SUnS 07

Scene Understanding Symposium
February 1-2, 2007
MIT
Website: suns.mit.edu

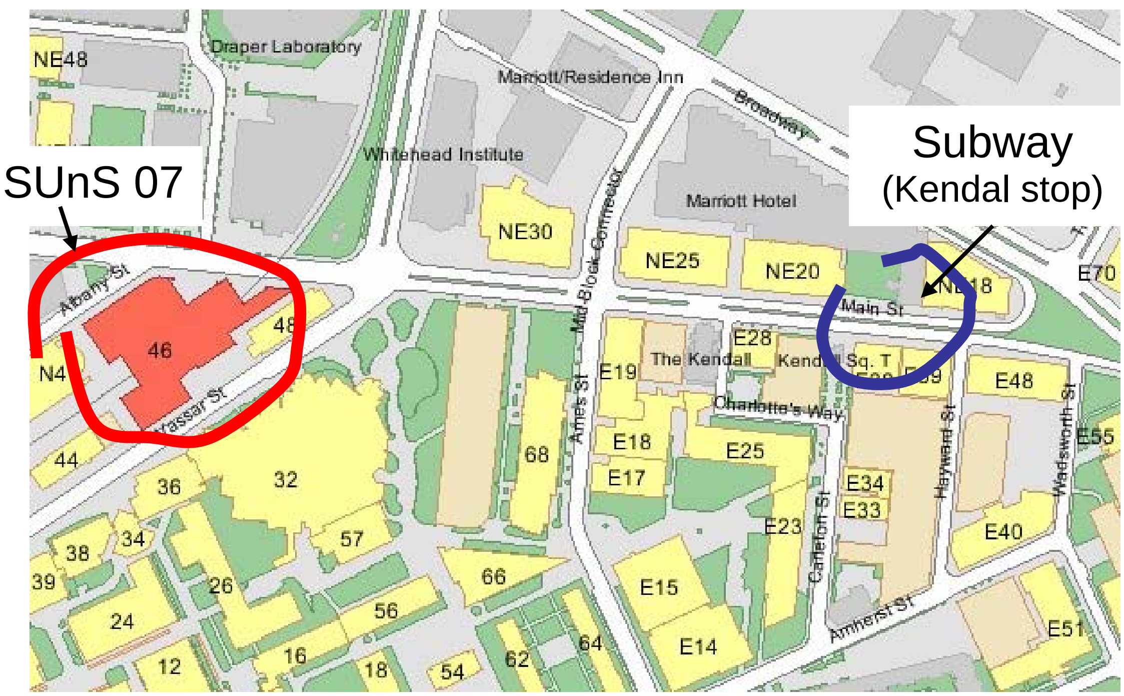
Organizers

Aude Oliva (oliva@mit.edu), Thomas Serre (serre@mit.edu), Antonio Torralba (torralba@mit.edu)

Sponsored by the McGovern Institute for Brain Research at MIT, the Department of Brain and Cognitive Sciences at MIT, the National Institutes of Health (T32 EY013935 and T32 MH020007), the National Science Foundation (IIS-Career Award) and the MIT Open Course Ware.

Schedule Thursday 1 February (1 pm to 6:20 pm) and Friday 2 February (9 am to 5 pm)

Location MIT Department of Brain and Cognitive Sciences, 43 Vassar Street, Cambridge, MA 02139 – Bldg 46-3002



