



Global-local articulation pattern-based pedestrian detection using 3D Lidar data

Shuangli Du, Bingbing Liu, Yiguang Liu & JianGuo Liu

To cite this article: Shuangli Du, Bingbing Liu, Yiguang Liu & JianGuo Liu (2016) Global-local articulation pattern-based pedestrian detection using 3D Lidar data, Remote Sensing Letters, 7:7, 681-690, DOI: [10.1080/2150704X.2016.1177239](https://doi.org/10.1080/2150704X.2016.1177239)

To link to this article: <http://dx.doi.org/10.1080/2150704X.2016.1177239>



Published online: 28 Apr 2016.



Submit your article to this journal [↗](#)



Article views: 40



View related articles [↗](#)



View Crossmark data [↗](#)



Global–local articulation pattern-based pedestrian detection using 3D Lidar data

Shuangli Du^a, Bingbing Liu^b, Yiguang Liu^a and JianGuo Liu^c

^aComputer Science and Technology, Sichuan University, Chengdu, China; ^bInstitute for Infocomm Research (I2R), Agency for Science Technology and Research (A*STAR), Singapore, Singapore; ^cDepartment of Earth Science Engineering, Imperial College, London, UK

ABSTRACT

Highly variable human poses and pedestrian occlusion make light detection and ranging (Lidar)-based pedestrian detection challenging. This letter proposes a novel framework to address these issues. Other than dividing humans into arbitrary number of parts and using the same features for all part detectors, we represent humans with global–local articulated parts and formulate new features relying on each part's own character. Articulated parts are effective because they each usually maintain a relatively consistent shape across a broader range of body poses. In addition, to extract visible human segments from cluttered surroundings with the presence of pedestrian occlusion, both 3D information and 2D spatial information are used in a coarse-to-fine manner, making the interaction of human part and its neighbouring objects better analysed. The algorithm is evaluated over a busy street dataset and is shown to be competitive with the state-of-the-art Lidar-based algorithms. Remarkably, even at long distances, up to 20 m, it can handle pedestrian occlusion efficiently and effectively.

ARTICLE HISTORY

Received 6 November 2015
Accepted 3 April 2016

1. Introduction

Robustly detecting pedestrians is an important and fundamental problem for mobile robots and autonomous vehicles. Popular sensors for this task are cameras and range finders. Some researchers are working on using one type of sensor (Oliveira and Nunes 2010; Hegger et al. 2013; Mikolajczyk, Schmid, and Zisserman 2004; Tang, Andriluka, and Schiele 2013). Recently, many researchers focus on fusing information provided by two or more types of sensors, such as Zivkovic and Kröse (2007) and Premebida and Nunes (2013). This letter uses a light detection and ranging (Lidar) system, which provides a set of 3D points, point cloud. The contribution idea of this letter is to represent humans with global–local articulation pattern, which is efficient in modelling diverse occlusion patterns. In addition, the interaction of human parts and its neighbouring objects is considered to increase the pattern's robustness.

Many 2D range data-based work reduce the problem of people detection into a leg detection task. In Xavier et al. (2005), leg segments were considered to be 'arc-like' shapes, and a circle detection technique called inscribed angle variance (IAV) was

presented. In Arras, Mozos, and Burgard (2007), they applied adaptive boosting algorithm (AdaBoost) to learn a robust leg detector from a set of statistic features. The same classification strategy was used in Zivkovic and Kröse (2007). Besides, it took into account the possible distance between the person's legs. In Mozos, Kurazume, and Hasegawa (2009), a multi-layer strategy was used. After training a classifier for each layer, they combined all layer classifiers in a probabilistic way to create a final robust detector. It is also popular in computer vision to use part-based representations and multi-layer strategies in object recognition tasks (Mikolajczyk, Schmid, and Zisserman 2004; Srinivasan and Shi 2007; Felzenszwalb Pedro et al. 2010).

