Mathematische Probleme des künstlichen Sehens aus der Perspektive spezifischer Anwendungen

Erhardt Barth

Institut für Neuro- und Bioinformatik

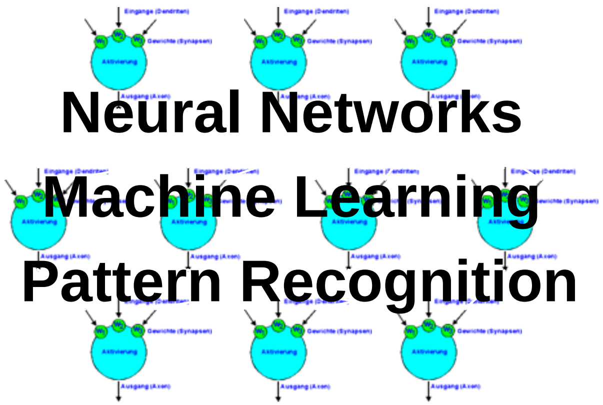
Universität zu Lübeck

barth@inb.uni-luebeck.de

Overview

- INB
- Visual information processing in the brain
  - ***intrinsic dimension*** and endstopping
  - motion selectivity
  - predictive coding
- *Complex motion (LOCOMOTOR, DFG SPP)*
- *Prediction and guidance of eye movements (Modkog/Itap, BMBF)*
- Applications in the car (SMI, BMW, VW)
  - tracking: ***fusion of information***
- Organic computing (DFG, EC)
  - ***emergent learning***

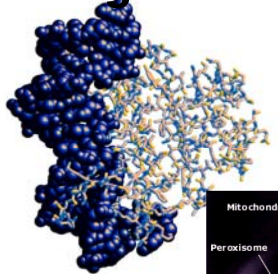
Institute for Neuro- and Bioinformatics

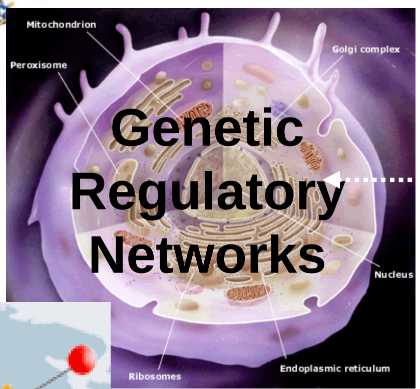


Neural Networks
Machine Learning
Pattern Recognition

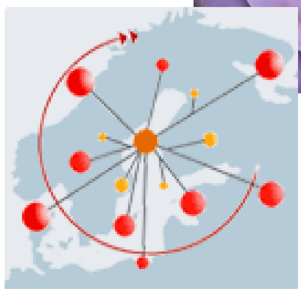
Organic computing

Binding Site Recognition



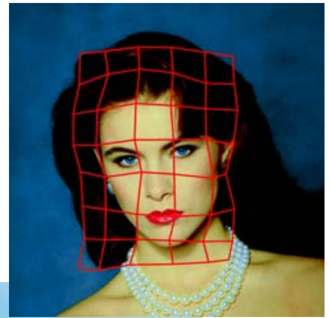


Genetic Regulatory Networks



Decisison Support

Face Recognition

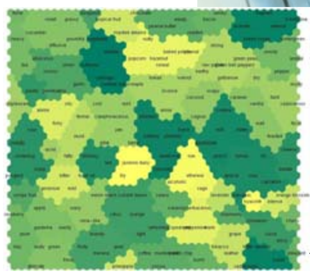


Automated diagnosis

Computer Vision
Attention Control

COGAIN
Communication by Gaze Interaction

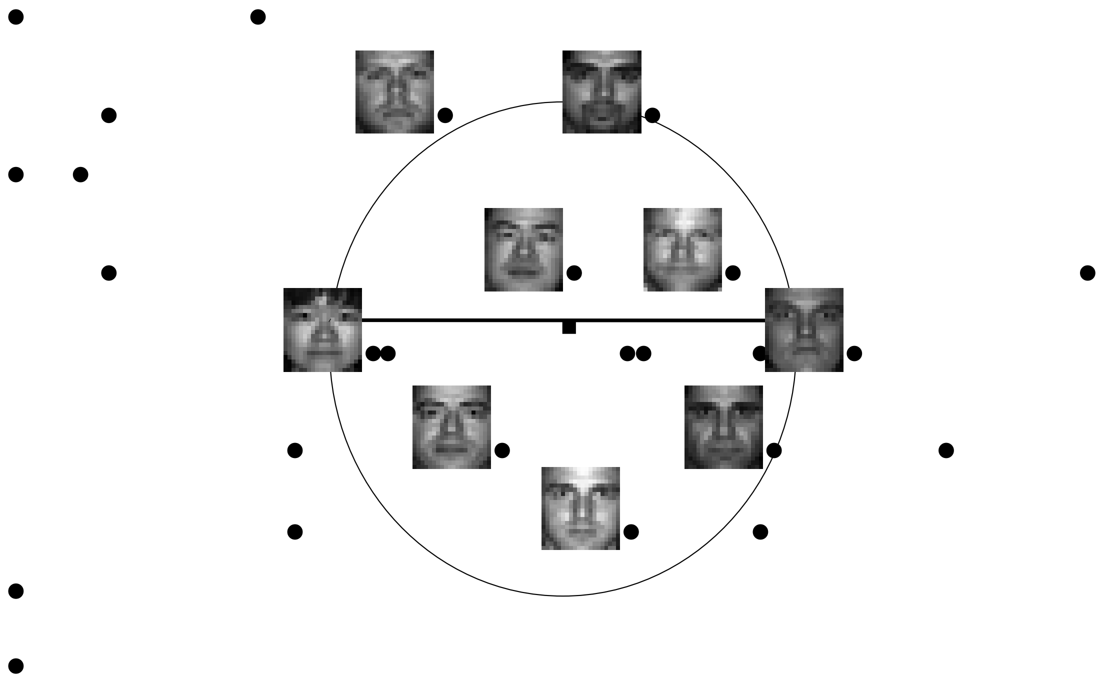
Information Society Technologies



Perception Space Analysis

INB

Face recognition



Where are the faces?

Seeing with sound

Video is transformed to sound

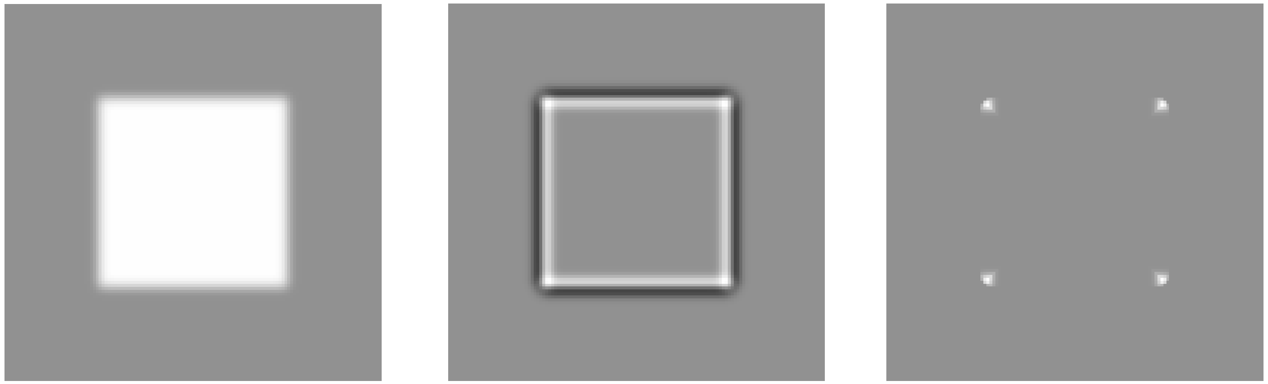
Blind people listen to the sound and learn to „see“

Mathematical problem: how can a sequence of images (three-dimensional signal) be transformed to sound (one-dimensional signal) with a minimal loss of information and under given biological constraints.

Current issues

- Information
  - what is redundant?
- Complex motion
  - overlaid motions
- Tracking
  - tracking of humans still difficult
- Learning
  - emergent learning
  - learning of features
- How the brain works

Endstopped neurons



A significant number of V1 and V2 neurons are endstopped, i.e., they do not respond to straight edges or lines, but only to corners and line ends.

