An attention-based approach to content-based image retrieval

A Bamidele, F W M Stentiford and J Morphett

Mark Weiser's vision that ubiquitous computing will overcome the problem of information overload by embedding computation in the environment is on the verge of becoming a reality. Nevertheless today's technology is now capable of handling many different forms of multimedia that pervade our lives and as a result is creating a healthy demand for new content management and retrieval services. This demand is everywhere; it is coming from the mobile videophone owners, the digital camera owners, the entertainment industry, medicine, surveillance, the military, and virtually every library and museum in the world where multimedia assets are lying unknown, unseen and unused.

The volume of visual data in the world is increasing exponentially through the use of digital camcorders and cameras in the mass market. These are the modern day consumer equivalents of ubiquitous computers, and, although storage space is in plentiful supply, access and retrieval remain a severe bottle-neck both for the home user and for industry. 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. Such a system for content classification and access will be of great use in current and future pervasive environments where static and mobile content retrieval of visual imagery is required.

1. Introduction

Weiser's vision that ubiquitous or pervasive computing will overcome the problem of information overload [1] is on the verge of becoming a reality. 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. While 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 — and this requirement will grow as we continue moving forward in a world connected by both fixed and wireless networks.

It is envisaged that massive volumes of image and video content will be generated by the requirements for more pervasive applications. There is an increasing demand not only for reacting to people's requests, but also for monitoring people's intent and behaviour in a passive manner within intelligent spaces (iSpaces). Whether the visual material will describe a security status, the behaviour of a crowd, the emotion of a PC user, or the interests of shoppers, major advances in image interpretation are needed to make these applications viable.

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.

In addition to storing and interconnecting iSpaces, service and network providers need to be able to reduce costs in providing content and content management services to a range of devices. Doing so in a cost-effective manner, however, only makes sense when the effectiveness of the systems makes it attractive enough for consumers to want to pay for such services. It is posited that the potential lack of both effectiveness and efficiency in current image management systems prevent them from being a commercial alternative to free (albeit ineffective) text-based search engines. Here lies the proposed commercial benefit of the work. We are working jointly on making content classification, access and retrieval effective for pervasive computing users while at the same time, seeking to remove many of the costs associated with manual data entry, thereby making the proposition commercially viable.

The academic perspective in this paper stems from identifying what is perceived as relevant information to the user by the integration of mechanisms of content-based image retrieval (CBIR) and context-aware technologies [2, 3]. Visual content continues to represent the most important and most desirable communication medium and it is a challenge to deliver relevant visual data to users engaged in diverse and unpredictable activities.

Section 2 of this paper outlines relevant state-of-the-art research. Section 3 describes the current research and overviews the visual attention model. Section 4 describes an experiment using the model and presents the results. Section 5 briefly discusses the results with section 6 concluding the paper and suggesting future work.

2. State of the art

It is the job of an image retrieval system to produce images that a user wants. In response to a user's query, the system must 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 key words. 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.

Colour, texture, local shape and spatial layout in a variety of forms are the most widely used features in image retrieval. Such features are specified by the user in the 'direct query on descriptions' retrieval method [4]. This approach makes great demands on the user who must be aware of the technical significance of the parameters that are being used during the search.

Swain and Stricker [5] measured the similarity of images using colour histograms and the Manhattan metric. The PICASSO system [6] proposed by Del Bimbo and Pala, uses visual querying by colour perceptive regions. Colour regions were modelled through spatial location, area, shape, average colour and a binary 128 dimensional colour vector. A single region characterises the image with a colour vector retaining the global colour attributes for the whole image. Similarity between two images is then computed based on the modelled colour regions. Jain and Vailaya [7] utilised colour histograms and edge direction histograms for image matching and retrieval.

The MARS project [8] used a combination of low-level features (colour, texture, shape) and textual descriptions. Colour is represented using a 2-D histogram of hue and saturation. Texture is represented by two histograms, one measuring the coarseness and the other one the image directionality, and one scalar for contrast. It was later enhanced using a shape-matching similarity algorithm [9], although invariant to transformational effects in image content, it was deficient in taking account of perceptual similarity between images.

Phillips and Lu [10], address the problem of the arbitrary boundaries between colour bins, which can mean that closely adjacent colours are considered different by the machine. They applied a method of perceptually weighted histograms to weaken this effect in other approaches.

One of the first commercial image search engines was QBIC [11] which executes user queries against a database of pre-extracted features. The Virage system [12] generates a set of general primitives such as global colour, local colour, texture and shapes. When comparing two images a similarity score is computed using the distance function defined for each primitive. Weights are needed to combine individual scores into an overall score and the developer is left to select the weights appropriate to his application [13].

