Where is the bark, the tree and the forest, mathematical foundations of multi-scale analysis

What we observe is not nature itself, but nature exposed to our method of questioning.
-Werner Heisenberg

M. Boucher[1]

[1]School of Computer Science
McGill University maxime.boucher@mcgill.ca

Brief summary of Comp601's readings, 2006

Outline

Important Questions that Will Be Addressed

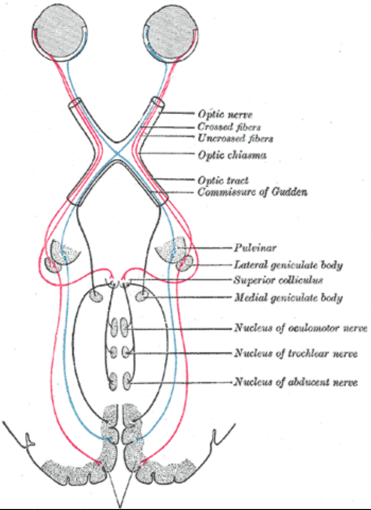
1. Given a certain feature detector (e.g. : edge detector), how is it possible to get the strongest and the most accurate answer on the location and the presence (or not) of the feature.
2. Given that features only exist at certain scales, is it possible to parcellate the space in order to make sure that every feature has a chance to be detected.
3. How can we optimize the probability to detect a feature and the accuracy of spatial localization? In other word, given an image, how can the number of detected features be minimized in order to concentrate the energy of each detected feature among only a few localized regions.

Early vision
An Example

1. When a person open his eyes, it takes only a fraction of a second to recognize the scene, identify what is present, and where is the important information and put it in context.
2. This represent too much information processing to cover everything during this talk!

Early vision

The human visual system



1. Image is acquired on the retina
2. Corresponding region of the left and right visual field is matched to the appropriate region of the V1 area.
3. Early vision refer to the processing of the image that has happened from the retina to the visual cortex

Early vision

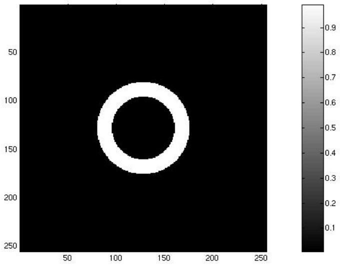
Low and high level vision

Early vision can be separated into low and high level vision.

- Low level vision: Feature detection based on models. Mainly edge detection and motion detection.
- High level vision: Feature detection based on inference using broader knowledge

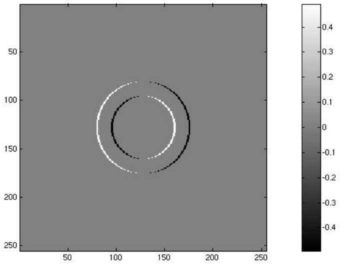
The focus of this talk is on optimization of low level vision.

Edge Detector
Feature detection example



We start by a simple case of feature detection. We would like to find the edges of this annulus.

Edge Detector
Feature detection example

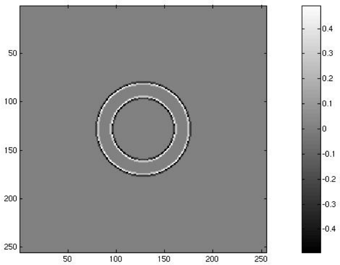


Lets look at the amplitude of the first derivative, in one direction, of the image given by the finite difference $\frac{\partial f}{\partial x} = f(x+\Delta x) - f(x+\Delta x) + \mathcal{O}(\Delta x^2)$.

Edges oriented in a certain direction are clearly identified as giving the greatest or the smallest answer within a region!

Edge Detector

Feature detection example



An extremum is then detected by looking at the point where the next derivative crosses zero. By combining the output of other oriented edge detector, it is possible to obtain a global view of the image. This gives a zero crossing representation of the image.

