Visual Clustering of Image Search Results

Trystan G. Upstill[a], Rajehndra Nagappan[a] and Nick Craswell[b]

[a]The Australian National University, Canberra, Australia
[b]CSIRO Mathematical and Information Sciences, Canberra, Australia

ABSTRACT

This paper presents a novel method for visualizing the results of an image search. Current approaches to visualizing WWW image searches rank results in a linear list and present them as a sorted thumbnail grid. The method outlined in this paper visually clusters images based on the user's search terms. To accomplish this, a flexible image retrieval method which incorporates a combination of content-based and textual image matching is used. A new information visualization is used to display the search results.

In our model multiple types of partitioning and querying can occur concurrently, thereby creating a multi-dimensional display of image properties. The display groups similar images, enabling users to quickly scan for the most relevant images. This visualization allows users to exploit the location of images as their guide to what an image contains and use thumbnails to preview potentially relevant images. Through the identification of relevant images users can locate relevant areas in the visualization. It is then possible for users to focus their attention on one area of the visualization using a zooming function. The user's interaction with the system is explored using new evaluation metrics based on Information Foraging theory.

Keywords: Information Visualization, Visualization Evaluation, Information Foraging, World-Wide Web Image Searching

1. INTRODUCTION

The number of digital images is expanding rapidly and the World-Wide Web (WWW) has become the dominant medium for their transferral. Consequently, there exists a requirement for effective WWW image retrieval. While several systems exist, they lack the facility for expressive queries and provide an uninformative non-interactive grid interface.

Thumbnail grids are often used for viewing image search results. Thumbnail grids are linear lists of images split horizontally between rows, a process which is analogous to words wrapping on a page of text. Images positioned horizontally next to each other are adjacent in the ranking. Thus, although the grid is two dimensional, thumbnail grids only represent a single dimension — the system's ranking of images.

WWW image searching can be thought of in terms of a vector space model. When using a vector space model with a thumbnail grid visualization, vector evidence is discarded. Figure 1 depicts a hypothetical thumbnail grid retrieved by an image retrieval engine for the query "clown, circus, tent". In the example there are three clusters, each containing multiple images, located in different directions from the query vector. When compressing this evidence the ranking algorithm selects images in order of their proximity, spiralling out from the target vector until the linear list is full. This discards image vector details, and leads to a thumbnail grid where similar images are not adjacent.

Further author information: (Send correspondence to T.G.U.)
T.G.U.: E-mail: upstill@ozemail.com.au
R.N.: E-mail: Raj.Nagappan@cs.anu.edu.au
N.C.: E-mail: Nick.Craswell@cmis.csiro.au

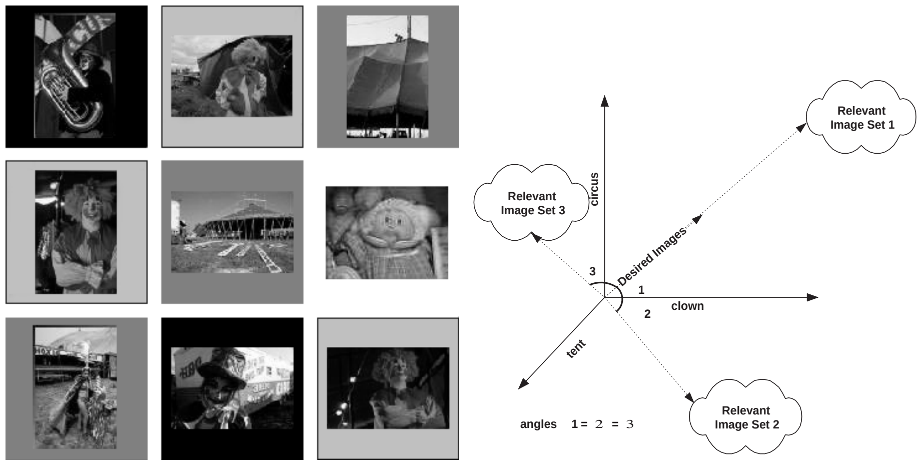

Figure 1. Example image grid & corresponding vector space. This example image grid is generated for the query "clown, circus, tent". Black images contain pictures of "circus clown"s, dark grey images contain pictures of "circus tent"s and light grey bordered images contain pictures of "clown tent"s. Similar images are not adjacent in the thumbnail grid. The corresponding vector space is illustrated on the right. The image set 1 contains the black images, image set 2 contains the dark grey images and image set 3 contains the light grey bordered images. The vector evidence is lost when compressing the ranking into a grid.

