

# A Road Sign Detection and Recognition System for Mobile Devices

Bo Xiong, Ozgur Izmirli  
Department of Computer Science, Connecticut College  
Email{bxiong,oizm}@conncoll.edu

## ABSTRACT

We present an automatic road sign detection and recognition service system for mobile devices. The system is based on a client-server architecture which allows mobile users to take pictures of road signs and request detection and recognition service from a centralized server for processing. The preprocessing, detection and recognition take place at the server end and consequently, the result is sent back to the mobile device. For road sign detection, we use particular color features calculated from the input image. Recognition is implemented using a neural network based on normalized color histogram features. We report on the effects of various parameters on recognition accuracy. Our results demonstrate that the system can provide an efficient framework for locale-dependent road sign recognition with multilingual support.

**Keywords:** Road Sign Detection, Road Sign Recognition, Machine Learning

## 1. INTRODUCTION

Automatic sign recognition can help orient people in unfamiliar environments. These systems can be utilized in many situations such as humanitarian rescue missions and signage translation for way finding. Automatic sign recognition involves two processes: (1) segmentation and detection of the sign within a larger image and (2) lighting, shadow and background invariant recognition of the sign. One application of this general problem is road sign detection which has been an important and active research area in both computer vision and artificial intelligence. Road signs provide drivers with vital information such as the speed limit, state of the road and direction in order to guide, orient and warn them. Correct interpretation of the road signs is crucial to the drivers' and pedestrians' safety. Automatic road sign detection and recognition are also essential to robotic vehicles that can drive automatically on the road.

In this paper, we present a fast mobile recognition system for road signs that, in general, can be useful in any signage related task. It may serve the needs of pedestrians and drivers in a foreign country or help student drivers. It may also be utilized in automotive warning systems. We define an easy to use sign recognition service for mobile users. A pedestrian or driver can take a snapshot of the road sign with the mobile phone. The phone will then preprocess the image by reducing the image resolution to save bandwidth and send the data to a remote central server over the wireless data network. After receiving the request with the image, the remote central server recognizes the road sign in the database specific to that country or region and sends the result back to the phone. Independent of the locale, the recognition result can be displayed in the chosen language. The user can verify the accuracy of the recognition by checking the returned sign prototype that is kept in the server's database as well as retransmit the query by manually framing the sign in question for a second try.

In our study, we used an Android mobile device as the client side due to its capabilities to capture video, images and geo-location information as well as perform simple processing of image data and communicate with a remote server through its wireless data network. There are several advantages to using this framework. The storage capacity and processing power of mobile phones are limited and applications need to be efficient in terms of power consumption. A remote server can perform the heavy processing to detect and recognize road signs. This can lift significant computational limitations on the mobile phones and leverage the quality of mobile services in general. This framework also has the advantage of employing a centralized database and associated recognition system which enables algorithmic improvements to its learning system as well as adaptation based on new data from incoming requests.

Smart phones are now ubiquitous and serve as low cost mobile information acquisition and transmission devices. Other researchers have developed similar systems that use the phone as a visual input device and the server to store the database and process data for search applications. For example, Tsai et al. [1] presented a mobile product recognition system. Chen et al. [2] presented a mobile system that retrieves information about books on a bookshelf.

### 1.1 System Overview

In our system, the user takes a picture of the road sign with the Android device and sends the picture to the server for recognition. Before sending the picture to the server, the user can optionally mark the region of the road sign in the picture in order to reduce the search space, save bandwidth and improve efficiency as well as accuracy.

In order to recognize the road sign, two steps are performed on the server: detection and recognition. The detection stage finds the position of the road sign in the picture and then extracts features from the road sign from a reduced resolution version of the image. The recognition stage uses the extracted features to recognize the road signs in the database. Finally, the server sends the result to the phone which displays the result. Our system also supports multiple languages and can be configured to use different databases according to the current locale determined based on geo-location data provided by the phone.