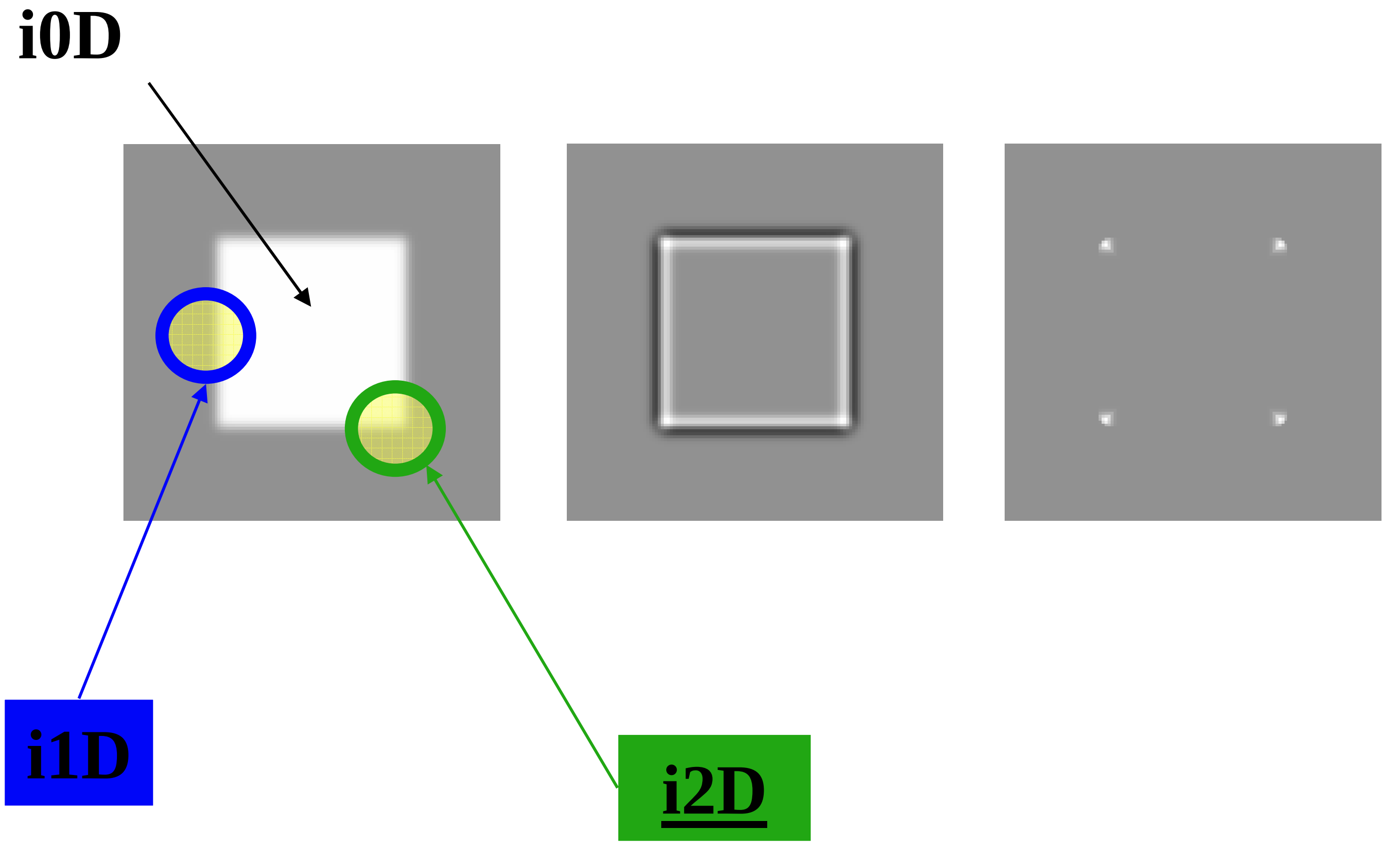
Intrinsic dimension in 2D

i0D constant in all directions: $f(x,y) = \text{const.}$

i1D constant in 2 directions: $\boxed{f(x,y) = g(\xi)}$

i2D no constant direction: $\boxed{f(x,y) = g(\xi,\zeta)}$

Intrinsic dimension in 2D



Intrinsic dimension in 3D

i0D: constant in all directions: $f(x,y,t) = const.$

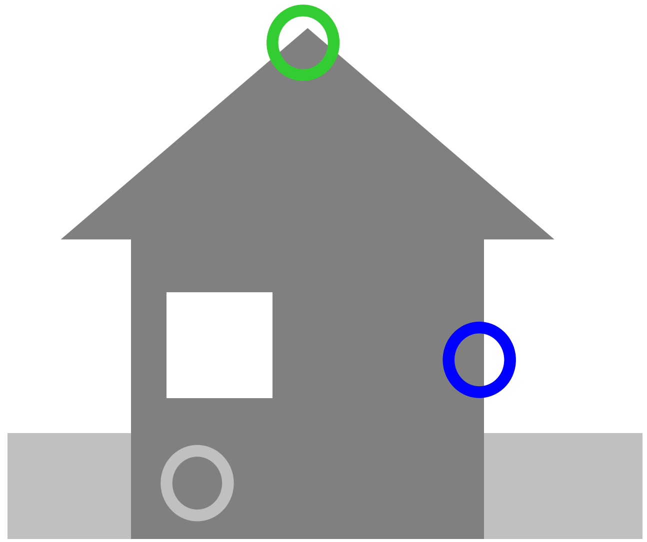
i1D constant in 2 directions: $\boxed{f(x,y,t) = g(\xi)}$

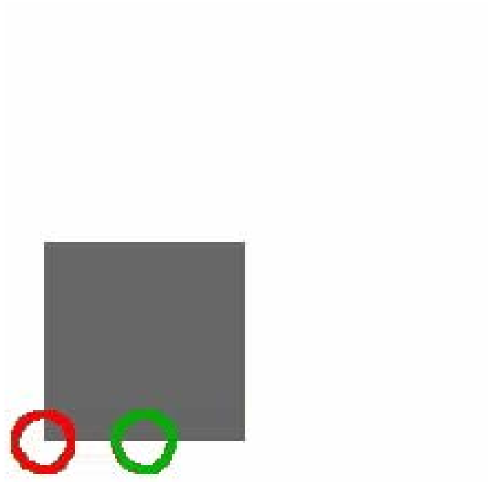
i2D constant in one direction: $\boxed{f(x,y,t) = g(\xi,\zeta)}$

i3D no constant direction: $\boxed{f(x,y,t) = g(\xi,\zeta,\tau)}$

Intrinsic dimension

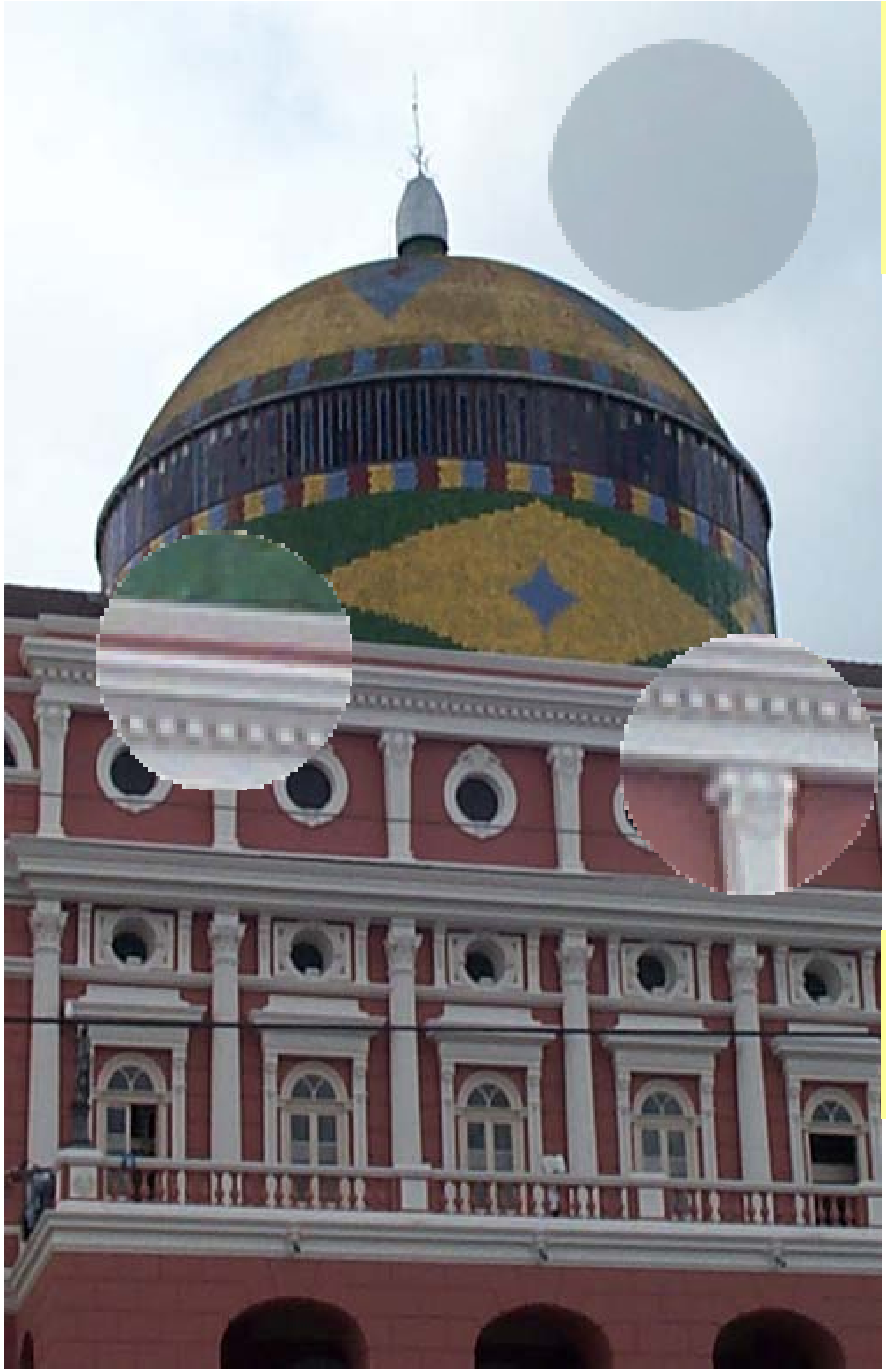
i2D and *i3D* regions are the least frequent but most significant.





○ *i0D* ○ *i1D* ○ *i2D* ○ *i3D*

Intrinsic dimension ...

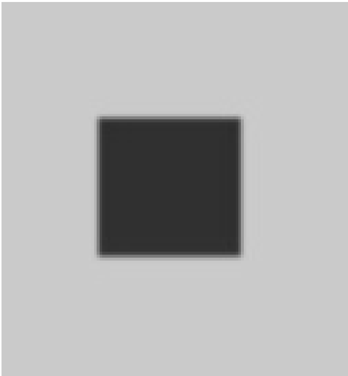


Corners and junctions are the least frequent but most significant image features.

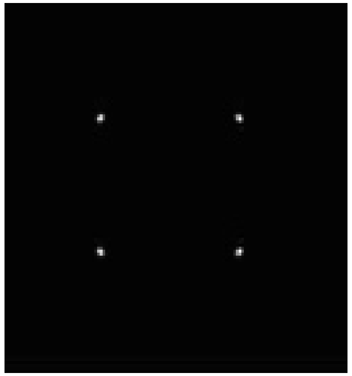
"A compact surface is determined by its curved (i2D) regions."

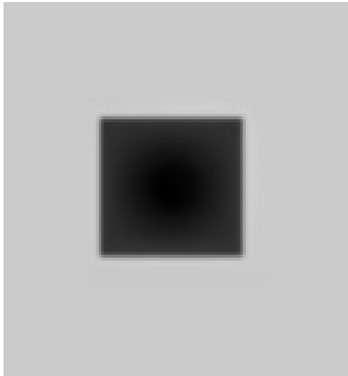
C. Mota and E. Barth: *On the uniqueness of curvature features.* Proc. Artificial Intelligence, vol.9: 175-8, 2000.

