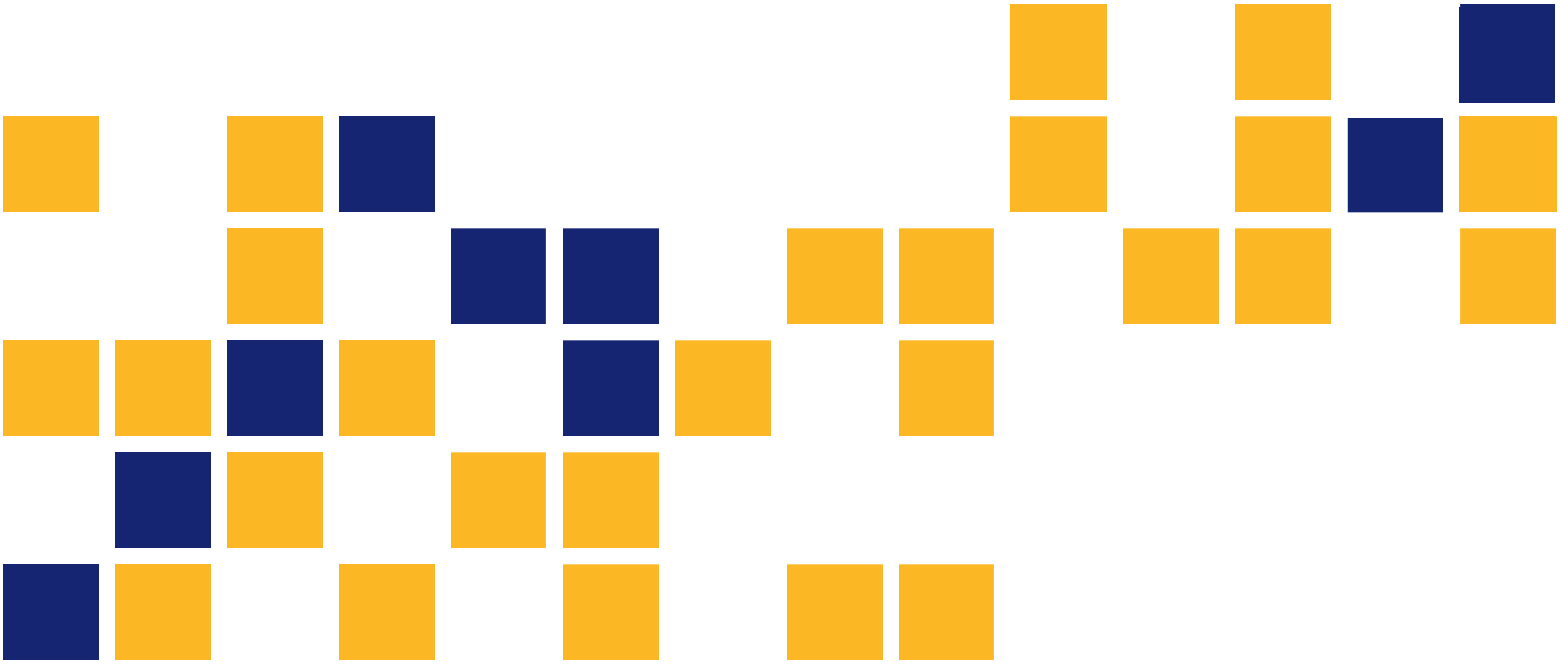
Report No. KS-09-9 ▪ FINAL REPORT ▪ June 2012

Final Report on Video Log Data Mining Project

Anupama Krishnan
Chris Lewis
Dwight Day
Kansas State University



Kansas Department of Transportation

1 Report No. KS-09-9	**2 Government Accession No.**	**3 Recipient Catalog No.**
4 Title and Subtitle Final Report on Video Log Data Mining Project		**5 Report Date** June 2012 **6 Performing Organization Code**
7 Author(s) Anupama Krishnan, Chris Lewis, and Dwight Day		**8 Performing Organization Report No.**
9 Performing Organization Name and Address Kansas State University Department of Electrical and Computer Engineering 2061 Rathbone Hall Manhattan, Kansas 66506-5204 Telephone: (785) 532-5600		**10 Work Unit No. (TRAIS)** **11 Contract or Grant No.** C1809
12 Sponsoring Agency Name and Address Kansas Department of Transportation Bureau of Materials and Research 700 SW Harrison Street Topeka, Kansas 66603-3745		**13 Type of Report and Period Covered** Final Report July 2008–September 2009 **14 Sponsoring Agency Code**

15 Supplementary Notes

16 Abstract

This report describes the development of an automated computer vision system that identities and inventories road signs from imagery acquired from the Kansas Department of Transportation's road profiling system that takes images every 26.4 feet on highways throughout the state. Statistical models characterizing the typical size, color, and physical location of signs are used to help identify signs from the imagery. First, two phases of a computationally efficient K-Means clustering algorithm are applied to the images to achieve over-segmentation. The novel second phase ensures over-segmentation without excessive computation. Extremely large and very small segments are rejected. The remaining segments are then classified based on color. Finally, the frame to frame trajectories of sign colored segments are analyzed using triangulation and Bundle adjustment to determine their physical location relative to the road profiler. Objects having the appropriate color, and physical placement are entered into a sign database. To develop the statistical models used for classification, a representative set of images was segmented and manually labeled determining the joint probabilistic models characterizing the color and location typical to that of road signs. Receiver Operating Characteristic curves were generated and analyzed to adjust the thresholds for the class identification. This system was tested and its performance characteristics are presented.

17 Key Words Sign Inventory, Road Profiling, Video Log, Data Mining		**18 Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161	
19 Security Classification (of this report) Unclassified	**20 Security Classification (of this page)** Unclassified	**21 No. of pages** 62	**22 Price**

Form DOT F 1700.7 (8-72)

Final Report on Video Log Data Mining Project

Final Report

Prepared by

Anupama Krishnan
Chris Lewis
Dwight Day

Kansas State University

A Report on Research Sponsored by

THE KANSAS DEPARTMENT OF TRANSPORTATION
TOPEKA, KANSAS

June 2012

NOTICE

The authors and the state of Kansas do not endorse products or manufacturers. Trade and manufacturers names appear herein solely because they are considered essential to the object of this report.

This information is available in alternative accessible formats. To obtain an alternative format, contact the Office of Transportation Information, Kansas Department of Transportation, 700 SW Harrison, Topeka, Kansas 66603 or phone (785) 296-3585 (Voice) (TDD).

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the views or the policies of the state of Kansas. This report does not constitute a standard, specification or regulation.

FINAL REPORT ON
VIDEO LOG DATA MINING PROJECT

BY

ANUPAMA KRISHNAN
CHRIS LEWIS
DWIGHT DAY

Department of Electrical and Computer Engineering
KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Abstract

This report describes the development of an automated computer vision system that identifies and inventories road signs from imagery acquired from the Kansas Department of Transportation′s road profiling system that takes images every 26.4 feet on highways through out the state. Statistical models characterizing the typical size, color, and physical location of signs are used to help identify signs from the imagery. First, two phases of a computationally efficient K-Means clustering algorithm are applied to the images to achieve over-segmentation. The novel second phase ensures over-segmentation without excessive computation. Extremely large and very small segments are rejected. The remaining segments are then classified based on color. Finally, the frame to frame trajectories of sign colored segments are analyzed using triangulation and Bundle adjustment to determine their physical location relative to the road profiler. Objects having the appropriate color, and physical placement are entered into a sign database. To develop the statistical models used for classification, a representative set of images was segmented and manually labeled determining the joint probabilistic models characterizing the color and location typical to that of road signs. Receiver Operating Characteristic curves were generated and analyzed to adjust the thresholds for the class identification. This system was tested and its performance characteristics are presented.

Table of Contents

List of Figures

List of Abbreviations

BA - Bundle Adjustment

pose - Position and orientation

SIFT - Scale Invariant Feature Transform

KDoT - Kansas Department of Transportation

ROC - Receiver Operating Characteristics

PDF - Probability Density Function

TPR - True Positive Rate

FPR - False Positive Rate

TP - True Positive

FP - False Positive

TN - True Negative

FN - False Nevative

Chapter 1

Introduction

This report describes development of an automated machine vision based road sign inventorying system that catalogs road signs from Kansas Department of Transportation's, (KDoT), imagery. KDoT's road profiler takes true color 1200x1600 pixel pictures every 26.4 feet along highways through out the state of Kansas. Many vision researchers have worked on building intelligent transportation systems addressing similar problems. Most utilize some means of segmentation process followed by a classification process that identifies the segments of a desired class.[1–5] Our system uses a novel difference image technique to efficiently achieve over-segmentation. It is also unique in that the 3D physical location of objects is ascertained and used to aid in classification. Being able to determine the 3-D location of imaged objects may permit database users to verify that road sign regulations are met as well as provide a myriad of other data mining opportunities. Our design is constrained by the massive number of images acquired yearly, and by the lack of control over lighting, viewing geometry and background. Algorithms are restricted to those that are computationally quick and memory efficient.

Processing begins by over-segmenting each image using two phases of a low cluster count K-Means algorithm. The first phase operates on the original image and the second on a difference image. Next overly large and small segments are rejected eliminating the vast majority from further consideration. Next, a statistical measure is used to classify segments as potential signs based on their color. Next, feature matching is performed to identify

sign colored segments that recur in consecutive frames. Recurring features are triangulated to estimate their 3D location relative to the vehicle's path. In the first triangulation pass, straight line motion between frames is assumed. The residual triangulation error is evaluated to determine if trajectory adjustment is warranted. When the error is large, usually along curved roads, Bundle Adjustment(BA) is used to refine motion estimates and facilitate better localization. Finally, another statistical classification process is performed using the 3D location. This technique takes advantage of the fact that signs are usually placed adjacent to the road at prescribed heights and distances from the road. A Neyman-Pearson hypothesis test is used to justify the use of Receiver Operating Characteristic curves for adjusting classifier thresholds. This report details the segmentation and classification process and presents results from an example set of highway images to demonstrate the efficiency and performance of the system.

Chapter 2

Segmentation

2.1 K-Means Segmentation

The goal of segmentation is to divide an image into clusters of like pixels. Ideally each cluster represents a single physical object. Like many image processing systems, we desire a segmentation that is slightly over segmented such that segments do not span multiple physical objects, but sometimes physical objects form multiple segments. K-Means is a computationally quick iterative algorithm that finds K natural clusters. It has seen widespread use for image segmentation for example, in medical applications such as[6] and[7]. The K-means algorithm iteratively finds the centers for k natural clusters in a set of data. The procedure for K-Means consists of the following five steps[8]:

Step 1: Initialize cluster centers $\mu_1, \mu_2, ...\mu_k$ and number of required clusters, k

Step 2: Classify each sample based on the nearest μ_i

Step 3: Recompute μ_i as the mean of the samples in each of the new classifications obtained from step 2

Step 4: Repeat steps 2 and 3 until there is no change in μ_i

Step 5: Return $\mu_1, \mu_2, ...\mu_k$

The number of clusters, k, is critical. Larger k results in more segments and increased computational complexity, while smaller values results in under-segmentation, but are computationally quick. For the images obtained from KDoT, sixteen clusters provided an appropriate trade off between adequate segmentation and computational complexity.

2.2 Two Pass Segmentation

Fig. 2.1, depicts the results with $k = 16$ with all pixels in each segment's color shifted to the closest cluster center's color. Even with only 16 colors, the image represents the original well. Note that the yellow road sign is only slightly shifted in color since k-means selected its shade of yellow as one of the 16 cluster centers.

Figure 2.1: *This figure shows the original image and the color segmented image obtained using K-Means Clustering on an image with a yellow road sign. In the color segmented image the pixels belonging to the yellow portion of the road sign are grouped into a segment.*

However, under-segmentation also occurs as shown in Fig. 2.2 where k-means assigns the small road sign to the cluster color associated with similarly colored background objects resulting in the merging of the two physical objects into one segment. This occurs often with green or white signs as vegetation or white sky are common backdrops. This deficiency is alleviated using a difference image technique.

Figure 2.2: *Here, the green sign's color is not one of the 16 color centers. Its pixels are shifted to the closest cluster center resulting in under segmentation.*

2.3 Difference Image Segmentation

Usually in cases where under-segmentation occurs with K-Means, smaller objects undergo large shifts since none of the cluster centers are that close to the object's original color. This phenomenon led to the use of a difference image which is formed as the euclidean distance in color space between the original image and the K-Means color segmented image for each pixel. The difference image accentuates objects receiving large color shifts. Proper segmentation of these problematic objects is achieved by reapplying K-Means to the difference image. Proper segmentation of the green sign from Fig. 2.2 shown in Fig. 2.3.

