# Camera Calibration for Urban Traffic Scenes: Practical Issues and a Robust Approach

**Karim Ismail** [*], M.A.Sc.
Research Assistant
Department of Civil Engineering
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
karim@civil.ubc.ca

**Tarek Sayed**, PhD. P.Eng.
Professor, Dept of Civil Engineering
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
604-822-4379
tsayed@civil.ubc.ca

**Nicolas Saunier**, PhD.
Assistant Professor, Department of Civil, Geological and Mining Engineering
École Polytechnique de Montréal
Montréal, Québec
(514) 340-4711 (#4962)
nicolas.saunier@polymtl.ca

*** Corresponding Author**

**Word Count:**

*Manuscript:* 5081 words
*Figure:* 9
*Tables:* 1

*Total:* 7581 words

## ABSTRACT

Video-based collection of traffic data is on the rise. Camera calibration is a necessary step in all applications to recover the real-world positions of the road users of interest that appear in the video. Camera calibration can be performed based on feature correspondences between the real-world space and image space as well as appearances of parallel lines in the image space. In urban traffic scenes, the field of view may be too limited to allow reliable calibration based on parallel lines. Calibration can be complicated in the case of incomplete and noisy data. It is common that cameras monitoring traffic scenes are installed before calibration was undertaken. In this case, laboratory calibration, which is taken for granted in many current approaches, is impossible. This work addresses various real world challenging cases, for example when only video recordings are available, with little knowledge on the camera specifications and setting location, when the orthographic image of the intersection is outdated, or when neither an orthographic image nor a detailed map is available. A review of the current methods for camera calibration reveals little attention to these practical challenges that arise when studying urban intersections to support applications in traffic engineering. This study presents the development details of a robust camera calibration approach based on integrating a collection of geometric information found in urban traffic scenes in a consistent optimization framework. The developed approach was tested on six datasets obtained from urban intersections in British Columbia, California, and Kentucky. The results clearly demonstrated the robustness of the proposed approach.

## 1. BACKGROUND

A research stream that is gaining momentum in traffic engineering strives to adopt vision-based techniques for traffic data collection. The use of video sensors to collect traffic data, primarily by tracking road users, has several advantages:

1. Video recording hardware is relatively inexpensive and technically easy to use.
2. A permanent record of the traffic observations is kept.
3. Video cameras are often already installed and actively monitoring traffic intersections.
4. Video sensors offer rich and detailed data.
5. Video sensors cover a wide field of view. In many instances, one camera is sufficient to monitor an entire intersection.
6. Techniques developed in the realm of computer vision makes automated analysis of video data feasible. Process automation has the advantage of reducing the labour cost and time required for data extraction from videos.

In a typical video sensor, observable parts of real-world objects are projected on the surface of an image sensor, in most cases a plane. An unavoidable reduction in dimensionality accompanies the projection of geometric elements (points, lines, etc.) that belong to a 3-dimensional Euclidian space (world space) onto a 2-dimensional image space. Camera calibration is conducted to map geometric elements, primarily road user positions, from image space to the world space in which metric measurements are possible. The recovery of real-world tracks of road users supports several applications in traffic engineering. Examples are the analysis of microscopic road user behavior, e.g. measuring temporal and special proximity for traffic safety analysis (1; 2), measurement of road user speed (3; 4; 5), and traffic counts (6). In addition, conducting road user tracking in real-world coordinates can improve tracking accuracy by correcting for perspective effect and other distortions due to projection on the image plane. Camera calibration enables the estimation of camera parameters sufficient to *reproject* objects from the image space to a pre-defined surface in the real-world space. A camera can be parameterized by a set of extrinsic and intrinsic parameters. Extrinsic camera parameters describe camera position and orientation. Intrinsic camera parameters are necessary to reduce observations to pixel coordinates.

Three major classes of camera calibration methods can be identified. First are traditional methods based on geometric constraints either found in a scene or synthesized in the form of a calibration patterns. The second class contains self-calibration methods that utilize epipolar constraints on the appearance of features in different image sequences taken from a fixed camera location. Camera self-calibration is sensitive to initialization and can become unstable in case of special motion sequence (7) and in case intrinsic parameters are unknown (8). Active vision calibration methods constitute the third kind of method. They involve controlled and measurable camera movements.

Only the first class of methods lends itself to traffic monitoring in which cameras have been fixed with little knowledge of their intrinsic parameters and control over their orientation, as is the case with many already installed traffic cameras. Other approaches include: linear and non-linear, explicit and implicit (9). Non-linear methods enable a full recovery of intrinsic parameters, as opposed to linear methods. Both methods may be combined, e.g. in (10), by obtaining approximate estimates using linear methods with further refinements using non-linear methods. Inferring camera parameters from implicit transformation matrices obtained using implicit methods is susceptible to noise (11). Limiting calibration to extrinsic parameters gives rise to the topics of pose estimation (12).

Despite the numerous studies of the topic of camera calibration, several challenges can arise due to particularities of urban traffic scenes.

1. Many of the photogrammetry and Computer Vision (CV) techniques available in the literature do not apply due to difference in context, hardware, and target accuracy. Powerful and mature tools such as self-calibrating bundle in the existing literature are not always possible to apply for relatively close-range measurements in urban traffic scenes, especially for images taken by consumer-grade cameras containing noisy or incomplete calibration data (13). In addition, other methods in photogrammetry and CV depend on observing regularization geometry or a calibration pattern. In the typical cases where video cameras are already installed to monitor a traffic scene, or when only video records are available, this procedure cannot be applied.
2. Many of existing techniques rely on parallel vehicle tracks, in lieu of painted lines, for vanishing point estimation (14) (5). Vehicle tracks can be extracted automatically using computer vision techniques. These methods are particularly useful for self-calibration of pan-tilt-zoom cameras used for speed monitoring on rural highways. However, the vehicle motion patterns in urban intersections are not prevalently parallel. An example is shown in Figure 1.
3. Much of the regularization geometry in traffic scenes comprises elements such as road markings that may be altered in many ways. In this study, one of the the monitored traffic sites “BR” was repainted after the orthographic image was taken, making point localization difficult. Using only point correspondences in this case can be unreliable.
4. A significant number of camera calibration methods rely on the observation of one of more sets of parallel co-planar lines. By estimating the points of intersection of these sets of lines, i.e. vanishing points located at the horizon line of the plane that contains these lines, camera parameters can be estimated. In urban traffic environments, the field of view of the camera can be too limited to allow the depth of view necessary for the accurate estimation of the location of the vanishing points. To achieve desirable accuracy, camera calibration must be based on additional geometric information.

5. In many cases, cameras monitoring urban traffic intersections are already installed. Many of these cameras function as traffic surveillance devices, a function that does not necessarily require accurate estimation of road user positions. Given the installation cost and intended functionality, in-lab calibration of intrinsic parameters, e.g. using geometric patterns, can be difficult.

As illustrated in Figure 1, 2 and Table 1, the proposed camera calibration approach was mainly motivated by issues encountered in case studies. These issues are the repainting of traffic pavement marking, the field of view is too limited or non-linear distortion is too strong to enable accurate estimation of vanishing point(s), and the analysis of video sequences collected by other parties. In addition, the geometric regularities abundant in traffic scenes offer geometric information besides the appearance of parallel lines that can increase the accuracy of camera calibration. The majority of the applications supported by this study involved the recovery of real-world coordinates of pedestrian tracks. Pedestrians move significantly slower than the motorized traffic, a characteristic that evidently required higher accuracy for camera parameters. Relying only on geometric information provided by parallel lines yielded camera parameters that provided unsatisfactory pedestrian speed estimates.

