ALFA BK UNIVERSITY
FACULTY OF INFORMATION TECHNOLOGIES
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
International Scientific Conference

UDK:004:343.982.323
DOI: 10.46793/ALFATECHproc25.264B
Original scientific paper

Person detection and re-identification in open-world settings of retail stores and public spaces

Branko BRKLJAČ[1], Milan BRKLJAČ[2]

Abstract: Practical applications of computer vision in smart cities usually assume system integration and operation in challenging open-world environments. In the case of person re-identification task the main goal is to retrieve information whether the specific person has appeared in another place at a different time instance of the same video, or over multiple camera feeds. This typically assumes collecting raw data from video surveillance cameras in different places and under varying illumination conditions. In the considered open-world setting it also requires detection and localization of the person inside the analyzed video frame before the main re-identification step. With multi-person and multi-camera setups the system complexity becomes higher, requiring sophisticated tracking solutions and re-identification models. In this work we will discuss existing challenges in system design architectures, consider possible solutions based on different computer vision techniques, and describe applications of such systems in retail stores and public spaces for improved marketing analytics. In order to analyse sensitivity of person re-identification task under different open-world environments, a performance of one close to real-time solution will be demonstrated over several video captures and live camera feeds. Finally, based on conducted experiments we will indicate further research directions and possible system improvements.

Keywords: person re-identification; computer vision; retail stores; public spaces; marketing analytics;

1 INTRODUCTION

Modern computer vision techniques are solving complex technical challenges posed by applications in open-world settings. This refers to scenarios in which the system must overcome uncontrolled and difficult to predict characteristics of the operating environment. This is also the case with person detection [1] and person re-identification [2] tasks, which have traditionally been researched in the context of autonomous vehicles [3], robotics [4], human-machine interfaces [5], crowd analysis [6], and most commonly security applications [7]. The concept of smart cities [8] and the integration of such vision systems [9]–[11] into complex decision-making frameworks [12]–[14] have introduced the need for efficient and robust solutions, with the ability to make necessary trade-offs between operational costs and system performance on a case-by-case basis. The common goal is providing necessary insights into persons' behaviour based on the ability to detect and possibly re-identify the same person at different points in time and space [15]. Such applications also raise the questions of ethical and privacy nature, which is a significant line of current research [16], [17]. Nevertheless, integration of behaviour analysis systems in retail stores and public spaces is gaining momentum [18] and is expected to have significant impact on insightful business operations and space management.

Approaching potential customers based on research analysis of their behaviour can be considered a key ingredient for designing successful marketing strategies and campaigns. It is also manifestation of the general trend of tailored campaigns and data-driven decisions emerging across different modalities of interaction with retail shoppers and public space users. In other words, persons' movement patterns and time spent in different parts of a store or elsewhere directly answer the questions: "How do consumers shop?", or "How people engage in public space?". Data driven insights into these questions are foundations for comprehensive understanding of their behaviour. Thus, behavioural constructs in different conceptual frameworks, like e.g. emotional regulation consumption or compensatory consumption [19], can be better related with the motivational and emotional aspects of the shopping process and public infrastructure engagement. It is also the kind of data that cannot be obtained from point-of-sale [20] or simple visitor statistics [21]. Therefore, spatial analysis of human behavior can be considered one of the complementary technologies for achieving optimal decisions and reducing the cost of adapting to individual needs of customers or users. In this paper, we provide a brief overview of methods and approaches for addressing design challenges for person detection and person re-identification in open-world environments and describe their use in retail sector and public spaces. We also present a demonstration of one cost efficient, close to real-time solution based on OAK-D lite [22] embedded vision platform [23] and discuss advantages and limitations of such hardware implementation and baseline algorithmic approach.

The rest of the paper is organized as follows. In Section II we break down the landscape of existing methods for person detection and person re-identification tasks and provide a structured overview of research directions that have been established by pursuing different solution architectures under the assumptions of specific imaging modalities, identification scenarios, robustness requirements and model learning strategies. In Section III we go into implementation details of developed low-cost solution and indicate advantages and limitations of the adopted fast prototyping system design approach. Section IV presents conducted experiments and discusses their implications for possible applications in retail marketing analytics and public space management. Finally, in Section V we indicate directions for future research and possible system improvements.

