Recognition and retrieval via histogram trees

Stuart Gibson and Richard Harvey
School of Information Systems,
University of East Anglia, Norwich, NR4 7TJ, UK.
Email:{s.e.gibson, r.w.harvey}@uea.ac.uk

Abstract

This paper explores a new method for analysing and comparing image histograms. The technique amounts to a novel way of backprojecting an image into one with fewer, statistically significant colours. When the method is tested with the COIL-100 and MPEG-7 data sets it is shown to have a performance that is as good as the best methods using fewer enteries than the original histogram. Therefore it offers the potential for extending the use of histograms into high dimensional feature spaces.

1 Introduction

Histograms are popular features for recognition and retrieval [20]. They form part of the forthcoming MPEG-7 standard for image and video metadata and, despite their known shortcomings as density estimators [26], they are popular because of their computational simplicity.

When using histogram-based techniques, important issues are the selection of an appropriate feature space, the quantization of the selected space, and the algorithm for comparing histograms. The choice of feature space is often problem specific (see [4] for an example where colour spaces are compared). The choice of histogram quantisation (bin size) is discussed in, for example, Scott [19][1]. This paper examines the last issue, the problem of comparing multidimensional histograms. We review the existing options and suggest a new method for representing and comparing multidimensional histograms.

There are several reviews of histogram comparison methods ([13, 16] are examples). The methods may be summarized as being based on inter-bin distances, intra-bin distances or feature distances. The inter-bin distances take the form

$$d(h,k) = g(\sum_i f(h_i,k_i))/W \tag{1}$$

where $h = [h_1,\ldots,h_n]$ and $k = [k_1,\ldots,k_n]$ denote the n-bin histograms, g and f are functions that vary from method to method, and W is a scaling factor. Table 1 summarizes these for a variety of inter-bin distances. For colour histograms formed in a non-invariant colour space the effect of lighting variation is to shift the modes of the underlying distribution. In this case inter-bin distances can be ineffective. The usual solution is to use an intra-bin distance, a perceptual colour space or both. The best known intra-bin distance

[1] From which one deduces that many Computer Vision systems often operate with undersmoothed histograms.

BMVC 2001 doi:10.5244/C.15.55

Distance type	$g(x)$	$f(h_i,k_i)$	W
Minkowski	$x^{1/p}$	$\lvert h_i-k_i\rvert^p$	1
Intersection	x	$k_i-\min(h_i,k_i)$	$\sum_i k_i$
Kullback-Leibler	x	$h_i \log h_i/k_i$	1
Jeffrey	x	$h_i \log \frac{h_i}{m_i}+k_i \log \frac{k_i}{m_i}$	1
χ^2	x	$\frac{(h_i-m_i)^2}{m_i}$	1

Table 1: Some inter-bin distances: the Minkowski distance [22–24]; histogram intersection [24]; Kullback-Leibler divergence [12]; the Jeffrey divergence [16] in which $m_i=\frac{h_i+k_i}{2}$ and the χ^2 distance [10].

is probably that used in the QBIC system [8] and allows for intra-bin similarity measures via a weighting matrix $\mathbf{A}$.

$$d_{QF}(h,k)=\left((\mathbf{h}-\mathbf{k})^T\mathbf{A}(\mathbf{h}-\mathbf{k})\right)^{1/2} \tag{2}$$

$\mathbf{A}$ is often chosen as $[\mathbf{A}]_{ij}=1-d_{hsv}(h_i,k_j)$ where d_{hsv} is the distance between the colour of the bin centres [21]:

$$d_{hsv}=1-1/\sqrt{5}\left((v_i-v_j)^2+(s_i\cos h_i-s_j\cos h_j)^2+(s_i\sin h_i-s_j\sin h_j)^2\right)^{1/2} \tag{3}$$

An alternative, for one-dimensional histograms, is to use a cumulative distance

$$d_M(h,k)=\sum_i |\hat{h}_i-\hat{k}_i| \tag{4}$$

where $\hat{h}_i=\sum_{j\le i}h_j$ is the cumulative histogram of h, and similarly for k. Equation (4) is related to the Kolmorgorov-Smirnov distance $d_{KS}(h,k)=\max_i(|\hat{h}_i-\hat{k}_i|)$ but for both distances there is no logical extension to two or more dimensions because there is no unique ordering of histogram bins. Note that all the distances defined so far assume that h and k have identical quantisations which can also be a significant restriction.