Reconstruction of a square



Original image



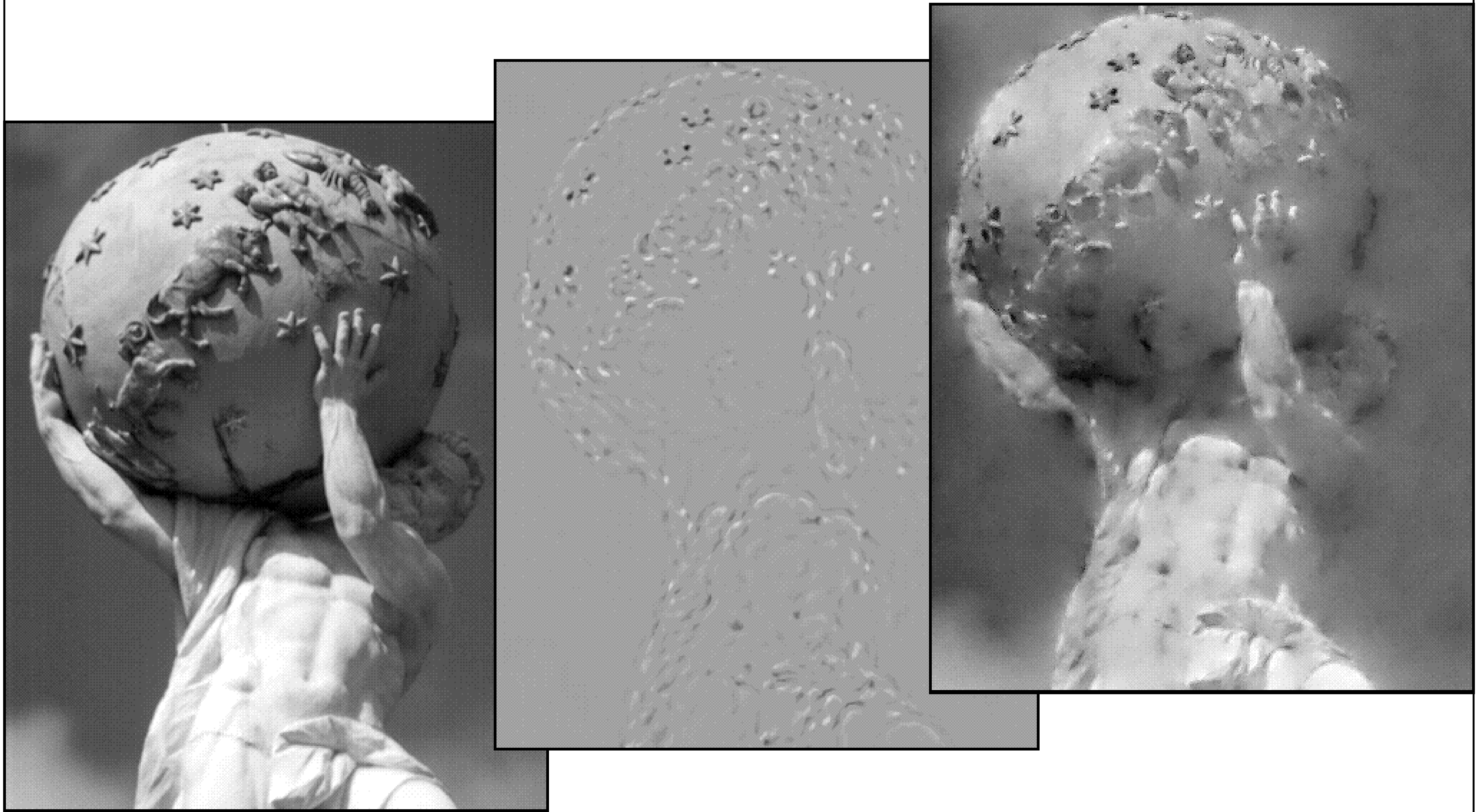


Reconstructed image

$$\boxed{f_y^2 f_{xx} - 2 f_x f_y f_{xy} + f_x^2 f_{yy}}$$

INB

Reconstruction of natural images from i2D features



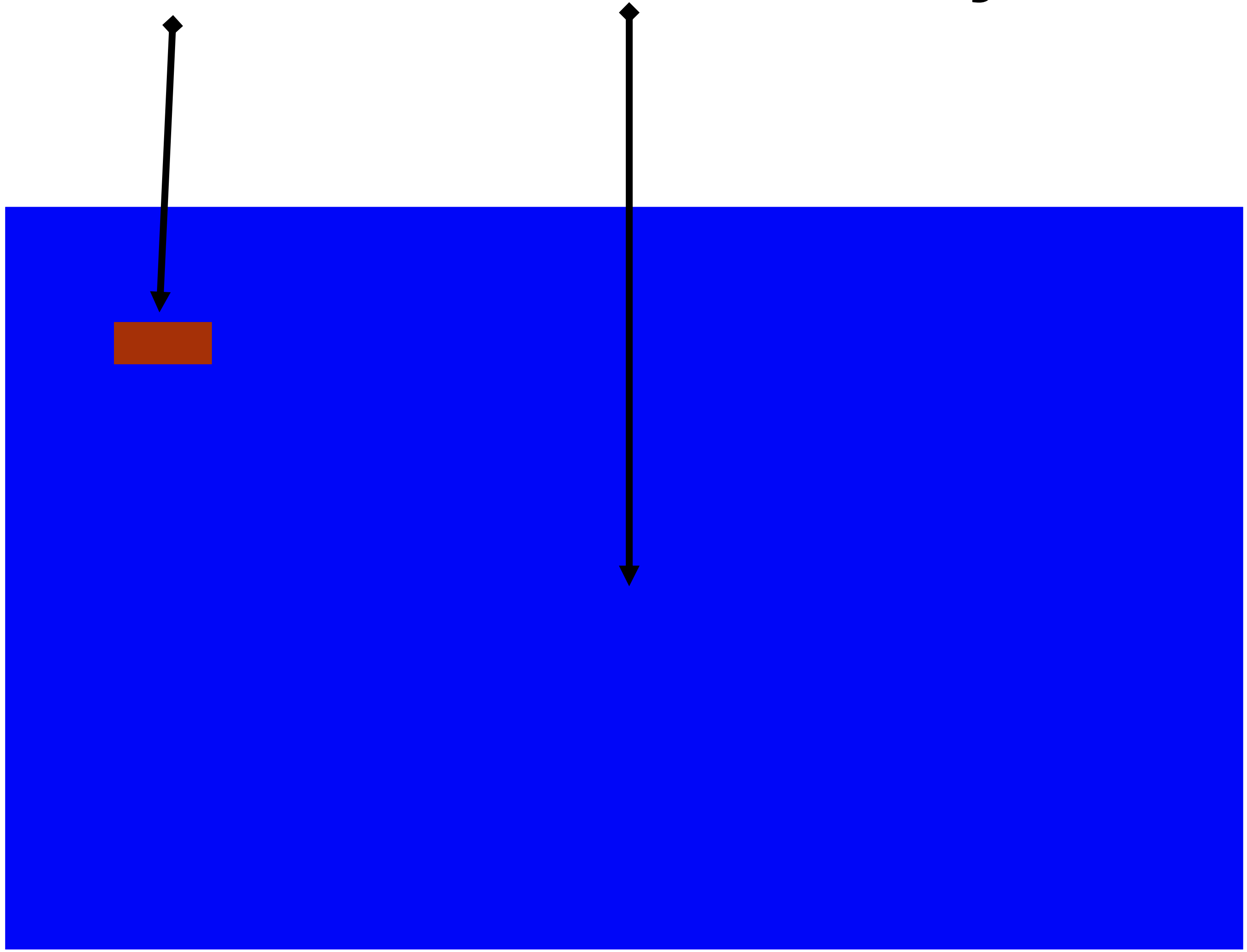
Intrinsic dimensionality in 2D

i0D: constant in all directions

i1D: constant in one direction

i2D: no constant direction

Linear and nonlinear systems



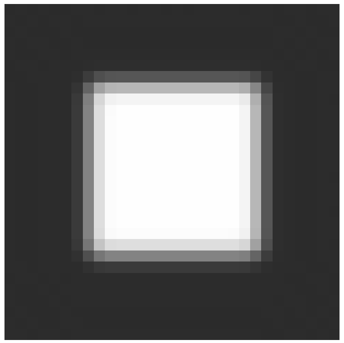
Images as surfaces

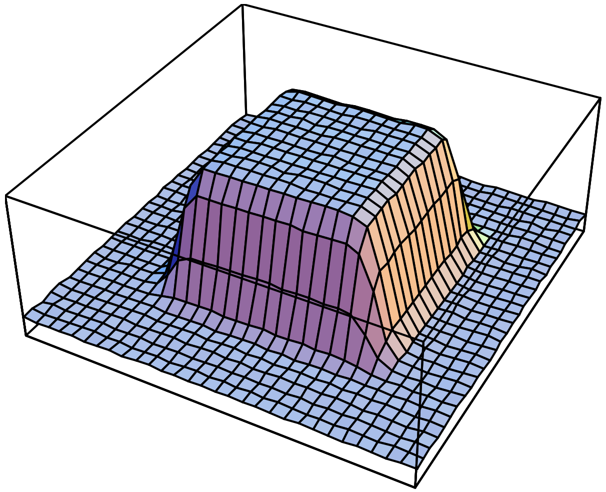
image intensity

$$f(x, y)$$

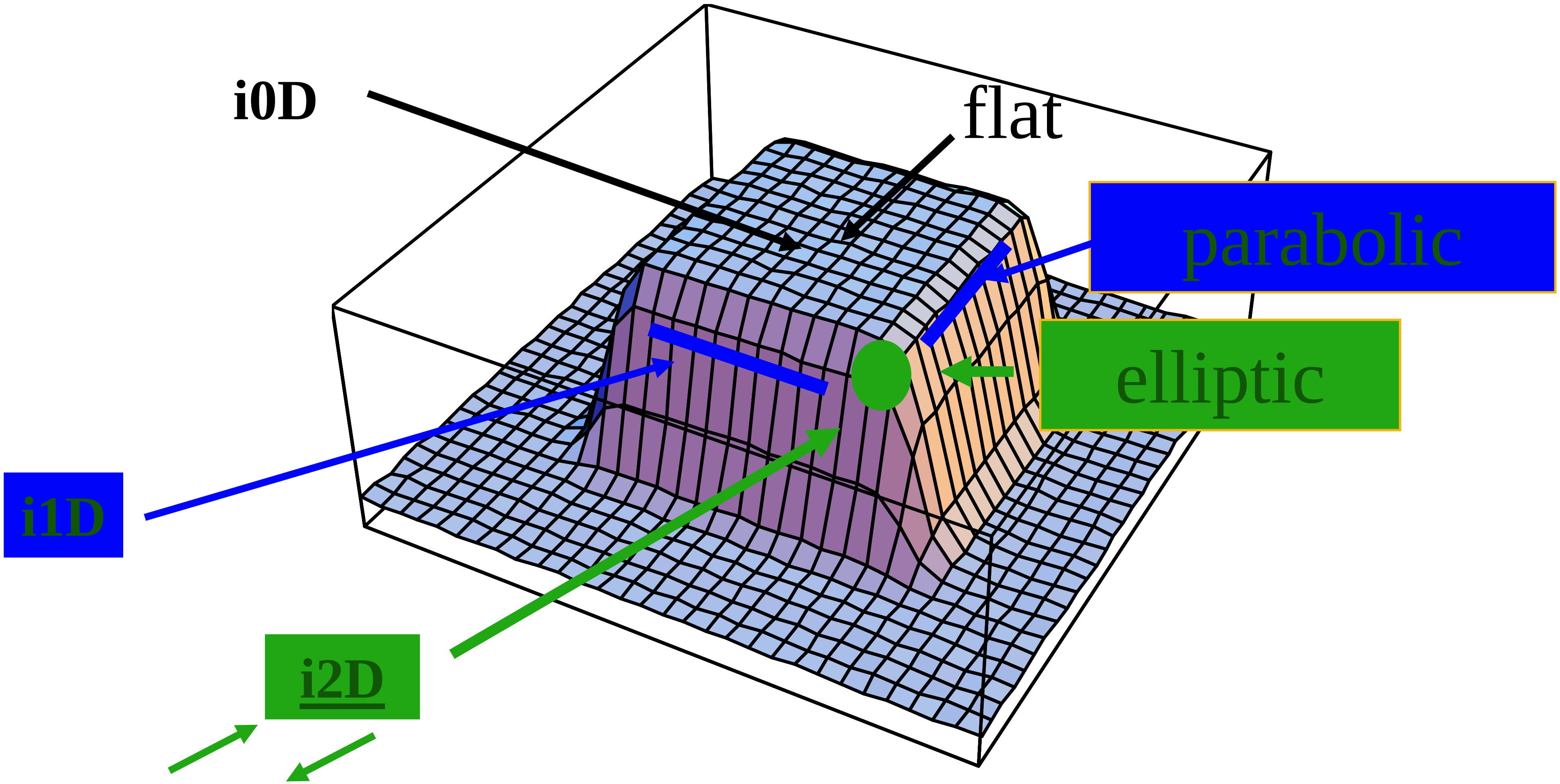
surface

$$(x, y, f(x, y))$$





Intrinsic dimension and curvature



$$\mathbf{R} \neq 0, \quad K \neq 0$$

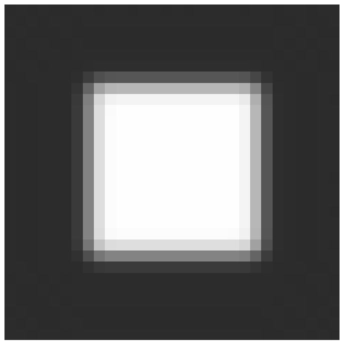
The visual input as a hypersurface

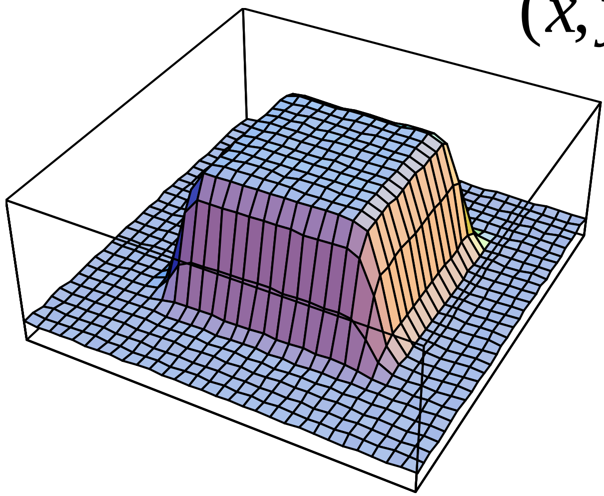
luminance $f(x, y, t)$

hypersurface $(x, y, t, f(x, y, t))$

Visualization of surfaces is easier:

