PHOTOGRAPHY MEDIA

Forum | Journal | Gallery | Contact | Site map | Bookstore | About

SHARE

**Photography Media Journal**
ISSN 1918-8153

# Image Indexing

Tomasz Neugebauer

March 2005

Pages > 1 2 3 **Print version (full-text)**

> "It is my intention to present—through the medium of photography—intuitive observations of the natural world which may have meaning to the spectators."[*]
> Ansel Adams (1902-1984)

## Abstract

The theoretical difficulties in indexing images include: 1) images do not satisfy the requirements of a language whereas textual materials do 2) images contain layers of meaning that can only be converted into textual language using human indexing 3) multi-disciplinary nature of the images where the terms assigned are the only access points. Theoretical foundations for image indexing consist of distinctions between classes of terms including 'of' and 'about', syntactic and semantic, specific and generic, and answers to the questions 'who?', 'what?', 'where?', and 'when?'. Content-based indexing can be used to generate terms for the color, texture and basic spatial attributes of images. Image searchers use textual descriptor search terms that require human description-based indexing of the semantic attributes of images.

> "To the complaint, 'There are no people in these photographs,' I respond, 'There are always two people: the photographer and the viewer.'"
> Ansel Adams

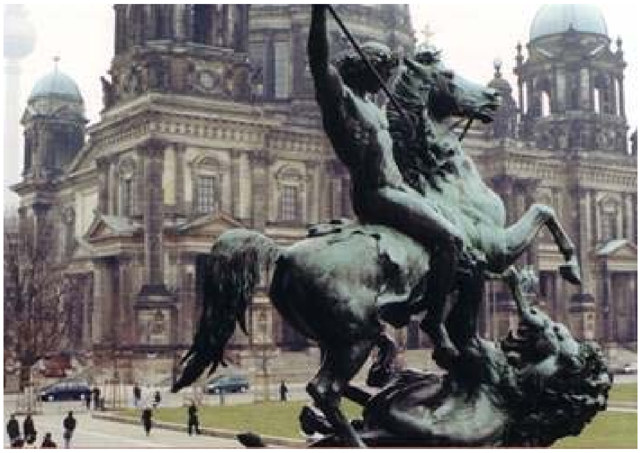
## Introduction

Jacobs argues that the importance of providing improved image access has increased due to the democratization of images as legitimate sources of information and education. Managed use of images “is no longer the fairly elite domain of art historians or specialized archives.” (Jacobs 119) The popularization of the image-rich World Wide Web as a communication and education medium certainly confirms this. Efficient image search tools by Google (http://images.google.com), MSN (http://search.msn.com/images) and Yahoo! (http://images.search.yahoo.com/) that index millions of freely available images online is clear evidence that image indexing and searching is common and widespread in today’s visual culture.

## Difficulties

The main difference and difficulty in indexing and classifying images as opposed to textual materials is that images do not satisfy the requirements of a language. Nelson Goodman defines the requirements of a language as satisfying at least “the syntactic requirements of disjointness and differentiation. “ (Goodman 226) Disjointness is the requirement that “no mark may belong to more than one character” (Goodman 133) whereas differentiation is defined as “For every two characters K and K’ and every mark m that does not actually belong to both, determination either that m does not belong to K or that m does not belong to K’ is theoretically possible.” (Goodman 136) It is especially this requirement of differentiation that is not satisfied in the case of images as opposed to text: “Nonlinguistic systems differ from languages, depiction from description, the representational from the verbal, paintings from poems, primarily through the lack of differentiation – indeed through density (and consequent total absence of articulation) – of the symbol system.” (Goodman 226) The disjointness and differentiation of textual language systems allows for simple decomposition into its component symbols: letters, words, paragraphs, and this represents at least a syntactic ease in indexing textual material. The semantic ambiguities inherent in indexing the meaning of a text ensure that indexing remains an art form without absolute answers, but at least the syntax lends itself to analysis. This is not the case with the syntax of images. Thus, as Jacobs observes, images “are usually absorbed holistically by the viewer.” (Jacobs 120)

