2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Multi-human Tracking using High-visibility Clothing for Industrial Safety

Rafael Mosberger, Henrik Andreasson and Achim J. Lilienthal

***Abstract*— We propose and evaluate a system for detecting and tracking multiple humans wearing high-visibility clothing from vehicles operating in industrial work environments. We use a customized stereo camera setup equipped with IR flash and IR filter to detect the reflective material on the worker's garments and estimate their trajectories in 3D space. An evaluation in two distinct industrial environments with different degrees of complexity demonstrates the approach to be robust and accurate for tracking workers in arbitrary body poses, under occlusion, and under a wide range of different illumination settings.**

I. Introduction

Reliably detecting and tracking workers from both human-driven and autonomous vehicles operating in industrial work sites is a crucial prerequisite for any on-board safety system aiming at preventing vehicle-pedestrian collisions. Systems that attempt to offer a solution to this problem are confronted with challenging requirements. They need to offer robust performance under different weather and illumination conditions and in potentially cluttered indoor and outdoor sites. Methods that address these challenges can be found in the well-studied area of pedestrian detection for road traffic, where a great deal of research has focused on detecting people in upright positions. However, the diversity of industrial environments and the variety of potential body positions in which workers appear (see Fig. 1 for some examples) prevent the direct application of existing pedestrian detectors in an industrial context.

In our previous work [1], we reported on a novel human tracking approach that targets industrial environments in which workers wear high-visibility clothing. Safety vests with highly reflective properties have become widely accepted throughout industry as an effective way to protect workers from accidents. The core idea that allows for reliable human tracking is to identify workers by detecting the retro-reflective material attached to their safety clothing. We use customized a camera system with infrared (IR) filter and active IR illumination to capture flash/non-flash image pairs in which reflective material appears significantly brighter in the image acquired with flash (refer to Fig. 3 for an example).

We tested a monocular camera setup with IR flash for the purpose of tracking a single upright person at distances ranging up to 10 meters and under different weather and illumination conditions [1]. A binary classifier, trained on local image feature descriptors proved satisfactory for discriminating the reflective stripes on a worker's safety vest from a limited set of other reflective objects. We further demonstrated that the distance to a human can be estimated with an accuracy of less than a meter using the same image features as for classification [2]. The results indicated reliable detection not only in favorable but also in challenging situations, such as under direct exposure to the sun. The estimation of the distance to a tracked target through supervised learning of visual features allowed to deploy the system as a compact single-camera setup. Yet, it requires costly acquisition and labeling of a considerable amount of training data.

Fig. 1. A selection of challenges faced by human detectors in industrial work environments, including partly occluded workers in non-upright body poses as well as varying illumination settings. Detections as obtained from our human tracker are indicated with yellow squares.

The camera system presented in this paper extends on our previous work [1]. We describe the necessary additions and modifications in order to perform simultaneous tracking of multiple humans. The new configuration uses a stereo setup to estimate the distance to a tracked person, thus rendering the training of the previously employed regressor superfluous.

This article makes the following main contributions: 1) We describe the extension of our single-target human tracking approach towards multi-target tracking. 2) We discuss the change from a point based to a contour based representation of interest regions and show how the change helps to address the data association problem in the tracking stage. 3) We present a modified hardware setup consisting of a near-infrared (NIR) stereo camera that extends the detection range from 10 to 20 meters. 4) We perform an extended evaluation in two distinct industrial environments and put particular emphasis on demonstrating our system's ability to track people in arbitrary body poses.

Rafael Mosberger, Henrik Andreasson and Achim J. Lilienthal are with the AASS Research Center, School of Science and Technology, Örebro University, S-70182 Örebro, Sweden firstname.lastname@oru.se

978-1-4673-6357-0/13/$31.00 ©2013 IEEE

II. Related Work

Robust human detection from vehicles and machines in industrial scenarios has so far attracted less attention than the field of pedestrian detection for road traffic. In the latter, the urgent demand to increase automotive safety has led to the development of on-board pedestrian protection systems (PPSs) that anticipate potential collisions, provide the driver with audible or visual warning signals, and if necessary even take automatic braking actions.

