AN ALGORITHM FOR PEDESTRIAN DETECTION IN MULTISPECTRAL IMAGE SEQUENCES

V. V. Kniaz[a,b], V. V. Fedorenko[a] *

[a] State Res. Institute of Aviation Systems (GosNIIAS), Moscow, Russia - (vl.kniaz, vfedorenko)@gosniias.ru
[b] Moscow Institute of Physics and Technology (MIPT), Russia

Commission II, WG II/5

KEY WORDS: pedestrian detection, optical flow, thermal vision, mobile robots

ABSTRACT:

The growing interest for self-driving cars provides a demand for scene understanding and obstacle detection algorithms. One of the most challenging problems in this field is the problem of pedestrian detection. Main difficulties arise from a diverse appearances of pedestrians. Poor visibility conditions such as fog and low light conditions also significantly decrease the quality of pedestrian detection. This paper presents a new optical flow based algorithm BipedDetet that provides robust pedestrian detection on a single-borad computer. The algorithm is based on the idea of simplified Kalman filtering suitable for realization on modern single-board computers. To detect a pedestrian a synthetic optical flow of the scene without pedestrians is generated using slanted-plane model. The estimate of a real optical flow is generated using a multispectral image sequence. The difference of the synthetic optical flow and the real optical flow provides the optical flow induced by pedestrians. The final detection of pedestrians is done by the segmentation of the difference of optical flows. To evaluate the BipedDetect algorithm a multispectral dataset was collected using a mobile robot.

1. INTRODUCTION

Pedestrian detection algorithms developed to date could be composed in a vast list. Such algorithms range from classical computer vision methods based on the feature detection to modern deep convolutional neural networks based methods. Modern algorithms such as DetectNet deep neural network provide performance comparable with a human operator. However such methods require a powerful GPU for a real-time detection. Thus a different approach is required to provide a robust performance on a system with a limited computational resources such as a mobile robot.

This paper presents a new optical flow based algorithm BipedDetect that provides robust pedestrian detection on a single-borad computer. Optical flow based methods prove to be a fast and robust solution for the problem of moving pedestrian detection. As the optical flow detection is based on motion it is robust against unknown pedestrian appearance. However most of optical flow estimation algorithms provide poor performance under low light conditions. The BipedDetect algorithm is designed to work with multispectral image sequences to improve the performance in degraded visual environments.

1.1 Related work

The following section presents related work on pedestrian detection. To analyse the requirements to a pedestrian detection dataset an overview over existing datasets for pedestrian detetction algorithms is provided.

The problem of pedestrian detection has received a lot of scholar attention. The first approaches to detect obstacles and pedestrians are dated to 2000s. In the paper by (Zheltov and Sibiryakov, 2000) a method of 3D-object detection based on orthophoto difference

analysis is proposed. The main idea of the detection method is as follows. If an analytical surface model is known, then orthogonal projections of the given surface are created on some convenient plane using the left and right stereopair images. The calculation of the difference of such projections leads to the appearance of characteristic geometric structures in the vicinity of 3D objects that do not belong to a given surface. Being relatively simple the differential methods prove to be a robust solution for object detection (Kniaz, 2014).

In (Dollár et al., 2010) a multiscale pedestrian detector operating in near real time (□6 fps on 640x480 images) with state-of-the-art detection performance was presented. To reduce the computational time of the detector it is proposed to approximate the feature responses of a broad family of features, including gradient histograms, using their feature responses at a single scale. The results have showed about 6 fps for detecting pedestrians at least 100 pixels high and 3 fps over 50 pixels.

Interesting method of pedestrian detection was published in (Shirazi and Morris, 2016). Suggested system of detection and tracking works only with necessary area including typical pedestrian paths in the frame. Tracking system uses two methods, optical flow and bipartite graph matching of detections. The presence of typical ways of pedestrians helps to remove false tracks.

