Bridging the Semantic Gap: Exploring Descriptive Vocabulary for Image Structure

Caroline Beebe

Abstract—**This poster summarizes the methodology and results from an experiment to collect terms used by subjects in the verbal description of images from the domains of abstract art, satellite imagery and photo-microscopy. The resulting natural language vocabulary was then analyzed to identify a set of concepts that were shared across subjects. These concepts were subsequently organized as a faceted vocabulary that can be used to describe the shapes and relationships between shapes that constitute the internal spatial composition -- or internal contextuality -- of images. This vocabulary has the potential to minimize the terminological confusion surrounding the representation of the content and internal composition of digital images in Content-Based Image Retrieval [CBIR] systems.**

Index Terms— **Image analysis, Image representations, Image shape analysis, Semantic networks**

I. Introduction

THIS research makes a methodological contribution to the development of faceted vocabularies and suggests a potentially significant tool for the development of more effective digital image retrieval systems. The research agenda applied an innovative experimental methodology to collect terms used by subjects in the description of images drawn from three domains. The resulting natural language vocabulary was then analyzed to identify a set of concepts that were shared across subjects. These concepts were subsequently organized as a faceted vocabulary that can be used to describe the shapes and relationships between shapes that constitute the internal spatial composition -- or internal contextuality -- of images. Because the vocabulary minimizes the terminological confusion surrounding the representation of the content and internal composition of digital images in Content-Based Image Retrieval [CBIR] systems, it can be applied to develop more effective image retrieval metrics and to enhance the selection of criteria for similarity judgments for CBIR systems.

CBIR is a technology made possible by the binary nature of the computer. Although CBIR is used for the representation and retrieval of digital images, from an end-user (searcher) perspective these systems make no attempt either to establish a basis for similarity judgments generated by query-by-pictorial-example searches or to address the connection between image content and its internal spatial composition. The disconnect between physical data (the binary code of the computer) and its conceptual interpretation (the intellectual code of the searcher) is known as the semantic gap. A user-generated descriptive vocabulary capable of representing the internal visual structure of images has the potential to bridge this gap by connecting physical data with its conceptual interpretation.

This research project addressed three questions: Is there a shared vocabulary of terms used by subjects to represent the internal contextuality (i.e., composition) of images? Can the natural language terms be organized into concepts? And, if there is a vocabulary of concepts, is it shared across subject pairs?

II. Methodology

A natural language vocabulary was identified on the basis of term occurrence in oral descriptions provided by 21 pairs of subjects participating in a referential communication task: one subject provided an oral description of an image that allowed her partner to produce a drawing of it. In this experiment, each subject pair generated oral descriptions for 14 of 182 images. These 182 images were drawn from the domains of abstract art (ART), satellite imagery (SAT) and photo-microscopy (MIC). Images in these domains can be considered non-application specific, and exhibit "ofness" but little "aboutness", that is, they contain minimal semantic components.

III. Results

Processing the verbal description data involved three major steps: transcribing, faceting, and tabulating. There were 107,581 natural language words from which 221 unique stop words were identified and removed, leaving 51,629 natural language words. These words were normalized both for variant form and semantic relationships (synonyms and near-synonyms) leaving 1319 unique terms. These terms were collapsed into 545 concepts which were subsequently organized into a faceted vocabulary. The hierarchical structure of the faceted vocabulary provides a framework for evaluating the concept categories of terms generated by subject pairs. The top level concepts are the four facets Object, Place, Property, and Spatial Location.

Manuscript received November 15, 2007.

C. Beebe is with the School of Library and Information Science, Indiana University, Bloomington, IN 47605 USA (phone: 812-335-0701; e-mail: beebe@indiana.edu).

The potential of a shared vocabulary is more readily evaluated at the concept level where synonymous and near-synonymous terms have been clustered and the specificity of a term has less import. The result of collapsing isolates (values or instances of a property) into concepts had the advantage of generalization in that it alleviated the specificity of many isolates (e.g., individual color values into Hue). The process of collapsing isolates resulted in 465 concept categories actually used by subjects.

