Interpreting Image Databases by Region Classification

Neill W. Campbell
Advanced Computing Research Centre
University of Bristol
Bristol, UK

William P.J. Mackeown
Dept. of Computer Science
King's College London
London, UK

Barry T. Thomas
Advanced Computing Research Centre
University of Bristol
Bristol, UK

Tom Troscianko
Dept. of Psychology
University of Bristol
Bristol, UK

Contact the first author at:

Dept. of Computer Science
Room 3.29
Merchant Venturers Building
Woodland Road
University of Bristol
Bristol
BS8 1UB
England, UK

Summary

This paper addresses automatic interpretation of images of outdoor scenes for image indexing and retrieval from databases. The method allows instances of objects from a number of generic classes to be identified: vegetation, buildings, vehicles, roads, etc., thereby enabling image databases to be queried on scene content. This is achieved using a powerful set of image features which are used to train a neural network classifier. A large database of high-quality colour images of outdoor scenes developed at Bristol University provides a ground-truth interpretation of the images, which has enabled a detailed quantitative analysis of the vision system performance.

The design of the feature set has been inspired by psychophysical models of vision. Texture is provided in the form of the magnitude response of a set of isotropic Gabor filters. Colour is represented by luminance, red/green and yellow/blue information. The shape of a region is represented by linear combinations of the principal modes of variation of an approximating polygon. The feature set also includes contextual information from an initial pixel classification. Progressive improvements in pixel classification are demonstrated by the addition of colour and texture information.

The image interpretation technique is very successful and correctly labels over 90% of the image area in our database of test images, for a wide range of object classes. Such a high recognition accuracy demonstrates that it is possible to interrogate large image databases by content.

Abstract

This paper addresses automatic interpretation of images of outdoor scenes. The method allows instances of objects from a number of generic classes to be identified: vegetation, buildings, vehicles, roads, etc., thereby enabling image databases to be queried on scene content. The feature set is based, in part, on psychophysical principles and includes measures of colour, texture and shape. Using a large database of ground-truth labelled images, a neural network is trained as a pattern classifier. The method is demonstrated on a large test set to provide highly accurate image interpretations, with over 90% of the image area labelled correctly.

Scene Understanding Object Recognition Classification
Image Segmentation Neural Networks

1 Introduction

The large and growing amount of digital image and video data on-line is stretching human ability to search this information by content. Tools for conducting such searches are still under development and are a long way from being powerful and flexible enough to be used as part of a general purpose system. Approaches to the image retrieval problem have employed mainly global descriptions of the scene, i.e. QBIC[(1)] and Photobook[(2)], and tried to identify other pictures in the database similar to a manually defined 'key' image. This paper proposes an alternative approach to image database classification and retrieval, that is the application of fully automatic segmentation and region classification steps. Thus, we aim to identify objects in the scene and index them by class. In our scheme, images are automatically analysed and indexed in addition to the database, so that their contents are available later for automatic retrieval.

To develop this approach, the Bristol Image Database [(3)] has been used. This database contains a large number of high quality, 24-bit colour images of outdoor scenes. A key component of the database is a human-generated ground-truth interpretation, that is, every pixel has been labelled as belonging to one of a large number of object categories: trees, bushes, stones, shadows, roads, cars, people, water, pavements, windows, etc. The database is being used to investigate many aspects of computer vision, including segmentation techniques and automatic image labelling. Previous work with this database has shown that a neural network trained as a pattern classifier using a ground-truth classification from the database and a set of region features, is capable of labelling regions automatically with high reliability on a large test set of images[3,4,5,6]. An important point is that the ground-truth information has enabled us to provide a very accurate, quantitative analysis of recognition performance on real images and has also allowed the contributions made by individual features and algorithms to overall recognition accuracy to be measured.

Outdoor scene analysis is a challenging problem in computer vision. Several different approaches have been used in previous work. Model-based approaches have met with some success in domains where objects can be well described using geometric primitives[7]. However, this approach is not ideal since it can be difficult to find good models for complex objects in outdoor scenes, e.g. trees. Knowledge-based approaches use a set of hand-coded rules to describe the properties and geometric relationships of each object to be recognised, together with rules for object-specific recognition strategies which reduce the overall complexity of the interpretation process. Examples are the SCHEMA vision system[8] and the region-based scene analysis system of[9], which have both been demonstrated to give good results in interpreting typical outdoor scenes. Drawbacks of these approaches are the significant effort required to develop the knowledge base and the computational complexity of the inference engine.

More recently, statistical approaches to outdoor scene analysis based on region classification have been demonstrated[4,10-12]. In[11], forward-looking infra-red images are automatically segmented using texture features based on the co-occurrence matrix, and then classified using a decomposition of the co-occurrence matrix as a feature vector for training a neural network. The method is reported to work very well for a five-class problem on an image sequence containing hundreds of images.

In previous work using the Bristol Image Database [4], a neural network system for region classification and a set of features based on properties of a region such as colour, brightness, texture, shape, topological properties, etc, were developed. The system achieved a mean classification accuracy of 86.3% per image by area on ideally segmented images. Although the performance of this scene analysis system is impressive, the use of an ideal segmentation model, whose implementation is largely manual, is unrealistic. In[5], a K-means segmentation technique was used to provide closed-region segmentations automatically. The same classification technique, when used on these regions, was shown to give a classification accuracy of 81.4% per image by area for the identical 11-class problem. For such a complex task, this classification rate is very impressive given the fully automatic nature of the segmentation and labelling steps.

