AUTOMATED REAL-TIME OBJECT DETECTION AND RECOGNITION ON TRANSPORTATION FACILITIES

Final Report, an Exploratory Project

Principal Investigator: Kelvin CP Wang

Research Staff: ZHIQIONG HOU AND WEIGUO GONG

CIVIL ENGINEERING

UNIVERSITY OF ARKANSAS

Project Number: MBTC DHS 1103

February 2010

Prepared for
Mack-Blackwell Rural Transportation Center
National Transportation Security Center of Excellence
University of Arkansas

This material is based upon work supported by the U.S. Department of Homeland Security under Award Number: "2008-ST-061-TS0003

Disclaimer: The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

AUTOMATED REAL-TIME OBJECT DETECTION AND RECOGNITION ON TRANSPORTATION FACILITIES

Kelvin C.P. Wang, Zhiqiong Hou, and Weiguo Gong

Civil Engineering, University of Arkansas, USA

Final Report, an Exploratory Project

Partially Supported by

U.S. Department of Homeland Security under Award Number: "2008-ST-061-TS0003".

February 2010

Abstract: *Inventory of road signs are part of asset management systems for a roadway agency. Detection, recognition, and positioning of road signs are critical components of a roadway asset management system. In this research a stereo vision based system is developed to conduct automated road sign inventory. Such techniques may be also used to detect other objects on the road or by the roadside. The system in real-time integrates and synchronizes the data streams from multiple sensors of high-resolution cameras, Differential Global Positioning System (GPS) receivers, Distance Measurement Instrument (DMI), and Inertial Measurement Unit (IMU). Algorithms are developed based on data sets from the multiple positioning sensors to determine the positions of the moving vehicle and the orientation of the cameras. The key findings from the research include feature extraction and analysis that are applied for automated sign detection and recognition in the Right-Of-Way (ROW) images, implementing a tracking algorithm of the candidate sign region among the image frames so the same signs are not counted more than once in an image sequence, and implementing stereo vision technique to compute the world coordinates of the road sign from the stereo-paired ROW images. Particular techniques are employed to conduct all data acquisition and analysis in real-time on board the vehicle. This system is an advanced alternative to traditional inventory methods in terms of safety and efficiency. It is anticipated that future studies may employ techniques developed in the research to automatically detect the presence of man-made objects around roadway areas for security purposes.*

Contents

1 Introduction 4
2 System Components 5
2.1 Digital Camera 5
2.2 GPS-IMU System 6
2.3 DMI 7
2.4 Synchronizer 7
2.5 Software Solutions 7
3 Stereo Vision 7
3.1 Triangulation 7
3.2 Current Implementation 9
3.2.1 Virtual coordinates C1 10
3.2.2 Vehicle coordinates C2 10
3.2.3 World coordinates C3 11
3.3 Calibration 11
3.4 Stereo Camera Configuration 12
3.5 Sign Extraction and Tracking 12
4 Field Test 14
4.1 Calibration 14
4.2 Preliminary Test 15
4.3 Road Test 16
5 Conclusion 19
Reference 20

1 Introduction

Many government and private agencies have the need to perform roadway asset inventory on a regular basis. This practice has become increasingly common due to advancement of technology and relatively lower cost for data collection and processing. Particularly, many decades ago inventory of roadway signs and related assets was conducted manually with paper and pen. Along with the application of 16-mm and 35-mm film, VHS and S-VHS video, analog Laserdisc came with vehicular based acquisition systems. Today, nearly all data collection devices for roadway inventory assets are digital, vehicular based and operated at highway speed. This technology is commonly referred to as Right-of-Way (ROW) imaging, or simply as photo logging.

In recent years, there has been wide deployment of satellite based positioning systems for ROW imaging. For instance, GPS receivers are used to provide x, y, and z references. IMU is also frequently used to guarantee signal integrity during GPS outages, further improve positioning accuracy, and at times provide 100 Hz or higher update rate which is normally not available in many industrial grade GPS receivers. Integrated multi-sensor systems are increasingly used to provide cost-effective, robust solutions to the challenges of rapid collection and storage of the geo-referenced roadway imagery.

Even though ROW imagery can be well referenced with positioning data, the position of a road sign in an image is not directly known. Obtaining the sign positions requires hardware and software efforts to extract 3D positioning information for the objects (signs) present in the imagery. There are generally two solutions to obtain the object positions in the ROW imagery. Laser ranger (Laflamme et. al., 2006) shoots a low-power laser beam to the surrounding area and determines the position and distance of the object based on reflected laser. The advantages of the laser ranger are: 1) it has a wider view than a camera has, and 2) it gives accurate distance between the vehicle in motion and the object. The disadvantages of laser ranger are: 1) it requires expensive hardware and tremendous effort for system integration, and 2) it lacks visual appearance information and it still has to be used in conjunction with the images from the ROW imaging sub-system in the vehicle.

Stereo vision technique is the second solution. This technique requires no additional hardware, therefore can be cost-effective, and much simpler in terms of hardware integration and maintenance. The stereo vision technique, a method to extract the 3D position from the 2D images, was started in the photogrammetric community (Slama, 1980). It is frequently used in computer vision today. Many companies have developed hardware and software solutions which can be used in a wide range of industrial inspection tasks. Stereo vision applications are found in a variety of scientific, engineering, industrial, even cultural disciplines, including archaeology, architecture, e-commerce, forensics, geology, planetary exploration, movie special effects, and virtual and augmented reality (Faugeras, 1993).

