Vehicle Dimensions Estimation Scheme Using AAM on Stereoscopic Video

Robert Ratajczak, Tomasz Grajek, Krzysztof Wegner,
Krzysztof Klimaszewski, Maciej Kurc, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznań University of Technology
Polanka 3 street, 60-965 Poznań, Poland
{rratajczak,tgrajek,kwegner,kklima,mkurc}@multimedia.edu.pl

Abstract

Recently, stereoscopic video systems are studied in the context of many new applications. This paper addresses the question how the road traffic surveillance may benefit from stereoscopic video analysis. In particular, we propose a novel method for estimating vehicle dimensions using stereoscopic video analysis. This method is based on active appearance model (AAM), a technique already used, e.g. for face detection. Experimental results are provided in the paper in order to demonstrate satisfactory precision of the measurements.

1. Introduction

This paper presents a novel technique for automatic estimation of vehicle dimensions based on stereoscopic video analysis. The goal of the work was to build a versatile system capable of extracting different dimensions of vehicles. Information from such a system is valuable for many applications, such as automatic traffic management, traffic measurements or statistical data gathering, to name a few. One of the possible applications may be warning drivers and the authorities about oversized vehicles. It can also be used in vehicle admission control systems or other security systems.

The system developed at Poznan University of Technology is able to perform fully automatic analysis of stereoscopic traffic images and extract physical dimensions of vehicles seen. Apart from typical length, width and height, also specific dimensions, like distance between axles or tire diameter can be provided by our system.

Most of the contemporary systems aim at vehicle localization and recognition, for example [10]. They exploit information from a single camera only to detect and recognize a particular type of the vehicle [11]. Such an approach cannot provide the real dimensions of the localized vehicle.

Nowadays road traffic is observed by multiple cameras (for example one camera per lane). This can provide more information about a vehicle than in the case of a single camera, including real vehicle dimensions.

2. Main Idea

The main purpose of our system is to measure vehicle dimensions. In order to do that, we must choose two points (1) between which distance will be measured (Figure 1).

$$\boldsymbol{M_H} = [X_H\ Y_H\ Z_H\ 1]^T \quad \boldsymbol{M_E} = [X_E\ Y_E\ Z_E\ 1]^T \tag{1}$$

Let us assume for the moment that chosen points are the head $\boldsymbol{M_H}$ and the end $\boldsymbol{M_E}$ of the vehicle, and thus we would like to calculate length of the vehicle. The estimated length L is simply Euclidean distance (norm) between those two points (2)

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

Unfortunately we don't know the 3D location of the selected points. All we have is an image acquired by the camera, which is a 2D projection of the vehicle being measured [1]. Acquisition of the image by the camera can be considered as projection of points from 3D location onto the image plane (3)

$$z\boldsymbol{m} = \boldsymbol{P} \cdot \boldsymbol{M} \tag{3}$$

where $\boldsymbol{M}$ is a [4x1] vector representing location of the point in 3D space in homogenous coordinates. $\boldsymbol{P}$ is a [4x3] projection matrix representing a camera intrinsic parameters and its spatial location. The vector $\boldsymbol{m}$ is a [3x1] vector representing location of the projected point onto the image plane in homogenous coordinates. Scalar z expresses distance between the camera and the analyzed 3D point and it is often called depth of the point.

From the acquired image we can get 2D coordinates of the selected characteristic points on the vehicle body. Inverting (3) gives us a possibility to get 3D location of abovementioned points (4)

$$\boldsymbol{M} = \boldsymbol{P}^{-1} \cdot (\mathrm{z}\,\mathbf{m}) \tag{4}$$

So the distance between selected points ($\boldsymbol{M_H}$ and $\boldsymbol{M_E}$) can be calculated from (5):

$$\begin{aligned} L &= \|\boldsymbol{M_H} - \boldsymbol{M_E}\| = \\ &= \|\boldsymbol{P}^{-1} \cdot (z_H \boldsymbol{m_H}) - \boldsymbol{P}^{-1} \cdot (z_E \boldsymbol{m_E})\| = \\ &= \|\boldsymbol{P}^{-1} \cdot (z_H \boldsymbol{m_H} - z_E \boldsymbol{m_E})\| \end{aligned} \tag{5}$$

The only unknown variables are the depth values of the selected points.

