Computer vision based on physiological aspects of human visual system

Maria Haase

Melanie Strauch

Workshop Selbstorganisation und Komplexität
25.-31.8.2007
Ameland, Niederlande

Human Vision vs. Computer Vision

Why is automated recognition such a difficult problem
while humans perform visual tasks with ease?

Can computer vision benefit from efficient concepts
and optimal strategies of human visual system?

Human visual system

- ❑ is able to extract essential information of a scene in minimal time
- ❑ developed specific strategies for specific visual tasks

Focus on two aspects

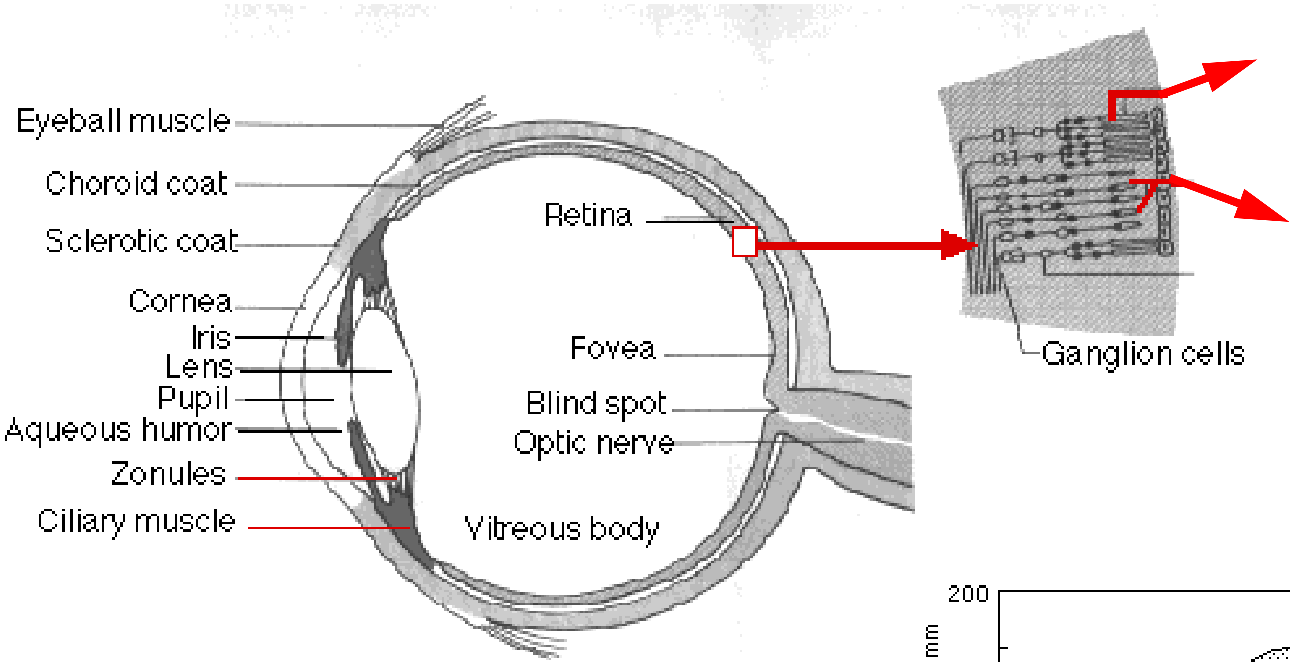
- ❑ Inhomogeneous resolution of photoreceptors on retina
- ❑ Saccadic eye movements

Not considered here:

- ✘ Pattern recognition based on associative memory (Haken, Fuchs, 1987)
- ✘ Synchronization of neural pulses (Haken, 2007)

Inhomogeneous resolution of photoreceptors on retina

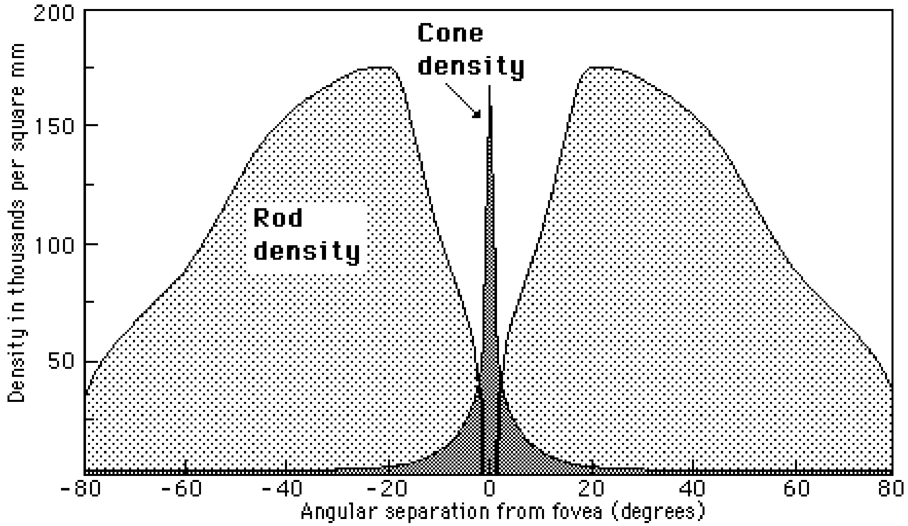
Human eye



Rods: black and white vision, light-sensitive, periphery

Cones: color vision, center of retina

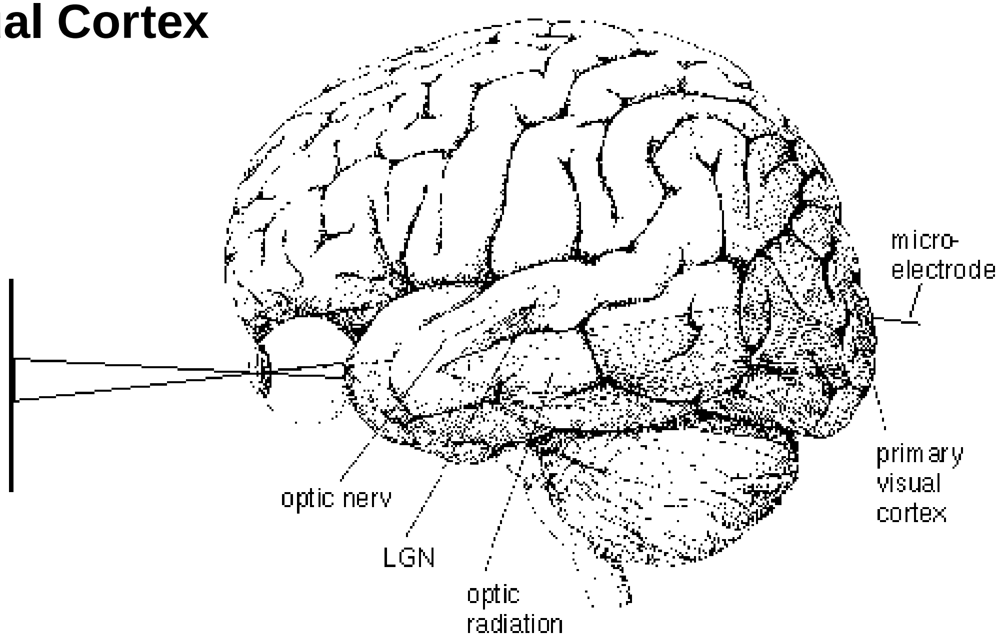
Radial distribution of rods and cones on retina