MetaSeek [14] also uses colour and texture for retrieval, but matching is carried out by other engines such as QBIC [11] and MARS [8]. MetaSeek uses a clustering approach for the locally extracted colour and texture features. The system was intelligently designed to select and interface with multiple Web-based image search engines by ranking their performance for different classes of user queries. Kulkami [15] used extracted texture feature values to formulate specific user-defined queries.

Region-based querying is favoured in Blobworld [16] where global histograms are shown to perform comparatively poorly on images containing distinctive objects. Similar conclusions were obtained in comparisons with the SIMPLIcity system [17]. VisualSEEk [18] determines similarity by measuring image regions by using both colour parameters and spatial relationships and obtains better performance than histogram methods that use colour information alone. NeTra [19] also relies upon image segmentation to carry out region-based searches that allow the user to select example regions and lay emphasis on image attributes to focus the search. 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 [20].

Vinod [21], proposed an interactive method to identify regions in images, which can represent a given object based on colour features. Regions of interest are extracted based on sampling using a square window. This technique increased the efficiency of search by concentrating on the most promising regions in the image. The approach focused on just the upper bound of histogram intersection and assumed all matching was the same across all focused regions.

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 [22]. In Quicklook [23] and ImageRover [24] items identified by the user as relevant are used to adjust the weights assigned to the similarity function to obtain better search performance. PicHunter [25] has implemented a probabilistic relevance feedback mechanism that predicts the target image based upon the content of the images already selected by the user during the search. Related work is reported by Jose [26]. This reduces the burden on unskilled users to set quantitative pictorial search parameters or to select images that come closest to meeting their goals, but it does require the user to behave consistently as defined by the machine. Retrieval should not require the user to have explicit knowledge of the features employed by the system and users should not have to reformulate their visual interests in ways that they do not understand.

Conventional approaches suffer from some disadvantages. Firstly there is a real danger that the use of any form of predefined 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.

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 [27, 28].

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 (ROIs) in an image with a view to improving the performance of image retrieval. The work will also involve the study of new database configurations that accommodate new metadata attributes and their associated functionality. The work may yield new metadata vocabularies and attributes that as yet are not encompassed by the MPEG-7 multimedia standards [29].

3. Current research

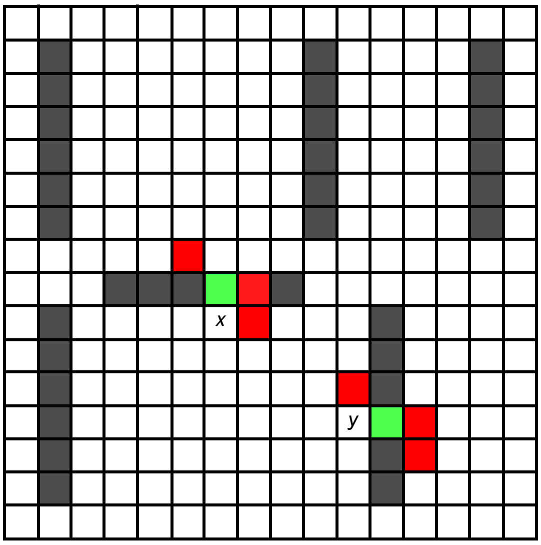
The use of models of human visual attention in problems of visual search is attractive because it is reasonable to believe that this is the mechanism people actually use when looking for images [30]. The model [31] used in this paper is favoured for its simplicity and the ease of implementation both in software and potentially in hardware. Initial work has concentrated upon demonstrating that knowledge of perceptually significant areas in an image improves search performance and that this can be automated through the application of an attention model.

Related work [32, 33] using an eye tracker is exploring gaze behaviour and is using an attention model to anticipate users' search intentions during CBIR. This work will validate and at the same time refine the visual attention model for specific applications.

3.1 Visual attention model

The visual attention mechanism used in this paper is based upon ideas that have their counterpart in surround suppression in primate V1 [34]. Petkov and Westenberg [35] use a version of this model and confirm qualitative explanations of visual pop-out effects. This model assigns high values of visual attention when neighbouring pixel configurations do not match identical positional arrangements in other randomly selected neighbourhoods in the image. This means that textures and other features that are common in an image will tend to suppress attention values in their neighbourhood.

