Limitations of Vehicle Length Estimation Using Stereoscopic Video Analysis

Tomasz Grajek, Robert Ratajczak, Krzysztof Wegner, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology
ul. Polanka 3, 60-965 Poznań
e-mail: {tgrajek,rratajczak,kwegner}@multimedia.edu.pl

***Abstract*—In the paper vehicle length estimation based on stereoscopic video analysis is considered. The accuracy of the parameters of the stereoscopic video acquisition system influence accuracy of the estimated vehicle length. The respective analysis is provided in the paper. The main result is the formula for estimating the accuracy of the vehicle length measurement.**

Keywords—video; surveillance; vehicle dimension; car detection; stereo

I. INTRODUCTION

At present, since the traffic intensity is constantly increasing, both in the cities, as well as in their outskirts, the issue of automatic traffic analysis has become a very important problem. The quality and accuracy of collected data is an equally important issue. In this paper, we focus on possible to achieve accuracy of vehicle length estimation. Knowledge about vehicle dimensions is important for such applications as:

- automatic traffic analysis,
- distinguishing and categorizing the type of vehicle,
- detection of oversize vehicle violating restrictions,
- automatic vehicle driving by detection of the size of other road users,
- automatic warning system for drivers of large vehicles, approaching narrow places or tight curves.

Most of the available methods of 3D scene geometry estimation base on various types of radiation [1]. Other methods use single camera image [2]. Unfortunately these approaches require and assume partial knowledge of the geometry of the observed 3D scene. This paper focuses on vehicle length estimation with the use of stereoscopic video sequences.

The following sections present the methodology of vehicle length estimation, the accuracy of the proposed estimation method and the experimental results including comparison with theoretical considerations.

II. VEHICLE LENGTH ESTIMATION

In our system, a pair of parallel cameras with baseline b records the city road scene with moving vehicles (Fig. 1). Camera system is well-calibrated, synchronized and all necessary camera parameters are known. Any deviation form parallelism of the cameras is removed by image rectification [3]. Both acquired images are analyzed and all the vehicles are detected. The length is estimated, based on the detected location of the vehicle in the captured images and camera system parameters.

The following convention for math notation will be maintained further in the paper. Subscript 1 means that the given symbol relates to the first camera and subscript 2 refers to second camera. Bold symbols like $\boldsymbol{M}$ describe vectors or matrixes. Capital letters X, Y, Z express Euclidean coordinates in 3D space while u, v are reserved for coordinates in the images. $\boldsymbol{M}^{-1}$ means inversion of the square matrix $\boldsymbol{M}$ or pseudoinversion of non-square matrices so that $\boldsymbol{M}^{-1} \cdot \boldsymbol{M} = \boldsymbol{I}$, where $\boldsymbol{I}$ is identity matrix.

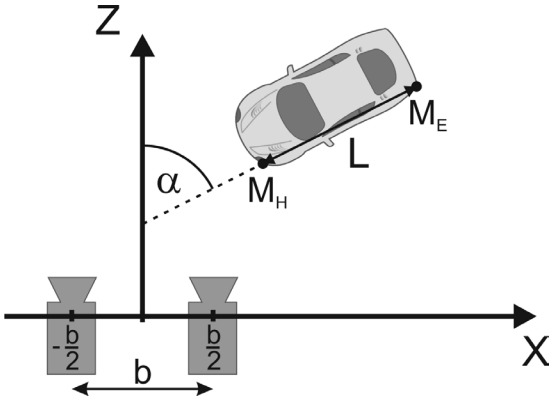

Fig. 1. Experimental setup of the system in order to assess accuracy of the system proposed

In order to estimate vehicle length we have to define two characteristic points of the vehicle: the head $\boldsymbol{M_H}$, and the end $\boldsymbol{M_E}$ (Fig. 1). Position of each of those two points in 3D space is given by [1x4] vector in homogenous coordinates (1).

$$\boldsymbol{M_H} = \begin{bmatrix} X_H \\ Y_H \\ Z_H \\ 1 \end{bmatrix} \quad \boldsymbol{M_E} = \begin{bmatrix} X_E \\ Y_E \\ Z_E \\ 1 \end{bmatrix} \tag{1}$$