Other Types of Features

Edges are not the only kind of feature we may be interested in. Examples:

- Ridges are represented as a maximum in the second derivative
- Orientation is represented as a maximum in different oriented filters.
- A statistical difference in a specific location between two groups can be represented as a maximum in distance.

Formal Mathematical Definition of Low Level Feature Detection

A local feature detector g is a function that

- is a linear function of the input. Hence, it is possible to compute the result of f after the application of g by a convolution

$$f * g = \int f(x-u)g(u)du \qquad (1)$$

- is localized in space (no Fourier basis)

$$\begin{aligned} \|g\|_{2\mu} &= \int xg(x)^2 dx \\ (\|g\|_{2\sigma})^2 &= \int x^2(g(x)-\mu)^2 dx < \infty \end{aligned} \qquad (2)$$

It applies to images, volumes, surfaces or any space where these concepts are defined

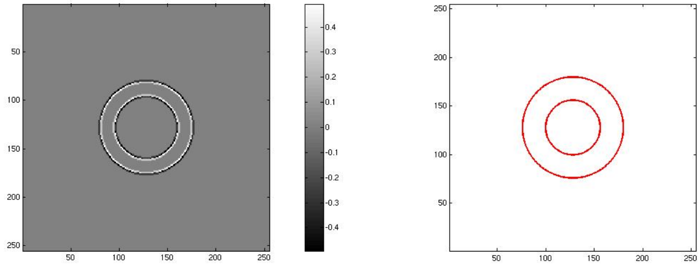
Examples of Acceptable Feature Detector

Here is a few acceptable feature detector

- Derivative: $\frac{\partial}{\partial x} \approx \delta_{\Delta x} - \delta_{-\Delta x}$
- Linear Projections: $< f, g >= \int f(x)g(x)dx$
- Distance functions: $\|f - g\|_2^2 = \|f\|_2^2 + \|g\|_2^2 - 2 < f, g >$

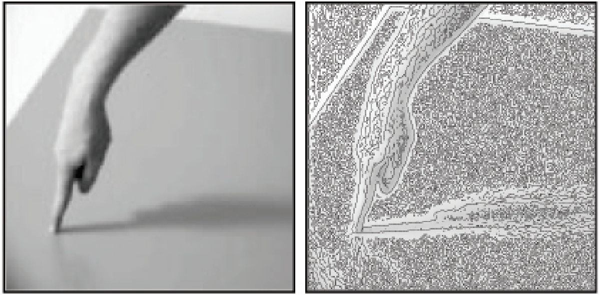
Zero-crossing representation of images

Features are afterward localized as a maximum in the output of a set of feature detector. This give a zero-crossing representation.



Is the story over?

Example of an edge detection attempt.



Problem with scale? How can we remove unimportant maximum and only detect those that most likely represent an edge?

Concept of Scale

Let two scales $\sigma_1 > \sigma_2$. Principles of maximization of detection using different scales:

- Maximum in derivatives should not "appear" as a result of the process. If a feature is present at a coarse scale σ_2, then it was present at the finer scale σ_1.
- Causality: The output at scale σ_2 can be computed knowing only the output at scale σ_1.

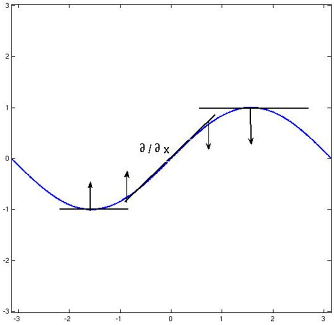
Concept of Scale

Theorem: *Gaussian filters* $W_{\Delta t}(x) = \frac{1}{\sqrt{2\pi\Delta t}} exp[-\frac{(x-\mu)^2}{2\Delta t}]$, *($\Delta t = \sigma^2$) which are the solution of the diffusion partial differential equation* $\frac{\partial f}{\partial t} = -\nabla^2 f$ *for a time step of* Δt *works for both principles.* Idea of the proof:

- Causality: Direct

Idea of the proof

No creation of zero crossings. Let a region on f where it allows an upper bound and a lower bound. Those corresponds to local maximum and local minimum.



