Similarity-Based Image Browsing

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
Cardiff Wales
CF24 3XF, UK

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

Abstract

Digital images and videos have an increasingly important role in today's telecommunication and our everyday life in modern information society. The past few years witnessed a proliferation of content-based image retrieval techniques. Images are typically characterized by intrinsic attributes of images such as color, texture, and shape. However, the potential of integrating these techniques with visualization and data-mining techniques has yet been fully explored. Users should be able to explore images in a database or video clips by visual similarities. In this article, we explore the synergy between Pathfinder networks and content-based information retrieval techniques. Salient structures of images are revealed through visualization models derived from features extracted from images. Visualizations are generated from three feature classes of the well-known QBIC system: color, layout, and texture.

1. INTRODUCTION

Digital images and videos have an increasingly important role in today's telecommunication and our everyday life in modern information society. The past few years witnessed a proliferation of content-based image retrieval techniques. An image is typically characterized by intrinsic attributes of its content, ranging from simple, low-level features such as color and texture, to more complex, relatively higher-level features such as shape. However, the potential of integrating these techniques with visualization and data-mining techniques has yet been fully exploited to enable users explore images in a database or video clips by visual similarities.

Content-based image retrieval has been a highly active area of research in the computer vision community [1, 2]. A number of image retrieval systems have been developed over the last few years, notably, IBM's QBIC [3], PhotoBook [4], ImageRover [5], and Webseek [6]. On the other hand, several issues must be understood before image retrieval is viable. For example, it is important to understand what *retrieving relevant images* entails. If users are provided with a spatial user interface in which content similarity between images can be intuitively conveyed by their spatial proximity, then such interfaces may help users to benefit more from a given image database.

The advances of information visualization and data mining techniques now allow users to explore 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.

Structuring and visualizing digital images based on their content similarities, however, is not as mature as its text-based counterpart. Currently, many content-based image retrieval techniques have been developed to incorporate higher-level feature extraction capabilities, but a lot of work remains to be done. 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 work described in this article was originated in our experience in organizing an image database concerning the design of various information visualization displays. We have collected nearly 300 images of various information visualization systems [11]. Researchers and practitioners in information visualization often need to find an optimal way to arrange various visualization displays 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.

Generalised Similarity Analysis (GSA) is a generic framework developed for structuring and visualizing information spaces [12, 13]. 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-frequency, hypertext reference links, author co-citation profiles, and browsing trails of users. A key element in GSA is the use of the Pathfinder network scaling technique to extract the most salient links and eliminate redundant or counter-intuitive links [14]. Pathfinder has some desirable features over techniques such as multidimensional scaling (MDS), for example, Pathfinder networks present a more accurate local structure. In this article, our aim is 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 organized 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. 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 characterized by a variety of visual properties known as features. It is common to compare images by color, texture, and shape, although these entail different levels of computational complexity. Color 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.

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 mainstream feature-extraction algorithms is explained as follows.

2.1 Color

Swain and Ballard [15] matched images based solely on their color. The distribution of color was represented by color 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 *normalized histogram intersection*. This approach became very popular due to the following advantages:

Robustness. The color histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled [15]. 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 color histogram is a straightforward process, including scanning the image, assigning color values to the resolution of the histogram, and building the histogram using color components as indices.

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

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

2.2 Texture

A common extension to color-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 color-based feature extraction, He and Wang's approach [16] can be used to generate a histogram of texture, which is called the *texture spectrum*. More recently, the circular co-occurrence matrix, a modified version of the standard co-occurrence method, is developed to produce a texture histogram with an additional degree of rotation invariance [17].

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 challenge 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 color histograms.

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

Edge Orientation. Jain and Vailaya combined edge orientation histograms with color histograms [18]. 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 [19]). In addition, minor errors in the edge map have little effect on the edge orientation histograms. Unlike color 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 [20]. 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 [21].

To improve the stability of the distance transform, Rosin and West [22] 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.

The distance values can be represented in histograms once the SDT has been performed. These histograms will respond differently to different types of shapes. There is the crude distinction between cluttered, complex scenes and simple sparse scenes, which will result in different ends of the histogram being heavily populated. Thus the profile of the distance histograms provides an indication of image complexity, along the lines of Kawaguchi and Taniguchi's [23]. However, rather than return a single complexity measurement, the shape of the histogram will indicate more subtle distinctions between shapes.

3. PATHFINDER NETWORKS

Pathfinder network scaling is a structural modeling technique originally developed for the analysis of proximity data in psychology [14]. We have adapted this modeling technique to simplify and visualize 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:

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

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 [24].