Then length L of the vehicle is simply Euclidean distance between those two points (2).

$$L = \|\boldsymbol{M_H} - \boldsymbol{M_E}\| \tag{2}$$

Of course from the captured stereo pair we know only projective positions of the head $(\boldsymbol{m_{H1}}, \boldsymbol{m_{H2}})$ and the end $(\boldsymbol{m_{E1}}, \boldsymbol{m_{E2}})$ of the vehicle in the first and the second image plane. This 2D positions can be expressed as [1x3] vectors in homogenous coordinates (3).

$$\begin{aligned} \boldsymbol{m_{H1}} &= \begin{bmatrix} u_{H1} \\ v_{H1} \\ 1 \end{bmatrix} \quad \boldsymbol{m_{H2}} = \begin{bmatrix} u_{H2} \\ v_{H2} \\ 1 \end{bmatrix} \\ \boldsymbol{m_{E1}} &= \begin{bmatrix} u_{E1} \\ v_{E1} \\ 1 \end{bmatrix} \quad \boldsymbol{m_{E2}} = \begin{bmatrix} u_{E2} \\ v_{E2} \\ 1 \end{bmatrix} \end{aligned} \tag{3}$$

Assuming pinhole camera model [4] the process of projection of 3D points onto the image plane can be modeled as (4) for a given point $\boldsymbol{M}$ in 3D space, and its 2D projection position $\boldsymbol{m}$ in the image.

$$z\boldsymbol{m} = \boldsymbol{P} \cdot \boldsymbol{M} \tag{4}$$

Scalar z is called the depth of the point and expresses the distance from the camera to the point $\boldsymbol{M}$. $\boldsymbol{P}$ is the projection matrix given by (5) where $\boldsymbol{A}$ is a [3x3] matrix of intrinsic parameters, $\boldsymbol{R}$ is a [3x3] rotation matrix, $\boldsymbol{T}$ is a [1x3] translation vector.

$$\boldsymbol{P} = \boldsymbol{A} \cdot [\boldsymbol{R}|\boldsymbol{T}] \tag{5}$$

Parameters $\boldsymbol{A}, \boldsymbol{R}, \boldsymbol{T}$ of the cameras can be estimated during system calibration and detailed description can be found in [4].

Inverting (4) and putting it into (2) results in obtaining vehicle length estimation from positions of the head and the end of the vehicle in the first camera image (6).

$$\begin{aligned} L &= \|\boldsymbol{M_H} - \boldsymbol{M_E}\| = \\ &= \left\|\boldsymbol{P_1}^{-1} \cdot (z_H \boldsymbol{m_{H1}}) - \boldsymbol{P_1}^{-1} \cdot (z_E \boldsymbol{m_{E1}})\right\| = \\ &= \left\|\boldsymbol{P_1}^{-1} \cdot (z_H \boldsymbol{m_{H1}} - z_E \boldsymbol{m_{E1}})\right\| \end{aligned} \tag{6}$$

Of course similar consideration can be done for the second camera. Depth of the head z_H and the end z_E of the vehicle can be derived using epipolar geometry [4], which for well calibrated stereo pair is simplified to trigonometric relation of the projection points position in captured stereo pair (7):

$$z = f\frac{b}{d} \tag{7}$$

where f is focal length of the camera, b is the baseline of the stereo pair, and the d is disparity i.e. difference in projective position of the same point in stereo pair (8):

$$d = |u_1 - u_2| \tag{8}$$

Disparity can be estimated using various techniques. Extensive review of such techniques can be found in [5].

Having (7), (6) can be rewritten to (9) which can be directly used for vehicle length estimation.

$$\begin{aligned} L &= \left\|\boldsymbol{P_1}^{-1} \cdot (z_H \boldsymbol{m_{H1}} - z_E \boldsymbol{m_{E1}})\right\| = \\ &= \left\|\boldsymbol{P_1}^{-1} \cdot \left(fb\frac{1}{d_H}\boldsymbol{m_{H1}} - fb\frac{1}{d_E}\boldsymbol{m_{E1}}\right)\right\| = \\ &= fb\left\|\boldsymbol{P_1}^{-1} \cdot \left(\frac{1}{d_H}\boldsymbol{m_{H1}} - \frac{1}{d_E}\boldsymbol{m_{E1}}\right)\right\| \end{aligned} \tag{9}$$

III. INACCURACY SOURCES

Although (9) allows quite precise vehicle length estimation, the accuracy is limited. Term precision means how consecutive results differ from each other, while the term accuracy means how calculated value differs from its true value (often true value is unknown).