There is less related work on pedestrian detection in 3D data. In Spinello et al. (2010), they extended the 2D-based methods in Mozos, Kurazume, and Hasegawa (2009) and Arras, Mozos, and Burgard (2007) to 3D. They see 3D point cloud as a collection of several 2D point clouds at different heights, and create a part detector by AdaBoost to decide whether each 2D point cloud is a part of a human body. Pedestrians are found by integrating the estimation results of all parts. However, such approaches are ineffective when human posture suffers from high variability. In Navarro-Serment, Mertz, and Hebert (2010), the 3D point cloud in the target was divided into three parts, i.e., the trunk and the legs. In addition, 2D histograms on two principal planes were used to represent a 3D pedestrian shape. The approach has low computational cost but its performance degrades a lot at longer distances. In Kidono et al. (2011), two features were introduced to improve the classification performance. One is the slice feature, which represents the profile of a human body by widths at the different height levels. The other is the distribution of the reflection intensities of points measured on the target. However, either Navarro-Serment, Mertz, and Hebert (2010) or Kidono et al. (2011) cannot deal well with the problem of pedestrian occlusion.

As mentioned, in addition to pedestrian occlusion and highly variable human poses, low spatial resolution at a long distance makes pedestrian detection more challenging. When a target is far from the sensor, the point cloud reflected from the target is sparse, making the extracted statistical and geometrical features unreliable. Pedestrian occlusion usually appears in complicated scenes, such as when people walk shoulder-to-shoulder in a lineup or push prams. In those scenarios, it is difficult to extract human segments. So the estimated features from that segment is disturbed by noisy points, and do not reflect the silhouette of humans. If features are estimated with neighbouring points on a 2D scan line, disturbance will be more evident. To identify humans with partial occlusion and model various poses, previous methods divide humans into different height layers or parts, and create a classifier for each layer. Finally, all classifiers are combined using a voting technique to find humans. The number of layers can be randomly given. There is no discussion about how many layers should be used. Humans have different sizes, body shapes and poses. Especially when walking, the appearance of lower body is highly variable. Therefore, to create an effective part classifier, a large number of training datasets are required, especially for parts in the lower body. Moreover, when humans are occluded partially, it is difficult to extract visible human segments and decide which layer the segment belongs to.

This letter proposes a pedestrian detection framework to handle these matters. The contribution in this letter is highlighted as follows.

- (1) We use global–local articulated parts to describe humans instead of dividing humans into arbitrary number of parts. The main benefits are: articulated body parts can model diverse body shapes and poses. More importantly, each part maintains a relatively consistent shape, resulting in more forceful part detectors.
- (2) To guarantee part detectors' robustness, we aim to build some slice features to describe the outlines of body parts, which have proved to be more effective when input points are noisy or few in number.
- (3) One challenge is to identify humans with partial occlusion. It requires us to extract visible human segment from cluttered surroundings. For this, we segment the raw point cloud by a coarse-to-fine technique. In the segmentation stage, the interaction of a visible human segment with its neighbouring objects is analysed and is first used for identification.

Figure 1 is the outline of the proposed detection system. An impressive articulated model is adopted, that is, a whole human body is divided into head, torso and legs. But the appearance of two legs is diverse, thus, a powerful pattern is needed, which should be able to characterize almost all shape types. For this, a set of articulated parts as used in somatology (the study of human body structure) is used. This is discussed in Section 3.2. The experiments show that the proposed system can handle pedestrian occlusion efficiently at long distances, up to 20 m.

2. Coarse-to-fine pre-processing 3D Lidar data

In real scenarios, when walking, people often carry backpacks, push prams or walk side by side. In this way, people appear in point clouds with partial occlusion. Considering the position in 3D space, human segments are near to other objects, and are often seen as part of others. But, from the top-down view, they are isolated. Based on this fact, a coarse-to-fine segmentation strategy is proposed to extract human segments.

