Visual Knowledge

For millennia philosophers have debated the precise definition of the concepts of *knowledge* and *seeing*. Around 350 BC Aristotle said, in effect, to *see* is to *know what* is *where* in an *image*. Today we would like to tighten up these definitions, so that a scientist can design visual knowledge into a robot and can test a robot to discover if it possesses visual knowledge. Later we may refine these definitions as our practical experience of constructing and testing robots demands. Others might care to tighten the definitions so that they apply to all possible sentient beings; but we are concerned only with hacking out definitions that are robust enough to suit the practical task of constructing a robot with visual knowledge.

Firstly, let us examine the concept of an *image*. We all know what images are. Images form on our retinas, appear in mirrors, cameras, photographs, televisions, computers, paintings, drawings, tattoos, printed things, embossed things, carvings, holographs, chromatographs, sonographs, siesmographs, embroidery, mosaics, formal gardens, plant and animal markings and, no doubt, many more things that do not spring immediately to mind. The common element to all of these is that there is a spatial arrangement of some quantity. For example, the density of black ink in a newspaper photograph varies across the page, and the intensities of red, green and blue light vary across a television screen. The former of these is what physicists call a *scalar field* and, at moderate spatial resolutions, the latter is a *vector field*. I would allow that any field, that is, any spatial arrangement of any kind of quantity, is an image. I am happy to agree that a human may see an image formed by light falling on the retina of the eye, and may equally well see an image formed by pin pricks arranged in an array on the back. I draw no philosophical distinction between the different physical processes that are involved in forming an image. Specifically, I will allow that the spatial distribution of electric charge in a camera, or in computer memory, forms an image which can be seen by a robot. For me, *any field is an image*, in the technical sense of the word “field” as used by physicists.

Secondly, let us examine the concept of *knowing what is in an image*. Figure 1 shows a drawing of a small section of a flower bed in a garden. Imagine being asked, “What is that?”, when the questioner indicates the part of the garden shown here at the centre of the black circle.

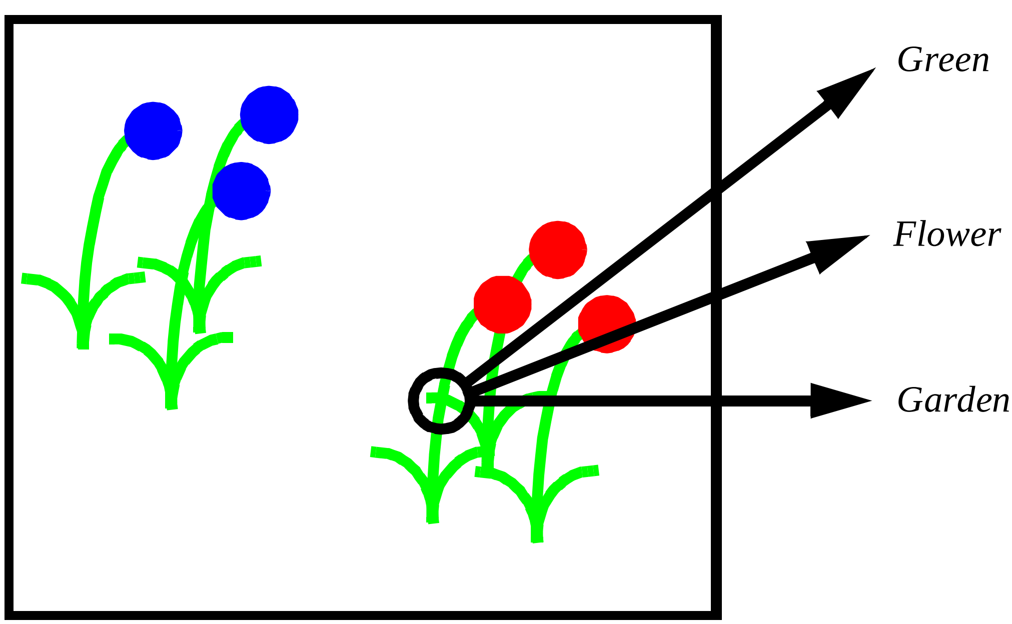

Figure 1: Knowing what is in an image.

