Query terms for art images: A comparison of specialist and layperson terminology

Daniel Iseemann
Dublin University
Trinity College
Dublin, Ireland
isemandi@scss.tcd.ie

Khurshid Ahmad
Dublin University
Trinity College
Dublin, Ireland
kahmad@scss.tcd.ie

Museum and art-gallery curators have long had specialist collection management systems to facilitate access to works of art that were not available to the general public. The increasing availability of aesthetic images on the World Wide Web, however, especially through museum and gallery web pages, went hand in hand with an increasing sophistication of search paradigms. Retrieval methods have, in a literal and metaphorical sense, become multi-faceted. Art image retrieval can, depending on the retrieval task, draw to a greater or lesser extent on a rich body of expert knowledge from the domain of art history. We have conducted an exploratory study analysing the use of query terms for the same set of target images among a group of art specialists and laypersons. The survey suggests systematic and identifiable similarities and differences between the two groups. We argue that this finding needs to be considered in the design of image retrieval systems that are tailored to art collections.

image query, fine art image retrieval, expert, specialist, layperson

1. INTRODUCTION

The recently launched ART PROJECT (compare <http://www.googleartproject.com>) allows for the virtual exploration of 17 art galleries and other places of heritage interest. It might be just the latest step in a line of development that has seen an ever widening public gain access to works of art that years ago would have been available only to the select few. While the ambulatory mode of exploration provided by ART PROJECT creates a novel experience for the user the main mode of digital image access remains keyword based, sometimes mediated through search templates. Indeed some of the galleries involved in ART PROJECT already provide digital access to their collections. The TATE COLLECTION and the STATE HERMITAGE MUSEUM have search interfaces that allow for multi-faceted search including subject search (compare <http://www.tate.org.uk> and <http://www.heritagemuseum.org>). The number of works of art that are accessible in this way is steadily rising. The HERMITAGE online collection, for example, is growing at an average speed of about five artifacts a day.¹

¹Calculated for the period from 4th of September 2008 to 22nd of July 2010 using the Advanced Search tool of the English language portal of the State Hermitage Digital Collection.

Unlike general purpose image retrieval, art and cultural artefacts present a special case in which a rich and traditional body of expert knowledge has been created. This needs to be utilised to meet the ever increasing interest and demand from novice users. With the development of an ontology-based prototype for semantic retrieval on richly marked-up collections of art images in mind we wanted to study user preferences in the currently predominant keyword-based paradigm. In this context we have been exploring the terminology deployed to retrieve digital copies of paintings by two groups — people specialising in art history and typical end-users such as college students of other specialisations. In a survey which looked at the use of query terms by the two groups we found systematic differences between specialists and laypersons and classified them according to an existing category system.

2. RELATED WORK

It has been said that the client queries submitted to a particular photographic archive in the 1980's often "fell into a 'no-man's land of categories'" (Enser 2008, p. 534). Since then several studies on image queries and image descriptions have tried to chart this *no-man's land*, often as part of the iterative development cycle of image retrieval systems.

Hollink et al. (2004) have developed a system of image descriptor classification that synthesises a number of other classifications, specifically a distinction between primitive, logical and abstract image features (Eakins (2002)), a ten-level model for image indexing (Jaimes and Chang (2000)) and a classification matrix based on Erwin Panofsky's work in iconography (Shatford (1986)).

The resulting unified model maintains a degree of 'backward' compatibility with the original systems.

There are similar studies in the domain of art images: Hastings (1995), Chen (2001). The latter of these has a sample of art history students and uses the categories employed in Enser and McGregor (1992), Fidel (1997) and Jörgensen (1998). At a more practical level, Chen (2007) compared the image needs and user behaviour of art historians and experts in related fields.

Fidel (1997) analysed 100 image requests using a modified version of Jörgensen's attribute classes (Jörgensen (1998)). She describes image retrieval tasks as being positioned on a continuum between two extreme poles, the *Data Pole* and the *Object Pole*: "At the Data Pole, images are used as sources of information, while at the Objects Pole, images are needed as objects" (Fidel 1997, p. 189). An example of retrieval tasks near the Data Pole are maps of a certain geographical area or medical slides of certain aspects of human anatomy. Examples of retrieval at the Objects Pole are requests for "a very specific kind of image of a person or event, or for any image that represents a specific idea or object" (ibid.).

