

Head-mounted mobility aid for low vision using scene classification techniques

M R Everingham¹, B T Thomas¹ and T Troscianko²

¹Advanced Computing Research Centre, University of Bristol,
Bristol BS8 1UB, UK

²Perceptual Systems Research Centre, University of Bristol,
Bristol BS8 1UB, UK

¹everingm@acrc.bris.ac.uk

ABSTRACT

This paper describes a new mobility aid for people with severe visual impairments which combines technology from the field of virtual reality with advanced computer vision techniques. A neural-network classifier is used to identify objects in images from a head mounted camera so that scene content specifically important for mobility may be made more visible. Enhanced images are displayed to the user on a head mounted display using a high saturation colour scheme where each type of object has a different colour, resulting in images which are highly visible and easy to interpret. The object classifier achieves a level of accuracy over 90%. Results from a pilot study conducted using people with a range of visual impairments are presented in which performance on a difficult mobility-related task was improved by over 100% using the system.

Keywords: Low Vision, Mobility Aids, Head Mounted Display, Object Recognition, Neural Networks

1. INTRODUCTION

Many people who are registered as blind nevertheless retain some residual vision, and are said to have “low vision”. Examples of conditions resulting in low vision are cataracts, diabetic retinopathy, age-related maculopathy, and retinal detachment. The precise form of visual impairment varies according to the particular medical condition it results from, but often a person with low vision experiences an extreme loss of perception of high spatial frequencies resulting in gross blurring of the visual scene over a significant area of their field of view. The visual impairment is typically such that the person is unable to be mobile without some form of assistance such as a guide dog. In recent years, with the development of head mounted displays and cameras for the virtual reality field, work has attempted to apply these devices to the needs of users with low vision in combination with computer image processing techniques. A common factor in the image processing techniques used is that since they work with no knowledge of the semantic content of an image, they tend to enhance noise and irrelevant detail in a scene equally to important visual information, resulting in cluttered images which are difficult to interpret, and no use has yet been made of the colour capabilities of modern head mounted displays. The Computer Science Department at the University of Bristol, in conjunction with the Psychology Department and Bristol Eye Hospital, is developing a mobility aid for low vision which uses artificial intelligence techniques to recognise objects in a scene so that visual enhancement may be carried out taking into account which components of a scene are important and which may be considered irrelevant detail or noise. This paper describes how such content-driven enhancement may be carried out, and presents results of a pilot study in which the concept was evaluated by a number of people with a variety of low vision conditions.

1.1. Structure of Paper

The outline of this paper is as follows. Section 2 reviews previous work in applying technologies from virtual reality and computer vision techniques to the needs of people with low vision. Section 3 describes the novel image enhancement approach used by our mobility aid and discusses its advantages over conventional approaches. Section 4 describes in detail the scene classification technique used. Section 5 presents results of the performance of the classification technique and the results of a pilot study evaluating the concept of the new mobility aid. Finally, in section 6 some conclusions are made and scope for further work discussed.

2. BACKGROUND

2.1. Use of Virtual Reality Technologies in Visual Impairment

As technology for virtual reality applications has advanced, several researchers have investigated the application of this technology to the needs of people with impairments affecting mobility, in particular the use of head-mounted displays.

Prothero (Prothero, 1993) used a head-up display, in which virtual imagery is overlaid on the real scene, to provide visual cues which improved the mobility of patients with akinesia due to Parkinson's disease. Massof and Rickman (Massof and Rickman, 1992) produced a system combining binocular monochrome LCD displays with three video cameras in a head mounted unit which provides magnification and contrast enhancement. Development of the device was continued in conjunction with NASA (NASA, 1993) and it is now commercially available as the Low Vision Enhancement System ("ELVIS"). Peli (Peli, 1995) implemented a device using a binary-output head mounted display to display binary images with adjustable brightness and contrast enhancement of high 1-dimensional spatial frequencies, and a 2x zoom facility. Goodrich and Zwern (Goodrich and Zwern, 1995) used a commercial head mounted display to provide variable magnification and contrast adjustment of images.

2.2. Application of Computer Vision to Visual Impairment

Traditional aids for low vision such as those described above have utilised two image enhancement techniques – magnification and contrast enhancement. This is the same approach which has been used by conventional desktop devices to make printed material accessible to people with visual impairments. The more advanced devices such as that by Peli (Peli, 1991) have used variable contrast amplification in specific spatial frequency bands rather than a uniform contrast amplification. This type of enhancement can be achieved using techniques from the field of image processing. The aim here is to make important image features, typically edges, more easily visible. Such operations are easy to achieve using video electronics which can currently be implemented in real-time, but this approach has fundamental limitations. As an example, Figure 1 shows an image of a typical urban scene, and the same image enhanced by the adaptive filtering technique described in Peli (1991), which amplifies the contrast of high spatial frequencies using an adaptive method to adjust to varying low frequency luminance across an image. As one can see, the technique does something to improve the visibility of important edges in the scene such as the outlines of pavement and cars, but this is at the expense of introducing much noise by the amplification of textural properties of the image which are not necessary for interpreting the scene, for example the cloud texture in the sky, and the noisy road appearance. This occurs because the technique, in common with all others using basic image processing, is unable to differentiate between image content which is semantically important or merely noise or textural detail. Such techniques are in general unable to perform specific enhancement of important high-level scene properties, for example dangerous objects such as vehicles or important visual properties such as the boundary between road and pavement.