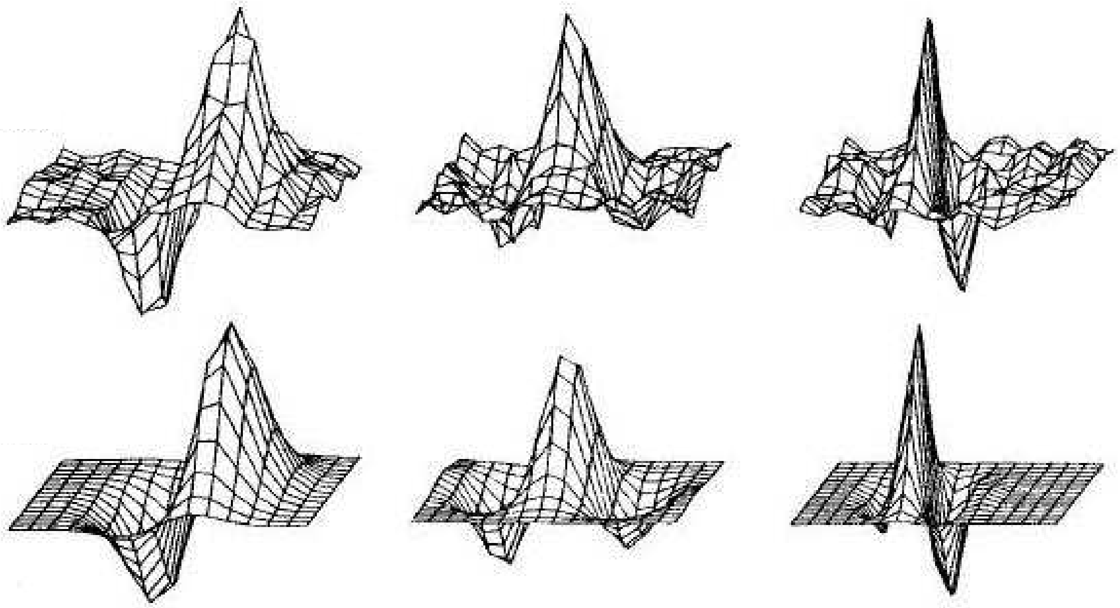
A retina with uniform resolution equal to highest fovea resolution would require ~ **10 000** more photoreceptors

Mapping of Receptive Fields in Visual Cortex

Simple cells in the primary visual cortex have receptive fields (RFs) which are orientation-selective and localized in space and frequency. These cells act as edge detectors.



Hubel & Wiesel, 1977
(cats and monkeys)



2D receptive field profiles

(measured for cats by
Jones & Palmer, 1987)

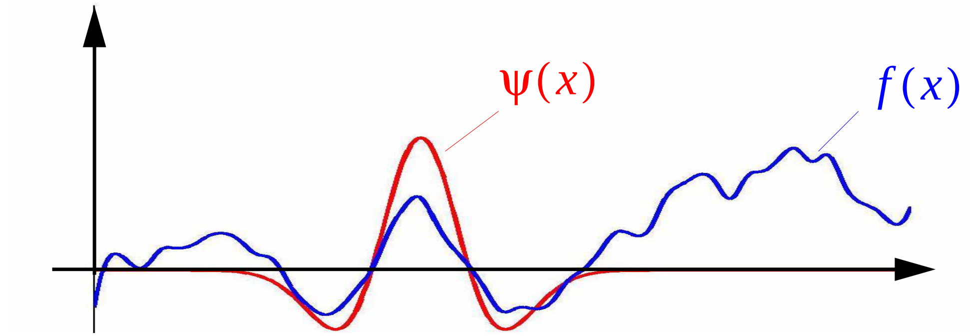
2D Gabor wavelets
(Daugman 1980)

Continuous Wavelet Transform

Decomposition of $f(x)$ into elementary space-scale components :

$$Wf(a,b) = \frac{1}{a}\int_{-\infty}^{\infty} f(x)\,\overline{\psi}\left(\frac{x-b}{a}\right) \mathrm{d}x \quad (a, b \in \boldsymbol{R}, a > 0)$$

highly redundant



Reconstruction of $f(x)$:

$$f(x) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{0}^{\infty} Wf(a,b)\,\psi\left(\frac{x-b}{a}\right)\frac{\mathrm{d}a}{a^2}\,\mathrm{d}b \quad \text{where} \quad C_\psi = \int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega}\mathrm{d}\omega < \infty$$

Admissibility condition: $\int_{-\infty}^{\infty} \psi(x)\,\mathrm{d}x = 0$ → name **"Wavelet"**

Discrete Wavelet Transform

Discrete parameters a, b

$$a = a_0^m, \; b = n\, b_0 a_0^m$$

$$(a_0 > 1,\ b_0 > 0,\ m, n \in Z)$$

Wavelet coefficients:

$$Wf_{m,n} = \int_{-\infty}^{\infty} f(x)\, \psi_{m,n}\, \mathrm{d}x = \langle f, \psi_{m,n} \rangle$$

$$\psi_{m,n}(x) = a_0^{-m/2}\, \psi(a_0^{-m} x - n b_0)$$

Frames:

redundant discrete systems

Orthonormal bases of wavelets:

not redundant, not shift invariant

→ Multiresolution analysis

→ Fast Wavelet Transform

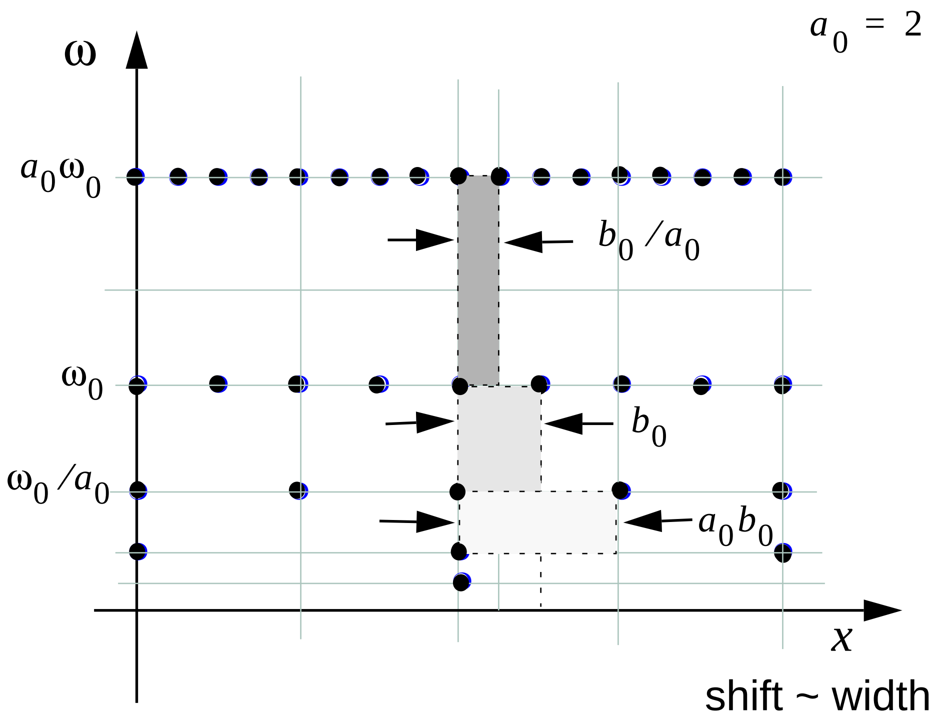
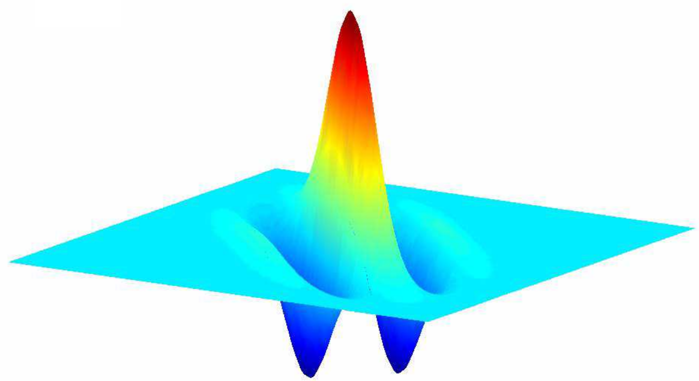


