Originally published in:
Proceedings of the 11th IEEE International Conference on Information Visualisation (IV 07), Zurich, Switzerland, 04-06 July 2007, pp. 431-435

A Non-Expert Organised Visual Database: a Case Study in Using the Amazon Metric to Search Images

Theodor G Wyeld
Media, H&SS, The University of Adelaide, Australia
{theodor.wyeld@adelaide.edu.au}

Abstract

In a previous paper the notion of "using the Amazon metric to construct an image database based on what people do, not what they say" was introduced (see [1]). In that paper we described a case study setting where 20 participants were asked to arrange a collection of 60 images from most to least similar. We found they organised them in many different ways for many different reasons. Using Wexelblat's [2] semantic dimensions as axes for visualisation in conjunction with the Amazon metric we were able to identify common clusters of images according to expert and non-expert orderings. This second study describes the construction of a visual database based on the results of the first case study's non-expert participants' organising strategies and rationales. The same participants from the first study were invited to search for 'remembered' images in the visual database. A better understanding was gained of their detailed reasonings behind their choices. This led to the development of a non-expert organised visual database that proved to be useful to the non-expert user. This paper concludes with some recommendations for future research into developing a non-expert, self-organising, visual, image database using multiple thesauri, based on these core studies.

Keywords: 3D visual database, image database, qualitative, non-expert, Amazon metric.

1. Introduction

In a previous paper the notion of "using the Amazon metric to construct an image database based on what people do, not what they say" was introduced (see [1]). In that paper how the problem with current image databases was described including how they are ostensibly organised by 'expert' categories following objective metadata schemas which may not help the novice or non-expert user to find what they seek [3, 4]. 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. Moreover, descriptors are usually textual. The problem with using textual surrogates to search images is 'how best to describe the object's contents' [5]? In a case study setting, it was found that asking a group of 20 people to arrange a collection of 60 images of fence spikes, they organised them in many different ways for many different reasons.

The paper goes on to describe how Wexelblat (1992) addresses the problem of visualization based on textual descriptors by identifying the elemental components of a semantic space with semantic dimensions as absolute and relative axes of visualisation. Within these relative semantic dimensions idiosyncratic subdivisions or associations among objects can be placed at will.

The problem remained, 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? The previous paper discussed the notion of displaying averaged user preferences in a dataset using what we called the Amazon.com metric – an averaging algorithm applied to the results based on how some *post facto* commonality can be detected.

This paper summarises the first case study and introduces a second. The second case study describes the construction of a visual database based on the results of the first case study's non-expert participants' organising strategies and rationales. The same participants from the first study were then invited to search for 'remembered' images in the visual database. A better understanding was sought of their detailed reasonings behind their choices and whether a non-expert organised visual database could be useful to the non-expert user.

2. Case Study 01 in brief

The first case study compared the expert categorising of images with non-expert categorising from the results of their sorting by a group of 20 participants. In the expert system, metadata was assigned to the images based on their objective features and content descriptions. The non-expert ordering was less clearly defined.

11th International Conference Information Visualization (IV'07)
0-7695-2900-3/07 $20.00 © 2007 IEEE

2.1 The Expert System

The expert system identified certain features in the group of images which are common to many. There were at least 14 different features, characteristics, aspects, or attributes of the spikes in the images.

The textual references to these features were then used to organise a visual database in a traditional manner. Two systems were implemented, Wexelblat's [2] 3D semantic space, and a self-organising network graph. Both demonstrated closeness due to similarities of features as determined by the expert categorisation.

2.2 The Non-Expert System

20 participants were interviewed and asked to arrange the same images (in the form of small cards) from most to least similar. Once this was complete, they were asked why they arranged them the way they did. It was found that they used many different strategies for organising the images, and they verbalised many different rationales for arranging them in they way they did. These were summarised as 10 ranges of types using 19 different terms. Which were further refined to 3 super themes, and their combinations.

