# Roadway Surveillance Video Camera Calibration Using Standard Shipping Container

Ruimin Ke

Department of Civil & Environmental Engineering
University of Washington, Seattle, WA, U.S.A
ker27@uw.edu

Zewen Pan

College of Resource and Civil Engineering
Northeast University, Shenyang, P.R.China
panzewen_neu@163.com

Ziyuan Pu

Department of Civil & Environmental Engineering
University of Washington, Seattle, WA, U.S.A
ziyuanpu@uw.edu

Yinhai Wang*

Department of Civil & Environmental Engineering
University of Washington, Seattle, WA, U.S.A
yinhai@uw.edu

***Abstract*—Surveillance video cameras have been increasingly deployed on roadway networks providing important support for roadway management. While the information-rich video images are a valuable source of traffic data, these surveillance video cameras are typically designed for manual observation of roadway conditions and are not for automatic traffic data collection. The benefits of turning these surveillance cameras into data collection cameras are obvious, but collecting traffic data would normally require the development of a cost-effective method to efficiently and accurately calibrate surveillance video cameras. This paper proposes such a robust and efficient method that calibrates surveillance video cameras using standard shipping container as the reference object. The traditional camera calibration model can be simplified and camera parameters can be recovered with precise mathematical derivation. After solving for all the camera parameters, the 3D object world coordinates can be reconstructed from 2D image coordinates, thus enabling the collection of a variety of traffic data using surveillance video camera data.**

***Keywords*—Camera calibration, Coordinate reconstruction, Surveillance video camera, Shipping container, Traffic data collection**

## I. INTRODUCTION

Over the past few decades, many new traffic detection technologies have been developed. One technology that has become increasingly prevalent in recent years is the video image processing (VIP) system, which typically consists of one or more cameras, a microprocessor-based computer for digitizing and processing the imagery, and software for interpreting the images and converting them into traffic flow data [1]. The performance of a VIP system depends largely on the quality of camera calibration, which establishes a relationship between the 3-dimensional (3D) world coordinate system and a 2-dimensional (2D) image coordinate system. Extensive research has been completed on video calibration problems in the past few decades, largely motivated by computer vision applications such as robot vision and unmanned vehicle driving.

In summary, existing methods can be roughly divided into non-linear [2-4] and linear methods [5-7]. Although different solution strategies are used in the calibration equations, nearly all such methods require the placement of calibration points or calibration objects in the camera's field of view. While using reference objects for video camera calibration may not be a big problem for many applications, it does create extra difficulties in traffic data collection because placing calibration objects/marks on an operating roadway can be disruptive to traffic. Moreover, due to differences in hardware, video image resolution, and target accuracy, most computer vision and photogrammetry techniques cannot be directly applied to roadway surveillance cameras [8].

Because of these difficulties and the differing objectives of camera calibration between the transportation field and those of computer vision, approaches for collecting traffic data using uncalibrated cameras have been proposed, which skip the complicated camera calibration procedure [9-10]. Sometimes camera is not fully calibrated because camera extrinsic parameters keep changing due to the constant motion of the camera platform [21-23]. But without calibrating all camera parameters, normally just certain types of traffic data can be acquired in certain scenarios.

Recently, transportation researchers have designed some typical methods for roadway surveillance camera calibration. The most common traffic scene used for calibration is lane marking, which are generally parallel to one another in the 3D world but appear to converge in a 2D image. The intersection of a group of parallel lines in a 2D image is called vanishing point, and the lines are known individually as vanishing lines. Vanishing points and lines have some useful properties that can be applied to camera calibration. Parallel lane markings are a set of vanishing lines in the video camera images, which offers a train of thought for calibrating surveillance cameras [11-12]. However, lane markings are sometimes not visible in video images due to camera resolution, weather conditions, or

**978-1-5386-2524-8/17/$31.00 ©2017 IEEE**

degraded marking quality. Vanishing points can also be estimated using parallel vehicle tracks that are extracted using image processing techniques [13-14]. This approach can be applied in scenarios where lane markings are not clearly visible, but in some cases vehicle motion patterns cannot be assumed to be parallel [15].

Substantial work has been completed as well using pedestrians or vehicles for surveillance video camera calibration. Researchers have tried to take advantage of motion information in video data when using pedestrians, but most existing work focusing on pedestrians requires an accurate pedestrian detection and even precise foot-head localization [16-18]. Quite often, these two requirements are challenging due to noise, shadows, and low video resolution. Furthermore, when a surveillance camera is placed in a scene with rare pedestrian occurrences, pedestrian-based calibration is not possible. For vehicles, vehicle track and vehicle itself can be used for reference in calibration. For instance, Zhang et al. [20] developed a very effective algorithm using vanishing points obtained from vehicles and standing pedestrians. However, little work has utilized vehicle size/shape alone for calibration thus far.

