VEHICLE SIZE ESTIMATION FROM STEREOSCOPIC VIDEO

Robert Ratajczak, Marek Domański, Krzysztof Wegner

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology,
email: robert.ratajczak@doctorate.put.poznan.pl, domanski@et.put.poznan.pl,
krzysztof.wegner@doctorate.put.poznan.pl

ABSTRACT

The scope of the paper is stereoscopic CCTV with applications to road traffic surveillance. The main idea is to exploit stereoscopic video analysis for automatic estimation of dimensions of moving vehicles. The major steps of the proposed technique include stereoscopic video correction, identification of characteristic points of vehicle contours in video frames, depth estimation and vehicle physical length calculation. The technique has been tested using an experimental system described in the paper. The paper also comprises experimental results obtained in real environment for cars moving along a city street. The obtained results prove sufficient accuracy of the technique.

Index Terms — video surveillance, road traffic monitoring, vehicle size estimation, stereoscopic video

1. INTRODUCTION

Recently, significant progress in stereoscopic video processing has been achieved. Therefore a question arises if stereoscopic video processing may provide new opportunities to video surveillance. In this paper, we are going to prove that stereoscopic video processing brings substantial new possibilities to video surveillance. This will be shown for fully automatic vehicle length estimation that is important for:

- automatic road traffic analysis,
- detection of vehicles that violate restrictions,
- warning drivers of long vehicles (e.g. approaching tight curves) etc.

In this paper, we propose a system that calculates physical length of vehicles from stereoscopic video sequences. The main advantage of the proposed system is non-invasiveness of the solution. The system exploits stereoscopic video processing only and does not illuminate an object using any radiation (radar, infrared etc.). Further advantages of the system are: low cost, simplicity of installation and no need to incorporate any sensors into the road.

The novelty of the solution proposed is related to a system as a whole and its application rather than to individual techniques, which are common ones.

2. THE MAIN IDEA OF THE PAPER

The purpose of the proposed system is to estimate physical length of a moving vehicle from stereoscopic video. Such estimation is based on measurements of the distances from a camera to characteristic points on the body of the vehicle.

Firstly, a vehicle is identified in the left and in the right view sequences. In order to do that, image segmentation and shape recognition are used. Next, the shape is analyzed and a front point $\mathbf{M_1}$ and a back point $\mathbf{M_2}$ are automatically identified on the body of the vehicle (Fig. 1). Those points correspond to the head and the end of the vehicle. The length of the vehicle is calculated as the distance between this two points:

$$Length = \left\| \mathbf{M}_1 - \mathbf{M}_2 \right\|, \tag{1}$$

where $\|\cdot\|$ is the Euclidean distance in the 3D space, $\mathbf{M_1}$, $\mathbf{M_2}$ are the spatial coordinates in the real 3D world.

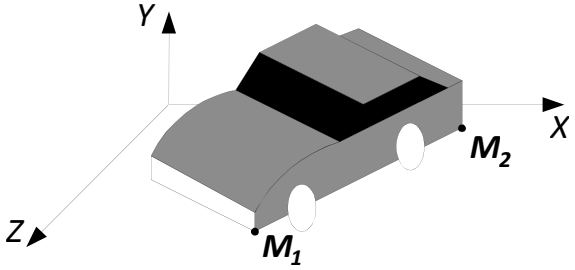

Figure 1. Characteristic points in a body of the vehicle.

In order to calculate the physical length of the vehicle we need to know the spatial location of the head ($\mathbf{M_1}$) and the end ($\mathbf{M_2}$) of the vehicle. The formulas needed to estimate and calculate the vehicle length may be easily derived from the simple camera model [2]. The real-world coordinates may be estimated as

$$\mathbf{M} = \mathbf{P}^{-1} \cdot z\, \mathbf{m}, \tag{2}$$

where $\mathbf{P}$ is a projection matrix (derived from calibration); $\mathbf{m} = [m_x,\ m_y]^{\mathrm{T}}$ is a vector of coordinates of the point on image plane (derived from image), T is matrix transposition; z is a distance from camera (called also depth).

For individual pixels in the image, the depth is estimated from a stereo image pair.

3. ALGORITHM

Once the system is installed in a new location, it requires calibrating. The calibration consists of the measurement of the parameters of the two-camera system. As the system operates, the accurate values of the parameters are used for compensation of the imperfections of the camera system in the process of rectification of the stereo video. The exact rectification is crucial for high accuracy of the vehicle size estimation.

