People Detection in Heavy Machines Applications

Manh Tuan Bui, Vincent Fremont, Djamal Boukerroui, Pierrick Letort

People Detection in Heavy Machines Applications

M. Bui[1,2], V. Frémont[1,2], D. Boukerroui[1,2], P. Letort[3]

***Abstract*— In this paper we focus on improving the performance of people detection algorithm on fish-eye images in a safety system for heavy machines. Fish-eye images give the advantage of a very wide angle-of-view, which is important in the context of heavy machines. However, the distortions in fish-eye images present many difficulties for image processing. The underlying framework of the proposed detection system uses Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM). By analyzing the effect of distortions in different regions in the field-of-view and by adding artificial distortions in the training process of the binary classifier, we can obtain better detection results on fish-eye images.**

***Index Terms*— Heavy machines, pedestrian detection, fish-eye, radial distortion, histogram of oriented gradients, machine learning, support vector machine.**

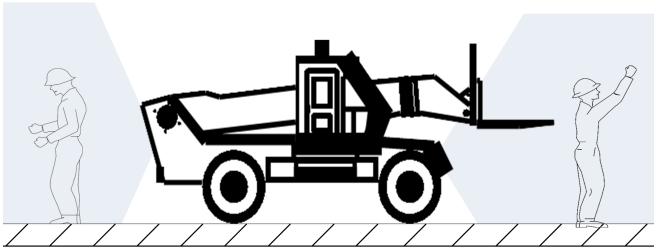

Fig. 1: The proposed configuration of cameras on a heavy machine: close areas in front and at the back of the machine are covered with fish-eye cameras.

I. Introduction

Construction sites are considered as a high risk working environment. People who work near heavy machines are constantly at risk of being struck by a machine or its components. Accidents between machines and people represent a significant contribution to construction health and safety hazards. It is hard for the drivers to keep watching all around their vehicle and fulfill their productive task at the same time, due to the complicated shape of these machines. It is therefore mandatory to develop an advanced driver assistance systems (ADAS) to help the driver watching the surrounding area and being able to raise a pertinent alarm when people are threatened. Notwithstanding many years of progress, safety system for people working around heavy machine is still an unresolved issue.

Various kinds of sensors have been tested and compared, individually or combined, but each one has some drawbacks. Range sensors, like radar, Light Detection And Ranging (Lidar) and ultrasonic, which have good performance in detecting obstacles, are unable to distinguish between objects and people. Heavy machines often work in complicated terrains with a lot of nearby objects. 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. Radio-frequency identification (RFID) technology is a much more popular sensor used on heavy machines and it actually very useful [1]. The only drawback is the management of RFID tags. Only people with the tag are protected. It is not always the case with open construction sites where the access is not controlled. The obligation of keeping the tag on them can be also an issue with the employees. The last commonly used sensor is the camera. It offers the best option as a low-cost and polyvalent sensor. Image processing provides the ability to recognize various kinds of objects, including people.

To our knowledge, most existing vision systems in this context do not integrate recognition functions. For example, Caterpillar develops an "Integrated object detection system" on their machines which is claimed to work on very harsh condition. Briefly, it is an obstacle detection system by radar with cameras assistance for visualization[1]. Camera-based systems are also provided by other manufacturers like Motec, Orlaco, Waeco. To the best of our knowledge, there is only one product on the market that provides vision-based assistance for obstacle and human detection on construction machine: the Blaxtair system from Arcure SA[2]. It is a stereo vision based system that detect obstacles using the depth map. In order to reduce the complexity and computation resource, the recognition algorithm is applied only on one of the images and only in regions of interest (ROIs). The ROIs correspond to the positions of the detected obstacles. This kind of system is widely used in the automobile sector.

