Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part C 15 (2007) 361–379
TRANSPORTATION RESEARCH PART C
www.elsevier.com/locate/trc

Using geometric primitives to calibrate traffic scenes

Osama Masoud, Nikolaos P. Papanikolopoulos *

Department of Computer Science and Engineering, University of Minnesota, 4-192 EE/CS Building, 200 Union Street SE, Minneapolis, MN 55455, United States

Received 30 January 2006; received in revised form 28 May 2007; accepted 30 May 2007

Abstract

In this paper, we address the problem of recovering the intrinsic and extrinsic parameters of a camera or a group of cameras in a setting overlooking a traffic scene. Unlike many other settings, conventional camera calibration techniques are not applicable in this case. We present a method that uses certain geometric primitives commonly found in traffic scenes, such as straight and curved lanes, lane markings, and poles in order to recover calibration parameters. We show experimentally that these primitives provide the needed redundancy and are capable of achieving accurate results suitable for most traffic monitoring applications.
Published by Elsevier Ltd.

Keywords: Calibration; Multi-view stereo; Multi-view reconstruction; Application systems; Traffic monitoring

1. Introduction

Images of natural and man-made environments exhibit certain regularities that should be utilized. One of these regularities is the presence of geometric entities and constraints that bind them together. Structure-from-motion methods that only use low-level geometric entities (or features) such as points and lines without additional constraints suffer from two main disadvantages. First, they usually require a large number of features to achieve robustness; and second, because there are no constraints among the features, errors in localizing these features in the image propagate to the structure unnoticed. It is therefore no surprise that primitive-based approaches for reconstruction and camera calibration are on the rise (Bartoli and Sturm, 2003; Bazin, 2000; Bondyfalat et al., 1998; Cipolla and Robertson, 1999; Criminisi et al., 2001; Debevec et al., 1996; Grossmann et al., 2001; Gurdjos and Payrissat, 2000; Liebowitz and Zisserman, 1998; Wilczkowiak et al., 2003). It is a very effective way to make use of the a priori knowledge in natural and man-made scenes. The primitives used can be planes, cubes, prisms, etc., and the relationships can be parallelism, orthogonality, coincidence, angle, distance, and so on.

This paper presents a primitive-based approach that targets traffic scenes. Traffic monitoring applications have long been and are still interested in computer vision techniques. Unfortunately, the input data available

* Corresponding author. Tel.: +1 612 625 0163; fax: +1 612 625 0572.
E-mail addresses: masoud@cs.umn.edu (O. Masoud), npapas@cs.umn.edu (N.P. Papanikolopoulos).

0968-090X/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.trc.2007.05.005

to these applications comes from cameras that are already mounted in an outdoor setting with little known information about the camera parameters (e.g., height, zoom, tilt, etc.). Calibrating these scenes is essential to produce measurements needed by traffic monitoring applications. For instance, a vehicle classification application that depends on estimating the real-world dimensions of vehicles (width, length, and height), will need some form of transformation that can map these real-world distances to image space. Similarly, applications that perform vehicle tracking to estimate real-world trajectories often employ stochastic estimation algorithms (e.g., the Kalman Filter). Having a world-to-image mapping not only makes it possible to estimate real-world quantities (e.g., vehicle locations, speeds, etc.) but also, through the propagation of error from the measurement space (the image space), facilitate the generation of estimation errors in these quantities.

Camera calibration is an important problem that has received a considerable amount of attention in the literature. Accurate camera calibration requires the use of especially designed patterns to be placed in the field of view of the camera. However, in many cases, such as in a traffic situation, this is not practical or even possible since one would need a very large calibration pattern let alone having to place it on the road.

The calibration requirements for traffic monitoring applications depend both on the application itself and on the scene. For instance, in a scene of a highway segment, where vehicle motion is not expected to vary significantly, and captured by a camera with a certain pose, partial calibration, e.g., measuring the change in scale in the general direction of motion, may be sufficient for specific tasks such as measuring average speeds. In other cases, such as at an intersection, where vehicles can move freely in many different directions, full calibration would be necessary. In this paper, we focus on the problem of full calibration which, as will be shown, requires only a small number of primitives that can be found in most traffic scenes. In addition to traffic scenes, the presented method is applicable in other situations (e.g., an environment of mobile robots) where similar geometric primitives can be found.

Depending on the application at hand, primitive-based methods select an appropriate set of relevant primitives (Bartoli and Sturm, 2003; Cipolla and Robertson, 1999; Criminisi et al., 2001; Debevec et al., 1996; Grossmann et al., 2001; Gurdjos and Payrissat, 2000; Liebowitz and Zisserman, 1998). In a similar manner, we select primitives commonly found in a traffic scene. Fig. 1 shows a depiction of a typical traffic scene and camera layout. The proposed primitives (straight lane structure, curved lane structure, point-to-point distances, normal, horizontal, and parallel lines) are usually either obvious in the scene, are previously known properties of the scene (e.g., lane width), or as in the case of point-to-point distances, can be measured. Our method then solves for camera parameters and scene structure by minimizing reprojection errors in the image.

A number of methods (Bazin, 2000; Bondyfalat et al., 1998; Wilczkowiak et al., 2003) have been proposed that addressed the primitive-based structure-from-motion problem as a theorem-proving and/or constraint propagation problem. These methods can accept arbitrary geometric constraints involving points, lines, and planes, provided as a grammar. The flexibility in such methods makes them suitable for large size

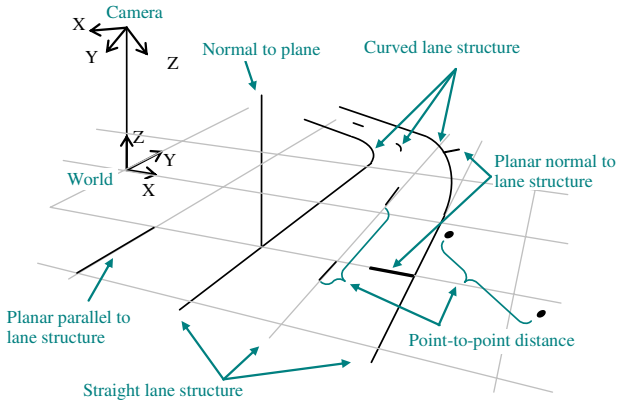

Fig. 1. Common traffic scenes geometric primitives; the figure also shows camera and ground plane coordinate systems.

problems such as architectural modeling. However, these methods still need to deal with one or more of a number of issues, such as the guarantee to find a solution, computational cost, and problems arising from the presence of redundant constraints. In our case, the primitives we deal with are well defined and therefore we can choose the parameters optimally.

In Worrall et al. (1994), an interactive method was proposed to perform traffic scene calibration. Although very intuitive, it relies on the user's visual judgment rather than actual measurements. Other approaches that utilized traffic scene geometry include (Bas and Crisman, 1997; Ferrier et al., 1994; Schoepflin and Dailey, 2003; Worrall et al., 1993). Our approach differs in that it utilizes more primitives and provides a natural way to combine them.

The contributions of this paper are: (i) A method for calibrating traffic scenes from primitives, such as straight or curved lanes, selected from a single image and multiple images; (ii) An error analysis of the effectiveness of using the proposed primitives by comparing our calibration results to those of a robust calibration method.

This paper is organized as follows. Section 2 discusses the camera parameters and assumptions. In Section 3, the geometric primitives used in our method are presented. The cost function and the optimization procedure are presented in Section 4 followed by a discussion of initial solution generation in Section 5. Section 6 discusses the extension to multiple cameras. The results are then presented in Section 7 followed by the conclusion in Section 8.

2. Scene and camera modeling

The objective of traffic scene calibration is to obtain a mapping between the image plane and the world. A common characteristic of traffic scenes is the presence of a road segment with well defined boundaries or markings. In this paper, we restrict our scene model to planar (but possibly curved) road segments, which are the most common. Although our method is in general still applicable to nonplanar roads by incorporating a parameterization of the road into the model, this has been left as a future work. Under this assumption, the geometric primitives that we use in this paper have one thing in common: they are related through coincidence or orthogonality relationships to a plane representing the ground (see Fig. 1). This is similar to the *ground-plane constraint* (GPC) of Worrall et al. (1994). Although apparently planar roads and intersections are usually not perfectly planar (e.g., they bulge upward to facilitate drainage), this is still a valid assumption as the deviation from planarity is insignificant (e.g., relative to camera height).