After a very small time, the bounds will tighten, decreasing the value of the derivatives. Then, the same reasoning can be applied to bounds on this derivatives, they will tighten, further reducing the amplitude of the second derivative, and so on.

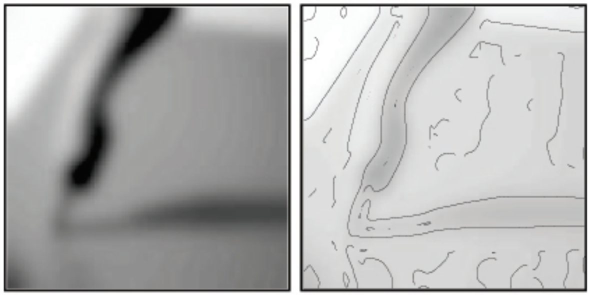
Different Scale Representation through Diffusion Smoothing

Using the commutativity of convolution, it is possible to obtain any desired scale σ by blurring the initial signal.

$$f * (g * W) = (f * W) * g$$

Is the story over?

Example of an edge detection attempt.



Problem with scale? Most unimportant edge disappeared as well as some important edges.

How is it possible to get the strongest and the most accurate answer on the location and the presence (or not) of the feature.

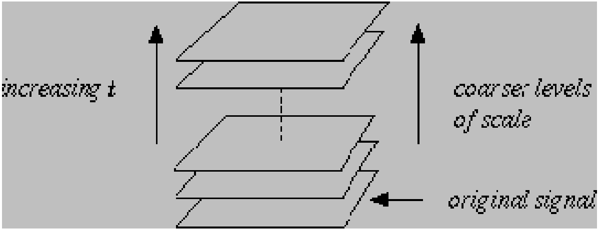
An edge, is a maximum in first spatial derivative AND scale derivative. Using normalized spatial derivatives

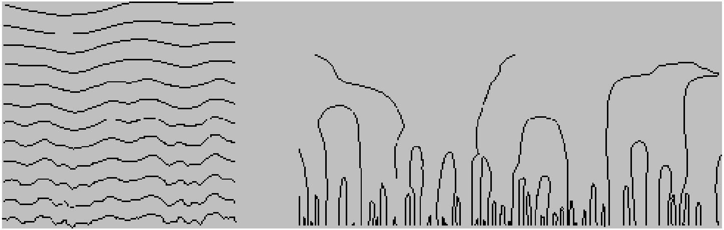
$$\partial_\xi = \sqrt{t}\partial_x$$

then an edge can be described as a set of conditions

- Maximum in spatial derivative: $\frac{\partial^2 f}{\partial \xi^2} = 0$ $\frac{\partial^3 f}{\partial \xi^3} < 0$
- Maximum in scale derivative: $\frac{\partial f}{\partial t} = 0$ $\frac{\partial^2 f}{\partial t^2} < 0$

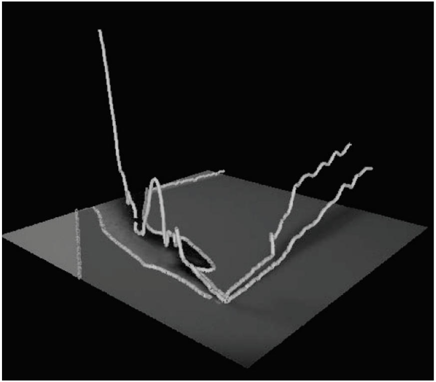
Principles of automatic scale selection





Is the story over?

Example of an edge detection attempt.



10 most significant edge extracted after scale selection. Height shows the scale at which the maximum answer was registered in terms of σ.

Recapitulation

How is it possible to get the strongest and the most accurate answer on the location and the presence (or not) of a feature?

