What do we see in images?

Peter Jörgensen,
University at Buffalo,
SUNY
peter@jorg2.cit.buffalo.edu

ABSTRACT

This paper describes preliminary results from two image classification experiments. Ss created their own ad hoc categories in the first. Pre-existing categories were provided in the second. In both of these tasks objects, especially people, were the predominant feature used to categorize the images.

RÉSUMÉ

Cette communication décrit les résultats préliminaires de deux expériences d'image de classification. *Ss* a crée leurs propres catégories *ad hoc* dans la première expérience. Des catégories préexistantes étaient fournies dans la deuxième expérience. Dans les deux tâches les objets, surtout les gens, étaient la particularité utilisée pour catégoriser les images.

INTRODUCTION

Information retrieval systems rely on indexes for their operation. (Baeza-Yates 1999) Indexing images presents a particularly difficult problem. (Jörgensen 1995) While promising work is being done in the area of non-textual indexing and querying of image databases (see, for instance, (Baxter 1996, Flickner 1995, Jörgensen 1993), little is known about the way images communicate their messages and meanings. This paper reports on one experiment which attempted to bring more clarity to the question of how people categorize images.

One type of *image* is a human-made pattern of light that remains substantially unchanged for a period of time measurable in minutes. This includes such things as photographs, paintings, and drawings as well as electronically produced analogues of these that may appear on a light emitting screen or projected onto a viewing surface. Although naturally occurring scenes could be thought of as *images* they will not be considered here as we are interested in the meanings conveyed by images, and it is arguable that *meaning* can be assigned to phenomena of the natural world. Nor will series of images that simulate motion, as in motion photography or videography be considered as these present an added capacity for information transmission (the ability to directly encode time and movement information) and generally have associated audio. Therefore, an *image*, in this work, shall mean a two-dimensional display that has been produced by human intelligence to be viewed by humans.

An underlying assumption of this definition is the communicative function

that images perform. In as much as there are a number of reasons for human communication, images are created to fulfil one (or more) of these functions. In order to understand how images serve as communicative channels we must understand how meaning is encoded in images.

Meaning can exist on a number of levels: iconic, indexical and symbolic (Peirce 1991); pre-iconographical, iconographical and iconological (Panofsky 1955); sign, picture, symbol (Arnheim 1969) and sign, pictograph, ideogram (Jacob 1996), among others. (Osgood 1960) concluded that there is a high degree of cross-cultural agreement in the assignment of certain attributes to specific concepts (229). On the other hand, Messaris points out that the field of advertising is littered with examples of images that at best were meaningless in another culture and at worst insulting (Messaris 1997).

Central to successful automated meaning extraction from images is the idea that objectively measurable features of a visual stimulus produce uniform ascriptions of meaning across a useful range of image seekers. This idea is supported by the hypothesis that many fundamental pattern recognition capabilities of the human brain have been the result of evolutionary processes, rather than cultural ones. These processes have resulted in neuro-algorithms (Vandervert 1998) which are the basis for much of our perception and interpretation of our environment. This implies that my perception of the "red" of a rose *really is* the same as your perception of the color of that same rose. Furthermore, this is, for all practical purposes, a perception of a wavelength of 620NM, i.e. something that can be mechanically measured. Likewise, Vandervert argues that more than just simple perceptions (like color) are shared by virtue of these evolutionarily derived neuro-algorithms. It could be argued that since the beginning of civilization these neuro-algorithms have been the most important evolutionary force acting on homo sapiens. If this were indeed the case, then it is reasonable to assume that although relatively little time has elapse (on an evolutionary time scale) since humans first organized themselves against the elements a significant amount of cognitive evolution may have taken place.

That different people see different things, i.e. perceive different meanings, from the same picture is explained by Arnheim (1980, 176) by the role of internal *form patterns* which vary not only between individuals, but within an individual at different times, an observation supported by Seloff's (1990) study of indexers. These *form patterns* are what the mind uses to organize the visual stimuli into meaningful concepts. He goes on to observe that this process does have natural limits that "offer resistance" to certain perceptions while offering less to others (177). Goode (1972) found six primary factors responsible for perceived similarity among works of art. These six factors were grouped into three factor pairs: *emotion, motif* and *composition.* Ruth (1974) et al found six factors as well: I Aesthetic-qualitative evaluation, II Emotional tone, III Symbolism, IV Dynamics, V Clarity and VI Stillness.