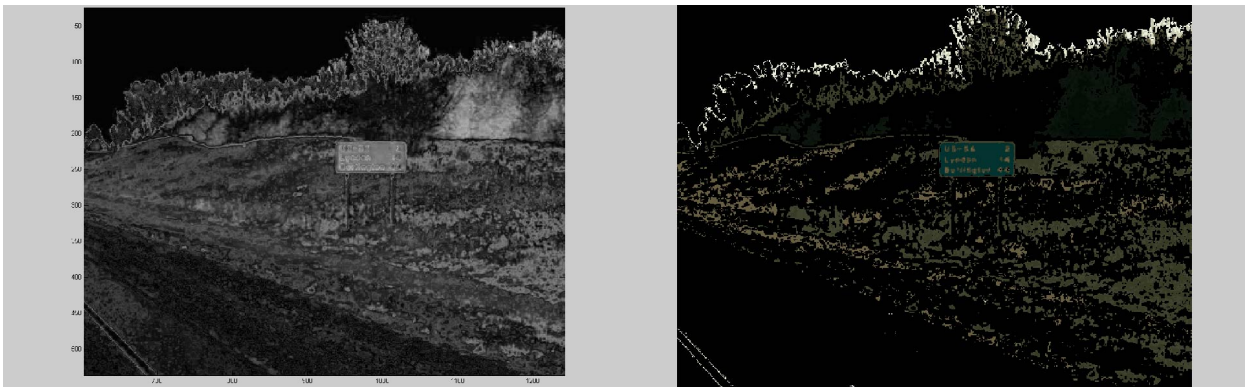


Figure 2.3: *Significant color shifts are accentuated in this gray scale difference image, (left). K-Means applied to the color difference image properly segments the green sign.*

Chapter 3

Classification

3.1 Elimination of Large and Small Segments

A large number of segments are obtained from the segmentation phase. Most of them are very small, a few pixels in size. Others are very large, such as segments of the sky. Neither need further consideration and are eliminated by applying size thresholds. For the current process, all segments that are either smaller than 50 pixels or larger than 20000 pixels are removed. Fig. 3.1 shows the segments remaining from Fig. 2.2 after thresholding.

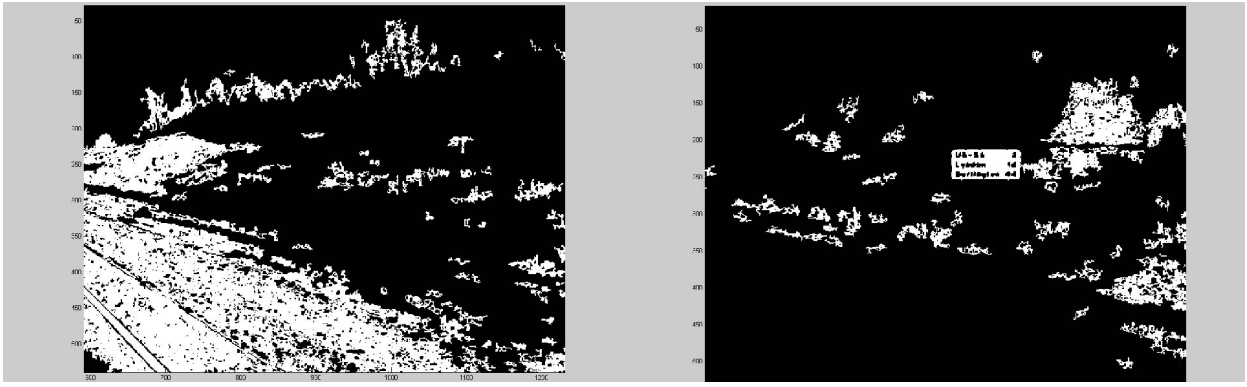


Figure 3.1: *Black and white depiction of remaining segments after size threshold is applied to the color segmented image shown in Fig. 2.2 and difference image shown in Fig. 2.3.*

3.2 Color based Classification

Road signs are more uniformly colored than typical objects. They have a unique correlation between their red, green and blue components even with differences in lighting. The joint color distributions for different sign types were determined by manually extracting numerous road signs of every type from a set of images. To help automatically classify segments its proximity to each of these joint color distributions is determined using the Mahalanobis distance measure.[9–11] The Mahalanobis distance is given by Eq. (3.1).

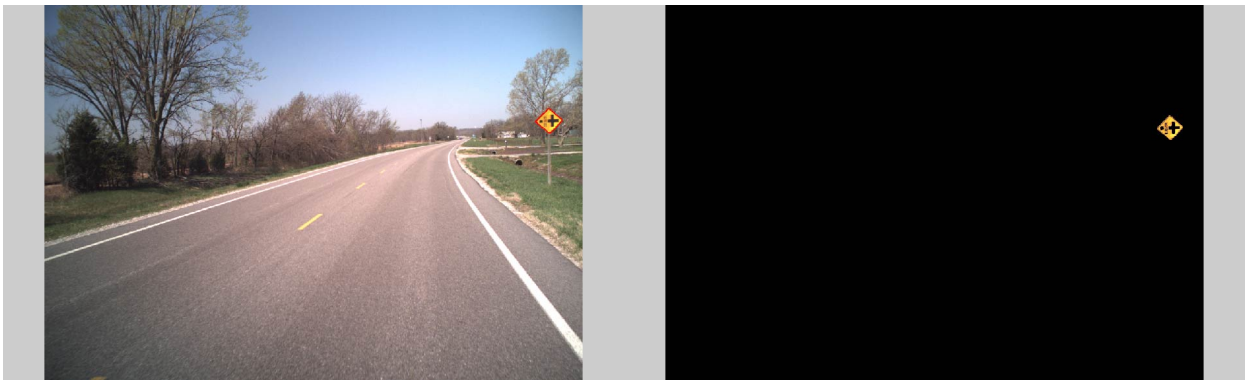
$$r = \sqrt{(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})} \tag{3.1}$$

where $\vec{\mu}$ is the mean vector and Σ is the covariance matrix of the prior joint color distribution and $\vec{x}$ is an input point in multidimensional space[8]. For this research, $\vec{x}$ contains the average values of the Red, Green and Blue components of all the pixels belonging to the segment being analyzed. An optimum distance threshold is practically determined from the knee of the Receiver Operating Characteristic (ROC) curves. These curves are generated using a test set of 25 images containing signs of each color. The knee of each of these curves corresponds to the threshold value that gives the best trade off between false positives rate and true positive rate. A Neyman-Pearson Hypothesis test provides the theoretical justification for setting thresholds in this manner.

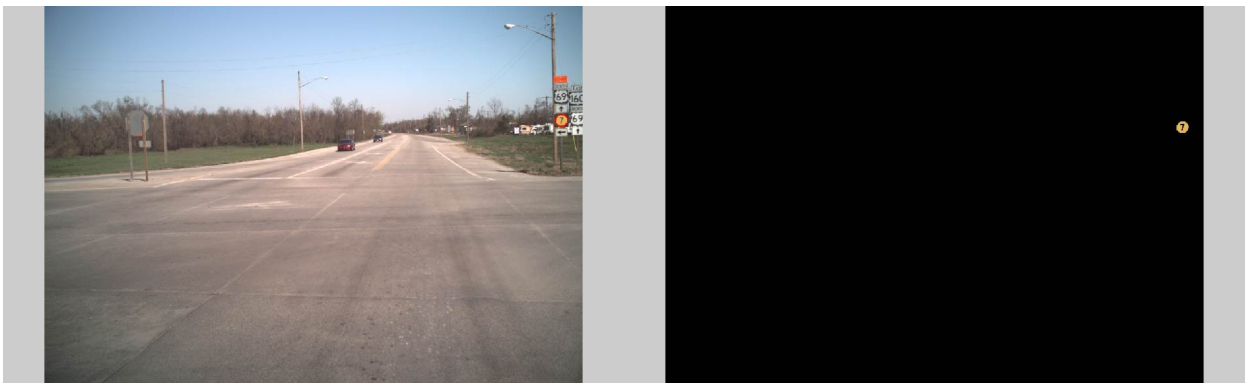
Few natural objects have a brilliant color similar to yellow road signs and hence the color classifier is nearly ideal for identifying segments of this type as shown in Fig. 3.2. For other colors such a green and white signs, this is unfortunately not the case.

For example, Fig. 3.3 displays some reasonably sized segments having similar color to green road signs, but which are not signs. Another example of this case is illustrated in Fig. 3.4, where road stripes were falsely classified as road signs due to their similarity in color distribution. Therefore, additional classification steps are necessary to accurately identify most sign types.

(a)



(b)



(c)

Figure 3.2: *The color classifier is able to identify yellow signs with very high precision.*

Figure 3.3: *Results of color based classification on segments obtained in 3.1. The highlighted regions depict the segments with a color distribution similar to that of road signs.*

Figure 3.4: *A case of false positives for yellow signs.*

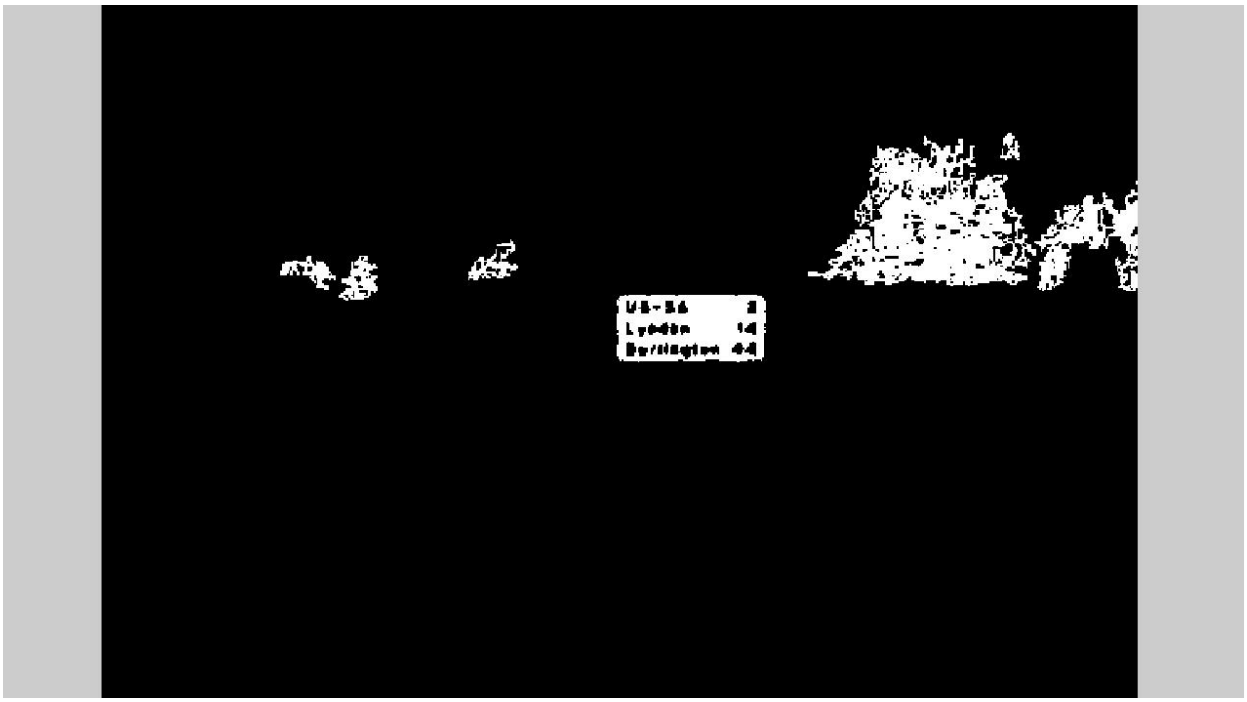
Chapter 4

Shape Classifier

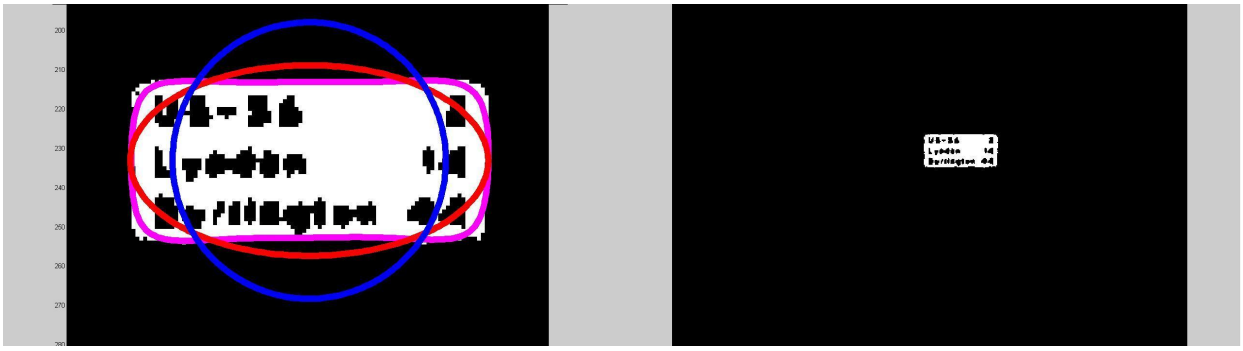
Fourier descriptors are a common tool used for pattern recognition and shape discrimination.[12] They are determined using the periphery of the object. The advantage of using the Fourier descriptors is their invariance to the starting point of the boundary, scale and rotation.[13,14] Fourier Descriptors begin by representing the boundary of a region as a sequence of complex numbers. By taking the Fourier transform of this sequence and retaining only the lower frequencies, an approximation of the shape can be obtained. For recognition purposes the ratio of the magnitude of the upper frequencies with the first frequency, provides a set of scale and rotation invariant features. Since, signs are generally simple polygons with few sides; only a few terms are required to represent their shape. For this work, the number of fourier descriptors chosen was 6. An example of the shape approximation is shown in Fig. 4.1. The blue boundary shows the best fitting circle, red shows the best fitting ellipse and magenta shows the shape approximation with six Fourier descriptors for the two segments.

