Archives of
Transport System
Telematics

Volume 6

Issue 3

September 2013

3D objects modelling and description from a traffic flow

M. BOJMIR[a], R. HUDEC[a], P. KAMENCAY[a], R. PIRNIK[b]
[a] UNIVERISTY OF ZILINA, Faculty of Electrical Engineering, Department of Telecommunications and Multimedia, Univerzitna 8215/1 010 26 Zilina, Slovakia
[b] UNIVERISTY OF ZILINA, Faculty of Electrical Engineering, Department of Control and Information Systems, Univerzitna 8215/1 010 26 Zilina, Slovakia
EMAIL: martin.bojmir@fel.uniza.sk

ABSTRACT
Contemporary research activities in the area of transportation systems usually utilize computer vision techniques. These activities are mostly oriented to the analysis of traffic flow. Key parts of the analysis consist of detection, extraction, modelling and recognition of objects in the traffic flow. Additional information can be obtained by the object tracking, too. The main aim of this contribution is to propose a new method for modelling 3D objects that are moving in the traffic flow. At the beginning of this paper, basic methods of object detection and extraction are shortly described. Moreover, the modified algorithm based on the object extraction and 3D models creation by mesh is also proposed. That mesh model is generated from a depth map. Finally, the results of proposed method are presented in the last part of this paper.

KEYWORDS: vehicle detection, vehicle extraction, depth map, 3D modelling polygonal mesh

1. Introduction

In recent years the area of transportation worldwide faced a crisis due to a limited capacity of main traffic flows. The density of vehicles has been continually growing, particularly in personal transportation. This state has caused more frequent traffic jams and accidents that notably affect the fluency of traffic flows. Because of that novel approaches to the traffic flow analysis have been proposed [1]. In general, these approaches belong to two different areas. The first, called invasive, requires a physical disturbance to the road surface and utilizes some type of electromagnetic sensors. The second, called non-invasive, for the data acquisition utilizes only camcorders placed close to the traffic flow. The video data provide more detailed information and specifications of the traffic flow. One of the non-invasive approaches applies image processing and particle analysis to indicate an increase in traffic. In the first part, colour plane extraction, image reversal and morphological function are performed. Consequently, grayscale thresholding and image masking are utilized. Finally, image rectification on the number plate is performed. This approach allows sending an alert about a traffic jam in order to inform about a critical density of traffic flow [2]. Another proposed approach uses the video data registered by camcorders. The video data are subsequently processed and the obtained information serves for object modelling [3]. A successive creation of a 3D vehicle model based on combining single parts of the object has been applied. Sequentially, 2D features are mapped on a 3D generic model. 3D features are adaptively aggregated over frames in order to fill up a 3D model. Another proposed approach utilizes a probabilistic 3D scene model. This model is based on objects detection, tracking their position in the traffic flow and on 3D geometrical relationships between them [4].

2. System overview

The basic system works with the video data, captured by a 3D camcorder that is placed above the traffic flow. The distance between two lens of camcorders is approximately 4cm. The source data contains a high-definition video with a resolution of 1920x1080/25i. The data preprocessing is performed before any image processing. This step is exercised to reduce the computing demand. Preprocessing involves deinterlacing and adjusting the video resolution to 720p. In the first step of processing, the background subtraction around moving objects is performed. This processing consists of some filters that allow reducing the noise and tiny unwanted motions, for example leaves motion. After the object extraction, a depth map can be created.

2.1 Enhanced system

The enhanced system includes additional steps of image processing and utilization of two 2D camcorders instead of one 3D camcorder. The centre between two camcorders represents the centre of traffic flow. We tested various camcorders optics distances, namely 3.5m, 4m and 5m. If the distance between two camcorders is bigger, a deeper space sense is obtained. We assume that this modification can provide better results in the phases of detection and subsequent processing of single objects. Supplemental steps include removing the shadows around the vehicle, as well as noise and blobs suppressions that are undesirable in the further image processing. Then, the object extraction is performed from the images. In the next step, a hybrid segmentation is calculated. The result of segmentation is processed by a stereo matching SAD (Sum of Absolute Differences) algorithm following the disparity/ depth map computing. A 3D model based on the data from the depth map is depicted. Also a 3D mesh model is generated from the depth map.

