Video Mensuration Using a Stationary Camera

Feng Guo[1] and Rama Chellappa[2]

[1] Computer Science & Center for Automation Research,
University of Maryland, College Park, MD 20742, USA
fguo@cfar.umd.edu
[2] Electrical and Computer Engineering & Center for Automation Research,
University of Maryland, College Park,
MD 20742, USA
rama@cfar.umd.edu

Abstract. This paper presents a method for video mensuration using a single stationary camera. The problem we address is simple, i.e., the mensuration of any arbitrary line segment on the reference plane using multiple frames with minimal calibration. Unlike previous solutions that are based on planar rectification, our approach is based on fitting the image of multiple concentric circles on the plane. Further, the proposed method aims to minimize the error in mensuration. Hence we can calculate the mensuration of the line segments not lying on the reference plane. Using an algorithm for detecting and tracking wheels of an automobile, we have implemented a fully automatic system for wheel base mensuration. The mensuration results are accurate enough that they can be used to determine the vehicle classes. Furthermore, we measure the line segment between any two points on the vehicle and plot them in top and side views.

1 Introduction

Mensuration in image and videos has been studied as an interesting problem with many applications. It has two stages: spatial localization and estimation of dimension. The spatial localization stage estimates an object's position relative to the environment, e.g., the ball's position with respect to the goal post can be used to determine whether a goal has been scored in a soccer game [1]. Since the object's location changes with time, generally multiple concurrent views are needed. Determining the dimension involves the estimation of the distance between two points on the same rigid object [2]. It has received more attention since the results can be used to recognize or identify the object itself. As the two points are invariant relative to the object in the world system, evidences accumulated using multiple frames always improve the mensuration result.

Mensuration requires less calibration information compared to 3D reconstruction problems. A common setup includes one or more parallel reference planes. In stationary surveillance scenarios, the reference plane generally refers to the ground. Minimal calibration [3], defined as the combination of the vanishing line of the reference plane and vertical vanishing point, is assumed to be available.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 164–176, 2006.

The objective is to estimate the ratio of lengths between two parallel line segments in the world system. If one of them has a known length, the length of other can be calculated.

Unlike existing approaches, we consider the lengths of line segments that are *nonparallel* in multiple frames under minimal calibration. Our method can be briefly described in three steps: (1) moving the line segments in parallel to share a common point; (2) fitting ellipses through the other end points under concentric constraints; and (3) calculating the ratio of lengths from the fitting parameters. We then extend the method to measure any line segments that do not lie on the reference plane.

Surveillance applications require the recognition of vehicles as well as human beings. Since different types of vehicles appear similar (e.g., BMW 3 and BMW 5), distinguishing among vehicle classes is very important. We have built a mensuration system for vehicles using the result of a wheel detection and tracking algorithm. Based on the color difference between the black tire and the silver wheel cover, wheels are extracted using an intensity threshold. An algorithm for measuring the wheel base, the distance between two wheel centers on the same side, is implemented and the result is used for mensuration of other parts of the vehicle.

The rest of the paper is organized as follows: Section 2 discusses related work. Section 3 introduces the basic idea of mensuration, the algorithm and error analysis. Section 4 describes the mensuration system for determining the wheel base of vehicles. Section 5 presents real video experiments results.

2 Related Work

Mensuration and related problems have been studied for more than ten years. It is well known that the ratio of two line segments can be recovered when the camera's parameters (intrinsic and extrinsic) are known. To simplify the problem, many assumptions have been made, such as: unit aspect ratio, zero skew and coincidence of principal point and the image center. Caprile and Torre, in their classical work, developed an algorithm to compute the focal length and projection matrix using the properties of vanishing points from a single view [4]. Liebowitz and Zisserman presented a two-step algorithm to rectify the perspective images of planes [5]. The first step estimates the vanishing line by detecting vanishing points, and transforms the image from projective to affine. The second step transforms the image from affine to metric using three constrains. Triggs presented a systematic error analysis [6] for autocalibration from planar scenes. Criminisi, *et al.* considered the estimation of height from an uncalibrated image [3] using projective geometry and cross-ratio, which have become popular in height estimation. In their subsequent work [2], the idea is extended to multiple reference planes. Kim, *et al.* calibrated a camera using a pattern of concentric circles [7]. Chen rectified a planar metric by obtaining the absolute conic from a projected circle and the vanishing line. [8]. Given two sets of perpendicular lines, Wang *et al.* measured the line segments on the reference plane without

rectification [9]. Moons *et al.* presented an algorithm for recovering the 3D affine structure from two perspective views taken by a camera undergoing pure translation [10]. Assuming that the focal lengths are constant, Hartley developed an algorithm based on matrix factorizations for self calibration [11]. In the context of video mensuration, Lv *et al.* proposed an algorithm for calibrating a stationary camera from a video using walking humans [12]. Stauffer *et al.* built a linear model to normalize objects size in a video [13]. Bose and Grimson rectified the plane by tracking objects moving on a straight line with constant speed [14].

