Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems, Washington, DC, October 3 - 6, 2004.

Adaptive Real-Time Road Detection Using Neural Networks

Mike Foedisch

Intelligent Systems Division
National Institute of
Standards and Technology
Gaithersburg, MD 20899
mike.foedisch@nist.gov

Aya Takeuchi

Advanced Technology
and Research, Corp.
Burtonsville, MD 20866
ayako.takeuchi@nist.gov

Abstract

We have developed an adaptive real-time road detection application based on Neural Networks for autonomous driving. By taking advantage of the unique structure in road images, the network training can be processed while the system is running. The algorithm employs color features derived from color histograms. We have focused on the automatic adaptation of the system, which has reduced manual road annotations by human.

1 Introduction

The Intelligent Systems Division of the National Institute of Standards and Technology has been engaged for several years in developing real-time systems for autonomous driving. Road detection systems have been developed using a number of different algorithms, but real-time processing is always the key. Algorithms for paved road detection [2, 4] have been extensively developed. Most are based on detecting road markings and lanes and cannot be applied to roads without any markings or lanes.

We present a real-time road detection application using Neural Networks [7], which is independent of road markings and lanes. During a short initialization step, feature data is automatically collected based on the conforming road structure in images captured from the driver's point of view (road images). Then, the new Neural Network is trained and applied to the new environment. This procedure allows the system to detect road adaptively.

In this paper, we will discuss the Neural Network-based road detection in Section 2. Section 3 explains the adaptability and real-time processing of the system for detecting various road and terrain environments. Section 4 gives details of the system implementation followed by the results in Section 5 and future work in Section 6.

2 Overview of the Neural Network Based Road Detection

Our adaptive real-time Neural Network-based road detection approach is an enhancement of the work done by Rasmussen [6] and Conrad [3]. Their approach consists of two steps: network training and the application of the network to road detection.

2.1 Training of the Neural Network

Figure 1 gives an overview of the Neural Network training step. From a number of road images, which are acquired as RGB images, features are extracted. As introduced in [6], an "independent" color histogram consisting of 8 bins per channel is used [9]. Additionally, we put the normalized x and y position values of the current point of consideration into the set of features. As shown in [3], the integration of x and y has improved classification results. Each feature vector is then manually annotated as either road or non-road area in the image.

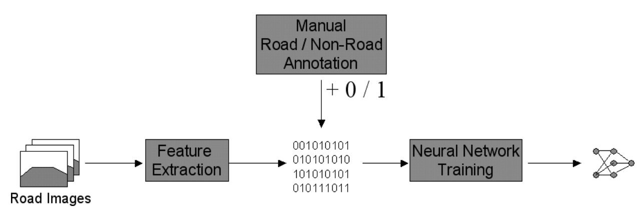


Figure 1: Neural Network Training Overview

In the next step, a Neural Network is trained based on the extracted set of feature vectors and the annotated road/non-road labels as targets.

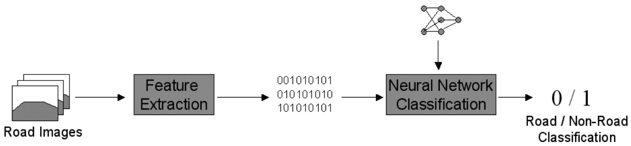


Figure 2: Neural Network based Road Detection Overview

2.2 Road Detection

Figure 2 shows the structure of the road detection algorithm. Each extracted feature vector is processed by the Neural Network, which was trained in the previous step. Finally, the Network computes the result value, which is interpreted as either the road class or non-road class. Figure 3 depicts a sample result where the area classified as road is drawn with white squares and the area classified as non-road shown with black dots.

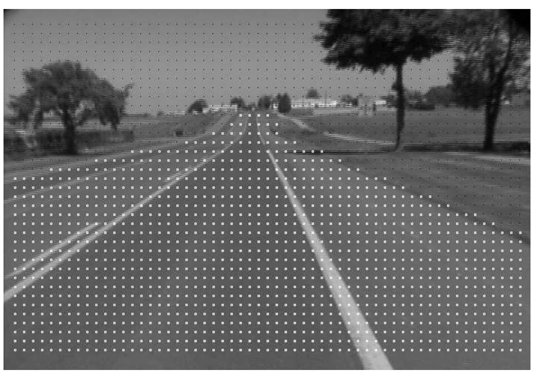


Figure 3: Sample Result of the Road Detection

Although the road detection step using an existing network is processed in real-time, the system does not have the ability to adapt to new environments. Currently, the system requires manual annotation by humans, which halts the dynamic real-time processing.

