

Wearable Mobility Aid for Low Vision Using Scene Classification in a Markov Random Field Model Framework

M. R. Everingham

B. T. Thomas

Department of Computer Science
University of Bristol

T. Troscianko

Perceptual Systems Research Centre
University of Bristol

This article describes work on a novel approach to vision enhancement for people with severe visual impairments. This approach utilizes computer vision techniques to classify scene content so that visual enhancement of the scene can identify semantically important concepts. The mediated view of a scene presented to the user is in the form of a highly-saturated color image in which distinct colors represent important object types in the scene. The effectiveness of this scheme was demonstrated in a pilot study participated in by people with a range of visual impairments. The scene classification technique uses an artificial neural network classifier within the framework of a Markov random field model, and the accuracy and robustness of this technique using low quality video images from a hand-held camera is demonstrated.

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 visual impairment is typically such that the person is unable to be mobile without some form of assistance (e.g., a guide dog). In recent years, as real-time image processing has become feasible, and devices such as head-mounted displays suitable for the implementation of wearable computers have become available, an increasing amount of research has attempted to harness this new technology for the

M. R. Everingham is now at the Department of Engineering Science, University of Oxford.

Requests for reprints should be sent to M. R. Everingham, Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, U.K. E-mail: me@robots.ox.ac.uk

needs of people with low vision. In this article, we first review the traditional approaches which have been used, then describe work on a novel wearable vision aid being developed at the University of Bristol.

1.1 Previous Work

Over the last 15 years an increasing amount of research has attempted to apply techniques from computer vision to the needs of people with low vision. Peli and co-workers (Peli, Lee, Trempe, & Buzney, 1994) examined contrast enhancement in spatial frequency bands important to specific tasks such as face recognition. Massof and Rickman (1992) produced a head-mounted system providing magnification and real-time contrast enhancement using a single camera and dual displays. Development of this device was continued in conjunction with NASA (NASA, 1993). Peli (1995) implemented a device using a single display to show binary images with adjustable brightness and contrast enhancement of high one-dimensional spatial frequencies, and a 2× zoom facility. Goodrich and Zwern (1995) used a commercially available color head-mounted display to provide variable contrast and magnification. Common to these approaches is the use of a person's residual vision to convey information to them, as typically this remains the highest bandwidth pathway for people with low vision, and because they tend to depend more on their hearing than sighted people it is desirable not to affect this sense. Other approaches include those of Molton, Se, Brady, Lee, and Probert (1999), who describe techniques for detecting obstacles, curbs, and steps with applications to people with complete loss of vision or low vision, although the exact method of conveying this information to the user is left as an open question.

1.2. Enhancement By Classification

We have concentrated on the domain of mobility within a typical urban environment, an activity that presents many problems for people with low vision. A fundamental limitation of the majority of previous approaches is that they operate without any semantic knowledge of the content of the scene being perceived, thus they are unable except heuristically to determine which visual aspects of the scene need be conveyed to the user, and which may be discarded. The common assumption has been that the contrast of high spatial frequencies should be increased, since people with low vision typically have reduced perception of high spatial frequencies, and these often correspond to edges conveying the outline of objects. However, in the absence of semantic knowledge, such a device may amplify irrelevant information such as noise or textural detail, while failing to amplify semantically important information (e.g., the distinction between pavement and road). In our approach, we have used high-level computer vision techniques to enable the vision aid to selectively enhance the visibility of important scene aspects. Our device therefore works in the mode of an intelligent reality mediator, selectively filtering the user's visual perception of the world instead of applying simple image processing techniques.



FIGURE 1 Image enhancement by scene classification.

Figure 1 shows an original scene and how it is ideally presented to the user by our mobility aid. The principle is to classify pixels of the image into one of a small set of object types, shown in the key, and to use distinct high saturation colors to represent each object type. In this way, objects which are visually very similar but semantically distinct may be transformed to become visually distinct also. For example, the road and pavement are very similar in appearance but the distinction between them is important. In the classified image they are colored black and white respectively, making the distinction very clear.