A critical step in stereo vision technique is to establish correspondence across the stereo images. In this particular problem, the road sign recognition module in the system not

only locates the sign region and identifies the sign type, but also provides a group of feature points which can be used as stereo correspondences. Road sign recognition is a computer vision process concerned with pattern recognition and classification. It was found that the road sign recognition has been commonly separated into two phases in the literatures: Detection and Recognition (or Classification). Detection phase aims at decreasing the search space in the image by cropping out the candidate region. Recognition phase aims at recognizing whether the candidate region is a sign and identifying the sign type. Artificial Neural Network (ANN) has been a major researched technique for the classification. Research based on ANN has been particularly active during the last decade (Lafuente-Arroyo et al., 2005). Cross correlation has been another basic classification technique used in road sign classification problem (Paclik et al., 2006). In recent years, new techniques based on invariant feature extraction have gained more attention (Pierre and Pietro, 2005). Feature matching is then conducted using various methods such as: conditional random field classifier (Weinman et al., 2004), pseudo-likelihood cross-validation (Paclik et al., 2000), Matching Pursuit filter (Hsu and Huang, 2001), and Support Vector Machine (Silapachote et al., 2005; Cyganek, 2007 and 2008).

Additional related work conducted recently includes using Kalman filter and wavelet techniques for traffic forecasting (Xie, Zhang, and Ye, 2007), a new vision algorithm for sign detection (Hu and Tsai, 2009), and laser scanning based techniques for geometric modeling and health monitoring (Cai and Rasdorf, 2007; Park, Lee, Adeli, and Lee, 2007)

Even though stereo vision itself is not a new technique in automated imaging, the proper implementation for sign inventory requires developing proper design, hardware integration, and software algorithms. In this paper, the development of a stereo vision based road sign inventory system is presented. The research focuses on the feasibility, reliability, and the precision of the stereo vision technique used in the road sign inventory system. The paper also addresses the critical issues on the integration of the multiple sensors and the instantaneous feature extraction in the images.

2 System Components

The physical ROW imaging system is part of the Digital Highway Data Vehicle (DHDV) and shares common positioning sensors with other sub-systems in the DHDV. The conceptual design of the automated stereo vision system is that the ROW imaging system produces digital images at known coordinates, linear reference and pointing angle based on GPS, IMU, and DMI data sources. With the use of the stereo vision technique, the coordinates of the objects in the images can be subsequently calculated. Consequently, the collection of objects representing roadside asset is accomplished. The system integrates the following sensors:

2.1 Digital Camera

One or multiple professional digital camcorders can be used to capture the ROW images. The camcorder has a resolution of 1920*1080, or 1080p resolution. Its iris can be

automatically adjusted under adverse lighting environment. The images are streamed from the camcorder to the computer system memory at the default rate of 30 frames per second. The camcorder can be mounted anywhere on the vehicle, as long as it has a clear view of the roadway and is on the right side where signs are located. Typically it is mounted on a roof rack with an environmentally protected housing or on the dashboard. Digital imagery is recorded in real time into JPEG format on a hard drive at a recording frame rate that is pre-determined in the control software in the DHDV by the operator. The recording frame rate is normally less than 30 frames per second and calculated based on a fixed distance interval for each image. The positioning information of images is acquired from various sensors, synchronized in real-time, and saved into the on-board database.

2.2 GPS-IMU System

One crucial element of the system is the integration of GPS receivers and IMU. GPS and onboard IMU are complementary positioning devices. GPS provides the position of the vehicle. However, GPS signals can degrade sometimes due to various issues in the atmosphere even during clear view of the sky. In addition, GPS signal may degrade in circumstances where obstructions are present, such as an overhead bridge, trees and forest, a tunnel, hills, or skyscrapers. IMU consists of a triad of accelerometers and a triad of gyroscopes and continuously monitors position and acceleration of the vehicle. When the outputs of these six sensors are integrated with respect to time, the displacement and attitude are determined. However, it suffers from biases, drifting errors, and scale factor errors that cause the solution to degrade in time or over distance. Kalman filter is commonly used to improve the position measurement made from both components of GPS receiver and IMU. It is also able to estimate states where it has no direct measurement. For example, position and velocity are compensated directly, but other measurements like accelerometer bias, have no direct measurements. The Kalman filter tunes these parameters so that the GPS measurements and the inertial measurements match each other as closely as possible (Scherzinger, 2003). After integration, GPS data can be used to correct IMU errors with its long term stability and no error-growth characteristics. Any GPS outage or signal degradation can be alleviated with IMU data sets as well. In this system, the integrated GPS/IMU system is used to provide a direct geo-referencing for the vehicle, consequently the cameras (three position components and three attitude angles). The update rate of the IMU is 100 Hz, which provides more than one positioning data point for each foot of traveling distance at 60MPH (about 100KPH). The positioning accuracy of the standard GPS receiver is improved by either of the two satellites based differential correction services. These are Satellite Based Augmentation System (SBAS) and OmniStar. SBAS services, such as WAAS and EGNOS, are wide area differential corrections provided for free. They provide an accuracy of about 1.2m Circular Error Probability (CEP). Therefore, the integration of the GPS receiver and IMU improves the data positioning accuracy. Data post-processing can be also conducted using the base station coordinates which can improve the final positioning accuracy to centimeter level.

2.3 DMI

DHDV uses DMI to provide the linear reference along the route. DMI works as a vehicle odometer. For accurate measurement of distance, a pulse generator and an electronic interface amplifier are required to work together with a DMI. An electrical impulse is generated by sensors when the vehicle is traveling. The generated pulses are then sent to the electronic interface amplifier. The electronic interface amplifier divides and amplifies the pulses into the suitable working rate. The pulses from the electronic interface amplifier are then sent to trigger the cameras and various laser sensors for other sub-systems in DHDV. The pulse reading of each dataset can be mathematically converted to distance with a calibrated ratio.

