Today’s Lecture

- Computational Vision
  - Images
  - Image formation in brief
  - Image processing: filtering
    - Linear filters
    - Non-linear operations
    - *Signatures*
    - *Edges*
  - *Image interpretation*
    - *Edge extraction*
    - *Grouping*
  - *Scene recovery*
  - *Recognition*

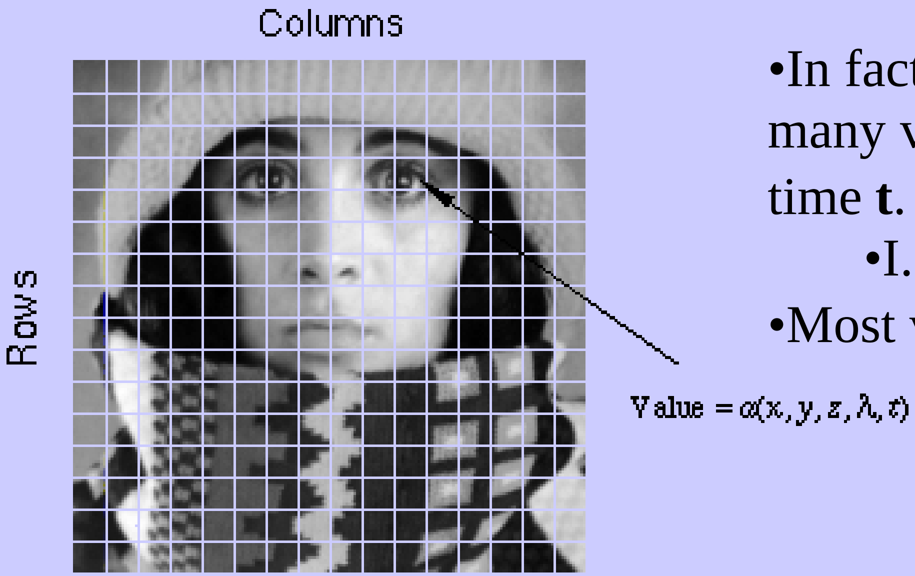
```
Items in blue
will (may) be
Covered later
```

What's an image?

- The ideal case:
  - We have an a continuous world (at macroscopic scales).
  - We have continuous images to that world.
  - Images are 2-dimensional projections of a three-dimensional world.
    - In addition, there are other **key** factors in the world that determine the image:
      - **Object reflectance properties (white or gray shirt?)**
      - **Light source position**
        - » **Alters intensities (day/night, shading & chiaroscuro)**
        - » **Alters shadows**
- In practice:
  - Discrete sampling of "the" continuous image.
  - Each pixel is an average.
  - *The image is a big array of numbers representing intensities, or sometimes triples representing colors too.*

Digital images: synopsis

- 2D continuous image a(x,y) divided into N rows and M columns.
  - The intersection of a row and a column termed a pixel.
  - Value assigned to integer coordinates **[m,n]** with **{m=0,1,2,...,M-1}** and **{n=0,1,2,...,N-1}** is **a[m,n].**



•In fact, in most cases **a(x,y)** is a function of many variables including depth **z**, color λ, and time **t**.

•I.e. we really have **a(x,y,z,y, λ,t)**.

•Most vision deals with the case of 2D, monochromatic ("black and white"), static images.

Image formation

Key processes:

1. Light comes from source
2. Hits objects in scene
3. Is reflected towards
    A) other objects [return to step 2]
    B) the camera lens
4. Passes through lens being focussed on imaging surface
5. Interacts with transducers on surface to produce electrical signals

In the eye, the lens is that of the eye, the imaging surface is the **retina**, and the transducers are **rods** and **cones**.

In a camera, the imaging surface is a **CCD** (charge-coupled device) made up of transistors (or related devices).

[several other types of video camera technology are commonly available, including newer “cheap and cheerful” CMOS cameras, and older “opticons”.]

The Human Visual System

- Digression of the biology of the early human visual system.

(On blackboard: no notes available.)

Vision, image processing, and AI

- Image processing:
  - Image -> image
  - Often characterized as data reduction, filtering
  - Signal-level transformation
- **Traditional** AI:
  - predicates -> predicates
  - What's hidden in the data: inference, "data mining"
  - Symbol-level transformation
    - ("Non-traditional" AI: neural nets, uncertainty, etc.)
- Vision
  - A) Images -> scenes
  - B) Images -> predicates
  - Getting more out than what's there: an **inverse problem**.

Vision problems

A) Images -> scenes

Known as "**shape-from**" or "**shape from X**"

Examples:

- Recovery of scene structure from a sequence of pictures: **shape-from-motion**
- Recovery of scene structure from shading: **shape-from-shading**
- Recovery of scene structure from how shadows are cast: **shape-from-shadows** (actually called "**shape-from-darkness**")
- Recovery of shape of changes in texture: **shape-from-texture**

B) Images -> predicates

Several variations, generally less mature.

Object recognition, functional interpretation, support relations.

What we want/need is (B), but (A) seems "easier" or is a natural prerequisite.

What is vision?