When the Amazon metric averaging algorithm (which clusters images identified by most participants as being similar) was applied to how they actually arranged them (what they did), as against their rationales (what they said), it was found that more than half of the participants arranged more than half of the images in a similar sequence, although they gave different reasons for doing so. From this, we found the two most paired images.

2.3 Discussion

We found that the two most often paired images in the non-expert system were not paired in the expert system. The expert system assigned metadata based on 14 common object features and content descriptions (see table 1). As such, it followed a categorising system that was divorced from the actual images. These were abstracted textual descriptions of the images which were then used to organise the actual images. The non-expert ordering, on the other hand, was derived directly of the images. Individual images were compared with each other by different participants, hence subjective decisions could be made about their similarity. This is something the expert system does not allow for.

Only when the expert system is implemented can the array of images be seen and any clustering detected. The visualisation schema chosen, Wexelblat's [2] semantic dimensions representation, addressed notions of place; where the different images were first associated to a place-holder or class. There were classes that did not have any images/objects in them and some images or objects that share multiple classes. Hence, it could be plotted in a multi-dimensional space [6]. A three-dimensional plotting was used. Under the expert, feature-based, system the chances that any two images will be 'near' each other is based on how many similar features they share with other images. The dimensions of similarity in the expert system were arbitrarily assigned ranges based on the most divergent yet most populated set of dimensions derived from the physical features of the fence spikes (this expert ordering and the terms used were provided by the author who has a background in design). As the participants chose their own criteria for organising the images, in terms of most to least similar, then the expert system can be thought of as just one possible permutation. The fact that the most commonly paired images are not paired in the expert system suggests the expert system is an uncommon permutation.

The expert metadata schema would not assist the group of 20 participants used in the first study to find the two most similar. This suggests the non-expert should be consulted when sorting images for a visual database searchable by non-experts. It was the semantic dimensions 'thrown up' by the application of the Amazon metric (a behavioural dimension) that caused the clustering of similar images in the non-expert system.

2.4 Case Study 01 Conclusions

From the first study it can be surmised that a non-expert organised system should better support the non-expert user. As such, the next phase was to address the question: "can the non-expert find what they are looking for in a non-expert organisation of images in a visual database?" To investigate this, what the original participants said and did in the first study was incorporated into a searchable visual database.

3. Case Study 02

From the first case study we found many participants used similar terms with different meanings, or different terms with similar meanings. Overall, despite the large variations in their rationales for why they sorted the images in the ways they did there was a lot of commonality. When we compared this commonality with the expert system there appeared to be little or no correlation. From this, a new research question was proposed (see above). To investigate this new research question the same participants were asked to remember and describe the features of four different images from the first case study exercise, and then find them in a non-expert organised visual database.

The second study investigates the detailed rationales underpinning what the participants said and did. These detailed rationales and idiosyncrasies were used to shed light on the specificity of the use of particular query terms. This proved to be a fruitful approach, as this group's first study rationales can be compared with their more detailed rationales in the second study to check for consistency. Gross data recorded included: what criteria or category was chosen; and, the strategy adopted in the browsing process.

3.1 Setting

All participants were separately interviewed. They had not seen the visual databases beforehand. They were introduced to the various interfaces and given minimal instruction on how to use them. The original image cards were provided at the end of each session as a control of sorts.

Participants were not able to view transcripts of what they had said from the first study. The user interface included a pull-down menu that contained all 19 terms that all the participants uttered as organisational rationales in the first study. When they selected one of the terms listed on the pull-down menu a new screen was launched which contained all the images arranged in an optimised order which followed how those people who used that term ordered the images in the first study.

All participants were asked the same questions:

- "Can you remember an image that stood out for you from the last time we met and I asked you to sort a collection of images or cards?" and,
- "Can you describe it to me?"

This was repeated four times to identify four different images before proceeding. All interviews were recorded for later analysis. They were then presented with the web interface with the simple pull-down menu containing the 19 terms for each of the one, two, and three-dimensional databases. They were asked to:

- "choose a category from the drop down menu"; and,
- (once the images were loaded) "search the collection of images and identify the image you described to me earlier."