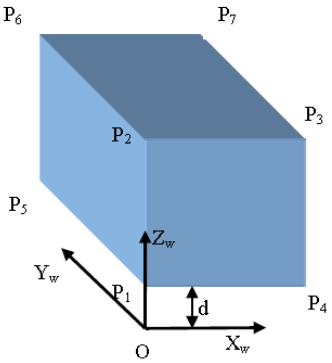
Among existing methods, some require the position of the video camera as input [19-20] while others do not. In practice, a surveillance video camera's position information such as height or tilt angle is not easy to determine in most cases, so calibration models requiring position parameters are not often practical.

Considering that some moving objects in a roadway system have known dimensions and are easily recognizable, surveillance video cameras may be calibrated using these standard moving objects.. In this paper, the authors propose a method for calibrating roadway surveillance video cameras using standard shipping containers with known dimensions and recognizable corner points. Shipping container has standard dimensions and are much more recognizable than other road users or lane markings due to its large size and bright color. This is the first time that shipping container used as the reference object for surveillance camera calibration. Based on a popular camera model, this paper developed an efficient and robust solution for camera calibration using shipping container. The camera parameters can be fully recovered with no other inputs.

## II. METHODOLOGY

### A. General Description

The method proposed here uses standard shipping container to calibrate roadway surveillance cameras. Standard shipping container is a rectangle of dimension (height × width × length) that must fall in one of four standard sizes: 8.5'× 8.0' × 20', 8.5'× 8.0' × 40', 9.5'× 8.0' × 40', or 9.5'× 8.0' × 45'. Let us denote its dimension height × width × length = h × w × l, as shown in Fig. 1. The distance between the underside of a shipping container and ground surface is assumed to be d, which is often known for specific truck types. In addition, the origin O is assumed to be under $P_1$, with $OP_1 = d$. The $X_w$ axis is parallel to $P_1P_4$, $Y_w$ is parallel to $P_1P_5$, $Z_w$ is parallel to $P_1P_2$. Thus, the 3D coordinates of $P_1$, $P_2$, …$P_7$ are shown in Fig. 1.




Truck corner points

$P_1$:(0, 0, d)$^T$

$P_2$:(0, 0, h+d)$^T$

$P_3$:(w, 0, h+d)$^T$

$P_4$:(w, 0, d)$^T$

$P_5$:(0, l, d)$^T$

$P_6$:(0, l, h+d)$^T$

$P_7$:(w, l, h+d)$^T$

Fig. 1. A model of standard moving object (international standard shipping container).

### B. Development of Necessary Equations

Based on Tsai's popular camera model [6], there are basically three steps to develop equations.

(1)From real-world coordinates to camera coordinates:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{1}$$

$$X_c = r_{xx}X_w + r_{xy}Y_w + r_{xz}Z_w + T_x \tag{1a}$$

$$Y_c = r_{yx}X_w + r_{yy}Y_w + r_{yz}Z_w + T_y \tag{1b}$$

$$Z_c = r_{zx}X_w + r_{zy}Y_w + r_{zz}Z_w + T_z \tag{1c}$$

(2)From camera coordinates to image coordinates:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix} \tag{2}$$

$$X = f \cdot \frac{r_{xx}X_w + r_{xy}Y_w + r_{xz}Z_w + T_x}{r_{zx}X_w + r_{zy}Y_w + r_{zz}Z_w + T_z} \tag{2a}$$

$$Y = f \cdot \frac{r_{yx}X_w + r_{yy}Y_w + r_{yz}Z_w + T_y}{r_{zx}X_w + r_{zy}Y_w + r_{zz}Z_w + T_z} \tag{2b}$$

(3)From image coordinates to pixel numbers:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} s_x \cdot X + u_o \\ s_y \cdot Y + v_o \end{bmatrix} \tag{3}$$

$$s_x X = u - u_0 \tag{3a}$$

$$Y = v - v_0 \tag{3b}$$

For each corner point $P_i$, with $X_{wi}$, $Y_{wi}$, $Z_{wi}$, $u_i$, $v_i$, $u_0$, $v_0$ known, we can get $s_xX_i$ and $Y_i$ from Eq. (3a) and (3b). Then, relationships between $s_xX_i$, $Y_i$ and $X_{wi}$, $Y_{wi}$, $Z_{wi}$ can be established as follows:

$$s_x X_i = s_x f \cdot \frac{r_{xx}X_{wi} + r_{xy}Y_{wi} + r_{xz}Z_{wi} + T_x}{r_{zx}X_{wi} + r_{zy}Y_{wi} + r_{zz}Z_{wi} + T_z} \tag{4a}$$

$$Y_i = f \cdot \frac{r_{yx}X_{wi} + r_{yy}Y_{wi} + r_{yz}Z_{wi} + T_y}{r_{zx}X_{wi} + r_{zy}Y_{wi} + r_{zz}Z_{wi} + T_z} \tag{4b}$$

Dividing Eq. (4a) with (4b) yields:

$$\frac{s_x X_i}{Y_i} = s_x \frac{r_{xx}X_{wi} + r_{xy}Y_{wi} + r_{xz}Z_{wi} + T_x}{r_{yx}X_{wi} + r_{yy}Y_{wi} + r_{yz}Z_{wi} + T_y} \tag{5}$$