This restriction may be removed via the Earth Mover's Distance (EMD) metric [17] which measures the distance between distributions by calculating the minimum amount of work to transform one into the other. EMD is based on a linear programming problem in which $P=\{(p_1,w_1),\ldots,(p_m,w_m)\}$ is the first signature of m clusters (for histograms the clusters are usually the bins), $Q=\{(q_1,w_1),\ldots,(q_n,w_n)\}$ the second signature with n clusters; and $D=[d_{ij}]$ a matrix where $d_{ij}=d(\mathbf{p}_i,\mathbf{q}_j)$ is the ground distance between clusters i and j. The method computes the flow $F=[f_{ij}]$, where f_{ij} is the flow between $\mathbf{p}_i$ and $\mathbf{q}_j$, that minimizes the cost $\sum_{i=1}^m\sum_{j=1}^n d(\mathbf{p}_i,\mathbf{q}_j)f_{ij}$ subject to the constraints: $f_{ij}\ge 0$; $\sum_{j=1}^n f_{ij}\le w_i$; $\sum_{i=1}^m f_{ij}\le w_i$; and $\sum_{i,j=1}^{m,n} f_{ij}=\min(\sum_{i=1}^m w_{\mathbf{p}_i},\sum_{j=1}^n w_{\mathbf{q}_j})$ where $1\le i\le m$ and $1\le j\le n$. The first constraint allows the movement of mass from P to Q and not vice versa. The second constraint limits the amount that can be sent by the clusters in P to their weights. The third constraint limits the clusters in Q to receive no more than their weights; and the fourth constraint limits the total flow. Having solved this transportation problem and calculated F, the EMD is defined as the resulting work normalized by the total flow:

$$d_{EMD}(P,Q)=\frac{\sum_{i=1}^m\sum_{j=1}^n d(\mathbf{p}_i,\mathbf{q}_i)f_{ij}}{\sum_{i=1}^m\sum_{j=1}^n f_{ij}} \tag{5}$$

In practice EMD is slow to compute [3] so tends to be restricted to histograms with small numbers of bins.

An alternative to using the full histogram is to compare features. One method uses features that are the sum of the weighted distances of the first three moments of a distribution [22]. Another computes the peaks of each histogram and compares these to give a candidate list of similar images [6]. A more sophisticated recent alternative is to build features based on modes in a local colour histogram [13]. An earlier method that has relevance to the proposal in this paper assigns a signature, $S(h)$, to every histogram h by finding bins that are local maxima and above some threshold [27]. $S(h)$ is the set of ratios of each local maximum to every other local maximum. The score for each comparison is created by a goodness function, $\Gamma(h_n,h_d) = \left(\sum_{\{a,b\}\in S(h_d)} \rho(a,b,h_n,h_d)\right) / \parallel S(h_d) \parallel$ where a and b denote bins in the histogram and ρ is called the ratio-match $\rho(a,b,h,h') = w(\ln r_{a,b}(h) - \ln r_{a,b}(h'))$ where $w(x)$ is a narrow Gaussian with a maximum of 1 centered at 0 and $r_{a,b}(h)$ is a ratio of the number of pixels in bin a to the number of pixels in bin b in histogram h. The Gaussian is used to assign a high goodness value to close matches. A confidence that a histogram h_n contains an object o is calculated as: $\text{conf}(h_n,o) = \max_{h_d\in H_o} \Gamma(h_n,h_d)$ where H_o is the set of histograms stored in the database as examples of object o.

A related development is the generation of features from compressed colour histograms [3] in which it is shown that, for a controlled database, retrieval performance for compressed colour histograms is similar to that with a full histograms.

2 Histogram trees

In [27], [6] and [13] it is argued that it is the extrema, especially maxima, of histograms that provide the most useful features. Thus any simplification of the histogram should not enhance existing extrema: the *scale-space* causality principle [11]. There are several possible systems that preserve scale-space causality in two-dimensions (see [9] for a review) but few that extend easily to $N > 2$ dimensions especially when the scale-space causality requirement is tightened to demand that no new extrema are introduced as scale increases[2]. It is known in graph morphology [18] that alternating sequential filters by reconstruction can be configured to satisfy the scale-space causality principle in N dimensions [2] and in these are termed *sieves*. Related filters are max/min-trees [18] and watershed-trees [25]. Here we concentrate on max trees since we are interested in representing the peaks of a probability distribution.

