1. Apertures and the notion of scale

Nothing that is seen is perceived at once in its entirety.
Euclid (~300 B.C.), Theorem I

1.1 Observations and the size of apertures

Observations are always done by *integrating* some physical property with a measurement device. Integration can be done over a spatial area, over an amount of time, over wavelengths etc. depending on the task of the physical measurement. For example, we can integrate the emitted or reflected light intensity of an object with a CCD (charge-coupled device) detector element in a digital camera, or a grain in the photographic emulsion in a film, or a photoreceptor in our eye. These 'devices' have a sensitive area, where the light is collected. This is the *aperture* for this measurement. Today's digital cameras have several million 'pixels' (*pic*ture *el*ements), very small squares where the incoming light is integrated and transformed into an electrical signal. The size of such pixels/apertures determines the maximal sharpness of the resulting picture.

An example of integration over time is sampling of a temporal signal, for example with an analog-digital converter (ADC). The integration time needed to measure a finite signal is the size of the *temporal aperture*. We always need a *finite* integration area or a *finite* integration time in order to measure a signal. It would be nice to have infinitely small or infinitely fast detectors, but then the integrated signal is zero, making it useless.

Looking with our visual system is making measurements. When we look at something, we have a range of possibilities to do so. We can look with our eyes, the most obvious choice.

We can zoom in with a microscope when things are too small for the unaided eye, or with a telescope when things are just very big.The smallest distance we can see with the naked eye is about 0.5 second of arc, which is about the distance between two neighboring cones in the center of our visual field. And, of course, the largest object we can see fills the whole retina.

It seems that for the eye (and any other measurement device) the *range* of possibilities to observe certain sizes of objects is *bounded* on two sides: there is a minimal size, about the size of the smallest aperture, and there is a maximal size, about the size of the whole detector array.

Spatial resolution is defined as the diameter of the local integration area. It is the size of the *field of view* divided by the number of samples taken over it. The spatial resolution of a Computer Tomography (CT) scanner is about 0.5 mm, which is calculated from the measurement of 512 samples over a field of view with a diameter of 25 cm.

The *temporal resolution* of a modern CT scanner is about 0.5 second, which is 2 images per second.

It seems that we are always trying to measure with the highest possible sharpness, or highest resolution. Reasons to accept lower resolution range from costs, computational efficiency, storage and transmission requirements, to the radiation dose to a patient etc. We can always reduce the resolution by taking together some pixels into one, but we cannot make a coarse image into a sharp one without the introduction of extra knowledge.

The resulting measurement of course strongly depends on the size of the measurement aperture. We need to develop strict criteria that determine objectively what aperture size to apply. Even for a fixed aperture the results may vary, for example when we measure the same object at different distances (see figure 1.1).

```
<< FrontEndVision`FEV`;
Show[GraphicsArray[
   {{Import["cloud1.gif"]}, {Import["cloud2.gif"]}}], ImageSize → 400];
```

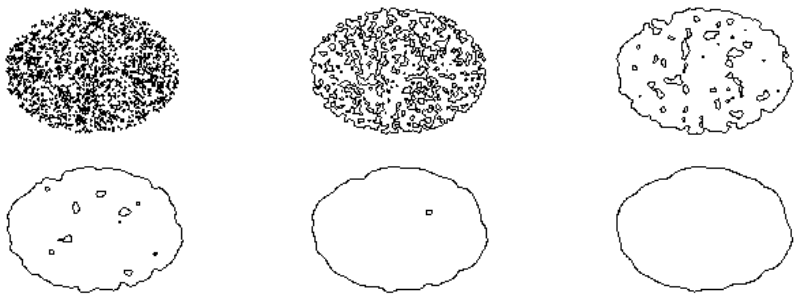

Figure 1.1 A cloud observed at different scales, simulated by the blurring of a random set of points, the 'drops'. Adapted from [Koenderink1992a].

