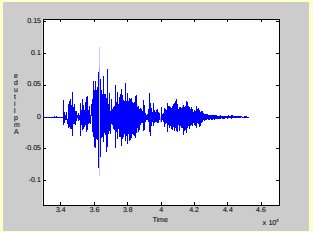
Signal, Image, Perception

Anne Guérin-Dugué
et C.Berrut
CLIPS – IMAG
Grenoble

Contents

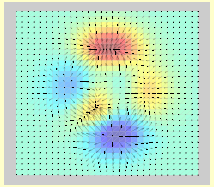
Signal

- Physical support of information
  - Temporal evolution
    - Physical sensors,
    - Speech, ...



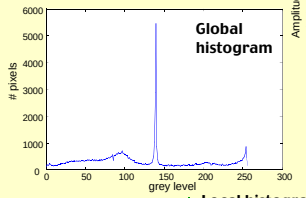
Image

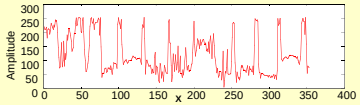
- Physical support of information
  - Surface, Spatial evolution
    - Visual scene
    - Representation of a physical phenomenon

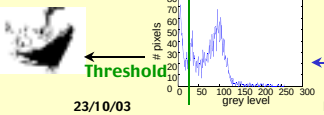


Image

- Grey level

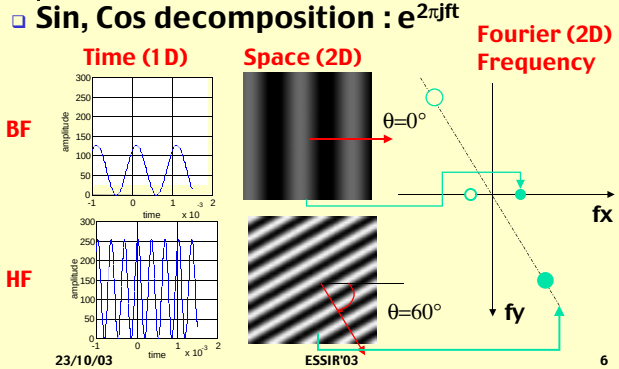






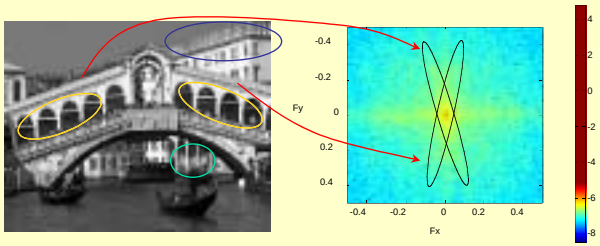
Tool : Fourier Transform

- Sin, Cos decomposition : $e^{2\pi jft}$



Tool : Fourier Transform

- Illustration



BF MF HF

Visual Perception

- « Process of acquiring knowledge about environmental objects and events by extracting information from light they emit or reflect »

S.E. Palmer, Vision Science, 2nd ed., 2002

Broad applications

- Search by association through an interactive choice
  - Browsing through a large set of images
  - Ill defined specific aim at start
- Search for one specific image (target)
  - Start : paper copy, a part, ...
- Search a category → Predefined labels, or Emergent clusters ?
  - to illustrate a document, a story
    - Semantic point of view
    - Esthetic point of view

A.W.M. Smeulders et al, I3E Pami, 22(12), 2000

Image domain (1/2)

- Narrow vs broad domain in image retrieval

	Narrow	Broad
Variance of content	Low	High
Source of knowledge	Specific	Generic
Semantics	Homogeneous	Heterogeneous
Ground truth	Likely	Unlikely
Scene and sensor	Possibly controlled	Unknow
Aimed application	Specific	Generic
Type of application	Professional	Public

Image domain (2/2)

- Narrow vs broad domain in image retrieval

	Narrow	Broad
Tools	Model-driven, specific invariants	Perceptual, cultural, general invariants
Interactivity	Limited	Pervasive, iterative
Evaluation	Quantitative	Qualitative
System architecture	Tailored database-driven	Modular interaction-driven
Size	Medium	Large to very large
A source of inspiration	Object recognition	Information retrieval

Content Based Image Retrieval

Computational vision → CBIR

Information Retrieval → CBIR

Visual Perception → CBIR

Natural images ?

