Dynamic Calibration of Pan–Tilt–Zoom Cameras for Traffic Monitoring

Kai-Tai Song, *Associate Member, IEEE*, and Jen-Chao Tai

Abstract—Pan–tilt–zoom (PTZ) cameras have been widely used in recent years for monitoring and surveillance applications. These cameras provide flexible view selection as well as a wider observation range. This makes them suitable for vision-based traffic monitoring and enforcement systems. To employ PTZ cameras for image measurement applications, one first needs to calibrate the camera to obtain meaningful results. For instance, the accuracy of estimating vehicle speed depends on the accuracy of camera calibration and that of vehicle tracking results. This paper presents a novel calibration method for a PTZ camera overlooking a traffic scene. The proposed approach requires no manual operation to select the positions of special features. It automatically uses a set of parallel lane markings and the lane width to compute the camera parameters, namely, focal length, tilt angle, and pan angle. Image processing procedures have been developed for automatically finding parallel lane markings. Interesting experimental results are presented to validate the robustness and accuracy of the proposed method.

Index Terms—Background segmentation, camera calibration, image measurement, image processing, lane-marking detection, traffic monitoring.

I. INTRODUCTION

CLOSED-CIRCUIT television (CCTV) cameras have been widely used for traffic monitoring and surveillance applications. For a vision-based traffic monitoring system (VTMS), the basic function is to extract automatically real-time traffic parameters, including flow rates, average vehicle speeds, traffic offense levels, etc., through image processing techniques [1]–[4]. Traffic parameters, such as vehicle speed, are very often obtained using image tracking techniques. The accuracy of traffic parameter estimation is thus affected by both the camera parameters and the tracking algorithm. VTMS works only if the cameras are calibrated properly, and its accuracy is very sensitive to the calibration results. Moreover, to obtain a flexible view and observation range, an increasing number of CCTV systems rely on movable cameras with adjustable pan/tilt and zoom settings. Proper calibration of the parameters of pan–tilt–zoom (PTZ) cameras plays an important role in vision-based traffic applications.

Most calibration methods [5]–[8] utilize known features in a scene to estimate the camera parameters, including tilt angle, pan angle, and focal length. In [9] and [10], sets of parallel lines of a hexagon are employed to estimate the camera parameters. Results from these presentations demonstrate that parallel lines can be employed to adequately determine the camera parameters. Effective algorithms have been developed for estimating the camera parameters using parallel lanes in a traffic scene [11]–[13]. Bas and Crisman [11] used the height and the tilt of the camera along with a pair of parallel lines in a traffic scene to calibrate the camera. Their approach, however, needs special manual operations to measure the tilt of the camera. In [12] and [13], multiple parallel lanes and a special perpendicular line were used to calibrate the camera parameters. The drawback of their design is that the perpendicular line seldom appears in a traffic scene. Moreover, the lane markings need to be manually assigned in the aforementioned methods. It will not be practical for a traffic monitoring system using PTZ cameras where manual operation should be avoided. Therefore, it will be necessary for the VTMS to possess the capacity of dynamic calibration of PTZ cameras.

In their recent presentation of dynamic calibration [14], Schoepflin and Dailey employed the trajectories as well as the bottom edges of vehicles to obtain two sets of parallel lines for PTZ camera calibration. The calibration procedure can be automated using the presented approach. However, the accuracy is considerably sensitive to the trajectories of vehicles in traffic imagery. Furthermore, to obtain reliable tracks with high quality, the system takes a longer time to capture more image frames for recording the recognizable tracks of vehicles. It becomes very time consuming and cannot meet the real-time requirement of the VTMS. For practical applications, a method to speed up the process and obtain stable results demands urgent attention.

In this paper, a novel focal length equation will be derived to estimate the PTZ camera parameters. The derivation requires only a single set of parallel lane markings, the lane width, and the camera height. Compared with existing approaches, the proposed method has the advantage of requiring neither the camera tilt information nor multiple sets of parallel lines. Furthermore, an image processing algorithm is also proposed to automatically locate the edges of the lane markings. Using lane-marking edges and the derived focal length equation, one can estimate the focal length and the tilt and pan angles of a PTZ camera.

K.-T. Song is with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: ktsong@mail.nctu.edu.tw).

J.-C. Tai is with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. and also with the Department of Mechanical Engineering, Minghsin University of Science and Technology, Hsinchu 300, Taiwan, R.O.C.
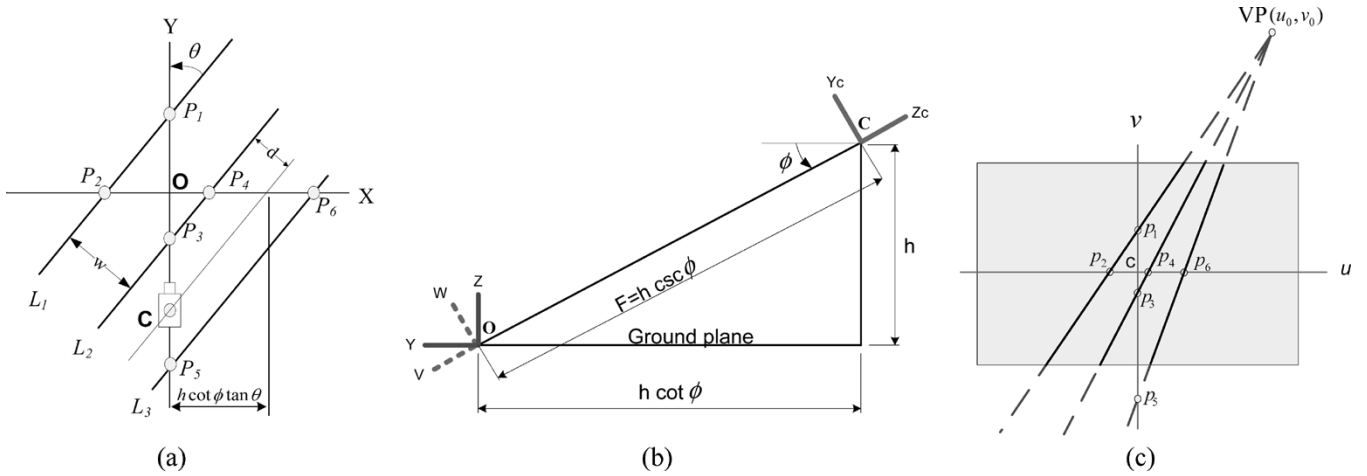
Fig. 1.   Coordinate systems used in the PTZ camera calibration. (a) Top view of the road map on the world coordinate system. (b) Side view of the camera setup and its coordinate systems used in calibration. (c) Road schematics used in the pixel-based coordinate system.

The rest of this paper is organized as follows. Section II presents the derivation of camera calibration equations for the focal length, as well as the pan and tilt angles. Section III describes the image processing algorithms for lane-marking detection. Synthetic sensitivity analysis and experimental results of the camera parameter estimation are presented in Section IV. Section V summarizes the contribution of this paper and future developments. Detailed derivation of the focal length equation is presented in Appendix A. Appendix B describes the conversion between pixel coordinates and world coordinates.

II. PROPOSED CALIBRATION METHOD

The objective of camera calibration is to determine all the required parameters for estimating the world coordinates from the pixel coordinates (u, v) of a given point in an image frame. In the following presentation, it is assumed that a change in the camera height and intrinsic parameters, except for the focal length, is negligible as the camera view changes. These parameters can be considered as fixed in vision-based traffic applications and are calibrated only once during the PTZ camera installation. A method for computing the changeable camera parameters, including the focal length and the pan and tilt angles, will be presented below.

Fig. 1 illustrates three coordinate systems utilized in the derivation, namely: 1) the world coordinate system (X, Y, Z); 2) the camera coordinate system (X_c, Y_c, Z_c); and 3) the camera-shift coordinate system (U, V, W). Fig. 1(a) depicts the top view of a ground plane in the world coordinate system. Lines L_1, L_2, and L_3 represent parallel lane markings, and point O is the origin of the world coordinate system on the road plane. The pan angle θ is defined by the angle between the Y axis and lane markings, f is the focal length, and w is the width between parallel lanes. The symbol d denotes a shift distance, which is a perpendicular distance between the projection of the principle point of the camera and L_3. Fig. 1(b) depicts the side view of the road scene, which is used to describe the geometrical relation between the ground plane and the camera; the direction of vector $\overrightarrow{CO}$ is perpendicular to the image plane. In Fig. 1(b), ϕ is the tilt angle of the camera,

h is the installed camera height, and F is the length of vector $\overrightarrow{CO}$. In this paper, the counterclockwise rotation is positive in expressing the sign of the angles.

