



**HAL**  
open science

## Multi-sensors people detection system for heavy machines

Manh Tuan Bui, Vincent Frémont, Djamel Boukerroui, Pierrick Letort

► **To cite this version:**

Manh Tuan Bui, Vincent Frémont, Djamel Boukerroui, Pierrick Letort. Multi-sensors people detection system for heavy machines. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Oct 2014, Qingdao, China. pp.867-872. hal-01083998

**HAL Id: hal-01083998**

**<https://hal.science/hal-01083998v1>**

Submitted on 18 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multi-sensors people detection system for heavy machines

M. Bui<sup>1</sup>, V. Frémont<sup>1</sup>, D. Boukerroui<sup>1</sup>, P. Letort<sup>2</sup>

**Abstract**—In this paper, we propose a multi-sensors system to detect people in the context of construction sites using heavy machines. The system includes a LIght Detection And Ranging (Lidar) sensor and a fisheye camera. We present an effective method to determine regions of interest (ROI) on fisheye images using Lidar data, which can be used with different sensors configurations. A Deformable Part Model (DPM) approach is adapted and used as the main people detector on image. We also present a specific dataset built using the multi-sensor system mounted on a heavy machine for evaluation.

**Index Terms**—Heavy machines, sensor fusion, pedestrian detection, deformable part model, fisheye, histogram of oriented gradients.

## I. INTRODUCTION

Heavy machines in construction sites constitute a high risk of serious injuries for people working in their surroundings. Due to their large different functionalities and shapes, accidents caused by heavy machines are various. An investigations of accidents in the period starting from 1997 to 2008 for France was conducted and published in the EPICEA database<sup>1</sup> [18]. From 2157 accident cases of all kinds, there are about 15% caused by collisions between machines and people. In view of these worrying statistics, efforts have been made to improve the safety for people working around heavy machines. To better prevent than cure, research has moved toward active safety intelligent systems which are able to predict dangerous situations and anticipate the accidents. They are referred as advanced driver assistance systems (ADAS), in the sense that they help the driver by providing warnings, assistance to take decisions and even taking automatic evasive actions in extreme cases. Notwithstanding many years of progress the problem of detecting obstacles and recognizing people from other ordinary objects is still a tough issue.

### A. Related works

To solve the problem of people detection, different sensors have been used and combined. The most popular type of sensor for people detection is the camera. It offers the best trade-off as a low-cost and polyvalent sensor. Recently, following the Euro New Car Assessment Program (NCAP) recommendations, people detection systems in automobile

using cameras known important progresses [8], [10], [12]. Although the objective is similar in both automobiles and heavy machines fields, we can clearly distinguish between the two. Indeed, in the automobile context, cars need to stop if there is an obstacle, no matter if it is a pedestrian or not. The task of recognizing people is more important for heavy machines where the main requirement is human's safety and since heavy machine can pass through specific obstacles (rocks, holes, woods, etc.). Besides, cars often operate at a higher speed and on straight ways. While it is important for the system on automobile to be able to detect people at far distances, heavy machines need preferably a larger field of view (FOV) to cover the nearby area that constitutes the main dangerous zone. Moreover, construction machines often have a complicated shape and a larger size, which can also benefit from the large FOV. These features reveal the necessity to use wide angle sensors such as fisheye cameras as the main sensor for people detection and recognition [4].

However, objects detection in fisheye images is a challenging task. Indeed, unlike in perspective images, the objects appearances are strongly distorted. Wrapping the image into a local perspective image is the direct way to avoid non-perspective deformations but, besides adding computational load, this approach also creates undesirable effects [5], [7]. Recently, the Deformable Parts Model (DPM) [12] and its variants have gained a lot of attentions in object detection and recognition. The DPM represents an object model as different parts floating around their reference locations and finds the associated optimal part-configuration at every root position. This elegant way of representing complex shapes of objects, brings a lot of benefit in detecting people in distorted appearances. Still, the DPM approach has few drawbacks that can be stated here:

- The high performance of the DPM approach comes with a high computation cost and complexity.
- The DPM approach using the sliding windows detection method needs to scan all over the image to detect the person. Moreover, it is not invariant to the rotation of the person's appearance on image. This drawback is critical in fisheye images. For example, it might not be suitable to detect people in images from a camera that looks down from high positions (see Fig.6).