In the coarse segmentation strategy, we use an occupancy grid map (Kidono et al. 2011) to divide the acquired 3D point cloud into three clusters: separate pedestrian candidates, combined pedestrian candidates and other objects, based on their sizes (height h , width w and length l). In this letter, the empirical condition for separate

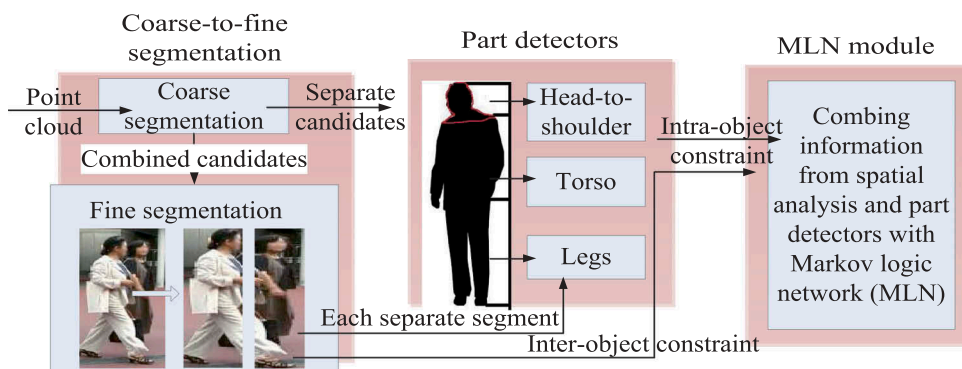


Figure 1. Outline of the proposed pedestrian detection system.

candidates are $0.8\text{m} \leq h \leq 2.0\text{m}$, $w \leq 1.2\text{m}$ and $l \leq 1.2\text{m}$. For combined candidates, $0.8\text{m} \leq h \leq 2.0\text{m}$, $w \leq 2.0\text{m}$ and $l \leq 2.0\text{m}$. Each combined candidate may contain multiple objects. For example, in the case when two people walk in a tight line or stand shoulder-to-shoulder in a lineup. With the side view, the two persons come together into a combined candidate. But, with the top-down view, they are separated. To properly segment the combined candidate, we perform a fine segmentation. First the combined candidate is segmented into small clusters by using the Euclidean cluster extraction method in Point Cloud Library (PCL) (<http://pointclouds.org/>). Subsequently, the small clusters are recombined into a set of new objects by incorporating 2D spatial information in the x - y plane into an iterative region growing process. The process gives an initial guess for the centroid point of each small cluster in the horizontal plane. Then, in each loop, a seed is defined as the centroid point of the largest one among the small clusters. If the distance from another centroid point to the seed is within a threshold, these two corresponding clusters are combined. This procedure iterates over the rest of small clusters until no cluster is left.

In addition to the object's own feature, its neighbouring objects' feature and their interaction provide some useful information about its type. The position of a human part relative to its environment is a useful cue about its type. An object size also provides some information about its type. Based on position and size, the new objects obtained in the fine segmentation stage are marked as different candidates. The details are listed below:

- the size of object 'o' meets the full body's constraints $\rightarrow$ 'o' is a full body candidate,
- the object 'o' is spatially close to other objects but there is no object above it, and its size is smaller than the full body's $\rightarrow$ object 'o' is an upper body candidate,
- the object 'o' is spatially close to other objects and it is on the ground, and its size is smaller than the full body's $\rightarrow$ 'o' is a legs candidate.

3. Global-local articulation patterns

To boost detection performance, an effective segmentation, coarse-to-fine framework, is required to find human segments with as little noise as possible. That segmentation yields candidates of whole body and body parts. This section focuses on how to create part detectors. Different from Spinello et al. (2010) using same features to construct each part detector, we adopt new features relying on global-local articulation patterns. That is, a head-to-shoulder is proposed for articulated head and torso; and to describe the highly variable appearance of legs, a label tree is created based on legs' articulation style.

3.1. Upper body

A person's head, neck and shoulders maintain a relatively consistent shape across a broader range of body poses/activities and are usually clearly visible in outdoor environments. Thus, the pattern of head-to-shoulder part is usually used for the people detection task (Kirchner, Alempijevic, and Virgona 2012). Due to occlusion or sensor's vertical field of view, the extracted upper body candidate may correspond to different regions, such as head, head-to-shoulder, shoulder-to-waist, etc. To decide which part a

candidate corresponds to, a new head-to-shoulder feature is created based on an assumption that the head breadth is less than the torso breadth.

The head-to-shoulder feature is a string over an alphabet $\Sigma = \{0, 1\}$. If a 2D line's breadth is larger than the pre-set threshold given by Otsu method (Otsu 1979), the value 1 is assigned; if it is lower, the value 0 is assigned. Given a feature string, it requires us to split it into two parts, and to decide whether the upper part is a head. If the upper is head shape, the lower tends to be torso.

