
Image Retrieval as Linguistic and Nonlinguistic Visual Model Matching

P. BRYAN HEIDORN

ABSTRACT

THIS ARTICLE REVIEWS RESEARCH ON HOW people use mental models of images in an information retrieval environment. An understanding of these cognitive processes can aid a researcher in designing new systems and help librarians select systems that best serve their patrons. There are traditionally two main approaches to image indexing: concept-based and content-based (Rasmussen, 1997). The concept-based approach is used in many production library systems, while the content-based approach is dominant in research and in some newer systems. In the past, content-based indexing supported the identification of "low-level" features in an image. These features frequently do not require verbal labels. In many cases, current computer technology can create these indexes. Concept-based indexing, on the other hand, is a primarily verbal and abstract identification of "high-level" concepts in an image. This type of indexing requires the recognition of meaning and is primarily performed by humans. Most production-level library systems rely on concept-based indexing using keywords. Manual keyword indexing is, however, expensive and introduces problems with consistency. Recent advances have made some content-based indexing practical. In addition, some researchers are working on machine vision and pattern recognition techniques that blur the line between concept-based and content-based indexing. It is now possible to produce computer systems that allow users to search simultaneously on aspects of both concept-based and content-based indexes. The intelli-

P. Bryan Heidorn, Graduate School of Library and Information Science, University of Illinois, 501 E. Daniel, Champaign, IL 61820

LIBRARY TRENDS, Vol. 48, No. 2, Fall 1999, pp. 303-325

© 1999 The Board of Trustees, University of Illinois

gent application of this technology requires an understanding of the user's visual mental models of images and cognitive behavior.

INTRODUCTION

To better understand the relationship between concept-based and content-based indexing in a volume such as this, it is useful to refocus and re-evaluate image indexing. An understanding of these techniques may be unified by examining how each relates to "visual mental models." From this perspective, image retrieval system work is an endeavor to create a concordance between an abstract indexing model of visual information and a person's mental model of the same information. All visual information retrieval research, from the computational complexity of edge detectors to national standards for museum indexing of graphical material, is an attempt to bring the indexing model and the user's mental model into line. All index abstraction, nonlinguistic or linguistic, may be classified by their success in matching the user's abilities. Borgman (1986) emphasizes that retrieval systems should be designed around "natural" human thinking processes. Index facet effectiveness is more dependent on the facets' harmonization of the facets with human cognition than on whether it is linguistic (concept-based) or nonlinguistic (content-based).

In describing the content of images in the realm of art, Panofsky (1955) distinguishes between pre-iconography, iconography, and iconology. Pre-iconographic content refers to the nonsymbolic or factual subject matter of an image. It includes the generic actions, entities, and entity attributes in an image. As an example, a pre-iconographic index may indicate that an image contains a stone (attribute), bridge (entity), and a river (entity). Iconographic content identifies individual or specific entities or actions. In the example, the bridge might be identified as the "Palmer Bridge" and the "Hudson River." The iconologic index would include the symbolic meaning of an image. The image might be indexed as "peaceful" or symbolizing "simpler times." The indexing that is appropriate depends on the type of subject matter that the searchers will eventually have in mind when they are doing a search.

This type of subject classification can be used to explain the strengths and weaknesses of content-based and concept-based indexing. Computers frequently perform content-based indexing. Computers can cost-effectively identify image attributes such as color, texture, and layout. Historically, limitations in computer algorithms have limited computer indexing to just a fraction of the pre-iconographics content. This, however, is changing, and the challenge for researchers and developers is to expand the functionality of the systems. Within limited contexts, computer indexing has been able to move into iconographic subject matter. For example, by exploiting information in picture captions in newspapers, a system may identify individuals in an image (Srihari, 1995). Other sys-

tems can identify and index objects such as trees or horses using low-level features such as texture and symmetry (Forsyth et al., 1996). Linguistic content-based indexing has traditionally been performed by humans. While it is expensive and time consuming, it is possible to create indexes for all three types of content matter described by Panofsky. Hastings (1995) demonstrated that, in some retrieval situations, searchers use a combination of both visual and verbal features. With current technology, this means the use of both content-based and concept-based techniques.

This article will focus on pre-iconographic indexing since this is the main area where content-based and concept-based techniques overlap. Content-based techniques may be used effectively where the computer can extract and synthesize features, attributes, and entities in images that are consistent with human understanding of the images. The computer must model the image in a way that is isomorphic (but not identical) to the human model of the image. Human indexers and searchers must also shape representations or mental models of the images if the indexer is to produce a functional index. In order to demonstrate the importance and pervasiveness of this process, this article will explore two aspects of indexing: color and object naming (shape). The first section will discuss the cognitive and social processes that give rise to the visual mental models that are shared by indexers and searchers. The next section explains what is meant by mental models in this context. Following this is a discussion of the representation of objects and shapes in visual mental models and then how both content-based and concept-based indexes capture (or neglect) aspects of these models. This is followed by a discussion of color in mental models and then discussion of the approaches to concept-based and content-based indexing by color.

IMAGE ACCESS AS A SOCIOCOGNITIVE PROCESS

Imagine an image of a bridge at sunset on a winter day. What color is the sky? Is there a name for the color? What objects are in the image? Are they important? Is the sun visible or has it already descended below the horizon? If you wanted to store this image with 100,000 others, how would you find it again? How would you describe it so that someone else could find it? Would words be enough? The answer to all of these questions depends on personal history and cultural expectations.

The act of indexing and accessing images from a database is a sociocognitive process grounded in both biology and experience. The term "sociocognitive" here means a combination of the social aspects of cognition as well as the individual aspects of mental life. Cognition refers to all processes involved in the perception, transformation, storage, retrieval, manipulation, and use of information by people. Of particular interest here will be those aspects of cognition that are called mental models. In a social context, we often wish to communicate our thoughts to

others. We frequently do this with language but also through our postures, gestures, or hand drawn illustrations or, for the gifted, through works of art. Communication between people is an act of one person referencing and changing the representations used in the cognition of another person, what they are thinking about, and even how they are thinking. In this context, indexing is a form of communication between the indexer and the people who will search for images in a collection. The indexer must rely on both shared cognitive heritage and social conventions to represent salient aspects of an image in the indexing scheme. The searchers, in using the index, must express their interests in the same language that was used by the indexers.