There are different ways of modeling the mapping from the world to the image plane. The mapping from the ground plane to the image plane (and vice versa) can be accomplished via a homography transformation. As discussed above, it is also desirable to be able to map other points, not necessarily on the ground plane. The latter can be achieved using a planar homology transformation as was done in Criminisi et al. (2001). This homography–homology pair of transformations is therefore sufficient in mapping any world point to the image plane. This approach is attractive because it does not require knowledge about intrinsic camera parameters. However, this comes at the risk of making the model more underdetermined and hence at a cost to accuracy. To see why this is, we must state that the geometric primitives that are available in a typical traffic scene are limited in number. These primitives are far from being a calibration pattern but rather a small set of sparse features. It is therefore imperative to make the most use of these features. A more constrained model (i.e., with fewer degrees of freedom) is preferable for this reason. To demonstrate this, consider a case where the features consist of three pairs of parallel lines: two pairs coincident with the plane and are normal to one another, and the third pair normal to the plane. A homography is uniquely defined from the ground plane lines (up to two scales). But this homography is useless as it is in mapping points not on the plane. To be able to achieve that, a homology needs to be defined using the third pair (normal to the plane). But there is a clear decoupling due to the large number of degrees of freedom: if the third pair of lines is changed, only the homology is affected. Contrast this with a Euclidean model where certain assumptions are placed on the camera's intrinsic parameters (known aspect ratio, zero skew, and known principal point). In this case only the lines on the plane are sufficient for full reconstruction (up to a scale). The third pair of lines (if available) will place more constraints and can only improve the solution. This advantage is due to the fact that the latter model has four degrees of

freedom less than the former. For this reason and for reasons explained below, we use the latter approach in this paper.

Our approach involves the recovery of the camera's intrinsic and extrinsic parameters. These parameters describe the formation of an image point from the projection of a 3D point **P** onto the image plane. Other effects on image formation, such as radial distortion, can be corrected for using certain methods (e.g., Devernay and Faugeras, 1995) prior to using the approach described in this paper. In a pinhole camera, this process can be expressed in homogeneous coordinates as

$$[x \quad y \quad w]^{\mathrm{T}} = \mathbf{ATP} \tag{1}$$

where $\mathbf{T} = [\mathbf{R}|\mathbf{t}]$ relates the world coordinate system to that of the camera through a rotation **R** and a translation **t**. The matrix **A** describes the camera's intrinsic parameters which in the case of a zero-skew camera is given by

$$\mathbf{A} = \begin{bmatrix} f & 0 & u_0 \\ 0 & \tau f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The parameter f is proportional to the camera's focal length (and is referred to as the focal length in this paper). The aspect ratio is τ and the principal point is (u_0, v_0). In this paper, we make a *natural camera* assumption (i.e., zero skew angle and known aspect ratio). It is a matter of practicality to make this assumption since these two parameters rarely differ from zero and one (respectively) anyway. Moreover, of all intrinsic parameters, only the focal length changes during camera operation due to changing zoom. Therefore, other parameters could be calibrated at the laboratory if needed. The principal point is also assumed to be known (the center of the image). It has been shown (Ruiz et al., 2002) that the recovery of the principal point is ill-posed especially when the field of view is not wide (which is the case in many traffic scenes).

We attach a coordinate system to the ground plane whose origin is the point closest to the camera (Fig. 1) and whose Y-axis depends on the main primitive used (see below). Therefore, in our model, if we let $\mathbf{R} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$,

$$\mathbf{T} = [\mathbf{R}| - h\mathbf{r}_3], \tag{3}$$

where h is the camera height. There are four degrees of freedom that relate the camera's coordinate system to the ground plane coordinate system. These may be understood as the camera's height, roll, pitch, and yaw. With the addition of focal length, this makes the total number of parameters to be estimated equal to five plus any parameters specific to the primitives. The primitives are essentially independent from one another and the only thing that relates them is the ground plane. Therefore, they are independently parameterized with respect to the ground plane coordinate system. It is worth noting that the equivalent ground-to-image homography given this model can be written as

$$\mathbf{H} = \mathbf{A}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad -h\mathbf{r}_3]. \tag{4}$$

3. Geometric primitives

3.1. Straight lane structure

Central to a traffic scene is what we refer to as a lane structure. Based on the nature of the road segment visible in the scene, a straight or a curved lane structure constitutes this main primitive. A straight lane structure consists of a set of parallel lines coincident to the ground plane with known distances among them. The Y-axis of the ground plane coordinate system is arbitrarily set to be parallel to the straight lane structure. Given this ground plane coordinate system, we can fully specify a lane structure with exactly one variable: the X-intercept of one of its lines (see Fig. 1). The straight lane structure with three or more lines is a powerful primitive as it encodes four constraints on the five camera parameters discussed above. As will be shown in Section 5.2, a reparameterization of the five camera parameters consists of two vanishing points and a scale.

The straight lane structure provides one vanishing point. The ratios between the lane widths provides the vanishing line and one of the widths itself provides a scale. The only remaining parameter is an offset on the vanishing line specifying the second vanishing point. This remaining ambiguity can be resolved if a planar normal to the lane structure is available or if one or more point-to-point distances, are available. Examples of this primitive can be seen in Figs. 1, 3, 7 and 10.

3.2. Curved lane structure

A curved lane structure consists of a set of concentric circles with known difference in radii (lane widths). Concentric circles comprise a good model since road design guidelines dictate that curved roads be constructed as circular arcs. The Y-axis of the ground plane coordinate system is arbitrarily set to point to the center of these circles. Given a ground plane coordinate system, a curved lane structure can then be fully specified by the distance to the center along the Y-axis and the radius of one of the circles (e.g., the innermost circle). This primitive is also a powerful one even more than the straight lane structure. In fact, a two-circle curved lane structure is sufficient to recover all the camera parameters. This is discussed in more detail in Section 5.4. Because the road is modeled as circular, its projection in the image is a set of conic sections. Therefore, five user-specified points are sufficient to describe each curve in this lane structure. Examples of this primitive can be seen in Figs. 1, 11 and 12.

3.3. Ground plane point-to-point distances

These primitives can be obtained from knowledge about the road structure (e.g., longitudinal lane marking separation) or by performing field measurements between landmarks on the ground. Another *creative* way of obtaining these measurements is by identifying the make and model of a vehicle from the traffic video and then looking up that model's wheelbase dimension and assigning it to the line segment in the image connecting the two wheels. A fixed length segment connecting two points can be fully specified in the ground plane coordinate system by three parameters: a 2D point (e.g., the midpoint) and a rotation angle (e.g., off the X-axis). Examples if this primitive can be seen in Figs. 1 and 3.

3.4. Line segments that are normal to the plane, planar parallel to lane structure, or planar normal to lane structure

These can represent poles, building corner edges, and pedestrian crossings, among other things. The planar primitives are primarily related to a lane structure. Normals to the plane can be specified by a single 2D point on the ground plane. In the case of a straight lane structure, planar normal (resp. parallel) lines can be specified by a Y (resp. X) coordinate. In the case of a curved lane structure only planar normals are meaningful. These pass through the circles' center and can be specified by an angle with respect to the Y-axis. Examples of these primitives can be seen in Figs. 1, 3, 7, 10, and 11.

4. Cost function and optimization

The cost function is the sum of squared reprojection errors in the image. In the case of point features (such as in point-to-point distances), it is straightforward what this means. In the case of line features, there is no universally agreed upon comparison error function. Many techniques that performed structure-from-motion using line features used one form or another for comparing the model and feature lines (Bartoli et al., 2003; Spetsakis and Aloimonos, 1990; Taylor and Kriegman, 1995; Zhang, 1994). In our case, we consider the error in a line segment as the error in the two points that specify the line segment. Consequently, the reprojection error for a line segment becomes the square of the two orthogonal distances from the end points to the reprojected model line. This is advantageous since it makes it possible to combine the errors from point and line features together in one cost function. This is also advantageous because the certainty about the location of a line is implicit in the segment length. Therefore, if only a short segment of a line is visible in the image, the user should only specify the endpoints of the visible part and not extrapolate. In a similar fashion the error

associated with curved features is the sum of square distances from the control points that were specified to the reprojected circles. Since the reprojection of a circle is a conic section, we simply find point on the conic section closest to the control point.