Such a tree may be defined in any number of dimensions by considering a histogram to be a set of connected voxels with their connectivity represented as a graph, $G = (V,E)$ where the set of vertices, V, are the voxel labels and the set of edges, E, represent the adjacencies. If the set of connected subsets of G containing r elements is $C_r(G)$ then the set of connected subsets of r elements containing a particular vertex x may be defined as,

$$C_r(G,x) = \{\xi \in C_r(G) | x \in \xi\}. \tag{6}$$

Morphological operators and functions may be defined on these connected regions. For a histogram $h(x)$ and for each integer $r \geq 1$, an operator $\psi_r : Z^V \mapsto Z^V$ can be defined over

[2]The usual, looser, requirement is that existing extrema should not be enhanced.

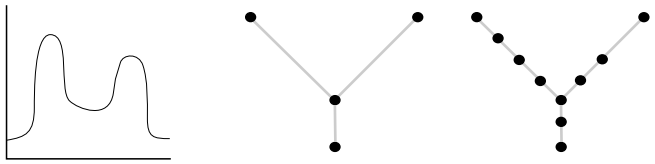


Figure 1: A one-dimensional histogram (left) and its associated watershed tree (center) and max-tree (right)

the graph, G, where

$$\psi_r h(x) = \max_{\xi \in C_r(G,x)} \min_{u \in \xi} h(u) \tag{7}$$

Applying this operator in a serial or recursive structure in which the output at a scale r, h_r is

$$h_{r+1} = \psi_{r+1} h_r \tag{8}$$

gives progressively simplified histograms. The differences between successive outputs $d_r = h_r - h_{r+1}$ have support regions that are connected sets of scale $r-1$. The boundaries of these support regions are contours of the histogram. As scale increases the contour expands so that contours of a given scale will always be contained by a larger scale contour. This containment may be represented by the edges of a scale tree [14, 18]. Figure 1 shows a one-dimensional example of the max tree. The leaf nodes represent the maxima. Parents of these leaf nodes represent successively lower slices of the two modes. The node with two children represents the slice where two peaks are no longer separated (a watershed). The root node represents the complete signal. Also shown is a watershed tree [25] which contains a subset of the nodes in the max tree but does not contain enough information to reconstruct the original signal (unlike the trees in this paper which are a transform of the signal). Inverting the signal or using the inverse to the ψ operator would produce a

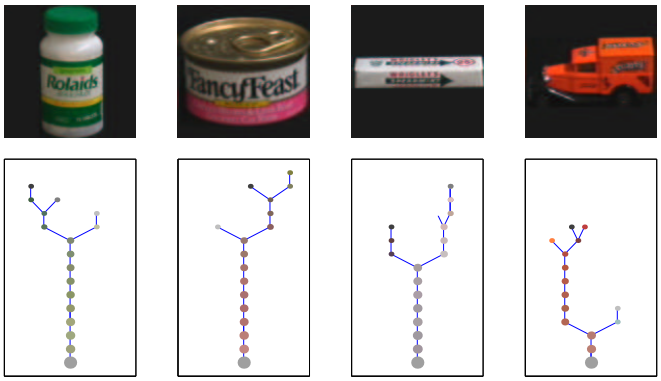


Figure 2: Images of Objects 5, 32, 67 and 100 from the Coil-100 database and their associated scale trees.

min-tree. When applied to images, max and min trees are less robust than trees built using the bipolar $\mathcal{M}$- or $\mathcal{N}$-filters [9].

Figure 2 shows images of four objects taken from the Columbia Database (Coil-100) [15] and their corresponding schematic max trees. Here the scale tree was computed via an immersion simulation from a $4 \times 4 \times 4$-bin RGB colour histogram. The scale and colour of each node in the tree represents the scale and mean colour of the flat-zone associated with the node. Thus the root of the tree is the large grey sphere and the leaves are shown as small coloured dots. The schematic version of the tree shown in Figure 2 is a two-dimensional simplification of the original three-dimensional max-tree. This is created by determining the number of nodes on a particular depth and creating a spacing table, which means that the edges of the tree visualisation do not overlap.

