35th International Symposium on Automation and Robotics in Construction (ISARC 2018)

Construction Worker Detection and Tracking in Bird's-Eye View Camera Images

M. Neuhausen, J. Teizer and M. König

Chair of Computing in Engineering, Ruhr-University Bochum
E-mail: marcel.neuhausen@ruhr-uni-bochum.de, jochen.teizer@ruhr-uni-bochum.de, koenig@inf.bi.rub.de

Abstract -

Construction sites are continuously changing environments where construction workers have to adapt to dynamic situations while executing their work tasks safely and efficiently. Simultaneously, they are surrounded by heavy construction machinery and massive crane loads which they have to be aware of at any time. Frequent interruptions may often lead to a loss in productivity and also causes hazardous conditions or even incidents, injuries or fatalities. Tracking workers' paths on site can be used to approach these issues. Recorded tracks can be used to identify close calls of inexperienced or distracted workers. To date, pedestrian workers may participate in customized trainings in order to overcome individual deficits. Machine operators can be assisted to mitigate hazardous situations by warning them from construction workers approaching their machines.

Since surveillance cameras are already existent on most construction sites, a video-based detection and tracking system can be implemented at low costs. Relying on video streams, the detection of workers becomes similar to pedestrian detection. Some effort has already been made to elaborate those methods to the needs of construction worker detection. However, in contrast to the frontal view supposed in most pedestrian detection approaches, cameras on construction sites commonly provide oblique or bird's-eye view perspectives. This complicates the detection task as most body parts of a worker are occluded. Hence, we evaluate the applicability of pedestrian detection approaches in terms of the camera settings at hand. Ensuing, we propose a concept for the detection and tracking of construction workers which allows to improve the productivity and safety on construction sites.

Keywords -

Cascaded Classifier, Detection, Monitoring, Productivity, Safety, Tracking.

1 Introduction

Construction sites constitute highly complex, dynamic environments. Often, workers execute their working routines in collaboration with heavy machinery operating in their own workspace. For example, excavators and other construction vehicles cross the paths of construction workers and cranes lift massive loads over their heads. A variety of hazardous situations may arise from that such as incidents involving material and equipment or injuries and fatalities happening to workers or persons standing nearby. Hence, construction workers have to be trained to be sensitized for potential hazards. Hazardous situations such as close calls can directly be identified from the workers' actual paths across the construction site. Recording these could facilitate the identification of the necessity of further trainings as well as the customization of trainings with respect to the workers' deficits.

Concurrently, machine operators should be assisted in recognizing potential hazards involving workers in advance so that accidents can be avoided. Augmenting their field of view with the positions and walking path trajectories of workers as sketched in Figure 1 could further improve the safety on site as it enables operators to perceive distracted workers even if they are laboring in blind spots. Warning machine operators of such hazards could mitigate such situations which otherwise could end fatal.

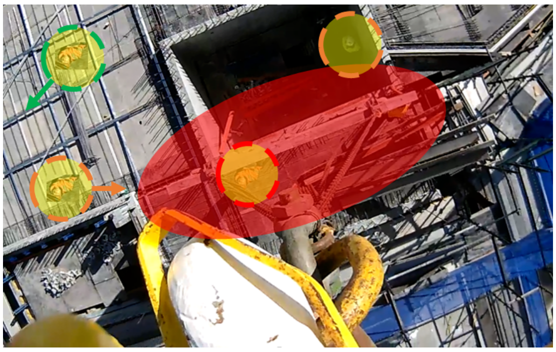

Figure 1. Augmented field of view of a crane operator. Yellow circles represent detected construction workers color-coded according to their level of hazard. Green highlighted workers are in sufficient distance to the hazard zone of the crane's load (red ellipsoid) and are predicted to walk away from it. Workers are marked orange if they approach the hazard zone or already stand near to it. Workers exposed to the hazard are marked red.