Furthermore, as Besser points out, textual material is usually written with a clearly defined purpose that is explicitly stated, summarized and abstracted by the publishers in introductions, prefaces and book covers, whereas images are not (Besser 788). Images are “decidedly multidisciplinary in nature: they contain a variety of features, each of which may be of potential interest to researchers from somewhat diverse fields of study” (Baxter & Anderson). For example:



A photograph such as the one above may be of interest to a student of sculpture, but it may also be “useful to historians wanting a snapshot of the times, to architects looking at buildings, to urban planners looking at traffic patterns or building shadows, to cultural historians looking at changes in fashion, to medical researchers looking at female smoking habits, to sociologists looking at class distinctions, or to students looking at the use of certain photographic processes or techniques" (Besser 788) This requires a high number of access points for each image, depending on the audience, purpose of the database, and the user characteristics. The cultural historian will not be using attributes such as composition, lighting and perspective, whereas a student of art history might. These difficult requirements are compounded by the symbolic and allegorical meaning of images that is highly subjective and interpretive, leading to low levels of interindexer consistency (Baxter & Anderson).

Presumably when indexing images for a particular database, we have a notion of who the end-users are so that we can try to anticipate the most appropriate access points for each image. However, image collections are especially multidisciplinary and targeting the indexing for a subset of the end-users truly excludes the others. The image is not textually expressive in itself and the assignment of terms is a purely interpretive exercise by the indexer that ends up being the most important access point for retrieval by the user. In the case of textual information, the user can always switch to natural language full-text searching if frustrated by exclusive descriptors, whereas this option is simply not available in image searching.

## Theory

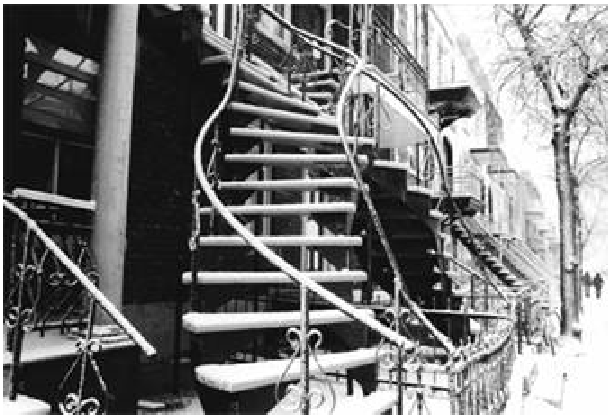
Shatford extends Panofsky's theory(see Erwin Panofsky, Studies in Iconology: Humanistic Themes in the Art of the Renaissance. New York: Harper & Row, 1962.) of three levels of meaning in a work of art to general subject analysis of pictorial work (Shatford 1986: 43). Only the first and second levels of Panofsky's theory are to be indexed, since the third seems exemplified by the content of a subjective critical review. Jacobs interprets Panofsky's third level as requiring extensive interpretation and cultural knowledge that is reserved for art history and similar domains. (Jacobs 120). Panofsky's first two levels are the basis for Shatford's generic and specific levels of description.

Table 1. Panofsky’s Levels of Meaning (Shatford 1986)[‡].

| Level of Meaning | Examples | Panofsky’s term | Shatford’s term |
|---|---|---|---|
| **first** | **Factual:** man, woman, tree **Expressional:** grief, peacefulness, happiness, anger | pre-iconography (description) – “primary or natural subject matter” requiring “everyday familiarity with objects and events.”<br>***factual*** \| ***expressional*** | generic description of the objects and actions represented. The **factual** are descriptions *Of* , **expressional** represent the *About.* |
| **second** | fat jolly man sitting in lotus position is *of* the Buddha *about* compassion. | iconography (analysis) – “secondary or conventional matter”, requiring knowledge of culture. | specific objective meaning descriptions *Of* and mythical abstract or symbolic *About* |
| **third** | Zanussi’s Illumination (film) is an instance of cinema-verite, and about science, ethics, and enlightenment. An image of a grey brick wall is *about* socialism. | iconology (interpretation) – “intrinsic meaning of content”, requiring synthesis, knowledge of artistic, social and cultural setting | not intended to be indexed except in highly specialized domains since agreement on meaning at this level is too difficult to achieve. |