This paper shows how fully automatic labelling of images may be achieved. It improves on previous methods by developing a new, more powerful feature set, based, in part, on a psychophysically plausible model of image information. The ability to label images is useful in the task of automatic indexing and retrieval in database applications.

2 Background

The work presented here has used colour images of outdoor road scenes from the Bristol Image Database. The database consists of over 350 images of a wide range of urban and rural scenes. The images were digitised using a calibrated digitiser from small-grain 35mm colour transparency film to produce high-quality 36-bit colour images. The statistics of the image content and acquisition conditions have been carefully controlled. Approximately half the images are rural and half are urban, spanning a wide range of viewpoints. Environmental conditions during image capture were dry, fully overcast and good atmospheric visibility (at least 1km). For all images, the camera was focussed at infinity and an aperture no larger than f/8 was used in order to ensure an adequate depth-of-field. The declination of the line of sight was approximately $4°$. The images have been hand-labelled to identify the objects in the scenes, thus providing a ground truth about the image contents. This database is a key resource in our work on quantifying vision system performance.

Throughout the work in this paper, features extracted from individual pixels and/or regions will be the input to a classifier. The classification is achieved by training an artificial neural network, a multi-layer perceptron, to act as a pattern classifier. The network has one hidden layer, and is optimised using a conjugate gradient descent algorithm. Non-overlapping training, validation and test sets, consisting of many thousands of examples, are used throughout.

The feature set used in[3-6] consisted of 29 features per region. These features, while proving to be very successful, do not take full advantage of the shape, texture and contextual information available in the images, and in this paper we therefore investigate several extensions and improvements to the feature set. As part of this improvement it has been felt that an insight into psychophysical processes may aid the development of a new feature set, which in turn may be used to explore psychophysical issues. Therefore, this paper discusses a more powerful feature set, capable of higher recognition performance, which uses some psychophysical concepts for the basis of its features.

In Section 3, the task of labelling single pixels rather than entire regions is considered. A feature sub-set is used which consists of only those features relevant to single pixels, such as pixel colour, or the responses of filters centered around the pixel, e.g. texture measurements. The simple features used for this purpose allow some important conclusions about the role of colour and texture, for this particular task, to be made.

In Section 4, the feature set is extended to allow regions to be classified by using features which include shape, size and context. The recognition accuracy of the new feature set is discussed.

Finally some conclusions are made, and scope for further work discussed.

3 Pixel Classification

In this section, the design of several features for pixel classification is discussed. Here, pixel classification involves training a multi-layer perceptron to label each pixel in the image as belonging to one of n classes, using only local information, such as the colour of the pixel and the response of filters centered around the pixel. Obviously, at the pixel level, more complex features such as region shape and size are not applicable.

It seems unreasonable to expect a system based only on pixel information to perform accurate, multi-label classification, so here we shall be interested in only four separate labels; sky, vegetation, road-like and 'others'. Previous work suggested that these were likely to be useful, generic classes of objects[4,6]. The labels 'sky' and 'vegetation' are self-explanatory, while the label 'road-like' is meant to describe any road surface, road marking or pavement pixels.

The specific features that will be used here are intensity, colour and texture. The luminance of a pixel at position (x, y), is defined as:

$$C_{lum} = \frac{3R_{(x,y)} + 6G_{(x,y)} + B_{(x,y)}}{10} \quad (1)$$

where $R_{(x,y)}$, $G_{(x,y)}$ and $B_{(x,y)}$ are the colours of the pixel, scaled in the range $[0, 1]$. The coefficients approximate the contribution of the three colour separations to that of luminance. To achieve a 3D, psychophysically plausible definition of colour[13,14], we also use:

$$C_{r-g} = \frac{R_{(x,y)} - G_{(x,y)} + 1}{2} \tag{2}$$

and

$$C_{y-b} = \frac{R_{(x,y)} + G_{(x,y)} - 2B_{(x,y)} + 2}{4} \tag{3}$$

These are the opponent red/green and yellow/blue colour-difference signals.

We shall also use a psychophysically plausible model of texture, based on Gabor filters. Gabor filters have become widely used for texture segmentation, since they appear to model some aspects of the human visual system[15,16] and have several attractive computational properties, such as optimal localisation in both the spatial domain, $f(x, y)$ and the frequency domain, $F(u, v)$[17]. A set of Gabor filters may be designed to sample the entire frequency domain of an image, by varying the shape, bandwidth, centre frequency and orientation parameters.

In the spatial domain, a Gabor filter takes the form of a complex sinusoidal grating oriented in a particular direction, modulated by a two-dimensional Gaussian. The convolution filter located at (x_0, y_0) with centre frequency, ω_0, orientation with respect to the x-axis, θ_0, and scales of the Gaussian's major and minor axes, α, β, is defined by[18]:

$$g_{(x,y)} = e^{-\pi\left[(x-x_0)^2/\alpha^2 + (y-y_0)^2/\beta^2\right]} \; e^{-2\pi i[u_0(x-x_0)+v_0(y-y_0)]} \tag{4}$$

This filter has a modulation of (u_0, v_0) such that $\omega_0 = \sqrt{u_0^2 + v_0^2}$ and the orientation of the filter $\theta_0 = \tan^{-1}(v_0/u_0)$.