- Given a certain feature detector, select maximums over all possibles scales.

Problems with Scale-Space Feature-Detection

Scale selection is an optimization process. Is this fit for statistical analysis?

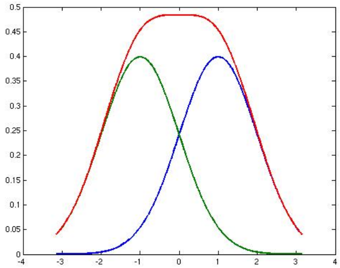
- Problem of Multiple Comparisons
- Features might be missed if we don't look at the appropriate scale

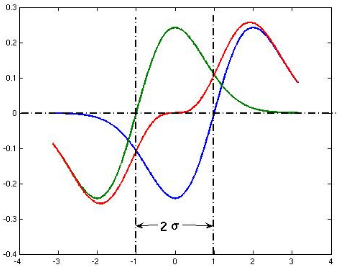
Question: Is it possible to make sure every feature has a chance to be detected? (Is it possible to cover the entire range of possible location and possible scale?)

The Question of Resolution

Corollary Question: Given a certain kernel width σ, what is the smallest distance at which two maximums can be detected?

- At least 2σ is required.





Thus, if we use a feature detector of size σ, we only need to sample the spatial domain at a sampling rate of σ.

Heisenberg Uncertainty Principle

- Heisenberg Uncertainty Principle: It is impossible to know at the same time with infinite precision the position and the frequency of a feature.
- If a feature g has a width of σ_x in the spatial domain, then the uncertainty in the frequency domain in given by σ_ω

$$\sigma_x \cdot \sigma_\omega \geq \frac{1}{2} \tag{3}$$

with equality when g is a gaussian.

Heisenberg Uncertainty Principle

Proof...

- Equality is easy to sketch using a Fourier transform.
- Inequality:

$$\sigma_x = \|xg\|_2 \, , \, \sigma_\omega = (2\pi)^{-1/2}\|\omega\hat{g}\|_2 = \|\frac{\partial g}{\partial x}\|_2 \tag{4}$$

Then, the last part is to show that $\|xg\|_2\|\frac{\partial g}{\partial x}\|_2 \geq - < xg|\frac{\partial g}{\partial x} >= \frac{1}{2}$

Maybe after the talk.

Heisenberg Uncertainty Principle in Real Life Situation

Two diapasons hit at the same time with very little difference in their first mode of vibration.

$$\begin{aligned}
f(x) &= cos(\omega x + \Delta\omega x) + cos(\omega x - \Delta\omega x) \\
f(x) &= cos(\omega x)cos(\Delta\omega x) - sin(\omega x)sin(\Delta\omega x) \\
&\quad +cos(\omega x)cos(\Delta\omega x) + sin(\omega x)sin(\Delta\omega x) \\
&= 2cos(\Delta\omega x)cos(\omega x)
\end{aligned} \tag{5}$$

Impossible to make a distinction with an amplitude modulated diapason or two diapasons!

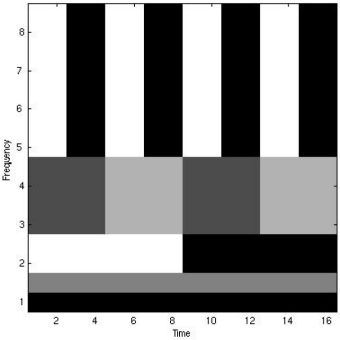
Given that features only exist at certain scales, is it possible to parcellate the space in order to make sure that every feature has a chance to be detected.

The principle to partition the space is the following: We need a spatial window that will cover at least one cycle of an oscillation before detecting a change.

- High Frequency: Operator that can detect a sharp change spatially
- Low Frequency: Operator that can discriminate between change at different scales.