Stereo matching and optical flow estimation benchmark was presented in (Geiger et al., 2012). It comprises 194 training and 195 test image pairs at resolution of 1240 x 376 pixels with semi-dense ground truth. 3D visual odometry dataset consists of 22 stereo sequences. All the objects were manually labelled in 3D point clouds.

A big multispectral pedestrian dataset was introduced in (Hwang et al., 2015). This dataset includes thermal image sequences of regular traffic scene as well as color image sequences. It has both

---

* Corresponding author

are color and thermal image pairs and contains the nighttime traffic sequences which are rarely provided.

An interesting approach for an optical flow dataset generation was proposed by (Menze and Geiger, 2015). It includes two steps: firstly, the static background of the scene is recovered by removing all dynamic objects and compensating for the vehicle's egomotion. After that, dynamic objects are inserted in the scene by fitting their detailed CAD models to the point clouds for each frame.

Recently a number of large datasets for pedestrian detection such as the Caltech pedestrian dataset (Dollár et al., 2009, Dollár et al., 2012b) and various datasets in KITTI Vision Benchmark Suite (Geiger et al., 2012) were developed. However no multispectral datasets that include both pedestrian labelling and ground truth optical flow were developed to date. To evaluate the developed algorithm a new dataset BipedFlow was generated.

2. DATASET

To provide a reliable basis for analysis of the developed algorithm a multispectral dataset was developed. A brief description of the dataset is presented in the following section.

2.1 Dataset design

The dataset was collected using a Hercules mobile robot equipped with a FLIR ONE thermal imaging camera. The robot was moved along various trajectories: a straight line, a sinusoid and a curve. To generate the ground truth optical flow a 3D modelling was used. Models of scenes in which the robot moved were generated using laser scanning. The internal orientation of the camera was recovered using the method proposed by (Gschwandtner et al., 2011). The dataset includes 20 sequences with an average duration of 30 seconds. The ground truth optical flow was generated using ray tracing (Wulff et al., 2012) in Blender 3D creation suite.

Scenario	Trajectory	Illumination	# frames
Pedestrian stands Robot stopped	Straight line	Day	500
Pedestrian stands Robot moves	Straight line	Day	500
Pedestrian walks Robot stopped	Straight line	Day	500
Pedestrian walks Robot moves	Straight line	Day	500

Table 1. Scenarios, trajectories and illumination conditions used in the BipedDetect dataset.

2.2 Mobile robot

To prepare the test sample, a wheeled robot equipped with a multispectral vision system was used. The robot is based on the Hercules mobile platform and a Raspberry Pi single-board computer (figure 1). The FLIR ONE multispectral camera was used as a video sensor. The camera is connected to a smartphone, that transfers the video sequence to a remote computer. The resolution of the camera in the visible range is 640x480 pixels. The resolution in the infrared range is 160x120 pixels. Complete technical specifications of the FLIR ONE camera are presented in table 2. The robot is controlled from a remote computer via a WiFi network using a dedicated protocol.

Figure 1. The mobile robot with the FLIR ONE multispectral camera

FLIR ONE Specifications	
Scene Temperature Range	-20° C to 120° C
Operating Temperature	0° C to 35° C
Weight	29 grams
Dimensions (L x W x H)	72 mm x 26 mm x 18 mm
Sensitivity	0.1° C
Visible Camera	640x480 pixels
Thermal Camera	320x240 pixels
Field of view	46°x35°

Table 2. FLIR ONE specifications

3. BIPEDDETECT ALGORITHM

The BipedDetect algorithm is designed to work with multispectral image sequences to provide a reliable performance under low light conditions. The optical flow is estimated by phase correlation in the frequency domain. The correlation is performed separately in visible and infrared channels. Thus two optical flow images based on infrared and visible channels are produced. The image fusion is based on a correlation coefficient that is provided for each pixel by a phase correlation.

3.1 Coordinate systems

Four coordinate systems are defined. Object coordinate system $O_oX_oY_oZ_o$ is related to some object of interest in the observed scene and defined as follows: $O_oX_oZ_o$ plane is normal to gravity vector, the Y_o axis is normal to X_o, Z_o axes. The point of origin is related to some point of the observed scene and is selected appropriately for a given problem (figure 2).

