Automatic Camera Calibration for Traffic Understanding

Markéta Dubská[1]
idubska@fit.vutbr.cz

Jakub Sochor[1]
isochor@fit.vutbr.cz

Adam Herout[12]
herout@fit.vutbr.cz

[1] Graph@FIT
Brno University of Technology
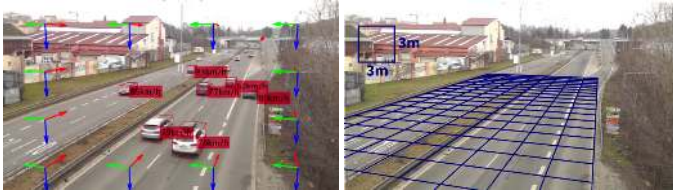Czech Republic

[2] click2stream, Inc.

Figure 1: We automatically determine 3 orthogonal vanishing points, construct vehicle bounding boxes (*left*), and automatically determine the camera scale by knowing the statistics of vehicle dimensions. This allows us to measure dimensions and speed (*right*) and analyze the traffic scene.

This paper proposes a method for fully automatic calibration of traffic surveillance cameras. Our method allows for calibration of the camera – including scale – without any user input, only from several minutes of input surveillance video. The targeted applications include speed measurement, measurement of vehicle dimensions, vehicle classification, etc.

The first step of our approach is camera calibration by determining three vanishing points defining the stream of vehicles (Fig. 2, [3]). The second step is construction of 3D bounding boxes of vehicles (Fig. 3) and their measurement up to scale. In the third step, we use the dimensions of the 3D bounding boxes for calibration of the scene scale (Fig. 4).
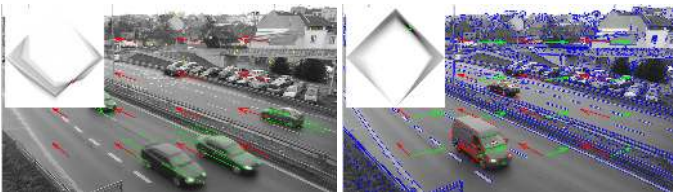

Figure 2: (*left*) Tracked points used for estimation of the 1st VP. Points exhibiting a significant movement (green) are accumulated. (*right*) Accumulation of the 2nd vanishing point. Only edges excluding the vertical ones and those with their direction towards the first VP (green) are accumulated to the diamond space.

Our method for VP detection uses Hough transform based on parallel coordinates [2], which maps the projective plane into a finite space referred to as the *diamond space* by a piecewise linear mapping of lines.
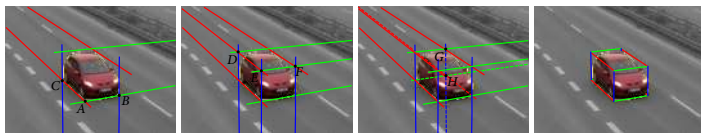

Figure 3: Construction of vehicle's 3D bounding box. From left to right: tangent lines and their relevant intersections A, B, C; derived lines and their intersections E, D, F; derived lines and intersection H; constructed bounding box.

The next step of our approach is construction of 3D bounding boxes of the observed vehicles (Fig. 3). We assume that the vehicle silhouettes can be extracted by background modeling and foreground detection and that the vehicles of interest are moving from/towards the first vanishing point. The 3D bounding box is constructed using tangent lines from vanishing points to the blob's boundary.

Having the bounding box projection, it is directly possible to calculate the 3D bounding box dimensions (and position in the scene) up to precise scale. By fitting the statistics of known dimensions and the measured data from the traffic, we obtain the scale of the scene (Fig. 4).

Camera orientation together with a know distance enables for measuring of vehicle speed/size or distances in the scene. We measured several
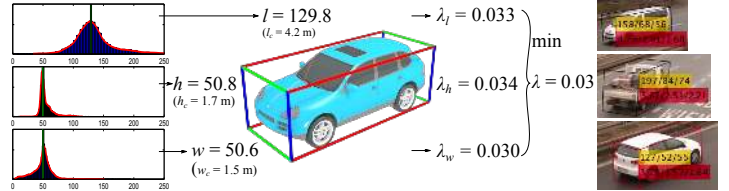

Figure 4: Calculation of scene scale. (*left*) Median (green bar) for each dimension is found in the measured data. (*middle*) Scales are derived separately based on known median car size and the final scale is derived as the minimum from these three scales. (*right*) Examples of relative size of the vehicles (yellow) and real dimensions in meters after scaling.

distances on the road plane and evaluated the error in measurements by our approach. Similar evaluation was provided by Zhang [5], who report average error of measurement "less then 10%". Our average error is 1.9% with worst case 5.6%, (Tab. 1).

dist	1.5 m	3 m	3.5 m	5.3 m	6 m	all
#	85	32	15	16	15	163
mean (%)	1.8	1.7	2.0	2.8	1.5	**1.9**
worst (%)	3.6	3.9	5.5	5.6	3.3	**5.6**

Table 1: Percentage error of absolute distance measurements. The error is evaluated as $|l_m - l_{gt}|/l_{gt} * 100\%$, where l_{gt} is the ground truth value and l_m is the distance measured by the presented algorithm. For each distance we evaluate the average and worst error. The numbers in the row labeled '#' are the number of measurements of the given length (from 5 videos).

When measuring the vehicle speed (Tab. 2), we take into account one corner of the bounding box which lies directly on the road). Vehicles in the video are tracked and their velocity is evaluated over the whole straight part of the track. The average speed of the vehicles was $75 \frac{km}{h}$ and therefore 2% error causes $\pm 1.5 \frac{km}{h}$ deviation. A similar evaluation was provided by Dailey [1] who used distribution of car lengths for scale calculation and reached average deviation $6.4 \frac{km}{h}$ or by Grammatikopoulos [4] whose algorithm has accuracy $\pm 3 \frac{km}{h}$ but requires manual distance measurements to obtain the scale.

	a (5)	b (3)	c (5)	d (5)	e (5)	f (5)	all (28)
mean (%)	2.39	2.90	1.49	1.65	1.31	2.58	**1.99**
worst (%)	3.47	3.63	3.18	3.77	2.40	4.26	**4.26**

Table 2: Percentage error in speed measurement. For obtaining the ground truth values, we drove cars with cruise control and get the speed from GPS. The error is evaluated as $|s_m - s_{gt}|/s_{gt} * 100\%$, where s_{gt} is speed from GPS and s_m is speed calculated by presented algorithm. The number in parentheses stands for the number of evaluated measurements for given video.

[1] D.J. Dailey, F.W. Cathey, and S. Pumrin. An algorithm to estimate mean traffic speed using uncalibrated cameras. *IEEE T-ITS*, 2000.

[2] M. Dubská and A. Herout. Real projective plane mapping for detection of orthogonal vanishing points. In *BMVC*, 2013.

[3] M. Dubská, A. Herout, J. Sochor, and R. Juránek. Fully automatic roadside camera calib. for traffic surv. *To appear in IEEE T-ITS*, 2014.

[4] L. Grammatikopoulos, G. Karras, and E. Petsa. Autom. estimation of vehicle speed from uncalibrated video seq. In *ISMTEPPGRF*, 2005.

[5] Z. Zhang, T. Tan, K. Huang, and Y. Wang. Practical camera calib. from moving objects for traffic scene surveill. *IEEE T-CSVT*, 2013.