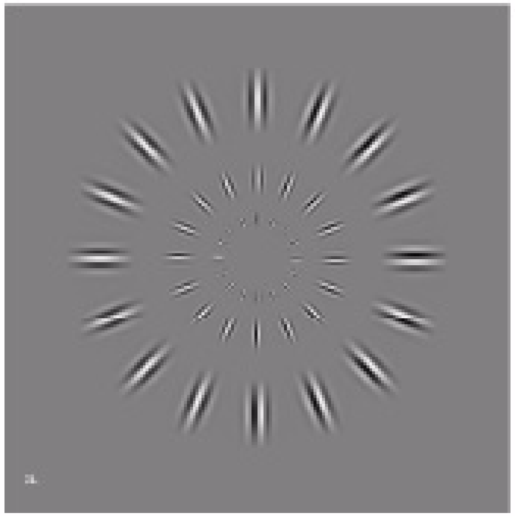
Image analysis with 2D Gabor Wavelet Frames

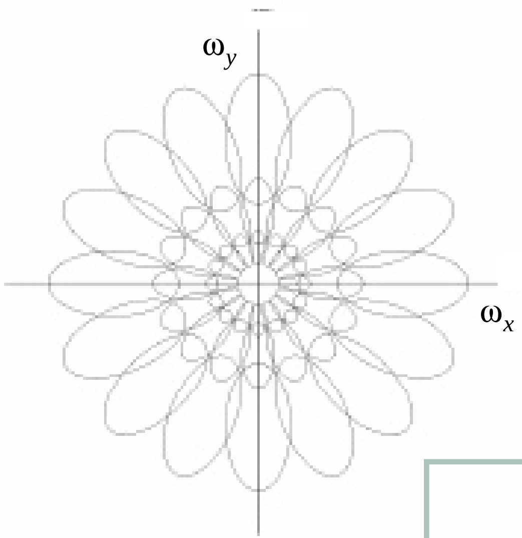
Orientation-sensitive Gabor Wavelets can be used as simple models for information processing in the primary visual cortex.

The increase of density of photoreceptor cone cells corresponds to an increase of visual acuity as daylight stimuli come closer to the fovea. Eye movements (saccades) direct the fovea to image regions of interest.



Symmetric Gabor wavelet
(real part)





Ensemble of Gabor Wavelets
in frequency plane

Basic wavelet

$$\psi(x, y) = e^{-\left[\left(\frac{x - b_x}{a_x}\right)^2 + \left(\frac{y - b_y}{a_y}\right)^2\right]} e^{i(\xi_0 x + \eta_0 y)}$$

T.S.Lee: IEEE Trans. Patt.Anal. Mach.Intel.18 (1996)

Frames:

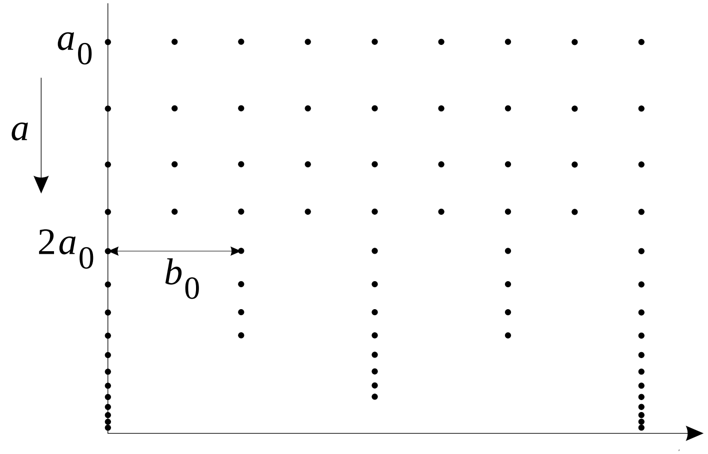
Unique and stable reconstruction of $f(x)$ if

$$A\|f\|^2 \le \sum_{m,n} |\langle f, \psi_{m,n}\rangle|^2 \le B\|f\|^2$$

$A = B = 1\ldots$ orthonormal basis

$B/A = 1\ldots$ tight frame (easy reconstruction)

$B/A \approx 1\ldots$ snug frame (suboctave scales)



Frame with 4 voices

Rotated Gabor Wavelet:

$$\psi(x', y') = \exp\left[-\left(\frac{x'-b_x}{a_x}\right)^2 - \left(\frac{y'-b_y}{a_y}\right)^2\right] \exp[i(\xi_0 x' + \eta_0 y')]$$

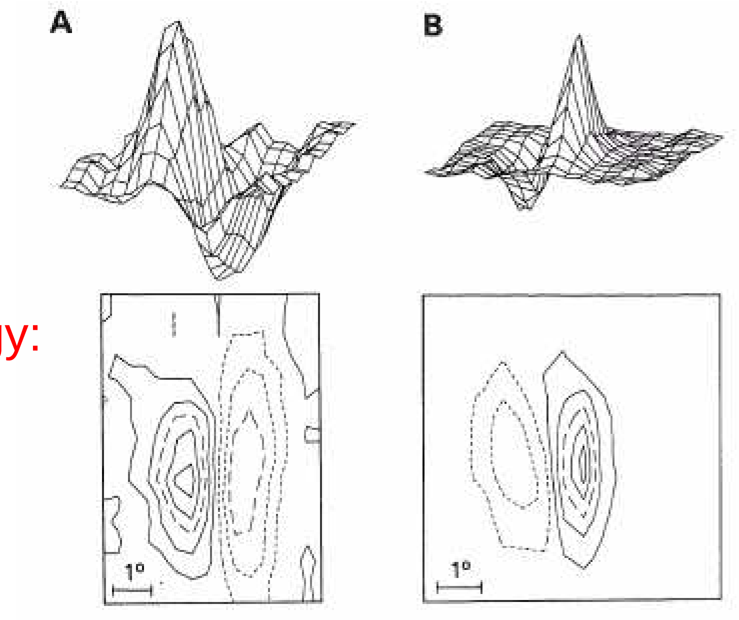
with $\boldsymbol{x}' = \boldsymbol{R}(\Theta)\boldsymbol{y}'$