With the fourth remembered image, participants were asked to search the original cards. They were then asked to find the same images in the expert system. Finally, they were asked to describe how each system helped or hindered their search for the remembered image.

3.2 Creating a Visual Database

To develop the visual database, the initial methods for physically sorting the cards employed by the participants in the first study was revisited. Four main strategies for organising their image cards was found:

1. a continuous row;
2. interconnected rows and columns;
3. satellite or interconnected clusterings with axial relationships; or,
4. combinations of these.

These organising strategies can be redefined as one, two or three dimensional sorting schemas. These 'dimensionalised' sorting schemas were used to reconstruct the super themes identified earlier, as discrete visual databases. Their five verbalised rationales were further summarised as the dimensional ranges: those images considered outside all categories are still included as they are part of the super-set nonetheless. Taking into account the participants' one, two, and three-dimensional sorting schemas, their one, two, and three-dimensional rationales can be combined to create one, two, and three-dimensional visual databases for searching. As a one-dimensional schema, any of the one-dimensional rationale ranges can be used to generate a string of images for browsing. As a two-dimensional schema, any two of the one-dimensional rationale ranges can be combined; and, as a three-dimensional schema, each of the three can occupy one of the axes with ranked order in three spatialised directions.

The purpose of the second case study was for the original participants to identify images that they could remember and describe from the first case study for searching in these dimensionalised visual databases. As they were searching images remembered from the first study, it follows that they should also be able to choose a categorisation that best fits their expectations based on what they said in the first round. Hence, the vocabulary of terms they could choose from, when searching each of the visual databases, included the same 19 terms they uttered when rationalising "why they arranged the images in the way they did" from the first study. The dimensionalised range each term refered to was derived of the most common association made by participants who used those descriptors and arranged their images along those dimensional ranges from the first study.

In the second case study, the 19 terms were available for participants to choose from a drop down menu which, when a single term was clicked on, launched a collection of images arranged according to that criteria or category. The resultant collection of images may not be exactly as they originally arranged them but rather an averaged form using the Amazon metric across all participants' arrangements against one of the three super schemas. Only a single term was required to activate either the one, two, or three-dimensional visual databases as not all the participants employed multiple terms for sorting their images in the first study. This study was more interested in the correlation between individual terms and descriptors from the first and second studies than which searchable visual database was used *per se*. The use of the three different types of searchable visual databases was conducted as a sub investigation into the relative efficacies of each system, and also to reflect the different dimensional strategies that were employed by the participants in the first study. In turn, this also allowed identification of any correlation between a participant's original sorting strategy and their preferred search interface.

As images were arranged according to the participants' original use of one or more of the three main dimensionalised strategies, each participant was asked to search for their remembered image from each of the one, two, and three dimensional visual database displays. As a control of sorts, a fourth remembered image was sought from the original, randomly sorted, collection of cards. This further establish the relative efficacies of the strategies used in the first study compared to the systems used in the second study.

Finally, the same participants were asked to search for the same images using the expert system.

3.3 Searching for Remembered Images

Participants had trouble remembering specific images but they were able to describe some features. Many different images were described, and many different rationales were used for the choice of descriptors and pull-down menu terms. Of these, the five most common terms used across all participants from the first and second studies were isolated. Their rationales behind why these terms were used were studied and analysed in detail. Findings included:

- some terms and descriptors were substituted for terms or descriptors not available from the 19 given terms, yet returned similar results to the use of the substitute term by others who understood its more common meaning;
- many common terms and descriptors were used with different meanings, yet returned similar results; and,
- some terms were interchangeable, both in meaning and result.

3.4 Comparing Terms Used

When synonymous terms were compared with those used with different tenses it was found that, despite the wide variety of rationales and meanings, there was some commonality in results returned.

