Image Visualization: *Generalized Similarity Analysis* Revisited

Chaomei Chen[1], George Gagaudakis[2], Paul Rosin[2]

[1]Department of Information Systems and Computing
Brunel University
Uxbridge, UB8 3PH, UK

[2]Department of Computer Science
Cardiff University
Newport Road, Cardiff Wales,
CF24 3XF UK

E-mail: Chaomeni.Chen@brunel.ac.uk , {George.Gagaudakis , Paul.Rosin}@cs.cf.ac.uk

Abstract
In conventional image retrieval systems, images are typically characterized by a range of features such as color, texture, and shape. A fundamental question is to what extent these low-level features effectively convey visual similarities of images to human users and enable users to explore image databases intuitively. In this article, we examine the potential of Generalized Similarity Analysis (GSA) in image visualization based on such features. Salient structures of images are visualized according to features extracted from color, texture, and shape orientation. Implications for visualizing and constructing hypermedia systems are discussed.

Keywords

Information visualization, content-based image retrieval, Pathfinder networks

1. INTRODUCTION

Content-based image retrieval has been a highly active field of research [1, 2]. A number of widely known image retrieval systems have been developed over the last few years, notably, IBM's QBIC [3], PhotoBook [4], ImageRover [5], and Webseek [6]. In these systems, images are typically characterised by attributes known as features, ranging from simple, low-level ones such as color and texture, to more complex, relatively higher-level ones such as shape and other semantically rich query classes. Little is known whether it is viable to combine these feature-extraction algorithms with information visualization techniques so that users can explore images in a digital library intuitively.

Ultimately, feature-extraction techniques, combined with other techniques, are expected to narrow down the gap between relatively primitive features extracted from images and high-level, semantically-rich perceptions by humans so that users will be able to find the right images more easily and intuitively.

The advances of information visualization and data mining techniques now allow users to explore an information space organized through a variety of metaphors, such as an information landscape or an information galaxy [7, 8]. Many of these visualizations are based on interrelationships derived from textual information, typically using classic information retrieval models such as the vector space model [9], Latent Semantic Indexing (LSI) [10], or other variants. Mapping the structure of a document collection into a high-dimensional vector space also lent itself to some information visualization and layout generation techniques, notably the use of the spring-embedder model and other physical systems. There has been a steady increase in the interest in this type of layout and visualization techniques, which tend to place similar objects near to each other and separate dissimilar objects far apart in the visualization space.

The work described in this article extends our earlier work in structuring and analyzing the design of various information visualization displays. We have gathered computer-generated images of a variety of information visualizations [11]. In particular, we have visualized image networks based on similarity measures produced by IBM's QBIC system [12], including color, layout, and texture.

Researchers and practitioners in information visualization often need to find an optimal way to arrange various visualization images so that design patterns and trends will become apparent. Ideally, images of similar layouts, spatial properties, or overall shapes should be closely grouped together. Users should be able to explore and compare images within such structures.

Generalized Similarity Analysis (GSA) is a generic framework developed for structuring and visualizing information spaces [13, 14]. Applications of GSA include visualization of university websites, online conference proceedings, and journals in digital libraries according to a variety of similarity measures, such as term-frequences, hypertext reference links, author co-citation profiles, and browsing trails of users. A key element in GSA is the use of Pathfinder network scaling technique to extract the most salient links and eliminate redundant or counter-intuitive links [15]. Pathfinder has some desirable features over techniques such as multidimensional scaling (MDS), for example, Pathfinder networks present a more accurate local structure.

