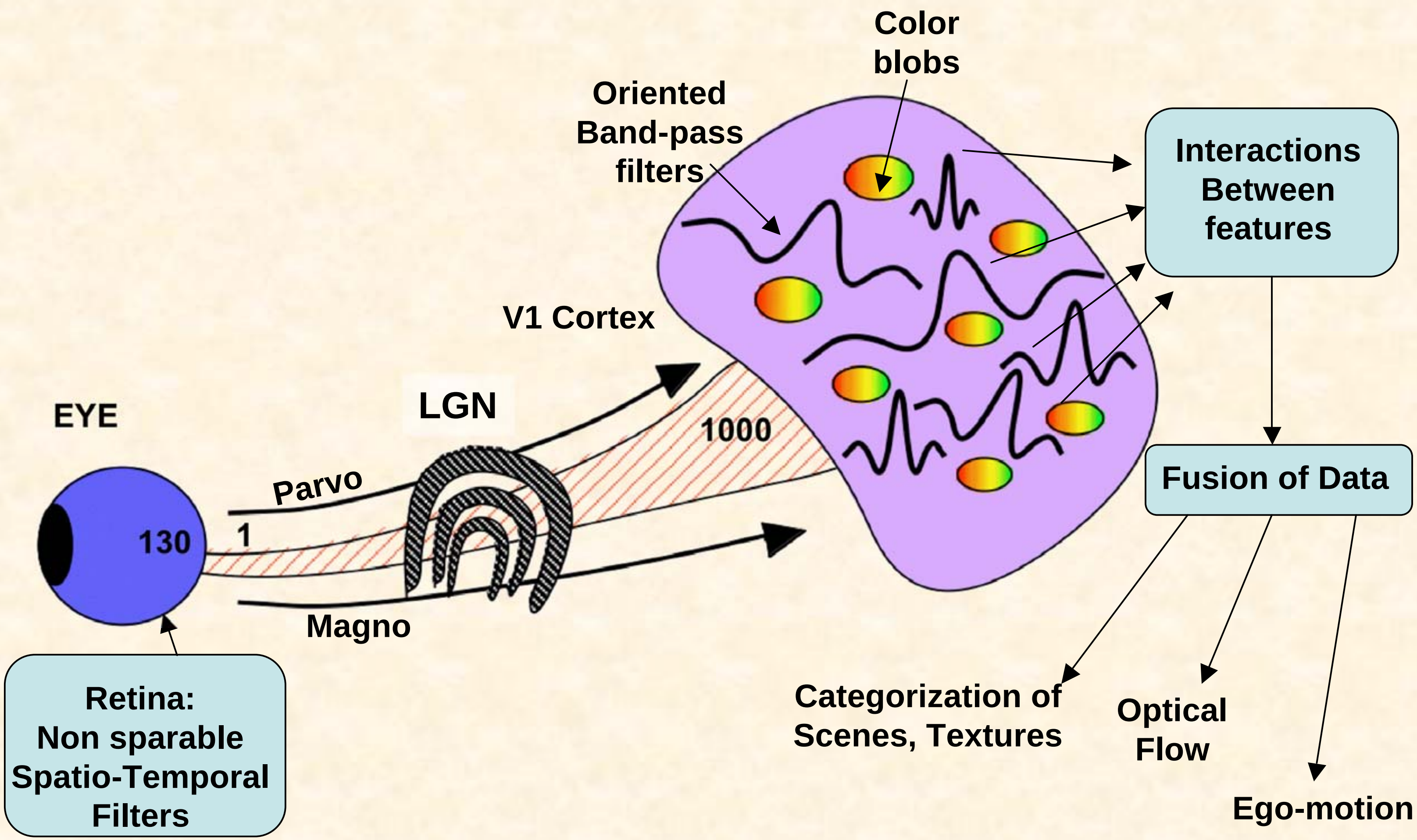
The Visual System: Reducing Variability to Better Categorize

Jeanny Herault

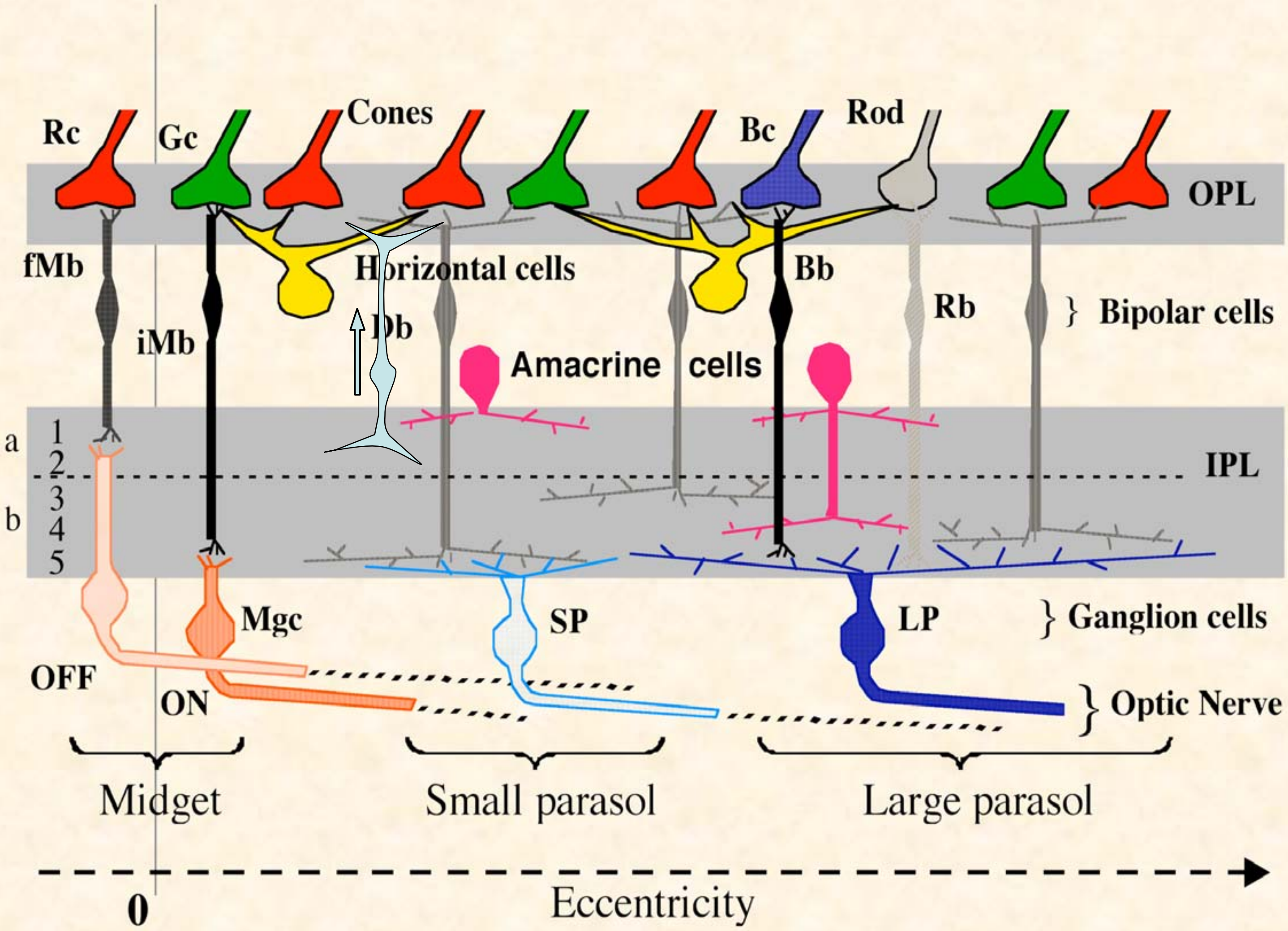
GIPSA-lab, Dept. Images & Signal,
Perception team, Grenoble

STATIM - Paris - January 2009

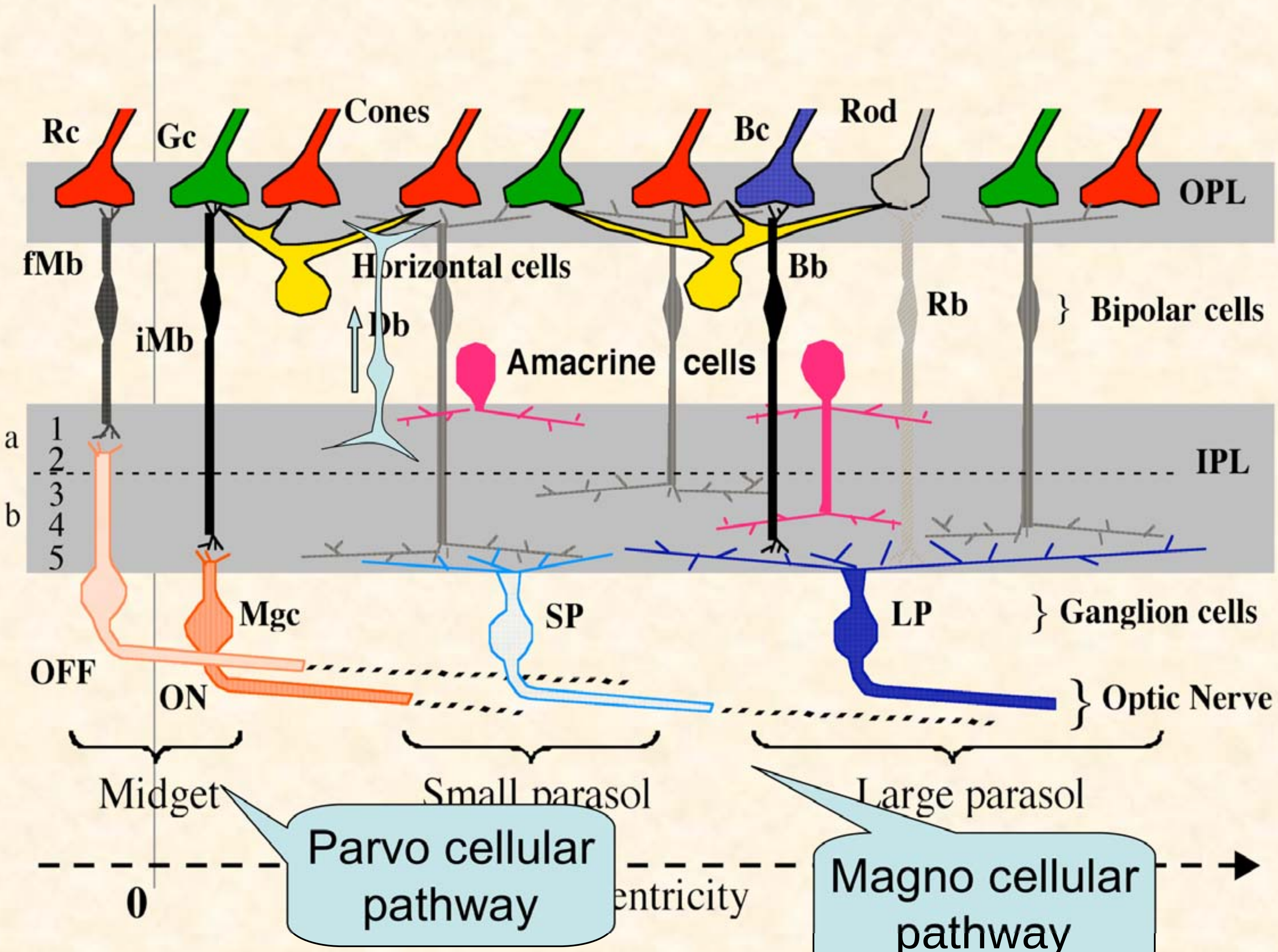
Visual System: Anatomy

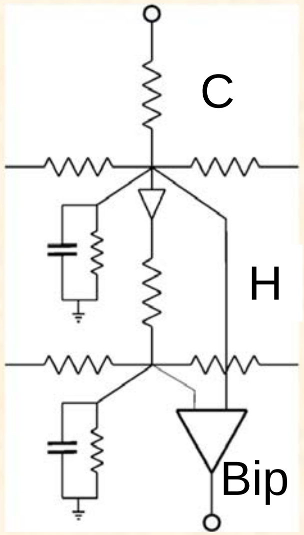


Retina: Linear Model (basics)



Retina: Linear Model (basics)



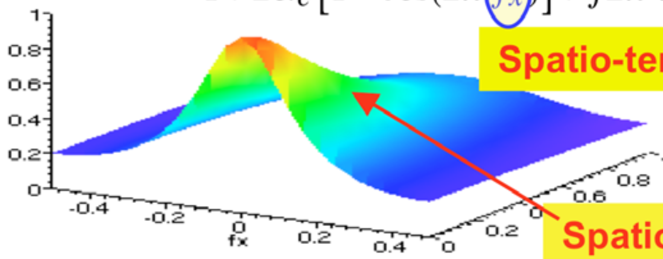


Retina: Parvo Cellular Pathway

Transfer function (properties)

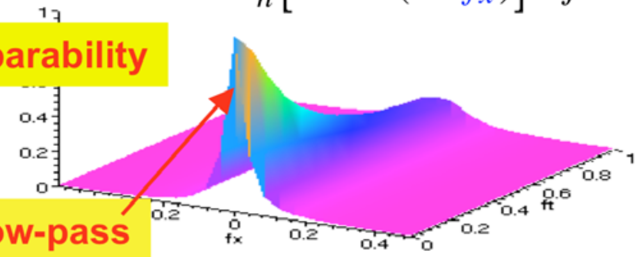
Photoreceptors

$$G_c(f_x,f_t)=\frac{1}{1+2\alpha_c\left[1-\cos(2\pi f_x)\right]+j2\pi\tau f_t}$$



Horizontal cells

$$G_h(f_x,f_t)=\frac{1}{1+2\alpha_h\left[1-\cos(2\pi f_x)\right]+j2\pi\tau f_t}$$