In the first paragraph of this section of the article, you were asked, through natural language, to create a "visual mental model" or "image" in your mind. Each reader's image is different, but certainly there are aspects of the image that are shared among readers. Some of these aspects may be based on the shared biology of our vision systems (most of us can imagine color), and some shared aspects may be attributable to our shared experience. We all know what bridges are without having been born with that knowledge. Some aspects of the visual mental model are easily described with natural language or verbal tags. Other aspects seem to defy simple linguistic description. "Although grammars provide devices for conveying rough topological information such as connectivity, contact, and containment, and coarse metric contrasts such as near/far or flat/globular, they are of very little help in conveying precise Euclidean relations: a picture is worth a thousand words" (Pinker & Bloom, 1995, p. 715).

This linguistic versus nonlinguistic contrast parallels concept-based and content-based indexing techniques. Understanding these mental models of images and how we can communicate information about them can enlighten us regarding content-based and concept-based indexing. Shera (1965) identified prerequisites for constructing a framework for indexing (an indexing vocabulary). These include an understanding of language and the communication process as well as an understanding of the relationship between human thought and mechanisms for recording thoughts such as language (p. 56). Indexers and system designers need to understand human cognition and communication in order to produce good indexes. The shared cognitive abilities and shared experience serve as the basis for this communication. These shared attributes may also arise from general world experience as in the earlier sunset example. Other attributes may arise from specialized training such as when an architect uses the *Art and Architecture Thesaurus* (Barnett & Petersen, 1989) to access a cultural heritage image collection or when a botanist uses the language in an identification key to label a specimen. In both cases, these cognitive attributes are learned in a social context.

In this discussion, the term “sociocognitive” is intended in its broadest sense. The social context here includes the conventions that allow indexers and searchers to learn common terminology, the natural and synthetic ontologies for image description. It is these aspects of the social environment that exist in a deep interplay with the shared cognitive abilities, biases, and frailties of the image access community. Cognitive abilities include not only a “higher” cognitive process but also the perceptual experience that is often the object of the “higher” cognitive processes. In this article, we do not focus on the social processes that indexers participate in to create indexing standards, although this is certainly important. The focus here is on the social environment that gives rise to the indexer’s thoughts about images.

Jacob and Shaw (1999) introduce a sociocognitive perspective on representation. From their perspective and the perspective of this article, language and communication influence the organization of knowledge at both the individual and social level. Social processes lead to the creation of a shared vocabulary to describe a field. However, the Jacob and Shaw treatment is primarily limited to linguistic constructs: “[R]epresentation is primarily linguistic, the development of truly effective systems of retrieval must include a thorough appreciation of how language is used in the social processes of communicating knowledge” (p. 131).

When describing images, however, content-based indexing techniques introduce nonlinguistic forms of indexing (and communication), so this sociocognitive perspective must be extended to include nonlinguistic processes (such as color and texture maps). For images, it is clear that descriptions are grounded first in the perceptual abilities of indexers and searchers.

This does not diminish the critical role of natural language in image description. The creation of a vocabulary to describe images is a Darwinian adaptation and is universal to the species. This language learning is a sociocognitive process. For example, the perception of color is physical, but the color names are arrived at through a social process. There are millions of colors that people can distinguish (Bruner, Goodnow, & Austin, 1956, p. 1) but only some are named. An information retrieval system designer must decide if a collection should be indexed using the unlabeled colors (i.e., color histograms) or using labeled category names such as “red,” “green,” or “blue.” The designer may choose to use both nonlinguistic and linguistic approaches. The decision must be made on both sociocognitive and technical grounds. In the mental image of a bridge at sunset, it might be reasonable to apply the label “red” for the sky. However, the colors in an actual sunset, or in our mental images, may defy our language skills. Figure 1, “Sunset, Palmer Bridge, New York”¹ is a digital image from the American Memory Collection at the Library of Congress (Detroit Photographic Co., c1900).² In this image, the sky’s color does

not have a name with which many people would agree. The designer must decide if the users have a word to describe the particular shade of a sunset that is needed to complement the color of a car in an automobile advertisement. Nonlinguistic, content-based color retrieval is provided in current commercial and research image database systems such as Virage (Gupta et al., 1997), VisualSEEK (Smith & Chang, 1996a), QBIC (Niblack et al., 1992; Flickner et al., 1995), and Photobook (Pentland, 1993). These include, among others, color swaths, color mixing interfaces, perceptually significant coefficients, and color similarity matching as discussed in the section on models of color.



Figure 1. "Sunset, Palmer Bridge, New York."

MENTAL MODELS OF IMAGES

When a person is searching for an image in a collection, they may be thought of as searching for images that match a mental model of the image being sought. The mental model of the target may change during the course of the retrieval session, but this does not influence the fact that there is a dynamic mental model or how the model is constructed. If the collection is small enough, the searcher may browse the images looking for one that matches the mental model. When the collection becomes too large for efficient browsing, other search strategies must be employed. In the realm of image databases, the searcher may use an index. The appropriate nature of the index is governed by the nature of the mental representation. All current indexing techniques, both manual and auto-

matic, linguistic and nonlinguistic, are attempts to make aspects of the mental representation explicit and match these aspects to the images in the collection. As depicted in Figure 2, aspects of the visual world are abstracts by the searcher and the indexer. The indexer must select aspects of the abstraction that are shared by the indexer and searcher and code them into the index so that the index itself is an abstraction of the visual world. Because of the nature of this matching process and the complexity of the visual mental models, neither concept-based nor content-based indexing alone is sufficient to support an effective retrieval system. The best aspects of these approaches to indexing need to be identified and integrated.

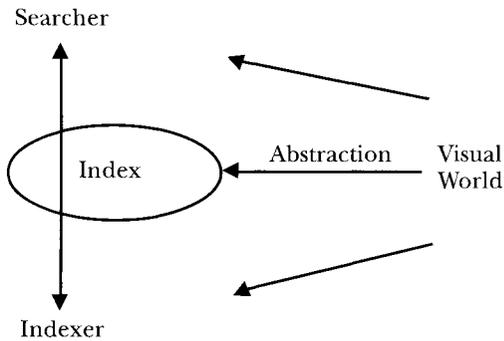


Figure 2. Index as Communication.