- Coming from our environment
- We see it every days
- Our visual cortex is « optimized » for such visual stimuli
- « Natural images » represent for us a very large set but are in fact a « very sparse subset of all possible images »

➔ Real-world images

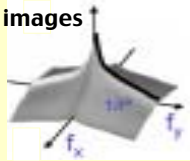
Statistics of Real-World Images

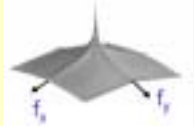
- Average Power Spectrum
  - Second order correlation function

Man-made scenes

Natural scenes

Average on more than 3500 images





Statistics of Real-World Images

- Average Power Spectrum
  - Model of scaling invariance

$S(f) = 1/f^a$ — Isotropic — First study : Deruigin, 1978

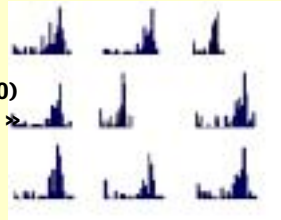
$S(f) = A(\theta)/f^{a(\theta)}$ — Anisotropic — Agreement with neurophysiological data : [De Valois, De Valois, 1988]

		Horizontal	Oblique	Vertical
Natural	a	1.98 (0.58)	2.02 (0.53)	2.22 (0.55)
	A	0.96 (0.40)	0.86 (0.38)	1 (0.35)
Man-Made	a	1.83 (0.58)	2.37 (0.45)	2.07 (0.52)
	A	1 (0.32)	0.49 (0.24)	0.88 (0.29)

Statistics of Real-World Images

- Grey level local histograms

L(x,y)=ln(I(x,y)/I0)
« Log_Contrast »



Similar shape, Non gaussian, Heavy tails.

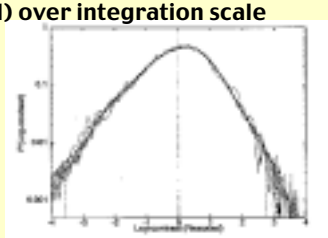
Statistics of Real-World Images

- Spatial correlations
  - Average log(grey level) over integration scale

N=2

L(x,y)=ln(I(x,y)/I0)
« Log_Contrast »

Integration scale



Scaling of log-contrast histograms over scale 1, 2, 4, 8, 16 and 32

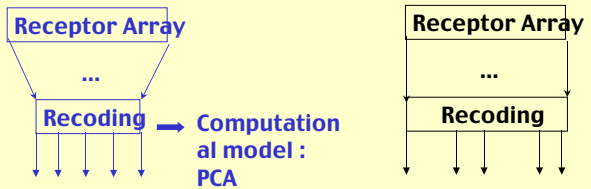
Statistics of Real-World Images

- Summary
  - Non gaussian distribution
  - Correlation between pixels
  - Scaling invariance
    - Power spectrum : $1/f^a$ law
    - High order correlation : integration scale invariant « shape »
- On filtered images
  - Long tailed histograms

Consequence of a sparse coding — D.J. Field, Neural Comput., 1994

Different type of coding

- Compact vs Sparse dispersed coding



Compact : Minimize the number of detectors

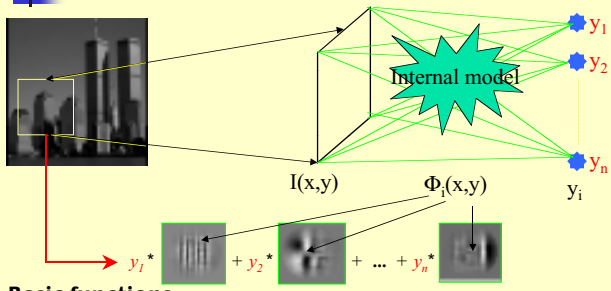
Sparse : Relatively few coding units are active at any one time

Dispersed : in average, all the available units will contribute equally to coding.