**TABLE 1** Summary of Case Studies

| Case Study Code | Site / City | Application | Issues Encountered | C[1] | D[2] | A[3] | E[4] |
|---|---|---|---|---|---|---|---|
| BR-1 | Downtown – Vancouver | Pedestrian Walking Speed (3) | Outdated orthographic map<br>No convergent lines | 13 | 6 | 4 | 0 |
| BR-2 | | | | 11 | 12 | 6 | 0 |
| BR-3 | | | | 5 | 10 | 5 | 0 |
| BR-4 | | | | 9 | 10 | 3 | 0 |
| PG | Downtown – Vancouver | Automated study of Pedestrian-vehicle conflicts (2) | No convergent lines | 22 | 2 | 2 | 0 |
| OK | Chinatown - Oakloand | Automated before-and-after study of pedestrian-vehicle conflicts | Camera unaccessible and not set by authors | 14 | 2 | 9 | 34 |
| K1 | Kentucky | Automated analysis of vehicle-vehicle conflicts | Camera unaccessible and not set by authors<br>Video quality is low<br>Strong non-linear distorsion<br>No orthographic image | 0 | 7 | 2 | 30 |
| K2 | | | | 0 | 7 | 2 | 39 |

[1] The number of point correspondences availabel for calibration.
[2] The number of line segments annotated in the image space with known real-world length.
[3] The number of annotated pairs of lines in the image space the angle between which is known in world space.
[4] The number of line segments annotated for equi-distance constraints. The endpoints of each line segment are annotated at two locations in the camera field of view.

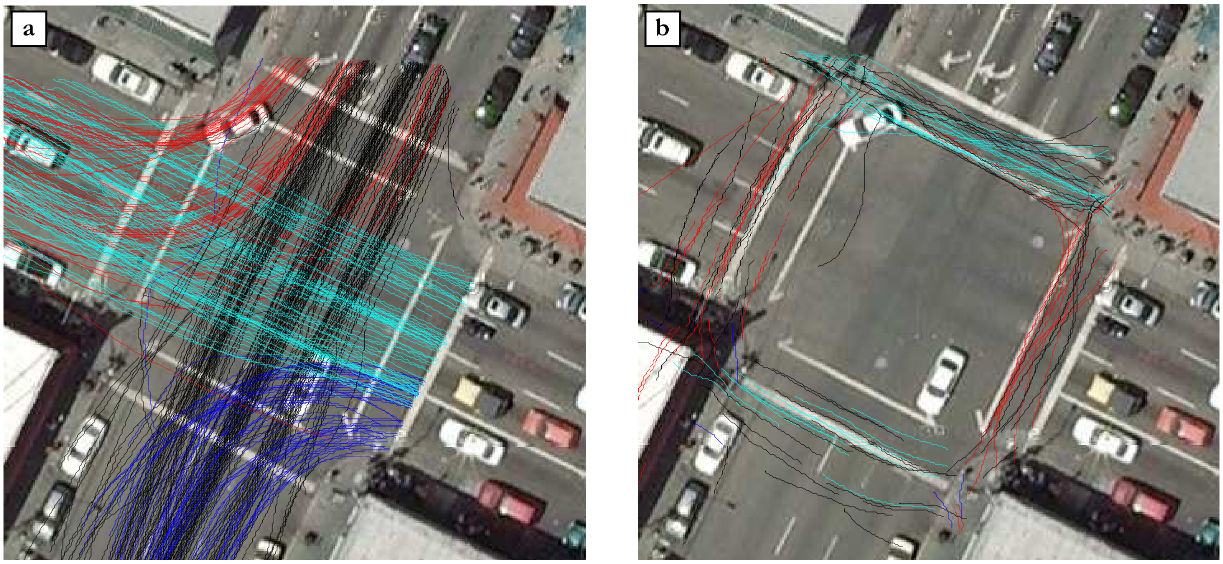


**Figure 1** The difficulty of relying on the automated extraction of road user tracks. Figure **a)** shows the motion patterns of vehicles at a busy intersection in China Town, Oakland-California (sequence OK). Reliance on vehicle tracks for vanishing point estimation is challenging because vehicle tracks do not exhibit enough parallelism. Many patterns representing turning movements and lane changing. Parallel vehicle tracks have to be hand-picked which is tantamount to manually annotating lane marking. Figure **b)** shows pedestrian motion patterns. It is evident that pedestrian tracks do not exhibit prevalent parallelism within crosswalks.

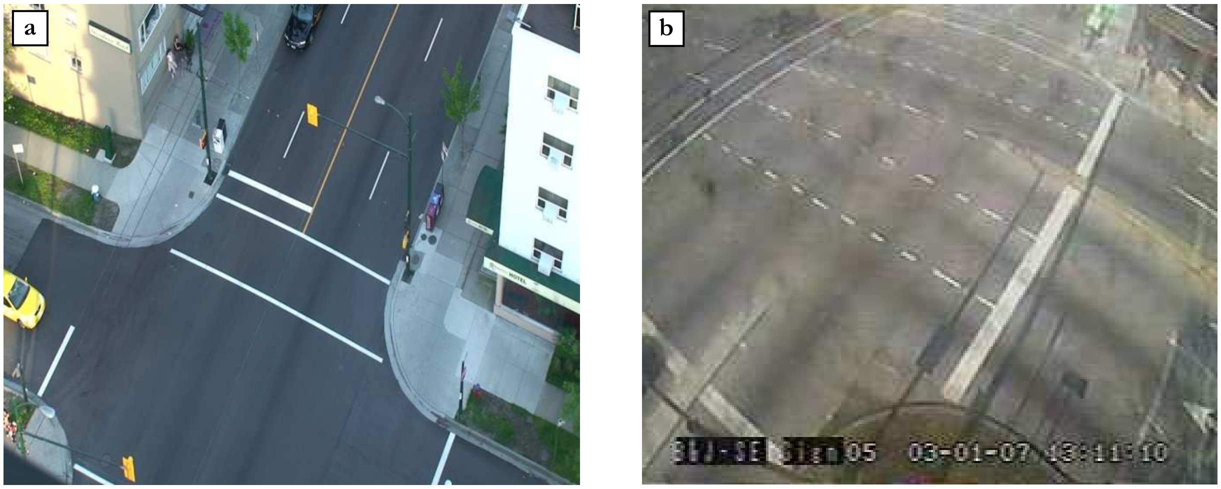


**Figure 2** An illustration of camera calibration issues that arise in urban traffic scenes. Figure **a)** shows a frame taken from video sequence from video sequence BR-1 shot at Vancouver-British Columbia. The estimation of the vanishing point location based on lane marking was unreliable. The obtained camera parameters were initially not sufficient to measure pedestrian walking speed in adequate accuracy. The integration of additional geometric constraints enhanced the estimates of the camera parameters and met the objectives of this application. Figure **b)** shows a sample frame from video sequence K1 of traffic conflicts shot in Kentucky. Significant radial lens distortion is observed at the peripheries of the camera field of view. A reliable estimation of the vanishing point location requires the consideration of line segments that extend to the peripheries of the camera field of view. The curvature of parallel lines was significant in these locations and the estimation of the vanishing point was challenging.

This study describes a robust camera calibration approach for traffic scenes in case of incomplete and noisy calibration data. The cameras used in this study were commercial-grade cameras; most were held temporarily on tripods during the video survey time, others were already installed traffic cameras. A strong focus of this study is on the positional accuracy of road users, especially pedestrians. This was possible by relying on manually annotated calibration data, not vehicle tracks as is the case in automatic camera calibration, e.g. (5).