Parameter reduction according to neurophysiology:

$a = a_x = 2a_y,\ \ \eta_0 = 0,\ \ \xi_0 \sim 2.5/a$

12 discrete orientations

3 voices, $B/A = 1.079$ --> snug frame



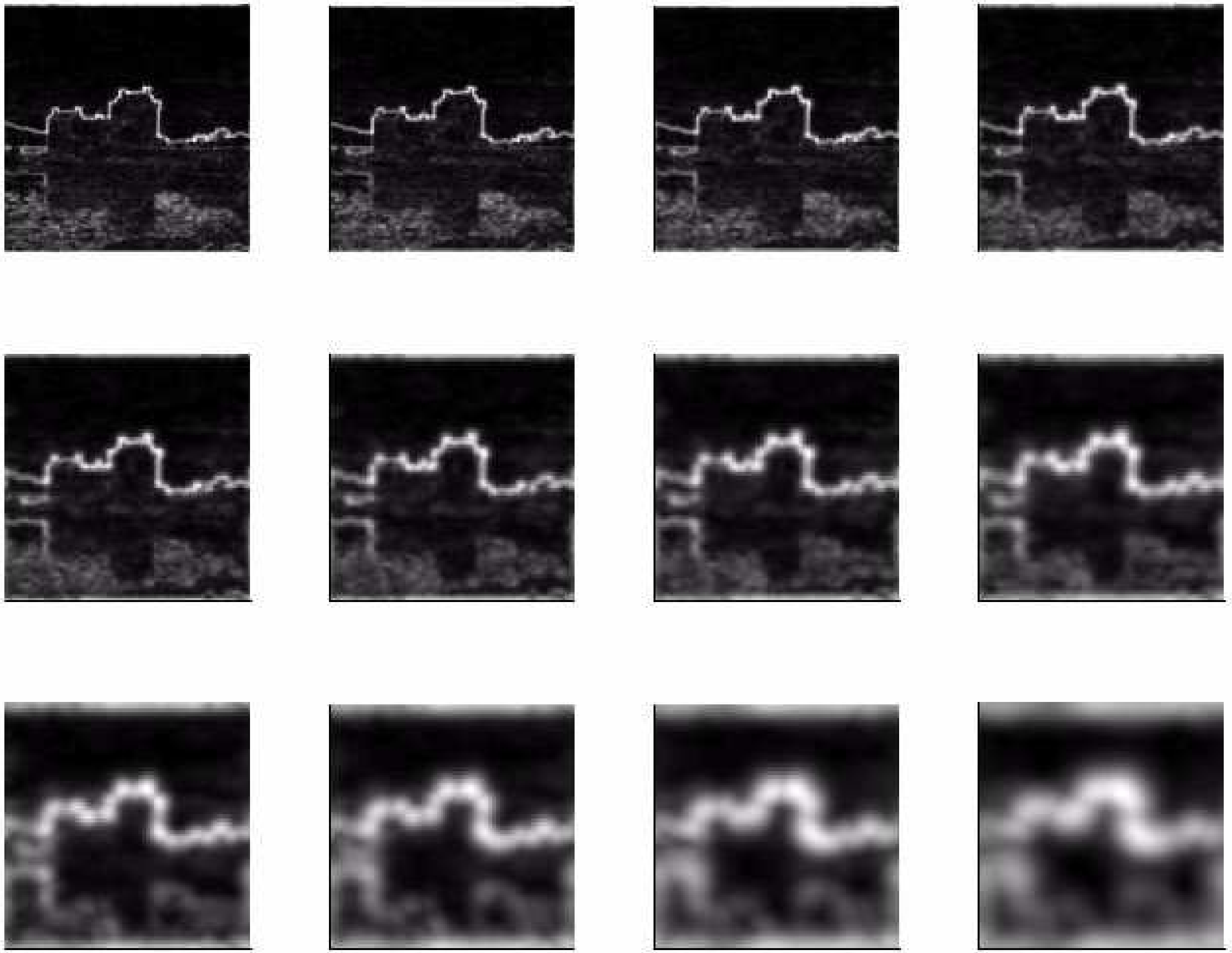
two receptive field measurements

Jones & Palmer, 1987

Image Processing with Gabor wavelets:

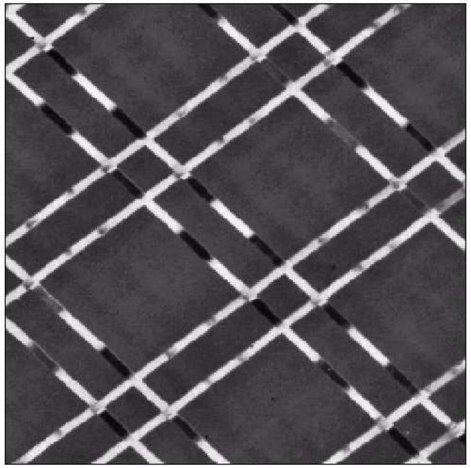
Original

a) Edge detection



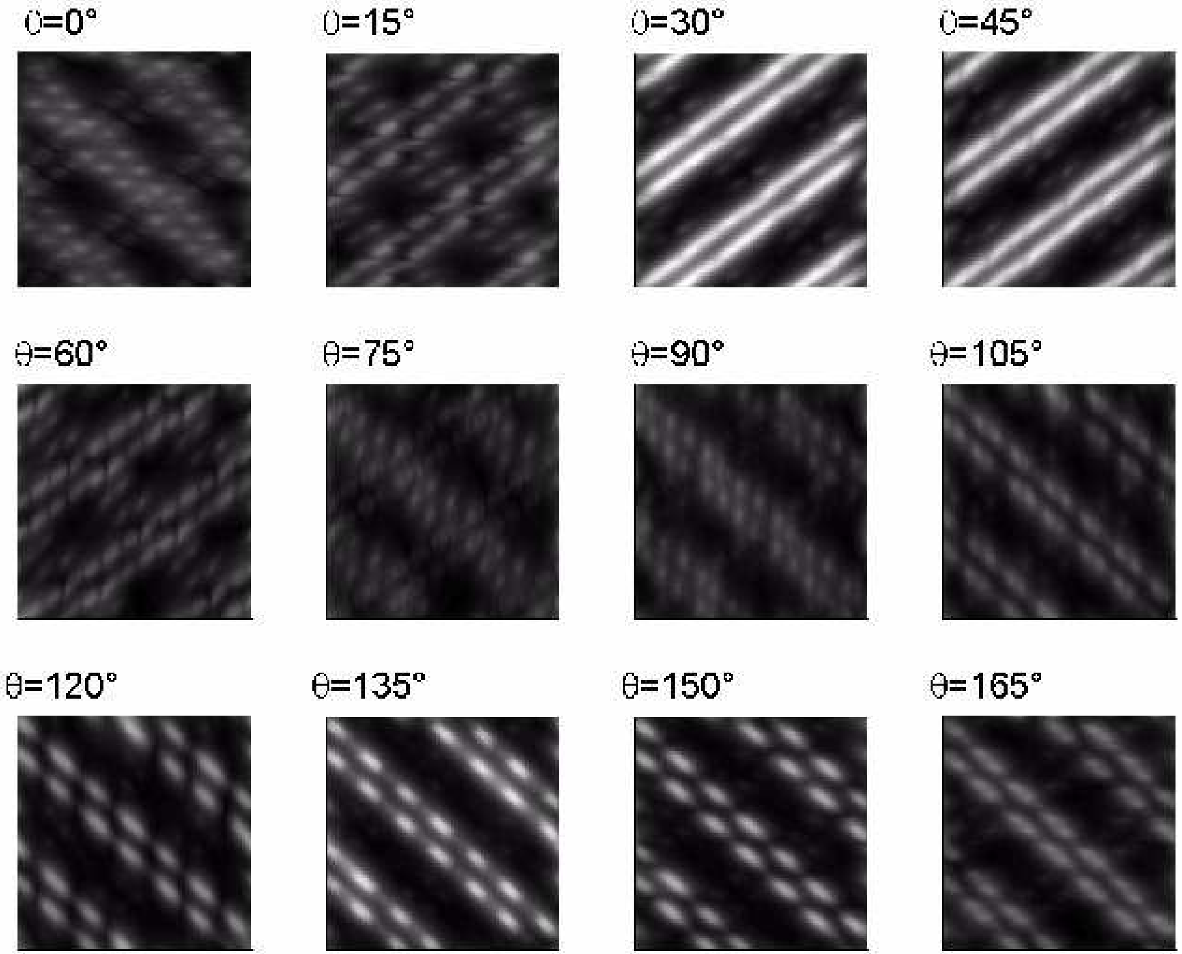
Wavelet coefficients $\left| Wf_m(b_x, b_y) \right|$ for different scales (average over directions)