Assuming $X_i' = s_xX_i$, $Y_i' = Y_i$, Eq. (5) can be expressed as:

$$\frac{X_i^{'}}{Y_i^{'}} = s_x \frac{r_{xx}X_{wi} + r_{xy}Y_{wi} + r_{xz}Z_{wi} + T_x}{r_{yx}X_{wi} + r_{yy}Y_{wi} + r_{yz}Z_{wi} + T_y} \tag{6}$$

The origin of the object coordinate planes is on the ground. $T_y$ is set to be nonzero by putting the origin of the object world coordinate away from the $Y_c$ axis of the camera coordinate system. Thus, rearranging Eq. (6) yields:

$$Y_i^{'}X_{wi}\frac{s_x r_{xx}}{T_y} + Y_i^{'}Y_{wi}\frac{s_x r_{xy}}{T_y} + Y_i^{'}Z_{wi}\frac{s_x r_{xz}}{T_y} + Y_i^{'}\frac{s_x T_x}{T_y} - X_i^{'}X_{wi}\frac{r_{yx}}{T_y} - X_i^{'}Y_{wi}\frac{r_{yy}}{T_y} - X_i^{'}Z_{wi}\frac{r_{yz}}{T_y} = X_i^{'} \tag{7}$$

The seven unknowns in Eq. (7) are $\frac{s_x r_{xx}}{T_y}$, $\frac{s_x r_{xy}}{T_y}$, $\frac{s_x r_{xz}}{T_y}$, $\frac{s_x T_x}{T_y}$, $\frac{r_{yx}}{T_y}$, $\frac{r_{yy}}{T_y}$, and $\frac{r_{yz}}{T_y}$.

The establishment of seven equations from the seven non-collinear points gives us a platform on which to solve these unknowns.

According to Eq. (7), for each calibration point $i$ with $(X_{wi}, Y_{wi}, Z_{wi})$ as the 3D world coordinate and $(X_i', Y_i')$ as the image coordinate, we can set up the following linear equation:

$$\left[Y_i^{'}X_{wi}, Y_i^{'}Y_{wi}, Y_i^{'}Z_{wi}, Y_i^{'}, -X_i^{'}X_{wi}, -X_i^{'}Y_{wi}, -X_i^{'}Z_{wi}\right]\begin{bmatrix}\frac{s_x r_{xx}}{T_y}\\ \frac{s_x r_{xy}}{T_y}\\ \frac{s_x r_{xz}}{T_y}\\ \frac{s_x T_x}{T_y}\\ \frac{r_{yx}}{T_y}\\ \frac{r_{yy}}{T_y}\\ \frac{r_{yz}}{T_y}\end{bmatrix} = X_i^{'} \tag{8}$$

Then, when we apply this equation to a shipping container, a linear equation set can be set up as follows:

$$\begin{bmatrix}0,0,dY_1^{'},Y_1^{'},0,0,-dX_1^{'}\\ 0,0,(h+d)Y_2^{'},Y_2^{'},0,0,-(h+d)X_2^{'}\\ wY_3^{'},0,(h+d)Y_3^{'},Y_3^{'},-wX_3^{'},0,-(h+d)X_3^{'}\\ wY_4^{'},0,dY_4^{'},Y_4^{'},-wX_4^{'},0,-dX_4^{'}\\ 0,lY_5^{'},dY_5^{'},Y_5^{'},0,-lX_5^{'},-dX_5^{'}\\ 0,lY_6^{'},(h+d)Y_6^{'},Y_6^{'},0,-lX_6^{'},-(h+d)X_6^{'}\\ wY_7^{'},lY_7^{'},(h+d)Y_7^{'},Y_7^{'},-wX_7^{'},-lX_7^{'},-(h+d)X_7^{'}\end{bmatrix}\begin{bmatrix}\frac{s_x r_{xx}}{T_y}\\ \frac{s_x r_{xy}}{T_y}\\ \frac{s_x r_{xz}}{T_y}\\ \frac{s_x T_x}{T_y}\\ \frac{r_{yx}}{T_y}\\ \frac{r_{yy}}{T_y}\\ \frac{r_{yz}}{T_y}\end{bmatrix} = \begin{bmatrix}X_1^{'}\\ X_2^{'}\\ X_3^{'}\\ X_4^{'}\\ X_5^{'}\\ X_6^{'}\\ X_7^{'}\end{bmatrix} \tag{9}$$

To solve Eq. (9), we have to prove that the coefficient matrix in this equation has full rank.

### *C. Proof for the Linear Independence of the Rows of the Coefficient Matrix in Eq. (9)*

Firstly, we should emphasize that among the seven corner points, no three points are collinear on the image plane. During an object's motion through the view range of a video camera, there are a large number of snapshots can be collected in different frames. In order to get an available snapshot for calibration, it is reasonable to assume that on the image plane, any two points among the seven corner points are not collinear with the origin, and no point is on X axis or Y axis.