Existing mensuration approaches rely on parallel line segments, which is not practical in video. Although sometimes mensuration can be solved by other methods such as rectification, those algorithms estimate global parameters, which may not be optimal to a specific object. The simplest problem of video mensuration can be stated as:

Problem 1. Given a reference length on the reference plane in multiple frames, estimate the length of any line segment on the reference plane (in multiple frames).

This problem cannot be solved using a single frame. It can be proved that at least three reference lengths are needed. First, we present the solution based on three frames and then show how more frames can be used. Combining the probe and reference lengths together, we obtain the ratio between the two lengths by fitting two concentric ellipses; our solution is computationally efficient and optimal. Using the results of wheel centers detection and tracking algorithm, mensuration results are provided for measuring the wheel base of cars.

3 Algorithm

In the rest of the paper, we use upper case letters to indicate points in the world system and the corresponding lower case letters for their images. A line segment with endpoints $\mathbf{P}$ and $\mathbf{Q}$ is denoted as $\mathbf{PQ}$ and its length is denoted as $\|\mathbf{PQ}\|$. The reference plane is labelled as $\mathcal{R}$, its vanishing line as $\mathcal{L}$. The vertical vanishing point is denoted as $\mathbf{z}$.

3.1 Lemmas on Projective Geometry

We present two lemmas:

Lemma 1. *Given four collinear points* $\mathbf{m}$, $\mathbf{n}$, $\mathbf{p}$ *and* $\mathbf{q}$ *in the image plane and the vanishing point* $\mathbf{v}$ *along this direction. Let* $d_{\mathbf{k}}$ *stand for* $\|\mathbf{vk}\|$ *for* $\mathbf{k} = \mathbf{m}$, $\mathbf{n}$, $\mathbf{p}$ *and* $\mathbf{q}$. *The ratio between two line segments* $\|\mathbf{MN}\|$ *and* $\|\mathbf{PQ}\|$ *can be written as:*

$$r = \frac{\|\mathbf{MN}\|}{\|\mathbf{PQ}\|} = \frac{d_{\mathbf{p}}}{d_{\mathbf{m}}} \frac{d_{\mathbf{q}}}{d_{\mathbf{n}}} \frac{(d_{\mathbf{m}} - d_{\mathbf{n}})}{(d_{\mathbf{p}} - d_{\mathbf{q}})} \tag{1}$$

Lemma 1 can be easily proved using the property of cross-ratio.

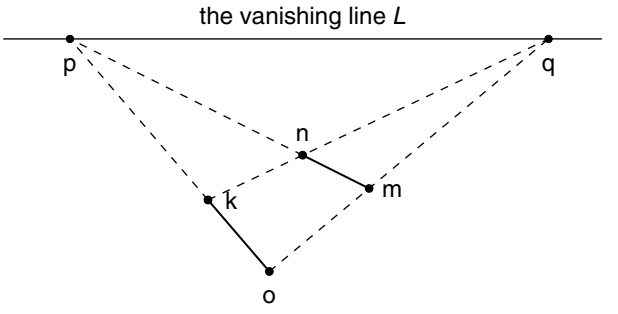


Fig. 1. An illustration of Lemma 2. **mn** is parallel moved to **ok**, where **o** is given and **k** is unknown. **p** and **q** are two points on the vanishing line.

Lemma 2. *Given a line segment* **mn** *and a point* **o**, *let* **mn** *and* **mo** *intersect the vanishing line* $\mathcal{L}$ *at points* **p**, **q** *respectively. Denote the intersecting point of lines* **nq** *and* **op** *as* **k** *(see Fig. 1). In the world system,* **OK** *is parallel to* **MN** *and* $\|\mathbf{OK}\| = \|\mathbf{MN}\|$.

Proof. Because **p** and **q** are vanishing points, **OP** and **MN** are parallel, and **NQ** and **MN** are parallel. **MNKO** forms a parallelogram, so $\|\mathbf{OK}\| = \|\mathbf{MN}\|$ □

k can be represented using the dot product as:

$$\mathbf{k} = \frac{(\mathbf{n} \bullet \mathcal{L})(\mathbf{o} \bullet \mathcal{L})\mathbf{m} - (\mathbf{m} \bullet \mathcal{L})(\mathbf{n} \bullet \mathcal{L})\mathbf{o} - (\mathbf{o} \bullet \mathcal{L})(\mathbf{m} \bullet \mathcal{L})\mathbf{n}}{(\mathbf{n} \bullet \mathcal{L})(\mathbf{o} \bullet \mathcal{L}) - (\mathbf{m} \bullet \mathcal{L})(\mathbf{n} \bullet \mathcal{L}) - (\mathbf{o} \bullet \mathcal{L})(\mathbf{m} \bullet \mathcal{L})} \tag{2}$$

Lemma 2 enables the parallel move of a line segment **MN** in the world system so that **M** maps to a given point **O** and **N** to an unknown point **K**. The image of **K** can be localized in the image plane using the image of other three points.