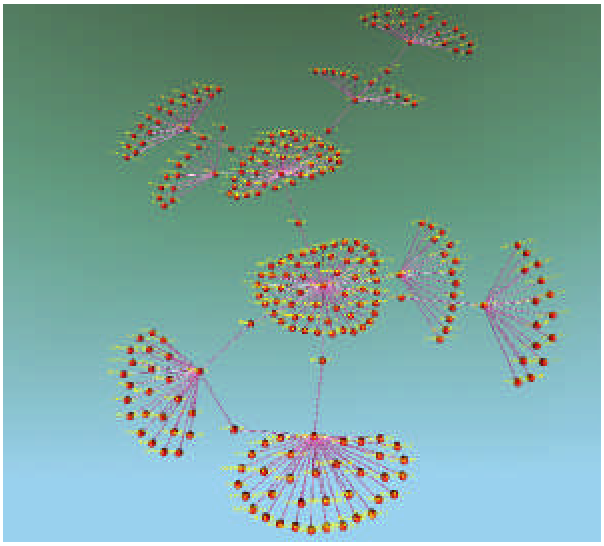

Figure 1. GSA and citation analysis. This is a network of authors derived from their citation profiles based on *International Journal of Human-Computer Interaction* over the last 10 years.

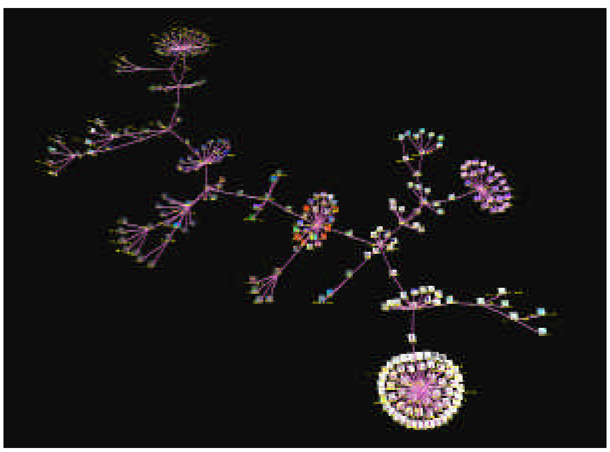

Figure 2. An earlier example of using GSA to visualize a smaller collection of images based on features extracted by IBM's QBIC system.

In this article, we revisit GSA in order to explore a synergy between Pathfinder network scaling and CBIR techniques to enable users to explore a collection of images according to their content similarities.

The rest of this article is organised as follows. First, the feature-extraction techniques to be used are introduced in more detail. Second, a brief history of using Pathfinder networks in information visualization is provided to form a wider context. Then, three sets of search results are included to illustrate the effects of four feature-extraction techniques. Subsequently derived Pathfinder networks are examined and discussed. Finally, implications of the synergy for visualizing and constructing hypermedia systems are discussed.

2. CONTENT-BASED RETRIEVAL

The key issue in CBIR is how to match two images according to computationally extracted features. Typically, the content of an image can be characterised by a variety of visual properties known as features. It is common to compare images by colour, texture, and shape, although these entail different levels of computational complexity. Colour histograms are much easier to compute than a shape-oriented feature extraction.

Most content-based image retrieval techniques fall into two categories: manual and computational [2]. In manual approaches, a human expert may identify and annotate the essence of an image for storage and retrieval. For example, radiologists often work on medical images marked and filed manually with a high degree of accuracy and reliability.

Figure 3. Manually clustered 279 computer-generated images.

Computational approaches, on the other hand, typically rely on feature-extraction and pattern-recognition algorithms to match two images. Feature-extraction algorithms commonly match images according to the following attributes, also known as query classes:

- color
- texture
- shape
- spatial constraints.

A robust CBIR technique should support a combination of these query classes. Ideally, users should be able to use high-level and semantically-rich image query classes, such as human facial expressions, in their image retrieval. However, the reliability of today’s feature-extraction techniques has yet to reach such a level of satisfaction. This is partially why simpler, and relatively low-level feature-extraction techniques are still being widely used and continuously developed. The background of the four feature-extraction algorithms to be used in our study is explained as follows.

2.1 Colour

Swain & Ballard [16] matched images based solely on their colour. The distribution of colour was represented by colour histograms, and formed the images' feature vectors. The similarity between a pair of images was then calculated using a similarity measure between their histograms called the *normalised histogram intersection*. This approach became very popular due to the following advantages:

Robustness. The colour histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled [16]. It is also insensitive to changes in image and histogram resolution and occlusion.

