Originally published in:
Proceedings of the Tenth International Conference on Information Visualisation (IV'06), London, England, 05-07 July 2006 / E. Banissi et al. (eds.), pp. 74-79

Using the Amazon Metric to Construct an Image Database based on what people do, not what they say.

Theodor G Wyeld, Robert M Colomb
School of Information Technology and Electrical Engineering, The University of Queensland, Queensland 4072 Australia
{twyeld@itee.uq.edu.au, colomb@itee.uq.edu.au}

Abstract

Current image database metadata schemas require users to adopt a specific text-based vocabulary. Text-based metadata is good for searching but not for browsing. Existing image-based search facilities, on the other hand, are highly specialised and so suffer similar problems. Wexelblat's semantic dimensional spatial visualisation schemas go some way towards addressing this problem by making both searching and browsing more accessible to the user in a single interface. But the question of how and what initial metadata to enter a database remains.

Different people see different things in an image and will organise a collection in equally diverse ways. However, we can find some similarity across groups of users regardless of their reasoning. For example, a search on Amozon.com returns other products also, based on an averaging of how users navigate the database. In this paper we report on applying this concept to a set of images for which we have visualised them using traditional methods and the Amazon.com method. We report on the findings of this comparative investigation in a case study setting involving a group of randomly selected participants. We conclude with the recommendation that in combination, the traditional and averaging methods would provide an enhancement to current database visualisation, searching, and browsing facilities.

1. Introduction

Current image database metadata descriptors include: size, file format, bit-depth, pixel ratio and so on. Such descriptions have easily determined parameters. Content is often more difficult to describe. Metadata associated with the content of an image database is often described in equally objective terms. If it is a photograph of a sunset it is called a sunset. Other metadata may be added such as clouds, what type, how many, location, who it was taken by, colours, identifiable objects in the photograph and so on. This metadata may follow one of the accepted metadata schemas such as the Dublin Core standard or the MPEG-7 multimedia standard, among others [3, 1].

While the different types of descriptors offered by the various metadata standards are easily identifiable they may not necessarily help users find a particular image in an image database. Indeed, images can be categorised in as many different ways as there are people to do the categorising. Not everyone sees the same thing in an image. For example, an image may evoke 'feelings' that are not readily applied as metadata to an object. However, when the user tries to retrieve a particular image from the database their feelings may be the only thing they remember about it.

Descriptors are usually textual. Hence, searching of databases is via an abstract textual query on a textual surrogate for the image, sometimes called a caption. The problem with a textual surrogate is how best to describe the object's contents [2]? If we ask a group of people to arrange a collection of images they will invariably organise them in different ways for different reasons. They may even give clear rationales for why they arranged them in a particular order.

Wexelblat [4] attempts to address the problem of visualization based on textual descriptors by identifying the elemental components of a semantic space with semantic dimensions of two broad kinds, absolute and relative. An absolute semantic dimension can be represented as an axis in a visualisation. The axes organise a system of places into which objects may be put. Relative semantic dimensions are either systems of places that are idiosyncratic subdivisions of particular larger places or, more interestingly, associations among objects not associated with place at all. The example given in Wexelblat [4] Figure 9.3 shows a collection of courses arranged in a space with two absolute dimensions, quarter and professor, with in addition a prerequisite structure shown. The prerequisite structure is a relative semantic dimension associating objects rather than places. One course is a prerequisite of another even if they are both moved in quarter and professor.

The problem remains, however, how best to represent the way a diverse group of people arrange a collection of images for their own idiosyncratic reasons yet make this information available to other users? This paper reports on a comparative study conducted to address this dilemma. It is organised into three main sections. The first discusses the background problem and the notion of displaying averaged user preferences in a dataset (Amazon.com). It then outlines a specific problem where we can apply an averaging algorithm. As a control, we propose initially a traditionally constructed image database and visualisation schema. The second

Proceedings of the Information Visualization (IV'06)

section outlines the case study used. It includes how the test was conducted, analysis of the interviews, and an overview of the results of the interviews. We then apply the averaging algorithm to the results. The third section visualises the averaging results. In the fourth section we compare the visualised averaging results with their visualisation using the traditional method we started with. This paper concludes with a brief discussion on the outcomes of this system and what can be gleaned from it that is not available in the traditional system alone including future directions, such as scalability and the efficacy of a combined system..