The advertising industry has, of course, long been concerned with the messages conveyed by images. Among the less concrete message types that images were recognized to convey, Larned lists: "the story of a service performed, its convenience, its profit its utilitarian advantages."; "To provide essential 'atmosphere'... unexpected aristocracy."; to employ a "...skilful play upon emotions." and "To dramatize the undramatic."(1925, 2). Among the useful characteristics of images that are highlighted are composition, angle of perspective, use of "heroic size", glorification, atmosphere, animation (as in bringing inanimate things to life), melodrama, and humor.

Sharp-angled shapes (stars, triangles, etc.) rank high on potency and activity dimensions compared to rounded shapes (circle and ellipse) (Messaris, 61). Shapes can also convey meaning iconographically by association. Two lines forming an acute angle can convey movement and direction (as an arrow) or sharpness or danger (as a knife).

Research done by Jörgensen (1995, 1996, 1997), demonstrates that when users are not constrained by their conceptions of the capabilities and limitations of systems they tend to include interpretative terms (such as those concerned with story or mood) in their descriptions of (non-photographic) images.

Indexing of images has, until recently, been done exclusively by hand. The massive volumes of information now readily available (*e.g.* on the World Wide Web) has created a need to find ways to index images automatically. In 1997 it was estimated that it would cost $400 million and 400 staff-years to adequately index the 3.5 million images in Berkley's Bancroft Library (Collins 1997, 23) The greatest success has been achieved in identifying objects and shapes. For instance, Lu (1997) describes a method to identify and code shapes using chain codes that are independent of the size and orientation of the shapes.

METHODOLOGY

System Description and Development

A computer-based image categorizing and sorting system was created to facilitate the collection of data in the human-subject side of the experiment. The system was created using HyperCard 2.4 running under MacOS 8.5.1 on a Macintosh G3. The system allowed subjects to sort and categorize images.

The system interface consisted of icons (image thumbnails) representing categories displayed on the left side of the screen, and sets of images which the subject would categorize or sort (by dragging onto the icons) displayed on the right side of the screen. The image icons representing categories were presented to the S as thumbnails (70 pixels wide by 60 pixels high) along the left hand side of the screen. The images to be categorized were selected randomly from the database and

presented nine at a time to the S (250 pixels wide by x150 pixels high) on the right hand side of the screen. Visual feedback was provided by highlighting a 2 pixel border around the thumbnail over which the S was currently dragging an image. A rectangle the size of the image being dragged followed the mouse on the screen while the S was dragging it.

At any time during the experiment the S could click on a thumbnail icon to see an enlarged version of the category image. A Question Mark icon was also displayed with the thumbnails to provide a category for images that the S felt didn't belong in any of the other categories.

For both experiments, an undo button was provided which, when clicked, removed the most recently categorized image from the category to it had been assigned and redisplayed it on the screen, allowing the subject to undo a mistake or make an immediate change (the undo only worked for the most recently categorized image).

The system was also programmed to provide automatic data capture and output. At the beginning of each session a new log was created containing the date and time. Each action of assigning an image to a category or creating a category was recorded in the log with a time stamp accurate to the second. When the S was done the log was written to a tab-delimited text file for analysis.

Data

The image database (a total of 542 jpeg images) consisted of two sets of images, some collected from a random sample of images on the World Wide Web (about 2/3 of the database) and a subset of images from the MPEG-7 CD data collection (CDs 6, 7, 8, 9, 12 and 13). The web images consisted of a range pertaining to personal interests (such as personal photos), informational images and some commercial images. Examples of the scenes depicted are airplanes, touch football or other outdoor games, typical travel scenes such as natural formations and landscapes, aerial photographs, photos of people (having a good time), maps, graphs, other scientific images, and promotional materials for movies, books, etc. Other factors entering into choice of images were to provide as wide a range of images as possible within the constraints of decency and suitability for display at a maximum resolution of 250 x 150 pixels.

Subjects

Subjects were recruited from among the student body at the Department of Library and Information Studies (DLIS), School of Information Studies, at the University at Buffalo, and from among acquaintances and friends of the research assistants. The student body at DLIS is quite varied as many students in this program are starting second or third careers. It is largely female but not exclusively. As it is a graduate program, all students have completed an undergraduate degree; these vary but are primarily from the humanities, social sciences, and professional programs, with a

smaller number of science degrees. Some students have other Masters degrees or even Ph.D.s. Students have often previously had extremely varied career experiences. Careers of other subjects were mostly white-collar; some high school students also participated.