2. BACKGROUND

The visualization presented in this paper is a variation of the spring-based graph drawing algorithm. In the spring-based model documents are separated using document *discriminators*.[1] Each discriminator attracts matching documents as though attached by springs — the degree of attraction is proportional to the degree of match. This clusters the documents according to common discriminators. In this model the dimensions are compressed using spring attraction, with each spring direction representing a dimension. Systems that use this model include the VIBE system,[2–4] WebVIBE,[5–7] LyberWorld[8,9] and Mitre.[10] Two similar systems that deal with image searching are ImageVIBE[11] and MageVIBE.[12] Further evaluation is required to evaluate the effectiveness of these systems.

An example of current WWW image retrieval systems is the AltaVista Image Photo and Media Finder.[13] This image retrieval engine provides a simple text-based interface to an image collection indexed from the general WWW community and AltaVista's image database partners. AltaVista's image retrieval engine is based on their text document search engine. Modifications to this architecture have been made to associate sections of Web page text to images, in order to obtain image descriptions.

The following points are WWW image retrieval visualization problems. The problems are similar to the bibliographical information retrieval visualization problems outlined by Korfhage.[14]

- **System Heterogeneity**
  Executing a query over multiple search engines, or repeatedly over the same search engine, typically retrieves differing search results. This is due to continual changes in the image collections and differing confidential ranking algorithms used by WWW engines. Further, these algorithms sometimes vary according to system load. This variability can lead to confusing inconsistencies in image search results.

- **No Transparency**
  The linear result visualizations used by WWW image retrieval engines do not transparently reveal why images are being retrieved.[14,15] This limits the user's ability to successfully refine their query expression. Confusion is amplified if the meta-data upon which the ranking takes place is misleading.

- **No Relationships**
  Horizontal adjacency in the thumbnail grid result visualization used by WWW image retrieval engines represents

the linear ranking of images and vertical adjacency is meaningless (meaning only a gap of n ranks, where n is the width of the grid). Similar rank does not imply similar content. Therefore, once a relevant image has been found, the identification of more related images can be an arduous task.[14]

- **Reliance on Ranking Algorithms**
  The algorithms used to compress multi-dimensional query-document relationship information into a thumbnail grid are not well understood by users. Further, algorithms that incorporate different types of evidence, e.g. a combination of text and content analysis are difficult to develop and can cause further user confusion.[16,14,15]

- **Coarse Grained Interaction**
  In providing a search service over the Internet, current WWW image retrieval systems are limited to providing coarse grained interaction. Users submit a query, retrieve results and then choose either to enter a new query or perform a *find similar* search. Searching is an iterative process, requiring continual refinement and feedback.[15,17] These interfaces do not facilitate a high degree of user interaction during the image retrieval process.

- **Lack of Potential for Foraging Interaction**
  Information Foraging is a theory developed to understand the usage of strategies and technologies for information seeking, gathering, and consumption. To enable *effective* information foraging, a result visualization must allow users to locate patches of relevant information and then perform detailed analysis of the information contained within a patch.[18] In current WWW image retrieval engines, there is no grouping of like images, preventing the formation of patches which are high in relevant material. Further there is no way for users to view a subset of the retrieved information. Thus information foraging is not encouraged through the visualization.

3. A VISUALIZATION OF IMAGE SEARCH RESULTS (VISR)

3.1. System Overview

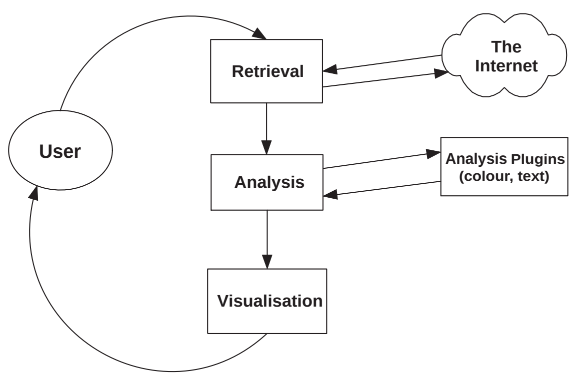

Figure 2. VISR system overview

Figure 2 contains an abstract overview of the VISR system architecture. Following a user image query AltaVista is used to retrieve the initial image set. The image set and associated text is downloaded and forms the image domain. This domain is analyzed using the image plugins and the VISR visualization is generated. A screenshot of the VISR visualization for the query "eiffel 'object oriented' book" can be seen in figure 3. There are four distinct clusters in this visualization, that of the "eiffel tower", that of " 'object oriented' book"s, that of " 'object oriented' signs" and a group in the middle of "eiffel 'object oriented' books".