Frequency of occurrence and domain distribution were tallied for each term and concept of the vocabulary, see example in Table I. However, while frequency counts may indicate common usage, they do not demonstrate common usage across domains. Evaluation of domain distributions of individual concepts used by subjects may indicate that a concept was used in multiple domains but it does not indicate that subjects were actually using a shared vocabulary of concepts.

TABLE I
DOMAIN DISTRIBUTION OF CONCEPTS IN FOUR TOP FACETS

(N= 465 concepts)	ART	MIC	SAT	Total
Object	165	152	110	227
Place	40	38	79	90
Property	100	93	90	108
Spatial Location	40	39	39	40

Note: Domains are not mutually exclusive so the total does not represent the sum of the three domains.

Each concept subsumes all associated semantically coherent terms actually used by subjects.

A shared-ness rating scale was devised to measure subject agreement on concept use. Rank ordering of concepts by shared-ness measure demonstrated which concepts were more broadly shared across subject pairs. To determine if the concepts generated by subject pairs were used consistently by each pair across the three domains the subjects were considered to be "judges" and the Spearman rank correlation was computed to indicate inter-rater reliability. Although most discussions of interrater reliability focus on consensus, Stemler [1] contends that "consistency estimates of interrater reliability are based upon the assumption that it is not really necessary for two judges to share a common meaning of the rating scale, so long as each judge is consistent in classifying the phenomena according to his or her own definition of the scale."

Correlation analysis indicated that subject pairs tended to agree in the extent to which they used certain concepts across multiple domains and 14 concepts with the highest shared-ness sums would form the heart of a shared vocabulary: Line, Angle, Similarity, Linguistic Quantity, Number, Size, Degree, Presentation, General-part, Extremity, Inside-of, On, Vertical, Horizontal. Given the shared-ness ratings for these concepts, it is not unreasonable to assume that subjects would use them to describe images from all domains.

IV. CONCLUSION

This faceted vocabulary can contribute to the development of more effective image retrieval metrics and interfaces to minimize the terminological confusion and conceptual overlap that currently exists in most CBIR systems. For both the user and the system, the concepts in the faceted vocabulary can be used to represent shapes and relationships between shapes (i.e., internal contextuality) that constitute the internal spatial composition of an image. Representation of internal contextuality would contribute to more effective image search and retrieval by facilitating the construction of more precise feature queries by the user as well as the selection of criteria for similarity judgments in CBIR applications. In addition, reliance of subjects on the use of analogy (recalling that the images were selected based on a lack of readily identifiable objects, note the magnitude of Object distribution total in Table I) to describe images suggests that the faceted vocabulary of terms and concepts could be used to provide both the user and the CBIR system with a link to the visual shape represented by a verbal construct. Developing a visual vocabulary of shapes and relationships could be an important application of the controlled vocabulary that emerged from this study. Verbal access to concepts could serve as entry points leading into the visual vocabulary where shapes would be paired with specific low-level terms.

There are many different properties that are addressed in CBIR and image research. But such research is generally based on assumptions about which properties to study, and too often researchers create their own idiosyncratic category names. Categories are rarely identified by the research process or organized in a logical structure as was attempted in this study. The result of idiosyncratic naming is category overlap and terminological confusion. The faceted vocabulary could facilitate a narrowing of the semantic gap and alleviate terminological confusion by providing CBIR researchers with a structured approach to the terminology that they are attempting to link to image evaluation metrics. Consideration of frequency counts and shared-ness ratings in the faceted vocabulary could also provide focus for CBIR research regarding similarity metrics and for research exploring relationships among image parts. Because a faceted approach allows vocabulary to evolve and new concepts to be added, coordinating the various efforts of CBIR researchers would make an important contribution to bridging the semantic gap.

ACKNOWLEDGMENT

The author thanks Dr. Elin K. Jacob, School of Library and Information Science, Indiana University, for her reviews of this work, and gives her credit for coining important concept phrases: Internal contextuality, semantic organizers, semantic coherence, and the shared-ness measure.

REFERENCES

[1] S.E. Stemler. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Available: http://pareonline.net/getvn.asp?v=9&n=4.