Figure 1. (a) Original image, (b) Enhanced by Peli method

Peli has recently proposed a new "wide-band" technique (Peli, 1998) which detects "edge" and "bar" features and superimposes these as thickened white lines on the original image using a head-up display. Quantitative results have not been presented, but the technique is likely to be limited by the same factors affecting other techniques based on image processing.

Given the limitations of image enhancement using simple image processing operations, some recent work has concentrated on the task of mobility, and used more advanced computer vision techniques to recognise obstacles in scenes explicitly rather than to simply enhance their visibility for the user. Such techniques are of course applicable to users with no residual vision. Molton et al (Molton et al, 1998) are developing a portable system using stereo vision to find obstacles in the path of a mobile user, though no discussion is made of how this information is to be presented to the user. Snaith et al (Snaith et al, 1998) describe work on a system using computer vision techniques to facilitate centre-path travel and recognise doorways.

3. MOBILITY AID USING SCENE CLASSIFICATION

3.1. System Structure

Figure 2 shows the proposed structure of our mobility aid. Images are taken from a head mounted camera, digitised and enhanced using a processing unit, and output to a head mounted video display unit. Due to the nature of our image enhancement technique we need only a single video camera, and a display unit in which the same image is presented to both eyes. We anticipate that in a final implementation the whole system could be implemented in a single head

mounted unit. In current prototypes a portable computer is being used with a conventional low-cost video camera and colour VR display unit.

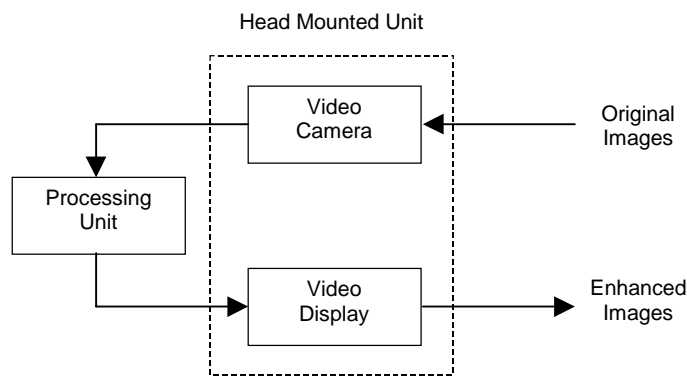


Figure 2. Structure of Mobility Aid

3.2. Image Enhancement Technique

The approach being developed at Bristol differs from other techniques in that it attempts to identify the semantic content of an image such that images can be enhanced in a content-driven manner, taking into account which features of an image are important, and which may be disregarded as noise or unnecessary detail. In our approach, scene classification forms the basis for image enhancement. In our enhanced images, each object in the scene is coloured using a solid high saturation colour corresponding to the type of object, for example the road is black, pavement white, vehicles bright yellow and so on. In current work we use a scheme whereby all objects are classified as one of a set of eight object classes, so for example the “obstacle” class contains such things as bounding walls, lamp-posts, pillar-boxes etc. We judge that it is easy to distinguish eight separate colours when they are high saturation and high contrast, and that the information present in the scene using this set of object classes is sufficient for mobility in a typical urban environment. Figure 3 shows as an example the image from Figure 1 and its appearance when enhanced by our method. Due to lack of colour printing facilities, figures have been reproduced in monochrome here and it should be noted therefore that the contrast between object types in these images is lower than in the on-screen colour images.



Figure 3. (a) Original image, (b) Enhanced by Bristol method

There are many advantages to using a scheme such as this which uses scene classification as the basis for image enhancement. In our enhanced images, the colour of an object is independent of its original visual properties which may make the object less visible, so that here for example cars are coloured bright yellow regardless of their actual colour which may be hard to see. Irrelevant detail such as textural properties of an object is also removed by classification, which gives the images a very simple and uncluttered appearance which is easy to interpret. The form of output is easy to customise to a particular user’s requirements; in this case we have used a predefined set of high saturation colours, but these colours may be customised by a user to improve visibility according to his or her particular visual impairment. Other modifications to the visual presentation scheme are equally easy to make, for example overlaying of edges on the image, or causing particular object types such as obstacles to flash when they are large in the scene to attract attention to them. Additionally, because our system gathers semantic information about the content of the scene, it is not even necessary to use a visual output medium – the information could instead or additionally be presented using existing touch, sound or speech output devices.