Let K be the coefficient matrix in Eq. (9), then K can be written as:

$$K=\begin{bmatrix}0 & 0 & dY_1^{'} & Y_1^{'} & 0 & 0 & -dX_1^{'}\\ 0 & 0 & (h+d)Y_2^{'} & Y_2^{'} & 0 & 0 & -(h+d)X_2^{'}\\ wY_3^{'} & 0 & (h+d)Y_3^{'} & Y_3^{'} & -wX_3^{'} & 0 & -(h+d)X_3^{'}\\ wY_4^{'} & 0 & dY_4^{'} & Y_4^{'} & -wX_4^{'} & 0 & -dX_4^{'}\\ 0 & lY_5^{'} & dY_5^{'} & Y_5^{'} & 0 & -lX_5^{'} & -dX_5^{'}\\ 0 & lY_6^{'} & (h+d)Y_6^{'} & Y_6^{'} & 0 & -lX_6^{'} & -(h+d)X_6^{'}\\ wY_7^{'} & lY_7^{'} & (h+d)Y_7^{'} & Y_7^{'} & -wX_7^{'} & -lX_7^{'} & -(h+d)X_7^{'}\end{bmatrix}$$

Let $M_i$ be the $i^{th}$ row of K, it is to be shown that the necessary and sufficient condition for $\sum_{i=1}^{7} a_i M_i = 0$ is that $a_i = 0$ $(i = 1,2,\ldots,7)$. The sufficiency is obvious in this case, so we move on to the necessity.

From columns 2 and 6, $a_5 lY_5^{'} + a_6 lY_6^{'} + a_7 lY_7^{'} = 0$ and $a_5 lX_5^{'} + a_6 lX_6^{'} + a_7 lX_7^{'} = 0$. Note that $l, w, h, d$ are all positive, we get:

$$a_3(X_3^{'}, Y_3^{'}) + a_3(X_3^{'}, Y_3^{'}) + a_7(X_7^{'}, Y_7^{'}) = 0 \tag{10}$$

If $a_5 \neq 0, a_6 \neq 0$ and $a_7 \neq 0$, then $(X_5^{'}, Y_5^{'}) = -\frac{a_6}{a_5}(X_6^{'}, Y_6^{'}) - \frac{a_7}{a_5}(X_7^{'}, Y_7^{'})$, which is impossible since $P_5$, $P_6$, $P_7$ are non-collinear points. Therefore, at least one of $a_5, a_6$, or $a_7$ is zero. Due to symmetry, it suffices to take $a_5 = 0$. Now, if $a_6 \neq 0$ and $a_7 \neq 0$, it is also impossible since we assumed no two points have a linear relationship with the origin of the image coordinate. Hence, due to symmetry again, we take $a_6 = 0$, then it is obvious $a_7 = 0$.

Similarly, from column 1 and column 5, we have:

$$a_3(X_3^{'}, Y_3^{'}) + a_4(X_4^{'}, Y_4^{'}) + a_7(X_7^{'}, Y_7^{'}) = 0 \tag{11}$$

then $a_3 = 0, a_4 = 0$.

Finally, consider the first two rows. From columns 3 and 4, we have:

$$\begin{cases} a_1 dY_1^{'} + a_2(h+d)Y_2^{'} = 0 \\ a_1 Y_1^{'} + a_2 Y_2^{'} = 0 \end{cases} \tag{12}$$

Eliminating $a_1$ yields $a_2 hY_2^{'} = 0$, which means $a_2 = 0$. And based on $a_1 Y_1^{'} + a_2 Y_2^{'} = 0$, we get $a_1 = 0$. Hence, the necessary and sufficient condition for $\sum_{i=1}^{7} a_i M_i = 0$ is that $a_i = 0$ for $i = 1, 2, \ldots, 7$, which implies the linear independence of the rows of the coefficient matrix in Eq. (9).

### *D. Solving the surveillance camera parameters*

As the matrix has full rank, the seven unknowns, i.e. $\frac{s_x r_{xx}}{T_y}$, $\frac{s_x r_{xy}}{T_y}$, $\frac{s_x r_{xz}}{T_y}$, $\frac{s_x T_x}{T_y}$, $\frac{r_{yx}}{T_y}$, $\frac{r_{yy}}{T_y}$, and $\frac{r_{yz}}{T_y}$ can then be solved.

