Beebe, C. (2007). Bridging the semantic gap: Exploring descriptive vocabulary for image structure. 18th Annual ASIS SIG/CR Classification Research Workshop doi: 10.7152/acro.v18i1.12866

BRIDGING THE SEMANTIC GAP: EXPLORING DESCRIPTIVE VOCABULARY FOR IMAGE STRUCTURE

This research makes a methodological contribution to the development of faceted vocabularies and suggests a potentially significant tool for the development of more effective image retrieval systems. The research project 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, 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 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? 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. In this experiment, each subject pair generated oral descriptions for 14 of 182 images drawn from the domains of abstract art, satellite imagery and photo-microscopy. Analysis of the natural language vocabulary identified a set of 1,319 unique terms; these terms were collapsed into 545 concepts which were subsequently organized into a faceted vocabulary. Frequency of occurrence and domain distribution were tallied for each term and concept of the vocabulary. 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

ISSN: 2324-9773

computed to indicate inter-rater reliability. 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.

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 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.

ISSN: 2324-9773

Beebe, C. (2007). Bridging the semantic gap: Exploring descriptive vocabulary for image structure. 18th Annual ASIS SIG/CR Classification Research Workshop doi: 10.7152/acro.v18i1.12866

Introduction

Content-Based Image Retrieval (CBIR) is used for the representation and retrieval of digital image resources. In its most fundamental application, CBIR image analysis algorithms create indexes of images by comparing ranges of colors, arrangements of colors, and relationships among pixels. There is a disconnect, however, between the binary code (physical data) of the computer and the intellectual code (conceptual interpretation) of the searcher – a "lack of coincidence between the information that one can extract from visual data and the interpretation that same data have for a user in a given situation" (Smeulders, *et al.*, 2000, p. 5). This disconnect is referred to as the "semantic gap" (Smeulders *et al*., 2000; Stenvert, 1992). Identification of a descriptive vocabulary that is capable of representing internal visual structure of an image has the potential to bridge this gap by connecting the physical data to its semantic interpretation. This research addresses the question: Is there a non-specialist, natural language vocabulary that is appropriate for describing the internal contextuality – the internal visual structure – of digital images?

With the proliferation of consumer grade digital photography, billions of new images are created every year. Assigning descriptors to images to support text-based discovery of images is time-consuming and often requires subject matter expertise. Documentation of the collection-building process (Besser & Yamashita, 1998; Eakins & Graham, 1999) indicates that it takes ten to forty minutes per image for an art history expert to add access terminology to a single image. Application of CBIR technology has the potential to cut the time spent indexing images by generating automatic representations based on pixel configurations.

Adherence to accepted standards of controlled vocabularies and normalization of the language that people use in everyday communication provides for consistency and predictability in the indexing of resources. E. K. Jacob (personal communication, July 1998) presents a representation model consisting of physical description; conceptual description; and contextual description. *Physical description* includes both administrative data (e.g., access rights, object location, copyright holder, file type) and biographical characteristics (e.g., creator, title). The focus of *conceptual description* is both the naming of objects, or *ofness* (e.g., girl, artist, studio), and the interpretation of image reference, or *aboutness* (e.g., is it civil war or guerrilla warfare?).

ISSN: 2324-9773

Contextual description provides a context for the referent of the image (e.g., the relationship of an object to other objects in an image).

CBIR indexes expand the notion of physical description to include the pre-semantic physicality of the pixel relationships, or the physical visual structure. However, this dimension of physical characteristics may be tied more closely to the conceptual description of an image than to either its biographical or administrative characteristics. The interrelationship between the physical characteristics of the image and its conceptual and contextual descriptions affects *ekphrasis* – the conceptual interpretation and verbal representation of the image.

Typically, CBIR applications use some form of query-by-pictorial-example (QBPE) as a search interface. QBPE methods rely on searcher identification of one or more images that are similar to the query. The CBIR system then analyzes the pixel configuration of the example image(s) and uses the result to identify images with similar pixel configurations. In general, however, the system is not informed of similarity criteria used by the searcher and the searcher is not supplied with an explanation of the similarity measures used by the system to generate the result set. An emerging direction in CBIR research builds on the assumption that it is more useful for CBIR interfaces to help users correlate semantics with perceptual clues than to have the computer automatically identify perceptual similarities and attempt to match them to semantic queries (Goodrum *et al*., 2001; Santini & Jain, 1998; Smeulders *et al*., 2000). Research comparing perceptual judgments of image similarity produced by searchers and similarity ratings generated by CBIR methodologies indicates that there is a correlation between the metrics of image features and semantic description (Rogowitz *et al*., 1998) and is a growing research trend in CBIR (Datta *et al*., 2005).