2.4 Synchronizer

GPS, IMU, DMI and the camera(s) collect their data at different frequencies. The synchronization process is needed to integrate the information from multiple sensors. In DHDV, an electronic control device, Control Chassis, was developed to integrate information and synchronize the signals from the different sensors. Precise time registration, i.e. instantaneous geo-referencing is found to be a challenge. A high resolution clock with a 1000 pulse/ second frequency is used in the Control Chassis. The signal acquisition time can be interpolated with a rate of 1/1000 second. Trigger signal is send to the cameras based on DMI signal. Geo-reference data for each image is obtained by interpolating the closest available GPS/IMU data points.

2.5 Software Solutions

Software programs include two modules. One is geo-referenced data acquisition to synchronize and integrate the data sources from multiple sensors. Another module is asset extraction module which allows manual and automated asset extraction from the image data. It enables the determination of the position of the objects in the imaging environment.

3 Stereo Vision

3.1 Triangulation

The basic element of the stereo vision theory is triangulation (Wong, 1975). As shown in Figure 1, a 3D point can be reconstructed from its two projections by computing the intersection of the two space rays corresponding to it. The 3D location of that point is restricted to the straight line that passes through the center of projection and the projection of the object point. Binocular stereo vision determines the position of a point in space by finding the intersection of the two lines passing through the center of projection and the projection of the point in each image.

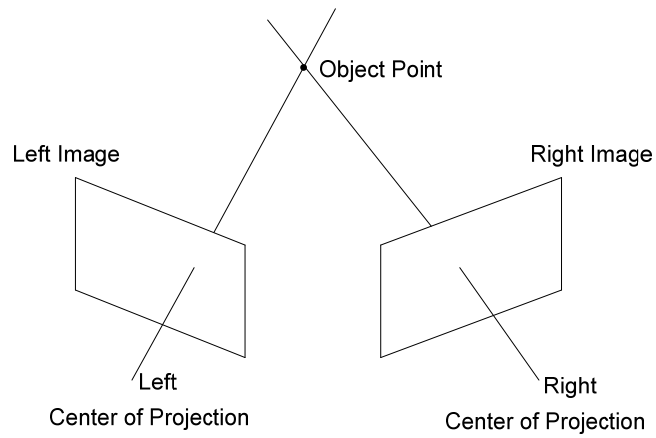


Fig. 1. Triangulation (Wong 1975).

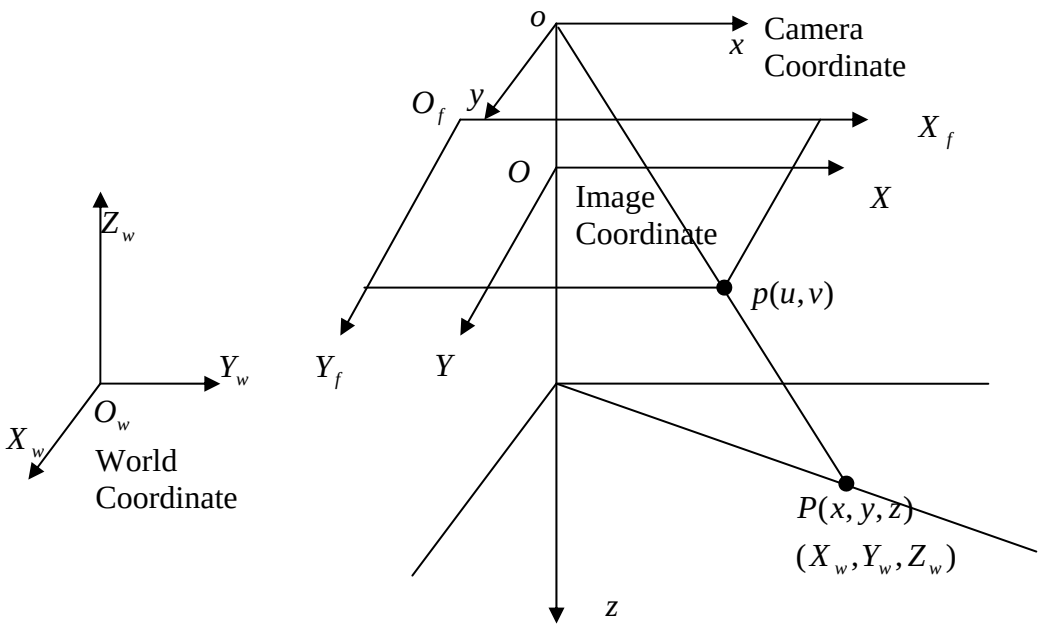


Fig. 2. The relation among the coordinates (Tsai 1987).

Triangulation describes the ideal relationship of the images and the objects. To build a mathematical model between the location of the object in the images and the 3D position of the object point, several coordinates' conversion (Figure 2) is involved, namely image coordinate ($X_f O_f Y_f$), camera coordinate (xoy), and world coordinate ($X_w O_w Y_w$).

Considering the triangulation and the coordinate's conversion, the relation between a 3D point P and its image projection p is given by

$$s\tilde{p} = A[R \quad t]\tilde{P} \tag{1}$$

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$P = [X, Y, Z]^T \tag{3}$$

$$p = [u, v]^T, \ \tilde{p} = [u, v, 1]^T \tag{4}$$

$p = [u, v]^T$ is a 2D point, $\tilde{p} = [u, v, 1]^T$, s is a scale.

A is camera intrinsic matrix. Intrinsic parameters characterize the inherent properties of the camera optics, including the focal length, the image centre, the image scaling factor and the lens distortion coefficients.

$P = [X_w, Y_w, Z_w]^T$ is a 3D point. Its vector form is written as $\tilde{P} = [X_w, Y_w, Z_w, 1]^T$ in equation (1).

$[R \quad t]$ is extrinsic parameters. It is the rotation and translation which relates the world coordinate system to the camera coordinate system.