1.2 Mathematics, physics, and vision

In *mathematics* objects are allowed to have no size. We are familiar with the notion of points, that really shrink to zero extent, and lines of zero width. No metrical *units* (like meters, seconds, amperes) are involved in mathematics, as in physics.

Neighborhoods, like necessary in the definition of differential operators, are taken into the limit to zero, so for such operators we can really speak of *local operators*. We recall the definition for the derivative of $f(x)$: $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, where the limit makes the operation confined to a mathematical point.

In *physics* however this is impossible. We saw before that objects live on a bounded range of scales. When we measure an object, or look at it, we use an instrument to do this observation

(our eye, a camera) and it is the range that this instrument can see that we call the scale range. The scale range is bounded on two sides:
- the smallest scale the instrument can see is the *inner scale*. This is the smallest sampling element, such as a CCD element in a digital camera, rod or cone on our retina;

- the largest scale the instrument can see is the *outer scale*. This is the field of view. The dimension is expressed as the ratio between the outer scale and the inner scale, or how often the inner scale fits into the outer scale. Of course the bounds apply both to the detector and the measurement: an image can have a 2D dimension of 256 x 256 pixels.

Dimensional units are *essential* in physics: we express any measurement in dimensional units, like: 12 meters, 14.7 seconds, 0.02 candela$/m^2$ etc. When we measure (observe, sample) a physical property, we need to choose the 'stepsize' with which we should investigate the measurement. We scrutinize a microscope image in microns, a global satellite image in kilometers. In measurements there is no such thing as a physical 'point': the smallest 'point' we have is the physical sample, which is defined as the *integrated* weighted measurement over the detector area (which we call the aperture), where area is *always finite*.

How large should the sampling element be? It depends on the task at hand in what scale range we should measure: "Do we like to see the leaves or the tree"? The range of scales applies not only to the objects in the image, but also to the scale of the features. In chapter 5 we discuss in detail many such features, and how they can be constructed. We give just one example here: in figure 1.2 we see a *hierarchy* in the range of scales, illustrated here for a specific feature (the gradient).

```
im = Import["Utrecht256.gif"][[1, 1]];
Block[{$DisplayFunction = Identity},
  p1 = ListDensityPlot[im];
  p2 =
   ListDensityPlot[Sqrt[gD[im, 1, 0, #]^2 + gD[im, 0, 1, #]^2] & /@ {1, 2, 4}];
Show[GraphicsArray[Prepend[p2, p1]], ImageSize → 500];
```

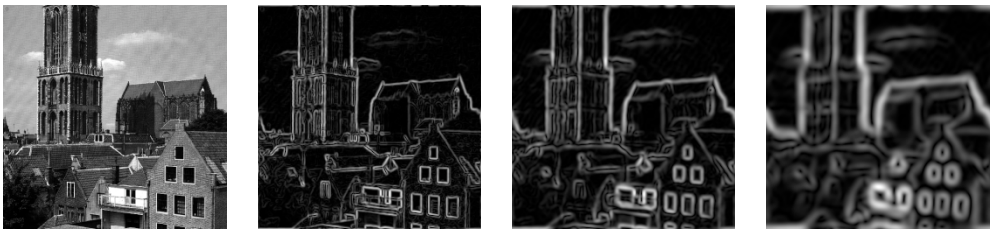

Figure 1.2 Picture of the city of Utrecht. The right three pictures show the *gradient*: the strength of borders, at a scale of 1, 2 resp. 4 pixels. At the finest scale we can see the contours of almost every stone, at the coarsest scale we see the most important edges, in terms of outlines of the larger structures. We see a *hierarchy* of structures at different scales. The *Mathematica* code and the gradient will be explained in detail in later chapters.

