A New Enhancement to Histogram-Based Approaches in Content Based Image Retrieval Systems

A Bamidele and F W M Stentiford

CONTENT UNDERSTANDING GROUP, UNIVERSITY COLLEGE LONDON, ADASTRAL PARK CAMPUS, IPSWICH, UK.

Abstract: The volume of visual data in the world is increasing exponentially through the use of digital camcorders and cameras in the mass market. Although storage space is in ample supply, access and retrieval remain a severe bottleneck both for the home user and for industry. In this study, we show that laying emphasis upon areas of images that attract high visual attention can improve retrieval performance. This paper describes an approach which makes use of a visual attention model together with a similarity measure to automatically identify salient visual material and generate searchable metadata that associates related items in a database. The saliency of images should play a major part in automated image retrieval and this paper exemplifies a way in which this can be achieved.

1 Introduction

The volume of digital images has been increasing dramatically in recent years and as a result a crisis is now taking place within a broad range of disciplines that require and use visual content. Whilst storage and image capture technologies are able to cope with huge numbers of images, poor image and video retrieval is in danger of rendering many repositories valueless because of the difficulty of access. Many disciplines and segments in industry including telecommunications, entertainment, medicine, and surveillance, need high performance retrieval systems to function efficiently. Visual searches by text alone are ineffective on images and are haphazard at best. Descriptive text simply does not reflect the capabilities of the human visual memory and does not satisfy users' expectations. Furthermore the annotation of visual data for subsequent retrieval is almost entirely carried out through manual effort. This is slow, costly and error prone and presents a barrier to the stimulation of new multimedia services. Much research is now being conducted into measures of visual similarity that take account of the semantic content of images in an attempt to reduce the human involvement during database composition. Indeed semantically associating related visual content will add value to the material by improving access and exposing new potential benefits to a wider market.

2. State of the Art

An image retrieval system must produce images that a user wants. In response to a user's query, the system has to offer images that are similar in some user-defined sense. This goal is met by selecting visual features thought to be significant in human visual perception and using them to measure relevance to the query. Many image retrieval systems in operation today rely upon annotations that can be searched using keywords. These approaches have limitations not least of which are the problems of providing adequate textual descriptions and the associated natural language processing necessary to service search requests.

A great deal of research [1] has been carried out in recent years and a few of the most relevant approaches are highlighted here. Colour, texture, local shape and spatial layout in a variety of forms are the most widely used features in image retrieval. Jain and Vailaya [2] utilised colour histograms and edge direction histograms for image matching and retrieval. The PICASSO system [3] used visual querying by colour perceptive regions. Colour regions were modelled through spatial location, area, shape, average colour and a binary 128 dimensional colour vector. The MARS project [4] used a combination of low-level features (colour, texture, shape) and textual descriptions. Colour was represented using a 2D histogram of hue and saturation. Texture was represented by two histograms, one measuring the coarseness and the other one the image directionality, and one scalar for contrast. Phillips and Lu [5] addressed the problem of the arbitrary boundaries between colour bins, which means that closely adjacent colours are considered different by the machine; moreover, they applied a method of perceptually weighted histograms to reduce this effect.

One of the first commercial image search engines was QBIC [6] which executes user queries against a database of pre-extracted features. Region based querying is favoured in Blobworld [7] where global histograms are shown to perform comparatively poorly on images containing distinctive objects. Object segmentation for broad domains of general images is considered difficult, and a weaker form of segmentation that identifies salient point sets may be more fruitful [8]. Relevance feedback is often proposed as a technique for overcoming many of the problems faced by fully automatic systems by allowing the user to interact with the computer to improve retrieval performance [9]. This reduces the burden on unskilled users to set quantitative pictorial search parameters or to select images that come closest to meeting their goals.

Conventional approaches suffer from a number of disadvantages. Firstly, there is a real danger that the use of any form of pre-defined feature measurements will preclude solutions in the search space and be unable to handle unseen material. Secondly, the choice of features in anything other than a trivial problem is unable to anticipate a user's perception of image content.