In this article, we will apply the Pathfinder network scaling technique on image similarity data computed based on color, layout, and texture feature classes from the QBIC system. We used our collection of images of information visualization displays. We call it the InfoViz database in this article. Most of them are computer-generated graphics from information visualization systems. The Pathfinder network scaling process will then take these similarity data as the input and generate Pathfinder networks. All the Pathfinder networks described in this article are minimum-cost networks, i.e. PFNETs (r=∞, q=N-1). These Pathfinder networks are subsequently rendered as virtual reality models in VRML (Virtual Reality Modeling Language) for examination and evaluation.

4. PATHFINDER NETWORKS OF IMAGES

Three Pathfinder networks of images were generated based on similarity data produced by corresponding image-matching schemes, namely, color, layout, and texture. We expected that images with similar structures and appearances should be grouped together in Pathfinder networks.

Figure 1 is the screenshot of the visualization of our InfoViz image database. The Pathfinder network was derived from similarities determined by color histograms. The layout of the visualization is visually appealing. It is apparent that several clusters of images have homogenous colors. The largest image cluster appears to be the one in the lower half of the screenshot. This cluster includes images typically with line-drawing-like diagrams and visualization displays.

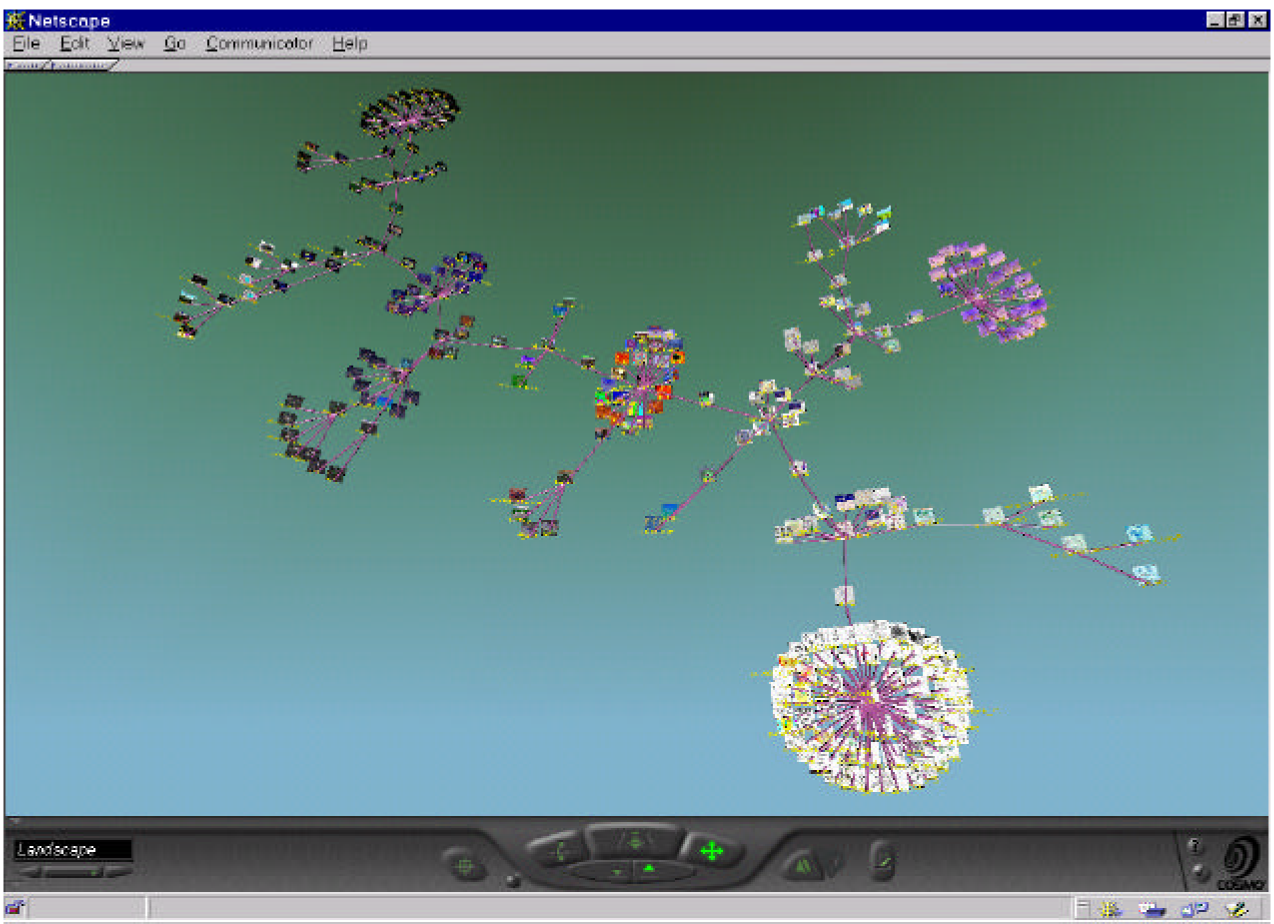

Figure 1: Images organized by color histogram.