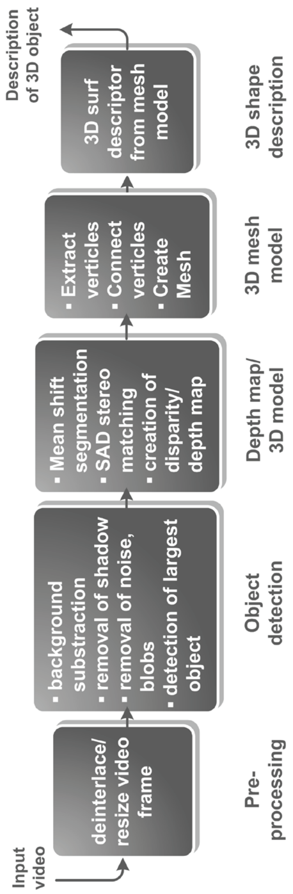

Fig. 1. Overview of proposed approach

3. Object detection and creation of depth map

3.1 Object extraction

This part of the proposed approach is focused on the sub-optimal background subtraction from the input image. In the first step, the time median for each pixel for a defined number of video frames is calculated. The number of frames in a queue is defined by a parameter called background adaptation. Moreover, the difference between the value of each pixel in the current frame and the time median of these pixels value in sequence is calculated. The image transformation to the HSV (Hue Saturation Value) colour space and filters that are used for dilation and erosion are also applied. These steps ensure appropriate detection of moving objects. To reduce the video data stream, a part which allows effective set intra/inter frame image coding, is embedded into the algorithm. A further step of the algorithm consists of shadows detection around moving objects. In general, this way of detection is based on the statistics of the intensity of luminance component of the image pixels. Calculation of the histogram and then calculation of the total average of the y-component of the overall image in the YCbCr colour space is performed. Consequently, the iterative process is performed, where the value of y-component of sliding window with average value y-component of the overall image is compared. The sliding window has an NxN size and it is iteratively reduced to the size of 3x3 pixels. In this approach, points detected as a shadow are progressively tagged.

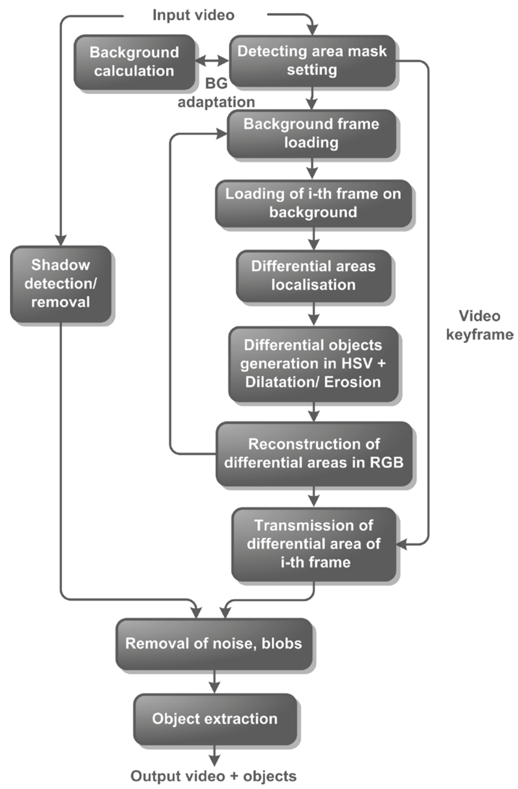

Fig. 2. Principal scheme of video object detection

Subsequently, the median filter is utilized to eliminate the noise from the filtering mask. The result of this step is a binary mask, which is applied to remove the shadow from the image. The next part of our proposed algorithm executes logical multiplication of the mask with detected shadows and the output image after removing the background. Then a filter removing the noise and blobs from this image is applied. Its size is defined by the threshold, resulting from executed experiments. Thus the image is appropriate for extraction of single objects. In Figure 2 the principal scheme of the object detection and extraction process from the input video is shown.

3.2 Creation of depth map

In this part, the algorithm for creation of a depth map is briefly described [5]. The extracted object is depicted in a stereo image captured by two 2D camcorders. At the beginning of algorithm it is necessary to apply a calibration matrix that is obtained from a properly calibrated camera. It is important to perform this step, because distortions can bring some inaccuracies to the whole process. A principal scheme of algorithm for creation of disparity/depth map is shown in Figure 3.

