A new computational model of how the primate brain recognizes objects creates a map “interesting” features (right) for a given image. The model’s predictions of which parts of the image will attract a viewer’s attention (green clouds, left) accord well with experimental data (yellow and red dots). Images courtesy of Sharat Chikkerur

(PhysOrg.com) -- Researchers at MIT's McGovern Institute for Brain Research have developed a new mathematical model to describe how the human brain visually identifies objects. The model accurately predicts human performance on certain visual-perception tasks, which suggests that it’s a good indication of what actually happens in the brain, and it could also help improve computer object-recognition systems.

The model was designed to reflect neurological evidence that in the primate brain, object identification — deciding what an object is — and object location — deciding where it is — are handled separately. “Although what and where are processed in two separate parts of the brain, they are integrated during perception to analyze the image,” says Sharat Chikkerur, lead author on a paper appearing this week in the journal Vision Research, which describes the work. “The model that we have tries to explain how this information is integrated.”

The mechanism of integration, the researchers argue, is attention. According to their model, when the brain is confronted by a scene containing a number of different objects, it can’t keep track of all of them at once. So instead it creates a rough map of the scene that simply identifies some regions as being more visually interesting than others. If it’s then called upon to determine whether the scene contains an object of a particular type, it begins by searching — turning its attention toward — the regions of greatest interest.

Chikkerur and Tomaso Poggio, the Eugene McDermott Professor in the Department of Brain and Cognitive Sciences and at the Computer Science and Artificial Intelligence Laboratory, together with graduate student Cheston Tan and former postdoc Thomas Serre, implemented the model in software, then tested its predictions against data from experiments with human subjects. The subjects were asked first to simply regard a street scene depicted on a computer screen, then to count the cars in the scene, and then to count the pedestrians, while an eye-tracking system recorded their eye movements. The software predicted with great accuracy which regions of the image the subjects would attend to during each task.

Don’t cross the streams

The software’s analysis of an image begins with the identification of interesting features — rudimentary shapes common to a wide variety of images. It then creates a map that depicts which features are found in which parts of the image. But thereafter, shape information and location information are processed separately, as they are in the brain.

The software creates a list of all the interesting features in the feature map, and from that, it creates another list, of all the objects that contain those features. But it doesn’t record any information about where or how frequently the
features occur.

At the same time, it creates a spatial map of the image that indicates where interesting features are to be found, but not what sorts of features they are. It does, however, interpret the “interestingness” of the features probabilistically. If a feature occurs more than once, its interestingness is spread out across all the locations at which it occurs. If another feature occurs at only one location, its interestingness is concentrated at that one location.

Mathematically, this is a natural consequence of separating information about objects’ identity and location and interpreting the results probabilistically. But it ends up predicting another aspect of human perception, a phenomenon called “pop out.” A human subject presented with an image of, say, one square and one star will attend to both objects about equally. But a human subject presented an image of one square and a dozen stars will tend to focus on the square.

Act locally

Like a human asked to perform a visual-perception task, the software can adjust its object and location models on the fly. If the software is asked to identify only the objects at a particular location in the image, it will cross off its list of possible objects any that don’t contain the features found at that location.

By the same token, if it’s asked to search the image for a particular kind of object, the interestingness of features not found in that object will go to zero, and the interestingness of features found in the object will increase proportionally. This is what allows the system to predict the eye movements of humans viewing a digital image, but it’s also the aspect of the system that could aid the design of computer object-recognition systems. A typical object-recognition system, when asked to search an image for multiple types of objects, will search through the entire image looking for features characteristic of the first object, then search through the entire image looking for features characteristic of the second object, and so on. A system like Poggio and Chikkerur’s, however, could limit successive searches to just those regions of the image that are likely to have features of interest.

John Reynolds, an associate professor in the Systems Neurobiology Laboratory at the Salk Institute for Biological Studies, finds Poggio and Chikkerur’s model intriguing because of its convergence with work that he and others have been doing on the physiology of the brain. “It holds the potential for linking underlying biology to information processing in a way that’s new and exciting,” Reynolds says. He points out, for instance, that while some neurological diseases, such as Alzheimer’s disease and schizophrenia, have physiological characteristics that can be studied independently, their relationship to the diseases’ cognitive effects is not always clear. “We’d like to be able to link those to failures of behavior and perception,” Reynolds says, “and those are naturally expressed in terms of these questions that Tommy’s raising, which is what computations are being performed.”

Reynolds speculates that, in the future, Poggio and Chikkerur’s model could be expanded so that, in the same way that it makes predictions about human eye movement, it could predict the cognitive deficits associated with disease. “It might suggest a way of reaching down into the biology and saying, ‘Look, this is the kind of mechanism that may be failing in those people,’” Reynolds says. “That could lead to treatments.”

Provided by Massachusetts Institute of Technology