The cost function is optimized iteratively using the Levenberg–Marquardt method. The search is performed on camera parameters (focal length and extrinsic, a total of five) and model parameters. The camera's rotation is represented in angle-axis form where the axis is represented in spherical coordinates. The model parameters are:

1. Straight lane structure: one parameter (X-intercept of an arbitrarily selected line).
2. Curved lane structure: two parameters (Y-coordinate of the circles' center and the radius of the innermost circle).
3. Point-to-point distances: three parameters each (as explained above), with the 2D point represented in polar coordinates.
4. Normals to plane, planar normals to lane structure, and planar parallels to lane structure: no parameters are needed because it is possible to compute a closed form solution in image space. The closed form solution is given in Appendix A. Note that in the case of curved lane structure, the same procedure can be applied with respect to planar normals, by making the circle center act as a vanishing point.

5. Initial solution

An initial solution close to the global minimum is needed to guarantee convergence of the above optimization. Since not all primitive types need to be specified by the user, the initial solution can be computed in different ways depending on which primitives have been supplied. If the lane structure supplied is straight, one vanishing point can be estimated. A second vanishing point can be estimated if two or more normals to the plane, or two or more planar normals to the lane structure are also specified. For the case of curved lane structure, we use a different approach to generate the initial solution as described below. We will now explain how we estimate the vanishing points and then the computation of the initial solution.

5.1. Vanishing point estimation

There are many methods for estimating the vanishing point from a set of convergent line segments. Many of these methods use statistical models for errors in the segments (Collins and Weiss, 1990; Kanatani, 1995; Liebowitz and Zisserman, 1998). Since the vanishing points we need are used in generating the initial solution, we instead estimate the vanishing point as simply the point with the minimum sum of square distances to all the lines passing through these segments. Let $\mathbf{u}_i$ be a unit normal to the line L_i that passes through segment i's endpoints $\mathbf{a}_i$ and $\mathbf{b}_i$. Given a point $\mathbf{p}$, the orthogonal distance from $\mathbf{p}$ to L_i is $|\mathbf{u}_i \cdot (\mathbf{p} - \mathbf{a}_i)|$ or $|\mathbf{u}_i \cdot \mathbf{p} - \mathbf{u}_i \cdot \mathbf{a}_i|$. Therefore, the sum of square distances from point $\mathbf{p}$ to a set of n lines can be written as

$$\sum_{i=1}^{n} (\mathbf{u}_i \cdot \mathbf{p} - \mathbf{u}_i \cdot \mathbf{a}_i)^2 \tag{5}$$

Minimizing this sum is equivalent to solving the linear system $\mathbf{Mp} = \mathbf{r}$ where $\mathbf{M} = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n]^{\mathrm{T}}$ and $\mathbf{r} = [\mathbf{u}_1 \cdot \mathbf{a}_1\ \mathbf{u}_2 \cdot \mathbf{a}_2\ \ldots\ \mathbf{u}_n \cdot \mathbf{a}_n]^{\mathrm{T}}$. If $\mathbf{M}$ is (or close to being) rank deficient, it is an indication that the vanishing point sought is at infinity (or very far). It is important to distinguish this case since it limits the usefulness of the vanishing point as we shall see below.

5.2. Initial solution using two vanishing points

If the input primitives include a straight lane structure and two or more normals to the plane, or two or more planar normals to the lane structure, two vanishing points are computed as above. If these two vanishing points are not infinite, they are sufficient to compute four of the five camera parameters. The remaining parameter can then be computed as a scale factor that makes model distances closest to what they should be. The following describes these steps in detail.

The problem of performing calibration from vanishing points has been visited repeatedly in the literature. For example, Caprile and Torre (1990) solve for the principal point and focal length from three vanishing points and Hartley and Kaucic (2002) solve for the focal length from a vanishing point and vanishing line through the remaining two points. Here, we solve for the focal length given two vanishing points and our assumptions on the knowledge of certain camera parameters. Without loss of generality, let $\mathbf{v}_y$ and $\mathbf{v}_z$ be the two vanishing image points corresponding to the ground's Y- and Z-axes. With the camera intrinsic parameters described in Eq. (2), we can compute $\mathbf{p}_y = \mathbf{A}^{-1}[\mathbf{v}_y^{\mathrm{T}}\ 1]^{\mathrm{T}}$ and $\mathbf{p}_z = \mathbf{A}^{-1}[\mathbf{v}_z^{\mathrm{T}}\ 1]^{\mathrm{T}}$ as the vectors in the camera coordinate system corresponding to $\mathbf{v}_y$ and $\mathbf{v}_z$, respectively (i.e., they are parallel to the ground's Y- and Z-axes, respectively). Since $\mathbf{p}_y$ and $\mathbf{p}_z$ are necessarily orthogonal, we write

$$\mathbf{p}_y \cdot \mathbf{p}_z = 0 \tag{6}$$

This equation has the following solution for the focal length:

$$f = \sqrt{-(\mathbf{v}_y - \mathbf{D})^{\mathrm{T}} \begin{bmatrix} 1 & 0 \\ 0 & \tau^{-2} \end{bmatrix} (\mathbf{v}_z - \mathbf{D})} \tag{7}$$

where $\mathbf{D} = [u_0\ v_0]^{\mathrm{T}}$ is the principal point. This solution is valid for finite vanishing points only. This is because a vanishing point at infinity corresponds to a vector parallel to the image plane. Given any orthogonal vector to this vector (representing another vanishing point), it can be easily seen that varying the focal length will not affect the orthogonality of these two vectors. Therefore, in cases where one of two vanishing point is infinite (or very large), we disregard it and assume that only one vanishing point is available (see Section 5.3 below).

The rotation matrix can now be formed from $\mathbf{p}_y$, $\mathbf{p}_z$ and $\mathbf{p}_x$ (the latter computed as the cross product of the former two) after normalization. Finally, the scale is determined as follows. We first assume a scale of one to complete the camera parameters. Primitives that involve distances (e.g., lane structure, point-to-point distances) are then projected from the image to the ground to produce *computed* distances on the ground plane. Let the original (measured) and the corresponding computed distances be specified as two vectors $\mathbf{m}$ and $\mathbf{c}$, respectively. The scale, s, is chosen to minimize $\|s\mathbf{c} - \mathbf{m}\|$. This is simply $s = \mathbf{c} \cdot \mathbf{m}/\|\mathbf{c}\|^2$.

5.3. Initial solution using one vanishing point

When there is only one finite vanishing point available, three remaining camera parameters still need to be determined: the focal length, a rotation about the vanishing direction, and camera height. Fortunately, the latter can dealt with as a last step as before. To solve for the former two, we try to match distance ratios between the measured distances with distance ratios between the computed distances. The optimization is done using the Levenberg–Marquardt method. The residual we try to minimize is completely dependent on the ratios among the scene measurements. If m_i and c_i represent the measured and the computed distances, respectively, we use one measurement, m_0, as a reference and relate all other measurements to it. The residual is then computed as

$$r = \sum_{i=1}^{n} \left(\frac{c_i m_0}{c_0 m_i} - 1 \right)^2 \tag{8}$$

We found that this converges rapidly and the choice of the initial values does not affect the convergence but there can be multiple solutions. The desired solution can always be found from any of these solutions (e.g., using a positive f).