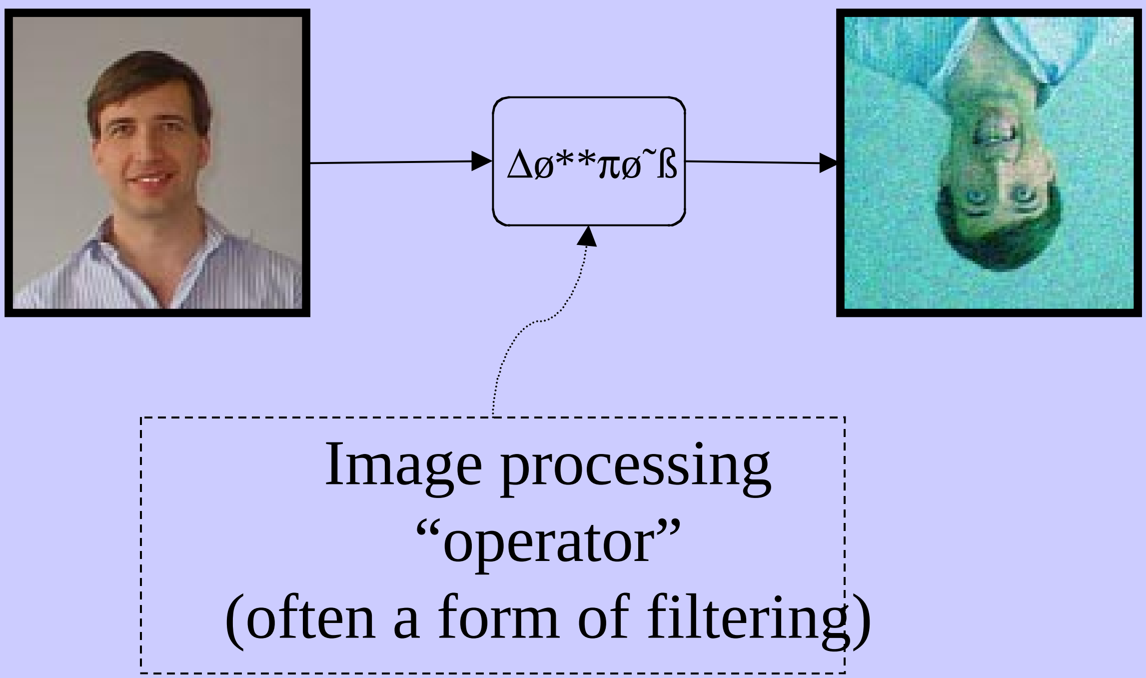
In general, vision involves the recovery of all those things that determine the world:

– Material properties, shadows, etc.

as well as the functional and categorical relationships between or pertaining to objects!

Image Processing

- Better understood than vision.
- Produce new images or arrays of data without worrying (much) about “interpretations.”



Filtering

- Given an input signal *f(t)* we can compute a transformed description *g(t)*.
  - Key requirement: the dimensionality of the domain and range is the same.
- This transformed signal is derived from f(t) by the application of either linear (multiplication/addition) or non-linear operators

E.g.

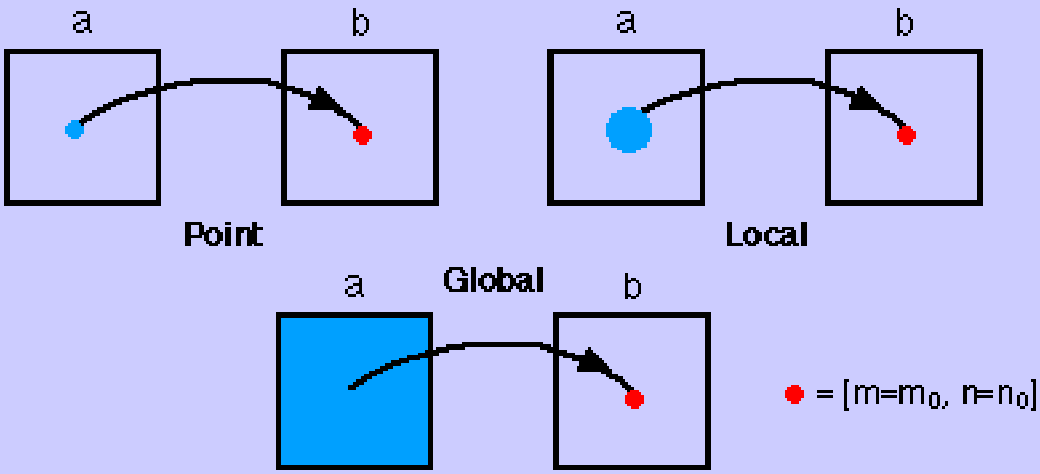
- $g(t) = f(t) + f(t-1) + f(t+1)$ [linear]
- $g(t) = f(t) * f(t-1) + f(t+1)^2$ [non-linear]
- *g(t) =* ***if*** *(f(t) >0) g(t)=1* ***else*** *g(t)-1* [non-linear]

Filtering in 2D

- Note that filtering applies in essentially the same way in
  - 1-D signals,
  - 2-D images,
  - or even higher dimensional spaces.

Filters: more flavors

- Additional key characterization is the the degree of locality of the filter.
  - Does it look
    - At a single point?
    - At a region…. And is the region symmetric?
    - At everything?



Convolution

- For vision and image processing, the most important class of filtering operation is **convolution.**
  - This is almost the same a correlation.
  - Convolution for 2 signals
  - c(t) = a(t) * b(t)
  - c(x,y) = a(x,y) * b(x,y)