The origin of the image coordinate system $O_iX_iY_iZ_i$ is located in the upper left pixel of the image, the X_i is directed to the right, the Y_i axis is directed downwards.

The origin of the robot's coordinate system $O_rX_rY_rZ_r$ is located in the center of the upped deck of the robot. The Y_r axis is directed towards forward motion of the robot, the Z_r axis is normal to the upper deck of the robot.

The origin of the camera coordinate system $O_cX_cY_cZ_c$ is located in the perspective center, the X axis is collinear with the X_i axis, the Y_c axis is collinear with the Y_i axis, the Z_c axis is normal to X_c and Y_c axes. The rotation of the camera coordinate system with respect to object coordinate system is defined using rotation matrix $\mathbf{R}_{oc}$:

$$\mathbf{R}_{oc} = \mathbf{R}_{\varkappa} \cdot \mathbf{R}_{\omega} \cdot \mathbf{R}_{\alpha}, \qquad (1)$$

where $\mathbf{R}_{\alpha}$ – rotation matrix around the axis Y, $\mathbf{R}_{\omega}$ – rotation matrix around the axis X, $\mathbf{R}_{\kappa}$ – rotation matrix around the axis Z.
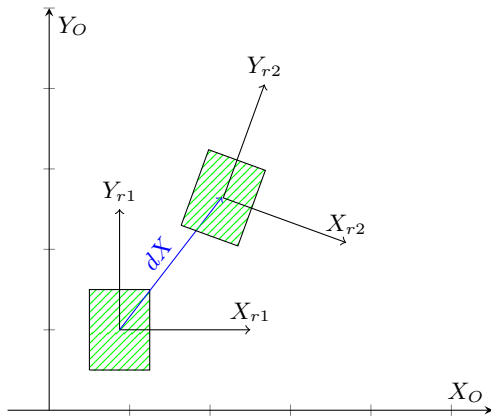

Figure 2. Coordinate systems

3.2 Algorithm pipeline

The algorithm consists of four main steps. Firstly optical flow is estimated using infrared and visible spectrum image sequences. Secondly robot's motion is estimated using the data from robot's encoders and the state-space model. The block diagram of the algorithm is presented in figure 3.

3.3 Multispectral optical flow fusion

The estimation of multispectral optical flow is performed separately for images in visible spectrum and infrared images. Various optical flow estimations algorithms were evaluated (Revaud et al., 2015, Liu, 2009, Farnebäck, 2003, Dosovitskiy et al., 2015). The FlowNet deep neural network based algorithm showed the best performance during the evaluation of the algorithms using the ground truth optical flow. However only computational complexity of the algorithm proposed in (Farnebäck, 2003) was low enough for implementation on a single-board computer.

Various methods of optical flow fusion were studied. Firstly the most simple fusion model based on the selection of maximum flow among the channels were considered. Let $\mathbf{u}_{TV} = (u_{TV}, v_{TV})$ be the visible spectrum optical flow, $\mathbf{u}_{IR} = (u_{TV}, v_{TV})$ infrared optical flow. Then the maximum fusion method is given by

$$u = \max(u_{TV}, u_{IR}), v = \max(v_{TV}, v_{IR}) \qquad (2)$$

Also fusion method based on the mean value of the optical flows was studied. The mean fusion method is given by

$$u = \frac{(u_{TV} + u_{IR})}{2}, v = \frac{(v_{TV} + v_{IR})}{2} \qquad (3)$$

Finally the adaptive relay threshold method introduced in (Man et al., 2007) was studied. The fused optical flow for the relay method is given by