Accuracy of the estimated vehicle length is limited by several factors:

- Accuracy of the stereo pair calibration.
  - Accuracy of the intrinsic and the extrinsic camera parameters estimation.

    Focal length and principle point localization have always limited accuracy. Those values are commonly estimated using some camera calibration procedure, while theirs true values remain unknown. The use of distorted value introduces errors in back-projected position of the 3D points of the vehicle, resulting in length estimation inaccuracy.
  - Accuracy of the rectification.

    Rectification is numerical procedure which transforms one image into another but always in digital form (pixel accuracy). This way rounding of the points position during transformation results in errors in projective position of the points in the images. Moreover, rectification is based on camera parameters of a limited accuracy, which also introduces errors.
- Accuracy of the system.
  - Digital nature of the image.

    The accuracy of positions of points in digital images is always limited. The position of the given point is usually represented with up to 1 pixel accuracy. However, some algorithms can estimate point's position with higher, but still finite accuracy. The same is for disparity, which can be estimated with up to 2 pixels accuracy (1 pixel accuracy for both the first and the second image pixel position).
  - Accuracy of the algorithms.

    Algorithms used for localizing the head and the end of the vehicle in the images have also limited accuracy. Of course more complex methods give better results but errors are inevitable. The same refers to the disparity estimation algorithms.

IV. ACCURACY OF THE VEHICLE LENGTH

In order to estimate accuracy of the vehicle length estimation (10) we use the total differential method on (9):

$$\begin{aligned} \mathrm{DL} &= \left|\frac{\partial \mathrm{L}}{\partial \mathrm{f}}\right| \Delta \mathrm{f} + \left|\frac{\partial \mathrm{L}}{\partial \mathrm{b}}\right| \Delta \mathrm{b} + \\ &+ \left|\frac{\partial \mathrm{L}}{\partial \mathrm{u_{H1}}}\right| \Delta \mathrm{u_{H1}} + \left|\frac{\partial \mathrm{L}}{\partial \mathrm{v_{H1}}}\right| \Delta \mathrm{v_{H1}} + \left|\frac{\partial \mathrm{L}}{\partial \mathrm{d_H}}\right| \Delta \mathrm{d_H} + \\ &+ \left|\frac{\partial \mathrm{L}}{\partial \mathrm{u_{E1}}}\right| \Delta \mathrm{u_{E1}} + \left|\frac{\partial \mathrm{L}}{\partial \mathrm{v_{E1}}}\right| \Delta \mathrm{v_{E1}} + + \left|\frac{\partial \mathrm{L}}{\partial \mathrm{d_E}}\right| \Delta \mathrm{d}_E \end{aligned} \tag{10}$$

Thus, accuracy of the vehicle length estimation can be considered as a function of accuracies of the values used to

estimate the length. Moreover, actual accuracy of the particular vehicle length is also the function of the value used to its estimation (11).

$$\mathrm{DL} = \mathrm{g}(\mathrm{f}, \mathrm{b}, u_{\mathrm{H1}}, v_{\mathrm{H1}}, \mathrm{d_H}, u_{\mathrm{E1}}, v_{\mathrm{E2}}, \mathrm{d_E}) \tag{11}$$

Having the values necessary to estimate the vehicle length and theirs accuracies we can estimate vehicle length accuracy. However, (11) does not tell explicitly how the accuracy is related with the actual vehicle location in 3D space. The answer to that question would require estimated accuracy expressed as a function of 3D vehicle position and stereo pair parameters. From (3) we can define projective position of the head and the end of the vehicle as a function f_u, f_v of its 3D location (12):

$$\begin{aligned} u_{H1} &= f_u(X_H, Y_H, Z_H, f, b) \\ v_{H1} &= f_v(X_H, Y_H, Z_H, f, b) \\ u_{E1} &= f_u(X_E, Y_E, Z_E, f, b) \\ v_{E1} &= f_v(X_E, Y_E, Z_E, f, b) \end{aligned} \tag{12}$$

