

Dan Ports
2002/10/22
6.034 - TA: Stephen Larson

In *A Framework for Representing Knowledge*, Minsky proposes using frame structures to model human intelligence and learning. He presents a number of different applications for frames, including a representation of human vision. With a few key modifications, his frame-based model of human vision can also be used to represent an infant's vision.

The goal of vision is to transform an image obtained by the eye into a symbolic representation of the objects in view. Minsky proposes doing so by dividing an image into its component parts. Thus one must begin with an image, represented by a frame. As a next step, the image must be divided into its components: boundary lines and surfaces. To accomplish this, each line that is a boundary between two surfaces can be identified and represented as a subframe; these subframes are attached to the image-frame that they are components of. Next, these lines can be used to identify the surfaces they bound. Each surface can be represented by a frame connected to the line frames that define its boundaries. One can then observe and describe a number of properties about each surface: for example, its color, size, and texture, and most importantly its approximate three-dimensional position relative to the observer. Minsky's example of perception of a cube explains how this approximate position can be determined using parallax and depth perception to find the distance of each surface. Then the surfaces can be combined into objects based on their position: for example, the individual faces of a cube can be connected into a cube shape. This creates a shape-frame that is connected to the surface-frames that it is made up of. When the observer moves and his perspective changes, a new image frame is created. The same process of identifying boundaries and surfaces is repeated; if new boundaries and surfaces are found, new frames are created for them. It may then be possible to associate new surface frames with previously existing shape frames, just as Minsky describes in his example of rotating a cube to reveal a previously hidden face.

In this respect, the process of transforming a two-dimensional image into the three-dimensional shapes it represents is essentially identical for infant and adult humans. There are a few minor differences: for example, an adult would be able to easily move to a new location to change perspective and view previously hidden surfaces and objects, whereas for a young infant, not yet mobile, this will be a non-trivial task. However, the major difference between an adult and an infant is in translating physical shapes into an understanding of the objects that they represent.

Consider Minsky's example of the chair. Both an adult and a child can use the process described above to observe the physical shape of the chair: a horizontal seat, a vertical back, and a set of legs. The adult, however, will be able to immediately identify the object by name — it is called a "chair" — and conclude that it is a piece of furniture that can be sat on. He can do so because of his pre-existing knowledge of such objects. An infant, lacking this repository of knowledge and experience, will see the shape of the object but will not be able to associate it with a name, a purpose, or any other higher-level information. In the extreme case, the child may never have seen a chair before at all.

For the adult, vision largely involves using knowledge to understand objects. By contrast,

an infant is primarily going through the process of accumulating knowledge. This knowledge can also be represented by frames. Upon seeing an object, one can create an object frame to describe its properties. One critical property is the object's shape. This terminal links the object-frame with the shape-frame: the chair is an object with a certain particular shape. Many sorts of other information can be included in the object-frame, depending on the object: for example, its purpose, its composition, where it is usually found, etc. One could also represent a memory of an event in a frame, then associate the frame with the object-frames of any entities involved in the event. Thus, upon seeing a chair for the first time, the child can create an object-frame containing everything it has observed and learned about the chair. When he next comes across an object with the shape of the chair, he will be able to associate the shape with the chair object's frame, and recall everything he had learned before. He may also recall the experience-frames, if any, that represent the memory of any events he has associated with the chair object-frame.

Minsky's paper outlines how frames can be used to represent human vision by dividing an image into its component parts, then associating each with the objects they represent. For an infant, the process is largely the same. The critical difference is the increased emphasis on learning, as the infant does not already possess very much knowledge.