Lecture notes: What is machine vision?

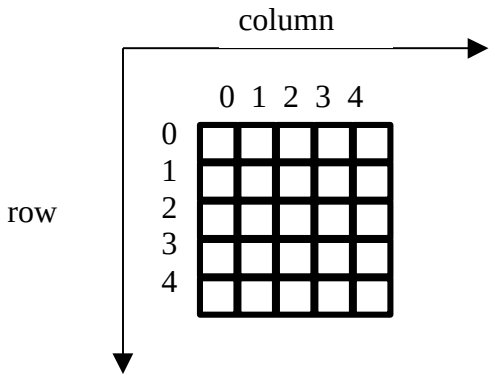
Computers are blind. Machine vision, also known as computer vision, concerns itself with providing sight to computers.

Why? Why would a computer need sight? In order to perform useful tasks, computers have to get input from somewhere. The familiar desktop machine gets its input from a human user, through a keyboard and mouse. Imagine if this was the only way you could get your input! How do you get your input? Eyes, ears, etc.

Sensor type doesn't matter. Machine vision can be accomplished by a variety of sensors, such as cameras, pressure sensors (e.g. microphone), sound sensors (e.g. ultrasound), etc. In this context, the modality of information is not relevant. We do not distinguish between audio, visual, tactile, etc., as these terms are merely human conveniences for describing our own sensing hardware. But machines are not people!

Imaging sensors are the most common. We tend to build image-scanning sensors, probably because vision is arguably the most powerful human sensing modality. In this course we will mostly concern ourselves with image-based machine vision. But we will also study motion sensors (accelerometers and gyroscopes), 3D sensors (range cameras), and others.

What is an image? An image is a 2D array of numbers stored in a raster. Each cell in the raster is called a pixel.



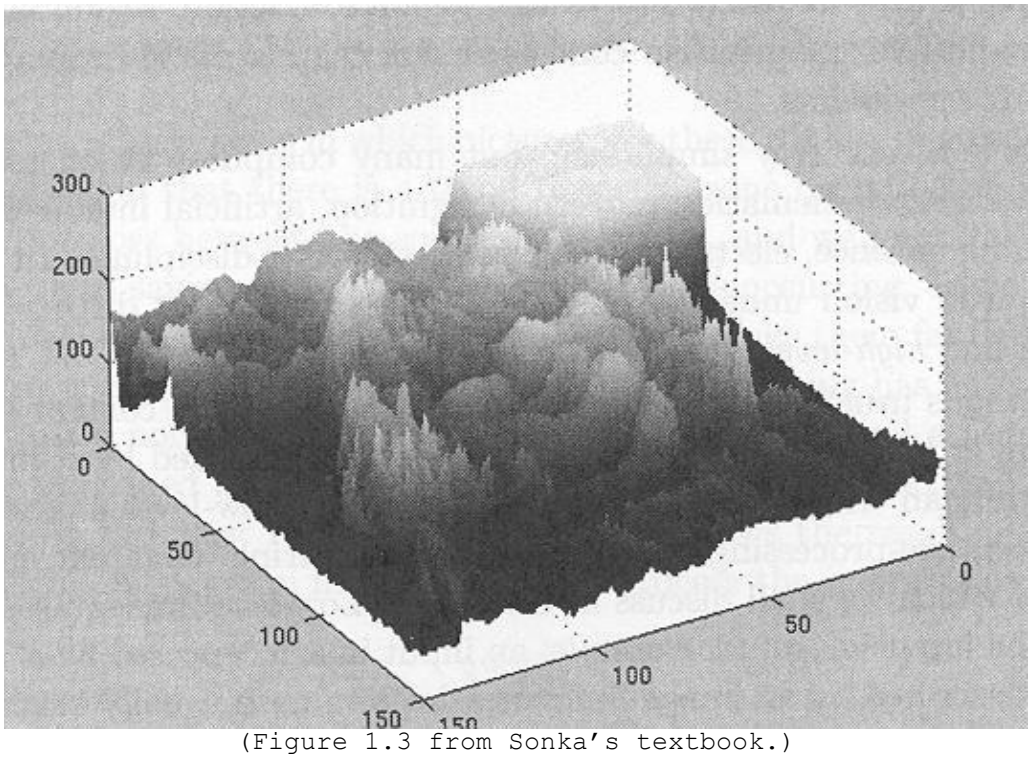
[fill in pixel values for an example]

Pixels hold measurements. Each pixel holds a measurement made by the sensor. For example, a common camera measures the amount (and possibly color) of the light hitting it over a spatial array (the image).