The basic task of a corresponding safety system for vehicles and machinery in an industrial scenario is similar. Yet, the diversity of environments, the potentially cluttered work space, as well as the variety of body positions of workers constitute a set of harsh challenges. Consequently, direct application of PPSs from road traffic to industrial applications does not lead to the performance demanded by the industry.

Several authors have proposed solutions that are tailored to the needs of industrial scenarios. Heimonen et al. [3] describe a stereo camera based human detection system for heavy industrial machinery that provides a framework for combining multiple human detection methods in order to increase the overall robustness. Dickens et al. [4] propose to fuse the information of thermal images (for detection) and a 3D range sensor (for depth estimation) for a vehicle-personnel collision system for the mining industry. Teizer et al. [5] propose a safety system for construction equipment based on radio-frequency identification (RFID). Compared to other approaches, it not only informs the machine operators about the presence of humans but also warns a detected human of the nearby vehicle. Therefore, each worker is equipped with a personal protection unit that provides auditory, visual and vibrating alarms in case of danger.

The idea of using the properties of high-visibility clothing in order to facilitate the detection of industrial workers has been applied by Park et al. [6]. Their vision-based approach specifically identifies the fluorescent color of safety vests by processing local color histograms extracted from the regions of interest. Even though the reflectivity of the vests is not exploited, the authors show that the distinctive color considerably contributes to effective detection.

III. System Description

The system presented in this paper is an extension of our earlier work on human tracking with a single flash camera [1]. Here, we adopt the same overall framework with a detection, classification and tracking stage. An overview of the different processing steps is presented in Fig. 2. The main objective is to extend the existing approach towards simultaneous tracking of multiple persons. Therefore, the system needs to undergo several modifications which primarily concern the tracking step. However, in order to address the problem of data association that every multi-object tracking system has to cope with, we also decided to modify parts of the segmentation process and opted for a shape based representation instead of the original feature point representation in [1] to describe the regions of interest.

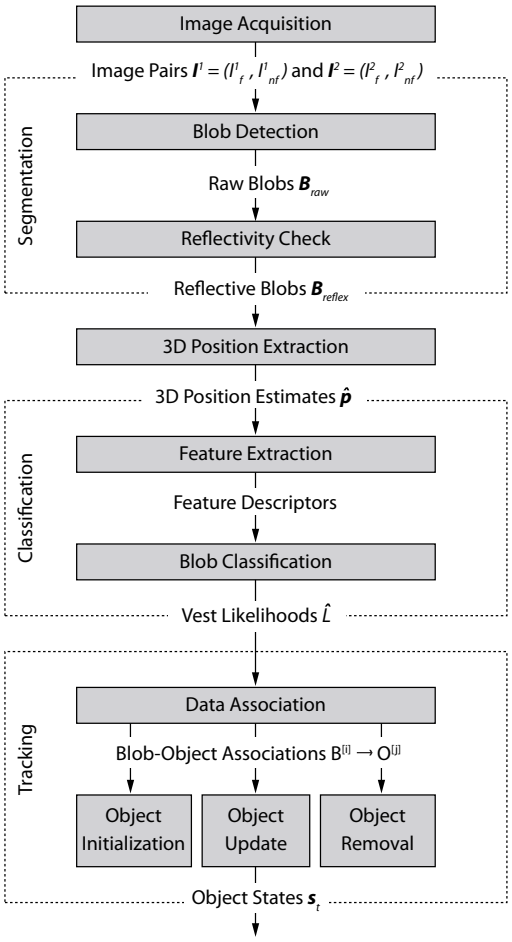


Fig. 2. Overview of the detection and tracking system

This change allows us to compute overlap measures between detected blobs and tracked objects and thus facilitate data association. We also replace the supervised regression based estimation of the distance to a detected person by stereo triangulation in order to reduce the costly acquisition and labeling of training data needed to train the regressor model.

As it was done in the previous system, we perform particle filtering to estimate a tracked object's position and velocity in a 3D space relative to the camera. The main difference of the multi-object tracker is that a separate particle filter is maintained for every of the simultaneously tracked objects and a data association step is introduced before the individual particle filters are updated.