Since cameras are low-cost sensors and since they exhibit high computation load, they can be combined with other sensors. For example, range sensors, like radar, Light Detection And Ranging (Lidar) and ultrasonic [1], which have good performance in detecting obstacles. Heavy machines often work in complicated terrains with a lot of nearby objects.

The authors are with <sup>1</sup>Université de Technologie de Compiègne (UTC), France. CNRS Heudiasyc UMR 7253, France. <sup>2</sup>Technical center for the Mechanical Industry (CETIM), France

<sup>1</sup>EPICEA is a data base on working accidents, collaboration between CNAMTS (Caisse nationale de l'assurance maladie), CRAM (les Caisses régionales d'assurance maladie) and INRS (Institut national de recherche et de sécurité) since 1988.

Sometimes they even need to crush these obstacles. In these situations, range sensors will trigger a permanent alarm, which is useless and annoying for the drivers. Range sensors are still very valuable in our context as they provide very precise and robust measurements. Therefore, such sensors are very good options in combination with cameras [2], [14], [20], [22].

### B. Article’s contributions and structure

The contributions of this paper are two folds. The first one lies in the combination of Lidar data to bypass the drawbacks of the DPM approach in people detection. The specific difficulties to process fisheye images in the context of heavy machines are taken into consideration. The proposed fusion system has various advantages:

- It helps to reduce the detection area on images resulting in a faster detection.
- It helps to give a lower false detection rate without sacrificing the detection performance.
- It helps to adjust the rotation of objects’ appearances caused by image distortions in different camera view angles.

The second contribution concerns the evaluation of our approach on real data acquired in the context of heavy machine.

The paper is organized as follows. First, two adaptive sensors configurations on heavy machines and details about our dataset are given in section II. Then, details on the data fusion algorithm and the projection process of the regions of interest (ROIs) on fisheye images are explained in section III. The evaluation protocol and the experimental results are discussed in section IV. Finally, section V presents the conclusions and the perspectives of our work.

## II. ACQUISITION SYSTEM AND DATASET

Datasets take a very important role in the process of the development of an obstacle detection algorithm. Indeed, a well-defined dataset is not only useful for evaluating the approach but it also takes part in the training process to improve the performance of the detector. To the best of our knowledge, there are no others available datasets which provides at the same time synchronized fisheye images and Lidar data in the context of heavy machines environments. Moreover, this context has some special features, such as the outdoor changing light conditions, the strong vibrations coming from the engine, and the brutal shocks that might make the detection process much harder. There are essential needs for a new dataset in order to identify conditions under which current detectors fail. We will focus our attention on these difficult cases.

### A. System Configuration

The danger is highly dependent on the action taken by the machine and its direction. Based on the survey of accidents caused by heavy machines (Tab. I), we can conclude that accidents rarely happen on the sides of the machines, except

Operation	Repartition of accidents (%)
Static	20
Move backward	42
Move forward	27
Not specified	11

Table I: Repartition of accidents cause by different heavy machine operations.

for “rotating-base machines”, such as excavators. The back and the front of a machine in motion are the most dangerous parts. Fisheye cameras mounted at the back and in front of the machine seem to be a good option in these cases.

We defined two positions for the fisheye camera denoted by index 1 and 2 in Fig.1 which led to two different sensors configurations of the system:

- Configuration 1: The two sensors are at position 1. In this configuration, the sensors are kept at low position (height  $h = 110cm$ ) and are parallel to the ground plane. This is the commonly used configuration of many people detection systems because the appearance of the person on the image is relatively the same at any relative position in the FOV of the camera. This advantage is not conserved in fisheye image since the object appearance distortions depends on its distance and its angle to the camera. A quantitative analysis of this distortion phenomenon is presented in [4].
- Configuration 2: The fisheye camera is mounted at position 2 (height  $h = 210cm$ , at the forks level), looking down with an angle of  $30^\circ$ . The Lidar is at the same position as in configuration 1. The high position of the camera is commonly used in the context of heavy machines. The reasons are to get a better coverage of blind angles around the machine and to avoid collisions that can damage the sensors. In our case, it gives an additional advantage in observing the whole people appearance at a very close range, even when the person touches the machine. The main drawback is the complicated transformation of the person’s appearance obtained on fisheye images.