The uniqueness of this study lies in the composition of the cost function that is minimized by the calibrated camera parameters. The cost function comprises information on various corresponding features in world and image spaces. The diversity of geometric conditions constituted by each feature correspondence enables the accurate estimation of camera parameters. Features are not restricted to point correspondence or parallel lines, but extend to distances , angles between lines, and relative appearance of locally rigid objects. After annotating calibration data, a simultaneous calibration of extrinsic and intrinsic camera parameters is performed, mainly to reduce error propagation (15).

The following sections describe in order, a brief review of previous work, the methodology of camera calibration, and a discussion of four case studies. Video sequences in these case studies were collected from various locations in the Downtown area of Vancouver, British Columbia, Oakland, California, and an unknown location in Kentucky.

## 2. PREVIOUS WORK

There is an emerging interest in the calibration of cameras monitoring traffic scenes, e.g. (16) (17) (18) (19) (5) (15). An important advantage of traffic scenes for this purpose is that they typically contain geometric elements such as poles, lane marking, and curb lines. The appearance of these elements is partially controlled by their geometry, therefore providing conditions on the camera parameters. Common camera calibration approaches draw the calibration conditions from a set of corresponding points, e.g. (10) (20), geometric invariants such as parallel lines (21), or from line correspondences (22).

These approaches however overlook other geometric regularities such as road markings, curb lines, and segments with known length. The use of geometric primitives is becoming more popular, e.g. in recent work (19) and citations therein. However, two main issues can arise in calibrating traffic scenes that cannot be addressed using existing techniques. First, most of the existing techniques construct the calibration error in terms of the discrepancy between observed and projected vanishing points. However, camera locations may be at significantly high altitude or its field of view too limited to reliably observe the convergence of parallel lines to a vanishing point. Finding initial guesses can be also challenging in such settings. Second, a detailed map or up-to-date orthographic image of the traffic scene may be unavailable. In this case, reliance on point correspondences is not possible. The proposed calibration approach draws the calibration information from the real-world lengths of observed line segments, angular constraints, and the dimension invariance of vehicles traversing the camera field of view.

# 3. METHODOLOGY

## 3.1. Camera Model

In this camera calibration approach, the canonical pinhole camera model is adopted to represent the perspective projection of real-world points on the image plane. A projective transform that maps from a point $\mathbf{X}\in R^{\mathrm{n}}$ to a point $\mathbf{Y}\in R^{\mathrm{k}}$ can be defined by a $(k+1)\times(n+1)$ full-rank matrix. In the case of mapping from 3-D Euclidean space to the image plane, $k=2$ and $n=3$. In homogeneous coordinates, the projective transform can be represented by a matrix $\mathbf{T}_{3\mathrm{x}4}$ and a normalization term $\omega$ as follows:

$$\omega\begin{bmatrix}\mathbf{Y}\\1\end{bmatrix}=\mathbf{T}\begin{bmatrix}\mathbf{X}\\1\end{bmatrix} \qquad \dots (1)$$

Similar to the column vectors in Equation 1, **T** is defined up to a scaling factor while containing 11 degrees of freedom. In theory, a total of 11 camera parameters can be recovered: 6 extrinsic and 5 intrinsic. However, 2 intrinsic parameters are primarily considered in the proposed approach. An additional non-linear parameter, radial lens distortion, is calibrated for using as an initial estimate of the calibrated linear camera parameters. The matrix **T** can be decomposed into two matrices such that: $\mathbf{T}=\mathbf{M}\times\mathbf{N}$, where matrix $\mathbf{N}_{4\mathrm{x}4}$ maps from world coordinates to camera coordinates, and matrix $\mathbf{M}_{3\mathrm{x}4}$ maps from camera coordinates to pixel coordinates. Knowledge of extrinsic camera parameters, comprising 3 rotation angles and a translation vector, is sufficient for generating **N**. Matrices **M** and **N** are calculated as follows:

$$\mathbf{N}=\begin{bmatrix}R & t\\0_{1\times3} & 1\end{bmatrix} \qquad \mathbf{M}=\begin{bmatrix}f_y & -f_x\cot\theta & u_o & 0\\0 & \frac{f_y}{\sin\theta} & v_o & 0\\0 & 0 & 1 & 0\end{bmatrix} \qquad \dots (2)$$

where $f_x$ and $f_y$ are respectively referred to as the horizontal and vertical focal length in pixels, $\theta$ is the angle between the horizontal and vertical axes of the image plane, and $(u_o, v_o)$ are the coordinates of the principal point. The principal point is assumed to be at the centre of the image in the video sequence. The second-degree form of the radial lens distortion is represented by the radial lens distortion parameter $k$ as follows:

$$\acute{x}=x*(1+kr^2)\ \&\ \acute{y}=y*(1+kr^2) \qquad \dots (3)$$

where $(x, y)$ are image space coordinates measured in pixels, $(\acute{x}, \acute{y})$ are the image space coordinate corrected for radial lens distortion and $r$ is the uncorrected distance in pixels from the principal point to a point on the image space.

## 3.2. Cost Function

There is no universally recognized cost function for errors in a camera model (19). There are stable formulations developed in the literature, e.g. in (23), for calibration data consisting of point correspondences. It is however more complicated to construct a proper cost function if the calibration error is based on different types of geometric primitives. A proper cost function should satisfy the following conditions:

1. Uniformly represent error terms from different geometric primitives, i.e. consistent weights and units. This is possible if the cost function is constructed in real-world coordinates.
2. Be perspective invariant, i.e. not sensitive to image resolution or camera-object distance.

It is also desirable that a cost function be meaningful in further image analysis steps so that keeping account of error propagation is possible. Satisfying condition one in a linear algebra, and without special mapping, entails some assumption and/or approximation. Following are the set of conditions proposed in this approach to represent a calibrated camera model:

1. Point correspondences. Matching features are points annotated in the image and world spaces. This condition matches the reprojection of points from one space to their positions in a current space. For unit consistency, point positions in world space are compared to the *back-projection* of points from the image space to the world space.
2. Distance constraints. This condition compares the distance between the back-projection of two points to the world space and their true distance measured from an orthographic map or by field measurements.
3. Angular constraints. This condition compares the true angle between the two annotated lines to that calculated from their back-projection to world space. Special cases are angles of 0° in case of parallel lines, e.g. lane markings or vertical objects, and 90° in case of perpendicular lines, e.g. lane marking and stop lines.
4. Equi-distance constraints. This condition compares the real-world length of a line segments observed at different camera depths. This condition preserves the back-projected length of a line segment even if it varies in the image due to perspective.

The following cost function is composed of four components, each representing a condition:

$$f(\mathbf{X}) = \sum_{i\in C, j\in D, k\in A, m\in E} ||\Delta\mathbf{P}_i||_2{}^2 + \left(\Delta s_j\right)^2 + \left(\bar{l}\tan\Delta a_k\right)^2 + (\Delta l_m)^2 \qquad \dots (4)$$

where,

- **X** is a vector of camera parameters,
- $C, D, A,$ and $E$ are respectively the sets of calibration point-difference, distances, and angular constraints, and equi-distnace constraints.
- $||\Delta\mathbf{P}_i||_2$ is the real-world distance between observed and back-projected calibration points in the $i^{th}$ set of point correspondences,
- $\Delta s_j$ is the difference between observed and projected distances in the $j^{th}$ set of distance correspondence,
- $\bar{\mathrm{l}}$ is the average length of the back-projected line segments on the pair of lines that defines the angular constraint,

- $\Delta a_{\mathrm{k}}$ is the difference between annotated and calculated acute angle between the $k^{th}$ back-projected pair of line segments that defines the angular constraint, and
- $\Delta l_m$ is the difference between the real-world length of a line segment calculated at two locations with different depth of view. This can be typically obtained by measuring the distance between two points on a vehicle traversing a traffic intersection.