3.2 Reference Length in Three Frames

We consider a simplified problem

Problem 2. Given the reference length $\|\mathbf{MN}\| = 1$ on the reference plane in three frames as $\mathbf{m}_i\mathbf{n}_i$ $(i = 1, 2, 3)$, estimate the length of any line segment **OP** on the reference plane using one frame.

$\|\mathbf{OP}\|$ can be acquired using the following steps:

1. Localize $\mathbf{k}_i$ on the image plane using the method in Lemma 2, such that $\mathbf{OK}_i$ is a parallel move of $\mathbf{M}_i\mathbf{N}_i$ in the world system.
2. Estimate the point $\mathbf{k}_{i+3}$ on the line $\mathbf{k}_i\mathbf{o}$ so that $\|\mathbf{K}_i\mathbf{O}\| = \|\mathbf{OK}_{i+3}\|$.
3. Fit an ellipse $\mathtt{E}$ passing through all the points $\mathbf{k}_i$ $(i = 1, 2, ..., 6)$. Denote the intersection between line **op** and $\mathtt{E}$ as $\mathbf{k}_0$.
4. Calculate $r = \|\mathbf{OP}\| : \|\mathbf{OK}_0\|$ using the cross-ratio as the mensuration result.

Proof. Using steps 1 and 2, we have $\mathbf{OK}_i = 1$ for $i = 1, 2, ..6$. Ellipse $\mathtt{E}$ is the image of a unit circle centered at point $\mathbf{O}$ in the world system. Thus $\mathbf{OK}_0 = 1$, and r is the length of the probe line segment. □

Assuming that the detected points have the same error distribution, we note that the mensuration result of a longer line segment is likely to be better than for shorter lines, because the estimated error compared to the length of the line segment is relatively small. This answers why the mensuration of horizontal line segments in the image plane is generally better than vertical ones. However, this is not always true. It depends on the angles of the reference lengths.

3.3 Reference and Probe Lengths in Multiple Frames

Now consider the mensuration of a probe length using multiple frames while the reference length also appears in multiple frames. This can be expressed as

Problem 3. Given a reference length $\|\mathbf{MN}\| = 1$ in multiple frames $\mathbf{m}_i\mathbf{n}_i$ $(i = 1, 2, ...f)$ and and a probe line segment $\mathbf{ST}$ also in multiple frames $\mathbf{s}_j\mathbf{t}_j$ $(j = 1, 2, ...g)$, measure $\|\mathbf{ST}\|$.

Arbitrarily select a point on the image plane as $\mathbf{o}$. Parallel move $\mathbf{m}_i\mathbf{n}_i$ to $\mathbf{op}_i$ and $\mathbf{s}_j\mathbf{t}_j$ to $\mathbf{oq}_j$; localize $\mathbf{p}_{i+f}$ on $\mathbf{op}_i$ so that $\|\mathbf{OP}_i\| = \|\mathbf{OP}_{i+f}\|$, same for $\mathbf{q}_{j+g}$.

Consider the relationship between points in the world system: $\|\mathbf{OP}_i\| = 1$ $(i = 1, 2, ...2f)$, $\|\mathbf{OQ}_j\| = r$ $(j = 1, 2, ...2g)$, and all points on the reference plane. Denote $\mathbf{O}$'s coordinate as $(x_\mathbf{o}, y_\mathbf{o})$, then $\mathbf{P}_i$ and $\mathbf{Q}_i$ are on two concentric circles $\mathtt{C}_p: \ (x - x_\mathbf{o})^2 + (y - y_\mathbf{o})^2 - 1 = 0$ and $\mathtt{C}_q: \ (x - x_\mathbf{o})^2 + (y - y_\mathbf{o})^2 - r^2 = 0$ respectively. Define a trivial circular conic as

$$\mathtt{C}_0 = \begin{pmatrix} 1 & 0 & -x_\mathbf{o} \\ 0 & 1 & -y_\mathbf{o} \\ -x_\mathbf{o} & -y_\mathbf{o} & x_\mathbf{o}^2 + y_\mathbf{o}^2 \end{pmatrix} \tag{3}$$

which represents the single point $\mathbf{O}$. We have

$$\mathtt{C}_p = \mathtt{C}_0 - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathtt{C}_q = \mathtt{C}_0 - r^2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

The 3×3 matrix from the world reference plane to the image plane is denoted as $\mathtt{H}$. Then the conic in the image plane is known as $\mathtt{E} = \mathtt{H}^{-\top}\mathtt{C}\mathtt{H}^{-1}$. Define $\mathtt{L} = \mathtt{H}^{-\top} diag(0, 0, 1)\mathtt{H}^{-1}$. The image of concentric circles is of the form

$$\mathtt{E}_p = \mathtt{E}_0 - \mathtt{L} \qquad \mathtt{E}_q = \mathtt{E}_0 - r^2\mathtt{L} \tag{5}$$

Let the vanishing line be denoted as $\mathcal{L} = (\mathcal{L}(1)\ \mathcal{L}(2)\ \mathcal{L}(3))$. $\mathtt{H}^{-1}$ can be written as

$$\mathtt{H}^{-1} = \alpha \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ \mathcal{L}(1) & \mathcal{L}(2) & \mathcal{L}(3) \end{pmatrix} \tag{6}$$