1.1 The Amazon Metric

Often the similarities we find in what is returned by users of an image database or other type of database differs greatly from user to user in the way they make their initial selections. Nevertheless, when we compare all collections or the groupings they make they seem to follow some logical ordering independent of the reasonings behind it. For example, when we search for a book to purchase on Amazon.com what is returned also is a small select range of books that Amazon.com suggest "customers who searched for [product title] ultimately chose:" also these others. We do not know why they chose the other books but it usually comes as little surprise that the secondary choices are indeed ones that we may be interested in too. They do seem to fit some sort of logical association. Although we have no idea what this is, other than perhaps there are like-minded people who enjoy reading the same or similar material to what we do. We will call this the Amazon.com Metric (AM). It suggests an algorithm that tracks how people search books and which ones are most often associated with others.

We can draw an analogy between the way the Amazon.com online book ordering system operates and the way someone may ask one what their Zodiac sign is, and when one tells them they say "aha, I knew it!" It is the apparent predictability in both the Amazon.com returned book list and the fortune teller's tale that indicates some tangible link between what we see or hear and how impossible this would have been to have known beforehand that evokes a sense of enigmatic intrigue. They both make an association where none objectively exists. What the fortune teller proposes, and the AM suggests, is an association not an *a priori* fact but a *post facto* interpretation of available information. Their *post facto* prediction is based on results that are interpretable, not really similar to what was know before in any objective way. In the case of the AM, it returned similar products based on how most people interact with the search engine. It is the intrigue of why what is returned appears similar that suggests they are – not any objective similarity. This raises the question: how could we apply this to an image database?

1.2 The Problem

To investigate the potential for the AM to help organise images for constructing an image database, we used a standing problem involving a collection of images based around a common theme – fence spikes. The problem was one which required a method for determining similarity between two or more images. One of the authors of this paper had been taking pictures of fence spikes over a number of years at different locations (see sample set below, figure 1). After awhile he noticed that it was getting increasingly difficult to identify whether he had seen a particular fence spike before or not. Indeed, with the many hundreds of images he had how could he quickly sort them to find an image that was similar or even the same? If he could do this it would save him the time of re-shooting the same fence spike. The challenge was to find a nearness algorithm for sorting the images.

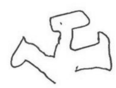
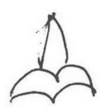
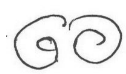
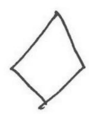
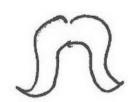
As discussed earlier, the construction of a traditional image database ordinarily starts by applying an expert knowledge of the images' content and context to the metadata that is thought to be most appropriate for one's needs. However, as our images are of an ambiguous nature (how many people take much notice of a fence spike?) we felt that while such an image database might satisfy our needs it may not be useful to others. We needed to make an image database that could be referenced by a wide variety of people with a potentially equally wide variety of ways of sorting and referencing its objects. To do this we needed to find how the way people organise collections of images compares with what they say about how they do it. By comparing what people say about how they arrange the images we are adding the reason behind the AM to the returned suggested book list.

1.3 Traditional Method

As a control study, we constructed a traditional image database for later comparison. The traditional method for assigning metadata to images when constructing an image database is to apply objective features and content descriptions. For this exercise, we concentrated on the content features of the objects depicted in our collection of 60 images of different fence spikes. We detected certain features which we could say were common to many. We can find at least 14 different features, characteristics, aspects, or attributes of the spikes in the images. We can organise these features into a table (see table 1).

Figure 1. Author's collection of sixty images used in the study.

Table 1. The fourteen main different features of the sixty fence spikes.

						
Spiky	Wavy	3-Ends	Leafy	Arrow	Ball	Blunt
						
Bottle	Cross	Curly	Diamond	Flowery	Moustache	Plain

Our visualisation of the traditional database is based on textual descriptions alone, created by an expert (see figure 4). The question remains: how might a diverse range of users organise the same image database? To address this question we asked 20 people to arrange the images in order of most to least similar and asked them why they arranged them the way they did, and finally compared this to our structured textual ordering.

2. The Case Study

2.1 Setting

None of the 20 participants involved in this exercise had ever seen the images before. They came from diverse backgrounds – international, interstate, and local students, researchers, professional people, and a home maker. Their ages ranged from 24-64. There were four females and sixteen males. For many, English was a second language.

