

Estimating the 3D Position of Humans wearing a Reflective Vest using a Single Camera System

Rafael Mosberger and Henrik Andreasson

Abstract This paper presents a novel possible solution for people detection and estimation of their 3D position in challenging shared environments. Addressing safety critical applications in industrial environments, we make the basic assumption that people wear reflective vests. In order to detect these vests and to discriminate them from other reflective material, we propose an approach based on a single camera equipped with an IR flash. The camera acquires pairs of images, one with and one without IR flash, in short succession. The images forming a pair are then related to each other through feature tracking, which allows to discard features for which the relative intensity difference is small and which are thus not believed to belong to a reflective vest. Next, the local neighbourhood of the remaining features is further analysed. First, a Random Forest classifier is used to discriminate between features caused by a reflective vest and features caused by some other reflective materials. Second, the distance between the camera and the vest features is estimated using a Random Forest regressor. The proposed system was evaluated in one indoor and two challenging outdoor scenarios. Our results indicate very good classification performance and remarkably accurate distance estimation especially in combination with the SURF descriptor, even under direct exposure to sunlight.

1 Introduction

People detection is an important task in both autonomous machines and human operated vehicles equipped with driver assistant technology. Especially when it comes to applications where machines operate in industrial workspaces shared with humans, it plays a crucial role towards improved safety for the operators. Different sensor modalities have been commonly used in people detection including laser scanners, thermal cameras and vision-based systems. All approaches suffer from drawbacks

Rafael Mosberger and Henrik Andreasson, e-mail: firstname.lastname@oru.se
AASS Research Centre, School of Science and Technology, Örebro University, Sweden

in safety critical applications. Thermal cameras are expensive and their use depends on the ambient temperature. Laser scanners are also expensive and can fail under extreme conditions such as direct sunshine into the sensor. Vision-based systems offer appealing solutions since they can be inexpensive but require that the ambient illumination is neither too strong nor too weak. Yet, for the application in safety systems for industrial environments, reliable people detection in a variety of different conditions is critical.

In many industrial workplaces such as manufacturing areas, construction sites, warehouses or storage yards the wearing of a reflective safety vest is a legal requirement. In contrast to more general approaches, the work presented in this paper therefore takes advantage of the enhanced visibility of a person due to the reflective vest to facilitate the detection. Andreasson et al. [2] introduced a people detection system based on a single camera unit which was successfully used to detect humans wearing a reflective vest. The core principle of the detection system is to take two images in short succession, one with and one without IR flash, and to process them as a pair. The processing scheme identifies regions with a significant intensity difference between the two images in order to detect locations where reflective material appears.

The system proposed in this paper is an extension of the work presented in [2]. The extended system allows not only for detection of people wearing reflective vests but also adds estimation of the 3D position for individual vest features. A machine learning approach is applied in order to estimate the position of a reflective vest based on the description of an image patch extracted from the neighbourhood of the location where the vest was detected in the image.

This paper is organised as follows. Sect. 2 briefly discusses related work in the field of vision-based people detection. In Sect. 3, the complete vest detection and position estimation system is described in detail. The section is divided into a part dedicated to the detection of reflective vest features (Sect. 3.1) and a part describing the estimation of a 3D position corresponding to each detected vest feature (Sect. 3.2). In Sect. 4, the performance of the system is evaluated in different environments and conclusions and an outlook on future work are presented in Sect. 5.

2 Related Work

Vision-based people detection for non-stationary environments has been extensively studied for applications in robotic vehicles, (semi-)autonomous cars, driver assistant systems and surveillance. State-of-the-art techniques mainly rely on either the detection of individual body parts or the analysis of templates. Both techniques are commonly used in combination with machine learning techniques. The state-of-the-art method of template based techniques uses the Histogram of Oriented Gradients (HOG) descriptors [8] computed on a dense grid of uniformly spaced cells. The descriptors are fed into a detection system consisting of a binary classifier obtained by supervised learning. Body part based detection systems follow a different approach

by representing the body as an ensemble of individual parts pairwise connected with a spring-like link. In [9], the different body parts are represented using a simple appearance model and arranged in a deformable configuration to obtain a pictorial structure which is then matched to the images to be observed.