The scene from Figure 1 is shown again in Figure 2, but blurred by a large amount to simulate how the scene might appear to someone with low vision in the form of very low spatial acuity. We can see that in the original image no detail is visible, while in the classified image concepts such as the distinction between road and pavement, location of vehicles, and even small obstacles (such as the lamppost on the left) are clearly visible. In a pilot study conducted in the Bristol Eye Hospital (Everingham, Thomas, & Troscianko, 1999) in which 16 registered-blind participants with a variety of visual impairments were examined, we showed that use of our technique consistently improved performance on tasks such as recognizing vehicles, obstacles, and road-pavement boundary, with over 100% improvement on the more difficult tasks.

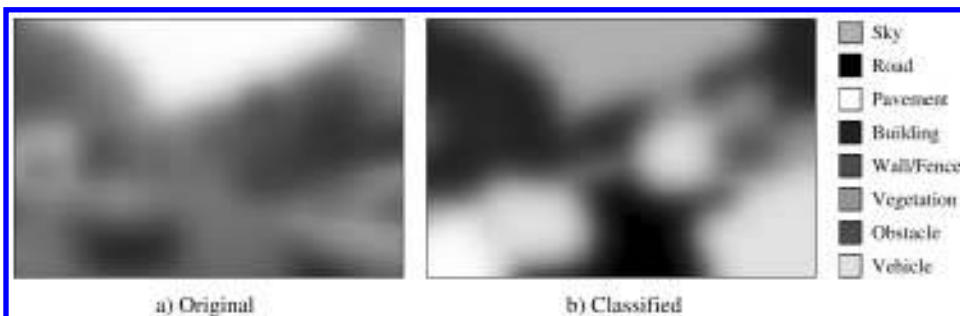


FIGURE 2 Simulated low-vision appearance.

2. SCENE CLASSIFICATION TECHNIQUE

In order to produce colored images (such as in Figure 1), we need to classify each pixel in an image as belonging to one of the chosen object classes. In previous work at Bristol (Campbell, Mackeown, Thomas, & Troscianko, 1997; Mackeown, 1994), techniques were developed for classification of static scenes using an approach of segmentation into regions, feature extraction, and classification using an artificial neural network (ANN). This was extended and adapted for the application under consideration (Everingham et al., 1999). In applying this approach to video sequences captured with low quality cameras, problems were encountered with the stability of segmentation and associated classification. Therefore, in this article we describe a modified approach based on a Markov random field (MRF) framework, which has proven more suitable for use with low-quality video sequences, at the cost of reduced image resolution.

Figure 3 shows the general architecture of the system. Original color images from a camera are first divided into blocks of pixels. Feature vectors describing the visual properties of a block are extracted, and an ANN is used to estimate the probability that the block arose from an object of each class. An MRF model then uses the ANN output probabilities and the class labels assigned to blocks in the previous frame to incorporate spatial and temporal relationships between blocks, producing a final labeling of each block used to color the output image. During a training phase a set of ground truth data in the form of hand classified images is used to train the ANN and initialize the MRF model.

2.1. Image Subdivision

In previous work (Everingham et al., 1999), we used an image segmentation algorithm to initially divide an image into regions of pixels which had similar visual properties, with the aim of isolating individual objects or object parts. When using low-quality images an algorithm could not be found which would reliably over-

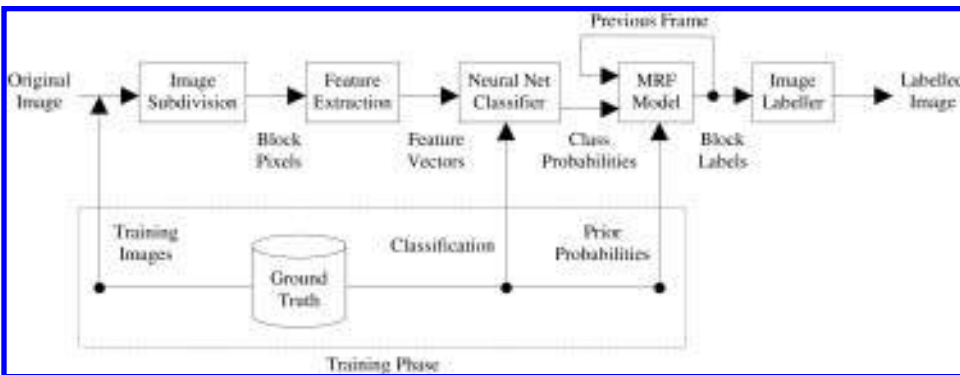


FIGURE 3 Mobility aid architecture.