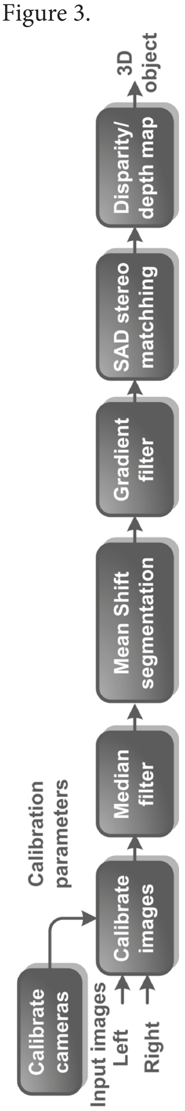

Fig. 3. **Principal scheme of 3D object creation [5]**

The application of image segmentation is the next step of the algorithm. In general, the segmentation is performed in order to simplify the representation of information from the image data. The main task is to divide the input image into single segments according to defined terms. A mean shift based segmentation algorithm is a nonparametric iterative segmentation method based on progressive calculation of the weighted average of pixels in the defined window. This average value is calculated according to how single pixels in the window are similar in relation to colour components. The mean shift includes two principal parameters. The first parameter defines the size of the surrounding area of the point, for which it is calculated. The second parameter represents the maximum colour distance between pixels of calculated window and the mean pixel. In this iterative process a new position is searched based on the previous two parameters, with which it is to work in the next step. A place with the highest kernel density is in the final position, whereby the kernel represents a radially symmetrical weighted function that determines the weights for particular pixels of calculation window by their distance from the centre and from the difference of colour components. After this process, the input image is divided into single parts that fill completely a full image.

A further step of image processing consists of stereo matching algorithm performance and consequently calculation of disparity map. Left and right segmented images contain pixels that belong to each other. These pixels are gradually found. Let the observed object is viewed through two camcorder's lenses that are at a distance of 2*l*. This distance is the same as the length of base B. From this point of view, the observed object appears to be in a similar position in both images. The object distance between left and right image designates parameter - disparity d. This parameter is defined by:

$$d = x_L - x_R = f\left(\frac{x_p + 1}{z_p} - \frac{x_p - 1}{z_p}\right) \tag{1}$$

$$z_p = \frac{2fl}{d} = \frac{fB}{d} \tag{2}$$

where x_L, x_R represent the coordinates x of the projected 3D coordinate onto the left and right planes I_L a I_R. In the case, that left and right image plane is located in the same plane, y - coordinates of the two images are the same.

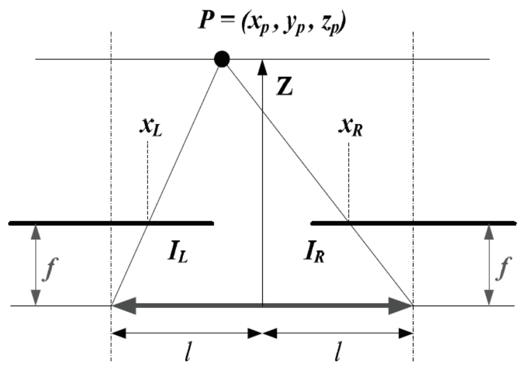

Fig. 4. **Principal scheme of stereo matching**

Known values of camera parameters and obtained values of disparities are used to calculate the depths of image segments. In the following step, a stereo matching algorithm that reconstructs the disparity map is calculated. Stereo matching algorithms use different mathematical methods to calculate disparity maps [6]. Usually utilized methods involve the sum of absolute differences (SAD). This method is one of the easiest methods to measure differences between two stereo images. The sum of absolute differences is defined as follows:

$$SAD(x,y,d) = \sum_{(i,j)\in W(x,y)}^{N} \left| I_L(i,j) - I_R(i-d,j) \right| \tag{3}$$

where I_L and I_R are functions of density pixels of left and right images. W(x, y) is a square window that surrounds the position of pixel (x, y).

The calculation of disparity by SAD (x, y, d) is repeated with x-coordinate frame in the line in the image, which is defined by 0 and the maximum possible disparity dmax of the scanned 3D image. The disparity is the inverse of the distance, therefore it can be easily transformed to a depth map.

3.3 Creation of mesh model

In this step of the proposed algorithm, the detection of vertices is performed. A window filter function was used to detect and consequently to extract the vertices. This filtering was used gradually through the rows and then through the columns. Finally, coordinates of vertices were obtained by logical multiplication of previous mentioned steps. After the extraction of vertices, 3D visualization of all vertices in a 3D space was realized. The next step should involve extraction of faces by connection of all adjacent vertices. A polygonal mesh model can be generated from coordinates of all vertices and faces.