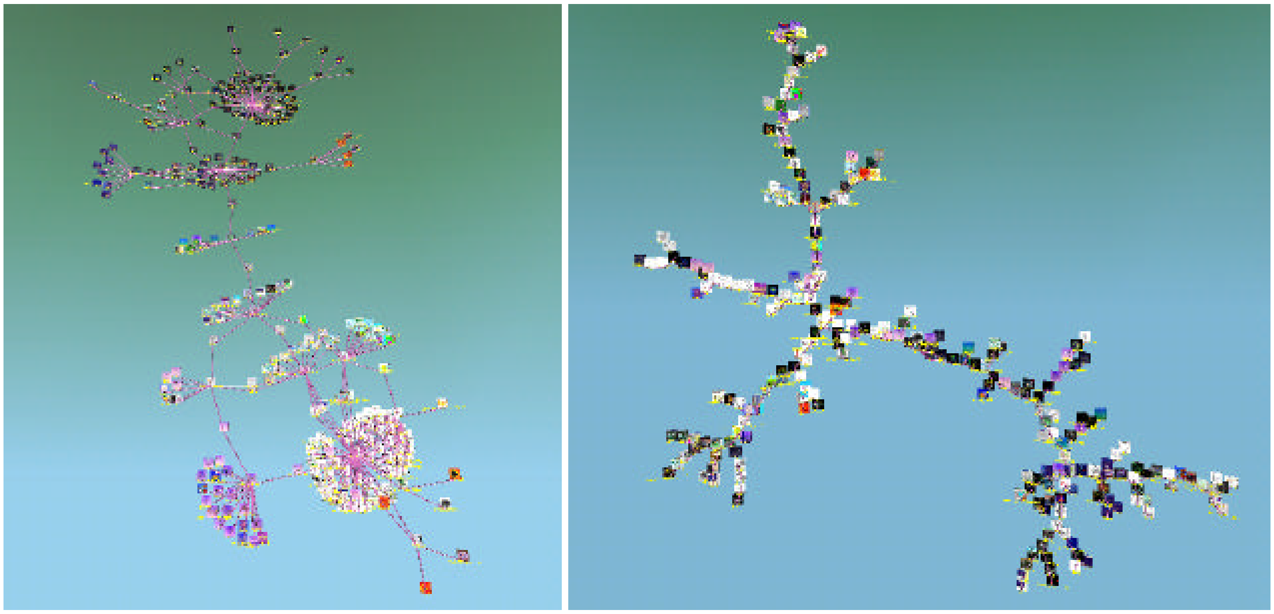

Figure 2: Pathfinder networks of images by layout (left) and texture (right).

A Pathfinder network consists of all the strongest connections among vertices as long as the triangle inequality condition is not violated. In our example, images are connected through such links. One can simply estimate the computed similarity between two images by estimating the minimum number of links connecting them. In principle, similar images should be placed near to each other, whereas dissimilar images should be placed further apart. This principle echoes the spring-embedder node placement model in a very intuitive way. This principle appears to be substantially realized for images based on color histograms.

Figure 2 shows the screenshots of two visualization models of the InfoViz image database. The left figure shows the Pathfinder network of images organized according to their similarities determined by layout as computed from the QBIC system. The right figure shows similar results based on texture similarities.

The overall structure of the layout-based visualization is different from the color-based visualization shown in Figure 1. This is expected due to the self-organizing nature of the spring-embedder model. On the other hand, visualizations based on the two schemes share some local structures. Several clusters appear in both visualizations.

Unlike the layout version, the texture-based visualization has a completely different visual appearance from the color-based visualization. To certain extent, this is understandable because the color histogram and color-layout schemes share some commonality in the way they deal with color. Figure 3 shows a closer view of some images in the texture-based visualization.

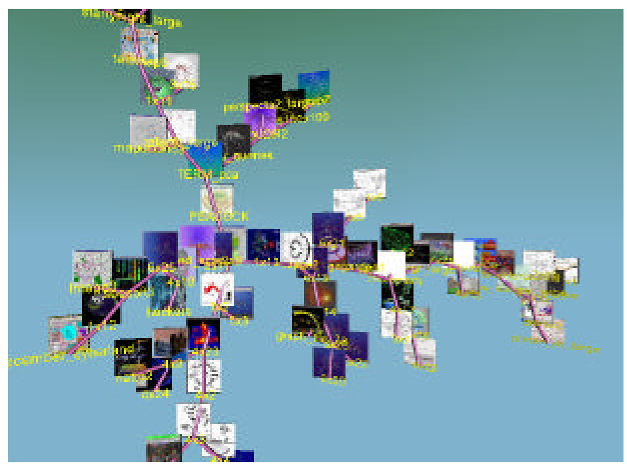

Figure 3: Images organized by texture.

