Cross-Modal Disambiguation

Kobus Barnard
Department of Computer Science
University of Arizona
kobus@cs.arizona.edu

Abstract

Semantic ambiguity is common when we attempt to extract meaning from data. In text, words like "bank" are ambiguous without a greater context. Similarly, image patches are indicative of scene elements, but typically require image context to choose among likely options. Thus there has been much work done to integrate world knowledge and various levels of context to disambiguate isolated cues. In this talk I will discuss extending context across modalities for disambiguation.

Multimodal data is rapidly becoming more available. Words with pictures, pictures with geotags, sound with video, and presentation slides with educational video represent some key frontiers. As an example of how multimodal data can held disambiguate meaning, consider an image associated with text with the word "bank". If the image can be classified as a city scene or an outdoor scene, then that can help resolve the financial institution sense versus the embankment (e.g., "river bank") sense. Similarly, the ambiguity of a blue patch in an image can be reduced if the image has an associated label suggesting sky or water. Further, parts of speech provide sub-modalities that can be exploited. For example, a visual adjective such a "red" in an image caption containing "red car" can help interpret the image data. In this talk I will provide some results in two domains: 1) words and pictures; and 2) using presentation slides to inform the analysis of educational video.

Biography

Kobus Barnard is an associate professor of computing science at the University of Arizona. He received his Ph.D. in computer science from Simon Fraser University in the area of computational color constancy. He then spent two years at the University of California at Berkeley as a post doctoral researcher working on image understanding and digital libraries. His current interests include stochastic modeling approaches to machine vision problems and applications to scientific data.