The camera-shift coordinate system can be obtained by rotating the world coordinate system an angle ϕ around the X axis. The relationship between the camera-shift coordinate frame and the world coordinate frame is given by

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{1}$$

By shifting the camera-shift coordinate frame from point O to point C along the vector $\overrightarrow{OC}$ and inversing the V axis of the camera-shift coordinate frame, one can obtain the camera coordinate system. The camera coordinates of any point on the road plane (where Z equals to zero) can be expressed as a function of the world coordinate via a coordinate transformation between the camera-shift coordinate frame and the world coordinate frame, i.e.,

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} U \\ W \\ -V - F \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \sin\phi \\ 0 & -\cos\phi \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ F \end{bmatrix}. \tag{2}$$

As given by the pinhole camera model [13], any point in camera coordinates has a perspective projection on the image plane. The relationship between pixel and camera coordinates can be written as

$$u = -f\frac{X_c}{Z_c} = -f\frac{X}{-Y\cos\phi - F} \tag{3}$$

$$v = -f\frac{Y_c}{Z_c} = -f\frac{Y\sin\phi}{-Y\cos\phi - F}. \tag{4}$$

The pixel coordinate system is shown in Fig. 1(c), in which the rectangular region represents the sensing area of the image sensor. Solid lines represent the lane markings that can be observed by the camera. Dashed lines denote the lane markings that are out of the field of view of the camera and cannot be observed. The parallel lines in Fig. 1(a) are projected onto a set

TABLE I
LIST OF VARIABLES FOR FOCAL LENGTH EQUATION

Variables	Meaning and Description
m	f^2
a	$B - \beta_2^2$
b	$B(u_0^2 + v_0^2) - \beta_1^2 - v_0^2 \beta_2^2$
c	$-v_0^2 \beta_1^2$
B	$\beta_1^2 \left(\dfrac{qw}{hu_4} \right)$
β_1	$\dfrac{\alpha v_1 v_3 - v_1 v_3}{\alpha v_1 - \alpha v_3}$
β_2	$\beta_2 = \dfrac{\alpha v_3 - v_1}{\alpha v_1 - \alpha v_3}$
q	$q = \dfrac{1}{1 - \alpha}$
α	$\dfrac{u_2}{u_4}$
u_2	u coordinate of point P_2
u_4	u coordinate of point P_4
v_1	v coordinate of point P_1
v_3	v coordinate of point P_3

of lines in Fig. 1(c) that intersect at a point known as vanishing point VP. The vanishing point lies at a position where the Y coordinate of (X, Y, Z) approaches infinity. The coordinate (u_0, v_0) of VP is given by

$$
\begin{aligned}
u_0 &= \lim_{Y \to \infty} u \\
&= \lim_{Y \to \infty} \left(-f \frac{X}{-Y \cos \phi - F} \right) \\
&= \lim_{Y \to \infty} \left(-f \frac{Y \tan \theta}{-Y \cos \phi - F} \right) \\
&= f \tan \theta \sec \phi
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
v_0 &= \lim_{y \to \infty} v \\
&= \lim_{y \to \infty} \left(-f \frac{Y \sin \phi}{-Y \cos \phi - F} \right) \\
&= f \tan \phi.
\end{aligned} \tag{6}
$$

In this study, we propose to use parallel lane markings to establish a geometrical relationship between a road plane and its camera view. As shown in Fig. 1(a), L_1, L_2, and L_3 intersect the X axis and Y axis at six points; these points are denoted by $P_1 - P_6$. From the perspective model, the corresponding coordinates of these points in the image plane can be obtained. Through geometrical analysis, it will be straightforward to derive the focal length equation as

$$
am^2 + bm + c = 0 \tag{7}
$$

where m is f^2. Table I summarizes the variables used in (7). The detailed derivation is given in Appendix A.

The solution f^2 of (7) must be positive. Accordingly, the focal length f is

$$
f = \sqrt{m}. \tag{8}
$$

Image Sequence

Background Segmentation

Edge Detection

Connected-Component Labeling

Lane-Marking Determination

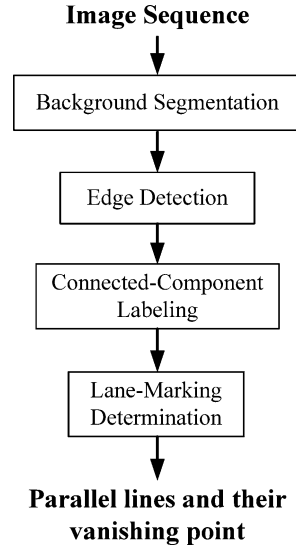**Parallel lines and their vanishing point**

Fig. 2. System architecture of image-based lane-marking determination.

Using (6), the tilt angle is given by

$$
\phi = \tan^{-1} \frac{v_0}{f}. \tag{9}
$$

From (5), the pan angle is expressed as

$$
\theta = \tan^{-1} \frac{u_0}{f \sec \phi}. \tag{10}
$$

If (7) has two positive roots, then the meaningful solution will be the one that satisfies (A24). Using the camera parameters, one can transform the pixel coordinates into their corresponding world coordinates $(X, Y, 0)$ [15]. The detailed procedure is given in Appendix B.

III. DETECTION OF PARALLEL LANE MARKINGS

In this section, an image processing procedure is proposed to automatically detect parallel lane markings in road imagery. The complete procedure consists of background segmentation, edge extraction, erosion, dilation, labeling, and lane-marking analysis. Fig. 2 shows the functional block diagram of the image processing procedure.

A. Background Segmentation and Edge Detection

Gaussian mixture model (GMM) approaches to obtaining reliable background images have gained increasing attention for VTMS in recent years [16]. GMMs feature effective background estimation under environmental variations through a mixture of Gaussians for each pixel in an image frame. For urban traffic, however, vehicles will stop occasionally at intersections because of traffic light or control signals. Such kind of transient stops will increase the weight of the nonbackground Gaussian and degrade the segmentation quality. In this paper, instead of exploiting conventional GMM, a histogram approach is proposed to solve the problem. The background intensity can be determined according to the maximum frequency in the histogram [17]. Moreover, the background image is segmented by

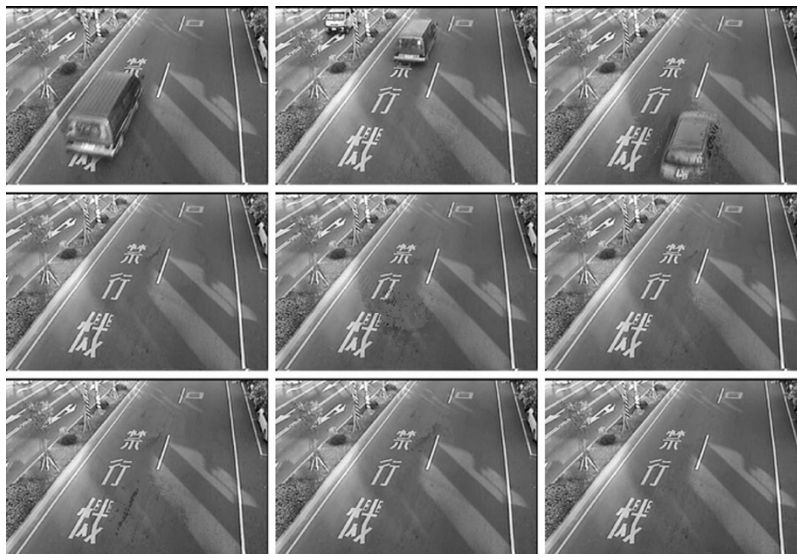Fig. 3.   Image sequence for background segmentation.