The tree in Figure 2 is a visualization tool – a property that has been exploited to visualize high-dimensional surfaces in signal processing [7]. Note that the topology of the tree encodes the histogram topology and so provides some invariance against lighting variations that alter the location but not number of modes.

Where the histogram has only a few narrow modes the tree will have only a few nodes but for broad under-smoothed histograms the tree may have many nodes. In these cases it is useful to have a pruning algorithm to reduce the size of the tree while retaining the important structures. For image trees such a pruning algorithm exists [14] but for histograms, where the tree nodes represent equi-probable contours, it is simple to devise pruning strategies that discard nodes on uni-modal branches until some specified tree size is reached. In this way the scale tree provides a controllable level of detail with the simplest trees identical to the watershed trees and the most detailed sufficient to reconstruct the original histogram.

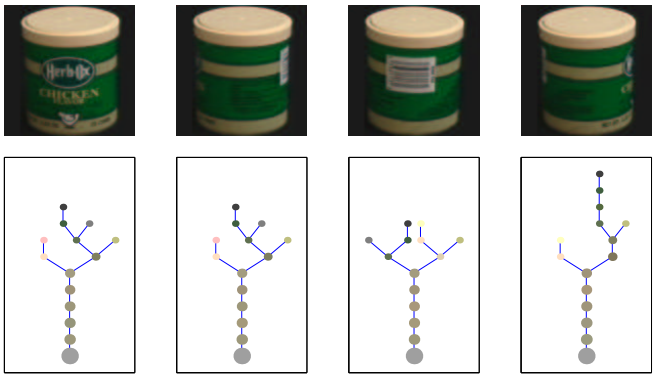
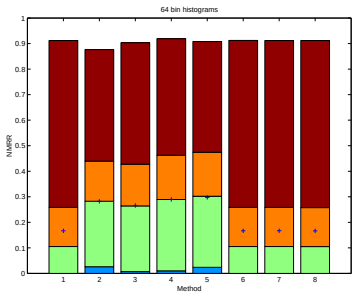


Figure 3: Top, object 26 at 0, 90, 180 & 270 degrees rotation and, bottom the associated scale trees.

3 Results

To measure the effectiveness of these trees for image retrieval we first use the Columbia Coil-100 Database, containing 7200 images. This database was originally created for testing viewpoint invariant object recognition [15]. The database contains images of 100

objects, each rotated through 360 degrees, with an image for every 5 degrees of rotation. Several authors have built systems to solve the COIL-100 problem usually with some degree of success (see [5] for an example). Figure 3 shows four images of object 26 from the Coil-100 database and the corresponding schematic scale trees.



Key	
1	χ^2
2	EMD in RGB space
3	EMD in Lab space
4	QBIC with $A = I$
5	QBIC with A from (2)
6	Tree of query image with χ^2 distance
7	Tree of database image with χ^2 distance
8	Tree metric with χ^2 distance

Figure 4: Showing the maximum, upper quartile, median, lower quartile and minimum NMRR computed over 100 queries for eight comparison methods using the COIL-100 set. Also shown, as crosses, is the mean NMRR. Note: The minimum NMRR is 0 in all cases; the lower quartile NMRR is too small to plot for methods 1,6,7 & 8.

This is essentially an image retrieval problem, so to quantify the retrieval performance we use the MPEG-7 Normalized Modified Retrieval Rank (NMRR) [1] in which the qth query has $N_G(q)$ ground truth images. The nth retrieved ground truth image is assigned a modified rank r_n where $r_n = 1$ indicates it is the top match, $r_n = 2$ the second best match and so on. A rank of $K+1$ is assigned to any ground truth retrievals that are not in the first K retrievals on the basis that images that are not retrieved are all equally useless. K is a number chosen to represent a reasonable depth into the database. Here we use the MPEG-7 recommendation $K = \min(4N_G(q), 2\max_q(N_G(q)))$. The NMRR is defined as

$$\text{NMRR} = \frac{\text{MRR}(q)}{K + 1/2 - N_G/2} \tag{9}$$

where $\text{MRR}(q) = \mu(q) - 1/2 - N_G(q)/2$ and $\mu(q) = \sum_{i=1}^{N_G(q)} r_i/N_G(q)$ and $N_G(q)$. A NMRR of zero indicates that all the ground truth images have been retrieved; an NMRR of unity indicates that none of the ground truth images have been retrieved. A query for each object was formulated by using the image of the object at zero rotation as the query image and all images of the object as the ground truth set when searching through the entire Coil-100 database.