Chu's (2001) analysis of image indexing and retrieval research indicates that there are two approaches to representation, which he describes as *content-based* (or low-level) and *description-based* (or high-level). Low-level metric features of the content-based approach have yet to be related to high-level semantic descriptions of the searcher, creating a "sensory gap" between computational description and the real-world object (Smeulders *et al.*, 2000, p. 4). Smeulders *et al.* (2000) indicate that the content-based and description-based approaches demonstrate an overlapping continuum between the development of semantic and syntactic representation methodologies. The traditional text-based approach – a high-level, user-centered approach – begins with conceptual interpretation of the perceptual content of the image/object

and moves toward broader or more general terminology. The computational approach used in CBIR digitizes the image/object and employs basic color physics to identify image syntax and object geometry, using the result to identify similar patterns in other images. Current CBIR research is concerned with the semantic categorizing of these patterns for subject identification, attempting to bridge the sensory gap. Text-based research, however, has yet to move past the perception of pattern similarity, resulting in a semantic gap with high-level features. Both approaches suffer from a lack of domain knowledge: CBIR does not understand user perceptions underlying conceptual semantics; and text-based approaches do not understand analysis based on visual syntax.

Current methods for automatic indexing using CBIR technologies can increase effective access to images but only by placing the onus of image description on the searcher and only if the searcher understands the dimensions of image physicality and can effectively close the semantic gap. There is the potential to enhance the precision of image retrieval if the searcher can be provided with a vocabulary for visual structure. For example, the vocabulary might describe both the color ranges of pixel groupings and their relative locations within the digital plane of the picture. The research presented here explored the possibility of developing a user-generated perceptual vocabulary that could be used by searchers and CBIR developers, to represent the visual characteristics of a two-dimensional image, incorporating both image syntax (or pre-semantic awareness of perceptual characteristics) and design characteristics (or internal contextuality). Such a vocabulary has the potential to facilitate more precise feature queries and to capture criteria for similarity judgments when CBIR technologies are involved, whether alone or in combination with high-level, concept-based representation, thereby supporting more effective image search and retrieval.

Three research questions were addressed in this experiment: Is there a shared vocabulary of terms that is used by subjects to represent the *internal contextuality* (e.g., the internal structural composition) of images? Can the natural language terms used by subjects be organized into concepts? If there is a vocabulary of concepts, is it shared across subject pairs?

For the purpose of this study, the following definitions have been adopted. An *image* is defined as an intentionally created two-dimensional artifact that stands for something else based on association and convention, but not necessarily on resemblance. The *internal contextuality* of an image refers to the shapes and the relationships between shapes that constitute the internal

spatial composition (structure) of an image. A *natural language vocabulary* consists of words or phrases, generated by an information searcher without recourse to a normalized or controlled vocabulary. A *word* or *phrase* is a natural language unit that represents a value. A *controlled vocabulary* is a set of mutually-exclusive and non-overlapping concepts (properties and values), each of which exists in a 1:1 relationship with a linguistic label. A *term* is an authorized label that may represent a set of synonyms and/or near-synonyms. A *stop word* is a word that is excluded from consideration as a preferred term due to lack of focus or specificity in the domain of the controlled vocabulary. A *property* (or *facet*) is a category that consists of a set of values grouped on the basis of similarity. A *value* (or *isolate*) is an instance of a property and may represent a single natural language term or a set of synonyms and/or near-synonyms. For example, *crimson*, *scarlet* and *rose* constitute a set of synonymous words that are represented by the *value* (term) *red* which is an instance of *color* (property). A *concept* is a category term. Within a controlled vocabulary, facets and isolates that represent more than a single natural language term are considered concepts, that is, categories of entities and/or properties. Terms are introduced into the controlled vocabulary as *conceptual antonyms* when it is necessary to make existing concepts meaningful; for example, *negation* is meaningless without *affirmation*, and *conceptual organizers* are introduced into the controlled vocabulary as superordinate terms organizing groups of related concepts.

Information systems

Much of the research on information systems reported in the scholarly literature focuses on human-computer interaction and system interfaces (e.g., Mostafa, 1994); on the searcher, the intermediary and the query process (e.g., Belkin, 1990; Kuhlthau, 1991); and on algorithmic methods for retrieving resources (e.g., with respect to image resources, Busch, 1994). Despite the integral relationship between the representation and its retrieval system, few attempts are made to link the document representation literature with the information systems design literature. Notable exceptions are Hirschheim, Klein, and Lyytinen (1995), who discuss data modeling as the extension of document representation, and Pollitt (1999), who equates relational database design with faceted classification. Perhaps one reason for the lack of such a linkage is that many text-based systems and databanks already exist and research problems involve re-design. Another reason, as suggested by Jacob and Shaw (1998), is reflected in the notion of anomalous

states of knowledge (ASK) (Belkin, Oddy, & Brooks, 1982), an approach founded on the assumption that the "traditional classification and indexing languages seem not to be designed as good representations of either [need or text], but rather as available, convenient intermediate languages" that are "bound to fail" (Belkin, 1977, p. 189).

The focus of development and research has thus been on the mechanics of the retrieval system itself, including the interface, because the representational structure has proven to be too difficult a problem. Jacob and Shaw (1998, p. 147) point out the need to (re-)consider "the problem of representation from a different perspective." Although the argument for reconsideration of representational systems was originally directed toward text-based information resources, it can be applied to the visual realm, given the dependence of retrieval systems on the "convenience" of verbal representation.