Effectiveness. There is high percentage of relevance between the query image and the extracted matching images.

Implementation simplicity. The construction of the colour histogram is a simple scanning of the image, to get the colour values, discretisation of the colour values to the resolution of the histogram, and building the histogram using colour components as indices.

Computational simplicity. The histogram computation has $\boldsymbol{O}(\boldsymbol{M}^2)$ complexity for images of size $\boldsymbol{M}\times\boldsymbol{M}$. The complexity for a single image match is linear, $\boldsymbol{O}(\boldsymbol{n})$, where $\boldsymbol{n}$ represents the number of different colours, or resolution of the histogram.

Low storage requirements. The colour histogram size is significantly smaller than the image itself, assuming colour quantisation.

Differentiating from the original proposal, towards a more compact colour representation, we used the 11 colour labels as obtained by the anthropological study of Berlin and Kay on colour terms in 100 different languages [17].

2.2 Texture

A common extension to colour-based feature extraction is to add textural information. There are many texture analysis methods available, and these can be applied either to perform segmentation of the image, or to extract texture properties from segmented regions or the whole image. In a similar vein to colour-based feature extraction, we modified the standard cooccurrence method in order to produce texture histograms with an additional degree of rotation invariance. The modified method, called the *circular cooccurrence* matrix, is described in [18].

In general, texture-based feature extraction tends to provide more spatial information than color histograms. In order to find out more about the content of an image, one may consider features associated with shapes. For example, the presence of edges, edge orientation, and edge distance may lead to a more accurate match of images.

2.3 Shape

Shape extraction remains a challenging to feature-oriented approaches. Several methods have been developed for detecting shapes indirectly. Whereas it tends to be extremely difficult to perform semantically meaningful segmentation, many reasonably reliable algorithms for low-level feature extraction have been developed. These will be used to provide the spatial information that is lacking in colour histograms.

Rather than attempt to directly measure shape we will calculate some simpler properties that are indirectly related to shape and avoid the requirement for good segmentation, providing a more practical solution.

Edge Orientation. Previous work in this area can be found in Jain and Vailaya's work [19]. They combined edge orientation histograms with colour histograms. These edge orientation histograms encode some aspects of shape information. As a result, image retrieval can be more responsive to the shape content of the images. Standard edge detection is sufficient for shape-oriented feature extraction (e.g. Canny's algorithm [20]). In addition, minor errors in the edge map have little effect on the edge orientation histograms. Unlike colour histograms, the orientation histograms are not rotationally invariant. Therefore the histogram matching process has to iteratively shift the histogram to find the best match.

A more important consideration is that the edge maps were thresholded by some unspecified means. For robustness an adaptive thresholding scheme should be used [21]. However, an alternative is to include all the edges and weight their contribution to the histogram by their magnitudes so as to reduce the contribution from spurious edges. This is the approach we take in the reported experiments.

Multi-resolution Salience Distance Transform. Another approach to including shape information is based on the distance transform (DT). The DT is a method for taking a binary image of feature and non-feature pixels and calculating at every pixel in the image the distance to the closest feature. Although this is a potentially expensive operation efficient algorithms have been developed that only require two passes through the image [22]].

To improve the stability of the distance transform, Rosin and West [23] developed an algorithm called the salience distance transform (SDT). In SDT, the distances are weighted by the salience of the edge, rather than propagating out Euclidean (or quasi-Euclidean) distances from edges. Various forms of salience have been demonstrated, incorporating features such as edge magnitude, curve length, and local curvature. The effect of including salience was to downplay the effect of spurious edges by soft assignment while avoiding the sensitivity problems of thresholding.

Segmentation by Thresholding. Partitioning based approaches as in [24] have been used to improve the performance of CBIR systems. Trying to avoid selection of rigid regions and true segmentation, we used the binary thresholding as a tool for partitioning.

The partitioning injects the spatial information into the analysis so that standard feature-based methods (e.g. non-spatial) can then be applied within each region. However, small changes in the threshold value may cause relatively large changes in resulting binary images. In order to overcome this potential drawback, we applied a *soft threshold* as introduced in [18] to generate similarity measures for the work reported in this article.