This information cannot be obtained by training on typical users because every user possesses a subtly different subjective perception of the world and it is not possible to capture this in a single fixed set of features and associated representations. Furthermore an approach to visual search should be consistent with the known attributes of the human visual system and account should be taken of the perceptual importance of visual material as well as more objective attributes [10].

This paper describes the application of models of human visual attention to CBIR in ways that enable fast and effective search of large image databases. The model employs the use of visual attention maps to define Regions-of-Interest (ROI) in an image with a view to improving the performance of image retrieval. This work will also involve the study of new database configurations that accommodate new metadata attributes and their associated functionality.

3. Current research

The use of models of human visual attention in problems of visual search is attractive because it is reasonable to believe that this is the mechanism people actually use when looking for images [11, 12]. The visual attention mechanism [13] used in this paper is based upon ideas that have their counterpart in surround suppression in primate V1 [14], and this is favoured for its simplicity and the ease of implementation both in software and potentially in hardware.

Let the colour histograms of images A and B be H_A and H_B each with n bins. The Manhattan global distance between the histograms is normalised by image area and is given by

$$d(H_A, H_B) = \sum_{i=1}^{n} |H_A(i) - H_B(i)| \quad \text{where} \quad H_\alpha(i) = \frac{number\ of\ pixels\ with\ hue\ i}{number\ of\ pixels\ in\ \alpha} \qquad (1)$$

A major disadvantage of the histogram and many other more sophisticated measures is their inability to distinguish foreground from background. This means that images with a dominant green background, for example, are very likely to be marked as similar regardless of the nature of the principal subject material which might be a tractor in one image and a horse in another. The visual attention mask is introduced to combat this problem.

Let the visual attention mask for image α be given by

$$M_\alpha(x,y) = \begin{cases} = 1\ if\ attention\,score\,at\ (x,y) \geq T \\ = 0\ otherwise \end{cases} \quad \text{where } T \text{ is a threshold} \qquad (2)$$

The attention histogram distance between the images A and B is defined as

$$d'(H_A, H_B) = \sum_{i=1}^{n} |H'_A(i) - H'_B(i)| \quad \text{where} \quad H'_\alpha(i) = \frac{numberof\ pixels\,with\,hue\,i\ and\ M_\alpha(x,y) = 1}{numberof\ pixels\,in\ \alpha\ and\ M_\alpha(x,y) = 1} \qquad (3)$$

The new attention based distance *d'* restricts the histogram calculation to pixels lying within areas that are assigned high values of visual attention by the model. This means that greater emphasis is given to subject material and hence retrieval performance should improve for those images possessing clear regions of interest, which are characterised by their colour histograms.

A similarity metric when applied to the images in a collection creates a network of associations between pairs of images each association taking the value of the strength of the similarity. More generally the associations can connect image regions to regions in other images, so that images may still be strongly related, if they contain similar objects in spite of possessing different backgrounds. Consequently it is this additional metadata that provides the information to enable a convergent and intelligent search path.

Images in a collection are processed offline to produce metadata that is stored in the relational database. Visual Attention (VA) analysis is applied to a query image and the similarity of ROIs to others in the database is determined. A rank ordered list of candidate retrieved images is returned to the user as illustrated in Figure 1. The precomputed network of similarity associations enables images to be clustered according to their mutual separations. Query images are matched first with ‘vantage’ images [15] in each cluster before selecting images from within the closest cluster groups.