Figure 4 gives the retrieval results for several conventional methods and three new methods using the tree. In the first tree method (method 6 in Figure 4) a $4 \times 4 \times 4$ bin RGB histogram built from the query image was mapped into a max tree. Each pixel in the image can be associated with a tree node by mapping histogram bins to tree nodes and mapping pixels to histogram bins. By this means, commonly called backprojection, an RGB image is converted into an indexed image in which the index is the tree-node label. Images may then be compared using a χ^2 distance between the one-dimensional histograms of the indexed images. A refinement (method 7 in Figure 4) is to repeat the process using the database image to build the tree. Taking the maximum of these two, gives a metric (method 8 in Figure 4).

The best performing methods for this task are χ^2 (method 1) and the tree schemes (methods 6,7 & 8), but note that the tree methods use fewer bins.

Building trees from simple images, such as those in the COIL-100 database is illustrative, but is not a realistic image retrieval task. However the formulation of realistic retrieval tasks is non trivial because the type of query can vary considerably, the type of image can vary considerably and the assessment of correct retrieval (ground truth) can vary considerably. Fortunately the MPEG-7 standards committee has addressed this task via the MPEG-7 common colour dataset and common colour queries [1]. This task uses a database of 5466 images and a set of fifty queries with predefined ground truth images. Figure 5 shows some ground truth images for two of the MPEG-7 queries plus a selection of randomly chosen images from the MPEG-7 common colour dataset. Figure 6 shows the results for a variety of retrieval metrics using both $4 \times 4 \times 4$ bin and $10 \times 10 \times 10$ bin histograms.

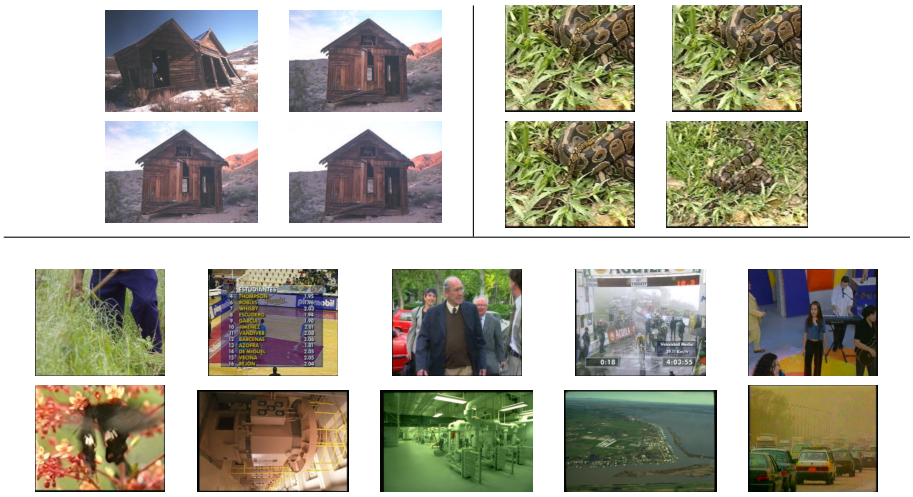


Figure 5: Showing, top left, four of the six ground truth images from Query 49. Four out the twelve ground truth images from Query 16 (top right). Ten randomly chosen images from the 5466 images in the common colour data set (bottom).

For the 64 bin histogram (left of Figure 6) in the best performing methods the lower quartile range is too close to zero to be visible. The best performing methods are conventional χ^2 and the tree methods followed by χ^2. However from Figure 7, the mean number of tree nodes is roughly 18, and the maximum number of tree nodes is 45 – a considerable reduction compared to the 64 entries in the original histogram. For the 1000 bin histogram the tree methods are marginally superior to χ^2 method and considerably superior to the others. We have not computed the EMD metric because the amount of computation is impractical. In this case the largest tree has 192 nodes. The mean number of nodes is 69. As the number of bins in the original histogram increases the compression grows. As an aside we mention that the correlation coefficient between the χ^2 performance and the tree methods is always above 0.83 and between the χ^2 and the tree methods always above 0.91, implying that some queries are difficult / easy for all methods.