Thus each Gabor filter is tuned to detect only a specific local sinusoidal pattern of frequency ω_0, orientated at an angle of θ_0 in the image plane. Each of these filters may be more conveniently represented in the frequency domain as a Gaussian bandpass filter, centered a distance ω_0 from the origin, with its minor axis at an angle θ_0 to the u-axis. A set of such filters, whose ratio of frequency bandwidth to centre frequency is constant, is shown in Figure 1.

In the experiments described here, in order to reduce the number of features used, we shall consider texture only as a function of frequency, and not of orientation. This isotropic measure of texture, for a particular frequency, is the mean response magnitude of all filters at this centre frequency. The texture measure for a particular pixel in the spatial domain is denoted by

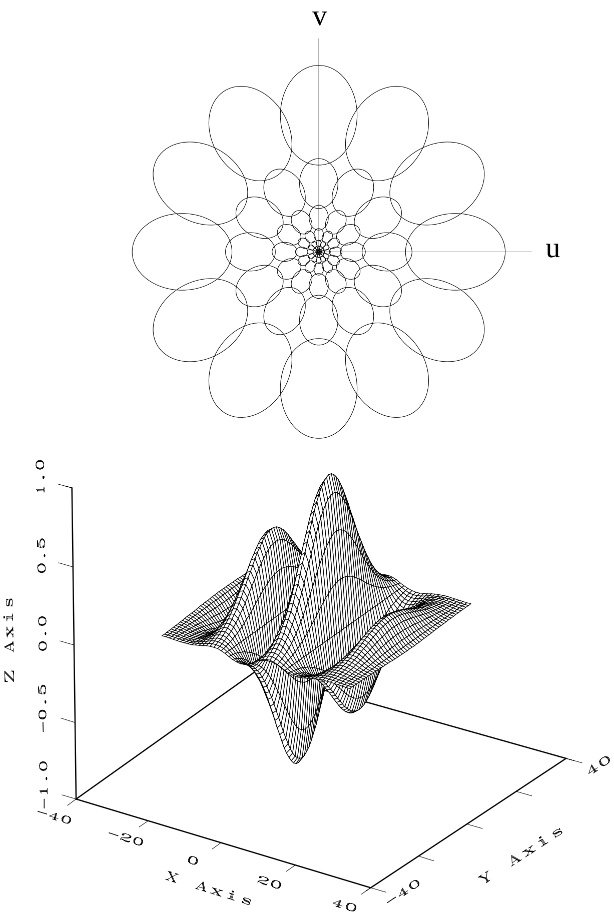

Figure 1: Gabor Filters : (Top) An example set of “daisy-petal” filters which sample the frequency domain. Here six orientations are shown. (Bottom) A spatial filter of 32 cycles/image.

G_n, where n is the frequency of the filter in cycles per image. This value represents the magnitude of the complex filter output.

3.1 Results

The results of the pixel classification experiments are summarised in Table 1. The table shows the features used for the four-label classification and

Feature(s)	**Accuracy**
Intensity	73.4%
Colour	85.0%
Colour + Texture	87.1%

Table 1: Pixel Classification Results using 4 Labels.

the corresponding accuracy obtained with those features. With only the single intensity feature, C_{lum}, the classification accuracy is 73.4%. With the addition of the two colour difference features, C_{r-g} and C_{y-b}, the accuracy increases significantly to 85.0%. The addition of texture in the form of eight isotropic Gabor Filters, $G_1, G_2, G_4, \ldots, G_{256}$, leads to the best result, where 87.1% of pixels are correctly classified.

An example of a pixel classification for a rural scene is shown in Figure 2. The trees are largely identified correctly despite wide variations in their colour and average intensity. Likewise, the road, sky and grass are all correctly identified. The classes chosen correspond to useful clusters of the feature data.

The performance of the pixel classification method has been thoroughly evaluated using approximately 200 images from the database. The results are very encouraging, but, as is already known, region information could be used to increase recognition performance. In the next section, several extensions are considered which combine the pixel classification information with region information, to enable more robust region-based classification.

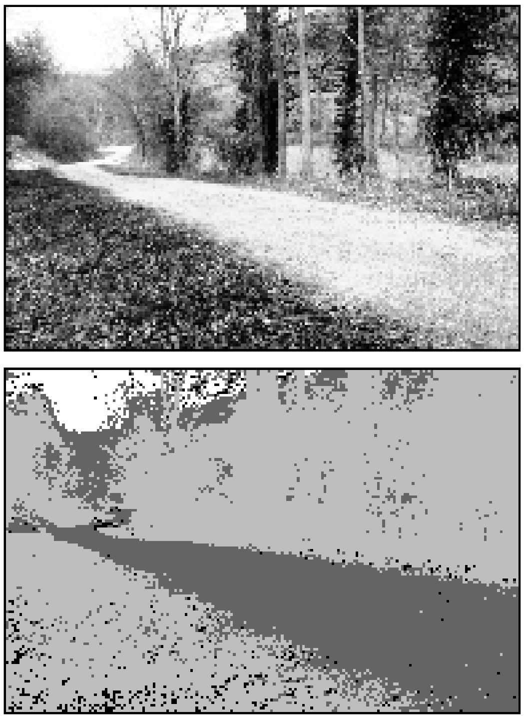

Figure 2: A Labelled Rural Scene. (Top) Monochrome representation of the scene. (Bottom) Labelled Rural Scene. Every pixel is classified according to its intensity, colour and Gabor texture response. White = Sky, Light Grey = Vegetation, Dark Grey = Road, Black = Other.