There are two types of correspondence that must exist between people and an image retrieval system—mental-model-to-index correspondence and cognitive-model-to-interface correspondence. The mental-model-to-index correspondence is the degree to which a particular indexing facet is in harmony with the cognitive/perceptual models and predispositions of the searcher. The cognitive-model-to-interface correspondence is the degree of agreement between the searcher's cognitive/perceptual models and the ability to express these in the interface. This applies not only to the representation of the index in the interface but also to the user's expectations and mental models about how interfaces work (Borgman, 1986).

It is important, then, to consider the nature of the visual mental representations and their relationship to the physical world. Mental models of images represent, at least, perceptible aspects of the world that they represent (Johnson-Laird, 1983, p. 157). For the purposes of this analysis, it does not matter whether the representation is like an image in one's mind (Kosslyn, 1980; Paivio, 1971), is propositional (Pylyshyn, 1973; Palmer, 1975), or both. In either case, these models are abstractions of the visual

world and are not actual images since this would require the existence of a homunculus to observe these models.

These mental models possess some isomorphic relationship to the visual world. When people imagine a bridge at sunset, they are constructing an active mental model in working memory out of long-term memory traces. The processes involved in perception determine the contents of a long-term model. That is, the model of an image begins with its perception. The stages of processing from the outside world to long-term memory include sensory detection, pattern recognition, short-term memory, and long-term memory. In the visual system, sensory detection is the conversion of light into nerve impulses. Only light of very particular wavelengths can be detected but, as discussed later in this section, these impulses can serve as the basis for distinguishing millions of colors. Long-term mental models may contain representations of these colors, and people may wish to search image collections based on them. Content-based indexing methods for color representation support this spectrum-like aspect of the mental model.

The next stage of perception is pattern recognition. Our visual systems are trained from birth to recognize patterns in our environment. We have physical apparatus and training which allows us to detect edges, surfaces, depth, motion, and other aspects of the environment. This recognition is sometimes associated with the linguistic label for the pattern, but linguistic labels are not necessary. So we may recognize a particular pattern as being a cat and apply that label (bringing with it an association to a "cat" category in memory). We can also recognize objects for which we have no name. For example, in a zoo or in a forest, we may see an animal that we have never seen before. The fact that we have no name for it does not mean that we do not recognize it and remember it. In fact, this type of pattern recognition is the basis for a significant application of image databases. It is possible to identify animals, plants, or archeological objects by finding like objects in an image collection. Concept-based indexing techniques may be used where an object or pattern is named. Content-based techniques may be used where no name is available for at least some of the database searchers. Most thesauri for graphical materials, such as the *Art and Architecture Thesaurus (AAT)* (1994) and the Library of Congress' *Thesaurus for Graphic Materials* (1995), are examples of the concept-based approach. In these resources, all objects and patterns have labels.

The next stage in visual processing is short-term memory or working memory. The human memory system is frequently conceptualized as having two components: short-term memory and long-term memory. Two main properties differentiate the storage mechanisms. Short-term memory is limited in both size and duration. It is the mechanism used to remember information that may be forgotten immediately after use.

This might include a phone number or a URL. In some situations, short-term memory is better named "working memory." It includes the mechanisms that allow us to manipulate mental representations including mental images (as discussed in the next paragraph). Short-term memory is the procedure used to combine information from a visual scene with long-term memories. Long-term memory does not have either the duration or size limits of the short-term memory. Long-term memory is, however, very susceptible to distortion. One particular memory of an event can easily "mix" with prior memories and expectations. From the information processing perspective, memories in long-term memory must be moved to working memory before one is able to act on the memory. During an image retrieval task, the searcher will form a mental model of the target image in working memory. This model will be dynamic. Information from sensory input and from long-term memory can move into working memory. The sensory input might alter details of the model as near misses are encountered or as the user interface suggests options. Likewise, details of a scene may be filled in from long-term memory as the need arises.

People activate visual mental models or construct them from memory and then use them as a basis for comparison of images in a database. In some situations, these models behave as if they were three-dimensional representations very close to those used in perception. There are retrieval mechanisms that exploit the image-like qualities of images. These mechanisms allow the use of image qualities directly without the intervention of linguistic labels. These include color-wheels, color spaces, texture menus, sketching shapes by hand, by example-based searching, and other techniques. Indexing techniques sometimes treat images as if they were lists of attributes, but the mental models of the users are more like pictures in the mind. Sometimes the searcher can "read-off" individual attributes from those mental images, but the mental image itself is a more integrated whole.

There is psychological evidence for this integrated view of visual mental models. Like physical objects, these models take time to rotate mentally (Cooper & Shepard, 1973; Shepard, 1978). When subjects are asked to compare rotated versions of the same object to verify that they are the same, the time required to do the comparison is proportional to the angular difference between the images. The larger the angular difference, the more time that is required to make the judgment. People also seem to scan these mental images as if they were images that their eyes are scanning. Kosslyn, Ball, and Reiser (1978) asked people to memorize highly schematic maps and then asked them to answer questions about the location of objects on the map. When the question arose about the location of an object, the time to reply varied with the distance that the object was from the prior location that they had been asked about. If the prior object

had been further away on the map, it took longer to answer than when the prior object had been nearby.

These mental models in the mind of an indexer or searcher can be descriptively sparse or rich depending on the situation. Some components of the model may be easily described linguistically, but other aspects might best be described or communicated by example or by images. The next four sections of this article will take a closer look at the attributes of shape and color in both mental models and in image indexes. Content-based and concept-based indexing will be related to each of these mental model attributes.

SHAPE IN VISUAL MENTAL MODELS

Psychologists have attempted to understand how perception gives rise to cognition and understanding. Many of these theories propose the existence of perceptual primitives. These are sometimes used in content-based approaches. In object recognition, these primitives may be generalized cones (Marr, 1982), simple geometric solids called "geons" (Biederman, 1987), or primitive features (Treisman & Gelade, 1980). When people are asked to describe objects, they often do so in terms of their parts (Tversky, 1989; Tversky & Hemenway, 1984). These primitive elements are combined in particular configurations to represent complex objects. In order to recognize and remember a bridge, the visual system breaks the scene into simple (more easily distinguished) parts. In early visual processing, a bridge may be broken into wedge-like supports and a rectangular prism for the deck. Once the shape has been recognized as a bridge, this fact may be added to long-term memory in abstracted form. The level of abstraction depends on the requirements of the task the person is working on. Information about the form of the bridge, including the shape of the arches, may be stored, or these details may be lost and only a strong trace of the general concept for "bridge" will be stored in long-term memory. For an image retrieval system, this means the indexer should index on just those properties that people have access to. These properties are dependent on both the shared memories of the users and the task parameters that accentuate particular aspects of these memories.

