Motion Segmentation:
A Biologically Inspired Bayesian Approach

Greg Corrado
gcorrado@stanford.edu

Motion Segmentation is the attribution of motion information to elements in a visual scene. Here we propose a motion segmentation algorithm that uses the properties of motion sensitive neurons as inspiration for its fundamental units. Each unit is treated as a local Bayesian estimator embedded in a larger belief network. The heuristics that we employ to drive segmentation in this network is that there should be at most one motion at each point in space.

Introduction:

Motion is one of the best understood aspects of visual processing in the brain. We know a lot about how motion is perceived by biological visual systems [Adelson & Movshon 1982]. We know a surprising amount about how motion information is represented in the brain [Albright 1993]. We even have some good ideas about how biological systems might extract motion information from the visual scene [Adelson & Bergen 1985, Simoncelli 1993]. But we know shamefully little about how motion information is combined across space and time to construct a "motion scene."

To interpret the visual world, any system must parse the continuous stream of sensory input into distinct salient elements. This is true in the simplest case of foreground background segregation all the way through the difficult task of object recognition. It is no different in the realm of motion vision. To estimate the motion of objects well, information must be combined across space and time - but not across sources. Motion elements must be segmented from each other and from the motion of the background to allow a system to make intelligent judgments about its environment. But how? Can we use what we know about biological vision to construct an algorithm?

A particularly compelling example of the need for this algorithm is the case of motion transparency. Motion transparency occurs whenever two motions appear in the same region of space. This might happen in a reflection on a pane of glass, or in the translating specularities of a moving reflective object, or in the jungle as the shadows of leave shimmer across the form of an approaching predator. In these cases it is the motion itself that allows us to separate one object from another. Because this is such a crisp example the challenges of motion segmentation, we will use a synthetic example of motion transparency to explore our algorithm. In addition we will use human perception of this example as inspiration for how to construct our algorithm.

Broadly, our approach will be to construct an interconnected network of local motion processors. The characteristics of each processing unit will be taken very directly from known properties of neurons in are MT of the primate visual cortex – the apparent seat of motion processing in the human brain. We will interpret the output of these unit as a Bayesian estimate of the probability of local motion given the images, and use belief propagation to reason about the motion scene.

The goal of this approach is really two fold. Not only might we construct a algorithm which successfully segments a motion scene, we might generate testable hypothesis for future neurophysiological studies of motion processing.

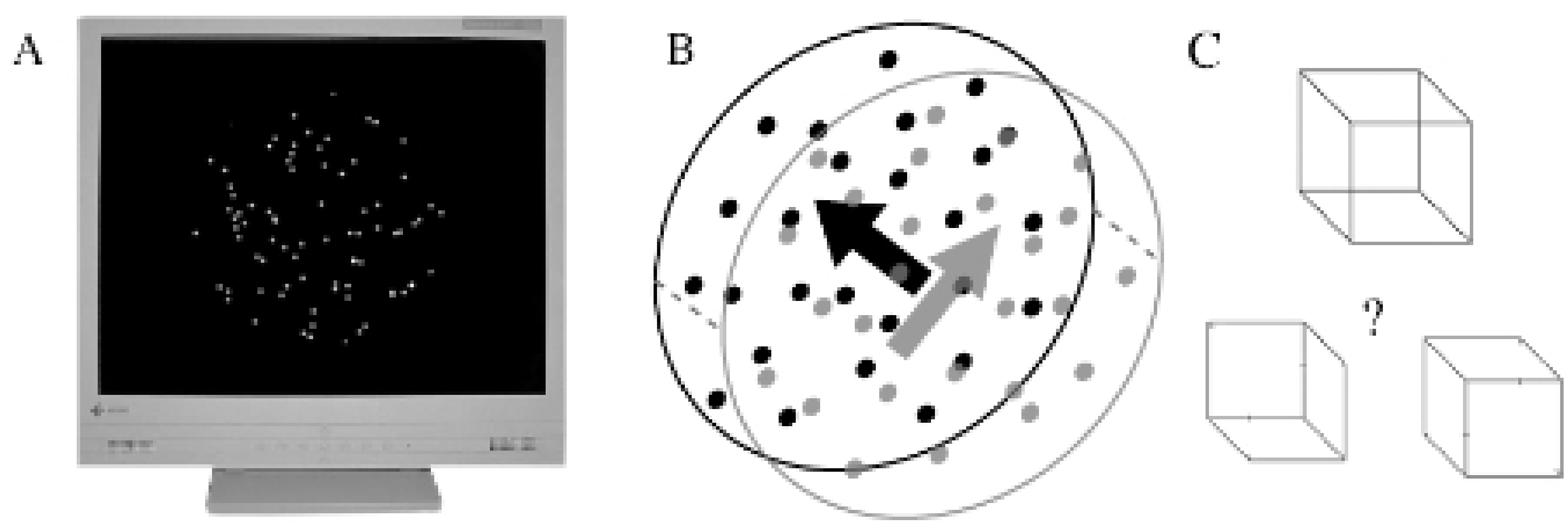
Introduction:

Consider the following synthetic motion transparency scene. We sparsely populate a region of an otherwise blank screen with high contrast Gaussian dots. [See Fig 1A.] At random we select half of the dots to drift with an arbitrary velocity – meaning that on each frame we displace each dot within this population by the same horizontal and vertical position. The dots move in this direction on each subsequent frame until they reach the end of the boundary of our region, at which time they are redrawn to appear at the opposite boundary. The other half of the dots, not yet accounted for drift in an different arbitrary velocity. [See Fig 1B.] We will call this motion presentation Transparent Random Moving Dots, or TRMD.

What is the percept human observers experience when viewing a TRMD? It is not a collection of dots, each moving independently. Rather, all the dots moving in a particular direction appear to grouped - intuitively belong to the same object. The two species of dots seem to coalesce into two distinct transparent textured sheets sliding across each other.

Interestingly, many observers report an apparent difference in the depth of these surfaces. The display is ambiguous however – there is no information in the visual scene that gives a cue as to which sheet of dots is in the foreground, and which is in the background. The inherent ambiguity of this display induces a bistabilty to this percept. In particular, though the apparent depth ordering of the dot planes may be stable for many seconds, it varies from presentation to presentation of an identical stimulus reminiscent of the famed Necker cube. [See Fig 1C.] We use this apparent crosstalk between motion processing and depth perception when viewing the TRMD as inspiration on how to segment transparent motion.

Figure 1



Now we turn our attention to some observations about motion processing in the brain in area MT. The neurons in area MT are some of the most well studied in the primate visual system, if not the entire brain. We can sketch some of their basic properties here to guide our thinking. Like many neurons in the visual system, they respond only to part of image, called the receptive field. These receptive fields tile the image in largely overlapping fashion, with neurons responding to adjacent receptive fields adjacent and most densely connected in the brain. [See Fig 2A.] Receptive fields tend not to have crisp boundaries, or perfectly symmetric shapes, but they can be well approximated as Gaussian masks through which the neuron views the world.

As a whole, we said that area MT processes visual motion - but what does that mean in terms of individual neurons? Each neuron appears to be a tuned nonlinear filter, responding most vigorously to a particular type of motion within its receptive field. In