The sixty images were printed onto 2x3 cm cards. The sessions were conducted in an isolated meeting room with a large table (4x2 metres). Only one participant and the researcher were in the session room at the same time. The test took an average 20-30 minutes to complete. A standard recording sheet was used in each interview session.

The test question is in the forma of a statement made at the beginning of the session: *Arrange these images in order of most to least similar.* The question asked at the end of the session was: *Why did you arrange them in the way you did?*

2.2 The Test

The cards were placed in front of the participants in no particular order. Most went straight into the task without hesitation. They started by making either a long row or small groupings of images in rows and columns. They spent their time 'getting to know' the images first by sifting through them trying to remember their distinguishing features. After awhile they abandoned the earlier rows they had created and moved onto new orderings. Towards the end, they quickly eliminated the remaining few images and announced they had finished after 20 to 30 mins. When asked the final question, most gestured to the images to help answer it. They often picked images out of the arrangements to make their point, or began to point towards the row-wise groupings. They then began to describe, in few words, the reason they were 'different'. Often they would point to the last image in a line-up and say why it was the most different. Most complained that the initial question was not logical; that there is no most or least similar – "apart from the ball" (image number 22).

2.3 Analysis of Interviews

We saw a wide variety of approaches taken by the participants. By looking at the four main typologies that emerged we can begin to identify how the task was approached:

- Background Contrast: There were participants that organised the images in terms of the quality of the image itself rather than what it contained. Hence, these participants concentrated on the contrasting features of the background. This suggests they may not have known what was depicted in the picture and focussed instead on what they could see on its surface (four participants followed this schema).
- Size Difference: The next typology followed a similar rationale. Instead of contrast levels of black and white, however, they focussed on how much of the picture's surface the object depicted in it filled. This did not require the participants to recognise the objects depicted in the picture just that it was an object with size (two participants followed this schema).
- Simple to Complex: By far the largest group of participants organised their images across a range of simple to complex values. Within this schema there were variations on what constituted complexity. The main defining character of this schema was the fact that they all focussed on the features of the objects depicted in the images suggesting they recognised what the objects were (seventeen participants followed this schema).
- Outside any Category: The final schema is really a sub-schema. It represents those images that were identified as not fitting a larger gross category. What is significant about this sub-grouping is that the objects depicted in the pictures were referred to explicitly suggesting recognition of the object itself (six participants followed this schema).

All four key typologies, or combinations of their subsets, could be represented in a given individual participant's organisational schema.

It is worth noting that while all participants attempted to create a continuous range of images almost all found this was not possible and separated out smaller groupings within a grand scheme. There was a large variation in their justifications for why they created the smaller groupings.

The next step of the exercise was to try to make sense of the variations between how participants arranged their images and what they said. We applied the AM algorithm in the form: how many times are any two images paired in a group across all groups? To do this we created a table in Access that included the following columns: ParticipantID, GroupID, and SpikeID. There are 1770 possible pairings.

2.4 Applying the AM to Results

Applying the AM to determine how many pairings are common across all of the 1770 possible pairings returns 1466 pairings. When we look at these results we find that more than half of the participants arranged more than half of the images in similar pairs across all groups. Despite the wide variety in what they said, within what they did we find there was more similarity. In other words, despite having different rationales for organising the images in the way they did most groups had at least a similar pair contained. The following SQL query was used to generate table 2.

```
SELECT T1.SpikeID, T2.SpikeID AS OtherSpike,
Count(*) AS OccursNo
FROM tables02 AS T1, tables02 AS T2
WHERE (((T1.SpikeID)<T2.SpikeID) And
((T1.GroupID)=T2.GroupID) And
((T1.ParticipID)=T2.ParticipID))
GROUP BY T1.SpikeID, T2.SpikeID
HAVING COUNT(*)>N
```

Table 2. This table shows those AM pairings that have a count greater than 12.

MostPairsCount		
SpikeID	OtherSpike	OccursNo
26	37	16
1	35	15
29	32	15
26	59	15

MostPairsCount		
SpikeID	OtherSpike	OccursNo
35	55	15
15	41	15
58	59	15
1	48	14
4	60	14
37	58	13
37	59	13
36	47	13
47	54	13
35	48	13
1	55	13
26	58	13
17	54	13
49	59	13

From table 2 we notice many instances where members of one pair share members of another pair and so on. If we take a sample of the pair which achieved the highest count across all groups and participants (26 and 37) we can correlate their rationales with where this pair occurs in their groupings. At a finer level of detail than the four super schemas identified earlier, from what they said we can summarise their responses as one of ten different sub-rationales (the number of participants using this schema in brackets):

- Number of prongs (one);
- Number of points (one);
- Skinniest to fattest (one);
- Simple to complex (four);
- Leaves curving out (two);
- Number of angles (one);
- Background contrast (two);
- Similar shapes (two);
- Soft to sharp (one); and,
- Spikes moving out (one).