	Thursday , 1 February
1:00	Opening Remarks
Session	**Scene Representation: A Behavioral Perspective**
1:05-1:30	**Perceiving, remembering & knowing in scene cognition: Where are the divisions?** **Helene Intraub** Department of Psychology, University of Delaware
1:35-1:50	**Transient attention when detecting pictures in RSVP search** **Mary C. Potter**, Rijuta Pandav, & Brad Wyble Brain and Cognitive Sciences, MIT
1:55-2:10	**Automatic and Implicit Encoding of Scene Gist** **Timothy Brady** & Aude Oliva Brain and Cognitive Sciences, MIT
2:15	**Coffee Break**
Session	**Scene Representation: A Cognitive Neuroscience Perspective**
2:30-2:55	**Neural systems for visual scene recognition** **Russell Epstein** Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania
3:00-3:25	**Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in scene perception** Soojin Park & **Marvin Chun** Department of Psychology, Yale University
3:30:-3:45	**Natural scene classification using distributed patterns of fMRI activity** **Li Fei-Fei**, D. B. Walther, E. Caddigan & D. Beck Computer Science Department, Princeton University
3:50-4:15	**Scenes, Contextual Associations and The Brain's Default Mode** **Moshe Bar**, Elissa Aminoff, Malia Mason, Mark Fenske Martinos Center, Massachusetts General Hospital & Harvard Medical School
4:20	**Coffee Break**
Session	**Capacity Limits on Scene and Object Processing**
4:40-5:00	**Robustness to Clutter and Diverted Attention in Ventral Visual Cortex** **Leila Reddy** & Nancy Kanwisher McGovern Institute for Brain Research and Brain and Cognitive Sciences, MIT

5:05-5:30	**Categorization of objects and scene context at various levels: can we always rely on fast visual processing?** **Michèle Fabre-Thorpe** CerCo, University Toulouse 3, CNRS, France
5:35-5:55	**Eye movements and high-level saliency effects in natural scenes** **Simon J. Thorpe** CerCo, University Toulouse 3, CNRS, France
6:00-6:20	**Rapid object categorization without conscious recognition: a neuropsychological study** **Muriel Boucart** University Lille 2, CNRS, Lille Hospital, France
6:25	**End**
	Friday, 2 February
8:30	**Breakfast**
Session	**Computational Approaches to Scene and Image Understanding**
9:00-9:20	**Object and scene recognition on tiny images** **Antonio Torralba**, Rob Fergus & William T. Freeman Computer Science and Artificial Intelligence Laboratory, MIT
9:25-9:45	**It's a 3D World: Toward a Qualitative 3D Representation of a Scene** **Alyosha Efros**, Derek Hoiem & Martial Hebert Computer Science Department, Carnegie Mellon University
9:50-10:10	**Observations from Parsing Images of Architectural Scenes** **Alexander C. Berg**, Floraine Grabler, Maneesh Agrawala & Jitendra Malik Computer Science Division, Berkeley
10:15-10:40	**Hierarchical statistical models for local and global structure in natural scenes** **Michael Lewicki** Computer Science & Center for the Neural Basis of Cognition, Carnegie Mellon University
10:45	**Coffee Break**
Session	**Scene Understanding: Role of Attentional Mechanisms and Cortical Feedback I**
11:10-11:30	**Neural Synchrony and Visual Search** **Robert Desimone** McGovern Institute for Brain Research, MIT

11:35-12:05	**Neurophysiological and behavioral effects of familiarity on visual search** **David Sheinberg** & Ryan Mruczek Department of Neuroscience, Brown University
12:15	**Lunch Break**
1:15	**Poster Session**
Session	**Scene Understanding: Role of Attentional Mechanisms and Cortical Feedback II**
2:15-2:35	**Modeling visual search in real scenes: what's the setsize?** **Ruth Rosenholtz** Brain and Cognitive Sciences, MIT
2:40-3:00	**Modeling attention to proto-objects in natural scenes** **Dirk B Walther** Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champagn
3:05-3:25	**Guidance of visual search by unlocalized scene properties** **Jeremy Wolfe** Brigham and Women's Hospital & Harvard Medical School
3:30	**Coffee Break**
3:55-4:25	**Feature-based attention dynamically changes shape representation in area V4** **Jack Gallant** Department of Psychology & Helen Wills Neuroscience Institute, Berkeley
4:30-4:50	**Immediate perception and feedforward models: what is next?** **Tomaso Poggio** Center for Biological and Computational Learning, Computer Science and Artificial Intelligence Laboratory and McGovern Institute for Brain Research, MIT
4:55-5:00	*Concluding Remarks*
5:00 - 6:00	**Reception**

Thursday, February 1

Scene Representation: A Behavioral Perspective

1:05-1:30

Perceiving, remembering & knowing in scene cognition: Where are the divisions?