Spatio-temporal inseparability

Spatio-temporal low-pass

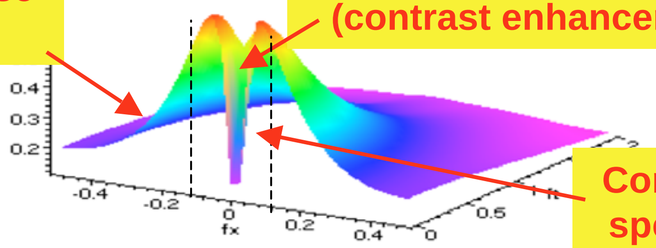
Bipolar cells :

\+ -

$$G_{OPL}(f_x,f_t)=G_c(f_x,f_t)\left[1-G_h(f_x,f_t)\right]$$

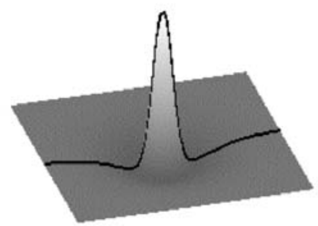
Structural Noise reduction

Spatio-temporal <u>high-pass</u> (contrast enhancement)



Compensates the 1/f spectrum of images (spectral whithening)

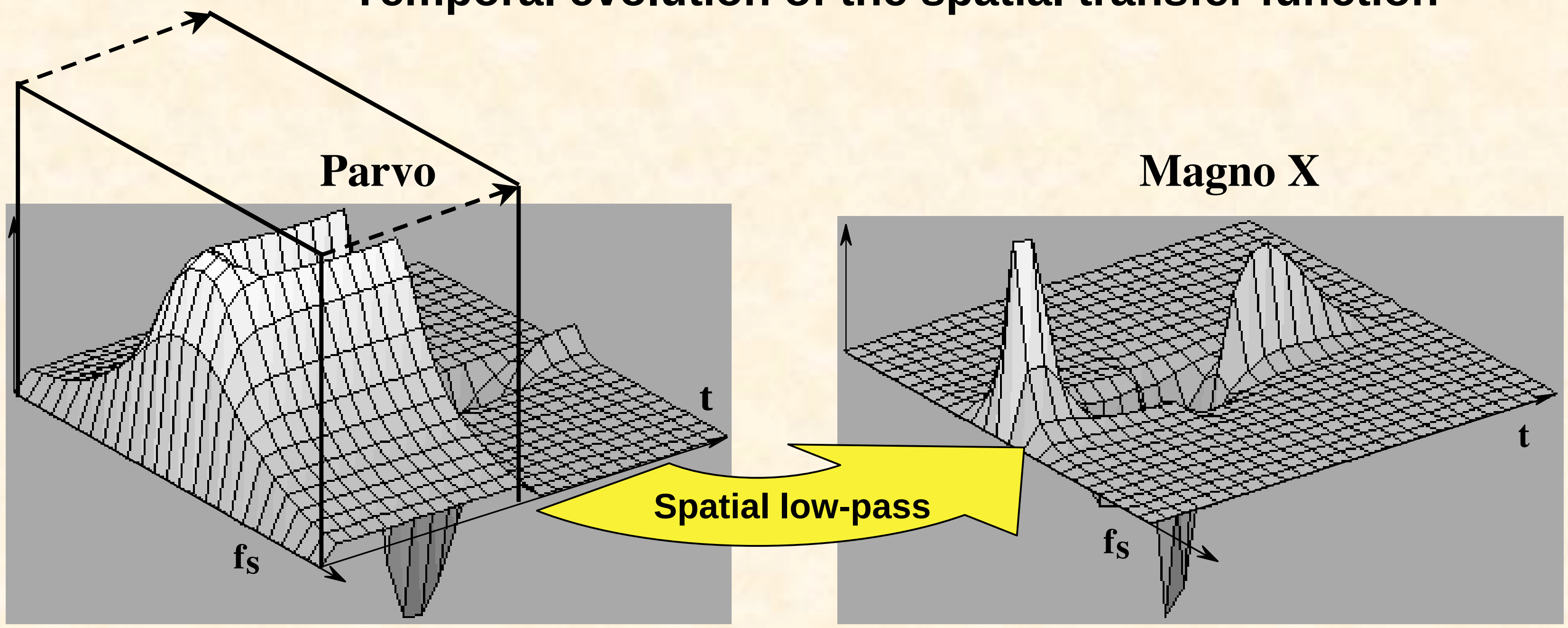
Receptive field



Beaudot Ph-D thesis (1994)

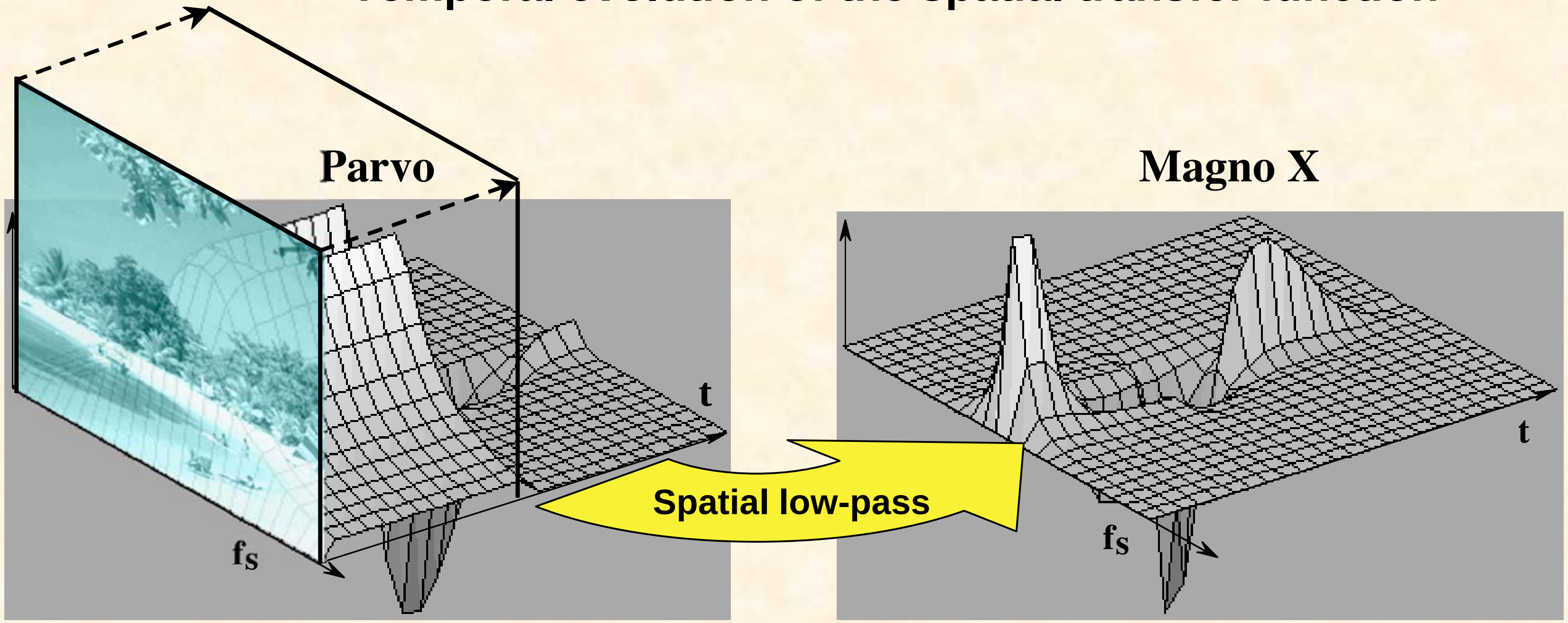
Parvo- & Magnocellular Pathways

- **Temporal evolution of the spatial transfer function**



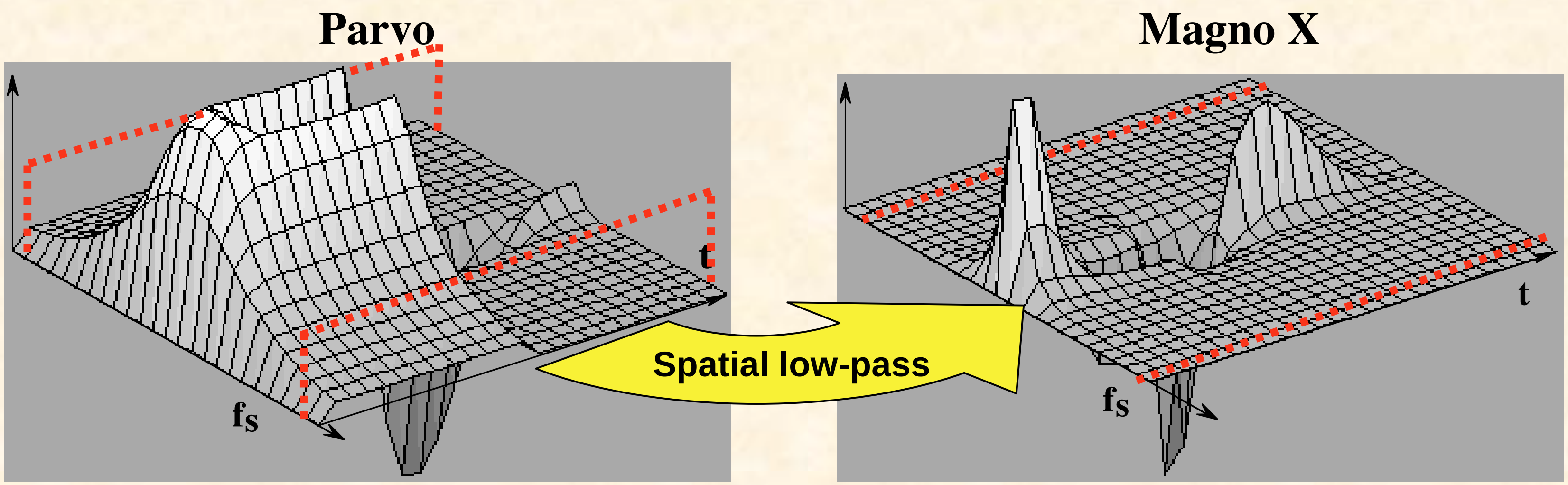
Parvo- & Magnocellular Pathways

- **Temporal evolution of the spatial transfer function**

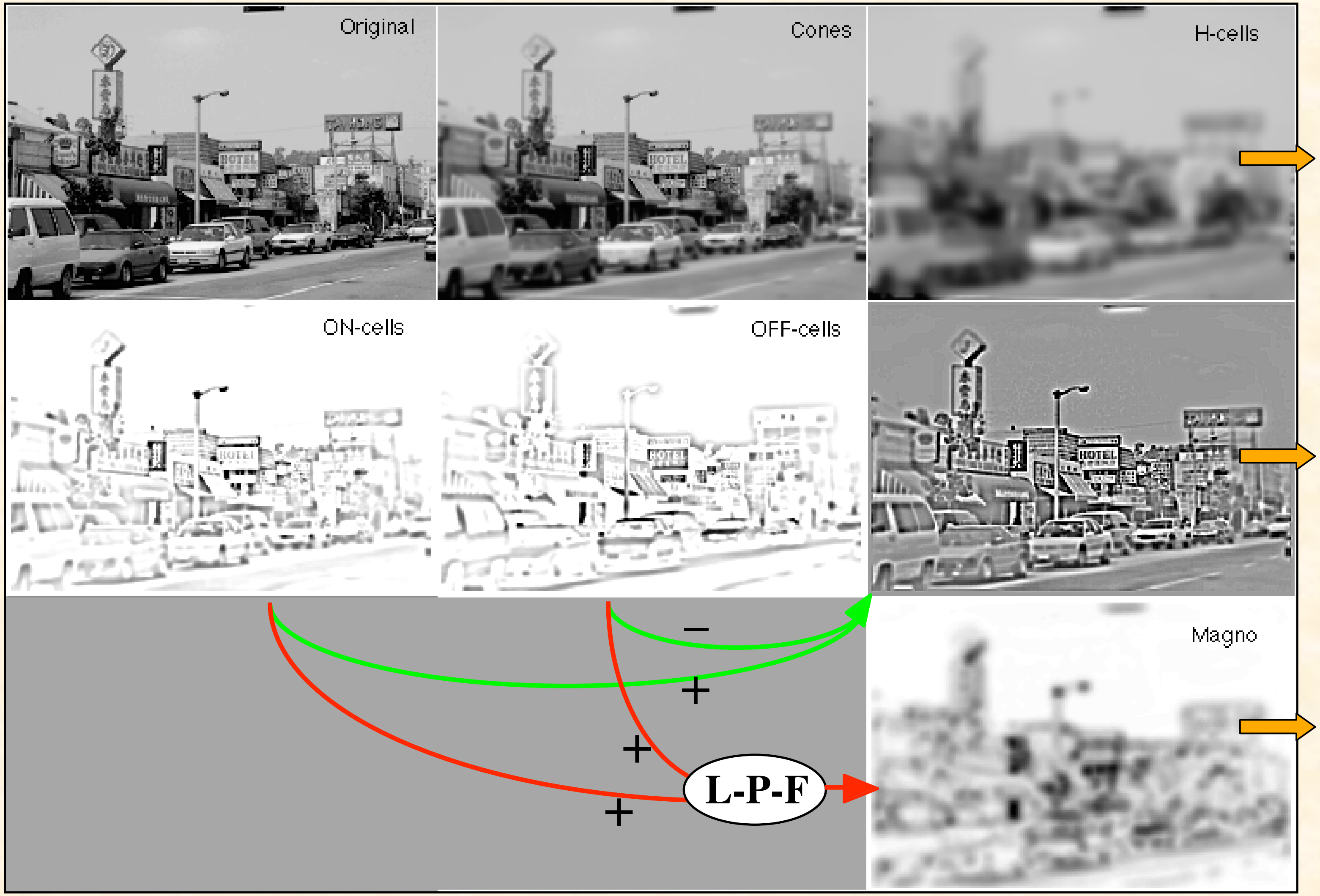


Parvo- & Magnocellular Pathways

- **Temporal evolution of the spatial transfer function**



Retina: Summary



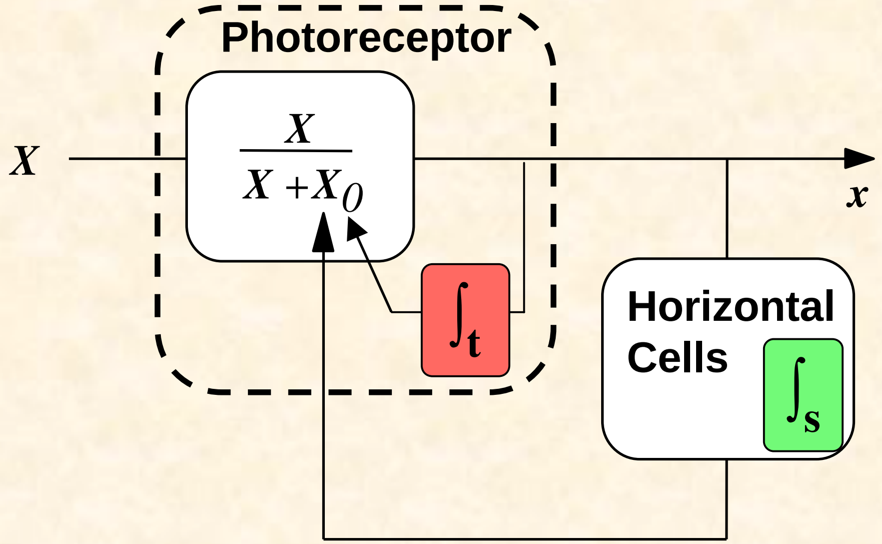
Equiv: «Local Mean»

