Does Organisation by Similarity Assist Image Browsing?

Kerry Rodden, Wojciech Basalaj
University of Cambridge Computer Laboratory
Pembroke Street, Cambridge CB2 3QG, UK
{kr205,wb204}@cl.cam.ac.uk

David Sinclair, Kenneth Wood
AT&T Laboratories Cambridge
Trumpington Street, Cambridge CB2 1QA, UK
{das,krw}@uk.research.att.com

ABSTRACT
In current systems for browsing image collections, users are presented with sets of thumbnail images arranged in some default order on the screen. We are investigating whether it benefits users to have sets of thumbnails arranged according to their mutual similarity, so images that are alike are placed together. There are, of course, many possible definitions of "similarity": so far we have explored measurements based on low-level visual features, and on the textual captions assigned to the images. Here we describe two experiments, both involving designers as the participants, examining whether similarity-based arrangements of the candidate images are helpful for a picture selection task. Firstly, the two types of similarity-based arrangement were informally compared. Then, an arrangement based on visual similarity was more formally compared with a control of a random arrangement. We believe this work should be of interest to anyone designing a system that involves presenting sets of images to users.

Keywords
Image retrieval, information visualisation, evaluation.

INTRODUCTION
Traditionally, graphic designers searching for images to accompany text have had to submit their requests to a trained intermediary at a photo library, receiving a selection of hand-chosen prints via courier some time later. Now, many image libraries allow direct searching of their collections via the Internet, such as CorbisImages.com, who have a collection of 1.6 million digital images available. Also, many companies provide smaller collections on CD-ROM, and newspapers and magazines may maintain their own digital photo archives. Designers can immediately download or copy images from these sources into their documents. It is therefore becoming increasingly important to provide good support for these users to search and browse digital image collections.

In most commercial collections, the photographs have been manually annotated with descriptive text, such as a caption and keywords, and they can be indexed and searched in much the same way as textual documents. Unlike textual documents, however, the content of an image can be taken in at a glance, and a large number of them can be presented to the user at once, using *thumbnail* (miniature) versions. This makes browsing easier, allowing users to simply scan an image set in order to find what they want. Browsing through images is also helpful when developing illustration ideas [11]. However, current systems concentrate more on providing support for querying than browsing.

The most common way for a system to present a set of images is in a two-dimensional grid of thumbnails [7], in some default order. Other researchers have found that information visualisation techniques can be used to make a set of textual documents easier to browse, giving a representation of the set a noticeable structure by clustering together mutually similar items [10]. One would therefore expect the same to be true for image collections, and this is what we investigate here.

IMAGE SIMILARITY
To create a visualisation of a set of objects, one first needs to define a way of measuring the similarity of a pair of the objects. If images are annotated with textual captions, one can use a measure based on these, which in theory should produce the most meaningful possible automatic arrangement of the image set, clustering together the photographs depicting the same people, places, or objects. For our experiments we have used information retrieval's vector model, with binary term weighting, and the cosine coefficient measure [19].

Captions must be manually assigned, which is expensive, and the results are often highly subjective. Image retrieval researchers have defined a number of measures based on automatically extracted visual features, with the aim of replacing or supplementing the conventional text-based measures. Of course, this is a very difficult problem, and current systems can only extract low-level features like colour and texture, with some limited segmentation of images into coherent regions. Query-based retrieval using purely visual measures has so far had limited success, and Rubner [17] was the first, as far as we are aware, to show that these measures could also be used in conjunction with information visualisation techniques, and others have since developed similar tools. The visual similarity measure we use is explained in more detail elsewhere [14], but, briefly, it is designed to take into account both global image properties (colours and textures) and the broad spatial layout of image

SIGCHI'01, March 31-April 4, 2001, Seattle, WA, USA.

regions. Other measures can be used with comparable effect [13].

CONSTRUCTING THE ARRANGEMENTS

The arrangements are created using multidimensional scaling (MDS) [3], a technique which treats inter-object dissimilarities as distances in some high dimensional space, and then attempts to approximate them in a low dimensional (commonly 2D or 3D) output configuration. We use our own MDS algorithms, developed and evaluated as part of separate research [2], although MDS functions are also available in many common statistical packages. To arrange a set of images, we first create a similarity matrix, containing the measured similarity of all pairs of images in the set. Once MDS has found a 2D configuration of points using the matrix (which takes about one second for 100 items, on an average PC), thumbnail versions of the images can be placed at these points, to produce an arrangement. The left-hand part of Figure 1 gives an example of one of these, using the visual similarity measure.

