Interconnected Recursive Filters in Artificial and Biological Vision

Aravind Battaje and Oliver Brock

Abstract— **We aim to find fundamental principles of robust vision and express them as interconnected recursive filters, which is a network capable of feed-forward and feed-back information. We demonstrate that certain visual illusions can be explained using interconnected recursive filters, while also serving as an algorithmic architecture to build robotic vision applications.**

Preferred type of presentation: Oral

I. Motivation and Problem Definition

Today's artificial vision is hardly as versatile as biological vision. Artificial vision works really well in human-designed niches, but are not robust in general situations. This hinders producing intelligent robotic solutions for day-to-day tasks. To achieve similar performance as biological vision then, one way is to identify the fundamental differences and import the qualities necessary for robust vision. However on the one hand, those qualities are not well known, and on the other, such qualities may not be easily expressible for artificial vision.

II. Related Work

Historically there have been many attempts to mimic neural mechanisms for machine vision. The most prominent work [1] in deep-learning cemented CNNs into common usage, but it still deviates in crucial ways from biological vision. For example, it is well known that information flows top-down and bottom-up in the visual cortex [2]. We believe CNNs are insufficient and fragile [3], and also don't allow mechanisms that are necessary for robust vision.

III. Own Approach and Contribution

Some important characteristics that have been identified in biological vision, but not fully leveraged yet in artificial vision are: multi-directional information flow [2], crossmodal fusion between different aspects of sensory information [4] and temporal coherence [5]. We believe interconnected recursive filters provide a language to express these qualities in a holistic way, because it is composed of a network of recursive (Bayes) filters that allow probabilistic fusion of multiple information sources while associating multiple priors (such as time consistency). Martín-Martín et al. [6] have already applied this to a robotic problem to detect and track kinematic degrees-of-freedom of arbitrary objects.

Both authors are with the Robotics and Biology Laboratory, Technische Universität Berlin, Germany and the Cluster of Excellence "Science of Intelligence" (SCIoI). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2002/1 "Science of Intelligence" - project number 390523135.

Given that we have a framework to express qualities necessary for robust vision, we follow an iterative strategy to accelerate our process: From hypotheses about certain characteristics of human vision, we build synthetic models using interconnected filters to verify those hypotheses. With the deviations and insights from such modeling, we perform psychophysical experiments on humans, which in turn reveals new strategies to model. Since we use interconnected filters during the process, it can easily transfer to robotic applications as demonstrated by [6].

We attempt such an iterative strategy to understand mechanisms in human vision related to shapes and color: We emulate two illusions using interconnected recursive filters, viz., "Filling-in afterimage between the lines" [7] elicits illusory colors in shapes as an aftereffect, and "neon-color spreading" [8] bleeds colors confined to contours of a shape. We explain such illusions by a constraint that arises from tightly coupling shape and color perception. Some may regard these illusions as glitches in human vision, but we believe this can be seen as taking advantage of statistics of natural image sequences [9]. We have still not completed a full verification with psychophysical experiments, but literature [9] already supports our view.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[2] M. Ahissar and S. Hochstein, "The reverse hierarchy theory of visual perceptual learning," *Trends in Cognitive Sciences*, vol. 8, no. 10, pp. 457–464, Oct. 2004.

[3] A. Barbu, D. Mayo, J. Alverio, W. Luo, C. Wang, D. Gutfreund, J. Tenenbaum, and B. Katz, "Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[4] H. MCGURK and J. MACDONALD, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, Dec. 1976.

[5] Á. Kristjánsson, K. Ó. Eyjólfsdóttir, A. Jónsdóttir, and G. Arnkelsson, "Temporal consistency is currency in shifts of transient visual attention," *PLoS ONE*, vol. 5, no. 10, p. e13660, Oct. 2010.

[6] R. Martín-Martín and O. Brock, "Coupled recursive estimation for online interactive perception of articulated objects," *The International Journal of Robotics Research*, p. 0278364919848850, May 2019.

[7] R. van Lier, M. Vergeer, and S. Anstis, "Filling-in afterimage colors between the lines," *Current Biology*, vol. 19, no. 8, pp. R323–R324, Apr. 2009.

[8] P. Bressan, E. Mingolla, L. Spillmann, and T. Watanabe, "Neon color spreading: A review," *Perception*, vol. 26, no. 11, pp. 1353–1366, Nov. 1997.

[9] N. Krüger and F. Wörgötter, "Multi-modal estimation of collinearity and parallelism in natural image sequences," *Network: Computation in Neural Systems*, vol. 13, no. 4, pp. 553–576, Jan. 2002.