Fig. 4.   Background image generation of Fig. 3 using a group-based histogram method.

using a group-based histogram to deal with unreliable intensity distribution caused by sensing uncertainties. The intensity that has the maximum group-based frequency is then selected as the background intensity $f_B(u, v)$ given by

$$f_B(u, v) = \arg \max_l \left\{ \sum_{r=-\sigma}^{\sigma} n_{u,v}(l+r); \ \ 0 \leq l+r \leq (L-1) \right\}$$

(11)

where $n_{u,v}(l)$ is the frequency of pixels at location (u, v) with intensity value l, σ is the predicted standard deviation of intensity, and L is the number of intensity levels. To obtain the intensity $f_B(u, v)$ efficiently, one only needs to calculate the frequencies of adjacent intensity levels. This method is very computationally efficient because it only uses addition and comparison.

To demonstrate the effectiveness of background segmentation, we show a test result using nine images of a traffic image sequence. Fig. 3 shows the original image sequence. Corresponding to Fig. 3, Fig. 4 depicts the generation of the background image using the group-based histogram method. Moving vehicles disappear in the extracted background image as expected. A video clip of the background image generation using the proposed method can be found at http://isci.cn.nctu.edu.tw/video/SMCB_PTZ/Attachment_1.mpg.

From the extracted background image, the edges of the lane markings can be obtained by adopting an intensity gradient method [18]. The detected edges of the traffic lane markings are depicted in Fig. 5(a). To verify the detection performance, we examine the background image together with the detected edges, as shown in Fig. 5(b). Note that two edges appear on both sides of the lane markings. In this design, only the right edges are selected for further calculation. A filter is designed to
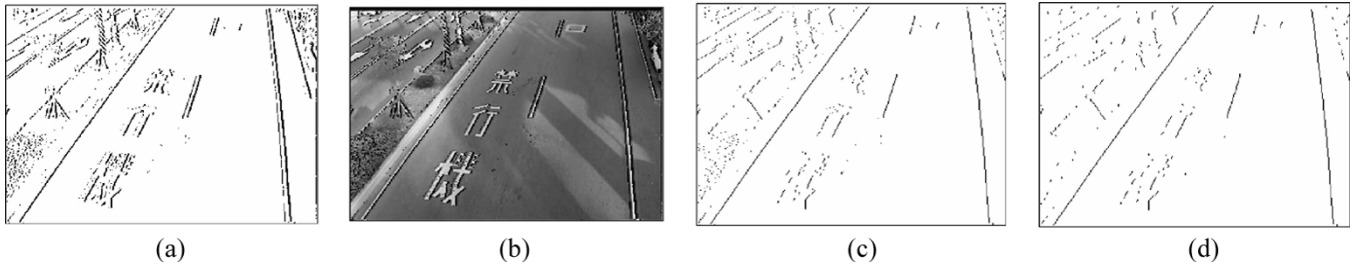
Fig. 5. Edge maps of the background image. (a) Edge map. (b) Background image and its associated edge map. (c) Right-side edge map. (d) Denoised edge map.
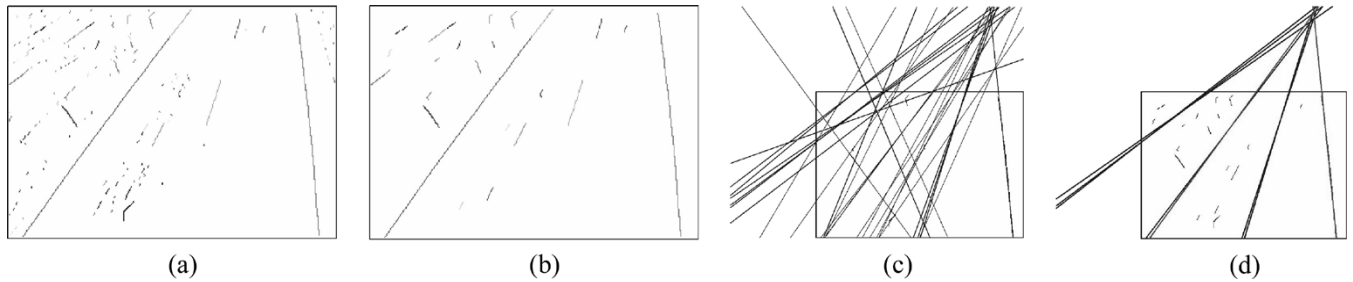

Fig. 6. Linear approximation of lane markings. (a) Labeled feature map. (b) Feature map of segments with higher counts. (c) Linear approximations map. (d) The lines that intersect the sidelines and are located within a vanishing-point region are reserved.

remove all adjacent edge pixels except those of the rightmost edge, i.e.,

$$P(u,v) = 0 \qquad \text{if} \sum_{j=u+1}^{u+5} P(j,v) \geq 1 \qquad (12)$$

where $P(u,v)$ is the binary value at (u,v) in the edge map. Fig. 5(c) depicts the filtered result. The left edges of the lane markings are removed as expected. An erosion operation is then employed to remove salt-and-pepper noise and shrink the detected edge [18]. Next, a dilation operation is applied to reconnect discontinuous features that belong to the same object [19]. Fig. 5(d) shows the final result of edge detection. It is clear that the salt-and-pepper noise is removed and the extracted edges are ready for lane-marking analysis.

B. Connected-Component Labeling and Lane-Marking Determination

As depicted in Fig. 5(d), the lane-marking segments are longer than the features generated by other objects, such as trees, bushes, guideposts, etc. Using a connected-component labeling operation [20], one can classify and label the pixels that are linked together. Fig. 6(a) shows the labeling result of the binary image in Fig. 5(d). The count (length) of the connected pixels can be used to determine whether the connected pixels are features of a lane marking or not. Only those with a larger count are preserved, whereas the rest will be removed. The result of this operation is illustrated in Fig. 6(b). On a multilane road, the lane markings of the road edges are normally indicated by solid lines, whereas the lane-divider lines are marked by broken lines. Based on this premise, the labeled segments that have the first and the second largest number are considered as the sides of a multilane road. They are termed as "sidelines."

Each sideline will then be represented by a linear polynomial equation

$$y = \lambda x + \rho \qquad (13)$$

where λ and ρ are real numbers. One can use a least square approximation to obtain λ and ρ. Accordingly, the intersection of the sidelines can be computed. It gives us the vanishing point of parallel lane markings.

Other segments are similarly processed to obtain their first-degree polynomial equations, as plotted in Fig. 6(c). Next, these lines are checked whether they parallel the sidelines in the real world. If a straight line is parallel with the sidelines, then the intersection of the line with the sidelines needs to be located within a vanishing point region Vr given by

$$\text{Vr} = \left\{ (u,v) : \left| (u - u_0)^2 + (v - v_0)^2 \right| \leq 13 \right\} \qquad (14)$$

where (u_0, v_0) is the vanishing point. Most lines, which are not parallel with the sidelines, will be removed by this intersection discrimination. As shown in Fig. 6(d), only those lines satisfying (14) are reserved.

To correctly locate all the lane markings on the road, the disconnected segments, which belong to a broken lane-divider line, must be merged into a line for obtaining a correct least-square linear representation. A criterion has been developed to find those lines that are near each other. Fig. 7(a) shows the result after merging such lines of Fig. 6(d). As shown in Fig. 7(a), although with reduced line numbers, there still might have extra lines existing in the image. Only the lane-divider lines that lie inside the sidelines need to be kept, and others must be removed as well. Exploiting the assumption that each traffic lane is of the same width on the road, one can apply a virtual horizontal line to intersect each candidate line to obtain its position information in the image plane. As
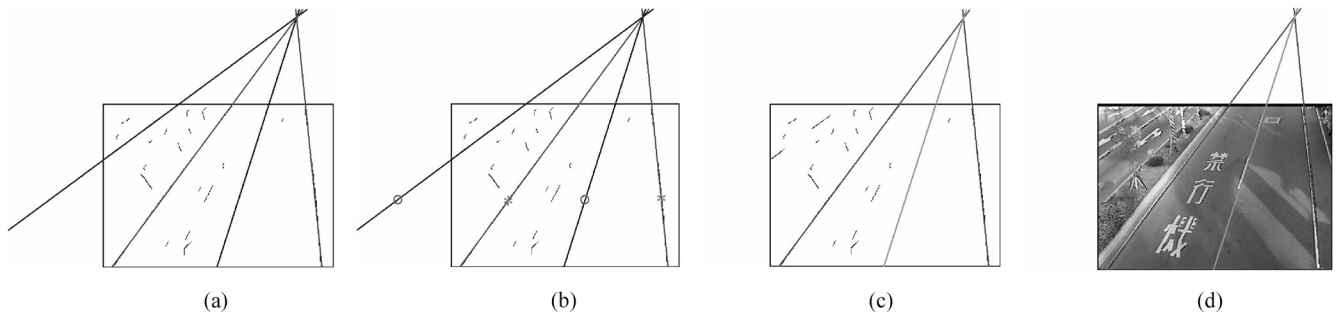
Fig. 7. Parallel lane markings and their vanishing point. (a) Parallel line map. (b) Location map of the parallel lines. (c) Parallel lane markings map. (d) Background image and its associated parallel lane markings.

shown in Fig. 7(b), circles are used to represent the positions of candidate lines and star symbols are used to represent the position of the sidelines, which have already been found. As shown in Fig. 7(b), the line with a circle that lies at the center of those with two stars will be the lane-divider line. Fig. 7(c) shows the detected lane markings and their vanishing point. Fig. 7(d) illustrates the background image together with the detected lane markings. This lane-marking detection algorithm is computationally efficient compared with popular Hough transform approaches. A video clip of the image processing steps for finding parallel lane markings can be found at http://isci. cn.nctu.edu.tw/video/SMCB_PTZ/Attachment_2.mpg.