3 Adaptive Real-Time Approach

This section discusses our solution to avoid the manual annotation of images by taking advantage of the conforming structure in the road images.

3.1 Conceptual Road Structure

The camera used as the sensor for detecting roads is positioned at the driver's viewpoint. The road structure in the images captured from this perspective typically forms the shape as illustrated in Figure 4. Each road image consists of the road area and non-road areas as well as named borders between these areas, e.g. road boundary and horizon. For the approach described here, we take advantage of the fact that the road usually covers a trapezoidal area, which is centered in the lower part of the image.

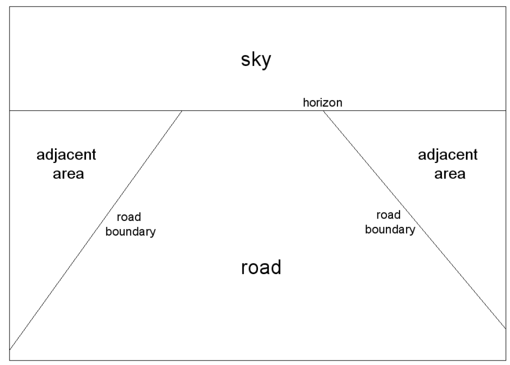


Figure 4: Conceptual Areas in Road Images

Based on the estimated road location in the image, feature vectors are collected from these areas and automatically labeled as either road or non-road (positive or negative areas respectively). Figure 5 shows three "positive windows" with white rectangles (the road) and "negative windows" with black and white rectangles (non-road areas).

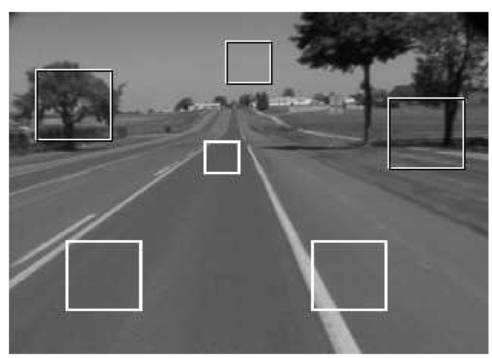


Figure 5: Areas of Positive and Negative Windows

3.2 Adaptive Feature Vector Acquisition

Figure 6 describes the improved process flow. The pre-defined positive and negative windows are used as filters, which are labeled as road or non-road area. Feature vectors are processed from these windows and annotated using the label of the window.

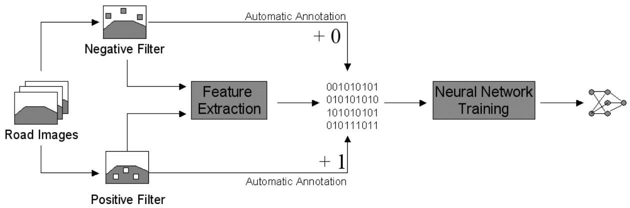


Figure 6: Neural Network Training with Automatic Annotation

3.3 Issues With Automatic Feature Acquisition

The acquisition of feature data can now be processed in an automatic way, e.g. the images can be taken every *n* frames. However, the road image does not always follow the structure described in Section 3.1. Figure 7 is a camera view when the vehicle is turning a corner. One of the positive windows covers non-road content and one of the negative windows covers partly road and non-road content. To let the Neural Network learn distinguished data for positive and negative areas, such contradicting data has to be avoided. Therefore, automatic acquisition of feature data cannot be applied all the time. As a result we allow the user to supervise and control the acquisition task.