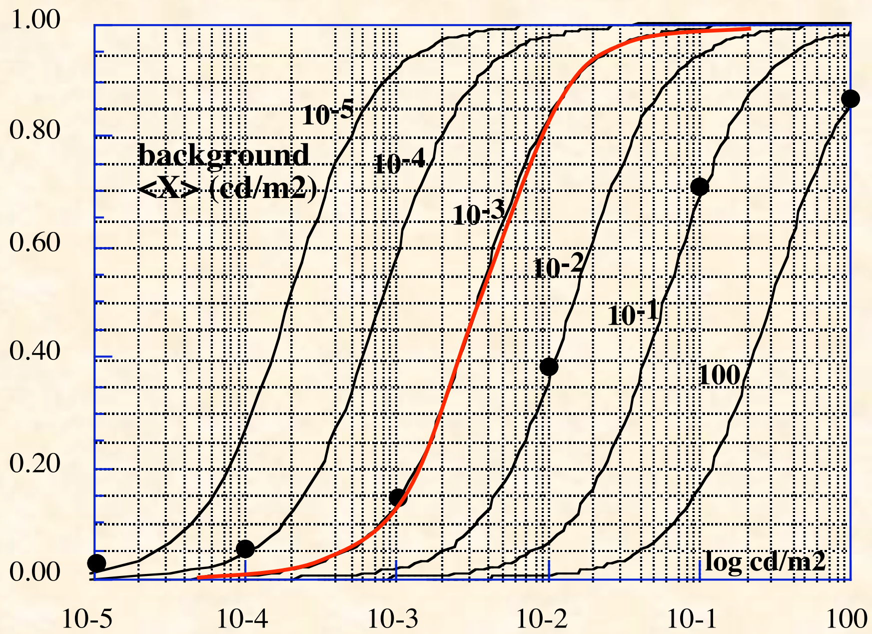
Equiv: «Local Centering»

Equiv: «Local Variance»

Retina: Non-Linearities

1. Photoreceptors' adaptive compression

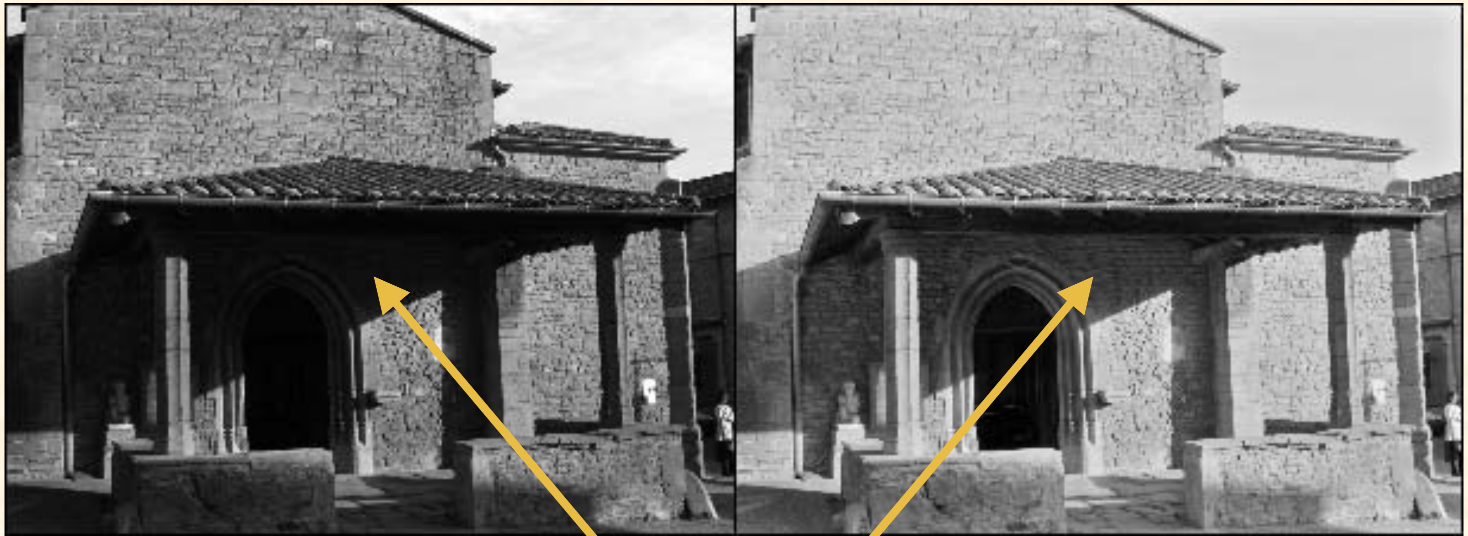




=> X_0 = Temporal and spatial Mean

Retina: Non-Linearities

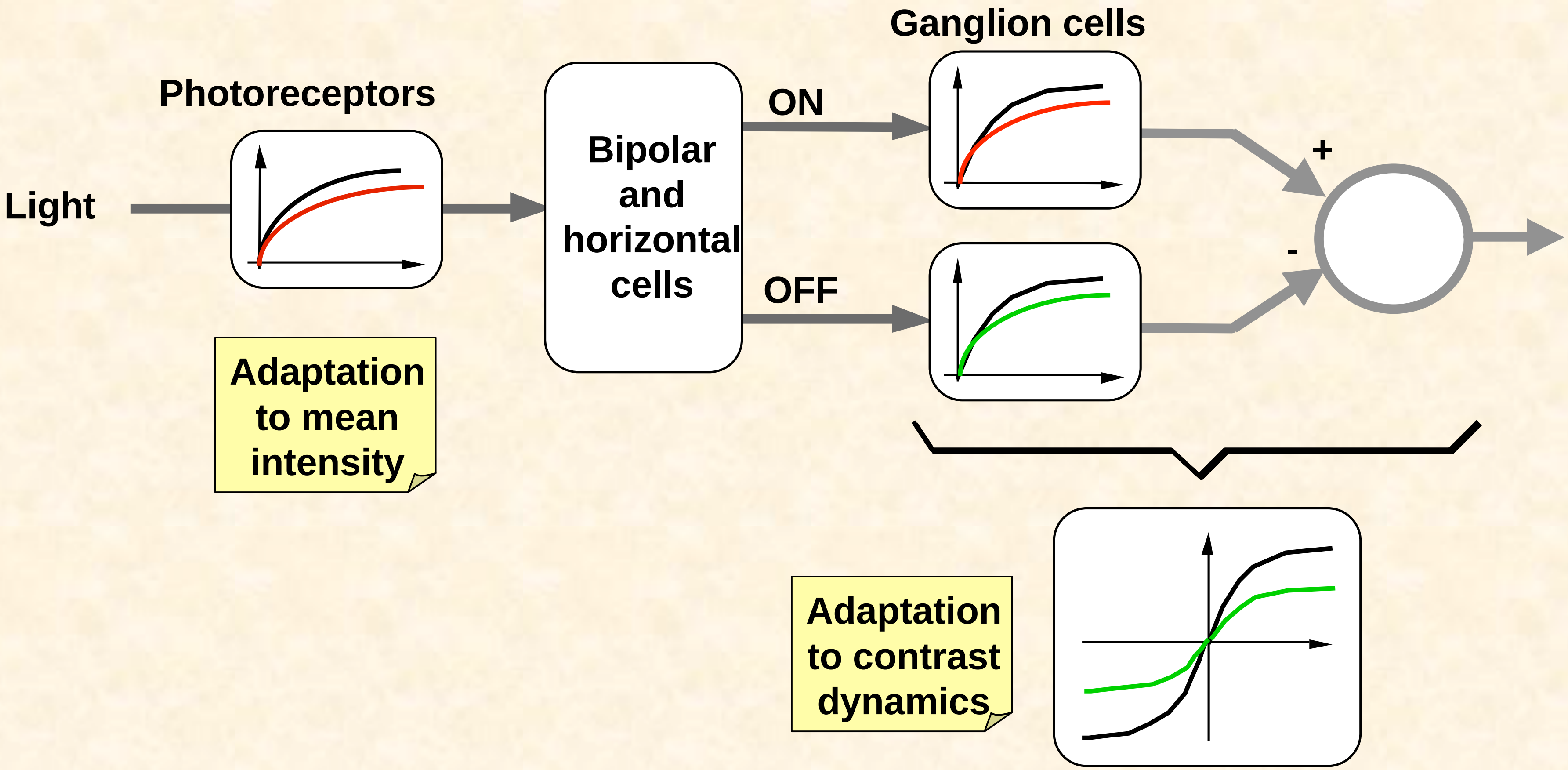
1. Photoreceptors' adaptive compression: application