We have primarily been investigating the use of the saturated colour output scheme, and Figure 4 demonstrates the effectiveness of this scheme. Here we see the image from Figure 1 in original form, enhanced by a Peli method (Peli, 1991), and by our method. In this case, all three images have been blurred by an equal amount in order to simulate a visual impairment causing reduced perception of high spatial frequencies.



Figure 4. (a) Blurred image, (b) Peli method, (c) Bristol method

Even without colour reproduction, Figure 4 clearly demonstrates the effectiveness of our method. Arguably the Peli method has reduced the visibility of the scene by reduction in the overall contrast of the image, whereas in the image enhanced by our method one can clearly see the boundary between pavement and road, which is invisible in the other two images, and the three vehicles which appear as bright yellow blobs. In the on-screen image it is also possible to see by colour the two lamp-posts in the scene, which are again invisible in the other images due to their poor visibility in the original scene.

3.2.1. Use of Virtual Reality to Train Users

We anticipate that the training time for a potential user of our mobility aid would be minimal because of the intuitive nature of the output, but certainly some training would be required to become acquainted with the presentation scheme used by the device, and also for the user to select an optimal choice of colours to represent each object type. Virtual reality offers a safe and cost-effective method for achieving this. Conventional street models developed for other applications may be very easily adapted for use in this application by labelling objects in the virtual world with their classification and utilising a flat shading model in which light sources are not considered in the rendering of the virtual environment.

3.2.2. Accessibility of Virtual Environments

Another possible application of our technique in addition to real-world image enhancement is in making virtual environments which simulate real world environments accessible to users with visual impairments. In the most general case, given sufficient robustness in the scene classification technique the same techniques can be applied to rendered images of virtual environments as to real world images, and it is not difficult to adjust the set of object classes used by the system to suit a given environment. Alternatively objects in a virtual world may simply be labelled with their object classification, and this information can then be used to colour objects by their class for use by people with low vision. Access to such environments could then be easily given to both sighted and visually-impaired users simply by a change in rendering technique.

4. SCENE CLASSIFICATION TECHNIQUE

4.1. System Architecture

Figure 5 shows a block diagram of the image enhancement system of our mobility aid, which uses a neural network to perform classification, building on work in outdoor scene classification at Bristol (Campbell et al, 1997). Input to the system comes from a head mounted camera, and output is displayed on a head mounted display. The *image segmentation* stage segments an image into a number of regions which are deemed to correspond to a single object or object part. The *feature extraction* stage takes the pixels of a region and calculates a set of numeric features or “feature vector”, which describes the visual properties of the region and its context in the image. The *neural net classifier* takes the feature vector of a region as input and its output corresponds to an object class to which the region belongs, for example road, vehicle etc. The final *image labeller* stage produces the coloured image in which the pixels of each region are coloured with the high saturation colour corresponding to the determined object class for the region. During a separate training phase, a *training algorithm* uses previously collected *ground truth* data in the form of hand-classified images to train the neural network. Each of these stages of processing is now discussed.

