Automatic Road Environment Classification

Isabelle Tang and Toby P. Breckon

Abstract—The ongoing development autonomous vehicles and adaptive vehicle dynamics present in many modern vehicles has generated a need for road environment classification—i.e., the ability to determine the nature of the current road or terrain environment from an onboard vehicle sensor. In this paper, we investigate the use of a low-cost camera vision solution capable of urban, rural, or off-road classification based on the analysis of color and texture features extracted from a driver's perspective camera view. A feature set based on color and texture distributions is extracted from multiple regions of interest in this forward-facing camera view and combined with a trained classifier approach to resolve two road-type classification problems of varying difficulty—{off-road, on-road} environment determination and the additional multiclass road environment problem of {off-road, urban, major/trunk road and multilane motorway/carriageway}. Two illustrative classification approaches are investigated, and the results are reported over a series of real environment data. An optimal performance of ∼90% correct classification is achieved for the {off-road, on-road} problem at a near real-time classification rate of 1 Hz.

Index Terms—Color classification, machine learning classifier, road-type classification, texture classification.

I. INTRODUCTION

A RANGE of vision-based vehicle systems have been developed to offer both autonomous guidance and human driver assistance [1]. These vary from on-road sign recognition [2] to off-road terrain classification [3], general obstacle detection [4], and highway lane detection [26], [31]. The use of radar obstacle detection is also being widely considered [5] for on-vehicle use. In all cases, the requirements to augment the driver's situational awareness, reduce driver workload, and automate all/part of the driving process are the keys.

In this paper, we specifically investigate the vision-based classification of differing road environments. Modern vehicle subsystems often include a capability to adapt traction, braking, and engine dynamics to the current type of terrain being traversed or the driving environment (e.g., urban/city or rural/highway). Work on visually detecting these terrain or road environment changes, possibly as part of a wider vehicle sensor suite, is limited [3], [6], [7].

Prior work in this area has investigated the use of color distributions as a method to classify differing off-road terrains into approximate classes [3], whereas other work in the field, which is targeted at road recognition, concentrates on the estimation of the geometric structure of the lane boundaries of a road [6]. Lane segmentation algorithms are now commonplace for road and traffic lanes analysis [5]. The watershed approach detailed in [6] for morphological segmentation is typical with the use of the contrast criterion for subsequent region combinations. A dynamic time filter of morphological opening and closing is then used on successive images of a sequence to identify separation between the traffic lanes [6]. This is essentially a color-driven approach. Alternative algorithms follow a texture-driven approach to this problem [7] via the classification of each individual image pixel based on texture analysis of their local pixel neighborhood. This facilitates the derivation of two texture clusters within the image (*road* and *off-road*) via neural network classification.

The recent work of Ess *et al.* [27] has also investigated the area of feature-based road scene understanding. This work concentrates on the understanding of the urban road environment in terms of road marking presence, on-road and road-side object occurrence, and road-type classification in terms of the occurrence of *{left, right,* cross-*road}* junctions, roundabouts, and road turns *{left, right}*. This work is driven by scene segmentation and limited to the understanding of the scene layout complexities of the urban road-way environment. This work is in itself related to the wider topic of scene classification in computer vision that addresses generalized scene categorization [28]–[30] of which road scenes generally appear as a single class.

By contrast, here, we investigate not the explicit segmentation or understanding of the road scene [4], [7] but instead look at advance work in terrain classification [3] to determining differing road driving environments. While primarily aimed at a driver assistance application, as a conduit to the automation of vehicle economy/performance dynamics, such a capability may also be of use in wider autonomous vehicle applications [1]. We specifically identify two classification problems of varying difficulty–our four- and two-class road environment classification problems. In the former, we consider four general environments (*off-road, urban, major/trunk road*, and *multilane motorway/carriageway*), whereas in the latter, recognizing the potential feature overlap in the previous on-road environments (in the four-class problem), we ground our investigation with a comparison to the more restricted two-class (*off-road* and *on-road*) problem. For clarification, we consider an urban environment to be a built-up city or town-type locale, a major or trunk road to be an single lane (per direction of travel) road interconnecting disjoint urban areas, and a motorway or multilane carriageway to be a multilane (per direction of travel) highway/expressway commonly found

road
road-edge
road-side

Fig. 1. Subregions of interest selected for feature extraction for the four road environments under consideration.

connecting major metropolitan areas (with vehicles operating up to the highest legal speed limit). An off-road environment is determined as being *"off the recognized public highway"* (U.K.) on an unpaved/nontarmac roadway. Examples of each environment class are shown in Fig. 1.