An ideal string in which no 1 exists between 0s (e.g., '000|111') is clearly made up of two parts. But, due to highly variable human poses or noisy input data, the ideal string seldom appears. Thus, accumulated ratio (the ratio of the number of continuous 0 to the current string length) is used. A string is divided according to the highest ratio. For instance, the accumulated ratios of the string '00011000111110011' are 3/5, 6/8, 6/13, 8/15 and 8/17, and the segmentation result is '00011000|111110011'. Then, the scan lines in the upper part are provided as an input of head detector, namely, size constraint. And the scan lines in the lower part are considered as an input of torso detector.

For each 2D scan line, slice features given in Kidono et al. (2011) (slice width, slice length) and point-to-point features given in Arras, Mozos, and Burgard (2007) (mean average deviation from median, boundary length, boundary regularity, mean curvature and mean angular difference) are computed. They form a vector $\mathbf{f} = (f_1, \dots, f_7)$, which describes the silhouette of a torso. Then, we train a torso detector from $\mathbf{f}$ using support vector machine (SVM) with a radial basis function (RBF) kernel.

3.2. Legs

The pattern of two legs is a distinctive human shape. It is, however, difficult to extract point-to-point features at a long distance, where the spatial resolution decreases. Hence, this letter exploits an 'arc-shape' detection procedure to detect legs like Xavier et al. (2005). Unfortunately, the points on a 2D scan line take on a weak-arc pattern in outdoor environments. As Figure 2(a) illustrates, a fitted circle is created by solving the unknowns, circle centre and radius in the least squares pattern. The leftmost and rightmost points, P_1 and P_3 , and the middle point P_2 can be located. If $\delta_1 \times d(P_1, P_3) \leq d(O, P_2) \leq \delta_2 \times d(P_1, P_3)$ with $d(\cdot, \cdot)$ the Euclidean distance, the 2D scan line is marked as a circle. The values of $\delta_1 = 0.1$, $\delta_2 = 0.7$ as given in Xavier et al. (2005). To further decide

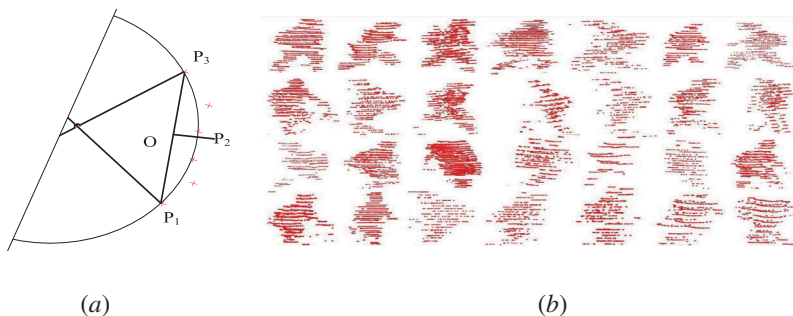


Figure 2. (a) A fitted circle, which is created by using least squares fitting technique on the red crosses. (b) Illustration of some example articulated patterns of two legs.

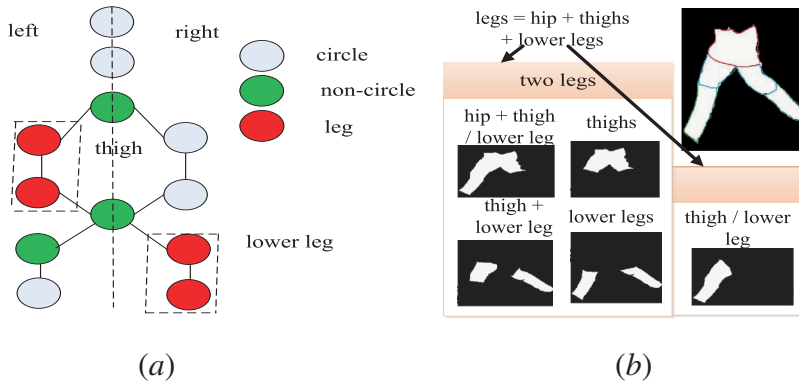


Figure 3. (a) Label tree is made up of three parts: the left, middle and right. There are three types of nodes: the circle, non-circle and leg segments. (b) Leg parse tree describes four cases for two legs and one case for one leg.

whether the circle is a leg segment, it requires us to check whether the distance $d(P_1, P_3)$ falls within the range of expected leg diameters (0.10 m–0.25 m).