However, this model is rather ideal. Calibration process has to be conducted to determine the internal (or intrinsic) parameters and external (or extrinsic) parameters.

3.2 Current Implementation

For this ROW application, each of the captured image sequences has been geo-referenced by using GPS/IMU integrated positioning device. The orientation parameters of each camera coordinate origin are determined with respect to a global coordinate system. By using techniques of photogrammetric intersection, the position of 3D object relative to the camera coordinates is achieved. To eventually calculate the world coordinates of the sign, the following coordinate transformation needs to be implemented, including C1, C2, and C3 coordinate systems:

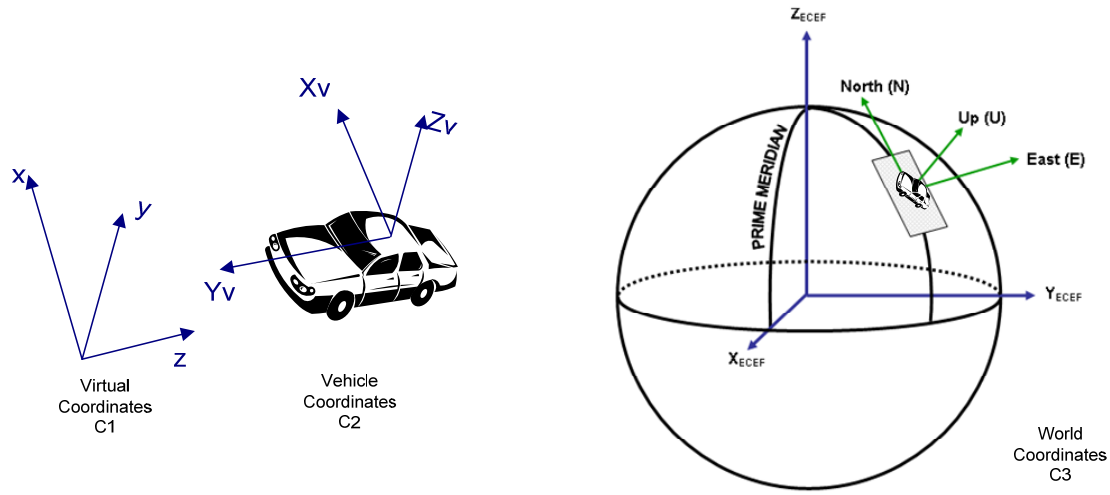


Fig. 3. Virtual coordinates, vehicle coordinates, and the world coordinates.

3.2.1 Virtual coordinates C1

This coordinate system is established during the calibration process. The origin of this coordinates is the origin selected in the calibration board. The points used in the calibration process are represented in this coordinate. Consequently, the location of the road sign is first represented in C1 coordinates from the 3D positioning function. The calibration board is purposely put in a plane perpendicular to the vehicle's y-direction (longitudinal direction). This is to exclude the rotation between the vehicle coordinates and the virtual coordinates. The offsets ($\Delta x, \Delta y, \Delta z$) between the origin in the virtual coordinates and the origin in the vehicle coordinates are measured during the calibration process.

3.2.2 Vehicle coordinates C2

While the vehicle is moving, its position is determined by the positioning sensors. The origin of the vehicle coordinates is a fixed point in the vehicle. For simplification, the location of the GPS receiver is set as the origin. The Y-axis is the forward direction (longitudinal) and the X-axis is point to the passenger's side (transverse). Once the sign location is obtained in the virtual coordinates C1, the task is then to convert it to the vehicle coordinates C2.

For example, if point P in the space can be represented as (x_1, y_1, z_1) in the C1 coordinates. It can be converted into C2 coordinates using the following equations:

$$\begin{aligned} X_v &= x_1 - \Delta x \\ Y_v &= -(z_1 - \Delta z) \\ Z_v &= (y_1 - \Delta y) \end{aligned} \tag{5}$$

3.2.3 World coordinates C3

With the heading, roll and pitch provided by IMU, the coordinates of point P, (X_v, Y_v, Z_v) in vehicle coordinates can be easily converted to (X'_v, Y'_v, Z'_v) in the ENU (local east, north, up) coordinates. The ENU coordinates can then be converted to ECEF (Earth Centered Earth Fixed) coordinates (X, Y, Z) using Equation 6 (Zhu, 1994).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\sin\lambda & -\sin\phi\cos\lambda & \cos\phi\cos\lambda \\ \cos\lambda & -\sin\phi\sin\lambda & \cos\phi\sin\lambda \\ 0 & \cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} X'_v \\ Y'_v \\ Z'_v \end{bmatrix} + \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} \qquad (6)$$

Whereas (X_G, Y_G, Z_G) is the ECEF coordinates of the GPS receiver, i.e. the origin of the vehicle coordinates or the ENU coordinates. λ and ϕ are the geodetic longitude and latitude. ECEF coordinates can be further converted to geodetic coordinates if needed (Zhu, 1994).

In the current implementation, the determination of relative distances and sizes of objects in the image pairs is an operation dependent only on the stereo cameras and its calibration.