In the fields of computer vision and image retrieval, systems are often devised in layers. Primitive features are extracted from a scene and then combined into more complex features. David Marr (1982) articulated a model of visual primitives and generalized cones that served as a basis for much current research. Marr attempted to model human vision from the retinal image to object recognition. The lower level features are used for the construction of the 2.5-dimensional sketch. This sketch contains attributes that allow for later processing without the necessity to deal with lower level features such as light variations and discontinuities attributable to occlusion.

SHAPE IN INDEXES

When people view shapes, the shapes are recognized independent of their location (translation), their orientation (rotation), and their scale. Recognition is relatively resistant to noise. Variations in lighting and small occlusion do not interfere significantly. If a bug is missing a leg, it is nonetheless a bug. Some features selected by the visual apparatus are considered more important than others in defining similarity. Ideally, content-based algorithms that define shape similarity should behave in a manner consistent with human expectations and with the techniques that people use to define shape. The problem for content-based indexing is that current computational techniques do not have all of these properties. They tend to be effective at finding individual visual features, but the features frequently are not the same ones that people would recognize. They also tend to be poor at integrating the features to classify or recognize more complex objects. They are effective in recognizing straight lines and arches but not at recognizing that a particular combination of lines, edges, and colors is a bridge.

Still, this type of processing is the goal of many research systems. Consistent with the machine-vision tradition, content-based image retrieval systems model low-level feature-based information such as color, texture, and rough shape. These are used as evidence for the existence of higher level features or objects. Mehrotra (1997) provides a framework for understanding the levels of abstraction that may exist between an image and the viewer. The graphic of this model is reproduced in Figure 3. At the lowest level, there are *image features*. In that model, these features include color histograms, boundary segments, texture, and other "simple" features.

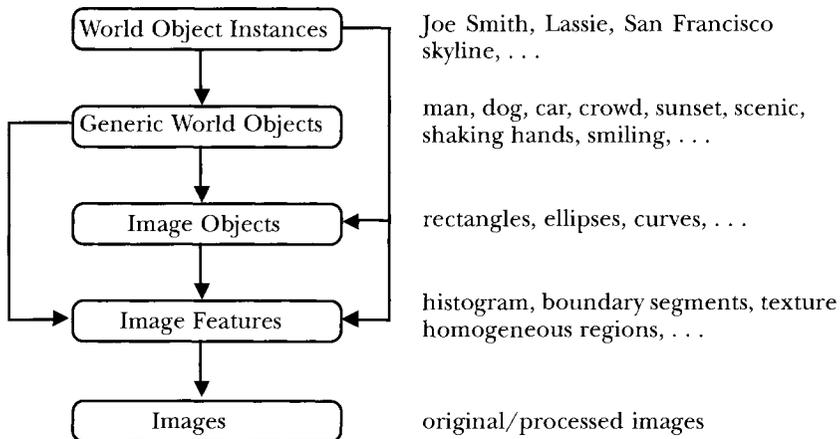


Figure 3. Levels of Abstraction in Museums (from Mehrotra, 1997).

Image objects, the next level of abstraction, are derived from collections of *image features*. Image objects include image regions, rectangles, and basic forms. The next level of abstraction is the *generic world object* such as man, dog, cat, or a smile. These include objects or categories to which many objects may belong. *World object instances* represent the next level of abstraction. These include objects for which there is one instance in the world that relates to the representation.

The concept-based approach to shape indexing focuses almost solely on generic world objects and world object instances. The indexer manually selects the relevant objects in an image and assigns keywords to the image. This linguistic tag approach is the primary means of image indexing in use today. The problems with this approach include expense, synonymy, and coverage. The manual operation requires a great deal of human effort to assign consistent tags and is therefore expensive. Another problem is that it is possible to describe an image many different ways. Even for textual material, it is difficult to select index terms that will be obvious to later searchers. This aspect of the problem is somewhat alleviated by the use of controlled vocabularies and thesauri, but then users are required to know that vocabulary. The problem is compounded in images. Sometimes users may wish to search for objects for which they have no name at all. This situation is not uncommon in that the image database is being used to facilitate object identification, as is the case with electronic field guides for the identification of plants and animals. Finally, the issue of coverage overlaps with that of expense. Indexers cannot normally create an entry for every object in an image. It is also very rare that an indexer has the time to index lower-level features such as the color or texture of an object or region. Consequently, when using the manual method of indexing, many objects and regions go unindexed.

These problems with content-based indexing can be demonstrated with the *AAT* by examining potential indexing options for the bridge in Figure 1 ("Sunset, Palmer Bridge, New York"). Part of the "IDNO 7836; TERM bridge" entry from the *AAT* is included in Figure 4. The index might be deepened by including the types of bridge that may apply but following the LINK entry "Bridge, stone" or adding entries for "IDNO 7838; TERM arch bridges" or "IDNO 7898; TERM single span bridges," if this is indeed a single span bridge. The parts of the object might be specified by following the related term of "RT <bridge elements>" from the "bridge" entry. Depending on the intended use of the index, the term "IDNO 994; TERM arch" could be included. The indexer would also need to decide which other objects in the image need to be indexed such as "IDNO 132410; TERM trees," "IDNO 8707; TERM river" (or perhaps "IDNO 8699; TERM stream" or "IDNO 11772; TERM water"), and "IDNO 133101; TERM winter." The correct index terms are not determined by the *AAT* but by the indexer's sociocognitive perspective on the intended

IDNO 7836
TERM bridges (built works)
ALT ALTERNATE bridge (built work)
BT <transportation structures by form>
RT <bridge elements>
SN SCOPE NOTE: Structures spanning and providing passage over waterways, topographic depressions, transportation routes, or similar circulation barriers
LINK bridges
LINK Bridges, aluminum
LINK Bridges, brick
LINK Bridges, concrete
LINK Bridges, iron and steel
LINK Bridges, masonry
LINK Bridges, plate-girder
LINK Bridges, prefabricated
LINK Bridges, stone
LINK Bridges, tubular
LINK Bridges, wooden

Figure 4. "Bridge" Entry in the *Art & Architecture Thesaurus*.

use. Even with this effort, these linguistic markers alone may be inadequate. Content-based techniques might facilitate some of the indexing and access.