The Mahalanobis distance between the computed Fourier descriptors for a segment and the distribution of the descriptors of the desirable shapes is calculated in a similar manner as in color classification. Again, a threshold is applied to discard segments with the wrong shape. The shape thresholds are also determined using the ROC analysis. The output after elimination using this shape classifier is shown in Fig. 4.1 and Fig. 4.2.

As fourier descriptors are strictly dependent on the boundary, they fail in instances where there are slight variations such as the case shown in Fig. 4.2. When tested on a

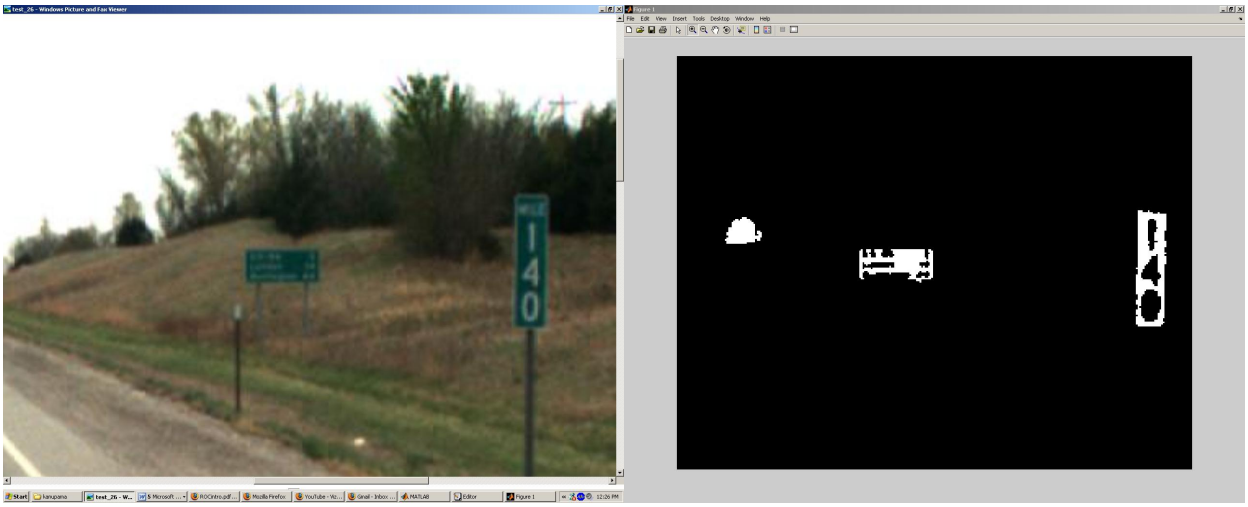


(a)

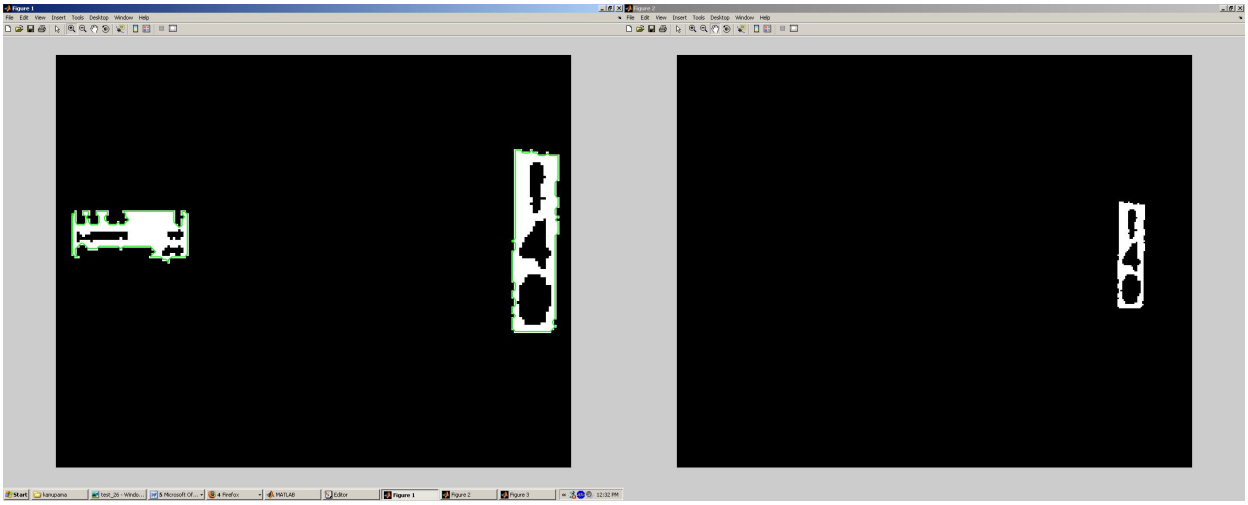


(b)

Figure 4.1: *This figures shows the results of classification based on shape of segments remaining from color classification in the example shown in Fig. 3.3. Fourier descriptors describe shape from the boundary of the segment. The best fitting circle (Blue), ellipse (Red) and a 6th order shape (Magenta) are shown for the two segments. The shape classifier applied to both sets of correctly colored segments, eliminates segments with improper shape but identifies the sign with the desired shape.*



(a) Image with green road signs, left, and resulting segments after color classification, right.



(b) Magnified image from (a) showing that one sign has an irregular boundary, (left). The shape classifier was able to identify the sign with the rectangular boundary but however misclassifies the sign with the irregular boundary as not sign.

Figure 4.2: *A case of false negative while using fourier descriptors for classification.*

set of 50 images, the shape classifier was observed to be 92% efficient at identifying road signs correctly but misclassified 8% of the road signs as not signs. This shortcoming was overcome by replacing the shape classifier with one based on the feature's frame to frame trajectory.

Chapter 5

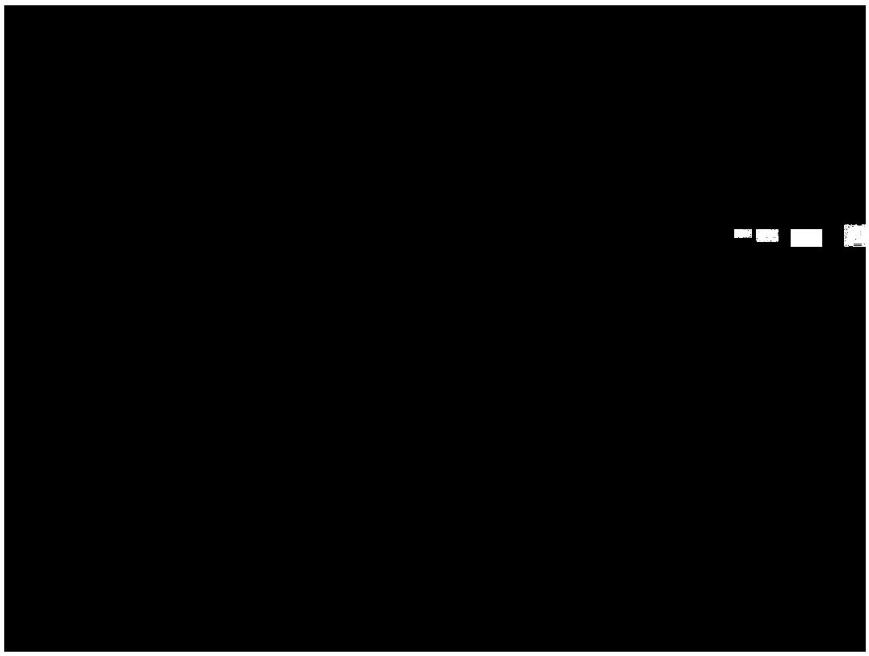
Location Classifier

Images are captured sequentially every 26.4 feet. Therefore, road signs are likely to appear in a sequence of several images and follow a characteristic trajectory. Matching features from segments appearing in consecutive images and comparing their trajectory through the field of view of the camera to that typical of road signs eliminates segments that are not road signs. In our system, the SIFT algorithm,[15] tracked features from frame to frame.

The 2-D trajectory of the features identified by SIFT was compared to a probabilistic model determined from a test set of sequential images containing road signs. Similar to color based classification where the feature was a 3 dimensional vector, components of which were the red, green and blue values of the pixels, the feature space for the location based classification are four dimensional vectors defining the motion of typical road signs. The components of each vector are the u and v pixel coordinates of the road sign and the change in location in u and v of that matched feature in terms of pixels in the consecutive image. An example illustrating the trajectory of a road sign is shown in Fig. 5.1 and Fig. 5.2.

This process removed problematic segments including billboards containing similar color distribution. Fig. 5.3 shows the resulting segments from the size and color based classification on an example pair of consecutive images. The outcome of location based classification on these segments is shown in Fig. 5.4. Typically, all the false positives from the color classifier are removed by the trajectory classifier. Frame to frame feature matching is an

Figure 5.1: *Four consecutive images obtained from KDoT's road profiler*



(a) Trajectory

(b) Magnified trajectory

Figure 5.2: *Positions of road sign in four images*

Figure 5.3: *Two subsequent images obtained from the video log. The highlighted regions indicate the segments that remain after size and color based classification.*

Figure 5.4: *Results of location based classification. All the segments that do not follow the trajectory of a typical road sign are removed.*

unstructured and computationally efficient form of triangulation.

This method worked well on straight roads as shown in Fig. 5.3 and Fig. 5.4, but failed on curves where trajectory of objects is altered through the field of view. The system now uses multiple images to triangulate objects and ascertain their three dimensional location relative to the vehicle. These 3D positions help distinguish potential road signs from objects not correctly positioned. An initial triangulation is carried out assuming straight line motion but the residual error is monitored to detect non-linear vehicle motion. In those relatively infrequent cases where the vehicle turns substantially, the vehicle's true motion is estimated using a process called Bundle Adjustment (BA), and the 3D locations are re-estimated for

the classifier.

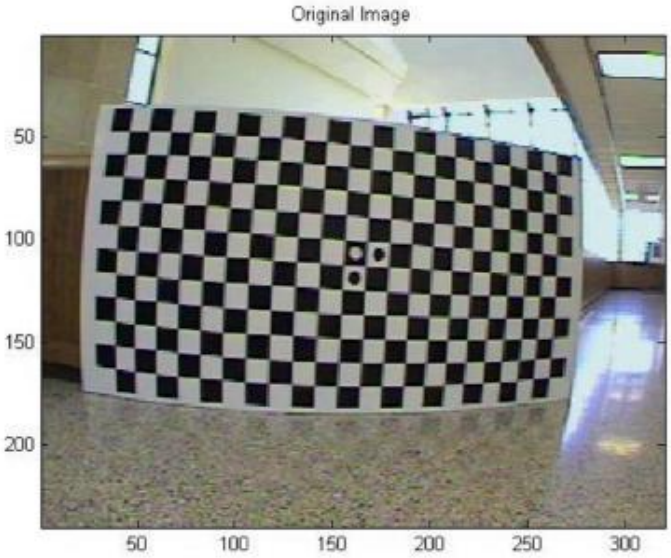
To implement a 3D location classification, features must be tracked frame to frame. As the vehicle moves, images of objects change in size. The Scale Invariant Feature Transform (SIFT) algorithm[15], is used by our system to track objects frame to frame because it is robust to changes in scale, rotation and lighting. SIFT is state-of the-art, and therefore commonly applied to video tracking applications.[16,17] An example of feature matching using SIFT is shown in Fig. 8.3. After identifying objects that appear in successive images using SIFT, it is important to remove lens distortion from the images to ensure a more accurate triangulation. The next chapter describes the procedure that was adopted to remove the distortion from the regions of interest identified by SIFT.