Shatford uses Frege’s distinction between sense and reference to show that “images may be defined as referents for the sense of the words used to describe them” (Shatford 1986: 46) A single image can be the referent for many different senses of words, or conversely and more conventionally: a picture is worth a thousand words.

The kinds of subjects to be indexed in pictures correspond to answers to the four questions: *who?, what?, when?, and where?*, and “each of these basic facets may then be subdivided into aspects based on *Of* in the specific sense, *Of* in the generic sense and *About*." (Shatford 1986: 48) To illustrate let us take the following photograph:



The first level represents the literal general meaning that requires little or no cultural knowledge: it is a photograph *of* staircases and *about* a city street in winter. The second level of meaning represents the symbolic and specific, requiring some cultural background to discern: it is a photograph *of* St. Urbain street in Montreal, and *about* a disappearing winter beauty of iron craftsmanship.

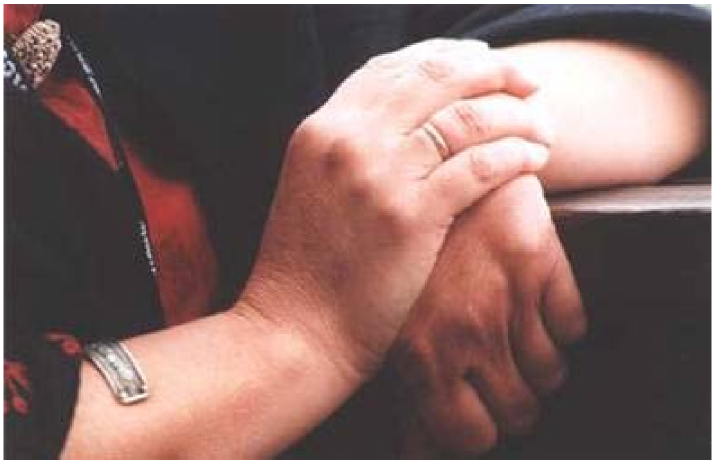
Shatford's who, what, when and where is the basis of a faceted classification of images, divided into answers to questions as to the *Of* and the *About* of a picture:

· "who or what, beings and objects, is this picture *Of*?"
· are these symbols, representations, abstractions or personifications, in other words, "what are they *About*?" (Shatford 1986: 50).

The *Who Of* is divided into the generic and specific. Berinstein gives the following illustrative example (86):

| Who of Specifically | Who of Generally | Who About |
|---|---|---|
| Nancy Garman | Woman | Editor |
| Starship Enterprise | Spaceship | Exploring space, the future. |

The “what” facet is broken down into *Of* (events and actions) and *About* (conditions and emotions). For example:



The above photograph is *Of* (general) a woman holding her hand, *Of (Specific)* Rossitsa Valkanova and *About* a film producer, support and perseverance.