The performance of vision-based techniques heavily depends on the presence of good visible structures in the images, and thus on a sufficient illumination of the observed scene. Their application is not suitable for dim or completely dark environments. Also, vision-based approaches typically struggle in cases where people have little contrast with the background. For these reasons, existing people detection approaches are not directly applicable in a safety system supposed to operate under challenging conditions, such as rain, snow or direct exposure to sunlight.

The system presented in this paper focuses on the detection of people wearing a reflective vest using active IR illumination. The detection of retro-reflective material has been successfully applied in motion capture systems where passive markers are used in combination with an array of IR or visible-light LEDs mounted around the lens of one or several cameras in order to detect selected spots on the human body [7]. Yet, to the best of the author's knowledge, there exists no people detection system that makes use of the reflective vest properties in the detection process.

Instead of analysing single images as it is done in most of the related work, our system processes a pair of images one of which is taken with an IR flash and one without. The proposed algorithm exploits the fact that the IR flash is very strongly reflected by the vest reflectors to detect locations in the image where a large intensity difference exists between the two images. Andreasson et al. show in [2] that especially in the mid- and long-range people detection where spatial resolution decreases rapidly in the image, their approach clearly outperforms a state-of-the-art people detection algorithm (Histogram of Oriented Gradient, HOG) applied to a single image.

3 System Description

The reflective vest detection and position estimation system presented in this paper is described in two parts. Sect. 3.1 is dedicated to the detection of reflective vest features in the input images while Sect. 3.2 describes the estimation of a 3D position for each detected vest feature. For a schematic overview of the complete system, individual processing steps as well as the data flow in the system see Fig. 1.

3.1 Vest Detection

The upper part of Fig. 1 depicts the detection scheme employed to detect persons wearing reflective vests. The detection system works by comparing two images, one acquired with IR flash, I_f , and one taken without, I_{nf} . A feature detector is used to