Specifically, we make use of four discrete subregions of the forward-facing camera view from the vehicle (see Fig. 1) empirically chosen for the varying environmental context information they contain. These are analyzed, following the combined concepts of [3] and [7], to produce a combined texture and color distribution-based feature representation of a given driving environment. Feature classification using three classifier variants is then investigated with regard to the identified two- and four-class road-type classification problems. A k-nearest neighbor (k-NN) and an artificial neural network (ANN) approach are

investigated, with optimal performance achieved using the ANN approach (see Tables II and III). The use of k-NN is aimed at demonstrating the underlying data discrimination properties of these two multiclass problems. Overall, we achieve near real-time performance (1 Hz) based on an input from a video input from a forward-facing dashboard/windshield camera suitable for integration into a vehicle management system.

In the remainder of this paper, we outline the chosen image subregions (see Section II) and the color and texture feature descriptors used as classifier inputs (see Section III), and we give a brief overview of the classification approaches investigated (see Section IV). Results over defined test sequences are presented with subsequent analysis (see Section V) and conclusions (see Section VI).

II. ROAD IMAGE SUBREGIONS

A set of three subregions of interest is selected from the forward-facing road image: 1) *road*; 2) *road edge*; and 3) *road side* (see Fig. 1).

As shown in Fig. 1, these subregions have been selected to capture differing properties of the driving environment. The *road* region captures properties of the road surface itself, whereas the *road-side* region captures the nature of the general driving environment off of the road surface itself. The remaining *road-edge* region captures the unique properties of the road surface to road-side transition including the specifics of any road side markings.

These subregions are fixed within the forward view image using a fixed-position windscreen-mounted camera. The position of this camera is calibrated at installation time.

III. FEATURE EXTRACTION AND REPRESENTATION

Within the process itself, each image frame (from the continuous real-time video input) is represented by a feature vector made up of color and texture features extracted from the subregions of interest. This vector forms the input to the latter stage of classification.

A. Color Features

A choice of color space representation is made to provide both a spread of color distribution features and robustness to variations in illumination [8]–[10] that support prior work in the related field of obstacle identification/avoidance [11] and traversable pathway determination [12].

A unique combination of original RGB, HSV [10], and YCrCb [8], [9] is selected for each of the identified subregions of interest with reference to both established practice within this area [8]–[10], prior successful classification work within the road domain [11], [12], and related work on other domains where color variance isolation is of primary importance [13].

With reference to these works [8], [10] and from empirical preliminary statistical investigations [12] on sample road imagery, a unique combination of color channel inputs was chosen for each of the subregions of interest, as shown in Fig. 1. This selection is detailed in Table I with a summary

TABLE I
COLOR CHANNEL SELECTION FOR EACH SUBREGION OF INTEREST

Region of Interest	Colour Representation / Channel	Supporting Statement
road-edge	- RGB = {R, G} (red and green)	- red/green combination empirically gives a strong, yet differentiable response to both *yellow* and *white* (road-edge markings) in addition to capturing road to road-side transition properties in urban and rural environments.
	- HSV = {H} (hue)	- illumination invariance [10].
road-side	- RGB = {G} (green)	- strong response indicator of general urban/rural environmental differences.
	- HSV = {H} (hue)	- illumination invariance [10].
road	- HSV = {H} (hue)	- illumination invariance [10].
	- YCrCb = {Y} (illumination)	- isolated illumination component [8, 10].

supporting statement for each. For definitions of the RGB, HSV, and YCrCb color spaces, see [14].

In general, a set of color features is derived from each selected color channel for each of the three subregions of interest (see Table I). Here, we represent each such channel of each subregion as the normalized histogram distribution (probability distribution) for value occurrence together with the mean, standard deviation, and entropy of color values present. Within this context entropy, E [14], giving an effective measure of the information conveyed within a given color channel of a given region is defined as follows:

$$E = -\sum_{k=1}^{L} p_k \log_2 p_k \qquad (1)$$

where a given color value (indexed $k = 1, \ldots, L$) occurs with probability p_k.

Each color channel is summarized as a color feature vector of combining the histogram (quantized to 10 "bins") and three summary features computed on the pixel values (mean, standard deviation, and entropy). The concatenation of this 13-value feature vector over all color channels for all three subregions of interest results in a 91-D color descriptor for each image frame.

B. Texture Features

In addition to the color feature descriptor, we additionally devise a corresponding texture descriptor based on the use of two established texture measures, i.e., grey-level cooccurrence matrix (GLCM) statistics [15] and Gabor filters [16], with modern relevance and interpretation [17], [18].