2 RELATED WORK

Person detection [1], [10], [11], alongside face detection [24], [25], is probably one of the most studied object detection tasks in the literature [26]. Similarly, person re-identification (ReID) [2], alongside face recognition [27], is one of the long-standing recognition problems in vision [28], [29]. Its goal is to retrieve information whether the specific person has already appeared in video or over multiple camera feeds in the past. In such case, re-identification is considered as successful if we can correctly assign person's identity with some of the identities already present in the gallery, or create a new identity entry in the case of novel person in the scene. Described working scenario corresponds to the open-set ReID task, which is different from the closed-set setting in which it is assumed that the query object already exists in the gallery. These two problems, detection and re-identification, can be approached either traditionally, i.e. separately [30], [31], in a sequential manner, or jointly, i.e. in an end-to-end manner [32], [33], which is a characteristic of more recent ReID approaches. In the following we provide a brief overview of methods utilized in these two specific tasks of interest.

2.1 Detection task

In general, besides the problem of defining object categories [34], object detection approaches solve two fundamental problems in vision: localization and classification of object instances inside the scene. Consequently, binary classification of persons in the image can be formulated as a two stage: localization and detection problem, or as a single stage, i.e. one step regression task [26]. As there are numerous approaches for person detection, their applicability depends on the cost-performance trade-off put in front of system designer. The role of the detector in the broader context of the system in which it operates is also important. In this sense, the choice of an adequate working point on the detector's receiver operating characteristic (ROC) is also a matter of compromise and the requirements set regarding the subsequent processing of generated information. Thus, precision and recall, or the relationship between detector's sensitivity (hit rate) and specificity (false positives rate) play a crucial role in the design choice of accompanying or downstream processing elements that are relying on detector's decisions. However, in some cases, greedy approaches in which the possibility of a miss is avoided by increasing the sensitivity of the detector, despite a higher number of false alarms, are also justified if it is possible to adequately use such results. Such techniques may include the use of detection or confidence scores in the case of still images, spatial filtering based on the temporal history of the object's locations in the case of video signal, or additional side information obtained from other sources. In the given context of public spaces monitoring and retail analytics, it is interesting to note that usual detection paradigm can be modified in order to overcome some of the difficulties posed by the application scenario. Thus, a new object detection paradigm called Locount is proposed in [35], with the goal of localizing groups of objects of interest and estimating number of instances within the group. It is similar to the type of problems considered in [6], which indicates that the task of person detection is still of active research interest. This is especially true in the case of open-world environments, where adaptation and resilience to various aggravating circumstances such as illumination variation, image noise, presence of occlusions, and variations in scale or resolution of persons in the video are often needed. It is achieved through advanced learning techniques, data augmentation, or combining of different signal modalities.

2.2 Re-identification task

According to [36], image-based person re-identification methods can be broadly grouped by: 1) signal modality (e.g. visible, infrared [37], thermal [38], or cross-modal [39]); 2) learning approaches (e.g. whether the recognition is based on metric learning [40], [41] or more complex deep ReID architectures and learning frameworks [31], [42]); and 3) how generalization across multiple domains, i.e. cross-domain ReID is addressed [43]. Video-based person ReID methods [44]–[46] that exploit temporal information in the learning process are also seen as a separate line of research, with a similar division of methods with respect to the above-mentioned properties. Other aspects in relation to which ReID solutions are considered include different types of challenges where standard methods exhibit certain weaknesses, leading to classes of special, specific methods that effectively overcome the limitations of standard solutions. Thus, the main characteristics of effective ReID solutions aare invariance to different variations of the input signal and robust hierarchical feature representations that lead to accurate identity matching.

Resolution mismatch [47], when a low resolution query image is matched against the high resolution gallery samples, is a common example of ReID problem that is affected by the scale of the object and image acquisition setup. Opposite situations of high resolution query images are also common, giving rise to resolution independent features [48] and multiscale pyramid representations [45].

The most common person ReID goal is a short-term identification, where it is assumed that the pedestrian's clothing remains unchanged over time. While such assumptions are fully justified in the case of described retail application scenarios, a more general solution would be the invariance of ReID to changes in person's clothing [49] and appearance [50], when captured during different activities. Thus, a step forward ideal long-term or lifelong person ReID would be invariance to cloth-changing, which could be regarded as more applicable in real-world scenarios. In [51] authors claim that the most reliable discriminative characteristic for above-mentioned invariance is unclothed body shape, i.e. effective representation of the 3D shape and texture of the human body. It is closely related to other biometric identification methods based on human gait recognition [4], [52], which are complementary to classical image- or video-based ReID [53].