In this model, digital images are represented as a set of pixels, arranged in a rectangular grid in Fig 1. The process of calculating the VA score for a pixel *x*, begins by selecting a small number (*m*) of random pixels in the immediate neighbourhood (radius, ε) of *x*. Then another



1. create a random neighbourhood at *x*
2. select a random second pixel *y* and compare
3. increment score and repeat from 2 for a mismatch
4. repeat from 1 for a match

Fig 1 Neighbourhood at *x* mismatching at *y* ($m = 3$, $\varepsilon = 1$).

pixel *y* is selected randomly elsewhere in the image. The pixel configuration surrounding *x* is then compared with the same configuration around *y* and tested for a mismatch. If a mismatch is detected, the score for *x* is incremented and the process is repeated for another randomly selected *y* for *t* iterations.

If the configurations match, then the score is not incremented and a new random configuration around *x* is generated. The process continues for a fixed number of iterations for each *x*. Regions obtain high scores if they possess features not present elsewhere in the image. Low scores tend to be assigned to regions that have features that are common in many other parts of the image. Such features may be dependent upon colour, shape or both.

The visual attention estimator has been implemented as a set of tools that process images and produce corresponding arrays of attention values. The attention values are displayed in Fig 2 as a map where VA scores are represented as false colours with the highest scores shown in green and lower scores as darker shades of red. This map is used as a mask to indicate which areas of the image are to be analysed for comparison purposes thereby suppressing background pixels from the computation.

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:

$$d(H_A, H_B) = \sum_{i=1}^{n} \left| H_A(i) - H_B(i) \right|$$

where $H_\alpha(i) = \dfrac{\text{number of pixels with hue } i}{\text{number of pixels in } \alpha}$

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 *a* be given by:

$$M_\alpha(x, y) \begin{cases} = 1 \text{ if attention score at} (x, y) \geq T \\ = 0 \text{ otherwise} \end{cases}$$

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

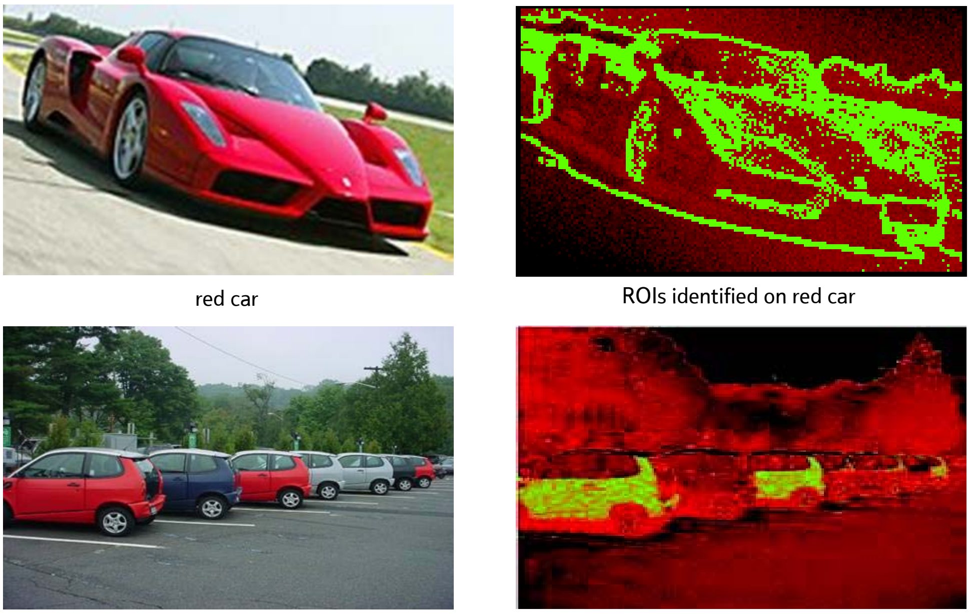
$$d'(H_A, H_B) = \sum_{i=1}^{n} \left| H'A(i) - H'B(i) \right|$$

$$H'_\alpha(i) = \frac{\text{number of pixels with hue } i \text{ and } M_\alpha(x, y) = 1}{\text{number of pixels in } \alpha \text{ and } M_\alpha(x, y) = 1}$$

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.

3.2 Process model

A similarity metric when applied to the images in a collection creates a network of associations between pairs of images each 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



red car

ROIs identified on red car

red cars and trees

ROIs identified on red cars

Fig 2 Images and corresponding VA map.

similar objects in spite of possessing different backgrounds. It is this additional metadata that provides the information to enable a convergent and intelligent search path.