5.4. Initial solution using curved lane structure

When instead of a straight lane structure, a curved one is supplied, a different method is used to compute the initial solution. A rectifying homography is a mapping between image points and the ground plane up to a similarity. Such a homography would encode eight constraints, two more than the six degrees of freedom in a Euclidean transformation. These two remaining constraints can be used as constraints on intrinsic parameters. Since our model has one unknown intrinsic parameter, only one such homography is required to recover all

our camera parameters. Let $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3]$ be this homography, the two constraints can be expressed as in Sturm and Maybank (1999),

$$\mathbf{h}_1^{\mathrm{T}}\omega\mathbf{h}_1 - \mathbf{h}_1^{\mathrm{T}}\omega\mathbf{h}_2 = 0 \qquad \mathbf{h}_1^{\mathrm{T}}\omega\mathbf{h}_2 = 0 \tag{9}$$

where $\omega = \mathbf{A}^{-\mathrm{T}}\mathbf{A}^{-1}$ is the image of the absolute conic. There are two ways to solve for ω, and subsequently $\mathbf{A}$. Since there are two constraints, we can assume the aspect ratio τ is unknown, resulting in two unknown variables. Once the solution f' and τ' is found, f can be re-estimated as

$$f'' = \arg\min_f (f - f')^2 + (\tau f - \tau' f')^2, \tag{10}$$

where τ is the known aspect ratio. The problem with this approach is the presence of many degenerate and common configurations in which it is not possible to find a solution (Sturm and Maybank, 1999). The alternative is to assume that only f is unknown and thus solve an overdetermined linear homogeneous system in the least squares sense as outlined in (Sturm and Maybank, 1999). The only degeneracy associated with this approach is when the camera plane is parallel to the ground plane. This configuration, in which the focal length and scale have the same effect, is quite rare in traffic scenes and can be handled as a special case of the camera model. Once f is computed, $\mathbf{R} = [\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3]$ can be computed as follows

$$\begin{aligned} \mathbf{r}_1 &= \frac{\mathbf{A}^{-1}\mathbf{h}_1}{\|\mathbf{A}^{-1}\mathbf{h}_1\|}, \\ \mathbf{r}_2 &= \frac{\mathbf{A}^{-1}\mathbf{h}_2}{\|\mathbf{A}^{-1}\mathbf{h}_2\|}, \\ \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2. \end{aligned} \tag{11}$$

We thus far described how the camera parameters (all but the scale, which we will recover below) can be extracted if a rectifying homography is available. Kim et al. (2005) have recently shown that it is possible to compute the two imaged circular points from the projection two concentric circles. The imaged circular points are two complex conjugate points $\mathbf{x}_1 \pm i\mathbf{x}_2$ on the vanishing line that are shared by all images of circles. Once found, a rectifying homography can be constructed by setting

$$\mathbf{h}_1 = \mathbf{x}_1 \quad \text{and} \quad \mathbf{h}_2 = \mathbf{x}_2 \tag{12}$$

We will represent a conic section in 2D using a matrix $\mathbf{C} \in \Re^{3\times 3}$ such that a 2D point $\mathbf{X}$ in homogeneous coordinates belongs to this conic if and only if

$$\mathbf{X}^{\mathrm{T}}\mathbf{C}\mathbf{X} = 0. \tag{13}$$

It has been shown in Kim et al. (2005) that if $\mathbf{C}_1$ and $\mathbf{C}_2$ are the projections of two concentric circles, there is rank 2 matrix $\mathbf{\Delta} = \beta\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1}$ which is the image of the conic dual to the circular points. After computing this matrix by solving for the generalized eigenvalues of $\mathbf{C}_1^{-1}$ and $\mathbf{C}_2^{-1}$, the imaged circular points can be extracted from $\mathbf{\Delta}$ in a straightforward manner.

The initial solution is still incomplete; the camera height and two parameters specific to the curved lane structure (the Y-coordinate of the circles' center and the radius of the innermost circle) still need to be estimated. Also, the transformation computed thus far has an arbitrary orientation (the Y-axis may not be pointing towards the circles' center). To resolve all these issues, we fit a model consisting of a set of concentric circles (equal to the number of user-specified conics) to the backprojection of supplied control points on the ground plane. It is important to fit the model to the resultant points as opposed to the backprojected conics that these points describe since these points signify anchor point with high confidence on the possibly very long perimeter of the conics. The equation of a circle centered at (u, v) having a radius r_i can be written as

$$x^2 + y^2 - 2uxw - 2vyw + w^2(u^2 + v^2 - r_i^2) = 0, \tag{14}$$

where $\mathbf{p} = (x, y, w)^{\mathrm{T}}$ is the homogeneous coordinates of a 2D point on the circle. By letting $c_i = u^2 + v^2 - r_i^2$, we can rewrite (14) as

$$\mathbf{M}\mathbf{x} = 0, \tag{15}$$

where, $\mathbf{M} = [x^2 + y^2 - 2xw - 2yww^2]$, $\mathbf{x} = \lambda[1\ u\ v\ c_i]^{\mathrm{T}}$ and λ is a scale factor. Assuming there are n concentric circles, each having m control points, and if $\mathbf{p}_{ik} = (x_{ik}, y_{ik}, w_{ik})^{\mathrm{T}}$ is the backprojection of the kth control point belonging to the circle, we define

$$\mathbf{M}_{ik} = \left[x_{ik}^2 + y_{ik}^2 \quad -2x_{ik}w_{ik} \quad -2y_{ik}w_{ik} \quad \underbrace{0 \quad \cdots \quad w_{ik}^2}_{i} \quad \underbrace{\cdots \quad 0}_{n-i} \right].$$

We can now redefine

$$\begin{aligned} \mathbf{M} &= \left[\mathbf{M}_{11}^{\mathrm{T}} \quad \cdots \quad \mathbf{M}_{1m}^{\mathrm{T}} \quad \cdots \quad \cdots \quad \mathbf{M}_{n1}^{\mathrm{T}} \quad \cdots \quad \mathbf{M}_{nm}^{\mathrm{T}} \right]^{\mathrm{T}}, \\ \mathbf{x} &= \lambda \left[1 \quad u \quad v \quad c_1 \quad c_2 \quad \cdots \quad c_n \right]^{\mathrm{T}} \end{aligned} \tag{16}$$

to make Eq. (15) describe the complete ensemble of concentric circles. This homogeneous linear system can be solved in the least squares sense producing a fit for the backprojection of all control points. Because the cost being minimized is algebraic (as opposed to a geometric one), we normalize each $\mathbf{p}_{ik}$ to have a unit length to improve stability. A 2D rotation can now be applied to $\mathbf{R}$ to make the circles' center lie on the Y-axis. Furthermore, since part of the curved lane structure includes the lane widths, we can now estimate the scale that makes the difference in the recovered radii above closest to the supplied values.

6. Multiple cameras

When images of the scene from multiple cameras are available, the following additional information can be used:

1. *Lane structure constraints.* These constraints are not required but will be used if provided:
   (a) Straight lane structure: parallelism or orthogonality. This happens if the lane structures in two or more images correspond to the same road or to two roads intersecting at right angles.
   (b) Curved lane structure: common center of concentric circles. This happens if the lane structures in two or more images correspond to the same curved road.
2. *Point correspondences.* These may or may not be part of the points used to specify the primitives in the individual images. Depending on whether lane structure constraints are supplied, a minimum of one or two point correspondences is required between each pair of cameras.

Note that we have not used any other correspondences among primitives across cameras (other than the coincidence of the ground plane and the direction of the lane structure). One reason is that imposing correspondence of what seems to be the same primitive may not be a good idea. Consider for example a marker pen in Fig. 2. The left edge of the same marker as seen in the two images corresponds to two different lines in space because the marker does not have a zero radius. In practice, it is easier for the user to concentrate on marking primitives in a single image without worrying about performing correspondences with other images.

The mapping between two ground plane coordinate systems can be uniquely defined using a 2D translation and a 2D rotation. Constraint 1(a) removes the need for the rotation while constraint 1(b) removes the need for the 2D translation. Either way, these parameters are added to the set of parameters to be solved for during multiple camera optimization. This optimization is done as a final step after each camera is optimized independently. During this final step, the parameters optimized are the parameters for all cameras, the primitives in each image, and the parameters relating ground planes (2D rotations and translations, based on the supplied constraints on lane structures). The cost function is the same as before but now also includes reprojection errors from point correspondences. So if a point $\mathbf{p}_{\mathrm{A}}$ in camera A's image corresponds to a point $\mathbf{p}_{\mathrm{B}}$ in cameras B's image, $\mathbf{p}_{\mathrm{A}}$ is projected to the ground plane and reprojected onto B's image where the distance to $\mathbf{p}_{\mathrm{B}}$ can be computed. The same is also done in reverse.