Significant class of adaptation methods also includes occluded person ReID [54]–[57]. A common strategy is to steer ReID toward decisions based on features that comprehend or do not take into account the hidden part of the person. Such examples include learning the dynamic prototype mask [54], feature attention mask [58], saliency-guided patch transfer [56], using multiple views of the same person [55], or relying on part based representations [59] and fine-grained image analysis [60], see e.g. [61], [62]. There have also been attempts to partition feature space [63], while the standard learning strategy for person ReID has been the triplet loss framework [64], [65] and its extensions [66]. Alternatives include various deep metric learning approaches [40] and ranking optimization at inference stage [42].

2 PROPOSED PERSON ReID IMPLEMENTATION

In order to analyze the sensitivity of person re-identification task under open-world setting of retail stores and public spaces we have implemented one cost efficient, close to real-time solution. It is based on OAK-D lite [23] device and pre-trained person detection and person ReID models made using OpenVINO [67] hardware acceleration deep learning framework. The corresponding camera device is shown in Fig. 1 and its main characteristics are summarized in Table I.

Table 1 Main characteristics of OAK-D lite platform

Feature	Specification
Processor	Intel Movidius Myriad X VPU
Cameras	1x 13 MP RGB (IMX214, rolling shutter) 2x 0.31 MP mono (OV7251, global shutter)
Stereo baseline	75 mm
AI performance	4 TOPS (1.4 TOPS for AI)
Video output	Up to 4K@30fps (H.264/H.265/MJPEG)
Depth perception	Stereo, 300,000-point, 200+ FPS
Connectivity	USB 3.1 Gen1 Type-C
Dimensions	(WxHxD), 91 mm x 28 mm x 17.5 mm
Weight	61 g
Variants	Fixed-focus / Auto-focus (RGB cam)
Mounting	1/4”-20 tripod, VESA (7.5 cm, M4)

As can be seen from Table I, it is a versatile and heterogeneous computing platform. It is suitable for applications in different vision tasks that require real time stream processing and advantages of using embedded hardware implementation, such as high level of integration, small form factor, modular design, low power consumption and relatively low cost, taking into account that the device integrates both imaging devices and signal processor. Besides providing the options of choosing between the global and rolling shutter, on board cameras are also designed to be geometrically precisely aligned in a linear arrangement to facilitate calibration and accurate distance measurements. Nevertheless, in the given experiments we have relied only on the main high resolution RGB camera and multiple stream processing by frame grabbing from created camera source in the constructed processing graph.

Figure 1 Embedded vision platform used in experiments

OAK-D Lite platform integrates multiple specialized processing units and sensors that collaboratively handle different computational tasks, optimizing performance and efficiency. Thus, Movidius Myriad X Vision Processing Unit (VPU) integrates several system components two of which are of main interest for implemented person ReID. First one is Image Signal Processor (ISP), on-chip hardware accelerator responsible for different camera pipeline functions and high-throughput image pre-processing tasks such as denoising and color correction. The second one is Neural Compute Engine (NCE), neural network hardware accelerator, which enables efficient deployment of trained deep learning models or any other computational graphs that are possible to define in supported ONNX [68] or OpenVINO [67] network exchange formats. Its advantage is small power consumption, which is especially suitable for edge processing and deployment in environments that require continuous system operation. Note that although the main processing (person detection and person ReID) was done on embedded platform, experimental setup also required a host device (CPU) for higher-level program control and application logic (visualization, communication and I/O interface).

For software implementation, besides the device drivers and other utilities provided by Luxonis corporation under the MIT license in official DepthAI code repository [69], we have used OpenCV [70] and Python programming language. We note that there are also C++ device interfaces and code implementations, available in [71]. The main algorithmic solution was taken from OpenVINO repository and consists of the pre-trained person detection model, available in [72], and person ReID model available from [73]. As a starting point for developing additional functionalities we have utilized demonstration script in [74], from which the provided code implementation was made: https://github.com/brkljac/personReID. Possible alternative for baseline implementation could also be some general purpose ReID toolbox [75].