The obtainment of the construction workers' trajectories in construction-related areas of a site and in the surrounding of machinery, thus, allows to significantly decrease safety issues. Further application areas may also benefit from this, such as checking workers for wearing appropriate safety equipment, site access control systems, or even the improvement of productivity. A reasonable tracking of workers during their working routines tackles multiple issues on site and is already in the scope of research. Depending on the surrounding, common approaches rely on different radio signal emitters like radio-frequency identification (RFID), ultra-wideband (UWB), and global navigation satellite system (GNSS) tags. These are attached to the gear of each construction worker which involves high costs [1] and causes discomfort [2]. In contrast, a camera based tracking of workers on construction sites provides a cost-saving and less obtrusive approach.

Relying on video streams, the task becomes similar to pedestrian detection which is a well researched topic today. There are already approaches adapting methods of this field for construction worker detection. Nevertheless, crucial differences between the application areas have to be taken into account. Due to the actual scope of application in advanced driver assistance systems, pedestrian detection approaches commonly assume a frontal view. In contrast, surveillance cameras on construction sites as well as cameras for crane operator assistance are usually mounted in heights to existing structures resulting in bird's-eye view perspectives recording the workers sometimes even over far distances. This complicates detection since most parts of the bodies are occluded by the workers' heads and shoulders or physical obstructions. Therefore, we evaluate previous pedestrian detection approaches with respect to their applicability in bird's-eye view images. Based on this research objective, we propose a conceptual system which improves over previously made approaches tracking construction workers for safety and productivity issues. For this purpose, we apply a background subtraction method which identifies regions of interest in the camera image. A cascaded classifier investigates these regions and detects construction workers. In order to track the detected workers through multiple camera frames, we apply Kalman filtering.

The remainder of this paper is structured as follows: in Section 2, we motivate our work by surveys and studies concerning safety and productivity issues. We investigate previously made approaches of pedestrian detection in Section 3 and discuss these with regard to the limitations in the video data in the field of construction monitoring. Based on the findings, a conceptual system for tracking construction workers on sites is proposed (see Sec. 4). Finally, we discuss the results and conclude on our concept in Section 5.

2 Motivation

High rates of injuries and fatalities in construction are often explained by its complex, dynamic, and continuously changing work environment. Whereas cranes play a central role in construction operations, federal labor statistics in many countries relate about 15-25% of all fatal construction workplace accidents to too close proximity of pedestrian workers to construction equipment or hazardous materials [3]. Struck by moving parts of crane equipment or hit by falling objects are some of the most frequent causes of crane-related construction accidents [4]. Their outcome is often fatal [5], which distinguishes crane-related accidents from the majority of other construction accidents where the outcomes are minor (e. g., cut in finger).

Cranes come typically in numerous configurations to fit unique sites [6]. Cranes carrying loads over, into, and/or around workers' environment add yet another dimension of risk to an already complex workspace. Though safe crane design, several on-board safety devices and operational procedures exist, large problems remain to operate them safely. Research studies state:

- About 16% of all construction fatalities relate to cranes [7]
- 33% of all construction casualties and permanent disabilities relate to cranes [8]
- 87% of crane-related deaths occur among workers and do not involve operators [9]
- Few in the transient construction workforce have operating or rigging experience [5]
- Little information about the causal factors or environments leading to the accidents or close call events is known [10]

Few safety statistics from around the world exist that explain the problem in detail. The Center for Construction Research and Training analyzed data collected by the U.S. Bureau of Labor Statistics in the years 1992-2006 [11]. In this time period, 632 crane-related deaths were identified which occurred in 610 crane incidents. These numbers equal to an average of 42 crane-related deaths per year. Whereas mobile or truck cranes (at least 71%) were the main types of cranes that have been associated with crane-related fatalities, tower cranes (5%), floating or barge cranes and overhead cranes, and other/unspecified cranes are the other crane types (24%). Of the total 632 crane-related deaths, 157 (25%) were caused by overhead power line electrocutions, 132 (21%) deaths were associated with struck by crane loads, 89 (14%) involved crane collapses, 78 (12%) involved a construction worker being struck by a (i.e. falling) crane boom/jibs, 56 (9%) included falls from cranes/crane baskets/crane loads, 47 (7%) were struck by crane or crane parts, 30 (5%) caught-in between, and 43 (7%) deaths were from other causes. The activities immediately preceding the workers' deaths related to deaths from struck by crane loads (132 between 1992-206) were: 32% of the workers were not involved with the crane, 32% loaded/unloaded, 15% performed other crane-related work, 14% were flagging/directing/guiding, and 7% operated the crane. The majority of the workers belonged to laborers (191 deaths), while the others trades were heavy equipment operators (101), supervisors/managers/administrative (86), ironworkers (42), mechanics (41) and other trades (171).

