An Obstacle Detection Method Based on Binocular Stereovision

Yihan Sun[1], Libo Zhang[2(✉)], Jiaxu Leng[1], Tiejian Luo[1], and Yanjun Wu[2]

[1] University of Chinese Academy of Sciences, Beijing 101408, China
[2] Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
zsmj@hotmail.com

Abstract. As the main tasks of Advance Driver Assistance Systems (ADAS), obstacle detection has attracted extensive attention. Traditional obstacle detection methods on the basis of monocular vision will lose its effect when new obstacles appear or the obstacles have severe occlusion and deformation, so this paper proposes an obstacle detection method based on disparity map, which can detect all obstacles on the road accurately. We first determine the disparity of the road in V disparity map through an approach based on weighted least square method. Then we obtain the disparity map contains only the obstacles on the road, and generate corresponding Real U disparity map by projection. Finally, obstacles are detected in Real U disparity map. Experiments show that the proposed method can not only precisely detect the obstacles at greater distances and the obstacles with a large area of occlusion, but also accurately calculate the distance information according to the disparity of the obstacle.

1 Introduction

In recent years, computer vision has made breakthroughs in many fields such as saliency detection [1,2], target tracking [3] and so on. Obstacle detection can improve the safety in driving and reduce traffic accidents as the main tasks of ADAS. Now obstacle detection methods based on vision have attracted wide attention of researchers, for its low cost and the ability to provide a wealth of scene information. In the past few years, different kinds of methods have been proposed. One commonly used method is to segment the image into image fragments, and then employ the similarity measure to classify these fragments. Some methods based on exhaustive search according to object characteristics, or contour information have also been put forward. At present, R-CNN is the most popular method in the field of monocular detection, which extracts feature by CNN, and then utilizes SVM to classify the feature vectors. Deep MANTA train the obstacle model through three layers of network, and then analyze the input image so as to detect the obstacles.

Above methods are based on monocular camera. Although this kind of methods have already made great progress, there are still two problems. (1) These methods implement the process of data acquisition and model training

B. Zeng et al. (Eds.): PCM 2017, LNCS 10736, pp. 571–580, 2018.
https://doi.org/10.1007/978-3-319-77383-4_56

in advance, therefore, once the obstacles in the scene are seriously blocked or the new obstacles appear, the detection algorithm will be invalid. (2) Monocular methods are difficult to calculate the distance of the obstacle accurately, for the reason that its calculation depends on the recognition of objects. In view of the above problems, obstacle detection methods based on binocular vision are beginning to emerge. Labayrade [4] proposed the V-disparity concept aimed at simplifying the process of separating obstacles from road surfaces. Therefore, extraction of 3D road surface and obstacles will be simplified as 2D linear extraction [5]. Hu proposed an U-V-disparity concept based on Labayrade's work, to classify the 3D road scene into relative surface planes and characterize the features of road, roadside, and obstacles [5]. The obstacle detection method based on U-V disparity map can detect obstacles accurately without training the model, and the change of the obstacles'shape does not affect the results.

Inspired by this method, we propose a fast obstacle detection method based on binocular stereovision. The main contributions are as follows: (1) In V-disparity map, a filtering method based on road features is proposed. (2) In V-disparity map, a road extraction method on account of the Hough transformation based on least square method is put forward. (3) For the problem that the obstacles at greater distances are difficult to be detected, we come up with a method to generate Real U disparity map, which can restore two-dimensional information to real spatial information. (4) To solve the problem of the deviation of the disparity, we raise a fast and effective method based on Real U disparity map.

2 Binocular Stereovision Model

Figure 1(a) is the geometric model of binocular stereovision model, which consists of three coordinate systems R_w, R_{cr} and R_{cl}. They represent the world coordinate system established with the ground level, right camera and left camera respectively. θ is the angle between the optical axis of the cameras and the ground level. h represents the height of the cameras above the ground. b is the distance between the cameras (baseline length).