3. PATHFINDER NETWORKS

Pathfinder network scaling is a structural modelling technique originally developed for the analysis of proximity data in psychology [15]. We have adapted this modelling technique to simplify and visualise the strongest interrelationships in proximity data. The resultant networks are called Pathfinder networks (PFNETs).

The key to Pathfinder is the so-called triangular inequality condition, which can be used to eliminate redundant or counter-intuitive links. Pathfinder network scaling particularly refers to this pruning process.

The topology of a PFNET is determined by two parameters *r* and *q* and the resultant Pathfinder network is denoted as PFNET(*r*, *q*). The weight of a path is defined based on Minkowski metric with the *r*-parameter. The *q*-parameter specifies that the triangle inequality must be maintained against all the alternative paths with up to *q* links connecting nodes n_1 and n_k:

$$w_{n_1 n_k} \le \left(\sum_{i=1}^{k-1} w_{n_i n_{i+1}}^{r}\right)^{\frac{1}{r}} \qquad \forall\, k = 2, 3, \ldots, q$$

The least number of links can be achieved by imposing the triangular inequality condition throughout the entire network (q=N-1). In such networks, each path is a minimum-cost path.

Pathfinder network scaling is a central component of the GSA framework. GSA provides a flexible platform for us to experiment with a variety of structures, such as the vector-space model, LSI, and author co-citation networks [25].

3.1 Image Database

In this article, we use a collection of 279 information visualization images. A considerable number of these images are computer-generated graphics included in [11]. We apply the Pathfinder network scaling technique on image similarity data computed based on color labels, texture, shape orientation, and a combined feature classes. These similarity data are submitted to Pathfinder network scaling. All the Pathfinder networks described in this article are minimum-cost networks, i.e. PFNETs (r=∞, q=N-1). These Pathfinder networks are rendered as virtual reality models in VRML (Virtual Reality Modeling Language) for examination and evaluation.

4. PATHFINDER NETWORKS OF IMAGES

Five Pathfinder networks of images were generated based on similarity data derived from color labels, color with spatial injection through soft thresholding, texture, shape orientation, and the combined similarity scheme. In this article, we expanded QIBC-derived similarity measures reported in [12], to include relatively higher-level features such as shape orientation. We expected that images with similar structures and appearances should be grouped together in Pathfinder networks.

Figure 2 shows a screenshot of image visualization based on a combination of color labels, texture histogram, and shape orientation. The layout reveals 7 apparent clusters. Images within each cluster appear to be homogenous, except the largest cluster, in which the color patterns of images appear to be mixed.

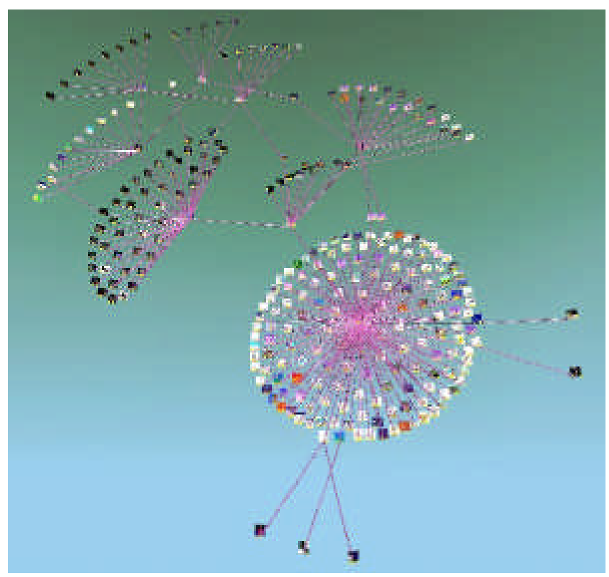

Figure 4. A Pathfinder network of the same 279 images generated by similarity measures drawn from a combination of color labels, texture, and shape histograms.

