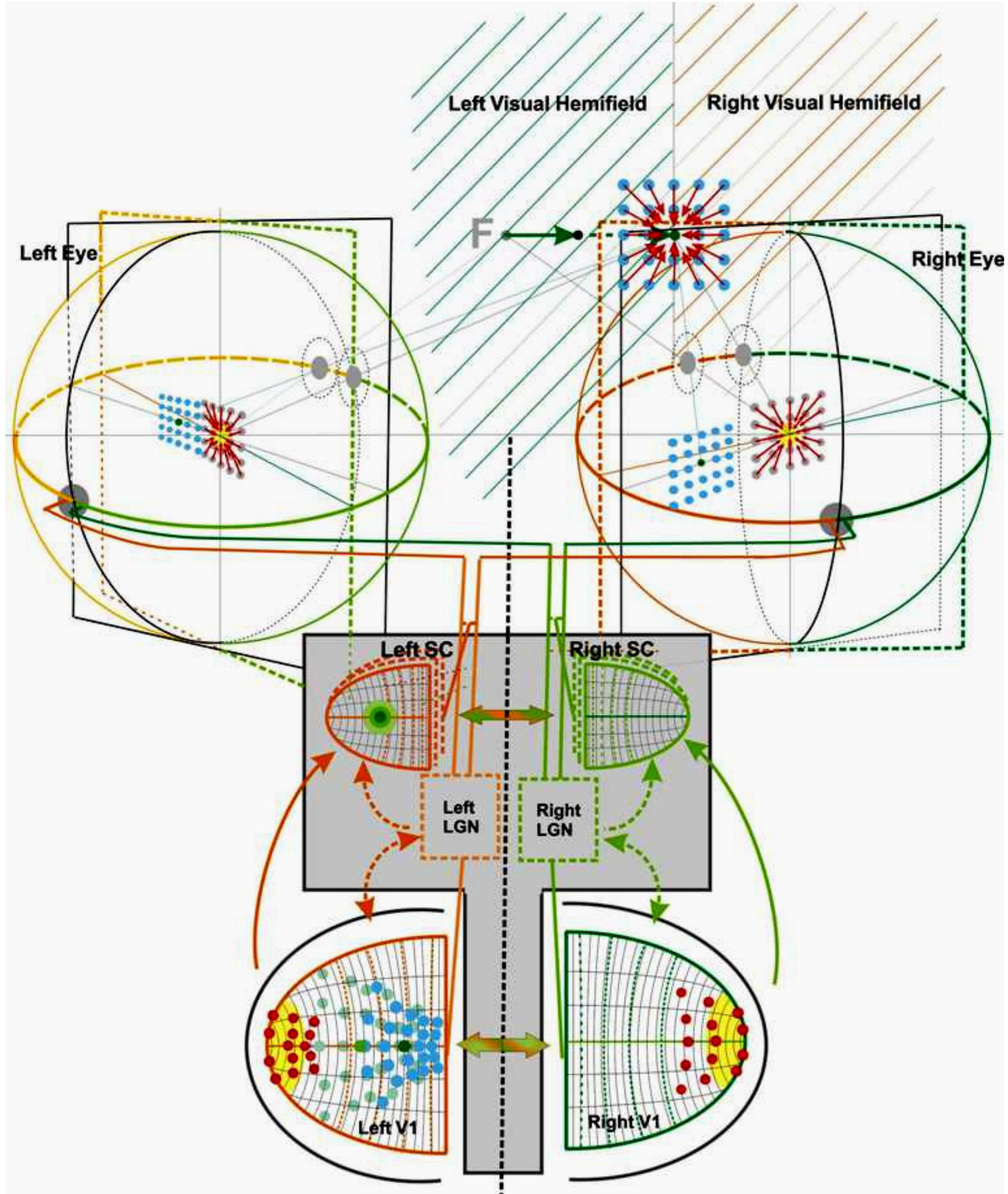
LGN = Lateral Genuculate Nucleus

V1 = Primary Visual Cortex

PFE = Parietal Eye Field

FEF = Frontal Eye Field




Left Visual Hemifield
Right Visual Hemifield
F
Left Eye
Right Eye
Left SC
Right SC
Left LGN
Right LGN
Left V1
Right V1

Conclusion

Projective Fourier analysis of the Conformal Camera provides the *efficient*, *physiologically realistic* image processing framework for the Binocular *Scanpath Vision*

Future Research

1. Develop "true" foveal and peripheral image representations. There is accumulated evidence that the foveal and peripheral image processing are supported by different neural processes.
2. Implement the binocular model of the scanpath vision we have proposed to test the perceptual continuity across fixations.
3. Both space and time undergo perisaccadic compression. It was suggested that it is similar to the contraction in special relativity given by Lorentz transformations. Since the ***SL***$(2,C)$ group is the double cover of the Lorentz group, the conformal camera could model in *a natural way* this **space-time perisaccadic compression**.

References

Mathematics

[1] Turski, J. Geometric Fourier Analysis of the Conformal Camera for Active Vision, *SIAM Review*, **46**, 230-255, 2004.

[2] Turski, J. Geometric Fourier Analysis for Computational Vision, *Journal of Fourier Analysis and Applications,* **11,** 1-23, 2005.

Implementation in Binocular Vision

[3] Turski, J. Computational Harmonic Analysis for Human and Robotic Vision Systems, *Neurocomputing*, **69**, 1277-1280, 2006.

Neuroscience

[4] 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.

Computational Neuroscience

[5] VanRullen, R. A simple translation in cortical log-coordinates may account for the pattern of saccadic localization errors. *Biological Cybernetics*, **91**, 131- 137, 2004.