Essentially, the GLCM measures the relative cooccurrence of image values by recording how often different combinations of gray (or, in general, any color representation value; see Table I) adjacently occur within the image given a localized orientation [15]. Traditionally, this localized orientation is defined with reference compass directions as $\{N, S, E, W, NW, NE, SW, SE\}$. The GLCM itself is not a texture feature descriptor in its own right but is more correctly a 2-D statistical record

of cooccurring value variation within the image [15]. In the original work of Haralick *et al.* [15], texture features were in turn calculated from the normalized cooccurrence matrix as a set of 14 summary statistics. The work of Arvis *et al.* [17] (in line with the subsequent review of [18]) down-selected a subset of five summary statistics based on the realization of the intercorrelation within the original set of 14. This subset of five measures ($entropy, energy, contrast, correlation, localhomogeneity$) [17] is used in this paper as a workable representative texture feature descriptor derived from the original GLCM formulation [15]. Each is described as follows:

$$Entropy = -\sum_{i=1}^{cols}\sum_{j=1}^{rows} M(i,j)log_2\left(M(i,j)\right) \qquad (2)$$

$$Energy = \sum_{i=1}^{cols}\sum_{j=1}^{rows} M(i,j)^2 \qquad (3)$$

$$Contrast = \frac{1}{\text{number of grey levels} - 1}$$
$$\times \sum_{i=1}^{cols}\sum_{j=1}^{rows} (i-j)^2 M(i,j) \qquad (4)$$

$$Correlation = \frac{1}{\sigma_I \sigma_J}\sum_{i=1}^{cols}\sum_{j=1}^{rows}(i-\mu_I)(j-\mu_J)M(i,j) \qquad (5)$$

$$Localhomogeneity$$
$$= \sum_{i=1}^{cols}\sum_{j=1}^{rows} M(i,j)\frac{1}{1+(i-j)^2} \qquad (6)$$

where $M(i,j)$ is the (i,j)th entry in GLCM M with dimension ($cols x rows$), horizontal standard deviation and mean (σ_I, μ_I), and vertical standard deviation and mean (σ_J, μ_J).

Entropy essentially measures the randomness (or information variation) within the GLCM [(2), cf., (1)]. The energy (3) measure represents a scalar quantity invariant with respect to rotations of space and measures textural uniformity (i.e., occurrence of pixel pairs). In contrast, (4) measures the local variations within the image, whereas correlation (5) describes the strength and direction of the linear relationship between the rows and columns of the GLCM. Local homogeneity (6) essentially measures differences in value pairs over the matrix. For further details, see [17] and [18].

In addition to GLCM, we also employ a subset of the Gabor filter texture measures devised in [16]. Gabor filters allow the study of the localized spatial distribution of the texture via a method of image convolution [14] with a given Gabor filter represented as a 2-D convolution kernel with a specified spatial frequency and orientation [19], [20]. As a result, the magnitude of the Gabor filter response identifies varying local texture frequencies and orientations in the image. This is useful for the extraction of more gradual (low-frequency) textures [19] and more generally for use in creating discriminative texture descriptors (e.g., [21]). In this paper, we combine the subset of five GLCM features from [17] computed in two of the four principle horizontal and vertical orientations (i.e., $\{N, E\}$) with a $60°$ orientated Gabor filter empirically matched in scale to the primary road edge within the image. Both GLCM and Gabor

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS

descriptors are computed upon the gray-scale conversion of the original (RGB color) image frame [14]. The use of only the (N, E) GLCM directions is driven from the symmetrical nature of the (N, S) and (W, E) pairings within this visual discriminatory context [17], [18]. The resulting magnitude response image of the Gabor filter in use is itself summarized as a quantized histogram (10 "bins"), mean, standard deviation, and entropy (1) as per the color descriptor of Section III-A. Overall, this results in a 23-value feature vector for each of the three subregions of interest (see Fig. 1), resulting in a 69-D texture feature descriptor for each image frame.

C. Edge-Derived Features

In addition to the general color and texture feature descriptors (i.e., feature vectors) derived for general use over all of the subregions of interest, we additionally define specific edge-based features relating to contour presence in the *road-edge* subregion (see Fig. 1). In this specific sub-region edge information such as edge presence, cohesiveness, consistency and orientation are specific intuitive indicators of off-road and onroad environments [26].

An established approach to the detection of such features is followed using a combination of Hough-transform-driven straight line fitting [22] and contour tracking [23] both operating on Canny edge detector output [24]. The set of edges [24], connected contours [23], and straight line detected [22] is then summarized by the entropy of the resulting Canny edge image, together with the number of contour and straight line instances detected (i.e., three additional edge-derived features specific to the *road-edge* subregion). The parameters for [22] and [24] are empirically chosen in relation to the dimensions of the *road-edge* subregion within the overall image frame (calibration/setup specific).