Point back-projection, i.e. mapping from image space to world space, is performed efficiently using the homography matrix $\mathbf{H}$ that corresponds to a set of camera parameters $\mathbf{X}$. A least square estimation of the homography matrix is conducted using four points selected from $C$, using $\mathbf{X}$. If the non-linear camera distortion parameter is estimated, back-projection using the homography matrix is not accurate. In this case, back-projection is cast as a minimization problem, such that the projection of the estimated world-space position, from world space to image space, achieves a minimum difference from the annotated image position. The initial estimate of this minimization problem is the world-space position of a point using homography. A basic Quasi-Newton non-linear optimization is sufficient for accurate estimation of the world-space position.

The cost function component that represents angular constraints has the useful property of being proportional to the length of the annotated line segments that define the angular constraint. This assigns larger weight to angles more precisely defined using long edges.

The previous cost function describes linear discrepancies between observed and back-projected geometric primitives, all expressed in real-world unit distance. This construction of the cost function clearly meets the previously proposed conditions. It is noteworthy that the construction of the cost function in pixel coordinates, commonly adopted in the literature, is significantly cheaper to compute than the proposed cost function. In the latter case, point projection to image space is a closed-form operation. The proposed camera calibration approach is designed as an accurate one-time operation to support data extraction from video surveys in which computational efficiency is of lesser importance. In addition, the expression of the projection error in pixel coordinates is implicitly biased toward features closer to the camera (represented by more pixels). This may not be desirable in all applications. For example, the case study based on the video sequence K1, shown in Figure 2 b, focuses on events that take place in the furthest intersection approach.

### 3.3. Implementation Details

The intrinsic camera parameters optimized under calibration are focal length, skew angle, and radial lens distortion. The extrinsic parameters are the translation and rotation (six parameters) of the camera coordinate system from the world coordinate system. The selection of these camera parameters yields more accurate results than if optimization is conducted for each element of the transformation matrices $\mathbf{M}$ and $\mathbf{N}$ (Equation 2).

The minimization of the cost function in Equation 3 over the camera parameters is performed using the Nelder-Mead (NM) simplex algorithm. This algorithm was selected over the commonly used Levenberg-Marquardt (LM) which failed in some cases to converge when the initial estimate of the camera parameters was not accurate. When both converged, NM was consistently more computationally expensive. Computational cost is of lesser importance for the applications targeted by this approach.

The initial guesses for the case study described below were obtained using an estimate of the camera position in an orthographic map of the monitored traffic intersections, of the camera height, and of the location of the back-projection of the principal point on the road surface. The estimate for the focal length was found using previous information and assuming away perspective. Obtaining an accurate initial estimate of the focal length and camera height proved difficult and was in most cases far from the calibrated value. A similar issue was encountered for estimating the camera height of video sequences that were not collected by the authors (sequences K1, K2, and OK). The calibrated camera height for K1 and K2 were 11.5 m and 10.9 m respectively, while their initial estimate was 5.5m.

The implementation of this method was conducted in MATLAB (24). A toolbox was developed to annotate the calibration data, find initial estimates, conduct the camera calibration and visualize the calibration results. The following section provides a review of four case studies in which the proposed camera calibration approach provided adequate estimates of camera parameter. The intended applications were carried out successfully using the obtained camera parameters (2) (3).

## 5. CASE STUDIES AND RESULTS

The four case studies analysed using the proposed camera calibration approach are summarized in Table 1. Camera calibration is conducted for video sequences collected from the downtown area of Vancouver, British Columbia (video sequences 1-4 from site BR and sequence PG), Chinatown in Oakland, California (OK), and an unidentified intersection in Kentucky (K1 and K2). When possible, real-world data was extracted from an orthographic image from Google Maps and in-field distance measurements.

### 5.1. Annotation of Calibration Data

Corresponding points are annotated in image and world spaces. The real-world coordinates of points in the image space can be calculated from their position on the world map. The true length of line segments that constitute distance and equi-distance conditions is calculated from the orthographic image. In case of sequences BR-1:4, the real-world length of line segments was collected by in-field measurements (total of 21 in-field measurements using a measuring wheel). This was necessary to obtain camera calibration with accuracy that supports the measurement of pedestrian walking speed (refer to Table 1). Pairs of lines that constitute the angular constraints are annotated in the image space. These lines are parallel lane markings, parallel light poles and road-side signs, and perpendicular road markings. Figure 3 shows the calibration data for sequence BR-2.

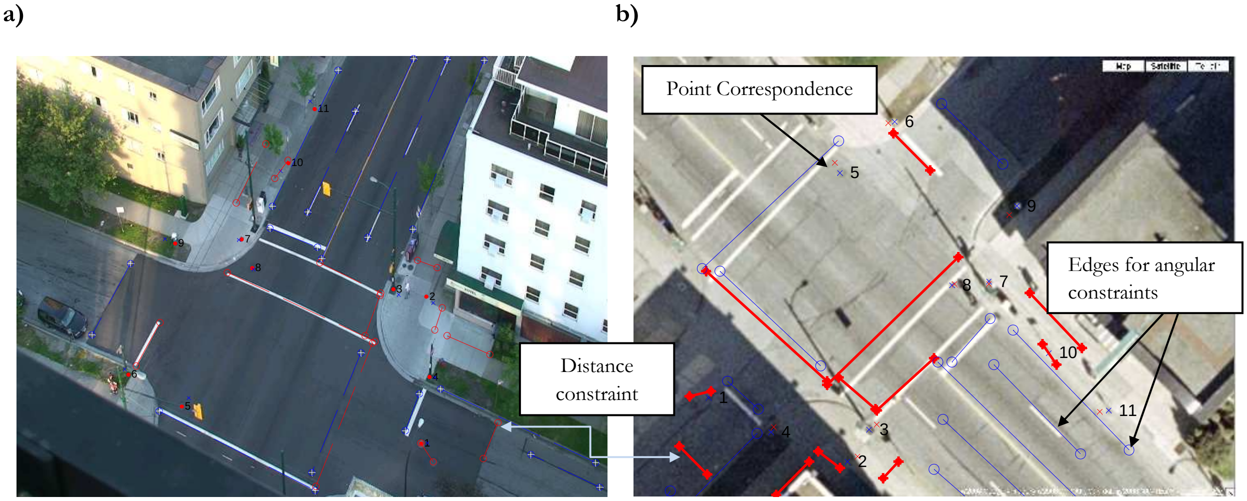


**Figure 3** Calibration data for video sequence BR-2. Point correspondences are annotated with their serial numbers. Points marked with red are calculated and points in blue are annotated. The segments in red define the distance conditions. The segments in blue define pairs of lines for angular conditions. Figure **a)** shows the calibration data (points, and lines) in the image space. Figure **b)** shows the back-projection of the calibration data to world-space.

