Chapter 2
Diffusion-based Image Filtering

Variational Methods in Imaging

Feb 2021

Yvain QUÉAU
GREYC-CNRS
ENSICAEN - Université de Caen Normandie

GREYC

Overview

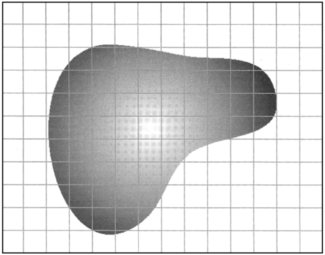
1. **Basic Digital Image Filtering**
2. **Derivative Filters**
3. **Partial Differential Equations**
4. **Nonlinear Diffusion**

Overview

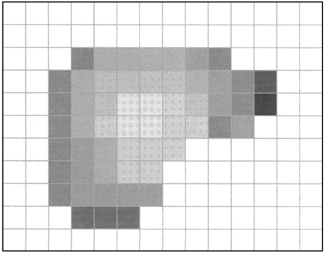
1. **Basic Digital Image Filtering**
2. Derivative Filters
3. Partial Differential Equations
4. Nonlinear Diffusion

Continuous versus Discrete

Digital images are discrete, both in space and in their values. Nevertheless, one can represent and analyze them in a continuous setting.



continuous



discrete
(sampling & quantization)

Continuous versus Discrete

- There are different levels of discretization:
  - Discretization in color or brightness space (=quantization)
  - Discretization in (physical) space
  - Discretization in time (for videos)
- Continuous representation: $f : (\Omega \subset \mathbb{R}^n) \to \mathbb{R}^d$
- $n = 2$: 2-dim. images,
  $n = 3$: volumetric images or 2-dim. videos,
  $n = 4$: volume + time,...
- $d = 1$: brightness images,
  $d = 3$: color images,
  $d > 1$: multispectral images
- Discretization (width W, height H):

$$f(x,y) \longrightarrow \begin{bmatrix} f(1,1) & f(2,1) & \cdots & f(W,1) \\ f(1,2) & f(2,2) & \cdots & f(W,2) \\ \vdots & \vdots & \ddots & \vdots \\ f(1,H) & f(2,H) & \cdots & f(W,H) \end{bmatrix}$$

Continuous versus Discrete

- Advantages of discrete representations:
    - Digital images *are* discrete, and their processing in a computer will ultimately require a discretization.
    - No numerical approximations in modeling the transition from discrete to continuous.
    - For various problems there exist efficient algorithms from discrete optimization.
- Advantages of continuous representations:
    - The world observed through the camera is continuous.
    - There exists abundant mathematical theory for the treatment of continuous functions (functional analysis, differential geometry, partial differential equations, group theory,...).
    - Certain properties (rotational invariance) are easier to model because artefacts of discretization can be ignored.
    - Continuous models correspond to the limit of infinitely fine discretization.

Linear Filters

- The term filtering is derived from frequency space methods where a spatial smoothing of the brightness values corresponds to a signal transform where high-frequency components are filtered out.
- An operator T is called linear if the following properties hold:
  1. $T(f+g) = T(f) + T(g) \quad \forall$ images f, g.
  2. $T(\alpha f) = \alpha T(f) \quad \forall$ images f, scalars α.
- For linear operators, the output brightness values are linear combinations of the input brightness values. Among the linear transformations is the convolution:

$$g(x,y) = \iint w(x',y')\, f(x-x', y-y')\, dx'\, dy'.$$

  In a spatially discrete setting, this corresponds to a weighted sum:

$$g(i,j) = \sum_{m,n} w(m,n) f(i-m, j-n).$$

Linear Filters

- In practice this summation extends over a certain neighborhood, often called window. The matrix of weights $w(m,n)$ is called a mask.