Visual images as graphic language documents were explored by Beebe and Jacob (1998) from the perspective of the document's structure as a spatially-oriented object. Drawing upon the Bauhaus concept that form as structure follows function and design principles derived from Gestalt theory, they explicated an intertwined relationship between structure and function in visual images: "The Bauhaus suggests there exists a unity of theory, material and idea that cannot be captured by merely verbalizing descriptors that would represent each of these aspects" (p. x). They concluded that geometry and representation have a similar connection and each contributes to the texture of description.

Data Collection Methodology

The current research employed a quasi-experimental equivalent materials design in order to collect words and phrases used by subjects to describe the physical structure of images. The test materials consisted of images from three domains: abstract art (ART), satellite imagery (SAT), and photo-microscopy (MIC). Independent judges selected 182 test images from an original set of 235 based on a lack of readily identifiable semantic objects or scenes.

Forty-four volunteers were recruited for participation in the study. Two eligibility requirements were clearly stated: volunteers must be native English speakers and volunteers could not have taken any college-level courses in art, architecture, photography, or geographic mapping. The subjects ranged in age from 18 to 56, with an average age of 30. The 14 men and 30 women were randomly assigned into pairs.

Vocabulary was generated using a describe-draw model derived from referential communication studies (Fussell & Krauss, 1989a, 1989b; Galantucci, 2005). In this model, a subject must formulate a message that allows a listener to identify the intended referent. Twenty-two subject pairs were asked to generate oral descriptions for 14 images: one subject would provide an oral description of an image that would allow her partner to produce a drawing of it. Subject pairs took turns performing describe-draw tasks, with seven descriptions per subject. Subjects were allowed to discuss the description and drawing after each task.

Method of Analysis

The vocabulary resulting from the describe-draw task was normalized for word variance, term identification, and conceptual organization. Frequency occurrences were tabulated for every word, every term, and every concept, and for the distribution of use across the three image domains. In order to determine if subject pairs did or did not agree on a shared vocabulary of concepts for describing images across the three domains, an interrater reliability test was used. Evaluation focused on concepts because of the specificity occurring at the term level. However, frequency counts and subject pair usage data could not be subjected to a significance test because concept distribution across domains did not form mutually exclusive categories: for every subject-pair, a concept could be used in 1, 2, or 3 domains. For this reason, the results were assessed on the basis of overall consensus and consistency in concept usage.

Frequency scores and subject pair distribution can identify concepts with the largest usage scores as well as concepts used by all subject pairs, but cannot indicate if a concept was used in all domains. To identify a shared vocabulary across subject pairs, it was necessary to determine if subject pairs agreed on concepts for usage in all three domains. This is a problem of an interrater reliability. Interrater reliability refers to the degree of agreement between judges asked to rate or score something, such as behavior, performance or open-response items on tests.

In this research, each subject pair was treated as a judge generating natural language descriptions in an “open-response” situation. These natural language descriptions consisted of terms (and concepts), each of which could be interpreted as a rating. Since each image has a pre-determined domain designation, each concept can be identified as being used in a specific domain when used to describe a given image. For each concept used (rated) by a subject pair (judge), a shared-ness score was calculated to reflect its use across image domains. The shared-

ness scale used in this analysis is a scoring rubric similar to that used when judges are knowingly performing a rating task. It represents the distribution of concept use across image domains by a subject pair during their whole describe-draw task session. The scale was constructed as a continuum extending from 0, indicating that the concept was not used by the subject pair, to 3, indicating that the concept was used by the subject pair to describe images in all three domains.

Using this scale for concept shared-ness across domains, higher scores indicate a greater degree of cross-domain use for a concept. While the concepts shared across all three domains could be identified through analysis of frequency data, the shared-ness scale takes into account those concepts which were used in varying levels across subject pairs (e.g., a concept that was not used by all pairs but was used in all domains by one or more subject pairs).

To determine the existence of a shared vocabulary of concepts, the Spearman's rank correlation coefficient for interrater reliability was computed to determine consistency across subject pairs and is reported as the mean correlation of subject pairs agreeing on use of a concept across domains. Although most discussions of interrater reliability focus on consensus (e.g., Cohen, 1988), Stemler (2004) 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." In this study, the question was whether subjects (judges), who are not necessarily aware of an image's domain, are assigning concepts (rating scale) to images based on a shared understanding of concepts. Agreement across subject pairs would indicate the presence of a shared vocabulary of concepts that was applicable across image domains.

Processing the Data

Processing the verbal description data involved three major steps: transcribing, faceting, and tabulating. Transcription of the audiotapes involved iterative listening and rule-making for transcription consistency across descriptions. For example, partial words and utterances were not transcribed, but all repetitions were included, and fractions and numbers were transcribed as numerals.

