Chromatic Statistics and Information in Natural Images

P. Garrigan, C. Ratliff, P. Sterling, D.H. Brainard, V. Balasubramanian

University of Pennsylvania

To enable color vision, the human photoreceptor mosaic has three cone types sensitive to long (L), medium (M), and short (S) wavelength light. Two outstanding puzzles about this mosaic are that the S cones are rare (<10%), and that L/M ratio varies enormously between individuals. We measured how mosaics with differing proportions of cone types encode natural images and found that these two features of the human retina are consistent with a design that maximizes the encoded information.

We acquired a database of natural images with a calibrated camera. The image values at each pixel represented estimates of the L, M, and S cone quantal absorption rates. We measured 2- and 3-point spatial correlations, both within and across cone channels. The within-channel 2-point correlations were considerable (>.25) up to separations of 100 pixels. The cross-channel 2-point correlations were also large, even for the L and S cones. We found that transforming the LMS cone images into a luminance channel and three cone-difference channels (cone absorption rate at each pixel minus the local mean luminance) led to a representation in which the four resulting channels were largely decorrelated from one another. In addition, while the luminance channel contained substantial 2-point spatial correlations at large separations, the 2-point correlations within the cone-difference channels dropped rapidly as separation increased. These latter correlations were close to zero for separations larger than 10 pixels.

If an image ensemble contains no structure beyond that captured by the 2-point correlations, then the 3-point correlations are derivable from the 2-point correlations. We found that 90% of the 3- point correlations were derivable from underlying 2-point correlations. We therefore estimated information in natural images on the assumption that the 2-point correlations adequately characterized image spatial structure. We asked what fraction of the cone mosaic should be L, M, and S to maximize information transfer from these images.

Because of long-range 2-point correlations, the information in the luminance channel grows slowly with the size of the cone mosaic (proportional to N^0.75 where N is the total number of cones.) Because of the more rapid falloff in their 2-point correlations, the information in the cone-difference channels grows nearly linearly with N. For this reason, the total information encoded by large arrays is dominated by the contribution from the cone-difference channels. Information encoded by single cone-difference response was estimated as I = 1/2 log(1 + SNR), with SNR estimated for each cone type (L, M, or S) in terms of corresponding channel variance compared to photon and dark noise. If there were no spatial correlations, the optimal mosaic would contain only the cone type with the highest value of I, but consideration of spatial correlations favors a mixed mosaic. We incorporated within-channel spatial correlations into our calculations. The results showed that the optimal mosaic is dominated by L and M cones, but that the amount of information encoded is largely indifferent to the precise L/M cone ratio. This matches the observed sparseness of S cones and the large variation in L/M ratio within the human population.

Acknowledgments
We thank L. Seyfarth for collecting images and J. M. Klein for assistance with camera calibration. This work was supported by NSF IBN-0344678, NIH EY10016, & NIH EY08124.