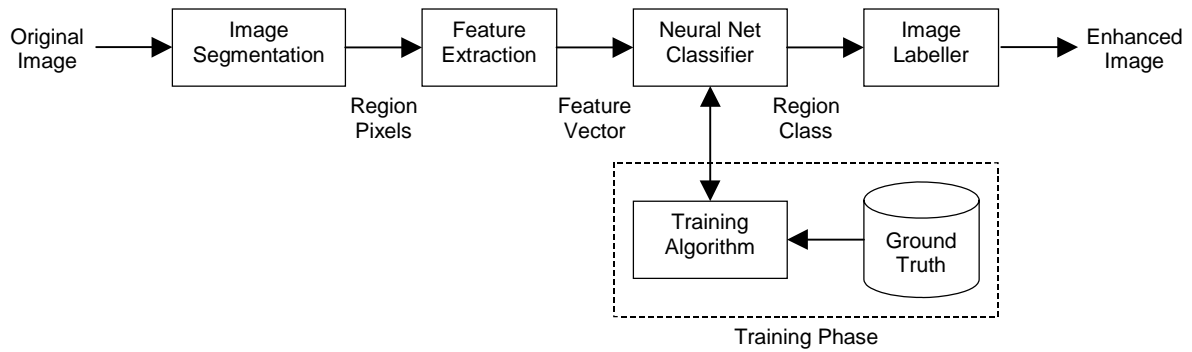


Figure 5. Architecture of Image Enhancement System

4.2. Segmentation

The aim of image segmentation is in general ill-defined, but in this context we will characterise it as being to divide an image into a set of connected regions such that each region corresponds to a single object or object part. Figure 6a shows an edge-map of an approximately ideal segmentation of the original image from Figure 1 performed by a skilled human operator. Automatic segmentation of general outdoor scenes remains an unsolved problem with much research being carried out in this field, see (Pal and Pal, 1993) for a review. Most techniques regard a region as being homogenous with respect to some visual properties such as colour or texture, and we have tried many and diverse techniques using a variety of homogeneity measures. We find that none give near-perfect results, but that good results can be achieved using a surprisingly simple technique, based only on homogeneity of grey-level luminance across a region.

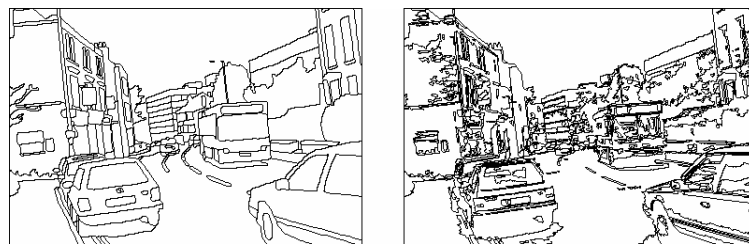


Figure 6. (a) Ideal Segmentation, (b) Automatic Segmentation

Figure 6b shows the quality of segmentation obtainable using this technique. Segmentation is performed using the well-known K-means clustering algorithm which is an iterative clustering technique which converges to a local minimum (Selim and Ismail, 1984). We define that regions are of homogenous grey-level, and cluster all pixels in the image using grey-level alone. Individual regions are then extracted using a connected-components algorithm such that all connected pixels assigned to the same cluster belong to the same region. The K-means algorithm requires only a single parameter, the number of clusters, and this was set to 5 for all images in our work, on an empirical basis. Clustering noise is removed by merging regions with area less than 16 pixels (determined empirically) to a neighbouring region. The result is typically a moderate over-segmentation of the scene into 100-400 regions.

4.3. Feature Extraction

In order to classify regions into an object class using a neural network classifier, the raw image information in the form of pixels of a region must be reduced to a small set of numeric *features* which describe the region sufficiently well for it to be classified. This is achieved by the feature extraction stage. We have used features which are intuitively appealing, corresponding to visual concepts which we as human beings typically use to describe objects, and where possible we have used features which have some psychophysical basis. The main features we extract in this manner are Colour, Shape, and Texture, and these are described here.

4.3.1. Colour

Colour is intuitively useful for recognising objects in a scene, for example grass is green, and roads are grey. Colour can also be used in cases where it might be expected to be less useful, for example cars are produced in many colours, but typically they have in common that the colour is of high saturation. Many colour spaces exist (Foley et al, 1990) which are candidates for generating colour features. We have experimented with a number of these including RGB, YUV, HSV and $L^*u^*v^*$, and found that colour spaces which separate the luminance and chromaticity components of a

colour are more successful. We found the CIE L*u*v* colour space (Wyszecki, 1982) to be most successful overall, and this is an appealing result as it is a space which has a good perceptual basis, in the sense that colours which are equidistant in the space are approximately equidistant in terms of perceived difference. In order to calculate three colour features for a region, the mean RGB value of the pixels in the region is taken from the original image, and converted to L*u*v*, with the L, u and v values being used as features.

4.3.2. Colour Constancy

When using colour to recognise objects, we should be careful to note that the colour of pixels in an image *do not* correspond to the actual colour of an object, but it's perceived colour under a general unknown illuminant, for example daylight. In addition, different image capture systems produce images with different colour properties. If we want our classifier to be robust to such changes as different weather, lighting or cameras, we must attempt to recover the *actual* colour of an object, rather than its colour under a particular illuminant and using a particular image capture system. The human visual system exhibits this invariance (Wyszecki and Stiles, 1982) known as colour constancy. We attempt to achieve this property by using a colour correction technique to normalise images. The key is to approximate the colour skew caused by illumination/capture properties and remove this. Our technique is based on the "grey world" model (Buchsbbaum, 1980) which determines colour skew by assuming that the mean reflectance in a scene is achromatic i.e. has equal RGB components. Given mean RGB and grey-level components of an image, each of the RGB components is scaled to make the individual means equal to the overall grey-level mean using a tri-diagonal matrix:-

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = \begin{pmatrix} I_\mu / R_\mu & 0 & 0 \\ 0 & I_\mu / G_\mu & 0 \\ 0 & 0 & I_\mu / B_\mu \end{pmatrix} \begin{pmatrix} R_u \\ G_u \\ B_u \end{pmatrix}$$

where R_c, G_c, B_c are the corrected RGB components of a pixel, R_u, G_u, B_u are the uncorrected RGB components, R_μ, G_μ, B_μ are the mean RGB components across the image, and I_μ is the mean grey-level across the image, equal to $(R_\mu + G_\mu + B_\mu)/3$. In addition to correcting colour skew in this manner we also normalise mean brightness and contrast by scaling grey-levels to full range and moving the mean to half range. By disregarding the top and bottom 5% of the grey-level histogram in this process, an approximate invariance to varying daylight conditions is obtained.