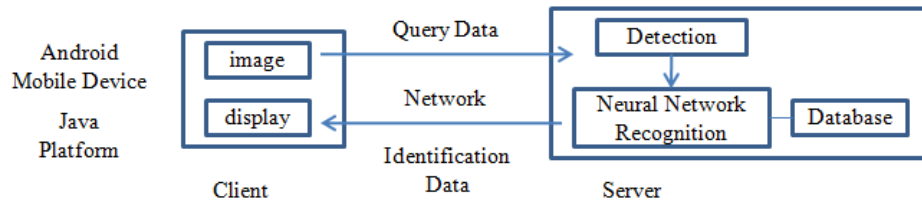


Figure 1: A client-server mobile recognition system for road sign detection and recognition

### 1.2 Related Work

Road sign detection is a challenging problem because it involves identifying an object from a generally unknown background. Researchers have developed different approaches to solve this problem. Some researchers use color information of the road sign to detect the position. For example, Hsien et al. [4] presented a method that uses the projection of specific colors of the road signs to detect their position in the image. They used the HSV color space. Hsu and Huang [3] presented a method that finds the relative position of a road sign in the original image by using a priori knowledge, shape and color information to capture a closer view image. Lorsakul and Suthakorn [6] used edge detection in order to find the position of the road sign.

For road sign recognition, most commonly used methods are template matching, Hidden Markov Models, neural networks and Support Vector Machines. In [6] and [7], the authors trained neural networks with vertical and horizontal grayscale histograms of the road sign images as features. Hsien et al. proposed a system that recognizes the road signs with a Hidden Markov Model. Another group of researchers [8] used support vector machines to classify the shapes of the road signs.

### 1.3 Database Description

In our study, we collected our own images of road signs using an Android phone and augmented the data set with images obtained from Google Maps Street View. We gathered 386 images spanning 18 different road signs. We trained and tested a neural network using this database. We performed cross validation using three quarters of the set to train and the remainder to test for all folds. Figure 2 shows a sampling of the road signs used in our study.



Figure 2: Database of the road signs used. Only one example from each type is shown.

## 2. ROAD SIGN DETECTION

In this section we describe the algorithm we use to detect and extract the road sign from the original image. We use a similar method to the one proposed in [4] to detect road signs. In their work, they first use the specific HSV color projection to detect the position of the road sign and extract the sign from the original image. Secondly, the extracted sign is normalized to the size of 96 by 96 pixels. Finally, they replace the background with magenta, a color not usually present in road signs.

Our detection algorithm uses RGB color ratio features instead of HSV color space ranges to detect the position of the road sign. Our database consists of signs that have a variety of colors including red, yellow and blue. We set constraints on the ratios of color intensities R/G, R/B, G/B in order to detect the color in the image. The thresholds for color ratios vary according to the target color. Once the image is detected, it is resized to N by N pixels (typically 40x40) and the background pixels are replaced with black. This is a more suitable representation since we do not want any contribution from the background pixels to the features used for recognition.

When we look at the vertical and horizontal projections of these features (See Figure 3), we observe that they have continuous characteristics. Thus, we can use this property to detect the position of the road sign. We detect the road sign based on continuity in both the vertical and horizontal projections. To find the horizontal position we first apply a set of color ratio constraints to all pixels in the image in order to obtain a binary image. We obtain the vertical projection by counting the number of pixels with value 1 for each column. Note that the index of the column corresponds to the horizontal position in the image. Thus, position  $x$  can be selected as a starting point of a sign candidate provided that the number of pixels with value 1 starting at the  $x^{\text{th}}$  column is over a threshold ( $1/20$  of the image width in our study). The region for the sign in the image is chosen by the longest spread of the particular projection feature. That is, we would like to ignore small objects and find the largest object that conforms to the properties being sought. The collection of the consecutive columns after the  $x^{\text{th}}$  column with 1s is a candidate for the road sign in the horizontal direction. Let  $C_i(x)$  represent the candidate which is the function equal to the number of consecutive non-zero columns starting at column  $x$  for color index  $i$ . We choose the horizontal starting position,  $p_{h,i} = \arg \max C_i(x)$ , and the width of the sign  $n_{h,i} = \max C_i(x)$ . Similarly, the vertical position and height,  $(p_{v,i}, n_{v,i})$ , are found independently. Since the road signs have three different colors, we apply the above procedure for all three colors and each time with different color ratio constraints to select the target colors. Then we pick the candidate that has the largest sum  $n_{h,i} + n_{v,i}$ , simultaneously determining the effective width, height and dominant color.

