2005 SLIS Doctoral Research Forum
Caroline Beebe, School of Library and Information Science, Indiana University, Bloomington
beebe@indiana.edu

Bridging the Semantic Gap: Exploring Descriptive Vocabulary for Image Structure

Content Based Image Retrieval [CBIR] is an emerging technology made possible by the binary nature of the computer. It is being applied to the representation and retrieval of digital image resources. There is a disconnect, however, between the binary code of the computer and the intellectual (conceptual) code of the searcher. This is referred to as the "semantic gap." Identifying a descriptive perceptual vocabulary for the structural components of the physical image has the potential to establish a connection between these two codes.

Graphic resources are proliferating on the World Wide Web, and industry estimates indicate that 80 billion new images are created every year. Verbal description has been the standard method of image representation. Assigning descriptors to images, however, is time-consuming and often requires subject matter expertise. Indexers have created descriptive vocabularies, such as the Library of Congress Subject Headings or the Art and Architecture Thesaurus, but these vocabularies are broad in scope, have little user input, and don't address the physicality of images upon which CBIR similarity judgments are based. This failure to represent the physical composition of images is indicated by exploratory research of Jorgensen et al. (2001) that image retrieval descriptions provided by both indexers and naïve users frequently included attributes at the physical level.

Typically, CBIR applications use some form of query-by-pictorial-example [QBPE] as a search interface. The pixel configuration of the example image is used to identify images with similar pixel configurations. The system, however, is not informed of the searcher's similarity criteria and searchers are not supplied explanations of the similarity measures used to generate QBPE results. An emerging direction in CBIR research is for interfaces to help users correlate semantics with perceptual cues rather than to have the computer automatically identify perceptual similarities and then attempt to match them to semantic queries.

This research proposes to generate a user-centered controlled vocabulary for the perceptual cues in images. The vocabulary is taken directly from natural language descriptions made by naïve users. The research design is modeled after a study by Krauss & Fussell (1991) constructing shared communicative environments. For image structure research, the shared environment involved a searcher and a human surrogate for the CBIR search mechanism. Ten pairs of subjects were asked to generate verbal descriptions for a set of images under the condition where one partner generates a representational drawing for each description. Images from three content areas were selected for their lack of readily identifiable semantic objects or scenes.

Currently in the data analysis phase, terms used in the verbal descriptions are being extracted to generate a faceted perceptual vocabulary through concept analysis, word frequency, and type-token ratios (number of different words to total words used by each subject). Preliminary facets, frequencies, and problem areas will be discussed. This vocabulary has potential use in the design of CBIR interfaces supporting image query formulation and similarity judgments. The facets can potentially provide CBIR developers with target concepts, based on user need, for prioritizing and focusing their binary techniques.

Jorgensen, C., Jaimes, A., Benitez, A. B., & Chang, S.-F. (2001). A conceptual framework and empirical research for classifying visual descriptors. *Journal of the American Society for Information Science and Technology, 52*(11), 938-947.
Krauss, R. M., & Fussell, S. R. (1991). Constructing shared communicative environments. In L. B. Resnick & J. M. Levine & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 172-200). Washington, DC: American Psychological Association.