The shape of two legs is a highly variable human feature. In this letter, articulated pattern of two legs is first considered to describe the variety. More specifically, a label tree is created to describe it. When people walk, there are large gaps between the left leg and the right leg at some height levels, while not at other levels. It means that some 2D scan lines can be split into two sub-segments, while others don't. If the maximum distance between two consecutive points exceeds a threshold d_{ϵ} , a new segment is produced. Subsequently, for the new obtained sub-segments and indivisible lines, the circle detection and leg segment detection are performed successively. Finally, we produce a label tree by arranging the detection results line by line like Figure 3(a). Some statistical and geometrical features are extracted from the tree to create a leg detector. First, we define a term, 'connected region'. In each of the three part of the tree, if there are two or more same nodes connected, they form a connected region. For each type of nodes (e.g., leg segments) of each part (e.g., the left part), a vector $\mathbf{g} = (g_1, g_2, g_3, g_4)$ is formed as below.

- $g_1 = \frac{n}{N}$ with n the number of leg segments in the left part, N the number of all 2D scan lines in the lower body part.
- $g_2 = \sum_{i=1}^M \frac{m_i}{r_i^2}$ with M the number of connected regions of leg segments, m_i the number of leg segments in region i , r_i the distance between the centre points of the i_{th} region and the leg candidate.
- $g_3 = \frac{\sum_i l_i}{n}$ with l_i the slice length of the i_{th} leg segment
- $g_4 = \frac{\sum_i w_i}{n}$ with w_i the slice width of the i_{th} leg segment.

Features g_1 , g_3 and g_4 are segments' own intrinsic characteristics, while g_2 describes the spatial relation between each connected region and the whole lower body candidate. Similarly, vector $\mathbf{g}$ can be defined for the other two nodes (circle and non-circle) in the

three parts. Then, we group them together into a feature vector with 36 elements, and train a leg detector from the feature vector using SVM with a RBF kernel.

Once a leg candidate is identified as human leg, to further analyse whether two legs or one leg it includes, the articulated pattern of two legs is exploited. As illustrated in Figure 3(b), there are four cases for two legs and one case for one leg. We just use the connected regions to check with the leg parse tree.

4. Combing information from spatial analysis and part detectors

How to combine information from spatial analysis and part detectors is discussed here. The relationship of the part detectors as well as contextual relations of inter-objects is modelled as first-order logic (FOL) in an Markov logic network (MLN) framework (Singla and Domingos 2005), which makes it possible to compactly specify probability distributions over complex relational domains. A MLN is a first-order knowledge base which is formed by formula (or clause) with each attached a weight. The FOL formulas are context rules describing the relations of intra-object and inter-objects. They are summarized in Table 1.

Formulas 1–3 in Table 1 embody a usual pattern with a pedestrian being detected with a ‘head-shoulder’. That is, if some parts of the lower body are occluded by other objects, then a detected ‘torso’, ‘leg’ or other object is necessary to validate it as a pedestrian. Similarly, formulas 4, 5 and 6 represent patterns with a pedestrian being detected with a leg. That is, if some parts of the upper body are structured by a potential occlusion, then a detected ‘torso’, another ‘leg’ or other object is necessary to validate it as a pedestrian.

5. Experiments

The data sets Polyterrasse and Tannenstrasse in Spinello et al. (2010) are used to train and test the proposed method. They are collected by a Velodyne HDL 64E S2 laser scanner with a frequency of 5 Hz.

In the proposed pedestrian detection framework, the part detectors of torso and leg are trained by SVM with 100 background objects and 700 persons extracted from the Polyterrasse data set, and $d_e = 10$ cm. The MLN context detector is trained on 100 combined pedestrian candidates and 300 separate pedestrian candidates. The weights of an MLN can be learned or hand-crafted. We use the discriminative learning approach (Singla and Domingos 2005) provided by the open source package, Alchemy (<http://alchemy.cs.washington.edu/>) (Kok et al. 2010), which is demonstrated

Table 1. First-order formulas.