In their recommendation to prevent crane accidents from happening in the future, CPWR recommends the following actions to take: crane operators should be certified; crane riggers and signalpersons should be adequately trained; crane inspectors should be qualified; cranes should be inspected; only qualified and competent persons should assemble, modify or disassemble a crane; cranes should not be allowed to pass over street traffic; and more thorough investigations and immediate follow-ups should be performed. Despite the poor safety performance of cranes in construction and several recommendations that are already part of many construction safety leaders' best practices, no proactive approach has been taken towards detecting and resolving the identified crane-related hazards [12]. For example, when a crane load swings over an active worker environment, pedestrian workers and crane operator should be warned of the risk of falling objects or being struck-by.

In Germany, construction occupational safety and health is embodied in and shaped by numerous laws, regulations and ordinances with a view to ensuring the safety and health of construction workers in the workplace. Technical Occupational Safety and Health includes all areas that affect the safety of workers at work. The Safety and Health at Work Act (ArbSchG) regulates the underlying occupational safety and health duties of the employer, the duties and rights of workers, and the monitoring of occupational safety and health in accordance with this Act. In the control hierarchy (a) technical, (b) organizational, and (c) personal measures are typically embedded in an organization's safety culture. These respectively and whenever possible, (a) avoid hazards in the first place by replacing hazardous work practices with safer ones and separate workers from hazardous workspaces, (b) limit the exposure time to hazards, and (c) provide personal protective equipment (PPE) and instruct personnel [13]. Again, existing regulations, rules or best practices on safety in construction do not envision proactive solutions other than education, training, and enforcement. For example, by means of using technology, pedestrian workers and operators could be automatically warned in real-time from nearby or approaching crane loads. However, such technology does not exist today [14].

3 Detection Methods

The safety of construction workers on site can be improved by focusing on their working behavior. By tracking their paths across the site, hazardous situations like close calls can be identified. Additionally, operators can be warned from potentially distracted workers approaching their machines too closely. For tracking workers on construction sites different technologies have already been applied. Most prominent are radio signal emitters like RFID, UWB, and GPS tags which are attached to the gear of the construction workers. Besides the imposed costs for equipping each worker on the site [1], theses tags are perceived to be obtrusive, resulting in discomfort accompanied by a decrease in motivation [2]. Video-based techniques overcome these deficiencies and allow for a uniform method of detecting and tracking workers all over the site. As nowadays cameras are ubiquitous and not primarily meant to control workers, they can be considered to be less obtrusive. Moreover, by making use of already existent surveillance cameras only few cameras have to be additionally installed, keeping new investments and maintenance costs low.

Identifying workers in video streams is similar to pedestrian detection. By now, this is a well understood field of research in the computer vision area which already provides a variety of satisfying methods. Especially the automotive industry continuously advances the current methodology. Due to the usual application area in advanced driver assistance systems, approaches in this field assume frontal images of pedestrians. Some approaches already adapt those methods to construction worker detection [15, 16]. Referring to this, Park and Brilakis [16] propose a two-parted detection approach. They learn shape features using a support vector machine (SVM) to identify people in frontal view images. Using color features these detections are further processed by a k-nearest neighbor (k-NN) classifier to detect construction workers by their safety vests. On construction sites, however, frontal view images are merely an exceptional case. Surveillance cameras are usually mounted to high posts, scaffolds, or on nearby building facades or roofs. In particular, for assisting crane operators at lifts, cameras have to be mounted to the jib or at least high on the crane tower. Detection and tracking of workers on construction sites, thus, have to be done in bird's-eye view images. This complicates the detection task as the workers' bodies are barely visible as can be seen in Figure 2. For this reason, identifying construction workers on site requires a robust detector which yields reasonable results despite sparse indications for the presence of a worker caused by the challenging perspective.

