Using CBIR and Pathfinder Networks for Image Database Visualisation

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

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

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

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

Abstract

Digital images and videos have an increasingly important role in today's telecommunication and our everyday life in modern information society. In this paper, we explore the synergy between content-based information retrieval techniques and Pathfinder networks. Salient image features, based on colour, texture, edge orientation, and edge distance, are extracted. The structural modelling capabilities of Pathfinder are then applied to simplify and visualise the strongest interrelationships in the image database.

1. Introduction

Generalised Similarity Analysis (GSA) is a generic framework developed for structuring and visualising information spaces [3]. Applications of GSA include visualisation of university websites, online conference proceedings, and journals in digital libraries according to a variety of similarity measures, such as term-frequencies, hypertext reference links, 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 [12]. Feature-extraction techniques and information visualisation have a great potential to benefit tremendously from each other. In this paper the integration of CBIR methods and Pathfinder networks is the first step towards the development of a comprehensive framework of visualising and exploring hypermedia networks. 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, and therefore we are investigating the clustering of images from video clips.

2. Content-Based Image Retrieval

Typically, the content of an image can be characterised by a variety of visual properties known as features. The key issue in CBIR is how to match two images according to these computationally extracted features. It is common to compare images by colour, texture, and shape, although these entail different levels of computational complexity. For instance 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. 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: colour, texture, shape, and 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 have yet to reliably reach this level of capability. 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 [10] 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 since it is robust, simple and efficient to compute, and relatively effective.

Furthermore to increase the efficiency of the colour histogram towards lower storage requirements and reduced sensitivity to illumination changes, we adopted the colour labels that resulted of the pioneering study of Berlin and Kay [2]. The colour space can be partitioned into a maximum of eleven basic classes representing colours, as well as black, white, and grey, which we use as our histogram bins.

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, a modification of the spatial grey-level dependence matrices [5] method is used that is rotation invariant, which we call circular co-occurrence matrices.

The first step is to define a digital circle of radius d, which defines a set of intensity pairs. Then all the pixels of the image are considered as centres (apart from the ones where the circle goes of range) defining more sets of intensity pairs. These pairs are finally histogrammed to obtain the 2D circular co-occurrence matrix, as shown in the following formulae:

$$C(i,j \mid d) = \#\left\{ [\boldsymbol{c},\boldsymbol{p}] \left| \begin{array}{l} f(\boldsymbol{c}) = i, \\ f(\boldsymbol{p}) = j, \\ \boldsymbol{p} = \boldsymbol{c} + (d\cos\varphi, d\sin\varphi) \end{array} \right. \right\}$$

where $\boldsymbol{c}$ defines the centre of the circle and $\boldsymbol{p}$ the location of the point on the digital circle at an angle φ.

In general, texture-based feature extraction tends to provide more spatial information than colour 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 matching 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.

2.3.1. Edge Orientation. Previous work in this area can be found in Jain and Vailaya's work [1]. 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 [9]). 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. For robustness, thresholding is avoided; all the edges are included, and their contribution to the histogram is weighted by their magnitudes so as to reduce the contribution from spurious edges.

2.3.2. Multi-resolution Salience Distance Transform. Another approach to including shape information is based on the distance transform (DT) which efficiently calculates the distance from every non-feature to the closest feature [7]. To improve its stability Rosin and West [11] weighted the distances by various factors (e.g. edge magnitude, length) over a range of scales.

The distance values can be histogrammed once the SDT has been performed. These histograms will respond differently to different type of shapes. For instance, 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.

2.4. Spatial Information

Partitioning based approaches as in [13] have been used to improve the performance of CBIR systems. Trying to avoid selection of rigid regions and true segmentation, we used binary thresholding as a tool for partitioning.

The partitioning injects some spatial information into the analysis so that standard feature-based methods (e.g. non-spatial) can then be applied within each region. In fact we only consider each of the two classes, black and white, as single composite regions rather than treat each individual region separately.

A potential drawback of this method is that the process can sometimes be sensitive such that small changes in the threshold value produce large changes in the resulting binary images. To overcome this we introduced a *soft threshold* [8].

2.5. Image Similarity

Having generated histograms based on the different properties (colour, texture, and shape), the similarity between two images is computed in two steps. For each property the Euclidean distance between the corresponding feature histograms is calculated. Next, the distances over the different features are combined using the geometric mean to provide a single similarity rating.

3. Pathfinder Networks

Pathfinder network scaling is a structural modelling technique originally developed for the analysis of proximity data in psychology [6]. 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:

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

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 [4].

In this paper, we will apply the Pathfinder network scaling technique on image similarity data computed using histograms, or combinations (see Table 1.), representing image aspects extracted using the methodology described earlier.

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 paper 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. Experimental Results

We tested the CBIR system using a database of 42 clips (378 images) sampled from sequences of broadcasted TV signal (approximately 1 image every 2 seconds). The recall and precision metrics were used to quantify the performance of each histogram combination. Recall represents the proportion of 'correct' images in the best matches set, while precision represents the distance between the first and last 'correct' image (low values indicate better results). In this paper we exploited the integration of 4 histogram combinations, as shown in the following table, with the pathfinder network, investigating the clustering of the images. In Figure 1 snapshots of the virtual worlds produced are illustrated, showing clusters of clips, using colour and texture histogram distances. They demonstrate that images belonging to the same clip are clustered together, verifying the effectiveness of the image similarity measures.