Recently, pedestrian detection system on automobile, which share a lot of characteristics with the context of heavy machines, has known important progresses [2], [3] . Although the problematic is similar in both contexts, we can clearly distinguish the two. In the automobile field, cars need to stop if there is an obstacle, no matter if it's a pedestrian or an object. The task of recognizing people is more important for heavy machines where the main requirement is human's safety. Besides, cars often operate at a higher speed. While it is important for the system on automobile to be able to detect people at far distances, heavy machines need a larger field of view (FOV) to cover the nearby area. Construction machines often have a complicated shape and large size, which can also

The authors are with [1]*Université de Technologie de Compiègne (UTC)*, France, [2]*CNRS Heudiasyc UMR* 7253*, France,* [3]*Technical center for the Mechanical Industry (CETIM), France*

[1]http://safety.cat.com/cda/layout?m=154441&x=7&f=399105
[2]http://www.arcure.net/2-33976-BLAXTAIR.php

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.

benefit from the large FOV. One other essential difference is that when cars run on the road, even in worse case, there will always be a dominant plane. Hypothesis about this plane are important to detect obstacles and region of interest (ROI) in view frames. This hypothesis is not always true for the heavy machinery environment. Harsh working conditions are also a challenge but it is not covered in this work.

Based on the survey of accidents caused by heavy machine in France for the period from 1997 to 2008 [4], the system requirements and the configuration of sensors are defined. *Table* I shows that the danger is highly dependent on the action taken by the machine and its direction. Accidents rarely happen on the sides of the machines, except for "rotating-base machines", such as excavators. The back and the front of a machine in motion are the most dangerous parts. Fish-eye camera mounted at the back and in front of the machine seem to be a good option (see Fig.1).

Our contribution in this paper lies at the analysis of the influences of radial distortions in fish-eye images on the people detection algorithm. An approach using enhanced training dataset is proposed to bypass these influences. The paper is organized as follows. First, a brief description of the characteristics of fish-eye cameras is discussed in section II. Section III presents our vision-based people detection, the underlying assumptions together with the analysis on the distortion of human appearance. The proposed solutions are then details in section IV. Results of the experiments are shown in section V followed by the conclusion section.

II. Fish-eye cameras

A. Distortion model

Wide-angle cameras have noticeable geometric distortions. While these distortions may be artistically interesting, it is generally desirable to remove them in many applications in computer vision. The geometric distortions include two major components: radial and tangential. Radial distortion causes image points to be translated by an amount proportional to their radial distance to the optical center. Tangential distortions (or decentering distortion) are generally less significant than radial distortions and are produced by the misalignment of the optical centers of various lens elements.

Given a real point $\boldsymbol{P} = (X, Y, Z)^T$, the undistorted point projected on the image sensor will be represented as $\boldsymbol{p} = \begin{pmatrix} u & v \end{pmatrix}^T = \begin{pmatrix} X\frac{f_x}{Z} & Y\frac{f_y}{Z} \end{pmatrix}^T$ with $f_x$and f_y the focal length of the camera optic. In the case of a wide-angle camera, the position of the distorted point on image is given by:

$$\tilde{\boldsymbol{p}} = \begin{pmatrix} \tilde{u} \\ \tilde{v} \end{pmatrix} = \boldsymbol{p} + \delta\boldsymbol{p} + \boldsymbol{p_0} \tag{1}$$

where $\delta\boldsymbol{p} = \begin{pmatrix} \delta u \\ \delta v \end{pmatrix} = \begin{pmatrix} \delta u^{(r)} + \delta u^{(t)} \\ \delta v^{(r)} + \delta v^{(t)} \end{pmatrix}$ is the approximated distortion and $\boldsymbol{p_0} = (u_0, v_0)$ is the principal point of the camera. $\delta u^{(r)}$, $\delta v^{(r)}$ represent radial distortions and $\delta u^{(r)}$, $\delta v^{(r)}$ are the tangential distortions along the two image axes.