To expand the range say of our eye we have a wide armamentarium of instruments available, like scanning electron microscopes and a Hubble telescope. The scale range known to

humankind spans about 50 decades, as is beautifully illustrated in the book (and movie) "Powers of Ten" [Morrison1985].

```
Show[Import["Powersof10sel.gif"], ImageSize → 500];
```

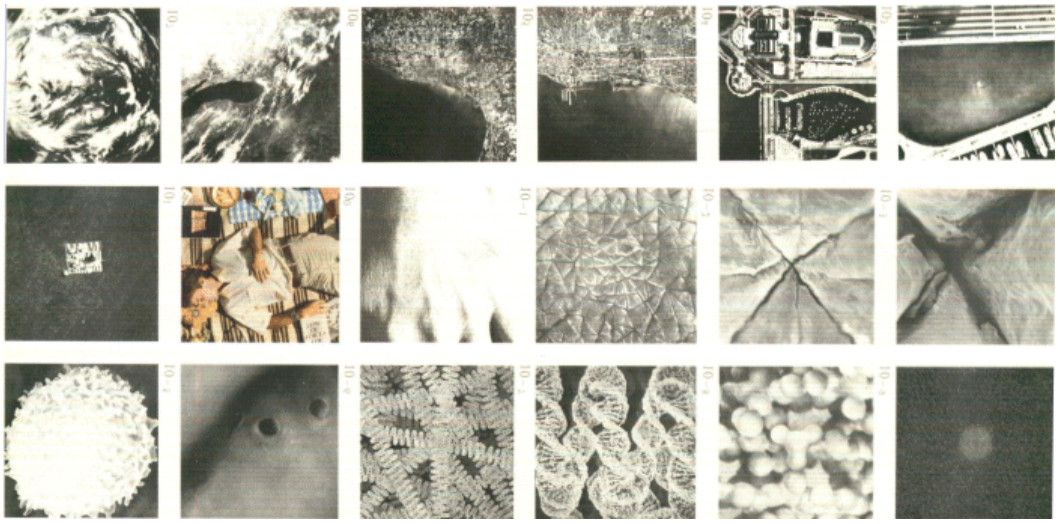

Figure 1.3 Selection of pictures from the journey through scale from the book [Morrison1985], where each page zooms in a factor of ten. Starting at a cosmic scale, with clusters of galaxies, we zoom in to the solar system, the earth (see the selection above), to a picknicking couple in a park in Chicago. Here we reach the 'human' (antropometric) scales which are so familiar to us. We then travel further into cellular and molecular structures in the hand, ending up in the quark structure of the nuclear particles. For the movie see: http://www.micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/index.html.

In *vision* we have a system evolved to make visual observations of the outside world. The *front-end* of the (human) visual system is defined as the very first few layers of the visual system. Here a special representation of the incoming data is set up where subsequent processing layers can start from. At this stage there is no memory involved or cognitive process.

Later we will define the term 'front-end' in a more precise way. We mean the retina, lateral geniculate nucleus (LGN, a small nucleus in the thalamus in our mid-brain), and the primary visual cortex in the back of our head. In the chapter on human vision we fully elaborate on the *visual pathway*.

The front-end sampling apparatus (the receptors in the retina) is designed just to extract multi-scale information. As we will see, it does so by applying sampling apertures, at a wide range of sizes simultaneously.

There is no sampling by individual rods and cones, but by well-structured assemblies of rods and cones, the so-called '*receptive fields*'.

In chapters 6 - 9 we will study the neuroanatomy of the human front-end visual system in more detail. The concept of a receptive field was introduced in the visual sciences by

Hartline [Hartline1940] in 1940, who studied single fibre recordings in the horseshoe crab (Limulus polyphemus).

Psychophysically (psychophysics is the art of measuring the performance of our perceptual abilities through perceptual tasks) it has been shown that when viewing sinusoidal gratings of different spatial frequency the threshold modulation depth is constant (within 5%) over more than two decades.

This indicates that the visual system is indeed equipped with a large range of sampling apertures. Also, there is abundant electro-physiological evidence that the receptive fields come in a wide range of sizes. In the optic nerve leaving each eye one optic-nerve-fibre comes from one receptive field, not from an individual rod or cone.