Before the multiple camera optimization can start however, an initial assignment for the 2D rotations and translations is necessary. This can be easily done using one or two of the supplied point correspondences (depending on lane structure constraints provided).

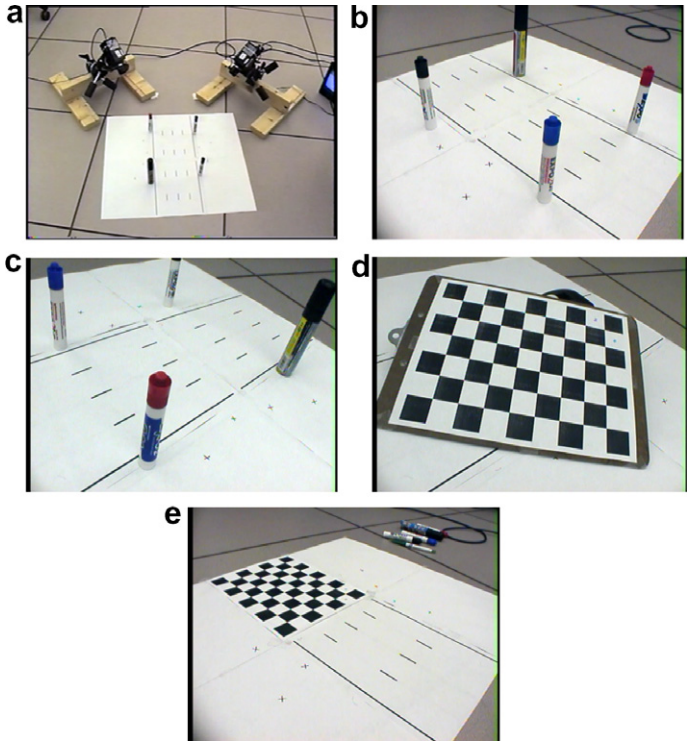

Fig. 2. Experimental setup and calibration procedure.

7. Results

In this section, we present results from lab experiments as well as actual traffic scenes. In order evaluate the quality of the calibration parameters that our method produces, we constructed a mini-road scene, which is a scaled down version of a typical road in all its aspects (e.g., lane widths, lane markings lengths, markings paint widths, etc.). The scale is approximately 1:78. We also used two cameras A and B. Fig. 2a shows a snapshot of the setup. Marker pens standing on their flat end were used to represent vertical poles in the scene. Fig. 2b and c show the images captured by the two cameras. The cameras are standard CCD with 6 mm lens giving them a horizontal field of view of about 60 degrees. The images are captured at a 640×480 resolution.

The reason for doing this scaling down is that it allows us to perform very accurate calibration of the cameras using a robust method, something that is not possible had we used an image of an actual traffic scene. This is essential in order to be able to produce a quantitative comparison. The robust method we used is by Bouguet (2005). In order to calibrate the cameras with this method, several images of a pattern (e.g., Fig. 2d) are first collected. We collected nine such images, one of which had the pattern carefully placed on the road and aligned with it (see Fig. 2e). This makes it possible to relate the coordinate system of the road with that of this pattern in order to be able to compute reprojection errors of all patterns. In Bouguet (2005), the user has the flexibility to choose which intrinsic parameters to optimize. We chose to estimate the focal length (2 parameters, assuming unknown aspect-ratio), the principal point (2 parameters), and lens distortion (4 parameters, a 4th order radial distortion model with a tangential component). The availability of many calibration images enables us to do this. The cameras are then simultaneously calibrated to refine all parameters. The RMS reprojection error was on the order of approximately 0.3 pixels for both cameras. We also repeated the process but this time with the restriction on intrinsic parameters that our method uses (i.e., a constant aspect ratio, a known principal point (image center) and no distortion model). This was done to give an idea of the expected lowest error when using a method that enforces these restrictions like ours. The results are shown in Table 1. Using an elaborate intrinsic model has an advantage but the restricted model is still acceptable with errors being less than one pixel.

Table 1
RMS reprojection errors using all patterns (in pixels)

Camera A	Camera B	Combined
Unconstrained intrinsic model		
0.27	0.31	0.29
Restricted intrinsic model		
0.86	0.92	0.89

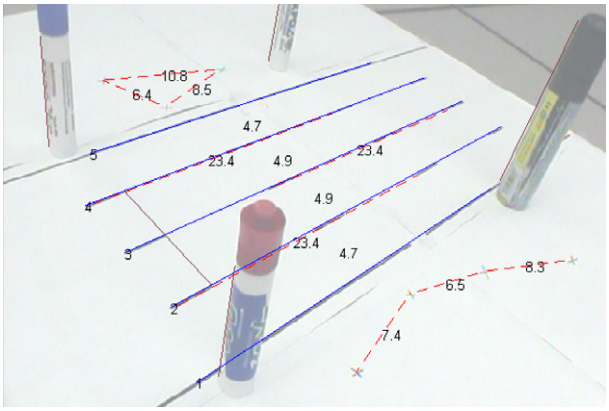

Fig. 3. Specification of primitives.

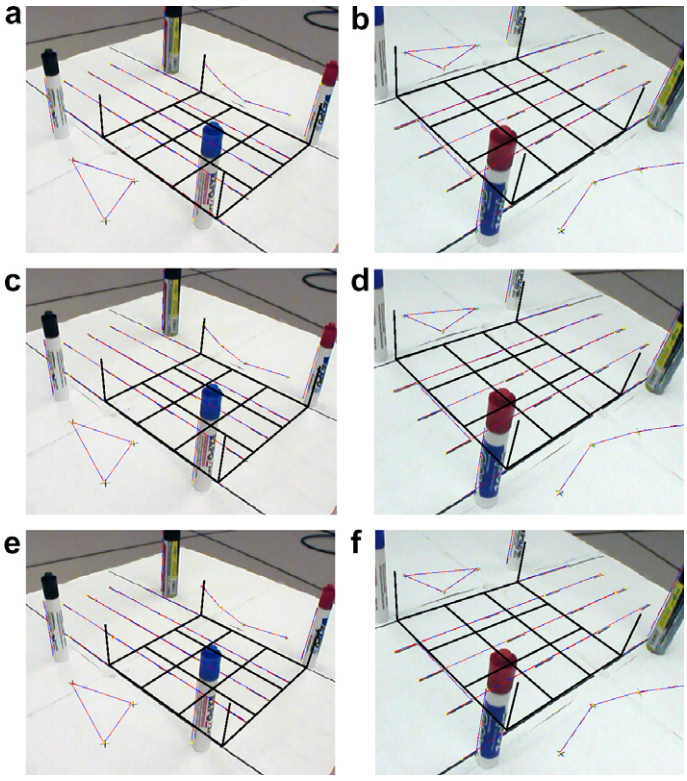

Fig. 4. Calibration stages for cameras A and B. (a,b) Results after initial solutions; (c,d) single image optimization; (e,f) results after cross-camera optimization.

To model this scene, we used a 5-line straight lane structure, 9 point-to-point distances, and 4 normals to the plane in each of the cameras. Camera B had an additional primitive: a planar normal to lane structure. These primitives are shown graphically (on a shaded background for clarity) in Fig. 3 for camera B. After generating the initial solution, the optimization was performed on each camera individually. Then the two cameras were optimized simultaneously using 8 correspondence points and a parallelism constraint on the two lane structures. Fig. 4 shows the results after each stage. The differences from one stage to the next are subtle (e.g., see parallelism to the planar normal in camera B's image). This is because the initial solution was already

Table 2
RMS reprojection errors using primitives (in pixels)

Camera A		Camera B		Combined	
Model	Pattern	Model	Pattern	Model	Pattern
Primitives: single image					
0.56[a]	1.67	1.08[a]	3.6	0.87[a]	2.83
Primitives: stereo					
0.98	2.91	1.47	2.17	1.25	2.56

[a] Do not include point correspondence errors.