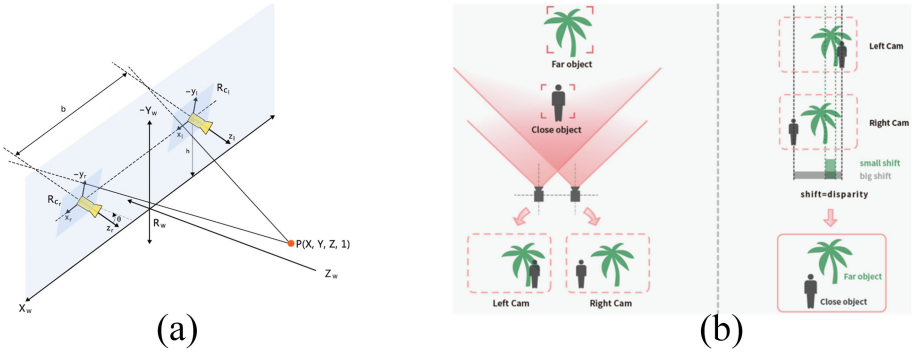

Fig. 1. (a) The model of binocular stereovision. (b) The generation of disparity.

In R_w, if a point is $P(X, Y, Z)$, its corresponding representation in homogeneous coordinate is $P(X, Y, Z, 1)$. Then, we can get the homogeneous coordinate of the imaging positions of point P in cameras and denote it as $p_i(u_i, v_i, 1)$:

$$\lambda \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = MT_iR \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (i = l \text{ or } r) \tag{1}$$

where λ is scale factor, M is parameter matrix inside camera, T is translation matrix, and R is rotation matrix. u_i and v_i is the coordinate of p_i.

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$T_i = \begin{bmatrix} 1 & 0 & 0 & k_i\frac{b}{2} \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

In Eq. (3), k_i is -1 if i equal to r, otherwise is 1.

The parameter matrix M in Eq. (1) is defined as:

$$M = \begin{bmatrix} f_u & 0 & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

in which (u_0, v_0) is the central coordinate of the image, f_u and f_v represent the focal length in u and v direction respectively. According to Eqs. (1), (2), (3) and (4), we can generate the imaging positions of P: (u_l, v_l) and (u_r, v_r).

$$\begin{cases} v_{l,r} = \frac{(f_v\cos\theta + v_0\sin\theta)(Y+h) + (v_0\cos\theta - f_v\sin\theta)Z}{Z\cos\theta + (Y+h)\sin\theta} \\ u_{l,r} = u_0 + f_u\frac{X \pm b/2}{Z\cos\theta + (Y+h)\sin\theta} \end{cases} \tag{5}$$

As shown in Fig. 1(b), we can calculate the disparity of P by Eq. (5).

$$\Delta = u_l - u_r = \frac{f_u b}{Z\cos\theta + (Y+h)\sin\theta} \tag{6}$$

Equations (5) and (6) describe the basic relationship between image coordinate (u, v) and disparity Δ, which can be simply represented by the following equations:

$$\begin{cases} u = \Phi_u(X, \Delta) \\ v = \Phi_v(Z, \Delta) \end{cases} \tag{7}$$

Particularly, $Y = 0$ denotes the road surface, and we can generate Eq. (8):

$$v\cos\theta = \frac{h}{b}\Delta - f_v\sin\theta \tag{8}$$

3 The Proposed Method

The first step of obstacle detection is the determination of the road area. Most of the existing monocular based methods require huge amount of computation and are difficult to meet the requirements of real-time. Therefore, we apply the method based on binocular stereovision to determine the road in V-disparity map. We propose a novel method on account of the Hough transformation based on weighted least square method, which can accurately remove the disparity representing the road and outside the road. Besides, the detection of obstacles at greater distances is always a difficulty, for the reason that the disparity points of the those obstacles are few in number and not continuous, thus the missed detection and false detection are much more apt to occur. To solve this problem, we put forward a method to generate Real U disparity map, which can restore two-dimensional information to real spatial information.

Figure 2 demonstrates the flow diagram of our algorithm. The algorithm starts with calculating the V disparity map according to the disparity map obtained by stereo matching, then extracting the road in V disparity map, and removing the disparity representing the road and outside the road. Next, U disparity map is generated via projection, and then mapped to the real space x-z to create Real U disparity map. Finally, the detection is completed in the Real U disparity map. This method is simple and effective, and is characterized by low computation workload and fast speed due to the use of the disparity map based on edge. In addition, it can detect obstacles at greater distances, because of the application of the convertion from disparity map to the real space.