IV. EXPERIMENTAL RESULTS

To demonstrate the performance of the proposed calibration algorithm, we first use synthetic traffic data to carry out a sensitivity analysis and then validate the calibration results using actual traffic images.

A. Sensitivity Analysis

In actual applications, there might exist intrinsic or extrinsic errors of camera calibration that cause measurement errors. For example, the principle point might vary with zooming [21] and will make pixel coordinates incorrect. Radial distortion also affects the accuracy of parallel lane detection. Tilt or pan operation will change the height of the image sensor and cause an error in focal length estimation. To assure the robustness of the proposed calibration method, we present a sensitivity analysis on the focal length estimation using synthetic data containing intrinsic and extrinsic errors.

Fig. 8(a) illustrates the synthetic traffic scene with two parallel lanes. The circle in the figure represents the camera. The camera view in Fig. 8(a) is constructed according to (3) and (4); the result is shown in the rectangular region of Fig. 8(b). Three lines of the image view intersect one another at the vanishing point, which is denoted by a cross in Fig. 8(b). The camera parameters, such as tilt angle, pan angle, and focal length, are calculated from the synthetic data using the calibration method described in Section II. In the simulation, the tilt angle is changed from $30°$ to $60°$, and the pan angle is changed from $-20°$ to $20°$. These angles reflect most situations in actual VTMS applications. To emulate the effect caused
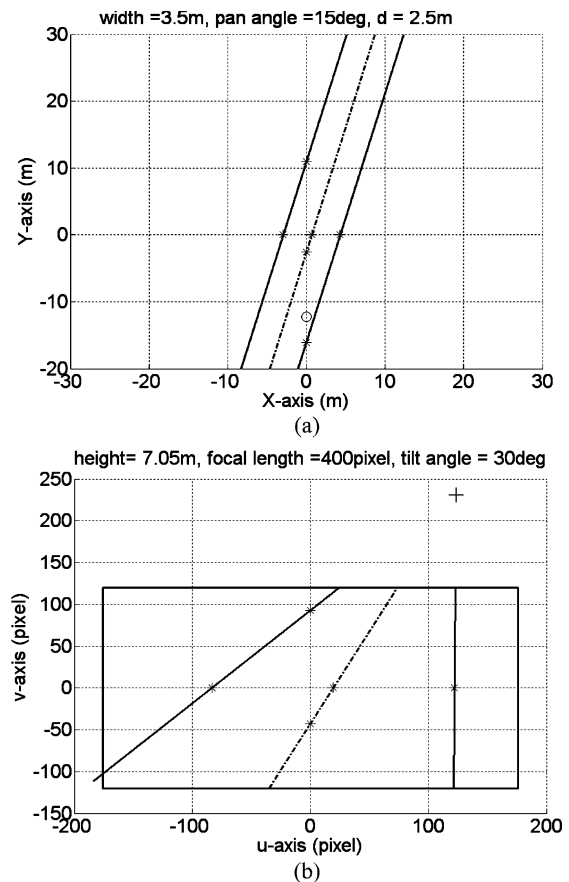

Fig. 8. Synthetic traffic scene for simulation. (a) Top view of a road scene. (b) Road view in image plane.

by radial distortion or incorrect principle-point position, the vanishing point is shifted two pixels upward, rightward, and diagonally, respectively. New parallel lines are established in accordance with the new vanishing point and the intersections of the original lines and the u axis. Using (7), focal length is calculated in accordance with these new parallel lines. In the simulation, the height of the camera is set to 7.05 m and the focal length is set to 430 pixels. The maximum error rates calculated for the condition of translational error in horizontal, vertical, and diagonal directions are presented in simulations 1–3 of Table II, respectively. The absolute error rates of focal length estimation are within 4.1%. Furthermore, the result reveals that focal length estimation is more sensitive to vertical translational error than the horizontal one.

TABLE II
MAXIMUM ERROR RATES OF FOCAL LENGTH, TILT ANGLE, AND VERTICAL POSITION UNDER DIFFERENT SIMULATED ERRORS

Simulation index	Focal length (pixel)	Height (m)	Simulated error			Maximum error rates *1		
			du (pixel)	dv (pixel)	dh (m)	Focal length	Tilt angle	Vertical position*2
1	430	7.05	2	0	0	0.8%	0.7%	0.4%
2	430	7.05	0	2	0	3.3%	3.5%	0.4%
3	430	7.05	2	2	0	4.1%	4.1%	0.7%
4	430	7.05	0	0	-0.02	1.7%	1.4%	1.1%
5	430	7.05	2	2	-0.02	5.6%	5.5%	1.7%
6	330	7.05	2	2	-0.02	6.6%	6.7%	2.3%
7	530	7.05	2	2	-0.02	4.9%	4.7%	1.3%
8	730	7.05	2	2	-0.02	4.0%	3.8%	0.8%
9	430	6.05	2	2	-0.02	5.8%	5.7%	1.9%
10	430	8.05	2	2	-0.02	5.4%	5.3%	1.6%
11	430	10.05	2	2	-0.02	5.1%	5.1%	1.4%

*1 : In each simulation, the tilt angle is changed from 30 to 60 deg and the pan angle is changed from -20 to 20 deg.
*2 : measurement at pixel coordinate (150,120)

It is observed from the sensitivity analysis that tilt angle estimation is also more sensitive to vertical translational error. The absolute error rates of the tilt angle are within 4.1%. Because the estimated parameters will be employed to estimate the position of the vehicle in VTMS, the error rates of the position in the image frame are calculated accordingly. The absolute error rates of the vertical position are within 0.7% at pixel coordinate (150, 120).

To investigate the effect of inaccurate camera height, we introduced a height error of −0.02 m into the simulation. Traffic view is generated according to the true height (7.05 m); focal length is then estimated using the inaccurate height data. Simulation 4 of Table II shows that the three kinds of error rates are all within 1.7%. Detailed error rate results are presented in Fig. 9 to examine the estimation of focal length, tilt angle, and position. It is observed that for a tilt angle smaller than 30° or an absolute pan angle greater than 20°, the error rates will become unacceptable. This phenomenon is mainly caused by the fact that in the image plane the parallel lanes and their vanishing point will deviate more seriously due to radial distortion and incorrect principle-point position. In addition, diagonal and height errors are also simulated. The error rates are all within 5.6%, as shown in simulation 5 of Table II.

As for the case of translational and height errors, simulations 6–8 of Table II show the simulation results with a focal length of 330, 530, and 730 pixels, respectively. The absolute error rates of the focal length and tilt angle are within 6.7%, and the absolute error rates of the vertical translazional position are within 2.3%. The results reveal that the larger the focal length, the less the error rate.

Finally, for cases with translational and height errors, simulations 9–11 of Table II show the simulation results for a height of 6.05, 8.05, and 10.05 m, respectively. The absolute error rates of the focal length and tilt angle are within 5.8%. The absolute error rates of the vertical position are within 2%. The results reveal that the higher the camera, the less the error rates. Furthermore, the effect of height change is not obvious in the test. From the synthetic analysis, the error rates introduced by extrinsic and intrinsic errors are within 6.7% (focal length = 330 pixels). These error rates of position measurement are acceptable for traffic monitoring.

