

Fusion in Dissimilarity Space Between RGB-D and Skeleton for Person Re-Identification



Md Kamal Uddin, Amran Bhuiyan, Mahmudul Hasan

Abstract: Person re-identification (Re-id) is one of the important tools of video surveillance systems, which aims to recognize an individual across the multiple disjoint sensors of a camera network. Despite the recent advances on RGB camera-based person re-identification methods under normal lighting conditions, Re-id researchers fail to take advantages of modern RGB-D sensor-based additional information (e.g. depth and skeleton information). When traditional RGB-based cameras fail to capture the video under poor illumination conditions, RGB-D sensor-based additional information can be advantageous to tackle these constraints. This work takes depth images and skeleton joint points as additional information along with RGB appearance cues and proposes a person re-identification method. We combine 4-channel RGB-D image features with skeleton information using score-level fusion strategy in dissimilarity space to increase re-identification accuracy. Moreover, our propose method overcomes the illumination problem because we use illumination invariant depth image and skeleton information. We carried out rigorous experiments on two publicly available RGBD-ID re-identification datasets and proved the use of combined features of 4-channel RGB-D images and skeleton information boost up the rank 1 recognition accuracy.

Keywords: Re-identification, RGB-D features, Skeleton information, Video surveillance.

I. INTRODUCTION

Person re-identification is the task of recognizing an individual across the distributed camera views. Retrieval of a specific person from a large number of camera views has broad applications including multi-camera tracking, access control, public security event management, and sports analytics. Person re-identification is a very challenging task because same individual appears in different views on different cameras. However, sometimes RGB appearance-based camera cannot capture the video of an individual under extremely low lighting conditions. Moreover, illumination conditions fluctuate frequently from one camera location to another, also different times of the day. As a result, when same individual passes through one

camera to another, it is much more difficult to recognize them in the camera network due to the aforementioned constraints (see in Fig. 1).



Fig. 1. Example image sequences illustrate (a) Viewpoint variations and (b) illumination variations where sample images are taken from MSMT17 [12] and IAS-Lab RGBD-ID [8].

In recent years, some Re-id researchers have been proposed several methods to address the problems mentioned above. Most of the existing approaches use RGB appearance cues captured by traditional RGB cameras [1-4]. After invention of modern RGB-D sensors (e.g. Intel RealSense Depth Camera and Microsoft Kinect), Re-id researchers proposed some methods to overcome illumination problem [5-10]. Those approaches also gained high recognition accuracy using multi-modal features because RGB-D sensors simultaneously provide RGB, depth and skeleton information [11] (see in Fig. 2). Although existing approaches use multi-modal features, such as RGB, depth and skeleton joint points as additional information, most of them cannot properly utilize them. In [6], We et al. use two modalities: depth shape descriptors and skeleton information. Imani et al. [10] consider local shape descriptor and anthropometric measure to tackle the illumination problem. Both methods use score-level fusion techniques to combine depth and skeleton features. In some Re-id methods, depth images are converted to point cloud, and each point cloud frame is warped up with skeleton standard poses [7, 8]. In [9], Imani et al. extract histogram oriented local vector pattern features from both RGB and depth images, and skeleton features are estimated by calculating Euclidean distances among skeleton joint points. Then, extracted features from different modalities are combined with score-level fusion. Igor Barbosa et al measures two features: Skeleton based features and Surface-based features from depth data to construct soft biometric cues which are illumination invariant descriptors. However, performance of the above-mentioned methods are not promising because they haven't utilized all the modalities properly to construct robust features, which help to increase the re-identification performance.

Manuscript received on October 16, 2021.

Revised Manuscript received on October 27, 2021.

Manuscript published on October 30, 2021.

* Correspondence Author

Md Kamal Uddin*, Department. of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh. Email: kamal.uddin.nstu@gmail.com

Amran Bhuiyan, Department. of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh. Email: amran.apece@gmail.com

Mahmudul Hasan, Department. of Computer Science and Engineering, Comilla University, Comilla, Bangladesh. Email: mhasanraju@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