Let assume that the Lidar is always parallel to the ground. In the targeted application, this constraint will be broken whenever the heavy machine is moving on a non-flat ground, or simply due to large vehicle vibrations with respect to the ground. In practice, these artifacts are minor when we use fisheye camera. A way to avoid this constraint is to use a multi-layers Lidar, along with a method to track the vehicle pitch.

### B. The Acquisition System

The proposed dataset consists of images captured from one fisheye camera (*Point Grey Firefly MV USB2.0*), one conventional camera (*Sony PlayStation Eye for PS3*) and one range-sensor (*LIDAR Hokuyo UTM-30LX-EW*). The sampling frequency is  $10Hz$  for both cameras and  $40Hz$  for the

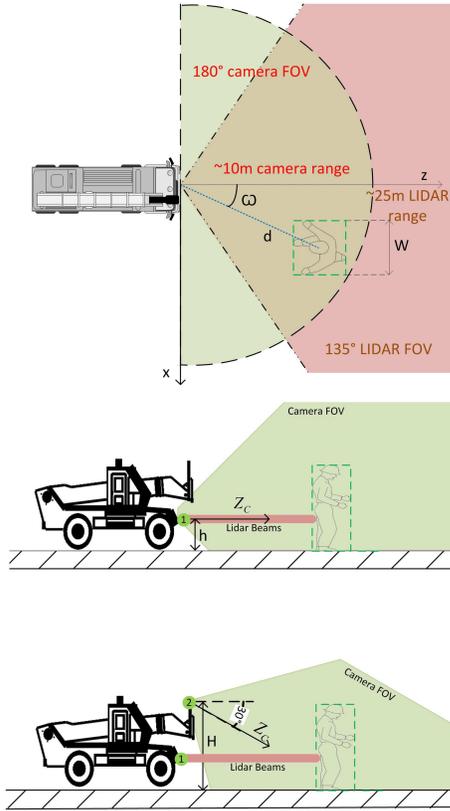


Figure 1: Sensor configurations map. Sensor heights are  $h = 110\text{cm}$  and  $H = 210\text{cm}$ .

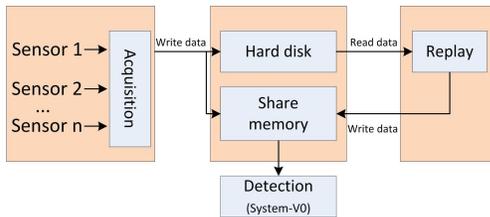


Figure 2: Architecture of the acquisition system.

Lidar. The middleware Pacpus<sup>2</sup> is used to manage the data acquisition process. Pacpus can handle data coming from different sensors in real time and save them to the hard disk driver (HDD) for replay and post-processing. These data are timestamped and synchronized so they can be used later to simulate real-time processing (Fig.2).

These data are taken on board of a telescopic-forklift, namely a Bobcat-TL470, as shown in Fig.3 (a professional driver has been recruited to operate the machine during the experiments). The experiments are divided into 4 scenarios for each configuration. All of the scenarios are pre-defined and took place under security control. In most of the data sequences, the machine moves on a clear terrain. The scenarios defined within the experiments aim to simulate frequently meet situations on a construction site. People wore

<sup>2</sup>Pacpus is the experimental platform on Intelligent Vehicles (IV) owned by Laboratory Heudiasyc. See <http://www.hds.utc.fr/pacpus>

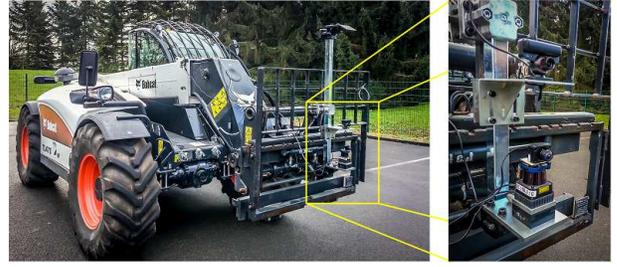


Figure 3: Real image of the acquisition system setup in the configuration 1.