No.	First-order-logic	Description
1	$\text{headShoulder}(o_1, \text{Head}) \wedge \text{near}(o_1, o_2) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has head-to-shoulder and is nearby other object, it is a pedestrian.
2	$\text{headShoulder}(o_1, \text{Head}) \wedge \text{hasTorso}(o_1, \text{Torso}) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has head-to-shoulder and torso, it is a pedestrian.
3	$\text{headShoulder}(o_1, \text{Head}) \wedge \text{hasLeg}(o_1, \text{Leg}) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has head-to-shoulder and one leg, it is a pedestrian.
4	$\text{hasTorso}(o_1, \text{Torso}) \wedge \text{hasLeg}(o_1, \text{Leg}) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has torso and one leg, it is a pedestrian.
5	$\text{hasLeg}(o_1, \text{Leg}) \wedge \text{near}(o_1, o_2) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has one leg and is near by any other object, it is a pedestrian.
6	$\text{hasLeg}(o_1, \text{Leg}) \wedge \text{hasLeg}(o_1, \text{Leg}) \Rightarrow \text{person}(o_1)$	‘ o_1 ’ has two legs, it is a pedestrian.

Table 2. Results for pedestrian detection at different ranges: 0–10 m, 10–15 m and 15–20 m.

Detector	0–10 m		10–15 m		15–20 m	
	TP	FP	TP	FP	TP	FP
Head	64.3	16.5	55.7	14.8	53.2	17.9
Torso	57.6	24.7	47.2	32.4	45.1	35.1
Leg	76.1	13.4	68.5	23.5	57.6	23.6
MLN	96.2	3.6	87.2	9.8	84.4	9.6

‘TP’ is true positives and ‘FP’ is false positives.

to outperform generative learning approaches. After training, an inference needs to be performed. The most widely used method for approximate inference in MLN is Markov chain Monte Carlo (MCMC). We use MC-SAT (Poon and Domingos 2006) provided by Alchemy, which is a combination of MCMC and SampleSAT algorithms.

The Tannenstrasse data set is used to validate the performance of the proposed method. A lot of experiments have been done. To save space, only comparisons with the state-of-the-art Lidar-based algorithms (Spinello et al. 2010; Kidono et al. 2011) are given. First, the relation of the detection performance with distance from the sensor is evaluated for each individual part detector and the MLN context detector. The results are summarized in Table 2. As the distance range increases, the performance of each detector decreases. But, even at 15–20 m, the MLN context detector obtained a quite good performance. The results demonstrate the benefits of the MLN detector compared to each individual detector. This means that the part detectors are complementary. The MLN detector makes use of the complementarity, improving the detection performance. In addition, due to the high variability of leg shape, the leg detector has a high false alarm rate. Figure 4(a) illustrates the overall precision–recall graph of the proposed method applied to the Tannenstrasse data set. The equal error rate (EER) for ranges between 0–10 m, 0–15 m and 0–20 m are 93.2%, 81.3% and 73.8%, respectively. Compared with the corresponding EER, 95%, 76% and 63% obtained in Spinello et al. (2010), it is clear that the proposed approach improves the classification performance at a long distance. Figure 4(b) shows an example frame, where all people are correctly detected, even when they are walking shoulder-to-shoulder. For the Tannenstrasse data set, 2525 pedestrians are detected, 679 of them are partial occlusion. This demonstrates that the proposed method can deal with pedestrian occlusion efficiently.

