Research Statement

Bosco S. Tjan

I study the human visual system; in particular, the neural computations that underlie the perception of form, a domain that includes object recognition, scene perception, and reading. Although my research addresses a broad spectrum of topics from basic questions in object recognition to clinical applications and I employ a number of different methodologies (psychophysical experimentation, fMRI, and mathematical modeling), the general theoretical framework motivating much of this work is that of optimal computation.

Some history: my graduate advisors at the University of Minnesota, Gordon Legge, and Daniel Kersten in Psychology, and William Thompson in Computer Science, afforded me an intense opportunity to develop skills in visual psychophysics and computational neuroscience as well as their applications to problems as diverse as low-level vision and reading. I developed additional skills in cognitive and computational neuroscience during a two-year post-doctoral fellowship with Heinrich Bülthoff at the Max-Planck Institute for Biological Cybernetics in Tübingen, Germany, and a two-year research associate position with John Oliensis and David Jacobs in the computer-vision group at the NEC Research Institute in Princeton, NJ. In 2001, I joined USC as an assistant professor in Psychology and established the Laboratory for Functional and Computational Vision. The basic research program in my lab is funded in part by the National Institutes of Health. A translational component is funded by the National Institute on Disability and Rehabilitation Research.

Despite many years of research, the understanding of form vision remains a challenging problem. My approach to the topic begins with two sorts of theoretical questions: (1) what is the optimal computational strategy for performing a given task with a given set of stimuli, and (2) what are the limitations that an otherwise optimal observer (termed an ideal observer and formulated in a Bayesian framework) must be subjected to in order to account for the empirical data when humans perform the task? I have developed a computational framework for (1) measuring the information content of a complex input, (2) describing how computations of such inputs are to be characterized, and (3) structuring experimentation to uncover the underlying computations in human observers. Most importantly, I have extended the ideal-observer analysis to tasks where feature complexity and invariance had previously precluded this approach. (Invariance refers to a processor's ability to give the same response, such as the identity of an object, when presented with physically distinct inputs, such as the object viewed at a different position, size, orientation, and illumination direction.) I have also developed a number of novel empirical techniques in psychophysics and fMRI to test the predictions of the computational models. I will review some of my major findings and future directions.

Object Recognition and Representation

The brain excels in general purpose object recognition, a task that is both complex and ill defined. Visual processing in general, including object recognition, is thought to be a hierarchical process of feature extraction and composition, resulting in features that are progressively more complex, invariant, and task-relevant. In a series of studies, beginning with my Ph.D. work, I have investigated this feature extraction process, from early low-level features, such as edges, to object representations are that adaptive to task demands.

Low-level features. An edge, which marks the location of a contrast discontinuity, is often considered an important primitive for object representation and the starting point of visual processing for object recognition. By comparing performance between human and a newly formulated ideal observer across image rendering modes (shaded objects, line drawings, silhouettes) and different tasks (detection vs. identification), I concluded that the visual system utilizes a very limited number of features for object recognition, which are at or near regions of contrast discontinuity. This limitation is in part a result of the visual system trying to achieve invariance to an object's orientation and position, even under conditions when neither the invariance nor the feature placement were optimal for the task and stimuli (Tjan, Braje, Legge, & Kersten, 1995; Braje, Tjan, & Legge, 1995). These results showed that the visual system is rather inflexible in choosing low-level features even in tasks for which such a choice would be advantageous.

Intermediate-level features. There are infinitely many ways to combine a set of low-level features to form intermediate- or higher-level features. How do we identify the intermediate-level features used by the visual system? I postulated that for an intermediate feature to be useful, it must be both common among objects and diagnostic about the identity of objects. If this is the case, then it follows from a Bayesian ideal-observer analysis that a behavioral signature for a feature X to be an intermediate feature used by the visual system is that it is hard to discriminate X from a small deviation of X, even though the detection efficiency for X may be high. With Zili Liu, a long-term collaborator, we found that this is the case for bilateral symmetry after controlling for stimulus information (Tjan & Liu, 2005) and for the natural constraints on human-body parts, such as the length of the forearms (Lu, Tjan & Liu, 2006). I believe this dissociation between discrimination and detection efficiencies can be a useful tool for uncovering the intermediate-level features used by the visual system for a given task.

Representational demand and adaptive representation. Invariance is a hallmark of general-purpose object recognition. How does the visual system represent objects to achieve invariance? I have performed a computational analysis on this issue using an ideal-observer framework. I found that the extent to which an object must be represented, as quantified by a complexity index, varies hugely across different ensembles of objects. Object ensembles with a high complexity index are also those that humans find difficult to learn and to generalize from one view to another (Tjan & Legge, 1998). I also found the complexity index for an object strongly depends on what the other objects are in the ensemble. These two findings jointly led me to argue that if the visual system computes optimally, it must have more than one representation for each object; furthermore, it must have a way to dynamically adapt or select the representation according to the task and the object context (Tjan, 2002). I propose that a rather simple mechanism could lead to flexible and adaptive representation, taking advantage of the fact that along a visual processing pathway, features are progressively more complex, invariant, and abstract, which in essence provides a full range of representations at different levels of abstractions (Tjan, 2001). I am planning experiments to test this idea empirically.

Form vision in the periphery

Understanding the qualitative and quantitative differences between central and peripheral vision in object recognition is critical for clinical vision research because a significant cause of visual impairment is the loss of central vision. I have devoted a great deal of efforts recently to studying form vision in the visual periphery.