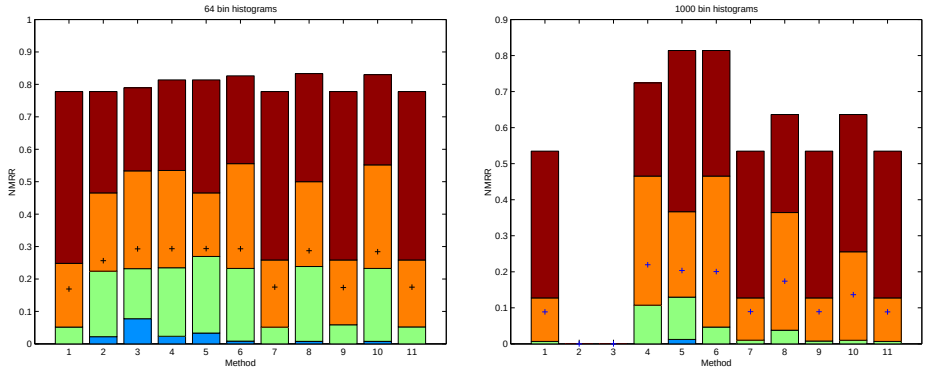


Figure 6: Showing the maximum, upper quartile, median, lower quartile and minimum NMRR computed over 50 queries for eleven comparison methods. Also shown, as crosses, is the mean NMRR. The methods are: 1) χ^2, 2) EMD in LAB space, 3) EMD in RGB space, 4) QBIC with $\mathbf{A} = \mathbf{I}$, 5) QBIC with $\mathbf{A}$ from (2), 6) Tree built from query image and Minkowski distance, 7) Tree built from query image with χ^2 distance, 8) Tree built from database image and Minkowski distance, 9) Tree built from database image with χ^2 distance, 10) Tree metric with Minkowski distance, 11) Tree metric with χ^2 distance.

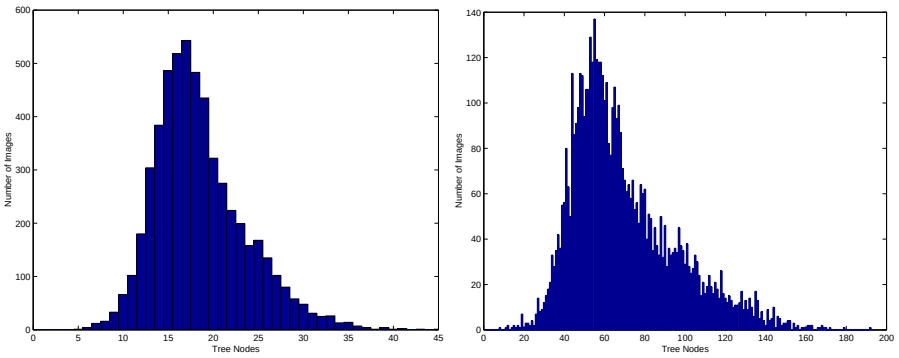


Figure 7: The number of tree nodes required to represent the $4 \times 4 \times 4$ histograms (left) and $10 \times 10 \times 10$ (right) of all the images in the MPEG-7 Common Colour Dataset.

4 Conclusions

This paper discussed a new scale-space based method for representing, analyzing and simplifying multidimensional histograms. The method uses a tree based representation that converts a multidimensional histogram into a data structure with fewer data entries than the original histogram. There are several options for building the tree. Here we built a max tree using an immersion simulation but other possibilities are multidimensional watersheds or sieves. Of these, the later are known to have attractive computational complexity – an important consideration for practical implementations. A further possibility is that trees built using $\mathcal{M}$- or $\mathcal{N}$ filters (as opposed to the openings or closings used here) can represent maxima and minima in the same data structure. At the moment we analyse

the histogram via only its maxima, but in the future we would like to know if the minima matter.