IV. Feature Classification

The combination of the texture, color, and edge-derived features from Section III gives rise to a 163-D combined feature vector per image frame. This is extracted from the identified subregions of interest within the image (see Fig. 1).

The problem of feature classification over these feature descriptor vectors is then formulated as a supervised machine learning problem [25] with labeled classes for the two- and four-class classification problems. Two classification approaches are utilized to illustrate both the fundamental discriminative properties of the derived feature representation (k-NN) and the potential performance of an optimized noise tolerant classification approach (ANN) [25].

The k-NN approach is one of the simplest available classifier approaches based on classifying given instances (feature vectors) according to the majority vote of the nearest k instances within the N-dimensional feature space (here 163-D space). By contrast, an ANN is a multilayer network of interconnected perceptrons trained to discriminate between specified output classes via a training methodology derived from gradient descent (i.e., backpropagation [25]). As these are established techniques, see [25] for further details.

The set of (manually) labeled examples used for training both of the classification approaches is made up of 800 image frames (200 per class, for four classes). In general, we pose two specific classification problems:

1) four-class problem: classes $= \{off - road, urban, major/trunk\ road, multilane\ motorway/carriageway\}$;
2) two-class problem: classes $= \{off - road, on - road\}$;

where in the latter case the set of training examples for $\{on - road\} = \{\{urban\} \cap \{major/trunk\quad road\} \cap \{multilane\ motorway/carriageway\}\}$ (i.e., the set union of all non-off-road examples from the total set of 800). Although this introduces an inherent bias into the two-class problem toward on-road classification, we empirically show that this appears to successfully counter any affect of in-class variance present within this larger *"on-road"* class.

V. Experimental Results

Initial experimental testing was performed over specifically constructed test video sequences of concatenated 10-s video segments corresponding to different classes. Video sequence 1 (40-s duration) consists (in total) of 10-s segments of each of *{off-road, urban, major/trunk road, multilane motorway/carriageway}*, whereas video sequence 2 (50-s duration) consists of 10-s segments of *{urban, major/trunk road, multilane motorway/carriageway}* and 20-s segments of *{off-road}*. Considering a 30-fps video, this results in approximately 600 test image frames for each of the three *{urban, major/trunk road, multilane motorway/carriageway}* classes (20 s at 25 fps per class) and 900 test image frame for the *{off-road}* class (30 s at 25 fps). Each of the 10-s segments contains multiple varying instances of the specific class in the form of varying road environments (locations). They are designed to test not only per class classification but class transition and the stability of classification within a given road environment as well. Further testing was also carried out over extended sequences based on the optimal parameters identified using these test sequences.
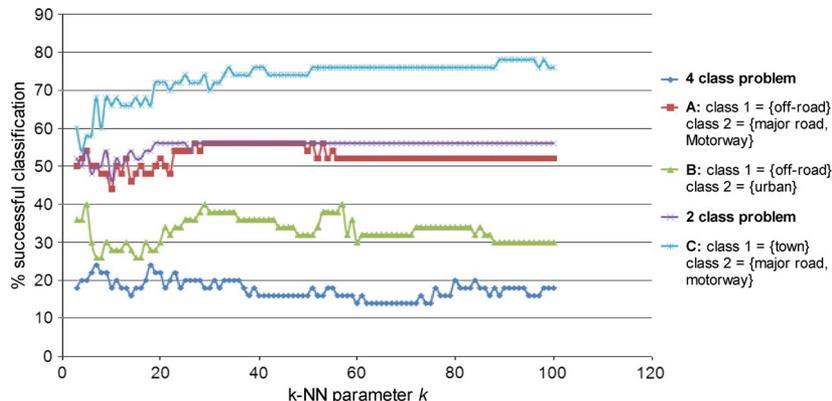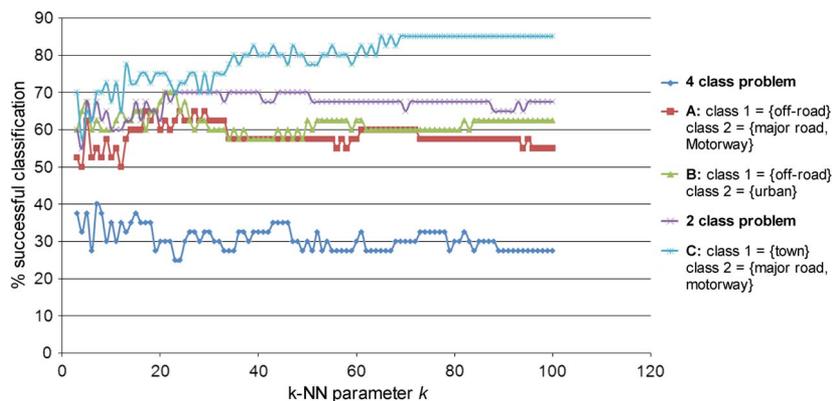
All of the road environments are geographically located within Bedfordshire/Hertfordshire, U.K., and the vehicle speeds are within (or at) the national legal speed limit for the class of road being considered (e.g., speed ≤ 70 mi/h on U.K. motorways for *motorway* class). Off-road speeds are within a (safety) 10- to 20-mi/h range.