We denote that $(m_1, m_2, m_3, m_4, m_5, m_6, m_7) = (\frac{s_x r_{xx}}{T_y}, \frac{s_x r_{xy}}{T_y}, \frac{s_x r_{xz}}{T_y}, \frac{s_x T_x}{T_y}, \frac{r_{yx}}{T_y}, \frac{r_{yy}}{T_y}, \frac{r_{yz}}{T_y})$. Note that the rotation matrix which contains $r_{xx}$, $r_{xy}$, …, $r_{zz}$ is orthonormal, there

are three properties we will use later:

1) $r_{ix}^2 + r_{iy}^2 + r_{iz}^2 = 1, r_{xi}^2 + r_{yi}^2 + r_{zi}^2 = 1,\ (i = x, y, z)$

2) $(r_{ix}, r_{iy}, r_{iz}) \cdot (r_{jx}, r_{jy}, r_{jz})^T = 0,\ (r_{xi}, r_{yi}, r_{zi}) \cdot (r_{xj}, r_{yj}, r_{zj})^T = 0,$
$(i, j = x, y, z; i \neq j)$

3) $\det \begin{vmatrix} r_{xx}, r_{xy}, r_{xz} \\ r_{yx}, r_{yy}, r_{yz} \\ r_{zx}, r_{zy}, r_{zz} \end{vmatrix} = 1$

Then, camera parameters $(r_{xx}, r_{xy}, \ldots, r_{zz}, T_x, T_y, T_z, s_x, f)$ can be computed in following steps.

Given that $m_5^2 + m_6^2 + m_7^2 = \frac{1}{T_y^2}$, and $m_1^2 + m_2^2 + m_3^2 = \frac{s_x^2}{T_y^2}$, we get $s_x = \sqrt{(m_1^2 + m_2^2 + m_3^2) T_y^2}$.

In order to determine the sign of $T_y$, we pick an object point whose computer image coordinate $(u_i, v_i)$ is away from the image center $(u_0, v_0)$. We assume the sign of $T_y$ is positive, and then compute parameters $r_{xx}, r_{xy}, r_{xz}, r_{yx}, r_{yy}, r_{yz}, T_x, X_{ci}, Y_{ci}$ in the following way. $r_{xx} = m_1 T_y / s_x$, $r_{xy} = m_2 T_y / s_x$, $r_{xz} = m_3 T_y / s_x$, $r_{yx} = m_5 T_y$, $r_{yy} = m_6 T_y$, $r_{yz} = m_7 T_y$, $T_x = m_4 T_y / s_x$,
$X_{ci} = r_{xx} X_{wi} + r_{xy} Y_{wi} + r_{xz} Z_{wi} + T_x$,
$Y_{ci} = r_{yx} X_{wi} + r_{yy} Y_{wi} + r_{yz} Z_{wi} + T_y$

Note that $f > 0$ and $Z_c > 0$ in $\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix}$. If $X_{ci} X_i > 0$ and $Y_{ci} Y_i > 0$, the sign of $T_y$ is positive. Otherwise, the sign of $T_y$ is negative. When the sign of $T_y$ is known, another 7 unknowns are solved. So far, only $r_{zx}$, $r_{zy}$, $r_{zz}$, f, and $T_z$ remain to be solved.

With property 1), we can derive the value of $|r_{zx}|$, $|r_{zy}|$, and $|r_{zz}|$. For example, $|r_{zx}| = \sqrt{1 - r_{yx}^2 - r_{xx}^2}$.

With property 2), every two column vectors are orthonormal, so $r_{zx} r_{zy} = -r_{xx} r_{xy} - r_{yx} r_{yy}$. Similarly, we can derive $r_{zx} r_{zz}$ and $r_{zy} r_{zz}$.

Thus, $| r_{zx} r_{zy} r_{zz} | = \sqrt{(r_{zx} r_{zy}) \cdot (r_{zy} r_{zz}) \cdot (r_{zx} r_{zz})}$.

To determine the sign of $r_{zx}, r_{zy}, r_{zz}$, just assume $r_{zx} r_{zy} r_{zz} > 0$, then we get $r_{zx} = r_{zx}^{'}, r_{zy} = r_{zy}^{'}, r_{zz} = r_{zz}^{'}$. With property 3), if the determinant of $\begin{bmatrix} r_{xx}, r_{xy}, r_{xz} \\ r_{yx}, r_{yy}, r_{yz} \\ r_{zx}^{'}, r_{zy}^{'}, r_{zz}^{'} \end{bmatrix}$ is positive, $r_{zi} = r_{zi}^{'} (i = x, y, z)$, otherwise $r_{zi} = -r_{zi}^{'} (i = x, y, z)$.

Finally, we come to the point of solving $f$ and $T_z$. Pick the seven container corner points $P_i$ (i=1,2,…,7) in the world coordinate and read its image coordinate $(X_i, Y_i)$ out from corresponding video snapshot. Since $s_x$ has already been solved, set up seven equations in the following pattern using these seven points, $X_i = f \frac{r_{xx} X_{wi} + r_{xy} Y_{wi} + r_{xz} Z_{wi} + T_x}{r_{zx} X_{wi} + r_{zy} Y_{wi} + r_{zz} Z_{wi} + T_z}$. Transpose all the equations to get seven inconsistent equations, and then solve them using Least Squares Method to get approximate $f$ and $T_z$.

### *E. Image Reconstruction*

To retrieve traffic data, three-dimensional information in world coordinates is usually necessary. However, with two-dimensional information obtained from an image, only two-dimensional information in world coordinates can be calculated. Only if data in another dimension were acquired could the world coordinates be reconstructed completely.