In a human eye there are about 150 million receptors and one million optic nerve fibres.So a typical receptive field consists of an *average* of 150 receptors. Receptive fields form the elementary 'multi-scale apertures' on the retina. In the chapter on human vision we will study this neuroanatomy in more detail.

1.3 We blur by looking

Using a larger aperture reduces the resolution. Sometimes we exploit the blurring that is the result of applying a larger aperture. A classical example is *dithering*, where the eye blurs the little dots printed by a laser printer into a multilevel greyscale picture, dependent on the density of the dots (see figure 1.4).

It nicely illustrates that we can make quite a few different observations of the same object (in this case the universe), with measurement devices having different inner and outer scales. An atlas, of course, is the canonical example.

```
Show[GraphicsArray[{Import["Floyd0.gif"], Import["Floyd1.gif"]}],
  ImageSize -> 330];
```

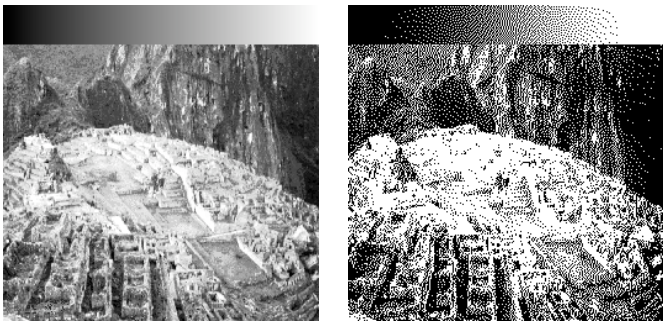

Figure 1.4 Dithering is the representation of grayvalues through sparse printing of black dots on paper. In this way a tonal image can be produced with a laserprinter, which is only able to print miniscule identical single small high contrast dots. Left: the image as we observe it, with grayscales and no dithering. Right: Floyd-Steinberg dithering with random dot placements. [From http://sevilleta.unm.edu/~bmilne/khoros/html-dip/c3/s7/front-page.html].

A priori we have to decide on how large we should take the inner scale. The front-end vision system has no knowledge whatsoever of what it is measuring, and should be open-minded with respect to the size of the measurement aperture to apply.

```
Show[Import["wales-colordither.gif"], ImageSize -> 400];
```

Figure 1.5 En example of color-dithering in image compression. Left: the original image, 26 KByte. Middle: color dithering, effective spreading of a smaller number of color pixels so that the blurring of our perception blends the colors to the same color as in the original. Filesize 16 Kbyte. Right: enlargement of a detail showing the dithering. From http://www.digital-foundry.com/gif/workshop/dithering.shtml.

As we will see in the next section, the visual front-end measures at a multitude of aperture sizes *simultaneously*. The reason for this is found in the world around us: objects come at all sizes, and at this stage they are all equally important for the front-end.

```
Show[Import["Edlef Romeny - cherry trees.jpg"], ImageSize -> 280];
```

Figure 1.6 In art often perceptual clues are used, like only coarse scale representation of image structures, and dithering. Painting by Edlef ter Haar Romeny [TerHaarRomeny2002b]. Owned by the author.

```
im = Import["mona lisa face.gif"][[1, 1]];
imr1 = Table[Plus @@ Plus @@ Take[im, {y, y + 9}, {x, x + 9}],
  {y, 1, 300, 10}, {x, 1, 200, 10}];
imr2 = Table[Plus @@ Plus @@ Take[im, {y, y + 14}, {x, x + 9}],
   {y, 1, 290, 15}, {x, 1, 200, 10}];
DisplayTogetherArray[ListDensityPlot /@ {im,
    Join @@ Table[MapThread[Join, Table[imr2 imr1[[y, x]], {x, 1, 20}]],
{y, 1, 30}]}, ImageSize → 250];
```