Since $\mathtt{E}_0$ stands for a single point $\mathbf{o} = (x_\mathbf{o}, y_\mathbf{o})$, it can be written as $a_0(x - x_\mathbf{o})^2 + b_0(x - x_\mathbf{o})(y - y_\mathbf{o}) + c_0(y - y_\mathbf{o})^2 = 0$ (we ignore the restriction $b^2 < 4ac$). Ellipse $\mathtt{E}_p$ and $\mathtt{E}_q$ then can be written as

$$a_0(x - x_\mathbf{o})^2 + b_0(x - x_\mathbf{o})(y - y_\mathbf{o}) + c_0(y - y_\mathbf{o})^2 - \alpha \mathtt{vl} = 0 \tag{7}$$

$$a_0(x - x_\mathbf{o})^2 + b_0(x - x_\mathbf{o})(y - y_\mathbf{o}) + c_0(y - y_\mathbf{o})^2 - \beta \mathtt{vl} = 0 \tag{8}$$

where $\beta = \alpha r^2$ and $\mathtt{v} = (x^2 \;\; xy \;\; y^2 \;\; x \;\; y \;\; 1)$

$$\mathtt{l} = (\mathcal{L}(1)^2 \;\; 2\mathcal{L}(1)\mathcal{L}(2) \;\; \mathcal{L}(2)^2 \;\; 2\mathcal{L}(1)\mathcal{L}(3) \;\; 2\mathcal{L}(2)\mathcal{L}(3) \;\; \mathcal{L}(3)^2)^\top$$

Define $\mathtt{e} = (a \;\; b \;\; c \;\; \alpha \;\; \beta)^\top$, a vector with components

$$\mathtt{u}_p = ((x - x_\mathbf{o})^2 \;\; (x - x_\mathbf{o})(y - y_\mathbf{o}) \;\; (y - y_\mathbf{o})^2 \;\; \mathtt{v}(p)\mathtt{l} \;\; 0 \;\;)$$

$$\mathtt{u}_q = ((x - x_\mathbf{o})^2 \;\; (x - x_\mathbf{o})(y - y_\mathbf{o}) \;\; (y - y_\mathbf{o})^2 \;\; 0 \;\; \mathtt{v}(q)\mathtt{l})$$

The coefficient matrix is $\mathtt{M} = (\mathtt{u}_{p1} \;\; \mathtt{u}_{p1} \ldots \mathtt{u}_{pf} \;\; \mathtt{u}_{q1} \ldots \mathtt{u}_{qg})^\top$, and it becomes a linear fitting problem:

$$\mathtt{Me} = 0 \tag{9}$$

The nontrivial least square solution is:

$$\mathtt{e} = argmin\|\mathtt{Me}\| \quad \text{subject to } \|\mathtt{e}\| = 1$$

It can be solved by performing a Singular Value Decomposition (SVD), $\mathtt{M} = \mathtt{USV}^\top$ where $\mathtt{e}$ is the singular vector corresponding to the smallest singular value of $\mathtt{M}$ [15].

After determining α and β, the length of the probe line segment can be calculated directly from

$$r = \sqrt{\beta/\alpha} \tag{10}$$

This method also works for mensuration of multiple line segments by estimating α, β, and γ, etc.

Algorithm in noisy environments. The above algorithm assumes that reference (probe) lengths are equally important during the fitting process. The real case is that the noise from each detected endpoint is independent identically distributed. The perturbation of $\mathbf{k}_i$ due to the noise can be derived using (2), which is decomposed into two parts: the one tangent to the ellipse, t_i, the one normal to the ellipse, n_i. Since t_i does not affect the fitting result, only n_i is considered. The mensuration algorithm can then be modified as follows:

1. Fit the two ellipses using $\mathtt{M}$ as described before.
2. Calculate the perturbation of the reference and probe lengths along with the normal direction of the fitted ellipses.
3. Estimate the total errors of each probe, denoted as n_i
4. Refit the two ellipse using a new matrix $\mathtt{M}'$. Each row of $\mathtt{M}'$ is constructed by the row of $\mathtt{M}$ times a factor $f_i \sim 1/n_i$.

The algorithm actually gives a weight to each point during fitting, which is inversely proportional to the perturbation. The Expectation-Maximization (EM) algorithm may be applied: the E-step is for estimating the errors from the ellipses; and the M-step for fitting the ellipses to minimize the errors. However, experiments have shown that the EM algorithm may not be necessary as one iteration often produces acceptable results.

3.4 Arbitrary Line Segment Mensuration

We briefly discuss the mensuration of line segments not lying on the reference plane, given a line segment on the reference plane as a reference. Since the reference lengths can be freely parallel moved on the reference plane, the problem can be restated as:

Problem 4. Given a reference line segment $\mathbf{m}_i\mathbf{n}_i$, estimate the length of $\mathbf{m}_i\mathbf{p}_i$.