3.3 Calibration

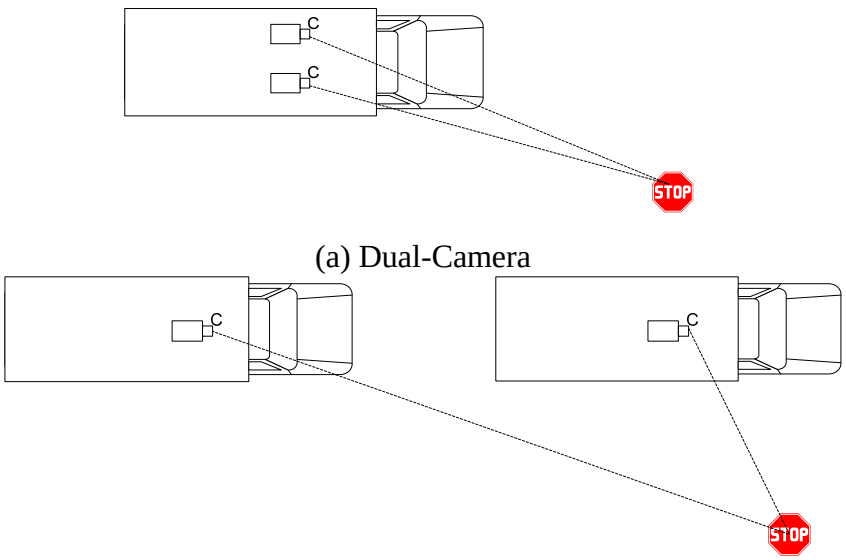
A study of the current calibration methods is conducted to search for the most proper method for implementation.

The techniques found in the literature for camera calibration can be broadly divided into photogrammetric calibration, self-calibration and something in between. There are three typical types of photogrammetric calibrations: 1) Linear methods: assume a simple pinhole camera model and incorporate no distortion effects. This method is non-iterative and fast (Abdel-Aziz and Karara, 1971; Wong, 1975; Ganapathy, 1984; Frugeras and Toscani, 1986). The limitation is the lens distortion effects can not be corrected. 2) Nonlinear methods: first the relationship between parameters is established and then an iterative solution is found by minimizing some error term (Brown, 1966; Haralick and Shapiro, 1993; Nomura et al., 1992). This category of methods requires a good initial guess to start the iteration. 3) Two-step techniques: it involves a direct solution of some camera parameters and an iterative solution for other parameters. This is the most commonly used approach to the problem (Tsai, 1987; Lenz and Tsai, 1988; Weng, 1992). Another category of calibration method is called self calibration. Techniques in this category do not use any calibration object. The calibration is conducted by moving a camera in a static scene (Zhang, 2000). Self calibration is more appealing to our problem due to its flexibility. However, the development of this method is not mature.

The method used in the research is based on Zhang's method, the details of which can be found in Zhang's paper (Zhang, 2000). This approach lies between the photogrammetric calibration and self-calibration. The reason for adopting this approach is that it is more

flexible than photogrammetric calibration and it gains considerable degree of robustness compared with self-calibration. It only requires the camera to observe a planar pattern shown at a few (at least three) different orientations. Either the camera or the planar pattern can be moved by hand. The motion does not need to be known.

3.4 Stereo Camera Configuration



(a) Dual-Camera

(b) Single-Camera

Fig. 4. Stereo pair-image formations.

The performance of the stereo vision can vary with different camera configurations. Figure 4 shows two different formations of the stereo pair-image. Figure 4a) shows a more conventional configuration, which is formed from two images taken by two cameras at the same time. Figure 4b) shows a single-camera stereo vision, which is sometimes also called structure from motion (SFM) configuration. The stereo pair-image is formed by two frames generated by the same camera and taken at different time. Finding structure from motion presents a similar problem as finding structure from stereo vision. In this project, both camera configurations have been tested.

3.5 Sign Extraction and Tracking

Vast amount of image data can be collected at highway speed. Therefore, rapid and accurate extraction of features of interest from the image stream is still a substantial challenge in both academic and industry circles. Manual feature extraction represents a bottleneck in the processing flow, and is the predominant practice today despite many years of research. The sign extraction module in this research includes capabilities of automatically determining presence of signs, classifying any number of road signs with different colors and shapes in a rapid fashion, and measuring sign dimensions in image sequences, all conducted in real-time.

First, raw roadway images are classified into different color bands based on color segmentation. The threshold used in the segmentation is obtained from the statistical test of the real images. After color segmentations, there are blobs generated in the binary images. Blobs meet the specific shape and color criteria will be detected as candidate region and trigger the recognition. The recognition is based on Principal Component Analysis (PCA) (Joliffe, 1986) and Support Vector Machine (SVM) (Chang, 2001). PCA algorithm is applied to extract the features of the regions of interest. These PCA features are input into the SVM model to do the classification.

PCA is a useful statistical technique that has found application in fields such as face recognition and image compression. Mathematically, it is an orthogonal linear transformation that transforms the data to a new coordinate system, by which the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA involves the computation of the eigenvalue decomposition or singular value decomposition of a data set, usually after mean-centering the data for each attribute. Then the PCA features, obtained as the first several principal components, can be used in image interpretation and classification. For road sign images, PCA deducts the dimension of the data set by retaining those characteristics of the data set that contribute most to its variance. This property makes it a good tool to extract the features of the road sign.

These features later are used as inputs into the SVM model to conduct the road sign classification. SVM model predicts whether a new example falls into one category or the other with a given set of training examples. A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall onto. As shown in Figure 5, the black dots and the white dots are the training examples which belong to two categories. The Plane H series are the hyperplanes to separate the two categories. The optical plane H is found by maximizing the margin value $2/\|w\|$. Hyperplanes H_1 and H_2 are the planes on the border of each class and also parallel to the optical hyperplane H. The data located on H_1 and H_2 are called support vectors.

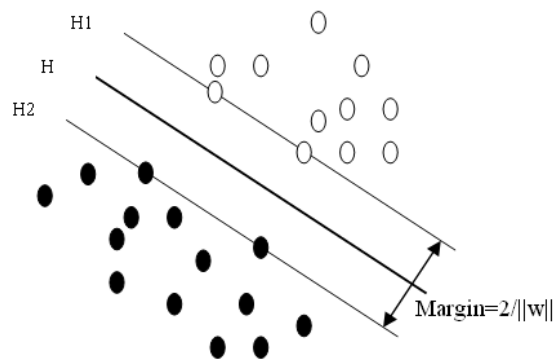


Fig. 5. The SVM binary classification.