More details in Shadows

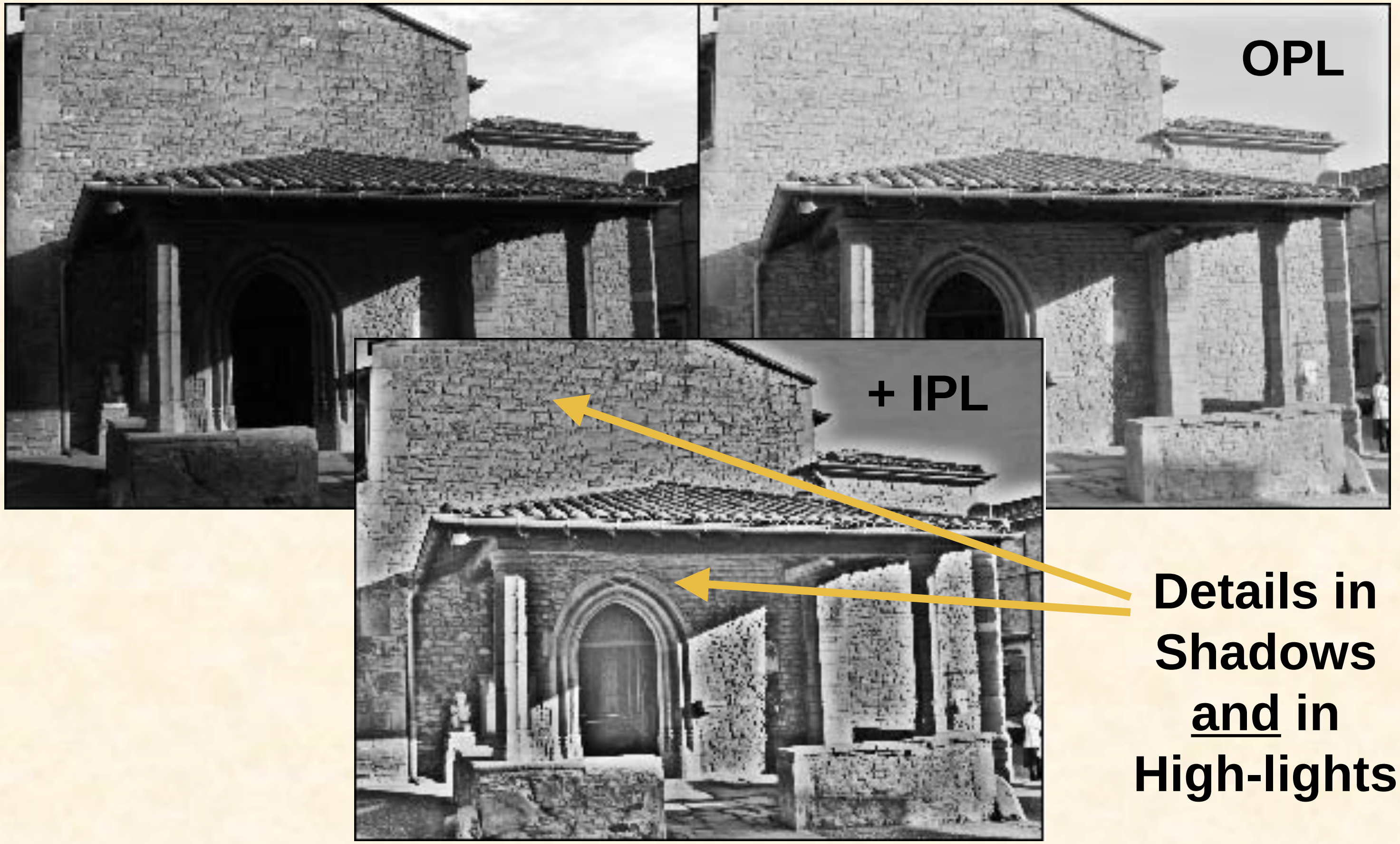
Retina: Non-Linearities

2. Compressive Adaptation in IPL



Retina: Non-Linearities

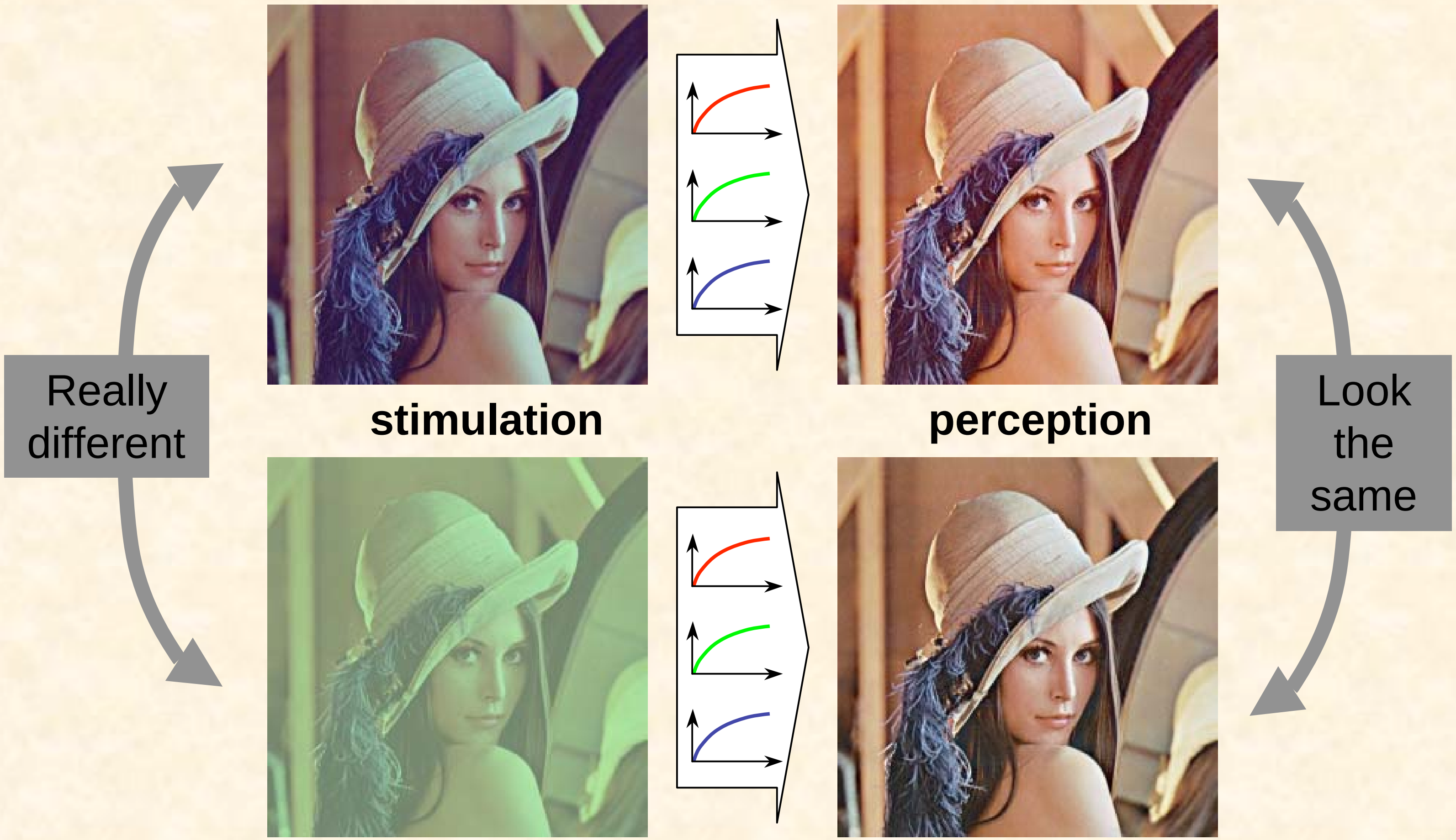
2. Compressive Adaptation in IPL: application



Details in Shadows <u>and</u> in High-lights

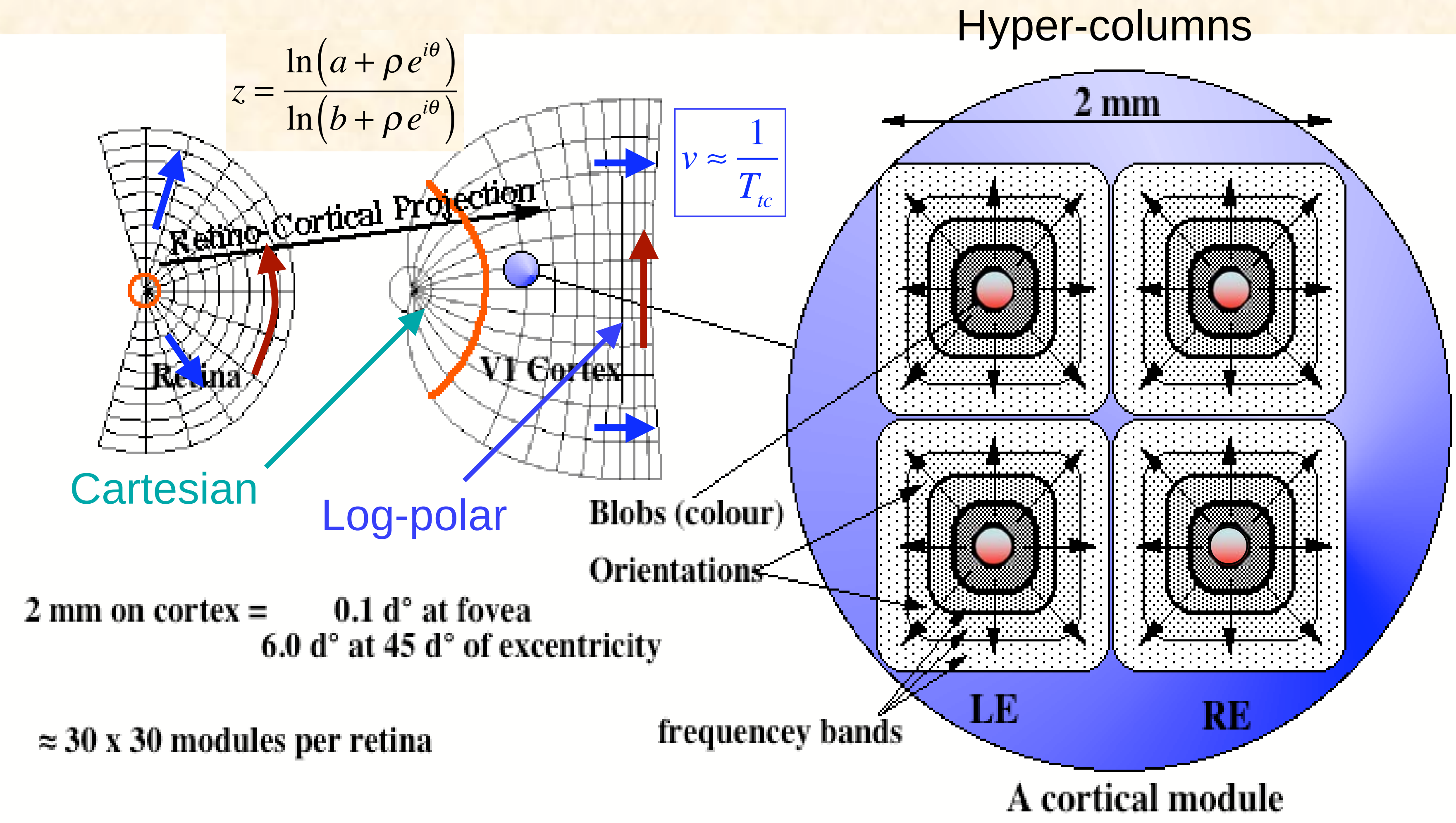
Non-Linearity and Color

3. Color Constancy => at the photoreceptor level

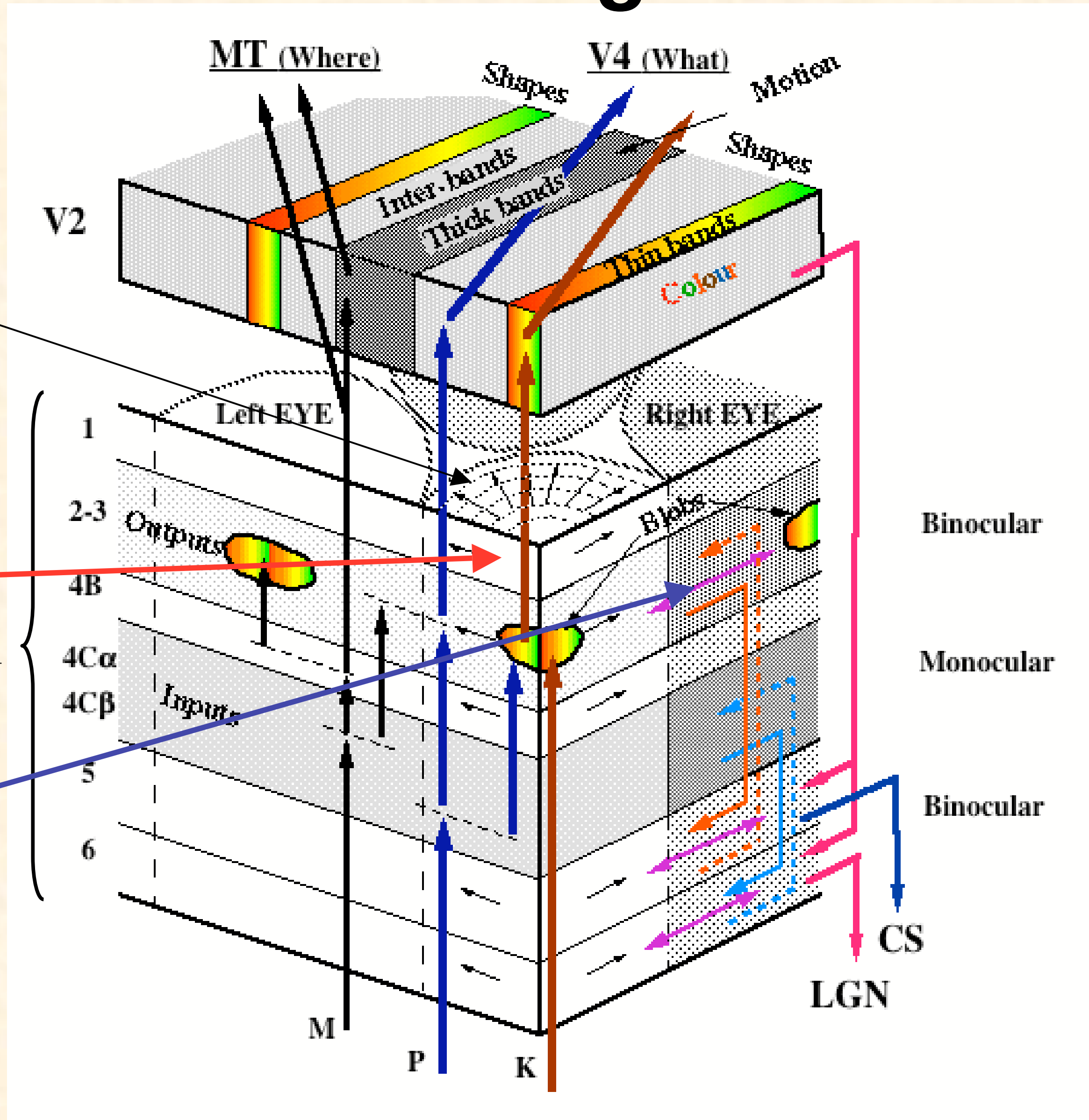


Alleysson Ph-D thesis (1999)

Retino-Cortical Projections



Primary Visual Cortex: Organization



Hyper-columns (Pin-wheel Organization)

Local interactions

Long range interactions

2nd Order Statistics and SPECTRAL ANALYSIS

Stat Autocorrelation: $R(x_1, y_1, x_2, y_2) = \mathbf{E}\left[i(x_1, y_1) \cdot i(x_2, y_2)\right]$ over all images, or by category

<u>H1: Stationary process</u> $\boxed{R(x, y) = \mathbf{E}\left[i(x_1, y_1) \cdot i(x_1 - x, y_1 - y)\right]}$

Th. Wiener-Kinchine: $F\left\{R(x, y)\right\} = S(f_x, f_y)$ ==> Spectral Density of Energy

<u>H2: Ergodicity:</u> $\boxed{R(x, y) = \gamma(x, y) = \iint_{x_1, y_1} i(x_1, y_1) \cdot i(x_1 - x, y_1 - y)\, dx_1\, dy_1}$

Statistical Autocorrelation = Spatial Autocorrelation

Properties: $S(f_x, f_y) = F\left\{\gamma(x, y)\right\} = \left|F\left\{i(x, y)\right\}\right|^2$ *The amplitude of the frequency spectrum is independent of the image position*

==> <u>One image is considered as a particular sample of a stochastic process</u>