Because many of the points in the 2D configuration are separated by less than the size of a thumbnail, there is usually some overlap of images. In one of our early experiments [14], many participants stated that they disliked the overlap, and also found the irregular structure difficult to scan. We therefore adapted our layout algorithms to fit the configuration into a square grid [2], as illustrated in Figure 1. Different sizes of grid can be used, depending on the trade-off one wishes to make between accuracy and thumbnail size: sparse grids provide a closer reflection of the structure of the original (continuous) configuration, but dense grids allow the thumbnails to be larger.

EVALUATION

As with other types of visualisation, the arrangements do look appealing, but there has been little evaluation of their actual effectiveness. Our early experiments [14,15] confirmed that arranging an image set according to visual similarity could make users faster at locating a given image, or a group of images matching a generic requirement. These experiments were low-level in nature, with simplified tasks, and we were interested in carrying out an evaluation in a more realistic situation.

In the area of information retrieval, Borlund and Ingwersen [4] have proposed the *simulated work task situation*, where experiment participants are assigned requirements that are as close as possible to those they might have in a real situation, including a described role and context. Jose et al. [9] applied this to image retrieval, using graphic designers as experiment participants, and asking them to imagine that they were required to do some picture selection as part of a freelance job. We adopted a similar approach.

Studies of requests submitted to photograph collections (e.g. [1,11]) have shown that users most commonly want to look for photos of specific things, perhaps named people or places. To satisfy this type of requirement, the photographs must be accurately annotated, which is generally the case in commercial image collections. We therefore expect that many image searches will begin with the user entering a textual query for the required item (or navigating a hierarchy to a named category), and then browsing the resulting set of thumbnails. We wanted to investigate whether arranging this set according to similarity would help the user to find the "right" image(s) within it.

THE TASK

The task was the same for both experiments. Participants were given the following written description of it:

> You have been asked to choose photographs to illustrate a set of "destination guide" articles for a new "independent travel" World Wide Web site. Each article will be an overview of a different location, and is to appear on a separate page. The articles have not yet been written, so all you have are short summaries to indicate the general impression that each will convey. You also have 100 photographs of each location, and your task is to choose 3 of the photos (to be used together) for each article. It is entirely up to you to decide on the criteria you use to make your selections—there are no "right" answers, and you are not bound by the given summaries.

The 100 images represent either the contents of a category or the top 100 results of a query for the place to be illustrated. Markkula and Sormunen [11] state that the journalists in their study were willing to browse through large numbers of thumbnails, only reformulating their queries if they had more than about 100 results. The photographs we used were drawn from the *Corel Stock Photo Library* (collections 1 and 2), where each image has a one-line caption associated with it. We chose places where the available images were of good technical quality, and corresponded well to the pieces of text. These were abridged from an existing online travel guide, to a length of approximately 200 words each. Where possible, we chose passages that focused on creating an impression of the place, rather than describing particular landmarks, in order to encourage participants to decide for themselves what they should look for.

A single search proceeded as follows. First, the name of the place was displayed on-screen, and participants read its associated text, which was provided on paper. They then pressed a button to cause the image set to appear. Participants were free to select or deselect images until they were satisfied with the chosen three, at which point they pressed a button marked "Done".

To display the images to the participants, and allow them to make their selections, we used a custom Visual C++ program, a version of which is shown in the left-hand part of Figure 3. All 100 images were displayed in thumbnail form, with the area currently under the user's mouse pointer shown at 3x magnification in another part of the window, along with the caption of the current image. An image was selected or deselected by clicking on it with the mouse. Selected images were highlighted, and copied to an area at the bottom left of the screen. If more than one arrangement of the image set was available, radio buttons allowed the user to switch between

Figure 1: Three arrangements of 100 images of Kenya, based on visual similarity. On the left is the continuous MDS arrangement (with overlap), in the middle a 12x12 grid (which removes the overlap while preserving some of the structure), and on the right a 10x10 grid (which maximises the thumbnail size). (See color plate on page 000.)