B. Experiments With Actual Imagery

The proposed algorithm has been tested with image sequences recorded from a main road near our university. The camera used in the experiments is a SONY EVI-D31 digital camera. The image sequences were captured with a resolution of 352 × 240 pixels. For traffic monitoring, the camera was installed at a height of 7.03 m, and the width between parallel lane markings is 3.52 m. In the experiments, the background image was first segmented and then used for the lane-marking detection. The vanishing point and the camera parameters, such as focal length, tile angle, and pan angle, were calculated using (7), (9), and (10), accordingly. To validate the estimated parameters, 12 sample features were assigned in a traffic scene for distance measurement, as shown in Fig. 10. The sample distances were measured manually and compared with the estimated distances for evaluation. The estimated distances were computed based on the calibrated camera parameters.

Fig. 10. Sample features selected for image measurement in road imagery.

and standard deviation of the tilt angle are $27.45°$ and $0.33°$, respectively. The estimated mean and standard deviation of the pan angle $8.01°$ and $0.07°$. These experimental results show that the error rates of absolute mean and standard deviation are within 2.39% and 1.49%. The proposed calibration algorithm gives satisfactory accuracy and is robust against zoom changes.

To evaluate the robustness of lane detection with respect to environmental illumination variation, we took imagery at various hours of a sunny day. Six sets of image sequences are presented to show different illumination conditions, as shown in Fig. 12. The intensity values of the lane markings vary in these image frames, but the lane markings always have higher intensities than their adjacent region. The gradient can be successfully used to detect the edge of the lane markings, as discussed in Section III. The results reveal that the lane-detection method performs satisfactorily under different lighting conditions. This robustness partly results from the fact that the SONY PTZ camera has autoexposure and backlight compensation functions to ensure that the subject remains bright even in harsh backlight conditions.

Finally, the algorithm is evaluated with a fixed zoom under different pose settings (A–D) of the PTZ camera, as shown in Fig. 13. Table IV shows the experimental results of the estimation of feature sizes, as depicted in Fig. 10. The mean and standard deviation of the estimated focal lengths are 417.08 and 9.56 pixels, respectively. The mean and standard deviation of the absolute error rates among these measurements are within 2.32% and 1.58%, respectively.

The experimental results of the different zoom and view settings show that the maximum calibration error of distance measurement is within 5%, which is comparable to the results achieved in [11] and [12]. However, our method offers improved autonomy and efficiency. A video clip of the image processing sequence for traffic parameter estimation can be found at http://isci.cn.nctu.edu.tw/video/SMCB_PTZ/Attachment_3.mpg.

V. Conclusion

A novel algorithm has been proposed for the automatic calibration of a PTZ camera overlooking a traffic scene. Focal
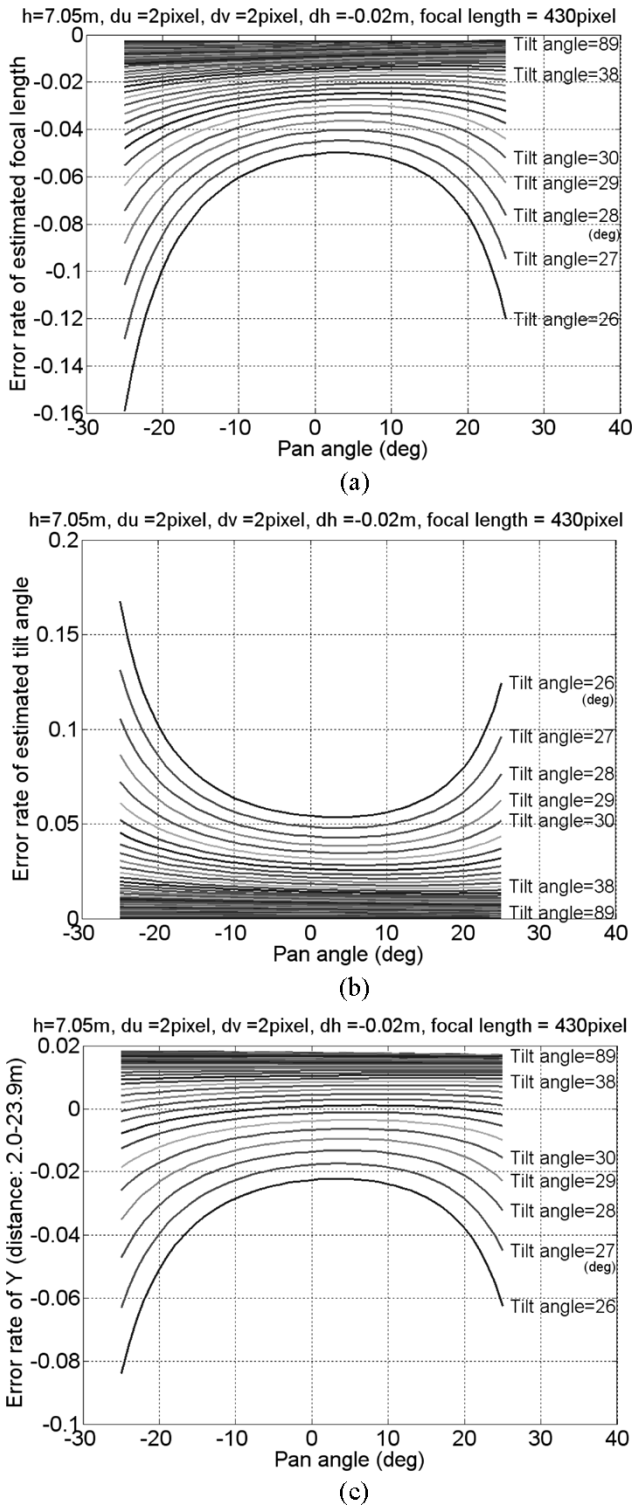

Fig. 9. Sensitivity analysis of translation and height errors.

In the first experiment, traffic image with different zoom settings (A–D) were captured to demonstrate the robustness to radial distortion, as shown in Fig. 11. The principle point of the camera is practically fixed for these zoom settings. The experiment results are listed in Table III. The focal lengths for the four zoom settings are estimated to be 452.91, 496.70, 542.54, and 592.36 pixels, respectively. Using the focal length, the pan angles and tilt angles are calculated. The estimated mean
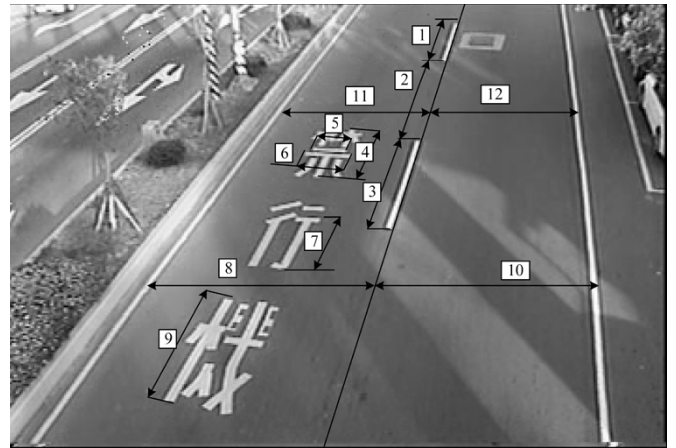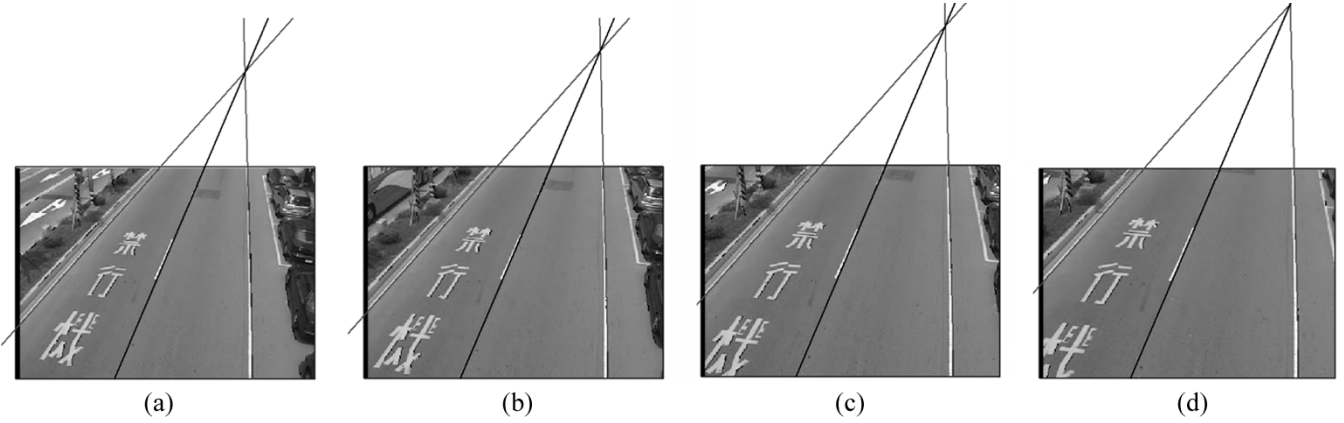
Fig. 11.   Traffic images captured under different zoom settings. (a) Image of zoom setting A. (b) Image of zoom setting B. (c) Image of zoom setting C. (d) Image of zoom setting D.