Among different distortion models, the standard polynomial model is the most popular [5]. In this paper, the standard polynomial model at third degree is used:

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

$$\begin{pmatrix} \delta u^{(t)} \\ \delta v^{(t)} \end{pmatrix} = \begin{pmatrix} 2p_1\tilde{u}\tilde{v} + p_2(r_d^2 + 2\tilde{u}^2) \\ p_1(r_d^2 + 2\tilde{v}^2) + 2p_2\tilde{u}\tilde{v} \end{pmatrix} \tag{3}$$

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. It is a prerequisite for any accurate geometric measurements from image data. Most of calibration methods use known geometric patterns, such as corners, dots, circles, lines and other features that can be easily extracted from images. The method described in [6] and implemented in [7] was used in our work

B. Warping fish-eye images

Given calibration information of a camera system, it is possible to remove distortions and to apply any operator in a local perspective plane. Wrapping a fish-eye image into a local perspective image is the direct way to avoid non-perspective deformations. Unfortunately, besides adding computational load, this approach also creates undesirable effects. It take about $60ms$ to warp an image at the VGA resolution (640×480)[3], which cannot be ignored in real-time applications. The non-uniform compression of the image structures (stretching effect) is a consequence of the wide-angle lens mechanism. There are more details about the scene on the center than on the edges of the image. In other words, the image sampling frequency decreases proportionally with the distance to the image center. The image rectification process wraps the distorted wide-angle image into a local perspective plane. As a result, the boundaries of a rectified image will contain vacant pixels. These pixels are not directly mapped from the original image but often deduced from the neighbor pixels by performing interpolation method. Image information at the boundaries of a rectified image is very low because of blurred. Basically, image processing operators are directly and uniformly applied on the whole rectified image without consideration of these non-uniform degradations. The advantages of the fish-eye camera lie at the large field-of-view (FOV) so these boundaries areas are very important. A

[3]Measured on a Windows 64-bit PC (CPU inter core i5 2.5Ghz), by using the library OpenCV 2.4.

drop in performance is observed in our experiments (section V-B).

Daniilidis [8] and Bulow [9] are two of the first researchers who argued that the warping of wide-angle images should be avoided. Recently, there are others researchers who proposed approaches to increase matching rate of Scale-invariant feature transform (SIFT) for wide-angle images [10], [11]. Using fish-eye cameras in recognition applications turn out to be an interesting research subject.

III. People detection on fish-eye image

A. Related works

There is a quite large literature on people detection, dating back to [12], [13], [14]. Better features or combination of features, classifiers, prior knowledge and dataset have helped people detection algorithms to become much more reliable. The readers can refer to the recent surveys [15], [16] for more details. People detectors typically follow a sliding window paradigm which entails feature extraction and binary supervised classification. Nearly all recently detectors use some forms of oriented gradient histograms features. Support vector machines and boosting are used in majority of the cases for classification. Recent works are more focused on combining multicue, multi-sensors [17], [18] or on handling the pose variation and occlusion problems [19], [20]. To the best of our knowledge, there have been few works exploiting fish-eye cameras for people detection.

The underlying people detector used in this paper was implemented as described in [14] and [21]. Histogram of oriented gradients (HOG) features were extracted using the integral histogram method [22]. Training and prediction processes are done by light SVM [23]. The detection is done by sliding a window with dense multiscale scanning. A comparison of this detector to others key methods of people detection is shown in experimental section (V-B).

B. Distortion analysis

The distortion of a wide-angle camera is not identical over all the image area. It particularly affects the detection at close range and at the image boundaries. The measurement of geometric distortion of fish-eye images and how it affects the detection performance is not well studied in the literature. We tried to illustrate these distortions in function of the relative position between the fish-eye camera and a person of an average width and height $W \times H$ (see Fig.2).

The center point $\boldsymbol{P_0}$ of a person in the coordinate of the camera is projected on the image plane as described in section II-A. The distortion of one point is computed relatively to the center point as $\delta p_0 - \delta p_{ij}$. The mean square error (MSE) of all points belonging to the rectangular $\Re$ of size $W \times H$ around the person is:

$$MSE_{\boldsymbol{P}_0} = \frac{\sum(\delta p_0 - \delta p_{ij})^2}{w \cdot h} \quad i,j \in \Re \qquad (4)$$

Although, the optic is open up to $90°$ on each side, a person is not fully visible on the image for angles higher than $60°$ and a distance closer to $0.7m$. The curves in Fig.2