Testing dataset	Number of sequences	Positive images	Positive samples
Configuration 1	7	5747	13045
Configuration 2	4	3570	9148

Table II: The testing dataset statistics.

different kind of clothes, including helmet, reflective vests and civil clothes. Different situations of occlusions were also simulated.

The fisheye images are partially annotated and used for testing purpose. The annotations for the ground truth of these image sequences are done using the labeling tool of Dollár et al [9]. This tool requires drawing the bounding box around objects in some key-frames and provides linear interpolation to infer the bounding boxes of the same object in intermediate frames. The objects can be labeled, in our case as: “person”, “person\_sitting” and “occluded”. In the evaluation, only “person” label are considered for instance. Tab.II summarizes mains characteristic of image sequences used in our tests.

### III. FROM LIDAR POINTS TO IMAGES REGION OF INTEREST

The sliding-window is the simplest candidate generation method [6] which makes the bridge between feature extraction and classification modules by scanning the entire spatial support of the image at different scales. One of the key problems of this popular method is the huge amount of negative candidates which are potentially false alarms. Rather than applying sliding windows detection techniques over the whole image, we propose to use a range sensor (here a Lidar) to roughly localize all potential obstacles inside the field-of-view (FOV) of the camera. The amount of negative candidates are significantly reduced without missing the positive candidate. Fig.4 shows the Lidar-based sensor fusion architecture used within our system. In this section, we present the whole process starting from the Lidar segmentation, through obstacles definition, coordinate transformation and ROIs localization. Preprocessing is the decisive phase since failures during this stage will strongly affect all the subsequent modules.

#### A. Lidar Data Segmentation

The Lidar data segmentation process is the primary stage for obstacle detection. An entity of interest that we call here

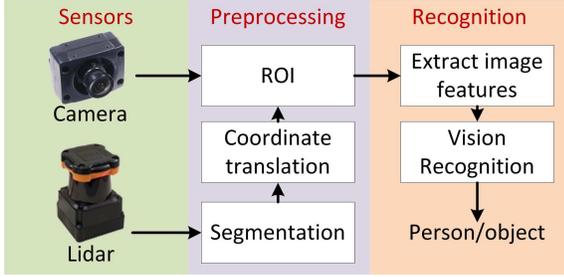


Figure 4: The Lidar-based sensor data fusion architecture.

an obstacle, can be anything: buildings, vehicles, people, etc. The goal of our system is to localize an obstacle and to classify it as a “person” or a “non-person”. The localization is done by performing a segmentation on the Lidar data.

The raw Lidar data is a 3D cloud of points. A single 2D scan is a sequence of  $N_S$  laser measurements  $S = \{(\theta_l, d_l) | l = 1, \dots, N_S\}$  where  $(\theta_l, d_l)$  denotes the polar coordinates of the  $l^{th}$  scan point. Each obstacle is characterized by a cluster of Lidar points. The cluster  $S_k$  is obtained by a convolution operation of the points in  $S$  with a kernel mask  $[-1, 1]$  as proposed in [21]. Those points whose distances to neighbor points are smaller than a given threshold  $\Theta$ , belong to the same cluster. A cluster  $S_k$  can be expressed as  $S_k = \{(\theta_m, d_m) | m = b_k, \dots, b_k + n_k\}$  where  $n_k$  and  $b_k$  are respectively the number of points and the initial scan point of the cluster  $S_k$ .

Now, an obstacle is represented as a cluster  $S_k$  in the Lidar frame. In practice, we are interested only in the centroid point  ${}^{\mathcal{L}}\mathbf{C}_k = (\theta_{\mathcal{C}_k}, d_{\mathcal{C}_k})$  and the size  $D_k$  of the cluster  $k$ . In the rest of the paper, we assume that the size of a cluster is represented as the Euclidean distance between its endpoints.