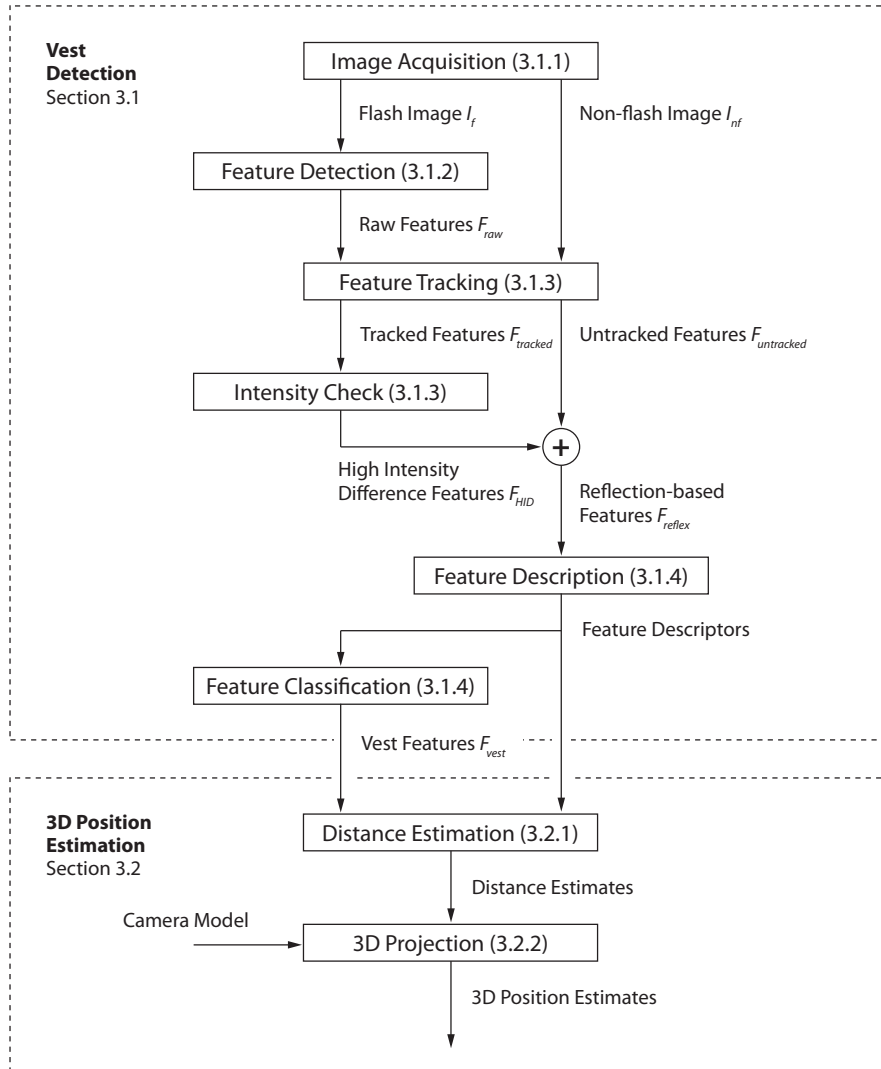


Fig. 1 Overview of the reflective vest detection and position estimation system

identify the set F_{raw} of high intensity blob-like interest points in the image I_f . Subsequently, the features detected in I_f are tracked in I_{nf} and, based on the output of the tracker, a subset of features is discarded as not belonging to reflective material and thus not originating from a reflective vest. Features are discarded if they can be tracked and if the intensity difference between the two images is below a set threshold. This pre-selection process is further described in Sect. 3.1.3. Finally, a binary random forest classifier, trained by supervised learning is used to discriminate vest features from non-vest features.

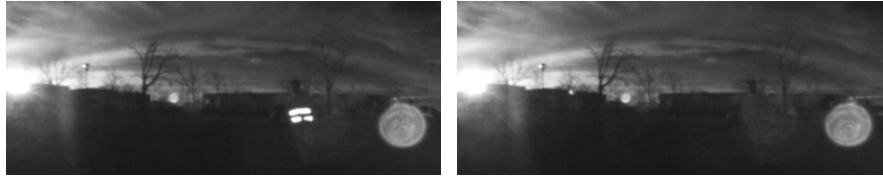


Fig. 2 Example of an image pair taken in short succession. The image on the left was taken with IR flash and the image on the right without. The images show a panoramic view which is obtained by unwrapping of the raw fish-eye images. The difference in intensity values at locations where the reflective vest appears is clearly visible. The filled white circle at the bottom right represents a lens artifact originating from direct sunshine into the camera. It may be noted that the overall brightness of the images is very low due to the use of the IR band pass filter in the camera system.

3.1.1 Hardware and Image Acquisition

The camera unit consists of a standard monochrome CMOS sensor with a resolution of 752×480 pixels and a fish-eye lens with an approximate FOV of 180 degrees. 8 IR LEDs with a wavelength of 850 nm are placed in a ring around the camera. The orientation of the LEDs assures a wide and relatively uniform illumination of the scene in the camera's FOV. A band pass filter with a center wavelength of 852 nm and a full width at half maximum of 10 nm is mounted between the lens and the sensor. The filter corresponds to the dominant IR wavelengths of the IR LEDs.

The image acquisition involves taking a pair of images, one with IR flash and one without. An exemplary pair of panoramic images, obtained from the raw fish-eye images is depicted in Fig. 2. The time increment t_a between the acquisition of the two images is kept as short as possible in order to minimize the difference between the two images due to changes in viewpoint and changes in the scene. The raw fish-eye images are unwrapped to represent a panoramic view containing the area of interest for the reflective vest detection. The unwrapped flash image I_f and the unwrapped non-flash image I_{nf} form the image pair $\mathbf{I} = (I_f, I_{nf})$ on which all the post-processing is based.

3.1.2 Feature Detection

The reflection of the IR light by the reflectors of a vest results in high intensity blob-like regions at locations where the vest appears in the image I_f . Shape and size of the high intensity regions depend heavily on the distance between the camera unit and the person wearing the vest as well as on the body pose of the person. Especially at short distances, the reflective markers of a vest appear as elongated regions rather than as circular blobs.

The first step in the vest detection process consists in identifying locations in the image I_f where such high intensity regions appear. It was shown in [2] that the STAR algorithm by Konolige et al. which is a speeded-up version of the CenSurE feature detector [1] yields good results. The STAR detector produces a set of raw

features, named F_{raw} , in which every feature is described by the image coordinate pair $\mathbf{u} = (u, v)$ indicating the location in the image I_f where it was detected.