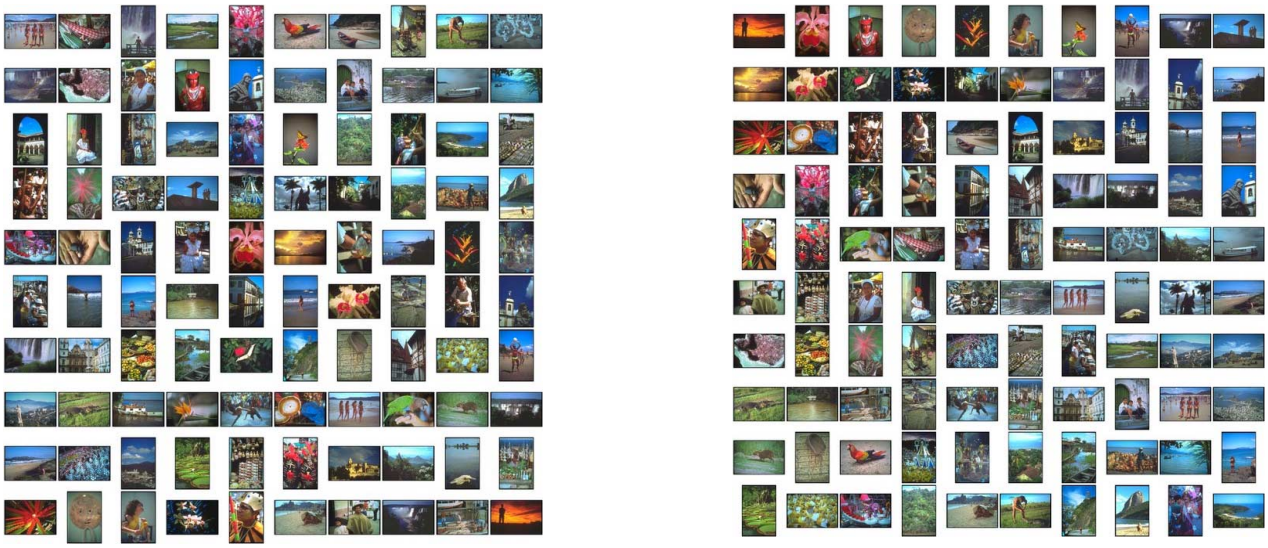

Figure 2: Two 10x10 grid arrangements of 100 images of Brazil, as used in Experiment 2. On the left, they are arranged randomly, and on the right, they are arranged according to visual similarity. (See color plate on page 000.)

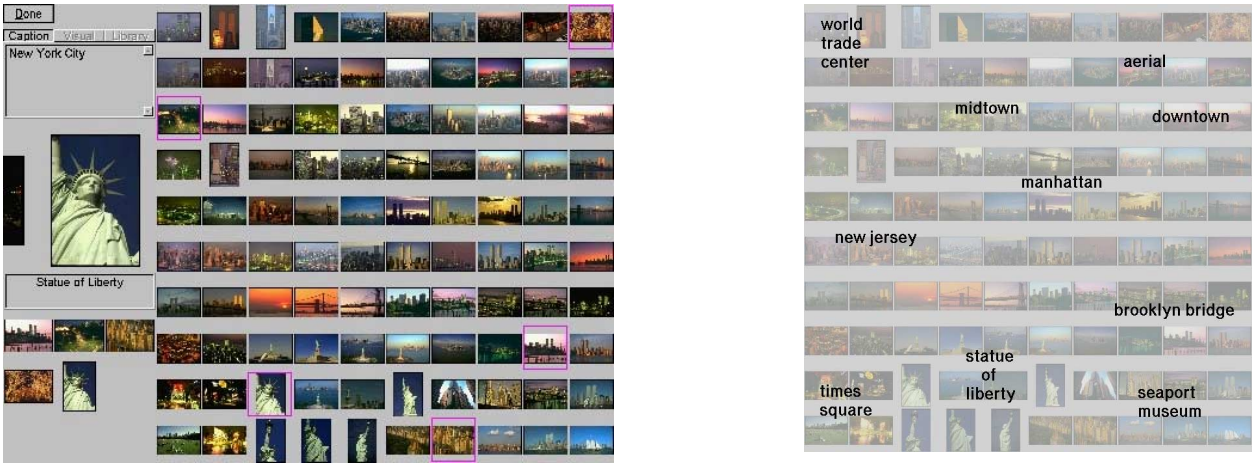

Figure 3: On the left is a version of the experiment software, showing a 10x10 caption-based arrangement of 100 images of New York. The area under the mouse pointer is displayed at 3x zoom (here, the Statue of Liberty). On the right is a mock-up of the same arrangement, with superimposed keywords (manually extracted from the captions), intended to make the structure more obvious. (See color plate on page 000.)

these. The highlight on selected images was retained after switching, and this allowed users to see how the surrounding context of their selections changed. User actions (selections, switches, and mouse movements) were logged and timed.