4 EXPERIMENTAL RESULTS AND DISCUSSION

The experiments included visual analysis of system performance in various operating environments. We have analyzed ability of system to cope with person ReID task in outdoor and indoor space, under different illumination and in the presence of varying number and dynamics of people. General conclusions is that in all considered scenarios system managed to achieve close to real time operation, but never more than $\approx$ 12 frames per second (fps). On the other hand, under certain difficult conditions, the fps dropped even to $\approx$ 4 fps. This included low-light conditions, but mainly the presence of very dynamic background noise in the scene. For example, in an outdoor scene where a significant portion of the background was covered by leaves, in the presence of wind, the background dynamics were very pronounced and caused additional slowdown. However, even in such cases the final ReID result was quite good and without any false detections in the described background area. The most challenging ReID were cases where the same person changed orientation towards the camera while walking or when entering and exiting the scene. This led to the creation of multiple identities for the same person, which is a common problem with person ReID. During experiments in an indoor retail space, it was found that certain object arrangements cause false person detection and initialize a correct but false person ReID in subsequent frames. This indicates that person detection module could be subject to further improvements in terms of robustness.

Most of the situations described above are illustrated in Fig. 2 and Fig. 3, while the corresponding video files are available in the given code repository, Section III.

The observed variability in people’s appearance could be overcome by model training on datasets like [76], [77], which contain a sufficiently diverse range of appearances (positions of the person relative to the camera, changes in illumination and different human activities). Another research direction are more frequent ReID model updates through better use of historical appereances and without re-indexing of images in the gallery [78].

5 CONCLUSIONS

In this paper, we analyzed solutions for person ReID from a technical and application perspective. Conducted experiments revealed the need for improvement of the existing baseline solutions in order to make them fully applicable as analytic tools in the considered retail stores and public spaces. They also revealed the need to combine signal modalities in order to filter and improve existing ReID in indor space. Therefore, perceptual features like

Figure 2 (a) succesful ReID, but with low fps rate due to dynamic background noise; (b) succesful ReID under low light conditions, but with identity loss after change of person's orientation at the end of sequence; (c) retail store application.

Figure 3 (a) crowded indoor space: successful ReID, but with low fps; (b)~false person detection, but with correct ReID; (c)~low light operation.

image depth [79], sound or other sensing devices [80] could bring complementary side information. As already mentioned, [6], [35], crowded scenes and cluttered object detection sometimes require alternative formulations of standard vision tasks.

On the side of possible applications, further analysis of the data collected by ReID systems could reveal a more complex behavior patterns [81], [82], i.e. go beyond simple statistics like time spent in different parts of a retail store. On the other hand, unsupervised representation learning [83] seems to be the key for described challenges and increasing the robustness of ReID. Nevertheless, existing person ReID solutions are already at the level that makes them applicable in the given context.

Acknowledgements
The authors would like to especially thank the conference organizers for their patience and support throughout the manuscript preparation process. This research was partially supported by the Science Fund of the Republic of Serbia, project" Multimodal multilingual human-machine speech communication - AI SPEAK", grant no. 7449; and by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (contract no. 451-03-137/2025-03/200156); and by the Faculty of Technical Sciences, University of Novi Sad, through the project "Scientific and artistic research work of researchers in teaching and associate positions at the Faculty of Technical Sciences, University of Novi Sad 2025" (no. 01-50/295).

6 REFERENCES

[1] Dollar, P.,Wojek, C.,Schiele, B. & Perona, P. (2011). "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern AnalysisandMachineIntelligence*,vol.34,no.4,pp.743–761.

[2] Gong, S., Cristani, M., Shuicheng, Y. & Change Loy, C. (2014). "Person re-identification,", ISBN978-1-4471-6295-7.

[3] Yurtsever, E.,Lambert, J., Carballo, A. &Takeda, K. (2020). "A survey ofautonomous driving: Common practices and emerging technologies,"*IEEE access*, vol. 8, pp. 58 443–58 469.

[4] Kwon, J., Lee, Y. &Lee, J. (2021). "Comparative study of markerless vision- based gait analyses for person Re-identification," *Sensors*, vol. 21, no. 24, p. 8208.

[5] Behera, N. K. S., Behera, T. K., Nappi, M. Bakshi, S.&Sa, P. K. (2021)."Futuristic person re-identification over internet of biometrics things(iobt): Technical potential versus practical reality," *Pattern Recognition Letters*,vol.151,pp.163–171.

[6] Chen, S., Fern, A. & Todorović, S. (2015)."Personcountlocalizationinvideosfrom noisy foreground and detections," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1364–1372.

[7] Gong, S., Loy, C. C.&Xiang, T. (2011). "Security and surveillance," in *Visual analysisofhumans:Lookingatpeople*.Springer,pp.455–472.

[8] Yang, H. F.,Cai, J., Liu, C., Ke, R. &Wang, Y. (2023)."Cooperative multi- camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning," *Transportation Research partC:EmergingTechnologies*,vol.148,p.103982.