Helene Intraub
Department of Psychology, University of Delaware

When testing the early time course of scene cognition questions often arise as to the locus of an effect (e.g., "Did this occur during perception or memory?"). These questions become paramount if we think of scene cognition solely in terms of a visual representation of the stimulus (usually a picture). In contrast, if we think of it in terms of a surrounding 3D spatial framework in which an input (e.g., visual, haptic, imagined, etc.) is set, then the questions change. Taking boundary extension (BE) as an example, I will report new research in which a fleeting interruption (42 ms; commensurate with a saccade) during presentation of a view was sufficient to elicit BE errors. For example, the pre-interrupt view was remembered as being slightly more panoramic than the currently visible view -- although they were identical. I will argue that a 3D spatial model of scene perception provides a more parsimonious explanation of this and related observations than does a purely visual one and that it may provide a means of eliminating the perception/memory dilemma.

1:35-1:50

Transient attention when detecting pictures in RSVP search

Mary C. Potter, Rijuta Pandav, & Brad Wyble
Brain and Cognitive Sciences, MIT

When searching for two target in a rapid serial visual presentation (RSVP) viewers often miss the second target (T2) if it appears soon after T1, an effect termed an attentional blink. When T2 appears immediately after T1, however, it is relatively easy to report (lag 1 sparing). Wyble, Bowman, & Potter (2007) propose that detection of a target in an RSVP sequence generates a brief burst of transient attention lasting about 150 ms that benefits both the target and any immediately following stimulus appearing at the same location. Most previous experiments used simple stimuli such as letters, digits, or words; we asked whether lag 1 sparing would be found for more complex targets such as pictures of objects in a specified category (e.g., fruit, birds). Supported by MH047432

1:55-2:10

Automatic and Implicit Encoding of Scene Gist

Timothy Brady & Aude Oliva
Department of Brain and Cognitive Sciences, MIT

One of the primary goals of the visual system is to extract statistical regularities from the environment to build a robust representation of the world. Recent research on visual statistical learning (VSL) has demonstrated that human observers can implicitly extract joint probabilities between objects during streams of visual stimuli (Fiser & Aslin, 2002). In the real world, temporal predictability between scenes and places exists at both exemplar and categorical levels: whatever office you are in, the probability that you will step out in a zoo is much lower than the probability that you will enter a corridor. In a series of experiments, we tested to what extent people are sensitive to the learning of categorical temporal regularities based on the *gist* or semantic understanding of natural scenes. Our results suggest that the gist of a scene is automatically and implicitly extracted even when it is not task-relevant, and that implicit statistical learning can occur at a level as abstract as the conceptual gist representation.

Scene Representation: A Cognitive Neuroscience Perspective

2:30-2:55

Neural systems for visual scene recognition

Russell Epstein

Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania

FMRI studies have identified a set of cortical regions that respond more strongly to visual scenes such as landscapes, cityscapes, and rooms than to other complex stimuli such as faces or decontextualized objects. This network includes the parahippocampal place area (PPA), retrosplenial cortex (RSC), and the transverse occipital sulcus (TOS). These findings are particularly intriguing when considered in the light of behavioral results that indicate that human observers can interpret complex visual scenes very rapidly. Thus, one possibility is that these regions form a network for scene recognition, which might be complementary to other cortical networks dedicated to face and/or object recognition. But what kind of scene recognition? Recent data from our laboratory suggest that these regions might be particularly involved in location identification -- the use of visual information to identify one's location and orientation within the larger spatial environment -- rather than other scene recognition tasks such as scene categorization. Furthermore, our data argue for a certain amount of specialization within the PPA-RSC-TOS cortical network, with PPA and TOS more concerned with the perceptual representation of the currently-visible scene and RSC more concerned with linking the current scene to stored representations of the broader environment that extends beyond the current horizon. Thus, the preferential response to scenes in these regions seems to reflect the engagement of mechanisms that support the transformation of visual input into spatial codes useful for navigation.