Chapter 6

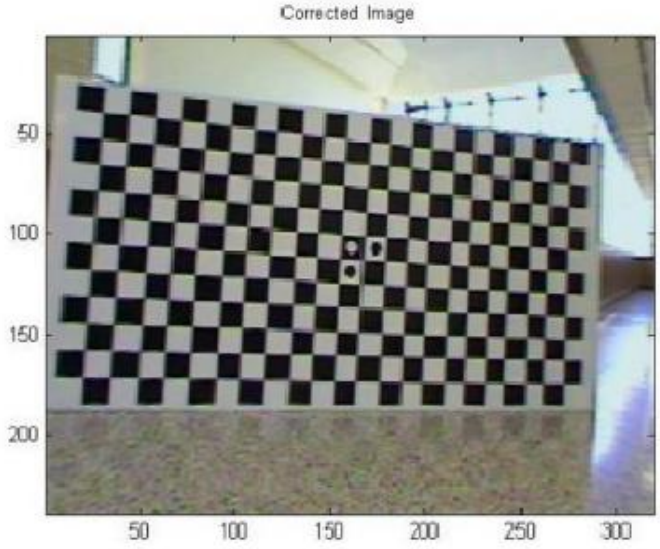
Camera Calibration

Real lenses distort images and corrupt the triangulation process. For our purposes, it is not necessary to undistort the entire image. Rather, to minimize computations, lens distortion effects are taken into account only for the features tracked frame to frame. The imaging model for a camera including lens distortion is determined through camera calibration.[18] Calibration parameters include the focal lengths, f_x and f_y, the principal point, (c_x, c_y), and the skewness.[19] To estimate these parameters, this effort used a Calibration toolbox for Matlab to process multiple images of a large checker board pattern printed on a 4x8 sheet of structural foam. The calibration process is explained in[18] and[20] and the tool itself is available on line.[21] To permit accurate triangulation, these calibration parameters are used to shift the location of features to where they would appear in the focal plane of an ideal pin hole camera. Photographs of the calibration target, with and without distortion, are shown in Fig. 6.1.

The next chapter describes the triangulation process adopted to address the research problem.



(a) With lens Distortion



(b) Without lens Distortion

Figure 6.1: *Checkerboard pattern used to calibrate for lens distortion, left. Unwarped image shown on right.*

Chapter 7

Triangulation

Triangulation is a well known photogrammetric process for determining the 3-D location of features that appear in two or more images taken from different vantage points. Several methods exist for triangulation[22]. The optimal triangulation method involves the evaluation of the roots of a sixth degree polynomial to compute the best estimate[22]. However, this is computationally expensive and hence a very computationally simple method was employed to obtain an estimate for the 3-D location of features in a pair of successive images. We use an analytic solution where the location estimate is the midpoint of the line segment passing between the points of closest approach between two rays, one from each camera. The pin hole camera model depicted in Fig. 7.1 defines each ray, $\vec{q}_1$ and $\vec{q}_2$, emanating from the focal points of each camera, passing through the image plane at the pixel coordinates u, v where the feature was observed in each image. Ideally these two rays intersect at the feature. Noise in camera position, orientation or in lens distortion results in these rays being skew. The shortest line segment between the two rays is mutually perpendicular to both and its direction is defined by their cross product as follows,

$$\vec{q}_3 = \vec{q}_1 \times \vec{q}_2 \tag{7.1}$$

A vector loop equation involving these rays defined as follows

$$\vec{C}_1 + \gamma_1 \vec{q}_1 + \gamma_3 \vec{q}_3 = \vec{C}_2 + \gamma_2 \vec{q}_2 \tag{7.2}$$

This may be rewritten as:

$$\vec{C}_1 - \vec{C}_2 = [-\vec{q}_1 \ \vec{q}_2 \ -\vec{q}_3] \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \tag{7.3}$$

and then solved to determine the unknowns γ_1, γ_2 and γ_3. The midpoint of the line segment

$$\vec{f}_{xyz} = \vec{C}_1 + \gamma_1 \vec{q}_1 + \frac{\gamma_3}{2} \vec{q}_3 \tag{7.4}$$

is our analytic triangulation estimate. Back projecting this midpoint into both images and comparing to the observed image provides the measure of residual triangulation error as follows:

$$e = (u_1 - \hat{u}_1)^2 + (v_1 - \hat{v}_1)^2 + (u_2 - \hat{u}_2)^2 + (v_2 - \hat{v}_2)^2 \tag{7.5}$$

where $\hat{u}$ and $\hat{v}$ are back projected image estimates given by:

$$\begin{bmatrix} \hat{u}_i \\ \hat{v}_i \end{bmatrix} = \frac{f}{z} \begin{bmatrix} x \\ y \end{bmatrix} \tag{7.6}$$

where x,y, and z are the coordinates of f_{xyz}, expressed the i^{th} camera's frame.

The initial triangulation is performed assuming straight line motion without any rotation because images are usually taken on straight highways with little curvature. In this case the difference between the vectors $\vec{C}_1$ and $\vec{C}_2$ is accurately known. However, on curved roads the straight line assumption is not met and triangulation error is high as illustrated in Fig. 7.2. For these cases, the motion estimate is refinement using BA.

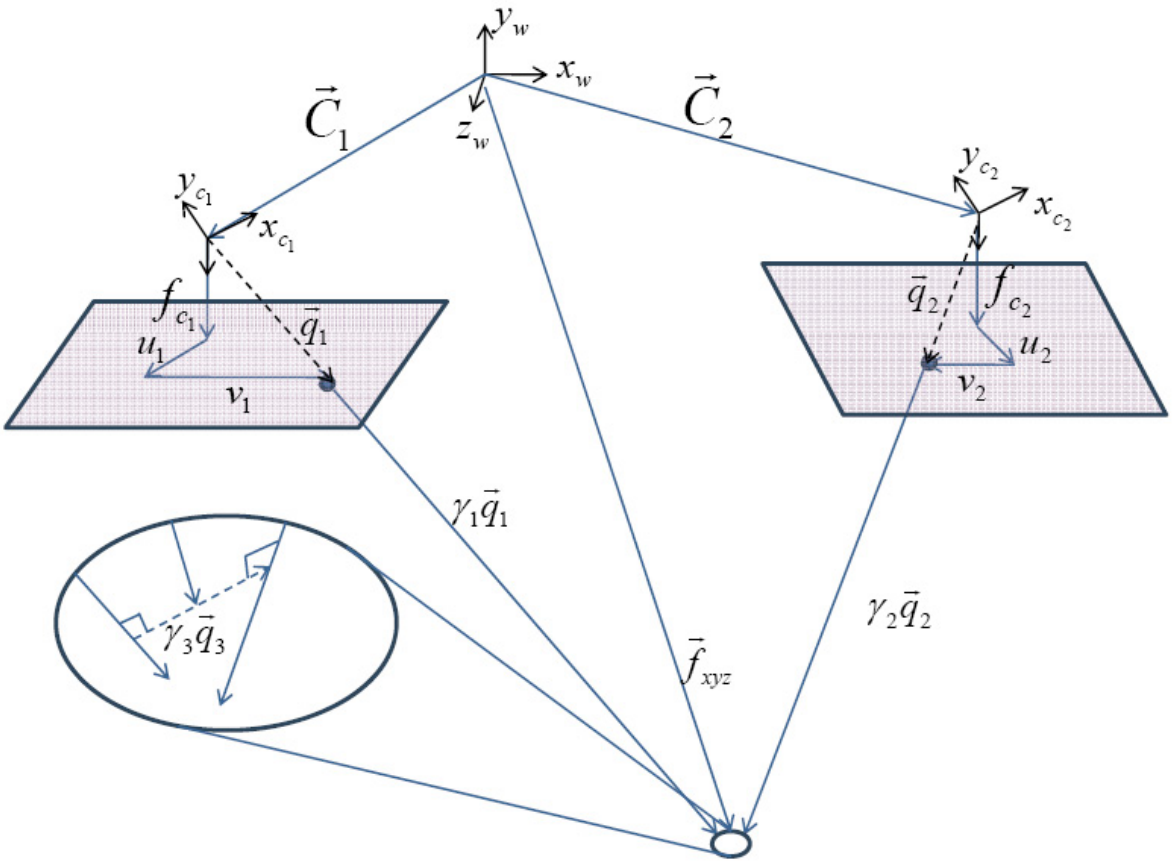


Figure 7.1: *A feature is imaged by two cameras located at $\vec{C}_1$ and $\vec{C}_2$. The rays $\vec{q}_1$ and $\vec{q}_2$ from these locations are defined by the feature's u and v pixel coordinates and the associated focal lengths f_1 and f_2. Ideally, these two rays intersect at the feature, but are more typically skew. Triangulation is the process of estimating the feature's location $\vec{f}_{xyz}$.*

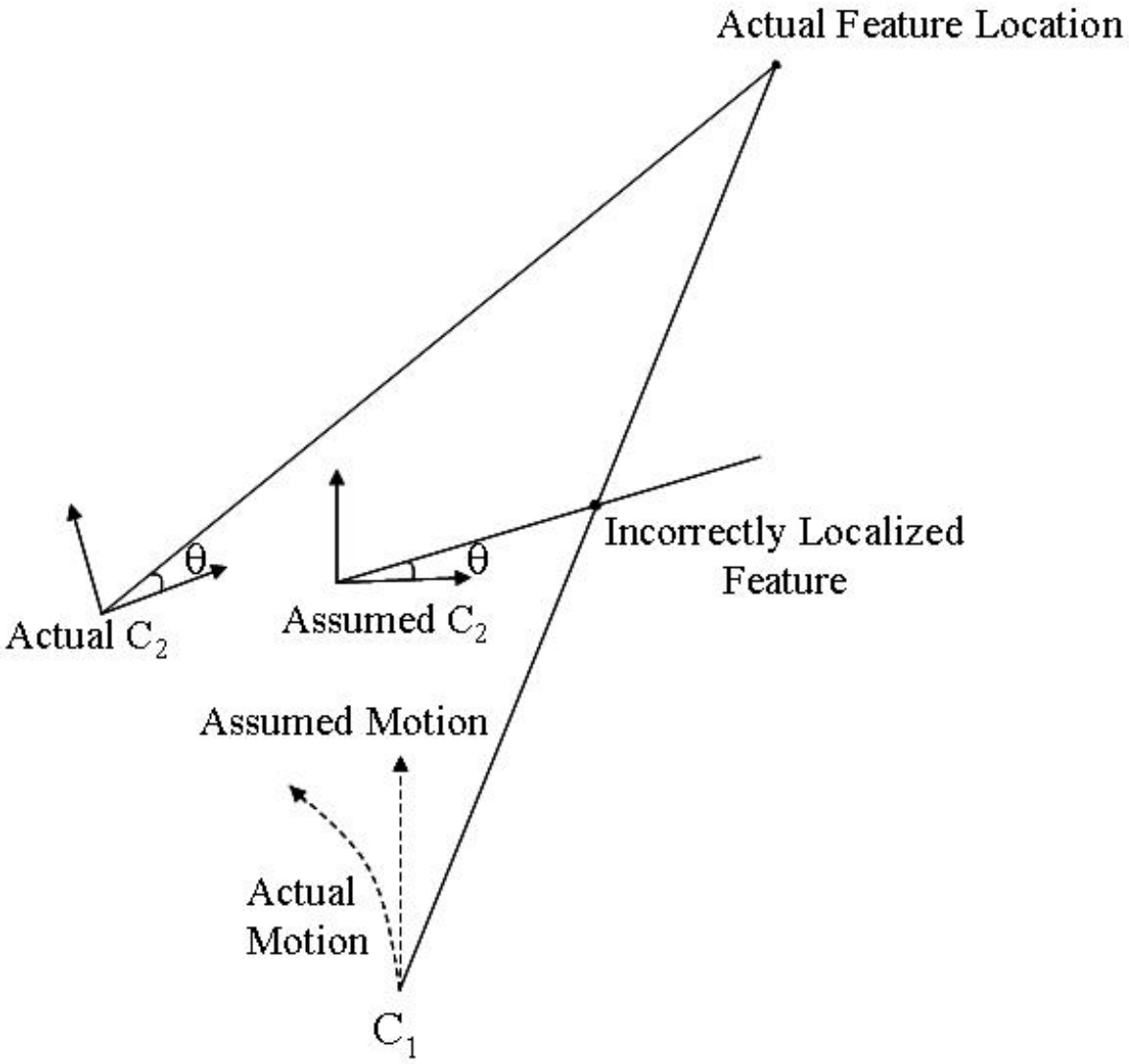


Figure 7.2: *Triangulation error is large with road profiling system traverses a curved path*