An exemplary result of the feature detection is given by the ensemble of crosses in the upper image of Fig. 3. The example illustrates that under the influence of the IR illumination from the flash and the sun, the detected feature set F_{raw} includes many features that do not originate from a reflective vest. Also it is worth mentioning that due to the STAR algorithm's sensitivity to circular shapes, one reflective vest marker can be detected more than once (cf. Fig 3), especially when its shape appears elongated.

3.1.3 Feature Tracking and Intensity Check

The detected features in the set F_{raw} originate either from a reflective vest or from another bright object in the FOV of the camera. As the images I_f and I_{nf} were taken in short succession, the appearance of non-vest features is assumed to change little from one image to another. In contrast, this *brightness constancy* assumption is not valid for features originating from a reflective vest since the intensity values in the vicinity of a vest feature differ considerably for the image pair I . Based on this property, the first processing step to eliminate non-vest features consists in tracking raw features, detected in image I_f , in the corresponding image taken without IR flash, I_{nf} , and perform a check on the intensity difference between the image patches surrounding the detected and the tracked feature locations.

The tracking of the features is performed using a pyramidal implementation of the Lucas-Kanade feature tracker [4]. The tracker is based on the assumption that the temporal increment between two consecutive images is small enough such that the location of a feature changes little from one image to another. As the images I_f and I_{nf} are taken in very short succession, this is the case. Furthermore, the tracker assumes brightness constancy.

In the case of vest features, the tracker is typically unable to find any suitable match in the image I_{nf} because the brightness constancy assumption does not hold true. Thus, features that failed to be tracked are added to the set of reflection based features F_{reflex} . It is worth noting that in contrast to the standard application of a feature tracker, we are not only interested in features that can be successfully tracked. We specifically identify features that cannot be tracked as possible vest features.

In the case of non-vest features, where the brightness constancy assumption holds true, the tracker typically finds the corresponding location in the non-flash image I_{nf} and an intensity difference check within a square window of size w_i surrounding the feature can be performed. If the average difference between the pixels in the window is above a threshold t_i the feature is declared reflection based (and exceptionally tracked) and added to the feature set F_{reflex} . Otherwise, this is not the case and the feature is considered to originate from an area without reflective material and will not be further processed.

3.1.4 Feature Description and Classification

The set F_{reflex} typically contains features that originate from the reflection of the IR light on a reflective material. Yet, other reflective objects than the reflective vest markers can appear in the scene, such as metallic surfaces, windows, mirrors or reflective floor and wall marking tape. An additional processing step therefore aims at classifying the features collected in the set F_{reflex} into a set of vest features F_{vest} and a set of non-vest features $F_{nonvest}$.

The additional processing step is also motivated by the fact that extreme camera movements that cause strong motion blur can result in a high number of detected features that can not be successfully tracked in the image I_{nf} . In such cases, the set F_{reflex} , supposed to contain mainly features representing reflective materials, would contain many other undesired features.

The classification is not performed by directly evaluating the raw intensity values of the image. Instead, a local image descriptor is extracted from the neighborhood of each detected feature in F_{reflex} and serves as input for the classifier. The descriptor is extracted from a square image patch of size w_d centered at the location where a reflection based feature was detected in image I_f . Requirements for an appropriate descriptor include robustness to illumination changes, motion blur and noise as well as computational efficiency of the extraction process. State-of-the-art feature descriptors that were found appropriate include SURF [3] and BRIEF [6].

A random forest classifier [5] is then applied to classify all features in the set F_{reflex} . The forest is an ensemble of n_t binary decision trees with a randomized selection of descriptor variables on which a tree splits. Thus, the classification of a feature descriptor with the random forest classifier provides n_t individual votes, one per each tree in the forest. The probability that a reflection based feature represents a reflective vest can be inferred by dividing the number of trees voting for a reflective vest by the total number of trees n_t in the forest. In a supervised learning task, the random forest classifier is trained on a set of descriptors that are manually labelled with a tag indicating whether the descriptor corresponds to a vest feature or not.

3.2 3D Position Estimation

The lower part of Fig. 1 depicts the two steps the system performs in order to estimate a 3D position for features that were detected in I_f and classified as belonging to the set of vest features F_{vest} . First, the system estimates the distance of a vest feature with a machine learning approach before exploiting the intrinsic camera model together with the distance estimate to produce a 3D position estimate.



