Research Statement

Bosco S. Tjan

November, 2007

I study the human visual system by exploring neural computations that underlie the perception of form, a domain that includes object recognition, scene perception, and reading. My research addresses a broad spectrum of topics from basic questions in object recognition to clinical applications, all of which I approach from the theoretical framework of *optimal computation* – by considering the computation and behavior of a mathematically optimal observer (an ideal observer) for the given stimuli, task, and known limitations of the human visual system. The tools that I use include psychophyiscal experimentation, fMRI, and mathematical modeling.

How the brain perceives forms and recognizes objects is a challenging problem. My approach begins with answering two questions: (1) what is the optimal computational strategy (ideal observer) for performing a given visual task, and (2) what are the limitations that must be imposed on the optimal computation in order to account for the human data? I have developed a computational-empirical framework for (a) measuring the information content of complex stimuli in the context of a task, (b) describing the corresponding optimal computation and the probable deviations from optimality, and (c) designing effective experiments to uncover the underlying computations in human observers. I have extended the ideal-observer analysis to complex tasks and stimuli that had previously precluded this approach. I have also developed a number of novel empirical techniques in psychophysics and fMRI. The basic research program in my lab and its translational components are funded by the National Institutes of Health and the National Institute on Disability and Rehabilitation Research. 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, and object recognition in particular, 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 along the visual-processing hierarchy.

Low-level features. Contrast discontinuity, or edge, is often considered an important primitive for object representation and the starting point of visual processing for form vision. By comparing object-recognition performance between human and ideal observers across different types of stimulus presentation (shaded objects, line drawings, silhouettes) and tasks (detection vs. identification), I found that the visual system utilizes a limited number of features for object recognition, and that these features are extracted at or near regions of contrast discontinuity. Further, the visual system imposes these limitations in an effortful insensitivity to stimulus orientation and position, even under circumstances when neither the invariance nor the feature placement is appropriate for the task at hand (Tjan, Braje, Legge, & Kersten, 1995; Braje, Tjan, & Legge, 1995). These results showed that the visual system is surprisingly inflexible in choosing low-level features on-demand.

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 should be moderately common among objects and rare among random clutter or texture. By formalizing this intuition in terms of a Bayesian prior, it can be shown that a visual system optimized for object recognition will have difficulty discriminating small deviation in these features, even though the detection efficiency for such features remain high. With Zili Liu, a long-term collaborator, I 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.

Representational demand and adaptive representation. How does the visual system represent objects such that they can be identified regardless of viewpoint or illumination? I invented a Bayesian ideal-observer method to measure the extent and details to which an object must be represented for a given task and stimulus set. I found that the demand of representation varies hugely across different ensembles of objects. The representational demand for the same object depends strongly on what the other objects are in the ensemble. Object ensembles with a high representational demand are also those that humans find difficult to learn and to generalize from one view to another. (Tjan & Legge, 1998). These findings 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 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. This mechanism takes advantage of the fact that features become progressively more complex, invariant, and abstract along a processing pathway – that is, the visual system provides a full range of representations at different levels of abstractions (Tjan, 2001). I plan 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. I use English letters to study form vision in the periphery because they are simple and also because they have real-world importance: reading is a very difficult task for patients with central vision loss. I developed an ideal-observer model for letter identification, using data collected by my collaborators Susana Chung and Gordon Legge (Chung, Legge & Tjan, 2002). The ideal-observer model showed that both central and peripheral visual systems use features of sizes appropriate to their respective spatial resolution and the size of the stimuli. In other words, if one controls for the differences in spatial resolution (as characterized by the contrast sensitivity functions), the spatial tuning for letter identification is optimal for both central and peripheral vision! The same surprising finding was obtained in the amblyopic fovea (Chung, Tjan & Levi, 2002), and in crowding (Chung & Tjan, 2007). Recently, my laboratory developed a classification-image method to visualize the perceptual templates and features used by human observers in peripheral vision. We found no qualitative difference between central and peripheral vision in terms of their first-order perceptual templates and second-order features, when a letter is presented alone (Tjan & Nandy, 2006; Nandy & Tjan, 2007). These results suggest that the features used for letter identification are qualitatively similar in central and peripheral vision. Nevertheless, peripheral form vision remains severely impaired. We hypothesize that this deficit could result from the manner in which features are integrated. We sought proof of this hypothesize by investigating the phenomenon of "crowding."