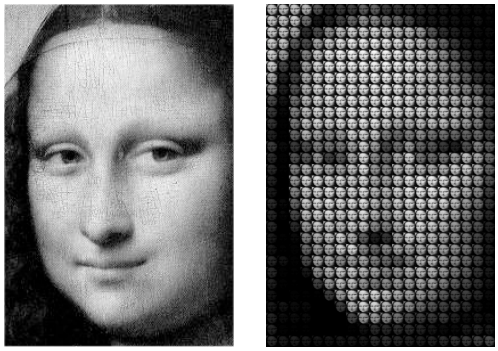

Figure 1.7 Image mosaic of the Mona Lisa. Image resolution 200x300 pixels. The image is subsampled to 20x30 samples, whose mean intensity modulates a mosaic of the subsampled images.

And that, in a natural way, leads us to the notion of multi-scale observation, and multi-scale representation of information, which is intrinsically coupled to the fact that we can observe in so many ways. The size of the aperture of the measurement will become an extra continuous measurement dimension, as is space, time, color etc. We use it as a *free parameter:* in first instance we don't give it a value, it can take any value.

▲ Task 1.1 Experiment with dithering with circular disks of proper size in each pixel. Calculate the area the disk occupies. Some example code to get started:
```
Show[Graphics[Table[Disk[{x,y},.3+im[[y,x]]/2048],{y,1,128}
,{x,1,128}]],AspectRatio→Automatic];
```

▲ Task 1.2 Experiment with dithering with randomly placed small dots in each pixel.

Mosaics, known since Roman times, employ this multiresolution perceptive effect. There is also artistic interest in replacing a pixel by a complete image (see e.g. figure 1.7). When random images with appropriate average intensity and color (and often intensity gradient) are chosen the technique is called an *image mosaic*.

▲ Task 1.3 One can play with other graphical elements, e.g. text (BasicBlock:→ (Text["FEV", #1,#2]&)) etc. Note that despite the structure in the dithering elements, we still perceive the large scale structure unchanged in depth.

It turns out that there is a very specific reason to *not only* look at the highest resolution. As we will see in this book, a new world opens when we consider a measurement of the outside world at all these sizes simultaneously, at a whole range of sharpnesses. So, not only the smallest possible pixel element in our camera, but a camera with very small ones, somewhat

larger ones, still larger ones and so on. It turns out that our visual system takes this approach. The stack of images taken at a range of resolutions is called a *scale-space*.

Another interesting application of dithering is in the generation of random dot stereograms (RDS), see figure 1.8.

```
Options[RDSPlot] = {BasicBlock :> (Rectangle[#1 - #2, #1 + #2] &)};
RDSPlot[expr_, {x_, xmin_, xmax_}, {y_, ymin_, ymax_}, opts___] :=
  Block[{pts = 120, periods = 6, zrange = {-1, 1}, density = .4, depth = 1,
    basicblock = BasicBlock /. {opts} /. Options[RDSPlot], guides = True,
    strip, xpts, ypts, dx, dy, xval, yval, zmin, zmax, exprnorm},
   {zmin, zmax} = zrange; {xpts, ypts} = If[Length[pts] == 2, pts, {pts, pts}];
   dy = (ymax - ymin) / ypts; dx = (xmax - xmin) / xpts; strip = Floor[xpts / periods] dx;
   exprnorm = (.25 depth (xmax - xmin) / (periods (zmax - zmin))) *
     (Max[zmin, Min[zmax, expr]] - (zmax + zmin) / 2);
   Graphics[{RDSArray[basicblock, {dx, dy} / 2, Flatten[Table[If[Random[] < density,
         Thread[{rdsimages[exprnorm /. y → yval, {x, xval, xmax, strip}], yval}], {}],
       {yval, ymin + .5 dy, ymax, dy}, {xval, xmin + .5 dx,
        rdsimage[exprnorm /. y → yval, {x, xmin, strip}], dx}], 2]],
     If[guides, makeguides[{.5 xmax + .5 xmin, 1.1 ymin - .1 ymax}, .5 strip], {}]],
    Sequence @@ Select[{opts}, ! MemberQ[First /@ Options[RDSPlot], First[#]] &]]];

rdsimage[expr_, {x_, xval_, dx_}] := xval + dx - N[expr /. x → xval + dx / 2];
rdsimages[expr_, {x_, xval_, xmax_, dx_}] := | If[xval ≤ xmax,
    Prepend[rdsimages[expr, {x, rdsimage[expr, {x, xval, dx}], xmax, dx}], xval], {}];
makeguides[pos_, size_] := Apply[Rectangle,
   Map[pos + size # &, {{{-1.1, -.1}, {-.9, .1}}, {{.9, -.1}, {1.1, .1}}}, {2}], 1];
Unprotect[Display]; Display[channel_, graphics_?(! FreeQ[#, RDSArray] &)] := (Display[
   channel, graphics /. (RDSArray[basicblock_, dims_, pts_] :> (basicblock[#, dims] & /@ pts))];
  graphics); Protect[Display];
```