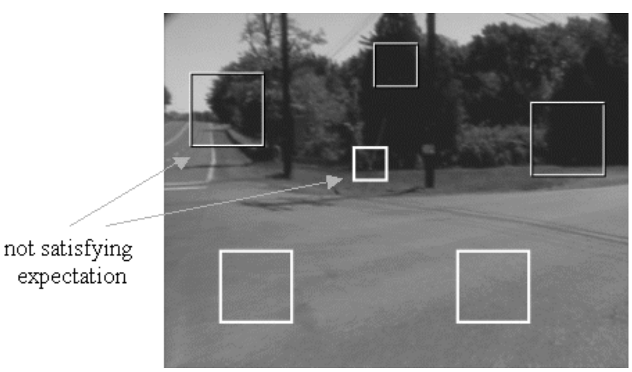


Figure 7: In certain situations windows might move out of their assumed content

After the Neural Network is trained over the collected feature data, it is applied in real-time to road detection as described in Section 2.2.

3.4 Post Processing

Road segmentation with Neural Network provides relatively reliable results in a real-time environment. However, the accuracy is enhanced with the following two steps as post processing. First, noise is reduced by erosion and dilation methods. Second, since the road does not change its shape drastically between two consecutive images, the road segmentation information of previously processed images (approximately prior to 8 images) can assist to detect false positive areas to construct the current road shape. After each frame has been processed, the value of each pixel (-1 for non-road area and 1 for road area) is added in the matrix. Each index in this matrix contains a value between –3 (most negative area) and 5 (most positive area). The result is displayed (e.g. Figure 3) based on the value in the matrix: positive values are represented as road, and negative as non-road.

4 Implementation

The application was implemented on a Dell Latitude laptop computer with 2.2 GHz Intel processor and Red Hat 9 Linux operating system[1]. The camera is a Unibrain Fire-i (320x240 pixels) firewire camera, which produces 30 frames per second. The application was written in C++ using a C++ based Neural Network library [8]. The Neural Network receives 26 inputs (24 RGB histogram bins plus *x* and *y* coordinates) and consists of three layers. The first two layers contain four neurons each. The last layer is composed of one neuron, which provides the output. The Neural Network employs back propagation learning [7].

In order to allow the process to run in real time, the application extracts every other pixel for processing, giving a 160x120 image. Moreover, for feature extraction samples are taken from every third pixel and comprise 7x7 pixel windows around the pixel. In other words, a 7x7 pixel area is treated as one unit window, which produces one set of input features (24 RGB color histogram bin), and which overlaps the next window by 3 pixels. The size of this window was determined for optimization purposes by the number of iterations and the size of the window unit. Currently, the processing time of road segmentation is approximately 60 milliseconds per frame.

[1] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily best for the purpose.

5 Results

We compared the system's performance with manually annotated frames of video files in order to measure the accuracy. This allowed us to compute the false positive and false negative ratios. False positives refer to actual non-road areas in the image, which were classified by the system as road, while false negatives refer to actual road areas classified as non-road.

Followings are the results of four videos with different road scenes processed by the system. Each section consists of four images and a graph. We trained the Neural Network on selected frames of the first four seconds of the video represented by the first image. The other three images are captured during the course of process.

The results are depicted by a graph in which we show the classification errors – false negatives, false positives and the sum of both – for every 25th frame of the video.

5.1 Example "Straight Road"

The road scene of this example (see Figure 8) consists mainly of a straight road showing the same appearance regarding pavement and lane markings. Some parts of the video show shadows on the road.

Figure 8: Sample Frames for Example "Straight Road"

The Neural Network training frames did not contain any shadows on the road. The classification results summarized in Figure 9 show a very low misclassification rate for most parts of the video.

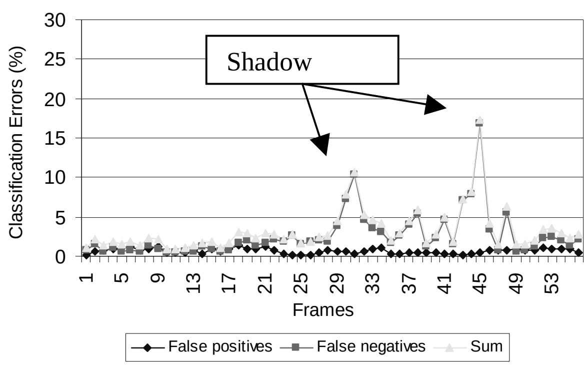


Figure 9: Sample Frames and Classification Results for Straight Road Driving Scene

Overall, the two peaks are mainly due to false negatives. This is caused by shadow on the road, which was not trained by the Neural Network and is therefore classified as non-road.