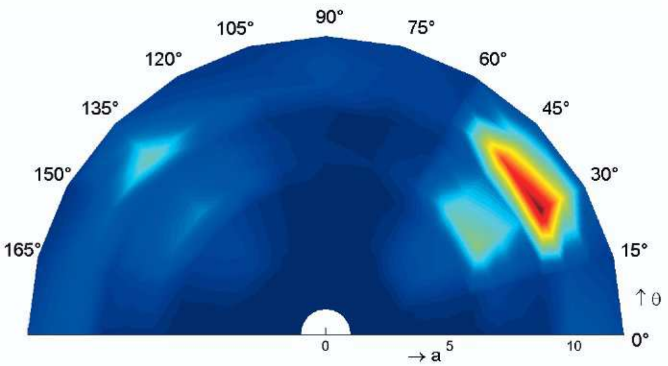
Image Processing with Gabor wavelets:



Original

b) Orientation sensitivity



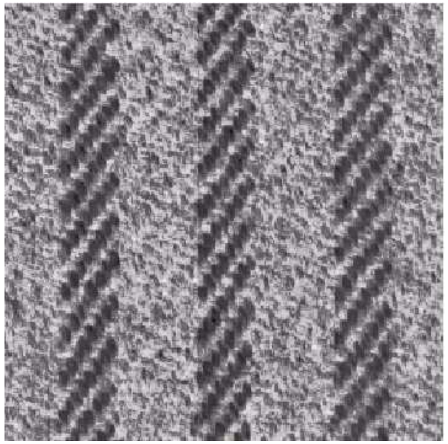


Variance of wavelet coefficients

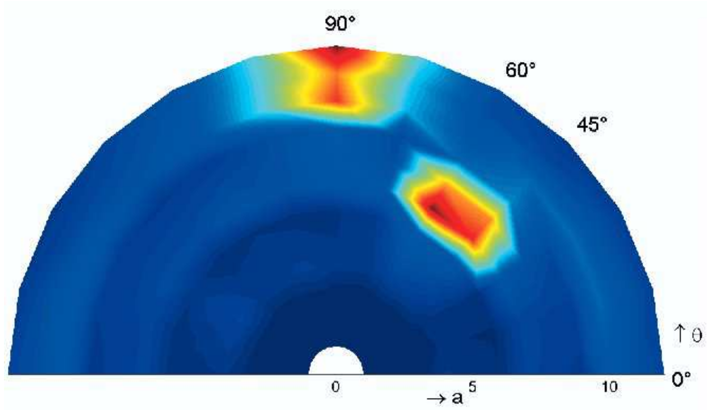
$\left|Wf_l(b_x, b_y)\right|$ for different orientations (average over scales)

Texture analysis

Field of attention



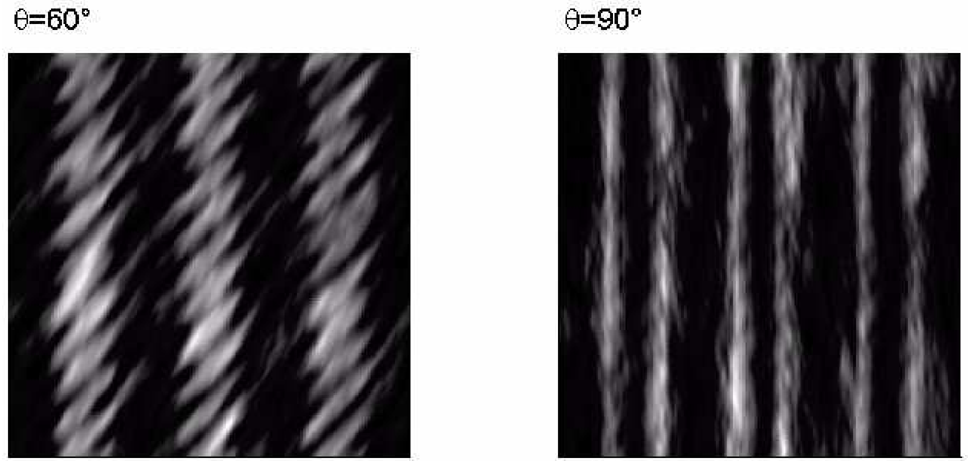
Original



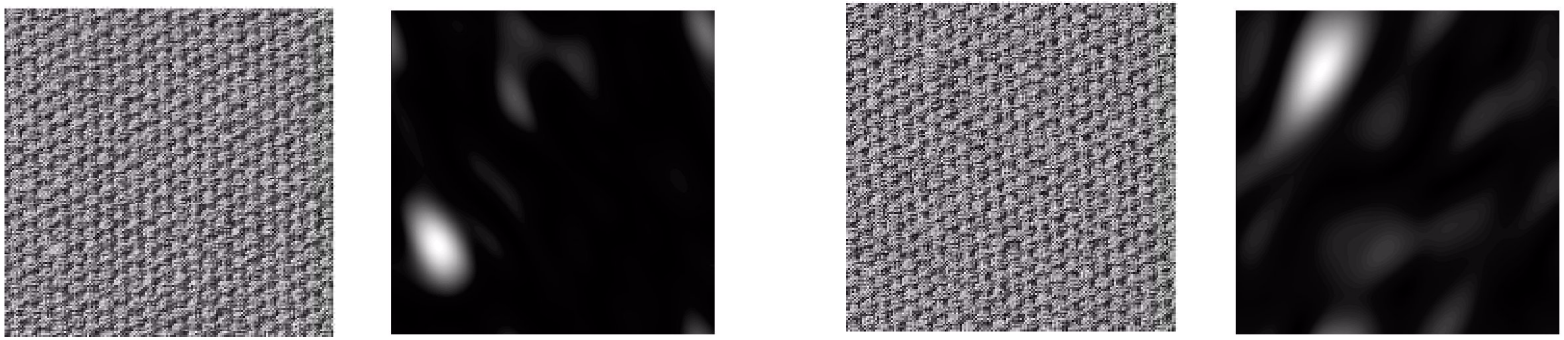
Variance of wavelet coefficients
(average over directions)