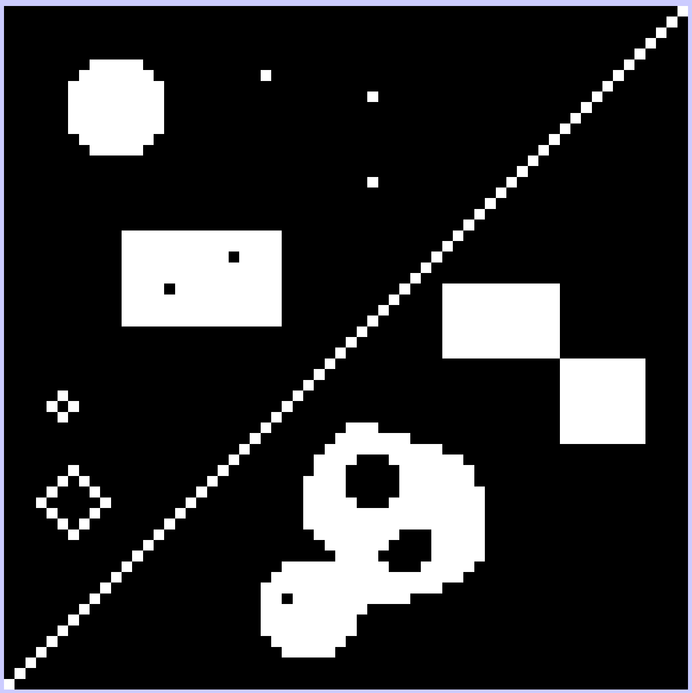
$$c(x,y) = a(x,y) \otimes b(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} a(\chi,\zeta)\, b(x-\chi, y-\zeta)\, d\chi\, d\zeta$$

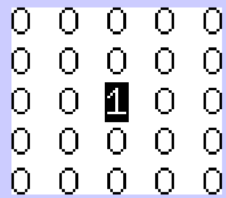
  - For discrete signals

$$c[m,n] = a[m,n] \otimes b[m,n] = \sum_{j=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} a[j,k]\, b[m-j, n-k]$$

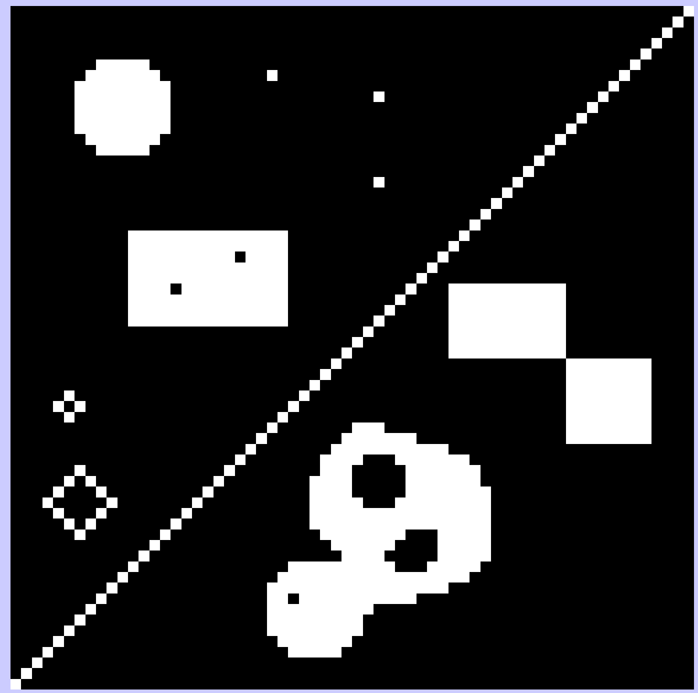
Convolution: specifics

- Typically we **convolve** a signal with a **kernel**
- Note that convolution is distributive, associative, and commutative.





Impulse

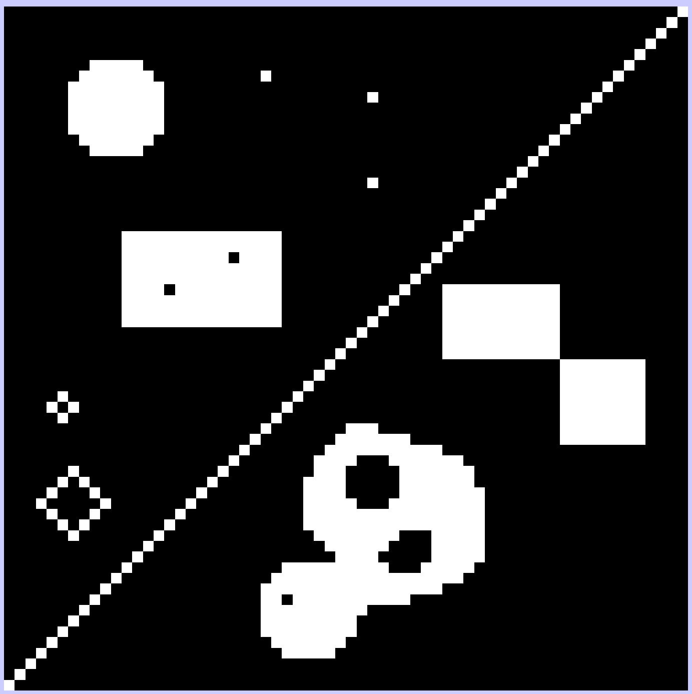


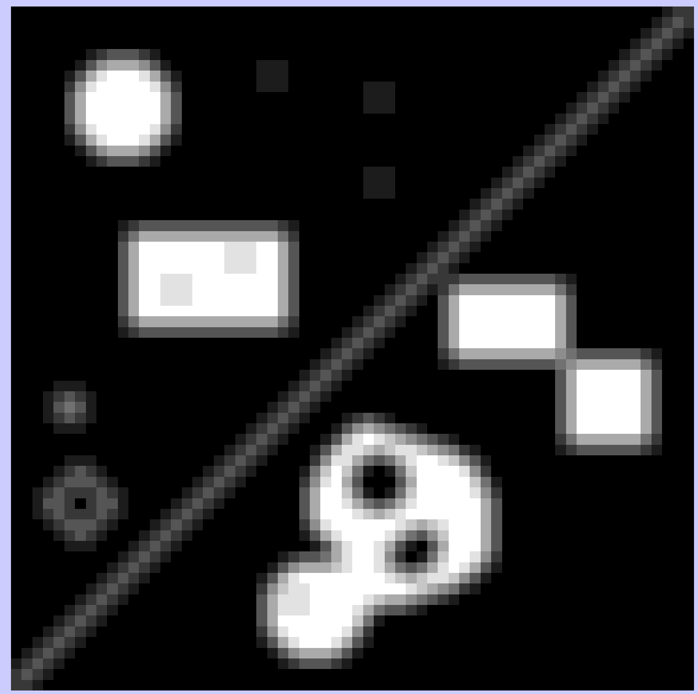
E.g. Blurring

- Convolve the input with a kernel that combines information from a range of spatial locations.
- What is the precise shape of the kernel?
- Why might this be useful?