Let $\mathbf{Q}$ be the projective point of $\mathbf{P}$ onto $\mathcal{R}$. From the results of last section, $\mathbf{mq}$ can be measured using $\mathbf{mn}$. The mensuration of $\mathbf{pq}$ can be obtained using Criminisi's method [3]. $\|\mathbf{MP}\|$ can be achieved using the Pythagoras's rule: $\|\mathbf{MP}\| = \sqrt{\|\mathbf{MQ}\|^2 + \|\mathbf{PQ}\|^2}$. This problem reduces to localizing the projective point $\mathbf{q}$ on the image plane.

Simplified Case. We first study a simplified version of the problem by adding a constraint to point $\mathbf{P}$:

Problem 5. Given a reference line segment $\mathbf{mn}$, estimate the length of $\mathbf{mp}$, under the constraint that the plane formed by $\mathbf{M}$,$\mathbf{N}$, and $\mathbf{P}$ is perpendicular to the reference plane $\mathcal{R}$.

Let the lines $\mathbf{mn}$ and $\mathbf{pz}$ interest at $\mathbf{q}$, where $\mathbf{z}$ is the vertical vanishing point. Point $\mathbf{q}$ is the projective point because (1) the line $\mathbf{PQ}$ is perpendicular to $\mathcal{R}$ as it passes through the vertical vanishing point and (2) $\mathbf{Q}$ lies at the intersection of plane $\mathbf{MNP}$ and plane $\mathcal{R}$.

General Case. To solve problem 4 at least two frames are required.

Theorem 1. *Given reference lengths* $\mathbf{m}_i\mathbf{n}_i$ *(*$\|\mathbf{M}_1\mathbf{N}_1\| = \|\mathbf{M}_2\mathbf{N}_2\|$*) and a point* $\mathbf{q}_1$ *on the reference plane* $\mathcal{R}$*, a point* $\mathbf{q}_2$ *can be localized so that* $\|\mathbf{Q}_2\mathbf{M}_2\| = \|\mathbf{Q}_1\mathbf{M}_1\|$ *and* $\angle\mathbf{Q}_2\mathbf{M}_2\mathbf{N}_2 = \mathbf{Q}_1\mathbf{M}_1\mathbf{N}_1$.

Proof. Draw an ellipse $\mathtt{E}_\mathbf{m}$ so that for any point $\mathbf{x}$ on $\mathtt{E}_\mathbf{m}$, $\|\mathbf{XM}_2\| = \|\mathbf{Q}_1\mathbf{M}_1\|$; ellipse $\mathtt{E}_\mathbf{n}$ so that for any point $\mathbf{x}$ on $\mathtt{E}_\mathbf{n}$, $\|\mathbf{XN}_2\| = \|\mathbf{Q}_1\mathbf{N}_1\|$. Denote the intersection point of the two ellipses as $\mathbf{q}_2$. We have $\triangle\mathbf{M}_2\mathbf{N}_2\mathbf{Q}_2 \simeq \triangle\mathbf{M}_1\mathbf{N}_1\mathbf{Q}_1$, thus $\mathbf{q}_2$ satisfies the requirement. □

Theorem 1 actually suggests a rotation operation on the reference plane. We can then obtain $\mathbf{q}$ from two frames by: rotating frame 1 to 1' so that $\mathbf{m}'_1 = \mathbf{m}_2$ and $\mathbf{n}'_1 = \mathbf{n}_2$. The vertical point $\mathbf{z}$ then becomes $\mathbf{z}'$. The intersection point of $\mathbf{z}'\mathbf{p}'_1$ and $\mathbf{zp}_2$ is $\mathbf{q}_2 = \mathbf{q}'_1$ (see Fig. 2), since during the rotation, $\mathbf{q}$ is always on the line $\mathbf{pz}$. A mensuration method from multiple frames can then be described as:

1. Denote **MN**'s midpoint as $\mathbf{K}$, find $\mathbf{K}_i$ in each frame.
2. Properly rotate each frame, so that $\mathbf{k}'_i = \mathbf{k}_0$ and $\mathbf{m}'_i\mathbf{n}'_i$ is parallel to x-axis. Here $\mathbf{k}_0$ is any fixed point.
3. Localize the point $\mathbf{q}_0$ which minimizes the distance to $\mathbf{z}'_i\mathbf{p}'_i$:

$$\mathbf{q}_0 = argmin_{\mathbf{q}}\ dist^2(\mathbf{q}, \mathbf{z}'_i\mathbf{p}'_i).$$

4. Project $\mathbf{q}_0$ to $\mathbf{z}'_i\mathbf{p}'_i$ as $\mathbf{q}'_i$.
5. Estimate $\|\mathbf{M}'_i\mathbf{P}'_i\|$ through $\mathbf{q}'_i$ and combine the result.