[9] Almazan, J., Gajić, B., Murray, N. & Larlus, D. "Re-ID done right: Towards good practices for person re-identification," *arXiv preprintarXiv:1801.05339*.

[10] Shili, M., Sohaib, O. & Hammedi, S. (2024). "Yolo v5 and deep simpleonline and real-time tracking algorithms for real-time customer behaviortracking and retail optimization," *Algorithms*, vol. 17, no. 11, p. 525.

[11] Del Carpio, A. F. (2024). "Analyzing computer vision models for detecting customers: A practical experience in a Mexican retail," *International JournalofAdvancesinIntelligentInformatics*,vol.10,no.1,pp.131–174.

[12] Quintana, M., Menendez, J. M., Alvarez, F.,& Lopez, J. P. (2016). "Improv- ing retail efficiency through sensing technologies: A survey," *Pattem RecognitionLetters*,vol.81,pp.3–10.

[13] Liu, X., Jiang, Y., Jain, P. & Kim, K.-H. (2018). "TAR: Enabling fine-grainedtargeted advertising in retail stores," in *Proceedings of the 16th Annual InternationalConferenceonMobileSystems,Applications,andServi*

ces, pp.323–336.
[14] Ghose, A., Li, B., Li, R. & Xu, K. (2022). "Real-time purchase prediction using retail video analytics," in *ICIS Proceedings*, vol. 6, pp. 1–17. [Online]. Available: https://aisel.aisnet.org/icis2022/ data analytics/dataanalytics/6
[15] Martini, M., Paolanti, M. & Frontoni, E. (2020). "Open-world person re- identification with RGBD camera in top-view configuration for retailapplications," *IEEE Access*, vol. 8, pp. 67 756–67 765.
[16] Van Rooijen, A., Bouma, H., Pruim, R., Baan, J., *et al.* (2020). "Anonymizedperson re-identification in surveillance cameras," in *Counterterrorism, CrimeFighting,Forensics,andSurveillanceTechnologiesIV*,vol.115 42. SPIE, pp.63–67.
[17] K. Hashemifard, F. Florez-Revuelta, G. Lacey *et al.*, "A fallen person detector with a privacy-preserving edge-AI camera," pp. 262–269,2023.
[18] Pietrini, R. (2019). "Deep understanding of shopper behaviours and interactions in intelligent retail environment,", doctoral thesis, Universita` PolitecnicadelleMarche,Ancona,Italy.[Online].Available:http://hdl.handle.net/11566/274602
[19] Lee L., *et al.* (2015). "The emotional shopper: Assessing the effectiveness of retail therapy," *Foundations and Trends® in Marketing*, vol. 8, no. 2, pp. 69–145.
[20] Yamashiro, T., Honma, Y.,Togo, R., Ogawa, T. & Haseyama, M. (2023). "Customerinterestestimationmethodinrealstoreusingre-identification and 3D posture estimation models," in *International Workshop on Advanced Imaging Technology (IWAIT) 2023*, vol. 12592. SPIE, pp.42–47.
[21] Becattini, F., Becchi, G., Ferracani, A., Bimbo, A.D., Presti, L.L., Mazzola, G., *etal.* (2022)."I-mallaneffectiveframeworkforpersonalizedvisits: Improving the customer experience in stores," in *Proceedings of the 1st WorkshoponMultimediaComputingtowardsFashionRecommendation*, pp.11–19.
[22] Luxonis (2025). OAK-D lite. [Online]. Available: https://docs.luxonis.com/hardware/products/OAK-D%20 Lite (Accessed2025-01-11).
[23] Luxonis(2022). OAK lite datasheet. [Online]. Available: https://github.com/luxonis/depthai-hardware/tree/master/ DM9095 OAK-D-LITE DepthAI USB3C/Datasheet (Accessed2025-01-11).
[24] Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). "A survey on face detection in the wild: Past, present and future," *Computer Vision and Image Understanding*,vol.138,pp.1–24.
[25] Feng, Y. ,Yu, S.,Peng, H., Li, Y.-R. &Zhang, J. (2022). "Detect facesefficiently: A survey and evaluations," *IEEE Transactions on Biometrics, Behavior, andIdentityScience*,vol.4,no.1,pp.1–18.
[26] Zou, Z., Chen, K., Shi, Z., Guo, Y. &Ye, J. (2023)."Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276.
[27] Wang, X., Peng, J., Zhang, S., Chen, B., Wang, Y. & Guo, Y. (2022). "A survey offacerecognition,"*arXivpreprintarXiv:2212.13038*.
[28] Qian, Y. , Barthelemy, J., Karuppiah, E., & Perez, P. (2024). "Identifying re-identification challenges: Past, current and future trends," *SN Computer Science*,vol.5,no.7,p.937.