A. Hardware and Image Acquisition

The camera setup deployed for detecting reflective material (cf. Fig. 4a) combines a 1 megapixel monochrome CMOS sensor with high NIR sensivity, a wide-angle lens, an NIR bandpass filter, and a flash unit consisting of 16 highpower NIR LEDs. The center wavelength of both filter and flash is 940 nm and the filter has a bandwidth of 10 nm. Using two identical cameras of the described type, we build a stereo camera unit with a base line of 200 mm (cf. Fig. 4b). The stereo rig further features a color camera that is used purely for visualization purposes.

Fig. 3. Outdoor scene containing a total of 3 persons (left) with corresponding input images as captured by one of the NIR cameras using IR flash (center) and without using flash (right). Reflective regions as identified in the segmentation process are indicated with a red contour.

A synchronized image stream is acquired from the stereo camera while alternately using the IR flash for every second image capture. The captured flash/non-flash input image pairs of both cameras, denoted $\boldsymbol{I}_f^1 = (I_f^1, I_{nf}^1)$ and $\boldsymbol{I}_f^2 = (I_f^2, I_{nf}^2)$ for the first and second camera respectively, are the only input to the tracking system. An example of a captured flash/non-flash image pair is depicted in Fig. 3.

B. Segmentation

The goal of this stage is to extract regions of interest from the input images that correspond to reflective objects. This is achieved by identifying areas that are significantly brighter in the image I_f captured with IR flash than in image I_{nf} captured without active illumination.

1) Blob Detection: A first step in the extraction of reflective objects is to identify bright regions in the image I_f taken with flash. To do so, we apply local adaptive thresholding to I_f with a threshold computed as the average intensity in a square local neighborhood subtracted by an offset. Using contour following applied to the thresholded image, we then extract a set of raw blobs $\mathcal{B}_{\text{raw}}$ in which each blob is characterized by its respective contour Λ and the centroid $\boldsymbol{c} = [c_x, c_y]$ of its position:

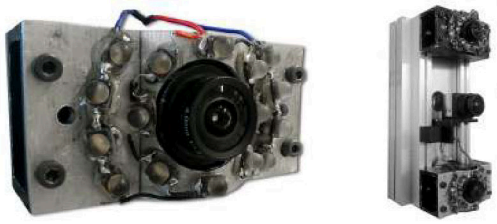
$$\mathcal{B}_{\text{raw}} = \left\{ B^{[i]} = \left\langle \Lambda^{[i]}, \boldsymbol{c}^{[i]} \right\rangle \mid i = 1, ..., N_{\text{raw}} \right\} \qquad (1)$$

2) Reflectivity Check: In order to specifically identify reflective objects, we submit all blobs extracted from I_f to a verification process. We verify whether the intensity values in the surrounding area of a blob are similar in the images I_f and I_{nf}, or if they are distinctly higher in image I_f. The first case is an indication that the blob corresponds to an object that appears bright due to background illumination, for example by the sun. Blobs that match this criterion are rejected. In contrast, the second case indicates that the corresponding object has highly reflective properties and therefore the blob is retained. We refer to the set of blobs that pass this reflectivity check as the set $\mathcal{B}_{\text{reflex}}$. For a detailed explanation of the verification process, refer to [1].

C. 3D Position Extraction

For all blobs that passed the reflectivity check we estimate the position of the corresponding reflective object in a 3D space relative to the camera. To do so, we make use of the stereo input and apply dense stereo matching to regions around each blob. Subsequently, we compute a single median disparity value per blob. As illustrated in Fig. 5, disparity extraction is unreliable in the white regions inside a blob due to the lack of texture, but produces consistent results in the regions near their border.

We therefore proceed in two steps: Using the stereo image pair (I_f^1, I_f^2), disparity is computed inside a rectangular local region around every reflective blob in $\mathcal{B}_{\text{reflex}}$. We then compute a mean disparity for each blob by taking into account all pixels with a distance smaller than s from the contour. Using the computed mean disparity value, the blob centroid $\boldsymbol{c}$ and the camera's inverse projection function, we finally obtain a 3D position estimate $\hat{\boldsymbol{p}}$ for every blob.



a) b)

Fig. 4. Hardware configuration used in our experiments: **a)** Near-infrared (NIR) camera equipped with a wide-angle lens, a NIR bandbass filter (not visible in the image), and a flash unit consisting of 16 high-power NIR LEDs. **b)** Stereo camera unit built using two identical NIR cameras (right). The additional color camera in the center of the stereo rig is not used by the tracking algorithm and purely serves for visualization purposes.