To build the faceted controlled vocabulary, transcripts for each subject were analyzed for term identification (Batty, 1989); and these terms were used for tabulation of term frequencies as

established by research on category norms (Battig & Montague, 1969; Hunt & Hodge, 1971). The process of building the controlled vocabulary and facet creation had three iterative phases: identification of stop words, syntactic normalization, and semantic normalization. The concept of stop words was applied but with a slight variation. The frequency of some words which carry little description of image physicality, such as pronouns, justified their elimination from the tabular analysis of vocabulary. Other traditional stop words, such as the numbers one through ten, proved to be meaningful in the representation of internal contextuality and were therefore retained in the word list. Non-traditional stop words included draw commands, describe-draw process dialog, indexical pointers, and verbal placeholders (e.g., self-dialog). Syntactic normalization involved basic concept identification and the grammatical analysis of variant forms of words with the same root in order to develop stemming guidelines. For example, rectangle, rectangular, and rectangularish were all reduced to the term rectangle by the guideline that noun forms take precedence. Semantic normalization followed guidelines for faceted thesaurus construction described by Batty (1989). The researcher identified synonym sets (synonyms and near-synonyms), determined an authorized value (isolate) to represent the set, grouped the values into concepts (facets), and created a hierarchical structure for each facet. For example, the synonyms beige, sandy, and tan have the value brown in the <PROPERTY> facet of <Color>, and brown appears after orange in the hierarchical structure.

As words were evaluated, first by normalizing variant forms, then by normalizing semantic referents and situating the resulting terms within the hierarchical structure of the faceted vocabulary, frequency of occurrence was tallied for each word, for each normalized form, for each grouping of synonymous and near-synonymous terms, and, finally, for each concept represented by an isolate or facet label. The units of analysis, therefore, are the terms and, more importantly, the superordinate concepts under which terms are grouped in the faceted vocabulary. Each term or concept had the potential to be used by 22 pairs of subjects in each of the three image domains. However, due to malfunctioning of the audiotape recorder, one subject pair was eliminated from analysis, leaving data for 21 subject pairs.

Results

Transcriptions of the subject descriptions generated a total of approximately 107,581 natural language words. An exact count of the natural language words was not maintained because the

development of processing guidelines was iterative during normalization, (e.g., the rule that utterances and partial words were not to be counted).). A set of 221 unique stop words accounted for approximately 50% of the total words (55,952) generated by subjects, leaving a total count of 51,629 natural language word occurrences. Once these natural language words had been normalized and variant forms eliminated, 2,075 unique words remained. After these 2,075 words had been semantically normalized by collapsing synonyms and near-synonyms, 1,319 unique terms remained. This final list of terms included 225 superordinate conceptual organizers and conceptual antonyms that were introduced during construction of the faceted vocabulary.

Totals for frequency of occurrence of the 1,319 terms ranged from zero to 3,695 per term, with zero frequency of occurrence indicating superordinates and antonyms introduced into the faceted vocabulary. Descriptions by individual subjects ranged from a total of 12 to 826 words per description, with a median of 155 words per description.

Terms with the highest frequency and pair count

Pair count was used as a criterion for selecting terms with the highest frequency counts. There are 60 terms that had high frequency counts and occurred in the descriptions of 20 to 21 (20/21) subject pairs (reported in Table 1). In limiting Table 1 to terms with a pair count of 20/21, the assumption was made that any difference between usage by 20 versus 21 pairs of subjects was due to the specificity of the image, subject individuality, or simple chance. The high frequency terms listed in Table 1 account for all the terms that occurred in 20/21 pairs, but they do not necessarily represent the highest frequency counts. For example, the term inch had a high frequency count (544) but it only occurred in the descriptions of 17 pairs. However, it can be assumed that the 60 terms used by 20/21 subject pairs indicate the potential for a shared vocabulary.

The high frequency terms with 20/21 pair occurrences were investigated for distribution across the three image domains. Table 2 indicates that there is general distribution of term occurrence across the three domains, with the ART domain having a slightly higher percent of total terms used. When considering the number of unique terms used in each domain, the distribution is fairly even. However, it must be remembered that the totals reported for unique terms by domain represent a simple count of terms used out of the total 1,319 unique terms possible and do not imply use of the same terms across domains.

ISSN: 2324-9773

There is considerable domain overlap in the 60 high frequency terms listed in Table 1. This is demonstrated in the domain frequency and rank columns in Table 1. However, each of the domains has high frequency terms that are not included here because they were not used by 20/21 pairs. For example: cell is a high frequency term in MIC; road, and curve are high frequency terms shared by ART and SAT; and section, 1/3, centimeter, and horizontal are high frequency terms in ART.

Collapsing to concepts

In order to determine the potential of a shared vocabulary of concepts, each isolate was collapsed into its superordinate facet. Collapsing was based on the assumption that isolates were recognizable as instances under the superordinate concept category. For example, blue is readily identifiable as an instance of the facet <Hue> and asterisk as an instance of the facet <Punctuation>. Thus, the frequency count for a superordinate facet is the sum of the frequency counts for all isolate terms nested under the superordinate plus the frequency count associated with the superordinate itself. Pair count for a superordinate facet is the union of the pair counts for all isolate terms and the pair count of the superordinate. There were 80 superordinate facets that had no subordinate isolates due to introduction during the construction of the faceted vocabulary. They retained the original pair value and frequency count of 0 and were removed from calculations, leaving a set of 465 concepts for evaluation.