[29] Zahra, A., Perwaiz, N., Shahzad, M. & Fraz, M. M. (2023). "Person re-identification: A retrospective on domain specific open challenges andfuturetrends,"*PatternRecognition*,vol.142,p.109669.
[30] Yaghoubi, E., Kumar, A., & Proença, H. (2021), "SSS-PR:Ashortsurveyofsurveys in person re-identification," *Pattern Recognition Letters*, vol. 143, pp. 50–57.
[31] Ming, Z.,Zhu, M., Wang, X., Zhu, J., Cheng, J.,Gao, C., Yang, Y. & Wei, X. (2022). "Deeplearning-basedpersonre-identificationmethods:Asurveyand outlook of recent works," *Image and Vision Computing*, vol. 119, p. 104394.
[32] Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y.& Tian, Q. (2017). "Person re-identification in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1367–1376.
[33] Xiao, T., Li, S., Wang, B., Lin, L. &Wang, X. (2017)."Joint detection and identificationfeaturelearningforpersonsearch,"in*Proceedingsofthe IEEE Conference on Computer Vision and Pattern Recognition, 2017*,pp. 3415–3424.
[34] Alexe, B. (2013). "Objectness and its use for visual learning," doctoral thesis, ETH Zurich, Switzerland. [Online]. Available:https://doi.org/10.3929/ethz-a-009906832
[35] Cai, Y., Wen, L., Zhang, L., Du, D. &Wang, W. (2021). "Rethinking object detection in retail stores," in *Proceedings of the AAAI Conference on ArtificialIntelligence*,vol.35,no.2,pp.947–954.
[36] Yadav A. & Vishwakarma, D. K. (2024). "Deep learning algorithms for person re-identification: State-of-the-art and research challenges," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 22 005–22 054.
[37] Lin, X., Li, J., Ma, Z., Li, H., Li, S., Xu, K., Lu, G.&Zhang, D. (2022). "Learning modal-invariant and temporal-memory for video-based visible-infrared person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20 973–20982.
[38] Li, Q., Zhang, C., Hu, Q., Fu, H. & Zhu, P. (2022). "Confidence-aware fusion using Dempster-Shafer theory for multispectral pedestrian detection," *IEEETransactionsonMultimedia*,vol.25,pp.3420–3431.
[39] Huang, N., Liu, J., Miao, Y., Zhang, Q., & Han, J. (2023). "Deep learning forvisible-infrared cross-modality person re-identification: A comprehen-sivereview,"*InformationFusion*,vol.91,pp.396–411.
[40] Zou, G., Fu, G., Peng, X., Liu, Y. , Gao, M., & Liu, Z. (2021). "Person re- identification based on metric learning: A survey," *multimedia tools andapplications*,vol.80,no.17,pp.26855–26888.
[41] Wojke, N. & Bewley, A. (2018)."Deep cosine metric learning for person re-identification," in *IEEE Winter Conference on Applications ofComputer Vision (WACV)*. IEEE, pp.748–756.
[42] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L. & Hoi, S. C. (2021). "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2872– 2893.
[43] Deng, J., Xu, D., Li, W. & Duan, L. (2023). "Harmonious teacher for cross- domain object detection," in *Proceedings of the IEEE/CVF conferenceoncomputervisionandpatternrecognition*,pp.23829–23838.
[44] Liu, C.-T., Chen, J.-C., Chen, C.-S. & Chien, S.-Y. (2021). "Video-basedperson re-identification without bells and whistles," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.1491–1500.
[45] Zang, X., Li, G., & Gao, W. (2022). "Multidirection and multiscale pyramid intransformer for video-based pedestrian retrieval," *IEEE Transactions onIndustrialInformatics*,vol.18,no.12,pp.8776–8785.
[46] Wu, L., Wang, Y., Gao, J. & Li, X. (2018). "Where-and-when to look: Deep siamese attention networks for video-based person re-identification,"*IEEETransactionsonMultimedia*,vol.21,no.6,pp.1412–1424.
[47] Jiao, J., Zheng, W.-S., Wu, A., Zhu, X., & Gong, S. (2018). "Deep low-resolution person re-identification," in *Proceedings of the AAAI Conference on ArtificialIntelligence*,vol.32,no.1.
[48] Zhang, L., Xu, Y., Zhao, L. & Qin, F. (2022). "Resolution independent personre-identification network," *IET Computer Vision*.
[49] Xu, P. & Zhu, X. (2023). "Deepchange: A long-term person re-identification benchmark with clothes change," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.