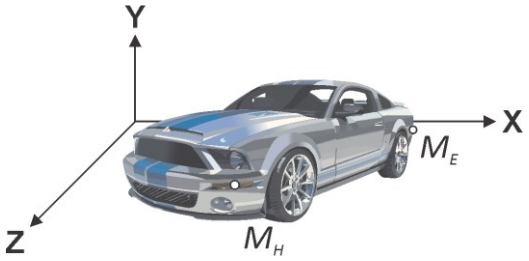


Figure 1: Location of selected points

There are many ways to estimate the depth values. One of the most common is to use disparity estimation from stereo pair for example like in [2]. For this reason, in our system we use two cameras which acquire two images of the vehicle at the exactly the same moment.

The basics are as follows. If we know the 2D locations of the same point in both acquired images, we could calculate disparity (6) which is a difference of the point's positions in both images (acquired by first and second camera):

$$d = \|(\boldsymbol{m_1} - \boldsymbol{m_2})\| \tag{6}$$

For rectified, well calibrated stereo images, disparity can be converted to depth value of the point by simple trigonometric relations (7):

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

where f is the focal length of the camera used, b is distance between both cameras and d is disparity of the given point. Therefore, inserting (7) into (5) results in (8):

$$\begin{aligned} L &= \|\boldsymbol{P}^{-1} \cdot (z_H \boldsymbol{m_H} - z_E \boldsymbol{m_E})\| = \\ &= \left\|\boldsymbol{P}^{-1} \cdot \left(fb\frac{1}{d_H}\boldsymbol{m_H} - fb\frac{1}{d_E}\boldsymbol{m_E}\right)\right\| = \\ &= fb\left\|\boldsymbol{P}^{-1} \cdot \left(\frac{1}{d_H}\boldsymbol{m_H} - \frac{1}{d_E}\boldsymbol{m_E}\right)\right\| \end{aligned} \tag{8}$$

Of course initially the locations of the selected vehicle points in both images are unknown, and we have to find it in both images. One approach is to use the information about depth. Vehicle measurement system with dense depth estimation approach can be found in [3]. In the method proposed in [3] the vehicle body is first detected in the images and then the depth is estimated for the vehicle and surrounding region of interest only. However it still requires a lot of time to estimate depths for the detected regions of interest.

In order to reduce the computation time of vehicle dimensions estimation, we propose to merge vehicle detection and depth estimation steps into a single step. To be able to combine those two steps, we need to find location of vehicle shape very precisely in both images and then calculate depth values trough disparity of the shape points (fig 2).

There are many shape detection algorithms that may be used to determine the location of vehicle shape in an image. There are many examples of exploiting Haar features [3, 4], Active Shape Model [5]. However, to achieve an acceptable accuracy of determining the location of vehicle shape, we propose a technique which was initially developed for face detection and analysis - an Active Appearance Model (AAM) [6]. Currently our proposal assumes that AAM is applied to both of the views independently. After a correct model matching, we can use coordinates of particular points from AAM to calculate disparity between model points (Figure 2). Once accurate coordinates of the selected points on the body of a car in the images from both cameras are found, the reconstruction of 3D coordinates of those points can be performed with satisfactory results.

3. AAM Model

The Active Appearance Model (AAM) is a technique which exploits deformable model matching into an object's image [6]. Originally it was developed for face detection but it has been proved useful for various kinds of objects [7]. The AAM consists of two parts: shape and appearance (texture). The shape $\boldsymbol{s}$ is defined by a set of points which are grouped into multiple closed polygons (9),

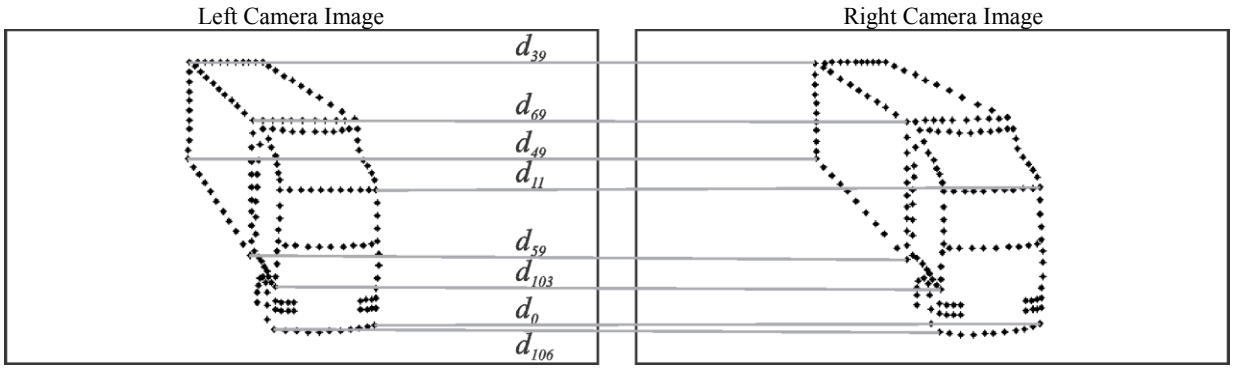


Figure 2: An exemplary AAM and location of fitted model in both images.