What is demonstrated in the list above is that, of the 16 participants who selected the same two as paired in a group (26 and 37), some reasons were given by more than one participant. This shows that the two 'obviously' go together for many different reasons. Thus, underscoring that the interpretability of the Amazon metric is *post-facto*, not *a priori*.

3. Visualisation of Results

We can visualise these results to gain a greater understanding of the visual relationship between the most paired images 26 and 37 at the local and global scale. For example, in figure 2, which presents a self-organising graph using the results of the AM for a count greater than 12, we can clearly see the similarity with the other spikes in the immediate locale. We also notice that the other groups included in this count similarly show an extraordinary internal visual similarity given the diversity of approaches and rationales to the task expressed by the participants.

When we expand this to include a subset visualisation of a count greater than 9 (which represents more than half of all participants and images) we notice an apparent clustering of highly interconnected similar types around our most counted pair, tapering off to more complex spikes (see figure 3). These other spikes, as we move away from our most counted pair, are not only more complex but tend to display leaves curving out. Both the shift to more complex and leaves curving out are features reasoned by at least six of the sixteen participants' schemas for placing the top pairing.

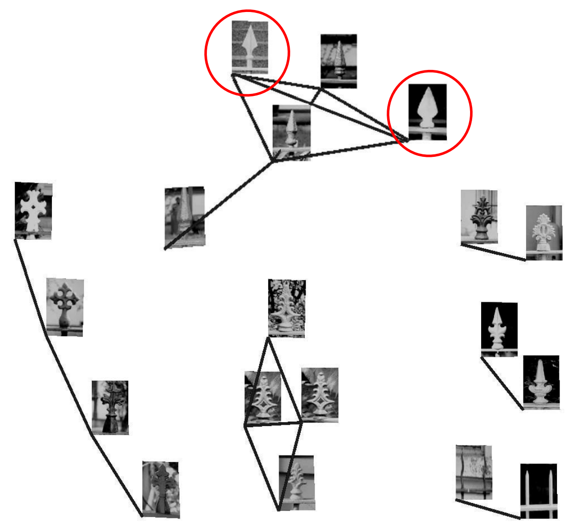


Figure 2. Visualisation of images where the AM pairings have a count greater than 12 using the images to display visual similarity (image 26 and 37 circled).

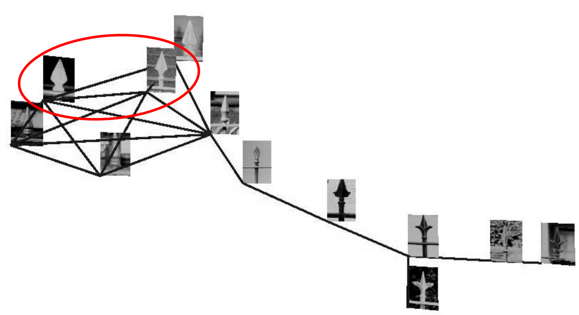


Figure 3. Visualisation of images where the AM pairings have a count greater than 9 (image 26 and 37 circled).

4. Comparing AM Results with the Traditional Method

The earlier traditionally structured image database we can say was addressing notions of place in

Wexelblat's terms; where the different images are first associated to a place-holder or class. We can have classes that do not have any images/objects in them and some images or objects can share multiple classes. Hence, we could plot this in a multi-dimensional space [5]. We chose to use a three-dimensional plotting (see figure 4, see also figure 5 for the non-dimensional network representation based on an 'expert' ordering).