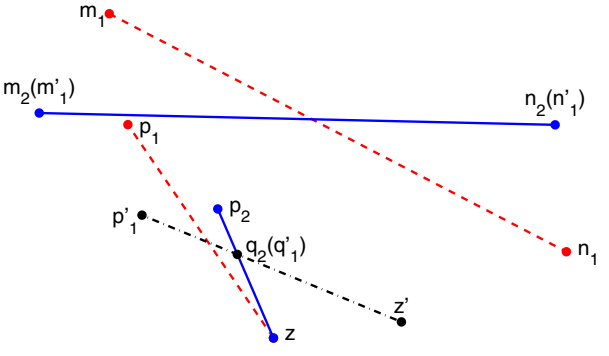


Fig. 2. Illustration of Theorem 1. Red dashed and blue solid indicate frames 1 and 2 separatively. The black dash dot line is the transformed frame 1.

4 System Implementation

In this section we discuss the implementation details for a wheel base mensuration system.

4.1 Wheels Detection and Tracking

After background subtraction [16], we estimate the vehicle's direction of motion using optical flow. We attach a system of skew coordinates to the moving vehicle (x', y'). x' is along the direction of motion, as shown in Fig. 3. Since the tire of a vehicle is always black and the wheel cover is silver or gray, even a simple

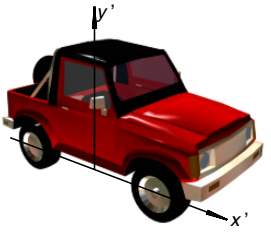


Fig. 3. Skew coordinate using vehicle's moving direction

intensity filter can separate the wheel cover from the tire. We gradually raise the threshold and remove pixels whose intensity is lower than the threshold from the foreground vehicle mask. The whole mask breaks into pieces step by step. The threshold stops increasing when two similar blobs appear at the bottom (with low y'). These two blobs are treated as the mask of wheel covers. The centers of the blobs are used as the detected wheel centers.

The algorithm to track wheels uses the same idea. To speed up the procedure, the initial threshold of the current frame is set as the threshold from the previous frame after subtracting a small value. The detected region focuses on the previous wheel center location and the shift of the vehicle center. After detection, the wheel centers can be tracked in near realtime.

4.2 Minimal Calibration

Without assuming the presence of parallel lines or a moving object along a straight line, the wheel base itself is used to estimate the vanishing line. Our algorithm includes two approaches.

Approach One. We first use a simplified camera model by assuming unit aspect ratio, zero skew, and coincidence of the principal point and the image center. Then the vanishing line is determined by the only unknown intrinsic parameter: the focal length f, and two other extrinsic parameters: the elevation angle θ and the rotation angle ϕ. The transform matrix from the reference plane to the image plane can be written as:

$$\mathtt{H} = \begin{pmatrix} \cos(\phi) & -\sin(\phi)\cos(\theta) & 0 \\ \sin(\phi) & \cos(\phi)\cos(\theta) & 0 \\ 0 & \sin(\theta) & 1 \end{pmatrix} \tag{11}$$

The vanishing line is

$$\mathcal{L} = (-\sin(\phi)\sin(\theta) \quad \cos(\phi)\sin(\theta) \quad -f\cos(\theta)) \tag{12}$$

Since the inverse transform can be written as $\mathtt{H}^{-1} = \mathtt{SAP}$, where $\mathtt{S}$, $\mathtt{A}$ and $\mathtt{P}$ are similarity, affine, and perspective transform matrices respectively, with structure as [5]

$$\mathtt{A} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathtt{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mathcal{L}(1) & \mathcal{L}(2) & \mathcal{L}(3) \end{pmatrix} \tag{13}$$

By comparison, (α, β) can be solved as

$$\alpha = \frac{\cos(\phi)\sin(\phi)\sin(\theta)}{\cos^2(\theta)\cos^2(\phi) + \sin^2(\phi)}, \quad \beta = \frac{\cos(\theta)}{\cos^2(\theta)\cos^2(\phi) + \sin^2(\phi)} \tag{14}$$

If the vanishing line $\mathcal{L}$ is known, ϕ can be obtained from $\tan(\phi) = -\mathcal{L}(1)/\mathcal{L}(2)$. $\mathtt{P}$ can be applied to the end points of the wheel base. Assume that the resulting wheel bases in two different frames are joined by (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}) $(i = 1, 2$

is the frame number). Using the constraint that the two wheel bases are of the same length in the world system, Liebowitz proved the following condition [5]: In 2D complex space with α and β as real and imaginary axes respectively, the point (α, β) lies on a circle with center $(c_\alpha, 0)$ and radius r.

$$c_\alpha = \frac{\delta x_1 \delta y_1 - \delta x_2 \delta y_2}{\delta y_1^2 - \delta y_2^2}, \quad r = \left| \frac{\delta x_2 \delta y_1 - \delta x_1 \delta y_2}{\delta x_1^2 - \delta t_2^2} \right| \tag{15}$$

where $\delta x_i = x_{i1} - x_{i2}$, $\delta y_i = y_{i1} - y_{i2}$. Using (14), θ can be solved by

$$\cos^2\theta = -\frac{\cos^2\phi - r^2 \sin\phi - 2c_\alpha \cos\phi \sin\phi + c_\alpha^2 \sin\phi}{\sin^2\phi - r^2 \cos^2\phi + c_\alpha \cos\phi \sin\phi + c_\alpha^2 \cos\phi}. \tag{16}$$