Smeulders et al. (2000) classify user aims in image retrieval into three broad categories: *target search*, *category search* and *search by association*. In target search the user wants to retrieve "a precise copy of the image in mind" (ibid., p. 1351), in category search the user is looking for an arbitrary image from a certain class of acceptable images and in search by association the user starts out with no specific aim in mind.

Most image retrieval studies vary in their experimental set-up, for instance in the task set to participants, and also in the system they employ for classifying image descriptors. In cases where a limited comparison between different studies is possible some of the results in the literature seem to be contradictory ((Hollink et al. 2004, p. 624f.), (Enser 2008, p. 534)). In our opinion little research has been done which allows for a direct comparison of lay and expert conceptualisations of aesthetic images.

3. SURVEY DESIGN

We have attempted to understand user behaviour during retrieval through their choice of keywords. For this we have conducted a survey to investigate query terms used by a group comprising art experts, art history students and laypersons for the same set of digital images of paintings. We have used the unified image descriptor classification developed by Hollink et al. (2004).

The survey was taken by 48 participants. At the time of the experiment 14 participants had received university level education in art history for at least one academic year. Another seven participants had already completed such an education. Two in the latter group held a Masters degree and one a Ph.D. in art history. These 21 participants, who had either already acquired a certain expertise in art history or chosen a specialisation that might eventually lead them to such expertise, are subsequently referred to as 'specialists'. All of the 27 'non-specialists' had completed secondary school, six were holding a first degree from a third level institution, six had a postgraduate degree and seven had a doctoral degree.

The setting was relatively straightforward. Participants were asked to give information on their experience with web search engines in general and image and art image search in particular. This was followed by a section in which participants were shown digital images of three paintings.

The three paintings were *Hercules and the Hydra* by Antonio del Pollaiuolo, *The Prophet* by Egon Schiele and *Road to Louveciennes* by Claude Monet (Figure 1). The images were selected to represent some of the diversity found in paintings in the Western tradition. They range from the 15th to the 20th centuries, from figurative to landscape art, exhibiting some of the characteristics of Renaissance, Expressionist and Impressionist art respectively.

Underneath each image a written instruction to participants read:

Please look at this image for a few seconds and then move on by clicking on the "Next" button.

While participants were asked to dwell on the image only for a "few seconds" it was ultimately their own decision to move on to the next screen and some of them might have examined the images for longer. Participants were subsequently prompted to enter keywords into a free-form text box by the following questions:

Figure 1: Images of paintings by Pollaiuolo, Schiele and Monet in the order in which they were shown to participants. After they were shown each image, participants were asked “What keywords would you use if you were looking for this image online”?

What keywords would you use if you were looking for this image online?

The retrieval task we set to participants falls into Smeulders et al.’s category of *target search*. Moreover, the retrieval task is located at Fidel’s so-called *Objects Pole*, because the image itself is the object of interest and not, for example, information contained within the image. In a real life scenario users might of course go on to extract information from such an image after they have retrieved it (e.g. information on the nature of Monet’s brush strokes). It is for this reason, that Fidel names the retrieval patterns of art historians as an example of retrieval that is typically situated between the *Data* and the *Objects Poles*. However, she emphasises the object nature of the initial retrieval: “to make [an] inference, [the art historian] wants to retrieve all images, all objects, and each image must be viewed as a whole, as an object” (Fidel 1997, p. 190).

We believe that target search and retrieval at or near the Objects Pole is relatively common in the domain of fine art images.

The survey concluded with a section in which participants had to provide demographic information on themselves.

4. ANALYSIS

The age of the 48 participants ranged from 18 to 60 years. The average age was 27.2 years, our age median was 24. At 25.9 years on average, our specialist group was slightly younger, than our non-specialists (mean age 28.2). One-third of our participants were male, two-thirds were female. Most of our participants were Irish (68.8%), followed by

Dutch (8.3%), German (6.3%) and participants of seven other nationalities (16.7%). Four participants were from a non-European background.