### B. Lidar Points Projection In the Camera Frame

The rigid body transformation between the Lidar  $\mathcal{L}$  and the camera  $\mathcal{C}$  frames is denoted as  ${}^{\mathcal{C}}[\mathbf{R}, \mathbf{t}]_{\mathcal{L}}$  where  ${}^{\mathcal{C}}\mathbf{R}_{\mathcal{L}}$  is the rotation matrix and  ${}^{\mathcal{C}}\mathbf{t}_{\mathcal{L}}$  is the translation vector. A point  ${}^{\mathcal{L}}\mathbf{p}$  in the Lidar frame is projected into the camera frame as:

$${}^{\mathcal{C}}\mathbf{p} = {}^{\mathcal{C}}\mathbf{R}_{\mathcal{L}} \cdot {}^{\mathcal{L}}\mathbf{p} + {}^{\mathcal{C}}\mathbf{t}_{\mathcal{L}} \quad (1)$$

The extrinsic parameter  ${}^{\mathcal{C}}[\mathbf{R}, \mathbf{t}]_{\mathcal{L}}$  can be obtained through the calibration process presented in [13] or deduced from direct measurements.

### C. Fisheye Camera Model

Wide-angle cameras show noticeable geometric distortions. In order to follow the pinhole projection model, it is generally desirable to remove them in many applications in computer vision. The geometric distortions include two major components: radial and tangential components. Radial distortions cause image points to be translated by a proportional amount to their radial distance from the optical center. Tangential distortions (or decentering distortions) are generally less significant than radial distortions and are mainly originated from the misalignment of the lens elements.

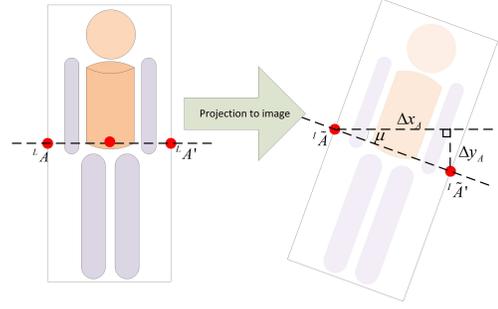


Figure 5: Computation of the ROI's angle on image

Given a 3D point  ${}^{\mathcal{C}}\mathbf{P} = (X, Y, Z)^T$  in the camera frame, the projected undistorted point on the image sensor will be represented as  $\mathbf{p} = (u, v)^T = \left(X \frac{f_x}{Z}, Y \frac{f_y}{Z}\right)^T$  with  $f_x$  and  $f_y$  the focal lengths of the camera optic. In the case of a fisheye camera, the position of the distorted point on the image is given by [15]:

$$\tilde{\mathbf{p}} = \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \mathbf{p} + \delta\mathbf{p} + \mathbf{p}_0 \quad (2)$$

where  $\delta\mathbf{p} = \begin{pmatrix} \delta u \\ \delta v \end{pmatrix} = \begin{pmatrix} \delta u^{(r)} + \delta u^{(t)} \\ \delta v^{(r)} + \delta v^{(t)} \end{pmatrix}$  is the approximated distortions and  $\mathbf{p}_0 = (u_0, v_0)$  is the principal point of the camera. The variables  $\delta u^{(r)}$ ,  $\delta v^{(r)}$  represent the radial distortions and  $\delta u^{(t)}$ ,  $\delta v^{(t)}$  are the tangential distortions along the two image axes.

Among different distortion models, the standard polynomial model [16] is commonly used. In this paper, we use a third degree polynomial model:

$$\begin{pmatrix} \delta u^{(r)} \\ \delta v^{(r)} \end{pmatrix} = \begin{pmatrix} \tilde{u}(k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6) \\ \tilde{v}(k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6) \end{pmatrix} \quad (3)$$

$$\begin{pmatrix} \delta u^{(t)} \\ \delta v^{(t)} \end{pmatrix} = \begin{pmatrix} 2p_1 \tilde{u}\tilde{v} + p_2(r_d^2 + 2\tilde{u}^2) \\ p_1(r_d^2 + 2\tilde{v}^2) + 2p_2\tilde{u}\tilde{v} \end{pmatrix} \quad (4)$$

The optimal values of  $(f_x, f_y, u_0, v_0)$  and  $(k_1, k_2, k_3, p_1, p_2)$  are estimated through a calibration process [3], [19].