The classification results for both k-NN and ANN classification approaches are first reported over both of these independent test sequences.

A. k-NN Classification

The results of k-NN classification are shown in Figs. 2 (for video sequence 1) and 3 (for video sequence 2). In both Figs. 2 and 3, we see the percentage of correct classifications (i.e., true positive) plotted against varying the k-NN query parameter k [25].

From the results, we can see that as k increases the performance of the classification improves for the two-class problem, achieving in general a peak successful classification of around 70% over both sequences (see Figs. 2 and 3). In addition,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TANG AND BRECKON: AUTOMATIC ROAD ENVIRONMENT CLASSIFICATION
5

Fig. 2. Video sequence 1. Classification results using k-NN varying parameter k.

Fig. 3. Video sequence 2. Classification results using k-NN varying parameter k.

classification subproblems posed as A (*{off-road} v {{major}* ∩ *{motorway}}*), B (*{off-road} v {urban}}*, and C (*{urban} v {{major} ∩ {motorway}}*) are also examined for comparison.

In general, we can see that the performance on the specified two-class and individual subproblems A–C (of the four-class problem), which examine the class separation of *off-road* (see Figs. 2 and 3, A and B) and *urban* (see Figs. 2 and 3, C) environment from other classes, outperform the four-class problem. In general, the performance of *urban* class isolation in Figs. 2 and 3 (subproblems B and C) is medium to acceptable with isolation against other on-road classes (i.e., *major/motorway*) performing the best (up to 85% success). The subproblem of combined major roads and motorways against off-road classification performs mildly worse (subproblem A).

Overall, the k-NN results illustrate that 1) class separation is possible within this feature space (e.g., two-class problem results; see Figs. 2 and 3), but in general, 2) the class separation is poor and possibly nonlinear in the dimension of the feature space in use. Classification is generally effected by noise, as expected (causing poor classification), although a subset of training examples (corresponding to certain values of k in k-NN) appears to offer satisfactory results. Linear class separation for the four-class problem appears to be significantly difficult.

B. ANN Classification

To this end, we further investigate with the use of a classifier that is both robust to noise and capable of nonlinear class separation—i.e., an ANN. We employ a classical two-layer network topology with H hidden nodes, one input node per feature vector entry (i.e., 163 inputs), and one output node per class (i.e., 2 or 4) [25]. The ANN is trained using I iterations of the backpropagation algorithm [25].

Considering the general range of parameter $H = \{10, \ldots, 60\}$ and $I = \{150, \ldots, 700\}$, we present a subset of the ANN results over these parameter ranges (for the two- and four-class problems) in Tables II and III based on testing the two multiclass video sequences previously discussed. These tabular results (see Tables II and III) are intended to summarize the general trend observed within the training cycle over varying hidden node H and backpropagation training iterations I.

Considering first the four-class problem results presented in Tables I and II, we can observe a general trend of improved classification performance as the number of hidden nodes (H) in the ANN topology and the number of training iterations (I) increase until a peak (or peaks) of performance is reached (see the entries highlighted in bold in Tables I and II), after which, a fall in classification performance is experienced. This behavior is typical of an ANN classifier [25] and attributable to the increased representational capability (H) and increased learning cycles (I) over the training set prior to the on-set of overfitting (as the number of both hidden nodes and/or training cycles reaches a given level) [25]. The results presented are curtailed prior to the further reduction in performance due to overfitting, and the peak value obtained is presented as the overall representative classification result for each of the

TABLE II
ANN CLASSIFICATION PERFORMANCE ON VIDEO SEQUENCE 1

Hidden Nodes (H)	Training Iterations (I)	4-Class Problem	2-Class Problem
10	150	74%	86%
	200	78%	86%
15	150	74%	**90%**
	200	74%	88%
	300	76%	88%
20	150	68%	88%
	200	68%	86%
30	150	78%	88%
	200	78%	**90%**
	300	78%	80%
40	150	76%	**90%**
	200	76%	88%
	300	74%	86%
	700	72%	88%
50	150	80%	88%
	200	80%	88%
	300	82%	88%
	500	**86%**	88%
	700	**86%**	88%
60	500	82%	88%