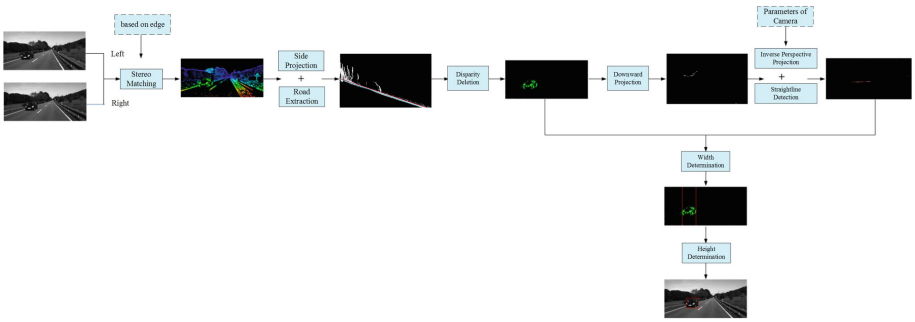

Fig. 2. Algorithm flowchart.

3.1 The Determination of Detection Area

Through the binocular stereovision model, the detection area of the real scenes can be projected to the corresponding disparity map. Taking the center of the center line of the cameras as the origin of coordinates, the detection area of real space is determined by four points: $(X_{left}, 0, Z_{min})$ $(X_{left}, 0, Z_{max})$

$(X_{right}, 0, Z_{min})$ $(X_{right}, 0, Z_{max})$. Do calculation according to the left two boundary points $(X_{left}, 0, Z_{min})$ and $(X_{left}, 0, Z_{max})$, and according to Eq. (9).

$$\begin{cases} \frac{Z}{b/2-x} = \frac{f}{u-w/2} \\ d = \frac{bf}{Z} \end{cases} \tag{9}$$

where w is the width of the image, Z is the real distance, b is the baseline length, x is the true horizontal distance, f is the focal length, u is the column of the image, and d is the disparity. Thus we can obtain a $u-d$ equation what represents the left boundary:

$$d = \frac{b(u-w/2)}{b/2-x} \tag{10}$$

Similarly, the right boundary equation can be computed. Now the transformation from the real space point (x, z) to (u, d) is achieved. According to the linear equations of left boundary and right boundary, we calculate the boundary disparity of each column in the disparity map. If the disparity of the current column is less than the disparity of road boundary, we set it to be zero to obtain a sub-disparity map containing only the disparity of detection area.

3.2 Road Extraction in V-disparity Map

The V disparity map can be obtained by function $G(D_\Delta) = (D_{v\Delta})$, which counts the number of the same disparity in each row of disparity map. D_Δ is disparity map and $D_{v\Delta}$ is V disparity map, in which the abscissa is disparity, while the ordinate corresponds to the original disparity map. The gray value of each pixel $D_{v\Delta}(v, d)$ is expressed as the number of the pixels with d disparity in the Vth row in original disparity map. To better detect the road, we design a morphological filtering method to filter the V disparity map and enhance the road.

The road in real scenes is usually an arch with lower sides and higher middle, instead of a flat shape. So, it is represented as a long strip with circular shape in V disparity map, and generally the approximately centered line in this strip area is considered as the expression of the road. The drawback of the Hough transformation is that only the straight line through the most points is taken into account. Thus, we put forward an improved road extraction method based on Hough transformation, which uses the least squares method to reduce the impact of noise. Specifically, the pixel points are weighted by the following methods:

$$w(d, v) = \alpha(1 - \frac{v}{h}) + \beta \begin{cases} \frac{n}{(1+\exp(-n))(1-\exp(-d))w} & d > T \\ 1 - \exp(\frac{-n}{wd}) & d \leq T \end{cases} \tag{11}$$

in which $w(d, v)$ denotes the weight of the pixel point (d, v) in the V disparity map. n is the number of current pixels, w and h are the width and height of the original disparity map respectively, α and β are weighting factors, and T is the threshold of disparity. From the equation we can get that the bigger value