$(x, y, f(x, y))$





Geometry and intrinsic dimension

- Hypersurface Geometry
  - mean curvature $H \neq 0$
    - **i1D++**
  - Riemann curvature tensor $R \neq 0$
    - **i2D+**
  - Gaussian curvature $K \neq 0$
    - **i3D**

The Riemann tensor components

$$R_{3131} = \frac{f_{xx} f_{tt} - f_{xt}{}^2}{1 + f_x^2 + f_y^2 + f_t^2}$$

$$R_{2121} = \frac{f_{xx} f_{yy} - f_{xy}{}^2}{1 + f_x^2 + f_y^2 + f_t^2}$$

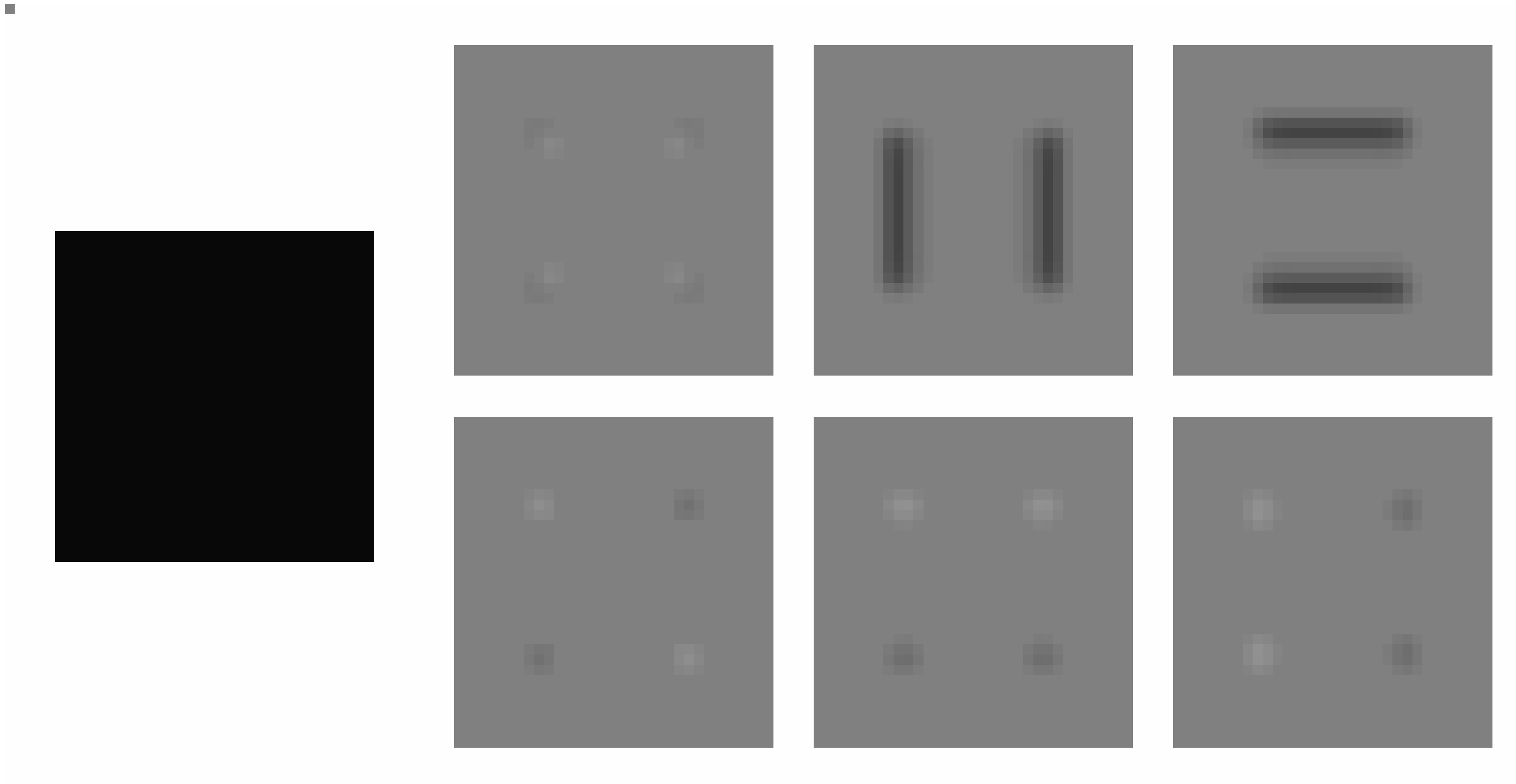
$$R_{3232} = \frac{f_{yy} f_{tt} - f_{yt}{}^2}{1 + f_x^2 + f_y^2 + f_t^2}$$

$$R_{3121} = \frac{f_{xx} f_{yt} - f_{xt} f_{xy}}{1 + f_x^2 + f_y^2 + f_t^2}$$

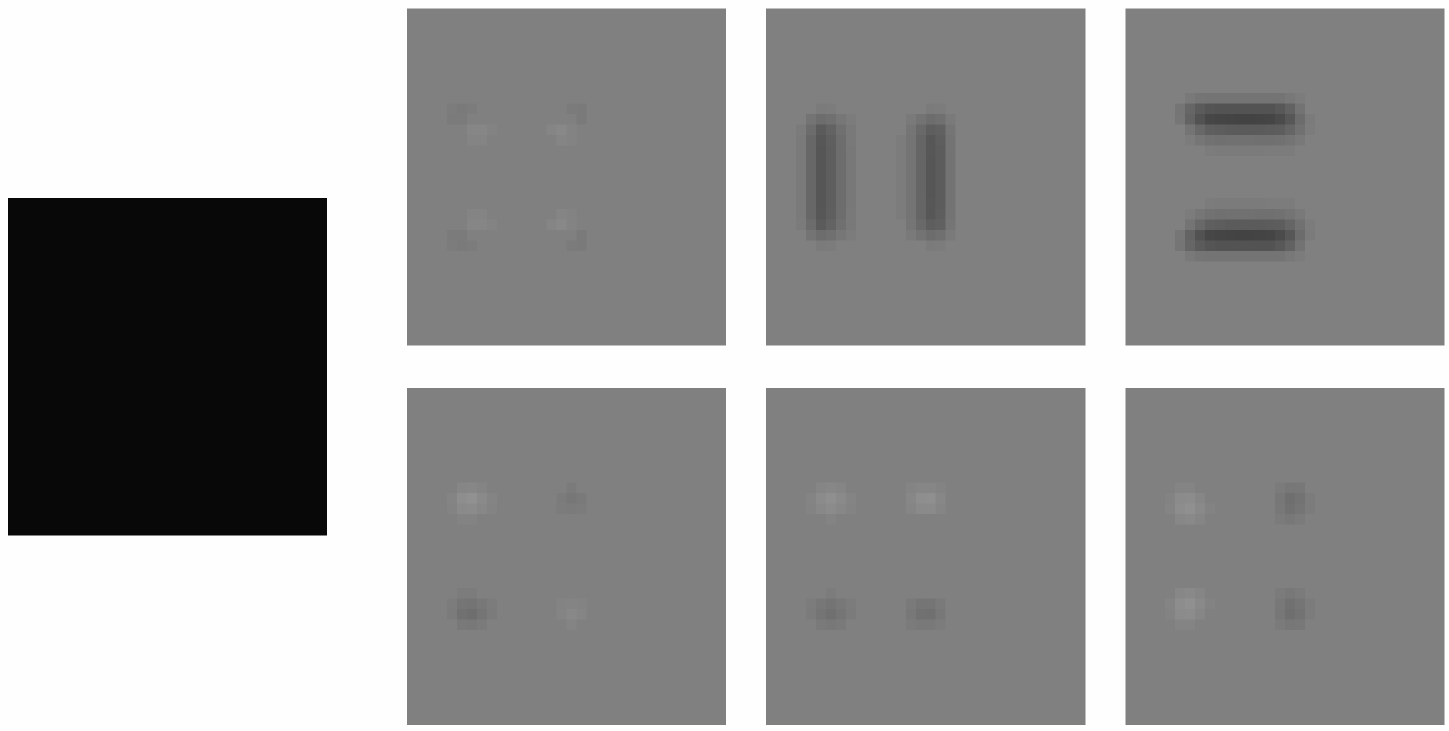
$$R_{3231} = \frac{f_{xy} f_{tt} - f_{xt} f_{yt}}{1 + f_x^2 + f_y^2 + f_t^2}$$

$$R_{3221} = \frac{f_{xy} f_{yt} - f_{yy} f_{xt}}{1 + f_x^2 + f_y^2 + f_t^2}$$

“Stationary” square



Square moves in different directions

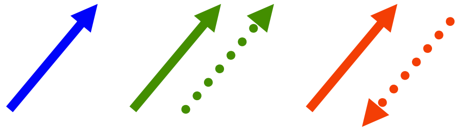


R **components and motion**

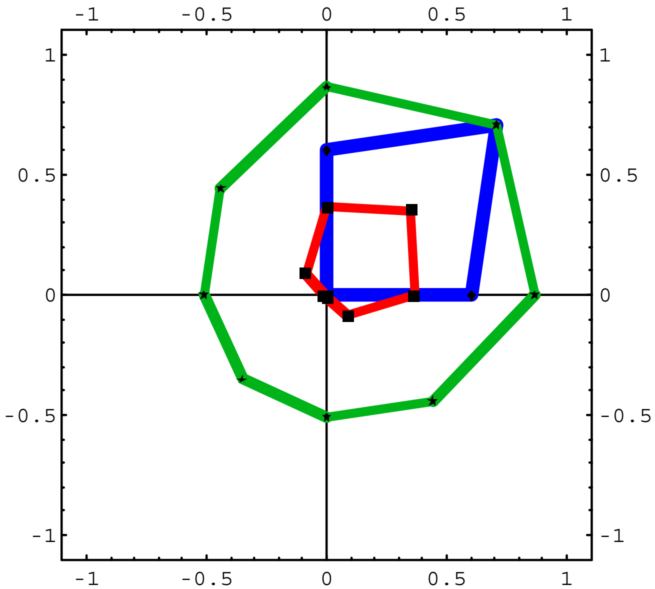
$$f(x,y,t) = f(x - tv_x, y - tv_y) \qquad \mathbf{v} = (v_x, v_y)$$

$$\mathbf{v}_1 = (R_{3221}, -R_{3121}) / R_{2121}$$
$$\mathbf{v}_2 = (R_{3231}, -R_{3131}) / R_{3121}$$
$$\mathbf{v}_3 = (R_{3232}, -R_{3231}) / R_{3221}$$
$$\mathbf{v}_4^2 = (R_{3232}, -R_{3131}) / R_{2121}$$

Multiple motions

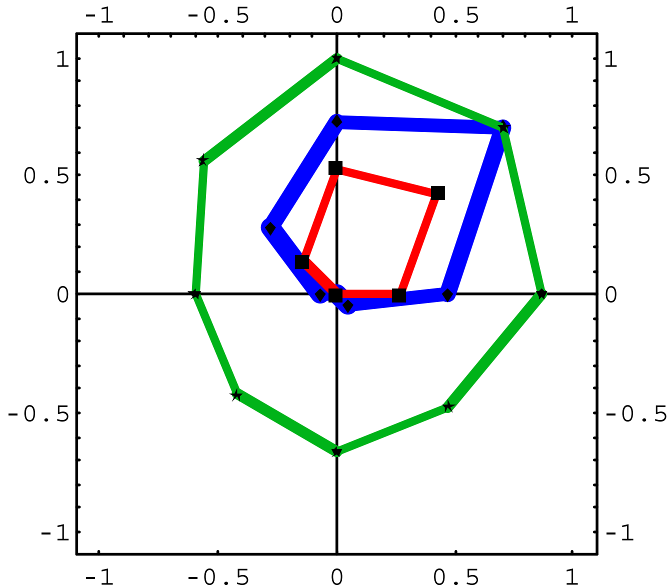


Analytical predictions based on **R** components



Barth & Watson, 2000

Typical MT neuron, macaque monkey



Recanzone, Wurtz, & Schwarz, 1997

Predictive coding

- Only deviations from an hypothesis are encoded
- Complexity of hypotheses increases:
  - intrinsic dimension is 0, 1, 2
    - everything is uniform
    - edges are straight
  - ...
  - the sun is above
  - ...