TABLE III
CALIBRATION RESULTS UNDER DIFFERENT ZOOM SETTINGS

Zoom settings	A	B	C	D
Estimated focal length (pixel)	452.91	496.70	542.54	592.36
Estimated tilt angle (deg)	27.19	27.21	27.51	27.89
Estimated pan angle (deg)	8.01	7.92	8.02	8.09
Sample feature index	Ground truth (m)	Residual error rate	Residual error rate	Residual error rate	Residual error rate
1	4	-0.96%	0.82%	*	*
2	5.8	-2.69%	-2.56%	-1.07%	-1.29%
3	4	-4.09%	-4.85%	-3.47%	-0.73%
4	2.5	-4.84%	-3.97%	-4.98%	-0.15%
5	0.75	0.72%	1.37%	-3.25%	2.07%
6	0.99	-2.84%	-3.07%	-2.90%	-1.32%
7	1.9	1.49%	-0.05%	1.78%	3.45%
8	3.52	-0.84%	-1.59%	-1.09%	-1.25%
9	2.42	-3.10%	-4.59%	*	*
10	3.52	-1.26%	-0.94%	-2.59%	-0.05%
11	3.52	-1.75%	-2.26%	-1.28%	-1.85%
12	3.52	-1.42%	-1.47%	-1.11%	-1.39%
Abs. mean		2.17%	2.34%	2.39%	1.29%
Abs. standard deviation		1.29%	1.49%	1.20%	1.00%

* : Feature sample is not measurable in the scene image.

length equation has been derived for camera calibration based on parallel lane markings. Subsequently, the pan and tilt angles of the camera can be obtained by using the estimated focal length. To locate parallel lane markings, an image processing procedure has been developed. Synthetic data and actual traffic imagery have been employed to validate the accuracy and ro-

bustness of the proposed method. The simulation results reveal that the error rates of position estimation are within 2.3% under the presence of reasonable translational and height errors. In actual experiments, 12 feature samples in a road scene were selected for distance measurement. The maximum error of the distance measurement is within 5%, and the absolute mean error is below 2.39%.

In the future, practical applications of PTZ cameras will be further studied. For instance, most vision-based traffic surveillance methods adopt a virtual window to detect vehicles [22]. If the view of the PTZ camera is changed, then the position and size of the window must be adjusted again manually. Using the dynamic calibration procedure developed in this paper, the detection window can be arranged automatically. On the other hand, the effect of lens distortion and nonfixed principal point needs to be handled to increase the accuracy of PTZ camera calibration.

APPENDIX A
DERIVATION OF FOCAL LENGTH EQUATION

Here, the focal length equation will be derived by using only two parallel lines in the image. As shown in Fig. 1(a), lines L_1 and L_2 are two parallel lines. L_1 and L_2 intersect the X axis and Y axis at P_1, P_2, P_3, and P_4; the corresponding coordinates of these points in the image plane are expressed as

$$u_2 = \frac{fX_2}{F} = -\frac{fY_1 \tan\theta}{h\csc\phi} = -\frac{f\alpha Y_3 \tan\theta}{h\csc\phi} \quad \text{(A1)}$$

$$u_4 = -\frac{fX_4}{F} = -\frac{fY_3 \tan\theta}{h\csc\phi} \quad \text{(A2)}$$

$$v_1 = \frac{fY_1 \sin\phi}{Y_1 \cos\phi + h\csc\phi} \quad \text{(A3)}$$

$$v_3 = \frac{fY_3 \sin\phi}{Y_3 \cos\phi + h\csc\phi} \quad \text{(A4)}$$

where X_2 is the X coordinate of P_2, X_4 is the X coordinate of P_4, Y_1 is the Y coordinate of P_1, and Y_3 is the Y coordinate of P_3. Dividing (A1) by (A2), we obtain $\alpha = (Y_1/Y_3) = (u_2/u_4)$. Writing r as $fY_3 \sin\phi$, s as $Y_3 \cos\phi$, and t as $h\csc\phi$, (A3) and
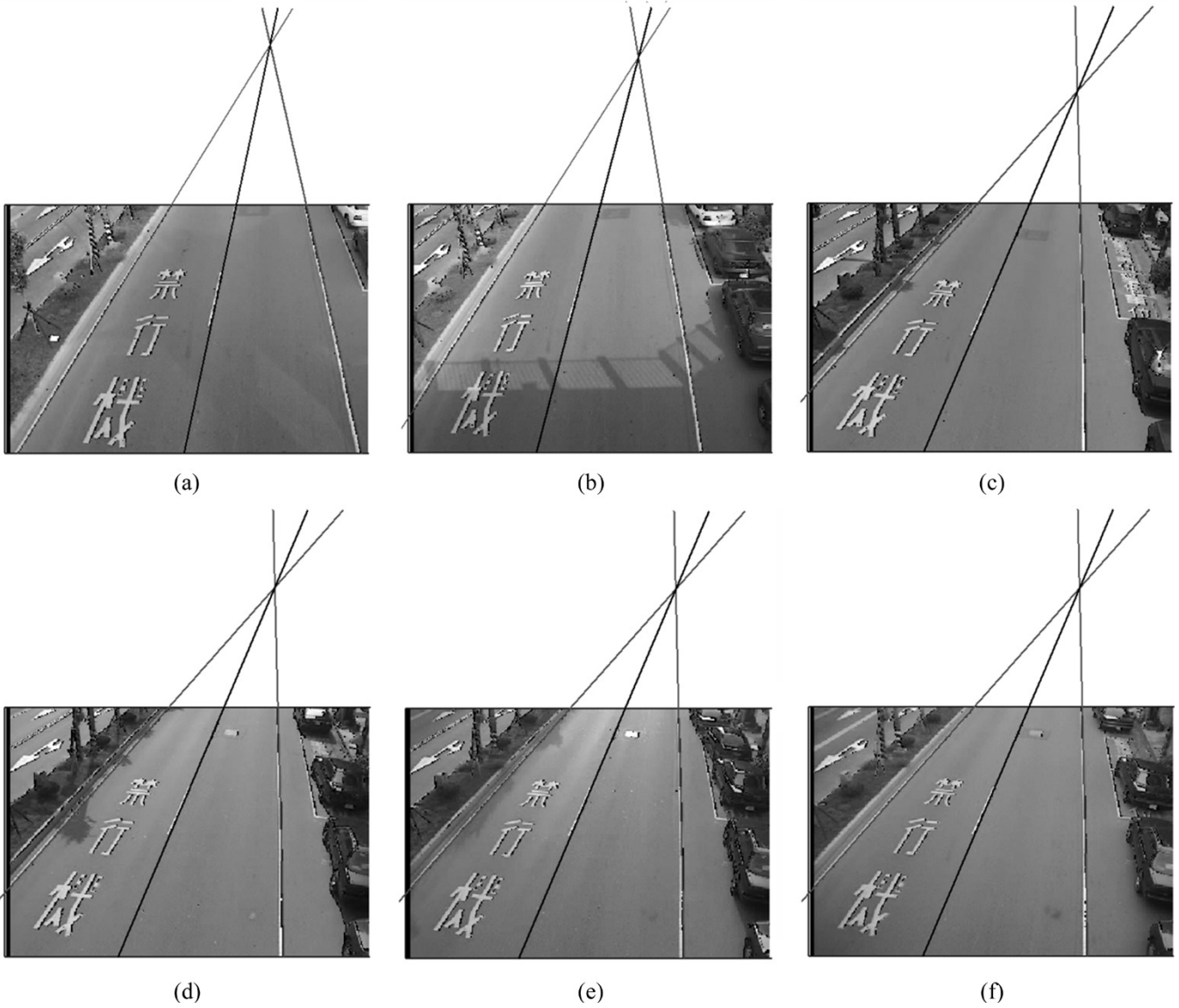
Fig. 12. Traffic images captured under different illumination conditions. (a) Image with weak shadow. (b) Image with strong shadow. (c) Image under bright illumination. (d) Image under soft illumination. (e) Image captured at sunset. (f) Image under darker illumination.
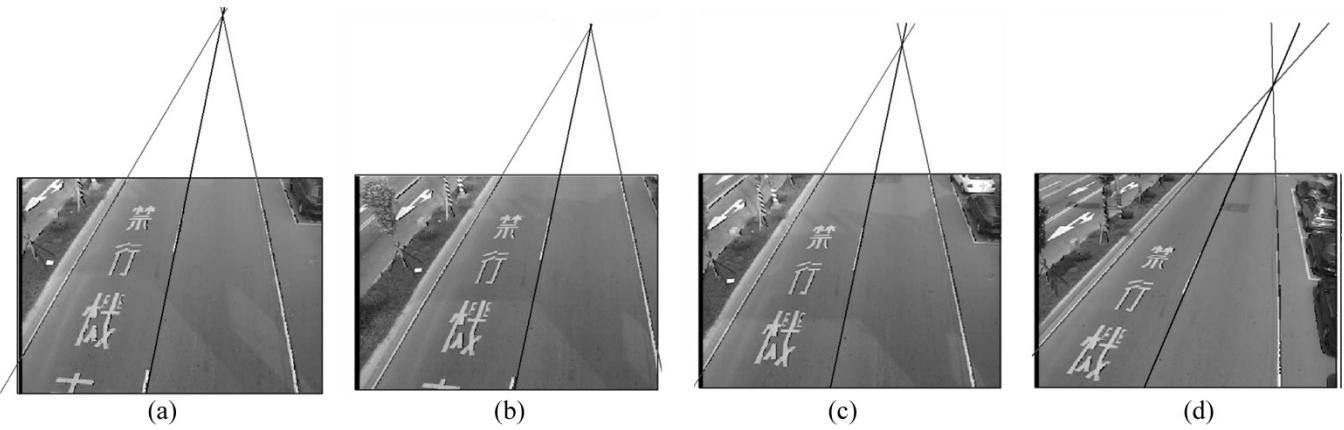