For each of the three paintings, Figure 2 shows two example sets of search terms that were submitted by participants. The punctuation mark, comma (,) between key words and phrases in Figure 2 represents incisions made by participants themselves in the form of punctuation marks or line breaks.

Pollaiuolo

Hercules, Hydra, Eurystheus twelve labours
Man, battle, lion

Schiele

Egon Schiele, Austrian Expressionism, psychological angst in art, the nude in art, figure drawing
Man, woman black and white, monkey

Monet

impressionist, evening, road, french country town, perspective, vanishing point
Landscape, street scene, Village, Sunset, Sunrise, Snow

Figure 2: Examples for query terms submitted by participants

We classified query terms into one of three broad categories: *nonvisual*, *perceptual* or *conceptual*. Nonvisual refers to image-external information, for instance, title and creator. The perceptual category covers physical elements of the image, for example, colour and texture. The conceptual category covers the visual elements that we see in an image due to our knowledge of the world and our experience: objects, individuals or even subjective or emotional

associations. The subcategorisation of the collected terms and phrases followed Hollink et al. (2004). The nonvisual level is comprised of twelve subcategories that form a subset of the VRA Core Categories (version 3.0, compare <http://www.vraweb.org/projects/vracore3>). These categories are *Creator*, *Culture*, *Date*, *ID Number*, *Location*, *Material*, *Measurements*, *Relation* (to other works), *Rights*, *Title*, *Source*, and *Style/Period*. The perceptual and conceptual levels are not subdivided into a flat list. Rather their subcategories are partially overlapping, lying along conceptually orthogonal axes. Tables 2 and 3 provide an overview of the subcategories of the perceptual and conceptual levels.

5. RESULTS

Two thirds of the query terms provided were at the conceptual level, one quarter referred to nonvisual properties and less than one in ten to perceptual properties of an image (Table 1).

Level	Count	%
Nonvisual	127	25.0
Perceptual	48	9.4
Conceptual	334	65.6
Total	509	100.0

Table 1: Distribution of image query terms over the three main category levels.

Nonvisual level. There are 12 categories of nonvisual descriptors in Hollink et al. (2004) but the respondents' selection of such terms was dominated by only two categories: *Style/Period* queries make up over half (50.4%); *Creator* accounts for a third (33.1%).

Perceptual level. The distribution of query terms over the perceptual level shows that half of the perceptual descriptors were queries for *Type/Technique* and almost 40 per cent were colour related queries. Next to none of the query terms referred to individual objects or regions in the image (see Table 2).

Conceptual level. The distribution of query terms over the conceptual level suggests over 80 per cent of descriptors were of a general nature, just over 10 per cent specific and only very few abstract. More than half of the conceptual queries made reference to an individual object in the image, just under half to the scene as a whole (see Table 3).

We have analysed the use of terms by specialists and laypersons. The striking difference between specialists and laypersons, in terms of the overall distribution over the three top-level categories, was that art historians tend to use more nonvisual terms,

laypersons by comparison more conceptual ones. The frequency of perceptual descriptions was nearly the same in both groups (see Table 4).

	Main levels			Total
	Nonvisual	Perceptual	Conceptual	
Specialist	31.8	9.1	59.1	100.0
Layperson	18.7	9.7	71.5	100.0
Overall	25.0	9.4	65.6	100.0

Table 4: Comparison of the distribution of specialist and non-specialist descriptors over the three most general categories (figures are in per cent). Bold figures indicate a statistically significant difference between the proportions of specialist and layperson terms in a given category.

The distribution within the nonvisual level was very similar with both groups showing percentages for *Style/Period* (49.4% vs. 52%) and *Creator* (33.8% vs. 32%) very close to the overall values. There were differences in the *Material* category (more used by specialists) and the *Relation* category (i.e. relation to other works) which was exclusively used by laypersons. Our observations at the perceptual level may suggest that that laypersons talk more about colour and specialists query more for type or technique related information. Finally, conceptual descriptors are similarly distributed in terms of the object/scene distinction (51.8% object, 48.3% scene, for specialists vs. 56.0% and 44.0% for non-specialists). Perhaps the noteworthy difference at the conceptual level is that art historians seem to use abstract descriptions (9.8%) more often than laypersons do (2.1%) (see Table 5).