D.J. Field, Neural Comput., 1994

B. Willmore et al, Perception, 29, 2000

Computational models



Basic functions decomposition :

$$I(x,y) \approx \sum_{i=1}^{N} y_i . \phi_i(x,y)$$

Computational models

- Sensory coding (y_i): Exploiting the structure of natural images
  - Reduction redundancy [Barlow, 1961, 1983]
  - Sparse decomposition [Olshausen, Field-1996]
  - Independent Compone Analysis [van Hateren, van der Schaaf- 199

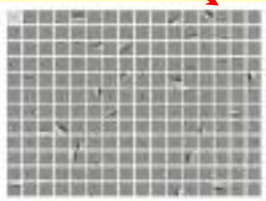
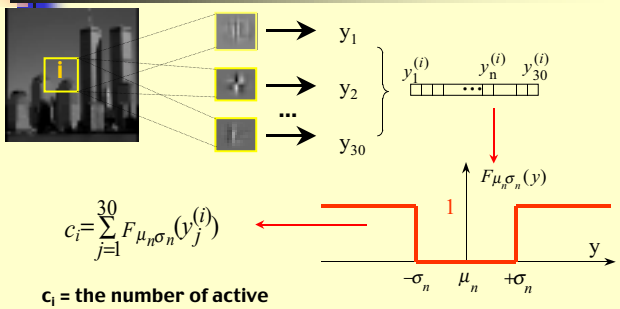


Illustration of sparse coding (1/2)



$$c_i = \sum_{j=1}^{30} F_{\mu_n \sigma_n}(y_j^{(i)})$$

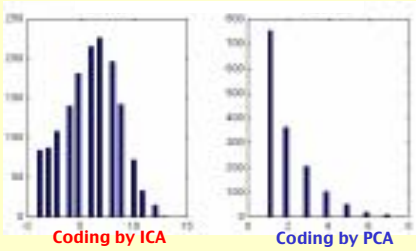
c_i = the number of active cells encoding the patch i.

Active cell if great valu

H. Le Borgne, A. Guérin-Dugué, Valgo, 2000

Illustration of sparse coding (2/2)

Histogram of the number of active cells (c_i) coding a patch.
Data set : collection of cityscapes,
#units : 30

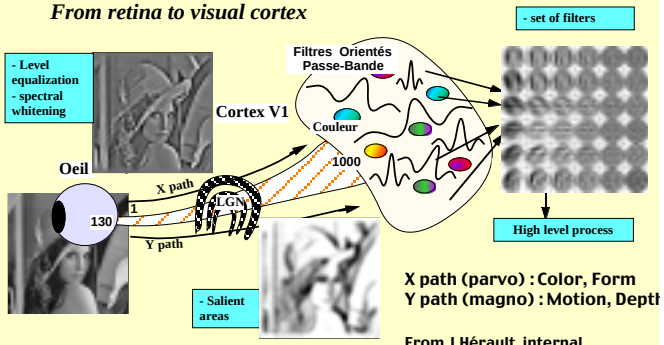


PCA : value closed to the average
➔ few active units
Compact coding : minimize the number of units
ICA : more active units which code great values
Sparse coding : minimize the number of active cells
Increase SNR ratio

H. Le Borgne, A. Guérin-Dugué, Valgo, 2000

Visual Pathways (1/3)