INB

Nonlinear analysis of multidimensional signals:

LOcal adaptive estimation of COmplex MOTion and ORientation patterns

LOCOMOTOR

http://www.math.uni-bremen.de/zetem/DFG-Schwerpunkt/

Gefördert unter Ba-1176/7

Cicero Mota

Why motion estimation

General motivation

Study the dynamics of image sequences to obtain new solutions and tools in the field of image processing and computer vision.

Computer vision applications that rely on motion estimation

Image compression

Environmental physics: nonlinear dynamics in wind waves

Denoising of medical image sequences: live x-ray, ultrasound, confocal fluorescence microscopy

Robot navigation, intelligent vehicles

Video surveillance

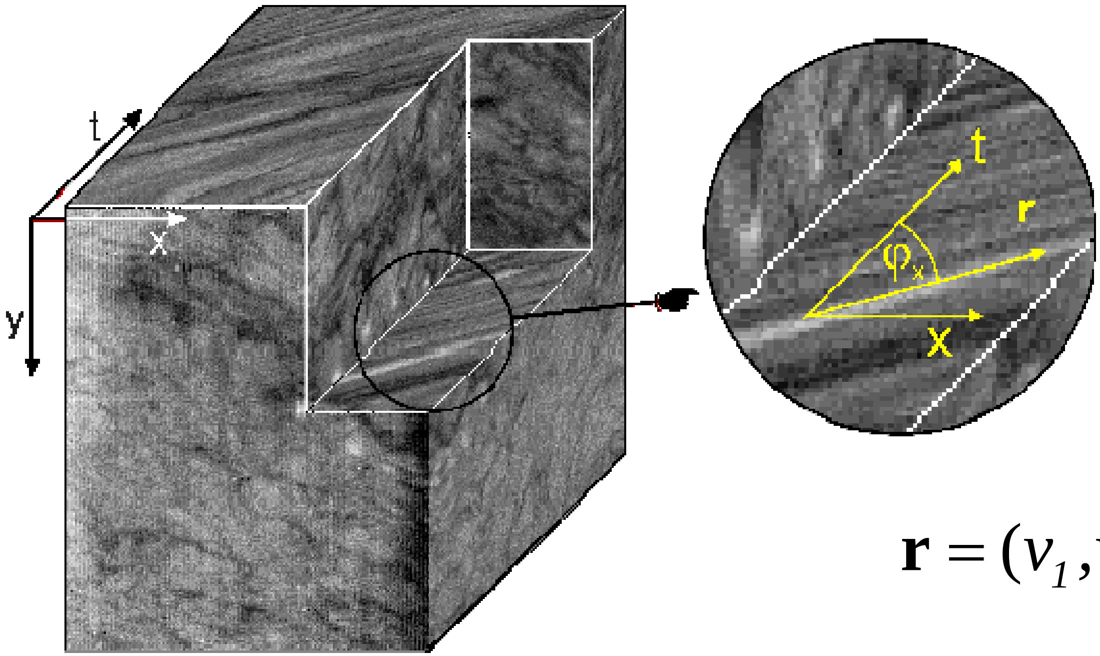
Human-computer interaction

The classical motion model

f(x,y,t) image-sequence intensity

$\mathbf{v} = (v_1, v_2)^T$ motion vector

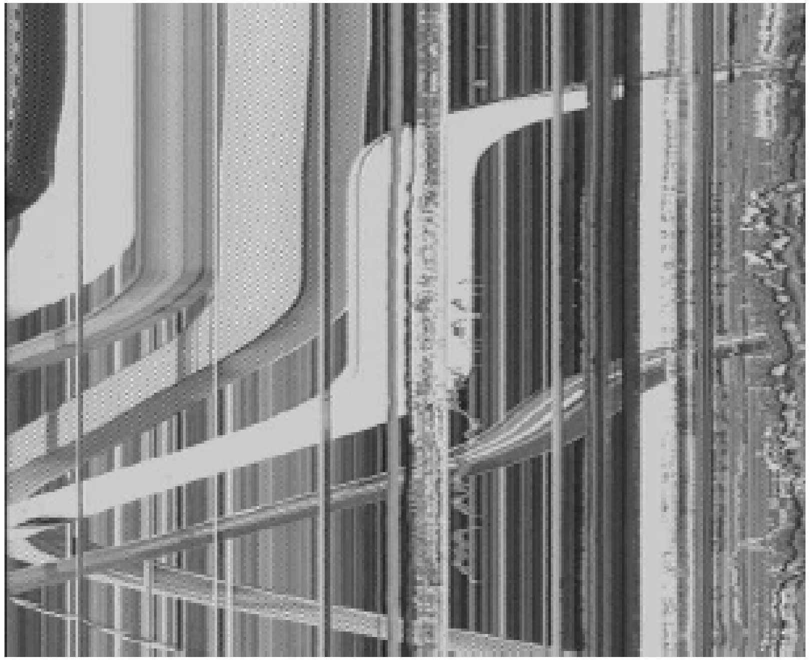
$$v_1 \frac{\partial f}{\partial x} + v_2 \frac{\partial f}{\partial y} + \frac{\partial f}{\partial t} = 0$$



$$\mathbf{r} = (v_1, v_2, 1)$$

Why complex motions?

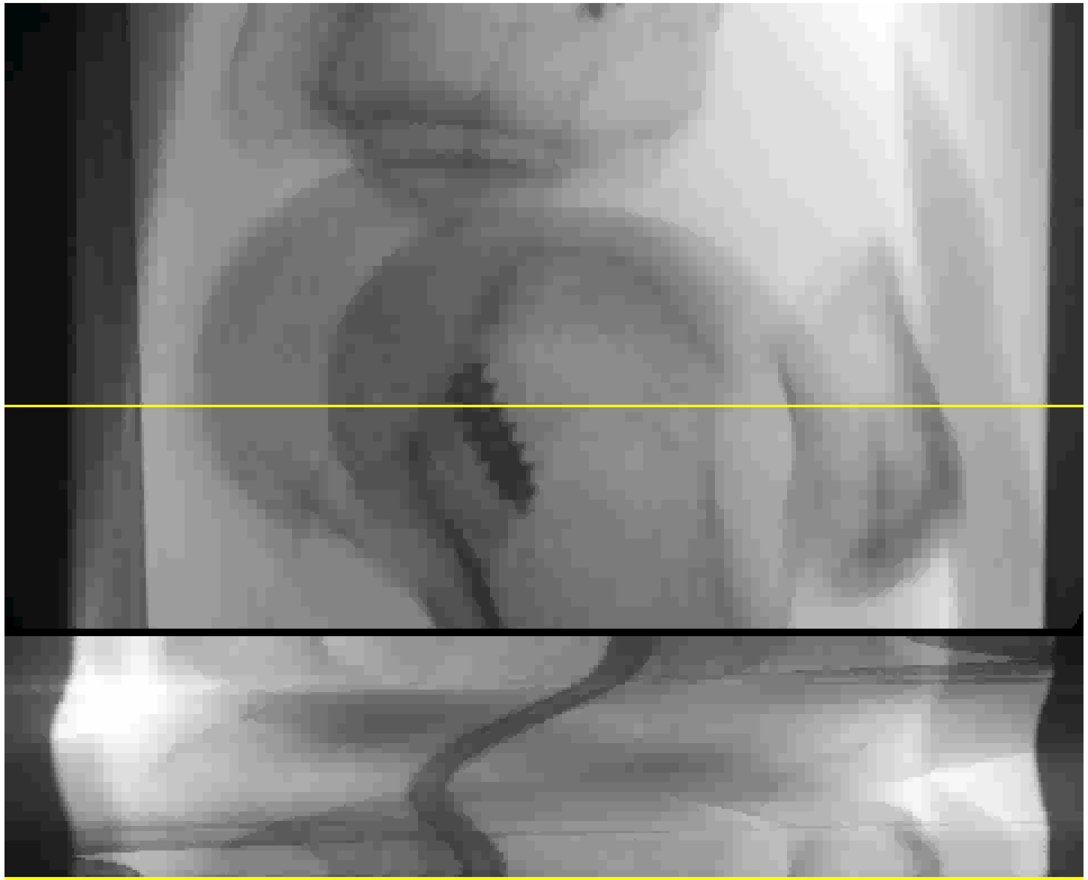
xy image



xt image

Basic problems with natural image sequences: multiple motions, occlusions, noisy data, features at different scales, ...

Transparent motions



Motion model for transparent motions

Optical flow for one motion

f(x,y,t) image sequence, $\mathbf{v} = (v_1,v_2)^T$ motion vector

$$\alpha(\mathrm{v})\mathrm{f} = v_1 \frac{\partial}{\partial \mathrm{x}}\mathrm{f} + v_2 \frac{\partial}{\partial \mathrm{y}}\mathrm{f} + \frac{\partial}{\partial \mathrm{t}}\mathrm{f} = \mathrm{V} \cdot \nabla \mathrm{f} = 0$$

with the derivative operator: $\alpha(\mathrm{v}) = v_1 \frac{\partial}{\partial \mathrm{x}} + v_2 \frac{\partial}{\partial \mathrm{y}} + \frac{\partial}{\partial \mathrm{t}}$

Optical flow for N motions: $\mathrm{f} = g_1(x-\mathbf{v}_1 t)+\ldots+g_N(x-\mathbf{v}_N t)$

$$\boxed{\alpha(\mathrm{v}_1)\alpha(\mathrm{v}_2)\ldots\alpha(\mathrm{v}_N)\mathrm{f} = 0}$$

INB

Mixed-motion parameters

Example for two motions **u, v** : $f(\mathbf{x},t)=g_1(\mathbf{x}-\mathbf{v}\ t)+g_2(\mathbf{x}-\mathbf{u}\ t)$

$$\begin{aligned}\alpha(\mathrm{v})\alpha(\mathrm{u})f &= u_1 v_1 f_{xx} + u_2 v_2 f_{yy} \\ &+ (u_1 v_2 + u_2 v_1) f_{xy} \\ &+ (u_1 + v_1) f_{xt} + (u_2 + v_2) f_{yt} + f_{tt}\end{aligned}$$

We define the *mixed-motion parameters* as:

$$\mathrm{c}_{\mathrm{xx}} = v_1 u_1 \qquad \mathrm{c}_{\mathrm{yy}} = v_2 u_2 \qquad \mathrm{c}_{\mathrm{xy}} = u_1 v_2 + u_2 v_1$$

$$\mathrm{c}_{\mathrm{xt}} = u_1 + v_1 \qquad \mathrm{c}_{\mathrm{yt}} = u_2 + v_2 \qquad \mathrm{c}_{\mathrm{tt}} = 1$$

Solving for the mixed-motion parameters

With the *mixed-motion parameters* we obtain:

$$\boxed{\begin{aligned}\alpha(v)\alpha(u)f &= c_{xx}f_{xx} + c_{yy}f_{yy} + c_{xy}f_{xy} \\ &+ c_{xt}f_{xt} + c_{yt}f_{yt} + c_{tt}f_{tt} \\ &= V \cdot L = 0\end{aligned}}$$

For one motion we had

$$V \cdot \nabla f = 0$$

The above constraint can be used in a number of ways to derive the mixed motion parameters in V from f, e.g. by defining the generalized structure tensor $\boxed{J_N V = 0}$ and solving the system $\boxed{J_N = h * L^T L}$

e.g. $V_i \propto (M_{im}, -M_{i(m-1)}, \ldots, (-1)^m M_{i1})$ $M_{ij}, i = 1, \ldots, m$ are the minors of $\mathbf{J}_N$

INB

Separation of the mixed-motion parameters

We interpret the motion vectors $\mathbf{u}$ and $\mathbf{v}$ as *complex numbers*
($\mathbf{v} = v_1 + i\, v_2$, $\mathbf{u} = u_1 + i\, u_2$) and observe that

$$\mathrm{u\,v} = \mathrm{A}_0 = \mathrm{c}_{xx} - \mathrm{c}_{yy} + i\mathrm{c}_{xy}$$

$$\mathrm{u} + \mathrm{v} = \mathrm{A}_1 = \mathrm{c}_{xt} + i\mathrm{c}_{yt}$$

A_0 and A_1 are homogeneous symmetrical functions of the coordinates $\mathbf{u}$ and $\mathbf{v}$ and therefore by Vieta's rule the coefficient of the complex polynomial

$$\boxed{\mathrm{Q}(z) = (z - \mathrm{u})(z - \mathrm{v}) = z^2 - \mathrm{A}_1 z + \mathrm{A}_0}$$

that has the complex roots $\mathbf{u}$ and $\mathbf{v}$.