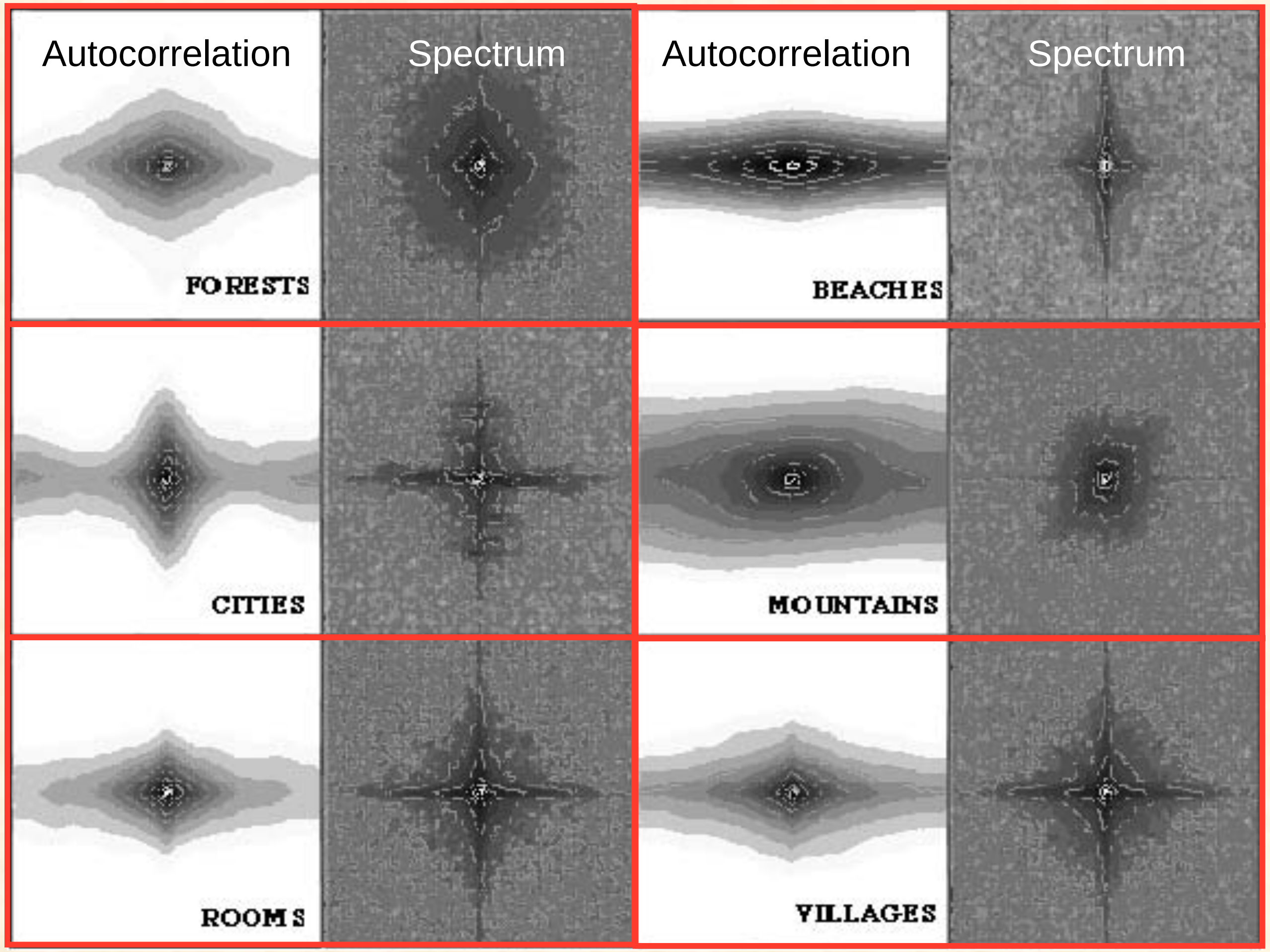
$\left|F\left\{i(x, y)\right\}\right|^2 \Leftrightarrow S_C(f_x, f_y)$ *The amplitude spectrum of an image is similar to that of its category*

Frequency Spectra of Scenes

1. Image data base

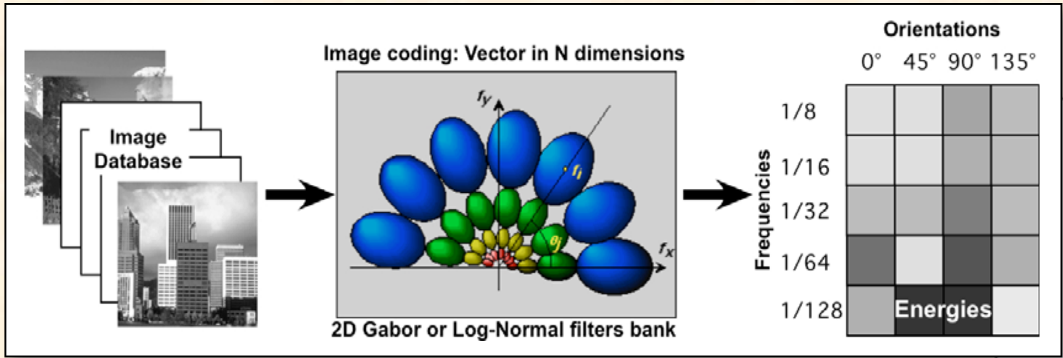
Frequency Spectra of Scenes

2. Mean Energy Spectra of image Categories



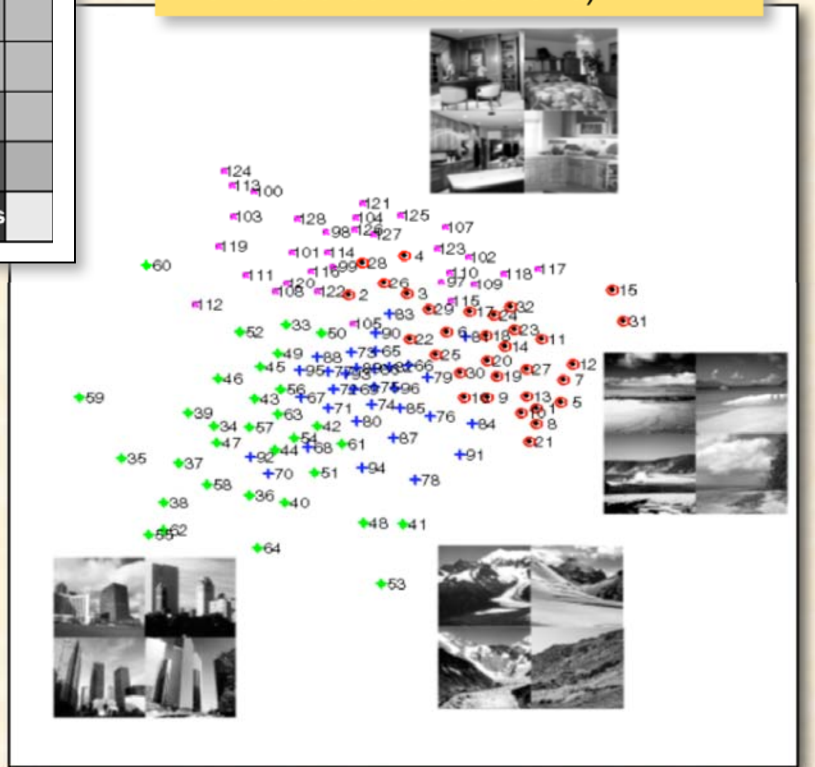
Cortical Model of Scene Analysis

Image Coding



Pin wheel models of V1 Complex cells

Scene Categorization by CCA: (Non-linear MDS with Metrics & Manifolds considerations)



1. **the N-dimensional space is not fully spanned by the image vectors**

2. **the N dimensions are non-linearly correlated**

N. Guyader Ph-D thesis (2004)

2D SPECTRAL ANALYSIS

2D image spectrum:

$$I(f_x, f_y) = \iint_{x,y} i(x,y)\exp\left[-j2\pi(x f_x + y f_y)\right]dxdy$$ ==> Cartesian in frequency

$$I(f,\theta) = \iint_{x,y} i(x,y)\exp\left[-j2\pi f(x\cos(\theta) + y\sin(\theta))\right]dxdy$$ ==> Polar in frequency

In Vector notation:

$$i(x,y) \xrightarrow[\mathbf{x}=[x,y]]{} i(\mathbf{x})$$

$$I(f_x, f_y) \xrightarrow[\underset{cartesian}{\mathbf{f}=[f_x,f_y]} = \underset{polar}{f\,e^{j\theta}}]{} I(\mathbf{f})$$

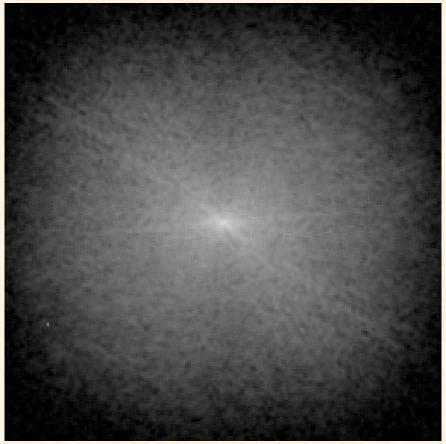
$$I(\mathbf{f}) = \iint_{\mathbf{x}} i(\mathbf{x})\exp\left[-j2\pi(\mathbf{x}^{\mathsf{T}}\mathbf{f})\right]d\mathbf{x}$$

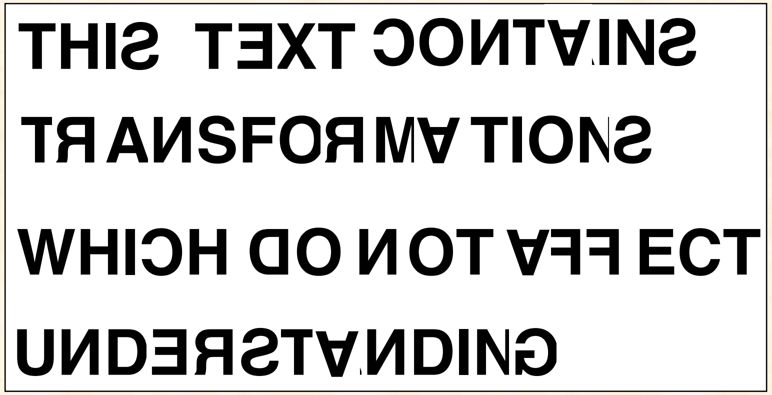
Remark:

$$\gamma(\mathbf{x}) = i(\mathbf{x}) * i(-\mathbf{x}) \quad \Rightarrow \quad \gamma(\mathbf{x}) = \gamma(-\mathbf{x}) \quad \Rightarrow \quad S(\mathbf{f}) = S(-\mathbf{f})$$

==> $S(\mathbf{f})$ is π periodic in θ

The Energy Spectrum is π Periodic



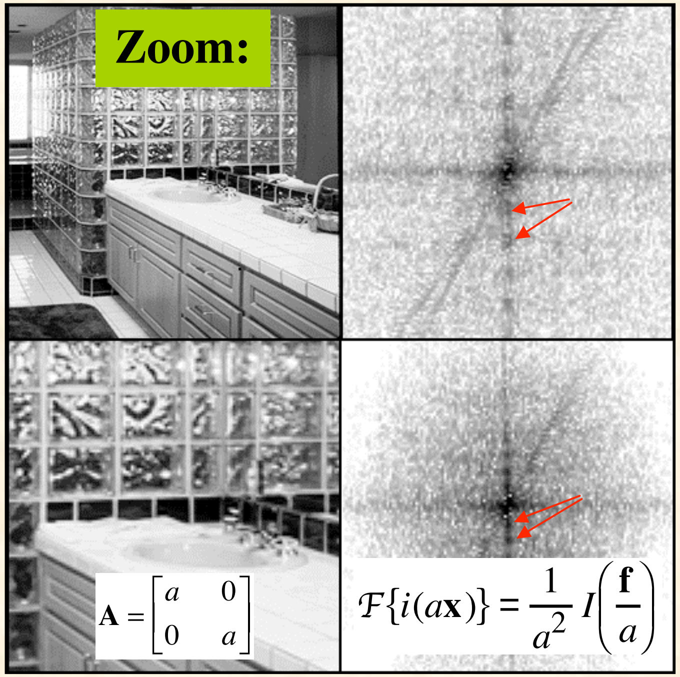


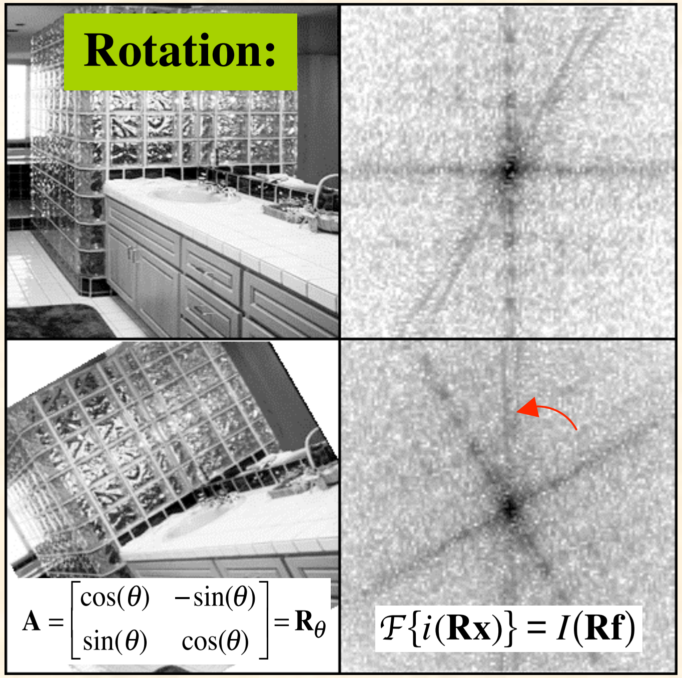
Variability within a same Category

spectrum of a 2D image transformation:

$$\mathcal{F}\{i(\mathbf{A}\mathbf{x})\} = \iint_{\mathbf{x}} i(\mathbf{A}\mathbf{x}) \exp\left[-j2\pi(\mathbf{x}^{\mathrm{T}}\mathbf{f})\right] d\mathbf{x} \quad \xrightarrow{\mathbf{x}' = \mathbf{A}\mathbf{x}} \quad \mathcal{F}\{i(\mathbf{A}\mathbf{x})\} = \frac{1}{|\det(\mathbf{A})|} \iint_{\mathbf{x}} i(\mathbf{x}') \exp\left[-j2\pi(\mathbf{x}'^{\mathrm{T}} \underbrace{\mathbf{A}^{-\mathrm{T}}\mathbf{f}}_{\mathbf{f}'})\right] d\mathbf{x}'$$

Mirror: $i(x,y) \rightarrow i(-x,y) \quad \Rightarrow \quad I(f_x, f_y) \rightarrow I(-f_x, f_y)$





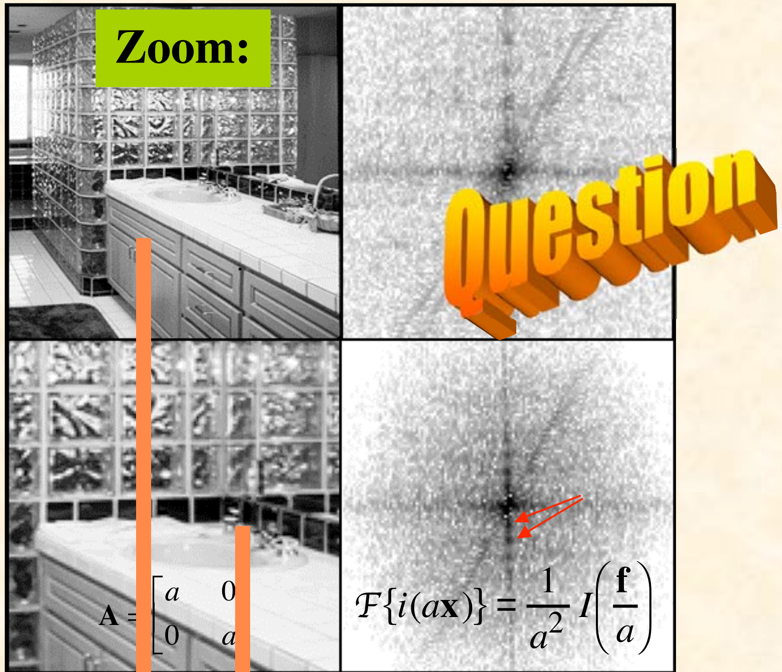
+ Perspective...