Figure 3 includes 6 sub-figures corresponding to 6 different clustering schemes, namely, manual, combined, color labels, color labels with spatial injection through soft thresholding, texture, and shape orientation. The combined scheme generated the best result, whereas the shape orientation did not reveal any clear sub-structures. Pathfinder network scaling on the shape orientation scheme along was not as effective as with the combined scheme.

Color labels with spatial injection appeared to generate a slightly better clustering pattern than the pure color label solution, in terms of the number of clusters and the homogeneity of clusters.

The Pathfinder network corresponding to the texture-based feature-extraction scheme consists of three huge clusters. A possible explanation is that most of these images are generated by computer; therefore, they may share texture patterns to a considerable extent.

In order to understand further about the nature of the clustering patterns in these Pathfinder networks, we compared the network structures corresponding to the 5 grouping schemes used. The results are summarized in **Table 1**. Given that all the networks consist of the same set of images, the focus of the comparison was on the number of links in common between a pair of network structures. The assumption is if two networks have more than their share of links in common, then this commonality indicates that these two structures together reveal some valuable information. On the contrary, if two networks only have a number of links in common more or less by chance, then it is unlikely that these networks contain any information valuable.

Apart from the manual scheme, pure color label scheme generated the largest number of links: 338. The shape orientation scheme generated the least: 227. It is particularly interesting to note that color labels with spatial injection through soft thresholding scheme has the highest overlap rate with the manual scheme, in terms of the information (16.074). This measuring scheme should be further investigated in the future.