3.2. Spring-based Image Position Calculation

The visualization is based on a spring model developed by Olsen and Korfhage.[2] This technique was adapted for *RadViz*.[19] In RadViz reference points are equally spaced around the perimeter of a circle. The data items are then distributed in the circle according to their attraction to the reference points.

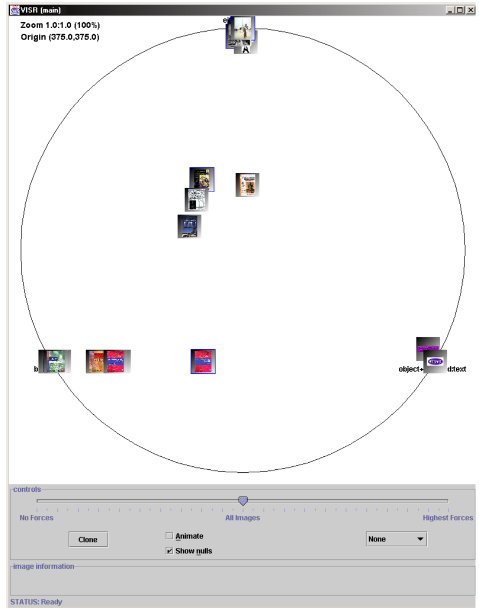

Figure 3. VISR search: "eiffel 'object oriented' book"

In VISR, the distribution occurs through having the user's query terms apply forces to the images in the collection. Springs are attached such that each image is connected to every query term and images are independent of each other. The query terms remain static while the images are pulled towards them according to how relevant the query terms are to the image. When these forces reach an equilibrium the images are in their final positions. The model of this visualization can be seen in figure 4.

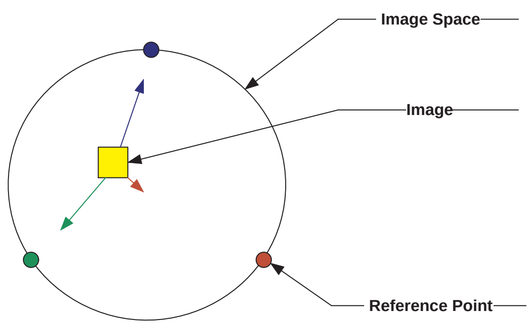

Figure 4. The spring-based visualization. In this example the image is attracted to all three reference points. It matches the black and light grey reference points more than the dark grey reference point. Arrows are not generated on the visualization, they serve in this diagram as a visual aid.

In the spring-based visualization metaphor images have no attraction to the center of the visualization, and are pulled freely towards whatever query terms they contain.

This can be modelled by the equation:

$$\overline{p}_s \text{such that} \sum_{i=1}^{n} a_i(|\overline{p}_s - \overline{q}_i|) = 0,$$

where $\overline{p}_s$ is the vector position of an image, n is the number of query terms, a_i is the scalar attraction to query term i, $\overline{q}_i$ is the vector position of query term i and $|\overline{p}_s - \overline{q}_i|$ is the net force $\overline{F}$. $\overline{F}$ moves $\overline{p}_s$ until $\overline{F}$ converges to 0. This is implemented using Eades' spring algorithm.[20] This gives the final value of $\overline{p}_s$.

3.3. Image Location Conflict Resolving

VISR incorporates two techniques to deal with overlapping images: Jittering, which shuffles overlapping images apart; and Animation, where overlapping images are animated with a specified delay from one overlapping image to the next. Additionally, zooming can be used to further alleviate the problems of image location conflicts (see section 3.5).