Moreover using (3) and (8) we can express disparity as a function f_d of 3D location and stereo pair parameters (13).

$$\begin{aligned} d_H &= f_d(X_H, Y_H, Z_H, f, b) \\ d_E &= f_d(X_E, Y_E, Z_E, f, b) \end{aligned} \tag{13}$$

Inserting (13) and (12) into (11) we can explicitly calculate accuracy based on the true 3D vehicle location and stereo pair parameters. Knowing true length L and the angle of the vehicle to the camera α (see Fig. 1) we can define head and the end of the vehicle as (14).

$$\mathrm{M_H} = \begin{bmatrix} \mathrm{X} - L \cdot Sin(\alpha)/2 \\ \mathrm{Y} \\ \mathrm{Z} - L \cdot Cos(\alpha)/2 \\ 1 \end{bmatrix} \quad \mathrm{M_E} = \begin{bmatrix} \mathrm{X} + L \cdot Sin(\alpha)/2 \\ \mathrm{Y} \\ \mathrm{Z} + L \cdot Cos(\alpha)/2 \\ 1 \end{bmatrix} \tag{14}$$

Combining all together results in explicit definition of the accuracy of the estimated length of the vehicle basing on his true 3D location, length and stereo pair parameters.

$$\mathrm{DL} = \mathrm{g}(\mathrm{f}, \mathrm{b}, u_{\mathrm{H1}}, v_{\mathrm{H1}}, \mathrm{d_H}, u_{\mathrm{E1}}, v_{\mathrm{E2}}, \mathrm{d_E}) = h(f, b, X, Y, Z, L, \alpha) \tag{15}$$

Explicit form of (15) is very complex but once obtained can be easily used to calculate accuracy of the measured vehicle length.

V. RESULTS

Taking (15) as a base we can evaluate the distribution of accuracy of the estimated vehicle length. For this evaluation, if not stated explicitly, we use accuracies and camera system parameters given in Table I (values taken from [6]). One of the possible approaches is to estimate accuracy with respect to distance of a vehicle from the camera system (Z) and the stereopair baseline (b) (Fig. 2). As we can see, accuracy drops along with the increase of vehicle distance from the camera, and decreases with the increase of baseline. Thus, to some extent, wider baseline assure more accurate results. Of course not always wide baseline is possible due to urban environment, but in general as wide baseline as possible should be used.

Moreover accuracy of the estimated length changes along with displacement of the vehicle in perpendicular direction to camera axis (Fig. 3). The most accurate results can be obtained when vehicle is directly at the front of the camera system. In other words, when vehicles are in the middle of the captured images, results have the highest accuracy and it decreases as the vehicle moves to the sides of the images.

Accuracy of the estimated length strongly depends on angle of the vehicle at the front of the camera system (Fig. 4). If the vehicle is perpendicular to the camera system (α=0°) the accuracy is the worst (vehicle is moving towards the camera or directly opposite). The best condition is when vehicle are moving from one side of the images to the other.

TABLE I. ASSUMED SYSTEM PARAMETERS AND THEIR ACCURACIES ALONG WITH 3D VEHICLE LOCATION.

$\Delta\mathbf{f}$	$\Delta\mathbf{b}$	$\Delta\mathbf{u_{H1}}, \Delta\mathbf{u_{E1}}$	$\Delta\mathbf{v_{H1}}, \Delta\mathbf{v_{E1}}$	$\Delta\mathbf{d_H}, \Delta\mathbf{d_E}$
40 px	0.01 m	2 px	2 px	5 px
f	**b**	**length**	**X**	**Y**
1756.72 px	1.445 m	4.2 m	0 m	0 m