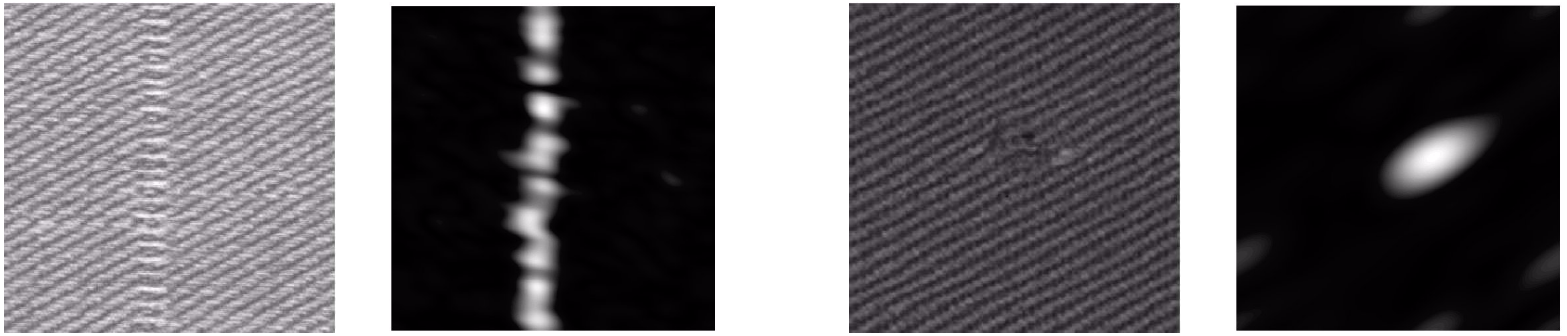
$\left| Wf_l(b_x, b_y) \right|$ for two orientations

(average over scales)

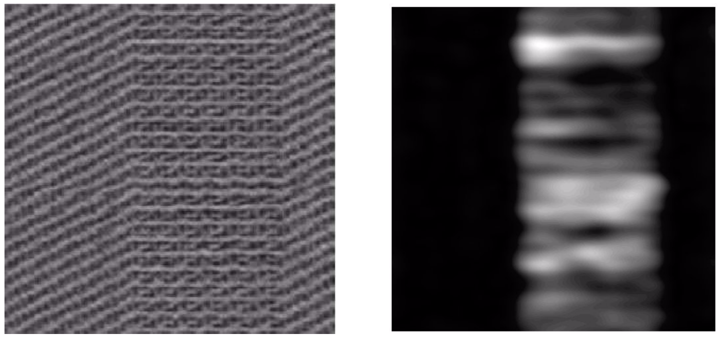


Fault detection (Quality control)





Original tissue



Wavelet coefficients for maximal variance

$$\left| Wf_{l_{\max}, m_{\max}} \right|$$

How much information can transmitted by neurons : ~ 3-4 bits?

Can image with 8 bit resolution be represented
by wavelet coefficients with 4 bit resolution ?

4 bits 3 bits 2 bits

Reconstruction of image ‘Lena’ with quantized wavelet coeffcients

Due to redundancy a tight wavelet frame allows for a robust reconstruction

T.S.Lee: IEEE Trans. Patt.Anal. Mach.Intel.18 (1996)

Active Vision Strategy:

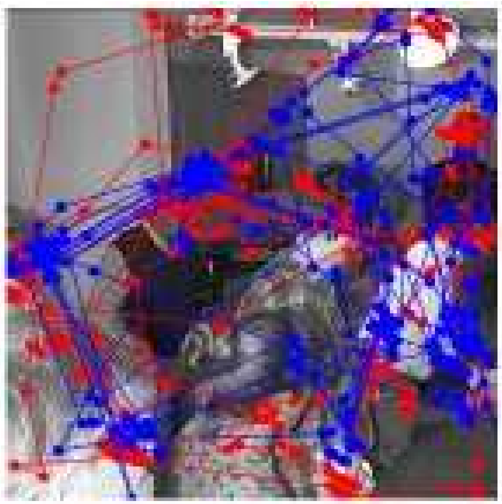
Saccadic eye movements

Human Scanpaths

Succession of saccadic eye movements
compensate loss of visual acuity at periphery of retina,
regions of interest are actively directed to fovea

realizations of a stochastic process

natural scene



2 typical scanpaths of the same person

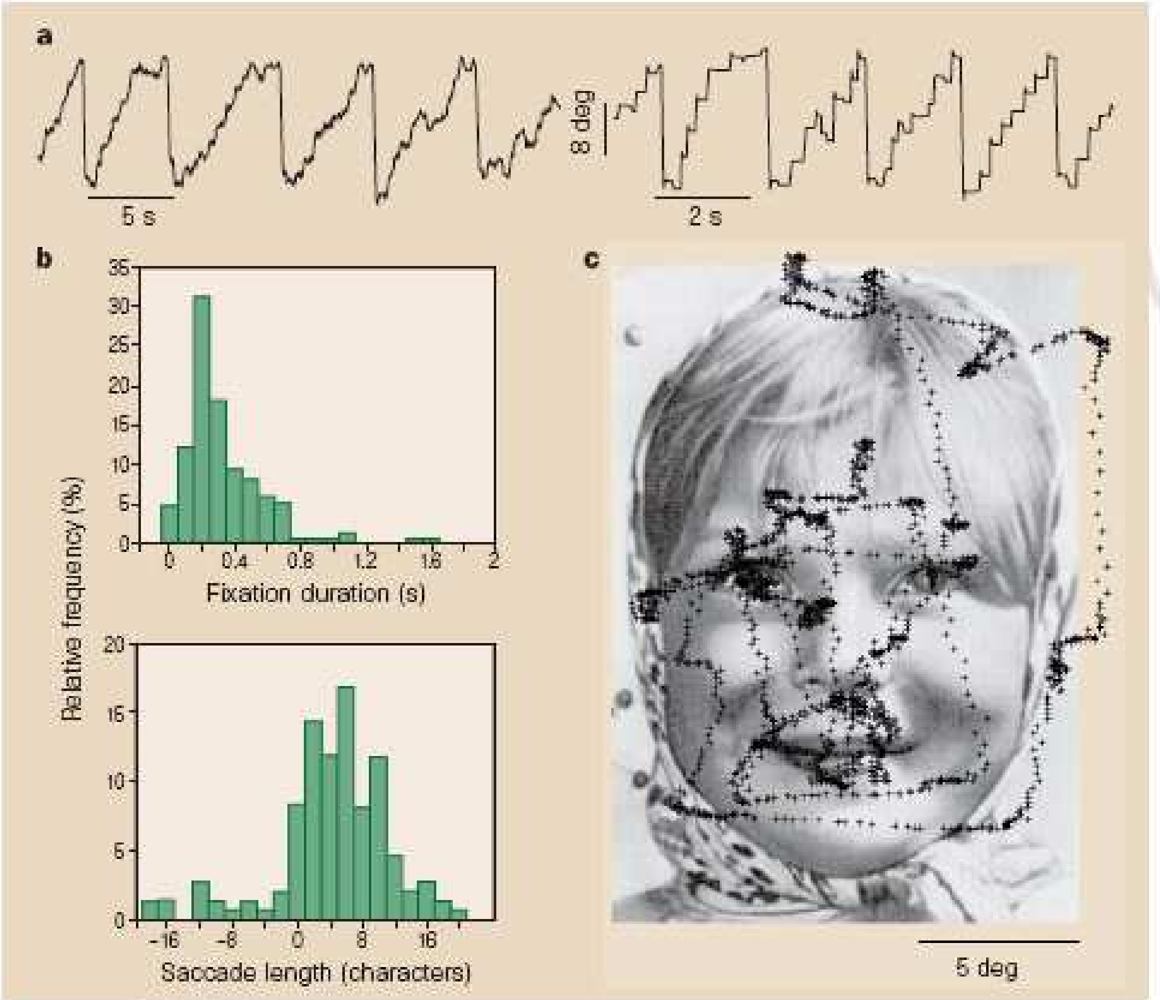
Compensation of oculomotoric deficit by rapid head rotations