The hierarchical structure of the faceted vocabulary provides a framework for evaluating the concept categories of terms generated by subject pairs. The top three levels of the hierarchy are presented in Table 5. The top level concepts are the four facets <OBJECT>, <PLACE>, <PROPERTY>, and <SPATIAL-LOCATION>. Table 3 shows the number of concepts used in each of the top level facets and their distribution by domain. Table 4 shows the frequency distributions of concepts across top level facets within image domains.

It is possible to compute concept frequencies at progressively subordinate levels of the faceted hierarchy, but discerning valuable information from such computations is problematic. At the second and third levels in the faceted hierarchy, there are 14 and 47 subordinate concepts respectively (see Table 5). However, these concepts are unevenly distributed across the top level facets and, at this point, are only of interest because they indicate the variety of unique terms used in each concept category. Counts between concepts cannot be computationally compared

ISSN: 2324-9773

between concepts at this level because of their uneven distribution; but they do provide insight into the distribution of concepts across subordinate levels in the hierarchy. More detailed data can be computed for each subsequent subordinate level, but such computations would only be of use in analyzing individual top level facets and could not be used for comparison across facets because of the hierarchical structure within each top level facet. Such analyses would be of primary interest for determining where the emphasis should be placed in the development of individual retrieval vocabularies for particular domains or user groups.

Shared concepts: Shared-ness rating

The purpose of the research was to assess the extent to which subjects used the same concepts to describe images from more than one domain. Evaluation of the domain and frequency 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. In order to determine if the concepts generated by subject pairs were used across domains, a rating scale was devised based on the actual use of each concept by each subject pair. This scale is referred to as the *shared-ness rating*. The shared-ness rating scale is a continuum from 0 to 3: 0 indicates that a given concept was not used at all by a subject pair; 1 indicates the concept was used in only one domain by a subject pair; 2 indicates usage in two domains; and 3 indicates usage in all three domains. This shared-ness rating captures the breadth of each concept's use by each subject pair.

Shared-ness ratings provide a general means for measuring subject agreement on concept use. For each concept, a shared-ness measure was computed by summing shared-ness ratings across all 21 subject pairs. For example, the concept <Triangle> was used by five subject pairs across all three domains, by ten subject pairs in two domains, by five subject pairs in one domain, and by one subject pair in no domain (i.e., the concept was not used by any pair). The resulting shared-ness measure for <Triangle> is 40 (i.e., 15 + 20 + 5 + 0 = 40). Concepts with different degrees of shared-ness can be arranged in a rank order according to shared-ness measure (i.e., from 0 to 63). The rank ordering of concepts by shared-ness measure (see Table 6) is important because it demonstrates which concepts were more broadly shared across subject pairs. The higher the shared-ness measure of a concept, the more likely it is that a concept is part of a shared vocabulary for describing the internal contextuality of images.

ISSN: 2324-9773

Determining if subject pairs agree on concepts used in all three domains is a problem of interrater reliability. Shared-ness ratings for concepts (i.e., 0 to 3) represent the distribution of concept use across image domains during the entire describe-draw session for a single subject pair and thus provides a scoring rubric similar to that used when judges are knowingly performing a rating task. Although the Pearson correlation coefficient is commonly used, it requires that data be normally distributed, which is not the case here since many concepts were used by on ly one subject pair. Thus the most appropriate indicator of interrater reliability that can be computed for this data would be the Spearman rank correlation coefficient (see Stemler, 2004).

A Spearman rank correlation coefficient was computed for each combination of two subject pairs. This produced a matrix of 210 correlations, one for each possible combination of the 21 subject pairs. Each correlation indicates the extent to which two subject pairs agreed on the overall use of concepts across domains. All correlations achieved statistical significance at the .001 level, indicating that subject pairs do tend to agree on the use of certain concepts across multiple domains. Correlations for concepts with high shared-ness measures, as reported in Table 6, indicate that most subject pairs used these concepts to describe multiple domains, thus pointing toward a shared vocabulary. The 14 concepts with shared-ness sums of 62 or 63 would form the heart of this shared vocabulary.

Future work

Having established the viability of a shared vocabulary of concepts for describing the internal contextuality of images, this vocabulary can be used to inform future research in the areas of image vocabulary development, identification of operators for image searching, construction of CBIR metrics for similarity judgments, and the design of interfaces for image retrieval systems.

The data collected in this project needs a more detailed analysis of the concepts that did not fall at the extremes of frequency counts and subject pair usage. If a concept were not in the high range of frequency use (i.e., 100 or more) or shared-ness rating (i.e., 2 or 3), then what was its importance in the natural language descriptions that generated the concept vocabulary? More sophisticated statistical methods need to be applied for this type of analysis, perhaps taking advantage of frequency rankings or domain usage variation. Many of the terms with mid-range frequencies may have been used because of the nature of the image being described: for

example, crystal did not emerge as a shared term but was used consistently to describe a single MIC image. This points to additional questions: Do individual pictures evoke the same terms from subjects? Does a particular image domain produce more descriptive words? And what level of frequency indicates that a term is domain specific?