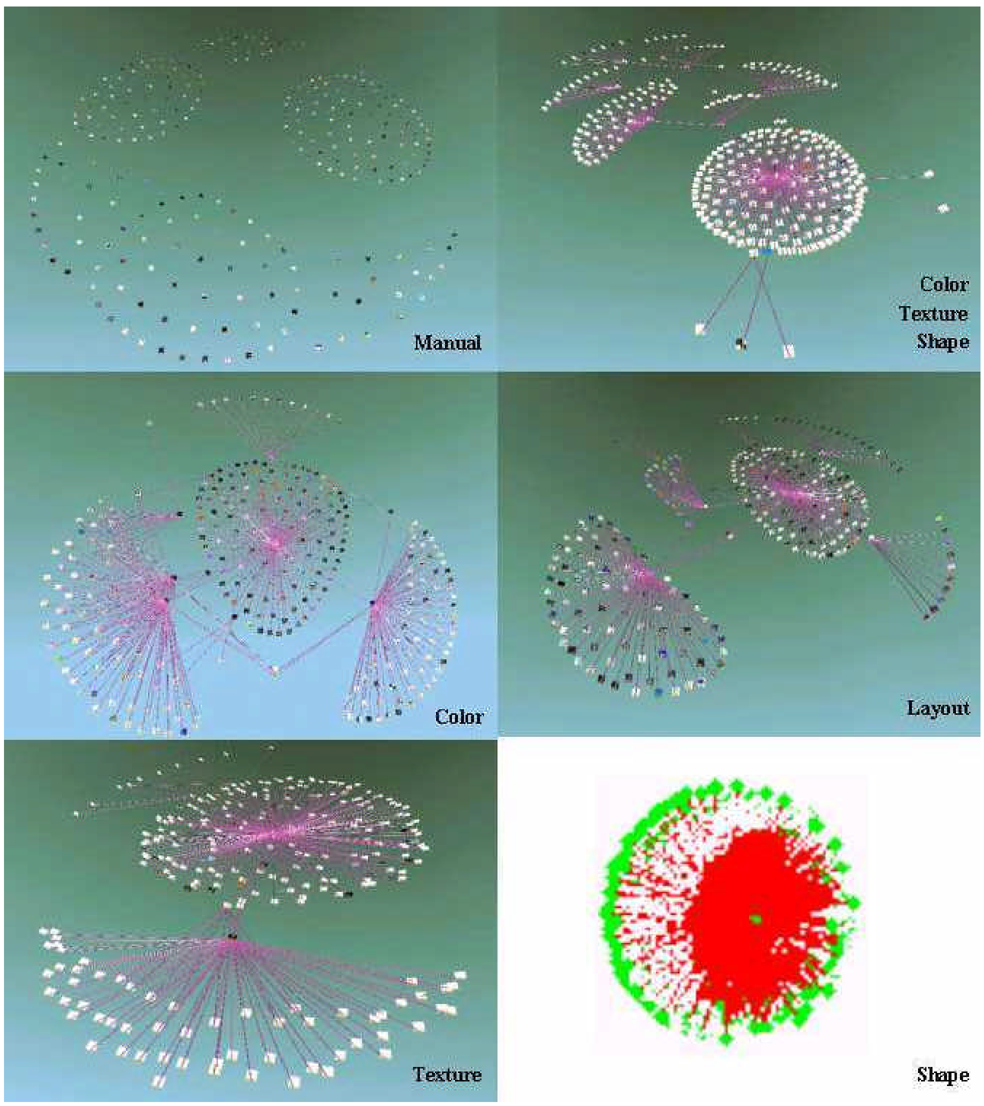

Figure 5. Pathfinder networks of the same 279 images by automatically extracted features.

5. DISCUSSION AND CONCLUSION

We have seen the results of applying the Pathfinder network scaling technique on various feature-extraction-based image matching schemes. On the one hand, incorporating shape-oriented feature-extraction algorithms appears to have improved the quality of image matching when combined with other schemes. We also identified that spatial injection to color label scheme yielded the highest overlap rate in terms of the network similarity.

In a long run, visualizing image clusters based on feature-extraction mechanisms remains a challenging field of research. In text-based information visualization, the groupings of concepts and documents can only be reliably verified by reading documents in questions. It would be difficult for users to look at an overview of a semantic network and tell us immediately whether a particular semantic network is meaningfully constructed. Unlike text-based information visualization, visualizing interrelationships among images has a unique advantage. Because humans can easily recognise visual patterns, it would be easier for users to detect discrepancies from a network of images than from a network of abstract concepts in text.

images 279	links		manual clusters	color texture shape	color	color layout	texture
manual clusters	8072						
color texture shape	284	common links	48				
		similarity	0.006				
		point probability	0.017				
		information	0.070				
color	338	common links	51	7			
		similarity	0.006	0.011			
		point probability	0.002	0.009			
		information	0.005	6.292			
color layout	280	common links	88	1	15		
		similarity	0.011	0.002	0.025		
		point probability	0.000	0.264	0.000		
		information	16.074	0.196	24.939		
texture	288	common links	46	10	139	1	
		similarity	0.006	0.018	0.285	0.002	
		point probability	0.008	0.000	0.000	0.260	
		information	0.026	14.055	716.538	0.190	
shape	227	common links	13	127	1	1	1
		similarity	0.002	0.331	0.002	0.002	0.002
		point probability	0.000	0.000	0.274	0.319	0.313
		information	0.000	729.229	0.211	0.308	0.292

Table 1. The similarity of network structures.

Compared to computational feature-extraction algorithms, human users may employ a much wider range of criteria to judge, compensate, or differentiate the similarity between two images. The integration of Pathfinder networks and some of the most commonly used feature-extraction schemes as presented in this article is only the first step towards the development of a comprehensive framework of visualizing and exploring hypermedia networks. Information visualization and feature-extraction techniques have the great potential to benefit tremendously from each other.

Clustering images has a wide range of potential applications, for example, data mining in remote sensing images and image retrieval from film and video archives. Most images in our sample are more likely to be different than similar. Such discreteness may obscure some otherwise obvious patterns in image groupings. We are applying this methodology to a sample of images with more continuous scenes, for example, video segments, in order to keep track of the impact of various feature-extraction techniques more closely (see examples in Figure 6).

Figure 6. The global view (left) and local view (right) of images sampled from a video clip.

Future work should address an optimal integration of feature-extraction techniques and other image indexing methods, especially meta-data approaches.

The integration of CBIR techniques and existing techniques in GSA so far provides additional tools for designers to organize images based on a variety of features for retrieval and browsing. Image indexing techniques described in this article have the potential to use generic visualization techniques to generate overviews of content-based image networks. Visualizations based on such content-based image indexing mechanisms may lead to more insights into emerging trends in information visualization.