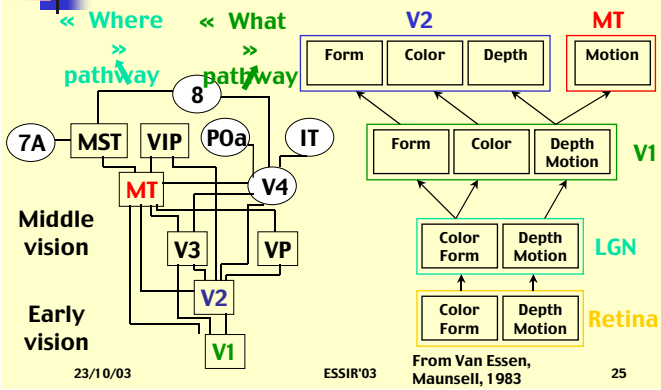
From retina to visual cortex



X path (parvo) : Color, Form
Y path (magno) : Motion, Depth

From J.Hérault, internal Scopie report, 2002

Visual Pathways (2/3)



From Van Essen, Maunsell, 1983

Visual Pathways (3/3)

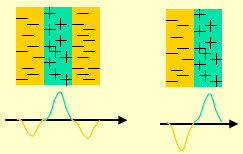
- Summary
  - Magno and Parvo Cells from the LGN structure

Color sensitivity	LOW	HIGH
Contrast sensitivity	HIGH	LOW
Spatial resolution	LOW	HIGH
Temporal resolution	FAST	SLOW
Receptive field	LARGE	SMALL

  - Early vision : V1, V2
    - Narrow receptive fields
    - Local operation (filtering) => Map of local features
      - Orientation, Edges elements, Depth, local, ...
  - Middle vision : V3, V4
    - Spatio-temporal Integration : Large receptive fields
      - Boundaries, edges completion, surface perception, ...

Line Edge Detector Theory

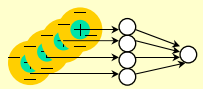
- Hubel, Wiesel 1959
  - V1 : Simple cells as line and edge detectors
  - Simple ➔ Complex ➔ Hypercomplex cells



Line detector Edge detector

(Idem for the inverse contrast)

- Assumption : Oriented pattern by spatial combination

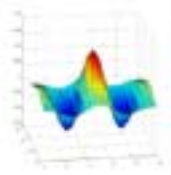


Spatial Frequency theory

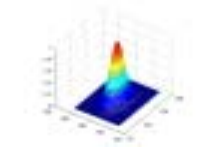
- Primitives in V1
  - Spatially extended patterns = sinusoïdal gratings
  - Sinusoïdal grating modulated by a gaussian function

Filtering model : 2D Gabor filter = pass-band oriented filter

Spatial domain



Frequential domain

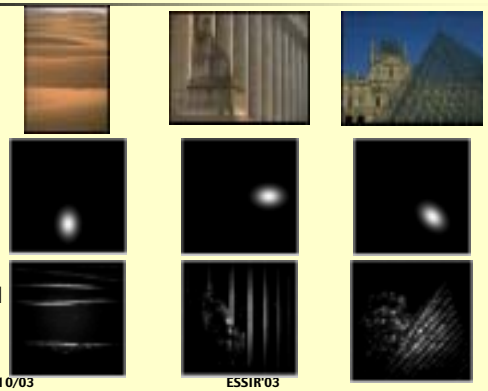


Illustration

Image

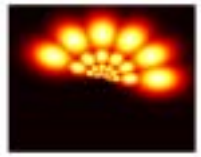
FT Gabor

Filtered Image



Spatial Frequency theory

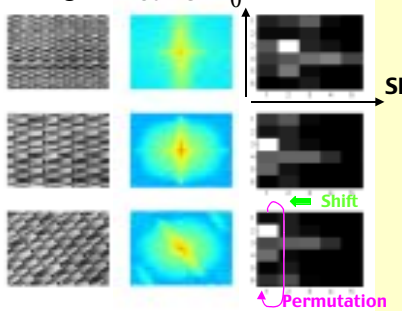
- Spatial frequency chanels
  - From LF to HF
  - All orientations (more at 0° and 90°)
  - Local interaction between cells (frequency, orientation)
  - Adaptation mechanism