In case of N motions

$$\boxed{\mathrm{Q}(z) = z^N - \mathrm{A}_{N-1} z^{N-1} + \cdots + (-1)^N \mathrm{A}_0}$$

Information technology for active perception: *Itap*

Institute for Neuro- and Bioinformatics (*INB*)
University of Lübeck

Partners:
Karl Gegenfurtner
SensoMotoric Instruments GmbH
Siemens

Seeing as an illusion: the door experiment

D. Simons, Harvard.

Seeing as an illusion: change blindness

INB

Demo Kevin O´Regan, Paris.

Seeing as an illusion: change blindness

Basketball count

INB

Visual communication today:
same image but different messages

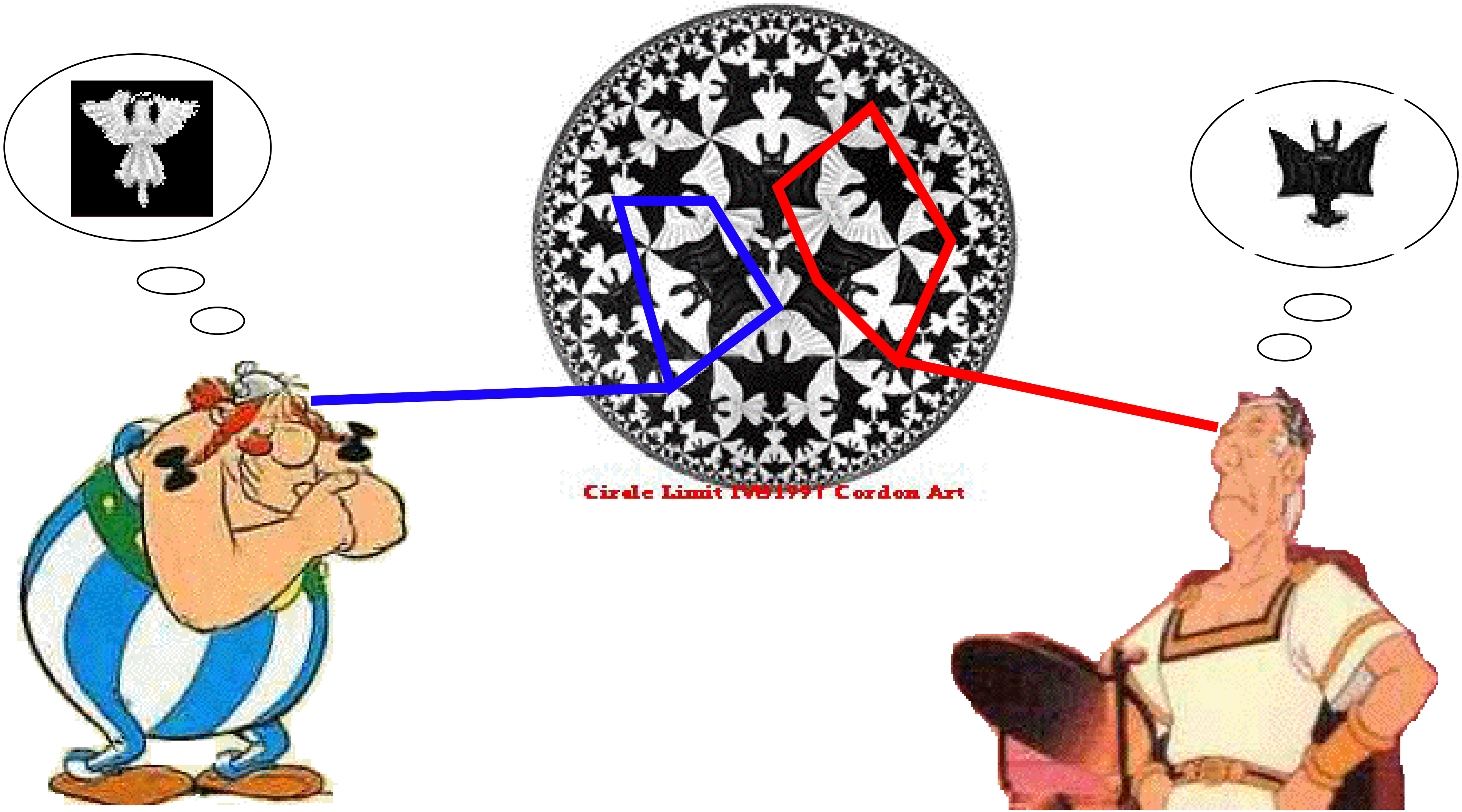


Figure by M. Dorr, INB.

Visual communication today

The message that is conveyed by an image depends very much on the *scan- path*, i.e, the sequence of eye movements that are used to look at an image.

Visual communication systems, however, are based on only the classical image attributes luminance and color.

Itap idea

The *scan path* and the active component of vision must become part of visual communication systems.

Therefore the *scan path* must be

recorded,

processed, and

“displayed“.

Major challenges

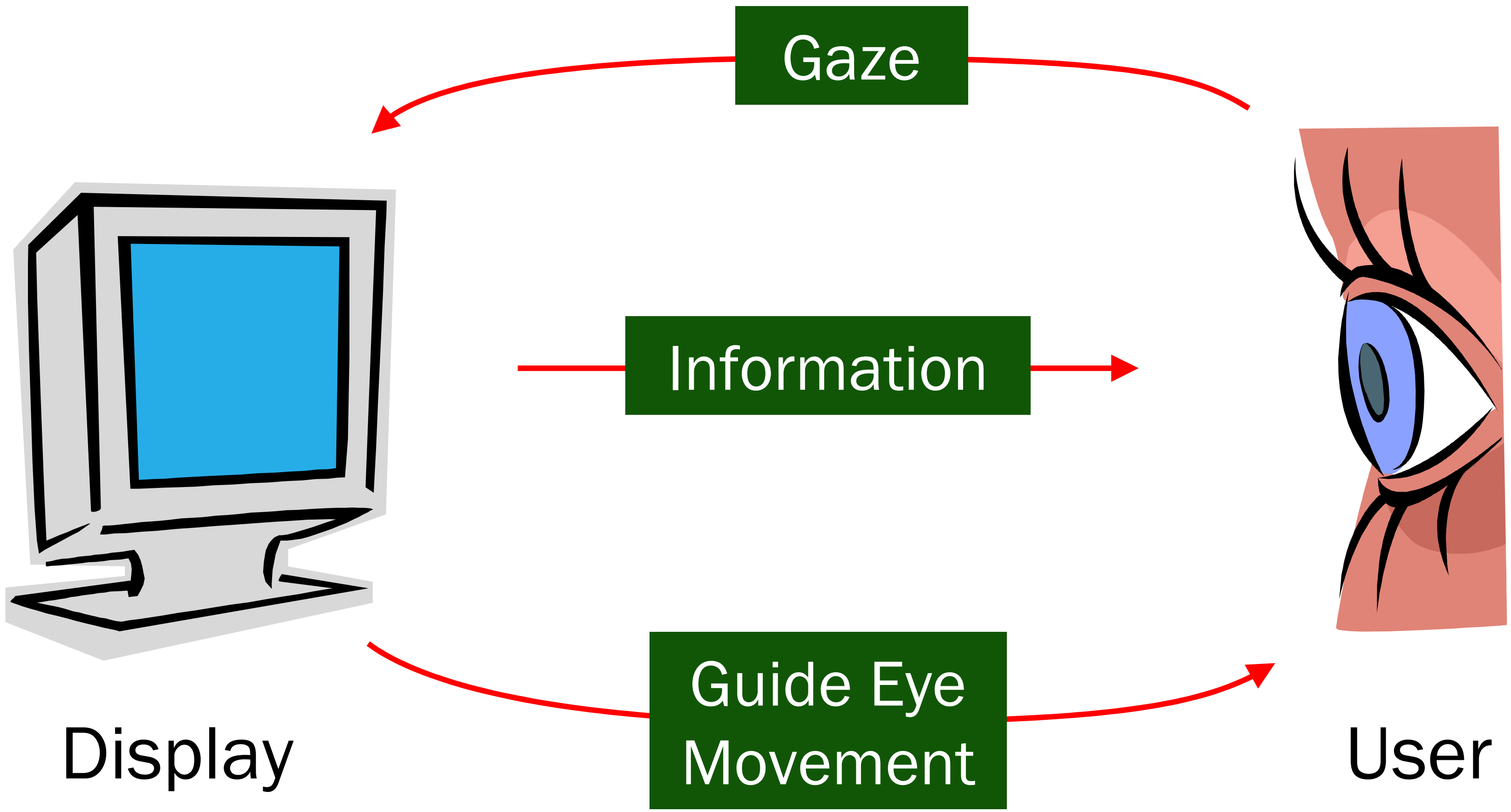
Remote, user-friendly eye tracking

Understanding of eye movements.

Scan-path guidance.

Development of gaze-contingent interactive displays (GCIDs).