A standard road sign image library is developed based on collected road sign images in the field. The images are captured under variant lighting conditions. Part of the images in the library is used for training and others for testing. The images are trained with a one vs. the other method. LIBSVM (a LIBrary for SVM) was employed for classifying training. The details of how to use LIBSVM can refer to the paper by Chang (2001). Once the SVM model is built up, proper class is assigned to each testing image.

Other than the automated sign extraction module, the software module also provides a capability for the user to manually extract the sign along a quality control function. Once sign extraction has been accomplished, positioning coordinates are assigned to each sign, and height and width measurements are made.

As contiguous image frames may contain same signs over a distance, it is important that same signs are not identified as separate multiple signs in the software module. A Kalman filter based tracking algorithm (Wang, 2006) is implemented in the software module to assure that single signs are correctly tracked and inventoried. The application of Kalman filter is to predict the location and size of the candidate region in future frames based on the sign and the size of the candidate region in the current frame. The Kalman filter technique includes two phases: time update and measurement update. The time update procedure is based on the dynamic equation which is derived from the spatial constraints from the two successive frames. The measurement update is based on the image processing location in the proximity of the predicted candidate region. This method tremendously reduces searching area in images, and decreases searching time.

4 Field Test

In this research, a trial of sign detection is conducted to evaluate accuracy of the system. During initial setup and any subsequent changes of camera relative positions to the vehicle and between themselves, calibration must be conducted.

4.1 Calibration

The calibration is conducted in the lab before data collection. The two cameras are mounted on a rack and the entire rack is fixed in the same position for data collection. A

checker board with a grid size of 2.5 inches is used in the calibration with the following calibration process:

1) Move the relative position of the camera to the calibration board four times and take four images T1, T2, T3, and T4.

2) Lock up the camera on the rack which is going to be mounted on the vehicle. Make the calibration board parallel to the X axis of the rack, perpendicular to the Y axis of the rack (C2 and Figure 3). Take the last image T5. This is to exclude the rotation angles between C1 and C2 (Figure 3) and simplify the intermediate calculation.

3) Use custom-made calibration software to detect the feature points on image T1-T5 and record the image coordinates of these points.

4) Determine the global coordinates of the feature points on the camera board. Use the bottom left feature point in the calibration board as the origin, right direction as X, up as Y, the direction toward camera is Z.

5) Input virtual coordinates and image coordinates of the feature points in the five images (T1-T5) to the calibration software. The output will be camera intrinsic matrix A and extrinsic parameters $\begin{bmatrix} R & t \end{bmatrix}$ (Equation 1).

A similar calibration is conducted for the single-camera configuration.

4.2 Preliminary Test

Upon completion of the calibration process, a preliminary test is conducted in the lab. The goal of the test is to examine the factors that will affect the accuracy of the stereo vision system with dual-camera configuration. An object in the lab with a certain size is put at different distance from the cameras. The distance and the size of the object are measured using the stereo vision algorithms. The measured results are compared with the true values and be listed in Table 1 and Table 2. It is found that: 1) the error increases as the distance between the object and the cameras increases. 2) the accuracy of the result is improved with longer baseline length for dual camera configuration. However, due to the limitation of the width of the vehicle body, the baseline can not be expanded as much as desired.

Table 1
The test result for two-camera configuration.
(Base line = 40.64 mm, Camera Angle = 7.9°)

Test No.	*Size (mm)*	*Distance (mm)*	*Size error*	*Distance error*
1	183.9	2901.9	1.19%	1.05%
2	183.9	4467.2	3.19%	2.79%
3	183.9	9347.2	6.09%	7.69%
4	183.9	12344.4	6.48%	9.39%

Table 2
The test result for two-camera configuration.
(Base line = 83.82mm, Camera Angle = 8.1°)

Test No.	*Size (mm)*	*Distance (mm)*	*Size error*	*Distance error*
1	183.9	3708.4	0.22%	0.60%
2	183.9	6064.2	1.41%	1.33%
3	183.9	9912.4	2.12%	3.01%
4	183.9	14300.2	3.36%	2.63%

4.3 Road Test

The stereo vision system, coupled with other positioning sensors in the DHDV, is used to conduct road sign inventory survey test shown in Figure 6. A road loop route near the research site in Fayetteville, Arkansas is chosen for the test. There are 52 signs in all along the route selected in the experiment. A selected 26 signs are shown in Figure 6 due to space limitation. Image data is collected while DHDV is driven at 25-45 MPH, largely following the speed limit on the road. The traveling path of the survey vehicle is rectangular in shape per the data gathered with the 100 Hz IMU. The pins in the Figure 6 indicate the road signs that are automatically detected and positioned using the developed system. The reference positioning coordinates of GPS format for all signs are obtained using a professional level handheld GPS unit with an accuracy of 0.3 meter in planar positioning. The difference in planar positioning between the handheld GPS results and the results obtained from the ROW imaging system are 5-18 meters from the dual-camera configuration and 1-3 meters from the single-camera configuration.

The results from the single-camera configuration are more accurate than those from dual-camera configuration. During the experiment, several parameters are found to be important to positioning accuracy, some of which may have influenced the outcome.

- The accuracy of the start position (GPS/IMU)
- Calibration of the internal parameters of the camera(s) in use (focal length, position of principal point, pixel size, pixel spacing, lens distortion, etc.)
- Calibration of the equipment configuration on the survey vehicle (camera orientation, distances from positioning system)

- Distance between the camera and the objects that are being measured or positioned
- The pixel spacing which is related to the zooming factors of the camera
- Synchronization capabilities of the acquisition system relative to image and geographic position data capture
- Timing control of the system clock in the acquisition system, which is used as a critical control factor along with results of DMI and IMU to determine longitudinal distance of traveling

For instance, the synchronization of image sequence from both cameras in the dual-camera configuration may have deteriorated the positioning accuracy due to timing control error in the operating system. Another possible contributing factor is the spacing of the two cameras may be too short in the lateral direction. A further improvement of the timing control is needed at less than 1 millisecond accuracy.