$$A_i = \sum_{j \in N} (u_{TVj}^2 + v_{TVj}^2)(u_{IRj}^2 + v_{IRj}^2) \qquad (4)$$
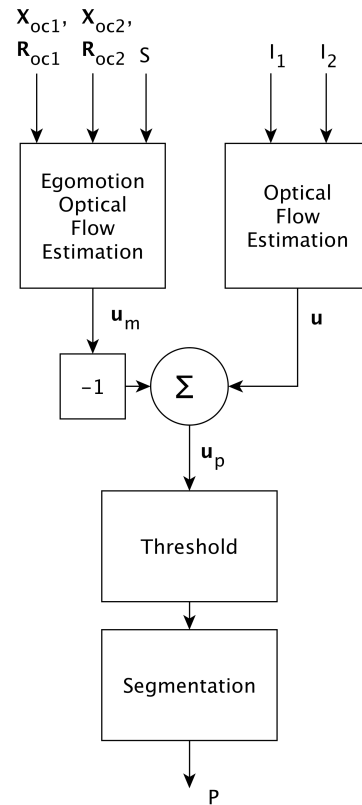

Figure 3. Block diagram of the BipedDetect algorithm

$$u_i = \begin{cases} max(u_{TV}, u_{IR}) & if\, A_i \geq T, \\ min(u_{TV}, u_{IR}) & if\, A_i < T \end{cases}$$

All three methods were implemented in the dedicated software and their performance was studied using the BipedFlow dataset. The evaluation have showed that the fusion based on the maximum has the minimal root mean square error.

3.4 State-space model of the mobile robot

A state-space model of the Hercules robot was developed during previous research (Kniaz, 2015, Kniaz, 2016). The state-space dynamic model of the robot is given by:

$$\dot{\mathbf{X}}(t) = \mathbf{A} \cdot \mathbf{X}(t) + \mathbf{B} \cdot \mathbf{U}(t) \qquad (5)$$

$$\mathbf{Y}(t) = \mathbf{C} \cdot \mathbf{X}(t) + \mathbf{D} \cdot \mathbf{U}(t) \qquad (6)$$

Where output vector $Y(t)$ of the robot consists of longitudinal speed v_x and rotational speed $\dot{\vartheta}$:

$$\mathbf{Y}(t) = \begin{bmatrix} v_x \\ \dot{\vartheta} \end{bmatrix}$$

The input vector $U(t)$ consists of average and difference motor speeds for left and right wheel motors:

$$\mathbf{U}(t) = \begin{bmatrix} a \\ d \end{bmatrix}$$

$$a = \frac{p_R + p_L}{2} \qquad (7)$$

Algorithm 1: BipedDetect

Input: Camera external orientation parameters for each frame $\mathbf{X}_{oc1}, \mathbf{X}_{oc2}, \mathbf{R}_{oc1}, \mathbf{R}_{oc2}$, images in visible and infrared spectrum $\mathbf{I}_{TV}^i, \mathbf{I}_{IR}^i$.

Output: Bounding boxes for detected pedestrians.

for $i := 1$ *to* N_{frames} **do** *Optical flow loop*

   Estimate optical flow $\mathbf{u}_{TV}^i(\mathbf{I}_{TV}^{i-1}, \mathbf{I}_{TV}^i), \mathbf{u}_{IR}^i(\mathbf{I}_{IR}^{i-1}, \mathbf{I}_{IR}^i)$;

   Perform the fusion of $\mathbf{u}_{TV}^i$ and $\mathbf{u}_{IR}^i$ using eq. (2) to obtain $\mathbf{u}^i$;

   Generate the synthetic optical flow $\mathbf{u}_m^i$ using eq. (9) and $\mathbf{X}_{oc1}, \mathbf{X}_{oc2}, \mathbf{R}_{oc1}, \mathbf{R}_{oc2}$;

   $\mathbf{u}_d^i = \mathbf{u}^i - \mathbf{u}_m^i$;

   Threshold $\mathbf{u}_d^i$ using Otsu's method to obtain $\mathbf{u}_t^i$;

   Perform the segmentation of $\mathbf{u}_t^i$ to obtain the bounding boxes for moving pedestrians;

end

$$d = \frac{p_R - p_L}{2} \qquad (8)$$

Where p_R, p_L - are motor speeds for right and left wheel motors.

