# Reverse-Engineering Computational Models for Human Visual Perception using Crowdsourcing

Karan Goel, 2011EE50555

September 15, 2015

**Abstract**

Crowdsourcing is a popular data-collection paradigm, exploited in particular to collect data for computer vision applications such as segmentation and object detection. Since data is collected from multiple human workers, it is inherently noisy and presents a challenge to researchers interested in purifying, or making inferences from this data. An interesting aspect of this is trying to incorporate human models of perception in understanding the kinds of perception mistakes made by humans, as well as generalities in cognitive human behavior when assigned a vision task. This is an interesting and challenging problem since the workers themselves are not directly accessible (or knowable), but only the data generated by them (and the task given to them) is available for examination. This proposal outlines the areas of visual cognition that the term paper will focus on, and the kinds of questions that the paper will hope to address, while drawing from the crowdsourcing and vision literature.

## Introduction

The human visual cortex is a incredibly complex system, and recent work by visual systems groups has attempted to reverse-engineer implementations of the human eye, and the visual cortex as a whole (Thorpe et al. [2000]). Other researchers have attempted to understand specific aspects of human visual behavior - how humans recognize faces (Sinha et al. [2006]), detect objects in environments (Oliva and Torralba [2006]), and looked at neurophysiological responses to object detection (Cox et al. [2004]) among a wide, multi-disciplinary body of work.

This proposal restricts itself to considering only *computational aspects* of human vision, and specifically towards understanding how researchers have attempted/can attempt to model the functions or behavior of the human visual cortex mathematically. Implicit in this consideration is that there is an underlying gold standard of comparison for researchers, which is the behavior of the visual cortex in practice, and whether it conforms to this model.

## Why is this an important problem?

Developing models of computation for visual behavior are essential for developing sophisticated robotic systems, such as self-driving cars (Lee et al. [2013]), and developing a prosthetic visual system (Cohen [2007]), among other applications. Computer Vision research has seen a long-term shift from purely computational approaches, to a more interdisciplinary approach informed by research in the cognitive sciences literature.

A major problem in this research agenda has been a lack of data, due to the difficulty of data collection from human subjects (Deng et al. [2009]). The advent of crowdsourcing has provided a new impetus to researchers across a variety of fields, including Computer Vision (Wah [2006]) and Natural Language Processing (Sabou et al. [2012]), among others.

There are broadly 2 ways in which crowdsourcing has been important in this regard (and which we will be concerned with),

- Easy collection of large amounts of vision data, for tasks such as image segmentation(Salvador et al. [2013]), object detection(Deng et al. [2013]), etc. as well as for general large-scale vision datasets (Lin et al. [2014]).
- Cheap access to a large number of different human subjects (termed *crowd-workers*) to get multiple responses for the same task. (Quinn and Bederson [2011])

However, due to the nature of crowdsourced work, it is rare for researchers making use of crowdsourcing to meet crowd-workers, much less have persistent access to them. Thus, while data acquisition becomes easy, and results in rich data collection, an important challenge is to be able to understand the *kinds of variations* in responses crowdsourced subjects may produce, and to develop *computational models* to account for them.

Specifically, for vision tasks, this presents an interesting challenge and opportunity, which adds to the focus of traditional computer vision, which is concerned with building accurate models, towards an additional focus on modeling perception as well. There is a new need to understand *how the data was collected* and to model *uncertainty and variation* in the responses collected. This can undoubtedly be done more systematically if the underlying differences between worker perceptions are well understood.

While in practice, researchers may not always invoke or make use of cognitive science, this proposal and term paper aim to advocate a shift towards a systematic modeling of human cognitive processes (for vision, in particular), and to demonstrate successes and failures, if any.

## Broad Outline

The major focus of the paper will be to examine the ways in which Crowdsourcing, as a data collection tool, can be leveraged to carry out interdisciplinary research on the human visual system. The paper will attempt to look at existing approaches, and after differentiating the success and failure cases, lay out ways in which researchers can attempt to make progress in this area.

The basic questions that will be addressed in the paper are,

- How does interface design effect the data collection process? Can we design interfaces in ways that can help us elicit valuable information about how the human visual system works?
- Concretely, how can one encode reliability into the visual interface design, so that given some data, and full information about the interface, we can confidently make inferences about the *intent* of individual workers, as well as the *perceptual differences* between worker responses?
- How can we resolve differences in perception between different humans in this setting into a single, computational model? Are the models reliable, and if so do they capture something particularly interesting about how humans see the world?
- Can these models be used to inform fundamental research into the human visual cortex?