That said, most traffic data have nothing to do with vehicle height. For instance, we can estimate the average speed of a vehicle by calculating different positions of the vehicle's projections from a camera to the ground at different moments. Therefore, the points on ground are most likely to be reconstructed since their height information is known to be zero and consistent in the model. Other points with known heights are also available and can be reconstructed when camera parameters are known. In other words, $Z_w$, which is the height of a point in the real world, can show information of a third dimension. Reconstruction can proceed by first rearranging Eq. (5) and thus yields:

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{bmatrix}^{-1} \left( \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} - \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \right) \tag{13}$$

Considering X, Y, and $f$ are known in $\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix}$, it is justifiable to assume $X_c = k_1 Z_c$ and $Y_c = k_2 Z_c$, where $k_1 = \frac{u - u_0}{s_x f}$ and $k_2 = \frac{v - v_0}{f}$. Just plug them into Eq. (13), we derive Eq. (14):

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{bmatrix}^{-1} \left( \begin{bmatrix} k_1 Z_c - T_x \\ k_2 Z_c - T_y \\ Z_c - T_z \end{bmatrix} \right) \tag{14}$$

Since $Z_w$, $T_z$ and Matrix R are known, $Z_c$ can be solved. Thus, world coordinate $(X_w, Y_w, Z_w)$ is finally obtained.

## III. Experimental Results

### *A. Parameters Solving*

The following video snapshot (Fig. 2) shows an example for calibrating our model from a roadway surveillance camera.


Fig. 2. A standard shipping container in two consecutive video frames.

A rough estimate for $d$ is 3.5' here, while it has only a little influence on Matrix $(T_x, T_y, T_z)$ and reconstruction. As shown in Fig. 3, we manually collect the seven corner points' pixel numbers and run the model in Matlab. Table I lists the recovered $R$, $T$, $f$, and $s_x$.

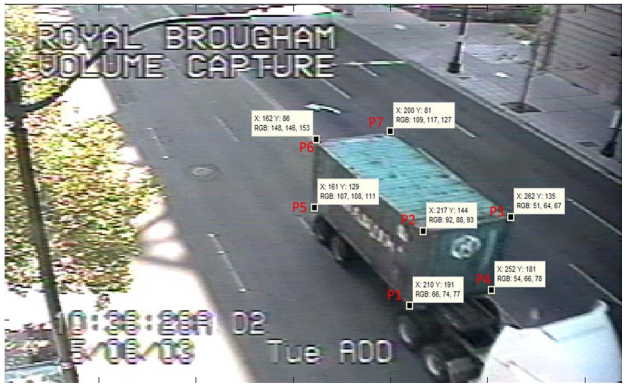


Fig. 3. Identify the seven corner points and obtain their pixel numbers for calibration.

TABLE I
RESULTS OF CALIBRATED CAMERA PARAMETERS

| Parameters | CAMERA CALIBRATION RESULTS |
|---|---|
| *Rotation matrix (R)* | 0.9166 -0.3991 0.0246<br>-0.1564 -0.4649 -0.8714<br>0.3432 0.8540 -0.4629 |
| *Translation matrix (T)* | 8.0751<br>13.8752<br>57.1012 |
| *Scale factor ($s_x$)* | 0.9475 |
| *Focal length (f)* | 371.7290 |

### B. Verification Using Rotation Matrix

We haven't completely made use of the property 3) of matrix R in solving camera parameters. For property 1) and 2) of matrix R, it is clear that not all the rows or columns were taken to compute the unknowns. With the properties that have been used before, the determinant of matrix R still cannot be around the true value if it is not solved properly. Hence, the results shown in this work are convincing since the determinant of matrix R turns out to be 1.0283, which is very close to 1.

### C. Verification Using Lane Markings

Four lane markings are selected and the pixel numbers of their endpoints are obtained as in Fig. 4. Running the reconstruction model to calculate the world coordinates of these eight endpoints; the results are shown in Table II.

According to Table II, the values of the $X_w$ coordinates of each pair of endpoints on the same lane marking are the same within an acceptable error range (point 1 and 2, 3 and 4, 5 and 6, 7 and 8 are four pairs of endpoints on the same lane markings), which is identical theoretically because the $X_w$ axis is approximately parallel to lane markings.

It should also be noted that the marking lines on the same section of a road should have the same length value. We can get these four marking lines' lengths approximately using the start and end point of each lane marking by simple arithmetic. Thus, the lengths are 12.8', 11.7', 12.5', 11.9', which are nearly the same within an acceptable error range.

Finally, we consider point 1, 2, 5, and 6. First of all, the $Y_w$ values of 1 and 5 are almost the same, as are the $Y_w$ values of 2 and 6, which is quite reasonable. Then, we estimate the lane width at 11' by calculating the average value of the distance between point 1, 5 and point 2, 6. The widths of vehicle lanes typically vary from 9 feet to 15 feet, while 10-12 feet is the most common range. 11' is a reasonably good estimate.