EXPERIMENT ONE

Initially, we were simply interested in investigating whether designers would find either of the similarity-based arrangements (visual or caption) useful for selecting images, and also if it was helpful to have both arrangements available. For example, in a caption-based arrangement of a set of photographs of New York, an image of the Statue of Liberty at sunset would appear next to all of the other Statue of Liberty photographs. However, in the visual arrangement the same photograph would be grouped with the other sunsets, giving the designer a different perspective, and perhaps also allowing him or her to easily avoid (or find) visually similar images when looking for images that work well together as a set.

Participants

To obtain access to a community of designers, we set up the experiment at the "infodesign 99" Information Design conference[1]. Our 18 participants were all attendees, who volunteered their time during the half-hour conference breaks, so this gave us a tight time constraint. All had normal or corrected-to-normal vision, with no colour blindness (self-reported).

Apparatus

We selected four places: New York, Paris, Kenya, and Alaska. For each of these, we created two 12x12 grid arrangements of the 100 images, one arranged according to visual similarity (as in the middle of Figure 1), and the other according to caption similarity. The thumbnail images were 96x64 pixels. We used two PCs, both running Windows NT 4, with 17-inch monitors set at 1600x1200 resolution. Each was used for half of the participants, and usually there were two people working at the same time.

Design and procedure

All participants had one practice search (New York), during which the software was demonstrated to them, and they were allowed to practise using it until they felt comfortable. The two arrangement types (named "caption" for caption similarity, and "image" for visual similarity) were explained, and participants were shown how to switch between them. They then did three "real" searches (Paris, Kenya, and Alaska). All six possible orderings of these three places were used.

The procedure was as already described. As there were two types of arrangement available, the participant had to choose which one to use first, by pressing its radio button. This caused the set of images to appear, and also started the timer for the search. Participants could flip between arrangement types as often as they wanted, and then press "Done" when they were happy with their three selections. Once they had completed the experiment, participants filled in a questionnaire, where they indicated their agreement or disagreement with a series of statements, and could write down further comments if they wished.

[1] in Cambridge, July 1999. See http://www.idu.co.uk/id99/

Results and discussion

We expected that both the caption-based arrangement and the visual arrangement would be regarded as useful in their own right, and also when used in combination.

Statement	*Agreement*					*Median*
	0	1	2	3	4	*score*
"the arrangement of photos by **caption** similarity was useful"	0	3	3	9	3	3
"the arrangement of photos by **image** similarity was useful"	3	4	3	6	2	2
"it was useful to have two different views of the same set of photos"	1	4	5	4	4	2

Table 1: Responses to three of the questionnaire items from experiment one. 0 represents "strongly disagree" and 4 represents "strongly agree".

Table 1 shows that two-thirds of the participants agreed that the caption-based arrangement was useful (and commented, for example: "it gave me a breakdown of the subject", "it helped link to the text"), but opinion was more divided on the usefulness of the visual arrangement. Eight people rated caption as more useful than visual, three gave visual a higher rating, and seven gave them the same score. The ties make statistical analysis difficult; a two-tailed Wilcoxon signed-rank test gives a p-value of 0.099, which is significant only at the $p<0.1$ level.

In 40 of the 54 searches, participants chose to look at the caption-based arrangement first, and on average spent 63% of their total search time using it, and 37% on the visual arrangement. These averages hide a large amount of variability in the time spent using each arrangement type. Seven participants heavily favoured the caption-based arrangement (using it for 85% of the time, or more; four used it exclusively), three heavily favoured the visual arrangement (using caption for 22% of the time, or less), and the remaining eight showed no obvious preference (spending between 39% and 69% of their time using the caption-based arrangement).

Six of the latter group also agreed that it was useful to have two different views of the same set of photos (eight agreed in total). One said "the caption mode I preferred to start with, that made sure there was enough diversity 'content-wise' (subject related) then I'd use the image mode to make sure they complemented each other as a set." In two-thirds of all searches there was at least one switch between arrangements. However, one participant did not feel this was useful, commenting that it "was distracting and broke concentration – I did it less and less as I went through."