### D. Adaptive ROI Projection

For each detected obstacle from the Lidar data segmentation, we project three characteristic points  $({}^{\mathcal{L}}\mathbf{C}_k, {}^{\mathcal{L}}\mathbf{A}, {}^{\mathcal{L}}\mathbf{A}')$  as  $({}^{\mathcal{I}}\mathbf{C}_k, {}^{\mathcal{I}}\tilde{\mathbf{A}}, {}^{\mathcal{I}}\tilde{\mathbf{A}}')$  on the fisheye image. The entities  ${}^{\mathcal{L}}\mathbf{A}$  and  ${}^{\mathcal{L}}\mathbf{A}'$  are two 3D real points that create a horizontal line through the obstacle following the projection process described in sections III-B and III-C. We denote  $(\Delta x, \Delta y) = {}^{\mathcal{I}}\tilde{\mathbf{A}} - {}^{\mathcal{I}}\tilde{\mathbf{A}}'$ . As depicted in fig.5, the angle between the obstacle appearance and the horizontal line in the fisheye image is therefore given by  $\mu = \arctan \frac{\Delta y}{\Delta x}$ .

In the proposed approach, we do not use the geometric characteristic of the cluster as a cue to recognize the obstacle. Only the cluster size helps to determine the size of the ROI in the fisheye image:

$$ROI \ width = \frac{D_k \times Frame \ width \times f_y}{d_{\mathcal{C}_k} \times Sensor \ width} \quad (5)$$

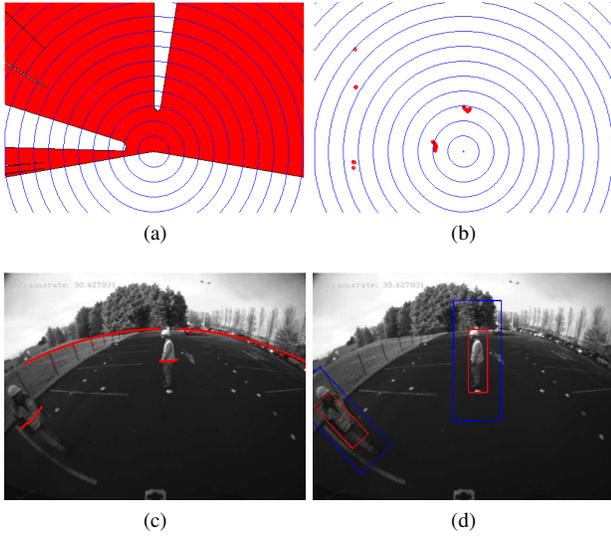


Figure 6: An example of the Lidar-camera people detection. (a) The Lidar data presented in Top-view angle. (b) Result of the Lidar segmentation. (c) Lidar points projected on image. The maximum range is limited to 10 m. (d) Obstacles' ROI are represented by a blue box on image. The result of the person recognition algorithm are represented with a red box inside the ROIs.

For any configuration of the Lidar/Camera sensors, one Lidar cluster defines one  $ROI = (\text{centroid point}, ROI \text{ width}, \mu)$ . This is very convenient for an image-based people detection algorithm and the requirement of a rotation invariant method is not necessary anymore.

#### IV. EXPERIMENTAL RESULTS

The dataset and the evaluation protocol have been presented in section II and IV-A. Our evaluation focus on the role of the Lidar in the fusion system with two different sensors configurations as stated in section II-A.

##### A. Evaluation protocol

The detection system takes an image and returns bounding boxes with corresponding scores or confidence indicators. A detected bounding box  $A$  and a ground truth bounding box  $B$  form a match if they have a sufficient overlap area. In the PASCAL challenge [11], the overlap criterion between the two bounding box  $A$  and  $B$  is defined as  $t = \frac{A \cap B}{A \cup B} > t_0$  where  $t_0$  is a given threshold. A value of  $t_0 = 0.5$  is considered as reasonable and is commonly used. The protocol of evaluation is adapted from the tool used in [9]. As the context of heavy machines requires reducing the false detection rate, the results are presented in miss-rate against false positive per image (FPPI). Only bounding boxes with a height more than 50 pixels are considered in the evaluation. Each detected bounding box may be matched once with the ground truth and redundant detections are considered as false positives (FP).