segment an image, causing problems such as road and pavement not being separated; so, here we instead divide the image into arbitrary square blocks for classification. We use 8×8 pixel blocks, where the image resolution is 360×288 pixels. The resulting classified image is thus of lower resolution than the original image, but we judge it to be of sufficient resolution for the task at hand, and note that people with low vision typically have reduced spatial acuity which means they are less affected by the lower resolution than people with normal vision. Figure 4 shows an ideal classification of an image, and the corresponding classification after subdivision into blocks.

2.2. Classification of Blocks

We formulate the classification problem as that of finding the assignment of a class, or “labeling” of each image block, which maximizes the posterior probability of the class labels given a set of observations on each block. We define the set of blocks $S = \{S_1, \dots, S_N\}$ and associate with each block a label X_S taken from the set $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ of eight labels shown in the key of Figure 4, and a corresponding set of observations on the image data $\{F_1 = f_1, \dots, F_N = f_N\}$, defined in Section 2.3. Abbreviating a particular assignment of labels $\{X_1 = x_1, \dots, X_N = x_N\}$ as $\{X = \omega\}$, and the set of all possible such assignments as Ω , the task is to find the configuration of labels $\hat{\omega} \in \Omega$ which maximizes the a posteriori probability $P(\omega | F)$. We initially make the assumption that a block’s label is dependent only on the observation at that block, and not the observations or labels at other blocks, to give

$$P(\omega | F) = \prod_{s \in S} P(x_s | f_s) \tag{1}$$

We use an ANN to directly estimate the posterior probability $P(x_s | f_s)$. The network has a standard fully-connected multi-layer perceptron architecture (Bishop, 1995) with one hidden layer of sigmoid units. Input to the neural network is the fea-

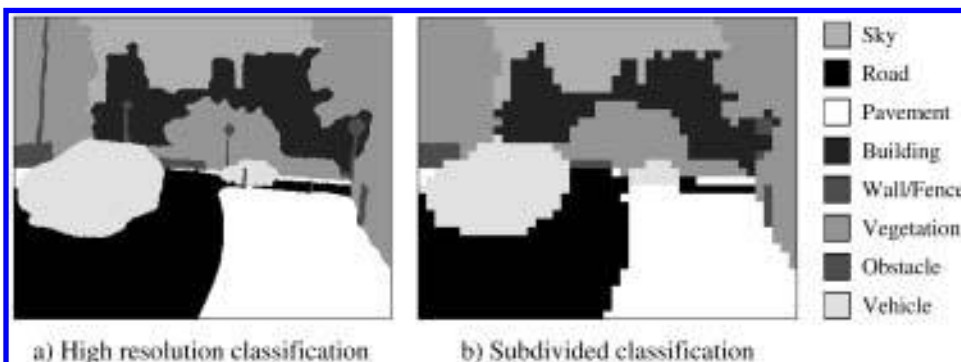


FIGURE 4 High resolution and subdivided classified image.

ture vector f_s of observations on a block. The network has outputs $q_1 \dots q_M$, one for each label. In order to be able to interpret the network outputs as probabilities $q_i = P(x_s = \lambda_i | f_s)$ we must ensure that they all lie in the range 0 – 1 and sum to unity. This is achieved by using the “softmax” activation function (Bridle, 1990) for the output units, defined as

$$q_i = \frac{\exp(a_i)}{\sum_{j=1}^M \exp(a_j)} \quad (2)$$

where a_i is the weighted sum of inputs from the hidden layer to output i .

2.3. Feature Set

Input to the neural network is a vector of real-valued features calculated from the pixels of an image block in the original color image. Table 1 shows the set of features used. The position of the block gives the ANN useful information since the set of views of a scene is restricted to those from a head-mounted camera used by a pedestrian. The distribution of color in the block is represented by mean and standard deviation across the block, calculated in the perceptually uniform CIE L*u*v* space (Wyszecki & Stiles, 1982). In addition, the mean output from a bank of complex-valued Gabor filters (Jain & Farrokhnia, 1991) is used to characterize the distribution of spatial frequency and orientation within the block, which gives information about textural properties and local structure. We used 16 filters at four frequencies and orientations, as detailed in Table 1. A total of 24 input features are presented to the ANN.