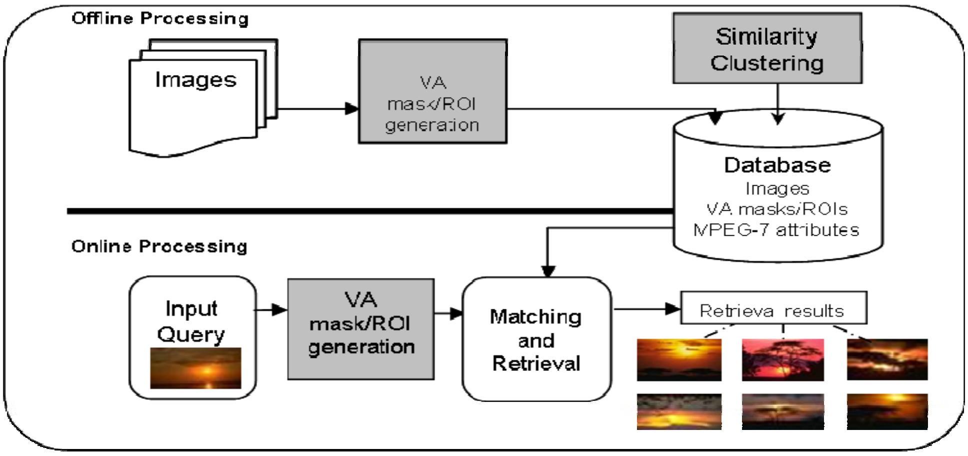

Figure 1 Image entry and retrieval system.

4. Results

The method is illustrated by application to a small set of 12 images consisting of 6 pairs that were clearly similar. VA maps were generated and mask arrays (2), extracted that indicated salient areas. The histograms are based upon the hue values at each pixel, which range from 1 to 360. The difference is due mainly to the different colour profiles of the background and foreground.

The distances d in Equation (1) and d' in Equation (2), between all 12 of the images using the global and attention based similarity measures were computed. In order to compare the global in Equation (4), and attention based histogram performances in Equation (5) on image *i* the distances between the pairs of subjectively similar images (*i, j*) were compared to those between all the others where

$$P_i = \left\{ \frac{\sum_{A \neq i,j} (d(H_A, H_i) - d(H_i, H_j))}{d(H_i, H_j)} \right\} \quad (4) \text{ and similarly} \quad P'_i = \left\{ \frac{\sum_{A \neq i,j} (d'(H_A, H_i) - d'(H_i, H_j))}{d'(H_i, H_j)} \right\} \quad (5)$$

The measure of performance, *P* is high, if on average the normalised distance between 'similar' images is much less than that between 'dissimilar' images; where Equation (4), is the measure of global histogram performance and Equation (5), the attention histogram performance.

The comparative performance is displayed in Figure 2 where (P_i - P'_i) is plotted for each image. Positive values indicate improvements in performance over the global similarity measure.

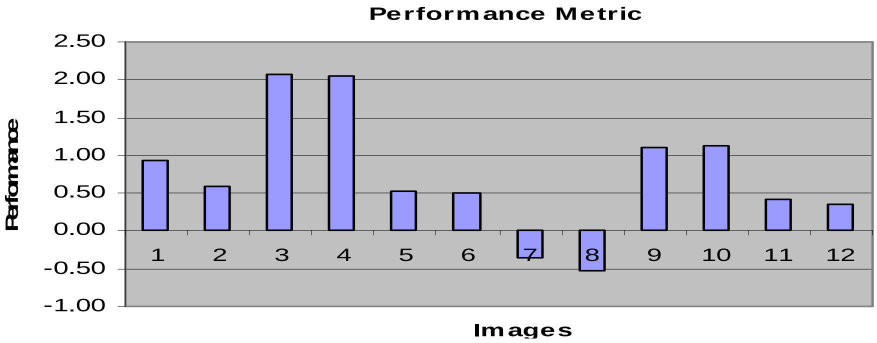

Figure 2 Image separation comparisons

5. Discussion

From Figure 2, an improvement in separation is seen in 5 of the pairs of images, but images 7 and 8 are not separated from images 3 and 4 as well as by the global histograms. This is because the visual attention masks cover a high proportion of white and grey areas in all four images at the same time as the background material being significantly different between the two pairs. The green background is treated as important by the global histogram but is suppressed by the attention mechanism.