$$\boldsymbol{s} = \{x_1, y_1, x_2, y_2, \dots, x_n, y_n\} \quad (9)$$

while the appearance (texture) $\boldsymbol{g}$ consists of all the pixels that lie inside the defined shape polygons (10).

$$\boldsymbol{g} = \{Y_1, Y_2, \dots, Y_n\} \quad (10)$$

The goal of matching the model with an image is to find such a set of shape and appearance (texture) parameters which would minimize the error between the model texture (appearance) and the underlying image.

In order to model a set of possible variations of the shape and texture, a statistical model of both, using PCA analysis [8], is constructed over the training set. After the PCA analysis each shape can be represented by a shape parameter $\boldsymbol{b_s}$ and mean shape $\bar{\boldsymbol{s}}$.

$$\boldsymbol{s} = \bar{\boldsymbol{s}} + \boldsymbol{P_s b_s} \quad (11)$$

$\boldsymbol{P_s}$ is a matrix describing a set of shape modes. Similarly each appearance (texture) of the object after warping [7, 9] to a mean shape $\bar{\boldsymbol{s}}$ can be represented by appearance parameter $\boldsymbol{b_g}$ and mean appearance $\bar{\boldsymbol{g}}$ (13). $\boldsymbol{P_g}$ represents main appearance modes (eigenvectors of texture samples representing main part of observed variations in a training set)

$$\boldsymbol{g} = \bar{\boldsymbol{g}} + \boldsymbol{P_g b_g} \quad (13)$$

Matching the AAM into a new image is performed by minimizing magnitude of the error vector $\boldsymbol{\delta g}$. The vector $\boldsymbol{\delta g}$ is a difference between the image warped into the mean shape and the model's appearance synthesized using current appearance parameter $\boldsymbol{b_g}$. Warp process is controlled by current shape modeled by shape parameters $\boldsymbol{b_s}$.

AAM matching is an iterative process. At each step an update of the model parameters $\boldsymbol{b_s}$ and $\boldsymbol{b_g}$ is calculated based on observed error vector $\boldsymbol{\delta g}$ throughout linearized relation between these vectors (14)

$$\left[\boldsymbol{\delta b_s}\ \boldsymbol{\delta b_g}\right]^T = \boldsymbol{A} \cdot \boldsymbol{\delta g} \quad (14)$$

The matrix $\boldsymbol{A}$ is called the search matrix. The process stops when further change of $\boldsymbol{b_s}$ and $\boldsymbol{b_g}$ does not minimize the error $|\boldsymbol{\delta g}|$.

4. AAM for stereoscopic video

A straightforward approach would be to use two AAMs, one for each image in stereo pair. However, this would require twice as much data as a single-image AAM. Moreover, vehicle seen from the second camera is only slightly rotated comparing to the first one. This rotation has a magnitude that can occur even in a single camera image when different vehicles pass the camera at a slightly different angle. That is why we proposed to develop one, view-independent AAM, which can be used to search vehicles in both of images of the stereo pair captured by two cameras. That approach makes AAMs less sensitive to possible changes caused by imperfection of two-camera system placement (optical axes of the cameras are never exactly parallel) and simultaneously less sensitive to direction of vehicles' movement. Such AAMs had to be prepared on the basis of stereoscopic images of the vehicles.

5. AAM-based depth estimation and vehicle dimensions estimation

Our goal was to determine depth value without involving full-frame depth estimation algorithms. Therefore we exploit AAM search shape fitting. We search for the same vehicle in both images independently using our joined view-independent AAM. We assume that after AAM search step, particular points of the fitted models indicate 2D locations of the same 3D point in both images. That allows us to simply calculate disparities between those corresponding model points fitted in both images.