## 5.2. Effect of Difference Cost Function Components

In order to investigate the effect of using a mix of geometric primitives, the cost function components in Equation 4 were incrementally introduced. The sizes of the different calibration datasets for each scene are shown in Table 1. Root Mean Square Error (RMSE) was calculated by leaving out one testing observation, from sets C and D, each at a time and adding up the error for each testing data point. The total number of iterations required for each scene is the maximum of the number of data points in sets C, D, and A. For example, the number of iterations is 13 for BR-1 and 12 for BR-2. The performance at scenes BR-3 and BR-4 is noteworthy given the limited number of calibration points available at these scenes. Figure 4a shows the reduction in back-projection error for sequences BR-1:4 and PG with the introduction of additional cost function components.

In order to investigate the effect of the equi-distance constraint, the video sequence OK was selected. This sequence has the largest number of calibration data points besides having the challenge of being observed from an unknown camera setting location. Figure 4b shows the back-projection error using different compositions of the cost function. The error was calculated in terms of the difference between the calculated and true lengths of a validation set of 12 line segments. These line segments were not included in the calibration data set.

There is a clear advantage of using calibration data in addition to estimates of point correspondences (four corner points which coordinates estimated based on an assumed lane width of 3.5 m) referred to as case 1 in Figure 4b. There is also an advantage over the use of angular constraints only (case 2) which is analogous to camera calibration based on vanishing point estimation. The addition of all cost function components provides (case 4) however only marginal improvement compared to using point correspondences only (case 3). This likely occurs because of the abundance of accurately localized point correspondences in this video sequence.

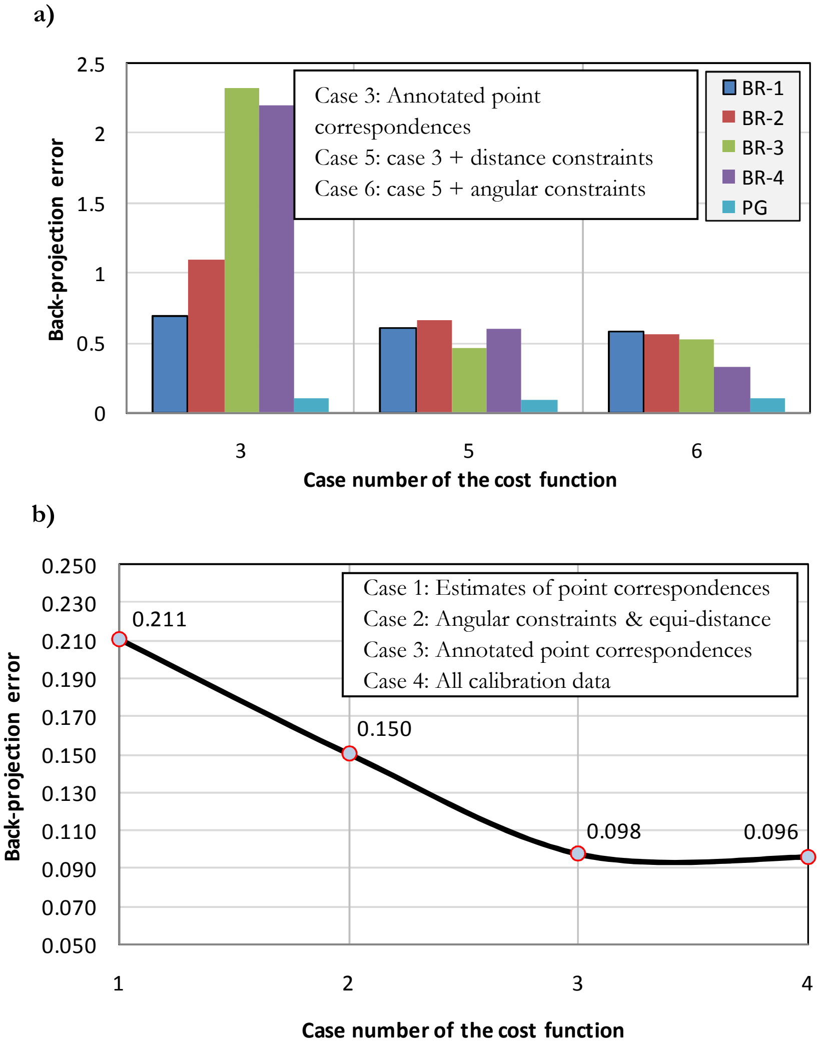


**Figure 4** An illustration of the reduction in camera calibration error due to the inclusion of various cost function components. Figure **a)** shows the RMSE error of test sets BR-1:4 and PG. The RMSE error is calculated based on the error back-projection of all calibration points and distances, each left out one at a time for validation. Figure **b)** shows the back-projection error in terms of the difference between the true and calculated lengths of 12 line segments in sequence OK. The 12 segments were not used in the calibration. The length difference is normalized by the segments length: $Back\ projection\ error = |L_{true} - L_{calculated}|/L_{true}$.

The effect of the addition of cost function components was more evident in sequences K1 and K2. The camera calibration for these sequences was the most challenging. The video sequence, collected from an unidentified site in Kentucky, contains a valuably large number of vehicle-

vehicle traffic conflicts that were analyzed in a different study. The effect of non-linear lens distortion was visible for almost all observed line segments. As shown in Figure 5a, there is a clear advantage of adding all cost function components. The back-projection error was calculated based on the difference in the calculated real-world length of line segments observed from two different cameras for the same site, corresponding to datasets K1 and K2. Figure 5b shows the validation results for camera calibration with complete cost function components (case 5).

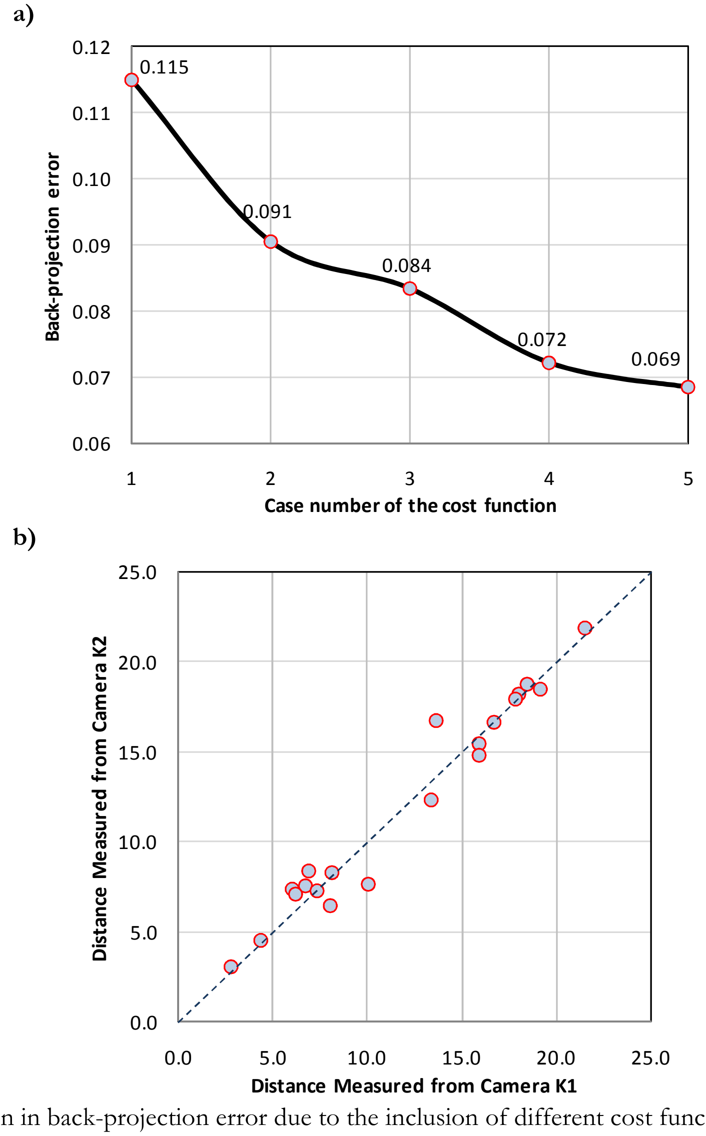


**Figure 5** The reduction in back-projection error due to the inclusion of different cost function components for video sequences K1 and K2. Figure **a)** shows the back-projection error measured as the difference between the real-world lengths of a total of 20 line segments calculated from two camera settings at K1 and K2. The discrepancy in the lengths of the validation line segments were normalized by each line segment length (average 12.57 m). Figure **b)** shows the lengths of the validation line segments for case 5. Refer to Figure 4 for the indication of cases 1:5.