4.3.3. Shape

Shape should intuitively be a very powerful feature for discriminating object classes. We use here a shape representation which is invariant to translation, rotation, and scaling. We consider such invariance important for obtaining robustness of classification. The representation we use is a Fourier shape descriptor, which considers the outer boundary of a region as a periodic sequence of two dimensional points, where transformation to the frequency domain allows the significant properties of the shape to be characterised in a small number of values. The particular formulation of Fourier descriptor used is that of Shridhar and Badreldin (Shridhar and Badreldin, 1984), who achieved very good results for hand-written character recognition using the descriptor. Formulation of the shape features for a region proceeds in three steps:-

First we use a novel modification of the algorithm which allows the shape descriptors to incorporate information about holes in a region in addition to their outer boundary. This is achieved by cutting the region silhouette using a minimum length set of cuts to link internal holes to another hole or to the area outside the region so as to form a *single* boundary which traverses both the original outer boundary and the boundaries of the holes in the region. The boundary of the region is then traced using a conventional boundary following algorithm to give an ordered, closed list of 2-D co-ordinates. Secondly, the two co-ordinate sequences corresponding to the x and y co-ordinates of the boundary points are separately transformed to the frequency domain using two 1-D Fast Fourier Transforms (FFT). Finally, the frequency domain components are transformed to give the required invariance:-

$$r(n) = \sqrt{|a(n)|^2 + |b(n)|^2}$$

where $a(n)$ is the FFT of the x co-ordinates and $b(n)$ is the FFT of the y co-ordinates. Discarding $r(0)$, $r(n)$ can be shown (Badreldin et al, 1980) to be invariant to translation and rotation. Dividing by $r(1)$ gives scale invariance:-

$$s(n) = r(n) / r(1)$$

Using a sub-set $s(2..k+2)$ of the k low-frequency coefficients of $s(n)$ gives a good approximation of the shape even when $k \ll L$ where L is the length of the boundary. In our work we have used 10 coefficients as shape features for a region. Figure 7 shows a region from an outdoor image and the approximation of it's outer boundary using 10 and 30 shape coefficients, showing how an increasing number of coefficients gives higher accuracy of shape approximation, particularly for objects with sharp corners.



Figure 7. (a) Cut region silhouette, Boundary reconstruction from (b) 10 coefficients, (c) 30 coefficients

4.3.4. Texture

Texture is also intuitively useful for discriminating between objects, for example brick walls have a highly periodic and directional texture where grass has a higher frequency more random texture. To obtain texture features for a region we use Gabor filters, which are a particular type of filter tuned to respond to a specified range of spatial frequencies and orientations. Gabor filters have been shown to have similar characteristics to simple visual cortical cells (Marcelja, 1980) and therefore sound psychophysical basis. Figure 8a shows an example of the real part of a complex-valued Gabor filter in the spatial domain; the imaginary part is equal but subject to a 90° phase shift. In the spatial domain, the filter has the form of a complex sinusoid modulated with a Gaussian oriented along the direction of the filter, while in the spatial frequency domain it is an orientated Gaussian. Using a complex-valued Gabor filter, the magnitude response is approximately constant given an input sinusoid image with frequency and orientation matching those of the filter. Jain and Farrokhnia (1991), and Dunn and Higgins (1995) provide good overviews of the mathematics of Gabor filters.



Figure 8. (a) Gabor filter in spatial domain (real part), (b) Filter bank in spatial-frequency domain (half peak)

Using a bank of n filters and taking magnitude output gives a vector of n real values characterising the texture content of an image at a particular pixel. We use the scheme of Jain and Farrokhnia (Jain and Farrokhnia, 1991) to determine the properties of our filter bank, in which the centre frequencies of the filters are in a power of two series, with 45° orientation step, and the frequency and orientation bandwidths are set to 1 octave and 45° respectively. Figure 8b shows this partitioning of the spatial-frequency domain, in which the half-peak magnitude responses of the filters are shown, with a total of 16 filters being used; the asymmetry in the spatial frequency domain is caused by the use of complex-valued filters. The magnitude response of a Gabor filter across a region of constant texture can be well characterised as a Rician distribution (Dunn and Higgins, 1995), but as estimation of the parameters of this distribution requires the costly solution of non-linear equations, we instead use a Gaussian distribution, which has been shown to be adequate in most cases (Dunn and Higgins, 1995). Thus to form our region texture features, we take the mean and standard deviation of each Gabor filter output across a region, giving 32 features in all. In practice however, we find that using means alone gives almost as good results, so we use a reduced set of 16 texture features formed by the mean magnitude responses.