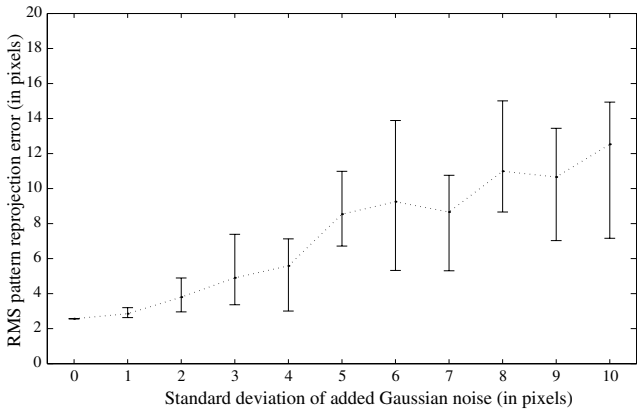

Fig. 5. Effect of adding noise to measurement coordinates. Graph shows minimum, maximum, and average RMS reprojection error after four trials at each point.

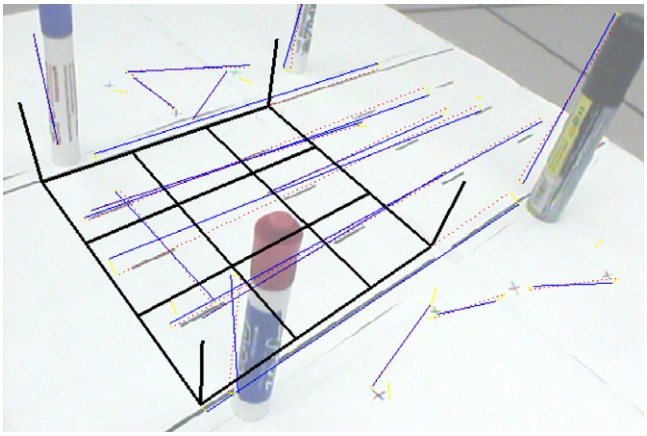

Fig. 6. Robustness under extreme noise (standard deviation 10). Solid lines are corrupted measurements and dotted lines represent the reprojection of the reconstructed model. The overlaid grid provides a visual indicator of reconstruction quality.

very good. Quantitative results are shown in Table 2. Values under "model" correspond to the RMS reprojection error resulting from projecting the geometric primitives to the image and computing the distances to the corresponding features. This is exactly the cost function being optimized and is therefore expected to be smaller than "pattern" error. The latter is the RMS reprojection error of all corners of all nine patterns. Notice that when calibrating multiple views simultaneously, the model error is higher than when using a single image. This is due to over-fitting noisy or otherwise insufficient features in the single image case. The combined pattern error, however, is decreased after simultaneous optimization, indicating an improvement over single-image optimization. This error is still three times larger than the best achievable but it is very small considering that a single pair of images was used to obtain it. As for the model error value of 1.25 pixels, it corresponds to

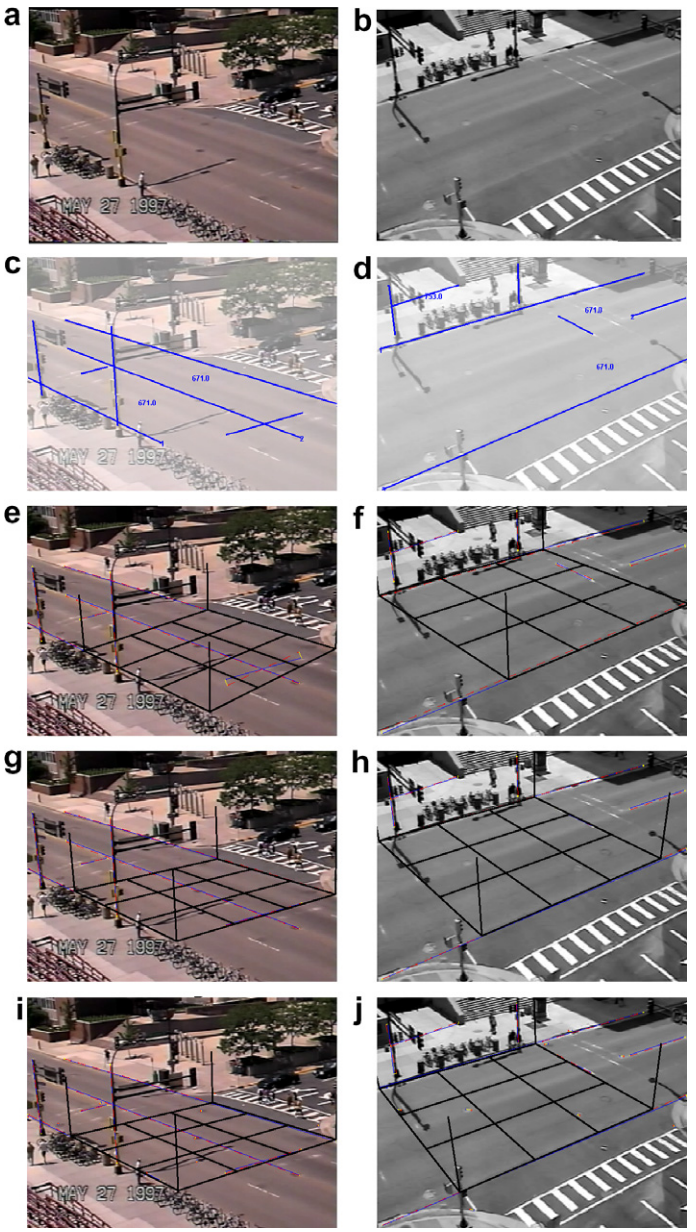

Fig. 7. Calibration of a real traffic scene using one and two cameras.

Fig. 8. Manually fitted cuboids on vehicles using the calibration parameters.

approximately 10 cm in the scaled-up version of this road at a point on the road near the center of the image. This is very acceptable in most traffic applications.

We also performed a sensitivity analysis by adding random Gaussian noise to the x- and y-coordinates of all image features. Fig. 5 shows how the RMS pattern reprojection error is affected. In all cases, a solution where the cameras are looking in the general right direction was found (see Fig. 6). We believe that the mere fact that the reconstruction did not fail is due to the inherent redundancy in the primitives. The figure also demonstrates graceful degradation of the solution quality.

Results from actual traffic scenes are now presented. Two images of the same traffic scene were captured by two different cameras A and B at 320×240 resolution. The primitives used were a 3-line straight lane structure and two normals to the plane. In addition, Camera A had two planar normals to lane structure while camera B had one such line and one point-to-point distance. The two images from the scene overlaid with these measurements are shown in Fig. 7a–d. Notice that the marked line segment corresponding to the middle line of camera B's lane structure is short. This is intentional since this was the only part that is clearly visible in the image and it is better not to extrapolate. The initial solution (Fig. 7e–f) is further improved after image-based optimization (Fig. 7g–h) but it still has problems as can be observed by noticing how parallelism between the overlaid grid and the shadow of the pole on the road progresses. The simultaneous optimization step uses nine point correspondences and the results from that look further improved (Fig. 7i–j). The RMS reprojection error is 2.0 pixels. This corresponds to approximately a 40 cm and a 20 cm distance on the road around the center of the images of camera A and B, respectively. From our experience, selecting more primitives and more accurate distances can further reduce this error. Fig. 8 gives a qualitative assessment of the results. The images shown are of the same time instant and the lines drawn are of manually placed cuboids whose bottom sides are coincident to the ground plane. It can be seen that the cuboids fit vehicles well including those visible in both cameras.

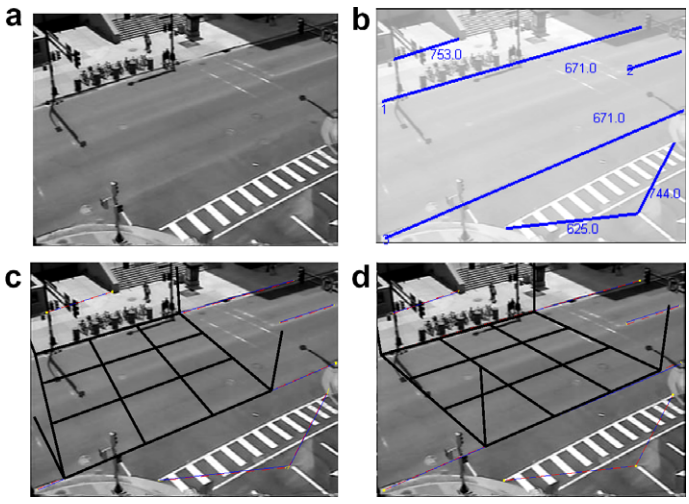

Fig. 9. Calibration of a traffic scene with only one measurable vanishing point.