The when facet includes specific linear and general cyclical time: “specific dates and periods (June 1885, Renaissance) and recurring time (Spring, Night)” (Shatford 1986: 53). The *about* aspect of when facet will rarely be used, it answers the question “Is the element of time represented in the picture a manifestation of an abstract idea? A symbol?” (Shatford 1986: 53). Berinstein interprets this as the “linking of ideas associated with time, such as ‘the end of the world’” (86) Similarly, the *About* aspect of the where facet will “only be present if the place symbolizes another place or an abstract idea (Heaven, Hell, Paradise)” (Berinstein 86). The where facet includes “geographic, cosmographic and architectural space” and is divided in the specific *Of* aspect that includes “locale, site, or place represented” and generic *Of* including “kinds of places, […] landscape, cityscape, interior, planet, jungle.” (Shatford 1986: 53)

## Two Separate Approaches

Jörgensen identifies some broad research areas in the domain of image access closely related to each other: indexing and classification, image users and uses, image parsing. (Jörgensen 1999: 294) In the area of indexing and classification, there is a fundamental question of the appropriateness of textual indexing for a non-textual medium.

Studies of the research literature in image indexing and retrieval confirm that there are two distinct approaches to the problem: content-based and description-based. (Chu 1017) The content based approach is the automated extraction of textual and image properties whereas description-based refers to human indexing using captions and keywords for artist and image data. (Chu 1011). Content-based techniques are primarily the domain of computer scientists developing technologies for “direct retrieval of visual image content such as shape, texture, color, and spatial relationship” (Jörgensen 1999: 300) In content-based retrieval systems, instead of entering search terms, for example, the user selects an area of an image in order to narrow down the search and retrieve related items. (Jörgensen 1999: 300) Alternatively, the system fetches similar items based on quantifiable aspects of the image or its parts. Although content-based research is popular, it has not produced systems that satisfy end-user information needs. This is because properties of an image such as shape, texture, and color contribute to our understanding of an image, but do not define it. Text-based search techniques remain the most efficient and accurate methods for image retrieval (Jörgensen 1999: 302).

## Indexing for Retrieval

In image database retrieval systems, the most common method of indexing and retrieval is the establishment of “fields containing bibliographic data (artist, photographer, title, date, etc.) together with a field containing some descriptive text to support the image field.” (Baxter 68) Shatford argues that although different image attributes need to be indexed depending on the type of collection and users, the general categories of: biographical, subject, exemplified and relationship attributes can be used as a checklist. (Shatford 1994: 584) The biographical attributes include those about the history of the image’s creation such as *creator attributes*, *creation date*, as well as its “travels […] Where it is now, where it has been, who has owned it, how much it costs or has cost, whether it has been altered in any way.” (Shatford 1994: 584) Exemplified attributes are distinct from the subject: they describe the visual object as an *instance-of*: an etching, photograph, poster or an 8-bit GIF. Relationship attributes are those *between different images* (for example, a digital image of a class hierarchy expressed in Universal Modeling Language and an image of the Graphical User Interface that is an implementation of that hierarchy), or *between images and texts* (e.g.: the list of functional requirements that were used to create the class hierarchy.)

Jörgensen summarizes research in cognitive psychology suggesting that users search for objects in images at a ‘Basic Level’, which is “neither the most specific nor the most abstract level but is rather an intermediate level” and carries out studies that confirm this (305). The layers and levels of indexing of images inevitably vary with each collection, its audience and purpose. Schroeder describes the General Motors Media Archive (GMMA) project dealing with over 3 million still photographs, and using a layered indexing approach consisting of the object (trees, sky, dirt road, etc.), style (glamour, documentary, engineering) and implication (purpose, unique qualities) (Schroeder, 1998). The implication layer is the greatest challenge to the indexer and of highest value to the searcher, as it provides the differentiating unique quality descriptions of the images, enabling an increase in precision.

The large number of terms necessary in image indexing has led to a large number of specialized thesauri-based systems. The Art and Architecture Thesaurus uses hierarchically arranged “terminology describing physical attributes, styles and periods, agents, activities, materials and objects” (Baxter & Anderson) ICONCLASS is another classification system used extensively to index images in the domain of art history (Baxter & Anderson). Thesauri-based systems raise issues of cost-effectiveness: it is time-consuming to translate the multitude of terms into structured vocabularies that are not standardized for the entire domain of images. The subjective elements of ‘aboutness’ of an image, formed by the viewer during the sensory visual experience of understanding an image require too many terms to be exhaustively indexed. This is the reason why content-based algorithmic techniques continue to be the focus of research and development.