Fig. 3 The figure shows an exemplary result of the vest detection process. Blob-like features are detected in the image I_f taken with IR flash (above). Detected features are represented by the ensemble of crosses in I_f . The detected features are then tracked in the image I_{nf} taken without IR flash (below). The detection area in I_f is restricted to the white bounding box to allow features to be tracked even in the case of quick rotational movements. Successfully tracked features are marked as white crosses in image I_f and the tracked locations are indicated by white crosses in image I_{nf} . All tracked features in the above example show a very low intensity difference and are therefore not considered as vest candidates. Features that failed to be tracked include detections on the reflective vest as well as on the metallic surface of the car standing right in front of the camera. All the untracked features are considered as vest candidates and classified by the random forest model. A black square is drawn around features that are finally classified as vest features.

3.2.1 Distance Estimation

The same local image descriptors used for the feature classification described in Sec. 3.1.4 are employed to estimate the distance of a reflective vest feature based on machine learning. Using supervised learning we train a random forest regressor on a set of descriptors that are labelled with the ground-truth distance between the camera and the reflective vest that caused the appearance of the corresponding vest feature. The trained model is then applied to obtain a distance estimate d for descriptors of unseen vest features.

A strong emphasis has to be made on the rotation and scale invariance of the feature descriptors adopted in the underlying system. The size of a reflective vest pattern in the image I_f decreases with increasing distance between the vest and the camera. As we aim to estimate the distance to the vest based on the local image descriptor, scale invariance is clearly undesired because it would make it impossible for the regressor to consider the size of the patch. On the other hand, rotation invariance would be beneficial for cases where the regressor has to estimate the distance of an untrained vest feature which is just a rotated version of a trained feature.

In the case of BRIEF the descriptor is neither scale nor rotation invariant but tolerates small amounts of rotation [6]. In contrast, the SURF descriptor is designed to be rotation and scale invariant but this property only holds true if the SURF descriptor is used in combination with the corresponding SURF feature detector which provides a scale and an orientation for every detected feature. The STAR feature detector used in our application does not provide any orientation for the detected features. Thus, we extract the descriptors within a window of fixed size w_d and constant orientation of zero degrees and obtain BRIEF and SURF descriptors that are neither scale nor rotation invariant.

3.2.2 3D Projection

The final step aims at estimating the 3D position relative to the camera for all features in the set F_{vest} . Therefore, an intrinsic camera model of the camera system is obtained by a calibration method dedicated to omni-directional cameras [10]. The method assumes that the image projection function can be described by a Taylor series expansion for which the coefficients are estimated in the calibration process.

Using the obtained camera model and given the image coordinates $\mathbf{u} = (u, v)$ of the location at which a feature in F_{vest} was detected, a ray in 3D space can be inferred on which the object that caused the appearance of the feature in the image must lie on. By further taking into account the distance d that was estimated for the corresponding feature, a 3D point on the ray can be located leading to the final position estimate $\mathbf{x} = (x, y, z)$ in the coordinate system fixed to the camera.

4 Experiments

The reflective vest detection and position estimation system has been evaluated in three different test scenarios as listed in Table 1. A sensor unit consisting of the camera system and a 2D laser range scanner (SICK LMS-200) both fixed to a solid mechanical frame was used for the data acquisition. An extrinsic calibration was carried out to obtain the position and orientation of the laser range scanner relative to the camera [11]. The sensor unit is mounted at a height of approximately 1.5 m on a mobile platform with four hard rubber wheels.

Scenario	Environment	Image Pairs I
1	Indoors, warehouse-like environment	400
2	Outdoors, car parking area, sunny weather conditions	380
3	Outdoors, car parking area, direct sunshine into camera	100

Table 1 Test scenarios with number of acquired image pairs featured in the system evaluation

Several training and validation data sets were acquired for each scenario by simultaneously recording the raw camera images and the 2D laser readings. During the acquisition of all data sets the mobile platform was moving at a speed of approximately 0.5 m/s. One data set per scenario was held back for evaluation purposes while the remaining sets served as training data. Table 2 summarizes the values of the different system parameters used in the evaluation setup.

All the acquired data sets were preprocessed to detect the set of raw features F_{raw} and to extract the corresponding local image descriptors. A BRIEF descriptor of 256 binary variables and a standard SURF descriptor of 64 floating point variables was extracted for every feature. A ground-truth label was manually assigned to each descriptor indicating whether it corresponds to a vest feature (label 1) or not (label 0). Furthermore, the ground-truth distance between the camera and the person wearing the vest was extracted from the laser readings and assigned to the descriptors.