4 Region-Based Classification

4.1 Segmentation

A K-means method is used to provide the closed-region segmentation required for region analysis. This segmentation approach has previously been shown to be the optimal method out of several segmentation methods for this task[5]. A segmented image is shown in Figure 3. This technique tends

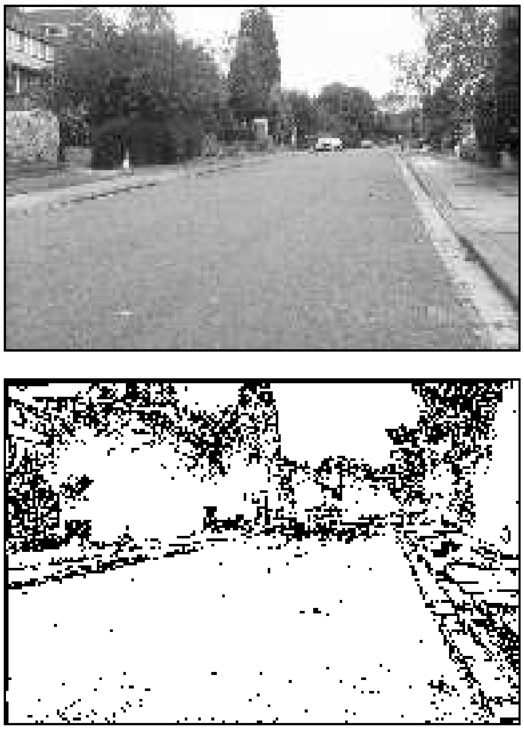

Figure 3: Segmentation of an image using K-means segmentation. (Top) Original monochrome image. (Bottom) Segmentation using K = 4.

to provide a slight over-segmentation of the image, and since it has only one parameter which needs to be set (the value of K itself), is easily tunable to provide the optimal segmentation for a large group of images. For the scenes in the Bristol Database, this generally results in around 100 – 500 regions per image being generated, each of which is then classified.

4.2 Region-Based Features

For region-based classification, the intensity, colour and texture features discussed above have been used (averaged over the entire region), with the addition of the size, position, shape and context of a region. A significantly more challenging task has been attempted than the four-label problem discussed above. Here classification into 11 classes is attempted, which allows direct comparison with the results in [3-5]. This section will discuss the full set of 28 features for use in solving this scene understanding task.

4.2.1 Contextual Features

That regions can be classified more accurately by reference to their surrounds is intuitively obvious. For instance, if an ambiguous object is surrounded by road, it is more likely to be a car. However, it is difficult to make these decisions automatically if contextual information such as the labels of adjacent regions is not available at the time of feature extraction. However, useful information concerning the identities of neighbouring regions is available from the pixel classification, as discussed above. The features used here to provide contextual information measure the frequency of occurrence of each of the four labels used in Section 3, in a neighbourhood of a region, R. In the pixel classification, a particular pixel j was assigned a probability $l_j(i)$ of belonging to the i^{th} class of object.

We define a contextual feature, L_i, as a measure of confidence that the region is surrounded locally by pixels labelled as belonging to class i. If N is the number of pixels in the image, I, and d_j is the distance from a pixel j to the centroid of the region R, then the definition of our context measures, $L_1 \dots L_4$, is given by:

$$L_i = \sum_{j \in I \setminus R}^{N} \frac{l_j(i)}{d_j^2} \tag{5}$$

Thus, $L_1 \dots L_4$ increase with the proportions of the neighbouring area assigned their respective labels, "sky", "vegetation", "road" and "others". The distance term in Equation (5) ensures that the dominant influence on these feature values is from the labels of nearby pixels, and not those far away.

4.2.2 Shape

Here, we use a shape representation which is invariant to planar affine transformations, such as translation, rotation and scaling. Every region is rotated so that its principle axis of inertia is horizontal. The shape of region i is then described as a one-dimensional vector of radii, $\mathbf{x}_i$:

$$\mathbf{x}_i = \{r_1 \dots r_{N_s}\} \tag{6}$$

where N_s is the number of radii, as shown in Figure 4. These radii are

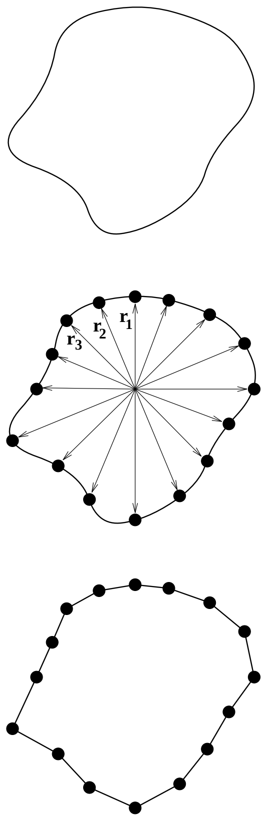

Figure 4: Shape Description : (Top) A typical region boundary. (Middle) Radii from the shapes centroid to its boundary. (Bottom) Polygon approximation of the original shape.