3.4 Description of mesh model

We proposed a mesh model, which will be described by a 3D SURF (Speeded Up Robust Features) descriptor. This descriptor involves a 2D surf descriptor with a special extension. 2D surf is based on sums of 2D Haar wavelet responses [7]. The extension to the third dimension in the first step includes voxelization of shape into a 3D cube using the intersection of faces with the grid-bins. Next the saliency for each grid-bin as determinant of the Hessian matrix is calculated. In the second step, a rotation and scale-invariant 3D SURF descriptor is computed around each point of interest obtained in previously mentioned steps [8].

4. Experiments

At the beginning of our experiments, image preprocessing was performed. Consequently, the algorithm for background subtraction, the algorithm for detection object shadows minimization, the noise and undesirable blobs were adjusted. In experiments, various threshold settings were tested.

Fig. 5. Input images from cam1 and cam2 at the distance of 5m

Input images from camcorders were pre-processed and next the object detection and extraction were performed. After this detection, objects were associated into several groups based on vehicle type. A disparity map was generated by a stereo matching algorithm. The created disparity map was transformed into a depth map and consequently 3D objects were extracted.

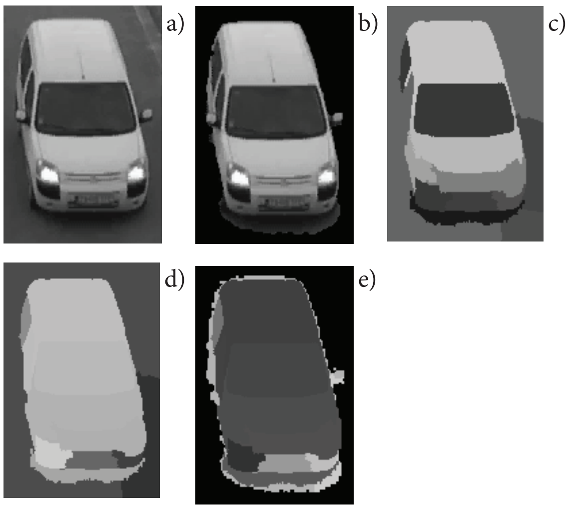

Fig. 6. Whole process of image processing: a) detected object, b) extracted object, c) segmented object, d) disparity map of the object, e) depth map without background

The whole process from the object detection to creation of a depth map is shown in Figure 6. In the next step of our proposed algorithm, the depth map was filtered and adjusted in order to improve a 3D representation of 3D objects. Mathematical operations such as morphological opening and closing were performed. Figure 7 presents the process of filtering, interpolation and final form of the depth map.

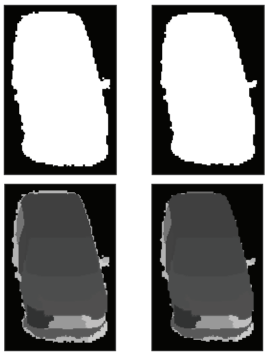

Fig. 7. **Adjustment and optimization of the depth map: a) origin contour of depth map, b) processed contour of depth map, c) origin of depth map, d) filtered depth map**

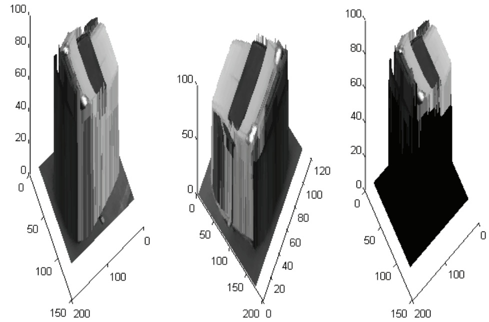

Fig. 8. **3D visualization of 3D model: a) view from the right side, b) view from the left side, c) view from the right side – the model with a removed background**

After optimization of the depth map, the texture from the input image was applied to the depth map. Next, the extraction of notable points from the depth map/object was performed. A window filtering function that checks values surrounding pixels around investigated pixel gradually through overall image is applied Final form achieved by logical multiplication of previous filtering was obtained. Thus extracted vertices from depth map and their 3D visualizations are shown in Figure 9.