Since the wheel base appears in multiple frames, each pair of two (if they are not parallel) can be used to solve for $\cos^2\theta$. If the vanishing line is correct, all the estimated results of $\cos^2\theta$ should be same or very close to each other; otherwise, the result should be different from each other. We use the variance to indicate how close the results are. The vanishing line can then be coarsely estimated using the following steps:

1. Select those wheel base pairs, which are unlikely to be parallel in the world system
2. Guess the vanishing line $\mathcal{L}$. Note that in a video sequence, as ϕ normally is close to 0, only the small slope of the vanishing line is considered.
3. Estimate $\cos^2\theta$ from each pair, and calculate its variance $var(\cos^2\theta)$
4. Choose the best vanishing line with minimized $var(\cos^2\theta)$ from all of them. Denote the vanishing line as $\mathcal{L}$

Although step 2 requires a lot of examples, this algorithm is very fast and obtains a good estimate of the vanishing line.

Approach Two. This does not require the simplified camera model assumption. First, we prove that the vanishing line can be obtained from the image of a circle and its center on the reference plane:

Lemma 3. *Given the image of a circle on the reference plane as* $\mathtt{E}$ *with parameter vector* $\mathsf{e} = (a\ b\ c\ d\ e\ f)^\top$ *and the image of the circle center as* $\mathbf{o} = (0, 0)$*, the vanishing line can be written as* $\mathcal{L} = (d\ \ e\ \ 2f)$.

Proof. As explained in the last section, the parameter vector of the ellipse can be written as $\mathsf{e} = \mathsf{e}_0 - \alpha \mathtt{l}$. By comparing the elements in each vector, we have

$$d = 2\alpha\mathcal{L}(1)\mathcal{L}(3), \quad e = 2\alpha\mathcal{L}(2)\mathcal{L}(3), \quad f = \alpha\mathcal{L}(3)^2 \tag{17}$$

which leads to $\mathcal{L} = (d\ \ e\ \ 2f)$. □

Using the above Lemma, we can refine the estimate of the vanishing line from a start point $\mathcal{L}_0$ using the following steps:

1. Let the iteration number $p = 0$
2. Parallel move the wheel base in frames $\mathbf{m}_i\mathbf{n}_i$ using $\mathcal{L}_p$ so that one of their end point shares the common point $\mathbf{o}$ and the other becomes $\mathbf{k}_i$.
3. Fit an ellipse $\mathtt{E}$ though $\mathbf{k}_i$.
4. Estimate the vanishing line, denoted as $\mathcal{L}_{p+1}$. If $\mathcal{L}_{p+1} \simeq \mathcal{L}_p$, output $\mathcal{L}_{p+1}$ and exit; otherwise, $p = p + 1$ and repeat from 2.

5 Experiment Results

The video sequences were captured using cameras located above 25 meters from the ground. Four type of vehicles: 2004 Toyota Camry, 2001 Honda Civic, 2004 Huyndai Elantra and 2004 Ford Explorer were imaged. The frame rate is 20 frame/second and each frame has 480×720 pixels. The sample frames are shown in Fig. 4. Samples of detected wheels are shown in Fig. 4 (e).

The probability of the mensuration result is shown in Fig. 5. From Table 1, although Civic and Elantra are not distinguishable, the Camry vehicle can be separated from them. The Explorer is different from Camry also. Thus the result can be used to determine the vehicles class.

Side View and Top View. By tracking feature points on the object using the KLT [17][18] tracker, the geometric properties of an object can be derived. We track the window corners and rack crossing points of the Ford Explorer and then measure their spacial locations using our method. The window corners are assumed to form a vertical plane passing through the two wheel centers (thus we can use the simplified method) and the rack points have no such constraint. The initial points are manually selected.