3.5 Synthetic optical flow generation

The synthetic optical flow is generated using ray tracing. Let $\mathbf{x}_p^i$ be a projection of a point of the scene in i-th frame. Then the optical flow is given by

$$\mathbf{u}_m^i = \mathbf{x}_p^i - \mathbf{x}_p^{i-1}, \qquad (9)$$

where $\mathbf{x}_p^i$ is given by

$$\mathbf{x}_p^{i-1} = \mu^{i-1} \cdot \mathbf{R}_{oc}^{i-1} \cdot (\mathbf{X}_p - \mathbf{X}_0^{i-1}), \qquad (10)$$

$$\mathbf{x}_p^i = \mu^i \cdot \mathbf{R}_{oc}^i \cdot (\mathbf{X}_p - \mathbf{X}_0^i), \qquad (11)$$

where $\mathbf{X}_p - $ 3D coordinate of a point of the scene, $\mu^i - $ scale factor, $\mathbf{X}_0^i - $ position of the camera. The difference of two last equations provides the final equation for the synthetic optical flow $\mathbf{u}_m^i$.
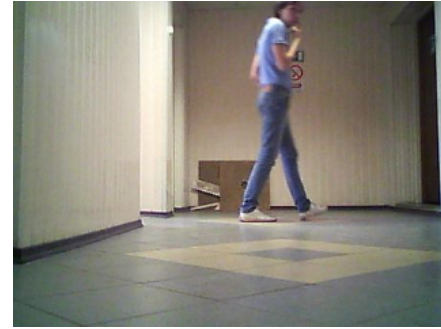
3.6 Pedestrian detection

With the given synthetic optical flow $\mathbf{u}_m^i$ and a real optical flow $\mathbf{u}^i$ their difference $\mathbf{u}_d^i$ is calculated

$$\mathbf{u}_d^i = \mathbf{u}^i - \mathbf{u}_m^i. \qquad (12)$$

As $\mathbf{u}_d^i$ doesn't contain optical flow induced by robot's egomotion pedestrians could be detected by thresholding the image using Otsu's method. The final bounding boxes for the pedestrians are provided by segmentation. The BipedDetect algorithm is summarised in the algorithm 1.

4. ALGORITHM EVALUATION

The algorithm was evaluated using the generated BipedFlow dataset. An example of pedestrian detection by BipedDetect algorithm is presented in figure 4.

(a) Visible spectrum image

(b) Infrared image

(c) Model optical flow

(d) Estimated optical flow

(e) Difference of (c) and (d)

Figure 4. Example of pedestrian detection using BipedDetect algorithm

A modified PASCAL metric (Everingham et al., 2009) proposed in (Dollár et al., 2012a) was used. The log-average miss rate for BipedDetect algorithm was 52% during the daylight and 43% during the night. The computational complexity of the algorithm is suitable for the realisation on a single-board computer with a CPU of 1.2 GHz for input images with the resolution of 160x120. Evaluation of the algorithm proved that it is robust against low light and high noise level.

5. CONCLUSION

The BipedDetect algorithm for pedestrian detection on mobile robots was developed. The algorithm is based on estimation of a multispectral optical flow and a synthetic optical flow. The multispectral optical flow is generated by fusion of the optical flow estimated for visible spectrum channel and infrared channel. Various optical flow fusion methods were evaluated. The evaluation proved that the best performance is obtained by the fusion based on the maximum of the optical flow.

A synthetic optical flow is generated using ray tracing and a simplified model of the scene. To estimate an interframe motion of the camera a state-space model is used. The difference of the synthetic optical flow and the multispectral optical flow provides the optical flow induced by the moving pedestrians and doesn't include optical flow induced by the camera's egomotion. The final detection of the pedestrians is performed by thresholding the difference of multispectral and synthetic optical flows and segmentation of the resulting optical flow.