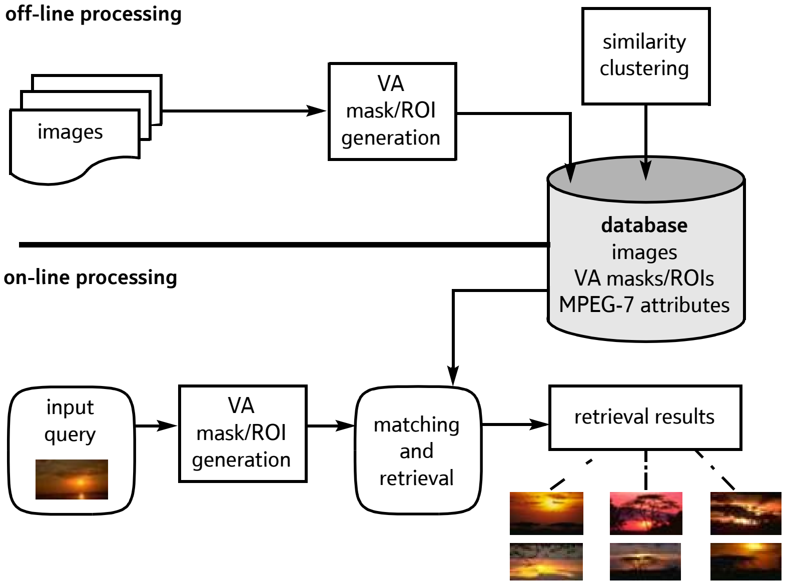
Images in a collection are processed off-line to produce metadata that is stored in the relational database. VA analysis is applied to a query image and the similarity of ROIs to others in the database is determined. A rank ordered list of candidate retrieved images is returned to the user as illustrated in Fig 3. The precomputed network of similarity associations enables images to be clustered according to their mutual separations. This means that query images are matched first with 'vantage' images [36] in each cluster before selecting images from within the closest cluster groups.

3.3 *Data model*

The data model encompasses regions of interest, images, clusters of images, and potentially a hierarchy of clusters of clusters. Similarity associations relate images and ROIs within clusters and images and ROIs in different clusters.

In addition, most images will be present in more than one cluster, for example, one on the basis of background content and another on the basis of foreground subject material.

Figure 4 illustrates an entity-relationship diagram (ERD) for the application. Two intermediate entities ('image to cluster mapping' and 'image to ROI



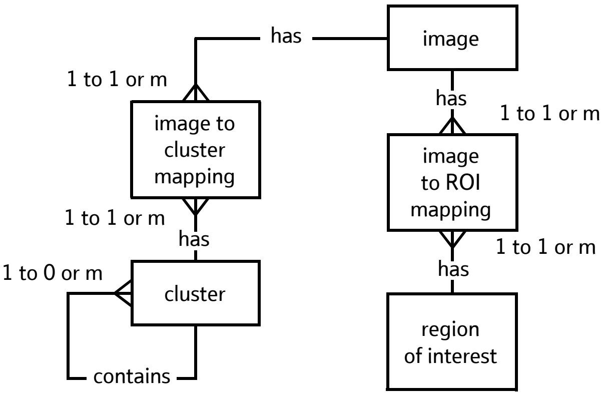


Fig 4 Application entity relationship diagram (ERD).

mapping') are inserted to break the potential many-to-many relationship between:

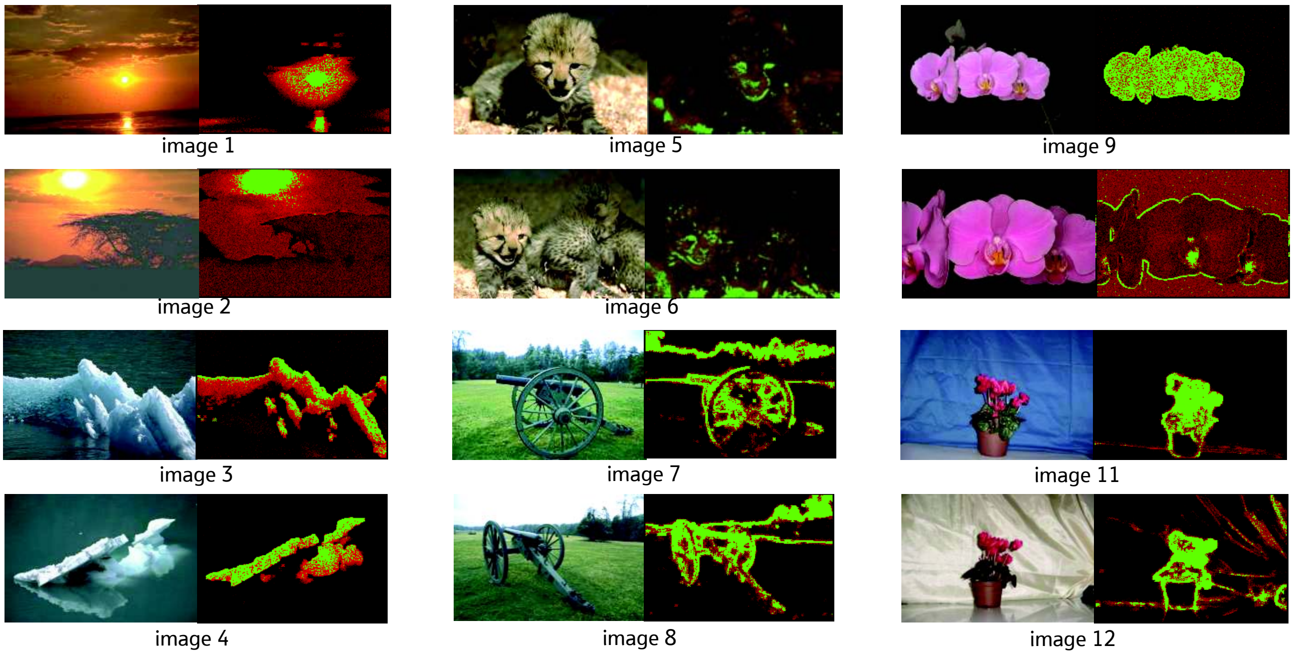
- 'cluster' and 'image',
- 'image' and 'region of interest'.

4. Results

The method is illustrated by application to a small set of 12 images consisting of 6 pairs that are clearly similar. Figure 5 shows the 12 images together with their VA maps.

The VA maps were obtained using the parameter values, t = 50, m = 1, and ε= 4. A mismatch is detected if any of the RGB values for the pixels being compared differ by more than 50. Each map yields a mask, which is used to construct the arrays $M_\alpha (x, y)$.

The histograms are based upon the hue values at each pixel, which range from 1 to 360. Examples of global and attention-based histograms for image 9 are shown in Fig 6. The difference is due mainly to the different colour profiles of the background and foreground.



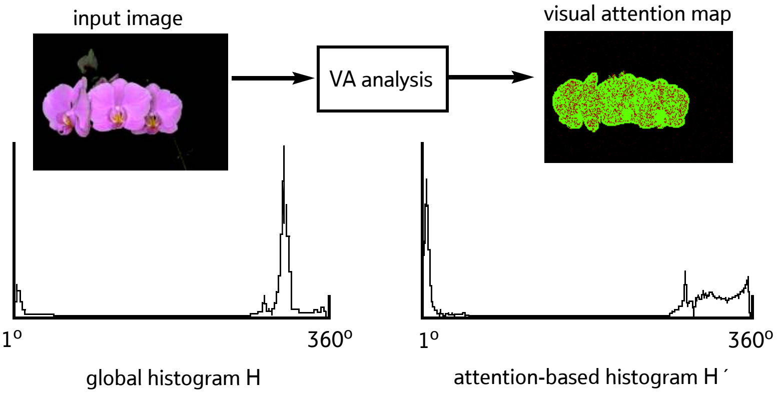


Fig 6 Colour histogram models.

The distances 100*d* and 100*d′* between all 12 of the images using the global and attention-based similarity measures were computed. In order to compare the global (P_i) and attention-based (P'_i) histogram performances on image *i*, the distances between the pairs of subjectively similar images (*i*, *j*) were compared to those between all the others where:

$$P_i = \left\{ \frac{\sum_{A \neq i, j} (d(H_A, H_i) - d(H_i, H_j))}{d(H_i, H_j)} \right\}$$

and similarly:

$$P'_i = \left\{ \frac{\sum_{A \neq i, j} (d'(H_A, H_i) - d'(H_i, H_j))}{d'(H_i, H_j)} \right\}$$

The comparative performance is displayed in Fig 7 where $100(P_i - P'_i)/10$ is plotted for each image. Positive values indicate improvements in performance over the global similarity measure.

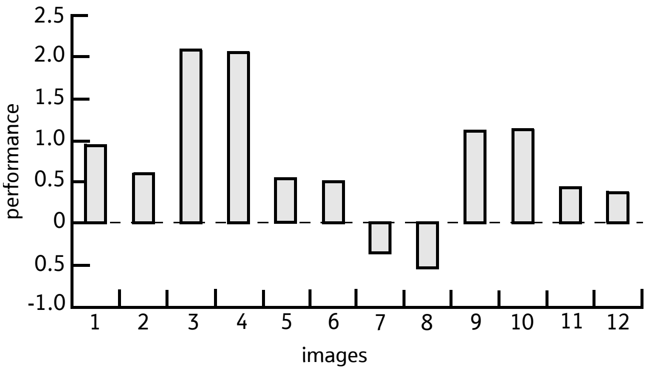


Fig 7 Image separation comparison.