Fig. 9 shows the same image that was calibrated in Fig. 7h but using different primitives. The primitives in this case (Fig. 9b) only include a 3-line straight lane structure and three point-to-point distances. In this case, only one vanishing point was used to obtain the initial solution (Fig. 9c). After image-based optimization, the solution (Fig. 9d) was improved and looks very similar to Fig. 7h.

In Fig. 10, the results from a different traffic scene are shown. The primitives used, shown in Fig. 10b, are a 3-line straight lane structure, three planar normals to the lane structure, and two normals to the plane. The initial solution is shown in Fig. 10c. Notice that because this solution used the two vanishing points from the lane structure and the normals to the plane, the remaining direction that was not used (the planar normal direction) had the most error. The error diminished after taking all input into consideration during the non-linear optimization step. Figs. 11 and 12 show examples of curved roads. Fig. 11b shows the measurements consisting of two conic sections each specified using five points. In addition, two normals to the plane are supplied. The initial solution in Fig. 11c was generated using only the two conics sections. Fig. 11d shows the final solution and a very good match between the measurements and the model. Fig. 12b shows that only a curved lane structure was supplied, also via five points per conic section. Fig. 12c and d show the initial and final solutions, respectively. The final solution shows almost a perfect match with the control points.

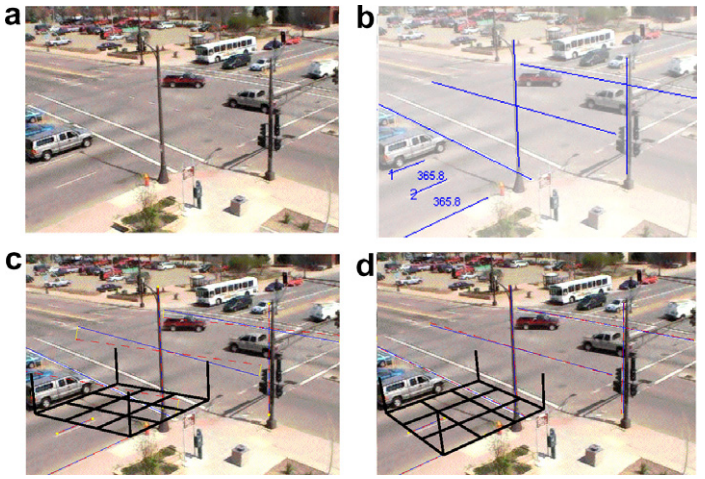

Fig. 10. Calibration of a 4-way traffic intersection.

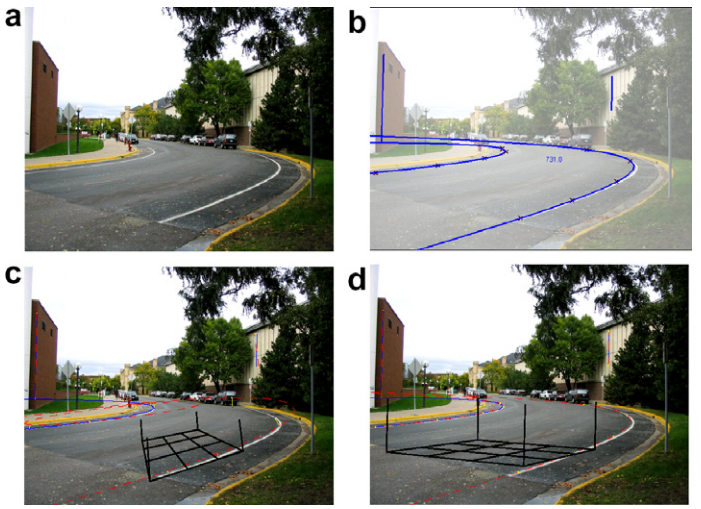

Fig. 11. Calibration of a curved road.

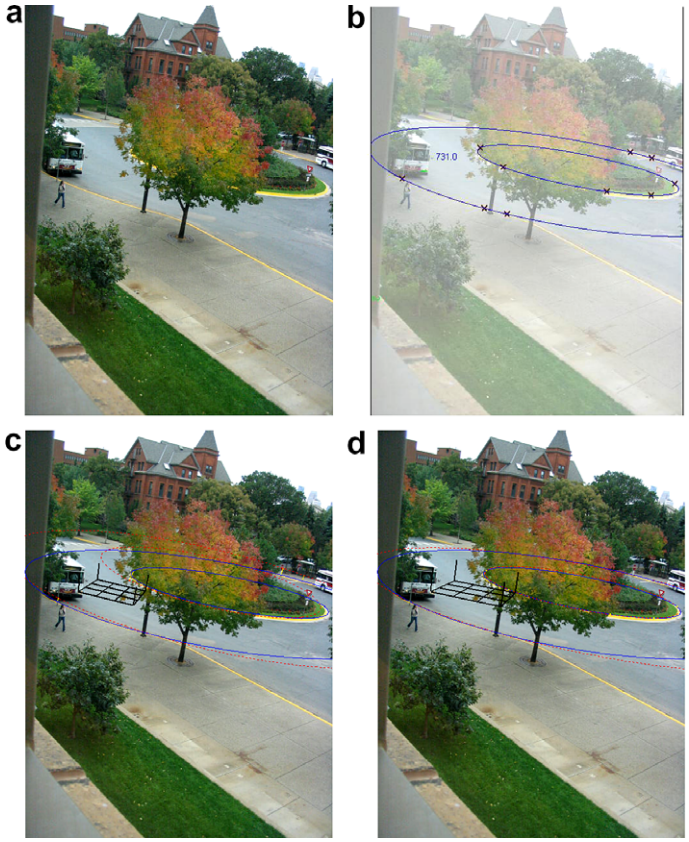

Fig. 12. Calibration of a curved road.

8. Conclusion

We presented a method to compute the camera intrinsic and extrinsic parameters given a single or multiple images of a traffic scene. The geometric primitives used have been carefully chosen to reflect actual primitives one would find in a traffic scene, including scenes with curved roads. We have shown that using these primitives our method is capable of achieving accurate results suitable for most traffic monitoring applications. Several issues need to be investigated. One such issue that would be useful to a user of this method is to have feedback indicating the type and location of the primitive that would be most decisive in improving the calibration accuracy. Such a feedback can guide users if they need to do field measurements. Another issue would be to extend the model to non-planar roads. By parameterizing the ground curvature, one should be able to estimate these parameters in a similar fashion to the methods described in this paper.

Acknowledgements

This work has been supported in part by the National Science Foundation through Grants #IIS-0219863, #CMS-0127893, #CNS-0324864, and #CNS-0224363, the Minnesota Department of Transportation, and the ITS Institute at the University of Minnesota. We would also like to thank the anonymous reviewers for their valuable comments.