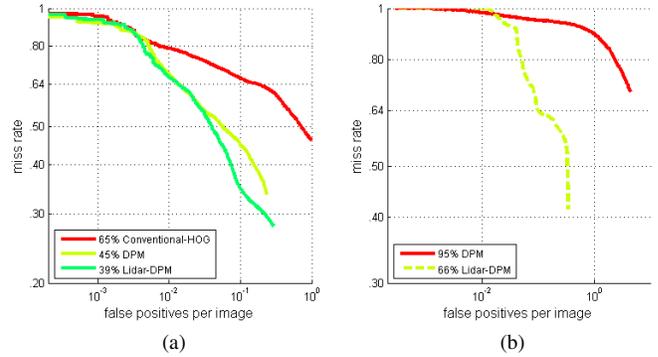


Figure 7: Comparing the detection performance of the vision-based approach versus the fusion system Lidar-fisheye camera using ROC curve (a) Configuration 1.(b) Configuration 2.

In Sect. IV-B, we plot the miss-rate versus FPPI (lower curves indicate better performance) by varying the threshold of the detector. Decreasing the threshold level results in reducing the miss-rate but will increase false positive per image. At a given FPPI, it is easy to compare the performance of different detectors. The log-average miss-rate is used to summarize detector performance. The log-average miss-rate is computed by averaging miss-rate at nine FPPI rates evenly spaced in log-space in the range  $10^{-2}$  to  $10^0$  (for curves that end before reaching a given FPPI rate, the minimum achieved miss-rate is used). When curves are somewhat linear in this range, the log-average miss-rate is similar to the performance at  $10^{-1}$  FPPI [9], [17]. The displayed legend entries are ordered by log-average miss-rate from the worst to the best case.

##### B. Results

1) *Configuration 1*: In terms of computational cost as expected, the combination of a Lidar with a fisheye camera improves significantly the speed of the people detection system. While the complexity of the obstacles detection in Lidar data and in the ROIs projection states are always the same, the speed of the people detection by DPM approach depends on the area of the ROIs. As a result, we noticed a speed factor of about 20 in detection time, measured using 300 images at VGA resolution on a desktop computer without any specific hardware acceleration. It is worth noting that the motions of heavy machines are normally much slower than automobiles and a detection frequency of 10Hz can be considered as enough.

The results shown in Fig.7a highlights the performance of 3 detectors: the standard HOG of [6], the DPM approach [12] and the proposed DPM approach used in combination with Lidar data segmentation. The DPM approach has a much better performance in people detection on fisheye images than the HOG detector, thanks to the flexible model of the deformable parts. Moreover, introducing a Lidar in a multi-sensor system apparently get an even better result by

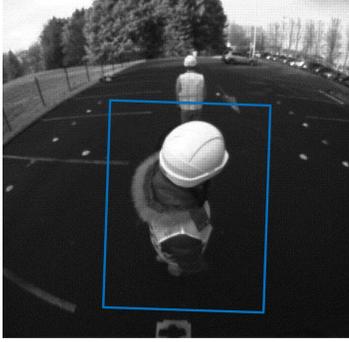


Figure 8: Extreme case where the DPM approach fails.

eliminating false detections in regions where it is impossible to have an obstacle. The measurements given by the Lidar are generally precise and robust, so the risk that the Lidar misses an obstacle is minor.

2) *Configuration 2*: Fig.7b shows the performance of the Lidar-Camera system compare to the Camera-only approach in configuration 2. People staying on the sides of the sensors are at the same time strongly distorted and rotated. Visual-based person detection methods which are not designed to take into account these rotations, have very bad performance. The system's miss detection rate reach 95% over the range of FPPI between  $10^{-2}$  and 1, which could be considered as totally failed. With the help of the Lidar, the ROIs corresponding to obstacle are rotated (detailed in section III-D) before launching the visual recognition. In this favorable case, the DPM approach performs much better. Most of the miss detections are due to the low resolution of the appearance on image when the person is staying at a long distance (over 5m).

There are also extreme cases where the fisheye camera gives a close to top-view figure of the person (Fig.8). The ROIs given by the Lidar is correct but the DPM detector cannot recognize the person. These specific problems can be solve by defining a mixture-model system where dedicated DPM model is trained in function of the position of the person.