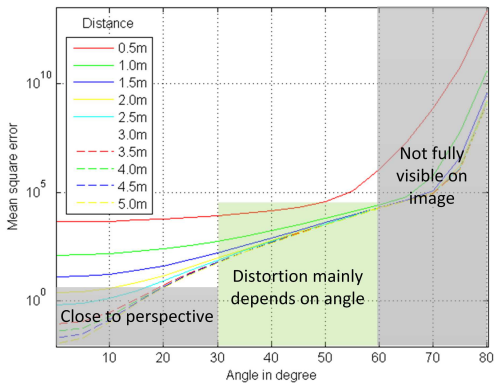

Fig. 2: Mean-square-error versus the relative position of a person to the camera.

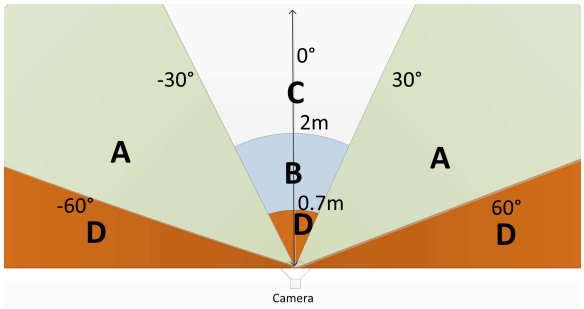

Fig. 3: Regions on the field-of-view of fish-eye camera:. In zone A and B, people appearance have significant distortion. In zone C, the distortion is not noticeable. In zone D, people are not fully visible on image.

show that the distortion of a person at a distance higher than $2m$ between $0°$ and $30°$ can be ignored. The most interesting range to be considered is therefore between $30°$ and $60°$ where the distortion does not depend much on the depth distance but only on the radial angle. We can briefly say that the difference between fish-eye and perspective images lies in region A and B (see Fig.3).

IV. Proposed training and detection algorithm

A. Distortion simulation

The direct approach to solve the problem of distortion is to divide the field-of-view of the camera into sub-regions, function of angles and distances. We assume that the supervised classifier (SVM in this case) can recognize distorted pedestrian appearance when it is trained with distorted sample images. The training samples are separated into sub classes, proportional to the amount of distortion. Wide-angle optics are very different from one to another. If a training dataset is built based on one kind of optic, it is not usable with the others. Because the building process of a training and a testing databases require a significant amount of time and resources, it will not be very practical. It is possible however to artificially deform existing image databases, taken by perspective cameras, in order to simulate

Fig. 4: Example of artificial distorted images at distance of 1m. From left to right: original image, distorted image at 0°, distorted image at -40°, distorted image at 40°.

wide-angle camera geometry. Given an approximate size ($W \times H$) of a person and its center position in the camera frame, a distorted appearance can be calculated using the camera projection equation and a given distortion model (see Eq.1). The algorithm is described in Algo.1 and example of distortion simulation is illustrated in Fig.4.

Algorithm 1 Distortion process

Input: - Height of the camera H_0 and average real size of a person $W \times H$
- Sample image of person at resolution $a \times b$
- Position $P_0 = (X_0, Y_0, Z_0)$

Output: - Distorted image correspond to position P_0 at resolution $a \times b$

1: ▶ Compute size on image of the person by Eq.1
2: ▶ Resize sample image to resolution $w \times h$
3: ▶ Each pixel $p = (i, j)$ from sample image correspond to a real point $P = (X, Y, Z)$ where

$$\begin{cases} X = X_0 \\ Y = (H_0 - H) + j\frac{H}{h} \\ Z = Z_0 - \frac{W}{2} + i\frac{W}{w} \end{cases} \quad with \quad \begin{matrix} i \in [0, w] \\ j \in [0, h] \end{matrix}$$

4: ▶ Project all points P on image as described in II-A
5: ▶ Crop and resize to original size $a \times b$

There is a trade-off between the image quality and the amount of distortion added to the training examples. This process has the drawback of introducing missing pixels that have to be filled by interpolation. This phenomenon is proportional to the amount of distortion and we believe that it may have an effect on the performance of the detector. In practice, the quality of the image samples simulated at angle superior to 60° are unusable because the samples loose most image details.