Figure 7.3: *Usually images are taken with straight line motion between frames, and the vectors $\vec{C}_1$ and $\vec{C}_2$ are accurately known. However, on curved roads the straight line assumption is not met and triangulation error is high.*

Chapter 8

Bundle Adjustment

Bundle Adjustment, (BA), modifies the camera's pose estimate including position, $\vec{C}_i$, and the orientation, (roll (R_r), pitch (R_p) and yaw (R_y)) along with the 3-D locations, $T_i = (x_i; y_i; z_i)$, of the features to reduce the triangulation error. In our implementation, the frame of reference is the first camera's frame. The six pose parameters of the second camera and three location parameters for each feature relative to this frame are optimized. The vector of parameters for m tie points is given by: $\vec{K} = \left[\vec{C}_2^t; R_r; R_p; R_y; T_1; \ldots T_m\right]$. The cost function optimized by BA is the sum of triangulation errors over all features. This non-linear cost function is solved with the iterative Levenburg-Marquardt algorithm[23]. This report only presents the basic idea behind Bundle Adjustment. Further elaboration on the concept of BA and comparison between different cost functions is given in[24] and[22]. Without constraints other than tie points scale is not determined. Constraining our solutions to match the known $||\vec{C}_1 - \vec{C}_2|| = 26.4$ feet of motion between images fixes scale and for our system BA converges providing initial pose estimates are reasonably accurate. Fig. 8.1 to Fig. 8.6 depicts the results of using triangulation and BA on some objects that appear in a pair of successive images obtained from KDoT's road profiling system. From the pair of images, the relative locations of various objects is determined. The physical location of objects is strongly tied to whether or not they are a sign.

Classification based on physical location is analogous to color based classification. To characterize the statistical distribution of sign locations, a database was created from 50

Figure 8.1: *Two consecutive images to be analyzed.*

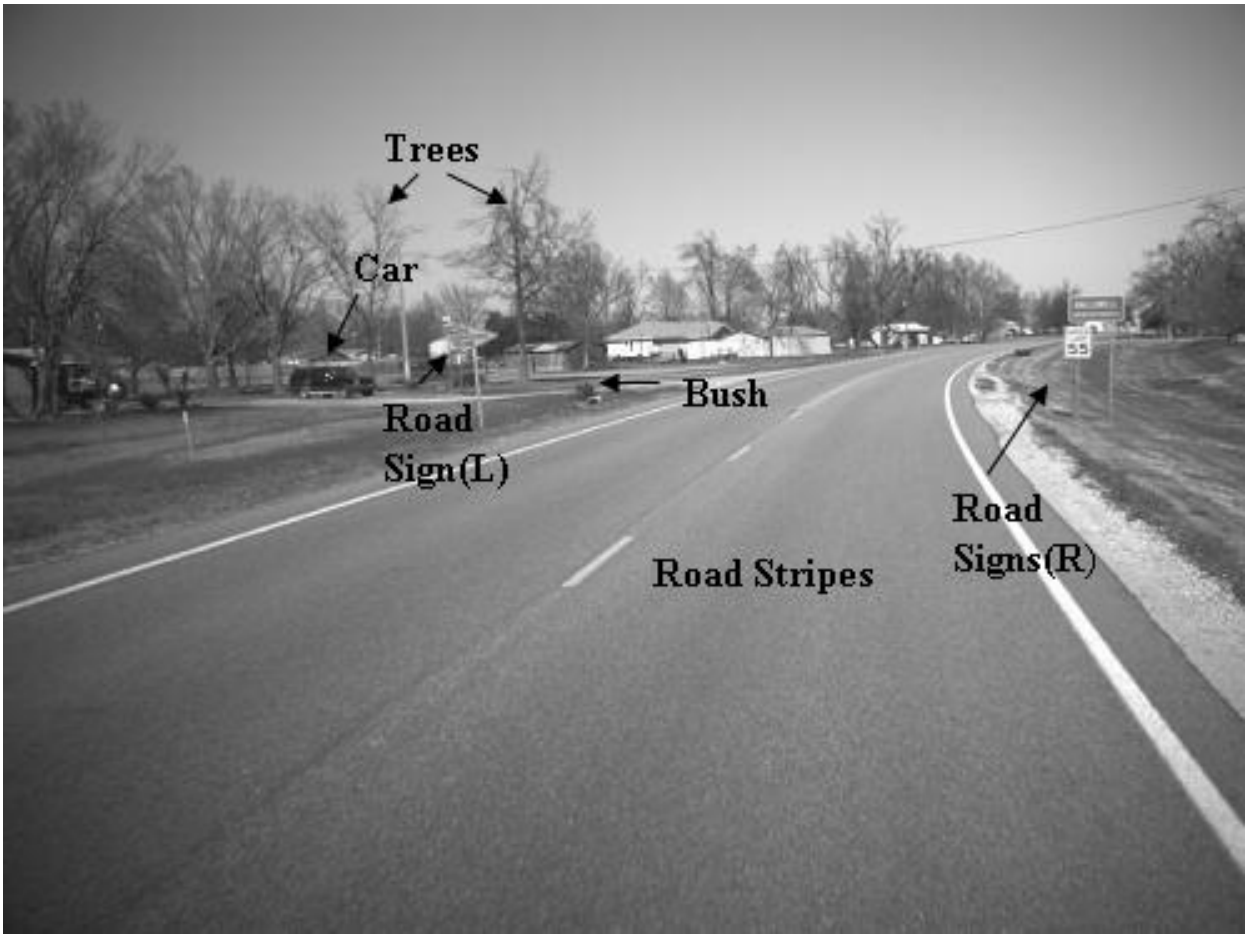


Figure 8.2: *Scene with various objects identified for triangulation. BA will reduce triangulation error and allow scene reconstruction.*

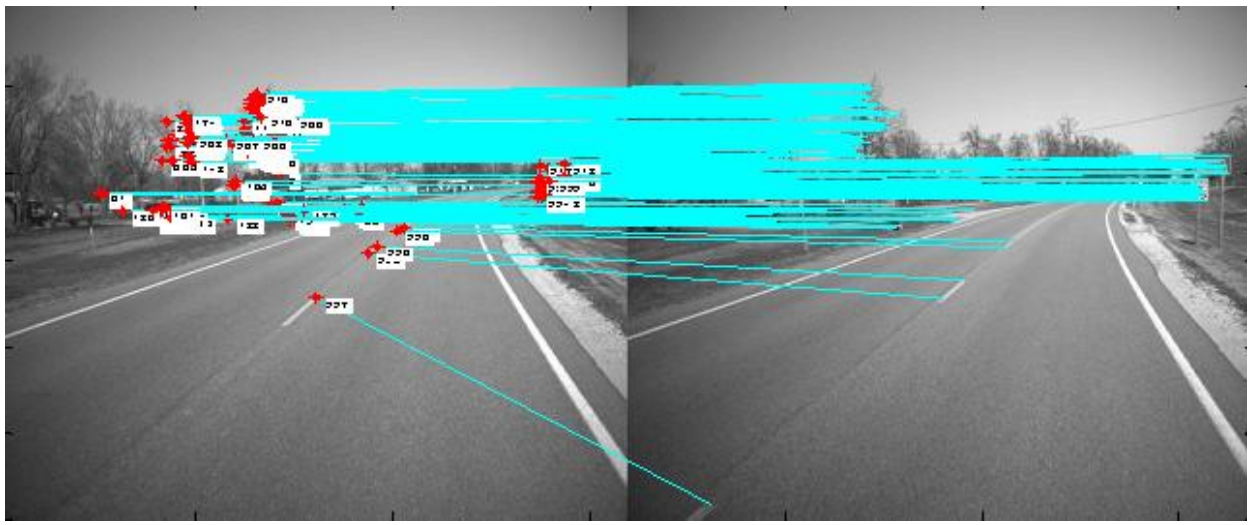


Figure 8.3: *For triangulation, features must be matched from frame to frame. SIFT is used to accomplish this.*

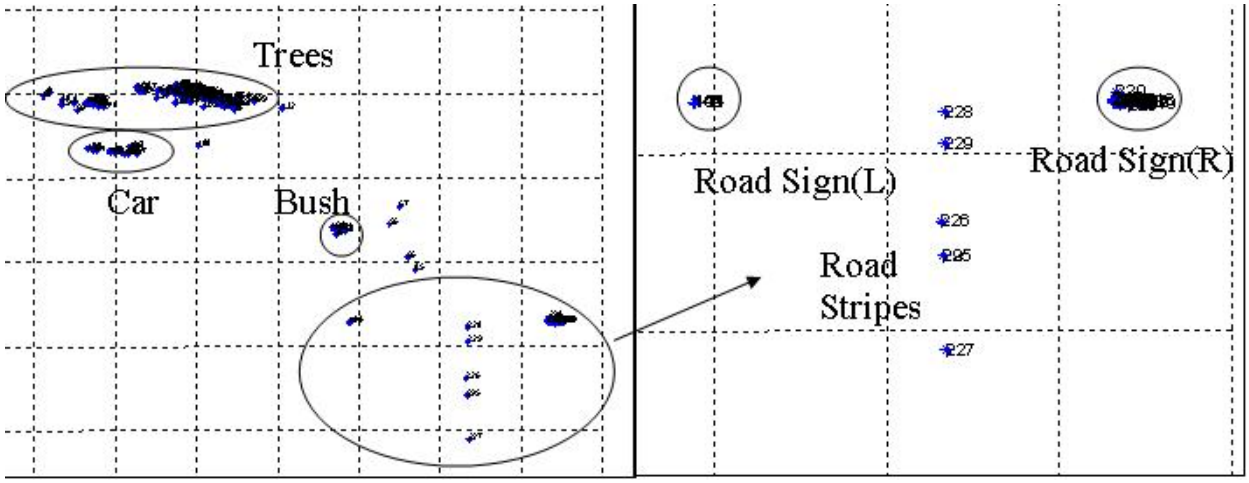


Figure 8.4: *The top view of the 3-D scene containing the labeled objects obtained from BA. Note how BA accurately computes the 3-D location of the features extracted from the road stripes. They appear in between the yellow road sign on the left and the green and white road signs on the right.*