$$g(i,j) = \sum_{m,n} w(m,n) f(i-m, j-n)$$

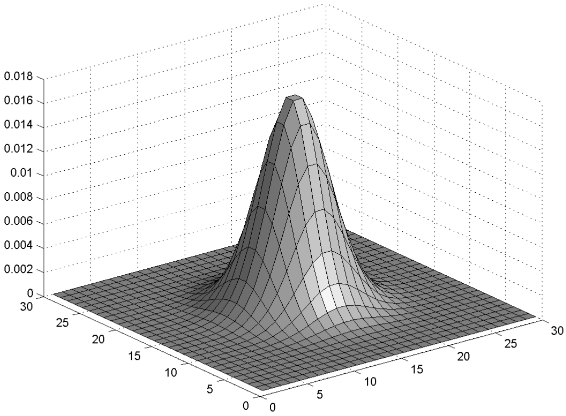
- For example, the 3×3 mask:

w(-1,-1)	w(0,-1)	w(1,-1)
w(-1,0)	w(0,0)	w(1,0)
w(-1,1)	w(0,1)	w(1,1)

- In the continuous representation the weight function $w(x', y')$ is called convolution kernel:

$$g(x,y) = (w * f)(x,y) \equiv \iint w(x', y')\, f(x-x', y-y')\, dx'\, dy'$$

Gaussian Convolution



"Everybody believes in the exponential law of errors: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation." (G. Lippman, in a letter to H. Poincaré)

Smoothing Filters

- Smoothing or low-pass filtering typically averages the brightness values in a certain spatial neighborhood.
- The most common example of smoothing kernel is the Gaussian kernel. It induces a weighted average of brightness values on the scale determined by the standard deviation σ:

$$w(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

- A multitude of alternative convolution kernels (or filter masks) is conceivable, for example box filters which are constant within the window:

$$w(i,j) = \frac{1}{9} \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$

- For pixels at the image boundary, the weight mask must be adapted appropriately.

Nonlinear Filtering

- The convolution of an input image $f(x)$ with a kernel $G(x)$:

$$g(x) = (G * f)(x) = \int G(x')f(x - x')dx'$$

is a classical example of a linear filter.
- Convolutions can be efficiently implemented in frequency space because in frequency space the convolution corresponds to a simple (frequency-wise) product and because the Fast Fourier transform allows a quick conversion to and from frequency space.
- In practice, however, linear filters are often suboptimal. In smoothing/denoising, for example, the Gaussian smoothing removes both noise and signal – semantically relevant structures tend to disappear along with the noise. Instead, one would like to remove noise in an adaptive manner such that semantically important structures remain unaffected. In principle this could be done with a Gaussian smoothing where the filter width σ is adapted to the local structure (larger in noise areas, smaller at important edges).

Image Filtering by Diffusion

- Formally this would amount to the following:

$$g(x) = \int G_{\sigma(f,x)}(x')f(x-x')dx',$$

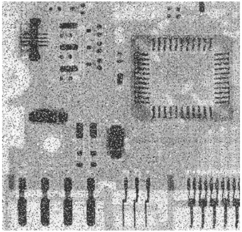
where now the width σ of the convolution kernel G depends on the brightness values in a local neighborhood.

- It turns out that there exist other more elegant solutions to model such adaptive denoising processes by means of Diffusion filtering.
- The key observation is that image smoothing can be modeled with a diffusion process. In this process, the local brightness diffuses to neighboring pixels due to differences in the local concentration of grayvalue.
- Mathematically diffusion processes are represented by partial differential equations (PDEs).

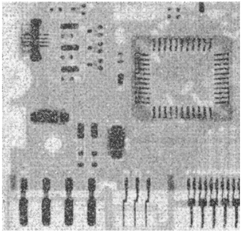
The Median Filter

- A specific class of nonlinear filters are the order statistics filters. For these filters, the brightness of the filtered image at a given pixel depends on the order of brightness values in a certain neighborhood.
- The best known example of an order statistics filter is the median filter. For this filter, each pixel is assigned the median value of brightness values in its neighborhood.
- Example: The median of the brightness values $\{1, 2, 2, 3, 4, 5, 20\}$ is 3, i.e. the central value after sorting.
- Median filters are particularly useful for reduction of impulse noise, also called salt-and-pepper noise, i.e. noise where some brightness values are randomly replaced by black or white values.
- Median filters typically induce less blurring than Gaussian or other linear smoothing filters.