Recordings of head (left) and eye (right) movements during text reading

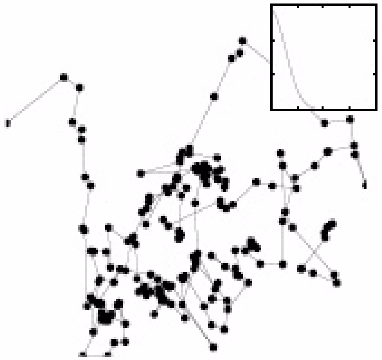
horizontal head / eye displacements

distribution of waiting times

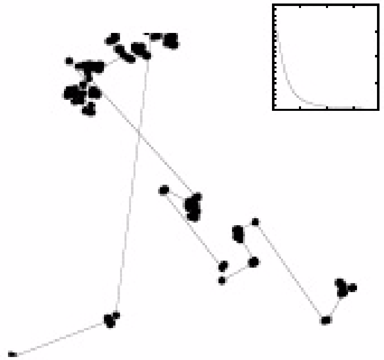
distribution of jump lengths



Findlay et al. Nature 1997



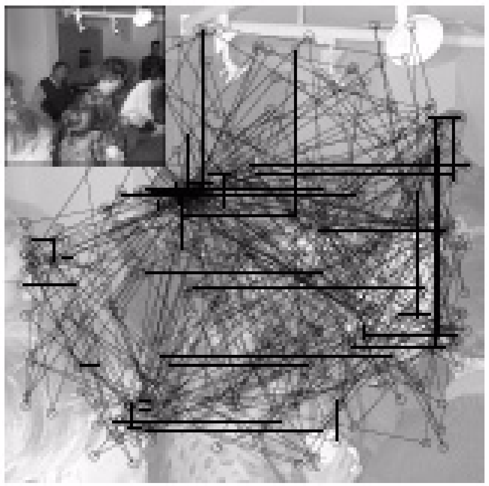
Normal diffusion: $\alpha = 2$



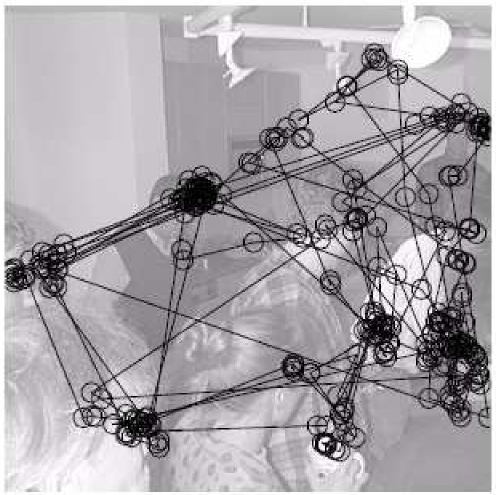
Scale-free Lévy flight: $\alpha = 1$

Growth of typical length

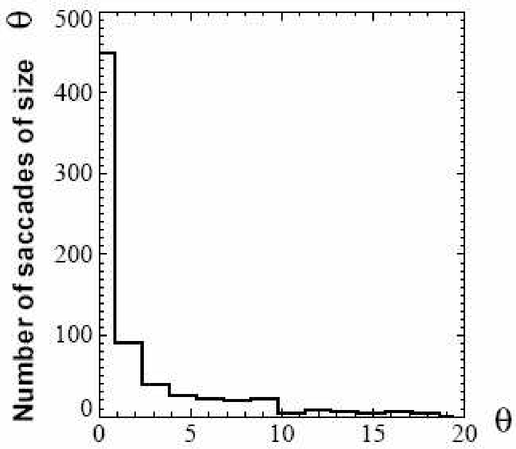
$$\langle |x| \rangle \sim t^{1/\alpha}, \quad \alpha \in (0, 2]$$



Finite variance random walk



Lévy flight



Saccadic magnitudes

T. Geisel, D. Brockmann (1997)

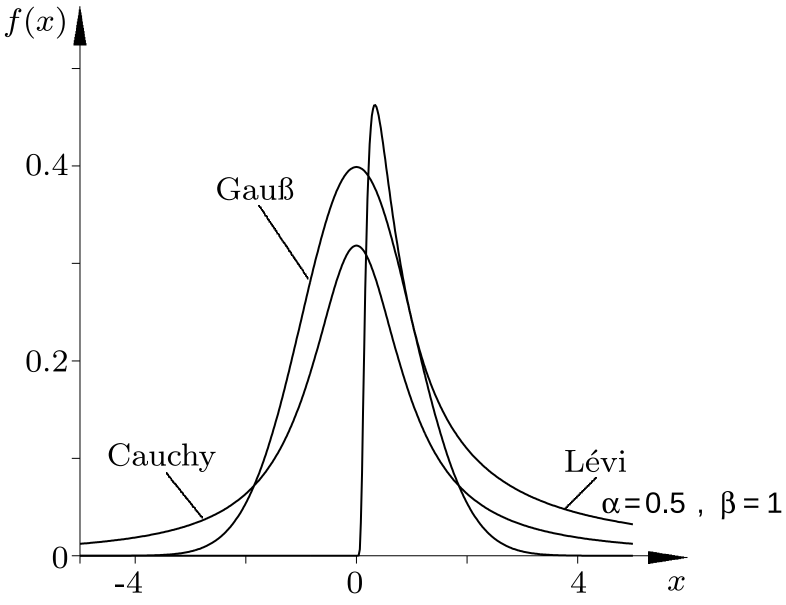
Lévy stable distributions

... only possible limit distributions for sums of normalized and centered i.i.d. random variables