formed around the centroid of the shape, and are each separated from their neighbours by an angle of $(2\pi/N_s)$ radians[19,20]. This vector is normalised so that the longest radius is of length 1. By using the centroid of the region and the normalisation of the radii, this descriptor is invariant to translation, scaling and rotation in the image plane, and approximates the shape as an N_s sided polygon.

To adequately model most regions, however, N_s must be large, which contravenes our desire for a compact feature set. Therefore a method for reducing the amount of data used to model a regions shape must be found. Here we have used principal component analysis, which has recently been successfully used for flexible template matching [21,22].

Principal components analysis allows a transformation from the $\mathbf{x}_i$ coordinate system to another, $\mathbf{y}_i$ in such a way that the new variables are uncorrelated. The new variables are sorted in such a way, however, that $\mathbf{y}_1$ accounts for the largest amount of variance of the original data, $\mathbf{y}_2$ accounts for the next highest percentage of variance and so on. A large training set of shapes is used in this analysis, so that common modes of variation will be available. Relatively few of the new variables $\mathbf{y}_i$ are capable of accurately approximating the shape. The shape features used here are simply the first eight values of the $\mathbf{y}_i$'s.

This is equivalent to saying that any shape may be approximated by a linear combination of the first eight principal modes of variation and the mean shape. The first eight modes of the shapes found in the segmented images of the database are shown in Figure 5. These are shown at the $+2\sigma$ and -2σ level for each mode, as applied to the mean shape.

4.3 Results

The full set of 28 features used for region classification is summarised in Table 2. Along with the features already discussed above, these include the size and position of a region, as well as its orientation. The orientation is expressed as the sine and cosine of the angle to avoid problems of discontinuity in the representation when training the neural network [4].

Using this feature set, a neural network with the 28 inputs and 11 label outputs, has been trained and optimised. The optimal network architecture, was found to have 24 nodes in the hidden layer. This network was used for classifying unseen data in a test set of over 3000 regions. A high overall accuracy was obtained, with 82.9% of the regions and 91.1% of the image area being correctly labelled. The difference is due to large regions being more reliably classified than small regions. These results have a standard

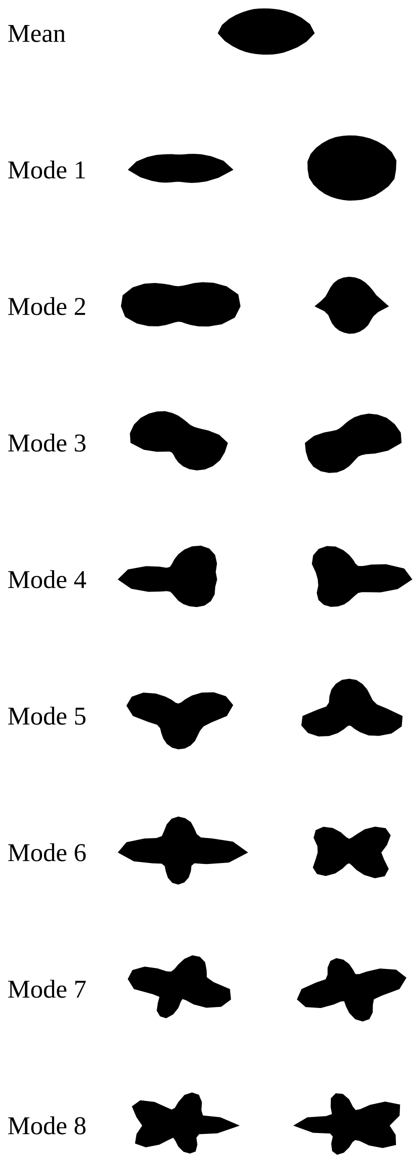

Figure 5: The mean and first eight principal modes of variation for the shapes generated by segmenting images in the database. The $+2\sigma$ (left) and -2σ (right) for each mode is shown as applied to the mean shape.

No.	**Features(s)**
1-3	Colour
	$C_{lum}, C_{r-g}, C_{y-b}$
4-11	Texture : Isotropic Gabor
	$G_1, G_2, G_4, \ldots, G_{256}$
12-19	Shape : Principal Modes
	$\mathbf{y}_1, \ldots, \mathbf{y}_8$
20-23	Label Context
	$L_1, L_2, \ldots, L_4$
24	Size
25-26	Rotation
27-28	Centroid

Table 2: Summary of Full Feature Set used for Classification.

error on the classification accuracy of approximately 1%.