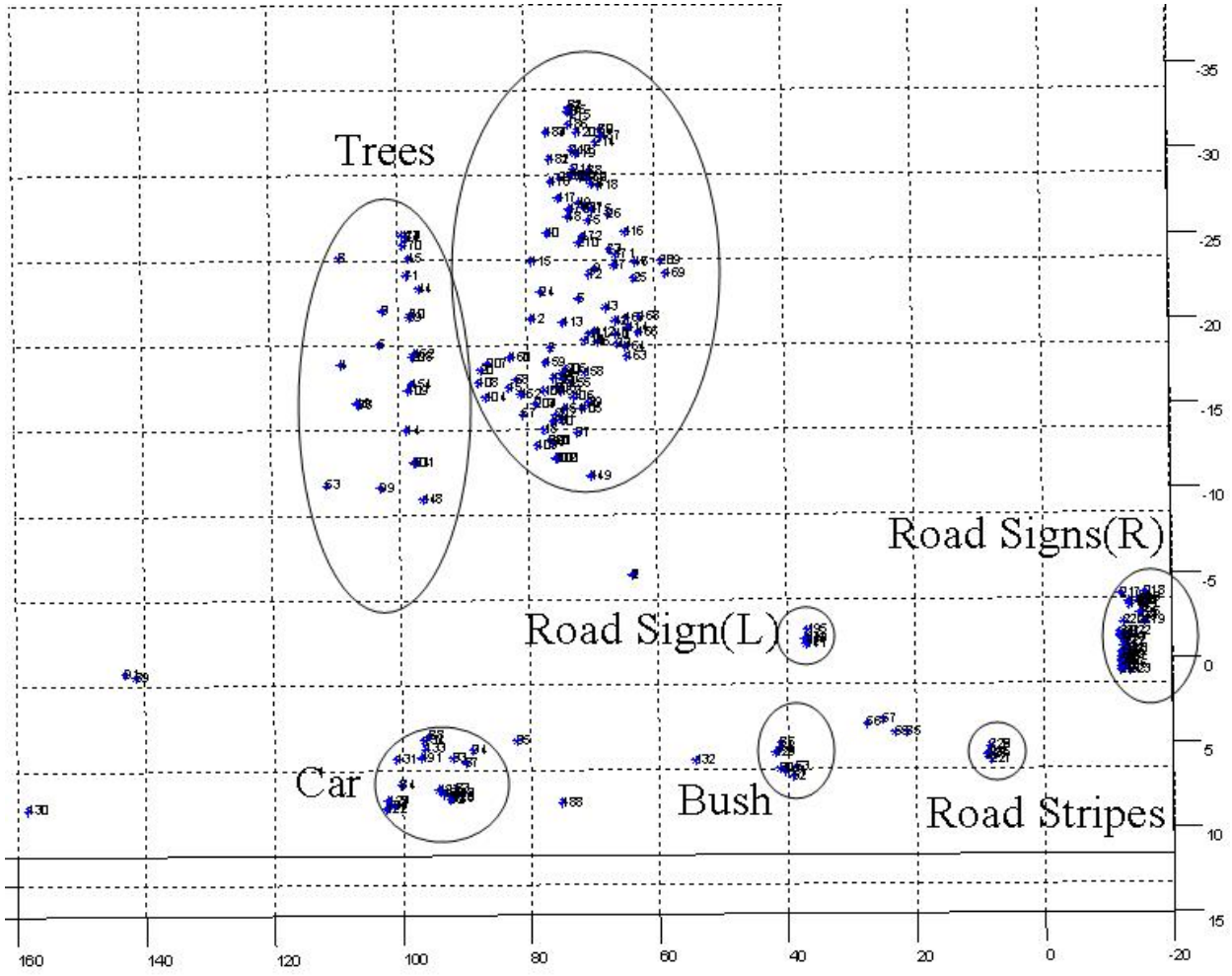


Figure 8.5: *Front View of 3-D plot of the locations of the matched features computed by BA.*

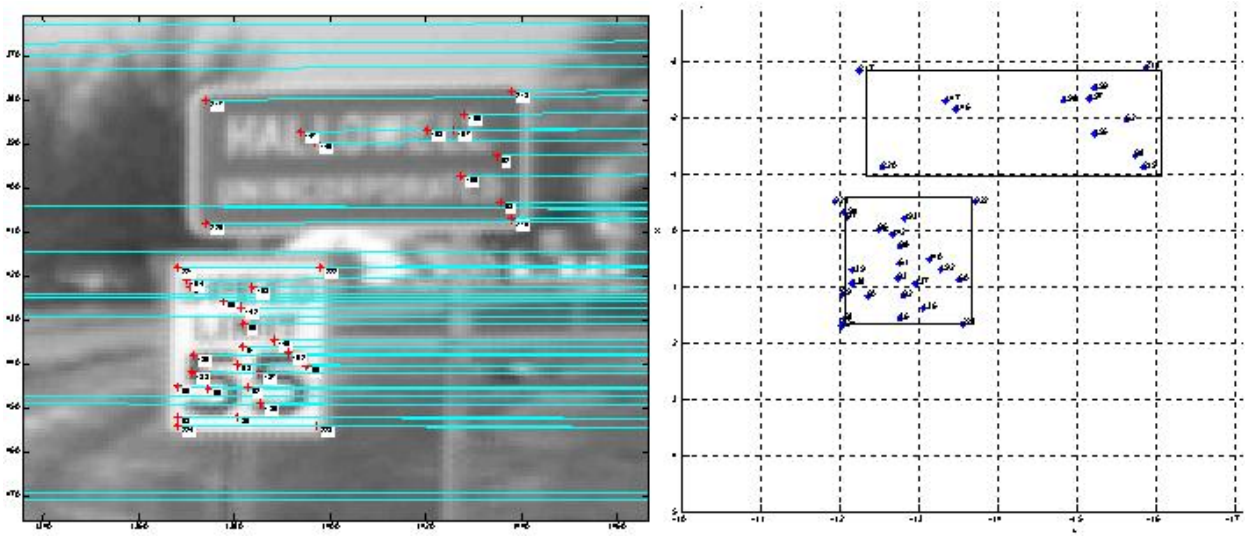


Figure 8.6: *Features matched from the two road signs on the right are shown on the left and the zoomed view of figure 9 showing the the front view of the corresponding 3-D location of the features obtained using BA is shown on the right. Each division is 1 foot and it can be seen that the larger road sign is about 3.5 feet wide and has a height of about 2 feet. Hence an estimate of the size of the road sign is obtained.*

images containing the 3-D location of features from road signs. Tie points were extracted using SIFT on just the regions of the image containing the road signs. Their 3D locations were determined using triangulation and BA. The mean (x, y) values of the positions of road signs was $(1.9783, 10.2391)$ in the first camera's frame, ten feet to the right and two feet above the camera. Note, since z is along the direction of travel, that it cannot be used. The Euclidean distance from this mean is now the Manhalanobis distance measure used for location based classification.

The Mahalanobis distance measures the proximity of a point to a distribution. This measure is used in both the color and location classifiers. Distance thresholds need to be determined for decisions to differentiate between objects being signs or not. The basis for selecting these critical threshold values is described in the next chapter.

Chapter 9

Optimal Threshold Value Selection

The selection of thresholds for classification can be developed as a hypothesis testing problem. The details on the various types of hypothesis tests and their comparisons are given in[25]. Here we apply the Neyman-Pearson test with a binary hypotheses for detecting white road signs. The problem can be defined as:

H_0: Not White Road Sign

H_1: White Road Sign.

The statistics of color are modeled using 50 images to visually construct a database of the red, blue and green values of all the pixels in objects that are and are not white road signs. Models of probability density functions are fit to the color histograms for both hypotheses. The PDFs are illustrated in Fig. 9.1. For red data, the PDF is modeled as:

$$p_0(y) = \frac{1}{\sqrt{2\pi * 1489}} exp\left\{\frac{-(y-118.337)^2}{2*1489}\right\} \tag{9.1}$$

$$\forall \quad 20 \leq y \leq 252.6$$

$$p_1(y) = 4.211 * 10^{-9} exp\left\{\frac{y}{\sqrt{235}}\right\} \tag{9.2}$$

$$\forall \quad 136.2 \leq y \leq 253.8$$

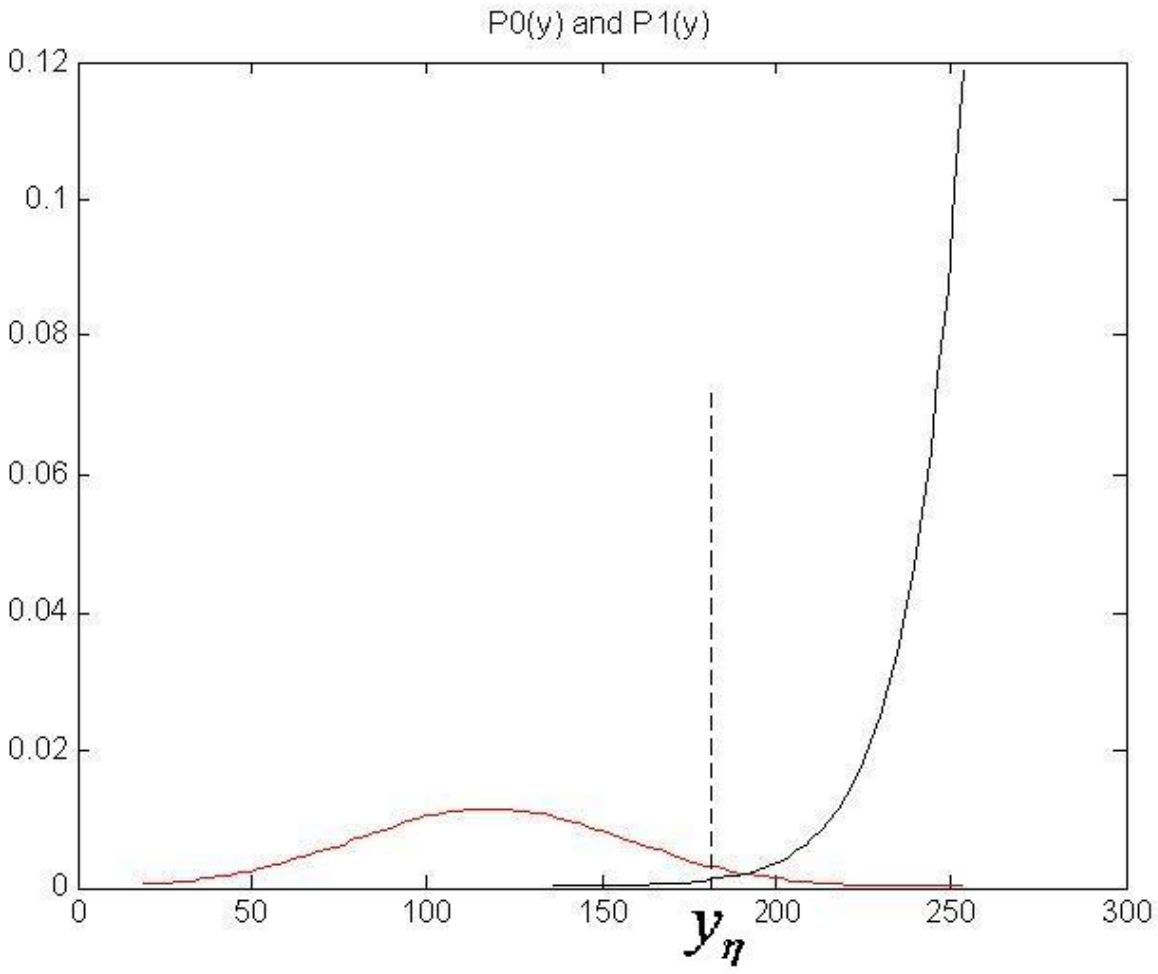


Figure 9.1: *PDFs for the two hypotheses:* $p_0(y)$ *and* $p_1(y)$

The likelihood ratio test is defined by:

$$choose \begin{cases} H_1 & \text{if } L(y) \geq \eta \\ H_0 & \text{if } L(y) < \eta \end{cases} \tag{9.3}$$

where $L(y)$ is the likelihood ratio function given by:

$$L(y) = \frac{p_1(y)}{p_0(y)} \tag{9.4}$$

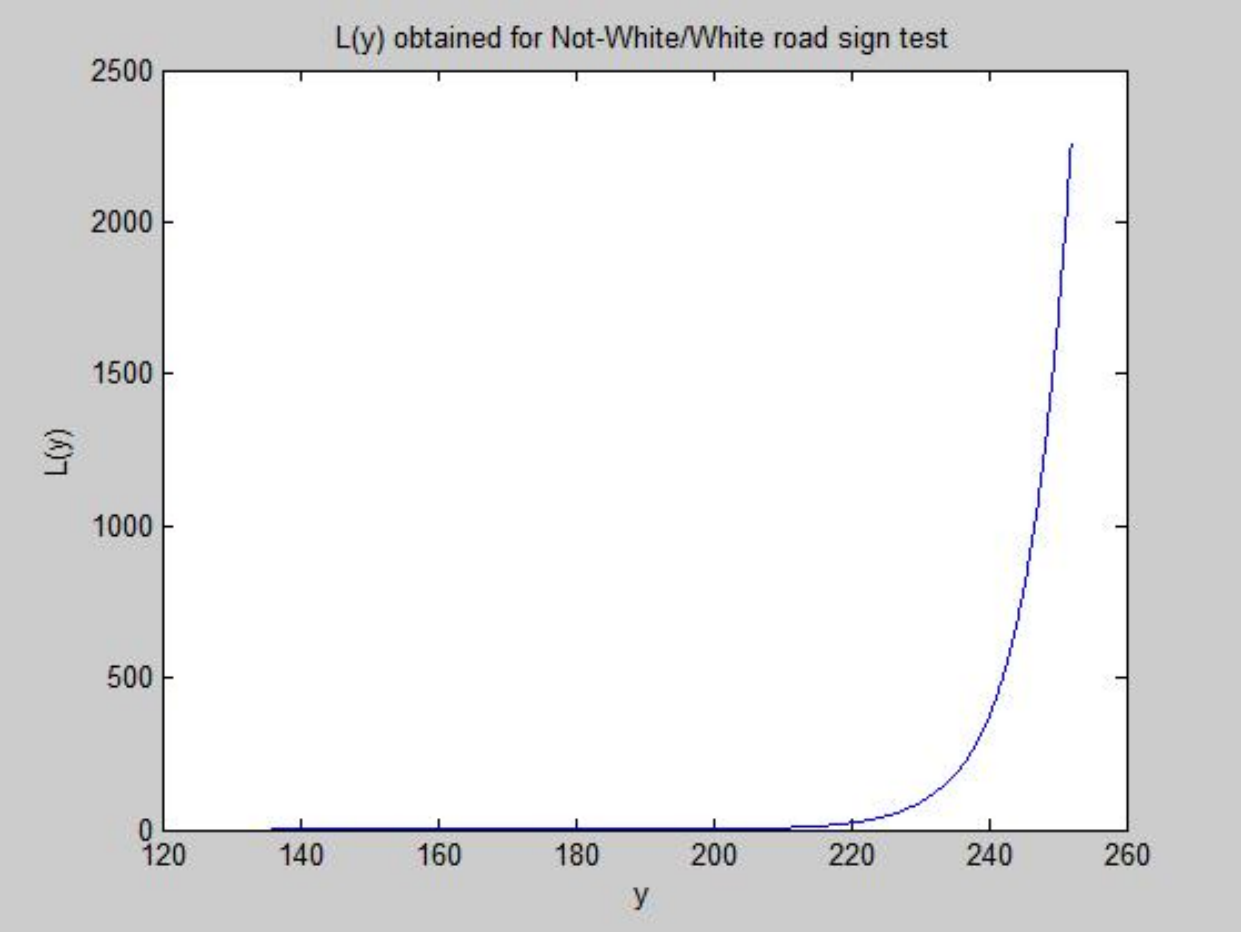


Figure 9.2: *Monotonic and continuous Likelihood function.*

Since in this case $L(y)$ is continuous and monotonic, as shown in Fig. 9.2, the relationship between $L(y)$ and η can be expressed as a relationship between y and some threshold y_η. Randomization is not necessary. Therefore, the decision rule for the Neyman-Pearson test

is:

$$\delta(y) = \begin{cases} 1 & \text{if } y \geq y_\eta \\ 0 & \text{if } y < y_\eta \end{cases} \tag{9.5}$$

The trade-off between probability of false alarm and probability of detection is attained by setting a bound on the probability of false alarm.

$$\max_{\delta} P_D(\delta) \;\; subject \;\; to \;\; P_F(\delta) \leq \alpha \tag{9.6}$$

Where $P_D(\delta)$ is the probability of detection, $P_F(\delta)$ is the probability of false alarm and α is the bound mentioned earlier and it is also known as the level or significance level of the test[25]. Probability of false alarm is computed as a function of y_η by finding the area under the curve $p_0(y)$ in the observation space of hypothesis H_1. This expression is equated to α. The decision threshold y_η is then expressed in terms of the false alarm α. P_D is then calculated as the area under the PDF $p_1(y)$ in the observation space of H_1. The expressions obtained after setting $P_F = \alpha$ for y_η and P_D are given as:

$$y_\eta = 118.337 - 54.57 erf^{-1}(2\alpha - 0.9995) \tag{9.7}$$

$$P_D = 6.4555 * 10^{-8} \left(15495907.18 - exp\left\{ \frac{y_\eta}{\sqrt{235}} \right\} \right) \tag{9.8}$$

The plot of P_D Vs P_F as shown in Fig. 9.3 is called the Receiver Operating Curve(ROC).

The effect of change in the value of alpha is illustrated in Fig. 9.4, Fig. 9.5 and Fig. 9.6. Increasing values of α lead to more false positives. Setting the threshold at the knee if the curve gives the best trade off between the probability of false alarm and probability of detection.

Hypothesis tests were not developed for all colors nor the location classifiers as some of the distributions were more complex. Instead, thresholds determined by generating ROC curves directly from the test data and setting the thresholds to the knee of the ROC curve. The next chapter describes how the ROC curves were estimated experimentally.

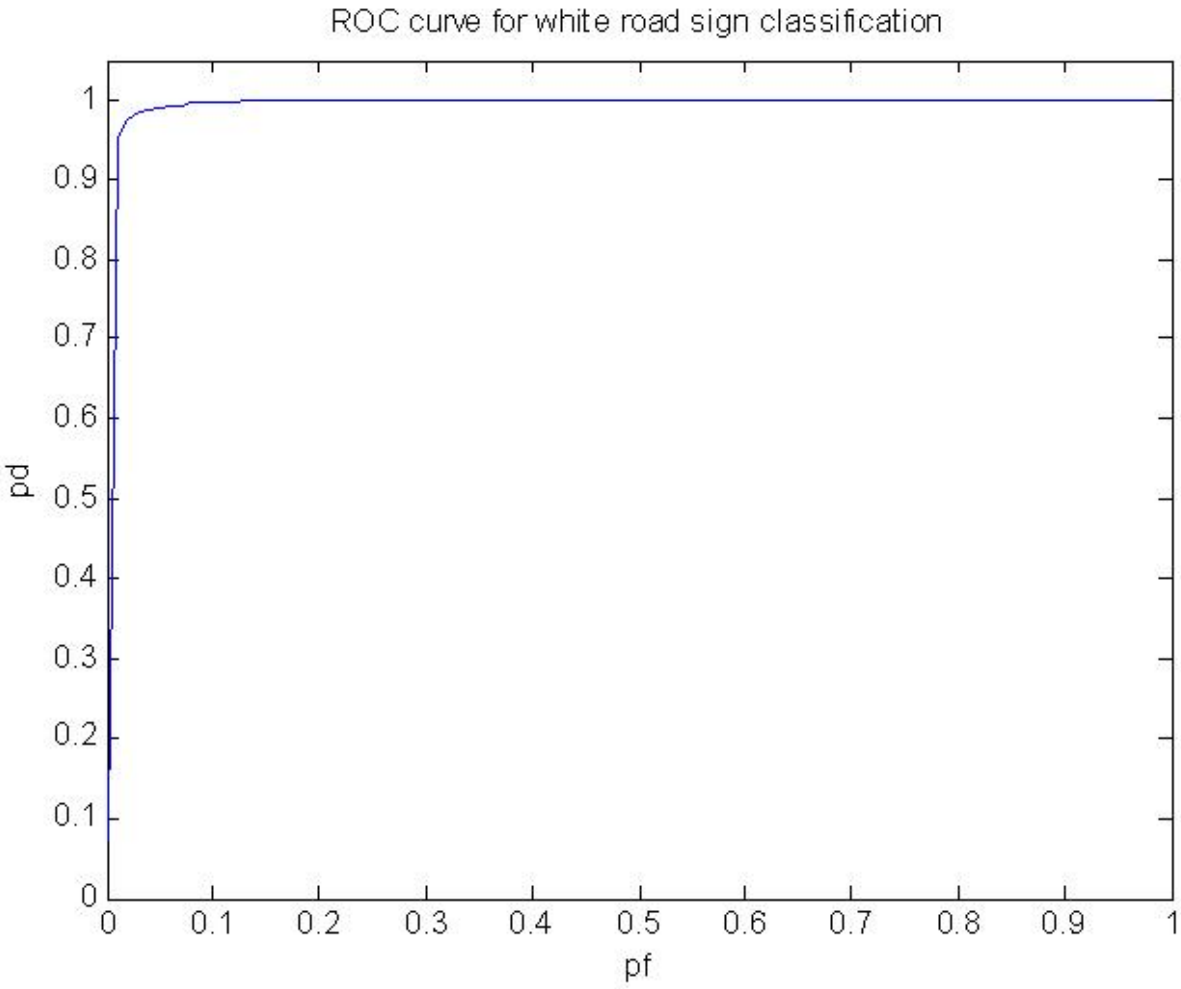


Figure 9.3: *ROC obtained for the test designed for Not-White/White road signs based on color*

Figure 9.4: *An image obtained from KDoT's road profiler.*

Figure 9.5: $\alpha = 0.1$,y_η *obtained is 172. More false positives. All the segments that have pixels whose Red, Green and Blue values are above* y_η *are classified as white road signs.*

Figure 9.6: $\alpha = 0.03$. *y_η obtained is 207. Lower value of alpha results in fewer false positives.*

Chapter 10

Receiver Operating Characteristic (ROC) Curves

10.1 ROC curve

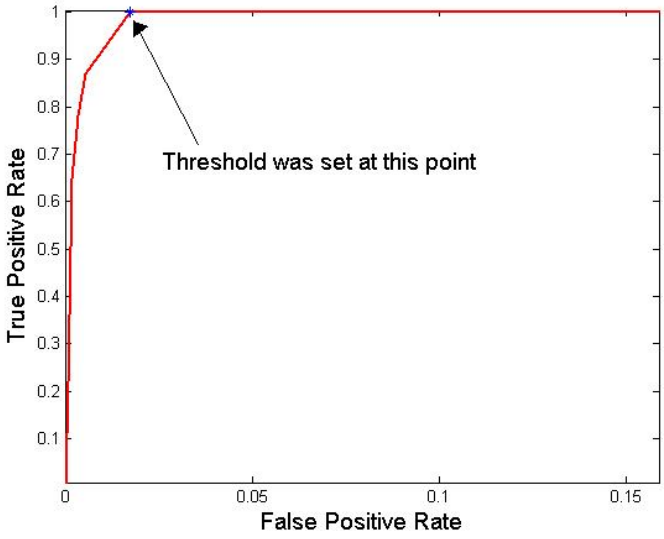
Classifiers produce four possible outcomes. If the classifier correctly identifies an object, it is a true positive (TP). If an object is incorrectly classified it is a true negative (TN). Objects of a class label not classified as such represent false negatives (FN). Likewise, objects not of the given type that are falsely identified as such are false positives (FP). The confusion matrix shown in Fig. 10.1b. ROC curves display the relationship between True Positive Rate (TPR) and the False Positive Rate (FPR) as the threshold is adjusted. The closer the curve is to the top left corner the better the performance[26]. The TPR and the FPR is calculated using:

$$TPR = \frac{TP}{TP + FN} \tag{10.1}$$

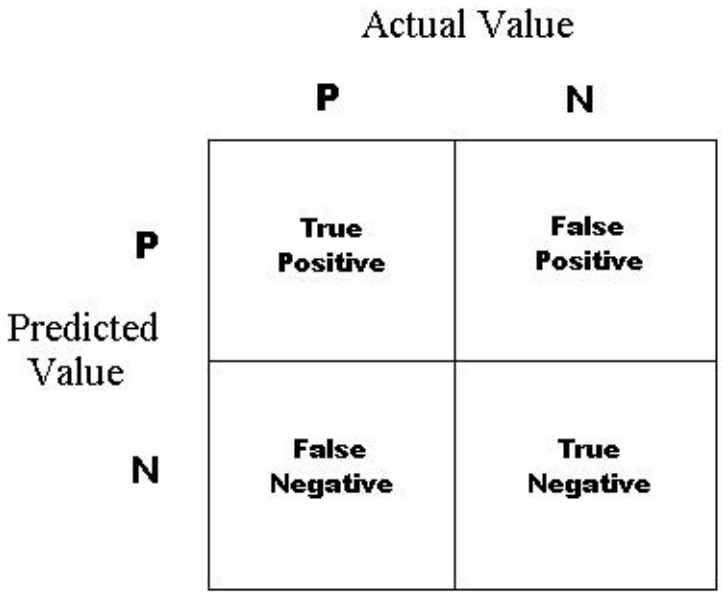
$$FPR = \frac{FP}{FP + TN} \tag{10.2}$$

10.2 Practical generation of ROC curves.

Using the database of manually labeled objects from 50 images, the ROC curves for both RGB color and XY distance thresholds and each sign color were constructed. An example



(a) *ROC curve relating the classification threshold to the TPR and FPR for the color classifier*



(b) *Confusion matrix*

Figure 10.1: *The arrow at the knee of the curves indicates an optimal threshold. A higher threshold minimally improves the TPR, but significantly increases the FPR. Note, since color performs well for classification the x-axis scale is magnified.*

ROC curve obtained for green colored road sign classifier is shown in Fig. 10.1a. In the example, with the threshold set at the knee as indicated all the segments belonging to road signs were identified and 1.8% of the non-sign segments were misclassified as road signs.