This research has the potential to inform CBIR developers and interface designers about user-generated vocabulary at both the levels of term and concept. Adopting a controlled vocabulary will lessen the semantic gap through use of standardized vocabulary to inform threshold settings for interface choices, CBIR similarity metrics, and relevance feedback. Identification of those concepts with high shared-ness ratings could inform CBIR researchers regarding image attributes that are prominent from the user's perspective and could be used to develop new image differentiation metrics. The faceted vocabulary itself offers an organizational structure that could facilitate the combination of CBIR research agendas through the coordination or differentiation of various attribute concepts. Furthermore, it suggests attributes that could be pursued for metric evaluation, such as the distinction between approximate and exact or operationalization of <SPATIAL-LOCATION> concepts.

Analysis of term frequency counts suggests that visual search operators should be explored for their potential to express relationships such as those represented by the terms and concepts nested within the facets <Gestalt> and <SPATIAL-LOCATION>. Using the high frequency shared concepts subsumed by <Gestalt> and <SPATIAL-LOCATION>, operators could be developed that would allow searchers to define relationships between elements when describing the internal contextuality of the desired image: A size and proportion tool could provide samples from which to select, an exact-approximate attribute could be offered; and a shape catalogue or visual vocabulary could provide both geometric and analogical shapes.

Developing a visual vocabulary of shapes and relationships would be an important application of the controlled vocabulary that emerged from this study. Although one objective of this study was to identify a vocabulary for representing the internal contextuality of an image, language does not always work well for describing images. However, the faceted vocabulary could serve as the basis for a visual vocabulary of shapes and relationships that would capture the visual implications of searchers' analogical use of terms.

References

Batty, D. (1989). Thesaurus construction and maintenance: A survival kit. *Database*, 13-20.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80*(3, Part 2), 1-45.

Beebe, C., & Jacob, E. K. (1998). Graphic language documents: Structures and functions. In W. M. el-Hadi, J. Maniez & A. S. Pollitt (Eds.), *Structures and relations in knowledge organization, proceedings 5th Int. ISKO Conference* (Vol. 6, pp. 244-255). Lille, France: Ergon Verlag.

Belkin, N. J. (1977). *The problem of 'matching' in information retrieval.* Paper presented at the Second International Research Forum in Information Science, London

Belkin, N. J. (1990). The cognitive viewpoint in information science. *Journal of Information Science, 16*, 11-15.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval Part 1. *Journal of Documentation, 38*(2), 61-71.

Besser, H., & Yamashita, R. (1998). *The cost of digital image distribution: The social and economic implications of the production, distribution and usage of image data*: Andrew W. Mellon Foundation.

Busch, J. A. (1994). How to choose among alternative technologies for physical access to art images. *Computers and the History of Art, 4*(2), 3-16.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ.: Erlbaum.

Datta, R., Li, J., & Wang, J. Z. (2005). *Content-based image retrieval - approaches and trends of the new age.* Paper presented at the MIR '05, Singapore.

Eakins, J. P., & Graham, M. E. (1999). *Content-based image retrieval: A report to the JISC Technology Applications Programme*. Newcastle: Institute for Image Data Research, University of Northumbria.

Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology, 25*, 203-219.

Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology, 19*, 509-525.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science, 29*, 737-767.

Goodrum, A. A., Rorvig, M. E., Jeong, K.-T., & Suresh, C. (2001). An open source agenda for research linking text and image content features. *Journal of the American Society for Information Science and Technology, 52*(11), 948-953.

Hirschheim, R., Klein, H. K., & Lyytinen, K. (1995). *Information systems development and data modeling: Conceptual and philosophical foundations*. Cambridge: Cambridge University Press.

Hunt, K. P., & Hodge, M. H. (1971). Category-item Frequency and Category-name meaningfulness (m'): Taxonomic Norms for 84 Categories. *Psychonomic Monograph Supplements, 4*(6), 97-121.

Jacob, E. K., & Shaw, D. (1998). Sociocognitive perspectives on representation. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 33, pp. 131-185). Medford, NJ: Information Today for the American Society for Information Science.

Kuhlthau, C. C. (1991). A process approach to library skills instruction. In *Information literacy: Learning how to learn* (pp. 35-40). Chicago: American Library Association.

Mostafa, J. (1994). Digital image representation and access. In M. E. Williams (Ed.), *Annual review of information science and technology (ARIST)* (Vol. 29, pp. 91-135). Medford, NJ: American Society for Information Science.

Pollitt, A. S. (1999). *Interactive information retrieval based on faceted classification using views*. Retrieved September 7, 1999, from www.hud.ac.uk/schools/cedar/dorking.htm

Rogowitz, B. E., Frese, T., Smith, J. R., Bouman, C. A., & Kalin, E. (1998, January 26-29). *Perceptual image similarity experiments.* Paper presented at the Human Vision and Electronic Imaging III, San Jose, CA

Santini, S., & Jain, R. (1998). *Beyond query by example.* Paper presented at the IEEE Workshop on Multimedia Signal Processing, Los Angeles, CA.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), -1-32.

Stemler, S. E. (2004). *A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability*. Retrieved June 15, 2006, from http://pareonline.net/getvn.asp?v=9&n=4

Stenvert, R. (1992). Bridging the gap between pixel and picture. *Computers and the History of Art, 2*(2), 19-24.