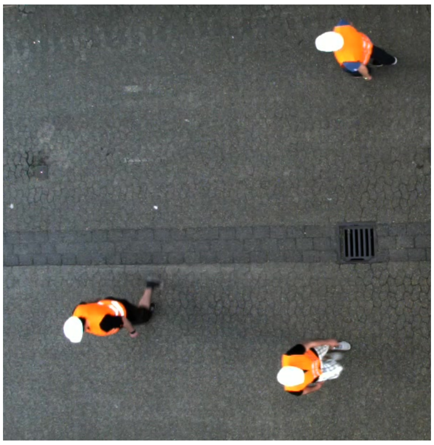

Figure 2. Construction workers recorded by a surveillance camera mounted 15 m above ground.

In comprehensive surveys Dollár et al. [17] and Benenson et al. [18] summarize the state-of-the-art pedestrian detection algorithms and evaluate their performance. Benenson et al. propose to categorize the approaches into deformable part-based models (DPM), deep learning, and decision forests.

DPM Approaches based on DPM subdivide an object into a star-structured part-based model consisting of a root object and multiple parts attached to it [19]. According to this, a latent SVM can be trained using a pyramid of histograms of oriented gradients (HOG) features in order to classify pedestrians by detecting their body parts [20, 21]. Advancing the part-based model towards a multi-resolution structure improves detection results [22]. Nevertheless, in our case the view is generally narrowed to the heads and shoulders of pedestrian construction workers wearing personal protective equipment (PPE). Approaches relying on the detection of silhouettes and body parts, consequently, are ineligible in this context.

Deep Learning Deep learning comprises approaches using large artificial neural networks. These can be used for object detection by extracting features from the image data. Sermanet et al. [23] apply a convolutional neural network which learns relevant features from the training data. Albeit the network yields fair results on similar data sets, it fails on generalization. Up to now, it has not been shown that deep learning approaches can be used to learn sufficient image features [18]. For other deep learning approaches [24, 25] features have to be predefined manually. It is doubtful if such heavy techniques are necessary to evaluate manually selected features. Furthermore, the latter approaches again pursue the part-based idea which is not applicable for bird's-eye view images. Since the results obtained by deep learning on pedestrian detection tasks are yet at the same level with DPM and decision forests, advantages of deep learning are still questionable [18]. Hence, simpler methods should be preferred instead.

Decision Forests Decision forests are ensembles of decision trees in which the nodes represent weak classifiers. Samples are classified by passing through the trees. Viola and Jones [26] proposed an approach using AdaBoost to train a pruned decision tree with Haar features as weak classifiers. Originally developed for face detection, they showed that it is also applicable to pedestrian detection [27]. Bourdev and Brandt [28] improved the method by promoting the confidence of each evaluated weak classifiers through the tree. Coupled with a generalized feature approach, Dollár et al. [29] showed that this method outperforms previous pedestrian detectors. This indicates that the detection results of decision forests improve with the development of features. By now, other pedestrian detectors achieve a similar detection quality compared to decision forests. Nevertheless, boosted decision trees usually outperform monolithic classifiers like SVM on most detection tasks [30]. Furthermore, boosting automatically selects the most suitable set of features from a given pool. This overcomes the need for evaluating features manually as it is commonly unclear which features qualify best for a certain task.

Whereas Benenson et al. mainly categorize the approaches by their machine learning algorithms, Dollár et al. focus on the sets of features. They found that gradient-based features like HOG [31] are most prominent. Besides this, shape features [32, 33] and motion features [27] are frequently used. While HOG feature approaches perform best in comparison to other single feature settings, even better detection results can be achieved when combined with multiple features providing complementary information. Accordingly, combining Haar-like features, shapelets, shape context, and HOG features outperforms any single feature approach [34]. In their study, Dollár et al. [29] focus on the choice of features and propose a framework to efficiently compute multiple features based on integral channels.