6. ACKNOWLEDGMENTS

This study was in part supported by the British research council EPSRC (GR/L61088 and GR/L94628).

7. REFERENCES

[1] M. Marsicoi, L. Cinque, and S. Levialdi, "Indexing pictorial documents by their content: A survey of current techniques," *Image and Vision Computing*, vol. 15, pp. 119-141, 1997.

[2] V. Gudivada and V. Raghavan, "Content-based image retrieval systems," *IEEE Computer*, vol. 28, pp. 18-22, 1995.

[3] M. Flickner, H. Sawhney, W. Niblack, J. Sahley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Computer*, vol. 28, pp. 23-32, 1995.

[4] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: Tools for content-base manipulation of image databases," Proceedings of SPIE Conference on Storage and Retrieval of Image and Video Databases II, San Jose, CA, 1994.

[5] S. Sclaroff, L. Taycher, and M. LaCascia, "ImageRover: A content-based image browser for thr World Wide Web," Proceedings of IEEE Content-Based Access of Image and Video Libraries, 1997.

[6] J. R. Smith and S.-F. Chang, "Searching for images and video on the World Wide Web," *Multimedia Systems*, vol. 3, pp. 3-14, 1995.

[7] J. A. Wise Jr., J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow, "Visualizing the non-visual: Spatial analysis and interaction with information from text documents," Proceedings of IEEE Symposium on Information Visualization '95, Atlanta, Georgia, USA, 1995.

[8] H. Small, "Update on science mapping: Creating large document spaces," *Scientometrics*, vol. 38, pp. 275-293, 1997.

[9] G. Salton, J. Allan, and C. Buckley, "Automatic structuring and retrieval of large text files," *Communications of the ACM*, vol. 37, pp. 97-108, 1994.

[10] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, pp. 391-407, 1990.

[11] C. Chen, *Information Visualisation and Virtual Environments*. London: Springer-Verlag London, 1999.

[12] C. Chen, G. Gagaudakis, and P. Rosin, "Similarity-based image browsing," Proceedings of the 16th IFIP World Computer Congress, International Conference on Intelligent Information Processing, Beijing, China, 2000.

[13] C. Chen, "Generalised Similarity Analysis and Pathfinder Network Scaling," *Interacting with Computers*, vol. 10, pp. 107-128, 1998.

[14] C. Chen and L. Carr, "Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998)," Proceedings of the 10th ACM Conference on Hypertext (Hypertext '99), Darmstadt, Germany, 1999.

[15] R. W. Schvaneveldt, F. T. Durso, and D. W. Dearholt, "Network structures in proximity data," in *The Psychology of Learning and Motivation, 24*, G. Bower, Ed.: Academic Press, 1989, pp. 249-284.

[16] M. Swain and H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11-32, 1991.

[17] B. Berlin and P. Kay, *Basic Colour Terms: Their Universality and Evolution*: University of California Press, 1969.

[18] G. Gagaudakis and P. Rosin, "Incorporating shape into histograms for CBIR," *Patern Recognition*, To Appear.

[19] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, pp. 1233-1244, 1996.

[20] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679-698, 1986.

[21] P. L. Rosin, "Edges: Saliency measures and automatic thresholding," *Machine Vision and Application*, vol. 9, pp. 139-159, 1997.

[22] G. Borgefors, "Distance transformations in digital images.," *Computer Vision,*

Graphics, and Image Processing, vol. 34, pp. 344-371, 1986.

[23] P. L. Rosin and G. A. W. West, “Salience distance transform,” *Graphical Models and Image Processing*, vol. 57, pp. 483-521, 1995.

[24] M. Striker and A. Dimai, “Special covariance and fuzzy regions for image indexing,” *Machine Vision and Applications*, vol. 10, pp. 66-73, 1997.

[25] C. Chen, “Visualizing semantic spaces and author co-citation networks in digital libraries,” *Information Processing and Management*, vol. 35, pp. 401-420, 1999.