Ages of subjects ranged from seventeen up through retirees (70s). There were 11 males and 30 females.

Experiment Logistics

The actual running of the subjects was done by two graduate research assistants from the Department of Library and Information Studies, under the supervision of Dr. Corinne Jörgensen. They were responsible for all scheduling, set-up, and running the experiments. Subjects were allotted one hour to complete the task, including preliminaries such as the consent form and general system instructions. For the consent form, subjects either read, or had the consent form read to them in its entirety. They were asked if they understood what they were signing, and it was emphasized to them that they could choose not to participate, or to stop, or to not have their data included in the study. No one had a problem and everyone fully cooperated. They were then brought them into the room where the computer was and the research assistants explained what they would be doing, and why, and the mechanics of the on-screen graphics were explained. They were told that the research assistants could not answer any questions on what an image was- it was whatever they perceived it to be- but that the RAs would be there to help with technical difficulties, should any arise. Subjects then proceeded with one of the tasks, which were assigned randomly.

Tasks

Two tasks were performed by subjects, a Sorting Task (sorting images into nine experimenter-defined categories) and a Categorizing Task (sorting images into up to 15 categories defined by the Subject during the task plus a sixteenth "throwaway" category). The general procedure consisted of choosing images from the right hand side of the screen to be sorted into categories displayed on the left side of the screen. More detail for each task is provided below. Some subjects performed one or the other task, and some subjects performed both tasks. If a subject performed both tasks, the Categorizing Task was performed first, so as not to influence the formation of categories by seeing our pre-defined categories. For those who did both tasks, the effect of having done the Categorizing Task first was greater familiarity with the images, and the net effect on the process was simply that the sorting took place slightly faster. There appeared to be no other difference in the results.

Categorizing Task

For the categorizing task, S created the categories themselves. In order to facilitate this, S were given a short preview of the image database. Images from the database

were presented to the S in random order at a rate of 1 per second in their natural sizes (from 70x60 up to 640x480) for up to three minutes (some S felt they had seen enough to start the task after approximately two minutes). This allowed the S to get a visual idea of what the database contained.

Next, a screen containing 15 generic file folder icons and a question mark icon was presented on the lefthand side of the screen. After any necessary instruction and orientation to the system the S clicked the start button. Subjects were to create up to fifteen of their own categories (based on previous research (Jörgensen, 1995) in which subjects sorted images into their own categories, the average number of categories created among the subjects was sixteen). As before, there was also a "miscellaneous" Question Mark category, into which unclassified images could be placed. The system then presented three rows of three images each, as in the sort experiment. The S then dragged each image onto either a generic folder icon, indicating that this should be representative of a new category, or onto an existing category's thumbnail. Each time a new category was created by dragging an image onto a generic folder icon the system would ask the S to confirm that s/he wanted to create a new category with that image. If the S answered "Yes" then that image would become that new category's thumbnail. If the S answered "No" then that image would return to the screen. As each image was assigned to a category or used to create a new category it would disappear from the screen. When all nine images were so categorized a new set of nine images would be presented. The process repeated until the same conditions as in the sorting experiment were met, i.e. the subject gave up, assigned at least ten images to each category or had seen half of the database. An undo button was provided as in the sorting experiment.

Sorting Task

For this task, the experimental team selected nine images to represent nine typical categories (based on previous research) and covering a range of semantic types. The categories were not represented by any text or label but simply by the image itself. After preliminaries (consent form and instructions), the S clicked on a button labelled "Start" to start the process of sorting images into these categories.

The system then presented three rows of three images randomly selected from the database. The S then proceeded to drag each image to the category thumbnail to which s/he thought it belonged, or to the Question Mark category. As each image was dropped onto a category thumbnail the image would disappear from the screen leaving the remaining images to be categorized. After the last image was dragged to a category nine more images would appear. The S then sorted these images into the categories.

This process was repeated until one of three conditions was met. First, the subject was allowed to stop at any time if they desired. Second, if all nine categories were assigned at least ten images each, the S was encouraged to stop. Finally, if the

S sorted half of the database s/he was encouraged to stop.