4 Concept

We evaluated previously made pedestrian detection approaches regarding the requirements and general conditions of construction workers detection in bird's-eye view images. In the following, we develop a concept for tracking construction workers on site considering the insights gained in Section 3. Figure 3 depicts the conceptual system in total.

According to the findings in Section 3, methods relying on DPM are not applicable for our purpose as the perspective does not allow for the detection of body parts. Also deep learning is not preferable in this context since it could not be shown that deep neural networks are advantageous over other approaches. In contrast, decision forests prove to be well suited for construction worker detection in bird's-eye view images.

For our conceptual worker tracking system, we propose a single classifier approach using a decision tree as this highly improves the speed over the two-parted detection system by Park and Brilakis [16]. We decide for the soft cascaded approach proposed by Bourdev and Brandt since its detection results exceed those of the Viola-Jones detector. Additionally, its likewise pruned cascading layout further improves the speed of classification by rejecting negative samples early in the cascade. Hereby, the entire cascade has to be processed for positive samples only; for negative samples only few image feature are evaluated before rejection. This eases real-time processing on video data. Moreover, since it is unknown which image features qualify for this task, boosting automatically selects the optimal set of features for our purpose. Accordingly, a manual selection as in the approach of Park and Brilakis [16] is not required.

We exchange the originally proposed thresholded Haar features by integral channel features. This further improves the detector as various image feature types can be efficiently computed using integral channels, and the combination of multiple feature types advances the detection quality over single feature approaches. Similar to Park and Brilakis [16], our concept relies on the identification of shape and color in order to detect construction workers. Since features responding to contours yield reasonable results on pedestrian detection, we draw on adjacent second-order integral channel features on grayscale images. These are equivalent to Haar features which act as edge detectors and, thus, indicate shape information. Apart from contours, construction workers are commonly characterized by their PPE including their helmets and safety vests. Thus, color features may improve detection by incorporating the prominent colors of these items. Color histograms can be efficiently computed by applying first-order integral channel features to quantified versions of each color channel separately. For both feature types, illumination has to be taken into account as it affects the features' responses. Grayscale images can be variance normalized to minimize the influence of different lighting conditions. In case of color histograms, the color space has to be chosen properly. In Figure 4 we compare RGB and HSV color spaces with respect to safety vests in different illumination. As can be seen, color histograms over red, green, and blue channels significantly change when altering the lighting conditions. As brightness is implicitly encoded in each channel of this color space, varying the illumination conditions directly affects the color information of these three color channels. Contrarily, in the HSV color space hue and saturation channels are invariant to illumination since brightness is explicitly encoded in the value channel so that only hue and saturation represent color information. Color spaces like HSL or YUV should also be considered as these offer equivalent channel characteristics.

Instead of scanning the entire image for construction workers we favor background subtraction beforehand. By identifying areas of motion within each video frame, classification can be limited to regions of interest. Albeit the speed of the soft cascaded classifier allows to scan images in real-time, restricting the scope of the classifier may significantly reduce false positive detection. Areas of motion can be identified by frame differencing. Assuming the previous video frame as background, subtracting it from the current frame reveals changes which imply movement (see Fig. 5). By thresholding the results, its sensitivity can be controlled:

$$|I_{t-1}(x, y) - I_t(x, y)| > \tau$$

where $I_{t-1}(x, y)$ and $I_t(x, y)$ denote the previous and current video frames and τ is the threshold.

The quality of the results highly depend on the choice of the threshold τ. Furthermore, image noise and fast illumination changes may be interpreted as motion. For a higher robustness we advance to learn a more complex background model. By averaging the background over multiple frames [35] a background model emerges which is insensitive to momentary changes between few frames.