Next, we extract the sign from the original image using the delineating points found above and normalize the image size to N by N pixels. Since the images of the road signs are taken from real road signs they contain a variety of backgrounds and we observed that using the images in this form reduced the accuracy of the classifier. We therefore replace the background with black pixels to make the representation of all signs equitable. More precisely, we scan the image from left to right and override the pixels with black until we find a pixel that has the dominant color of the road sign as determined above. We then repeat the process for a right to left scan. This method is very computationally inexpensive compared to other algorithms such as the flood-fill algorithm.

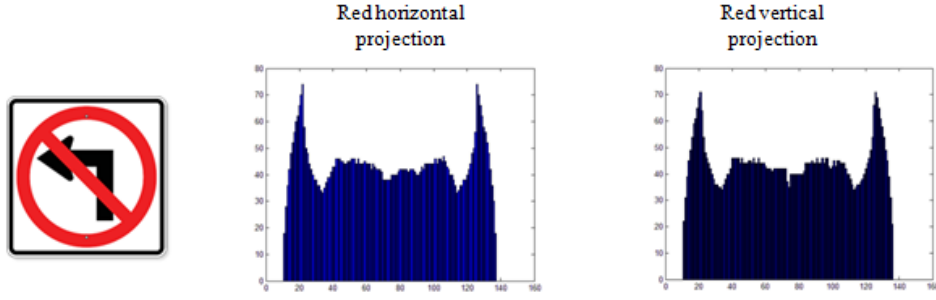


Figure 3: Horizontal and vertical projection for road sign

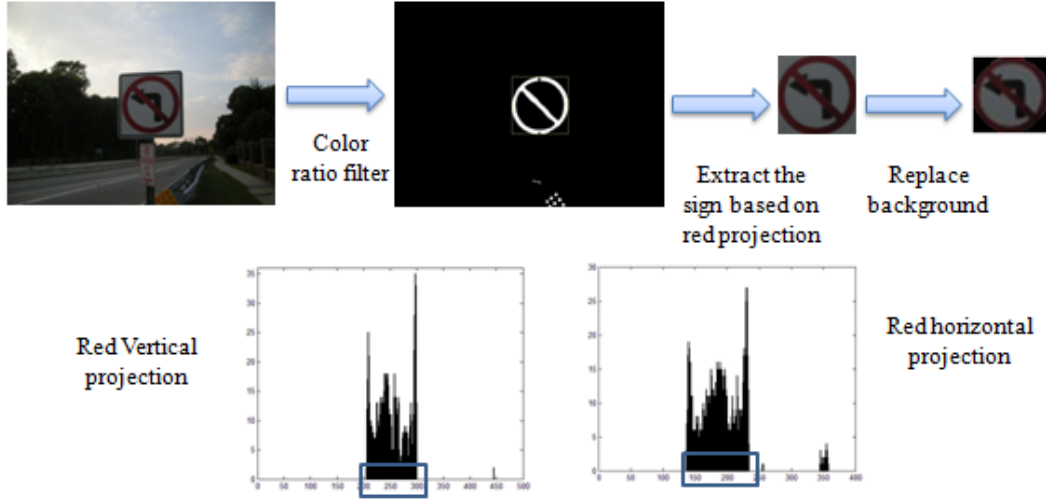


Figure 4: Stages of the detection algorithm: color ratio filter, sign extraction and background replacement.

### 3. ROAD SIGN RECOGNITION

In this section, we describe the algorithm for road sign recognition. As a result of the detection stage outlined in the preceding section, we obtain an extracted road sign image with a black background. Then, we train a neural network for road sign recognition with the training set ensuring that all signs have been segmented correctly. Other researchers [6], [7] have also trained neural networks for road sign recognition. In their work, they used 3 normalized average maximum pixel values and vertical and horizontal histograms of the gray scale representation of the road sign as feature vectors. However, we found that the RGB representation of the road sign outperforms gray scale representation.