Variability within a same Category

spectrum of a 2D image transformation:

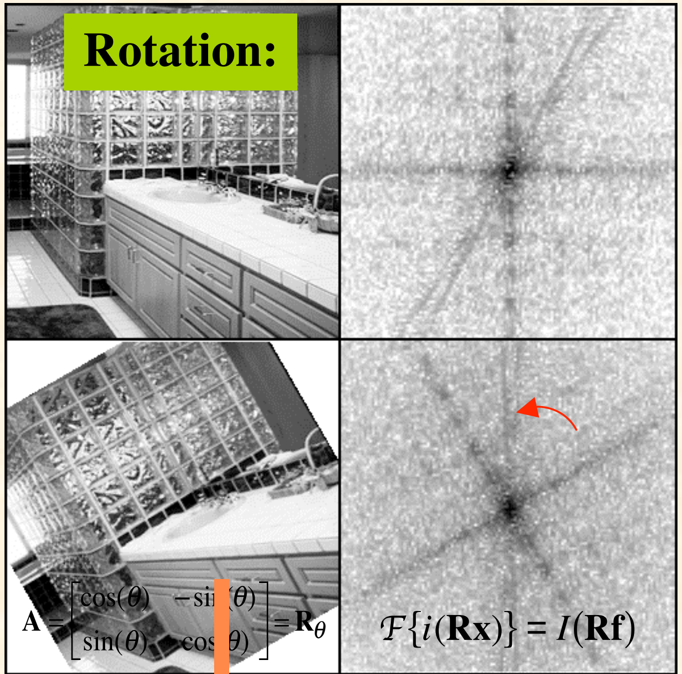
$$\mathcal{F}\{i(\mathbf{Ax})\}=\iint_{\mathbf{x}} i(\mathbf{Ax})\exp\left[-j2\pi(\mathbf{x}^{T}\mathbf{f})\right]d\mathbf{x} \quad \xrightarrow{\mathbf{x}'=\mathbf{Ax}} \quad \mathcal{F}\{i(\mathbf{Ax})\}=\frac{1}{|\det(\mathbf{A})|}\iint_{\mathbf{x}} i(\mathbf{x}')\exp\left[-j2\pi(\mathbf{x}'^{T}\underbrace{\mathbf{A}^{-T}\mathbf{f}}_{\mathbf{f}'})\right]d\mathbf{x}'$$

Mirror: $i(x,y)\rightarrow i(-x,y) \quad \Rightarrow \quad I(f_x,f_y)\rightarrow I(-f_x,f_y)$



$\mathbf{A}=\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$

$\mathcal{F}\{i(a\mathbf{x})\}=\frac{1}{a^2}I\left(\frac{\mathbf{f}}{a}\right)$



$\mathbf{A}=\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}=\mathbf{R}_\theta$

$\mathcal{F}\{i(\mathbf{Rx})\}=I(\mathbf{Rf})$

Same category ?

+ Perspective...

LOG-POLAR REPRESENTATION

Use of the Log-Polar frequency spectrum:

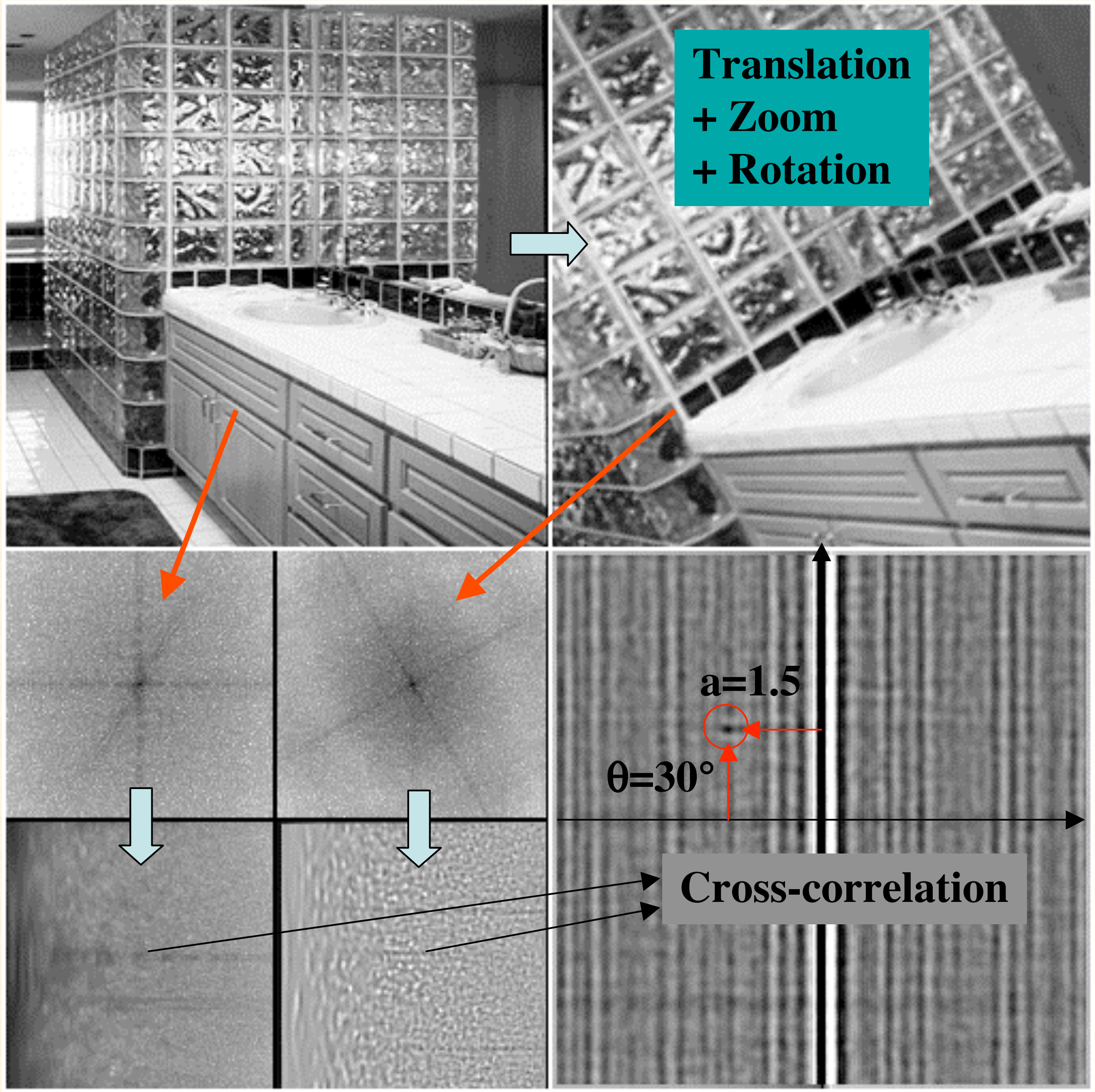
$$\mathcal{F}\left\{ i(a\mathbf{R}_\theta \mathbf{x}) \right\} = \boxed{\frac{1}{a^2} I(\frac{\mathbf{R}_\theta \mathbf{f}}{a})} \quad = \quad \boxed{\frac{1}{a^2} I(e^{\nu - \ln(a)}, \varphi - \theta)} \quad \text{if} \quad \nu = \ln(f)$$