5.2 Example "Road with Shadow"

The Neural Network applied for this scene was trained on frames where some road areas are covered with shadow (see Figure 10). This condition consistently appears throughout the video.

Figure 10: Sample Frames for Example "Road with Shadow"

A rise of the overall classification error as seen in Figure 11 is mainly due to a higher false positive rate.

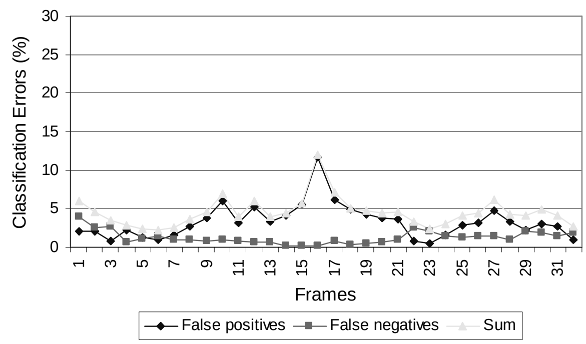


Figure 11: Sample Frames and Classification Results for Shadow Road Scene

Since the dark areas of shadow on the road and in non-road areas appear similar, the system tends to misclassify dark non-road samples as road.

5.3 Example "Slight Changes in Environment"

The road in this video follows a curvy course but is still simple in structure (see Figure 12).

Figure 12: Sample Frames for Example "Slight Chances in Environment"

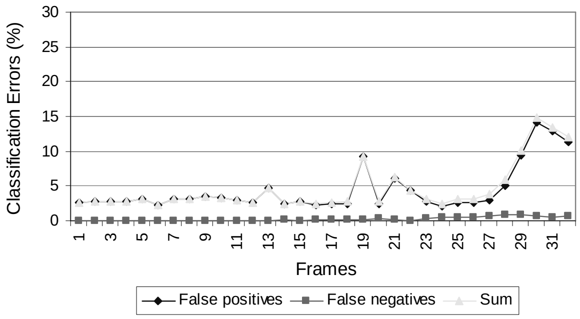


Figure 13: Sample Frames and Classification Results for a Scene with slight Changes

Smooth changes occur in the environment (non-road area) while the road itself appears similar. These changes and the changing direction of the sun cause an increase in false positives over the time.

5.4 Example “Road and Environment with drastic Variations”

In this video, the environment and the road itself change drastically (see Figure 14).

Figure 14: Sample Frames for Example "Road and Environment with drastic Variations"

The graph in Figure 15 shows frequent changes in false positives as well as false negatives.

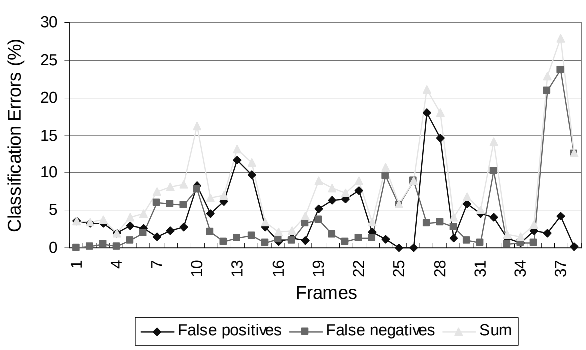
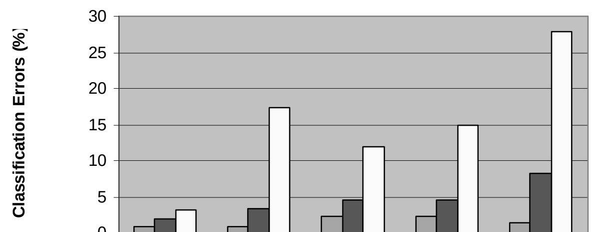


Figure 15: Sample Frames and Classification Results for a Scene with massive Changes

Since the Neural Network was trained in the beginning of the course, where it was exposed to limited types of road and scenery, variations of the road (e.g. concrete, regular pavement or gravel) cause higher false negative rates. Higher false positives rates are due to changes in the environment (buildings, trees etc.).