3.4. Query Term Moving

Users are able to move query terms around the circumference of the visualization circle. When placed near to but not on the circle, the query terms *snap to* the closest position on the circumference of the circle. The snap to location is established through the examination of the angle generated by the query term movement. The visualization is regenerated immediately after query term movement to reflect the new visualization. The movement of query terms can be used to compress dimensions through the placement of multiple query terms in the same location.

3.5. Zooming

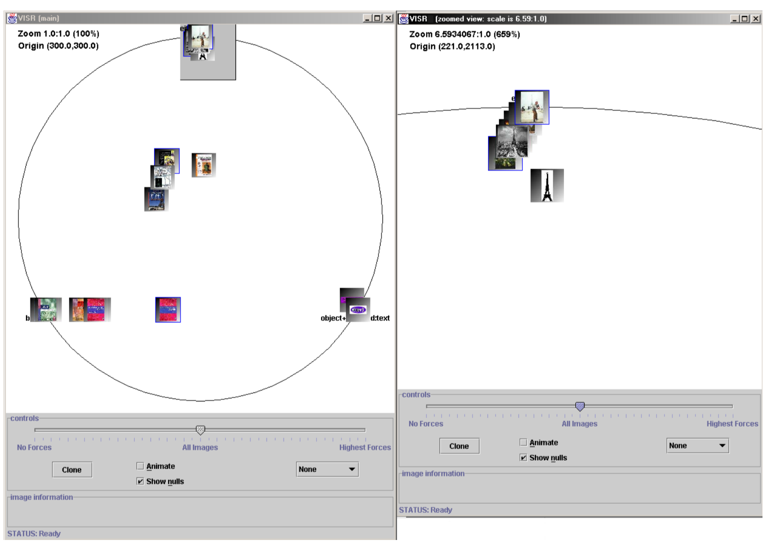

Figure 5. VISR search: "eiffel 'object oriented' book" - zooming

To zoom, the user selects an area of the visualization and a new window is created with the selected area maximized. The zoom factor is determined by the area selected by the user. The new origin of the visualization becomes the center of the box drawn by the user. When zooming, the image size is scaled by a lesser factor than the area. This provides increased separation between images in the selected area, while maintaining visualization accuracy. Figure 5 is a screenshot of zooming in VISR. In the example the user has selected the area surrounding the 'eiffel' query term in the left window. The selected area is highlighted and a new visualization window is created (the right window). The new visualization window is a zoom of the grey area and contains all the images in the selected area, magnified with a larger spread.

4. EVALUATION

4.1. Visualization Entropy

Visualization Entropy is used to gauge the consistency of a visualization after changes to the underlying document collection. An increase in entropy implies an increase in variation between visualizations.

The visualization entropy formula is:

$$E = \frac{\sum_{i=1}^{n} |v1_i - v2_i|}{n},$$

where E is the visualization entropy in terms of image positions moved, n is the number of images common to both visualizations, $v1_i$ is the position of image i in the first visualization and $v2_i$ is the position of image i in the second visualization.

Visualization entropy compares the consistency of VISR and thumbnail grid visualizations. A thumbnail grid and VISR visualization were generated for two document collections retrieved using the same query at different times, with a three month gap in between searches. The image collection indexed by the WWW image retrieval engine is continually changing, as such, the two retrieved document collections contained differing images.

Visualization Method	Visualization Entropy
Thumbnail Grid	7.2
VISR Visualization	0

Table 1. Summary of visualization entropy results.

The results of the visualization entropy experiment are given in table 1. The position of common images in the thumbnail grids changed, while remaining constant in the VISR visualization. In the VISR visualization all images are ranked independently, with image rankings not affecting each other. However, in the thumbnail grids when the position of one image changes, the change is propagated to all images below* that image. These results demonstrate VISR's consistent ranking of images compared to the thumbnail grid's volatile ranking.

4.2. Visualization Precision

Visualization Precision is an extension of the precision ranking measure in document retrieval, e.g: TREC evaluations.[21] Rather than measuring precision of relevant retrieved documents, this measure aims to gauge the precision of the clustering algorithm.