of the sample point and the smaller the disparity, the less likely it is to be the noise point. The closer the sample point to the bottom, the more likely it is the point represents road, so we give it a greater weight. After the pixels are weighted, we can detect the straight line representing the road. The red line in the graph in the third column of Fig. 2 is the straight line detected by Hough transformation, which is somewhat high and basically extracting the top of the arch. Our method is to detect a straight line by Hough transformation, then use weighted least square method to achieve the process of fitting to generate the second straight line as the road. Specifically, we use Eq. (12) to represent the line detected by Hough transformation:

$$v = kd + b \tag{12}$$

where k and b are parameters of the linear equation:

$$|v' - kd' - b| < \varepsilon \tag{13}$$

$$\min_{k',b'} \sum_{i=1}^{m} (v' - b' - kd') \tag{14}$$

By Eq. (14), we use all the effective pixels (v', d') satisfied Eq. (13), and their corresponding $w(d', v')$, to estimate the straight line representing the road exactly, which is shown as the blue line in the graph in the third column of Fig. 2. This method can not only avoid possible deviations from the road area caused by the drawback of Hough transformation, but also overcome the problem that ordinary least square method is sensitive to noise. By utilizing the detected line of road, we can determine the road's boundaries, and then apply the method in Sect. 3.1 to remove the disparities of the road and outside the road.

3.3 Obstacle Detection Based on Real U-disparity Map

After completing the above steps, we obtain the disparity map contains only the obstacles on the road, and then count the number of the same disparity in each column of disparity map to generate the U disparity map (top view), in which the abscissa corresponds to the original disparity map, while the ordinate represents disparity. The obstacle cannot be detected by detecting the horizontal line segments, because there is usually the errors in the stereo matching. In addition, the disparity points of the obstacles at greater distances are few in number greatly increases the difficulty of detection. To solve above problems, a method based on Real U disparity map is raised, which can revert the two-dimensional information to real spatial information. Here are details about how to transfer U disparity map to Real U disparity map, and how to detect obstacles:

(1) Map the U disparity map u-d to the spatial coordinate u-z, and calculate the distance between the obstacle and the binocular cameras according to $z = BF/d$, and then the pixels in a certain distance interval are counted together. So the obstacle, especially the nearby obstacle, can be compressed into a horizontal line segment, thus can avoid detecting an object as several objects.

(2) Proceed the speckle filtering in u-z graph for removing the hot pixel: Count the number of valid pixels and the sum of the valid pixel values, and then compare them with the corresponding threshold. Let this pixel equal to 0 when the two values are both smaller than the threshold, and equal to 255 otherwise.
(3) To stretch the pixels of the obstacles at greater distances, we map the u-z graph to the x-z graph to produce the Real U disparity map. Assume (x, z) to be one of the pixels in Real U disparity map and compute the corresponding points in the u-z disparity map according to Eq. (15).

$$u = u_0 + (x - B/2)F/(z\kappa) \tag{15}$$

in which B is the baseline length of binocular cameras, F is the focus distance, κ is the physical size of each pixel, and u_0 is the abscissa of the image center. Now we obtain the corresponding points of (x, z) in u-z graph, signed as (u, z). In actual scenes, it is necessary to fill it by interpolation method. A search scope is calculated based on current (x, z) and the given Δx, as shown in Eq. (16).

$$\begin{cases} u_l = u_0 + (x - \Delta x - B/2)F/(z\kappa) \\ u_r = u_0 + (x + \Delta x - B/2)F/(z\kappa) \end{cases} \tag{16}$$

where u_l and u_r denotes the left and right boundary respectively. By doing so, we can remove the hot pixel, and fill the missing disparity of obstacles. To detect horizontal line segments in the Real U disparity map, we move a valid pixel, which is taken as the starting point, to the right, and search the valid values horizontally. At the same time, set the threshold for the length of the line segment. Finally, merge the group of horizontal line segments.
(4) Determine the interval of columns and the scope of disparity of the obstacles according to the detected line segments in Real U disparity map. The pixels that satisfy the scope of disparity are compressed into one column, and the number of the valid disparities in each row is computed. Then search the obstacles from top to bottom. Finally, remove obstacles that are close to the ground according to the height.