Forsyth (1999) describes a system that represents a midpoint between content-based and concept-based approaches. This system uses a set of low-level image properties to infer the existence of objects. For example, an area of an image with a skin-like color, extended bilateral image symmetry, and nearly parallel sides might be a human limb. In a similar manner, it might be possible to build a bridge detector based on low level features. For example, the arched bridge in the image "Sunset, Palmer Bridge, New York" could be detected as a large dark area (stone) with a prominent arch(es) bounding the bottom and a near horizontal vertical line bounding the top. Based on the results of the bridge detector, "bridge" can be entered into the database along with a value representing the certainty of the classification. This technique borrows heavily from vision research and has the goal of being able to perform concept-based indexing at least within limited domains. The weakness of the technique is that detectors must be built for all objects of interest. The detector for arched bridges might not generalize to other bridge types, such as suspension bridges, requiring the construction of another detector. Detectors for rivers, trees, and sunsets would need to be constructed.³

In some situations, it may be possible to introduce a visual thesaurus. This type of thesaurus represents the choices visually rather than in natural

language as is the case with typical thesauri (Hogan et al., 1991). This allows people to "see" the visual indexing structure of a collection.

The techniques used in most content-based systems are aimed at a lower level in Mehrotra's hierarchy and stop with image features. The main techniques include template matching and edge abstraction matching. Most techniques of this type are two-dimensional projections of three-dimensional objects and suffer from perspective dependence. A query is constructed by using an example image from which the system may extract a shape outline or by hand sketching the desired shape. Rotation and scaling can also cause a mismatch. For example, the profile of a bridge looking from the road crossing is very different from the profile from the river the bridge crosses. Likewise, the same bridge from two different distances may produce different results. Current research is aimed at eliminating these types of limitations.

A general discussion of template matching can be found in Forsyth (in this issue of *Library Trends*). The System Query by Image Content (QBIC) (Barber et al., 1992; Niblack et al., 1992; Flicker et al., 1995) is a typical example of this approach. In template matching, a shape is normalized through translation, rotation, and scaling to produce an easily comparable standard form or template. These templates may be automatically extracted, but it is easier to have a user provide a sketch or outline of the desired object. The indexer, with the assistance of the computer, sketches the outline of objects of interest in an image. The system converts these outlines into templates by applying a standard rotation, scale, and translation. The system then stores the template as an index. In the sample image, the indexer would sketch the outline of the bridge. When users search the system, they may sketch the desired object. The system converts this sketch to a template and then compares this template to those in the index by counting the number of overlapping pixels. The greater the count, the higher the similarity. Allowing the indexer to add a name to the sketches could augment this technique. Unfortunately, the technique is sensitive to small variations in the images. In our sample image, the edges of the bridge are partly obstructed by trees. The indexer and searcher may choose different edge boundaries leading to a template mismatch. The same bridge from a different perspective would not be recognized or retrieved although scaling can be compensated for by QBIC. The advantage of the technique is that one algorithm applies to all objects. There is no need to create new detectors for each object of interest.

There are a number of edge abstraction techniques for classifying shape. These include, for example, turning angle descriptors, segmentation, and Fourier descriptors. Mehrotra and Gray (1995) describe a shape representation based on segmenting the edge of objects into straight-line segments. These segments are normalized for scale, rotation, and transla-

tion. Similarity is defined as the Euclidean distance between normalized points. The normalization helps to make the algorithm match human expectations, but the establishment of a start location and break points for the segmentation is problematic.

Another example of a boundary-based shape similarity approach is the Modified Fourier Descriptor (MFD) (Huang et al., 1997). This approach corrects faults in the Fourier Descriptor approach to produce a representation that is consistent in the face of transformations and noise.

All three of these approaches are weak in that they are not well-matched to human performance or expectations. They do not break objects into parts or other psychologically relevant features. Among these is the critical issue of dimensionality. Humans perceive two-dimensional images as three-dimensional. People combine the evidence in the image with long-term models in memory to produce three-dimension-like visual mental models (Hayward & Tarr, 1997). The QBIC template matching technique, the line-segment technique, and Fourier descriptors all act on two-dimensional projections of three-dimensional objects. Model-based computer vision research is focused on solving this projection problem.

A key issue is how any of these methods relate to users' mental models and how they operate at the interface level. If a user has a mental model and retrieval goal of a particular type of bridge at a particular orientation low-level feature, content-based techniques may be appropriate. If, however, these details are not relevant in the mental model or unspecified in the model then the concept-based approach is more appropriate. If the domain is narrow enough, this content indexing might be provided by automatic techniques such as those developed by Forsyth.

COLOR IN MENTAL MODELS

Color is not light of a particular wavelength but rather it is combinations of light of different wavelengths. It is possible to produce the same perceived color through many combinations of wavelengths and intensity of light. The perception of color derives from the relative activation of three types of color receptors in the human retina. These receptors have highest sensitivity to wavelengths corresponding approximately to red, green, and blue. Red and green act as opponent colors, as do combinations of red and green receptors (yellow) and blue. The activation of one opponent color leads to the inhibition of the other opponent color—e.g., the perception of yellow stems from simultaneous moderate activation of both the red and green receptors. Brightness or intensity is encoded separately as a combination output of red and green receptors. Sharp and Philips (1997) provide a brief discussion of the neural aspects of vision. Hendee (1997) provides a discussion of the cognitive interpretation of color.

In a retrieval environment, multiple levels of abstraction may exist in both the index and in the minds of the searchers. Objects at all levels of abstraction sometimes have linguistic labels that are available to the searcher or indexer. For example, at the image feature level, a particular color may or may not have a color name associated with it. In a content-based index, a region's color might be stored as a color histogram with no linguistic label. The viewers may possess no words in their mental model to describe the color. There are 7 million discernible colors. Categorization and naming allows us to reduce this complexity. We cannot name them all (Bruner et al., 1956). The number of named primary colors may vary (in very systematic ways) from culture to culture as discussed below. Indeed, the meaning of the label may vary with the object to which it is applied. For example, the location on a color map of a *red apple* is different from the label for red in *red skin* (Clark, 1992, p. 370). Conceptual indexing would work best for the first and content-based techniques for the latter.