Cartesian *Log-polar*

A zoom factor in the image space ==> a translation in log-frequency

A rotation in the image space ==> a translation in angular frequency

INTEREST OF LOG-POLAR SPECTRUM

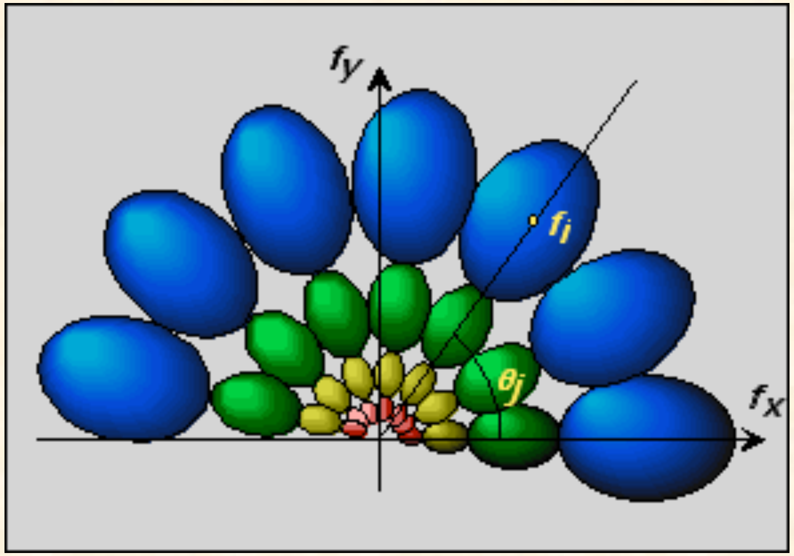


Amplitude Spectra

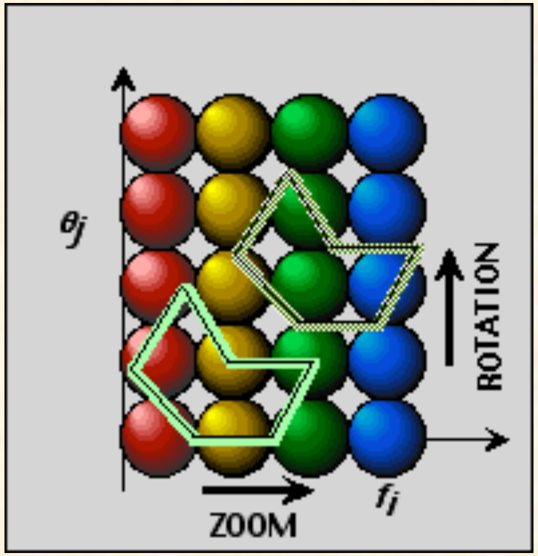
Log-Polar Spectra

Properties OF LOG-POLAR FILTERS

2D Filter bank



Log-Polar Transform



$$f_{i+1} = k\, f_i, \quad \theta_{j+1} = \theta_j + \frac{k'}{\pi}$$

$$\nu_{i+1} = \nu_i + \ln(k), \quad \theta_{j+1} = \theta_j + \frac{k'}{\pi}$$

- Image zoom and rotation correspond both to log-polar spectral translation

- Image perspective corresponds to a log-polar shear

Two kinds of LOG-POLAR Filter banks

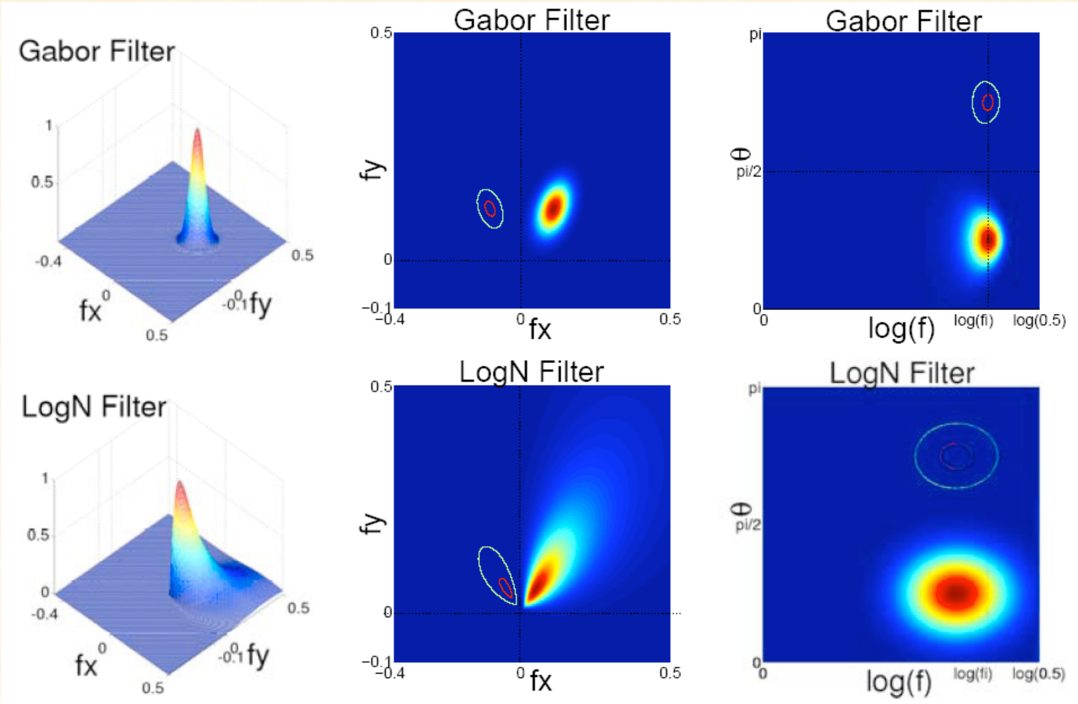
Gabor and Log-Normal filters

Gabor

$$G_i(f) = e^{-\frac{(f-f_i)^2}{2\sigma^2}}$$

Log-Normal

$$G_i(f) = \frac{1}{f} e^{-\frac{[\ln(f/f_i)]^2}{2\sigma'^2}}$$



Gaussian filters in log-polar domain => log-normal filters in cartesian domain

Two kinds of LOG-POLAR Filter banks

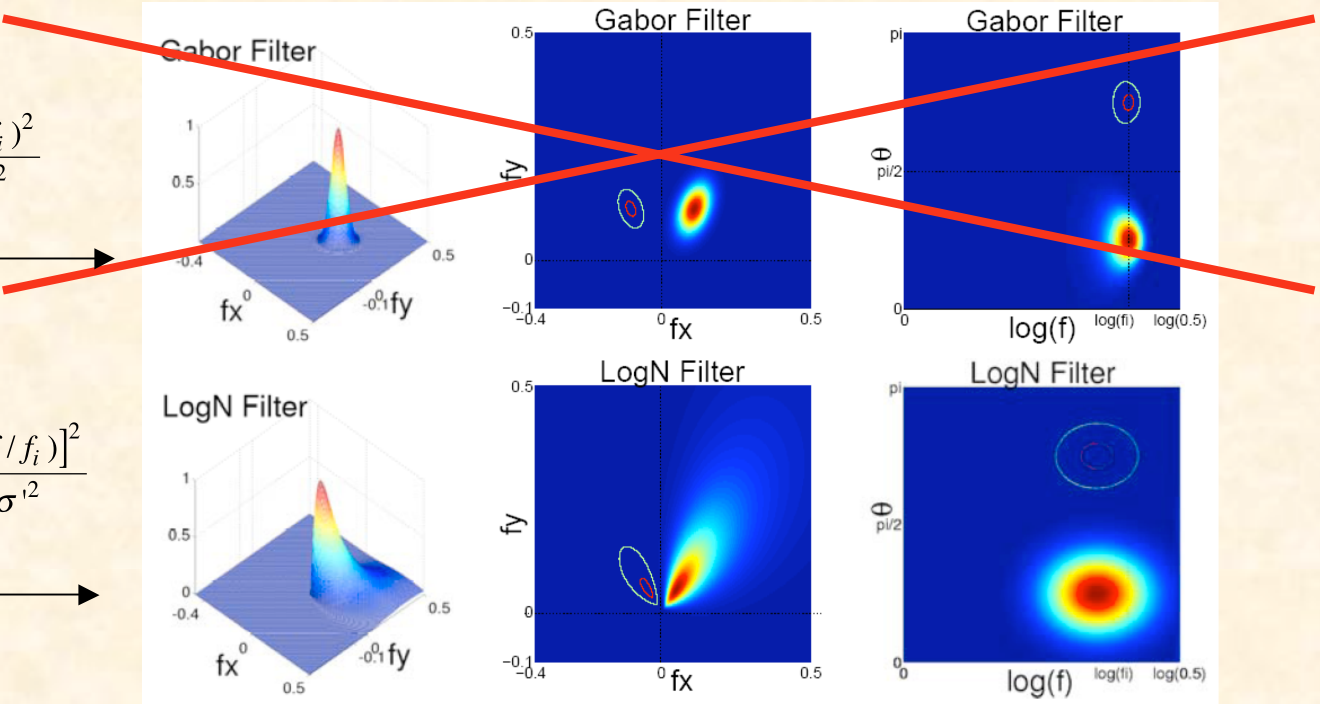
Gabor and Log-Normal filters

Gabor

$$G_i(f) = e^{-\frac{(f-f_i)^2}{2\sigma^2}}$$

Log-Normal

$$G_i(f) = \frac{1}{f} e^{-\frac{[\ln(f/f_i)]^2}{2\sigma'^2}}$$



=> **Log-normal filters are more suitable and more biologically plausible than classical Gabor filters**

Properties OF LOG-NORMAL FILTERS

2D filters with separable radial and angular variables:

$$|G_{ij}|^2 = |G_{i.}|^2 \, |G_{.j}|^2$$

Mean energy in each frequency band :

$$C_{ij} = \iint_{f,\theta} S(f,\theta) \boxed{\frac{1}{f^2} e^{-\frac{\ln^2(f/f_i)}{2\sigma^2}}} \boxed{\left(\cos\left(\theta-\theta_j\right)\right)^{2n}} \, f df d\theta$$

$G_{i.}$ are *Log-normal* filters: $\boxed{|G_{i.}| = \frac{1}{f} e^{-\frac{(\ln(f/f_i))^2}{4\sigma^2}}}$

$G_{.j}$ are π *periodic Gaussian-like* filters: $\boxed{|G_{.j}| = \left(\cos(\theta-\theta_j)\right)^{2n}}$

with: $\nu = \ln(f/f_0)$

$$C_{ij} = \iint_{\nu,\theta} S(e^{\nu},\theta) \, e^{-\frac{(\nu-\nu_i)^2}{2\sigma^2}} \left(\cos(\theta-\theta_j)\right)^{2n} d\nu d\theta$$

For zoom and rotation:

$$i(a\,\mathbf{R}_{\varphi}\mathbf{x}) \Leftrightarrow \frac{1}{a^2} I\left(\frac{1}{a}\mathbf{R}_{\varphi}\mathbf{f}\right)$$

$$C_{ij} \approx \frac{1}{a^4} \iint_{\nu,\theta} \boxed{S(e^{\nu-\ln(a)},\theta-\varphi)} \, e^{-\frac{(\nu-\nu_i)^2}{2\sigma^2}} \, e^{-\frac{(\theta-\theta_j)^2}{2\sigma'^2}} d\nu d\theta$$

$$C_{ij} \approx \frac{1}{a^4} \iint_{\nu,\theta} S(e^{\nu},\theta) \boxed{e^{-\frac{(\nu-\nu_i+\ln(a))^2}{2\sigma^2}} \, e^{-\frac{(\theta-\theta_j+\varphi)^2}{2\sigma'^2}}} d\nu d\theta$$

==> *The log-polar spectrum is regularly sampled by the filter bank in f_i and θ_j*
If the number of filters is odd, non-integer translations in f_i and θ_j can be coded

INTEREST OF LOG-NORMAL FILTERS (1)

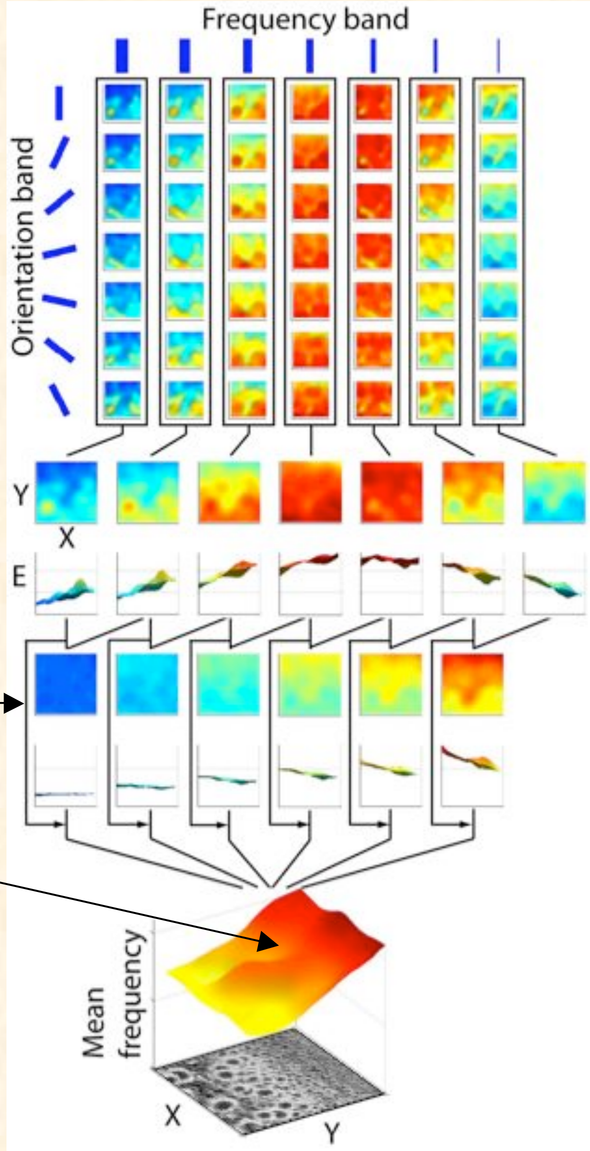
Perspective estimation through local frequency gradient

(similar to long-range inretactions in V1)

narrow-band local frequency

$$\frac{C_{i+1}(x,y)}{C_i(x,y)} = \frac{1}{\sqrt{f_i f_{i+1}}} \langle f \rangle_i (x,y)$$

wide-band local frequency



C. Massot Ph-D thesis (2005)

INTEREST OF LOG-NORMAL FILTERS (2)

Fourier transform of C_{ij}

$$\mathcal{F}\left\{C_{ij}\right\} = \frac{1}{a^4} \iint_{\nu_i,\theta_i} \iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-\frac{(\nu-\nu_i+\ln(a))^2}{2\sigma^2}}\, e^{-\frac{(\theta-\theta_j+\varphi)^2}{2\sigma'^2}}\, d\nu d\theta\, e^{-j2\pi(\xi\nu_i+\eta\theta_j)}\, d\nu_j d\theta_j$$

$$\mathcal{F}\left\{C_{ij}\right\}(\xi,\eta) = \frac{1}{a^4}\, \underbrace{e^{-(k_1\xi^2+k_2\eta^2)}}_{\text{Gaussian envelope}}\, \underbrace{e^{-j2\pi(\xi\ln(a)+\eta\varphi)}}_{\text{Zoom \& Rotation}} \iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-j2\pi(\xi\nu+\eta\theta)}\, d\nu d\theta$$

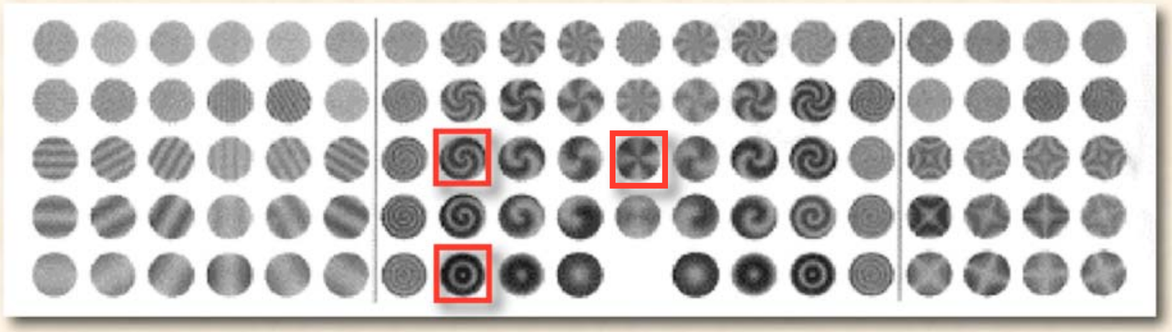
Gaussian envelope *Zoom & Rotation*

$\left|\mathcal{F}\left\{C_{ij}\right\}(\xi,\eta)\right|$ *independent of* (a,φ)

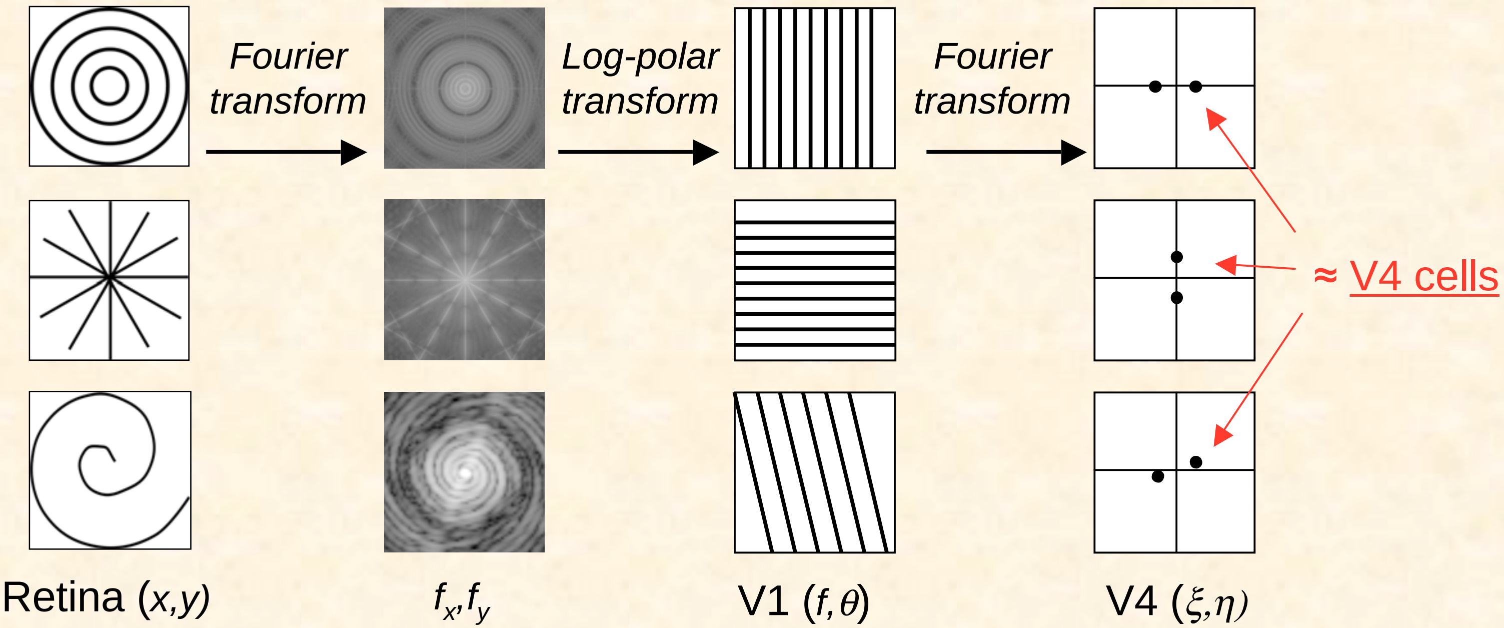
==> a possible operation done in visual area V4...