The misconception about algorithmic techniques is that they do not require human indexers. Modeling indexing activities for a computer requires the kind of sound theoretical foundation in indexing that only professional indexers and information professionals can provide. The separation of computer scientists working towards content-based indexing from traditional indexing theorists is not ideal. It seems unlikely that mathematical analysis of images as bit-depth digit maps alone will result in useable systems. It is the hybrid approaches which are most promising.

Besser’s proposed hybrid solution of “text-based cataloguing and indexing sufficient to allow the user to narrow a retrieval set to a reasonable size, coupled with some kind of procedure for browsing through the retrieved set of images” (790) anticipates the popular image search interfaces by Google, Yahoo! and MSN. A division into a pyramid of syntactic and semantic levels for visual descriptor terms has produced “consistent and positive results” for image representation and retrieval. (Jörgensen et al. 2001: 945) This approach satisfies Small’s Aristotelian “Principle Number One” for image indexing: “’Do not make your datum more accurate than it is.’ This principle may be rephrased as, ‘Preserve the Mess.’ Preservation of ambiguity, however, does not mean a lack of either organization or controls.” (Small 52)

The reasonable approach is to combine algorithmic approaches for syntactic attributes that can be extracted from images automatically, and human indexing for semantic attributes that require world knowledge. Jörgensen et al.’s pyramid contains the following syntactic elements that lend themselves to automatic (computer-generated) indexing: type/technique (e.g.: black & white, color), spectral sensitivity (color), frequency sensitivity (texture), image components such as dot, line, tone, spatial layout of elements. (Jörgensen et al. 2001: 940) Automated assignment of keywords saves time and money, so it should be done as much as possible. However, humans “mainly use higher level attributes to describe, classify and search for visual material” (Jörgensen et al. 2001: 940) and these semantic attributes cannot at this time be algorithmically assigned. The semantic attributes include: generic objects (e.g.: table, telephone, chair), generic scenes (e.g.: city, landscape, indoor, outdoor, portrait), specific objects (e.g.: Notre Dame Basilica, Bruce Lee), specific scenes (e.g.: Warsaw, Plains of Abraham), and abstract objects (e.g.: guilt, remorse) (Jörgensen et al. 2001: 940).

The popularity of the World Wide Web has emphasized the need to index images in a multimodal environment which lends itself especially to hybrid approaches. The MARIE-3 system, for example, uses Web page layout and word syntax to extract captions of photographs (accompanying text) from the rest of the text on the page (Rowe & Frew 1998). Extracted captions from multimodal documents can be used as an automatic content-based indexing strategy for improving search precision (Srihari et al.). With the addition of human semantic level indexing to the system using a structured set of categories for visual descriptors (see Jörgensen et al. 2001 pyramid) we have the kind of hybrid approach that results in useable image search applications for the web.

## Conclusion

This paper presents visual materials as distinct from the textual in that only the latter satisfy Nelson Goodman’s syntactic requirements for a language. When indexing images with textual descriptors, indexers are in fact *creating* linguistic interpretations of their subjective experience of the images. The end-users are similarly creative when searching for images using written language. The result of this is the need for general theoretical basis for crosswalks from these image-experiences to linguistic descriptions, because as Paula Berinstein points out “If your inquiring mind and that of the cataloger don’t meet, you won’t find the pictures you need.” (85)

The theoretical distinction between the *Of* and *About* of a picture has been used to create structured layers of visual descriptors. Syntactic attributes of images can be indexed algorithmically; however, human information seekers use more abstract terms for searching that require the indexing and image recognition abilities of the human mind. Computer scientists are continually improving algorithms for automatic content-based image analysis and indexing. The use of image captions in multimodal environments as a source of automatically generated index terms has proven to be particularly successful. However, this is nothing more than the *extraction* of human-generated indexing and description from multi-modal environments. End-users search for images using the kind of abstract concepts that require human processing and classification of the images. It is inevitable that in the absence of existing textual descriptions for images, the indexer will have to create these in order to provide access. Solid theoretical background in the types and classes of descriptors for the domain of images will improve inter-indexer consistency, but the interpretation of images will always be a creative process.