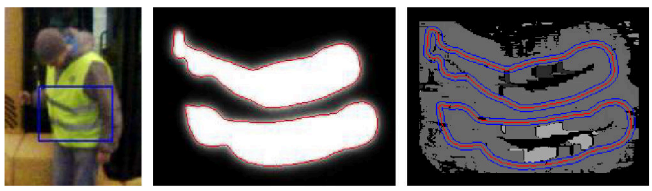


Fig. 5. The figure illustrates how disparity is computed for a detected blob: The scene (left) is captured with the NIR camera and the resulting image I_f (middle) is segmented to extract the contours of the reflectors, drawn in red. A disparity map (right) is then computed in the neighborhood of each contour by using the images I_f of both NIR cameras. Due to the lack of texture, disparities are sometimes not (black pixels) or erroneously (differing gray levels) computed in regions within the blob. However, consistent results are obtained from the regions near the contour and therefore only the zone delimited by blue lines is taken into consideration.

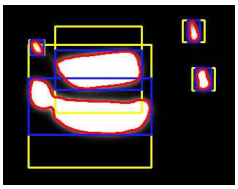


Fig. 6. U-SURF feature descriptors are used to describe the visual content in the neighborhood of the detected blobs (as outlined in red). The descriptors are extracted from a square region (yellow) with a size corresponding to the length of the blob's bounding box (blue).

D. Classification

Blobs that passed the reflectivity test originate either from a reflector on a high-visibility garment or from another reflective object in the scene. In order to avoid false alarms under the presence of such objects, we attempt to classify all blobs $B \in \mathcal{B}_{\text{reflex}}$ into vest- and non-vest objects, where the term *vest* refers to any kind of high-visibility clothing.

1) Feature Extraction: We observed in [2] that state-of-the art image feature descriptors such as SURF or BRIEF provide a powerful tool to describe the local neighborhood of a detected reflector. The new contour-based representation provides now additional flexibility in defining the exact locations from which the feature descriptors are extracted. As illustrated in Fig. 6, we extract SURF descriptors from a square area with the same center as the blob's bounding box and with a size corresponding to the bounding box's length.

It is worth noting that the SURF descriptors are extracted from non-SURF features. The characteristic feature orientation which is provided if features are detected with the SURF detector is not available here. In consequence, the extracted descriptors correspond to the upright, non-rotation invariant version of SURF, referred to as U-SURF.

2) Blob Classification: Classification of the feature descriptors is achieved using a Random Forest classifier. Each sample is individually classified by all the trees building the forest and we compute the estimated likelihood $\hat{L}$ that a reflector represents a vest reflector as the number of trees with a positive vote divided by the number of trees.

E. Tracking

As the ultimate goal of the application is to track individual persons, the tracking unit aims at associating single blob detections with a set of tracked objects and maintain a filtered estimate of the object's state. Thereby, we choose to first track all reflective objects in the scene regardless of their nature and then infer the likelihood of an object representing a person by aggregating the information provided with the individual blobs.

Therefore, let us assume that at a time instant t we are given a set $\mathcal{B}_{\text{reflex},t}$ of N_t blobs corresponding to all detected reflective items in the scene. For each of the blobs, characterized by the contour Λ and centroid $\boldsymbol{c}$, a 3D position estimate $\hat{\boldsymbol{p}}$ has been computed as well as an estimate $\hat{L}$ of the likelihood that the blob represents a reflector belonging to a high-visibility garment:

$$\mathcal{B}_{\text{reflex},t} = \left\{ B_t^{[i]} = \left\langle \Lambda_t^{[i]}, \boldsymbol{c}_t^{[i]}, \hat{L}_t^{[i]}, \hat{\boldsymbol{p}}_t^{[i]} \right\rangle \mid i = 1, ..., N_t \right\} \tag{2}$$