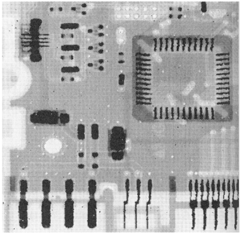
Median versus Gauss



noisy input



Gauss filtered



median filtered

In contrast to the Gaussian filter (center), the median filter better removes noise without blurring structures. Nonlinear methods are often more general and more powerful than linear approaches. A natural mathematical framework for this is **nonlinear diffusion**.

Overview

1. Basic Digital Image Filtering
2. **Derivative Filters**
3. Partial Differential Equations
4. Nonlinear Diffusion

Derivative Filters

- Derivative filters capture the spatial variations of brightness. In particular, they provide information about edges or corners in an image. In a simplified world of black objects on white ground, these brightness edges correspond to object boundaries.
- Mathematically the partial derivatives of the function $f(x, y)$ with respect to x is defined as:

$$\frac{\partial f(x, y)}{\partial x} = \lim_{\epsilon \to 0} \frac{f(x + \epsilon, y) - f(x, y)}{\epsilon}$$

- Taking $\epsilon = 1$, the continuous derivative can be approximated discretely by (forward) finite differences:

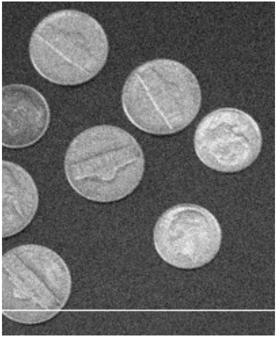
$$\partial_x f(x, y) \equiv f_x(x, y) \equiv \frac{\partial f(x, y)}{\partial x} \approx f(x + 1, y) - f(x, y)$$

- Alternatives:

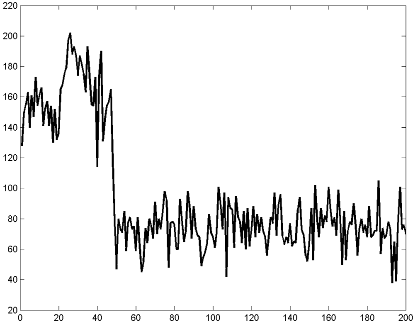
$$\partial_x f(x, y) \approx f(x, y) - f(x - 1, y) \quad \text{(backward difference).}$$

$$\partial_x f(x, y) \approx \tfrac{f(x+1,y)-f(x-1,y)}{2} \text{ (central difference)}$$

Example: 1D Brightness Profile

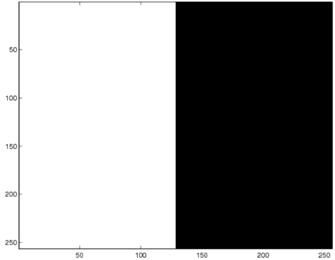


Input

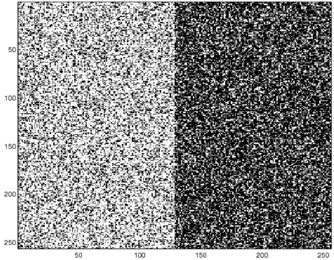


1D brightness profile

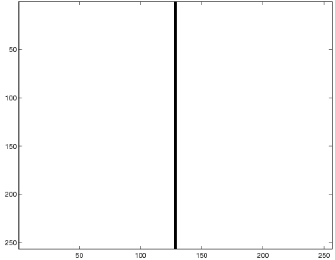
Example of the First Derivative



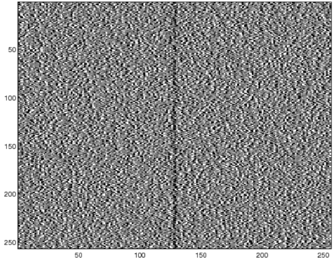
Input image



Input with noise

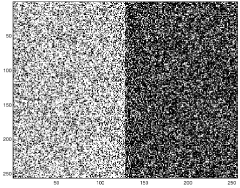


x-derivative

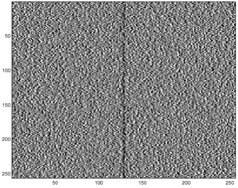


derivative of noisy image

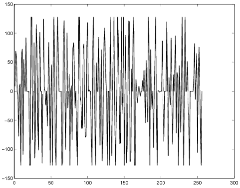
Noise Sensitivity of the Derivative



Input f



derivative f_x



f_x along horizontal

Observation:

- Vertical edges can be determined as maxima of the norm of the x-derivative.
- Horizontal edges can be determined as maxima of the norm of the y-derivative.
- This approach only allows to selectively determine horizontal or vertical edges.
- It is very sensitive to noise.