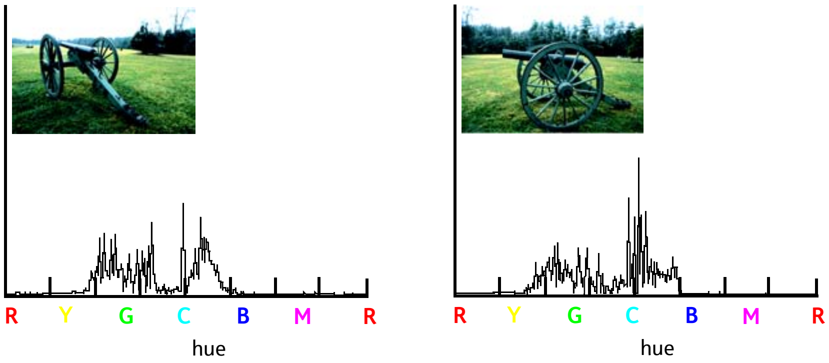
5. Discussion

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 (Fig 8) but is suppressed by the attention mechanism (Fig 9). 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 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.

Processing time on a 1.8 GHz machine for a 214 × 144 image is 543 ms for code written in C++. However, the score calculation in the VA algorithm is independent for each pixel and is therefore eminently suitable for parallel implementation.

6. Conclusions

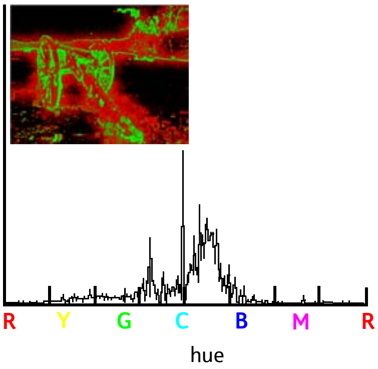
There is good reason to believe that the saliency of images should play a major part in automated image retrieval and this paper illustrates a way in which this might be achieved. The work has indicated that laying emphasis upon areas of images that attract high visual attention can improve retrieval performance. It seems reasonable that most image retrieval tasks will be largely determined by the principal subject material rather than the background content, although this will not always be the case. The results have also highlighted the well known failings of histogram-based metrics which take no account of image structure but



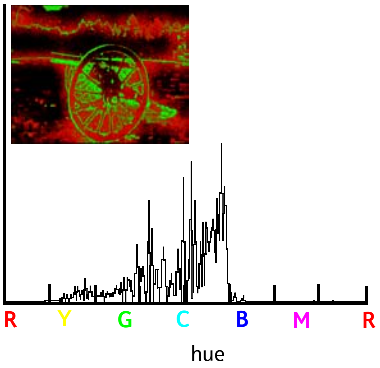
global histogram, H7

global histogram, H8

Fig 8 Global histogram of images 7 and 8.



attention histogram, H′7



attention histogram, H′8

have been used here as a vehicle for evaluating the effectiveness of measures of saliency in retrieval tasks.

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.

Future work on automating the selection of appropriate values for the VA parameters, t, m, and ε are continuing to develop the VA pipeline into a ‘black box’ suitable for commercial deployment. Furthermore, expanding the current implementation from being a single PC, multi-threaded process, to a multi-processor-based server cluster should see high returns as the algorithm is highly parallel in nature. This whole process will then need to be integrated with a commercially available content management system, such as ICF (Interactive Content Factory) from Trans World International, or Asset Manager from Asset House. Equally, content management workflows will need to be adapted to accommodate content ingestion from users and system analysis before distribution back to the user. Once achieved, the complete system will then be able to effectively serve a variety of ubiquitous devices in what we believe to be a cost-efficient manner.

A collaboration with Berkeley will provide a rich source of annotated images collected through mobile videophones on campus. This project [37] is revealing the communal benefits of mobile media creation, sharing and reuse. It takes advantage of previously annotated media to make educated guesses about the content of newly captured media. Visual attention technology promises to add value to the associations that can be automatically deduced from image content.

Content-based image retrieval technology, that can retrieve and even appear to anticipate users’ needs, will find huge application. It can be immediately applied to video retrieval through the medium of key frames. There is a growing demand for new video mobile services but which cannot become pervasive until accessibility bottle-necks are removed. Security and crime-prevention applications are increasingly hoping to rely on new technology to deliver visual content that is relevant to the moment, none of which is currently possible without heavy manual involvement. Natural and intuitive access to multimedia content is a vision for anyone and everyone in the communications industry.