Some aspects of a model may be easily nameable and others may be difficult to assign labels to. The indexer must select the technique that is appropriate for each type of image element. Color names follow cross-cultural patterns. The indexing must follow these patterns or run the risk of producing confusing results to a user's queries. Lakoff (1987, pp. 24-40) discusses research into the relationship between color categories and color names. It is possible to assume that the assignment of color names to the spectrum is arbitrary. Different cultures might focus on different colors that are relevant in their environment and assign them names. Contrary to the arbitrary color hypothesis, Berlin and Kay (1969) demonstrated that there is a set of basic colors shared across cultures. These colors tend to have shorter names, are used more frequently, and there are about eleven of them. These are *black, white, red, yellow, blue, green, brown, purple, pink, orange* and *gray*, in that order. Of course the actual word used to represent the color differs but the colors are the same. When a language has only two basic color names, they are *black* and *white*. The other basic colors are grouped under these terms as *dark* or *bright* colors. When a language has a third basic color named it is *red*. When there is a fourth basic color term in a language it is usually *yellow, blue, or green*. Any one of these may be added first. Languages with six color terms will have the equivalents of *black, white, red, yellow, blue, and green*. The seventh color is *brown*. *Purple, pink, orange, and gray* are added next in no particular order. The obvious question is "Why" and is beyond the scope of this article. Interested readers, however, might start with Kay and McDaniel (1978) or Lakoff (1987).

The significance of this research for indexing is that if one is going to use linguistic labels for color names, these are the ones to use first. As is discussed in the next section, these labels are indeed included in the *AAT*.

There is additional structure to the human perception of color. Rosch (1973) studied focal and nonfocal color patterns. The first observation is that there are best examples of colors that cross cultures. So the *red* in one culture is the same *red* in other cultures. This is even true in cultures that have no basic color name for red. These colors are easier to learn. The focal colors act as cognitive reference points (Rosch, 1975). While these colors may have a privileged status in mental models, experience can play an important role in identifying the meaning of a term like "red." The meaning of red varies with context, so the meaning of red is different for a red apple, red skin of a sunburn, or a red sunset. Focal red is located at the center of a neighborhood of meaning for the color red (Clark, 1992, p. 371).

COLOR IN INDEXES

Using the concept-based approach, it is natural to map color names to particular objects within an image. There are a number of available controlled vocabularies available for this purpose. The *ATT* has a hierarchical color naming system. "IDNO: 131648 TERM: chromatic colors" acts as a base term, these being *pink, red, orange, brown, yellow, olive, yellow green, green, blue, and purple*. This does not quite match the basic color names described earlier but it is close. These *AAT* color terms serve as base terms for less prototypical colors. For example, *blue* is the base term for "IDNO: 129787 TERM: <intermediate blues>" and in turn "<intermediate blues>" is the base term for "IDNO: 130602 TERM: violet" as well as other related colors.

Colors are further categorized using the "use for" (UF) fields. The chromatic color thesaurus entries (*pink, red, orange, and so on*) do not contain UF fields except *yellow, green* that has a UF of *green, and yellow*. "IDNO: 129645 TERM: pale blue" is a type of *blue* and the terms such as *dull greenish blue* are mapped to *pale blue* through the UF field. So the sky in the bridge at sunset picture might be coded as 129645. Individuals will vary on their definition of these more obscure color names (but not the focal colors).

There are also achromatic colors defined in *AAT*. These are colors without hue and include black, white, and grays. This is consistent with the color space discussion of content-based indexing later in this section.

There are thousands of color names defined in the *AAT*. It takes a good deal of effort on the part of both indexers and searchers to arrive at the same term for the color of a particular sky. There is some support for color similarity through the thesaurus. If a user enters a particular color term, the system should be able to search for images or objects in images with this label. Should the search fail, the system should be able to automatically relax the matching constraints by moving up to the base term (e.g., *pale blue* to *blue*) and perform a search with that high level

term. If that fails, the system might use OR to group all terms having the base term of *blue*. This is much to ask of a system, but similar systems have been built. There exist other color name standards, such as the *National Bureau of Standards Dictionary of Color Names*, which provides thousands of color names and the *National Bureau of Standards/NBIC Color System*, that maps all colors into a small set of over 200 names.

If a retrieval environment were to require the use of more than the primary colors, it would be unreasonable to expect either indexers or searchers to have names for them. It is even more unlikely that they would agree on the names. In these situations, it might be more reasonable to allow users to directly specify color. An analytic approach might allow users to select different levels of red, green, and blue from window slider bars for example. The combination of the colors would reproduce all other colors. There are multiple problems associated with deciding on a color indexing system—e.g., no one solution can fit all needs. Frequently, a combination of approaches is needed, drawing from both what has been defined as content-based and concept-based techniques. A few of the central issues include the decision of what to index, the determination of color averages and color naming (both qualitative and quantitative). For the content-based approach, the color histogram is the favored method for representing average color. Color histograms are discussed elsewhere by Forsyth in this issue of *Library Trends*. These may be used to represent the overall image color, the color of regions, or the color of objects in increasing level of difficulty.

Digital images are composed of a series of points. The color of a particular point may be represented in either qualitative or quantitative terms. Indexing on a pixel level is not very useful in most cases. As indicated in the Mehrotra model in Figure 3 above, these individual elements may be collected together into homogeneous regions as *image features*. Unlike concept-based indexing, these regions need not correspond to individual objects or even object parts. These regions or “blobs” may be indexed independently. Smith and Chang (1996a, 1996b) discuss one approach to this problem. Regions of similar color are labeled with their location and color.

There are many approaches to color, but the Smith and Chang approach is useful for the purposes of explanation since it demonstrates some basic ideas and is computationally tractable. In this approach, color is represented in the HVS color space: hue, value, and saturation. There are many other color spaces, and these will be discussed later. Hue is the tint of what is typically pure color. Saturation is the amount of color mixing where fully mixed red, green, and blue appears as white. Value is the lightness or intensity. Colors are quantized into a small number (166) of color regions: eighteen hues, three saturations, three values, plus four grays. Colors that fall anywhere in a region are considered the same for

indexing purposes and for identifying regions. This quantization of color is reminiscent of the categorization of color performed by humans. The choice of eighteen hues is interesting in that it does not correspond to the eleven basic colors identified by Berlin and Kay (1969) and Rosch (1974). Eleven hues might match human expectations better without adding computational complexity to the approach.