Chapter 11

Results

The thresholds for both the location and color classifiers were determined from the ROC curves. A new set of 100 images were tested with the algorithm. It was observed that all the road signs were identified with almost no false positives. Cases showing the classification of road signs in a pair of test images after location based classification are illustrated in Fig. 11.1 and Fig. 11.2.

Therefore, the 3-D location based classification is a more complete and optimal solution than the previously developed shape and 2-D location based classifiers.

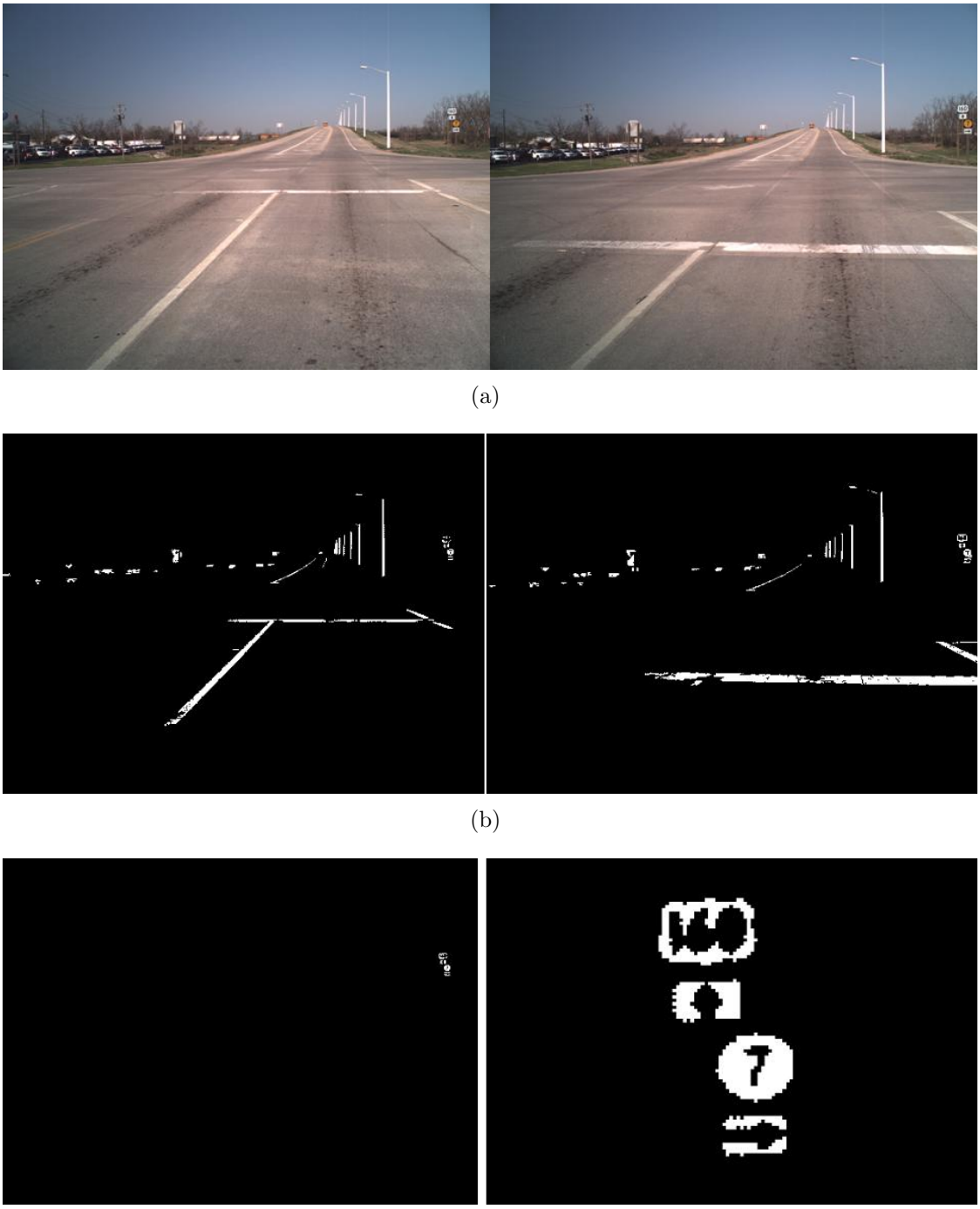
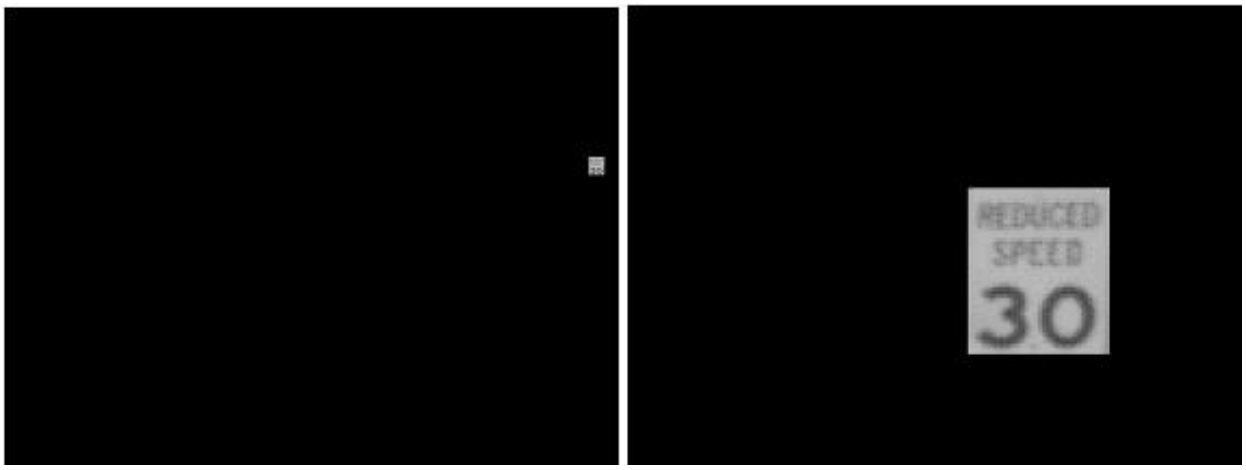


Figure 11.1: *Example of process: a. Original image pair., b. All objects classified as potentially road signs by their size and color., c. Location based classifier correctly distinguishes the signs from among all the remaining objects*

(a)



(b)

Figure 11.2: *With BA to estimate motion between frames, the triangulation and thus sign identification remains effective even on curved roads.*

Chapter 12

Future Scope

The current method can identify road signs, determine their type, 3-D location with respect to the camera and size. Future scope of the research is to determination of the distance between road signs of a particular type and investigate processes that take advantage of 3D information ascertained from triangulation such as measuring road easement profiles. Once the road signs are obtained, the size of the road signs can be estimated just as in the example case shown in Fig. 8.6. Future work will also include identification of the type of post on which the road sign is mounted with the possibilities being wood post, metal post or steel beam. This can be implemented with the prior knowledge of the location of the road sign and the color information.

Chapter 13

Conclusion

This paper describes the development of an automated system that identifies and inventories road signs from imagery obtained from KDoT's road profiling system of highways through out the state. Over-segmentation of images was achieved using K-Means clustering on first, the original image and then again on a difference image. This was followed by three stages of classification based on size, color and 3-D location of road signs relative to the camera. Triangulation and Bundle Adjustment were used to estimate the 3-D location of signs. Mahalanobis distance was used to measure the proximity of an object to being a road sign based on color and location. ROC curves were used to identify decision thresholds for the color and location classifiers. The system was tested and the corresponding results are documented. Future scope of this research is also presented.

Bibliography

[1] Y. L. Andrzej Ruta and X. Liu, *Real-time traffic sign recognition from video by class-specific discriminative features*, Elsevier Journal of Pattern Recognition, Received 1^{st} August 2007; revised 22^{nd} May 2009; accepted 26^{th} May 2009, Available online 8^{th} June 2009.

[2] C. Bahlmann, Y. Zhu, and M. Pellkofer, *A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information*, IEEE Intelligence Vehicles Symposium (*IV2005*), Las Vegas, 2005, p. 255-260.

[3] H. Fleyeh and M. Dougherty, *Road and Traffic sign detection and recognition*, 10th EWGT Meeting and 16th Mini-EURO, 2005.

[4] L. C. Gomez and O. Fuentes, *Color-Based Road Sign Detection and Tracking*, Image Analysis and Recognition (*ICIAR*), CA, 2007.

[5] M. Benallal and J. Meunier, *Real-Time Color Segmentation of road signs*, Canadian Conference on Electrical and Computer Engineering *IEEE CCECE*, p. 1823-1826, 2003.

[6] C.-C. C. Ming-Ni Wu, Chia-Chen Lin, *Brain Tumor Detection Using Color based K-Means Clustering Segmentation*, Third International Conference on International Information Hiding and Multimedia Signal Processing, *IIH-MSP*, Vol. 2, p. 245-250, 2007.

[7] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, *Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm*, *IEEE* Southwest Symposium on Image Analysis and Interpretation, p. 61-65, 2006.

[8] R. O. Duda, P. E. Hart, and D. G. Stock, *Pattern Classification*, volume 2nd Ed., Wiley - Interscience Publication, 2001.

[9] V. D and K. C, *Gaussian Mixture Modeling by Exploiting the Mahalanobis Distance*, *IEEE* Signal processing Transactions, 56, Issue 7, p. 2797-2811, 2008.

[10] V. A. Kremens, R. L. Ononye, and A. T. .C, *A Hybrid Contextual Approach to Wildland Fire Detection Using Multispectral Imagery, IEEE Geoscience and Remote Sensing Transactions, Volume 43, Issue 9, p. 2115-2126, 2005,*.

[11] V. F. N. C. E. V. Moraes. J. C. T. B., Seixas. M. O., *A real time QRS complex classification method using Mahalanobis distance*, Computers in Cardiology, p. 201-204, 2002.

[12] E. Persoon and K. Fu, *Shape Discrimination Using Fourier Descriptors*, *IEEE* transaction on systems, man, and cybernetics, Vol 7, p. 170-179, 1977.

[13] L. F. Costa and R. M. Cesar, Shape Analysis and Classification, Theory and Practice, CRC Press Boca Raton, Florida,, 2001.

[14] I. Kunttu, L. Lepisto, J. Rauhamaa, and A. Visa, *Multiscale Fourier descriptor for shape-based image retrieval*, pattern recognition, icpr, vol 2, p. 765-768, 2004.

[15] D. G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, p. 91-110, 2004.

[16] T. T. Y. A. Akira Suga, Keita Fukuda, *Object recognition and segmentation using SIFT and Graph Cuts*, *ICPR Pattern Recognition, p. 1-4, 2008,*.

[17] Z. Z. Xuelong Hu, Yingcheng Tang, *Video object matching based on SIFT algorithm*, Neural Networks and Signal Processing, p. 412-415, 2008.

[18] Z. Zhang, *Flexible camera calibration by viewing a plane from unknown orientations*, Seventh *IEEE* International Conference on Computer Vision, p. 666-673, 1999.

[19] Y. Sagawa, R.; Yagi, *Accurate calibration of intrinsic camera parameters by observing parallel light pairs*, *IEEE ICRA*, IEEE International Conference Robotics and Automation, p. 1390-1397, 2008.

[20] S. O. Heikkila, J., *four-step camera calibration procedure with implicit image correction.*

[21] J.-Y. Bouguet, *Camera Calibration Toolbox for Matlab*, *http* : *//www.vision.caltech.edu/bouguetj/calib*$_d$*oc/*, 2008.

[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, volume 2nd Ed., Cambridge University Press, 2003.

[23] C. Buckley, *Photomosaicing and automatic topography generation from stereo aerial photography*, Master's thesis, Department of Mechanical Engineering, Kansas State University, 2008.

[24] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, *Bundle adjustment - a modern synthesis*, springer-verlag, 2000.

[25] H. V. Poor, *An Introduction to Signal Detection and Estimation*, volume 2nd Ed., Springer Texts in Electrical Engineering, 1994.

[26] T. Fawcett, *An Introduction to ROC analysis*, Pattern Recognition Letters, special issue on ROC analysis, Vol. 27, p. 861-874, 2005.

AD ASTRA PER ASPERA

Kansas

Department of Transportation