The trees are a valuable visualization tool because they can map an N-dimensional histogram into a two-dimensional plot. A further advantage is that it is possible to control their detail so the resulting data structure is manageable. Testing using the MPEG-7 methodology reveals that the new method is at least as good as commonly used alternatives with the added benefit of using fewer entries than the original histogram. Initial experiments indicate that retrieval performance remains robust even when the trees are simplified further.

5 Acknowledgements

The authors would like to thank Dr Steven Hordley (UEA) for his help with EMD and Dr Emanuele Trucco (Heriot-Watt) for suggesting the use of the COIL database.

References

[1] Description of core experiments for MPEG-7 color/texture descriptors. Technical Report ISO/IEC JTC1/SC29/WG11/N2929, MPEG-7, October 1999.

[2] J.A. Bangham, R.W. Harvey, P.D. Ling, and R.V. Aldridge. Morphological scale-space preserving transforms in many dimensions. *Journal of Electronic Imaging*, 5(3):283–299, Jul 1996.

[3] J. Berens, G.D. Finlayson, and G. Qiu. Image indexing using compressed colour histograms. *IEE Proc.: Vis., Im. and Sig. Proc.*, 147(4):349–355, August 2000.

[4] Yi Chan, Richard Harvey, and J. Andrew Bangham. Using colour features to block dubious images. In *EUSIPCO-2000*, volume III, Tampere, Finland, Sept 2000.

[5] V. Colin de Verdiere and J.L. Crowley. Visual recognition using local appearance. In *European Conference on Computer Vision*, pages 640–654, Jun 1998.

[6] M. Das, E.M. Riseman, and B.A. Draper. Focus: Searching for multi-colored objects in a diverse image database. In *CVPR-97*, pages 756–761, June 1997.

[7] M. Fisher, D. Mandic, J.A Bangham, and R.W. Harvey. Visualising error surfaces for adaptive filters and other purposes. In *ICASSP-2000*, pages 3522–3525, 2000.

[8] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, 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*, 28(9):23–32, Sept 1995.

[9] R. Harvey, A. Bosson, and J.A. Bangham. Robustness of some scale-spaces. In *British Machine Vision Conference*, volume 1, pages 11–20, 1997.

[10] M.G. Kendall and A. Stuart. *The advanced theory of statistics Volume 2*, pages 422–425. Charles Griffin and Company, 1961.

[11] J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[12] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, Mar 1951.

[13] J. Matas. *Colour-based Object Recognition*. PhD thesis, University of Surrey, UK, 1996.

[14] K. Moravec, R. Harvey, and J.A. Bangham. Scale trees for stereo vision. *IEE Proceedings: Vision, Image and Signal Processing*, 147(4):363–370, August 2000.

[15] S.K. Nayar, S.A. Nene, and H. Murase. Real-time 100 object recognition system. In *ARPA Workshop on Image Understanding*, pages 1223–1227, Feb 1996.

[16] J. Puzicha, T. Hofmann, and J. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 267–272, June 1997.

[17] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In *ICCV-98*, pages 59–66, Jan 1998.

[18] P. Salembier and L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation and information retrieval. *IEEE Transactions on Image Processing*, 9(4):561–576, April 2000.

[19] D.W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, December 1979.

[20] A.W.M. Smeulders, M. Worring, S. Santini, and R. Gupta, A.and Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[21] J.R Smith and Shih-Fu. Chang. Visualseek a fully automated content-based image query system. In *ACM Multimedia*, pages 87–98, Nov 1996.

[22] M. Stricker and M. Orengo. Similarity of color images. In *SPIE Conference on Storage and Retrieval for Image and Video Databases IV*, volume 2420, pages 381–392, Feb 1995.

[23] M. Stricker and M.J. Swain. The capacity of color histogram indexing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 704–708, 1994.

[24] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[25] L. Vincent and P. Soile. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, June 1991.

[26] M.P. Wand. Data-based choice of histogram bin width. *Statistical Computing and Graphics*, 51(1):59–64, Feb 1997.

[27] L.E. Wixon. Real-time qualitative detection of multi-colored objects for object search. In *Image Understanding Workshop*, pages 631–638. Morgan Kaufmann, 1990.