We have subjected the observed differences between specialists and laypersons to statistical significance testing. For each pair of proportions of terms supplied, at each of the category levels by the two groups, we have tested the null hypothesis that the observed differences between these proportions are only due to chance variation (two-tailed t-test, significance level 0.05). We found the differences at the nonvisual level and at the conceptual level to be statistically significant. Within the conceptual level, we found the difference between the proportions of abstract descriptors to be statistically significant. These value pairs are printed in bold in Tables 4 and 5 respectively. For all other value pairs the null hypothesis (that the recorded differences stem from chance variation only) could not be rejected. This seems particularly surprising at the perceptual level where some of the differences appear to be stark (31.8% vs. 46.2% colour, 4.5% vs. 15.4% other descriptors). The reason for this lies in the fact that few perceptual terms were used by the participants

Scope	Object		Scene		Total	
Characteristic	Count	%	Count	%	Count	%
Colour	2	4.2	17	35.4	19	39.6
Shape	0	0.0	0	0.0	0	0.0
Texture	0	0.0	1	2.1	1	2.1
Composition	0	0.0	4	8.3	4	8.3
Type/technique	0	0.0	24	50.0	24	50.0
Subtotal	2	4.2	46	95.8	48	100.0

Table 2: Occurrences of perceptual categories in absolute numbers and percentages.

Level		General		Specific		Abstract		Total	
Scope	Characteristic	Count	%	Count	%	Count	%	Count	%
Object	Event	7	2.1	0	0.0	0	0.0	7	2.1
	Place	9	2.7	0	0.0	0	0.0	9	2.7
	Time	0	0.0	0	0.0	0	0.0	0	0.0
	Relation	1	0.3	0	0.0	0	0.0	1	0.3
	Uncharacterized	134	40.1	28	8.4	2	0.6	164	49.1
Subtotal		151	45.2	28	8.4	2	0.6	181	54.2
Scene	Event	20	6.0	5	1.5	1	0.3	26	7.8
	Place	49	14.7	2	0.6	1	0.3	52	15.6
	Time	40	12.0	0	0.0	1	0.3	41	12.3
	Relation	0	0.0	0	0.0	0	0.0	0	0.0
	Uncharacterized	18	5.4	3	0.9	13	3.9	34	10.2
Subtotal		127	38.0	10	3.0	16	4.8	153	45.8
Total		278	83.2	38	11.4	18	5.4	334	100.0

Table 3: Occurrences of conceptual categories in absolute numbers and percentages.

of our study. The small sample size of perceptual terms therefore limits the power of the statistical test.

6. DISCUSSION AND CONCLUSION

Comparing our results to those reported in Hollink et al. (2004) we found some similarities but also some marked differences. The distribution of conceptual descriptors over the three abstraction levels — general, specific and abstract — is similar. However our study showed a higher count of general descriptors at the expense of abstract ones. The ratio of conceptual descriptors to perceptual descriptors is nearly the same (8.3 vs. 7.3). Their respective shares in the total number of descriptors do however differ. This is due to the fact that nonvisual descriptors were much more frequent in our study than in Hollink et al. (2004) (27% vs. 0.9%). It does not seem unlikely that nonvisual descriptors play an important role in searching for art images.

Another big difference between our study and Hollink's was that perceptual descriptors supplied by our participants almost exclusively referred to the scene as a whole (97.5%) and hardly ever to individual objects (2.5%). The ratio was much more

balanced in Hollink et al. (2004) (54.4% *Scene*, 45.6% *Object*).

Our findings suggest categories which are of importance for art image retrieval and perhaps for semantic and ontological characterisations of works of art. At the nonvisual level the artist and the style or period of a work feature most prominently. We have made suggestions elsewhere on how the description of style and art historic period can be enriched and integrated with date records (Isemann and Ahmad (2009)).

Like Jörgensen (1998) and Hollink et al. (2004), we found that the conceptual level is by far the most frequently used. Our study would suggest that apart from detailed subject catalogues, geographic and temporal descriptions of the depicted scene and a systematic coverage of depicted events could be useful for improving image retrieval of art.

Perceptual queries by contrast were relatively rare. In light of the difficulties that content-based image retrieval has faced (compare Enser (2008), especially p. 537) this is perhaps not surprising, for it is at this level that a content-based approach would be expected to make the biggest impact.