$$\text{Show}\left[\text{RDSPlot}\left[-\frac{2}{\sqrt{2\,\pi}}\ x\ \text{Exp}\left[-\frac{x^2+y^2}{2}\right],\ \{x,\ -3,\ 3\},\ \{y,\ -3,\ 3\}\right],\right.$$
$$\left.\text{ImageSize} \to 400\right];$$

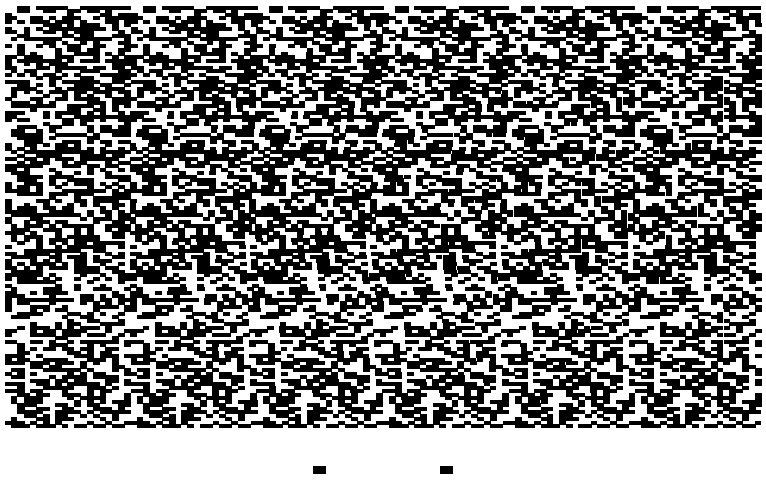

Figure 1.8 Random dot stereogram (of the first derivative with respect to x of the Gaussian function, a function which we will encounter frequently in this book). The dots are replaced by a random draw from the letters A-Z.
Code by Bar-Natan [Bar-Natan1991, www.ma.huji.ac.il/~drorbn/].
See also www.ccc.nottingham.ac.uk/~etzpc/sirds.html. Look with both eyes to a point behind the image, so the dots under the figure blend together. You will then see the function in depth.

See also the peculiar paintings of the Italian painter Giuseppe Arcimboldo (1527-1593). See www.illumin.co.uk/svank/biog/arcim/arcidx.html.

```
Show[Import["Vertumnus.jpg"], ImageSize -> 170];
```

Figure 1.9 Vertumnus (Rudolph II) by Giuseppe Arcimboldo (ca. 1590). Painting in the Skoklosters Slott, Stockholm, Sweden.