Quantization. A standard is 8 bits per pixel, so that pixel values range from 0 to 255. By convention, 0 represents black, 255 represents white, and the values in-between represent a greyscale. **Note: in this class, we will only be working with grayscale images.**

[live example of image, with magnification, showing pixel values]

What is machine vision? The primary problem in machine vision is interpreting pixel values. For example, look at the following plot of intensity values (Z-axis) of pixels in an image (X- and Y-axes).

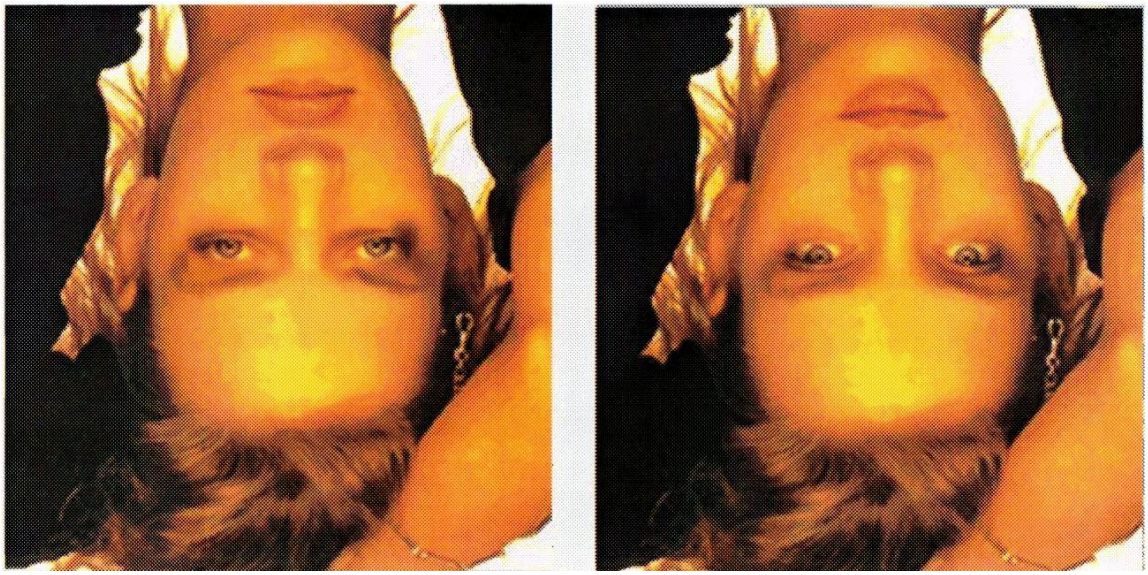


(Figure 1.3 from Sonka's textbook.)

Seems easy, right? There are hundreds of thousands of pixels in a standard-size image. In a medical or satellite image, there may be hundreds of millions of pixels. What does that plot of image data show?


(Figure 1.4 from Sonka's textbook.)