2.4. Results

The ANN was trained from a set of 50 hand-labeled images using the scaled conjugated gradient (SCG) algorithm (Moller, 1993). Twenty-five additional images were used as a validation set, with training being stopped when the classification error on the validation set failed to decrease. A further 75 images were used for testing. Table 2 shows the confusion matrix of percentage of blocks correctly classified

Table 1: Feature Set

<i>Feature</i>	<i>Description</i>
1–2	Block position: x, y
3–8	Color: L*u*v* mean, sd
9–24	Gabor filter mean magnitude output frequency = $1 / (i\sqrt{2})$ cycles/pixel: i=2, 4, 8, 16 orientation = 0°, 45°, 90°, 135° bandwidth = 1 octave, 45°

Table 2: Neural Network Confusion Matrix

	<i>Sky</i>	<i>Road</i>	<i>Pave</i>	<i>Build</i>	<i>Wall</i>	<i>Veg</i>	<i>Obst</i>	<i>Vehi</i>
<i>Sky</i>	90.12	0.00	0.00	2.19	0.04	6.49	1.09	0.07
<i>Road</i>	0.00	90.41	8.47	0.13	0.13	0.18	0.38	0.30
<i>Pave</i>	0.00	6.42	91.52	0.29	1.41	0.10	0.21	0.05
<i>Build</i>	1.11	0.70	1.43	65.68	13.82	14.52	2.48	0.26
<i>Wall</i>	0.00	0.68	5.94	9.73	71.48	10.69	1.34	0.13
<i>Veg</i>	0.56	0.10	0.13	2.45	3.74	91.96	1.00	0.07
<i>Obst</i>	2.22	4.46	7.74	12.11	19.55	25.82	24.99	3.10
<i>Vehi</i>	1.52	18.96	6.37	4.56	5.02	6.66	6.46	50.45

Note: Pave = pavement, Build = building, Veg = vegetation, Obst = obstacle, Vehi = vehicle. Boldface = correct classification.

using the ANN, with the correct classification shown in boldface rows corresponding to the correct class label. The class label assigned to a block was taken as that for which the network output $q(\lambda_i)$ was maximal. The mean percentage of each image correctly classified is 82.4%. A qualitative example of this classification accuracy can be seen in Figure 6b. We can see that the overall accuracy is quite high, but a significant number of misclassified image blocks remain, which appear as speckles in the classified image, and which are unstable across consecutive frames of video.

3. MRF FRAMEWORK

In Section 2.2., we assumed that the labeling of an image block is independent of all other blocks (Equation 1). We now extend the classification scheme using an MRF model (Geman & Geman, 1984) to incorporate information about the spatial and temporal relationships between blocks. For example, we expect that neighboring blocks are likely to have the same label, and that a block's label is likely to remain unchanged over a short period of time. We formulate the labeling problem as a Bayesian estimation problem, aiming to find the configuration of labels $\hat{\omega}$ which maximizes the posterior probability $P(\omega | F)$ defined by Bayes theorem as

$$P(\omega | F) = \frac{P(F | \omega)P(\omega)}{P(F)} \quad (3)$$

and model this as an MRF. We begin by defining a graph $G = \{T, E\}$ where $T = \{S_{t-1}, S_t\}$ is the set of nodes corresponding to image blocks defined in Section 2.2. at times $t-1$ and t respectively, and E is the set of edges connecting them. Consecutive values of t correspond to frames of a video sequence. Recalling that the assignment of labels to each block is denoted $\{X = \omega\}$, then X is an MRF with respect to G if

$$\begin{aligned} P(X = \omega) > 0 \quad \forall \omega \in \Omega \\ P(X_s = x_s | X_r = x_r, \forall r \neq s) = P(X_s = x_s | X_r = x_r, \forall r \in N_s) \end{aligned} \quad (4)$$

that is to say, the label assigned to a particular node is affected only by the labels assigned to its neighbours, and all assignments are possible. By the Hammersley-Clifford theorem it can be shown that these conditions are satisfied if $P(X = \omega)$ is a Gibbs distribution with respect to G having the form

$$\pi(X = \omega) = \frac{\exp\left[-\sum_{c \in C} V_c(\omega)\right]}{Z} \quad (5)$$

where C is the set of cliques defined on the graph G , V_c is a potential function defined on the clique c , and Z is a normalizing constant. A clique is a subset of G such that all its members are neighbors. The utility of this formulation is that we can now define the relationships between labels assigned to blocks in terms of clique potentials instead of attempting to estimate the label probabilities directly.