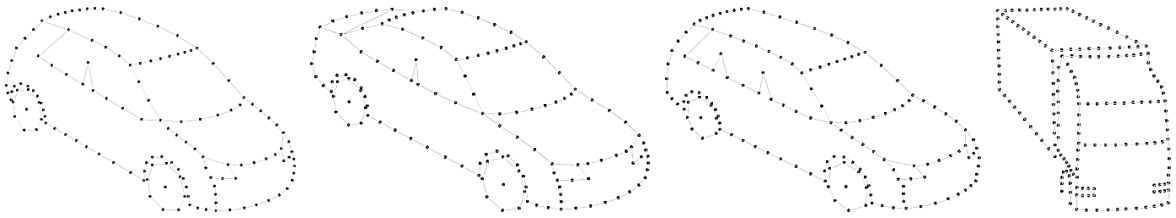
We performed experiments to determine the necessary level of complexity of our models that would ensure the best matching between the model and vehicle on the image. Initially, our models were quite simple and consisted of over a dozen nodes, but the matching results were not satisfactory. Therefore, we have updated our AAMs, expanding them with additional nodes (Figure 3). The details about the proposed models are presented in Section 7, and Table 1.

6. Vehicle database – experimental material

Unfortunately, we have no access to any database of traffic images captured by well calibrated stereoscopic cameras. Therefore, we created our own set of stereoscopic video shots. We have recorded more than 5 hours of traffic at a highway with the use of stereoscopic camera system. Our system consists of two Basler aviator avA1900-50gc cameras placed on tripods at 2.5 m baseline. We chose these cameras due to their parameters (1920x1080 pixels. resolution, 50 fps, RAW output format and mutual synchronization) and very high manufacturing quality. For the captured video material we have estimated camera system parameters (both intrinsic, and extrinsic, including scale) to avoid any 3D reconstruction ambiguities. All captured material have been rectified, based on estimated system parameters. A short summary of our test material can be found in Table 1.

7. Vehicles AAM development

Vehicles on roads differ in size and shape. We can distinguish for example: trucks, sedans, hatchbacks and wagons. For each of these vehicle types different points on their body must be precisely found, in order to calculate their physical dimensions. That is the reason why we have



a) b) c) d)

Figure 3: Vehicles' Active Appearance Models: a) hatchback (140 characteristic points), b) sedan (138 characteristic points), c) wagon (151 characteristic points), d) truck (178 characteristic points).

prepared several view-independent AAMs - one per each vehicle type. As for now, we have trained AAMs only for hatchback, sedan, wagon and truck but the scheme can be applied to any other type of vehicles.

To develop an AAM for each vehicle a number of training stereo image samples presenting single vehicle have been used. Vehicles in each stereo sample have been precisely manually labeled in both images. Stereo samples of given vehicle type have been used to estimate single view-independent AAM.

Finally, the proposed models consist of 140, 138, 151 and 178 shape points for hatchback, sedan, wagon and truck respectively (Figure 3, Table 1). Besides basic vehicle dimensions i.e. length, width, and height (Figure 4) our models allow automatic calculation of the distance between many others characteristic points on a vehicle body (e.g. wheel size or wheelbase).

8. Experimental results

In order to test our algorithm, we implemented software which performs automatic vehicle detection and measurement. Most important operations for our measurement system are listed below:

1. to search vehicle with use of AAM in both images,
2. to exploit stereo correspondences to find the three dimensional shape of the vehicle,
3. to perform the measurements,

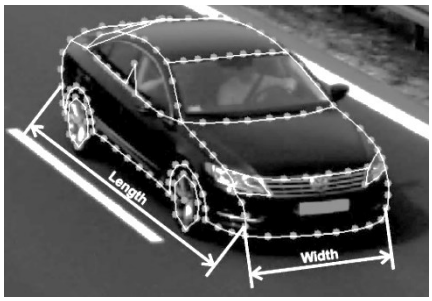
Our experiments were performed on images containing vehicles that were not used to train AAMs. Each stereo image was analyzed and AAM have been fitted to the vehicle found in both images. Next, disparities between corresponding points of the fitted model were calculated. Based on the disparities of the model points, appropriate depth values were calculated and vehicle dimensions was obtained through (8).

Unfortunately, semi-trailers have various shapes and sizes. Therefore we were unable to determine the measurement error of the whole truck (tractor with semi-trailer), due to the lack of data concerning its ground truth size.

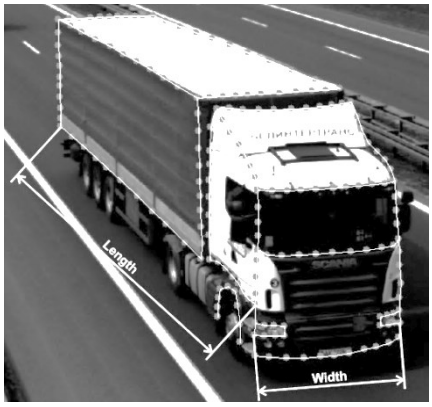
In order to compare our results with the ground truth ones, in Table 2 we present several sizes of the tractor's cabin taken from the particular tractor's model datasheet. Data are presented as follows: in first ground truth values, then estimated distances between particular AAM's nodes.