Figure 4 shows a screenshot of a combined user interface, in which the image visualization and the QBIC search engine are integrated. Users can explore the image space in the virtual world. One can select a query image for the standby QBIC search engine by clicking on a thumbnail image in the visualization model in VRML. The default binding was set as the Special Hybrid feature class provided by QBIC.

Figure 4: Searching images in QBIC through the layout-based visualization.

4.1 Pathfinder Network Structures

In order to further understand the impact of different feature schemes on visualized structures, we compared the Pathfinder networks across the three feature-based matching schemes. The number of links in each network and the number of links in common are used as the basis for network comparisons. The degree of similarity between two networks is determined by the likelihood that a number of common links are expected to be found given the total number of links in the networks involved.

4.1.1 Color versus Layout

Color- and layout-based visualization schemes turned out to have significantly similar structures (p = 0.000). The magnitude of structural similarity is 0.182. This suggests that these two visualizations reveal some salient characteristics of the image database.

4.1.2 Color versus Texture

Color-based and texture-based visualization networks are completely different. There are only 2 links in common between the two networks. This confirms our visual inspection of the networks. The network similarity is 0.004.

Number of Images	279
Links in PF by Color	271
Links in PF by Layout	319
Common Links	91
Expected Common Links	2.23
Point Probability	0.00
Information	406.94

Table 1. Comparison between color- and layout-based visualizations.

Number of Images	279
Links in PF by Color	271
Links in PF by Texture	284
Common Links	2
Expected Common Links	1.98
Point Probability	0.27
Information	0.76

Table 2. Comparisons of color- and texture-based visualizations.

4.1.3 Layout versus Texture

Given the comparison results so far, it is not surprising to find out that layout- and texture-based visualizations are completely different. There is only one common link between the two networks. The network similarity is 0.002.

Number of Images	279
Links in PF by Layout	319
Links in PF by Texture	284
Common Links	1
Expected Common Links	2.34
Point Probability	0.23
Information	0.14

Table 3. Comparisons of layout- and texture-based visualizations.

The color-based visualization has the least number of links (279). The layout-based version has the largest number of links (319).

4.2 Video

In addition to visualize the InfoViz image database, we also visualized a video clip. This was motivated by our observation that one cannot expect a similarity continuum of images from independent sources. On the other hand, the provision of a similarity continuum of images will facilitate the validation and evaluation of the visualization approach. Images from video clips are likely to satisfy this requirement. Frames in the same shot are likely based on the same scene.

Figure 5 shows an overview of the video network generated similarly based on color-histograms. As expected, the visual similarity of each image cluster is more easily to recognise.

Figure 5. Visualizing video frames.

5. DISCUSSION AND CONCLUSION

Organizing images in Pathfinder networks based on various types of image features allows us to compare and verify the reliability and correctness of a particular feature-extraction algorithm.

In a long run, organizing images based on low-level features remains a challenging field of research. 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 networks of images.

Visualization techniques described in this article have a wide range of potential applications, for example, data mining in remote sensing images and image retrieval from film and video archives. We are now considering applying this methodology on a sample of images with more continuous scenes, for example, video segments, so that we will be able to keep track of the impact of various feature-extraction techniques more closely.

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

The integration of CBIR techniques and existing techniques in GSA 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.

ACKNOWLEDGMENTS

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

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," presented at 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," presented at 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," presented at 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, "Generalised Similarity Analysis and Pathfinder Network Scaling," *Interacting with Computers*, vol. 10, pp. 107-128, 1998.

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

[14] 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.

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

[16] D. C. He and L. Wang, "Texture unit, texture spectrum, and texture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, 1990.

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

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

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

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

[21] G. Borgefors, "Distance transformations in digital images.," *Computer Vision, Graphics, and Image Processing*, vol. 34, pp. 344-371, 1986.

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

[23] E. Kawaguchi and R. I. Taniguchi, "The depth first picture-expression as an image thresholding strategy," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, pp. 1321-1328, 1989.

[24] C. Chen, "Visualising semantic spaces and author co-citation networks in digital libraries," *Information Processing and Management*, vol. 35, pp. 401-420, 1999.