The classification results for each label class are shown in a confusion matrix in Table 3. The rows of the table correspond to the true labels for

	Label	1	2	3	4	5	6	7	8	9	10	11
1.	Sky	**104**	9	0	0	0	9	0	0	0	0	0
2.	Vegetation	6	**1728**	0	23	30	35	13	0	2	0	9
3.	Road Marking	0	0	**24**	2	2	0	0	0	0	0	0
4.	Road	0	31	7	**510**	65	2	4	0	0	0	5
5.	Pavement	0	45	4	51	**213**	4	5	0	0	0	4
6.	Building	3	89	0	0	2	**406**	5	0	0	0	18
7.	Fence/Wall	0	35	0	1	4	21	**46**	0	0	0	11
8.	Road Sign	0	6	0	0	0	12	0	**0**	0	0	3
9.	Signs/Poles	0	4	0	2	1	5	1	0	**2**	0	2
10.	Shadow	0	2	0	1	1	3	3	0	0	**0**	6
11.	Mobile Objects	0	7	0	0	4	18	6	0	0	0	**75**

Table 3: Confusion Matrix for Region Classification for 11 Labels.

each of the 3751 regions, i.e. 1728 regions are correctly labelled as 'vegetation', 6 vegetation regions are incorrectly labelled as 'sky' and only 9 'sky' regions out of 111 are mis-labelled. Mistakes include the misclassification of road as pavement (and vice versa). Apart from the labels 'shadow' and 'road sign', for which there are very few examples available, all the labels are classified very accurately. The label 'sky' for instance is recognised correctly in over 93% of cases, and 'building', a conceptually complex class, is recognised correctly in over 77% of cases.

The eleven classes used here are a super-set of the 93 labels available in the ground-truth and are described more fully in previous work [6]. For instance, the label 'Mobile Object' is a combination of many labels such as 'Car-Body', 'Car-Window', 'Van-Body' etc. The large majority of the objects in the database can be grouped into these 11 classes. The remainder, together with objects too small to be classified, account for less than 3% of pixels. In training and testing the classifier we have only used regions corresponding to the 11 classes. In a system having a less general set of labels it would be necessary to include a further class for unknown objects, but this is not required here.

Figure 6 shows a test scene, and some of the classifications applied to it. It can be seen that the areas labelled as vegetation are largely correct, and that the 'difficult' regions of pavement and road have been correctly distinguished.

A major objective of this research is to allow retrieval of image data on the basis of content. For instance, to find rural rather than urban scenes, or to retrieve images which contain particular object classes. Figure 7 gives examples of images which have been automatically retrieved from the database on the basis of content. For this case the query was to find images containing more than one vehicle. The sensitivity of the system is clearly demonstrated by the fact that in two of the images, small regions corresponding to cars in the distance have been successfully identified.

5 Conclusions

This paper has shown how fully automatic interpretation of outdoor scenes can be achieved using a powerful set of region descriptors and an accuracy of 82.9% by region frequency, and 91.1% by image area has been achieved.

Given such high accuracies one can use this information as part of the data insertion and indexing process of an image database. Such knowledge of the image content can be used to sequence the database, or to partition it into groups of related images. More importantly, image content may be used to index specific images in the database for the purposes of rapid retrieval. In the example given in Figure 7, the segmented image and its classification labels have been preserved along with the image. This allows recovery of specific regions of images. For many purposes, however, it would suffice to index each image with the number of instances of each object class. Such an index would provide a mapping between the complete set of labels generated

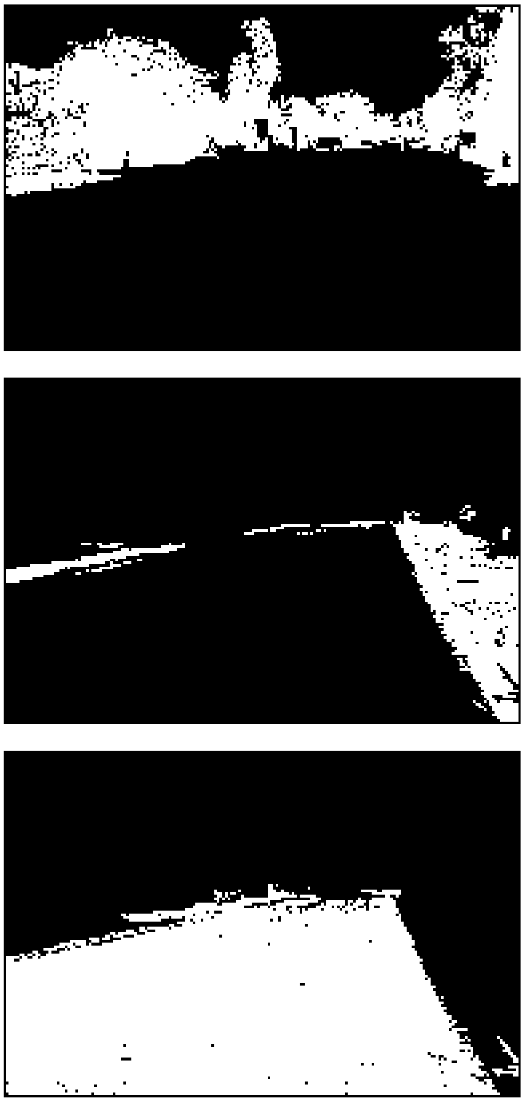

Figure 6: Test Scene. (Top) Those regions classified as 'vegetation'. (Middle) Those regions classified as 'pavement'. (Bottom) Those regions classified as 'Road'.