Fig. 4. Obtain the pixel numbers of lane endpoints.

TABLE II
RECONSTRUCTION RESULTS FOR LANE MARKINGS

| Point | u | v | $Z_c$ | $X_w$ | $Y_w$ | $Z_w$ |
|---|---|---|---|---|---|---|
| *1* | 76 | 78 | 103.5 | -10.3 | 58.5 | 0 |
| *2* | 84 | 97 | 92.5 | -10.7 | 45.7 | 0 |
| *3* | 108 | 145 | 73.0 | -10.6 | 22.9 | 0 |
| *4* | 125 | 181 | 63.0 | -10.8 | 11.2 | 0 |
| *5* | 111 | 71 | 109.0 | 0.9 | 60.4 | 0 |
| *6* | 122 | 88 | 98.1 | 0.5 | 47.9 | 0 |
| *7* | 223 | 68 | 113.9 | 36.0 | 52.0 | 0 |
| *8* | 243 | 83 | 103.5 | 35.3 | 40.1 | 0 |

### D. Verification Using the Same Shipping Container in A Different Snapshot

The same shipping container in a different frame can also be a verification for the results. Since the height of all seven corner points of the shipping container are known and do not change, we calibrate the seven points as shown in Fig. 5. Table III presents the reconstruction results.

In the ideal situation, Point 1, 2, 3 and 4 have the same $Y_w$ coordinates, as do Point 5, 6 and 7; Point 1, 2, 5 and 6 have the same $X_w$ coordinates, as do Point 3, 4 and 7. As we can see, Table III presents quite reasonable reconstruction results according to this principle.

In addition, as we know the true distance between every two container corner points, we can choose any two corner points and compute their reconstructed distance to see if the result is reasonable compared to the true value. For example, from the results in Table III, the reconstructed distance between Point 1 and 5 are estimated to be 20.1', which is very close to its true value 20'.

Further, if the model is established correctly, the shipping container should have the same size in different frames. The reconstructed length and width of the shipping container in this snapshot are estimated in Eq. (15) and Eq. (16), respectively.

$$\text{Length} = \frac{38.6+36.2+37.7}{3} - \frac{16.1+17.1+17.1+16.5}{4} \quad (15)$$

$$\text{Width} = \frac{9.8+9.2+9.0}{3} - \frac{0.0+0.8+0.2+0.8}{4} \quad (16)$$

The reconstructed length and width are 20.8' and 8.8', while the true values should be 20' and 8'. The relative errors here are respectively 4.0% and 9.1%.

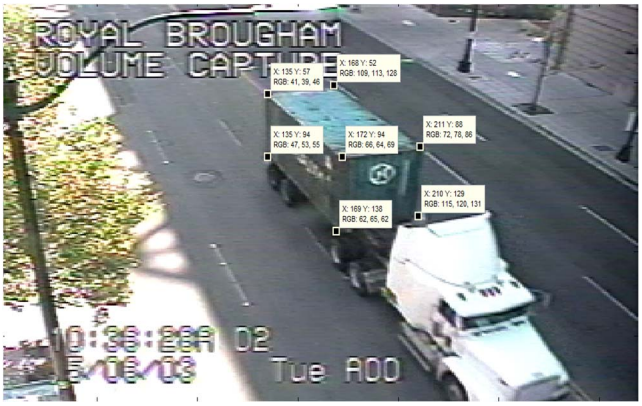



Fig. 5. Obtain the pixel numbers of corner points of the same truck in another snapshot.

TABLE III
RECONSTRUCTION RESULTS FOR CONTAINER CORNER POINTS

| Point | u | v | $Z_c$ | $X_w$ | $Y_w$ | $Z_w$ |
|---|---|---|---|---|---|---|
| *1* | 169 | 138 | 69.2 | 0.0 | 16.1 | 3.5 |
| *2* | 172 | 94 | 66.4 | 0.8 | 17.1 | 12.0 |
| *3* | 211 | 88 | 69.3 | 9.2 | 17.1 | 12.0 |
| *4* | 208 | 129 | 72.6 | 9.0 | 16.5 | 3.5 |
| *5* | 135 | 94 | 86.5 | 0.2 | 36.2 | 3.5 |
| *6* | 135 | 57 | 84.0 | 0.8 | 37.7 | 12.0 |
| *7* | 168 | 52 | 88.0 | 9.8 | 38.6 | 12.0 |

## IV. CONCLUSION AND DISCUSSION

Roadway surveillance camera calibration is a pivotal component of video-based traffic analysis. This paper proposed a cost-effective camera calibration method which has the potential to further facilitate the application of camera calibration in a real-world scenario. The proposed camera calibration method takes standard shipping container with known dimensions as reference objects. Experimental results show that the calibration method developed in this paper is able to reconstruct 3D coordinates from 2D images efficiently with high accuracy. Because of the robustness in the mathematical derivation and the properties of shipping containers (large size, bright color), the proposed method is not sensitive to environmental and roadway conditions such as weather condition, lane marking quality, and road type. Errors in the model mainly come from two sources: 1) the measuring error in collecting image pixel coordinates; 2) the rough estimate of parameter "d" in some situations, such as an unknown truck type. Future work on this model should focus on improving accuracy, however, for most traffic data collection purposes, these errors may not be among the critical factors to consider.