- Computational model
  - Set of Gabor wavelets
  - Very popular approach for 30 years
  - Texture segmentation → Scene interpretation

Texture Classification

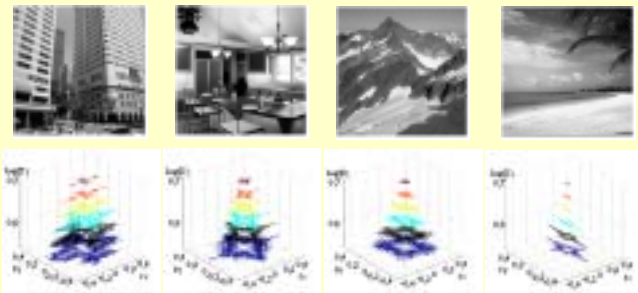
Image | Fourier T. | Gabor Energy | 30 Gabor filters



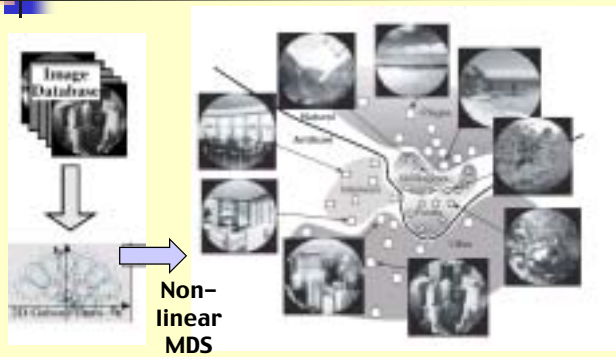
A texture : A vector into a multidimensional space (here 30)

Invariant to scale and rotation : Choice of a specific distance

Scene Interpretation (1/3)



Scene interpretation (2/3)



[Hérault, Oliva, Guérin-Dugué, ESANN'97]

Scene interpretation (3/3)

- Computational models
  - Gabor energy coefficients [Hérault, Oliva, Guérin-Dugué, ESANN'97]
  - Local Orientation [Guérin-Dugué, Oliva, PRL,2000]
  - Edge, DCT coef, Color [Vailaya, Jain, PR, 98]
- Questions ?
  - Which visual categories ?
  - Which semantic levels ?
  - How combine the « Gabor » channels ?

Psychophysical experiments

Visual categories

Psychophysical experiments using a collection of images

Psychophysical experiments

- Objectives
  - Understanding the wide semantic categories driving our visual perception
  - Analyzing human judgment about similarities between images
  - Modeling user similarities with system
  - Finding correlations between between the high-level semantics and low-level descriptors automatically extracted from the

[Rogowitz et al, Perceptual Image Similarity Experiments, SPIE, San José,

Example 1 (1/3)

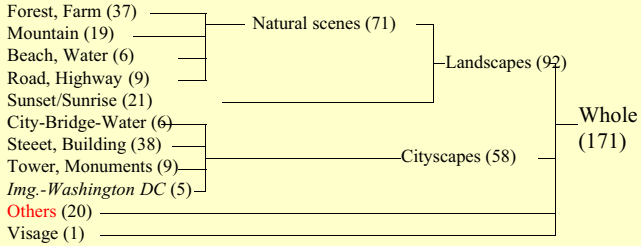
- Context : 171 images et 8 human subjects
- Objectives
  - To organize images into clusters
  - To explain and name each cluster
- Results
  - 1 to 2 hours per subject
  - Number of clusters : 7 to 17
  - Cluster coherence : variable ; more or less divided

Example 1 (2/3)

- Data analysis : between class dissimilarities
  - d(i,j) : +1 if the subject doesn't set i and j into the same cluster
    - $0 \leq d(i,j) \leq 8$ (number of subjects)
- D[i,j] = input matrix of a ascendant hierarchical classification

Example 1 (3/3)