Blur picture

0	0.00	0.00	0.00	0
0	0.33	0.33	0.33	0
0	0.33	0.33	0.33	0
0	0.33	0.33	0.33	0
0	0.00	0.00	0.00	0



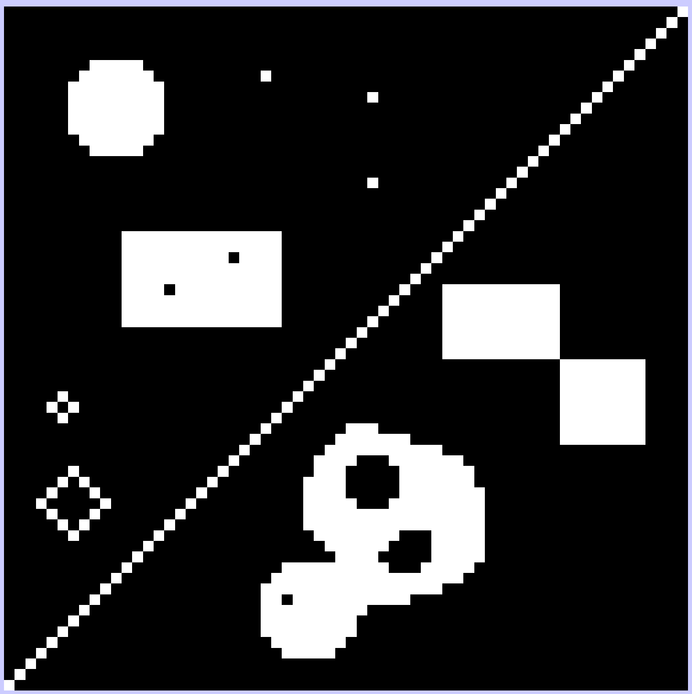


Edges

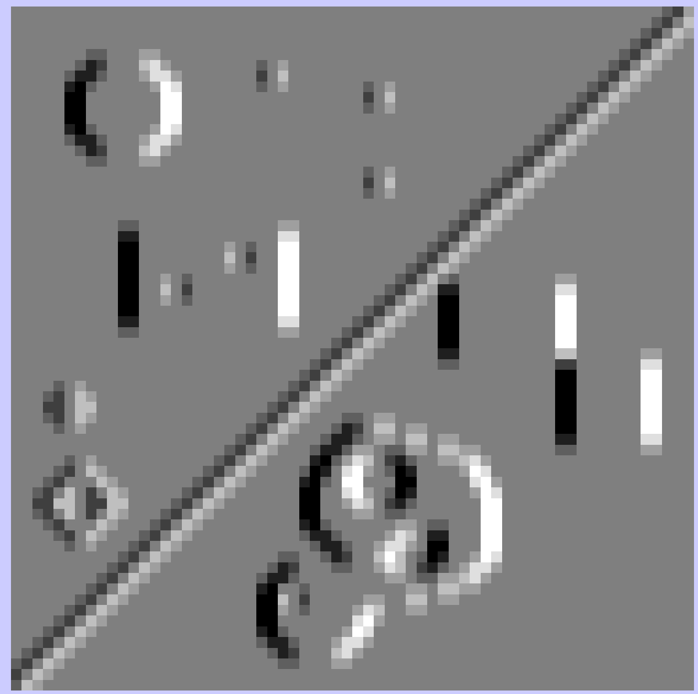
- Boundaries are thought to be critical to image interpretation.
  - Why do cartoons look as reasonable as they do?
  - Idea: detect the boundaries of objects in images.

The Sobel Edge operator

- Filter for horizontal and vertical edges, combine.