3:00-3:25

Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in scene perception

Soojin Park & **Marvin Chun**

Department of Psychology, Yale University

One of the biggest challenges for human visual system is to create a seamless continuous world from multiple snapshots. Mediating this ability, there are at least two scene-sensitive regions of the brain, the parahippocampal place area (PPA) and retrosplenial cortex (RSC). Whereas the PPA is sensitive to changes in viewpoint of the same scene (Epstein, Graham & Downing, 2003), RSC may encode a more abstract scene-based representation independent of specific viewpoint (Bar & Aminoff, 2003; Epstein & Higgins, 2006; Park et al., 2006). To directly test this hypothesis, we presented different snapshot views from panoramic scenes. These views overlapped with each other by only 30%, so that they represented clearly different views, but appeared to come from the same panoramic scene. Using fMRI adaptation, we tested whether the PPA and RSC treated these panoramic views as the same or different. In the panoramic condition, three different views from a single panoramic scene were presented. In the identical condition, three identical views were presented. If the activity within the scene-selective ROI reflects an integrated scene representation, then we should observe attenuation in the panoramic condition. On the other hand, if the scene activity is specific to views, then we should observe no attenuation for panoramic repetition. We did not find any attenuation for panoramic repeats in the PPA, replicating a previous study (N=17). In contrast, the RSC showed significant attenuation for the panoramic condition. Both the PPA and RSC showed significant attenuation for the identical repetition. These results demonstrate that the PPA and RSC play different roles in scene perception: the PPA focuses on selective discrimination of different views while RSC focuses on the integration of scenes under the same visual context. Such integration may facilitate perception of a continuous world from multiple snapshots arising from eye and head movements.

3:30-3:45

Natural scene classification using distributed patterns of fMRI activity

L. Fei-Fei *, D. B. Walther, E. Caddigan and D. Beck
**Computer Science Department, Princeton University*

Human observers are able to quickly and efficiently perceive the content of natural scenes (Potter, 1976). Previous studies have examined the time course of this rapid classification (Thorpe et al, 1996) as well as the brain regions activated when subjects categorize natural scenes (Epstein & Higgins, 2006). In this talk, we present some preliminary results on the classification of distributed patterns of fMRI activity associated with particular natural scene categories (beaches, mountains, forests, tall buildings, highways, and industrial scenes). fMRI data was acquired while subjects viewed 100 images from each of six categories. We tested several pattern recognition algorithms (e.g. Support Vector Machines, Gaussian Naive Bayes, etc.) in a leave-one-run-out procedure on selected voxels and found that all algorithms predict the natural scene category seen by the subject well above chance. Furthermore, prediction accuracy was still well above chance when retinotopic cortex was excluded from the analysis, suggesting that this multi-voxel analysis does not rely solely on differences in simple visual features or differences in the retinotopic representation of the stimuli. If time permits, we will discuss some ongoing experiments related to this question, including the testing of these pattern recognition algorithms under different visual stimuli transformations (e.g. scale changes or translations).

3:50-4:15

Scenes, Contextual Associations and The Brain's Default Mode

Moshe Bar, Elissa Aminoff, Malia Mason, Mark Fenske
Martinos Center, Massachusetts General Hospital & Harvard Medical School

A visual scene is made of individual objects and the spatial relations between them. Scene understanding is facilitated, if not afforded, by the associative processes that glue these constituents together. These associations are based on the regularities that are continuously extracted from the environment; by translating them to predictions we can rely on our memories to promote perception, cognition and action. These associations are the focus of our talk. We have previously shown in a series of studies that a specific cortical network, parts of which have been implicated in the past in diverse processes such as spatial processing and episodic memory, could be best explained as mediating associative processing. By broadening the functional description of this network, therefore, we provided a platform for reconciling findings from multiple, seemingly independent, domains. Here we take this evolving framework a step further. Associative processing has been long proposed to provide the vehicle of thought. We describe here observations from cognitive neuroscience that elucidate the neural processing that mediates this element. This account further allows a more specific attribution of a cognitive function to the brain’s default activity and mind-wandering. Put simply, we show that when people see visual scenes, or pictures of individual objects that elicit a relatively large number of associations (e.g., a traffic light), they activate cortical regions that are typically recruited when people are engaged in an unconstrained thought. This striking overlap between the contextual association network and the default network is presented as a manifestation of the degree to which associations provide the principle basis of thought.