A supervised learning process is applied to obtain the models of the feature classifier and the distance regressor. We trained a random forest on 45k extracted image descriptors and the corresponding labels to obtain the classifier described in Sect. 3.1.4. Likewise, we trained a random forest on 30k image descriptors labelled as vest features (label 1) and the corresponding ground-truth distance to the person to obtain the regressor described in Sect. 3.2.1.

The evaluation was then performed by processing the validation data set of each scenario and comparing the obtained results with the ground-truth labels assigned during preprocessing. The main processing steps of the system according to Fig. 1 are evaluated individually.

Parameter	Description	Value
f_a	Image pair acquisition rate	~ 14 Hz
t_a	Time delay between the acquisition of I_f and I_{nf}	~ 35 ms
w/h	Width and height of the unwrapped input images I_f and I_{nf}	600x240 Pixel
b	Feature detection window border size	40 Pixel
w_i	Window size for the intensity difference check	5 Pixel
t_i	Threshold for the intensity difference check	30.0
w_d	Window size for the descriptor extraction	24 Pixel
n_t	Number of trees in the random forest classifier/regressor	20

Table 2 Values of the various system parameters used for the evaluation setup

Feature Detection

To evaluate its performance, the feature detector (Sect. 3.1.2) is applied on each image I_f in the validation data sets resulting in a set of raw features F_{raw} . If a reflective vest is identified with at least one feature in F_{raw} the detection process for image I_f is declared successful. Images in which no reflective vest appears are not considered in the evaluation. The vest detection rate is defined as the ratio between images in which the vest is successfully detected and the number of images showing a vest.

Table 3 shows the result of the feature detection evaluation. The data shows that in nearly all images of scenarios 1 and 2 the detector reliably detects at least one raw feature per reflective vest. In scenario 3 the camera faces the sun resulting in lens artifacts frequently appearing in the images. The detector occasionally fails to detect features intersecting with the lens artifacts which leads to a detection rate decreased by approximately 10 %.

Feature Classification

In a second step we evaluate the system’s ability to correctly classify a set of detected features F_{raw} into a set of vest features F_{vest} and a set of non-vest features. The evaluation assesses the performance of several processing steps as a group, namely the feature tracking and intensity check (Sect. 3.1.3) as well as the feature description and classification (Sect. 3.1.4). Every set of raw features F_{raw} detected in the series of images I_f is processed to obtain a corresponding set of predicted vest features F_{vest} . The set of predicted non-vest features is defined as $F_{nonvest} = F_{raw} \setminus F_{vest}$. The result of the binary classification into vest and non-vest features is then compared to the ground-truth label manually assigned during preprocessing.

Table 4 shows the results of the evaluation in form of confusion matrices. Scenario 1 contains images acquired indoors where the only IR source was the flash of the camera system and where no other reflective object than the vest appeared. Thus, true negative and false positive rates are not defined. The results illustrate the effect of the feature description and classification described in Sect. 3.1.4. The false positive rate is decreased by rejecting features of other reflective material than reflective vests. By doing so, the classifier also erroneously discards some actual vest features that look unfamiliar, resulting in an increased false negative rate. Classification based on the SURF descriptor yielded the best trade-off between the two effects.

Scenario	Total Detected Features	Average Features per Image I_f	Vest Detection Rate
1	1612	4.03	97.50%
2	1540	4.05	97.84%
3	4953	49.53	88.37%

Table 3 Result of the feature detection process

			Scenario 1								
			Predicted			Predicted					
			0	1	0	1	0	1			
Actual	0	n/a	n/a	n/a	0	n/a	n/a	0	n/a	n/a	
	1	0.06%	99.94%	1	16.00%	84.00%	1	4.28%	95.72%		
			Scenario 2								
			Predicted			Predicted					
			0	1	0	1	0	1			
Actual	0	79.07%	20.93%	0	97.94%	2.06%	0	95.09%	4.91%		
	1	1.30%	98.70%	1	15.32%	84.68%	1	10.94%	89.06%		
			Scenario 3								
			Predicted			Predicted					
			0	1	0	1	0	1			
Actual	0	97.45%	2.46%	0	99.29%	0.71%	0	99.51%	0.49%		
	1	15.92%	84.08%	1	34.39%	65.61%	1	30.86%	69.14%		
			a)				b)				c)

Table 4 The table shows the confusion matrices of the binary feature classification into vest- and non-vest features for the case where a) the classification is based only on the feature tracking and intensity check ($F_{vest} = F_{reflex}$), b) the feature set F_{vest} is obtained by further classification based on the BRIEF descriptor and c) by further classification based on the SURF descriptor.