We felt that the wide range of preferences expressed by the participants in this experiment may simply have been due to individual differences. Information designers tend to work primarily with text, and only 11 of the 18 participants had prior experience of carrying out picture selection. This group tended to favour the visual arrangement more than those without experience.

The caption-based arrangement can of course only be as good as its captions, and two participants commented on its limitations for considering different levels of meaning. For example, in the "Kenya" set, the wildlife photographs were scattered around the screen, because the captions only contain the individual species names, such as "lion" or "giraffe", and no generic word like "animal" to connect them. Rose et al. [16] have experimented with using the WordNet lexical database to automatically expand image captions, and we believe that their approach would lead to better-structured caption-based arrangements.

EXPERIMENT TWO

We felt that the findings of this exploratory study were interesting enough to warrant a more formal experiment, this time directly comparing a similarity-based arrangement to a simple control: a random arrangement (something that was not possible in the first study). As well as finding out if designers would prefer a similarity-based arrangement to a random one, we were also interested to know if it would help them to carry out the given task more quickly.

For the similarity-based arrangement, we decided to use a visual measure rather than a caption-based one, primarily because the usefulness of the visual arrangements seemed less clear-cut, but also because of the inconsistencies in caption quality, as discussed above. We used the same task as before. Obviously, to avoid biasing the participants, we did not want to tell them that one of the arrangements was simply random, and so we instead named it "library", and said that the images were placed in the same order in which they were provided by the creators of the stock photograph library. The instructions were also carefully worded, to try to ensure that both arrangements were presented as being equally valid.

We wanted to consider individual differences in more depth in this experiment. Other information visualisation researchers [6,18] have found that the results from tests of spatial ability can be correlated with quantitative experiment results. We therefore asked our participants to take Set I (12 questions) of Raven's Advanced Progressive Matrices (APM) [12], a culture-free test of spatial reasoning that specifically evaluates the ability to think about abstract categories.

We also carried out a follow-up study, a month after this experiment, where two of its participants (numbers 04 and 05) were asked to perform similar tasks while thinking out loud, and were recorded on video. They again used visual and random arrangements, as well as a number of variations, including caption-based arrangements. We were aiming to gain more insight into their thought processes while searching than was possible with post-experiment written comments. A full analysis of the transcripts can be found elsewhere [13], but in the following results section we quote some of the comments that were relevant to this paper.

Participants

Our participants were ten students of graphic design from Anglia Polytechnic University, Cambridge. Eight were at the end of the second year (of three) and two were at the end of the first year. Again, all had normal or corrected-to-normal vision, with no colour blindness (self-reported). Each was paid a small amount for their participation. They were split into two groups, so that five people were doing the experiment at a time.

Apparatus

We selected nine places for the main part of the experiment (Brazil, Canada, Death Valley, Denmark, Ireland, Jamaica, Kenya, Nepal, and Yellowstone National Park), and one to use as a practice (Devon). For each of these, we created a random arrangement, and an arrangement based on visual similarity (see Figure 2). In both cases the 100 images were placed into 10x10 grids, rather than the 12x12 of the first experiment, as we had to use a lower screen resolution (1024x768) and wanted to maximise thumbnail size (75x50 pixels). We used five Windows 98 PCs, which all had 17-inch monitors.

Design

In the first experiment, participants always had both types of arrangement available. This time, we wanted them to use each one on its own, in order to fully appreciate its strengths and weaknesses, before going on to having a choice between the two. We also wanted to be able to directly compare the arrangements to each other in terms of how quickly selections were made. The experiment was therefore divided into two parts. In part one, the students chose photographs for six places, using one type of arrangement for the first three, and the other for the second three (a within-subjects design). Half of the participants used the visual arrangement first and half used the random arrangement first. Then, in part two, both types of arrangement were available; the students had to choose one of them to use initially, and could then switch between the two views as they wished.

For part one, we chose six of the places and divided them into two groups. Group X was Denmark, Jamaica, and Nepal, and group Y was Death Valley, Ireland, and Kenya. We then counterbalanced both the type of arrangement and the group of places, so that first type of arrangement (visual or random) alternated with every participant, and first group of places (X or Y) alternated with every second participant. We also varied the order of places within a group. The remaining three places (Brazil, Canada, and Yellowstone National Park) formed group Z, and this was used for every participant in part two, with both arrangements available.

The APM test was administered between part one and part two. This gave the students a break from the task, and was

intended to reduce the chance (in part two) of them simply favouring the arrangement type they used most recently.