RESULTS

Categorizing Task

Data from twenty Ss were suitable for analysis. The subjects saw and average of 292 images. The Ss created 20 major categories, some of which could be divided into subcategories. Each S created an average of 14.2 categories which is very close to the maximum of 15 which they could create. Ss sometimes created overlapping or even duplicate categories, especially in the more common categories. *People* was the most common category accounting for an average of 2.3 occurrences per S. On average, each S created almost 2 (1.8) *Art* categories and 1.5 *transportation* categories. Other commonly created categories included *Medical/Scientific*, *Buildings*, *Animals*, *Aerial Photos and Maps*, *Disrepair* (i.e abandoned machinery, buildings, ruins, etc.) and *Nature*. Some categories (e.g. *Food*) were only created by one S. And an occasional category was populated by only one image. The *People* category was often more precisely specified as *Active*, *Groups*, *Individuals*, or *Couples*, in that order. Similarly, the *Art* category appeared often as *Graphic Art*, *Fine Art*, or *Cartoon*. The mean number of images that were not assigned to any category was 24.8. 265 of the images were assigned to the *other* category by one or more subjects. An analysis of the voice recordings made during and after the trials revealed some categories that could not have been determined solely by examination of the data files. One such example was a category of "shape, if shape struck me." See table 1 for a summary of these results.

Sorting Task

In the following discussion the categories will be referred to by the name of the image which was used to "label" the category. For example, The category exemplified by the photo of a flying airplane (image p234.jpg) will be referred to as category p234. The category images can be viewed on the Web. Contact the author for more information.

Category	Number Produced	Category	Number Produced
People	51	**Aerial photos**	19
active	10	B&W recon	1
groups	8	maps	2
People	7	**Disrepair**	14
individuals	6	Stone ruins	1
couples	3	Shacks	1
Famous	1	**Nature**	14
Art	40	Landscape	6
Graphic Art	12	CloseUps	2
Fine Art	5	Leaves	3
Cartoons	3	Water	7
	3	from air	1
Transportation	34	**Objects**	9
airplanes	11	**Historical**	4
Ship	5	**Black & White**	1
cars	2	**News**	3
spacecraft	1	**Technology**	3
Medical/Scientific	27	**Food**	1
Scientific	10	**Misc**	9
Medical	10		3
Engineering	3	**Color**	5
Astronomy	2	**Travel**	2
Building	23	**Advertising**	2
historical	3		
Animals	22		
specific	3		

Figure 1
Summary of Categories

Data from 19 subjects was used in the analysis of the sorting task. Each image was seen by an average of 8.9 Ss (SD 2.122). Category p234 had the most images (22) uniquely assigned to it. These 22 images made up 30.0% of the (181) images that were assigned to this category. 30% of category P083's images were also uniquely assigned to it. Category c19 had a unique assignment of 25.9% and category p009 contained 14.9% unique images. The rest of the categories had relatively few unique images assigned to them. With the exception of categories p009, p083 and the "other" category about one third of all images presented were assigned to each category at least once. 286 of the images were assigned to the *other* category. See table 2 for a summary of these results.

Category	p234	C19	p060	p009	p083	p094	L31	p137	p230	other
Total # of images	542	542	542	542	542	542	542	542	542	542
Total in category	181	178	170	70	57	188	128	141	156	286
# Unique to this cat.	22	15	14	6	5	5	4	1	1	0
% Unique to this cat.	30.0 %	25.9 %	14.9 %	24.1 %	30.0 %	4.8%	9.4%	3.6%	3.0%	0.0%
% Assigned to this cat.	34.2 %	33.6 %	32.1 %	13.2 %	10.8 %	35.5 %	24.2 %	26.7 %	29.5 %	54.1 %

Figure 2
Summary of Sort Task

DISCUSSION

Categorization Task

Although a good deal of the categorization that was done could be called idiosyncratic, there was considerable agreement in some areas. These include *People*, *Art*, *Transportation*, and *Medical/scientific*. Additionally, this study bears out the observation made by others (Jörgensen, 1995) that people focus primarily on objects. The lack of affective response to these images was probably due, in large part, to the nature of the images themselves (very few of the images can be said to evoke strong feelings) as well as the nature of the instructions provided. It is clear that people have little trouble creating ad hoc categorization schemes. Less than 10% of the images were assigned to the *other* category by the average S. Some Ss expressed a desire to have the ability to create more categories.