Table 1: *Terms with high frequency counts that were used by 20/21 subject pairs, including frequency count and rank by domain.*

Pair Total	Total Freq. Count	Freq. Rank	Terms	ART Freq. Count	MIC Freq. Count	SAT Freq. Count	ART Freq. Rank	MIC Freq. Rank	SAT Freq. Rank
21	2670	1	similar	838	858	974	1	1	1
21	2174	2	a	836	485	803	2	3	2
21	1763	3	<Inside-of>	691	550	522	3	2	4
21	1524	4	left	638	349	537	5	5	3
21	1458	5	right	605	382	471	6	4	5
20	1286	6	<Line>	676	263	347	4	12	9
21	1152	7	top	455	313	384	9	8	6
21	1078	8	bottom	431	285	362	11	9	7
20	1066	9	<Approximate>	537	176	353	7	22	8
21	941	10	side	448	215	278	10	13	16
21	937	11	center	316	315	306	16	6	12
21	913	12	<On>	357	264	292	13	11	14
21	877	13	small	270	284	323	21	10	10
21	860	14	<Rectangle>	501	201	158	8	16	23
21	853	15	down	379	171	303	12	23	13
20	825	16	up	318	189	318	15	19	11
21	701	17	negation	294	184	223	17	20	17
21	690	18	corner	196	215	279	23	14	15
21	686	19	<Shape>	333	203	150	14	15	24
21	682	20	circle	164	314	204	30	7	18
21	637	21	edge	276	167	194	19	24	20
21	631	22	2	290	191	150	18	18	25
21	581	23	very	185	200	196	25	17	19
20	537	24	1/2	276	95	166	20	29	21
21	521	25	all	176	181	164	28	21	22
21	462	26	large	181	147	134	27	25	28
21	442	27	<Dot>	244	91	131	22	30	29
21	374	28	start	129	97	148	37	28	26
21	367	29	square	193	58	116	24	45	34
21	355	30	<Outside-of>	144	89	122	33	32	33

Note: Terms that are also concepts are indicated by < >. Table 1 continued on next page.

ISSN: 2324-9773

Table 1 continued.

Pair Total	Total Freq. Count	Freq. Rank	Terms	ART Freq. Count	MIC Freq. Count	SAT Freq. Count	ART Freq. Rank	MIC Freq. Rank	SAT Freq. Rank
21	354	31	<Part>	139	78	137	35	38	27
21	347	32	<Joined>	172	81	94	29	36	42
20	345	33	<Triangle>	182	66	97	26	41	39
20	325	34	almost	105	91	129	41	31	30
21	309	35	more	111	101	97	40	27	37
21	307	36	end	119	62	126	38	44	31
21	305	37	straight	154	46	105	31	50	35
21	285	38	<Exact>	78	82	125	45	35	32
21	285	39	<Cross>	138	50	97	36	48	40
21	278	40	1	142	69	74	34	40	45
21	271	41	<Off>	95	79	97	42	37	38
20	242	42	long	149	65	28	32	43	59
20	224	43	same	86	83	55	44	34	51
21	218	44	<Surrounding>	76	37	105	46	55	36
21	216	45	3	114	56	46	39	46	53
21	214	46	many	56	88	70	52	33	47
21	198	47	some	66	73	59	48	39	48
20	188	48	<Beside>	67	47	74	47	49	46
20	185	49	squiggle	35	113	37	55	26	55
20	163	50	below	87	46	30	43	51	57
20	151	51	<Between>	61	56	34	49	47	56
21	144	52	different	60	38	46	50	53	54
20	140	53	<Color>	46	66	28	53	42	58
21	132	54	<Whole>	42	38	52	54	54	52
20	130	55	<Area>	33	19	78	57	58	44
20	129	56	above	34	39	56	56	52	50
20	115	57	4	59	28	28	51	57	60
20	104	58	tiny	16	29	59	58	56	49
20	100	59	<Land>	1	1	98	59	59	41
20	89	60	aerial-view	1	0	88	60	60	43

Table 2. *Distribution of term frequencies across image domains.*

	ART	MIC	SAT	Total terms
Total term frequency	20,618	13,694	17,317	51,629
% of Total term frequency	40%	27%	34%	
Total unique terms	684	624	622	1,319
% of Total unique terms	63%	57%	57%	

Note. Total counts for unique terms do not equal the sum of unique terms for the three domains because the same term may be used in multiple domains.

Table 3. Concepts used in the four top level facets and their distribution across image domains.

Concepts	Total unique concepts (465)	Concepts used for ART	Concepts used for MIC	Concepts used for SAT
<OBJECT>	227	165	152	110
<PLACE>	90	40	38	79
<PROPERTY>	108	100	93	90
<SPATIAL-LOCATION>	40	40	39	39

Note. Total counts for unique concepts in each of the four top level facets do not equal the sum of the three domains because the same concept may be used in multiple domains.