4 Result

We evaluate our proposed method on KITTI dataset [6]. The reference images of obstacle detection in KITTI are divided into three categories: Car, Pedestrian and Cyclis, including 7481 images as training data and 7518 as test data. For each category, three levels are set according to the occlusion and truncation of the object: easy, moderate, hard. Our method can detect obstacles without knowing what kind of the obstacle is, therefore it does not require off-line training. The most authoritative method named oracle recall [7] is used to evaluate our method. We first compute the IoU corresponding to each obstacle, where

Table 1. Algorithm performance test.

Method	MHT	RU	Car		Cyclist		Pedestrian	
			Recall	Precision	Recall	Precision	Recall	Precision
ODDM-N2	No	No	0.62	0.78	0.59	0.71	0.51	0.68
ODDM-M	Yes	No	0.72	0.73	0.72	0.68	0.68	0.62
ODDM-R	No	Yes	0.58	0.88	0.55	0.8	0.53	0.79
ODDM-MR	Yes	Yes	0.81	0.95	0.78	0.89	0.75	0.85

IoU represents the degree of overlap between the bounding box and the ground truth. Then we set the threshold as T and receive three kinds of results: positive samples are detected as positive samples (TP), negative samples are detected as positive samples (FP), and the positive samples are detected as negative samples (FN). According to the results, we can calculate recall and precision: $Recall = TP/(TP + FP)$, $Presionl = TP/(TP + FN)$. Finally, we calculate a set of the recall and precision under different threshold T, further generating the R-P graph to compute the Average Precision (AP).

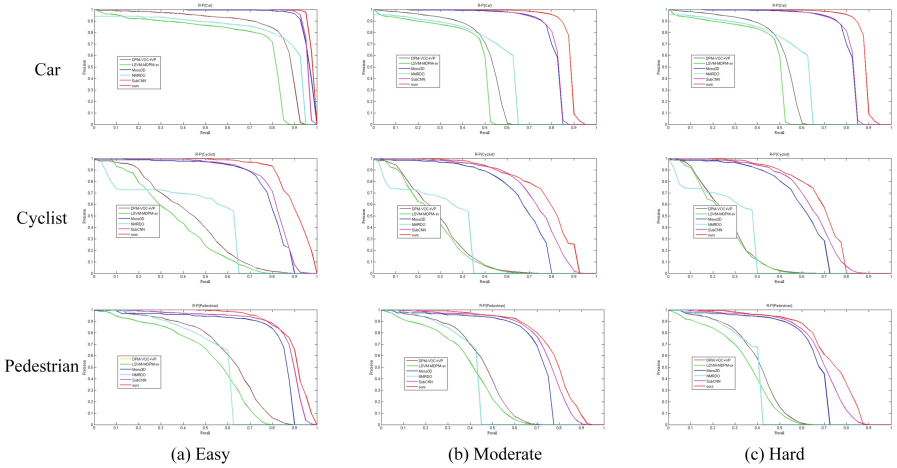

Fig. 3. Detection results of different method on KITTI dataset.

The result, when threshold T = 70%, is shown in Table 1. ODDM-N2 is the traditional obstacle detection method based on the disparity map, the experimental results show that this method has a large number of false detections and misses partial obstacles. ODDM-M applies the modified extraction method on the basis of ODDM-N2, thus the recall is obviously improved. Based on ODDM-N2, ODDM-R uses Real U disparity map to detect obstacles, which can better detect the obstacles at greater distances and further improve the precision.

Table 2. Average Precision (AP) (in %) on the test set of the KITTI object detection benchmark.

	Cars			Pedestrians			Cyclist		
	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
SS	75.91	60	50.98	54.06	47.55	40.56	56.26	39.16	38.83
3DOP	91.58	85.8	76.8	61.57	54.79	51.12	73.94	55.59	53
EB	83.91	67.89	58.34	46.8	40.22	33.81	43.97	30.36	28.5
R-CNN	-	-	-	61.61	50.13	44.79	-	-	-
pAUCEnsT	-	-	-	65.26	54.49	48.6	51.62	38.03	33.38
Regionlets	84.75	76.45	59.7	73.14	61.15	55.21	70.41	58.72	51.83
Faster R-CNN	86.71	81.84	71.12	78.86	65.9	61.18	72.26	63.35	55.9
Ours	94.01	93.15	90.28	88.25	85.81	80.42	80.53	77.86	75.24

ODDM-MR is our ultimate method, which has obvious advantages comparing with ODDM-N2. The recall of the Car, Cyclist and Pedestrian detection increased by 19%, 19%, and 24% respectively. The precision of the Car, Cyclist and Pedestrian detection increased by 17%, 18%, and 17% respectively.