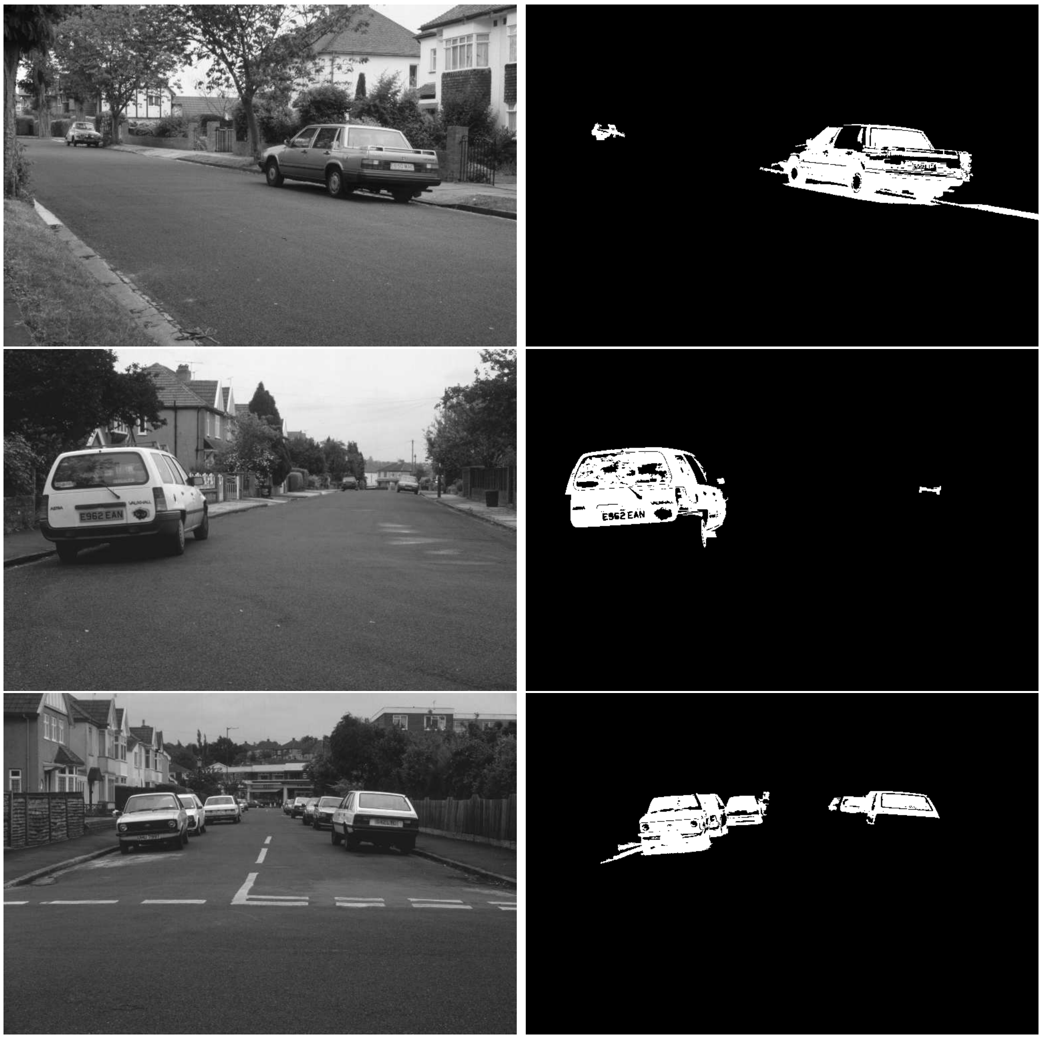

Figure 7: Some Images from the Bristol Image Database automatically segmented and classified as having multiple non-adjacent 'Mobile Object' regions. (Left) Monochrome version of original image. (Right) Regions from images classified as 'Mobile Object'.

by the neural network, into a set of unique image numbers. This can then be used to retrieve images containing specific objects, or can be utilised with high-level rules such as: rural scenes may contain cars and roads but have few buildings and are dominated by vegetation.

As the quantity of image data increases, it becomes more important to have a mechanism for classifying images as fully as possible prior to insertion in a database. Many of the current approaches to query by content depend upon global properties of an image, thereby allowing the overall similarity of scenes to be measured. These do not deal well with searches for specific objects in otherwise dissimilar scenes. The technique described here has been used to classify objects in outdoor scenes but would, however, recognise many of the objects present in completely different scenarios. A future goal of our research is to produce separate classifiers, each trained to recognise objects from specific scenarios, to allow a wider range of images to be categorised by content.

Acknowledgements

This project is supported by EPSRC (grant GR/J36662). We thank British Aerospace PLC, Sowerby Research Centre, Bristol for their support with the development of the database. Additional thanks are due to Phil Greenway and Andy Wright for useful discussions. We would like to thank the anonymous referees for their constructive comments.

References

[1] Flickner, M. *et al*, Query by Image and Video Content: The QBIC System. *IEEE Computer* 9, 23–32, Sept 1995.