Nonvisual				
	Style/Period	Creator	Other	Total
Specialist	49.4	33.8	16.9	100.0
Layperson	52.0	32.0	16.0	100.0
Overall	50.4	33.1	16.5	100.0

Perceptual				
	Type/Technique	Colour	Other	Total
Specialist	63.6	31.8	4.5	100.0
Layperson	38.5	46.2	15.4	100.0
Overall	50.0	39.6	10.4	100.0

Conceptual				
	General	Specific	Abstract	Total
Specialist	79.7	10.5	9.8	100.0
Layperson	85.9	12.0	2.1	100.0
Overall	83.2	11.4	5.4	100.0

Table 5: Comparison of the distributions of specialist and non-specialist descriptors over subcategories of the nonvisual, perceptual and conceptual levels (figures are in per cent). Bold figures indicate a statistically significant difference between the proportions of specialist and layperson terms in a given category.

Finally, similarities notwithstanding, our study suggests that art historians and laypersons emphasise different categories when querying for paintings. Laypersons focus more on conceptual information, while art historians use more descriptors, that are not directly visible in a painting and also more abstract categories.

These findings have influenced the design of a first prototype of an ontology-backed image retrieval system we have developed. The relative importance of the style/period category was reflected in the fact that our system allows users to specify temporal, geographic, stylistic and historic information (participation in historic events) independently or in conjunction with each other. Similarly rich biographical information on the creator, i.e. the artist, can be specified in our system. Preliminary experiments have shown that despite the considerably increased complexity of queries compared to conventional retrieval, both expert and lay users found the system intuitive and easy to use.

The differences we found between specialists and laypersons might be useful in personalising image retrieval systems, for example as far as the ordering or focus of a faceted search interface is concerned. Lay users might prefer a stronger focus on content or subject search, specialists on the other hand may

favour more refined options for specifying metadata categories.

7. REFERENCES

- Chen, H.-L. (2001) 'An analysis of image queries in the field of art history', *Journal of the Am. Soc. for Information Science and Technology*, vol. 52, no. 3, pp. 260-273.
- Chen, H.-L. (2007) 'An socio-technical perspective of museum practitioners' image using behaviour', *Electronic Library*, vol. 25, no. 1, pp. 18-35.
- Eakins, J. P. (2002) 'Towards intelligent image retrieval', *Pattern Recognition*, vol. 35, no. 1, pp. 3-14.
- Enser, P. (2008) 'The evolution of visual information retrieval', *Journal of Information Science*, vol. 34, no. 4, pp. 531-546.
- Enser, P. and McGregor, C. (1992) *Analysis of visual information retrieval queries*, British Library Research and Development Report 6104, British Library.
- Fidel, R. (1997) 'The image retrieval task: implications for the design and evaluation of image databases', *New Review of Hypermedia and Multimedia*, vol. 3, no. 1, pp. 181-199.
- Hastings, S. (1995) 'Query categories in a study of intellectual access to digitized art images', *Proceedings of the 58th ASIS Annual Meeting, Chicago*.
- Hollink, L., Schreiber, A. et al. (2004) 'Classification of user image descriptions', *International Journal of Human-Computer Studies*, vol. 61, no. 5, pp. 601-621.
- Isemann, D. and Ahmad, K. (2009) 'Navigating cultural heritage in style. Sketching an ontological representation of metadata: the example of artistic period and style', *Museum Ireland*, vol. 19, 149-155.
- Jaimes, A. and Chang, S.-F. (2000) 'A conceptual framework for indexing visual information at multiple levels', *IS&T/SPIE Internet Imaging 2000*, San Jose, CA.
- Jørgensen, C. (1998) 'Attributes of images in describing tasks', *Information Processing & Management*, vol. 34, no. 2/3, pp. 161-174.
- Smeulders, A. W. and Worring, M. et al. (2000) 'Content-Based Image Retrieval at the End of the Early Years', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380.
- Shatford, S. (1986) 'Analysing the subject of a picture: a theoretical approach', *Cataloging and Classification Quarterly*, vol. 6, no. 3, pp. 39-62.