In a first step we attempt to assign the blobs in $\mathcal{B}_{\text{reflex},t}$ to a set $\mathcal{O}_t$ of M objects being tracked at time instant t, using both 2D overlapping and 3D distance criteria. Then, based on the assignments, the object states are updated and new objects are initialized for blobs that failed to be assigned. Objects in $\mathcal{O}_t$ are characterized by their 3D position and velocity state $\boldsymbol{s}_t$, a tracking confidence measure c_t and an estimated likelihood L_t that the tracked object corresponds to a person:

$$\mathcal{O}_t = \left\{ O_t^{[j]} = \left\langle \boldsymbol{s}_t^{[j]}, c_t^{[j]}, L_t^{[j]} \right\rangle \mid j = 1, ..., M_t \right\} \tag{3}$$

with the state vector $\boldsymbol{s}_t$,

$$\boldsymbol{s}_t = [\boldsymbol{p}_t \; \dot{\boldsymbol{p}}_t]^\top = [x_t, y_t, z_t, \dot{x}_t, \dot{y}_t, \dot{z}_t]^\top \tag{4}$$

being recursively estimated by a particle filter. The confidence measure c_t of an object is used as a indicator of consistent detection and is incremented by 1 in every frame where one or several blobs are assigned to the object and decremented by 1 in the opposite case. Only objects that reach a confidence of $c_t = 3$ are considered. The estimated likelihood L_t that the observed object represents a person is calculated as the average of the corresponding estimates $\hat{L}$ of all individual blobs assigned to the object up to time t. Every object is further represented by a 2D bounding box that is computed from state $\boldsymbol{s}_t$ and that approximately delimits the image region where the tracked object is believed to be.

1) Data Association: Standardized high-visibility clothing (cf. Fig. 9) always comes with multiple reflectors attached to different areas of a garment and it is therefore most likely that multiple reflective blobs are detected for the same tracked person. On the other hand, a reflector cannot represent multiple objects at the same time. In consequence, the first tracking step consists of a many-to-one mapping that assigns blobs to currently tracked objects. For every potential assignment of a blob $B^{[i]}$ to an object $O^{[j]}$ we define a cost function $d(B^{[i]}, O^{[j]})$ that takes a low value if it is likely that the i-th blob represents a reflector of the j-th object and a high value in the opposite case.

The cost function is computed based on two criteria. First, the overlap of the blob area with the bounding box of an object in the 2D image plane should be high. This is expressed by the overlap cost function $d_\cap$, defined as

$$d_\cap(B^{[i]}, O^{[j]}) = 1 - \frac{A(B^{[i]}) \cap A(O^{[j]})}{A(B^{[i]})} \tag{5}$$

where $A(B^{[i]})$ denotes the area delimited by the contour $\Lambda^{[i]}$ of the i-th blob, and $A(O^{[j]})$ the area covered by the j-th object's 2D bounding box. At the same time, the absolute difference of the blob's 3D position estimate $\hat{\boldsymbol{p}}_t$ and the object's current 3D position $\boldsymbol{p}_t$ should be small. This is expressed by the distance cost function d_δ:

$$d_\delta(B^{[i]}, O^{[j]}) = 1 - \exp\left(-\alpha \times ||\hat{\boldsymbol{p}}_t - \boldsymbol{p}_t||\right) \tag{6}$$

Finally, we define the cost function as the weighted sum of $d_\cap$ and d_δ where the weights $w_\cap$ and w_δ allow to give a preference to one of the two criteria:

$$d(B^{[i]}, O^{[j]}) = w_\cap \times d_\cap(B^{[i]}, O^{[j]}) + w_\delta \times d_\delta(B^{[i]}, O^{[j]}) \tag{7}$$

Fig. 7. Wheel loader (left) and forklift (right) on which the camera setup was evaluated. The location of the sensor unit is indicated in red.