3.1 Clique Potential Definitions

Figure 5 shows the cliques defined on the graph of image blocks in the current and previous video frames. Observations are defined only on the blocks of the current frame, shown in gray. The spatial cliques are defined between adjacent blocks on an 8-connected grid. Each block in frame t also has a single temporal neighbor at the corresponding spatial position in frame $t-1$. In order to avoid any latency in the classification process, we do not have access to any future frames, so the model works in an online manner. We now discuss the definition of potentials associated with each kind of clique.

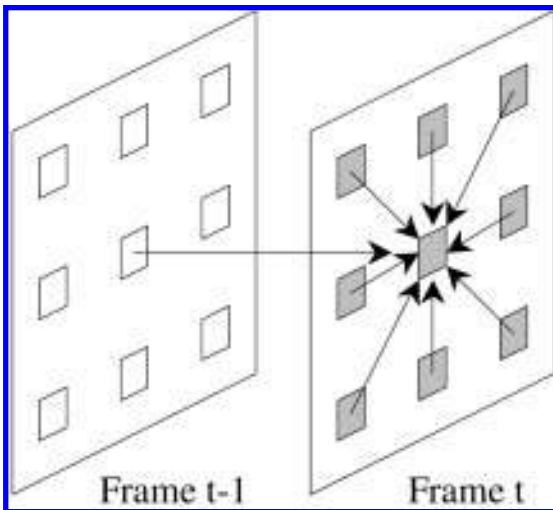


FIGURE 5 Spatial and temporal cliques.

Prior potentials. We first define clique potentials on each block in the current frame characterizing the prior probability of each label $P(x_s = \lambda)$, which we estimate from the training data:

$$V_p(x_s) = -\log[1 + P(x_s)] \tag{6}$$

Spatial potentials. The spatial clique potentials model the likelihood of a pair of adjacent image blocks in the current frame having particular labels, given the spatial relationship between them (e.g., given that one is above the other or to it's left). We estimate the probability of a block's label given a single neighbor's label, and form the clique potential

$$V_s(i, j) = -\log[1 + P(x_s)] \tag{7}$$

where i and j are adjacent image blocks at time t , and r_{ij} gives the relationship between them (e.g., "above"). An estimate of the conditional probabilities is obtained from the ground truth data using a conventional histogram technique. Table 3 shows an example of this distribution, for $x_i = sky$ and $r_{ij} = below$. In this case we can see that blocks below those labeled as "sky" are most likely to also be "sky", but also more likely to be "vegetation" than "obstacle."

Temporal potentials. The temporal clique potentials model the likelihood of a block label changing from one value to another in consecutive video frames. In the absence of complete hand classified image sequences, we adopt a particularly simple model in which a block's label is assumed to stay unchanged from one frame to the next, equivalent to assuming that an object's motion between frames will be small relative to the block size. This gives the clique potential

$$V_t(i, j) = \begin{cases} +1 & \text{if } x_i \neq x_j \\ -1 & \text{if } x_i = x_j \end{cases} \tag{8}$$

where i and j represent an image block at a single spatial position in frames $t-1$ and t respectively.

Table 3: Probability of Label Below "Sky"

<i>Sky</i>	<i>Road</i>	<i>Pave</i>	<i>Build</i>	<i>Wall</i>	<i>Veg</i>	<i>Obst</i>	<i>Vehi</i>
0.939	0	0	0.0129	0	0.0349	0.0131	0

Note: Pave = pavement, Build = building, Veg = vegetation, Obst = obstacle, Vehi = vehicle.