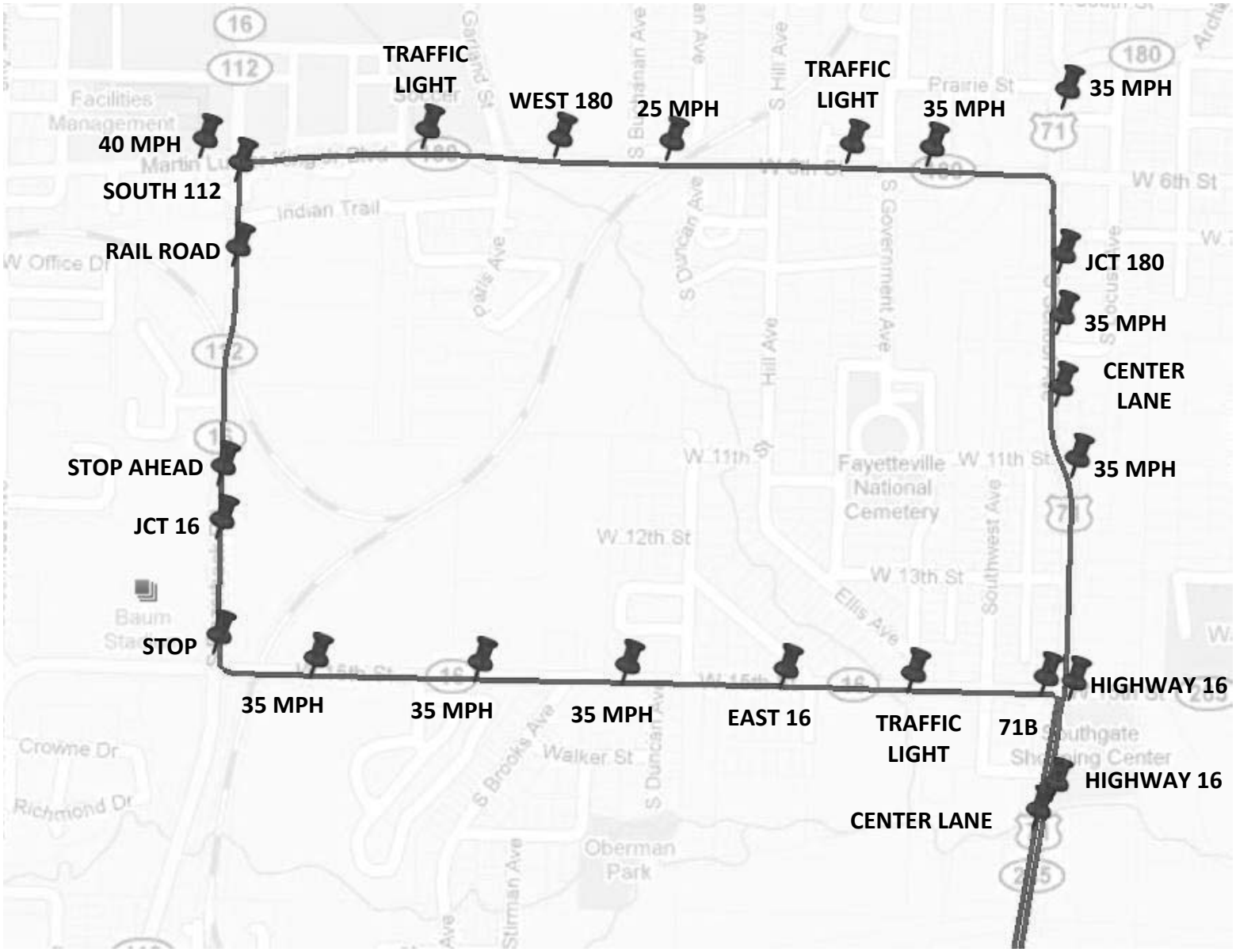


Fig. 6. The map for the test site with the detected road signs.

5 Conclusion

In recent years, the need to satisfy various asset management requirements prompted roadway agencies to start collecting various roadside manmade structures. Roadway signs are a primary component of these assets. However, since the start of using digital imagery for right-of-way survey in the 1990's, investment through both public and private sectors in developing automated asset inventory has stalled in recent years, despite substantial progresses have been made in the automation of road sign detection, recognition, and positioning. This paper represents a renewed effort to systematically design and integrate a real-time system for both acquisition and processing. A working level hardware system housed in the Digital Highway Data Vehicle (DHDV) has been developed and initial versions of calibration and processing software have been tested. The accuracy of the developed stereo vision system was evaluated via a case study by comparing them to locations measured by a handheld precision GPS receiver. This study concludes that the proposed stereo vision based automated road sign inventory system has achieved acceptable accuracy. The ultimate goal is to develop and implement a fully automated system to conduct sign inventory for all three objectives: detection, recognition, and positioning. In the future more extensive tests shall be conducted on larger road networks of both interstate highways and local streets. Precision and bias based on certain benchmarks relating to the three objectives will need to be established and studied. Reflectivity and conditions of road signs need to be evaluated as well. It should be pointed out that LIDAR technology has been experimented in recent years to detect presence of signs and other manmade objects on or near roads. It is envisioned that in the next few years, several types of fully automated systems may emerge in the market place for sign inventory. The techniques described in the research may be of significance to future studies which focus on the detection of the presence of man-made objects around roadway areas.