Appendix A

The closed form solution mentioned in Section 4 can be found as follows. Without loss of generality, let p_1p_2 be an image segment that was labeled as belonging to a normal line (see Fig. 13). All model normal lines

must go through a vanishing point, v, in the image. Given such a line L and according to our error function, the segment p_1p_2 would contribute an error of $h_1^2 + h_2^2$. We seek L that minimizes this quantity. Let $p_a = ap_1 + (1-a)p_2$ be the desired point on the segment that minimizes $C = h_1^2 + h_2^2$. To find the optimum value of a, we equate the derivate of C with respect to a to zero. Consider the following vectors:

$$\begin{aligned} \mathbf{v}_1 &= v\vec{p}_1, \\ \mathbf{v}_2 &= v\vec{p}_2, \text{ and} \\ \mathbf{v}_a &= v\vec{p}_a = a\mathbf{v}_1 + (1-a)\mathbf{v}_2. \end{aligned} \tag{A.1}$$

Projecting $\mathbf{v}_1$ on $\mathbf{v}_a$ allows us to compute h_1^2:

$$h_1^2 = |\mathbf{v}_1|^2 - \frac{(\mathbf{v}_1 \cdot \mathbf{v}_a)^2}{|\mathbf{v}_a|^2} \tag{A.2}$$

Computing h_2^2 in a similar fashion allows us to write C as follows:

$$C = |\mathbf{v}_1|^2 + |\mathbf{v}_2|^2 - \frac{(\mathbf{v}_1 \cdot \mathbf{v}_a)^2 + (\mathbf{v}_2 \cdot \mathbf{v}_a)^2}{|\mathbf{v}_a|^2} \tag{A.3}$$

Solving $\mathrm{d}C/\mathrm{d}a = 0$ results in the following positive solution for a (after simplification):

$$a = 1 - \frac{1}{1 + t + \sqrt{t^2+1}} \tag{A.4}$$

where

$$t = \frac{|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2}{2\mathbf{v}_1 \cdot \mathbf{v}_2} \tag{A.5}$$

We now consider the possible singularities while evaluating (A.4). First, note that the denominator in (A.4) cannot be 0. Also, although t is undefined when $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, Eq. (A.4) can still be evaluated in the limit and thus this can be handled as a special case:

$$a = \begin{cases} 1 & |\mathbf{v}_1| \geqslant |\mathbf{v}_2| \\ 0 & \text{otherwise} \end{cases} \tag{A.6}$$

More importantly, Eq. (A.4) is valid even when the vanishing point is at infinity (or very far). We can rewrite t as

$$t = \frac{|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2}{2|\mathbf{v}_1||\mathbf{v}_2|\cos\theta} = \frac{1}{2\cos\theta}\left(\frac{|\mathbf{v}_1|}{|\mathbf{v}_2|} - \frac{|\mathbf{v}_2|}{|\mathbf{v}_1|}\right), \tag{A.7}$$

where θ is the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$. This angle can be computed as the difference in orientation between the two lines l_1 and l_2 where $l_i = v \times [p_{ix}\ p_{iy}\ 1]^{\mathrm{T}}$ assuming v is given in homogeneous coordinates, $v = [x\ y\ w]^{\mathrm{T}}$. The terms $|\mathbf{v}_i|/|\mathbf{v}_j|$ can be written as

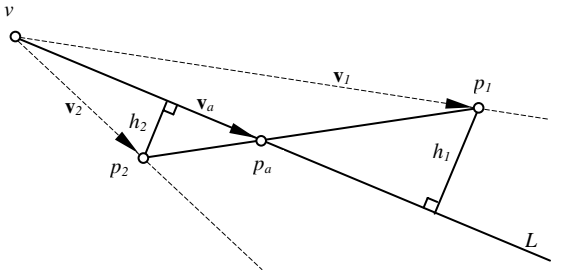

Fig. 13. Finding the best fit for a segment.

$$\frac{\sqrt{\left(\frac{x}{w}-p_{ix}\right)^2+\left(\frac{y}{w}-p_{iy}\right)^2}}{\sqrt{\left(\frac{x}{w}-p_{jx}\right)^2+\left(\frac{y}{w}-p_{jy}\right)^2}}=\frac{\sqrt{(x-wp_{ix})^2+(y-wp_{iy})^2}}{\sqrt{(x-wp_{jx})^2+(y-wp_{jy})^2}}, \tag{A.8}$$

which can be evaluated for any (even infinite) vanishing point as long as the segment endpoints do not coincide with it. In the latter case, C does not depend on a (i.e., a can be arbitrarily assigned 0.5). The errors h_1 and h_2 can now be computed by computing certain angles in a similar manner to how θ was computed:

$$h_i = |p_i - p_a| \sin(\angle v p_a p_i) \tag{A.9}$$

References

Bartoli, A., Sturm, P., 2003. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. IJCV – International Journal of Computer Vision 52 (1), 45–64.

Bartoli, A., Hartley, R.I., Kahl, F., 2003. Motion from 3D line correspondences: linear and non-linear solutions. In: CVPR '03, June 2003, pp. 477–484.

Bas, E.K., Crisman, J.D., 1997. An easy to install camera calibration for traffic monitoring. In: IEEE Conference on Intelligent Transportation System, pp. 362–366.

Bazin, P.L., 2000. A parametric scene reduction algorithm from geometric relations. In: Vision Geometry IX, SPIE's 45th Annual Meeting.

Bondyfalat, D., Mourrain, B., Papadopoulo, T., 1998. An application of automatic theorem proving in computer vision. In: Automated Deduction in Geometry, pp. 207–231.

Bouguet, J.Y., 2005, Camera calibration toolbox for Matlab [online]. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.

Caprile, R., Torre, V., 1990. Using vanishing points for camera calibration. IJCV – International Journal of Computer Vision 4 (2), 127–140.

Cipolla, R., Robertson, D.P., 1999. 3D models of architectural scenes from uncalibrated images and vanishing points. In: IAPR 10th International Conference on Image Analysis and Processing, September, Venice, pp. 824–829.

Collins, R., Weiss, R., 1990. Vanishing point calculation as a statistical inference on the unit sphere. In: International Conference on Computer Vision (ICCV'90), December 1990 Osaka, Japan, pp. 400–403.

Criminisi, A., Reid, I., Zisserman, A., 2001. Single view metrology. IJCV – International Journal of Computer Vision 40 (2), 123–148.

Debevec, P.E., Taylor, C.J., Malik, J., 1996. Modeling and rendering architecture from photographs. In: SIGGRAPH '96, August 1996, pp. 11–12.

Devernay, F., Faugeras, O., 1995. Automatic calibration and removal of distortion from scenes of structured environments. In: SPIE'95, pp. 62–72.

Ferrier, N.J, Rowe, S.M., Blake, A., 1994. Real-time traffic monitoring. In: IEEE Workshop on Applications of Computer Vision, 1994, pp. 81–88.

Grossmann, E., Ortin, D., Santos-Victor, J., 2001. Algebraic aspects of reconstruction of structured scenes from one or more views. In: BMVC '01, pp. 633–642.

Gurdjos, P., Payrissat, R., 2000. About conditions for recovering the metric structures of perpendicular planes from the single ground plane to image homography. In: ICPR '00, pp. 1358–1361.

Hartley, R., Kaucic, R., 2002. Sensitivity of calibration to principal point position. In: ECCV '02, pp. 433–466.

Kanatani, K. 1995. Statistical optimization for geometric computation: theory and practice. Technical report, AI Lab, Dept of Computer Science, Gunma University.

Kim, J.S., Gurdjos, P., Kweon, I.S., 2005. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (4), 637–642.

Liebowitz, D., Zisserman, A., 1998. Metric rectification for perspective images of planes. In: CVPR '98, pp. 482–488.

Ruiz, A., Lopez-de-Teruel, P.E., Garcia-Mateos, G., 2002. A note on principal point estimability. In: ICPR '02, pp. 304–307.

Schoepflin, T.N., Dailey, D.J., 2003. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. IEEE Transactions on Intelligent Transportation Systems 4 (2), 90–98.

Spetsakis, M., Aloimonos, J., 1990. Structure from motion using line correspondences. IJCV – International Journal of Computer Vision 4 (3), 171–183.

Sturm, P., Maybank, S., 1999. On plane-based camera calibration: a general algorithm, singularities, applications. In: CVPR '99, pp. 1432–1437.

Taylor, C.J., Kriegman, D.J., 1995. Structure and motion from line segments in multiple images. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (11), 1021–1032.

Wilczkowiak, M., Trombettoni, G., Jermann, C., Sturm, P., Boyer, E., 2003. Scene modeling based on constraint system decomposition techniques. In: 9th International Conference on Computer Vision (ICCV '03), pp. 1004–1010.

Worrall, A.D., Sullivan, G.D., Baker, K.D., 1993. Advances in Model-based Traffic Vision. In: Proceedings of 4th British Machine Vision Conference, 1993, pp. 559–569.

Worrall, A.D., Sullivan, G.D., Baker, K.D., 1994. A simple, intuitive camera calibration tool for natural images. In: 5th British Machine Vision Conference, 1994, pp. 781–790.

Zhang, Z., 1994. Estimating motion and structure from correspondences of line segments between two perspective images. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (12), 1129–1139.