To evaluate the performance of the BipedDetect algorithm a BipedFlow multispectral optical flow dataset was developed. It was generated using a mobile robot equipped with a multispectral camera. The BipedFlow dataset includes various scenarios captured during the daylight and in low light conditions. The evaluation of the BipedDetect algorithm proved that it is robust against low light level. The log-average miss rate for BipedDetect algorithm was 52% during the daylight and 43% during the night. The computational complexity of the algorithm is suitable for the realisation on a single-board computer with a CPU of 1.2 GHz for input images with the resolution of 160x120.

6. ACKNOWLEDGEMENTS

REFERENCES

Dollár, P., Belongie, S. and Perona, P., 2010. The Fastest Pedestrian Detector in the West. In: *Procedings of the British Machine Vision Conference 2010*, British Machine Vision Association, pp. 68.1–68.11.

Dollár, P., Wojek, C., Schiele, B. and Perona, P., 2009. Pedestrian detection: A benchmark. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 304–311.

Dollár, P., Wojek, C., Schiele, B. and Perona, P., 2012a. Pedestrian Detection - An Evaluation of the State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell. ()*.

Dollár, P., Wojek, C., Schiele, B. and Perona, P., 2012b. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell. ()* 34(4), pp. 743–761.

Dosovitskiy, A., Fischery, P., Ilg, E., usser, P. H., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D. and Brox, T., 2015. FlowNet: Learning Optical Flow with Convolutional Networks. *2015 IEEE International Conference on Computer Vision (ICCV)* pp. 2758–2766.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A., 2009. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* 88(2), pp. 303–338.

Farnebäck, G., 2003. Two-Frame Motion Estimation Based on Polynomial Expansion. *SCIA* 2749(Chapter 50), pp. 363–370.

Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. *CVPR* pp. 3354–3361.

Gschwandtner, M., Kwitt, R., Uhl, A. and Pree, W., 2011. Infrared camera calibration for dense depth map construction. *Intelligent Vehicles Symposium*.

Hwang, S., Park, J., Kim, N., Choi, Y. and Kweon, I. S., 2015. Multispectral pedestrian detection - Benchmark dataset and baseline. *CVPR* pp. 1037–1045.

Kniaz, V. V., 2014. A Fast Recognition Algorithm for Detection of Foreign 3D Objects on a Runway. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-3, pp. 151–156.

Kniaz, V. V., 2015. Fast instantaneous center of rotation estimation algorithm for a skied-steered robot. *SPIE Optical Metrology* 9528, pp. 95280L–95280L–11.

Kniaz, V. V., 2016. Real-time optical flow estimation on a GPU for a skied-steered mobile robot. In: N. Kehtarnavaz and M. F. Carlsohn (eds), *SPIE Photonics Europe*, SPIE, pp. 989706–989706–12.

Liu, C., 2009. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. PhD thesis, Massachusetts Institute of Technology.

Man, H., Holt, R. J., Wang, J., Martini, R., Netravali, R. and Mukherjee, I., 2007. A new optical flow estimation method in joint EO/IR video surveillance. In: G. C. Holst (ed.), *Defense and Security Symposium*, SPIE, pp. 65430P–11.

Menze, M. and Geiger, A., 2015. Object scene flow for autonomous vehicles. *CVPR* pp. 3061–3070.

Revaud, J., Weinzaepfel, P., Harchaoui, Z. and Schmid, C., 2015. EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow. *CoRR abs/1501.02565* cs.CV, pp. 1164–1172.

Shirazi, M. S. and Morris, B. T., 2016. Vision-based pedestrian behavior analysis at intersections. *Journal of Electronic Imaging* 25(5), pp. 051203–14.

Wulff, J., Butler, D. J., Stanley, G. B. and Black, M. J., 2012. Lessons and Insights from Creating a Synthetic Optical Flow Benchmark. In: *Computer Vision – ECCV 2012. Workshops and Demonstrations*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 168–177.

Zheltov, S. and Sibiryakov, A., 2000. Method of 3d-object detection based on orthophoto difference analysis. *International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B3. Amsterdam 2000.* pp. 1041–1047.