Session: Capacity Limits on Scene and Object Processing

4:40-5:00

Robustness to Clutter and Diverted Attention in Ventral Visual Cortex

Leila Reddy & Nancy Kanwisher
McGovern Institute for Brain Research, Brain and Cognitive Sciences, MIT

The pattern of fMRI responses across the ventral visual pathway to objects presented in isolation carries information about the category of the object. However, natural images usually contain multiple objects ("clutter"), a notorious challenge for distributed representations. Here we used pattern analysis methods to ask whether category information in the fMRI response is preserved under conditions of clutter and diverted attention, and whether the answer to these questions depends on the category of object viewed. We found that information in the spatial pattern of fMRI response about standard object categories (shoes and cars) is severely disrupted by clutter, and eliminated when attention is diverted. However, information about preferred categories in category-specific cortical regions (faces in the FFA and houses in the PPA) is undiminished by clutter, and partly preserved under diverted attention. These findings suggest that under natural viewing conditions, the spatial pattern of fMRI response provides robust information about object category only for a "special" categories that are coded in their own selective cortical regions.

5:05-5:30

Categorization of objects and scene context at various levels: can we always rely on fast visual processing?

Michèle Fabre-Thorpe
CerCo, Universite Toulouse 3, CNRS, France

About 10 years ago, we have shown that human subjects could mainly rely on a fast feed-forward flow of information processing to categorize photographs of natural scenes on the basis of whether they contained an animal or not. Early responses can be seen from 260 ms after stimulus onset. Such fast processing is not limited to the categorization of biological targets, it can also be used for artefactual targets such as means of transports for example, or when categorizing the gist of a scene (natural or artificial contexts). This first wave of processing appears to have an optimal speed. So far, it has proved impossible to speed up the processing flow, even when such an effect could be expected: when categorizing well known photographs, simple targets (circles vs. squares) or targets reported to be processed faster than other objects such as human faces for example. Moreover, human subjects appear to switch target category from testing block to another without any cost on processing speed. On the other hand, such fast processing can be disturbed. We report the effect of object saliency on contextual categorization and the surprising additional processing time needed to perform categorizations at the basic level (dogs, birds) when it has largely been reported that this level of categorization was the fastest accessible level.

5:35-5:55

Eye movements and high-level saliency effects in natural scenes

Simon J. Thorpe
CerCo, Universite Toulouse 3, CNRS, France

Last year, Kirchner and Thorpe reported that when two images are flashed left and right of fixation, subjects can make reliable saccades to the side where the image contains an animal from just 120-130 ms following stimulus onset. The result was a surprise, because the value includes the time for motor execution, implying that the underlying visual processing can be done in less than 100 ms. In this presentation, I will discuss evidence that this sort of very fast processing differs from the sort of manual response task that we have traditionally used. In particular, it appears that the tendency of subjects to saccade towards particularly significant objects such as animals and faces is not something that can be

easily controlled by top-down task-dependent factors. Instead, I will argue that it could reflect a form of high-level saliency effect. According to this view, certain key stimulus classes such as faces can produce strong, selective, and short latency neural activation. When this activity is unbalanced between the two hemispheres, this activity could be directly used to trigger saccades in the appropriate direction.

6:00-6:20

Rapid object categorization without conscious recognition: a neuropsychological study

Muriel Boucart

University Lille 2, CNRS, Lille Hospital, France

Kirchner and Thorpe (2006, *Vision Research*) showed that when two scenes are simultaneously flashed left and right of fixation human observers can reliably make saccades to the side containing the target (e.g., an animal) in as little as 120 msec. It is suggested that rapid object categorization can be performed without conscious recognition using highly automatic and data driven routines. This question was addressed in a neuropsychological study with neurologically impaired patients. Two patients with visual agnosia and control participants were presented with photographs of natural scenes or isolated objects. Two scenes, one containing the target (an animal) were simultaneously displayed for 3 sec 7° left and right of fixation. In separate sessions participants were instructed either to locate the target animal with a key press (left/right) or to make a saccade toward the target. Patients, who identified very few objects with unlimited exposure time were able to detect the target with 70% accuracy and saccade latencies of 250-300 ms in the categorization task. Performance was better with saccades than with manual response. For all observers accuracy was equivalent for objects in scenes and isolated objects but RTs were shorter for scenes. Altogether, the results support the hypothesis that rapid categorization can be accomplished without conscious recognition.