Acknowledgments

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

References

1 Weiser M: ‘The computer for the 21st century’, Sci Amer (September 1991).

2 Brown P J and Jones G J F: ‘Context-aware retrieval for pervasive computing environments’, IEEE International Conference Proceedings on Pervasive Computing, Zurich, Switzerland (August 2002).

3 Lee D L and Lee W-C: ‘Data management in location-dependent information services’, Proc IEEE International Conference on Pervasive Computing, Zurich, Switzerland (August 2002).

4 Vendrig J: ‘Filter image browsing: a study of image retrieval in large pictorial databases’, Master’s thesis, Dept Computer Science, University of Amsterdam, The Netherlands (February 1997).

5 Stricker M and Swain M: ‘The capacity of colour histogram indexing’, IEEE CVPR, Seattle, pp 704—708 (1994).

6 Del Bimbo A and Pala P: ‘Visual querying by color perceptive regions’, Pattern Recognition, 31, pp 1241—1253 (1998).

7 Jain A K and Vailaya A: 'Image retrieval using color and shape', Pattern Recognition, 29, No 8, pp 1233—1244 (1996).

8 Ortega M, Rui Y, Chakrabarti K, Mehrotra S and Huang T S: 'Supporting similarity queries in MARS', Proceedings of Fifth ACM International Multimedia Conference, Seattle, USA (1997).

9 Mehrotra S and Chakrabarti K: 'Similarity shape retrieval in MARS', IEEE International Conference on Multimedia and Expo, New York (2000).

10 Lu G and Phillips J: 'Using perceptually weighted histograms for colour-based image retrieval', Proceedings of 4th Int Conf on Signal Processing Proceedings, ICSP '98, (Vol 2) (1998).

11 Niblack W and Flickner M: 'Query by image and video content: the QBIC system', IEEE Computer, pp 23—32 (September 1995).

12 Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R and Shu C: 'The Virage image search engine: an open framework for image management', Proceedings of the SPIE Storage and Retrieval for Image and Video Databases IV, San Jose, CA, USA, pp 76—87 (February 1996).

13 Veltkamp R C and Tanase M: 'Content-based retrieval systems: a survey', (March 2001) — http://www.aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/

14 Beige M, Benitez A B, and Chang S F: 'MetaSeek: a content-based meta-search engine for images', Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases VI, San Jose, CA (January 1998) — http://ana.ctr.columbia.edu/metaseek/

15 Kulkami S: 'Interpretation of fuzzy logic for texture queries in CBIR', in: 'Vision, Video and Graphics', Prentice Hall and Willis (2003).

16 Carson C, Belongie S, Greenspan H and Malik J: 'Blobworld: segmentation using expectation-maximisation and its application to querying', IEEE Trans PAMI, 24, No 8, pp 1026—1038 (August 2002).

17 Wang J, Li J Z and Wiederhold G: 'SIMPLIcity: semantics-sensitive integrated matching for picture libraries', IEEE Trans PAMI, 23, No 9, pp 947—963 (September 2001).

18 Smith J R and Chang S-F: 'VisualSEEk: a fully automated content-based query system', Proc ACM Int Conf Multimedia, pp 87—98, Boston, MA (November 1996).

19 Ma W-Y and Manjunath B S: 'NeTra: a toolbox for navigating large databases', Multimedia Systems, 7, pp 184—198 (1999).

20 Smeulders A W M, Worring M, Santini S, Gupta A and Jain R: 'Content-based retrieval at the end of the early years', IEEE Trans PAMI, 22, No 12, pp 1349—1379 (December 2000).

21 Vinod V and Murase H: 'Focused color intersection with efficient searching for object extraction', International Conference on Multimedia Computing and Systems, Pattern Recognition, 30, No 10, pp 1787—1797 (1997).

22 Rui Y, Huang T S, Ortega M, and Mehrotra S: 'Relevance feedback: a power tool for interactive content-based image retrieval', IEEE Trans on Circuits and Video Technology, pp 1—13 (1998).

23 Ciocca G and Schettini R: 'A multimedia search engine with relevance feedback', Proc SPIE, 4672, San Jose (January 2002).

24 Taycher L, Cascia M La, and Sclaroff S: 'Image digestion and relevance feedback in the ImageRover WWW search engine', Proc 2nd Int Conf on Visual Information Systems, San Diego, pp 85—94 (December 1997).