Procedure

The procedure was largely the same as for experiment one. During the practice search, both types of arrangement were available, and were explained to the participants in turn, along with the software and its features. They were then asked to practise using it until they were comfortable, and had made three selections. Ideally, we would have introduced each arrangement type just before the participant was about to use it, but this was awkward because of the balanced design. To reiterate, in part one each participant did six searches: three with one arrangement (on its own) and then three with the other. They did this in their own time but were encouraged to be quick. Then they were given the APM test, and asked to complete that in their own time. In part two they had both arrangements available and could switch between them. At the end they were given a post-experiment questionnaire.

Results

In part one of the experiment, the visual and random arrangements were directly compared to each other. Our principal response (dependent) variable was *done*, the time taken from the start of the search until the "Done" button was pressed[2]. This was not normally distributed, so we applied a log transform to it prior to analysis. We chose the statistical package S-PLUS for data analysis, and used its linear regression features to construct a linear model for *done*. The following predictor variables were used: *subject* (the ID of the participant), *trial* (an ordered factor: the sequence number of the search, from 1 to 6), *place* (the ID of the place, from six possibilities), and *type* (the type of arrangement used, visual or random). Then the analysis of variance (ANOVA) function was used to extract information from the fitted model, as shown in Table 2.

	Df	F-value	Pr(F)
subject	9	7.19	4.6×10^{-6}
trial	5	5.44	6.9×10^{-4}
place	5	1.31	0.278
type	1	6.34	0.016

Table 2: ANOVA results for *done*.

The variable *type* is significant at a level of $p<0.05$. However, the direction of the difference is surprising: participants were **slower** with the visual arrangement, taking a mean time of 103.8 seconds (s.d. 56.0) to complete a search with visual, compared to 81.3 seconds (s.d. 41.4) with random. Also, *trial* is significant, at a level of $p<0.001$: in general, participants took less time over their selections as they went along. This could be because they became more practised, or started taking the task less seriously, or some combination of these. But because of the balanced design, this does not affect the significance of *type*.

In the post-experiment questionnaire, participants were asked to express their satisfaction with regard to their selections for each search (from 0 to 6, where 0 is "not at all satisfied" and 6 is "very satisfied"). Overall, the median satisfaction score for searches using the visual arrangement was 4.5, and for random it was 4. A simple comparison of these scores with a two-tailed Wilcoxon rank-sum test gives a p-value of 0.054, just missing significance at the $p<0.05$ level. If we consider the scores as totals per participant, six of them were more satisfied with the searches they did using visual, one was more satisfied with random, and the remaining three were equally satisfied with both. Statistical analysis of this data is problematic because of the three ties, but comparing the ten pairs of numbers with a two-tailed Wilcoxon signed-rank test gives a p-value of 0.053.

Table 3 shows the participants' median scores for three statements regarding the random and visual arrangements, and the results of comparing the sets of scores using two-tailed Wilcoxon signed-rank tests. The scores for visual are significantly higher for the first two statements, $p<0.05$ and $p<0.01$, respectively. For the third statement, the scores for visual are significantly higher only at a level of $p<0.15$, as despite the difference in median scores, three participants gave random a better rating than visual.

Statement	*Median score*		*p*
	visual	*random*	
"it was enjoyable to use"	4	2.5	0.019
"it made it easy for me to find the photos I wanted"	5	3	0.007
"it made it easy to find photos that complemented each other"	4.5	3	0.133

Table 3: Comparing the two arrangements. 0 represents "strongly disagree" and 6 represents "strongly agree".

In part two of the experiment, visual was chosen first in 21 of the 30 searches, and on average was used for 66% of the search time, with random being used for 34%. Five participants heavily favoured visual (using it 73% or more of the time), with two heavily favouring random (using visual for 22% of the time or less), and three favouring neither (using visual between 56% and 66% of the time).

Raw scores on the APM test were between 7 and 11 (out of 12), which is from "low average" to "high" for the 18–39 age range. However, no relationship was found between these scores and any of the experimental variables or questionnaire results.

Discussion

[The quotes in this section are either drawn from the post-experiment questionnaires, or the transcript of the follow-up session. In the latter case, the participant's ID number is

[2] We also considered the length (in grid cells) of a participant's mouse trail across the arrangement, but as this was significantly correlated with *done*, the results were very similar and are omitted here.

suffixed with an 'f'. The participants, of course, refer to the random arrangement as "library".]