Let a real signal f sampled N times at intervals of Δx. Its Fourier transform has $\frac{N}{2}$ independent complex components sampled at intervals of $\frac{1}{N\Delta x}$. The total size is $\frac{N}{2}$ and each window occupies an area of $\frac{1}{2}$. The total scale-space plane is covered using only N windows!

Example of a dyadic grid



Example of a such a representation: Gaussian pyramid with differential encoding

Optimizing the coverage of windows

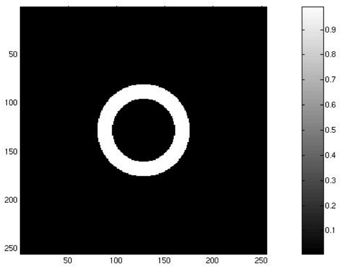
Up until now, assumptions is that images shows isotropic Fourier spectrum. Is it possible to do better?

- It would be interesting to parcellate space into similar regions.
- Maximize detection by selecting regions where they will always share the same output. (Either, feature is present or feature is not present).

Optimizing the coverage of windows

How can one partition the space?

- New strong hypothesis: Images are piecewise constant.
- Piecewise constant regions are delimitated by edges that forms a curve that one would like to detect.



Optimizing the coverage of windows

Perona and Malik suggested to adapt diffusion $c(x, y, t)$ speed to the local probability to find an edge. The idea is to diffuse as mush as possible in constant area and as less as possible across different areas.

- $\frac{\partial f}{\partial t} = \nabla(c\nabla f) = c\nabla^2 f + \nabla c \nabla f$
- The trick is: we don't where is the edge, however, approximation can be refined during the diffusion process.

Optimizing the coverage of windows

Anisotropic diffusion smoothing can *enhance* edge. It does not scale them because it would violate the no new information added principle of scaling function.

- 1D case: $\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\phi(\frac{\partial f}{\partial x}) = \frac{\partial \phi}{\partial x}(\frac{\partial f}{\partial x})\frac{\partial^2 f}{\partial x^2}$
- We are interested to look at what happen to the first derivative to perform edge detection $\frac{\partial^2 f}{\partial x \partial t}$.
  $\frac{\partial^2 f}{\partial x \partial t} = \frac{\partial^2 \phi}{\partial x^2}(\frac{\partial f}{\partial x})\frac{\partial^2 f}{\partial x^2} + \frac{\partial \phi}{\partial x}(\frac{\partial f}{\partial x})\frac{\partial^3 f}{\partial x^3}$
- An edge detector implies that $\frac{\partial^2 f}{\partial x^2} = 0$ and $\frac{\partial^3 f}{\partial x^3} << 0$ because it detects a maximum in magnitude.

Optimizing the coverage of windows

Thus,

- $\frac{\partial^2 f}{\partial x \partial t} = -\alpha \frac{\partial \phi}{\partial x}(\frac{\partial f}{\partial x})$ where $\alpha = \frac{\partial^3 f}{\partial x^3}$.
- It is possible to enhance edge if $\frac{\partial \phi}{\partial x}(\frac{\partial f}{\partial x}) < 0$!
- Example of edge enhancing function:$\phi(\nabla f) = -\frac{\|\nabla f\|}{\sigma^2} e^{-\frac{\|\nabla f\|^2}{2\sigma^2}}$

Optimizing the coverage of windows

For Further Reading I

- S.W. Zucker.
  Early Vision.
  Encyclopaedia of Artificial Intelligence, 1987.
- D.H. Hubel and T.N. Wiesel
  Brain Mechanisms of Vision.
  Scientific American, 1979.
- R.A. Hummel
  Representations based on Zero-crossings in Scale-Space.
  CVPR, 1990.

For Further Reading II

P. Perona and J. Malik
Scale-Space and Edge Detection using Anistropic Diffusion.
IEEE TPAMI, 1990.

T. Lindberg
Scale-space for Discrete Signals.
IEEE TPAMI, 1990.

T. Lindberg
Edge Detection and Ridge Detection with Automatic Scale Selection.
CVPR, 1996.