## ACKNOWLEDGEMENT

This research was supported in part by the Shenzhen Science and Technology Planning Project (Grant No. GJHZ20150316154158400).


## REFERENCES


[1] L. A. Klein, "Traffic flow sensor technologies," in *Sensor Technologies and Data Requirements for ITS,* Boston: Artech House, 2001, pp. 239-317.

[2] I. Sobel, "On calibrating computer controlled cameras for perceiving 3D scenes," *Artif. Intell.*, Vol. 5, pp. 185-198, 1974.

[3] W. Faig, "Calibration of close-range photogrammetry systems: mathematical formulation," *Photogramm. Eng. Remote Sens.*, Vol. 41, pp. 1479-1486, 1975.

[4] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 14, pp. 965-980, Oct. 1992.

[5] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Proc. IEEE Conf. CVPR*, 1986, pp. 364-374.

[6] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, Vol. 3, pp. 323-344, Aug. 1987.

[7] M. Ito, and A. Ishii, "A non-iterative procedure for rapid and precise camera calibration," *Pattern Recogn.*, Vol. 27, pp. 301-310, 1994.

[8] F. Remondino and C. Fraser, "Digital camera calibration methods: considerations and comparisons," *Int. Arch. Photogram. Rem. Sens. Spatial Inform. Sci.,* Vol.36, pp. 266-272, 2006.

[9] D. J. Dailey, F. W. Cathey, and S. Pumrin, "An algorithm to estimate mean traffic speed using uncalibrated cameras," *IEEE Trans. Intell. Transport. Syst.*, Vol.1, pp.98-107, Mar. 2000.

[10] G. Zhang, R. P. Avery, and Y. Wang, "Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras," *Trans. Res. Rec., J. Transp. Res. Board*, Vol.1993, pp.138-147, 2007.

[11] Y. Li, F. Zhu, and Y. Ai et al. "On automatic and dynamic camera calibration based on traffic visual surveillance," in *IEEE Intell. Veh. Sym.*, 2007, pp. 358-363.

[12] Z. Chen and P. Shi, "Efficient method for camera calibration in traffic scenes," *Electron. Lett.,* Vol.40, pp.368-369, Mar. 2004.

[13] T. N. Schoepflin and D. J. Dailey, "Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation," *IEEE Trans. Intell. Transport. Syst.,* Vol.4, pp.90-98, Jun. 2003.

[14] N. K. Kanhere, S. T. Birchfield, and W. A. Sarasua, "Automatic camera calibration using pattern detection for vision-based speed sensing," *Trans. Res. Rec., J. Transp. Res. Board*, Vol.2086, pp. 30-39, 2008.

[15] K. Ismail, T. Sayed, and N. Saunier, "A methodology for precise camera calibration for data collection applications in urban traffic scenes," *Can. J. Civil Eng.* Vol.40, pp. 57-67, Jan. 2013.

[16] N. Krahnstoever and P. R. S. Mendonca, "Bayesian autocalibration for surveillance," in *Tenth IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1858-1865.

[17] F. Lv , T. Zhao, and R. Nevatia, "Camera calibration from video of a walking human," *IEEE Trans Pattern Anal. Mach. Intell.*, Vol.28, pp. 1513-1518, Sept. 2006.

[18] B. Micusik and Pajdla,T. "Simultaneous surveillance camera calibration and foot-head homology estimation from human detections," in *IEEE Conf. CVPR*, 2010, pp. 1562-1569.

[19] E. K. Bas and J. D. Crisman. "An easy to install camera calibration for traffic monitoring," in *Proc. IEEE Conf. ITS*, 1997, pp. 362-366.

[20] Z. Zhang, T. Tan, and K. Huang et al, "Practical Camera Calibration from Moving Objects for Traffic Scene Surveillance," *IEEE Trans Circuits Syst. Video Technol.*, Vol.23, pp. 518-533, Mar. 2013.

[21] R. Ke, S. Kim, Z. Li, and Y. Wang, "Motion-vector clustering for traffic speed detection from UAV video," in *Proc. IEEE Conf. Smart Cities,* Guadalajara, Mexico, 2015.

[22] R. Ke, Z. Li, S. Kim, J. Ash, Z. Cui, and Y. Wang, "Real-time bidirectional traffic flow parameter estimation from aerial videos," *IEEE Trans. Intell. Transp. Syst.,* vol. 18, no. 4, pp. 890-901, Apr. 2017.

[23] R. Ke, J. Lutin, J. Spears, and Y. Wang, "A cost-effective framework for automated vehicle-pedestrian near-miss detection through onboard monocular vision," in *Proc. IEEE Conf. CVPRW*, 2017, pp. 898-905.