A given blob $B^{[i]}$ is then assigned to the object $O^{[j]}$ for which the lowest assignment cost was computed, under the restriction that this cost needs to be lower than a defined assignment threshold λ_d. If the lowest computed assignment cost is above λ_d, a blob is considered to belong to an object which is not yet being tracked and no assignment is made.

2) Object Update: Based on the blob-to-object assignments, the states of all tracked objects in $\mathcal{O}_t$ are updated. This is achieved by providing each object's particle filter with the respective blob detections and applying the motion and measurement model as described in detail in [1]. After updating the state $\boldsymbol{s}_t$, the 2D bounding box of an object is updated by centering it around the projection of the new filtered 3D position estimate $\boldsymbol{p}_t$ on the image plane.

3) Object Initialization and Removal: Blobs in the set $\mathcal{B}_{\text{reflex},t}$ that could not be associated to any tracked object are considered as candidates to initialize new objects. To do so, we extract connected regions of similar disparity from the previously computed disparity image and initialize a new object for every group of blobs that falls in the same cluster.

Finally, existing objects are removed either if their 2D bounding box falls out of the image border or if their confidence measure c_t reaches a value equal to 0.

IV. Experiments and Results

We evaluated our tracking system in two distinct environments and by mounting the camera setup on two different industrial vehicles as shown in Fig. 7. A synchronized data stream from the NIR stereo camera was recorded at 50 fps, leading to flash/non-flash image pairs available at 25 Hz. Images from a color camera were recorded at 10 fps for the purpose of visualization and to facilitate the manual annotation of the datasets. A binary Random Forest classifier was trained on 200k samples from reflective vest reflectors and 100k other reflective objects recorded in various environments. In addition to the camera system, we further equip the sensor unit with a Velodyne HDL-64E 3D LIDAR in order to extract ground truth positions of the tracked persons.

Fig. 9. The two types of reflective safety clothing used throughout the experiments: ANSI/ISEA safety class 2 vest with reflectors only around the body (right) and safety class 3 jacket with additional reflectors in the shoulder and arm area (left).

Two different types of reflective safety clothing were used throughout the experiments as depicted in Fig. 9. The two garments represent respectively safety class 2 and 3 of the ANSI/ISEA 107-2004 standard and are distinguished by the amount and spatial distribution of the reflective material.

The first test environment is an outdoor gravel loading pit where we mounted the camera system on the roof of a wheel loader (cf. Fig. 7, left). Typical loading and unloading scenarios were simulated, including sharp turns and alternate forward and reverse driving up to 30 km/h. Apart from the host vehicle, a second wheel loader and a hauler were present in the area and a total of four persons were either walking in the vicinity of the machines or operating them. Two test sequences from this environment are evaluated. Seq. #1 contains a total of 1800 frames corresponding to 3 minutes and was recorded in cloudy weather conditions. Seq. #2 contains 1200 frames corresponding to 2 minutes and was recorded in the evening when the sun was low and shining directly into the camera, leading to much higher background illumination in the images captured by the NIR cameras. Apart from the safety clothing, the only other reflective objects present in the environment are multiple cat's eye reflectors on the vehicles.

The second test environment is an indoor manufacturing and maintenance site for industrial vehicles. Numerous workers are present in the cluttered area, carrying out tasks in various body postures while often being partly occluded by objects. The sensor unit was mounted on the roof of a forklift (cf. Fig. 7, right) that was operated at speeds up to 20 km/h. A test sequence (Seq. #3) of 1200 frames (2 minutes) is evaluated in which a considerable amount of reflective objects other than the safety clothing appear.

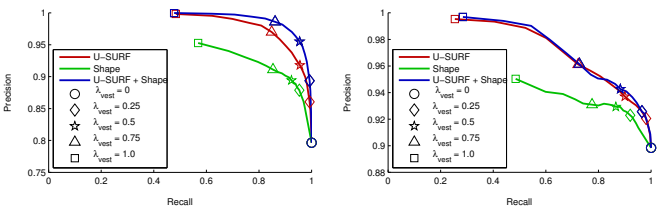


Fig. 8. Precision-recall curves obtained from classifying blobs into vest and non-vest reflectors based on the different descriptors. The graphs respectively represent Seq. #1 (left) and #3 (right).