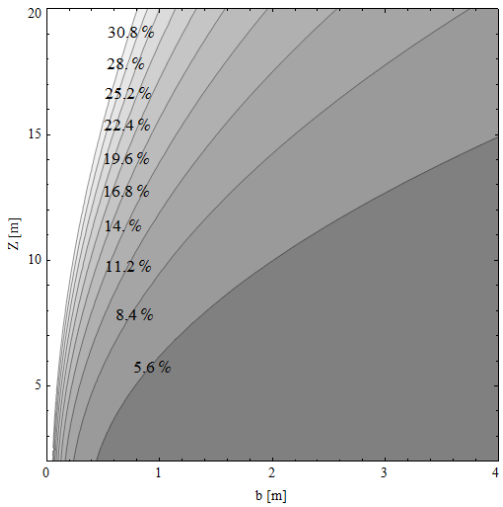

Fig. 2. Estimated accuracy with respect to distance of vehicle from the camera system (Z) and the camera baseline (b), assuming constant vehicle length L = 4.2 m and angle $\alpha = 0°$ (vehicle moving toward the camera).

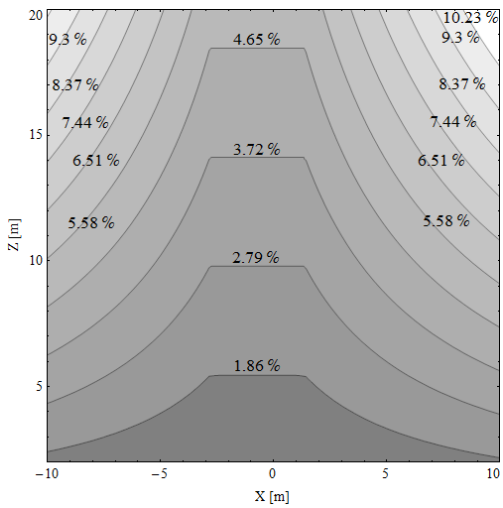

Fig. 3. Measurement accuracy with respect to distance of vehicle from the camera system (Z) and the position of the vehicle (X), assuming constant baseline b = 0.175 m, vehicle length L = 4.2 m and angle $\alpha = 90°$.

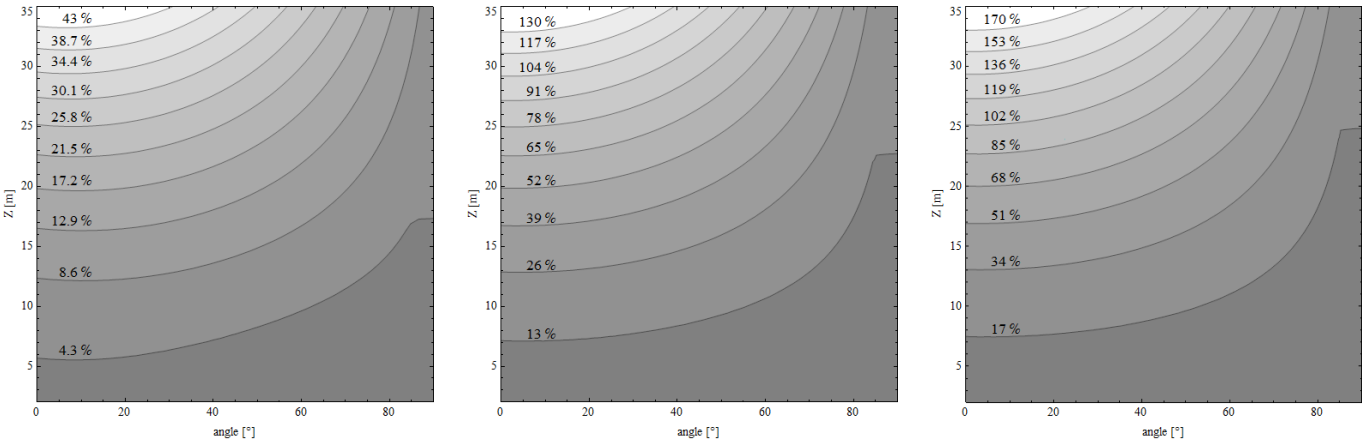

Fig. 4. Two-dimensional chart of the estimated accuracy with respect to distance of vehicle from the camera system (*Z*) and the angle of the vehicle (*angle*), assuming vehicle length L = 4.2 m and baseline b = 1.445 m (left), b = 0.228 m (center), b = 0.175 m (right).