Fig. 13. Traffic images captured under different camera pose settings. (a) Image of pose setting A. (b) Image of pose setting B. (c) Image of pose setting C. (d) Image of pose setting D.

TABLE IV
CALIBRATION RESULTS UNDER DIFFERENT CAMERA POSE SETTINGS

Pose settings		A	B	C	D
Estimated focal length (pixel)		421.82	413.27	427.50	405.74
Estimated tilt angle (deg)		35.63	34.06	31.68	30.86
Estimated pan angle (deg)		7.95	12.18	7.36	14.38
Sample feature index	Ground truth (m)	Residual error rate	Residual error rate	Residual error rate	Residual error rate
1	4	*	*	*	-1.68%
2	5.8	*	4.94%	4.33%	4.94%
3	4	0.00%	-0.98%	-2.38%	2.15%
4	2.5	-1.75%	0.38%	2.94%	-1.31%
5	0.75	-3.01%	1.61%	-3.99%	-0.25%
6	0.99	-2.99%	-0.92%	-4.85%	-3.40%
7	1.9	4.35%	2.93%	4.57%	3.23%
8	3.52	-0.78%	-3.11%	-1.20%	-2.55%
9	2.42	-1.45%	1.48%	-0.52%	2.77%
10	3.52	-1.59%	-2.76%	-0.54%	-2.84%
11	3.52	-0.22%	-3.00%	-1.90%	-1.64%
12	3.52	-1.10%	-4.10%	-1.34%	-2.88%
Abs. mean		1.68%	2.04%	2.32%	2.18%
Abs. standard deviation		1.27%	1.24%	1.58%	1.08%

* : Feature sample is not measurable in the scene image.

(A4) can be rewritten as

$$v_1 = \frac{\alpha r}{\alpha s + t} \tag{A5}$$

$$v_3 = \frac{r}{s + t}. \tag{A6}$$

Applying trigonometric function properties, one can easily find the relationship between r, s, and t. Computing $r^2 + f^2 s^2$ and $r^2 t^2$, one can obtain

$$r^2 + f^2 s^2 = f^2 Y_3^2 \tag{A7}$$

$$r^2 t^2 = (f Y_3 h)^2. \tag{A8}$$

Dividing (A7) by (A8), one has

$$\frac{r^2 + f^2 s^2}{r^2 t^2} = \frac{1}{h^2}. \tag{A9}$$

Solving for r and s in terms of t from (A5) and (A6), one has

$$r = \frac{\alpha v_1 v_3 - v_1 v_3}{\alpha v_1 - \alpha v_3} t = \beta_1 t \tag{A10}$$

$$s = \frac{\alpha v_3 - v_1}{\alpha v_1 - \alpha v_3} t = \beta_2 t. \tag{A11}$$

Substituting (A10) and (A11) into (A9) and using $t = h \csc \phi$, (A9) can be rewritten as

$$\frac{\beta_1^2 + f^2 \beta_2^2}{\beta_1^2 (h \csc \phi)^2} = \frac{1}{h^2}. \tag{A12}$$

Rearranging (A12), we obtain

$$\beta_1^2 + f^2 \beta_2^2 = \beta_1^2 \csc^2 \phi. \tag{A13}$$

Next, the relationship between $\csc \phi$ and $\sec \theta$ is derived. In Fig. 1, the world coordinates of the camera is $(0, -h \cot \phi, h)$ and the X coordinates of P_2 and P_4 are expressed as

$$X_4 = h \cot \phi \tan \theta - d \sec \theta \tag{A14}$$
$$X_2 = X_4 - w \sec \theta. \tag{A15}$$

The u coordinates of P_2 and P_4 in an image frame are expressed as

$$u_2 = \frac{f X_2}{F} = \frac{f(X_4 - w \sec \theta)}{F} \tag{A16}$$

$$u_4 = \frac{f X_4}{F}. \tag{A17}$$

Let $X_4 = q w \sec \theta$, then $X_2 = (q - 1) w \sec \theta$. Equations (A16) and (A17) can be rewritten as

$$u_2 = \frac{f(q - 1) w \sec \theta}{F} \tag{A18}$$

$$u_4 = \frac{f q w \sec \theta}{F}. \tag{A19}$$

From (A18) and (A19), q can be expressed in terms of u_1 and u_2 as

$$q = \frac{1}{1 - \frac{u_2}{u_4}} = \frac{1}{1 - \alpha}. \tag{A20}$$

Substituting $F = h \csc \phi$ into (A19), we obtain the relationship between $\csc \phi$ and $\sec \theta$ as

$$\csc \phi = \frac{q w}{h u_4} f \sec \theta. \tag{A21}$$

Substituting (A21) into (A13), we obtain

$$\beta_1^2 + f^2 \beta_2^2 = \beta_1^2 \left(\frac{q w}{h u_4} \right)^2 f^2 \sec^2 \theta. \tag{A22}$$

Using the vanishing point constraints and trigonometric function properties, one can proceed to derive equations that will determine the equation containing $\sec^2 \theta$ and, finally, the focal length equation. Squaring (5) and using $\sec^2 \phi = 1 + \tan^2 \phi$, we have

$$u_0^2 = f^2 \tan^2 \theta + f^2 \tan^2 \theta \tan^2 \phi. \tag{A23}$$

Inasmuch as $v_0 = f \tan \phi$, (A23) can be rewritten as

$$u_0^2 = \left(f^2 + v_0^2 \right) \tan^2 \theta. \tag{A24}$$

Using $\tan^2 \theta = (\sec^2 \theta - 1)$, (A24) yields

$$u_0^2 + f^2 + v_0^2 = \left(f^2 + v_0^2 \right) \sec^2 \theta. \tag{A25}$$

The equation containing $\sec^2 \theta$ is found. Dividing (A25) by (A22), we get

$$\frac{u_0^2 + f^2 + v_0^2}{\beta_1^2 + f^2 \beta_2^2} = \frac{f^2 + v_0^2}{\beta_1^2 \left(\frac{qw}{hu_4} \right)^2 f^2}. \tag{A26}$$

Rearranging (A26), we arrive at

$$am^2 + bm + c = 0 \tag{A27}$$

where m is f^2 and the other variables are listed Table I. This governing equation is presented in Section II as the focal length equation.