1.4 A critical view on observations

Let us take a close look at the process of observation. We note the following:

- Any physical observation is done through an aperture. By necessity this aperture has to be *finite*. If it would be zero size no photon would come through. We can modify the aperture considerably by using instruments, but never make it of zero width. This leads to the fundamental statement: *We cannot measure at infinite resolution*. We only can perceive a 'blurred' version of the mathematical abstraction (infinite resolution) of the outside world.

- In a first 'virginal' measurement like on the retina we like to carry out observations that are *uncommitted*. With uncommitted we mean: not biased in any way, and with no model or any a priori knowledge involved. Later we will fully incorporate the notion of a model, but in this first stage of observation *we know nothing*.

An example: when we know we want to observe vertical structures such as stems of trees, it might be advantageous to take a vertically elongated aperture. But in this early stage we cannot allow such special apertures.

At this stage the system needs to be general. We will exploit this notion of being uncommitted in the sequel of this chapter to the establishment of *linear scale-space theory*.

It turns out that we can express this 'uncommitment' into axioms from which a physical theory can be derived. Extensions of the theory, like nonlinear scale-space theories, follow in a natural way through relaxing these axioms.

- Being uncommitted is a natural requirement for the first stage, but *not* for further stages, where extracted information, knowledge of model and/or task etc. come in. An example: the introduction of feedback enables a multi-scale analysis where the aperture can be made adaptive to properties measured from the data (such as the strength of certain derivatives of the data). This is the field of *geometry-driven diffusion*, a nonlinear scale-space theory. This will be discussed in more detail after the treatment of linear scale-space theory.

```
Show[Import["DottedPR.gif"], ImageSize -> 380];
```

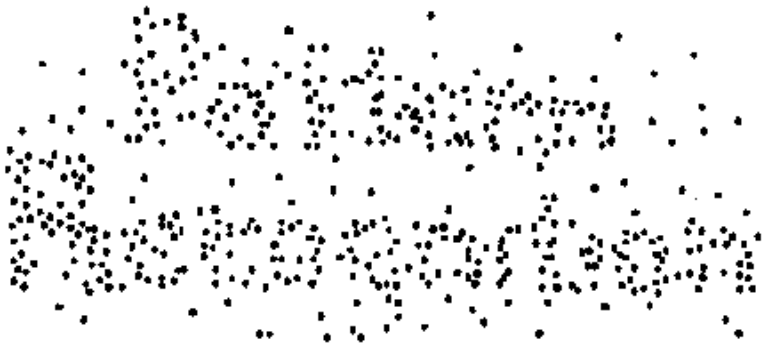

Figure 1.10 At different resolutions we see different information. The meaningful information ir this image is at a larger scale then the dots of which it is made. Look at the image from about 2 meters. Source: dr. Bob Duin, Pattern Recognition Group, Delft University, the Netherlands.

- A *single constant size* aperture function may be sufficient in a controlled physical application. An example is a picture taken with a camera or a medical tomographic scanner, with the purpose to replicate the pixels on a screen, paper or film without the need for cognitive tasks like recognition. Note that most man-made devices have a single aperture size. If we need images at a multiple of resolutions we simply blur the images after the measurement.

- The human visual system measures at multiple resolutions *simultaneously*, thus effectively adding scale or resolution as a measurement dimension. It measures a *scale-space* $L(x, y; \sigma)$, a function of space (x, y) and scale σ, where L denotes the measured parameter (in this case luminance) and σ the size of the aperture. In a most general observation *no* a priori size is set, we just don't know what aperture size to take. So, in some way control is needed: we could apply a whole range of aperture sizes if we have no preference or clue what size to take.