- Results



Others : cityscapes or landscapes : panoramic view

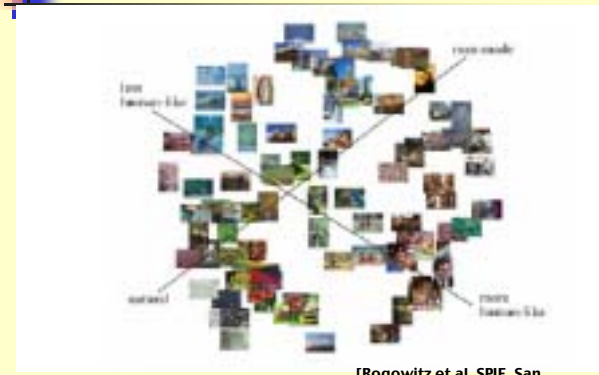
Example 2 (1/3)

- Objective :
  - To Build metric space in which images would be set according our visual
- 2 kinds of experiments
  - « Table scaling Experiment » (9 subjects)
  - « Computer scaling Experiment » (15 subjects)
- Data set : 97 images (color)

«Table Scaling Experiment» (2/3)

- Objective :
  - To physically organize all the images on a table
- Results :
  - Duration : 30 à 45 min per subject
  - Distance matrix :
    - d(i,j), with d(i,j) = Euclidean distance on the table
  - Average matrix over all the subjects (8) : input matrix of a linear MDS algorithm

Result (3/3)



Example 3 : Computer scaling Experiment (1/3)

Objective :
To capture perceptive similarities (Pre-attentive perception)

Task:
Select the closest image from a target (less than 1s) Estimate the perceptive

Database : 105 images –N&B

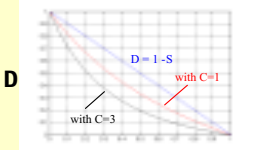
Example 3 (2/3)

- 48 subjects in order to have 12 full similarity matrices → average matrix

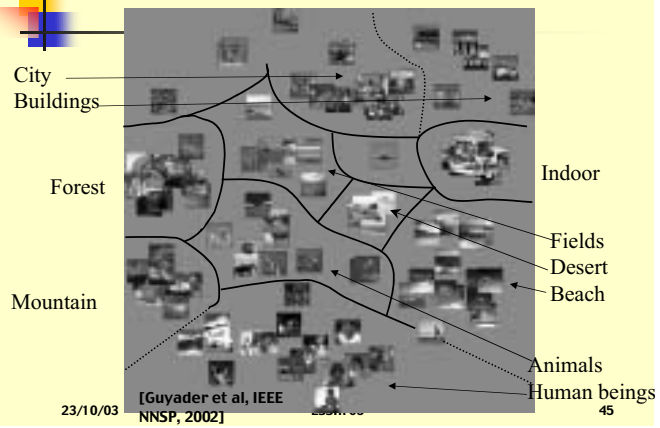
$$S(i,j) = \frac{S_1(i,j) + \frac{1}{8} S_2(i,j) + \frac{1}{27} S_3(i,j) + \frac{1}{64} S_4(i,j)}{(1+8+27+64)}$$

- Compute a distance matrix from the similarity judgment

$$D(i,j) ? \frac{\frac{1}{? ? S_T(i,j)?^{C}} ? 2^{C}}{1 ? 2^{? C}}$$



Example 3 :Non linear MDS (3/3)



Visual categories

- Set of categories :
  Portraits, People outdoors, People indoors, Outdoor scenes with people, Crowds of people, Cityscapes, Outdoor architecture, Technoscenes, Objects indoors, Waterscapes with human influence, Landscapes with human influence, Waterscapes, Landscapes with mountains, Sky/Clouds, Winter/Snow, Green landscapes, Landscapes with fields and foliage, Plants, Animals, Textures

[Mojsilovic, Rogowitz, Capturing image semantics with low-level descriptors, ICIP'2001]