11 196–11205.
[50] Zhang, G., Luo, Z., Chen, Y., Zheng, Y.& Lin, W. (2022).“Illuminationunification for person re-identification,” *IEEE Transactions on CircuitsandSystemsfor VideoTechnology*,vol.32,no.10,pp.6766–6777.
[51] Liu, F., Kim, M., Gu, Z., Jain, A. & Liu, X. (2023). “Learning clothing and pose invariant 3d shape representation for long-term personre-identification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 19 617–19626.
[52] Sepas-Moghaddam, A. & Etemad, A. (2022). “Deep gait recognition: A sur-vey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.45,no.1,pp.264–284.
[53] Li, W., Hou, S., Zhang, C., Cao, C., Liu, X., Huang, Y. &Zhao, Y. (2023). “An in-depth exploration of person re-identification and gait recognition incloth-changing conditions,” in *Proceedings of the IEEE/CVFConferenceon Computer Vision and Pattern Recognition*, pp. 13 824–13833.
[54] Tan, L., Dai, P., Ji, R. &Wu, Y. (2022).“Dynamic prototype mask for occluded person re-identification,” in *Proceedings of the 30th ACM internationalconferenceonmultimedia*,pp.531–540.
[55] Dong, N., Yan, S., Tang, H., Tang, J. & Zhang, L. (2024). “Multi-view informa- tion integration and propagation for occluded person re-identification,”*InformationFusion*,vol.104,p.102201.
[56] Tan, L., Xia, J., Liu, W., Dai, P., Wu, Y. &Cao, L. (2024). “Occluded person re-identification via saliency-guided patch transfer,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 5, 2024, pp. 5070–5078.
[57] Ning, E., Wang, Y., Wang, C., Zhang, H. & Ning, X. (2024). “Enhancement, integration, expansion: Activating representation of detailed features foroccluded person re-identification,” *Neural Networks*, vol. 169, pp. 532–541,2024.
[58] Ding, G., Khan, S., Tang, Z. & Porikli, F. (2020). “Feature mask network for person re-identification,” *Pattern Recognition Letters*, vol. 137, pp. 91–98,2020,learningandRecognitionforAssistiveComputerVision.
[59] Agarwal, S., Awan, A. & Roth, D. (2004). “Learning to detect objects inimagesvia a sparse, part-based representation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 11, pp. 1475–1490.
[60] Wei, X.-S., Song, Y.-Z., Mac Aodha, O., Wu, J., Peng, Y. , Tang, J., Yang, J. &Belongie, S. (2021). “Fine-grained image analysis with deep learning: Asurvey,” *IEEE Transactions on Pattern Analysis and Machine Intelli- gence*,vol.44,no.12,pp.8927–8948.
[61] Somers, V., De Vleeschouwer, C. & Alahi, A. (2023). “Body part-based repre- sentation learning for occluded person re-identification,” in*Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 1613–1623.
[62] Nguyen, B. X., Nguyen, B. D., Do, T., Tjiputra, E., Tran, Q. D. &Nguyen, A. (2021). “Graph-basedpersonsignatureforpersonre-identifications,” in *Proceedings of the IEEE/CVF conference on computer vision andpattern recognition*, pp.3492–3501.
[63] Zhai, Y., Guo, X., Lu, Y. &Li, H. (2019). “In defense of the classificationloss for person re-identification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp.1526–1535.
[64] Ding, S., Lin, L., Wang, G. & Chao, H. (2015). “Deep feature learning with relative distance comparison for person re-identification,” *Pattern Recognition*,vol.48,no.10,pp.2993–3003.
[65] Hermans, A., Beyer, L. & Leibe, B. (2017). “In defense of the triplet loss for person re-identification,” *arXiv preprint arXiv:1703.07737*.
[66] Ijjina, E. P., Medipelly, R., Beerukuri, S. K., Vinnakota, S. & Nelakurthi, V. C. (2024). “Person re-identification using vision transformer and cen-troid triplet loss,” *Multimedia Tools and Applications*, vol. 83, no. 29, pp. 73 777–73 788.
[67] Intel. (2025) OpenVINO™ Toolkit (Open Visual Inference and Neural Network Optimization). [Online]. Available: https://docs.openvino.ai/ 2025/index.html (Accessed2025-01-11).