Distance and Position Estimation

The trained model of the random forest regressor (Sect. 3.2.1) is used to obtain a distance estimate for every predicted vest feature in F_{vest} . The distance estimate is used together with the feature coordinates $\mathbf{u} = (u, v)$ and the intrinsic camera model to obtain a corresponding 3D position estimate according to Sect. 3.2.

Fig. 4 shows the results of the distance estimation. While the estimations based on the SURF descriptor show a widely stable accuracy over the whole distance range, the BRIEF descriptor only allows a reliable estimation for short range distances up to 7 m. The plots also report sporadic but large outliers indicating a distance estimation error of several meters. Even in the most extreme conditions with direct sunshine into the camera (Scenario 3) the system still gives accurate estimates up to 6 m distance. Under the influence of strong sunlight, the system fails to detect features for higher ranges and no distance and position estimations are available.

In the same way as for the distance estimation, we also evaluated the final position estimation error. The results are not shown here for lack of space. However, they indicate the same tendency as the results shown in Fig. 4.

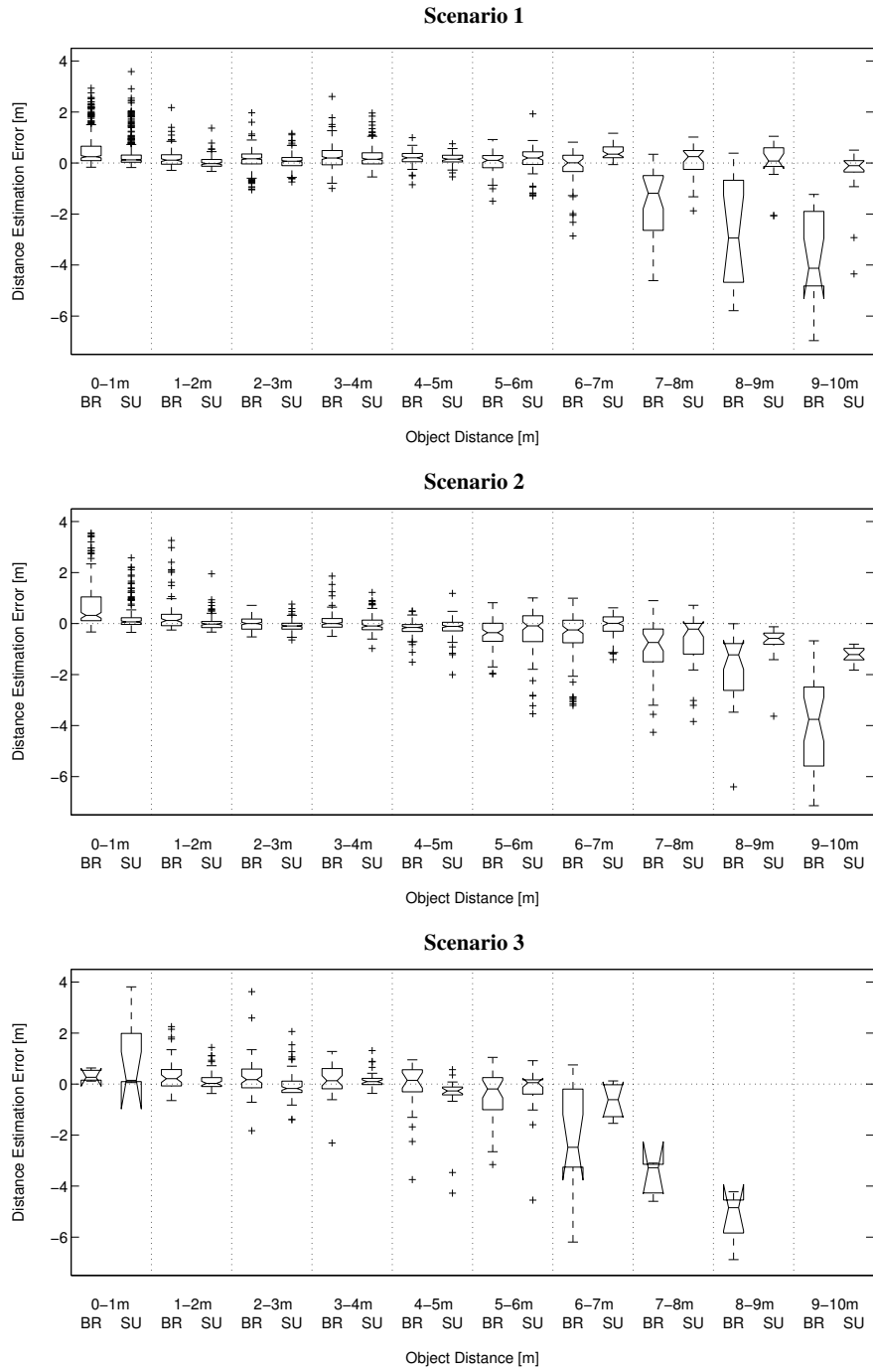


Fig. 4 The box plots show the distance estimation error for the scenarios 1–3. The indications BR (BRIEF) and SU (SURF) specify the image descriptor on which the estimation is based. Missing plots indicate that the vest detection failed and no distance estimation could be performed.

5 Conclusions and Future Work

In this paper we presented a system capable of detecting people wearing reflective vests and estimating their position in 3D space. The system has been evaluated in an indoor warehouse-like environment and outdoors in sunny weather conditions. The experiments show that the system gives accurate distance estimates for distances up to 10 m, with only sporadic outliers. Even under the extreme conditions of direct sunshine into the camera, the system still performs well for distances up to 6 m.

Future work includes the tracking of reflective vests over time using a particle filter which is continuously updated with the 3D position estimates of single vest features. Thus, vest detections will be maintained over several frames and the influence of outliers will be reduced. To allow for simultaneous detection and tracking of multiple persons, a clustering process will also be introduced.