0	0	0	0	0
0	−1	0	1	0
0	−2	0	2	0
0	−1	0	1	0
0	0	0	0	0



Edges

Edge linking

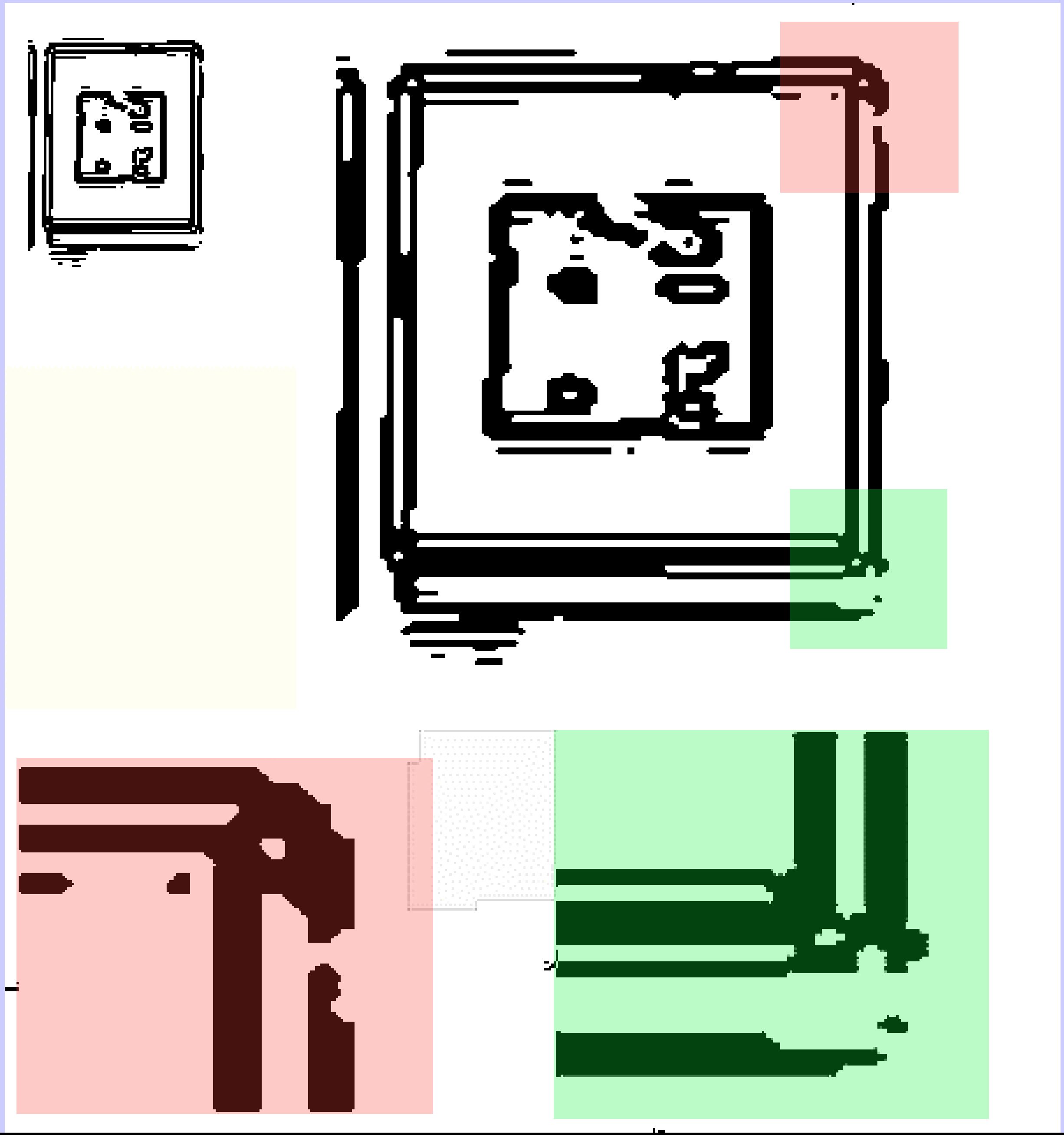
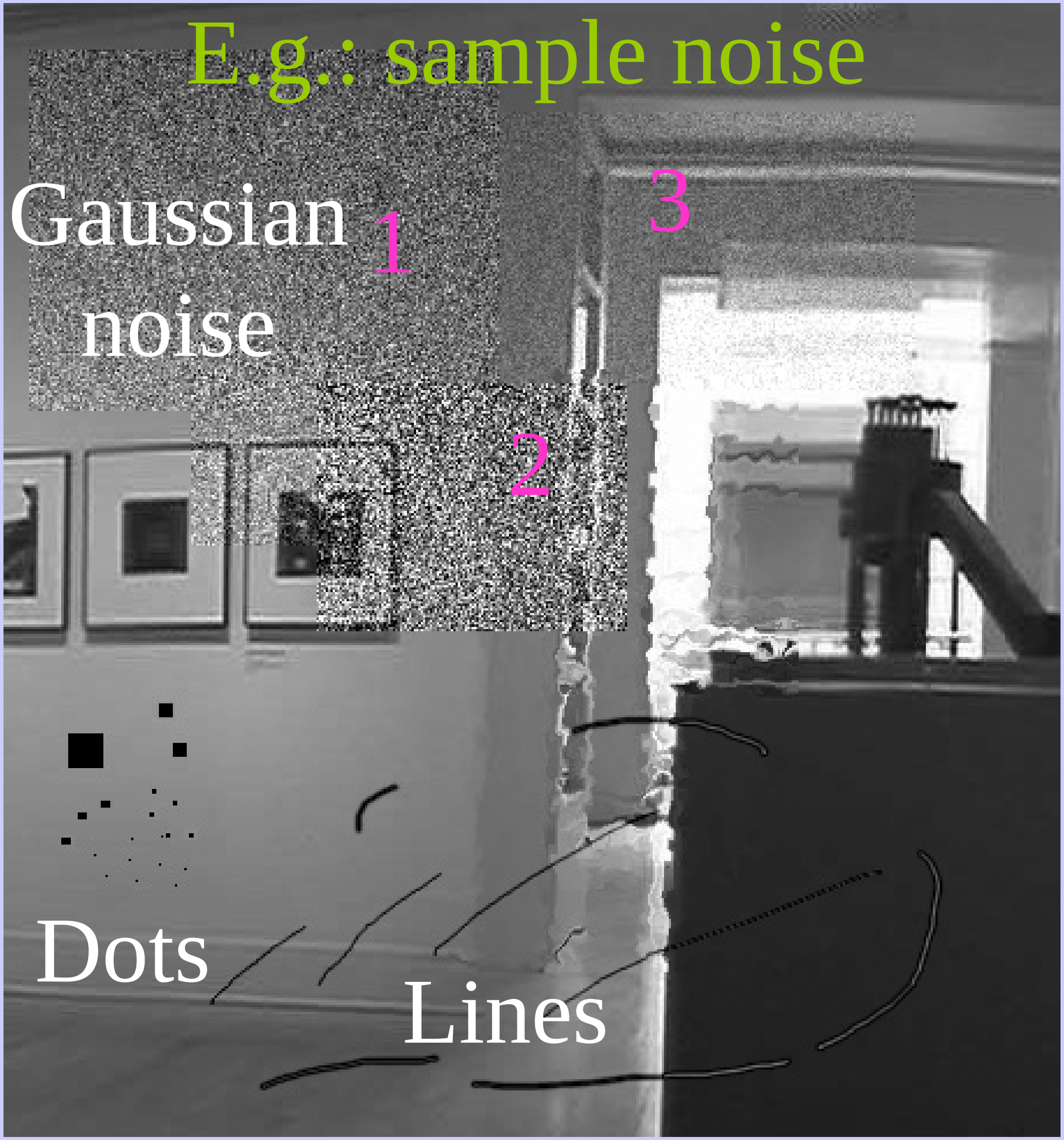