Next, let us discuss how the camera parameters affect the sign of coefficient a. For simplicity, the sign of at^2 is discussed instead of the sign of a, i.e.,

$$at^2 = Bt^2 - \beta_2^2 t^2 = Y_3^2 (\cos^2 \theta - \cos^2 \phi). \tag{A28}$$

Equation (A28) reveals that the magnitude of the tile and pan angles affects the sign of coefficient a. Details are listed as follows:

- $a > 0$ if $|\phi| > |\theta|$;
- $a = 0$ if $|\phi| = |\theta|$ or $Y_3 = 0$;
- $a < 0$ if $|\phi| < |\theta|$.

It is clear that the difference between the absolute value of tilt and pan angles determines the sign of coefficient a. When $a = 0$, the focal length equation becomes linear and the focal length can be estimated easily. This completes the derivation of the focal length equation.

When the vanishing point is far from the image center or disappears (for instance, as the tilt angle is equal to $90°$) in the image frame, (A27) cannot be used to find the focal length. Instead, the focal length can be easily obtained by the perspective projection equation given by

$$f = h \frac{w_p}{w} \tag{A29}$$

where w_p is the width between parallel lanes in the image frame.

APPENDIX B
CONVERSION BETWEEN PIXEL COORDINATES AND WORLD COORDINATES

Here, we derive the transformation between pixel coordinates and world coordinates. We will explain how focal length and tilt angle are used to obtain the world coordinates of a feature in the ground plane.

A pixel coordinate (u, v) is expressed as a function of the world coordinate (X, Y, Z) as

$$u = -f \frac{X_c}{Z_c} = -f \frac{X}{-Y \cos \phi - F} \tag{B1}$$

$$v = -f \frac{Y_c}{Z_c} = -f \frac{Y \sin \phi}{-Y \cos \phi - F}. \tag{B2}$$

Equation (B2) is rewritten as

$$vY \cos \phi + vh \csc \phi = fY \sin \phi. \tag{B3}$$

Dividing (B3) by $\cos \phi$, one can have

$$vY + \frac{vh \csc \phi}{\cos \phi} = fY \frac{\sin \phi}{\cos \phi}. \tag{B4}$$

Using $v_0 = f \tan \phi$, (B4) is rewritten as

$$vY + \frac{vh}{\sin \phi \cos \phi} = Y v_0. \tag{B5}$$

The solution of (B5) is

$$Y = \frac{h}{f} \frac{v}{\sin^2 \phi} \frac{v_0}{v_0 - v}. \tag{B6}$$

Substituting (B6) into (B1), it is easy to obtain

$$X = \frac{h}{f} \frac{u}{\sin \phi} \left(\frac{v}{v_0 - v} + 1 \right) = \frac{h}{f} \frac{u}{\sin \phi} \frac{v_0}{v_0 - v}. \tag{B7}$$

From (B6) and (B7), one can transform the pixel coordinates into their world coordinates.

REFERENCES

[1] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 2, pp. 108–118, Jun. 2000.

[2] O. Masoud, N. P. Papanikolopoulos, and E. Kwon, "The use of computer vision in monitoring weaving sections," *IEEE Trans. Intell. Transp. Syst.*, vol. 2, no. 1, pp. 18–25, Mar. 2001.

[3] J.-C. Tai, S.-T. Tseng, C.-P. Lin, and K.-T. Song, "Real-time image tracking for automatic traffic monitoring and enforcement applications," *Image Vis. Comput. J.*, vol. 22, no. 6, pp. 485–501, Jun. 2004.

[4] R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE Trans. Intell. Trans. Syst.*, vol. 1, no. 2, pp. 119–130, Jun. 2000.

[5] A. M. Sabatini, V. Genovese, and E. S. Maini, "Toward low-cost vision-based 2D localisation systems for applications in rehabilitation robotics," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, Lausanne, Switzerland, 2002, pp. 1355–1360.

[6] F.-Y. Wang, "A simple and analytical procedure for calibrating extrinsic camera parameters," *IEEE Trans. Robot. Autom.*, vol. 20, no. 1, pp. 121–124, Feb. 2004.

[7] S. Ying and G. W. Boon, "Camera self-calibration from video sequences with changing focal length," in *Proc. IEEE Int. Conf. Image Process.*, Chicago, IL, 1998, vol. 2, pp. 176–180.

[8] E. Izquierdo, "Efficient and accurate image based camera registration," *IEEE Trans. Multimedia*, vol. 5, no. 3, pp. 293–302, Sep. 2003.

[9] L.-L. Wang and W.-H. Tsai, "Camera calibration by vanishing lines for 3-D computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 370–376, Apr. 1991.

[10] T. Echigo, "A camera calibration technique using three sets of parallel lines," *Mach. Vis. Appl.*, vol. 3, no. 3, pp. 159–167, Mar. 1990.

[11] E. K. Bas and J. D. Crisman, "An easy to install camera calibration for traffic monitoring," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Boston, MA, 1997, pp. 362–366.

[12] C. Zhaoxue and S. Pengfei, "Efficient method for camera calibration in traffic scenes," *IEE Electron. Lett.*, vol. 40, no. 6, pp. 368–369, Mar. 2004.

[13] A. H. S. Lai and N. H. C. Yung, "Lane detection by orientation and length discrimination," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 4, pp. 539–548, Aug. 2000.

[14] T. N. Schoepflin and D. J. Dailey, "Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 2, pp. 90–98, Jun. 2003.

[15] Y. U. Yim and S.-Y. Oh, "Three-feature based automatic lane detection algorithm (TFALDA) for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 219–225, Dec. 2003.

[16] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, 1999, pp. 246–252.

[17] J.-C. Tai and K.-T. Song, "Background segmentation and its application to traffic monitoring using modified histogram," in *Proc. IEEE Conf. Netw. Sens. Control*, Taipei, Taiwan, R.O.C, 2004, pp. 13–18.

[18] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*. New York: McGraw-Hill, 1995.

[19] S.-T. Bow, *Pattern Recognition and Image Preprocessing*. New York: Marcel Dekker, 2002.

[20] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[21] S. Sinha and M. Pollefeys, "Towards calibrating a pan–tilt–zoom cameras network," in *Proc. 5th Workshop Omnidirectional Vis. Camera Netw. Non-Classical Cameras*, Prague, Czech Republic, 2004, pp. 42–54.

[22] J.-C. Tai and K.-T. Song, "Automatic contour initialization for image tracking of multi-lane vehicles and motorcycles," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Shanghai, China, 2003, pp. 808–813.

Kai-Tai Song (A'91) was born in Taipei, Taiwan, R.O.C., in 1957. He received the B.S. degree in power mechanical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 1979, and the Ph.D. degree in mechanical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1989.

From 1981 to 1984, he was with the Chung Shan Institute of Science and Technology. Since 1989, he has been Faculty Member and is currently a Professor at the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu. His research interests include mobile robots, image processing, visual tracking, sensing and perception, embedded systems, intelligent system control integration, and mechatronics.

Dr. Song served as the Chairman of the IEEE Society of Robotics and Automation, Taipei Chapter, from 1998 to 1999.

Jen-Chao Tai was born in Nantou, Taiwan, R.O.C., in 1962. He received the B.S. degree in power mechanical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 1985, the M.S. degree in control engineering from the National Chiao Tung University, Hsinchu, in 1992, and is currently working toward the Ph.D. degree in electrical and control engineering at the National Chiao Tung University, Hsinchu.

Since 1992, he has been a Lecturer at the Department of Mechanical Engineering, Minghsin University of Science and Technology, Hsinchu. His research interests include image processing, visual tracking, vision-based traffic parameter estimation, dynamic camera calibration, real-time imaging systems, and mechatronics.