Future work also includes a systematic evaluation of the system in a range of different weather conditions including rain, snowfall, and fog. Additional scenarios will be tested that were not addressed in this paper, such as persons that are partly occluded or lying on the floor (e.g. fainted persons) as well as different types of camera movements. An extended version of the camera system will include more powerful IR LEDs to extend the detection range to 20 m and beyond.

References

1. Agrawal, M., Konolige, K., Blas, M.R.: Censure: Center surround extremas for realtime feature detection and matching. In: D.A. Forsyth, P.H.S. Torr, A. Zisserman (eds.) ECCV (4), *Lecture Notes in Computer Science*, vol. 5305, pp. 102–115. Springer (2008)
2. Andreasson, H., Bouguerra, A., Stoyanov, T., Magnusson, M., Lilienthal, A.: Vision-based people detection utilizing reflective vests for autonomous transportation applications. IROS Workshop on Metrics and Methodologies for Autonomous Robot Teams in Logistics (MMART-LOG) (2011)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)* **110**, 346–359 (2008)
4. Bouguet, J.Y.: Pyramidal implementation of the lucas kanade feature tracker description of the algorithm (2000)
5. Breiman, L.: Random forests. *Machine Learning* **45**(1), 5–32 (2001)
6. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: Brief: Binary robust independent elementary features. In: K. Daniilidis, P. Maragos, N. Paragios (eds.) ECCV (4), *Lecture Notes in Computer Science*, vol. 6314, pp. 778–792. Springer (2010)
7. Chung, J., Kim, N., Kim, G.J., Park, C.M.: Postrack: A low cost real-time motion tracking system for vr application. *International Conference on Virtual Systems and MultiMedia* (2001)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *In CVPR*, pp. 886–893 (2005)
9. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient matching of pictorial structures. In: *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, pp. 66–73 (2000)
10. Scaramuzza, D., Martinelli, A., Siegwart, R.: A flexible technique for accurate omnidirectional camera calibration and structure from motion. In: *Proc. of The IEEE International Conference on Computer Vision Systems (ICVS)* (2006)
11. Zhang, Q.: Extrinsic calibration of a camera and laser range finder. In: *In IEEE International Conference on Intelligent Robots and Systems (IROS)*, pp. 2301–2306 (2004)