Table 4. *Frequency distribution of concepts across top level facets and their distribution across image domains.*

	Total frequency	ART	MIC	SAT
<OBJECT>	9826	4328	2850	2648
% Total Freq	19%	8.50%	5.50%	5%
<PLACE>	2092	266	287	1539
% Total Freq	4%	0.50%	0.50%	3%
<PROPERTY>	22832	9090	6159	7583
% Total Freq	44%	17.50%	12%	14.50%
<SPATIAL-LOCATION>	16879	6933	4399	5547
% Total Freq	33%	13.50%	8.50%	11%
Total freq. for all concepts	51629	20618	13694	17317
% Total Freq	100%	40%	26.50%	33.50%

ISSN: 2324-9773

Table 5. *Concepts in the top three levels of the hierarchy of the faceted structure.*

Concepts: Hierarchy Level 1	Concepts: Hierarchy Level 2	Concepts: Hierarchy Level 3
<OBJECT> (601)		
	<*Image*> (17)	
		<*Kind-of-image*> (10)
		<*Image-foundation*> (7)
	<*Non-living-thing*> (353)	
		<Figure> (105)
		<*Artifact*> (178)
		<*Mechanical-part*> (5)
		<Substance> (8)
		<*Naturally-occurring-phenomena*> (57)
	<Living-organism> (233)	
		<Animal-life> (55)
		<Plant> (33)
		<Body> (117)
		<*Aspects-of-living-thing*> (27)
<PLACE> (224)		
	<*Constructed-environment*> (130)	
		<*Water-based-environment*> (9)
		<*Land-based-environment*> (109)
		<*Locale*> (11)
	<*Natural-place*> (62)	
		<Sky> (2)
		<*Body-of-water*> (17)
		<Shore> (4)
		<*Land-water-formation*> (6)
		<Terrain> (20)
		<*Ecosystem*> (4)
	<*Sociopolitical-location*> (26)	
		<Continent> (2)
		<Country> (9)
		<State> (4)
		<*Municipality*> (9)
	<*Generic-place*> (14)	
		<Area> (5)
		<*Opening*> (4)
		<Joint> (4)

Note. Counts for individual concepts are indicated in parentheses following each concept and do not include counts for any subordinate concepts. Concept labels introduced by the researcher are indicated by *italics*. Table 5 continued on next page.

Table 5
continued.

Concepts: Hierarchy Level 1	Concepts: Hierarchy Level 2	Concepts: Hierarchy Level 3
<PROPERTY> (410)		
	<*General-concept*> (16)	
		<*Existence*> (3)
		<Domain> (9)
		<*Validation*> (3)
	<*Attribute*> (391)	
		<*Visibility*> (5)
		<*Gestalt*> (67)
		<Quantity> (79)
		<Comparison> (32)
		<*Condition*> (45)
		<*Judgment*> (30)
		<*Change-in-condition*> (14)
		<Action> (22)
		<*Art-and-craft-process*> (66)
		<Color> (38)
<SPATIAL-LOCATION> (85)		
	<*Format*> (3)	
	<Position> (65)	
		<*Indexical*> (23)
		<*Relational*> (42)
	<Direction> (10)	
		<Pointing-to> (1)
		<Vertical> (3)
		<Horizontal> (3)
		<Diagonal> (1)
		<Perpendicular-to> (1)
	<Compass-orientation> (9)	
	<Clock-orientation> (1)	

Table 6. *Concepts with highest sum of the domain ratings.*

	Pair	Freq.	Shared-ness rating				
Concept	**Count**	**Total**	**0**	**1**	**2**	**3**	**Sum**
<Similarity	21	3059	0	0	0	21	63
<*Linguistic-quantity*>	21	2174	0	0	0	21	63
<Vertical>	21	1823	0	0	0	21	63
<Inside-of>	21	1773	0	0	0	21	63
<Size>	21	1558	0	0	0	21	63
<Degree>	21	1157	0	0	0	21	63
<Horizontal>	21	3200	0	0	1	20	62
<*General-part*>	21	3167	0	0	1	20	62
<Line>	21	1647	0	0	1	20	62
<*Extremity*>	21	1345	0	0	1	20	62
<Number>	21	1021	0	0	1	20	62
<*Presentation*>	21	980	0	0	1	20	62
<On>	21	913	0	0	1	20	62
<Angle>	21	861	0	0	1	20	62
<Rectangle>	21	1227	0	0	2	19	61
<*Certitude*>	21	1361	0	0	3	18	60
<*Validation*>	21	701	0	1	2	18	59
<Shape>	21	686	0	0	6	15	57
<*Fraction*>	20	1053	1	1	2	17	56
<Outside-of>	21	355	0	0	7	14	56
<*Width*>	21	306	0	0	7	14	56
<*Vertical-perspective*>	21	292	0	1	5	15	56
<Piece>	21	315	0	0	8	13	55
<*Unequal*>	21	336	0	2	5	14	54
<*Extension*>	21	317	0	3	3	15	54
<Beside>	20	276	1	1	5	14	53
<Dot>	21	426	0	2	6	13	53
<Part>	21	354	0	1	8	12	53
<Length>	21	335	0	3	4	14	53
<*Hue*>	21	706	0	1	9	11	52
<Joined>	21	359	0	2	7	12	52
<Distance>	19	232	2	0	5	14	52
<*Rotated*>	19	283	2	2	3	14	50

Note. For each concept a shared-ness measure (Sum) has been computed by summing shared-ness ratings across all subject pairs.