B. Proposed people detection approach

Using distorted images generated by the method mentioned above (IV-A), we propose two people detection approaches in fish-eye images. To this end, the angles $\Theta = [-45°, 0°, 45°]$ and the distance $D = 1.5m$ were chosen to represent the critical areas A and B in the FOV as shown in III-B.

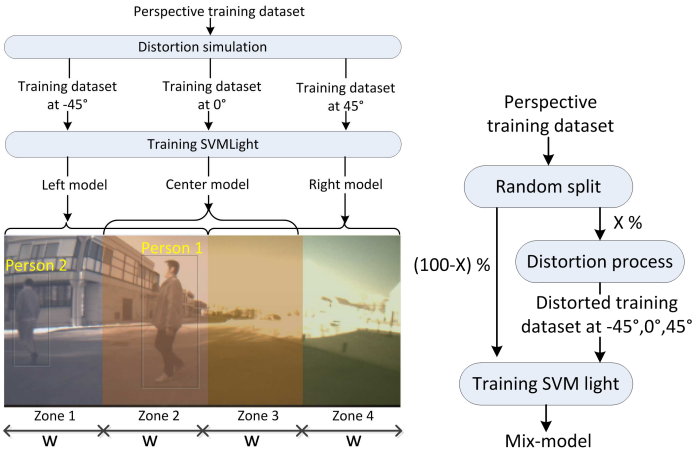

Fig. 5: Flow chart of proposed people detection algorithms: (a) Multi-angles approach. (b) Mix-training-dataset approach.

1) Multi-angles approach: 3 detectors (left, right and center) are trained by 3 different distorted datasets corresponding to the 3 angles in Θ. The detection follows a sliding-window paradigm with dense multiscale scanning (rescaling factor $e = 1.1$). The three specialized detectors operates on 3 overlapping image areas. The left-model takes care of the first and second zones; the center-model works on two center zones and the right-model operates in the third and forth zones (see Fig.5a). Overlapping the detection zones avoids occlusions when a person is at the frontier between two zones. This approach needs however a classifier fusion mechanism at the overlapping areas; here zone 2 and 3 as shown in Fig.5a. The winner takes all approach is adopted in our work.

2) Mix-training-dataset approach : In this approach we use one classifier only. The classifier is however trained on sample images without distortions and with simulated distortions at different rates. Starting from a training dataset without distortions, we replace randomly a percentage of undistorted images by distorted ones. The latter are simulated at different distortion angles. Therefore the total number of positive and negative sample images in the training dataset is the same in all cases. After training, the detector is applied on the whole image(see Fig.5b). Percentage of the distorted images in training dataset can vary and its effect on the performance of the detector is analyzed in section V-B.

V. Experiments and Results

A. Evaluation method

The detection system takes an image and return 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 [24] and the survey of Dollár [15] et al., the overlap criterion between the two bounding box A and B is $t = \frac{A \cap B}{A \cup B} > t_0$ where t_0 is a threshold. $t_0 = 0.5$ is

Label	"person"	"person-occluded"	"ignore"
Percentage of occlusion	<20%	>20% and <60%	>60%

TABLE II: Labels used in evaluation

considered reasonable and is commonly used. The protocol of evaluation is adapted from the tool of Dollár which was use in [15]. As the context of heavy machines requires to reduce false detection rate, the result is represented in miss rate against false positive per image (FPPI).

Only bounding boxes with height more than 50 pixels are considered. This is reasonable because the smallest sliding window used in our tests is of 48×96 pixels and there is no upsampling applied to detect smaller objects. Each detected bounding box may be matched once with the ground truth and redundant detections are consider false positive.