The Image Gradient $\nabla f(x, y)$

- The gradient of a function $f(x, y)$ is the vector:

$$\nabla f(x, y) = \begin{pmatrix} \partial_x f \\ \partial_y f \end{pmatrix} \equiv \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

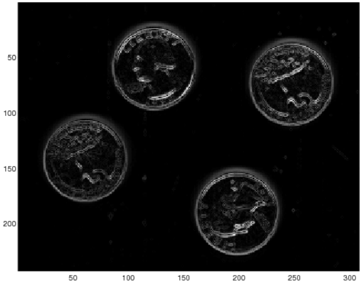
- The gradient norm (often also called “gradient”) is given by the Euclidean length of the gradient vector:

$$|\nabla f(x, y)| = \left| \begin{pmatrix} f_x \\ f_y \end{pmatrix} \right| = \sqrt{(f_x)^2 + (f_y)^2}$$

- The gradient norm is a nonlinear operator for detection of edges in arbitrary orientation.
- The gradient norm is rotationally covariant (sometimes called “rotationally invariant”). This means: The gradient norm of the rotated image is the same as the rotated gradient norm of the unrotated image. This implies that the performance of this operator does not depend on how the input image is rotated.

Example of the Image Gradient

Input image



Gradient norm

The Laplace Operator $\Delta f(x, y)$

- The divergence of a vector $v = (v_1, v_2)$ is defined as $\nabla \cdot v = \partial_x v_1 + \partial_y v_2$.
- The Laplace operator Δ is given by the concatenation of gradient and divergence:

$$\Delta f(x, y) = \nabla^2 f(x, y) = \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} \cdot \begin{pmatrix} f_x \\ f_y \end{pmatrix} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = f_{xx} + f_{yy}$$

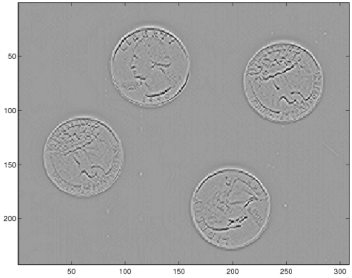
- The Laplace operator is linear:

$$\Delta(\alpha_1 f(x) + \alpha_2 g(x)) = \alpha_1 \Delta f(x) + \alpha_2 \Delta g(x) \quad \forall \alpha_1, \alpha_2 \in \mathbb{R}, \forall f, g$$

- Linearity has several practical advantages. Linearity implies that it does not matter whether one first sums images and then processes them or vice versa.

Example of the Laplace Operator

Input image



Laplace operator of the image

Discretization of Derivatives

There exist different discrete approximations of derivatives. In the following we shall denote height and width of a single pixel by h_x and h_y. Then the x-derivative of a brightness image f at pixel (i, j) can be approximated as:

1. Symmetric differences:

$$f_x(i,j) \approx \frac{f(i+1,j) - f(i-1,j)}{2h_x}$$

2. Forward differences:

$$f_x(i,j) \approx \frac{f(i+1,j) - f(i,j)}{h_x}$$

3. Backward differences:

$$f_x(i,j) \approx \frac{f(i,j) - f(i-1,j)}{h_x}$$

How do these masks differ? Which one is better?

Discretization of the Laplacian

0	1	0
1	−4	1
0	1	0

1	1	1
1	−8	1
1	1	1

Two masks showing discretizations of $\Delta f = f_{xx} + f_{yy}$.