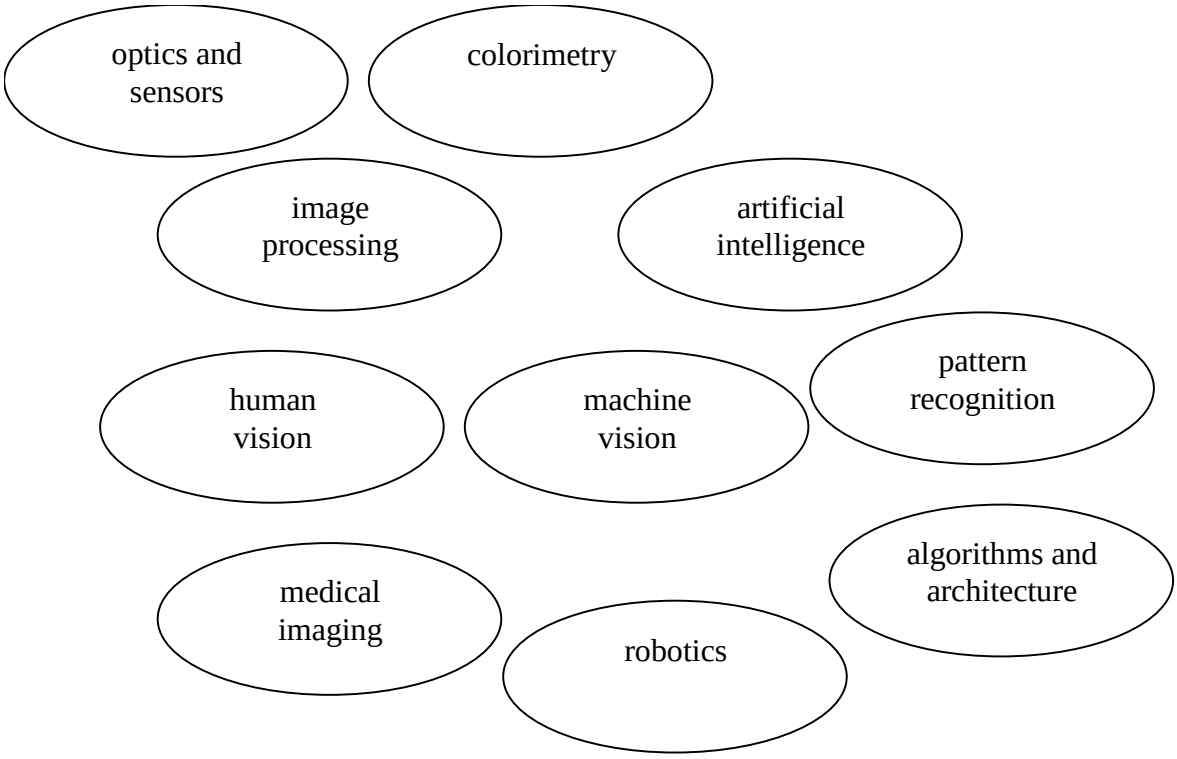
Here the values are displayed using a greyscale for interpretation. It is a testament to the human brain that we so easily understand what we see, because in fact what we are seeing is millions of measurements of light intensity.



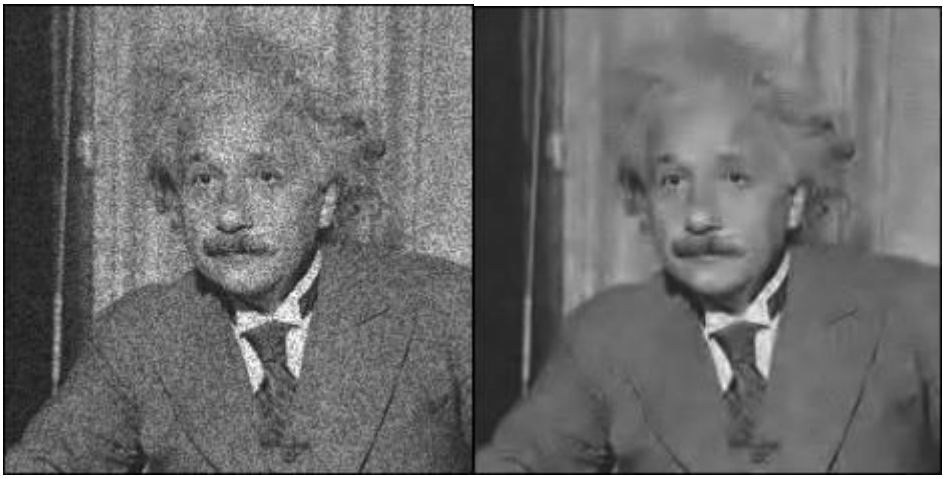
Here is another interesting example of how the human vision system works. In this orientation, it looks like two reasonable pictures of actress Uma Thurman, upside-down. But ...

In this orientation, it quite clearly has something wrong with it! In the image on the left, her eyes and mouth were cut and pasted to appear upside-down. When viewed in the context of a face upside-down, our "vision system" is not capable of seeing the distortion, but in the proper right-side-up orientation, we see it quite clearly. Why? Our vision system is hard-wired to encounter people in a vertical position, and if we see a face in another orientation, it "fixes it" so we can recognize the person (or at least gender, threat, etc.).

Related fields. Machine vision is closely related to and overlaps several other fields.



- Image processing. Usually stops short of understanding the image. Often employed for improved human consumption. For example, noise may be smoothed out, or contrast equalized.



Einstein image, with noise Einstein image, smoothed.