We have built a test dataset with 7 image sequences of 3200 images captured by a fish-eye camera (Firefly-MV from Pointgrey, angle-of-view is up to $180°$). The sequences include indoor and outdoor scenes with different backgrounds. The camera is held at the height of $90cm$ and parallel to the ground. They are not taken in a crowded place, there are maximum 3 or 4 people in a frame. The annotation for the ground truth of these image sequences are done by the labeling tool of Dollár et al.. This tool require marking the bounding box around objects in some key-frames and provide linear interpolation to infer the bounding boxes of the same object in intermediate frames. The object can be labeled, in our case as: "person","person-occluded" and "ignore" (see table II). In the evaluation, only "person" label are considered.

Each detector is trained by 15 660 positive and 20 000 negative sample images taken from the Daimler dataset [16].

B. Results

The first experiment involves the HOG-SVMLight detector (conventional detector), multi-angles detector (denoted by Full-distorted) and Mix-training-dataset detectors at different percentage of distorted images (denoted by Mix-model). We plot miss rate versus FPPI (lower curves indicate better performance) and use the log-average miss rate to summarize detector performance. The log-average miss rate 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 miss rate achieved is used). When curves are somewhat linear in this range, the log-average miss rate is similar to the performance at 10^{-1} FPPI but is more stable in general [25], [15]. The displayed legend entries are ordered by log-average miss rate from the worst to the best. Fig.6 show the full image evaluation of all the detectors. Fig.7 summary the performance of detectors versus percentage of distorted image in training dataset.

The multi-angles approach which was trained with only distorted images have the worst performance. The degradation of image quality during the distortion process affects remarkably the performance of the detection. Notice however that our proposition to train the SVM classifier with distorted and non-distorted images gives better results. In Fig.6, we also show the performance of HOG-SVMLight detector on rectified test sequences to a perspective plane. The results are even worse than applying the HOG-SVMLight directly on the fish-eye images. The approach of rectified distortion might work with a small amount of distortion but it is not adapted to fish-eye optics where the angle-of-view is too large.

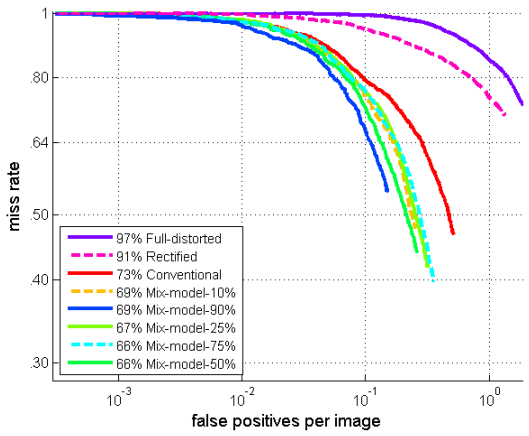

Fig. 6: Result of different detectors trained with different percentage of distorted samples on fish-eye test sequences.

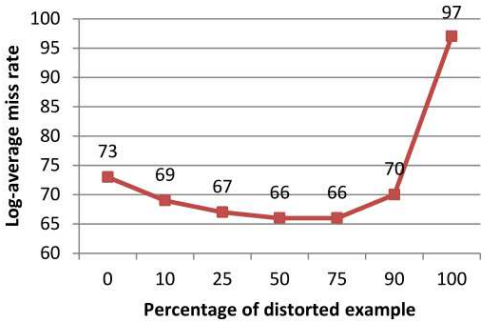

Fig. 7: Log-average miss rate versus the percentage of distorted image in training dataset.

Fig.8 shows the performances of all the detectors in function of the horizontal position of a person on the fish-eye images. Detection results are compared to the ground truth annotation on a region of 240×480 pixels. By sliding this region horizontally across the image we hope to see experimentally the effect of the distortion rate on people detection. The results are better at the center than at the boundaries of the images, which is proportional to the amount of distortion. The curves are not symmetric because people do not evenly appear across images in the test sequences. For the multi-angles detector, the performance is the same at all angles but it is hard to conclude anything because the log-average miss rate is over 95%, which is far worse than the rest.