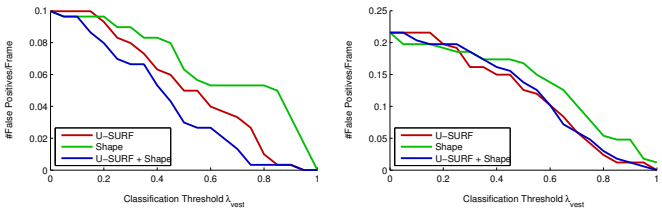


Fig. 10. The plots show the evolution of the number of false positives per frame for Seq. #1 (left) and #3 (right) depending on the classification threshold λ_{vest}.

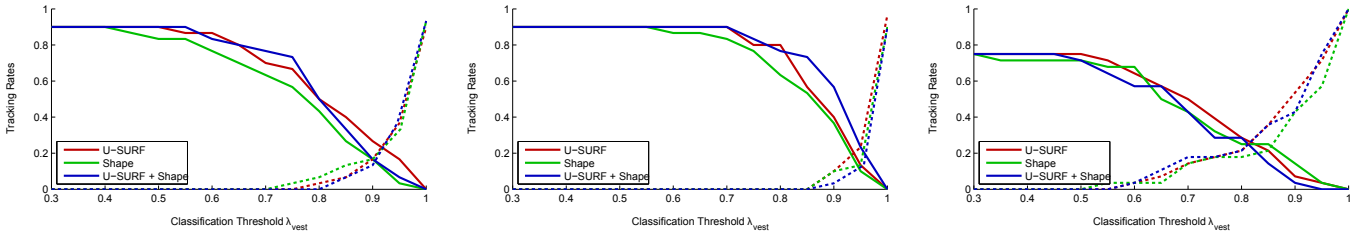


Fig. 11. Ratio of mostly hit (solid) vs. mostly missed (dashed) trajectories for Seq. #1–#3 depending on the classification threshold λ_{vest}.

A. Classification

Fig. 8 shows the performance of the Random Forest blob classification into vest and non-vest reflectors. To illustrate the superiority of SURF over shape-based descriptors for the type of objects to be classified in our application, we also indicate results for a descriptor combining the seven Hu Moments [7] with several other variables computed from the blob contour, namely the circularity, the area-to-perimeter ratio as well as the aspect ratio of the bounding box of the blob contour.

B. Tracking

Tracking performance is evaluated in the 2D image space using similar criteria as described in [8]. The trajectories of all humans are manually annotated in every 5th frame. A new trajectory is counted if a person is occluded during more than 10 consecutive frames. We define a trajectory as *mostly hit* if successful detections cover more than 80% of the frames in which the corresponding person is visible. Similarly, *mostly missed* trajectories are defined as being covered in less than 20%. Finally, we consider the number of *false alarms*, referring to tracked reflective objects that are mistakenly classified as humans. Only one false alarm is counted if the same object is repeatedly misclassified.

Tracking performance according to the above measures is shown in Figs. 11–10 while numerical results for $\lambda_{\text{vest}} = 0.7$ and a 64-dimensional U-SURF descriptor are summarized in Tab. I. Example tracking results are depicted in Fig. 12 whereas Fig. 13 illustrates different types of erroneous tracker outputs. The results indicate the capability of the algorithm to detect workers not only in upright but in arbitrary body poses and under partial occlusion. Failure modes include missed detections due to the occlusion of visible reflectors, missed detections of humans that are outside the detection range ($\approx 20m$) and false alarms through misclassification of detected blobs. Occasional groupings of two or more persons into a single tracked object were observed in cases where persons stand very close to each other. Though an undesirable effect, it is not crucial from a safety point-of-view as long as detection is maintained.

Seq.	Trajectories	Mostly Hit	Mostly Missed	False Alarms	Trajectory Coverage
#1	30	23	0	4	89.9%
#2	30	27	0	0	88.4%
#3	28	12	5	10	65.5%

TABLE I
QUANTITATIVE TRACKING RESULTS FOR SEQ. 1–3

Most trajectories that fail to be consistently tracked belong either to persons appearing outside the detection range of the system, or to persons under a high degree of occlusion. Especially in Seq. #3, occlusion was the predominant reason for poorly covered trajectories.