The cluster space is a minimum bounding box around all images in a visualization that are relevant to the user. A box is used to bound the cluster as it is analogous to the VISR zooming function. The cluster space may contain both relevant and irrelevant images. c^r is defined as the number of images relevant to a user in a cluster space, c^i is defined as the number of images irrelevant to a user in a cluster space

The visualization precision is: the number of relevant images in a cluster, c^r divided by the total number of images

*to the right or underneath

in the cluster $c^r + c^i$, or $V = \frac{c^r}{c^r+c^i}$. This is similar to the measure of document cluster precision by Pirolli and Card.[22]

This can be extended to include partial clusters. Given c^r as the total number of relevant images in the cluster space, c_p^r is now introduced as number of relevant images at a percentage p of the cluster space. The percentage of the cluster space is the percentage of all relevant images available that must be enclosed in the minimum bounding box. An example of the calculation of visualization precision for $p = 100\%$, $p = 80\%$ and $p = 50\%$ is illustrated in figure 6.

The revised formula for visualization precision is then:

$$V_p = \frac{c_p^r}{c_p^r + c_p^i},$$

where p percentage of relevant images, V_p is the visualization precision at percentage p, c_p^r number of relevant images at percentage p and c_p^i is the number of irrelevant images in the cluster at percentage p.

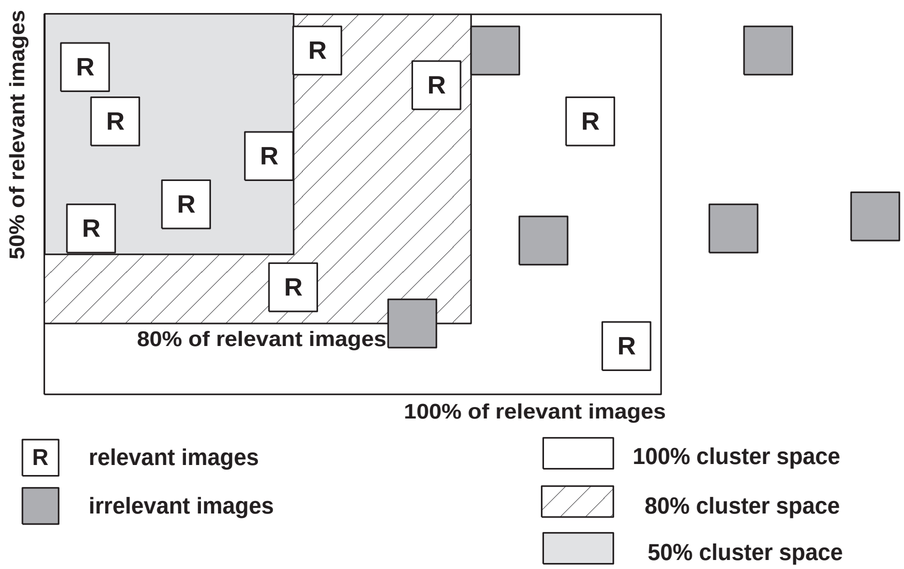

Figure 6. Cluster space example for $p = 100\%$, $p = 80\%$ and $p = 50\%$. Relevant images are represented by white boxes marked with an 'R', while irrelevant images are depicted as grey boxes. For $p = 100\%$ a minimum bounding box is drawn around all relevant images in the visualization. There are three irrelevant images in the cluster space at $p = 100$ — $c_{100}^r = 10$, $c_{100}^i = 3$, therefore $V_{100} = \frac{10}{13}$. For $p = 80\%$ a minimum bounding box is drawn around 80% of the relevant images, or 8 of the images in the visualization. There is only one irrelevant image in the cluster space at $p = 80\%$ — $c_{80}^r = 8$, $c_{80}^i = 1$ therefore, $V_{80} = \frac{8}{9}$. For $p = 50\%$ a minimum bounding box is drawn around 50% of the relevant images, or 5 of the images in the visualization. There are no irrelevant images in the cluster space at $p = 50\%$. $c_{50}^r = 5$, $c_{50}^i = 0$ therefore $V_{50} = \frac{5}{5} = 1$.

This measure is useful for determining the effectiveness of clustering on noisy data by shrinking the bounding box and discarding outlying images.

4.2.1. Multiple cluster evaluation

The multiple cluster evaluation measures the effectiveness of the VISR tool in clustering all the image groups in the retrieved collection.

Method:

1. A candidate image is picked from an image set
2. All other images in the set are judged as relevant or irrelevant to a user who finds the candidate image relevant
3. Both VISR and thumbnail grid visualizations are created
4. Bounding boxes for the specified percentage of relevant images in each visualization are drawn and precision is measured

This process is performed with multiple candidate images concurrently.