Friday, February 2

Computational Approaches to Scene and Image Understanding

9:00-9:20

Object and scene recognition on tiny images

Antonio Torralba, Rob Fergus, William T. Freeman

Computer Science and Artificial Intelligence Laboratory, MIT

The human visual system is remarkably tolerant to degradations in image resolution: in a scene recognition task, the performance of subjects is similar whether 32x32 color images or multi-mega pixel images are used. Accordingly, we consider the task of scene recognition in computer vision using very small (32x32 pixel) images. Small images force us to employ contextual reasoning and to explore new representations due to the lack of textural cues on which many existing algorithms are based. However, the small size of each image carries two important benefits: (i) it permits standard machine learning tools to be easily applied and (ii) huge image databases may be easily collected. Inspired by the data-driven nature of Google, we deploy simple data mining methods to a large collection of images (50 million).

9:25-9:45

It's a 3D World: Toward a Qualitative 3D Representation of a Scene

Alyosha Efros, Derek Hoiem and Martial Hebert

School of Computer Science. Carnegie Mellon University

Most of today's computational approaches to visual object recognition essentially reduce the problem to one of pattern classification, where rectangular image patches are independently compared to stored templates to produce isolated object labels. But the real world is, in fact, very much three-dimensional, with occlusion, perspective effects, complex 3D shapes, and interactions between objects in the 3D space of the scene. To make progress, we must somehow account for the three-dimensional nature of the real world. However, capturing quantitative 3D information may not be desirable, or even feasible, e.g., when the visual input consists of just a single image. Instead, we advocate *qualitative geometric reasoning* in terms of 3D spatial relationships between scene components, category-level object models, and global scene understanding, without explicit, quantitative 3D reconstruction. In this talk, I will briefly describe out first steps in this direction, and discuss some of the challenges that we are facing.

9:50-10:10

Observations from Parsing Images of Architectural Scenes

Alexander C. Berg, Floraine Grabler, Maneesh Agrawala, Jitendra Malik
Computer Science Division, Berkeley

Computational models for visual recognition show promise for some tasks. I will review our success in this area and show some information theoretic comparisons with our ongoing work on parsing scenes. For images of architectural scenes we have observed that very simple independent local features provide a great deal of information about what components -- building, sky, ground, etc. -- make up a scene. In addition a few carefully chosen image wide latent variables are added to the model then even more information is available. Finally given this coarse level parsing it is possible to effectively identify features such as windows and roof-lines that would be difficult to parse in isolation.

10:15-10:40

Hierarchical statistical models for local and global structure in natural scenes

Michael Lewicki
Computer Science & Center for the Neural Basis of Cognition, Carnegie Mellon University

Scene Understanding: Role of Attentional Mechanisms and Cortical Feedback I

11:10-11:30

Neural Synchrony and Visual Search

Robert Desimone
McGovern Institute for Brain Research, MIT

11:35-12:05

Neurophysiological and behavioral effects of familiarity on visual search

David Sheinberg & Ryan Mruczek
Department of Neuroscience, Brown University

We have previously shown that search efficiency depends on a subject's familiarity with the search context. In this talk, I'll describe recent findings indicating that the receptive field properties of cells in the temporal lobe selective for target images are dependent on the familiarity of the context within which these targets are shown. Specifically, the functional field of view of individual cells seems to expand in a way that mirrors behavior.

Scene Understanding: Role of Attentional Mechanisms and Cortical Feedback II

2:15-2:35
Modeling visual search in real scenes: what's the setsize?
Ruth Rosenholtz
Brain and Cognitive Sciences, MIT

Visual clutter concerns designers of user interfaces and information visualizations. This should not surprise visual perception researchers, since excess and/or disorganized display items can cause crowding, masking, decreased recognition performance due to occlusion, greater difficulty at both segmenting a scene and performing visual search, and so on. Given a reliable measure of the visual clutter in a display, designers could optimize display clutter. Furthermore, a measure of visual clutter could help generalize models like Guided Search (Wolfe, 1994) to real scenes by providing a substitute for "set-size" more easily computable on more complex and natural imagery. We will discuss measures of visual clutter which operate on arbitrary images as input, and explore the use of these measures as a stand-in for set-size in visual search models. We demonstrate that such measures of visual clutter correlate well with search performance and perceived clutter in complex imagery. This includes the search-in-clutter displays of Wolfe, Oliva, Horowitz, Butcher, & Bompas (2002), Bravo & Farid (2004), and other experiments.

