Gabor Mosaics: A Description of Local Orientation Statistics, with Applications to Machine Perception.

Javier R. Movellan and **Ram S. Prayaga**
Cognitive Science 0515
University of California, San Diego
La Jolla, CA 92093
{movellan, rprayaga}@cogsci.ucsd.edu

Our goal is to understand the role of natural signal processing by simulating it on machine perception tasks. We use Gabor filters as a first approximation to the operation of primate striate cortex. It has been proposed that besides modeling the receptive fields of simple cells in striate cortex (Jones & Palmer, 1987), Gabor receptive fields provide an efficient and sparse code for natural images (Field, 1994; Daugman, 1989). It has also been proposed that one of the functions of this sparse code may be to facilitate visual pattern recognition (Field, 1994). We test this prediction on a machine vision task: automatic, speaker independent lipreading of the first four digits in English.

Automatic visual speech recognition is being extensively studied in the machine perception and cognitive science communities. Besides its role in improving the intelligibility of acoustic speech, visual information automatically modulates the perception of auditory signals (McGurk & MacDonald, 1976; Massaro & Cohen, 1983). In recent years lipreading has also received attention for its application in the automatic recognition of speech in noisy environments and because it provides a tractable avenue to computational issues on intermodal integration (Yuhas, Goldstein, Sejnowski, & Jenkins, 1990; Bregler, Hild, Manke, & Waibel, 1993; Wolff, Prasad, Stork, & Hennecke, 1994; Movellan, 1995).

In this project, we use a biologically inspired approach to early visual processing. We model striate cortex as a bank of Gabor filter and propose a method to visualize local texture and orientation using these filters. We model later recognition processes as a bank of hidden Markov models, the most successful system for the automatic recognition of visual and acoustic speech(Rabiner & Juang, 1993). The recognition engine was trained and tested using the following database.

Training Sample: The training sample consists of 4504 images (456 movies) of 57 undergraduate students from the Cognitive Science Department at UCSD. Subjects were asked to talk into a video camera and to say the first four digits in English twice. Subjects could see their lips in a monitor and were told to position themselves so that their lips were relatively centered in the monitor. We tried a variety of illumination conditions and allowed the lips not to be entirely centered.

Current Directions: We investigate whether Gabor based representations do indeed create a sparser code in our database. We analyze the resistance of this representation to changes in illumination and present results comparing the performance of a pixel-based approach and the Gabor-based approach.

References

Bregler, C., Hild, H., Manke, S., & Waibel, A. (1993). Improving connected letter recognition by lipreading. *Proc. Int. Conf. on Acoust., Speech, and Signal Processing*, Vol. 1 (pp. 557–560). IEEE, Minneapolis.

Daugman, J. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering, 36*, 107–114.

Field, D. (1994). What is the goal of sensory coding? *Neural Computation, 6*, 559–601.

Jones, J. P., & Palmer, L. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*, 1233–1258.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 753–771.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Movellan, J. R. (1995). Visual speech recognition with stochastic networks. In G. Tesauro, D. Touretzky, & T. Leen (Eds.), *Advances in Neural Information Processing Systems*, Vol. 7. MIT Press.

Rabiner, L. R., & Juang, B.-H. (1993). *Fundamentals of speech recognition.* Englewood Cliffs, NJ: Prentice-Hall.

Wolff, G. J., Prasad, K. V., Stork, D. G., & Hennecke, M. E. (1994). Lipreading by neural networks: Visual preprocessing, learning and sensory integration. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), *Advances in Neural Information Processing Systems*, Vol. 6 (pp. 1027–1034). Morgan Kaufmann.

Yuhas, B. P., Goldstein, M. H., Sejnowski, T. J., & Jenkins, R. E. (1990). Neural network models of sensory integration for improved vowel recognition. *Proc. IEEE, 78*(10), 1658–1668.