Interactive Display



Interactive Display

Guiding of Eye Movements

Automobile

Training

Predict where gaze will fall and modify scene at those locations

⇨ Need saliency measure

Structure Tensor (1)

Compute saliency measure using structure tensor:

$$\mathbf{J} = \mathrm{w} * \begin{pmatrix} f_x^2 & f_x f_y & f_x f_t \\ f_x f_y & f_y^2 & f_y f_t \\ f_x f_t & f_y f_t & f_t^2 \end{pmatrix}$$

spatial smoothing
kernel

partial derivatives of
image-intensity function
$f(x, y, t)$

Structure Tensor (2)

Three saliency measures derived from structure tensor:

$$H = \frac{1}{3}\mathrm{trace}(\mathbf{J}) = \lambda_1 + \lambda_2 + \lambda_3 \quad (\geq \text{i1D})$$

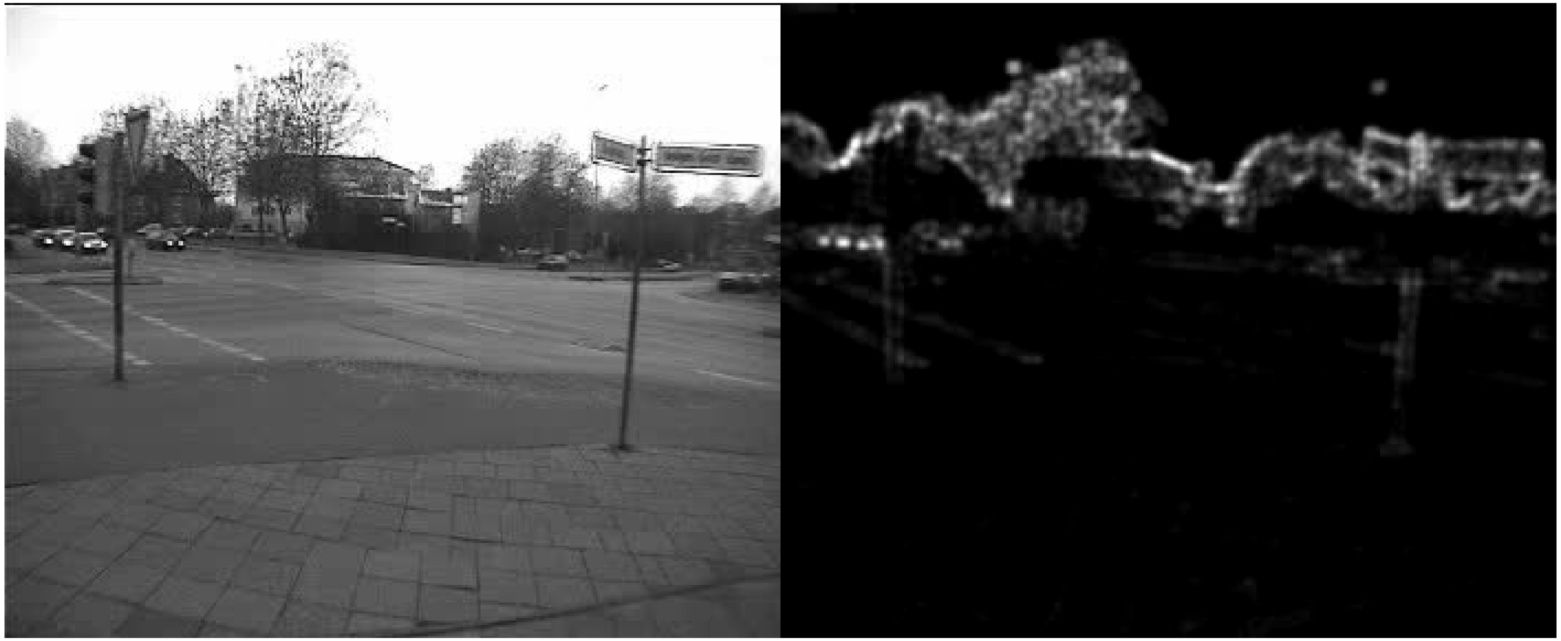
$$S = |\mathrm{M}_{11}| + |\mathrm{M}_{22}| + |\mathrm{M}_{33}| = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3 \quad (\geq \text{i2D})$$

$$K = |\mathbf{J}| = \lambda_1\lambda_2\lambda_3 \quad (\text{i3D})$$

$\lambda_1, \lambda_2, \lambda_3$: Eigenvalues of the structure tensor

$\mathrm{M}_{11}, \mathrm{M}_{22}, \mathrm{M}_{33}$: Minors of the structure tensor

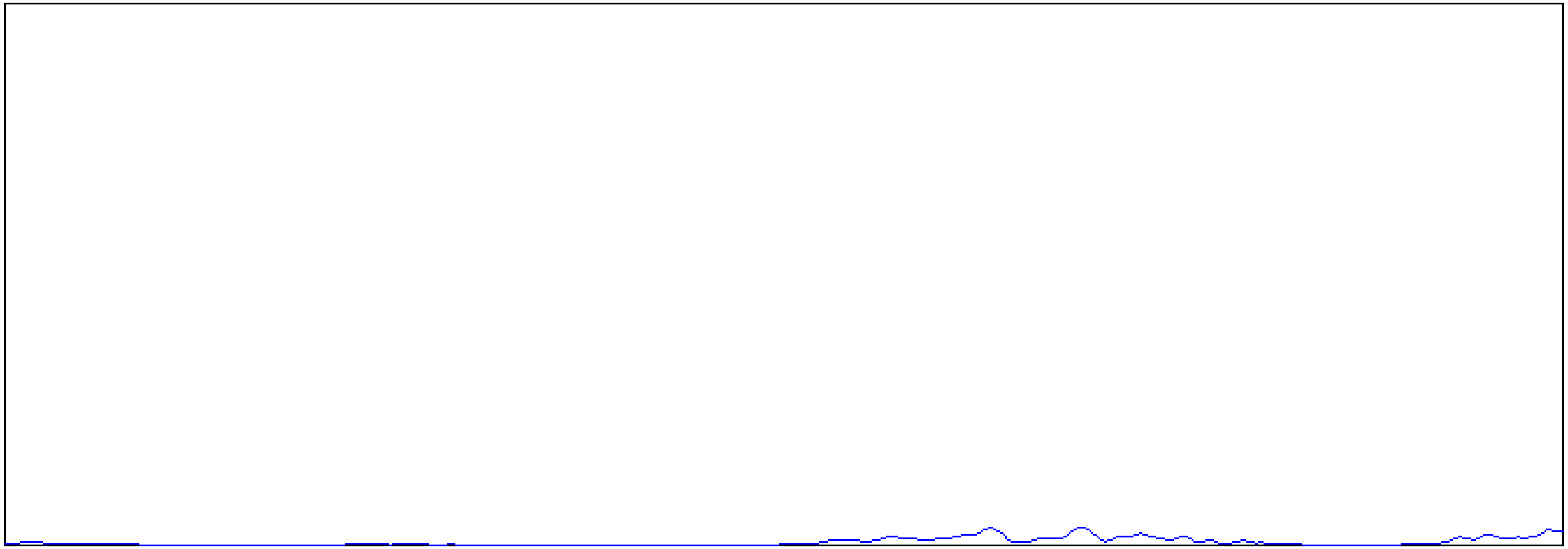
Saliency



Salient Location Extraction

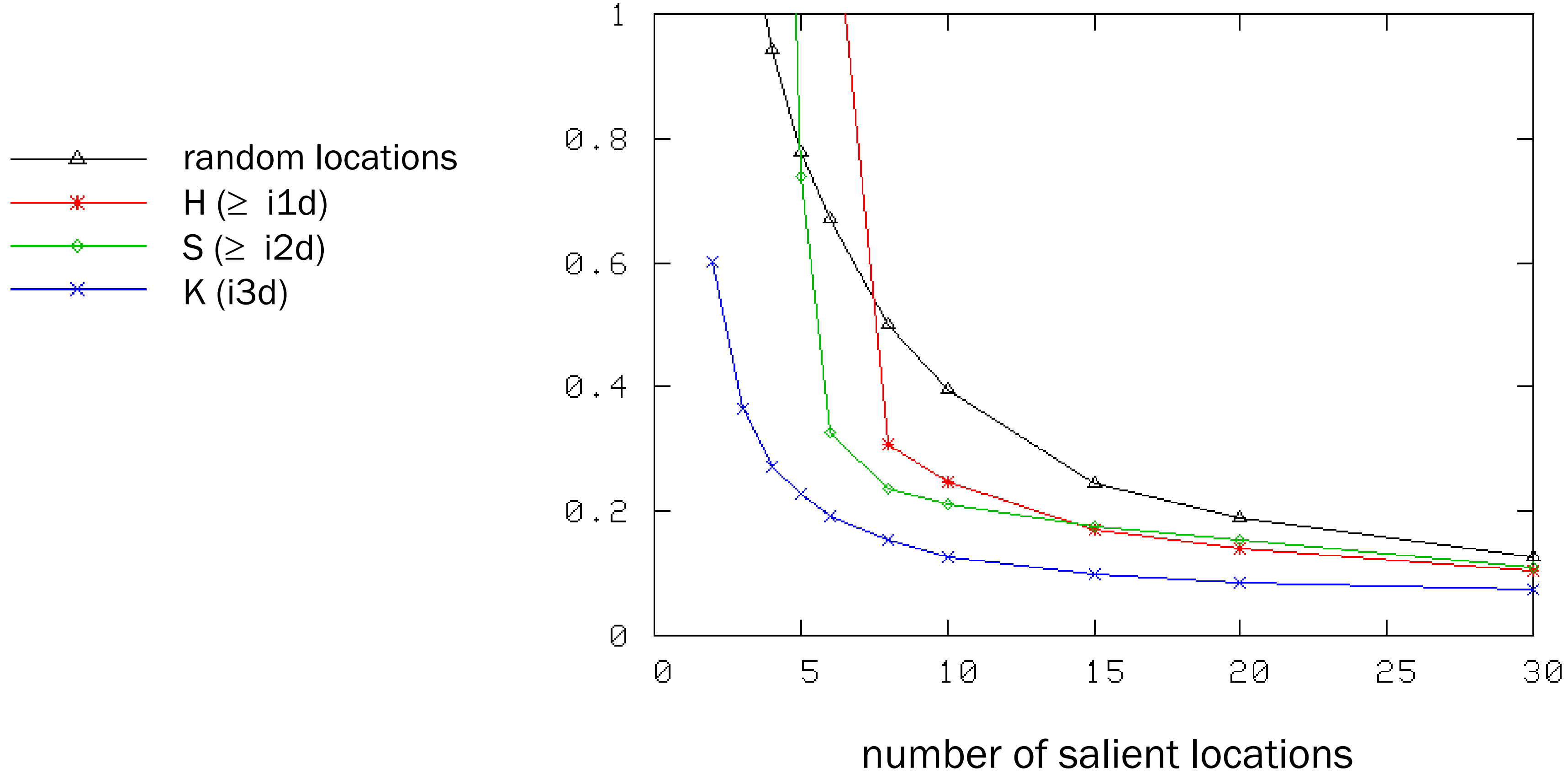
Find maximum

Attenuate using Gaussian



Prediction results

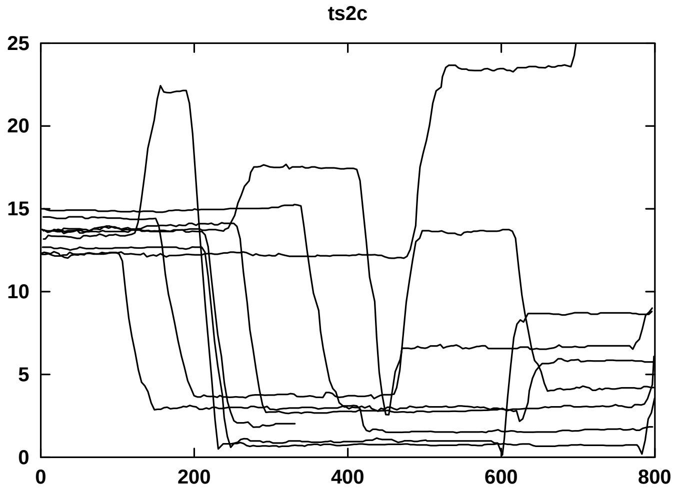
Average relative distance of saccade target to closest salient location



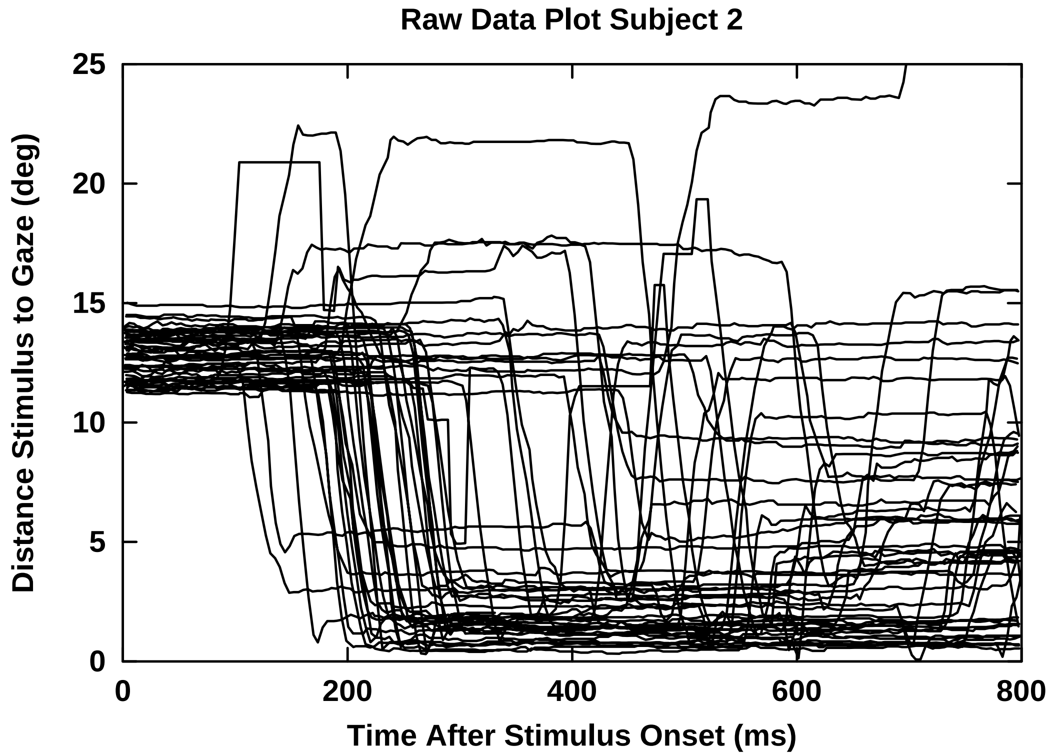
Results for 6 subjects with 3 runs each.