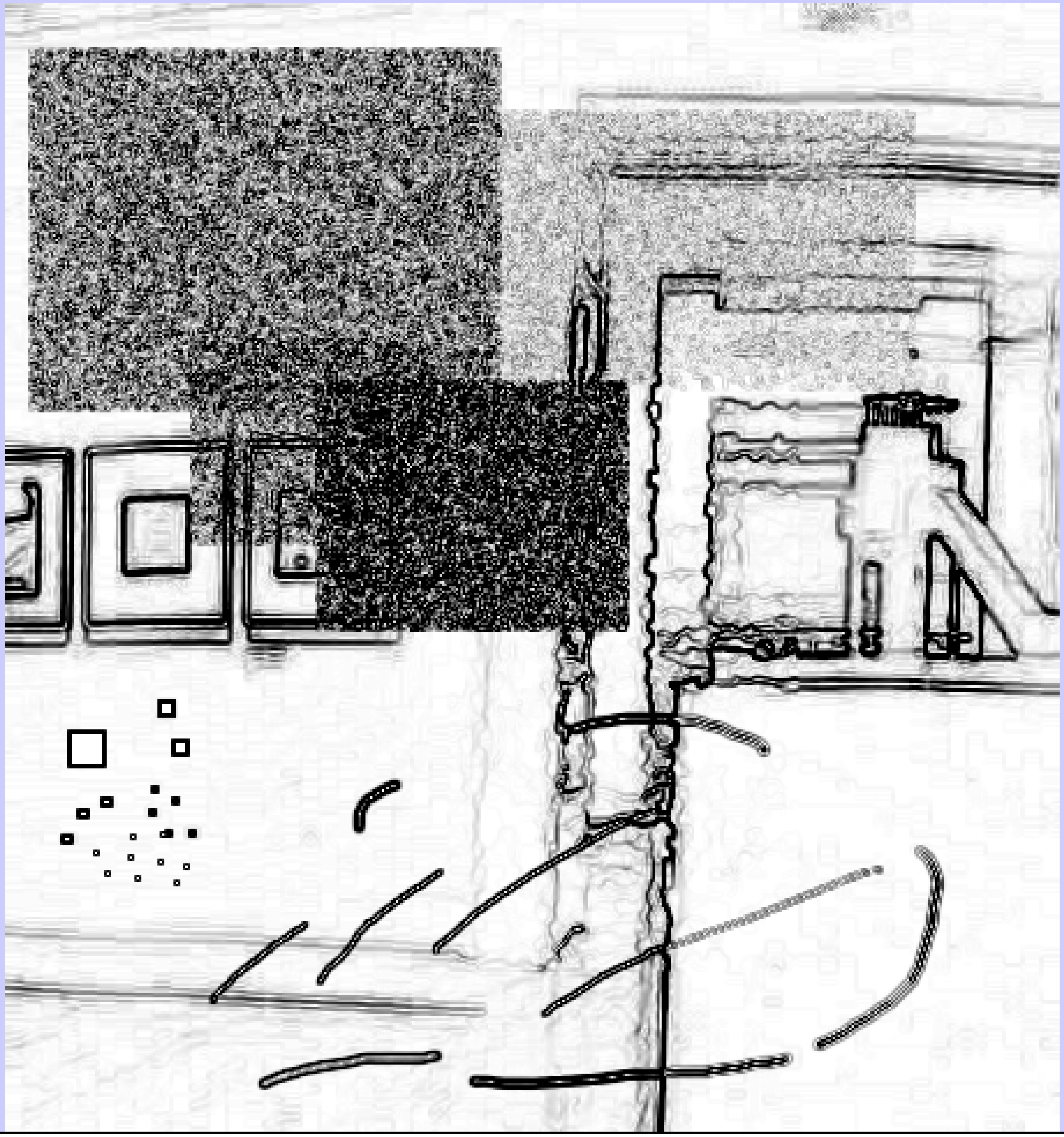
Image Noise

- Images usually are corrupted by several types of "noise"

- Digital noise
- Shadows
- Shiny spots (specularities)
- Camera irregularities
- Bad assumptions about what's being computed ("model noise").