In Table II we present the results of our previous experiments [6] concerning vehicle length estimation from stereoscopic video. The result are grouped according to the camera baseline values: 1.445 m, 0.228 m and 0.175 m (paths of vehicle's movement lied at the angles alpha: 75°, 63° and 60° respectively). Each vehicle was measured at a different distance from the camera system. Exact distance between vehicle and the camera system is also presented in the Table II. Estimated vehicle length and its true length is used to calculate measurement accuracy. Estimated accuracy can be compared with the measurement accuracy reported in [6].

TABLE II. RESULTS

Vehicle	*b* [m]	Distance from the camera [m]	Vehicle length [m]		Accuracy [%]	
			True	*Estimated*	*Measured*	*Estimated*
Fiat Panda II	0.175	34.291	3.560	**2.797**	21%	**90%**
Ford Focus I		27.487	4.292	**3.514**	18%	**50%**
Honda Concerto		26.970	4.415	**6.247**	41%	**47%**
Toyota Yaris		25.393	3.640	**4.217**	16%	**52%**
VW Caddy	0.228	19.539	4.405	**3.050**	31%	**51%**
Ford Focus I W		21.008	4.465	**3.434**	23%	**57%**
Skoda Octavia II W		19.204	4.578	**2.475**	46%	**50%**
Opel Meriva A		19.045	4.288	**2.117**	51%	**49%**
Alfa Romeo 147	1.445	8.865	4.223	**3.994**	5%	**4.3%**
Fiat 126p		8.540	3.054	**2.955**	3%	**5.1%**
Daewoo Tico		7.460	3.340	**3.393**	2%	**4.1%**
Opel Corsa		8.580	3.990	**3.863**	3%	**4.3%**
VW Caddy		8.560	4.405	**4.070**	8%	**4.0%**
Honda Concerto		8.260	4.415	**4.258**	4%	**3.9%**
VW Polo		8.310	3.916	**3.543**	10%	**4.2%**

VI. CONCLUSIONS

In the paper we have presented an approach to estimate accuracy of the stereo vision measurement system. Presented approach can be used to analyse distribution of accuracy in order to choose best acquisition parameters and conditions.

We have analyzed distribution of the accuracy depending on the distance of a measured vehicle from the camera system, stereo baseline and the angle of movement. Best results (higher accuracy) can be obtained with a wider baseline when measured vehicle is at the front of the camera system, relatively close to it. Also direction of movement strongly influences the accuracy. A vehicle moving to or from the camera system assures better accuracy comparing with the situation when a vehicle is passing across the image. This is an important factor for placing the camera system on the road. The best place would be on an axis of the road (for example above the road). Placing the system beside the road gives lower accuracy.

We have also compared theoretically estimated accuracy with the one obtained from experiments. The obtained accuracy can be successfully predicted with the proposed methodology.

ACKNOWLEDGMENT

Research project was supported by The National Centre for Research and Development, Poland. Grant no. NR02-0022-10/2011.

REFERENCES

[1] C. Wang, C. Thorpe, and A. Suppe, "LADAR-Based Detection and Tracking of Moving Objects from a Ground Vehicle at High Speeds" Proceedings. *IEEE Intelligent Vehicles Symposium*, 2003, pp. 416-421.

[2] A. Criminisi, J. Reid, A. Zisserman, "Single view metrology", *International Journal of Computer Vision*, vol. 40, no. 2, 2000, pp. 123-148.

[3] J. Stankowski, K. Klimaszewski, "Application of Epipolar Rectification Algorithm in 3D Television" Image Processing and Communications Challenges 2, Advances in Intelligent and Soft Computing, Springer, p. 345 - 352, 2010.

[4] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision (2nd edition)", *Cambridge University Press*, ISBN 0-521-54051-8, 2003.

[5] S. Smirnov, A. P. Gotchev, M Hannuksela, "Comparative analysis of local binocular and trinocular depth estimation approaches", *Proceedings SPIE 7724*, *Real-Time Image and Video Processing*, 2010.

[6] R. Ratajczak, M. Domanski, K. Wegner, "Vehicle size estimation from stereoscopic video", *19th International Conference on Systems, Signals and Image Processing (IWSSIP 2012)*, April 2012, pp: 405-408.