Computational Model

- Features
  - From computational vision (edge, color, region, ...)
- Proposition for a metric from the definition of a prototype for each semantic category
  - Π : « required features »

[Mojsilovic, Rogowitz, Capturing image semantics with low-level descriptors, ICIP'2001]

« Where » Pathway

- Computational Model
  - Saliency map
    - Bottom-up pathway from low level features
    - Selection by local maxima
    - Temporal exploration : sorting maxima
  - Evidence for top-down pathway
    - Global normalization
    - Context learning
- Experimental data
  - Analyzing Eye Movements with an « Eye tracker »

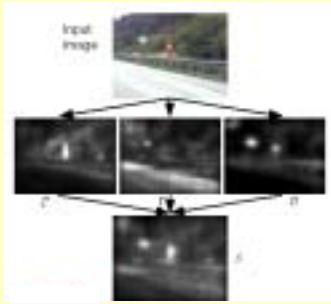
Saliency map : Bottom-Up (1/2)

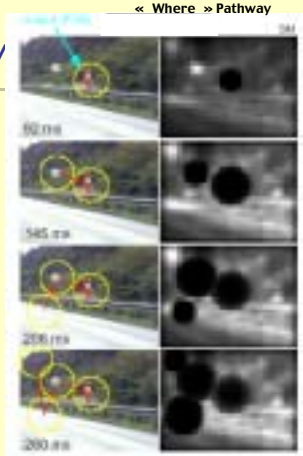
- Itti's Model

Bottom-up :

From Itti, Koch, Niebur, A model for saliency-based visual attention for rapid scene analysis, PAMI, 20(11), 1998

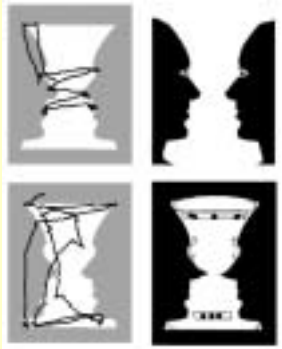
Bottom-Up (2/





ScanPath Theory

- Illustration of the « Scanpath theory »
  - Interaction « Bottom-Up and Top-Down pathways

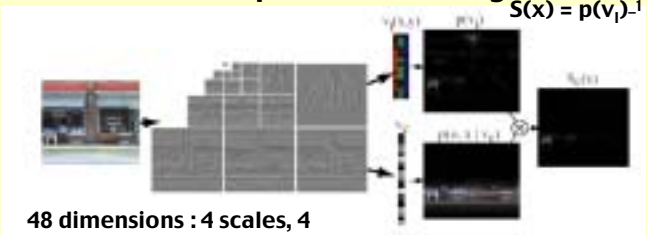


From Stark et al, Representation of human vision in the brain: How does human perception recognize images?, J.E.I, 10(1), 123-151, January 2001.

Saliency map : Top-Down control (1/2)

- Model with top-down learning

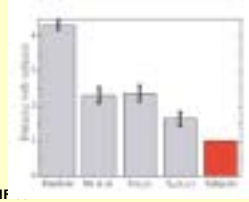
$S(x) = p(v_i)^{-1}$



48 dimensions : 4 scales, 4 orientations, 3 color chanels

From Oliva, Torralba, Castelhano, Henderson, Top-Down Control of Visual Attention in Object Detection, ICIP'03, Spain

Saliency map : Top-Down control (2/2)



From Oliva, Torralba, Castelhano, Henderson, Top-Down Control of Visual Attention in Object Detection, ICIP'03, Spain

Conclusion, Perspectives (1/2)

- Visual cortex : wide source of inspiration for computational vision
  - Efficiency, Robustness, ... for recognition
- Different theories
  - Scenes, Objects, Faces : Frequency, Primitive, Associative memory
- Focus on the frequencial model
  - Scene interpretation

Conclusion, Perspectives (2/2)

- For CBIR
  - A lot of research on separate bottom-up pathways (color, texture, shape)
- Challenge
  - Top-down pathways and feedback connections