4.3.5. Combined Feature Set

Combining colour, shape and texture features with other simple features describing a region's position, size and orientation, we arrive at the complete set of 35 features shown in Table 1.

Table 1. Feature Set

Feature	Description
1	Size (proportion of image)
2-3	Position (scaled x, y co-ordinates)
4-5	Orientation (sin & cosine of angle)
6-8	Colour (mean $L*u*v*$)
9-18	Shape (invariant Fourier coefficients)
19-35	Texture (mean Gabor magnitude)

4.4. Neural Network Classifier

A multi-layer perceptron (see Bishop, 1995) with one hidden layer is used as a classifier, with a 35-n-8 architecture, and with hidden and output units having logistic sigmoid activation functions. The 35 inputs correspond to the feature vector, and each of the 8 outputs indicates membership of a particular object class. A one-of-n coding scheme is used to encode membership of each class, and a winner-takes-all scheme is used to determine the object class chosen by the network. Inputs were normalised to have zero mean and unit variance across the entire training and test data sets. In our work we experimented with a number of hidden units between 5 and 40, finding 23 hidden units to be optimal.

4.4.1. Training Procedure

The neural network classifier is trained using the Scaled Conjugate Gradient method (Møller, 1993). Training data was taken from two sources, primarily the Bristol Image Database (Mackeown, 1994). This is a database of 200 high-quality outdoor suburban and rural scenes which have been hand-classified by a skilled human operator. A second database, the Bristol Blind Mobility Database, containing 10 scenes captured in more challenging urban situations was used in addition. The data was split randomly into 70% training regions and 30% testing regions.

5. RESULTS

5.1. Classifier Performance

Table 2 shows the classification performances achieved by our neural network classifier. Performance is quoted for the two image databases as percentage correct using two metrics – percentage correct by regions, and more meaningfully, percentage correct by area. In addition, the maximum *a priori* (MAP) probabilities of correct classification are shown with the corresponding most likely class. This information represents the maximum performance possible by a Bayesian classifier with no *a posteriori* information i.e. no feature data.

Table 2. Neural Network Classifier Performance

Database	% by Region	% by Area	MAP % by Region
Bristol Image	79.1	92.5	31.8 vegetation
Blind Mobility	70.1	80.4	36.7 vehicle

We see that overall the performance of the classifier is very high, with up to 92.5% of image area correctly classified using images from the Bristol Image Database, and 80.4% using the Bristol Blind Mobility Database. Figure 9 shows this level of performance qualitatively, which can be seen to be very good. The lower performance on the Bristol Blind Mobility database may be explained by the higher complexity of the scenes in this database, and the low overlap with the Bristol Image Database images which dominate the neural network training. This performance should be improved by collecting more training data.

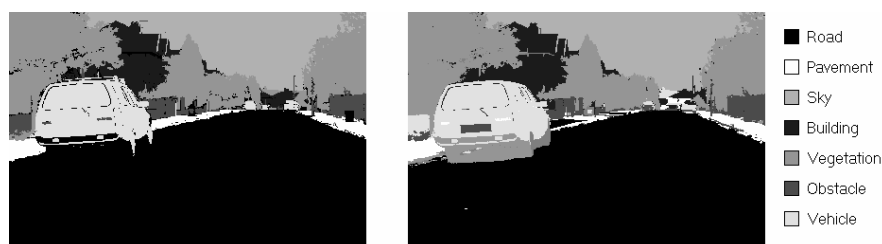


Figure 9. (a) Ideal Classification, (b) Neural Network Classification

5.1.1. Colour Constancy

Our technique for achieving colour constancy proves to be very effective. Figure 10 demonstrates qualitatively the results of using uncorrected or corrected colour features. Figure 10a shows the neural network classification of an image from the Bristol Image Database. The colour content of the image was then modified to give a very slight yellow tint, approximating the type of images obtained from our camcorder. Figure 10b shows the classification of the modified image by a network trained using uncorrected colour features, and Figure 10c shows the classification using corrected colour features. As can clearly be seen, the network trained using uncorrected colour features fails completely on most object classes due to the shift in the colour space which was not present during training, whereas the network trained using corrected colour features obtains a classification approximately as accurate as using the unmodified image.