[68] ONNX (2025). Open Neural Network Exchange [Online]. Available:https://onnx.ai (Accessed2025-01-11).
[69] Luxonis, DepthAI. (2025a). Python API utilities, examples, and tutorials. [Online]. Available: https://github.com/luxonis/depthai (Accessed 2025-01-11).
[70] Bradski, G., *et al.* (2025). OpenCV: Open Computer Vision library. [Online]. Available: https://opencv.org/ (Accessed2025-01-11).
[71] Luxonis,DepthAI. (2025b).C++library.[Online].Available:https://github.com/luxonis/depthai-core (Accessed 2025-01-11).
[72] OpenVINO. (2022a) Person detection model. [Online]. Available: https://github.com/openvinotoolkit/open model zoo/tree/master/models/intel/person-detection-retail-0013 (Accessed2025-01-11).
[73] OpenVINO.(2022b). Person ReID model. [Online]. Available:https://github.com/openvinotoolkit/open model zoo/tree/master/models/intel/person-reidentification-retail-0288 (Accessed2025-01-11).
[74] Luxonis. (2023). Depthai. experiment: Pedestrian reidentification. [Online]. Available: https://github.com/luxonis/depthai-experiments/tree/master/gen2-pedestrian-reidentification#pedestrian-reidentification (Accessed 2025-01-11).
[75] He, L., Liao, X., Liu, W., Liu, X., Cheng, P. &Mei, T. (2023). “Fastreid: A Pytorch toolbox for general instance re-identification,” in *Proceedings of the 31st ACM International Conference on Multimedia*, pp.9664–9667.
[76] Yıldız, S. & Kasım, A. N. (2024). “ENTIRe-ID: An extensive and diverse dataset for person re-identification,” in *18th International Conference on Automatic Face and Gesture Recognition*. IEEE, pp.1–5.
[77] Zheng, L.. Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S. & Tian, Q. (2016). “MARS: A video benchmark for large-scale person re-identification,” in *14th European Conference on Computer Vision, Proceedings, Part VI 14*. Springer, pp.868–884.
[78] Cui, Z., Zhou, J., Wang, X., Zhu, M. & Peng, Y. (2024). “Learning continual compatible representation for re-indexing free lifelong person re-identification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.16 614–16 623.
[79] Luna, C.A., Losada-Gutiérrez, C., Fuentes-Jimenez,D., &Mazo, M. (2021).“People re-identification using depth and intensity information froman overhead camera,” *Expert Systems with Applications*, vol. 182, p.115287.
[80] S.LiandR.Hishiyama,“Anindoorpeoplecountingandtracking systemusing mmWave sensor and sub-sensors,” *IFAC-PapersOnLine*, vol. 56,no. 2, pp. 7096–7101.
[81] Chophuk, P., Boonmee, P., Jiarasuwan, S., Jearasuwan, S.& Bookprakong, P. (2023). “Theft detection by patterns of walking behavior using motion-based artificial intelligence,” in *International WorkshoponAdvanced Imaging Technology*, vol. 12592. SPIE, pp.140–145.
[82] Abbattista, G., Chimienti, M., Dentamaro, V., Giglio, P. Impedovo, D., Pirlo, G. & Rosato, G. (2023). “A biometric-based system for unsupervised anomalybehaviourdetectionatthepawnshop,”*Cyber-PhysicalSystems*, vol.9,no.4,pp.338–356.
[83] Chen, H. C, Wang, Y., Lagadec, B., Dantcheva, A. & Bremond, F.(2022). “Learninginvariancefromgeneratedvarianceforunsupervisedpersonre-identification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 7494–7508.

Contact information:

Branko BRKLJAČ, Ph.D., associate professor
(Corresponding author)

Institution: Faculty of Technical Sciences, University of Novi Sad,
Dept. of Power, Electronic and Telecommunication Engineering
Postal address: Trg Dositeja Obradovića 6, 21000 Novi Sad, Republic of Serbia
M-Mail: brkljacb@uns.ac.rs
https://orcid.org/0000-0001-7932-6676

Milan BRKLJAČ, Ph.D., assistant professor
Institution: Faculty of Finance, Banking and Auditing, Alfa BK Univeristy
Postal address: Bulevar maršala Tolbuhina 8, 11070 Novi Beograd,
Republic of Serbia
E-Mail: milan.brkljac@alfa.edu.rs
https://orcid.org/0000-0002-0617-973X