[2] Pentland, A. and Picard, R. and Sclaroff, S., Photobook: Tools for Content-Based Manipulation of Image Databases. *SPIE Conf. on Storage and Retrieval of Image and Video Databases II*, 34–47, Feb 6-10, 1994.

[3] Mackeown, W.P.J. *A Labelled Image Database and its Application to Outdoor Scene Analysis.* PhD thesis, University of Bristol, 1994.

[4] Mackeown, W.P.J., Greenway, P., Thomas, B.T., and Wright, W.A. Road Recognition with a Neural Network. *Engineering Applications of Artificial Intelligence*, 7(2):169–176, 1994.

[5] Campbell, N.W., Mackeown, W.J., Thomas, B.T., Troscianko, T. Automatic Interpretation of Outdoor Scenes. *British Machine Vision Conference*, Birmingham, UK, Sep 1995.

[6] Mackeown, W.P.J., Greenway, P., Thomas, B.T., and Wright, W.A., Contextual Image Labelling with a Neural Network. *IEE Procs. Vision, Speech and Signal Processing.* 4:238–244, 1994.

[7] Brooks, R.A. Model-based 3D Interpretation of 2D Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):140–150, 1983.

[8] Draper, B.A., Collins, R.T., Brolio, J., Hanson, A.R., and Riseman, E.M. The Schema System. *The International Journal of Computer Vision*, 2:209–250, 1989.

[9] Ohta, Y. *A Region-Oriented Image-Analysis System by Computer.* PhD thesis, Kyoto University, March 1985.

[10] Wright, W.A. Image Labelling with a Neural Network. *In Procs.* 5^{th} *Alvey Vision Conference* Reading, UK, 227–232, 1989

[11] Haddon, J.F., Boyce J.F. Texture Classification of Segmented Regions of FLIR Images using Neural Networks. 1^{st} *International Conference on Image Processing*, Austin, Texas, USA, November 1994.

[12] Haddon, J.F., Boyce, J.F., Protheroe, S. and Hesketh, S. Neural Networks for the Texture Classification of Segmented Regions in Forward Looking Infrared Images. 4^{th} *British Machine Vision Conference*, Guildford, UK, 197–206, 1993.

[13] Glassner, A.S. *Principles of Digital Image Synthesis.* Morgan Kafmann, 1995.

[14] De Valois, R.L., De Valois, K.K. A Multi-Stage Color Model. *Vision Research*, 33(8):1053-1065, May 1993.

[15] Malik, J., Perona, P. Preattentive Texture Discrimination with Early Vision Mechanisms. *J. Opt. Soc. Am.*, A7:923-932, 1990.

[16] Griffiths, E., Troscianko, T. Can Human Texture Discrimination be Mimicked by a Computer Model using Local Fourier Analysis ? *Spatial Vision*, 6:149-157, 1992.

[17] Daugman, J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am.*, 2(7):1160-1169, 1985.

[18] Bovik, A.C., Clark, M., Geisler, W.S. Multichannel Texture Analysis Using Localized Spatial Filters. *T-PAMI*, 12(1):55-73, Jan 1990.

[19] Kashyap, R.L., Chellappa, R. Stochastic Models for Closed Boundary Analysis : Representation and Reconstruction. *IEEE Transactions on Information Theory*, 27(5):627-637, Sept. 1981.

[20] Gupta, L. and Srinath, M.D., Contour sequence moments for the classification of closed planar shapes. *Pattern Recognition*. 20(3):267-272, 1987.

[21] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J. Training Models of Shape from Sets of Examples. *British Machine Vision Conference*, 9–18, Leeds, UK, 1992.

[22] Cootes, T.F., Taylor, C.J. Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5):403–409, June 1995.

About the Author - Neill Campbell received his B.Eng and Ph.D. degrees from the Bristol University, UK in 1990 and 1995 respectively. His research interests include computer vision and computer graphics. He lectures in the Department of Computer Science at Bristol University.

About the Author - William Mackeown received a B.Eng. in Computer Systems Engineering with 1st Class Honours from Bristol University in 1988, and a Ph.D. in computer vision in 1994 following the award of an academic scholarship. He is the designer of the Sowerby (Bristol) Image Database. He was a junior research fellow at Bristol University before joining King's College London in June 1995 as a lecturer in computer science. His main research interests are in computer vision and pattern recognition. He is a member of the British Computer Society and the British Machine Vision Association.

About the Author - Barry Thomas received his B.Sc. , M.Sc. and Ph.D. degrees in physics from Imperial College, London, UK 1970, 1971 and 1975 respectively. He spent seven years in the USA at the Jet Propulsion Lab. and Caltech. For the past ten years he has been at Bristol University, where he is now a Professor of Computer Science. His research interests are in image processing, computer vision and neural networks. He is the author of over 60 refereed papers in international journals and conferences.

About the Author - Tom Troscianko obtained a degree in Physics at Manchester University and a PhD in Visual Science at The City University, London. He then worked in the Brain and Perception Laboratory in the Anatomy Department at Bristol University, and was a founder member of the Perceptual Systems Research Centre in the Department of Psychology. He spent periods at IBM UK Scientific Centre, Winchester, and Tubingen University Eye Hospital. His research covers many aspects of vision - using psychophysical, physiological, and computational approaches.