### 5.3. Visualization of Results

In order to visualize the accuracy of the estimated camera calibration parameters, a reference grid is depicted in Figure 6 for sequences BR-2, PG, and OK. The reference grids for sequences K1 and K2 are shown in Figure 7. For sequences K1 and K2, the calibrated radial lens distortion parameter could explain the apparent distortion of the boundaries of the closer sidewalk. The distortion at the further sidewalks could not be completely captured. This demonstrates that additional non-linear parameters are required to capture other types of image distortion evident in this video sequence.

Sample results of applications supported by the estimated camera parameters for these case studies are shown in Figures 8 and 9.

## 6. CONCLUSIONS AND FUTURE WORK

The use of video analysis techniques for transportation applications is on the rise. Camera calibration is necessary for recovering metric information from video sequences. Despite the development of successful methods, current approaches do not address critical issues that arise when monitoring traffic scenes, especially when high camera calibration accuracy is required.

This paper proposed a robust camera calibration approach that overcomes several of these issues. As supported by the reported results, the novel composition of the cost function that defines the error in the camera calibration parameters helps integrating clues from various geometric regularities in traffic scenes.

The formulation of this cost function in a linear algebra entails assumptions regarding the angular constraints. An important extension of this work is the reformulation of the cost function using geometric algebra in which different geometric elements can be uniformly represented. Further improvements to the method are the inclusion of additional non-linear parameters such as the tangential distortion that was evident in video sequences K1 and K2.

## 7. ACKNOWLEDGMENTS

The collection of video sequences at site BR was possible by the support of the following parties: The City of Vancouver and The Empire Landmark Hotel and Conference Centre. The video sequence PG was collected by support from Sam Young and staff of the MMM Group (Vancouver office). The video sequence OK was recorded in a previous study conducted at the UC Berkeley Traffic Safety Center and communicated by Jenna Hua and Prof. David Ragland. Video sequences K1 and K2 were obtained through Zuwahn Kim of California PATH and Ann Stansel of the Kentucky Transportation Cabinet. The authors would like to sincerely thank the previous parties and personnel. We would like to thank Varun Ramakrishna from the Indian Institute of Technology, Chennai for assisting in the annotation of calibration data for K1.

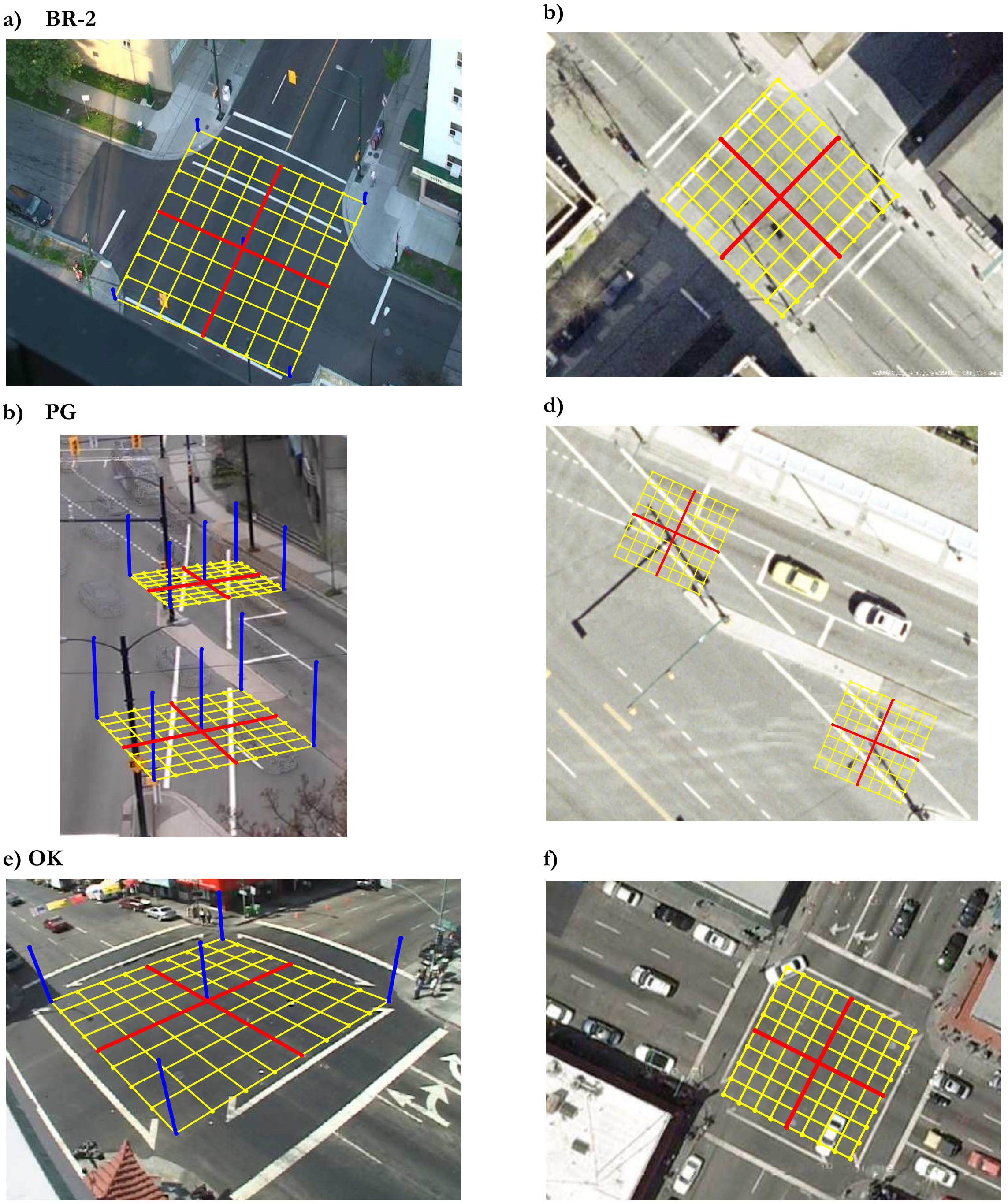
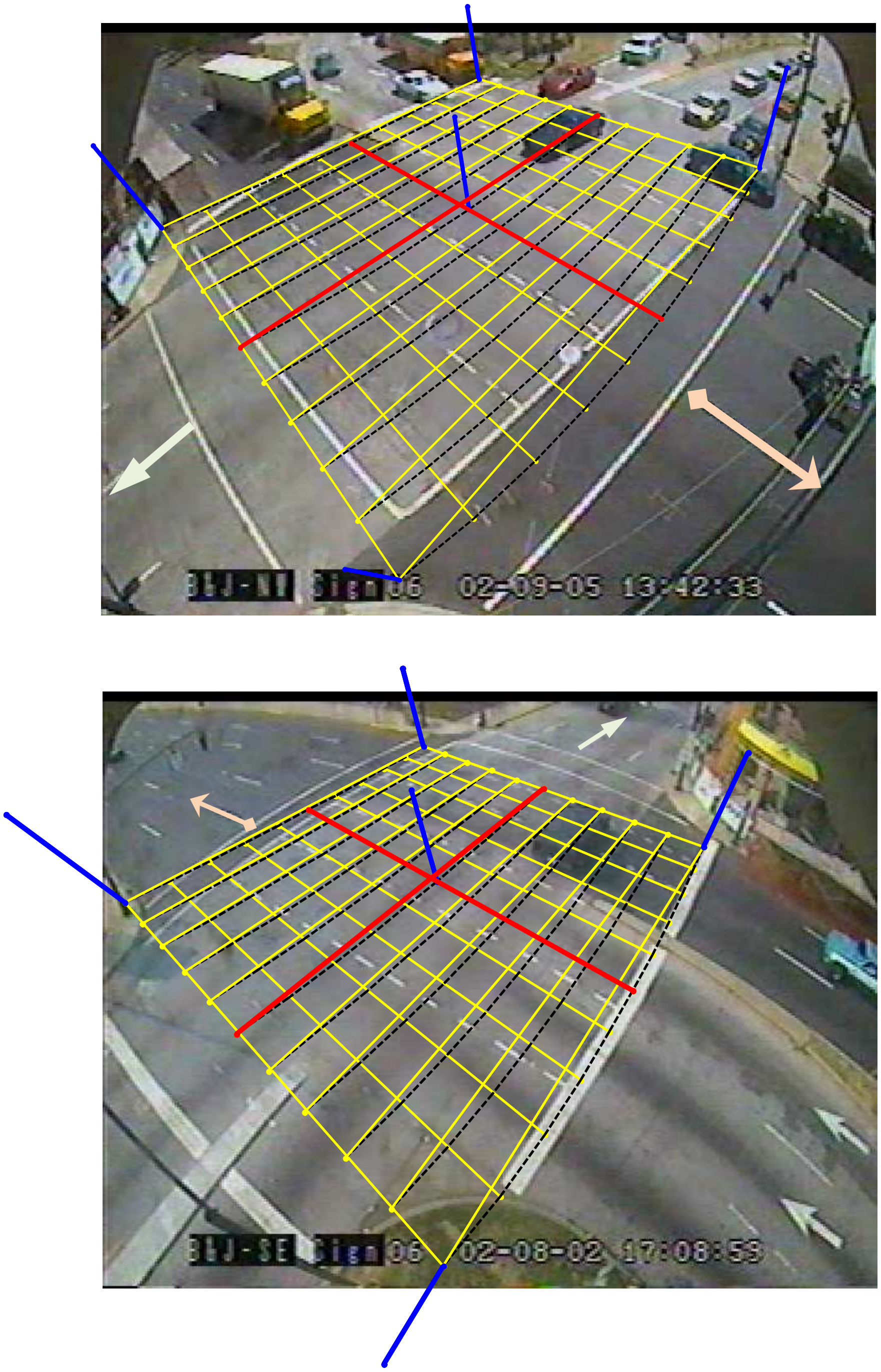