For estimation of the vehicle size, the video processing technique is used. This technique implements the algorithm that consists of the following steps (Fig. 2):

1. Rectification of the acquired stereoscopic video.
2. Detection of moving vehicle in the stereoscopic video.
3. Recognition of the vehicle shape and estimation of the front $\mathbf{M_1}$ and the rear $\mathbf{M_2}$ points of the vehicle.
4. Estimation of the region of interest (RoI) around the vehicle.
5. Estimation of depth maps in the region of interest only.
6. Calculation of the final vehicle length based on the vehicle location and the estimated depth maps.

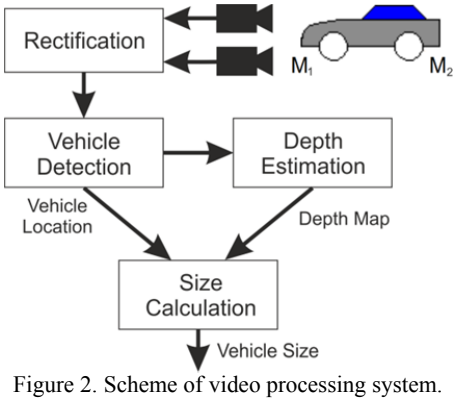

Figure 2. Scheme of video processing system.

4. SYSTEM CALIBRATION AND VIDEO RECTIFICATION

The exact camera system parameters can be estimated using checkerboard pattern with known dimensions. To estimate intrinsic matrix, translation vector and rotation matrix, the standard Zhang technique [2] may be used. In order to simplify further steps of stereoscopic image processing, like depth estimation, the input images have to be rectified [3].

The rectification accelerates depth estimation because corresponding pixels in both images lie on the same horizontal line. In case of a stereoscopic pair, accurate rectification is always possible [3]. Rectifying transform can be calculated using camera intrinsic and extrinsic parameters estimated during system calibration. It is also necessary to remove distortions caused by the imperfections of the camera lenses which is also done during system calibration. The color characteristics of the sensors in two cameras are usually different, and their correction is important for further depth estimation. The experience of the authors proves that this problem usually may be solved by matching the color histograms in the left and right pictures.

5. VEHICLE DETECTION

There are many shape detection algorithms that may be used to detect a vehicle in an image. In our implementation we used the well-known machine learning algorithm proposed by P. Viola and M. Jones [4], and extended by R. Lienhart and J. Maydt [5]. We used the typical rectangular feature set proposed in [5]. The abovementioned algorithm is based on a cascade of classifiers which allows hierarchical shape detection.

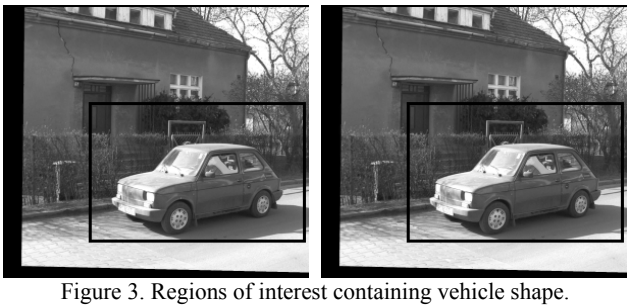

Figure 3. Regions of interest containing vehicle shape.

In the experiment, we have trained 4 cascades to detect a vehicle in 4 positions: front-left, front-right, rear-left, rear-right. This diversification allows us to easily find the end of the vehicle (point $\mathbf{M_2}$). In order to train cascades of classifiers we used over 4,000 positive samples (of size 60x45), containing a vehicle shape, and over 10,000 negative samples with a background. An outcome of vehicle detection is a rectangle containing the vehicle shape (Fig. 3).

Figure 4. Vehicle shape detected from the movement.

Moreover, for the region of interest (RoI), large moving objects are detected in the scene. During the absence of the car, the reference image is created. In order to detect motion, we use the reference image to cut out the moving objects (Fig. 4). The result of the motion detection helps us to find the end of the vehicle (point $\mathbf{M_2}$) as the most distant point. After post-processing using closing morphological operator, we obtain the vehicle shape.

6. DEPTH ESTIMATION

In order to calculate the vehicle length, the precise values of the coordinates of the characteristic vehicle points $\mathbf{M_1}$ and $\mathbf{M_2}$ are needed (see Eq. 1). The values of the coordinates x, y on the image plane can be taken directly from a rectified image (e.g. left one). The depth value (coordinate z) has to be estimated from a stereo image pair.