In order to predict the paths of construction workers, the detected workers have to be tracked throughout the video frames. For this, we suggest the application of Kalman filtering to the detections. Given position and velocity data of a detected construction worker, the Kalman filter predicts its future state ongoing. Subsequent measurements of position and velocity are fed into the Kalman filter to reduce the uncertainty of the predictions. This approach also supports the detector as it provides further regions of interest in which construction workers may be detected even if they are standing still.

5 Conclusion

On construction sites, workers are exposed to a variety of hazards. Construction machines cross their paths and loads are lifted over their heads. This often results in close calls or even accidents. Thus, construction workers have

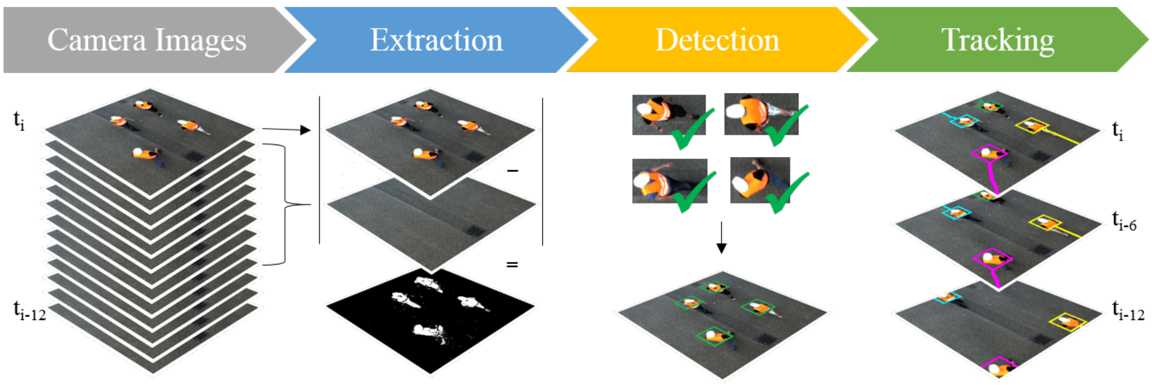

Figure 3. Conceptual system for tracking construction workers on site in bird's-eye view images. Initially a foreground extraction of the incoming camera images is made via background subtraction averaging the background based on a certain number of previous frames. Construction workers are, then, detected in the foreground by means of edge and color feature. Using a Kalman filter, these detections are tracked in the camera frames over time.

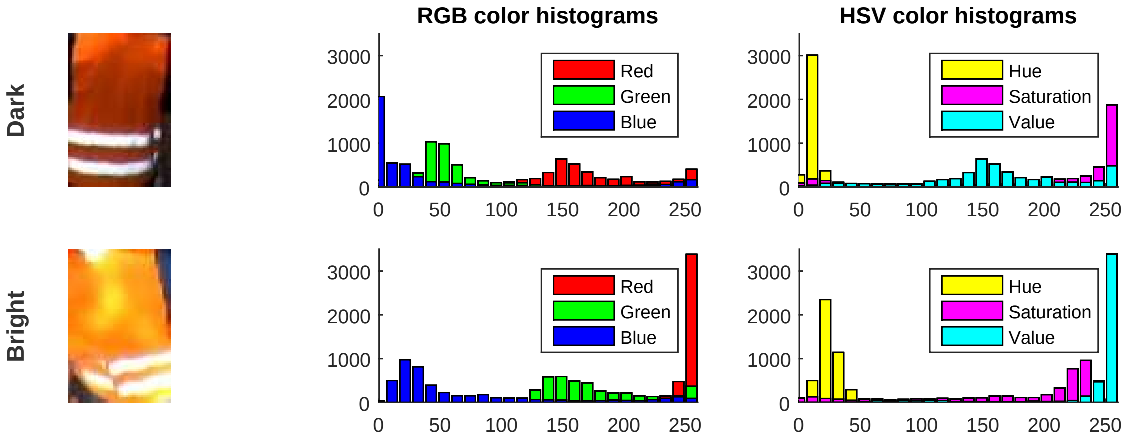

Figure 4. Comparison of histograms in different color spaces with respect to a safety vest (first column) in dark (first row) and bright (second row) illumination. In the RGB color space (second column) high variation occurs in the peak positions of each channels. On the contrary, hue and saturation channels of the HSV color space (third column) are significantly more robust to illumination changes.