The visualization precision is calculated at a cluster space of p=50-100%. This evaluation was performed for 3, 4 and 5 term queries using both VISR and the thumbnail grid. The results for this experiment are shown in figure 7. This graph shows that VISR had the best precision when using 3 query terms, with 81% precision at 100% of relevant images, and 100% precision at 50% of relevant images. The least effective visualization was the thumbnail grid, with 30% precision at 100% of relevant images, and 34% precision at 50% of images. In information foraging the profitability of a search is measured by the quality of results with regard to the amount of time spent. The plots from the VISR visualization reveal little differentiation between profitability and time spent, until 90-100% where performance suffers due to noise. For the evaluation of document collections, search profitability is maximized by bounding 90% of the relevant images. The grid, however, has a fairly random cluster structure, where the gradient oscillates implying that there is no optimally profitable search pattern.

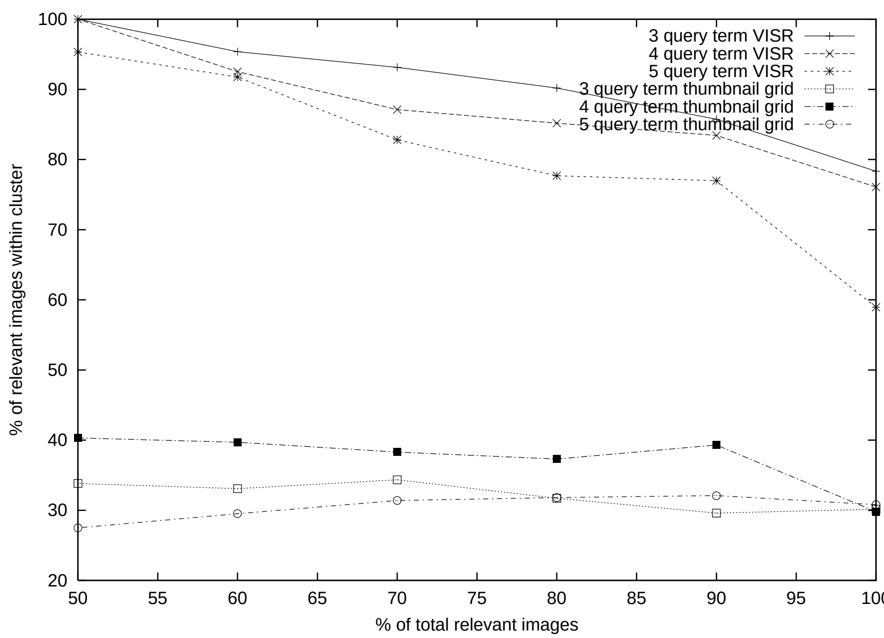

Figure 7. Multiple cluster evaluation results. Note that all gradients decrease between 90 and 100% indicating noisy images are present in the image collection.

4.3. User Study

It is difficult to objectively compare visualization techniques using user studies.[23] Aesthetic visualization properties make it hard to separate user subjective evaluations from objective analysis. As a result, much information visualization research neglects comprehensive user evaluation. Previous work has shown that testing user interaction with an interface is not a coherent measure of visualization clarity but rather interface usability.[7] Morse and Lewis evaluated the performance of core visualization features through the use of *de-featured interfaces* and had positive results.[5] These de-featured interfaces tested the underlying visualization metaphors through a paper-based user study. Users were not required to interact with the system.

The user studies for the VISR tool are paper-based. This decouples the examination of the visualization clarity from the interface effectiveness. This experiment provides a preliminary evaluation as to whether users can understand the visualization metaphor used in the VISR tool.

Three measures are tested for 3, 4 and 5 query terms:

- User determination of all of an image's query term matches (All).
- User determination of unrelated query terms to an image (Not).
- User determination of the single most related query terms to an image (Most).

Eleven representative users were given an open-ended survey with 9 VISR visualizations. All surveys were unique with three random images highlighted in each visualization. Users were asked to draw conclusions as to the above measures. Figure 8 contains a histogram of the results.

The preliminary results show that users' performance degrades with an increase in query terms. However, users were able to determine the most strongly matching query term for an image irrespective of query term number.