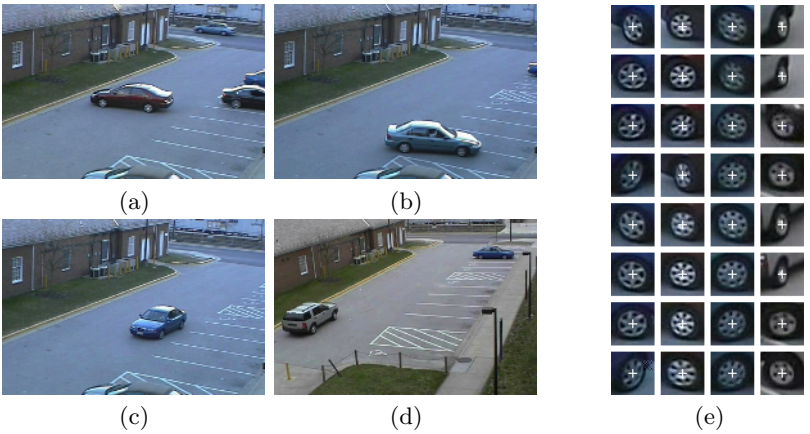
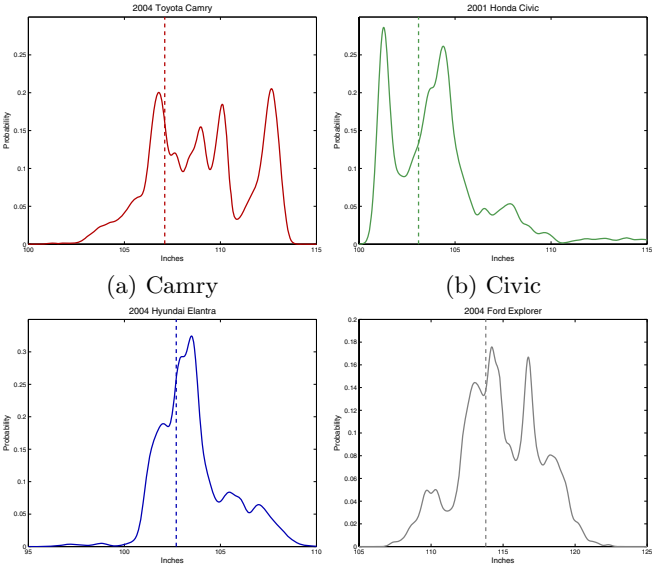


Fig. 4. Sample frames from outdoor sequences. (a) Toyota Camry (b) Honda Civic (c) Hyundai Elantra (d) Ford Explorer and (e) Sample of detected wheels. The detected wheel centers are indicated bye white crosses.

Table 1. Wheel base mensuration results

Year/Make/Model	Size/Category	Ground Truth(In.)	Mensuration result(In.)
2004 Hyundai Elantra	compact sedan	102.7	102.83 ± 1.75
2001 Honda Civic	compact sedan	103.1	103.47 ± 2.77
2004 Toyota Camry	midsize sedan	107.1	108.02 ± 2.45
2004 Ford Explorer	Large SUV	113.8	115.04 ± 2.32



(a) Camry (b) Civic

(c) Elantra (d) Explorer

Fig. 5. Wheelbase mensuration probability plots

(a) side view (b) top view

Fig. 6. Mensuration of the tracked points from reference wheel base

Fig. 6 compares our result with the ground truth images from side and top views. Major errors occur when the tracking points drift away as frames are processed.

6 Summary

We have presented a method for video mensuration. It measures any arbitrary line segments, including those with an angle to the reference plane. A fully

automatic system has been developed for vehicle mensuration. We recover the minimal calibration of the scene by tracking the wheel centers. Further, the line segments joined by any two points on the vehicle is measured using the reference wheel base in single or multiple frames. The wheel base mensuration is very accurate and can be applied to classify the size of vehicles. Other mensuration steps generate side and top views of a vehicle with decent accuracies.

References

1. Reid, I.D., Zisserman, A.: Goal-directed video metrology. ECCV II. (1996) 647–658
2. Criminisi, A., Reid, I.D., Zisserman, A.: Single view metrology. IJCV **40** (2000) 123–148
3. Criminisi, A., Zisserman, A., Gool, L.J.V., Bramble, S.K., Compton, D.: New approach to obtain height measurements from video. Investigation and Forensic Science Technologies **3576** (1999) 227–238
4. Caprile, B., Torre, V.: Using vanishing points for camera calibration. IJCV **4** (1990) 127–140
5. Liebowitz, D., Zisserman, A.: Metric rectification for perspective images of planes. CVPR. (1998) 482–488
6. Triggs, B.: Autocalibration from planar scenes. Lecture Notes in Computer Science **1406** (1998) 89–108
7. Kim, J.S., Gurdjos, P., Kweon, I.S.: Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. IEEE Trans. PAMI. **27** (2005) 637–642
8. Chen, Y., Ip, H.H.S.: Planar metric rectification by algebraically estimating the image of the absolute conic. ICPR (4). (2004) 88–91
9. Wang, G., Wu, Y., Hu, Z.: A novel approach for single view based plane metrology. ICPR (2). (2002) 556–559
10. Moons, T., Gool, L.J.V., Diest, M.V., Pauwels, E.J.: Affine reconstruction from perspective image pairs obtained by a translating camera. Applications of Invariance in Computer Vision. (1993) 297–316
11. Hartley, R.I.: Estimation of relative camera positions for uncalibrated cameras. ECCV. (1992) 579–587
12. Lv, F., Zhao, T., Nevatia, R.: Self-calibration of a camera from video of a walking human. ICPR (1). (2002) 562–567
13. Stauffer, C., e Kinh, Lily, T.: Robust automated planar normalization of tracking data. VS-PETS. (2003)
14. Bose, B., Grimson, E.: Ground plane rectification by tracking moving objects. VS-PETS. (2003)
15. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press (2004)
16. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Trans. PAMI **19** (1997) 780–785
17. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. IJCAI. (1981) 674–679
18. Tomasi, C., Kanade, T.: Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University (1991)