Observation potentials. Thus far we have only considered the a priori probability $P(\omega)$ in Equation 3. We now incorporate the probability density function (pdf) of the set of observations given the labels assigned to the image blocks $p(F | \omega)$. We assume that the observations at a block are independent of all others in space and time, giving

$$p(F | \omega) = \prod_{s \in S_t} p(f_s | x_s) \quad (9)$$

In Section 0 we showed how the ANN can be used to estimate $P(x_s | f_s)$. By applying Bayes theorem we now obtain

$$p(f_s | x_s) = \frac{P(x_s | f_s)p(f_s)}{P(x_s)} \quad (10)$$

We note that the pdf $p(f_s)$ does not depend on the labeling, and therefore obtain

$$p(f_s | x_s) \propto \frac{q_{x_s}(f_s)}{P(x_s)} \quad (11)$$

where $q_i(f)$ is the ANN output for label λ_i given inputs f . The prior probability $P(x_s)$ is estimated from the training data. Converting to a potential compatible with the MRF model, we obtain

$$V_o(s) = -\log \left[1 + \frac{q_{x_s}(f_s)}{P(x_s)} \right] \quad (12)$$

Energy function minimization. Combining Equations 6, 7, 8, and 12, we arrive at the following energy function, whose minimum gives the MAP labeling ω according to the MRF model:

$$U(\omega, F) = \sum_{s \in S_t} V_o(s) + \alpha \sum_{s \in S_t} V_p(s) + \beta \sum_{c \in C_s} V_s(c) + \gamma \sum_{c \in C_t} V_t(c) \quad (13)$$

where C_s is the set of spatial cliques at time t , and C_t is the set of temporal cliques. The parameters α , β , and γ control the relative contribution of prior, spatial, and temporal potentials. We use values of $\alpha = 1$, $\beta = 0.7$, and $\gamma = 0.1$, where these values were chosen experimentally. The results are relatively insensitive to the exact parameter values. To minimize $U(\omega, F)$ we use the Iterated Conditional Modes (ICM) algorithm (Besag, 1986), which is an iterative deterministic relaxation algorithm which converges to a local minimum of the energy function. The energy is minimized under the constraint that labels of sites in the previous frame are unchanged.

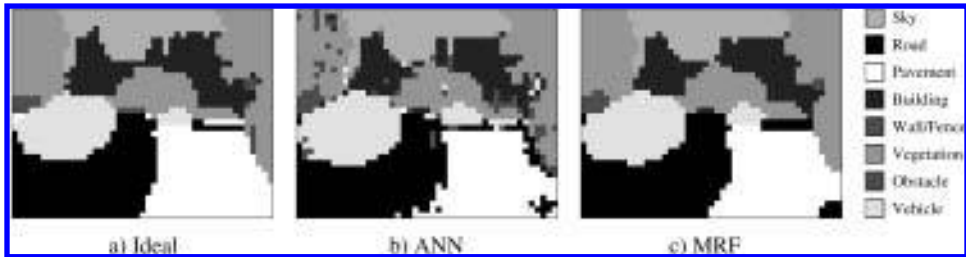


FIGURE 6 Classification results.

We found that the algorithm typically converges very quickly, requiring no more than 10 iterations.

4. RESULTS

Figure 6 shows qualitatively an example of the classification performance achieved. Figure 6a shows the ideal classification, Figure 6b the classification achieved using the neural network classifier alone, and Figure 6c shows the final classification after applying the MRF model. The results are subjectively very good, with the speckled misclassifications found when using the ANN alone removed. Figure 7 and Figure 8 are further examples of classification using the MRF model. These show two video