25 Cox I J, Miller M L, Minka T P, Papathomas T V and Yianilos P N: 'The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments', IEEE Trans Image Processing, 9, No 1 (January 2000).

26 Innes M and Jose J M: 'A personalised information retrieval tool', 26th Int ACM SIGIP Conf on Research and Development in Information Retrieval, Toronto (July—August 2003).

27 Stentiford F W M: 'An attention based similarity measure with application to content based information retrieval', SPIE, 5021, Storage and Retrieval for Media Databases, Santa Clara (January 2003).

28 Pauwels E J and Frederix G: 'Finding Salient Regions in images: non-parametric clustering for image segmentation and grouping', Computer Vision and Image Understanding, 75, Nos 1 and 2, pp 73—85 (August 1998).

29 Salembier P: 'Overview of the MPEG-7 standard and of future challenges for visual information analysis', EURASIP Journal on Applied Signal Processing (2002).

30 Bamidele A and Stentiford F W M: 'Image retrieval: a visual attention based approach', Postgraduate Research Conference in Electronics, Photonics, Communications and Networks, and Computing Science, Hertfordshire (April 2004).

31 Stentiford F W M: 'An estimator for visual attention through competitive novelty with application to compression', Picture Coding Symposium, Seoul (April 2001).

32 Oyekoya O K and Stentiford F W M: 'Exploring human eye behaviour using a model of visual attention', International Conference on Pattern Recognition, Cambridge (August 2004).

33 Oyekoya O K and Stentiford F W M: 'Eye tracking as a new interface for image retrieval', BT Technol J, 22, No 3, pp 161—169 (July 2004).

34 Nothdurft H-C, Gallant J L and Van Essen D C: 'Response modulation by texture surround in primate area VI: Correlates of 'popout' under anesthesia', Visual Neuroscience, 16, pp 15—34 (1999).

35 Petkov N and Westenberg M A: 'Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition', Biol Cybern, 88, pp 236—246 (2003).

36 Vleugels J and Veltkamp R C: 'Efficient image retrieval through vantage objects', Pattern Recognition, 35, pp 69—80 (2002).

37 Sarvas R, Herrarte E, Wilhelm A and Davis M, 'Metadata creation system for mobile images', 2nd International Conference on Mobile Systems, Applications and Services (MobiSys 2004), Boston (2004).

Adetokunbo Bamidele graduated in 2000 from Kings College London, with a BEng honours degree in Computer Systems and Electronics. As part of his degree he received a sponsorship for a year with the Omron Corporation in Japan working as a software research and development engineer. Following graduation he worked for a period as a systems support analyst with Business Financial Solutions and then as a universal multimedia messaging systems analyst with the Unisys Corporation.

He is now studying for an Engineering Doctorate in Telecommunications while based at the UCL Adastral Park Campus. His work is sponsored at UCL by the Understanding Visual Content group within BT's Broadband Applications Research Centre. He is an associate member of the IEE.

Fred Stentiford won a scholarship to study Mathematics at St Catharine's College, Cambridge, and obtained a PhD in Pattern Recognition at Southampton University. He first joined the Plessey Company to work on various applications including the recognition of fingerprints and patterns in time varying magnetic fields. He then joined BT and carried out research on optical character recognition and speech recognition. From 1983 he led a team developing systems employing pattern recognition methods for the machine translation of text and speech. This work led to the world's first demonstration of automatic translation of speech between different languages. He led research into the design of new dialogues for telephone services and managed the government funded collaborative Dialogues 2000 project which aimed to promote common standards in the spoken user interface in UK industry. He then returned to vision research and led a group developing new systems for analysing and delivering multimedia content over the telecommunications networks. He now holds a chair in Telecommuni-cations with UCL and leads a team at the Adastral Park Campus researching new technologies for understanding visual content, in close collaboration with BT's Broadband Applications Research Centre. He is a corporate member of the IEE and the BCS and has published over 50 papers and filed 15 patents on pattern recognition techniques.

Jason Morphett holds a PhD in Computer Science from the University of Nottingham. He graduated from the University of East Anglia with a BSC (1st) in Software Engineering in 1995 and joined BT at Adastral Park soon after.

He was until recently the Head of the Future Content Group, but now works on bridging the commercial and technical divide in bringing new media propositions through Research and Enterprise Venturing for BT Group CTO. He is currently engaged in bringing forward new media content delivery over broadband to non-PC devices.

He is a Chartered Engineer and a member of the Institute of Electrical and Electronic Engineers, has published over 20 papers, and holds a number of patents in the field of content management and distribution.