- When we observe noisy images we should realize that noise is always part of the observation. The term 'noisy image' already implies that we have some idea of an image with structure 'corrupted with noise'. In a measurement noise can only be separated from the observation if we have a model of the structures in the image, a model of the noise, or a model of both. Very often this is not considered explicitly.

```
im = Table[If[11 < x < 30 && 11 < y < 30, 1, 0] + 2 Random[], {x, 40}, {y, 40}];
ListDensityPlot[im, FrameTicks -> False, ImageSize -> 120];
```

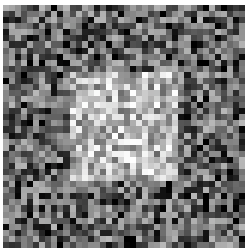

Figure 1.11 A square with additive uniform pixel-uncorrelated noise. Jagged or straigh contours? 'We think it is' or 'it looks like' a square embedded in the noise. Without a model one really cannot tell.

- When it is given that objects are human-made structures like buildings or otherwise part of computer vision's 'blocks world', we may assume straight or smoothly curved contours, but often this is not known.

- Things often go wrong when we change the resolution of an image, for example by creating larger pixels.

- If the apertures (the pixels) are square, as they usually are, we start to see blocky tesselation artefacts. In his famous paper "The structure of images" Koenderink coined this *spurious resolution* [Koenderink1984a], the emergence of details that were not there before, and should not be there. The sharp boundaries and right angles are artefacts of the representation, they certainly are not in the outside world data. Somehow we have created structure in such a process. Nearest neighbour interpolation (the name for pixel replication) is of all interpolation methods fastest but the worst. As a general rule we want the structure only to decrease with increasing aperture.

```
Show[Import["Einsteinblocky.gif"], ImageSize -> 120];
```

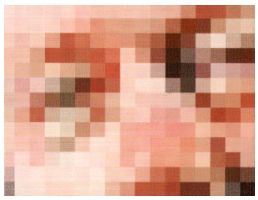

Figure 1.12 Spurious resolution due to square apertures. Detail of a famous face: Einstein Much unintended 'spurious' information has been added to this picture due to the sampling process. Intuitively we take countermeasures for such artefacts by squeezing our eyes and looking through our eyelashes to blur the image, or we look from a greater distance.

- In the construction of fonts and graphics *anti-aliasing* is well known: one obtains a much better perceptual delineation of the contour if the filling of the pixel is equivalent to the physical integration of the intensity over the area of the detector. See figure 1.13 for a font example.

```
Show[Import["anti_alias.gif"], ImageSize -> 250];
```

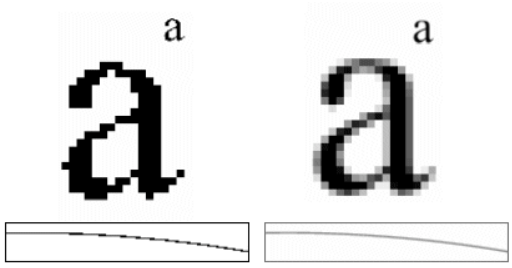

Figure 1.13 Anti-aliasing is the *partial volume effect* at the boundaries of contours. Wher making physically realistic test images for computer vision applications it is essential to take this sampling effect into account.

1.5 Summary of this chapter

Observations are necessarily done through a finite aperture. Making this aperture infinitesimally small is not a physical reality. The size of the aperture determines a hierarchy of structures, which occur naturally in (natural) images. With the help of instruments (telescopes, microscopes) we are able to see a scale-range of roughly 10^{50}. The visual system exploits a wide range of such observation apertures in the front-end simultaneously, in order to capture the information at all scales. Dithering is a method where the blending/blurring through an observation with a finite aperture is exploited to create grayscale and color nuances which can then be created with a much smaller palet of colors.

Observed noise is part of the observation. There is no way to separate the noise from the data if a model of the data, a model of the noise or a model of both is absent. Without a model noise is considered input which also contains structural geometric information, like edges, corners, etc. at all scales.

The aperture cannot take any form. An example of a wrong aperture is the square pixel so often used when zooming in on images. Such a representation gives rise to edges that were never present in the original image. This artificial extra information is called 'spurious resolution'. In the next chapter we derive from first principles the best and unique kernel for an uncommitted observation.