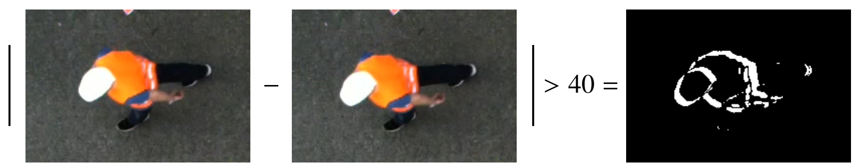

Figure 5. Frame differencing applied to two consecutive frames showing a construction worker with a threshold τ of 40. White areas in the resulting binary image indicate areas with motion.

to be trained individually in order to sensitize them for potential hazards. Tracking workers during their working routines in construction-relevant areas on sites may improve the customization of the trainings since these can be adjusted to individual deficits. Furthermore, the tracking of workers can be used to avoid accidents involving heavy machinery by assisting machine operators and providing position and walking path information. The optimal for communicating such valuable and potentially life-saving information has yet to be found for applications in construction.

A video-based tracking system offers a cost-efficient and uniform alternative compared to conventional tag-based methods. Although detecting construction workers is similar to pedestrian detection, the perspective of high mounted cameras complicates the task. Thus, we investigated common pedestrian detection approaches with respect to the applicability for construction worker detection in bird's-eye view images. We discussed detectors of different categories and found that decision forests qualify best. Using a multi-feature approach enhances common edge feature approaches with complementing information which further improves detection. For this reason, we prefer to apply a soft cascaded classifier in our conceptual tracking system. As weak classifiers, we propose edge features and color histograms of the integral channel feature approach. Background subtraction is used to focus the classifier on regions of interest and to reduce false positive detections. In order to track detected workers and to predict their future walking paths, Kalman filtering can be applied.

References

[1] E. Nasr, T. Shehab, and A. Vlad. Tracking in Construction: Applications and Comparisons. In Proceedings of the Annual Associated Schools of Construction International Conference, 2013.

[2] A. Juels. RFID security and privacy: A research survey. IEEE Journal on Selected Areas in Communications, 24:381–394, 2006.

[3] J. W. Hinze and J. Teizer. Visibility-related fatalities related to construction equipment. Journal of Safety Science, 49(5):709–718, 2011. doi: 10.1016/j.ssci.2011.01.007.

[4] A. Shapira and B. Lyachin. Identification and analysis of factors affecting safety on construction sites with tower cranes. Journal of Construction Engineering and Management, 135(1):24–33, 2009. doi: 10.1061/(ASCE)0733-9364(2009)135:1(24).

[5] R. L. Neitzel, N. S. Seixas, and K. K. Ren. A review of crane safety in the construction industry. Applied Occupational and Environmental Hygiene, 16(12): 1106–1117, 2001. doi: 10.1080/10473220127411.

[6] T. Cheng and J. Teizer. Modeling tower crane operator visibility to minimize the risk of limited situational awareness. Journal of Computing in Civil Engineering, 28(3), 2014. doi: 10.1061/(ASCE)CP.1943-5487.0000282.

[7] S. G. Pratt, S. M. Kisner, and P. H. Moore. Machinery-related fatalities in the construction industry. American Journal of Industrial Medicine, 32:42–50, 1997.

[8] D. V. MacCollum. Crane hazards and their prevention. American Society of Safety Engineers, 1993.

[9] OSHA. Strategic Plan FY1997-FY2000. Crane and Hoist Safety, 1996.

[10] O. Golovina, J. Teizer, and N. Pradhananga. Heat map generation for predictive safety planning: Preventing struck-by and near miss interactions between workers-on-foot and construction equipment. Automation in Construction, 71:99–115, 2016. doi: 10.1016/j.autcon.2016.03.008.