We made assumption that second characteristic vehicle point $\mathbf{M_1}$ is the nearest point (the highest depth value) from estimated region of interest (Fig. 5).

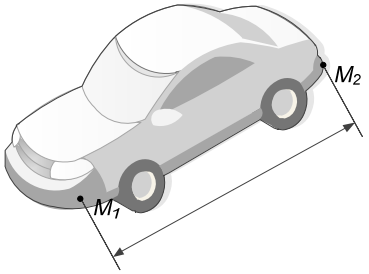

Figure 5. Locations of characteristic points $\mathbf{M_1}$ and $\mathbf{M_2}$.

Application of depth estimation is limited to the region of interest around the vehicle. Such an approach ensures that the amount of computations remains reasonable.

In the implementation, the state-of-the-art block matching technique has been used for depth estimation. The technique exploits sophisticated block matching algorithm which uses soft segmentation, along with graphs cuts optimization algorithm to produce disparity maps with sub-pixel accuracy [7]. This technique is implemented in Depth Estimation Reference Software (DERS) [6] used as a reference technique for 3D video standardization activities within the ISO Moving Picture Expert Group (MPEG).

The actual outcome of depth estimation are disparity values d for individual pixels.

7. VEHICLE LENGTH ESTIMATION

We assume that projection matrix $\mathbf{P}$ is given by:

$$\mathbf{P} = \mathbf{A} \cdot \left[\mathbf{R}|\mathbf{T}\right] \tag{3}$$

where: $\mathbf{A}$ is a [4x4] matrix of intrinsic camera parameters, $\mathbf{R}$ is a [3x4] camera rotation matrix, $\mathbf{T}$ is a [1x4] camera translation vector. From Eqs. (1) and (2) the physical dimensions of a vehicle may be calculated:

$$\begin{aligned} L &= \|\mathbf{P^{-1}} \cdot z_1 \cdot \mathbf{m_1} - \mathbf{P^{-1}} \cdot z_2 \cdot \mathbf{m_2}\| = \\ &= f \cdot b \cdot \left\|\mathbf{P^{-1}} \cdot \left(\frac{1}{d_1}\mathbf{m_1} - \frac{1}{d_2}\mathbf{m_2}\right)\right\| \end{aligned} \tag{4}$$

where: $\mathbf{P}$ is a projection matrix, f is a focal length (after rectification the same for both cameras), b is the baseline, d_1 and d_2 are the disparity values for points $\mathbf{M_1}$ and $\mathbf{M_2}$, respectively, $\mathbf{m_1}$, $\mathbf{m_2}$ are the projections of $\mathbf{M_1}$ and $\mathbf{M_2}$ points onto the image plane.

8. ESTIMATION OF MEASUREMENT PRECISION

In order to estimate the precision of the length L estimation the total differential method may be used (see Eq. 5):

$$\begin{aligned} \Delta L \approx\ & \frac{\delta L}{\delta f} \cdot \Delta f + \frac{\delta L}{\delta b} \cdot \Delta b + \frac{\delta L}{\delta d_1} \cdot \Delta d_1 + \frac{\delta L}{\delta d_2} \cdot \Delta d_2 \\ & + \frac{\delta L}{\delta x_o} \cdot \Delta x_o + \frac{\delta L}{\delta y_o} \cdot \Delta y_o \\ & + \sum_{i=1}^{2} \frac{\delta L}{\delta m_{ix}} \cdot \Delta m_{ix} + \sum_{i=1}^{2} \frac{\delta L}{\delta m_{iy}} \cdot \Delta m_{iy} \end{aligned} \tag{5}$$

where: x_o and y_o are coordinates of the principal point. The projection matrix $\mathbf{P}$ is represented by focal length f, baseline b and coordinates of the principal point x_o and y_o, and therefore it can be omitted in Eq. 5.

The theoretical considerations about precision of the proposed technique were made with the following assumption: actual vehicle length - 4000 mm, parallel position relative to the image plane, minimal disparity value - 1, image resolution - 1920x1080, focal length - 1756.72 (experimentally estimated for Canon XH-G1 used in the experiment). Using the results [2], camera precision parameters increase with the number of calibration shots. We have used twelve different calibration shots that allow us to estimate focal length of our camera system with precision of $\Delta f \approx 2\%$. The disparity is obtained with precision $\Delta d = 1$. Experiments show, that coordinates of the points $\mathbf{m_1}$ and $\mathbf{m_2}$ are obtained with precision $\Delta m_x = \Delta m_y = 5$ (i.e. 5 sampling periods). Analysis of the measurement precision yields that the minimal measurement error can be achieved when the baseline value is contained in the range of 1300 mm and 2800 mm.