TABLE III
ANN CLASSIFICATION PERFORMANCE ON VIDEO SEQUENCE 2

Hidden Nodes (H)	Training Iterations (I)	4-Class Problem	2-Class Problem
10	150	78%	88%
	200	78%	88%
15	150	78%	90%
	200	**80%**	**93%**
	300	**80%**	90%
20	150	73%	90%
	200	73%	**93%**
30	150	75%	88%
	200	75%	85%
	300	70%	83%
40	150	73%	**93%**
	200	73%	**93%**
	300	73%	90%
	700	73%	88%
50	150	78%	90%
	200	75%	88%
	300	75%	90%
	500	75%	88%
	700	75%	88%
60	500	63%	85%

sequences. A consistent peak of 86%/80% is obtained for the two test sequences in use for the four-class problem (see Tables II and III).

For the (semantically simpler) two-class problem, we see a similar behavior of the ANN classifier over varying hidden node topologies (H) and training cycles (I), resulting in several peaks in classification performance at 90%/93% for the two test sequences (see Tables II and III).

Notably, the set of classification results obtained is similarly bounded for over both sequences. In general, taking into account the general performance under different ANN conditions and potential feature bias in one or both of the data sets, we can conservatively conclude an approximate an 80% classification performance rate on the four-class problem and, similarly, an

approximate 90% rate for the two-class problem (see Tables II and III). The ANN outperforms the classification of the earlier k-NN approach on both four- and two-class problems (concretely) and by extrapolation of the four-class problem result, similarly on subproblems A–C.

Further testing over randomly selected video sequences taken from approximately 4–6 h of multiterrain and road environment footage captured as part of the project confirmed a general ANN performance level of ∼80%/90% for the two-/four-class problems over an extended test set. This extended testing was based on the same training set of image frames, as previously outlined for the Table II and III results.

A range of correct road environment examples is shown in Figs. 4 and 5. Here, we see a range of typical road environments and their successful classification using an ANN classifier.

C. Extended Sequence Results

In addition to these initial test sequences, further extended testing was performed over the full set of extended full video sequences (representing the complete set of viable data gathered over several hours in varying environments). This testing was performed using the ANN classification approach based a subset of parameters identified from the analysis on Sequences 1 and 2 (see Tables I and II). These results are presented in Table IV, where we see results consistent with the earlier sequence analysis of Tables II and III. A consistent peak of ∼86% is achieved for the four-class problem and ∼97% for the two-class problem. The mild difference in performance between this and the earlier short sequences is attributable to statistical differences in the duration/size of the test set.

D. Misclassification

As can be seen from the results shown in Tables II–IV misclassifications are also incurred by the illustrated approach, although these are limited. As implied by the earlier analysis using a k-NN classifier on the feature descriptor vectors, a number of feature overlaps occur between the classes in both four- and two-class problem spaces. Here, we examine some of these issues with reference to specific illustrative examples (see Figs. 6–9).

In the four-class problem, we see a misclassification as *off-road* in Fig. 6 due to the possible dominance of the road-side subregion of interest in this example. Similarly, in Fig. 7, a misclassification of an ambiguous case (in terms of the subregion of interest definitions in Fig. 1) appears to contradict the (ground truth) urban locale of this scene. This example suitably illustrates some of the inherent ambiguities on per-image frame classification within this problem domain.

Within the two-class problem, similar problems also exist with misclassification, and two examples are shown in Figs. 8 and 9. Here, it is notable that both have a significant presence of off-road type color/texture within the road-side and road-edge subregions of interest. The equal contribution of all of the subregions in the image frame feature vector and the inherent in-class variation (highlighted by our k-NN results) both contribute to these issues.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TANG AND BRECKON: AUTOMATIC ROAD ENVIRONMENT CLASSIFICATION 7

Fig. 4. Examples of successful ANN classification for four road environments (ANN configuration: $H = 15$ Nodes; $I = 200$).

Fig. 5. Examples of successful ANN classification for two road environments (ANN configuration: $H = 15$ Nodes; $I = 200$).

TABLE IV
ANN CLASSIFICATION PERFORMANCE ON EXTENDED FULL SEQUENCES

Hidden Nodes (H)	Training Iterations (I)	4-Class Problem	2-Class Problem
10	150	85%	**97%**
	200	**86%**	**97%**
15	150	85%	**97%**
	200	**86%**	96%
	300	84%	96%
20	150	84%	**97%**
	200	84%	**97%**
30	150	85%	96%
	200	**86%**	95%

E. Discussion

From the results shown in Tables II–IV (supported by examples Figs. 4–9), it clear that the automatic classification of varying road environments is achievable within a given bound of misclassification accuracy using an ANN-based classification technique. The inherent in-class variance within road environments is illustrated by the prior analysis with k-NN. From the results, better performance is achieved on the two-class {on-road, off-road} problem, which is potentially attributable to both the general feature space separation of these classes within the chosen feature space (or a subset of it) and, more specifically, to the inclusion of specific on-road/off-road edge-derived features (see Section III-C). Consideration

Fig. 6.   Misclassification as "off-road" instead of "major/trunk road."