Overview

1. Basic Digital Image Filtering
2. Derivative Filters
3. **Partial Differential Equations**
4. Nonlinear Diffusion

Review: Partial Differential Equations

- A partial differential equation (PDE) is an equation containing the partial derivatives of a function of several variables.
  Example — the wave equation:

$$\frac{\partial^2 \psi(x,t)}{\partial t^2} = c^2 \Delta \psi(x,t)$$

- For functions of a single variable we have the special case of ordinary differential equations (ODEs).
  Example — the pendulum:

$$m\frac{d^2 x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + kx(t) = 0$$

- Many natural phenomena can be modeled by partial differential equations. In most cases, one can derive the respective equation from a few basic principles. A solution of a differential equation is a function for which the differential equation is true.

Analytical Solutions

- A few PDEs can be solved analytically, i.e. the solution can be written in closed form.
- Example — The wave equation (in 1D):

$$\frac{\partial^2 \psi(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \psi(x,t)}{\partial x^2}$$

  has the (not necessarily unique) solution:
  $\psi(x,t) = \sin(x - ct)$
- If solutions are not unique one can impose additional assumptions boundary conditions or initial conditions, for example $\psi(x,0) = \psi_0(x)$.
- Example — The harmonic oscillator (without friction):

$$m \frac{d^2 x(t)}{dt^2} + kx(t) = 0$$

  has the (generally not unique) solution:

$$x(t) = sin(\omega t), \quad \text{with } \omega = \sqrt{k/m}.$$

Non-Analytical Solutions - The Finite Differences Method

Example: discretization of the 2D wave equation

$$\frac{\partial^2 u(x,y,t)}{\partial t^2} = c^2 \Delta u(x,y,t)$$

- Time discretization (backward, order 2):
  $\frac{\partial^2 u(\cdot,\cdot,t)}{\partial t^2} \approx u^{(t)} - 2u^{(t-1)} + u^{(t-2)}$
  $\Rightarrow u^{(t)} - 2u^{(t-1)} + u^{(t-2)} = c^2 \Delta u^{(t)}$
  $\Rightarrow \left(\mathrm{id} - c^2\Delta\right) u^{(t)} = 2u^{(t-1)} - u^{(t-2)}$
  (assume $u^{(-1)}$ and $u^{(0)}$ are known - initial condition)

- Space discretization (central, order 2):
  $\Delta u(x,y,\cdot) \approx$
  $u(x+1,y)+u(x-1,y)+u(x,y+1)+u(x,y-1)-4u(x,y)$
  (assume, e.g., Dirichlet boundary conditions)

$\Rightarrow$ Linear system $A^{(t)}u^{(t)}(:) = b^{(t)}$ to solve to obtain u^t

Example: numerical simulation of the 2D wave equation

Overview

1. Basic Digital Image Filtering
2. Derivative Filters
3. Partial Differential Equations
4. **Nonlinear Diffusion**

The Diffusion Equation

- Diffusion is a physical process which aims at minimizing differences in the spatial concentration $u(x, t)$ of a substance.
- This process can be described by two basic equations:
  - Fick's law states that concentration differences induce a flow j of the substance in direction of the negative concentration gradient:

    $$j = -g\nabla u$$

    The diffusivity g describes the speed of the diffusion process.
  - The continuity equation

    $$\partial_t u = -\mathrm{div}\, j$$

    where $\mathrm{div}\, j \equiv \nabla^\top j \equiv \partial_x j_1 + \partial_y j_2$ is called the divergence of the vector j.
- Inserting one into the other leads to the diffusion equation:

$$\boxed{\partial_t u = \mathrm{div}\,(g \cdot \nabla u)}$$

Solution of the Linear Diffusion Equation

The one-dimensional linear diffusion equation ($g = 1$)

$$\partial_t u = \partial_x^2 u.$$

with initial condition

$$u(x, t = 0) = f(x)$$

has the unique solution:

$$u(x,t) = (G_{\sqrt{2t}} * f)(x) = \int_{-\infty}^{\infty} G_{\sqrt{2t}}(x - x')f(x')dx',$$

where

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}},$$

is a Gaussian kernel of width $\sigma = \sqrt{2t}$.