9. EXPERIMENTAL RESULTS

9.1. Test set

Evaluation of the proposed system is not a straightforward task. First of all, in general, there is no stereoscopic test sequence created for that specific task available. We faced the problem of creating such a stereoscopic test sequences, that must be recorded in natural conditions, typical for the urban monitoring systems.

In order to record stereoscopic test sequences, we have used two HD Canon XH-G1 cameras, that captured uncompressed, 1080p stereoscopic video. This configuration emulates high quality video surveillance systems used nowadays.

For the experimental purposes, we have recorded sixteen sequences presenting several cars models with various camera baselines. Exemplary frames from the recorded test sequences are presented in Figure 6.

Figure 6. Examples of stereoscopic test sequences.

9.2. Results

In order to evaluate performance of the proposed system we used the recorded stereoscopic test sequences. We have used our algorithm for a stereoscopic pair, and have compared the estimated length of the recorded car with its true length taken from the car model datasheets.

Table 1. Results for example test sequences.

Sequence name	Distance range from the camera [m]	True vehicle length [mm]	Average estimated vehicle length [mm]	Average relative error [%]
Fiat 126p	> 13	3054	1678	47
	9 - 13		2530	18
	< 9		2955	4
Alfa Romeo 147	> 13	4223	2708	36
	9 - 13		3717	12
	< 9		3994	6
Daewoo Tico	> 13	3340	2948	12
	9 - 13		3150	6
	< 9		3393	2
Honda Concerto	> 13	4415	3137	29
	9 - 13		3899	12
	< 9		4258	4
Opel Corsa	> 13	3990	2816	29
	9 - 13		3545	12
	< 9		3863	3
Volkswagen Polo	> 13	3916	2841	28
	9 - 13		3390	14
	< 9		3557	6
Skoda Fabia	> 13	3960	2815	29
	9 - 13		3215	18
	< 9		3557	8
Volkswagen Caddy	> 13	4405	2710	38
	9 - 13		3436	22
	< 9		4070	8

In Table 1 the results are presented for the vehicle length estimation with the baseline b = 1445 mm. The results are grouped according to the distance between vehicle and camera system into 3 groups: far, middle and near. These particular groups refer to distances: over 13 meters, 9-13 meters and less than 9 meters, respectively. The measurements were carried out in successive frames of stereoscopic test sequences, and the results were averaged within the above mentioned distance intervals. The average relative error decreases with the distance from cameras. This confirms the theoretical considerations described in Section 8. This average relative error does not exceed 10 percent.

In practice, an increase in the baseline length results in the improvement of precision.

10. CONCLUSIONS

In the paper a novel approach to automatic estimation of the dimensions of moving vehicles is presented. The main new idea is to use the analysis of stereoscopic video sequences.

The system has been implemented and a series of tests for cars has been performed. It is shown that the length estimation accuracy may be below 10% for cars moving closely to the cameras.

The system exploits an original depth estimation technique proposed at Poznań University of Technology [6].

The system does not need sophistical installation but the cameras need to be of high resolution. Further studies will be done for wider group of vehicles and various baseline b distances.

11. ACKNOWLEDGEMENT

This work was supported by the public funds as a research project.

REFERENCES

[1] M. Ito, Y. Takada, T. Hamamoto, „Distance and Relative Speed Estimation of Binocular Camera Images Based on Defocus and Disparity Information”, 28th Picture Coding Symposium, Nagoya, Japan, p. 278 - 281, December 2010.

[2] Z. Zhang, “A Flexible New Technique for Camera Calibration”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22(11), p. 1330 - 1334, November 2000.

[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] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, USA, Vol. 1 p. 511 - 518, 2001.

[5] R. Lienhart, J. Maydt, “An Extended Set of Haar - like Features for Rapid Object Detection”, Proc. of IEEE Int. Conf. on Image Processing 2002, New York, USA, p. 900 - 903, 2002.

[6] O. Stankiewicz, K. Wegner, M. Wildeboer, "A Soft – Segmentation Matching in Depth Estimation Reference Software (DERS) 5.0" ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17049, Xian, China, October 2009.

[7] “Report on Experimental Framework for 3D Video Coding” JTC1/SC29/WG11 MPEG 2010 / N11631, Guangzhou, China, October 2010.