The background happens to be a distinguishing feature in this dataset. Images 9 and 10 yield a significant improvement because the central subject material is very similar. It should be observed in Figure 2 that the subjects in images 11 and 12 are identical but the background is substantially different. In this case the attention model has been able to focus on the important image components and detect a high value of similarity. By the same token Image 10 is a magnified and cropped version of image 9 and illustrates how an effective similarity measure might detect infringements of copyright in which parts of images have been replicated and distorted.

6. Conclusions and Future Work

There is good reason to believe that the saliency of images should play a major part in automated image retrieval and this paper illustrates a way in which this might be achieved. The work has indicated that laying emphasis upon areas of images that attract high visual attention can improve retrieval performance.

Future experiments will make use of a weighted VA mask, which will provide a better balance between the foreground and background areas in the computation of similarity scores. In addition attention mechanisms will be incorporated into more meaningful measures of similarity that take account of image structure and other features. More work is necessary on larger sets of images to obtain statistical significance in the results and we are working closely with other academic institutions on this.

The authors acknowledge the support of BT Exact and the EPSRC in this work.

References.

[1] R. C. Veltkamp and M. Tanase, ‘Content-based retrieval systems: a survey’, http://www.aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/ , March 2001.

[2] A.K. Jain and A. Vailaya, ‘Image retrieval using colour and shape’, Pattern Recognition, 29(8), pp 1233-1244 ,1996.

[3] D. Bimbo and Pala, ‘Visual Querying by Colour Perceptive Regions’, Pattern Recognition, Vol 31, pp. 1241-1253 ,1998.

[4] S. Mehrotra and K. Chakrabarti, ‘Similarity shape retrieval in MARS’, IEEE International Conference on Multimedia & Expo, New York, 2000.

[5] G. Lu and J. Phillips, ‘Using Perceptually Weighted Histograms for Colour-based Image Retrieval’, Proceedings on 4th Int. Conf. on Signal Processing Proceedings, 1998. ICSP '98.Vol: 2, 1998.

[6] W. Niblack and M. Flickner, ‘Query by image and video content: The QBIC system’, IEEE Computer, pp 23-32, September 1995.

[7] C. Carson, S. Belongie, H. Greenspan, and J. Malik, ‘Blobworld, ‘Segmentation using Expectation-Maximisation and its Application to Querying’, IEEE Trans. PAMI, Vol 24, No 8, pp 1026-1038,,August 2002.

[8] W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, ‘Content-Based Retrieval at the End of the Early Years’, IEEE Trans PAMI, Vol 22, No 12, pp 1349-1379 ,December 2000.

[9] J. Cox, M. L. Miller., T. P. Minka, T. V. Papathomas, and P. N. Yianilos, ‘The Bayesian image retrieval system, PicHunter: theory, implementation, and Psychophysical experiments’, IEEE Trans. On Image Processing, Vol 9, No 1, January 2000.

[10] F. W. M Stentiford, ‘An attention based similarity measure with application to content based information retrieval’, SPIE Vol 5021, Storage and Retrieval for Media Databases, Santa Clara, 22-24, January 2003.

[11] A. Bamidele and F. W. M Stentiford, ‘Image Retrieval: A Visual Attention Based Approach.’ Postgraduate Research Conference in Electronics, Photonics, Communications & Networks, and Computing Science, Hertfordshire, 5 - 7 April 2004.

[12] A. Bamidele, F. W. M Stentiford and J. Morphett, ‘An Attention-based approach to Content Based Image Retrieval.’ British Telecommunications Advanced Research Technology Journal on Intelligent Spaces (Pervasive Computing), Vol 22 No 3, July 2004.

[13] F.W.M Stentiford, ‘An estimator for visual attention through competitive novelty with application to compression’, Picture Coding Symposium, Seoul, 24-27 April, 2001.

[14] N. Petkov and M. A. Westenberg, ‘Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition’, Biol. Cybernetics., Vol 88, pp 236-246, 2003.

[15] Vleugels and R. C. Veltkamp, ‘Efficient image retrieval through vantage objects’, Pattern Recognition, Vol 35, pp 69-80 , 2002.