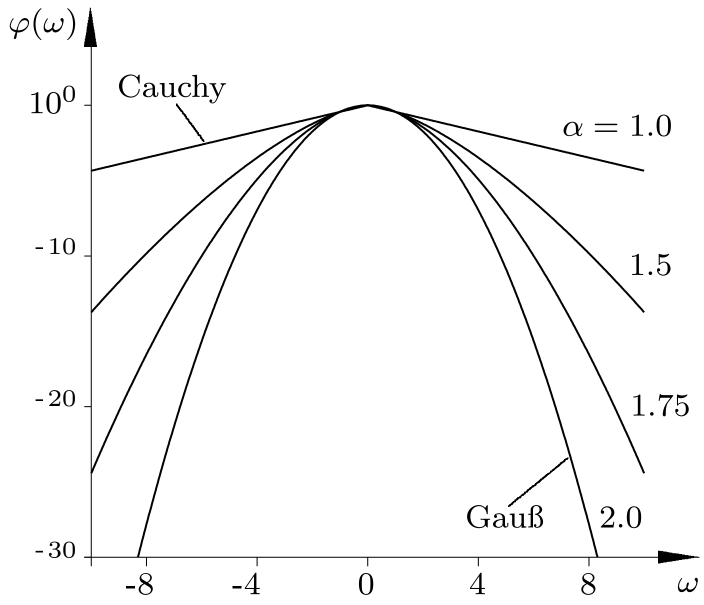
$$\varphi(\omega) = \exp[-|\omega|^{\alpha}(1 - i\beta\,\mathrm{sgn}(\omega)\phi)] \quad \text{where} \quad \begin{cases} \phi = \tan\frac{\pi\alpha}{2}, & \alpha \neq 1 \\ \phi = -\frac{2}{\pi}\log\omega, & \alpha = 1 \end{cases}$$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(\omega) e^{i\omega x} d\omega$$

Tail approximation: $f(x) \propto (1+\beta)x^{-(1+\alpha)}$



Probability density function



Characteristic function

Phenomenological model for human scanpaths

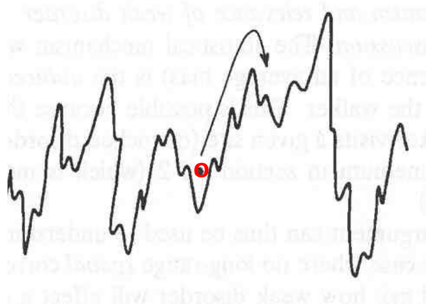
T. Geisel, D. Brockmann (1997)

Observations:

Scanpaths are realisations of a stochastic process in a random quenched saliency field

Fixations accumulate in visually salient regions

Short saccades are more frequent than long ones



Basic assumptions:

Saccade is random walk in quenched field of attention

Visual system minimizes typical time needed to process a visual scene

Model:

$$p(x,t) = \int p(x|y)\,p(y,t-1)\,dy$$

pdf for fixating a location x at disrete time t in a visual field of size L

$$p(x|y) \propto s(x)\,f(|x-y|)\ldots$$

transition pdf for a gaze shift from x to y

probability of generating a saccade of magnitude $|x-y|$

quenched saliency field

Distribution of saccade magnitudes: Cauchy or Gauss ?

Assumption: visual system attempts to reach stationary state as fast as possible

Two choices for $f|x-y|$

Gauss (normal diffusion): $f_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$

Cauchy (super-diffusive): $f_C = \frac{C}{\pi(C^2+x^2)}$

Calibration: both systems need the same relaxation time for constant $s(x) \rightarrow C = \frac{\pi\sigma^2}{L}$

Ensemble of natural scenes: $|\hat{s}(k)|^2 \propto k^{-2+\eta},\ \mu \ll 1$

Result: Cauchy system (Lévy-stable process) needs less time for relaxation to stationary state

T. Geisel, D. Brockmann (1997)

Efficient random search strategies in biological systems

What is the best strategy for maximizing encounter rate between searcher and target?

Biological systems: predator - prey, foraging,
mating partner, flower-pollinator,
parasite infection

Lévy flights are observed in many animal foraging movements like
wandering albatrosses, bees, deer,....

Best strategy depends on

the distribution of targets, whether they are fixed or moving,
differences of searcher and targets in size and velocity,
destructive or nondestructive search

Increasing target to searcher size ratio: Brownian search strategy better

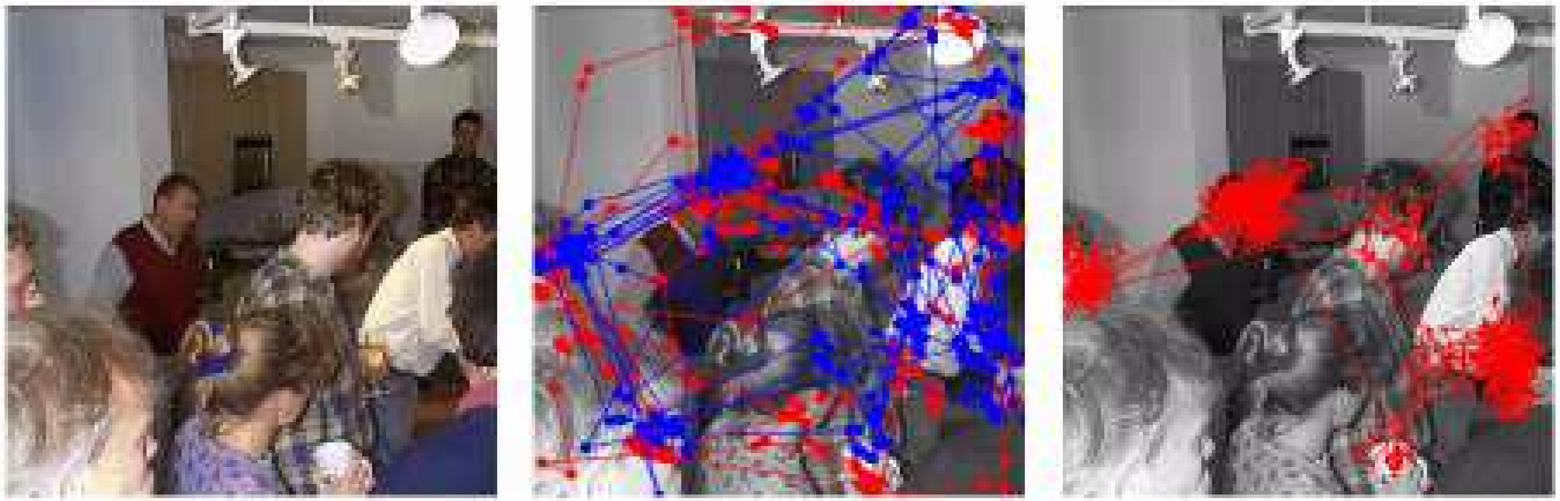
If target sites are sparse and can be revisited: Lévy flight index $\alpha_{opt} = 1$

can be visited only once: $\alpha_{opt} \to 0$

Viswanathan et al., Nature 401 (1999), da Luz et al. Physica A 295 (2001), Bartumeus et al. PRL 88 (2002)

Towards computer vision adopting concepts of human visual system

Human visual system	Computer vision - image processing
Natural scene	**Image**
Activation of photoreceptors on retina: overlapping multi-resolution receptive fields are mapped to primary visual cortex (synchronization of spike trains)	2D orientation-sensitive redundant multi-scale Gabor wavelet frames can be used to model of receptive fields
Brain dynamics, associative memory: selection of regions of attention according to specific visual task	Weighted (?) wavelet energy landscape $A(\boldsymbol{x})\left\|Wf_{l,m}(\boldsymbol{x})\right\|$ serves as saliency field
Oculomotor system actively controls eye movements (saccades): scanpaths are Lévy flights	Scan of the scene: Lévy walk in inhomogeneous quenched “wavelet energy field”



a) scene b) 2 scanpaths c) simulation: Lévy flight

T. Geisel, D. Brockmann (1997)