2:40-3:00
Modeling attention to proto-objects in natural scenes
Dirk B. Walther
Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champagn

Selective visual attention is believed to be responsible for serializing visual information for recognizing one object at a time in a complex scene. But how can we attend to objects before they are recognized? In coherence theory of visual cognition, so-called proto-objects form volatile units of visual information that can be accessed by selective attention and subsequently validated as actual objects. We propose a biologically plausible model of forming and attending to proto-objects in natural scenes. We demonstrate that the suggested model can enable a model of object recognition in cortex to expand from recognizing individual objects in isolation to sequentially recognizing all objects in a more complex scene. All code is available in a Matlab toolbox at: http://www.saliencytoolbox.net

3:05-3:25
Guidance of visual search by unlocalized scene properties
Jeremy Wolfe
Brigham and Women's Hospital & Harvard Medical School

When you walk into the lecture hall for this talk, you will search for a place to sit. Your search will not be random but will be guided by knowledge about seats, lecture halls in general and, perhaps, this lecture hall in particular. Form of guidance differs from other forms of attentional guidance in visual search. You would not be guided by target features (as in a search for red among green) nor would you be guided by specific location information (as in spatial cueing experiments). Rather you would be guided by what could be called "unlocalized scene properties" (e.g. seats are on horizontal surfaces, not walls). How does this guidance compare to other forms of guidance? In a new series of experiments, our observers viewed an unchanging loose pyramid of cubes. Each cube had three visible surfaces: the top and two sides. Cube tops did not form a coplanar surface. On each trial, red, yellow, and blue Ts and Ls were distributed randomly over these surfaces and observers searched for the letter T. There were five conditions. In the No Guidance condition, observers simply searched for a T. In the Color Guidance conditions, observers were told the color of the T in advance. This information could be

Blocked (e.g. Ts might be blue for the whole block) or cued randomly on each trial. In the Scene Guidance conditions, observers were told that the T, if present, would be on one of the three surfaces. Again this information could be Blocked (Ts always on top) or Random from trial to trial. The slope of the RT x Set Size function was 39 msec/item in the No Guidance condition. Blocked Color Guidance produced a slope of 12 msec/item, as would be expected if observers restricted search to items of the correct color. Random color guidance was less efficient (21 msec/item) as if the guidance was not as effective. At small set sizes, Scene Guidance, blocked and random, provided no apparent benefit. At larger set sizes, both forms of Scene Guidance appear to speed response compared to No Guidance conditions. Unlike guidance by attributes like color, Scene Guidance may take quite a long time to become effective. It would still aid an extended search for a seat but not a search for a T among just a few Ls.

3:55-4:25
Feature-based attention dynamically changes shape representation in area V4
Jack Gallant
Dept. of Psychology & Helen Wills Neuroscience Institute, Berkeley

Attention enables the selective filtering of sensory information to meet behavioral demands, but the mechanisms underlying this process remain unknown. Most physiological evidence supports the idea that attention highlights attended locations or features by enhancing firing rates of relevant neurons without altering their selectivity, but some theoretical work suggests that attention can alter the filtering properties of these cells. In two experiments we have demonstrated that feature-based attention alters the multidimensional tuning curves of many V4 neurons, thereby dynamically changing the way that shape is represented in these cells. This attentional modulation enhances responses to attended features, consistent with a matched filter model of visual attention. Our results demonstrate that V4 neurons do not act as “labeled lines,” and so have important implications for the way the visual network must be designed in order to permit correct decoding and interpretation of the information in V4 by higher visual areas. Taken together, our results suggest that memory and decision-making processes required for visual search are integrated into the basic processes of visual representation and distributed widely across the neocortex.

4:30-4:50
Immediate perception and feedforward models: what is next?
Tomaso Poggio
Center for Biological and Computational Learning, Computer Science and Artificial Intelligence Laboratory and McGovern Institute for Brain Research, MIT

I will briefly review hierarchical feedforward models of the ventral stream and their apparent success in explaining several physiological data and psychophysical results in scene categorization. I will then focus on the limitations of such models for object recognition, suggesting specific questions about the computational role of attention and about recognition tasks beyond scene classification.