**Figure 6** Refernce grid for video sequences BR-2, PG, and OK, overlaid over frames of the video sequence and orthographic images. The grid spacing is 1 m and the height of the vertical reference lines (depicted in blue) is 4.0 m. Sequences BR-1 and BR-3:4 are recorded at the same site (BR) with different field of views. Their results grid are similar to BR-2 and omitted for space limitation.


a) K1

BLJ-NW Sign 06 02-09-05 13:42:33

b) K2

BLJ-SE Sign 06 02-08-02 17:08:53

**Figure 7** Reference grids for video sequences K1 K2. The non-linear calibration parameters could capture the distortions at the closer sidewalk of sequences K1 and K2. This is evident by comparing the curvature of crosswalk boundaries and the grid side (black dashed). The grid spacing is 2.0 m and the height of the displayed vertical line segment (depicted in blue) is 4.0 m.

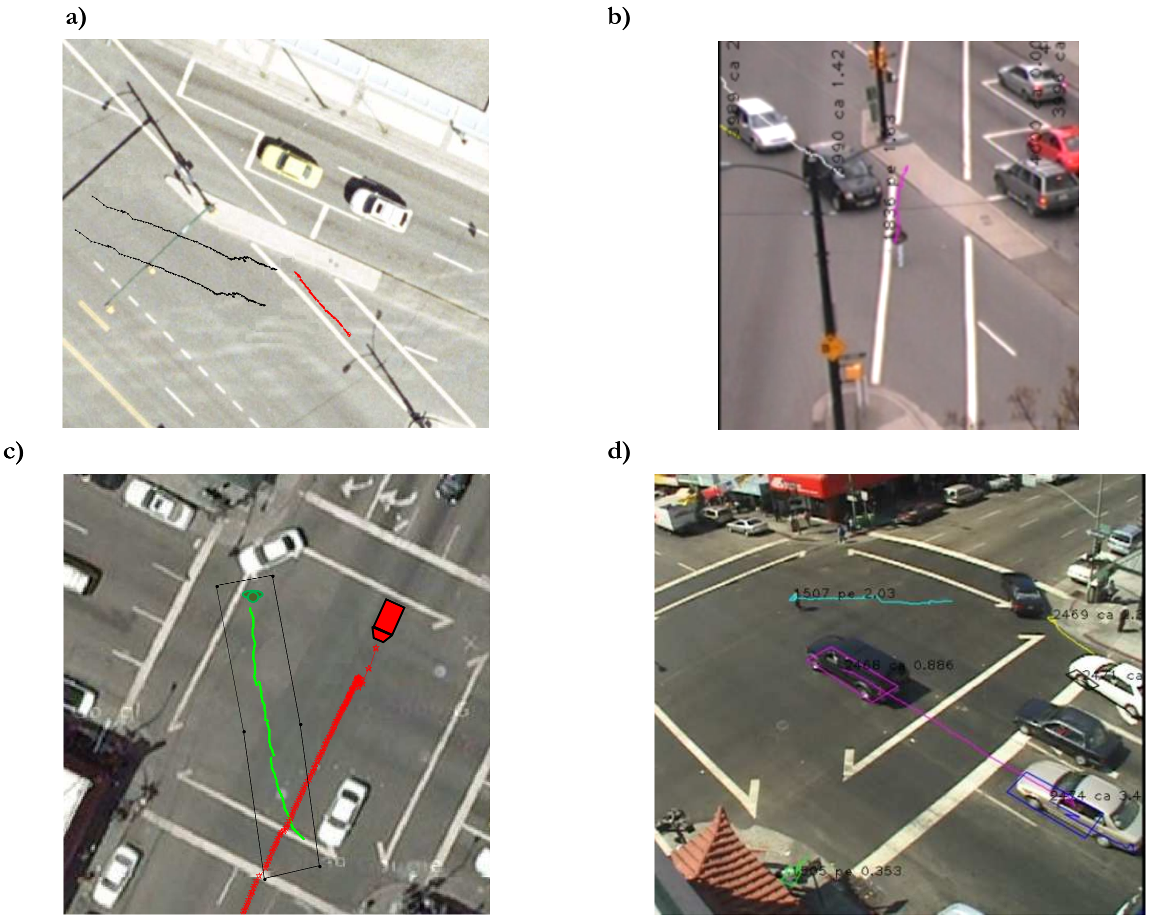


**Figure 8** In this traffic safety application, accurate road user tracks are required to measure their temporal and spatial proximity. Left are the back-projected pedestrian and motorist tracks. Right are the CV-based tracks of the interacting road users. Figures a and b show the world and image space of video sequence PG (the study was reported in (2)). Figures c and d show the world and image space of video sequence OK.