4. Discussion

From the correlations in the terms they used some were seen to be interlinked by a common notion. However, the individual meanings were quite different. Hence, the key finding from the second study was how the participants' interpretations of the various terms they found in the pull-down menu list of criteria were associated with other participants' different interpretations yet the images they eventually chose seemed to be related. For example, two participants may chose a similar term (although they did not know what it actually meant) to mean one thing returned images that express two quite different concepts of the same thing, yet both did reflect similar attributes of that thing, nonetheless. In another case, while most participants used the same term to refer to how an object's appearance, it was also used to refer to the emotions invoked by that particular image or its background. The variability in the way the same and different terms were used underscores the need for multiple thesauri to support this sort of search strategy (something current expert systems try to avoid). It is these variations in the use of the same term that builds the multiple thesauri of meanings, synonyms, and associations.

Despite having trouble remembering specific images and finding suitable terms, most participants were able to find what they were looking for in the non-expert organised visual database. When they selected a term for the expert system, on the other hand, there was exact, little, or no correlation between the non-expert and expert systems. The image they were looking for was often many steps away. Finally, although they had wildly divergent rationales, substitutions, and synonymous approaches, there was some commonality in their searched images by category (although this cannot be conclusively confirmed due to the size of the study group and dataset).

5. Conclusion

In the first study we looked for commonality in what participants said and did. We found they said many different things but often meant something quite similar. In the second study we looked for a more detailed rationale behind specific terms. The findings indicate that, the sheer complexity of their detailed rationales obscured any overall commonality. However, the images returned tended to show some localised similarity, despite using the same terms with different meanings, or different terms with similar meanings. This suggests there may be strong reasons why images in a non-expert organised visual database might be clustered in particular ways generated from very different core queries.

Where our non-expert system differs from the amazon.com or the Google 'I'm feeling lucky' metrics is in exposure of the reasoning behind the query parameters. In our case, a textual association could be made along with the visual reference as a clustering of similar images. This was demonstrated in the collection of clustered images from the first study – they showed strong correlations although it was not immediately clear why. For example, when asked to comment on why the two particular images most often paired were, most people responded that they 'just seemed right', or 'it seemed logical enough'. The expert system, on the other hand, did not show these apparent relationships.

This study demonstrates how it would be possible to populate a database of images with information from non-expert users that is useful, meaningful, and effective.

6. Future Directions

These studies demonstrate the importance of developing a schema for a non-expert, self-organising, visual database of images with metadata populated by non-experts.

Where the expert organised, feature-based, visual database returned the same image cluster (for the same term) each time, a non-expert, self-organised, visual database based on multiple thesauri would return a cluster of useful images for browsing, among which the image sought is most likely to be present. It is most likely to be present because many users have previously identified this cluster as having some common meanings. These common meanings may not relate strictly to any category that could be independently identified by an expert. This would make a non-export system both more useful for non-expert browsing and dynamically respond to user input rather than a fixed expert system's prior, rigid, categorisation.

For example, a visual database that returns a collection of images for browsing with an extended vocabulary which supports multiple thesauri would allow for one term to also mean something different (following a query which linked objects in the database to these two terms). This is because that is how some users have searched for these images in the past – using both terms as key terms despite their apparent disparity. This means that any objects in the database that are linked to the first term would tend to cluster with those that are also referred to by the other term, and vice versa. Unless both terms were defined as synonymous in the expert system, it would be unlikely to return the same or similar clusters of images.

In other words, if a term is used in a number of different ways by a number of different users then all of those different ways can be used to populate a database generating clustering of like images, such that when users see those clusterings they 'appear' ordered. This cannot be predicted, but would *appear to be so – post facto*.

References

[1] Wyeld, T. G., and Colomb, R, 2006, "Using the Amazon Metric to Construct an Image Database based on what people do, not what they say", in proc. of Information Visualisation 2006, London, England Jul05 -07

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

[3] 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.

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

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

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

11th International Conference Information Visualization (IV'07)
0-7695-2900-3/07 $20.00 © 2007 IEEE