Vehicle type	No. of nodes	No. of samples
Hatchback	140	> 100
Sedan	138	> 150
Wagon	151	> 100
Truck	178	> 200

TABLE 1 EXPERIMENTAL MATERIAL DESCRIPTION.



a)



b)

Figure 4 Exemplary vehicle dimensions for: a) sedan and b) truck.

Truck no.	Height		
	Ground Truth	Automatically Estimated	Relative Error
1	2.90	2.73	5.86 %
2	2.88	2.99	3.82 %
3	2.90	3.21	10.69 %
4	2.96	2.84	4.05 %
5	2.96	2.88	2.70 %
Average	-	-	5.43 %

Truck no.	Width		
	Ground Truth	Automatically Estimated	Relative Error
1	2.43	2.12	12.76 %
2	2.44	2.22	9.02 %
3	2.43	2.26	6.99 %
4	2.49	2.18	12.45 %
5	2.49	2.01	19.28 %
Average	-	-	12.10 %

Truck no.	Length		
	Ground Truth	Automatically Estimated	Relative Error
1	2.26	1.92	15.04
2	2.28	1.86	18.42
3	2.26	2.22	1.77
4	2.25	2.06	8.44
5	2.25	2.22	1.33
Average	-	-	9.00 %

TABLE 2 COMPARISON OF GROUND TRUTH DIMENSIONS WITH ESTIMATED DIMENSIONS.

For the results from Table 2, we have calculated a relative error of dimensions measurement. We also present average relative error for particular dimension and for particular truck. Moreover, we present total average error for the whole dataset. Obtained relative measurement errors are align with results obtain throughout dense depth estimation from [3]. Moreover there are close to the theoretical boundary obtained in [12].

9. Conclusions

We have presented fast and robust vehicle measurement algorithm operating on stereoscopic video. The algorithm can be utilized in urban surveillance and traffic control. The main advantage is omitting an expensive dense depth estimation step, and using vehicle localization step instead for the purpose of disparity estimation.

Moreover the proposed method allows calculating not only the basic vehicle dimensions i.e. length, width, and height but also the distance between many others characteristic points on a vehicle body (e.g. wheel size or wheelbase). Obtained results are align with the results obtained with use of dense depth estimation methods, and close to the theoretical precision boundary.

Acknowledgment

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

References

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

[2] Y.-C. Wang, C.-P. Tung, P.-C. Chung: "Efficient Disparity Estimation Using Hierarchical Bilateral Disparity Structure Based Graph Cut Algorithm With a Foreground Boundary Refinement Mechanism," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 5, pp. 784-801, 2013.

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

[4] H Bai, J. Wu, C. Liu: “Motion and Haar-like Features Based Vehicle Detection”, 12th International Multi-Media Modelling Conference Proceedings, 2006.

[5] Y. Li, L. Gu, and T. Kanade: “A robust shape model for multi-view car alignment”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 2466–2473, 2009.

[6] T. F. Cootes, G. J. Edwards, C. J. Taylor: „Active appearance models”, 5th European Conference on Computer Vision, H. Burkhardt and B. Neumann, eds., vol. 2, pp. 484-498, Springer, Berlin, 1998.

[7] J. Wang, Y. Xu, X. Zhang: "A Method to Recognize Ship Target in SAR Imagery Using an AAM", International Conference on Computer Science and Software Engineering, vol. 1, pp. 1025-1027, 2008.

[8] I.T. Jolliffe: "Principal Component Analysis (2nd edition)", Springer-Verlag, ISBN 978-0-387-95442-4, 2002.

[9] T. Beier, S. Neely: "Feature-Based Image Metamorphosis", Proceedings of SIGGRAPH '92, pp. 35-42, New York, 1992.

[10] J. Lou, T. Tan, W. Hu, H. Yang, and S. J. Maybank, “3-D model-based vehicle tracking,” IEEE Trans. Image Process., vol. 14, no. 10, pp. 1561–1569, Oct. 2005.

[11] Z. Zhang; T. Tan; K. Huang; Y. Wang, "Three-Dimensional Deformable-Model-Based Localization and Recognition of Road Vehicles," Image Processing, IEEE Transactions on , vol.21, no.1, pp.1,13, Jan. 2012.

[12] T. Grajek, R. Ratajczak, K. Wegner, M. Domański, „Limitations of Vehicle Length Estimation Using Stereoscopic Video Analysis,” 20th International Conference on Systems, Signals and Image Processing, 2013.