We implement a feed-forward artificial neural network, more specifically; multi-layer perceptions (MLP), the most commonly used type of neural network. Our MLP consists of an input layer, one hidden layer and an output layer. Each layer of MLP includes neurons that are directionally linked with the neurons from the previous and next layer. All the neurons in MLP are similar. Each neuron takes the output values from several neurons in the previous layer through its inputs and passes the response to neurons in the next layer. A weighted sum is calculated from the values obtained from the previous layer and a bias term is added. Finally, the sum is transformed using the activation function. In our study, we use the sigmoid function  $f(x) = (1 - e^{-x}) / (1 + e^{-x})$  as the activation function. For training the neural network, we use the random sequential back-propagation algorithm.

In our database, we have 18 different types of road signs and therefore we have the same number of output nodes of the neural network in which each node represents a different road sign. For the input layer of the neural network, we have

6xN nodes (for an NxN input image). We construct a vertical histogram and a horizontal histogram for each of the three RGB values. The input nodes consist of 3xN elements taken from the vertical histograms and 3xN elements taken from the horizontal histograms.

Since the lighting conditions from one image to the other vary significantly, we need to normalize the RGB values. Before feeding the values into the neural network we divide the sum of the histogram by the sum of the total RGB values. Let T be the sum of the RGB values of the entire image. Let R be the matrix of red values of the extracted road sign and let  $R_{i,j}$  denote the red value of the pixel in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column. Similarly, we define G and B for green and blue values of the image. Let I denote the vector representation of the road sign image features. It consists of the concatenation of the six normalized histograms resulting in a length of 6xN elements.

The calculation of the red histogram features are shown in the formula below. The blue and green histogram features are found in a similar manner and placed at locations starting at 2N+1, 3N+1 etc.

$$I_i = \frac{1}{T} \sum_{j=1}^N R_{i,j} \quad 1 \leq i \leq N \quad \quad I_j = \frac{1}{T} \sum_{i=1}^N R_{i,j-N} \quad (N+1) \leq j \leq 2N$$

## 4. RESULTS

In this section, we present the recognition accuracy of the system described above. Two important parameters are normalization size and the number of hidden layer neurons. The size of the normalized image affects the performance because if the size is too big, it will take a long time to train the neural network, take more processing time and have less translation tolerance. On the contrary, if the size is too small, the recognition rate will decrease due to lack of detail essential for classification. In our experiments some signs became visually unrecognizable with smaller image sizes. The other very important parameter of the neural network is the number of neurons in the hidden layer. If the number is too small, the neural network will not be able learn the data set in order to properly classify the road sign. If the number is too large, the neural network will have difficulty in converging and it will take a long time to train the neural network. We therefore report on the relationship between the normalization size of the image and the successful recognition rate as well as the number of hidden layer neurons to provide insight into the robustness of the system's performance with respect to the parameter set yielding the best performance.

Images in the database are randomly divided into 4 subsets (we also make sure that each subset contains all kinds of road signs). We perform 4-fold cross validation testing by training the neural network each time on three subsets and testing on the remaining subset. Finally, we report the average recognition rate for all four train-test scenarios in Table 1. Here, the successful recognition rate is the number of successfully recognized images divided by the number of total images.

Table 1: Recognition accuracy (percent) as function of number of hidden neurons in the neural network and input image size.

| Image Size \ Hidden Neurons | 4×4   | 5×5   | 10×10 | 20×20 | 30×30 | 40×40 |
|-----------------------------|-------|-------|-------|-------|-------|-------|
| 20                          | 62.95 | 67.36 | 86.79 | 90.16 | 78.24 | 76.17 |
| 30                          | 63.73 | 67.36 | 87.31 | 89.38 | 89.12 | 88.86 |
| 40                          | 65.80 | 71.76 | 86.27 | 89.12 | 90.41 | 86.27 |
| 50                          | 64.77 | 71.76 | 87.82 | 87.82 | 89.12 | 80.57 |
| 70                          | 59.07 | 67.62 | 88.34 | 89.38 | 88.86 | 81.34 |
| 80                          | 56.99 | 66.58 | 88.60 | 88.86 | 83.94 | 80.57 |
| 100                         | 54.40 | 53.87 | 44.30 | 58.81 | 39.90 | 45.08 |

From Table 1 we can see how image size and number of neurons in the hidden layer relate to the recognition rate. We specifically see that the parameter region centered around 40 hidden neurons and 30x30 image size provide the best recognition rate. The recognition rate remains almost the same until the normalization size drops below 10 by 10 pixels or the hidden neurons are more than 80. The recognition rate drops significantly when the normalization size is less than 5 by 5 pixels. We can conclude that, at least for the given data set, the system displayed stable learning and recall characteristics. The table also shows that the performance is not critically dependent on parameter choices suggesting the system is quite robust with good recognition capability.