E.g.: sample noise
Gaussian noise
1
2
3
Dots
Lines

Edge detection: trickier than it seems

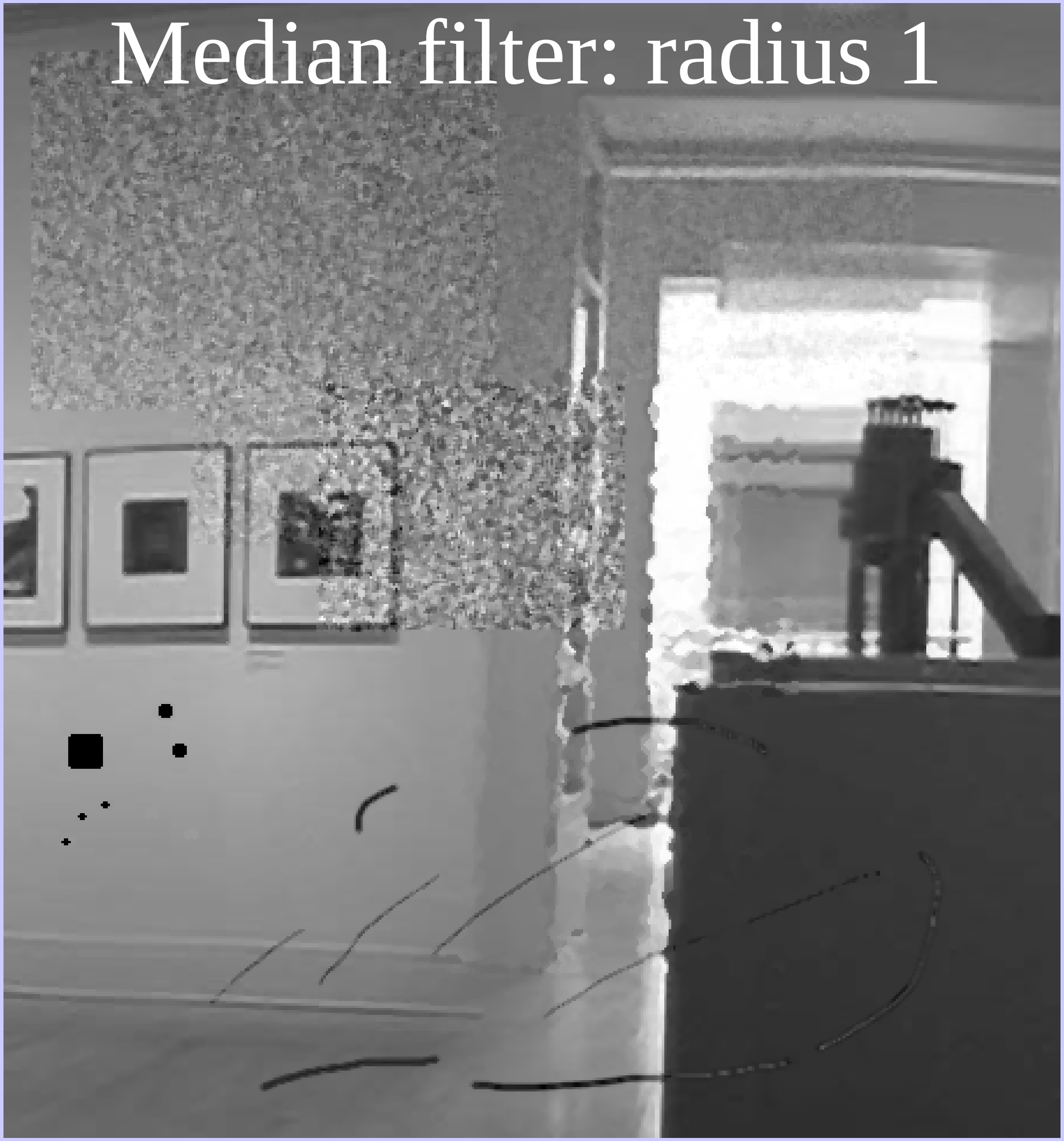


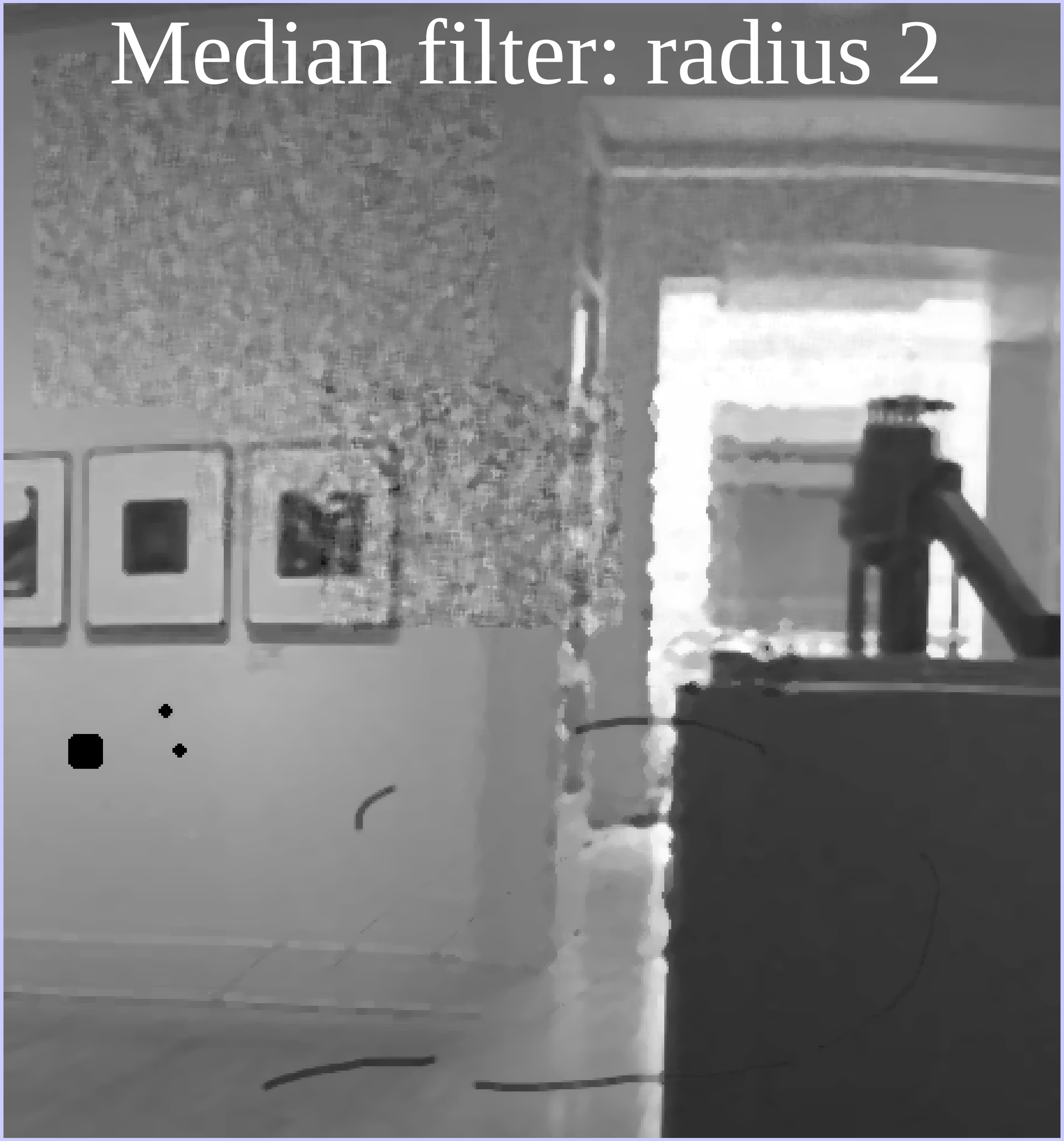
Example: Median Filtering

- A classical non-linear filter.

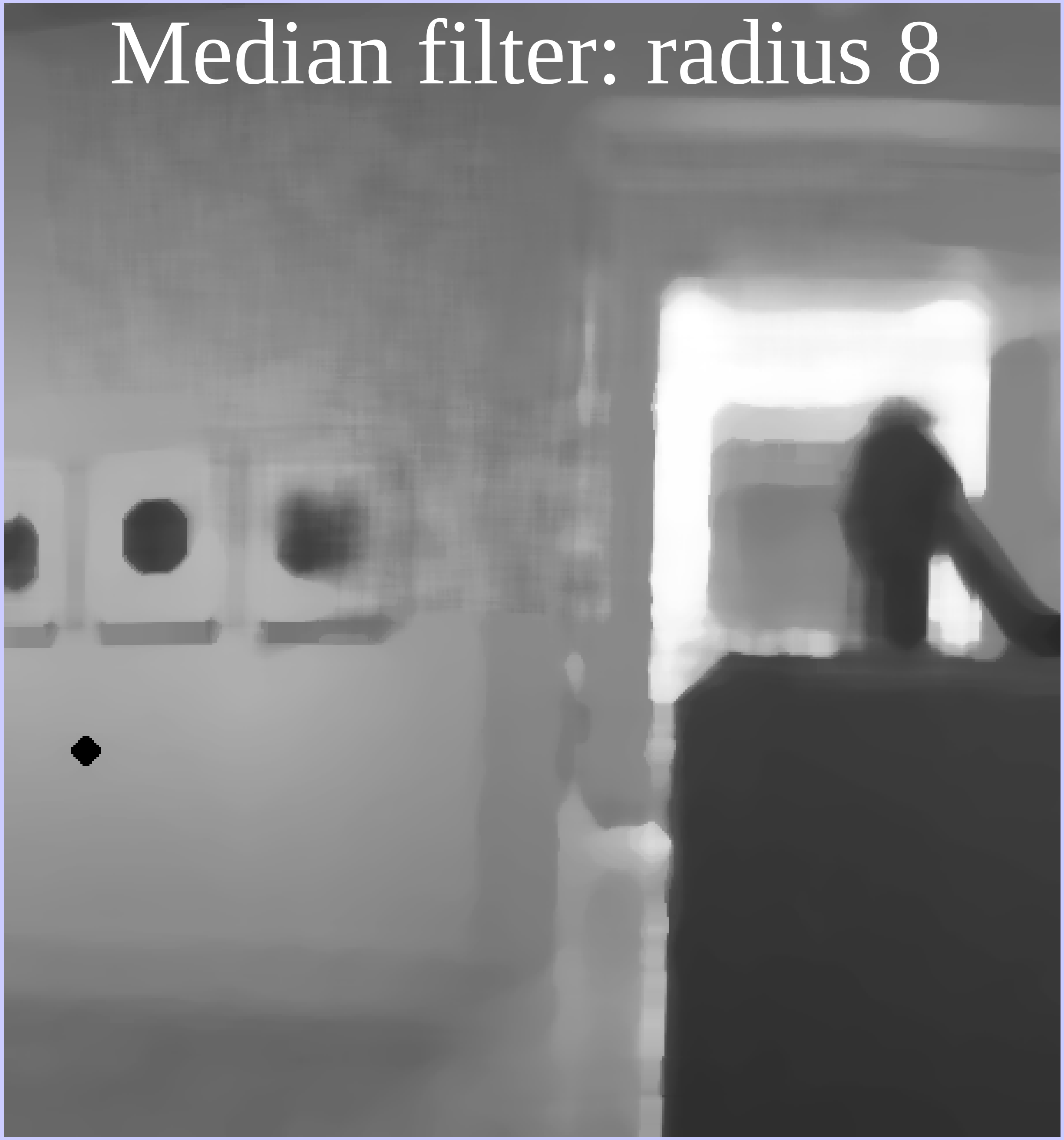
- Over a window, compute the median value of the signal.
- This is the value of the filter.

- This can be considered a non-linear form of averaging.
  – Note it never produces values that weren’t already in the input!


Median filter: radius 1


Median filter: radius 2


Median filter: radius 4


Median filter: radius 8

Notes