Model of V4

Stimuli and
Best responses
in area V4 cells

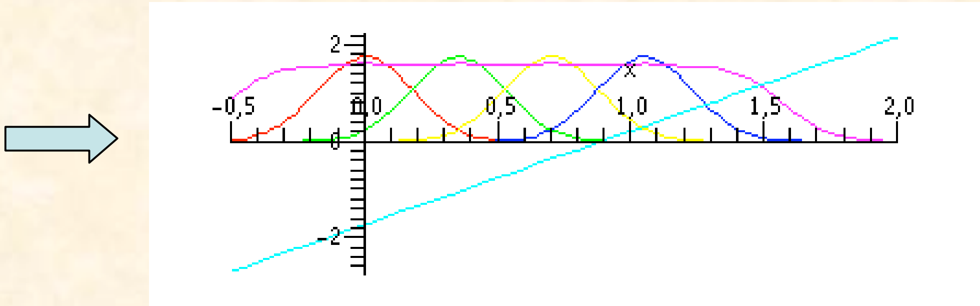


Model



CHOOSING BANDWIDTHS

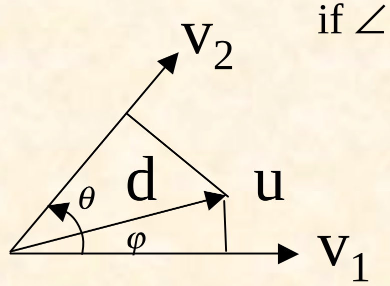
The log-spectrum is convolved by: $e^{-\frac{\left(\ln(f/f_i)\right)^2}{2\sigma^2}}$ Then sampled every $\Delta\nu = \ln\left(f_{i+1}/f_i\right)$



σ **is choosen according to Nyquist condition**
or spectral coverage

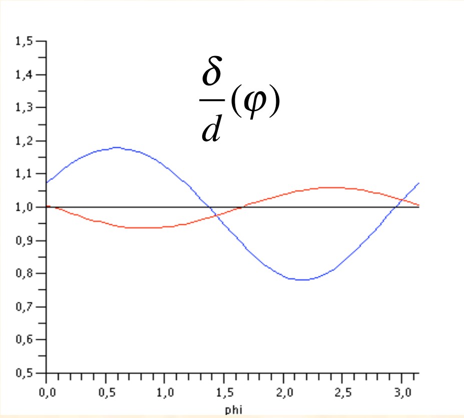
Pb: ORTHOGONALITY of FILTERS

$C_{ij} = \frac{1}{a^4} \iint_{\nu,\theta} S(e^{\nu},\theta)\, e^{-\frac{(\nu-\nu_i)^2}{2\sigma^2}}\, e^{-\frac{(\theta-\theta_j)^2}{2\sigma'^2}}\, d\nu d\theta$ **= dot product between a signal** $S(e^{\nu},\theta)$ **and functions** $\mathbf{v}_{ij} = \left|G_{ij}\right|^2$



if $\angle$ $v_i v_j = \theta \neq 90°$, the estimated "Euclidean" distance δ is not the real one d:

$$\delta^2 = d^2 \cos^2(\varphi) + d^2 \cos^2(\theta - \varphi)$$



example for various θ: $\angle$ C_1C_2=67°, $\angle$ C_1C_3=88.7°, $\angle$ C_1C_4= 89.98°

ORTHOGONALITY of FILTERS (2)

Possible solution: **linear combinations of filters**

$$C_i \rightarrow \frac{C_i - aC_{i-1} - aC_{i+1}}{1+2a} \qquad C_j \rightarrow \frac{C_j - a'C_{j-1} - a'C_{j+1}}{1+2a'}$$

a	$\angle C_1C_2$	$\angle C_1C_3$	$\angle C_1C_4$
0.18	85.75	96.07	89.66
0.21	89.21	97.09	89.42
0.23	91.54	97.71	89.21

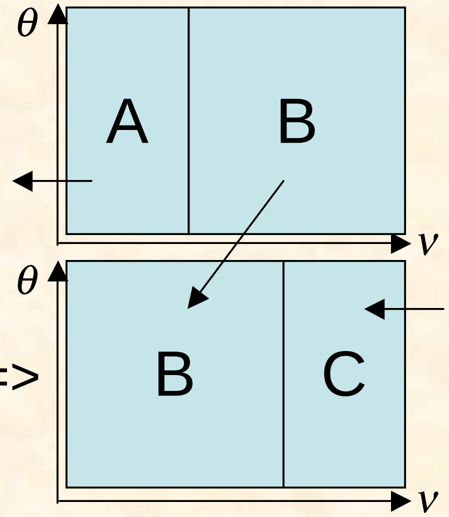
$$a \approx 0.2$$

similar to cortical interactions between filters **==> Euclidean distances are better approximated**

Problem with different Zoom Ratios

$I_1(\ln(f))$ =>

$I_2(\ln(f)-\ln(a))$ =>



Low frequencies (A) *disappear*

Frequency pattern (B) *translates*

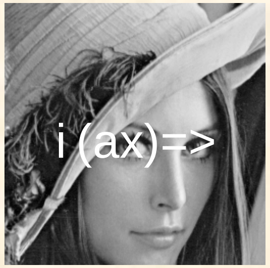
High frequencies (C) *appear*

Classical distance: $d_{1,2}^2 = \|I_1\|^2 + \|I_2\|^2 - \gamma_{1,2}(0) - \gamma_{2,1}(0) = \|A\|^2 + 2\|B\|^2 + \|C\|^2 - \gamma_{1,2}(0) - \gamma_{2,1}(0)$

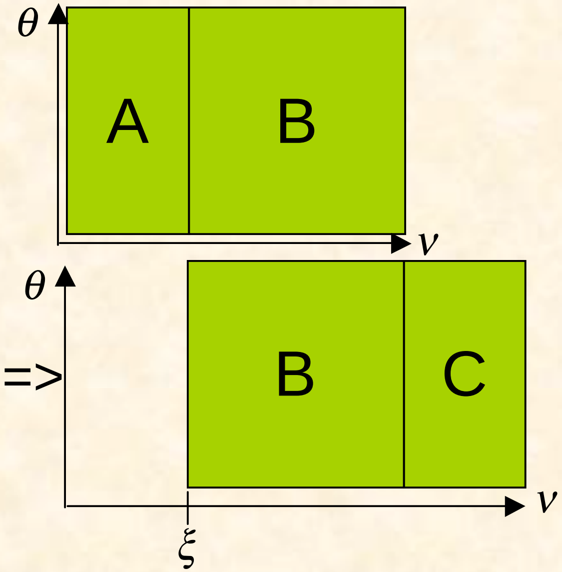
should disappear

Problem with different Zoom Ratios

$I_1(\ln(f))$ =>



$I_2(\ln(f)-\ln(a))$ =>

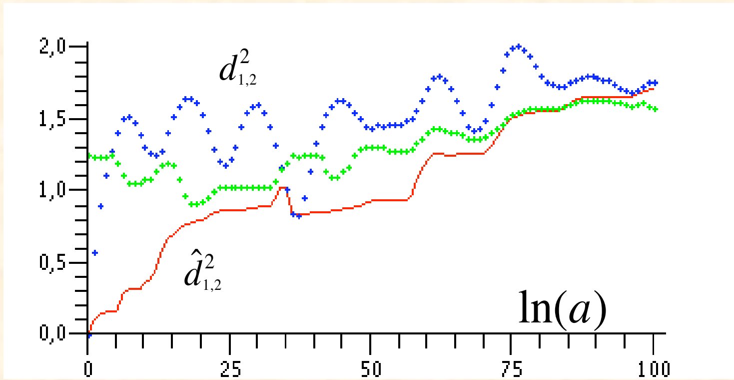


search for maximum intercorrelation

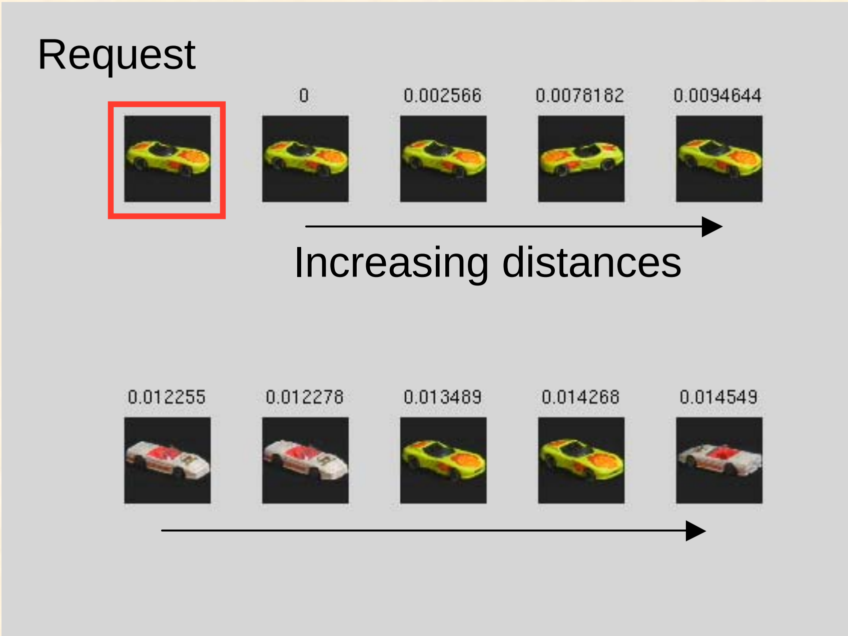
Better distance would be: $\hat{d}_{1,2}^2 = \|A\|^2 + \|C\|^2$

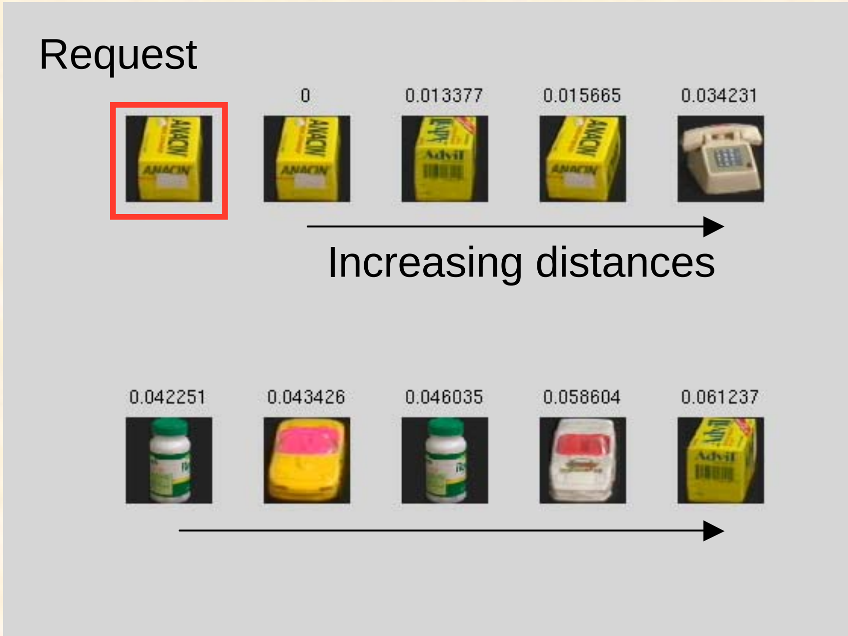
That is:

$$\hat{d}_{1,2}^2 = \underbrace{\|I_1\|^2 + \|I_2\|^2}_{\|A\|^2 + 2\|B\|^2 + \|C\|^2} - \underbrace{2\ Max\left(\gamma_{1,2}(\xi)\right)}_{2\|B\|^2}$$



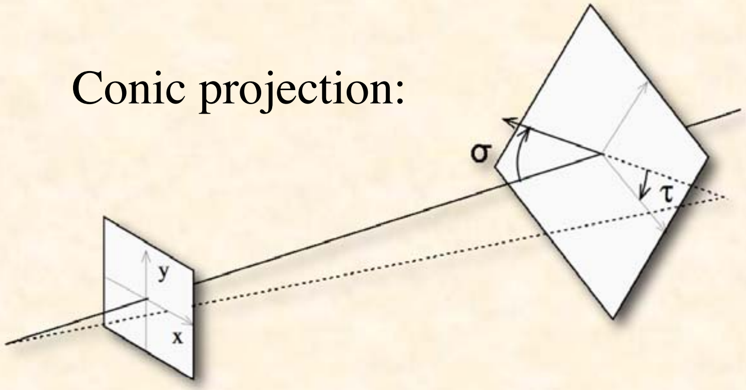
EXAMPLES


Request
0
0.002566
0.0078182
0.0094644
Increasing distances
0.012255
0.012278
0.013489
0.014268
0.014549


Request
0
0.013377
0.015665
0.034231
Increasing distances
0.042251
0.043426
0.046035
0.058604
0.061237

For PERSPECTIVE

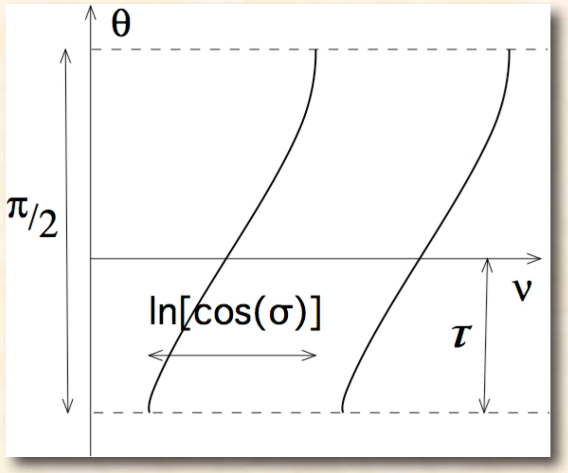
Conic projection:



slant: σ
tilt: τ

Maximum frequency expansion: ln[cos(σ)] at $\theta = \pi/2 + \tau$

Log-polar frequency plane



To be continued...

Thank you for your attention