If you had recently been discussing painting, you might answer, “It is green”. Alternatively, if you had been discussing plants, you might say, “It is a flower”. Or, if you had been discussing a house, you might say, “It is the garden”. Indeed you might have all of the concepts *green*, *flower*, and *garden* in mind when you answer in any of these ways. That is, all of these concepts might come to mind when that part of the image is indicated, but you might choose to speak about any, all, or none of them. What you say does not affect your knowledge of what is in an image, though a scientist would want some verbal, or other kind of report from you as evidence that you do know what is in the image.

The arrows in the figure indicate a *directed mapping* that runs *from* the part of the image in discussion *to* each of the concepts. The existence of even one mapping in your mind would demonstrate that you know what is in the image. The existence of many mappings would demonstrate that you know many things about the image of the garden. Without digressing into the theory of knowledge expounded by philosophers, let us admit that humans are often wrong in what they claim to know, and can know things other than facts. For example, you might cut the flower at the point indicated and say, “Its for you.” and smile sweetly. The mapping in your mind might then be from the image to knowledge of how to cut flowers and give gifts.

For our rough and ready purpose of designing robots with visual knowledge we may lump all kinds of knowledge, belief, assumption, and skill together in computer programs, or algorithms. That is, we may express the knowledge that the *green* thing is a *flower* or a *garden* in an algorithm that allows a robot to work with these thoughts. Equally an algorithm can perform the cutting of flowers and smiling. Thus an algorithm will suffice as a mathematical model of knowledge for the purpose of defining visual knowledge.

Thirdly let us examine the concept of *knowing where something is in an image*. Figure 2 shows where the concept *green* is portrayed in Figure 1. Green occurs at the precise point under discussion and in all the other points marked green in Figure 2. Thus we may imagine further arrows coming from the concept *green* to every green pixel in the Figure 2. However, this is a slightly mistaken impression of where green occurs in Figure 1, because one of the flowers is missing a leaf.

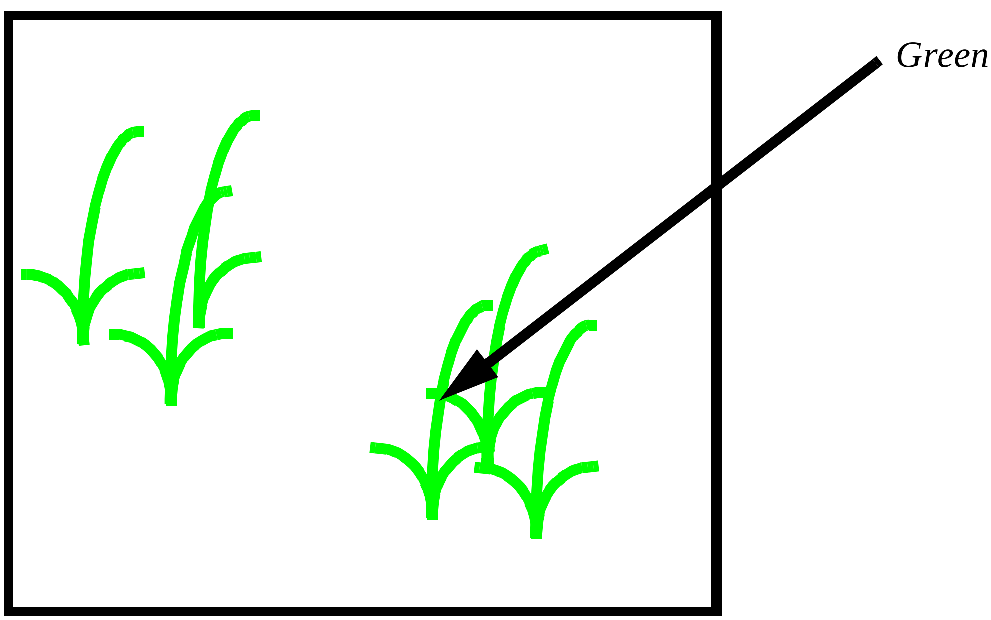

Figure 2: Knowing where *green* is in an image.