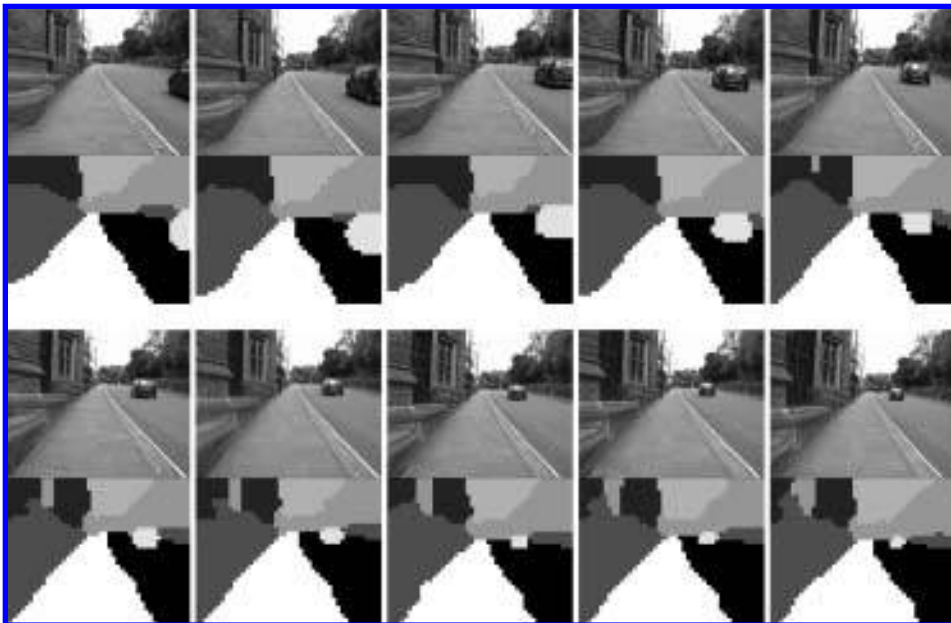


FIGURE 7 Classified video sequence.

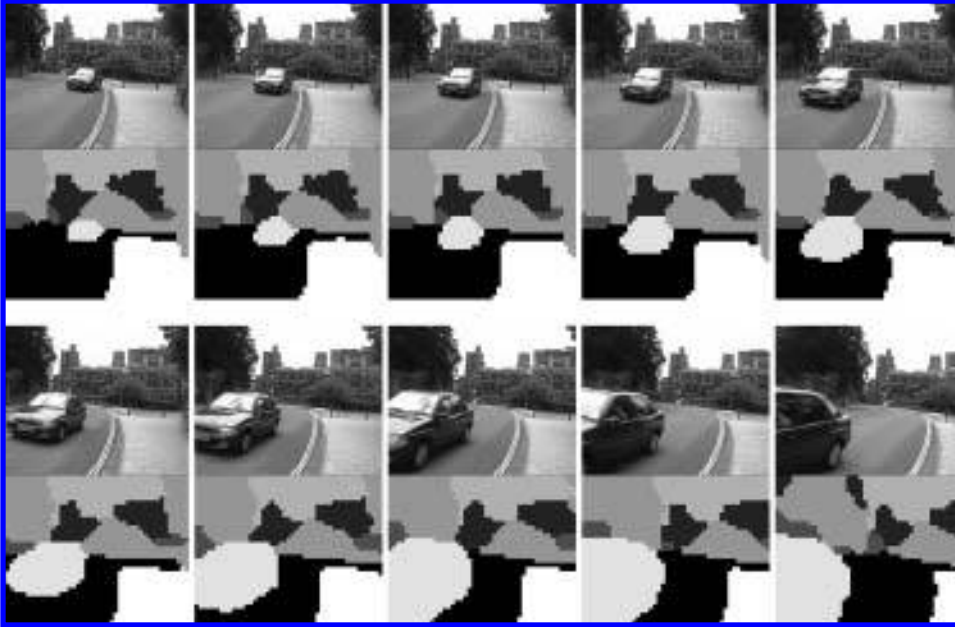


FIGURE 8 Classified video sequence.

sequences, where frames were captured at 25 frames per sec, and every fifth frame is shown. The sequences were captured using a hand-held domestic camcorder by a walking observer. The results demonstrate the robustness and stability of the technique to different viewpoints, low quality video, and a moving camera (further examples may be found on the world wide web at: <http://www.cs.bris.ac.uk/~everingm/demo/>).

Table 4 shows the confusion matrix of percentage of blocks correctly classified using the final MRF model, with the correct classification shown in boldface. The mean percentage of each image correctly classified is increased to 86.3%, com-