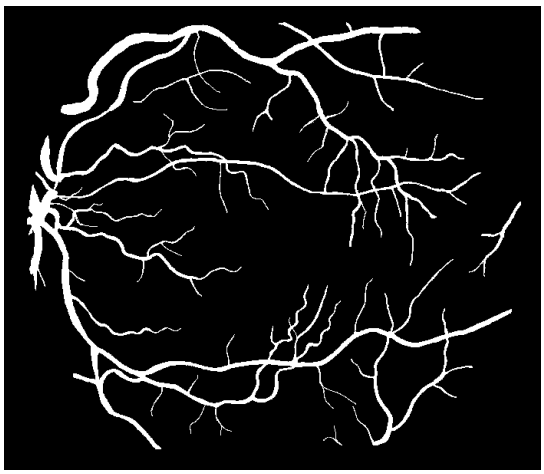
Note however that if a process helps a human see the image better, it probably also helps the machine see it better.

- Optics and sensors. How is the image captured? Can we model the capture process, to help remove noise?

- Colorimetry. How should color be modeled? RGB is just one possibility. Are three values enough? How many bits?

- Pattern recognition. Once a piece of an image is recognized, how is it classified? For example, we might want to find the vessels in an image of the retina:

retinal image



blood vessels

  Finding the vessels is a machine vision problem. Once it is found, we can measure their properties (tortuousity, color, size) and try to classify the vessels as normal or abnormal.

- Artificial intelligence. Once we classify the vessels, can we reason about the overall health of the subject?

- Algorithms and architecture. How long does all this take? Lots of pixels!

- Medical imaging. One of the biggest applications of machine vision is in medicine. The bottom line is automation, repeatable performance of a known level, which is desired in medicine.

- Robotics. What about a machine that can move? But in order to move, we would prefer the machine be able to see, for obvious reasons.

Spheres of influence. Although we will concentrate on machine vision, it is important to understand that all these fields overlap. In fact, part of being a graduate student is beginning to understand how knowledge may be modeled as "spheres of influence". As you investigate a subject, you should be able to discuss your problem within the larger body of knowledge.

Particularly when you write papers or give presentations.

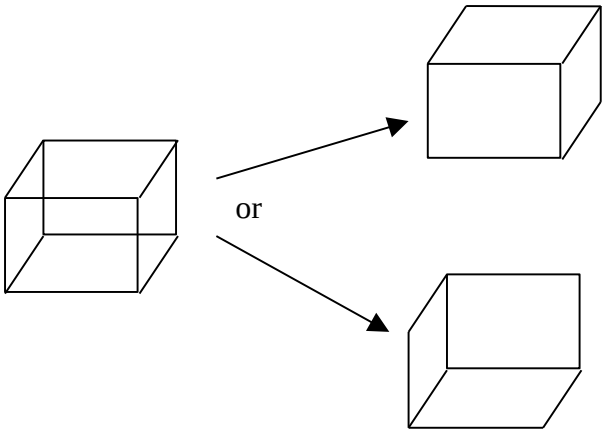
Bag of tools. There is no universal theory of vision. Modern knowledge consists of a collection of tools, methods of trying to understand image content. In this course we will study many of the more useful tools and approaches.

Vision paradigms. Although there is no universal vision theory, there are several paradigms under which research takes place. Each describes a school of thought for interpreting sensed data:

- **Marr's theory (reconstruction).** David Marr was a famous MIT researcher who proposed the first formal framework for vision. He advocated geometric interpretation of each image individually, according to three steps:

  image -> primal sketch (features based upon intensity changes)
  -> 2.5D sketch (depth image)
  -> 3D sketch (geometric model)

  This paradigm remains popular, even though it has been shown how many problems are ill-posed (for example, you simply can't get a unique 3D interpretation of an image):



(From the original either interpretaton is possible.)

- **Active vision.** Active vision proponents consider exploration and interpretation of an image to be intertwined. In this case, an important purpose of the interpretive process is to decide where to look next (to aid in further interpretation).

- **Purposive vision.** In the purposive vision framework, the task controls the interpretation. For example, it may not be important to understand all the contents of an image, but only that portion necessary to accomplish the task.

- **Qualitative vision.** Abandoning geometry, qualitative interpretation seeks only to develop a relational model of the information in an image (the door is next to the walls and floor).

Most of what we do in this class will fall under Marr's reconstructive theory, because that has been around the longest (and hence has the most results). We will however touch upon these other theories when we talk about active contours, object modeling, and object recognition.