Now consider where the flower is. Figure 3 shows the location of the flower under discussion. We would not accept any other flower as proving that a human observer knows where the flower is, because we expect humans to be able to indicate positions with sufficient accuracy to indicate which of the six flowers is under discussion. However, we know that humans are not infinitely accurate and might miss the considerable displacement that we could measure by placing a tracing of Figure 3 on Figure 1. The whole of Figure 3 is displaced upward and to the left, with respect to the black box forming its frame.

Now consider where the garden is. The observer might indicate the whole of the garden shown in Figure 1. We would allow this as proving that the observer knows where the garden is, even though we know that the figure shows only part of the garden - as stated in the introduction to these examples.

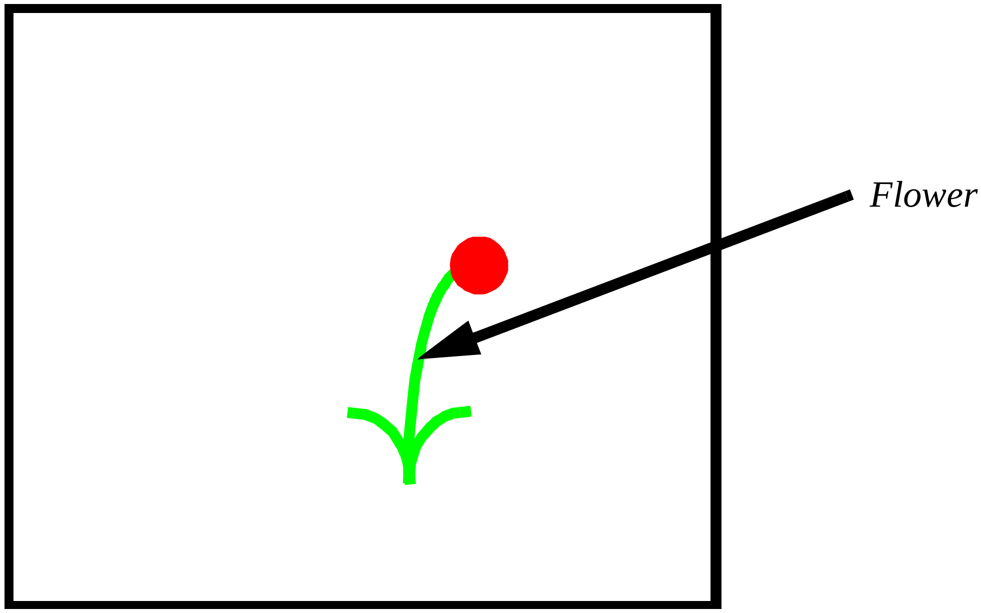

Figure 3: Knowing where the indicated flower is in the image.

By way of further illustration of this point, imagine standing in a garden and looking beyond your feet. What do you see? If you say, “the Earth”, I will agree with you, even though you can see only a tiny part of the Earth. No matter how tall you are, even to the extent of having your head in outer space, you cannot see the whole of the planet Earth. It is a brute mathematical fact that there is no where in three dimensional space where you can see the whole of the planet Earth at the same time. But seeing just part of an object, and identifying it, is held to be seeing the object.

Thus we have seen that, in human terms, the precise geometry of the thing indicated is not essential to knowing where it is. We know that we operate with limited accuracy, with limited memory, and limited perception. The defining characteristic of knowing where something is is to have a directed mapping in mind from the concept to some part, all, or more than the extent of the thing in an image. The *where* part of Aristotle’s definition of *seeing* requires that we can indicate a location in an image.

Drawing all of this together we tighten Aristotle’s definition by defining that *visual knowledge is knowledge which is in a two-way, directed mapping with an image*. Or, for those who prefer mathematical terminology, if not mathematical formalism, *knowledge is an algorithm* and *visual knowledge is an algorithm which is in a two-way directed mapping with a field.*

As one might expect, all computer vision programs ever written are able to answer the *what* question. Though they are all fallible! However, astonishingly few computer vision programs can answer the *where* question. But philosophers since Aristotle, and common sense, require that seeing involves being able to answer the *where* question. In our research on visual consciousness we shall invariably answer both the *what* and the *where* questions, but with varying degrees of detail.