Fig. 7.   Misclassification as "major/trunk road" instead of "urban."

Fig. 8.   Misclassification as "off-road" instead of "on-road."

of subproblems to the four-class problem (see Figs. 2 and 3) further supports the lesser result achieved on this problem with the ANN. From the misclassification examples presented in Figs. 6–9, it is also clear that this issue of potential per-frame environment ambiguity contributed by one or more of the (equally weighted) subregions of interest may also contribute substantially to the misclassifications suffered. The considera-

Fig. 9.   Misclassification as "off-road" instead of "on-road."

tion of alternative classification approaches, optimization of the selected subregions of interest, and selection of additional more discriminative features is an area for future work.

As presented, the approach detailed operates in near real time at approximately 1 Hz on an AMD Turion-64 laptop computer (software running as 32-bit on a single CPU). This computational limitation is attributable to the calculation of the texture feature descriptor elements of the overall feature vector used for classification. Input video frames are 640×480 resolution and subsampled (at 1 Hz) from a 30-fps video feed from a standard digital video camera (Sony Cybershot W-210) mounted onto the windscreen/shield. A rubber suction mount, in addition to the standard anti-vibration mounting of the windscreen itself, eliminates the majority of high-frequency vibration within the resulting video input. The approach has been demonstrated operating at 1 Hz (i.e., one frame classification per second) from video taken using an on-vehicle forward-facing camera.

VI. CONCLUSION

A methodology for the near real-time classification of the road environment has been presented based on the use of a combined color and texture feature vector extracted from multiple subregions of a forward-facing on-vehicle camera view. An ANN classifier gives ∼90%–97% successful classification for the two-class *on-road* and *off-road* determination problem, whereas a lesser ∼80%–85% result for the more complex *{off-road, urban, major road, motorway}* four-class determination problem. Prior analysis using a k-NN classifier implies the inherent feature overlap within the current feature space and the resulting difficulty of the classification problem itself—not in the least due to ground truth ambiguity for any given frame outside of the temporal context of the sequence (e.g., Fig. 7).

Future work will further investigate the use of temporal analysis over multiple frames and the effect of varying weather and lighting conditions on performance. In addition, principle-component-analysis-based feature optimization, alternative computationally efficient texture measures, and subregion optimization will all be considered as a conduit to full real-time performance. Further evaluation over a wider variation of international on-road and off-road environments can also be investigated.

Acknowledgment

References

[1] M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: State of the art and perspectives," *Robot. Auton. Syst.*, vol. 32, no. 1, pp. 1–16, Jul. 2000.

[2] M. L. Eichner and T. P. Breckon, "Integrated speed limit detection and recognition from real-time video," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 626–631.

[3] P. Jansen, W. van der Mark, J. C. van den Heuvel, and F. C. A. Groen, "Colour based off-road environment and terrain type classification," in *Proc. Int. Conf. Intell. Transp. Syst.*, 2005, pp. 216–221.

[4] D. Ferguson, M. Darms, C. Urmson, and S. Kolski, "Detection, prediction, and avoidance of dynamic obstacles in urban environments," in *Proc. Intell. Veh. Symp.*, 2008, pp. 1149–1154.

[5] A. Discant, A. Rogozan, C. Rusu, and A. Bensrhair, "Sensors for obstacle detection—A survey," in *Proc. 30th Int. Spring Semin. Electron. Technol.*, 2007, pp. 100–105.

[6] S. Beucher and X. Yu, "Road recognition in complex traffic situations," in *Proc. 7th IFAC/IFORS Symp. Transp. Syst.: Theory Appl. Adv. Technol.*, 1994, pp. 413–418.

[7] C. Fernandez-Maloigne and W. Bonnet, "Texture and neural network for road segmentation," in *Proc. Intell. Veh. Symp.*, 1995, pp. 344–349.

[8] J. Wu, Z. Yang, J. Wu, and A. Liu, "Virtual line group based video vehicle detection algorithm utilizing both luminance and chrominance," in *Proc. Conf. Ind. Electron. Appl.*, May 2007, pp. 2854–2858.

[9] Y. Shan, F. Yang, and R. Wang, "Color space selection for moving shadow elimination," in *Proc. Int. Conf. Image Graph.*, 2007, pp. 496–501.