Figure 10. (a) Classification, (b) Uncorrected Classification, (c) Corrected Classification

5.1.2. Processing Speed

As we are aiming eventually at a real-time implementation of our system, processing speed is important. At present, because of the use of Gabor filters for texture description, which are computationally expensive, performance does not approach real-time, with each image requiring 9.5s total processing time on a 300MHz Intel Pentium-II processor. However, discarding texture features we are able to achieve a classification accuracy of up to 86.5% by area in a processing time of just 300ms. We anticipate that with optimisation and current trends in increased processing speed, we may achieve true real-time performance in the near future.

5.2. Pilot Study

In order to test the effectiveness of our technique for enhancing images by classification, a pilot study was conducted in the Bristol Eye Hospital. 16 registered-blind subjects with a variety of visual impairments including age related macular degeneration, retinitis pigmentosa and optic atrophy participated in the experiment, ranging in age from 38 to 87 years with mean age 69 years. A repeated measures design was used with the subjects being shown a total of 45 images in each of three conditions – original images, image enhanced by a Peli technique (Peli, 1991) and ideally-classified images using our classification technique. The set of 45 images was divided into three groups of 15, and for each group a different task was set – pointing to at least two obstacles such as lamp-posts, pointing to at least two vehicles, and tracing the line between both pavements and the road. The order of the tasks was counterbalanced using a Latin Square technique. Images were displayed on a computer monitor with 50cm diagonal screen, at a distance subtending 54.6° of the subject’s visual field. To avoid effects due to difference in luminance between the image types, all images were normalised to have a mean luminance of 19.29cdm⁻². Percentage correct scores were recorded for each patient under each condition of image type and task.

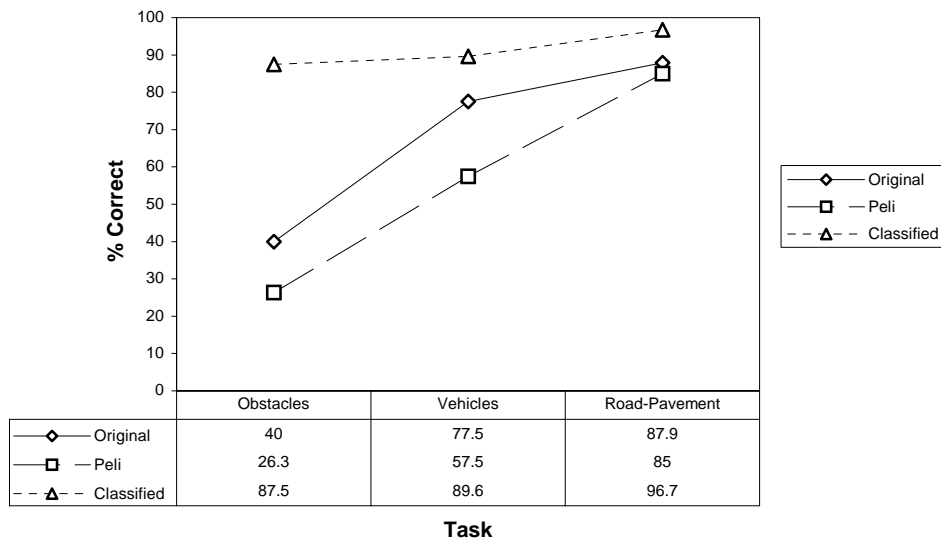


Figure 11. Mean % correct by Image Type and Task

Figure 11 summarises the main results of the experiment. A 2-way repeated measures ANOVA was performed on the individual percentage correct scores, using Image Type and Task as two factors, each with three levels. Two significant ($p < 0.001$) main effects were found. The main effect on Image Type indicates that the type of images did change performance on the tasks. The main effect on Task indicates that the tasks varied in difficulty. A significant interaction was found between Image Type and Task indicating that the type of image used improves performance on the different

tasks to varying degrees. Post-hoc analysis was performed using the Newman-Keuls technique and reveals the following significant ($p < 0.001$) results.

Mean performance of subjects is significantly better using our classification technique than using either original images or images enhanced by the Peli technique. On the most difficult task, recognition of obstacles, mean performance across subjects increased from 40% to 87.5% using our technique. Peli-enhanced images gave significantly worse performance on the first two tasks than the original images. There is also no statistically significant difference between performance over the three different tasks using our technique, indicating an improvement in performance on the tasks proportional to their difficulty.

In summary, our technique improved overall performance significantly and resulted in consistent high performance across tasks of varying difficulty. On the most difficult task, obstacle recognition, performance increased by over 100%.

6. CONCLUSIONS

This paper has described a novel content-driven approach to image enhancement for people with low vision in the context of mobility in an urban environment, combining technology from the field of virtual reality with modern computer vision techniques. The scene classification technique used as the basic model for image enhancement achieves a very high level of accuracy of up to 92.5% by image area. A pilot study has demonstrated the effectiveness of the approach, which improved object recognition performance by people with a range of visual impairments on a difficult mobility-related task by over 100%, and resulted in consistent very high performance over three tasks important for mobility.