Reference

Abdel-Aziz, Y., & Karara, H. (1971), Direct linear transformation into object space coordinates in close-range photogrammetry. Proceedings of the Symposium on Close-Range Photogrammetry, 1- 18.

Brown, D., (1966), Decentering distortion of lenses. *Photogrammetric Engineering Remote Sensing*, 444- 462.

Cai, H. and Rasdorf, W. (2007), "Modeling Road Centerlines and Predicting Lengths in 3-D Using LIDAR Point Cloud and Planimetric Road Centerline Data," Computer-Aided Civil and Infrastructure Engineering, 23:3, pp. 157-173.

Chang, C. & Lin, C. (2001), LIBSVM: a Library for Support Vector Machines, available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Cyganek (2008), B. , "Colour Image Segmentation with Support Vector Machines: Applications to Road Signs Detection," International Journal of Neural Systems, Vol. 18, No. 4, pp. 339-345.

Cyganek, B. (2007), "Circular Road Signs Recognition with Soft Classifiers," Integrated Computer-Aided Engineering, Vol. 14, No. 4, pp. 323-343.

Faugeras, O. (1993), *Three-Dimensional Computer Vision*. MIT Press.

Faugeras,O. & Toscani, G. (1986), Calibration problem for stereo, Proceedings of the International Conference on Computer Vision Pattern Recognition, 15- 20.

Ganapathy, S. (1984), Decomposition of transformation matrices for robot vision, Proceedings of the IEEE International Conference on Robotics and Automation, (New York: IEEE), 130- 139.

Joliffe, I. T. (1986), *Principal Component Analysis*, Springer-Verlag.

Haralick, R. & Shapiro, L. (1993), *Computer and Robot Vision*, **2** (Reading, Massachusetts: Addison-Wesley), 125- 178.

Hsu, S.H. & Huang, C.L. (2001), Road Sign Detection and Recognition Using Matching Pursuit Method. In *Image and Vision Computing*, **19**, 119–129.

Hu, Z. and Tsai, Y. (2009), "A Homography-Based Vision Algorithm for Traffic Sign Attribute Computation, "Computer-Aided Civil and Infrastructure Engineering," 24:6, pp. 385-400.

Laflamme, C., Kingston T., & McCuaig, R. (2006), Automated Mobile Mapping for Asset Managers. XXIII International FIG Congress, Munich, 8-13 October.

Lafuente-Arroyo, S., Gil-Jimenez, P., Maldonado-Bascon, R., Lopez-Ferreras, F. & Malsonado-Bascon, S. (2005), Traffic sign shape classification evaluation I: SVM using Distance to Borders, IEEE Proceedings of Intelligent Vehicles Symposium.

Lenz, R., & Tsai, R. (1988), Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**, 713- 720.

Nomura, Y., Sagara, M., Naruse, H., & Ide., A. (1992), Simple calibration algorithm for high-distortion lens camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 1095- 1099.

Paclik, P., Novovicova, J., & Duin, R. P. W. (2006), Building Road-Sign Classifiers Using a Trainable Similarity Measure. *IEEE Transactions on Intelligent Transportation Systems*, **7**(3), September.

Paclik, P., Novovicova, J., Pudil, P., & Somol, P. (2000), Road sign classification using laplace kernel classifier. *Pattern Recognition Letter*, **21** (13-14), 1165–1173.

Park, H.S., H.M. Lee, H.M., Adeli, H., and Lee, I. (2007), "A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning," Computer-Aided Civil and Infrastructure Engineering, 22:1, pp. 19-30.

Pierre, M. & Pietro, P. (2005), Common-Frame Model for Object Recognition. In Lawrence K. Saul, Yair Weiss, and Leon Bottou, editors, *Advances in Neural Information Processing Systems* 17. MIT Press, Cambridge, MA.

Silapachote, P., Weinman, J., Hanson, A., Weiss, R., and Mattar, M. A. (2005), Automatic Sign Detection and Recognition in Natural Sciences. IEEE Workshop on Computer Vision Applications for the Visually Impaired, San Diego, June.

Slama, C. (1980), *Manual of Photogrammetry*, 4th edition, American Society of Photogrammetry, Fall Church, Virginia, USA.

Scherzinger, B.M. (2003), *Precise Robust Positioning with Inertial /GPS RTK*. ION-GPS.

Tsai, R. (1987), A versatile camera calibration technique for high-accuracy 3D machine vision metrology using on-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation*, 323- 344.

Wang, K.C.P., Hou, Z., Gong, W.G., & McCann, R. (2006), A Kalman Filter based Tracking System for Automated Inventory of Roadway Signs, *Transportation Research Record*: Journal of the Transportation Research Board, **1968**, Washington, D.C.

Weinman, J., Hanson, A., and McCallum, A. (2004), Sign detection in natural images with conditional random fields. In Proc. Of IEEE International Workshop on Machine Learning for Signal Processing, Sao Luis, Brazil, September, 549-558,

Weng, J., Cohen, P., & Herniou, M. (1992), Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 965- 980.

Wong, K. (1975), Mathematical formulation and digital analysis in close range photogrammetry. *Photogrammetric Engineering Remote Sensing*, **41**, 1355- 1373.

Xie, Y., Zhang, Y., and Ye, Z. (2007), "Short-term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition," Computer-Aided Civil and Infrastructure Engineering, 22:5, pp. 326-334.

Zhang, Z. (2000), *A Flexible New Technique for Camera Calibration*, IEEE-PAMI **22** (11), 1330-1334.

Zhu, J. (1994), "Conversion of Earth-centered Earth-fixed coordinates to geodetic coordinates," *IEEE Transactions on Aerospace and Electronic Systems*, **30**, 957-961.