Finite Differences Discretization

Discretization of the 2D diffusion equation

$$\frac{\partial u(x,y,t)}{\partial t} = \text{div}\left(g \cdot \nabla u(x,y,t)\right)$$

- Time discretization (backward, order 1):
  $\frac{\partial u(\cdot,\cdot,t)}{\partial t} \approx u^{(t)} - u^{(t-1)}$
  $\Rightarrow (\text{id} - \text{div}\,(g \cdot \nabla))\, u^{(t)} = u^{(t-1)}$
  (assume $u^{(0)}$ known - initial condition)
- Space discretization (order 1 - forward for the gradient, backward for the divergence):
  $\Rightarrow$ Linear system $A^{(t)} u^{(t)}(:) = b^{(t)}$ to solve to obtain u^t

Example: numerical simulation of the 2D diffusion

With uniform diffusivity ($g \equiv 1$):

Example: numerical simulation of the 2D diffusion

With non-uniform diffusivity ($g \equiv 1$ except on the “walls”)

Nonlinear and Anisotropic Diffusion

- General diffusion equation:

$$\partial_t u = \operatorname{div}(g \nabla u)$$

- For $g = 1$ (or $g = \text{const.} \in \mathbb{R}$) the diffusion process is called linear, isotropic and homogeneous.
- If the diffusivity g is space-dependent, i.e. $g = g(x)$, the process is called an inhomogeneous diffusion.
- If the diffusivity depends on u, i.e. $g = g(u)$, then it is called a nonlinear diffusion because then the equation is no longer linear in u.
- If the diffusivity g is matrix-valued then the process is called an anisotropic diffusion. A matrix-valued diffusivity leads to processes where the diffusion is different in different directions.
- Note: In the literature this terminology is not used consistently.

Edge-preserving Diffusion

- Idea: Less diffusion (smoothing) in locations of strong edge information.
- Gradient norm $|\nabla u| = \sqrt{u_x^2 + u_y^2}$ serves as edge indicator
- Diffusivity should decrease with increasing $|\nabla u|$. For example (Perona & Malik, *Scale Space and Edge Detection using Anisotropic Diffusion, PAMI 1990*):

$$g(|\nabla u|) = \exp\left\{-|\nabla u|^2/\lambda^2\right\} \quad \text{or} \quad g(|\nabla u|) = \frac{1}{1 + |\nabla u|^2/\lambda^2}$$

- $\lambda > 0$ is called a contrast parameter. Areas where $|\nabla u| \gg \lambda$ will not be affected much by the diffusion process.
- The Perona-Malik model had a huge impact in image processing because it allowed a better edge detection than classical edge detectors (such as the Canny edge detector).

Image Smoothing by Diffusion

With uniform diffusivity ($g \equiv 1$):

Image Smoothing by Diffusion

With Perona-Malik's "anisotropic" diffusivity ($g = \exp\left\{-|\nabla u|^2/\lambda^2\right\}$), $\lambda = 5$

Image Smoothing by Diffusion

With Perona-Malik's "anisotropic" diffusivity ($g = \frac{1}{1+|\nabla u|^2/\lambda^2}$), $\lambda = 5$

Image Smoothing by Diffusion

With minimal surface diffusivity ($g = \frac{1}{\sqrt{1+|\nabla u|^2/\lambda^2}}$), $\lambda = 1$

Appropriately Setting the Contrast Parameter λ

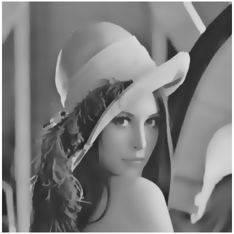
Example with $g = \frac{1}{1+|\nabla u|^2/\lambda^2}$

Nonlinear Diffusion a la Weickert

Lena original

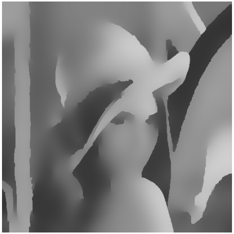
diffusion $t = 9$

diffusion $t = 25$



diffusion $t = 100$

diffusion $t = 400$



diffusion $t = 900$

Author: D. Cremers