Sort Task

The predefined categories resulted in greater agreement between Ss and also resulted in less images being assigned to the *other* category. The categorizations were highly object-driven. The most easily identified objects, i.e. airplanes, were most often grouped together. The least familiar image (p137 - an elliptical shape filled with irregular blue, red and black areas, possibly the ozone hole over Antarctica) attracted the least consistent members into its category. The image carrying possibly the greatest affective meaning (p083 - a "greeting card" style color drawing of two bears in an armchair) ranked high in the percentage of images uniquely assigned to its category (30%) and had the fewest images, over all, assigned to it (57).

This experiment provided little opportunity, in the way of appropriate images, to investigate affective communication by images. It did, however, support earlier findings that objects, especially familiar or easily identified ones, are of primary interest to people in image classification and sorting tasks.

REFERENCES

Arnheim, R. 1969. *Visual Thinking*: University of California Press.

Arnheim, Rudolf. 1980. Dynamics and Invariants. In *Perceiving Artworks*, ed. John Fisher, 3:166-184. Philadelphia: Temple University Press.

Baeza-Yates, R. and Berthier de Araujo Neto Ribeiro. 1999.
Modern information retrieval Ricardo Baeza-Yates, Berthier Ribeiro-Neto. Reading, Mass.: Addison-Wesley Longman.

Baxter, Graeme and Douglas Anderson. 1996. Image indexing and retrieval: some problems and propsed solutions. *Internet Research: Electronic Networking Applications and Policy* 6, no. 4: 67-76.

Collins, Karen. 1997. Providing Subject Access to Images: A Study of Users Queries. Masters, University of North Carolina.

Flickner, Myron, Hrpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. 1995. Query by Image and Video Content: The QBIC System. *Computer* : 23-31.

Goude, Gunnar. 1972. A multidimensional scaling approach to the perception of art.II. *Scand. J. Psychology* 13: 272-284.

Jacob, Elin K. and Debora Shaw. 1996. Is a Picture Worth a Thousand Words? Classification and Graphic Symbol Systems. In *Knowledge Organization and Change*, ed. Rebecca Green, 5:174-181. Frankfurt/Main: Index-Verlag.

Jörgensen, Corinne, Peter Jörgensen, and Matthew Hogan. 1993. The Visual Thesaurus: A Practical Application. In *Sixth International Conference of the Museum Documentation Association*, 223-228. Cambridge, England.

Jörgensen, Corinne. 1995. Image Attributes: An Investigation. PhD, Syracuse University.

Jörgensen, Corinne. 1996. The Applicability of Selected Classification Systems to Image Attributes. In *Knowledge Organization and Change*, ed. Rebecca Green, 5:189-197. Frankfurt/Main: Indekx-Verlag.

Jörgensen, Corinne. 1997. How People Describe Images: Continuing Research. In *ASIS '97*, ed. Candy Schwartz and Mark Rorvig, 34:375. Washington, DC: American Society for Information Science.

Larned, William Livingston. 1925. *Illustration in advertising*. New York: McGraw-Hill book company inc.

Lu, Guojun. 1997. An approach to image retrieval based on shape. *Journal of Information Science* 23, no. 2: 119-127.

Messaris, Paul. 1997. *Visual persuasion : the role of images in advertising*. Thousand Oaks,Calif.: Sage Publications.

Osgood, Charles E. 1960. The Cross-Cultural Generality of Visual-Verbal Synesthetic Tendencies. In *Language, Meaning, and Culture: The selected papers of C. E. Osgood*, ed. Charles E. Osgood and Oliver C. S. Tzeng, 1990:203-234. New York: Praeger.

Panofsky, Erwin. 1955. *Meaning in the visual arts*. Garden City N.Y.: Doubleday.

Peirce, Charles S. and James Hoopes. 1991. *Peirce on signs : writings on semiotic*. Chapel Hill: University of North Carolina Press.

Ruth, Jan-erik and Kyösti Kolehmainen. 1974. Classification of art into style periods; a factor analytical approach. *Scand. J. Psychology* 15: 322-327.

Seloff. 1990. Automated access to the NASA image archives,. *Library Trends* .

Vandervert, Larry R. 1998. How A Priori Image-Schematic, Simulative Neuro-Algorithms provide Us with Mental Universals which Parallel Physical World Principles. In *Systems Theories and A Prior Aspects of Perception*, ed. J. Scott Jordan, 126:259-287. Amsterdam: Elsevier.