Letter identification. English letters, because of their simplicity and real-world importance (reading for patients with central vision loss is very difficult), have been used as the main stimulus material in this portion of my research. In collaboration with Susana Chung and Gordon Legge, I developed an ideal-observer model for letter identification and demonstrated that both central and peripheral visual systems use features of sizes appropriate to their respective spatial resolution and the size of the stimuli (Chung, Legge & Tjan, 2002). In other words, if one controls for the differences in spatial resolution (as measured by the contrast sensitivity functions), the spatial tuning for letter identification is optimal for both central and peripheral vision! The same conclusion was obtained in the amblyopic fovea, which is believed to be functionally analogous to the visual periphery (Chung, Tjan & Levi, 2002). More recently, my student Anirvan Nandy and I used a classification-image method to visualize the perceptual templates (i.e. a sum of all features) used by human observers in a letter identification task and found no qualitative difference between central and peripheral vision in terms of the first-order classification images we obtained (Tjan & Nandy, 2006). All these results suggest that the features used for letter identification are qualitatively similar in central and peripheral vision. However, peripheral form vision remains severely impaired. The cause could be in how these features are integrated, which led me to investigate the phenomenon of “crowding.”

Crowding refers to the phenomenon that when flanked by other items, a target becomes unrecognizable, although it can be recognized when presented alone. It is most pronounced in peripheral vision, yet is almost non-existent in central vision - it encapsulates many unique nonlinear aspects of the form-vision deficits in peripheral vision. The mechanistic causes of crowding are still unknown. Through collaborative research with Susana Chung and my students, I have gathered a broad range of empirical data on crowding over the past few years. We found that: 1) spatial tuning for the crowded target is nearly optimal (Chung & Tjan, *in revision*); 2) crowding appears to be due to an increase in the internal noise of the visual system, induced by the flankers (Tjan, He, Chung & Schwartz, 2004 [abstract]; Nandy & Tjan, 2006 [abstract]); 3) to induce this internal noise, the flankers need to have a spatial frequency distribution similar to that of the target (Tjan & Dang, 2005 [abstract]); 4) the position uncertainty intrinsic to the periphery plays a catalytic role in causing crowding (Tjan, 2006, R01 proposal); and 5) the neural origin(s) of crowding appears to be “bottom-up” from the early visual areas V1-V3 (Arman, Chung & Tjan, 2006 [abstract]).

The coherent mechanistic picture emerging from these studies is that feature integration in the periphery is less discriminatory, probably due to a failure to prune forward-projecting excitatory connections during visual development. This “failure” may actually be optimal for the kind of tasks for which peripheral vision is normally specialized, such as motion detection. To capture the recent momentum in the field on this topic, I am co-editing (with Denis Pelli, Robert Desimone, Patrick Cavanagh and Anne Treisman) a special issue on crowding for the Journal of Vision. I am also co-organizing (with Pelli & Cavanagh) a symposium on crowding at the upcoming European Conference for Visual Perception.

Signal-in-noise methods for psychophysics and fMRI

There has been a long tradition in engineering to use noise to perturb an unknown system in order to decipher the system’s internal mechanisms. Such external noise methods for system identification have found a wide range of applications in visual psychophysics, particularly in conjunction with ideal-observer analysis. I have made significant advances in using noise to uncover the perceptual templates and high-order features used by the visual system and to reveal the ordering of information processing in the human cortex.

Higher order classification images. A classification image provides a method for visualizing the image patterns that lead to a particular response. The current classification-image technique is limited as a linear method, which can only reveal the image pixels that an observer used to perform a task averaged over trials. It cannot reveal the perceptual templates used by the visual system if the system is highly invariant or uncertain - that is, when many different input patterns are mapped to the same response (e.g. the same object seen from different views or positions). I have developed a novel method to overcome this limitation (Tjan & Nandy, 2006) and I have successfully applied the method to obtain classification images in the visual periphery, where the visual system has a high uncertainty about the position of a stimulus. I have also derived a second-order classification image method to infer the features (not just pixels) use by the visual system to perform a recognition task (Nandy & Tjan, 2006 [abstract]). I believe these techniques can be adapted to electrophysiology to reveal the visual features extracted by neurons in higher cortical areas.

Revealing information flow with BOLD fMRI. Large regions of the cortex outside of the primary visual areas are thought to be important for form vision; yet little is known about how visual information is processed in the higher cortical areas. I have combined ideal-observer analysis with fMRI to address this issue. My first goal was to derive an fMRI method that could reveal the forward direction of information flow between these cortical regions. As visual processing progresses along the forward direction of a visual pathway, visual features become progressively more invariant - or put another way, more physically different images will lead to the same neural response. For an ideal observer, an increase in invariance will lead to an increase in the log-log steepness of its psychometric function (% target detection vs. image signal-to-noise ratio). Based on this analysis, I proposed that the steepness of the BOLD response function (% BOLD vs. image SNR) can be used to index the sequential ordering of a cortical region: the steeper the BOLD response function, the higher the cortical region is in the visual processing pathway. We tested and confirmed this idea using visual areas with well-known functional organizations (the ventral pathway from V1 to the lateral occipital complex (LOC), and the dorsal pathway from V1 to V3a) (Tjan, Lestou & Kourtzi, 2006). This encouraging result suggests that other theoretic properties of an ideal-observer cascade can be used in conjunction with fMRI to reveal the computational architecture in the visual cortex. This topic is currently an intense subject of research in my lab.

Summary

My research concerns the neural computations that lead to form vision in humans. I have found it particularly instructive and fruitful to approach the research questions from the perspective of optimal computations. My research on object recognition addresses basic questions about form vision, while my research on peripheral vision seeks to translate the answers to clinical needs. These two lines of research will continue to interact in the foreseeable future, and I am motivated by the belief that both will benefit from the novel empirical and computational techniques that I have been developing.