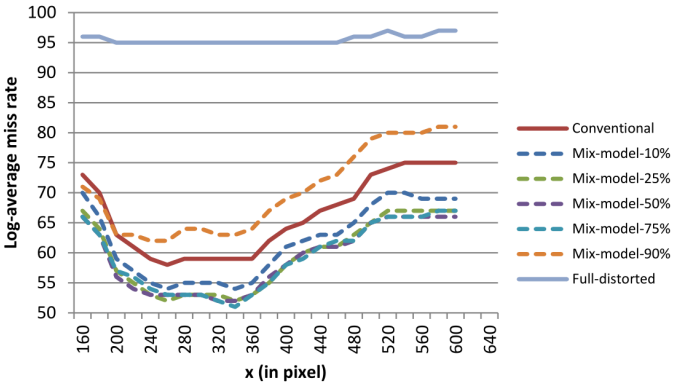

Fig. 8: Evaluation of detection performance along the horizontal axis of fish-eye images. Different detectors trained with different percentage of distorted images are compared.

VI. Conclusion

In this paper a novel approach to improve the performance of people detection on fish-eye images is proposed. It is demonstrated by the result of the experiments that enriching the training dataset can handle the distortion on the people's appearance. Such approach has the advantage of being more generic as it can be adapted to all camera optics with known distortion in order to simulate the camera distortions. Moreover, the increase of complexity is only on the training process without any influence on online detection speed.

In future work, the performance of the mix-training-dataset approach can be enhanced by increasing the quality of the distorted images. More precisely, a thorough analysis of the effect of interpolation during distorting process of sample images is needed. Additionally, a comparison of the HOG feature vectors of perspective image and distorted sample images might reveal a possibility to introduce the distortion directly on the HOG vectors without manipulating the sample image.

In order to improve the robustness of the detection, especially in the context of heavy machines, we plan to combine fish-eye camera with a range sensor (Lidar or ultrasonic). Indeed, range sensors are very helpful in accelerating the detection and reducing false positive in a complex texture background.

Acknowledgments

This work is supported by Technical Center for the Mechanical Industry (CETIM).

References

[1] S. Chae and T. Yoshida, "Application of rfid technology to prevention of collision accident with heavy equipment," *Automation in Construction*, 2010.

[2] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: Single-frame classification and system level performance," in *IEEE Intelligent Vehicles Symposium*, 2004.

[3] M. Enzweiler and D. Gavrila, "A multi-level mixture-of-experts framework for pedestrian classification," in *IEEE Transaction on Image Processing*, 2011.

[4] J. Marsot, P. Charpentier, and C. Tissot, "Collisions engins-piétons, analyse des récits daccidents de la base epicea," *Hygiene et Sécurité du Travail*, 2008.

[5] 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.

[6] 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.

[7] J. Bouguet, "Camera calibration toolbox for matlab," 2004.

[8] 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.

[9] T. Bülow, "Spherical diffusion for 3d surface smoothing," *IEEE Trans. Pattern Anal. Machine Intell.*, 2004.

[10] P. Hansen, P. Corke, and W. Boles, "Wide-angle visual feature matching for outdoor localization," *The International Journal of Robotics Research*, 2010.

[11] M. Lourenço, J. Barreto, and F. Vasconcelos, "srd-sift: Keypoint detection and matching in images with radial distortion," *IEEE Trans. Robot*, 2012.

[12] D. Gavrila, M. Kunert, and U. Lages, "A multi-sensor approach for the protection of vulnerable traffic participants the protector project," in *IEEE Instrumentation and Measurement Technology Conference*, 2001.

[13] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, 2005.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[15] 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.

[16] M. Enzweiler and D. Gavrila, "Monocular pedestrian detection: Survey and experimentsd," *IEEE Trans. Pattern Anal. Machine Intell.*, Dec 2009.

[17] D. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International Journal of Computer Vision*, 2007.

[18] L. Oliveira, U. Nunes, P. Peixoto, M. Silva, and F. Moita, "Semantic fusion of laser and vision in pedestrian detection," *Pattern Recognition*, 2010.

[19] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[20] X. Wang, T. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *IEEE International Conference on Computer Vision*, 2009.

[21] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[22] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[23] T. Joachims, "Making large scale svm learning practical," 1999.

[24] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, 2010.

[25] 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.