Crowding refers to the phenomenon that a target becomes unrecognizable when flanked by other items, even though it can be recognized when presented alone. It is pronounced in peripheral vision but almost non-existent in central vision. The mechanism of crowding is still unknown. In collaboration with Susana Chung, my laboratory has gathered a broad range of empirical data on crowding. Briefly, we found that: (1) spatial tuning for a crowded target is close to optimal (Chung & Tjan, 2007) – in other words, the proper set of low-level features are utilized by the visual system in the crowded condition. (2) Crowding appears to be due to an increase in the internal noise of the visual system, induced by the flankers. (3) The necessary and sufficient condition for inducing this internal noise is for the flankers to have a similar spatial-frequency distribution as that of the target; specifically, we found that a phase-scrambled flanker is as effective as a phase-intact one. (4) Spatial uncertainty in peripheral vision plays a catalytic role in causing crowding – it only leads to crowding when spatial resolution is low. (5) The neural origins of crowding appear to be "bottom-up" from visual areas as early as V2.

A 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 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 co-edited (with Denis Pelli, Robert Desimone, Patrick Cavanagh and Anne Treisman) a special issue on crowding for the Journal of Vision. I also co-organized (with Pelli & Cavanagh) a symposium on crowding at the 2006 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 higher-order features used by the visual system and to reveal the ordering of information processing in the human cortex.

Higher order classification images. The method of classification images provides a way to visualize the input image pattern that leads to a particular response. Most of the current classification-image techniques are linear methods. They cannot reveal the perceptual templates used by the visual system if many distinct input patterns are mapped to the same response (e.g. in an object identification task when objects are seen from different viewpoints). I have developed a novel method to overcome this limitation (Tjan & Nandy, 2006) and successfully applied it to obtain classification images in the visual periphery, where the visual system has a high degree of uncertainty about the spatial location of a stimulus. I have also derived a second-order classification-image method to infer the features the visual system uses when performing a letter-identification task (Nandy & Tjan, 2007). We use these techniques extensively in our current investigations of peripheral vision and amblyopia. I also 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 these 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 – they 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 log-log steepness of the BOLD response function (log %BOLD vs. log image SNR) is related to 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/B) (Tjan, Lestou & Kourtzi, 2006). This encouraging result suggests that other theoretical properties of an ideal-observer cascade may 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.

Translational Research

Wanting to help people with visual impairments motivated my research in peripheral vision. The same motivation underlies two current projects that develop assistive technologies for the blind and visual impaired.

Digital Sign System for Indoor Wayfinding. I invented a barcode-like sign and the corresponding computer-vision algorithm that are at the core of an indoor navigation system designed for the blind and the visually impaired (Tjan et al., 2005). The low-cost sign, the size of a credit card and printed on a special retro-reflective material, can be reliably detected and identified at 30 frames per second within a distance of 10 feet by a small handheld camera. Each sign contains a unique 16-bit number, which is used as an index into a spatial database that communicates the location-specific information to a user. In collaboration with Gordon Legge and a Minnesota company (AME), we have built a robust prototype as part of a SBIR project funded by NIH. We are testing the prototype with low-vision and blind subjects.

Mid-Level Vision Systems for Low Vision. I am leading the human testing and modeling efforts in a newly awarded Bioengineering Research Partnership grant. The project, coordinated by Norberto Grzywacz, is to develop image-enhancement systems for people with central-vision loss. It built on two new technologies: a neuromorphic contour extraction algorithm invented by Bartlett Mel, and a region-sensitive contrast enhancement algorithm developed by Gerard Medioni. These technologies allow us to parse a scene into meaningful regions of objects such that we can selectively enhance and/or simplify different regions in the display to minimize crowding and facilitate form perception in the visual periphery. My expertise in periphery vision informs the choices of enhancements. I have designed a battery of psychophyiscal tasks to evaluate and predict the efficacy of a given image-enhancement scheme. I will also develop a training regimen, assisted by machine learning, to both train individual users and to tune the system to the user's needs.

Summary

My research on object recognition addresses basic questions about form vision, while my research on peripheral vision seeks to translate the answers to clinical solutions. These two lines of research will continue to interact in the foreseeable future. I am motivated by the results we have obtained so far. My midterm goal is to develop a computer model that describes the early stages of form vision in the periphery. In addition to testing our theories about form vision, such an observer model will be instrumental in the development of image enhancement technologies for the visually impaired.