## V. CONCLUSIONS

Given that the DPM approach shows very good performance in detecting and recognizing objects in perspective images, we built a fisheye image dataset and evaluated the DPM approach in the context of people detection using fisheye images for heavy machines applications. It turned out that the deformable models can handle very well the strong fisheye distortions, however requires high computation cost and is not invariant to rotation of the person's appearance. Therefore, we proposed a multi-sensor system that consists of a fisheye camera and a Lidar in order to bypass these limitations. The results are promising both in term of speed and performances. Moreover, the fusion architecture and the Lidar data segmentation algorithm used in our system are sequential and very simple. In the near future, we are planning to integrate an object tracking algorithm to improve the

quality of the Lidar segmentation and to solve the occlusion problems during the people detection phase.

## REFERENCES

- [1] K.O. Arras, O.M. Mozos, and W. Burgard. Using boosted features for the detection of people in 2d range data. In *IEEE International Conference on Robotics and Automation*, pages 3402–3407. IEEE, 2007.
- [2] Pradeep K Atrey, M Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S Kankanhalli. Multimodal fusion for multimedia analysis: a survey. *Multimedia systems*, 16(6):345–379, 2010.
- [3] J.Y. Bouguet. Camera calibration toolbox for matlab. 2004.
- [4] M. T. Bui, V. Frémont, D. Boukerroui, and P. Letort. People detection in heavy machines applications. In *IEEE International Conference on Cybernetics and Intelligent Systems*, 2013.
- [5] T. Bülow. Spherical diffusion for 3d surface smoothing. *IEEE Trans. Pattern Anal. Machine Intell.*, 2004.
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [7] K. Daniilidis, A. Makadia, and T. Bulow. Image processing in catadioptric planes: Spatiotemporal derivatives and optical flow computation. In *IEEE Workshop on Omnidirectional Vision*, 2002.
- [8] P. Dollár, S. Belongie, and P. Perona. The fastest pedestrian detector in the west. *British Machine Vision Conference*, 2010.
- [9] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. Pattern Anal. Machine Intell.*, 2011.
- [10] M. Enzweiler and D.M. Gavrila. A multi-level mixture-of-experts framework for pedestrian classification. In *IEEE Transaction on Image Processing*, 2011.
- [11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 2010.
- [12] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Trans. Pattern Anal. Machine Intell.*, 32(9):1627–1645, September 2010.
- [13] V. Fremont, S. Rodriguez, and P. Bonnifait. Circular targets for 3d alignment of video and lidar sensors. *Advanced Robotics*, 28:2087–2113, 2012.
- [14] Fernando García, Arturo de la Escalera, José María Armingol, and Felipe Jiménez. Context aided fusion procedure for road safety application. In *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on*, pages 407–412. IEEE, 2012.
- [15] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
- [16] C. Hughes, M. Glavin, E. Jones, and P. Denny. Wide-angle camera technology for automotive applications: a review. *IEEE Trans. Intell. Transport. Syst.*, March 2009.
- [17] M. Hussein, F. Porikli, and L. Davis. A comprehensive evaluation framework and a comparative study for human detectors. *IEEE Trans. Intell. Transport. Syst.*, Sep 2009.
- [18] J. Marsot, P. Charpentier, and C. Tissot. Collisions engins-piétons, analyse des récits daccidents de la base epicea. *Hygiene et Sécurité du Travail*, 2008.
- [19] C. Mei and P. Rives. Calibration between a central catadioptric camera and a laser range finder for robotic applications. In *IEEE International Conference on Robotics and Automation*, pages 532–537. IEEE, 2006.
- [20] L. Oliveira and U. Nunes. Context-aware pedestrian detection using lidar. In *IEEE Intelligent Vehicles Symposium*, pages 773–778. IEEE, 2010.
- [21] Cristiano Premebida, Gonçalo Monteiro, Urbano Nunes, and Paulo Peixoto. A lidar and vision-based approach for pedestrian and vehicle detection and tracking. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 1044–1049. IEEE, 2007.
- [22] Cristiano Premebida and Urbano Nunes. Fusing lidar, camera and semantic information: A context-based approach for pedestrian detection. *The International Journal of Robotics Research*, 32(3):371–384, 2013.