So, in the content-based approach, regions of like color or "blobs" are indexed. At first glance, such a nonobject-oriented approach would not seem to correspond to the human experience of the same image. Humans, after all, recognize physical objects in images as in Mehrotra's "generic world objects" (man, dog, car, and so on). In practice, however, the technique is sometimes useful because, happily, the "blobs" do correspond to objects. In an image database of nature photography, yellow blobs in the middle of the frame frequently correspond to yellow flowers and yellow blobs in a collection of bird photos often correspond to yellow birds. There is, of course, a high error rate that increases with the heterogeneity of the image collection. The color quantization approach is useful for finding color regions, but an additional mechanism is needed to handle color similarity. In the concept-based or keyword approach to color matching, either the color of an image matches the color of the query or it does not. That is, if a sky is indexed with the keyword "blue," only the word "blue" in a query will match it. This does not match with human modeling of color. We know from Rosch's work on color prototypes that colors are not created equal, and that some colors may be better or worse members of the "blues" than others. "Blue" is closer to "light blue" and "bluish-green" than it is to "red." Content-based color similarity methods can be built which much more closely match these intuitions. Again using Smith and Chang's quantized color region approach as an example, the distance between two colors, or their similarity, can be defined as the number of steps that need to be taken in the quantized space to move from one color region to another. Hue is broken into eighteen regions. The first region might correspond to something like reds, the second region to oranges, and so on up to the last visible violet. Regions that are close to one another are close in color. The "orange" bin is distance one from "red," and the "violet" region is distance seventeen from "red." The same applies along the (3) saturation and (3) value axes. Color distance is the sum of the hue, saturation, and value distances. A user may search for a blue sky and have a relatively strong color match for the sky in the "Sunset, Palmer Bridge, New York" image.

There are other color spaces and similarity measures that more closely match human perception. Human color perception, however, has certain limits. Some wavelengths are simply not visible. This may be because the wavelength of light is beyond the range of our color receptors (retinal cones). Likewise, the intensity may be too low or too high. The designer

might choose to use a model that may represent all visible colors. The Commission Internationale de L'Eclairage Color Space (CIE) is such a representation. This is a three-dimensional color model that represents saturated colors (red, green, and blue) on outside edges of a bounded plane. Unsaturated colors are in the central area with white in the middle. Intensity is expressed on an axis orthogonal to the color plane. One boundary is black and the other full intensity. While this model represents all visible colors, it does not compensate for human processing of the RGB channels. The CIE-LAB model does bit mapping of the color space into complementary colors. There is a red-green axis, a yellow-blue axis, and a black-white axis as there is in the central processing of color in humans. *Munsell*, a U.S. standard, is another popular standard. The problem with these spaces is that it is sometimes difficult to map the standard RGB encoding used in monitors and scanners.

These color spaces have been constructed to capture important aspects of the human perception of color. Human and computer indexers may use them as a tool to describe aspects of an image. This use of the color spaces will be successful inasmuch as they are consistent with expectations and mental models of the users of the index.

CONCLUSION

Humans have evolved mechanisms that allow them to represent important aspects of the visual world. These visual mental representations are used on a daily basis to recognize objects and navigate through the world. Many aspects of these visual models predate the evolution of language. Language evolved to facilitate our ability to communicate with one another—i.e., facts about the world and our understanding of the world. Language has access to particular aspects of our visual mental models, allowing people to describe their interpretation of the world. In order for others to understand these descriptions, there must be a shared experience of the world and a shared vocabulary. The nature of both the visual mental models and the linguistic mechanism have a profound effect on how image retrieval systems should be built. Indexers may use language and this shared knowledge to create language-based descriptions of images in a collection. Computer algorithms are being developed that allow some parts of this linguistic indexing to be performed cost effectively by computers at least in narrow subject domains (Forsyth et al., 1996; Forsyth, 1999; Srihari, 1995, 1997). These computer systems are breaking down some of the distinctions that have existed between content-based and concept-based indexing.

Some aspects of the visual mental models are not easily described with natural language. As discussed in the section on color indexing, there are millions of human-discernible colors but relatively few color names. In some cases, content-based computational techniques can be

used to communicate information about these nonlinguistic aspects of the visual models. These techniques are used in systems such as Virage (Gupta et al., 1997), QBIC (Niblack et al., 1992; Flickner et al., 1995), VisualSEEK (Smith & Chang, 1996a), and Photobook (Pentland, 1993). Some systems, such as Photobook, attempt to select image properties that are particularly perceptually salient. Some of the mechanisms involved in the representation of shape and color are discussed in this article. No one content-based representational technique is likely to capture all of the important aspects of an image. The mental model of images has multiple aspects. The image features of different types are reflected in the different aspects of the mental models. Content-based and concept-based approaches to indexing are each better suited to different aspects of the models. Indexers may choose to use content-based or concept-based linguistic or nonlinguistic indexing depending on the demands of the tasks that will be performed by the users and what aspects of the visual mental models will be available to them.

NOTES

- ¹ Reproduced with permission from the Library of Congress, Prints and Photographs Division, Detroit Publishing Company Collection.
- ² It is useful to be able to refer to the color version of this image in the American Memory Collection. The image may be accessed through the Web by searching for the title at <http://memory.loc.gov/ammem/detroit/dethome.html>.
- ³ Interestingly, detectors for trees and sunsets have been constructed (see Forsyth in this issue of *Library Trends*).

REFERENCES

- Art and Architecture Thesaurus*. (1994). Getty Art History Information Program, 2d. ed. New York: Oxford University Press.
- Barber, R.; Cody, W.; Equitz, W.; Flickner, M.; Glasman, E.; Niblack, W.; & Petkovic, D. (1992). *Query by image content (QBIC) status as of 8/92*. Unpublished Technical Report RJ 89 (80237) September 14, 1992. IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099.
- Barnett, P. J., & Petersen, T. (1989). Subject analysis and AAT/MARC implementation. *Art Documentation*, 8, 171-182.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Borgman, C. L. (1986). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies*, 24, 47-64.
- Bruner, J.; Goodnow, J.; & Austin, G. (1956). *A study of thinking*. New York: Wiley & Sons, Inc.
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. Orlando, FL: Academic Press.
- Detroit Photographic Co. "Sunset, Palmer Bridge, New York." c1900 (Touring Turn-of-the-Century America: Photographs from the Detroit Publishing Company, 1880-1920). Retrieved November 10, 1999 from the World Wide Web: <http://memory.loc.gov/ammem/detroit/dethome.html>.

