

Reading Summarization

Yue Gao

February 10, 2011

1 Last Meeting Summary

1. For a given visual scenes, how people may use scene priors in their interpretation of scenes, how detailed these priors are, how they are structured, and how these priors may be learned over different time scales.
2. Hierarchical scene understanding . Generative model on high level categorization and discriminative model on low level,in class discriminations.
3. self-supervised learning for cross-modal clustering. How different sensors can communicate with each other and to use different type of modalities to achieve harder tasks. (lip reading) [1]

2 Paper Reading Feedbacks

1. The Eckstein paper (2006) [2] mainly talks about how context cues affect saccadic targeting. He proposed two models, limited attentional resources or differential weighting of information. The second model is more closer to the human results. The model assumes that each of the considered locations in the scene elicits an internal response for each target-relevant visual feature. All internal responses are subject to Gaussian independent neural noise. The model then calculates for each location a joint likelihood of observing the feature responses given that the target is present at that location and a joint likelihood of observing the feature responses given that the target is absent at that location. The ratios of these two likelihoods distributed over the scene create a likelihood-ratio map (a map of evidence for target presence). The evidence is weighted by a map of weights specifying the prior probabilities of the target appearing at each location.
2. The Elze paper (2011) [3] is a very interesting of designing the experiment. However, I am not so sure this is purely vision. Because Germany is a language using abstract notion while Chinese is a language which is like a figurative language. Very two different concept and two different ways of thinking.
3. The Kanan and Cottrell paper (2010) [5] uses a model that combines sequential visual attention using fixation with sparse coding.
4. The Xu and Chun paper (2009) [6]describes how human dealing with multiple objects.
5. The Epshtein et al paper (2008) [4] used Bidirectional Hierarchical model. the goal is to simulate the feed-forward processing. Therefore they want to model $\max P(C, P|F)$ where C is the object, P is a set of parts for the object and F is a set of features.

3 topics

1. natural scene understanding (prior, object spatial location)
2. How to learn new category labels
3. online learning on Vision (encode knowledge space into vision world), maybe not just hierarchical or using first order logic.
4. Reinforcement learning theory

References

- [1] M.H. Coen. Cross-modal clustering. In *Proceedings Of The National Conference On Artificial Intelligence*, volume 20, page 932. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.
- [2] M.P. Eckstein, B.A. Drescher, and S.S. Shimozaki. Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, 17(11):973, 2006.
- [3] T. Elze, C. Song, R. Stollhoff, and J. Jost. Chinese characters reveal impacts of prior experience on very early stages of perception. *BMC neuroscience*, 12(1):14, 2011.
- [4] B. Epshtein, I. Lifshitz, and S. Ullman. Image interpretation by a single bottom-up top-down cycle. *Proceedings of the National Academy of Sciences*, 105(38):14298, 2008.
- [5] C. Kanan and G. Cottrell. Robust classification of objects, faces, and flowers using natural image statistics. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2472–2479. IEEE, 2010.
- [6] Y. Xu and M.M. Chun. Selecting and perceiving multiple visual objects. *Trends in cognitive sciences*, 13(4):167–174, 2009.