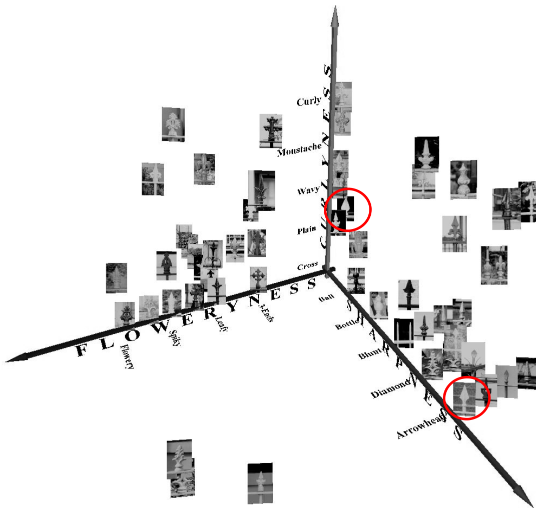


Figure 4. Wexelblat semantic dimensional representation of all images (image 26 and 37 circled).

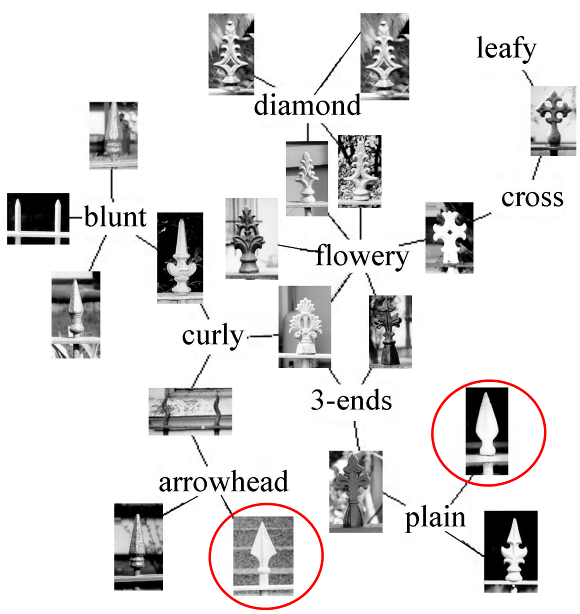


Figure 5. Two-dimensional representation using a non-dimensional self-organising graph using the same images as those in figure 3 (image 26 and 37 circled).

In the AM system we were not plotting nearness between places but nearness between the objects themselves. The AM nearness system has no ordinal sequence, hence there is no semantic sense between near objects other than what we impose *post facto*. However, as in the traditional model, the *post facto* ordination of the AM system allows us to navigate its places. Where the ordinal system in the traditional model comes from arbitrary relationships that we can map between its classes it relies on *a priori*, predictable, definable difference. In the AM system we can only define difference *after* the objects have been arbitrarily sorted.

When we return to the traditional method we find the most paired images (26 and 37) do not display any local similarity. We can conclude from this that as a database organising schema these two images are not likely to be paired. In other words, our expert metadata schema would not assist the group of 20 participants used in this study to find 26 and 37 as similar.

6. Conclusion

In conclusion we can say the AM creates very local orderings. Each is at its own centre and nearly everything else is out at infinity. Hence, we need the traditional method's large-scale geometry to see the whole space. Hence, the AM has proven to be a useful way to define similarity in this database and is different from the more structured traditional method. What this exercise highlights are the benefits of incorporating multiple schemas for organising an image database. This should support different kinds of information-seeking behaviour. The traditional method supports those users who have a good idea what they are looking for in a database. On the other hand, if the user wants to explore the possibilities they can browse a large collection of objects both of a global scale and locally supported by how other users browse the same collection. Future work would include evaluating the efficacy of the AM method for constructing an image database with a larger number of images.

References

[1] Chang, S.F., Sikora T. & Puri, A. (2001). Overview of the MPEG-7 Standard. IEEE Transactions on Circuits and Systems for Video Technology, 11 (6), 688-695.

[2] Colomb, R. M., 2002, *information Spaces: the Architecture of Cyberspace*, Springer, London.

[3] Dublin Core Metadata Initiative. (1999). Dublin Core Metadata Element Set, Version 1.1: Reference Description. DCMI Recommendation, 2 July. http://dublincore.org/documents/dces/.

[4] Wexelblat, A., 1992 "Giving Meaning to Place: Semantic Spaces", in (ed.) Benedict, M., *Cyberspace: First Steps*, MIT Press, Cambridge, MA.

[5] Wyeld, T. G., 2005, "3D Information Visualisation: an Historical Perspective", in proceedings of *Information Visualisation* '05, London, 6-8 July, 2005.

0-7695-2602-0/06 $20.00 © 2006 **IEEE**