As we mentioned earlier, we have assumed in these experiments that the user has already carried out some restriction of the image collection, by issuing a query, or opening a category, and that therefore all of the images in the displayed set are potentially relevant. Designers' ideas evolve during the search process [8], so that for example an initial search for images of New York may develop into a series of sub-requirements for the Statue of Liberty, a taxi cab, a night scene, etc. An arrangement of images based on visual similarity often has an obvious structure, helping the user to decide where to look, and this is probably why the visual arrangement scored significantly higher than the random arrangement for the statement "it made it easy for me to find the photos I wanted". Participants' comments about the visual arrangement support this conclusion:

"If you had an idea in mind, say a sunset, you have them all together, easier to make the choice." (02)

"Visual puts photos in groups, i.e. it is easier to find one type of photo. You have direct comparisons to make a choice of photo." (03)

"Good for finding genres, e.g. landscape, night-time, daytime etc." (09)

"That's now got me thinking of a temple shot. Which I'll click on to the visual for, see if there's any temples. Which there is, found that straight away." (04f)

"I prefer a darker one […] which I'm going to look for now on this side of the screen, which is easier, so I much prefer the visual for this." (05f)

However, arranging images by visual similarity does have a drawback that is not usually encountered in information visualisation: the fact that visually similar images are placed next to each other can sometimes cause them to appear to merge, making them less eye-catching than if they were separated, in the random arrangement, so that it becomes "easier to miss a photo completely" (09).

"They're all merging together, my eyes aren't concentrating enough on each one, I think they're too similar." (05f)

"My eye was often drawn to one set of colourfully interesting images while ignoring the rest." (10)

This may help to explain the two most surprising results of the experiment: the fact that searching was faster with the random arrangement, and that participants did not overwhelmingly favour the visual arrangement in part two. Because all of the images were potentially relevant, especially at the beginning of a search, often our participants did not really know what they were looking for, and simply wanted to scan the set to see what was there. In this case a random arrangement can be helpful as it may enable a strong image to stand out, rather than be lost among visually similar images (Figure 2 may be helpful in understanding this effect).

"[With library] you saw the whole selection, but mixed up so the ones you liked sprang out at you" (02).

"I'm going to click back on to the library again, because I just feel more comfortable with that in the first instance, just for skimming." (04f)

"[I switched to library] so I could look all over it and just see if something actually caught my eye." (05f)

So, when using the random arrangement in part one, participants could simply grab eye-catching images, but with the visual arrangement they were perhaps forced into thinking harder about what they were looking for, thus taking longer to complete their search. This may also be why they were more satisfied with their choices using the visual arrangement, although we could find no correlation between search time and satisfaction score.

Five of the ten participants said that they preferred having both arrangements available:

"The more variety the better. I preferred using each arrangement as a layout to 'scan' to see if the same images stuck out to me." (05)

"Both had merit. Useful with visual if I wanted to make sure [my selected] images were not too similar. Library was a good way to start and then switch to visual." (10)

Four preferred using visual on its own. One participant preferred library (random), but said that "visual was good for a quick browse".

With regard to individual differences, what we tested for was the person's ability to make sense of a visualisation. Given our results, it appears that this was less important than the amount of active searching they were prepared to do: perhaps those participants who found most benefit in a similarity-based arrangement were those who were more likely to attempt to narrow down their search, rather than being content to simply scan the set for eye-catching photographs.

The caption-based arrangement

This was only used in the follow-up study, not in the main part of the experiment. It does not have the drawback of making adjacent images appear to merge, although in experiment one we found that its usefulness is affected by the level of detail of the captions. Because the Corel captions tend to focus on names, the resulting arrangements usually do a good job of grouping together photos of a particular area, for example in the Czech Republic set:

"Now that I know this is the Prague area I'm sort of staying here, because I think it's got quite a lot more to offer, visually." (04f)

This was especially true with a 12x12 grid (as used in experiment one), as the extra space allowed the clusters to be more obvious than they were in the 10x10 grid. However, as the quote suggests, the main problem participants had was actually locating the right part of the arrangement, as the basis of its structure is not immediately obvious, unlike the visual arrangement.