## 5. CONCLUSION

We have presented an automatic road sign recognition system using mobile devices as the front end. The system utilizes the cost and mobility advantages as well as multimedia and connectivity capabilities of mobile devices. These are complemented by the processing and adaptability advantages of a central server. This system can be used to orient drivers and pedestrians or be adapted for automatic driving applications. It can also serve as a generalized sign recognition system with multilingual support for universal signs or alternatively for separate locales defined by their particular set of signs.

For road sign detection, we use particular color features calculated from the input image. We implement an artificial neural network for road sign recognition. Our evaluations have shown that the neural network has achieved good recognition performance while being relatively robust for a wide range of parameter choices. We have found that the RGB histogram representation of the road sign is a good feature for recognition. We also explored the relationship between the normalization size and successful recognition rate. With our database, the recognition rate started to drop when the normalization size was lower than 10 by 10 pixels. In order to keep a good recognition rate, the normalization size of the image should be at least 20 by 20 pixels.

In the future, we would like to explore different representations of the image for the neural network in order to further improve the performance of the neural network. Also, we want to see how the system scales when the database become very large. We are also interested in exploring the performance of the neural network when the database contains road signs with varying degrees of detail for differentiability such as signs with similar shape and color but different text.

## 6. ACKNOWLEDGMENTS

This work is funded by the NSF grant No. CCF-0939002 awarded for community building with Humanitarian Free and Open Source Software. We would also like to thank Maxwell Novak for programming part of the client side software and helping gather images.

## REFERENCES

- [1] S. Tsai, D. Chen, V. Chandrasekhar, G. Takacs, N.-M. Cheung, R. Vedantham, R. Grzeszczuk, and B. Girod, "Mobile product recognition", *ACM Multimedia (MM)*, October 2010.
- [2] D. Chen, S. Tsai, C.-H. Hsu, J. P. Singh, and B. Girod, "Mobile augmented reality for books on a shelf", *IEEE Workshop on Visual Content Identification and Search (VCIDS)*, July 2011.
- [3] S.-H. Hsu, C.-L. Huang, "Road sign detection and recognition using matching pursuit method", *Image and Vision Computing* 19, 119–129, 2001.
- [4] Jung-Chieh Hsien, Yi-Sheng Liou, Shu-Yuan Chen, "Road Sign Detection and Recognition Using Hidden Markov Model", *Asian Journal of Health and Information Sciences*, Vol. 1, No. 1, pp. 85-100, 2006.
- [5] Andrzej Ruta, Fatih Porikli, Shintaro Watanabe, Yongmin Li "In-vehicle camera traffic sign detection and recognition" *Machine Vision and Applications*, 2:359–375, 2011.
- [6] Auranuch Lorsakul, Jackrit Suthakorn, "Traffic Sign Recognition for Intelligent Vehicle/Driver Assistance System Using Neural Network on OpenCV", *The 4th International Conference on Ubiquitous Robots and Ambient Intelligence*, 2007.

- [7] R. Vicen-Bueno, R. Gil-Pita, M.P. Jarabo-Amores and F. L´opez-Ferreras, “Complexity Reduction in Neural Networks Applied to Traffic Sign Recognition”, *Proceedings of the 13<sup>th</sup> European Signal Processing Conference*, Antalya, Turkey, September , 2005.
- [8] Lafuente-Arroyo S, Gil-Jimenez P, Maldonado-Bascon R, Lopez-Ferreras F, Maldonado-Bascon S, "Traffic sign shape classification evaluation I: SVM using distance to borders ", *Intelligent Vehicles Symposium*, September, 2005.