HCI researchers have attempted to conduct experiments on human visual perception, to understand how to improve graphical visualizations (Heer and Bostock [2010], Kong et al. [2010]), learn about human color perception (McLeod [2014]), visual summarization (Rudinac et al. [2013], Robb et al. [2015]), etc. Others have attempted to leverage crowdsourced data to develop probabilistic models to create segmentation (Welinder and Perona [2010]), and visual recognition systems (Long et al. [2013]).

Taking cues from this work, the paper will focus attention on the task of color perception as a case study, perhaps with crowdsourced experiments to validate any claims that are made. For this particular task, there is a wide variety of work in the computational perception and cognitive sciences literature, which the paper will draw on (Davidoff [1991], Ullman and Humphreys [1996], Jacobs [2013], Kaiser and Boynton [1996], Papathomas et al. [1991], Robertson [2008]).

## Expected Outcomes

The aim of the paper will be to advocate the use of crowdsourcing by cognitive scientists to develop computational models. Thus, a major focus will be to comment on the *efficacy* of any claims made in the paper. While crowdsourcing has been successfully used by vision researchers, the further question that is implicit in this proposal is whether those models would align in principle with existing research in human perception by cognitive scientists.

An excellent conclusion would be that crowdsourcing has certain indispensable characteristics, which must be used to gain a wider, more general understanding of the visual system.

## References


Ethan D Cohen. Prosthetic interfaces with the visual system: biological issues. *Journal of neural engineering*, 4(2):R14, 2007.

David Cox, Ethan Meyers, and Pawan Sinha. Contextually evoked object-specific responses in human visual cortex. *Science*, 304(5667):115–117, 2004.

Jules B Davidoff. *Cognition through color*. Mit Press Cambridge, MA, 1991.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

Jia Deng, Jan Krause, and Li Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 580–587. IEEE, 2013.

Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 203–212. ACM, 2010.

Gerald Jacobs. *Comparative color vision*. Elsevier, 2013.

Peter K Kaiser and Robert M Boynton. Human color vision. 1996.

Nicholas Kong, Jeffrey Heer, and Maneesh Agrawala. Perceptual guidelines for creating rectangular treemaps. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):990–998, 2010.

Gim Hee Lee, Friedrich Faundorfer, and Marc Pollefeys. Motion estimation for self-driving cars with a generalized camera. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2746–2753. IEEE, 2013.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014*, pages 740–755. Springer, 2014.

Chengjiang Long, Gang Hua, and Ajay Kapoor. Active visual recognition with expertise estimation in crowdsourcing. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3000–3007. IEEE, 2013.

Ryan Nathaniel McLeod. A proof of concept for crowdsourcing color perception experiments. 2014.

Aude Oliva and Antonio Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155:23–36, 2006.

Thomas V Papathomas, Andrei Gorea, and Bela Julesz. Two carriers for motion perception: Color and luminance. *Vision Research*, 31(11):1883–1892, 1991.

Alexander J Quinn and Benjamin B Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1403–1412. ACM, 2011.

David A Robb, Stefano Padilla, Britta Kalkreuter, and Mike J Chantler. Crowdsourced feedback with imagery rather than text: Would designers use it? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1355–1364. ACM, 2015.

Alan R Robertson. Color perception. *Physics Today*, 45(12):24–29, 2008.

Stevan Rudinac, Matt Larson, and Alan Hanjalic. Learning crowdsourced user preferences for visual summarization of image collections. *Multimedia, IEEE Transactions on*, 15(6):1231–1243, 2013.

Marta Sabou, Kalina Bontcheva, and Arno Scharl. Crowdsourcing research opportunities: lessons from natural language processing. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, page 17. ACM, 2012.

Amaia Salvador, Axel Carlier, Xavier Giro-i Nieto, Oge Marques, and Vincent Charvillat. Crowdsourced object segmentation with a game. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, pages 15–20. ACM, 2013.

Pawan Sinha, Benjamin Balas, Yuri Ostrovsky, and Richard Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

SJ Thorpe, A Delorme, R Van Rullen, and W Paquier. Reverse engineering of the visual system using networks of spiking neurons. In *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, volume 4, pages 405–408. IEEE, 2000.

Shimon Ullman and Glyn W Humphreys. *High-level vision: Object recognition and visual cognition*, volume 2. MIT press Cambridge, MA, 1996.

Catherine Wah. Crowdsourcing and its applications in computer vision. *University of California, San Diego*, 2006.

Peter Welinder and Pietro Perona. Online crowdsourcing: rating annotators and obtaining cost-effective labels. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 25–32. IEEE, 2010.