C. Position Estimation

The accuracy of the estimated trajectories was assessed using Seq. #1. In Fig. 14, a selection of trajectories with ground-truth and estimated positions as projected to the ground plane is shown. A mean absolute positioning error of 0.34 m was reported over all trajectories.

V. CONCLUSION

In this article, we presented a multi-person tracking systems developed for industrial work environments in which humans wear high-visibility clothing with reflective markers. The system has been evaluated in industrial indoor and outdoor environments and the results indicate robust and accurate tracking performance whenever the reflectors of the safety garments are clearly visible to the camera. A considerable decrease in tracking coverage is observed in

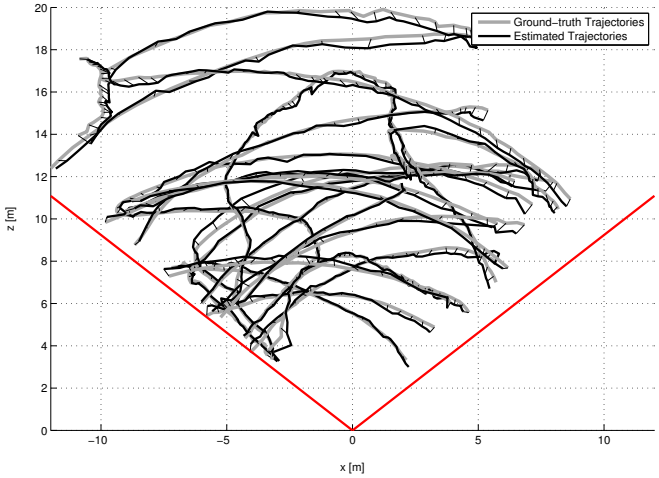


Fig. 14. The figure shows a selection of ground-truth and estimated trajectory pairs as projected on the ground plane. The camera's field-of-view is indicated with red lines.

SEQUENCE #1

SEQUENCE #2

SEQUENCE #3

Fig. 12. Example tracking results for Seqs. #1–3. Detections are indicated with a yellow square.

a) b) c) d)

Fig. 13. Typical erroneous outputs of the tracker: **a)** missed detection of a person (marked with a red circle) in a body position that hides all reflectors of the safety clothing, **b)** missed detection of a person that is tracked but misclassified as a non-human, **c)** false alarms, and **d)** occasional grouping of persons that stand close to each other.

highly cluttered environments where the reflectors are often hidden to the camera.

Among the advantages over conventional vision-based human detectors we emphasize the robustness to different illumination settings and the ability to detect humans regardless of the body pose, provided that a certain amount of reflective material is visible. A drawback of the approach lies in the fact that any kind of reflective material is detected. Even though the blob classifier attempts to extract only the relevant part of the reflectors, occasional false alarms are hard to prevent, especially if reflectors look similar in shape and size to the ones attached to the safety clothing.

REFERENCES

[1] R. Mosberger and H. Andreasson, "An inexpensive monocular vision system for tracking humans in industrial environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[2] ——, "Estimating the 3d position of humans wearing a reflective vest using a single camera system," in *International Conference on Field and Service Robotics (FSR)*, 2012.

[3] T. Heimonen and J. Heikkilä, "A human detection framework for heavy machinery." in *Proc. of the International Conference on Pattern Recognition (ICPR 2010), Istanbul, Turkey*, 2010, pp. 416–419.

[4] J. S. Dickens, M. A. van Wyk, and G. J. J., "Pedestrian detection for underground mine vehicles using thermal images," in *IEEE Africon 2011 Conference*, 2011.

[5] J. Teizer, B. S. Allread, C. E. Fullerton, and J. Hinze, "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system," *Automation in Construction*, vol. 19, no. 5, pp. 630 – 640, 2010.

[6] M.-W. Park and I. Brilakis, "Construction worker detection in video frames for initializing vision trackers," *Automation in Construction*, vol. 28, pp. 15 – 25, 2012.

[7] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.

[8] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "Robust multi-person tracking from a mobile platform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1831–1846, 2009.