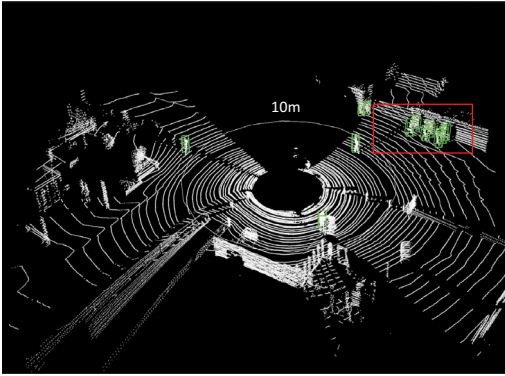
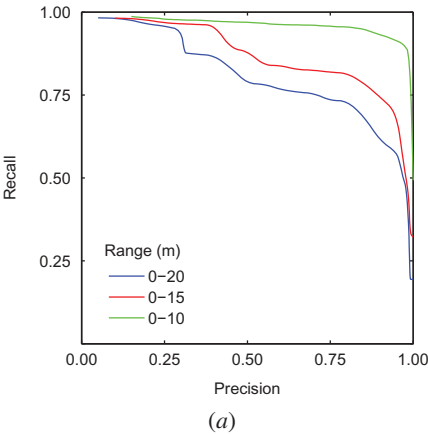


Figure 4. (a) Illustration of precision-recall graph at different ranges: 0–10 m, 0–15 m and 0–20 m. (b) Example frame showing detection results as green boxes.

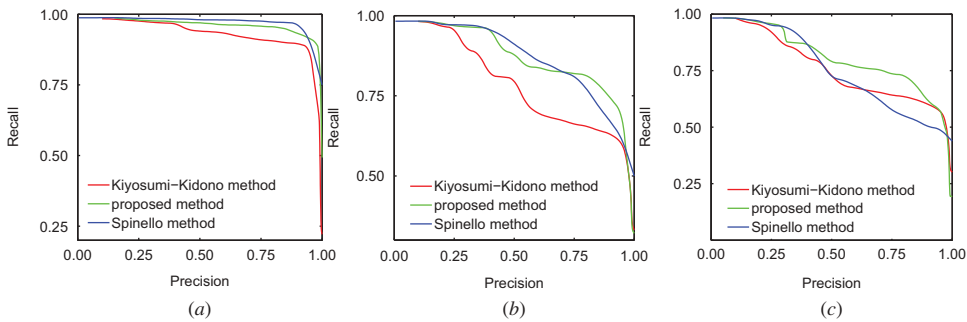


Figure 5. Precision–recall curves of the proposed system, Spinello method and Kiyosumi-Kidono method at different ranges. (a) 0–10 m, (b) 0–15 m and (c) 0–20 m. Our method shows better performance especially at the ranges 0–15 m with many pedestrians occluded.

The Kiyosumi-Kidono method (Kidono et al. 2011) is run on the Tannenstrasse data set for comparison. The result is shown in Figure 5. The proposed method generally shows better performance than the Kiyosumi-Kidono method, especially at the ranges 0–15 m. The main reason is that many people are partially occluded at the ranges 0–15 m. Our method can identify the partially occluded humans, while the Kiyosumi-Kidono method cannot.

6. Conclusions

An effective and novel Lidar-based pedestrian detection approach is proposed in this letter, which addresses difficulties caused by pedestrian occlusion and highly variable poses. We first represent humans with global–local articulated parts other than arbitrary number of parts, leading to better model various human poses. To efficiently find partially occluded humans, the interaction information of a human segment with its neighbouring objects is first considered. In addition, to make sure extracted features are robust, part detectors are built upon slice features instead of point-to-point features. Experimental results show that our approach clearly improves the detection performance compared with the state-of-the-art Lidar-based algorithms, and can handle pedestrian occlusion efficiently even at a long distance.

Acknowledgements

We thank the editor and reviewers for time and effort spent handling this letter.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the National Natural Science Foundation of China under [Grant 61571313, Grant 61173182, and Grant 61411130133] and in part by the Sichuan Province under [Grant 2014HH0048 and Grant 2014HH0025].