INB

Guidance results



Results for one movie



Results for all movies (4)

Applications

Communication systems

will be defined not only by brightness and colour, but will be augmented with a recommendation of what to see, of how to view the images. Thereby Itap will start to improve communication by considering the active role of the observer.

Augmented-vision systems

Attention is directed towards objects or features that have been detected by a computer-vision system.

Vision aids

COGAIN network EC-FP6

COGAIN is a network of excellence that aims at helping disabled people to communicate more effectively.

http://www.cogain.org/

Current automotive applications

Fatigue measurement

- Blink measurement
- PERCLOS measurements will be soon enforced by law in the US.

Head tracking for airbag control

Driver identification

Intelligent airbags

Problems:

- Deployment with kids and OoP (harm)
- Useless deployment (cost)

Solution: **video-based** control

OoP: Out of Position

SMI

INB

Intelligent airbags: OoP system

SMI

LAPI

INB

Blink and fatigue measurement

bvisual & Liamis

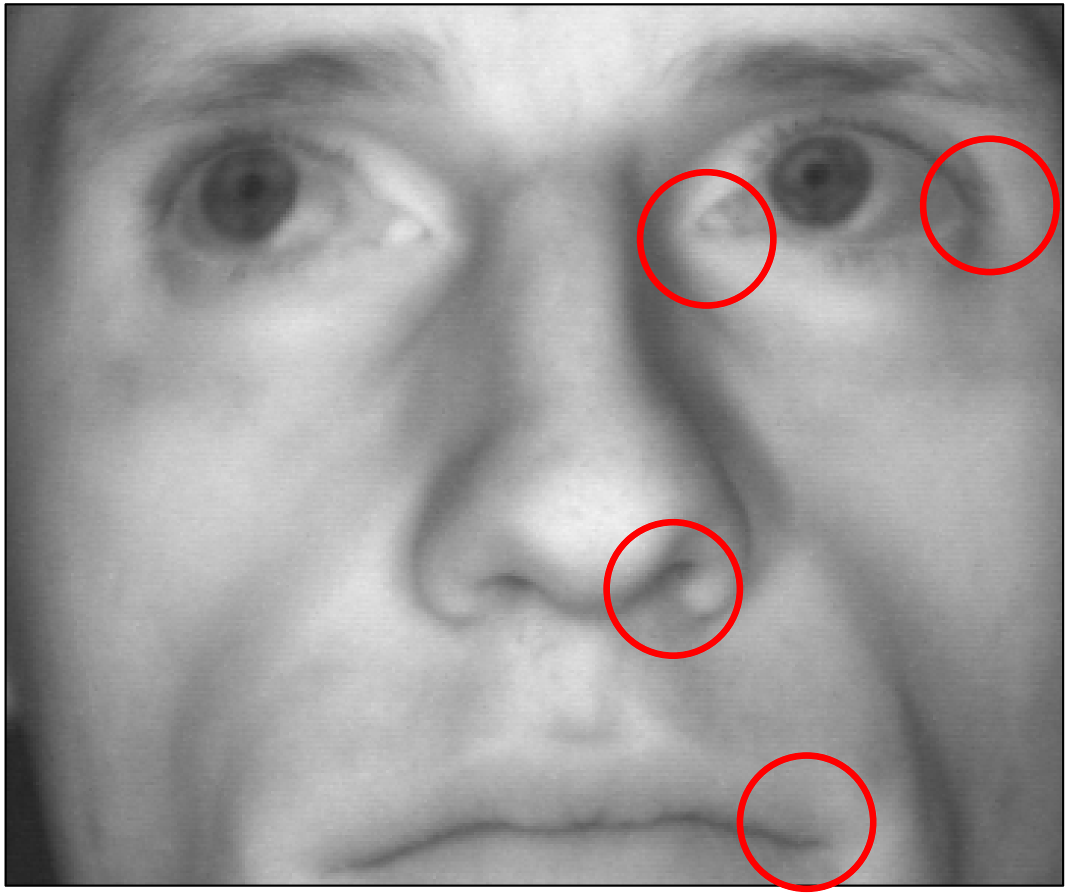
SMI

BMW

LAPI

INB

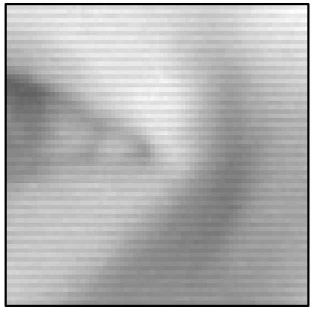
Tracking of humans is still difficult



Problems:

- Initialization
- Drift
- Lost features
- Accuracy

Where is the eye corner?



Coordination of subsystems

Subsystems are limited in scope, perspective and information. They need to cooperate in a given situation to form an integrated functional whole, taking into account, and making mutually consistent, all relevant information.

Organic systems reach global certainty as a consistent meshwork of many subsystems, although individually these may have started out with large uncertainty.

Organic Computing

http://www.organic-computing.org/

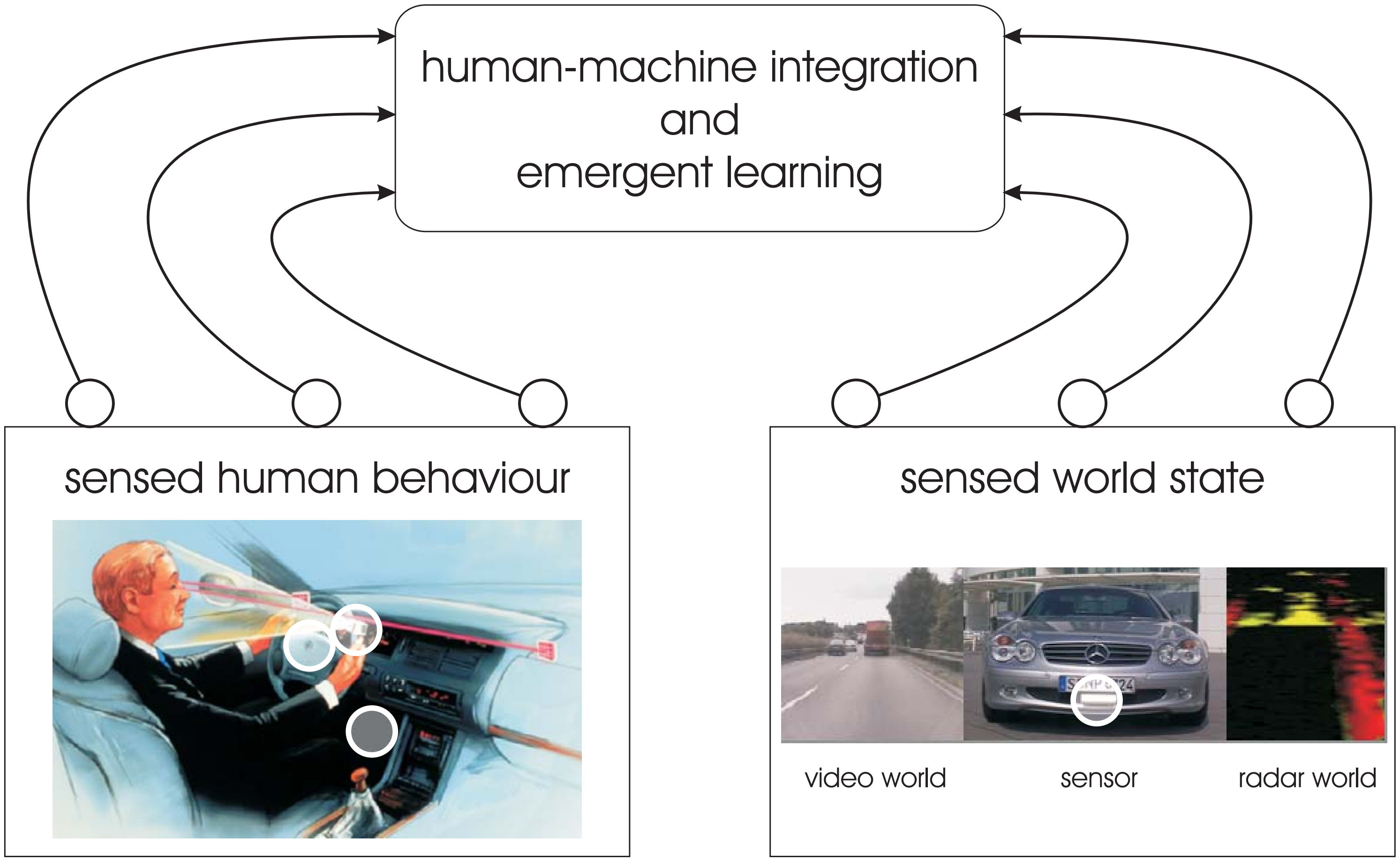
Organic Computing is a call to arms for a concerted intellectual effort towards a Science of Organization. Ultimately, it is a quest for understanding the complexity of Life.

http://www.organic-computing.de/

Ein „organischer Computer“ ist definiert als ein selbstorganisierendes System, das sich den jeweiligen Umgebungsbedürfnissen dynamisch anpasst. Organische Computer sind selbst-konfigurierend, selbst-optimierend, selbst-heilend und selbst-schützend.

DFG Schwerpunktprogramm 1183: "Organic Computing"

Organic Computing (INB)



The information sensed from the human is used to structure the process of selforganization. By assuming that all the relevant properties of the world will be somehow reflected in the behavior of the human, the integrated system will be able to build up knowledge about relevant states of the world.

Summary

- Early and mid-level vision leads to a sparse and increasingly efficient encoding of the visual input
- The encoding can be explained by the concept of intrinsic dimension
- Vision is an illusion and this leads to, e.g., communication problems
- *Itap* will improve visual communication by helping people see what they are meant to see; it will lead to new kinds of vision aids
- Computer vision still faces many challenging problems
- Organic Computing is a new attempt to handle complexity by using biology as inspiration

Acknowledgements

- Cicero Mota, Ingo Stuke, Til Aach
- Martin Böhme, Michael Dorr
- Thomas Martinetz

- Support from BMBF, DFG, SMI

INB