In future work, we aim to improve the classifier accuracy and its robustness to varying environmental conditions and situations. A further study is planned to investigate which errors made by the scene classifier are most detrimental to a mobility task so that these may be targeted in development of the classification technique. Work is proceeding in attempting to achieve a real-time implementation of the system in a collaboration with the Multimedia Appliances Group at Bristol, investigating real-time implications for both the hardware architecture and algorithms used. We also plan to run further trials of the technique using modern head mounted display devices as there are issues involving field-of-view and ergonomic questions which need to be addressed.

Acknowledgements: This work is supported by the National Eye Research Centre. We thank British Aerospace PLC, Sowerby Research Centre, Bristol for their support with the development of the Bristol Image Database. We also thank Nital Karia and Manoj Kulshreshta for collection of data for the pilot study.

7. REFERENCES

- A Badreldin, A K C Wong, T Prasad, M Ismail (1980), "Shape descriptors for N-Dimensional curves and trajectories", *IEEE Proceedings on Cybernetics and Society*, pp. 713-717.
- C M Bishop (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.
- G Buchsbaum (1980), "A spatial processor model for object colour perception", *Journal of the Franklin Institute*, **310**, pp. 1-26.
- N W Campbell, W P J Mackeown, B T Thomas, T Troscianko (1997), "Interpreting Image Databases by Region Classification", *Pattern Recognition*, **30**, 4, pp. 555-563.
- D Dunn, W E Higgins (1995), "Optimal Gabor Filters for Texture Segmentation", *IEEE Transactions on Image Processing*, **4**, 7, pp. 947-964.
- J D Foley, A van Dam, S K Feiner, J F Hughes, *Computer Graphics: principles and practice*, Addison-Wesley.
- G L Goodrich and A Zwern (1995), "From virtual reality to large print access: Development of a head-mounted virtual display", *Low Vision Shared Interest Group Meeting, American Academy of Ophthalmology*, Atlanta, Georgia 30th October.
- A K Jain and F Farrokhnia (1991), "Unsupervised Texture Segmentation using Gabor Filters", *Pattern Recognition*, **24**, 12, pp. 1157-1186.
- W P J Mackeown (1994), "A Labelled Image Database and its Application to Outdoor Scene Analysis", *PhD Thesis*, University of Bristol.
- S Marcelja (1980), "Mathematical description of the responses of simple cortical cells", *Journal of the Optical Society of America*, **70**, 11, pp. 1297-1300.
- R W Massof and D L Rickman (1992), "Obstacles encountered in the development of the low vision enhancement system", *Optometry and Vision Science*, **69**, pp. 32-41.

- Møller, M (1993), "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks*, **6**, 4, pp. 525-533.
- N Molton, S Se, J M Brady, D Lee, P Probert (1998), "A stereo vision-based aid for the visually impaired", *Image and Vision Computing*, **16**, 4, pp. 251-263.
- NASA (1993), "High-tech help for low vision", *NASA Tech Briefs*, **17**, 2, pp. 20-22.
- N R Pal, S K Pal (1993), "A Review on Image Segmentation Techniques", *Pattern Recognition*, **26**, 9, pp. 1277-1294.
- E Peli (1991), "Image enhancement for the Visually Impaired – Simulations and Experimental Results", *Investigative Ophthalmology and Visual Science*, **32**, 8, pp. 2337-2350.
- E Peli, E Lee, C Trempe and S Buzney (1994), "Image enhancement for the visually impaired: the effects of enhancement of face recognition", *Journal of the Optical Society of America*, **11**, pp. 1929-1939.
- E Peli (1995), "Head-mounted display as a low vision aid", *Unpublished manuscript*.
- E Peli (1998), "Wide-band image enhancement for the visually impaired", *Poster, ARVO Annual Meeting 1998*.
- J D Prothero (1993), "The treatment of akinesia using virtual images", *MSE Thesis, University of Washington*.
- S Z Selim and M A Ismail (1984), "K-means-type algorithms: A generalized convergence theorem and characterisation of local optimality", *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, **6**, 1, pp. 81-87.
- M Shridhar and A Badreldin (1984), "High accuracy character recognition algorithm using fourier and topological descriptors", *Pattern Recognition*, **17**, 5, pp. 515-524.
- M Snaith, D Lee and P Probert (1998), "A low-cost system using sparse vision for navigation in the urban environment", *Image and Vision Computing*, **16**, 4, pp. 225-233.
- G Wyszecki, W S Stiles (1982), *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Wiley, New York.