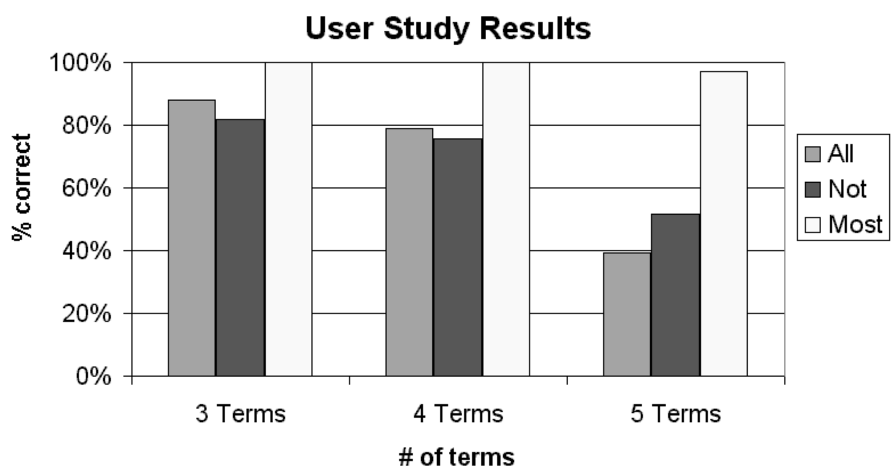

Figure 8. Bar graph of user study results. For each number of query terms users were able to determine the most relevant images. When dealing with over four query terms identification of relevant query terms dropped by 40%.

5. DISCUSSION

System Homogeneity
When the visualization is generated, images are distributed in a consistent manner. This implies that if two search engines returned the same image, the images would be co-located in the display. Through the use of consistent plugins for retrieval and analysis, and the transparent clustering visualization, the VISR tool reduces the effects of system heterogeneity. The visualization entropy experiment showed that common images were displayed in the same location after changes to the underlying image collection. In the VISR tool, documents are always ranked in the same manner and placed at the same position in the visualization.

Transparency
The visualization facilitates user understanding of why images are retrieved and which query terms matched which images. A key issue in image retrieval is how images are perceived by users.[15] Educating users about the retrieval process assists them to understand how the system is matching their queries, and thus how they should form and refine their queries. The visualization user study showed that users are able to interpret image collections using the VISR tool. A large percentage of users were successful in determining complete image associations for 3 and 4 query terms. Queries that contain more than 4 query terms can be viewed transparently through the movement of query terms which dynamically compresses dimensions.

Relationship Maintenance
The maintenance of image relationships in the visualization is achieved through clustering. This allows users to find similar images quickly. The effectiveness of clustering is shown through the multiple cluster evaluation. In this evaluation the VISR tool outperformed the grid approach clustering images with a visualization precision of over double that of the thumbnail grid.

Less Reliance on Ranking Algorithms
The display based on per-term ranking information reduces the reliance on ranking algorithms. In the visualization there is no combination of evidence. Our implementation included both color histogram and text analysis and demonstrates the combination of these techniques into the single visualization.

Fine Grained Interaction
Finer grained interaction is facilitated through local analysis and visualization. When interacting with the dynamic query interface the user's changes are reflected immediately in the visualization. All tasks that do not require new images to be retrieved are completed with low latencies. Thus features such as dynamic filters, query re-weighting and zooming can be implemented efficiently. Further evaluations are required to evaluate the effectiveness of the dynamic query interface.

Foraging Interaction Possibilities
Foraging interaction[18] is encouraged though the visualization's ability to cluster and zoom. Between-patch foraging, the search for a relevant cluster, is aided through the grouping of similar images. Within-patch foraging, the search for relevant images within a relevant cluster, is facilitated through the ability to zoom in and examine a single cluster in greater detail. Through zooming users are able to perform a more thorough investigation of the images contained within a cluster.

6. CONCLUSION

In an attempt to resolve WWW image search difficulties this paper presented a new visualization for image search results. System clarity was improved through a new result visualization, which elucidates why images are returned and how they matched the query. System control was improved by enhancing user-system interaction.

There were no existing metrics for the evaluation of such a system. This paper proposed two new evaluation measures: visualization entropy and visualization precision. Visualization entropy was created to measure visualization consistency. The visualization precision measure was created to determine cluster accuracy.