5.5 Summery

Figure 16 gives an overview of the minimum, average and maximum classification errors for the processed road scene samples. The left-most column represents the result from the straight road scene with no shadows only. The misclassification rate of this part is lower than the column second from left, which depicts the result for the entire straight road scene.



	Straight Road (No Shadow)	Straight Road	Shadow	Slight Changes	Massive Changes
Min	0.93506494	0.93506494	2.23376623	2.28571429	1.4025974
Average	1.9554731	3.27272727	4.51948052	4.62987013	8.21872864
Max	3.11688312	17.3506494	11.9480519	14.8571429	27.8961039

Figure 16: Overview over minimum, average and maximum classification errors.

Our approach proves less than 5 % misclassifications for the road with slight changes in environments (as shown in the left-most column), and can even track robustly roads with non-homogenous appearance (as in the shadow example).
In general, the classification error increases over time as new environments are entered. Variations in the non-road areas lead to more false positives and changes in the road area lead to more false negatives.

6 Future Work

Although the above-mentioned system adapts to new environments quickly, the Neural Network might become old as the vehicle proceeds into new environments. Constant update of the network with new feature data is desirable to maintain accurate road tracking as the vehicle proceeds and the terrain and sunlight change. Currently the Neural Network training requires approximately 600 milliseconds. In order to train the network constantly without halting the road detection process, parallel processing by two computers becomes the solution. Neutral Message Language [5] is employed for real-time data exchange between the computers.

We are currently working on two issues of adaptive real-time learning:

1. For the purpose of the integration of new feature data and the deletion of old data, our first approach employs a buffer working as a first-in first-out queue. Feature data put into the queue first will be removed first when the buffer reaches its capacity.
2. A solution must be found to the problem mentioned in Section 3.3. The solution can be either to hide wrongly positioned windows or to dynamically move the windows to ensure that they always cover the desired content of the image. Additionally, we are evaluating the use of motion information, which can be used to distinguish between driving a straight road and turning.

With the adaptive real-time learning approach, we expect the system to be more adaptive and less sensitive to the changes in the environment.

Another issue is the recognition of oncoming traffic which has been relatively successfully detected as non-road. To expand the application and to detect traffic as moving obstacles, several algorithms have been evaluated. Since the camera itself is moving and nothing stands still in the consecutive images, motion analysis based on color optical flow [1] algorithms will most likely be suitable for this task.

Acknowledgments

Thanks to Christopher Rasmussen, who originally developed the feature extraction system and determined the methods for SVM and neural Network generation, and Patrick Conrad, who has enhanced the system by adding more features. We are also grateful for the support of the Army Research Laboratory and the Defense Advanced Research Projects Agency MARS program.

References

[1] J. Barron and R. Klette, “Quantitative Color Optical Flow”, in Proc. *IEEE Int. Conf. 16th ICPR*, Volume: 4, 251 -255, 2002.

[2] M. Bertozzi, A. Broggi and A. Fascioli, "Vision-based Intelligent Vehicles: State of the Art and Perspectives," in *Robotics and Autonomous Systems* 32, 2000.

[3] P. Conrad and M. Foedisch, “Performance Evaluation of Color Based Road Detection Using Neural Nets and Support Vector Machines”, in Proc. *Applied Imagery Pattern Recognition Workshop*, Washington D.C., 2003.

[4] E. Dickmanns, "Vehicles capable of dynamic vision", *Proc. Int. Joint Conf. On Artificial Intelligence*, pp. 1577-1592, 1997.

[5] V. Gazi, M.L. Moore, K.M. Passino, W.P. Shackleford, F.M. Proctor and J.S. Albus, “The RCS Handbook”, Wiley Interscience, 2001.

[6] C. Rasmussen, “Combining Laser Range, Color, and Texture Cues for Autonomous Road Following”, *in Proc. IEEE Inter. Conf. on Robotics & Automation*, Washington D.C., 2002.

[7] B. Ripley, “Pattern Recognition and Neural Networks”, Cambridge University Press, 1996.

[8] A. Schmidt, “A modular neural Network architecture with additional generalization abilities for high dimensional input vectors” Master’s thesis Manchester Metropolitan University, September 1996.

[9] M.J. Swain and D.H. Ballard, “Color Indexing”, *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.