References

- Arras, K. O., Ó. M. Mozos, and W. Burgard. 2007. "Using Boosted Features for the Detection of People in 2D Range Data." In *IEEE International Conference on Robotics and Automation*, Roma, 10–14 April, 3402–3407. Piscataway, NJ: IEEE.
- Felzenszwalb Pedro, F., R. B. Girshick, D. McAllester, and D. Ramanan. 2010. "Object Detection with Discriminatively Trained Part Based Models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32: 1627–1645. doi:10.1109/TPAMI.2009.167.
- Hegger, F., N. Hochgeschwender, G. K. Kraetschmar, and P. G. Ploeger. 2013. "People Detection in 3d Point Clouds Using Local Surface Normals." In *RoboCup 2012: Robot Soccer World Cup XVI*. Vol. 7500, 154–165. Berlin, Heidelberg: Springer.
- Kidono, K., T. Miyasaka, A. Watanabe, T. Naito, and J. Miura. 2011. "Pedestrian Recognition Using High-Definition LIDAR." *Intelligent Vehicles Symposium (IV)*, Baden-Baden, June.
- Kirchner, N., A. Alempijevic, and A. Virgona. 2012. "Head-to-Shoulder Signature for Person Recognition." In *IEEE International Conference on Robotics and Automation*, Saint Paul, MN, 14–18 May, 1226–1231. Piscataway, NJ: IEEE.
- Kok, S., P. Singla, M. Richardson, and P. Domingos. 2010. *The Alchemy System for Statistical Relational AI: User Manual*. <http://alchemy.cs.washington.edu/user-manual/manual.html>.
- Mikolajczyk, K., C. Schmid, and A. Zisserman. 2004. "Human Detection Based on a Probabilistic Assembly of Robust Part Detectors." In *8th European Conference on Computer Vision (ECCV)*, 69–82. Berlin, Heidelberg: Springer.
- Mozos, O. M., R. Kurazume, and T. Hasegawa. 2009. "Multi-Layer People Detection Using 2D Range Data." In *Proceedings of the IEEE ICRA 2009 Workshop on People Detection and Tracking*, Kobe, Japan, May. Piscataway, NJ: IEEE.
- Navarro-Serment, L. E., C. Mertz, and M. Hebert. 2010. "Pedestrian Detection and Tracking Using Three-dimensional LADAR Data." *The International Journal of Robotics Research, Special Issue on the Seventh International Conference on Field and Service Robots* 29 (12): 1516–1528.
- Oliveira, L., and U. Nunes. 2010. "Context-Aware Pedestrian Detection Using LIDAR." In *Intelligent Vehicles Symposium (IV)*, San Diego, CA, 21–24 June, 773–778. Piscataway, NJ: IEEE.
- Otsu, N. 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems* 9 (1): 62–66.
- Poon, H., and P. Domingos. 2006. "Sound and Efficient Inference with Probabilistic and Deterministic Dependencies." In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, Boston, MA, 458–463. Palo Alto, CA: AAAI.
- Premebida, C., and U. Nunes. 2013. "Fusing LIDAR, Camera and Semantic Information: A Context-Based Approach for Pedestrian Detection." *The International Journal of Robotics Research* 32: 371–384. doi:10.1177/0278364912470012.
- Singla, P., and P. Domingos. 2005. "Discriminative Training of Markov Logic Networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*. Pittsburgh, PA. Palo Alto, CA: AAAI.
- Spinello, L., K. O. Arras, R. Triebel, and R. Siegwart. 2010. "A Layered Approach to People Detection in 3D Range Data." In *Proceedings of the Twenty Fourth AAAI Conference on Artificial Intelligence*, Atlanta, GA, 11–15 July, 1625–1630. Palo Alto, CA: AAAI.
- Srinivasan, P., and J. Shi. 2007. "Bottom-Up Recognition and Parsing of the Human Body." In *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, 17–22 June, 1–8. Piscataway, NJ: IEEE.
- Tang, S., M. Andriluka, and B. Schiele. 2013. "Detection and Tracking of Occluded People." *International Journal of Computer Vision* 110 (1): 58–69.
- Xavier, J., M. Pacheco, D. Castro, A. Ruano, and U. Nunes. 2005. "Fast Line, Arc/Circle and Leg Detection from Laser Scan Data in a Player Driver." In *Proceedings of the 2005 IEEE International Robotics and Automation*, Barcelona, Spain, April, 3930–3935. Piscataway, NJ: IEEE.
- Zivkovic, Z., and B. Kröse. 2007. "Part Based People Detection Using 2D Range Data and Images." In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Diego, CA, 29 October–2 November, 214–219. Piscataway, NJ: IEEE.