**Images:**

· All of the images used in this paper are by the author (Tomasz Neugebauer)

**Works Cited:**

Baxter, G. Anderson, D. "Image indexing and retrieval: some problems and proposed solutions" Internet Research 6.4 (1996). Research Libraries. ProQuest. McGill University Libraries. 1 Mar. 2005 <http://www.proquest.com>.

Berinstrein, P. "Do You See What I See? Image Indexing Principles for the Rest of Us" Online March/April (1999): 85-88. Research Libraries. ProQuest. McGill University Libraries. 1 Mar. 2005 <http://www.proquest.com>.

Besser, H. "Visual Access to Visual Images: The UC Berkley Image Database Project" Library Trends 38.4 (1990): 787-98.

Chu, H. "Research in Image Indexing and Retrieval as Reflected in the Literature." Journal of the American Society for Information Science and Technology 52.12 (2001): 1011-1018. Research Libraries. ProQuest. McGill University Libraries. 1 Mar. 2005 <http://www.proquest.com>.

Goodman, Nelson. Languages of Art : an approach to a theory of symbols. Indianapolis: Hackett Publishing Company, 1976.

Jacobs, C. "If a picture is worth a thousand words, then…." The Indexer 21.3 (1999): 119-121.

Jörgensen, C. “Access to Pictorial Material: A Review of Current Research and Future Prospects.” Computers and Humanities 33 (1999): 293-318. Elsevier Science Direct. McGill University Libraries. 1 Mar. 2005 < http://www.sciencedirect.com/>.

Jörgensen, C., Jaimes A., Benitez, A. B., Chang, S. “A Conceptual Framework and Empirical Research for Classifying Visual Descriptors.” Journal of the American Society for Information Science and Technology 52.11 (2001): 938-947. Research Libraries. ProQuest. McGill University Libraries. 1 Mar. 2005 <http://www.proquest.com>.

Rowe, N., C., Frew, B. “Automatic Caption Localization for Photographs on World Wide Web.” Information Processing & Management 34.1 (1998): 95-107. Elsevier Science Direct. McGill University Libraries. 1 Mar. 2005 < http://www.sciencedirect.com/>.

Shatford Layne, S. "Some issues in the indexing of images." Journal of the American Society for Information Science 45.8 (1994): 583-588. ACM Portal. Google Scholar. 6 Mar. 2005 <http://scholar.google.com>.

Shatford, S. “Analyzing the Subject of a Picture: A Theoretical Approach.” Cataloging & Classification Quarterly 6.3 (1986): 39-62.

Schroeder, K. “Layered indexing of images.” The Indexer. 21.1 (1998):11-15.

Small, J., P. “Retrieving Images Verbally: No More Key Words and Other Heresies.” Library Hi Tech 9.1 (1991): 51-60.

Srihari, R., K., Zhang, Z., Aibing, R. “Intelligent Indexing and Semantic Retrieval of Multimodal Documents.” Information Retrieval 2 (2000): 245-275. Elsevier Science Direct. McGill University Libraries. 1 Mar. 2005 < http://www.sciencedirect.com/>.

[*] Ansel Adams quotation source: The Most Notable Quotations: 1950-1988, Compiled by James B. Simpson. Originally published by Boston: Houghton Mifflin Company, 1988. 7 Mar. 2005. <http://www.bartleby.com/63/3/5803.html>

**Image Indexing**

by: Tomasz Neugebauer

March 2005

in this section:

1 2 3