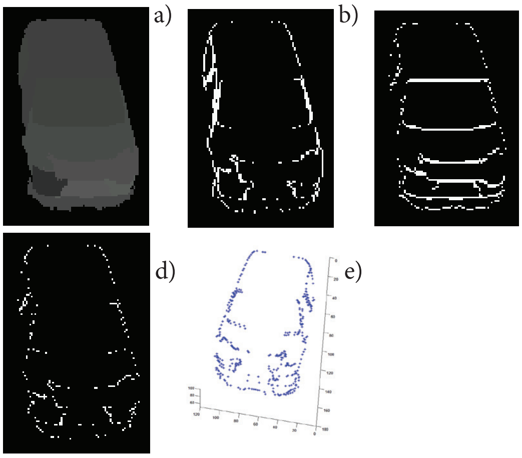

Fig. 9. **Vertices extraction: a) depth map, b) vertices extraction in rows, c) vertices extraction in columns d) final vertices extraction e) 3D visualization of vertices**

After vertices extraction, we have to connect all the vertices together. From the defined coordinates of vehicles and faces a 3D mesh model of a 3D object can be generated. This model can be described by a 3D descriptor that is mentioned in the previous part of this contribution.

Table 1. **Results of a 3D object detection and modelling from the traffic flow**

Vehicle class	Displayed objects	Detected and extracted objects	Correct creation of depth map	Correct extraction of vertices	Success rate (%)
General	20	18	15	15	75

In experiments, a comparison of the number of successful detected objects, the number of correct creations of the object depth map and the number of correct vertices extraction from the depth map was performed.

5. Conclusion

In this paper, a new approach to image processing for intelligent transportation systems, which consists of four main parts – the object detection and creation of 3D objects, the creation of a 3D model and the creation of a mesh model has been proposed. In these experiments, 20 different vehicles from a real traffic flow were tested. The overall success rate of correct extraction of vertices of a 3D vehicle model achieved 75 %. In the future, we want to focus on improving the creation of mesh model and also the description of general mesh model by a 3D descriptor. Moreover, the data analysis based on 3D descriptors of mesh models by a learning algorithm for classification will be used.

Acknowledgments

This paper was supported by the operational program "Research and development" of ASFEU-Agency under the project "Nové metódy merania fyzikálnych dynamických parametrov a interakcií motorových vozidiel, dopravného prúdu a vozovky". ITMS 26220220089 OPVaV-2009/2.2/03-SORO co financed by European fond of regional development.

Bibliography

[1] Directorate-General for Research, European Union: Intelligent transport systems, http://www.greendigitalcharter.eu/wp-content/uploads/2012/11/2010-European-Commission-Report-on-Intelligent-Transport-Systems.pdf, 2010

[2] KAMATH, A.: Image Processing and Particle Analysis for Road Traffic Detection, International Journal of Computer Applications (0975–8887) Volume 55–No.2, October 2012, http://research.ijcaonline.org/volume55/number2/pxc3882604.pdf

[3] GHOSH, N.: Unsupervised Learning for Incremental 3-D Modeling, IEEE Transactions on intelligent transportation systems, vol 11, no. 2, 2010 http://vislab.ucr.edu/RESEARCH/sample_research/AAAIworkshop'05 final paper.pdf, 0

[4] WOJEK, CH.: Monocular 3D Scene Modeling and Inference: Understanding Multi-Object Traffic Scenes, Lecture Notes in Computer Science Volume 6314, 467-481, 2010

[5] KAMENCAY, P.: Improved Depth Map Estimation from Stereo Images Based on Hybrid Method, Radioengineering 2012/4, Volume 21, Issue 1, Pages: 70-78, Part 1, http://www.radioeng.cz/fulltexts/2012/ 12_01_0070_ 0078.pdf

[6] NAVAB, N., UNGER, CH.: Stereo Vision II: Dense Stereo Matching, 2011 http://campar.in.tum.de/twiki/pub/Chair/TeachingWs11Cv2/3D_CV2_WS_2011_Stereo.pdf

[7] BAY, H., ESS, A., TUYTELAARS, T., VAN GOOL, L.: Speeded-up robust features (surf) Computer Vision and Image Understanding 110, 2008, 346–359

[8] KNOPP, J., PRASAD, M.: Hough Transforms and 3D SURF for robust three dimensional classication, Lecture Notes in Computer Science Volume 6316, Pages 589-602, 2010, http://link.springer.com/content/pdf/10.1007%2F978-3-642-15567-3_43.pdf