To better illustrate the advantages, we compare our method with the currently mainstream obstacle detection methods: Mono3D [8], NMRDO, SubCNN [9], DPM-VOC+VP [10] and LSVM-MDPM-sv [11]. Figure 3 is the R-P graph, which shows the functional relation between the recall and the precision of different methods, and we can find that our method is superior to other detection algorithms in all three categories, even if under the hard level.

We calculate the average precision under different conditions, then compare with other methods. As shown in Table 2, the average precision of our method have absolute advantage in every aspect compared with other methods (SS, 3DOP [12], EB, R-CNN, pAUCEnsT [13], Regionlets [14] and Faster R-CNN [15]). What's more, even under the hard level, our method still maintains good performance. For example, compared with Faster R-CNN, the average precision of our method in Pedestrian detection on different levels (easy, moderate and hard) increased by 7.3%, 11.31% and 19.16% respectively.

5 Conclusion

We propose an obstacle detection method based on binocular stereovision, which does not require off-line training and can detect all obstacles on the road. We first present a method based on weighted least square method in V disparity map to detect the road accurately. Then we propose an approach to project U disparity map to Real U disparity map. Finally, the position of the obstacle is determined by a rapid locating method. The experimental results show that our proposed method is effective and reliable, and can deal with complex scenes very well, especially in case of occlusion and severe deformation of obstacles.

References

1. Zhang, L., Sun, Y., Luo, T., Rahman, M.M.: Note: a manifold ranking based saliency detection method for camera. Rev. Sci. Instrum. **87**(9), 096103 (2016)
2. Zhang, L., Yang, L., Luo, T.: Unified saliency detection model using color and texture features. PLoS ONE **11**(2), e0149328 (2016)
3. Zhang, L., Cai, Y., Ullah, Z., Luo, T.: MLPF algorithm for tracking fast moving target against light interference. In: International Conference on Pattern Recognition, pp. 3939–3944 (2016)
4. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through "V-disparity" representation. In: Intelligent Vehicle Symposium, vol. 2, pp. 646–651 (2003)
5. Hu, Z., Uchimura, K.: UV-disparity: an efficient algorithm for stereovision based scene analysis. In: Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 48–54. IEEE (2005)
6. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3354–3361. IEEE (2012)
7. Hosang, J., Benenson, R., Dollar, P., Schiele, B.: What makes for effective detection proposals? IEEE Trans. Pattern Anal. Mach. Intell. **38**(4), 814–830 (2016)
8. Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., Urtasun, R.: Monocular 3D object detection for autonomous driving. In: Computer Vision and Pattern Recognition (2016)
9. Xiang, Y., Choi, W., Lin, Y., Savarese, S.: Subcategory-aware convolutional neural networks for object proposals and detection (2016)
10. Pepik, B., Stark, M., Gehler, P., Schiele, B.: Multi-view and 3D deformable part models. IEEE Trans. Pattern Anal. Mach. Intell. **37**(11), 2232 (2015)
11. Geiger, A., Wojek, C., Urtasun, R.: Joint 3D estimation of objects and scene layout. In: Advances in Neural Information Processing Systems, pp. 1467–1475 (2011)
12. Chen, X., Kundu, K., Zhu, Y., Berneshawi, A.G., Ma, H., Fidler, S., Urtasun, R.: 3D object proposals for accurate object class detection. In: Advances in Neural Information Processing Systems, pp. 424–432 (2015)
13. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Pedestrian detection with spatially pooled features and structured ensemble learning. IEEE Trans. Pattern Anal. Mach. Intell. **38**(6), 1243–1257 (2016)
14. Long, C., Wang, X., Hua, G., Yang, M., Lin, Y.: Accurate object detection with location relaxation and regionlets re-localization. In: Asian Conference on Computer Vision, pp. 260–275 (2014)
15. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137 (2016)