[10] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with HSV color information," in *Proc. IEEE Intell. Transp. Syst.*, 2001, pp. 334–339.

[11] L. M. Lorigo, R. A. Brooks, and W. E. L. Grimsou, "Visually-guided obstacle avoidance in unstructured environments," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 1997, vol. 1, pp. 373–379.

[12] I. Katramados, S. Crumpler, and T. P. Breckon, "Real-time traversable surface detection by colour space fusion and temporal analysis," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2009, pp. 265–274.

[13] S. Kewei, F. Xitian, C. Anni, and S. Jingao, "Automatic face segmentation in YCrCb images," in *Proc. Asia-Pacific Conf. Commun., 4th Optoelectron. Commun.*, 1999, vol. 2, pp. 916–919.

[14] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[15] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[16] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, no. 10, pp. 847–856, 1980.

[17] V. Arvis, C. Debain, M. Berducat, and A. Benassi, "Generalization of the cooccurrence matrix for colour images: Application to colour texture classification," *Image Anal. Stereology*, vol. 23, pp. 63–72, 2004.

[18] J. J. van der Sanden and D. H. Hoekman, "Review of relationships between grey-tone co-occurrence, semivariance and autocorrelation based image texture analysis approaches," *Can. J. Remote Sens.*, vol. 38, no. 3, pp. 207–213, 2005.

[19] A. Jain, N. Ratha, and S. Lakshmanan, "Object detection using Gabor filters," *Pattern Recognit.*, vol. 30, no. 2, pp. 295–309, Feb. 1997.

[20] T. P. Weldon, W. E. Higgins, and D. F. Dunn, "Gabor filter design for multiple texture segmentation," *Opt. Eng.*, vol. 35, no. 10, pp. 2852–2863, 1996.

[21] C. Wolf, J. Jolion, W. Kropatsch, and H. Bischof, "Content based image retrieval using interest points and texture features," in *Proc. Int. Conf. Pattern Recog.*, 2000, vol. 4, pp. 234–237.

[22] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[23] D. Douglas and T. Peucker, "Algorithms for the reduction of the number of points required for represent a digitized line or its caricature," *Can. Cartographer*, vol. 10, pp. 112–122, 1973.

[24] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[25] T. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[26] A. Broggi, "Vision-based road detection in automotive systems: A real-time expectation-driven approach," *J. Artif. Intell. Res.*, vol. 3, no. 1, pp. 325–348, Jun. 1995.

[27] A. Ess, T. Müller, H. Grabner, and L. Van Gool, "Segmentation-based urban traffic scene understanding," in *Proc. Brit. Mach. Vis. Conf.*, 2009.

[28] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, Apr. 2008.

[29] D. Gokalp and S. Aksoy, "Scene classification using bag-of-regions representations," in *Proc. Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.

[30] N. Rasiwasia and N. Vasconcelos, "Scene classification with low dimensional semantic spaces and weak supervision," in *Proc. Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–6.

[31] J. Melo, A. Naftel, A. Bernardino, and J. Santos-Victor, "Detection and classification of highway lanes using vehicle motion trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 2, pp. 188–200, Jun. 2006.

Isabelle Tang received the M.Sc. degree in computational and software techniques (specializing in digital signal and image processing) from Cranfield University, Bedfordshire, U.K., in 2008 and the Diploma degree in information systems engineering from INSA de Rouen, Rouen, France, in 2008.

She previously worked in the area of automatic road sign detection (VALEO, France) and currently works as a Systems Engineer for information systems support for the telecommunications industry (France). She is currently with INSA de Rouen.

Toby P. Breckon received the B.Sc. degree (Hons.) in artificial intelligence and computer science and the Ph.D. degree in informatics from the University of Edinburgh, Edinburgh, U.K., in 2002 and 2006, respectively.

He has held visiting positions with Northwestern Polytechnical University, Xi'an, China, and Waseda University, Tokyo, Japan. He is currently a Senior Lecturer with the School of Engineering, Cranfield University, Bedfordshire, U.K. In addition, he is a visiting member of faculty with Ecole Supérieure des Technologies Industrielles Avancées, France. His key research interests, in the domain of computer vision and robotics, are as follows: 3-D sensing and reasoning, 3-D visual completion, vision in built environments, sensor fusion, visual surveillance, and robotic deployment in hazardous environments.

Dr. Breckon is a Chartered Engineer and a member of the IET. In addition, he is an Accredited Imaging Scientist and an Associate of the Royal Photographic Society. He led the development of image-based automatic threat detection for the 2008 U.K. MoD Grand Challenge winning team [R.J. Mitchell Trophy, (2008), IET Award for Innovation (2009)]. He has a range of publications and leads several funded research projects in applied image processing and computer vision.