Table 4: Confusion matrix with MRF model

	<i>Sky</i>	<i>Road</i>	<i>Pave</i>	<i>Build</i>	<i>Wall</i>	<i>Veg</i>	<i>Obst</i>	<i>Vehi</i>
Sky	91.04	0.00	0.00	2.34	0.01	5.22	1.38	0.01
Road	0.00	91.70	6.42	0.14	0.34	0.39	0.57	0.43
Pave	0.00	6.82	90.64	0.22	1.50	0.25	0.47	0.10
Build	1.44	0.37	0.62	74.29	10.67	9.19	3.09	0.33
Wall	0.00	0.38	4.35	10.61	75.32	6.63	2.62	0.09
Veg	0.92	0.13	0.30	3.07	4.43	89.05	1.99	0.12
Obst	0.94	2.68	5.65	3.65	8.20	7.36	66.92	4.60
Vehi	0.84	9.45	2.12	2.00	0.97	3.40	3.85	77.37

Note: Pave = pavement, Build = building, Veg = vegetation, Obst = obstacle, Vehi = vehicle. Boldface = correctly classified.

pared to 82.4% obtained using the ANN alone. We can also see that the performance on certain individual classes has improved significantly; for example, the performance on obstacles has increased from 24.99% to 66.92%. This is an important result because this class represents a very important concept for users with low vision. In addition, we note that when the system makes confusions, they are likely to be of low importance. For example, the confusion of "wall" with "building" is unlikely to cause problems for the user. In the case of such misclassifications the user can often use his own domain knowledge to perceptually correct misclassifications made by the system, thus a synergy evolves between the user and computer.

5. DISCUSSION

In this article, we have described a novel approach to vision enhancement for people with low vision in the context of mobility, which is suitable for implementation as a wearable mobility aid. The technique of using scene classification to drive the vision enhancement provides flexibility impossible using conventional image processing techniques, and the potential robustness of the system has been demonstrated. In future work, we aim to improve further the accuracy and speed of the system, and a real-time wearable implementation is being developed exploiting the inherent parallelism in the scene classification technique. We additionally plan more research in conjunction with the Bristol Eye Hospital into ways in which the mediated view of the world presented by the device can be enhanced or customized to maximally utilize the residual vision of users with severe visual impairments. We note also that because the system gathers semantic information about a scene, it should be possible to use output modalities other than visual displays, such as speech or haptic output devices.

REFERENCES

- Besag, J. (1986). On the statistical analysis of dirty images. *Journal of the Royal Statistical Society B*, 48(3), 259–302.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Clarendon Press.
- Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. F. Soulie, & J. Hérault (Eds.), *Neurocomputing: Algorithms, architectures and applications* (pp. 227–236). New York: Springer-Verlag.
- Campbell, N. W., Mackeown, W. P. J., Thomas, B. T., & Troscianko, T. (1997). Interpreting image databases by region classification. *Pattern Recognition (Special Edition on Image Databases)*, 30(4), 555–563.
- Everingham, M. R. (1999). *Low vision mobility aid demonstrations*. Retrieved February 10, 2003 from <http://www.cs.bris.ac.uk/~everingm/demo/>
- Everingham, M. R., Thomas, B. T., & Troscianko, T. (1999). Head-mounted mobility aid for low vision using scene classification techniques. *International Journal of Virtual Reality*, 3(4), 3–12.

- Geman, D., & Geman, S. (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 721–741.
- Goodrich, G. L., & Zwern, A. (1995, October). *From virtual reality to large print access: Development of a head-mounted virtual display*. Paper presented at the Low Vision Shared Interest Group Meeting, American Academy of Ophthalmology, Atlanta, GA.
- Jain, A. K., & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12), 1167–1186.
- Mackeown, W. P. J. (1994). *A labelled image database and its application to outdoor scene analysis*. Unpublished doctoral dissertation, University of Bristol, UK.
- Massof, R. W., & Rickman, D. L. (1992). Obstacles encountered in the development of the low vision enhancement system. *Optometry and Vision Science*, 69, 32–41.
- Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6, 525–533.
- Molton, N., Se, S., Brady, M., Lee, D., & Probert, P. (1999). Robotic sensing for the partially sighted. *Robotics and Autonomous Systems*, 26(3), 185–201.
- NASA. (1993). High-tech help for low vision. *NASA Tech Briefs*, 17(2), 20–22.
- Peli, E. (1995). *Head-mounted display as a low vision aid*. Unpublished manuscript.
- Peli, E., Lee, E., Trempe, C., & Buzney, S. (1994). Image enhancement for the visually impaired: The effects of enhancement on face recognition. *Journal of the Optical Society of America*, 11, 1929–1939.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.