- Flickner, M.; Sawhney, H.; Niblack, W.; Ashley, J.; Huang, Q.; Dom, B.; Gorkani, M.; Hafner, J.; Lee, D.; Petkovic, D.; Steele, D.; & Yanker, P. (1995). Query by image and video content: The QBIC system. *Computer*, 28(9), 23-30.
- Forsyth, D.; Malik, J.; Leung, T.; Bregler, C.; Carson, C.; Greenspan, H.; Fleck, M. (1996). Finding pictures of objects in large collections. In P. B. Heidorn & B. Sandore (Eds.), *Digital image access & retrieval* (Proceedings of the 33rd Annual Clinic on Library Applications of Data Processing, March 24-26, 1996, University of Illinois at Urbana-Champaign). Urbana-Champaign: University of Illinois, Graduate School of Library and Information Science.
- Gupta A., & Jain R. (1997). Visual information retrieval. *Communications of the ACM*, 40(5), 70-79.
- Hastings, S. K. (1995). Query categories in a study of intellectual access to digitized art images. In T. Kinney (Ed.), *ASIS '95* (Proceedings of the 58th annual meeting of the American Society for Information Science, October 9-12, 1995, Chicago, IL) (pp. 3-8). Medford, NJ: American Society for Information Science.
- Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology-Human Perception and Performance*, 23(5), 1511-1521.
- Hendee, W. (1997). Cognitive interpretation of visual signals. In W. R. Hendee & P. N. T. Wells (Eds.), *Perception of visual information* (pp. 149-175). New York: Springer-Verlag.
- Hogan, M.; Jorgensen, C.; & Jorgensen, P. (1991). The visual thesaurus in a hypermedia environment: A preliminary exploration of conceptual issues and applications. In D. Bearman (Ed.), *Hypermedia & interactivity in museums* (Proceedings of an International Conference, October 14-16, 1991, Sheraton Station Square, Pittsburgh, PA). Pittsburgh, PA: Archives & Museum Informatics.
- Huang, T. S.; Mehrotra, S.; & Ramchandran, K. (1997). Multimedia analysis and retrieval system (MARS) project. In P. B. Heidorn & B. Sandore (Eds.), *Digital image access and retrieval* (Proceedings of the 33rd Annual Clinic on Library Applications of Data Processing, March 24-26, 1996, University of Illinois at Urbana-Champaign). Urbana-Champaign: University of Illinois, Graduate School of Library and Information Science.
- Jacob, E., & Shaw, D. (1999). Sociocognitive perspective in representation. *Annual Review of Information Science and Technology*, 33, 3-57
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- Kay, P., & McDaniel, C. (1978). The linguistic significance of the meanings of basic color terms. *Language*, 54(3), 610-646.
- Kosslyn, S. M. (1980). *Images and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M.; Ball, T. M.; & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of visual scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 47-60.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Library of Congress. (1995). *Thesaurus of graphic materials* (comp. & ed. the Prints & Photography Division, Library of Congress). Washington, DC: Library of Congress Catalog Distribution Service.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Mehrotra, R. (1997). Content-based image modeling and retrieval. In P. B. Heidorn & B. Sandore (Eds.), *Digital image access and retrieval* (Proceedings of the 33rd Annual Clinic on Library Applications of Data Processing held March 24-26, 1996, at the University of Illinois at Urbana-Champaign). Urbana-Champaign: University of Illinois, Graduate School of Library and Information Science.
- Mehrotra, R., & Gray, J. E. (1995). Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9), 57-62.
- Niblack, W.; Barber, R.; Equitz, W.; Flickner, M.; Glassman, E.; Petkovic, D.; Yanker, P.; Faloutsos, C.; & Taubin, G. (1992). The QBIC project: Querying images by content using color, texture, and shape. In A. A. Jamberdino & W. Niblack (Eds.), *IMAGE storage and retrieval systems* (Proceedings of the SPIE—the International Society for Optical Engineering) (vol. 1662, pp. 173-181). Bellingham, WA: SPIE.

- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
- Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman, D. E. Rumelhart, & LNR Research Group (Eds.), *Explorations in cognition*. San Francisco: Freeman Press.
- Panofsky, E. (1955). *Meaning in the visual arts*. Garden City, NY: Doubleday Anchor Books.
- Pentland, A.; Picard, R. W.; & Sclaroff, S. (1993). Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3), 233-254.
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13(4), 707-784.
- Pylyshyn, Z. W. (1977). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80, 1-24.
- Rasmussen, E. M. (1997). Indexing images. *Annual Review of Information Science and Technology*, 32, 167-196.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532-547.
- Sharp, P., & Philips, R. (1997). Physiological optics. In W. R. Hendee & P. N. T. Wells (Eds.), *Perception of visual information* (pp. 1-32). New York: Springer-Verlag.
- Shepard, R. N. (1978). The mental image. *American Psychologist*, 33, 125-137.
- Shera, J. H. (1965). *Libraries and the organization of knowledge*. Hamden, CT: Archon Books.
- Smith, J. R., & Chang, S-F (1996a). VisualSEEK: A fully automated content-based image query system. In *Proceedings of the ACM International Conference on Multimedia* (November 1996). Boston: Association for Computing Machinery.
- Smith, J. R., & Chang, S-F. (1996b). Tools and techniques for color image retrieval. In *Proceedings Storage & Retrieval for Image and Video Databases IV* (Vol. 2670). San Jose, CA: IS&T/SPIE.
- Srihari, R. (1995). Automatic indexing and content-based retrieval of captioned images. *IEEE Computer*, 38(9), 49-56.
- Srihari, R. (1997). Using speech input for image interpretation, annotation, and retrieval. In P. B. Heidorn & B. Sandore (Eds.), *Digital image access and retrieval* (Proceedings of the 33rd Annual Clinic on Library Applications of Data Processing held March 24-26, 1996, at the University of Illinois at Urbana-Champaign). Urbana-Champaign: University of Illinois, Graduate School of Library and Information Science.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Tversky, B. (1989). Parts, paronomies, and taxonomies. *Developmental Psychology*, 25(6), 983-995.
- Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, 113(2), 169-191.