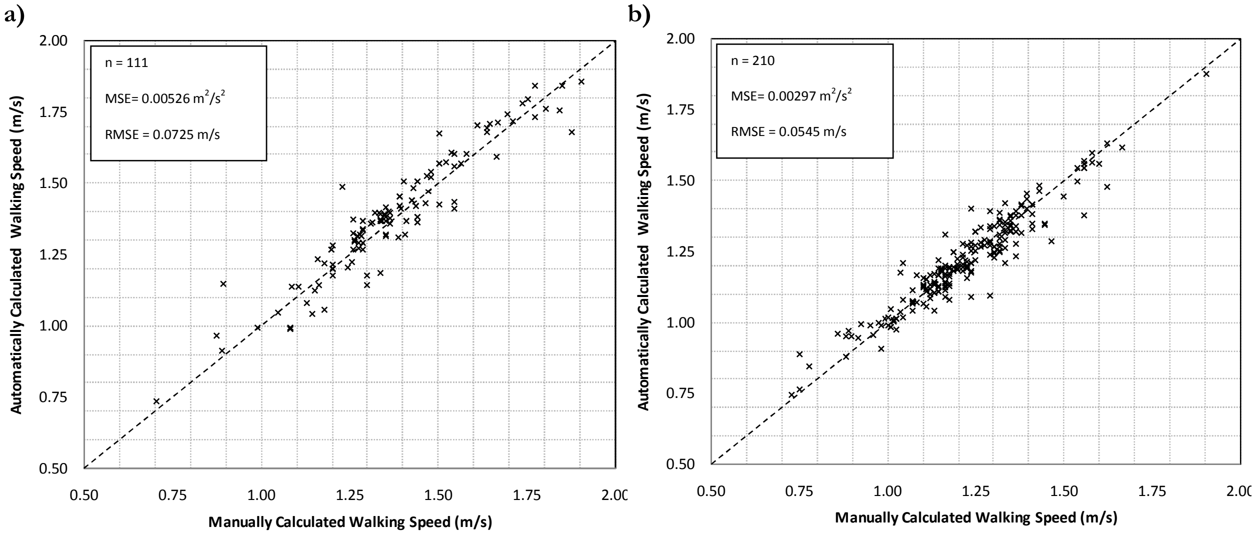


**Figure 9** Validation of walking speed measurements reported in (3). Horizontal axis depicts walking speed based on the time interval required to walk between two check lines. Vertical axis depicts the average walking speed within the same time interval based on automated pedestrian tracking. Figures **a)** and **b)** show the validation of walking speed measurements for two different sets, during day- and night-time respectively.

## 8. REFERENCES

1. *Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis.* **Saunier, N., Sayed, T. and Lim, C.** Seattle : 10th International IEEE Conference on Intelligent Transportation Systems, 2007.

2. *Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data.* **Ismail, K., et al.** Washington, DC : s.n., 2009, Transportation Research Record: Journal of the Transportation Research Board.

3. *Automated Collection Of Pedestrian Data Using Computer Vision Techniques.* **Ismail, K., Sayed, T. and Saunier, N.** Washington, DC : Transportation Research Board Annual Meeting , 2009.

4. *Video-Based Monitoring of Pedestrian Movements at Signalized Intersections.* **Malinovskiy, Yegor, Wu, Yao-Jan and Wang, Yinhai.** 2008.

5. *Automatic Camera Calibration Using Pattern Detection for Vision-Based Speed Sensing.* **Kanhere, N., Birchfield, S. and Sarasua, W.** 2008, Transportation Research Record: Journal of the Transportation Research Board, Vol. 2086, pp. 30-39.

6. *Real-Time Detection and Tracking of Vehicle Base Fronts for Measuring Traffic Counts and Speeds on Highways.* **Kanhere, N., et al.** 2007, Transportation Research Record: Journal of the Transportation Research Board, Vol. 1993, pp. 155-164.

7. *Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction.* **Sturm, P.** 1997. CVPR. p. 1100.

8. *From Projective to Euclidean Space Under any Practical Situation, a Criticism of Self-Calibration.* **Bougnoux, S.** 1989, ICCV, p. 790.

9. *Implicit and Explicit Camera Calibration: Theory and Experiments.* **Ma, G. Wei & S. De.** 5, s.l. : IEEE Trans. Pat. An. Mach. Int., 1994, Vol. 16, pp. 469-480.

10. *A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses.* **Tsai, R.** 1987. IEEE Journal of Robotics and Automation. Vol. 3, pp. 323-344. http://www.cs.cmu.edu/~rgw/TsaiCode.html. 4.

11. *Object pose from 2-D to 3-D point and line correspondences.* **Phong, T., et al.** 3, s.l. : Springer, 2005, IJCV, Vol. 15, pp. 225-243.

12. *Estimating motion and structure from comspondences of line segments between two perspective images.* **Zhang, Z.** 12, June 1994, IEEE Trans. Pat. An. Mach. Int., Vol. 17, pp. 1129-1139.

13. *Digital camera calibration methods: consideration and comparisons.* **Remondino, F. and Fraser, C.** B5, Dresden : s.n., 2006, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36.

14. *Dynamic Camera Calibration of Roadside Traffic Management Cameras for Vehicle Speed Estimation.* **Schoepflin, Todd N. and Dailey, Daniel J.** 2003. Vol. 4.

15. *Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking.* **Yu, Xinguo, et al.** 2009, Computer Vision and Image Understanding, Vol. 113, pp. 643-652. Computer Vision Based Analysis in Sport Environments.

16. *A Simple, Intuitive Camera Calibration Tool for natural Images.* **Worrall, A.D., Sullivan, G.D. and Baker, K.D.** 1994. 5th British Machine Vision Conference. pp. 781-790.

17. *Efficient method for camera calibration in traffic scenes.* **Pengfei, C. Zhaoxue and S.** 6, March 18, 2004, Electronic Letters, Vol. 40.

18. *On Automatic and Dynamic Camera Calibration based on Traffic Visual Surveillance.* **Li, Y., et al.** Istanbul : IEEE, 2007. Intelligent Vehicles Symposium. pp. 358-363.

19. *Using geometric primitives to calibrate traffic scenes.* **Masoud, O. and Papanikolopoulos, N. P.** 6, s.l. : Elsevier, December 2007, Transportation Research Part C, Vol. 15, pp. 361-379.

20. *A flexible new technique for camera calibration.* **Zhang, Z.** 11, 2000, IEEE Trans. Pat. An. Mach. Int., Vol. 22, pp. 1330-1334.

21. *Using vanishing points for camera calibration.* **Caprile, B. and Torre, V.** 2, 1990, International Journal of Computer Vision, Vol. 4, pp. 127-140.

22. *Combining Line and Point Correspondences for Homography Estimation.* **Dubrofsky, E. and Woodham, R.** 2008, Lecture Notes in Computer Science, Vol. 5359, pp. 202-213.

23. *Camera Calibration with Distortion Models and Accuracy Evaluation.* **J.Weng, P. Cohen, and M. Herniou.** 10, 1992, IEEE Trans. Pat. An. Mach. Int., Vol. 14, pp. 965-980.

24. [Online] 2009. http://www.mathworks.com/access/helpdesk/help/techdoc/matlab_product_page.html.