[11] M. McCann. Understanding crane accident failures: A report on causes of deaths in crane-related accidents. Presentation, www.elcosh.org/record/document/2053/d001029.pdf, 2010. last accessed July 18, 2018.

[12] J. Teizer. Right-time vs real-time pro-active construction safety and health system architecture. Construction Innovation, 16(3):253–280, 2016. doi: 10.1108/CI-10-2015-0049.

[13] Mechanische Gefährdungen – Maßnahmen zum Schutz vor Gefährdungen beim Verwenden von mobilen Arbeitsmitteln. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Ausschuss für Betriebssicherheit, 2015. TRBS 2111, Teil 1.

[14] A. Shapira, Y. Rosenfeld, and I. Mizrahi. Vision system for tower cranes. Journal of Construction Engineering and Management, 134 (5):320–332, 2008. doi: 10.1061/(ASCE)0733-9364(2008)134:5(320).

[15] J. Teizer and P. A. Vela. Personnel tracking on construction sites using video cameras. Advanced Engineering Informatics, 23(4):452–462, 2009. doi: 10.1016/j.aei.2009.06.011.

[16] M. Park and I. Brilakis. Construction worker detection in video frames for initializing vision trackers. Automation in Construction, 28:15–25, 2012. doi: 10.1016/j.autcon.2012.06.001.

[17] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):743–761, April 2012. doi: 10.1109/TPAMI.2011.155.

[18] R. Benenson, M. Omran, J. Hosang, and B. Schiele. Ten years of pedestrian detection, what have we learned? In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, editors, Computer Vision – ECCV 2014 Workshops, pages 613–627. Springer, 2015.

[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9): 1627–1645, 2010.

[20] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2008.

[21] J. Yan, Z. Lei, L. Wen, and S. Z. Li. The fastest deformable part model for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2497–2504. IEEE, 2014.

[22] D. Park, D. Ramanan, and C. Fowlkes. Multiresolution models for object detection. In Proceedings of the European Conference on Computer Vision, pages 241–254. Springer, 2010.

[23] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun. Pedestrian detection with unsupervised multi-stage feature learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3626–3633. IEEE, 2013.

[24] W. Ouyang and X. Wang. Joint deep learning for pedestrian detection. In Proceedings of the International Conference on Computer Vision, pages 2056–2063. IEEE, 2013.

[25] W. Ouyang and X. Wang. A discriminative deep model for pedestrian detection with occlusion handling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3258–3265. IEEE, 2012.

[26] P. Viola and M. J. Jones. Robust real-time face detection. International Journal of Computer Vision, 57(2):137–154, May 2004. doi: 10.1023/B:VISI.0000013087.49260.fb.

[27] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In Proceedings of the IEEE International Conference on Computer Vision, pages 734–741, Oct 2003. doi: 10.1109/ICCV.2003.1238422.

[28] L. Bourdev and J. Brandt. Robust object detection via soft cascade. In Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 236–243. IEEE Computer Society, 2005.

[29] P. Dollar, Z. Tu, P. Perona, and S. Belongie. Integral channel features. In Proceedings of the British Machine Vision Conference, pages 91.1–91.11. BMVA Press, 2009. doi: 10.5244/C.23.91.

[30] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In B. Michaelis and G. Krell, editors, Pattern Recognition. DAGM 2003. Lecture Notes in Computer Science, volume 2781, pages 297–304. Springer, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-540-45243-0_39.

[31] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 886–893. IEEE, 2005.

[32] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In Proceedings of the International Conference on Computer Vision, pages 90–97. IEEE, 2005.

[33] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2007.

[34] C. Wojek and B. Schiele. A performance evaluation of single and multi-feature people detection. In Gerhard Rigoll, editor, Proceedings of the Joint Pattern Recognition Symposium, pages 82–91. DAGM, Springer, 2008.

[35] M. Piccardi. Background subtraction techniques: a review. In Proceedings of the International Conference on Systems, Man and Cybernetics, volume 4, pages 3099–3104 vol.4. IEEE, Oct 2004. doi: 10.1109/ICSMC.2004.1400815.