The preliminary results using these new measures showed that the VISR tool improves upon traditional WWW image retrieval systems. The clustering evaluations using visualization precision show how VISR clusters images more effectively than the thumbnail grid. The visualization entropy experiment demonstrated the stability of VISR over changing data sets. A user study demonstrated that the spring-based visualization metaphor used by VISR is easily understood.

REFERENCES

1. D. Dubin, "Document Analysis for Visualization," in *Proceedings of the eighteenth annual international ACM/SIGIR conference on research and development in information retrieval*, pp. 199–204, 1995.

2. K. Olsen, R. Korfhage, M. Spring, K. Sochats, and J. Williams, "Visualization of a Document Collection with Implicit and Explicit Links: The VIBE System," *The Scandinavian Journal of Information Systems* , pp. 79–95, August 1993.
3. D. Dubin, *Structure in Document Browsing Spaces.* PhD thesis, University of Pittsburgh, 1996.
4. P. B. Heidorn, "Development and Testing of a Visual Information Retrieval Environment," tech. rep., Graduate School of Library and Information Science, University of Illinois, 1998.
5. E. L. Morse and M. Lewis, "Why Information Retrieval Visualizations Sometimes Fail," in *Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1680–1685, October 1997.
6. E. Morse, M. Lewis, R. Korfhage, and K. Olsen, "Evaluation of Text, Numeric and Graphical Presentations for Information Retrieval Interfaces: User Preference and Task Performance Measures," in *Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1026–1031, October 1998.
7. E. L. Morse, *Evaluation of Visual Information Browsing Displays.* PhD thesis, School of Information Sciences, University of Pittsburgh, 1999.
8. M. Hemmje, C. Kunkel, and A. Willett, "LyberWorld - A Visualization User Interface Supporting Fulltext Retrieval," in *Proceedings of the seventeenth annual international ACM/SIGIR conference on Research and development in information retrieval*, pp. 249–259, 1994.
9. M. Hemmje, "LyberWorld - A 3D Graphical User Interface for Fulltext Retrieval," in *Proceedings of the CHI'95 conference*, pp. 417–418, May 1995.
10. M. Konchady, R. D'Amore, and G. Valley, "A Web Based Visualization for Documents," in *Proceedings of NPIV'98*, pp. 13–19, 1998.
11. L. Cinque, S. Levialdi, A. Malizia, and K. A. Olsen, "A Multidimensional Image Browser," *Journal of Visual Languages and Computing* **9**, pp. 103–117, 1998.
12. R. R. Korfhage and V. Kolluri, "MageVIBE: A Multimedia Database Browser," tech. rep., Visual Information Retrieval Interfaces Lab, Department of Information Science, University of Pittsburgh, 1996.
13. AltaVista Company, "Altavista," 2000. http://www.altavista.com/ accessed on the 29th October 2000.
14. R. R. Korfhage, "To see, or Not to See - Is That the Query," in *Proceedings of the fourteenth annual international ACM/SIGIR conference on Research and development in information retrieval*, pp. 134–141, 1991.
15. T. Huang and Y. Rui, "Image Retrieval: Past, Present, and Future," in *Proc. of Int. Symposium on Multimedia Information Processing*, December 1997.
16. P. Aigrain, H. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review," in *Multimedia tools and applications*, vol. 3, pp. 179–202, 1996.
17. J. P. Eakins and M. E. Graham, "Content-based Image Retrieval: A report to the JISC Technology Applications Programme," tech. rep., Institute for Image Data Research, University of Northumbria at Newcastle, January 1999.
18. P. Pirolli and S. K. Card, "Information Foraging," in *Psychological Review*, January 1999.
19. P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley, "DNA Visual and Analytical Data Mining," in *Proceedings of the 8th IEEE Visualization '97 Conference*, 1997.
20. P. Eades, "A heuristic for Graph Drawing," *Congressus Numerantium* **42**, pp. 149–160, 1984.
21. TREC, "Trec overview," August 2000. http://trec.nist.gov/overview.html accessed on the 29th of October, 2000.
22. P. Pirolli and S. Card, "Information Foraging in Information Access Environments," in *Conference proceedings on Human factors in computing systems CHI'95*, 1995.
23. A. Globus and S. Uselton, "Evaluation of visualization software," *Computer Graphics* **29**(2), pp. 41–44, 1995.