Colour Labels	82.42	67.00
Spatialized Colour Labels (CL)	82.33	67.24
Colour Labels + Spatialized CL	84.52	70.62
Colour + Texture + Orientation	91.89	81.34
Colour + Texture + Orientation + Distance	91.63	81.17
Spatialized CL + Texture + Orientation	88.92	76.87
Spatialized CL + Texture + Orientation + Distance	89.21	76.80
Colour Labels + Spatialized CL + Texture + Orientation	89.08	77.56
Colour Labels + Spatialized CL + Texture + Orientation + Distance	89.37	77.53

Table 1. Method's Recall(%) and Precision(%) results.

5. Conclusions

We have described a system combing techniques from CBIR [8] and data visualisation [4]. The combination is attractive since it enables quick and easy validation of the CBIR system results by the designer. This is valuable since verification of CBIR systems is problematic: purely automatic methods do not capture all perceptually significant aspects, while human analysis of individual images is not practical for large databases. A second benefit is the visualisation provides a relatively concise, well-structured representation of the database that can aid browsing.

6. Acknowledgements

We would like to thank Matthew Seaborn for providing the look-up table of the colour labels.

The images were captured after permission of CRETA channel, Greece.

7. References

[1] A. K. Jain and A. Vailaya. "Image retrieval using color and shape". *Pattern Recognition*, 29, 8 (1996), 1233-1244.

[2] B. Berlin and P. Kay Basic *Color Terms: their universality and evolution*. University of California Press, 1969

[3] C. Chen. "GSA and Pathfinder Network Scaling". *Interacting with Computers*, 10, 2(1998), 107-128.

[4] C. Chen. "Visualising semantic spaces and author co-citation networks in digital libraries". *Information Processing & Management*, 35, 2 (1999), 401-420.

[5] D. H. Ballard and C. M. Brown *Computer Vision*, Prentice Hall, 1982

[6] F. Idris and S. Panchanathan, "Review of image and video indexing techniques", *Journal of Visual Communication and Representation*, 8, 2 (1997), 146-166.

[7] G. Borgefors. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34 (1986), 344-371.

[8] G. Gagaudakis and P L Rosin, "Incorporating Shape into Histograms for CBIR", *Journal of Pattern Recognition, Forthcoming.*

[9]J. Canny. "A computational approach to edge detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 6 (1986), 679-698.

[10] M. Swain and H. Ballard. "Color indexing". *International Journal of Computer Vision*, 7, 11 (1991), 11-32.

[11] P. L. Rosin and G. A. W. West. "Salience distance transform". *Graphical Models and Image Processing*, 57, 6 (1995), 483-521.

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

[13] M. Striker and A.Dimai. "Special covariance and fuzzy regions for image indexing". *Machine Vision and Applications*, 10, 2(1997), 66-73.

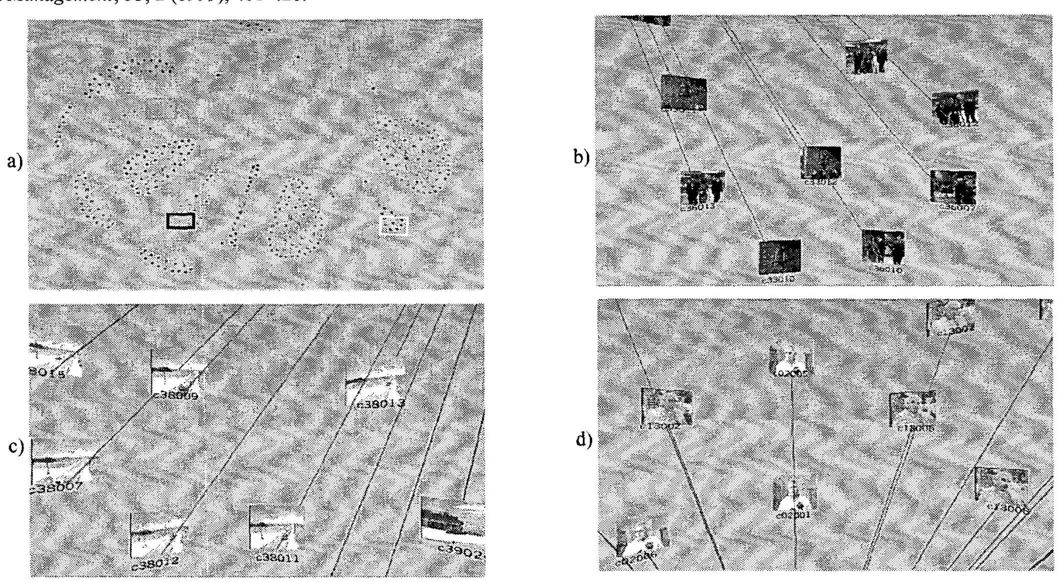

Figure 1. VRML views of the database using Colour Labels.