"I'm trying to work out how the captions go together. […] Is it in alphabetical order?" (05f)

We showed the two follow-up participants a mock-up that consisted of a caption-based arrangement of images, with short textual labels placed on top (see the right-hand part of Figure 3; compare it to the version without labels, on the left). These were manually extracted from the captions of the images in a particular "cluster". This elicited a number of positive comments, and we believe that if such "imageability features" [5] could be reliably created automatically, they would be of great help to users in understanding the structure of caption-based arrangements.

CONCLUSIONS

Automatically arranging a set of thumbnail images according to their similarity does indeed seem to be useful to designers, especially when they wish to narrow down their requirement to a particular subset. A caption-based arrangement helps to break down the set according to meaning, although its usefulness depends on the level of detail in the available captions. Labels may also be necessary to help the user understand its structure. An arrangement based on visual similarity helps to divide the set into simple genres, although it can also cause adjacent images to appear to "merge" (an effect not observed in other information visualisation applications), meaning that when the user does not have a particular requirement in mind, a random arrangement may be more useful, as strong images usually contrast to their neighbours and thus appear to stand out. There is also evidence that, for some people, having access to different arrangements of the same set of images is useful, although the source of the individual differences is yet to be conclusively pinned down.

It would be interesting to investigate the usefulness of these arrangements for other image browsing tasks, with different user populations, particularly if they were incorporated into a real system. For example, it is possible that arrangements based on visual similarity will be less useful when the images are of a lower technical quality than those used here.

ACKNOWLEDGEMENTS

We are grateful to Will Hill of APU, Roz McCarthy of the University of Cambridge, the organisers of infodesign 99, and our experiment participants. Rachel Hewson provided helpful comments on this paper. Kerry Rodden's work was supported by the UK EPSRC, and by AT&T Laboratories Cambridge.

REFERENCES

1. Armitage, L.H., and Enser, P.G.B. Analysis of user need in image archives. *Journal of Information Science* 23(4), 1997, 287–299.
2. Basalaj, W. Proximity visualisation of abstract data. PhD thesis, University of Cambridge Computer Laboratory, 2000.
3. Borg, I., and Groenen, P. Modern multidimensional scaling. New York: Springer-Verlag, 1997.
4. Borlund, P., and Ingwersen, P. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation* 53(3), 1997, 225–250.
5. Chalmers, M., Ingram, R., and Pfranger, C. Adding imageability features to information displays. *Proc. UIST'96*, ACM, 1996.
6. Chen, C, and Czerwinski, M. Spatial ability and visual navigation: an empirical study. *The New Review of Hypermedia and Multimedia*, 3, 67–89.
7. Combs, T.T.A., and Bederson, B.B. Does zooming improve image browsing? *Proc. Digital Libraries '99*, ACM, 1999.
8. Garber, S.R, and Grunes, M.B. The art of search: a study of art directors. *Proc. CHI'92*, ACM, 1992.
9. Jose, J.M., Furner, J., and Harper, D.J. Spatial querying for image retrieval: a user-oriented evaluation. *Proc. SIGIR'98*, ACM, 1998.
10. Leuski, A., and Allan, J. Improving interactive retrieval by combining ranked lists and clustering. *Proc. RIAO 2000*.
11. Markkula, M., and Sormunen, E. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval* 1(4), 2000, 259–285.
12. Raven, J.C. Raven's Advanced Progressive Matrices. Psychological Corporation, http://www.psychcorp.com.
13. Rodden, K. Evaluating user interfaces for image browsing and retrieval. PhD thesis, University of Cambridge Computer Laboratory, 2001.
14. Rodden, K., Basalaj, W., Sinclair, D., and Wood, K. Evaluating a visualisation of image similarity as a tool for image browsing. *Proc. InfoVis'99*, IEEE, 1999.
15. Rodden, K., Basalaj, W., Sinclair, D., and Wood, K. Evaluating a visualisation of image similarity. *Proc. SIGIR'99* (poster), ACM, 1999.
16. Rose, T., Elworthy, D., Kotcheff, A., Clare, A., and Tsonis, P. ANVIL: a system for the retrieval of captioned images using NLP techniques. *Proc. CIR2000*, BCS (http://www.ewic.org.uk), 2000.
17. Rubner, Y., Tomasi, C., and Guibas, L.J. Adaptive color-image embeddings for database navigation. *Proc. Asian Conference on Computer Vision*, IEEE,1998.
18. Swan, R.C., and Allan, J. Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. *Proc. SIGIR'98*, ACM, 1998.
19. van Rijsbergen, C.J. *Information Retrieval*. London: Butterworths, 1979.