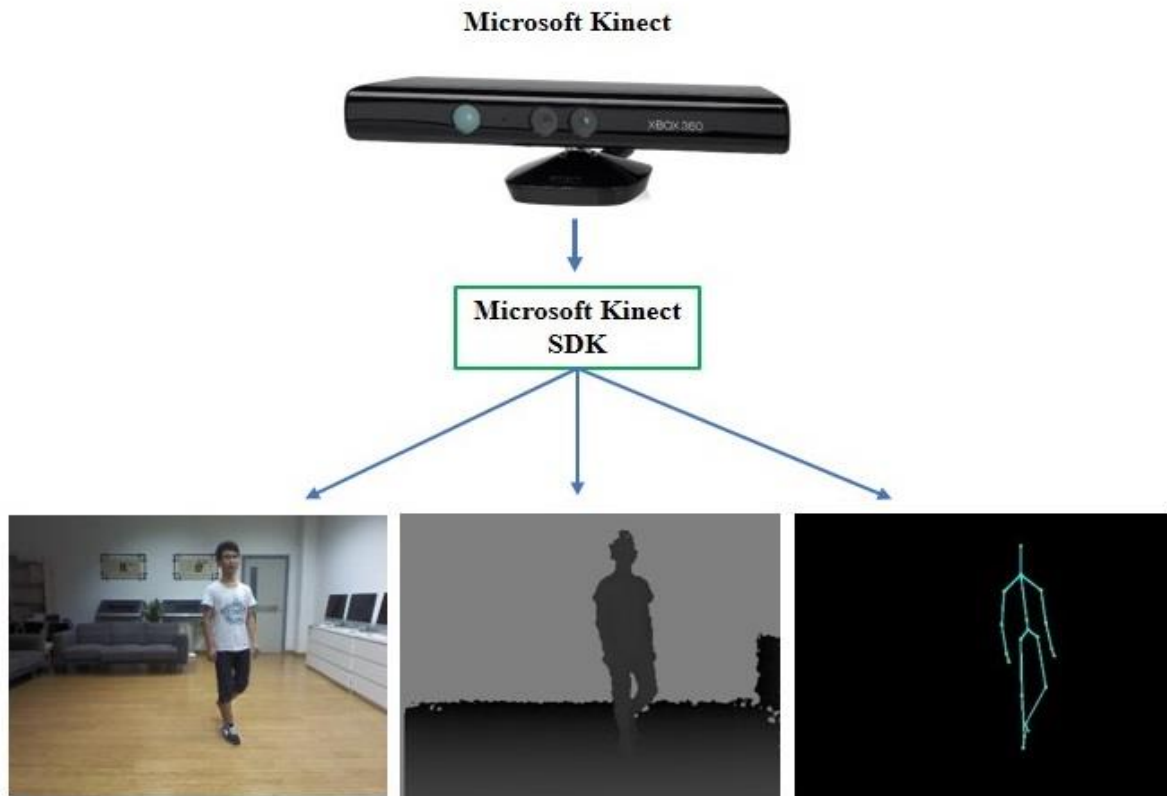


Fig. 2. RGB, Depth and Skeleton generation using Microsoft Kinect SDK

This work constructs 4-channel RGB-D image by allocating an extra channel for a single-channel depth image. Using 4-channel RGB-D images, we extract robust features using a deep CNN network. Beside this features, we also consider another anthropometric features, skeleton information of an individual, particularly which are illumination and color invariant. Then, we estimate anthropometric geometric characteristics from skeleton joints points by measuring Euclidean distance among the different body joint points. 4-channel RGB-D based CNN features and skeleton features are used to calculate the score separately between probe and gallery set. Finally, both scores are fused in dissimilarity spaces using score-level fusion strategy.

Our contributions are summarized as follows:

1. We propose a person re-identification method for fusing 4-channel RGB-D and skeleton information in dissimilarity space to gain high re-identification accuracy.
2. We performed extensive experiments on two publicly available dataset and gained high recognition accuracy, proving that our method is effective.

We organized the remaining section of our paper as follows. In section 2, we describe related works that are more relevant to our work, especially emphasize the RGB-D sensor-based person re-identification approaches. Our proposed approach is explained in section 3. Experimental analysis and dataset description are discussed in section 4. Finally, we summarized our work in section 5.

II. RELATED WORKS

This section reviews the background of RGB, depth and skeleton-based person re-identification methods that are most relevant to our work. RGB-D sensors can provide different

modalities simultaneously, which can be employed to recognize an individual under diverse illumination conditions. However, when an individual passes through distributed camera network, he may change his clothes. This problem can be addressed by using illumination and color invariant depth and skeleton modalities. Based on the different modalities, existing Re-id works can be divided into different categories: (i) RGB and Depth based Re-id (ii) RGB and Skeleton based Re-id (iii) Depth and Skeleton based Re-id and (iv) RGB, Depth and Skeleton based Re-id.

Some state-of-the-art deep learning methods have been proposed [13-19] based on the RGB and depth modalities. Ren et al. [13, 14] introduced a multi-modal uniform and variational deep learning method for person re-identification, where [13] define a network to combine features extracted from processed depth images and RGB images to a fusion layer. In Ref. [14], authors design an auto-encoder which finds a latent variable into a common subspace. Auto-encoder retains complete information extracted from RGB and depth images, and create a relationship between RGB and depth features. A multimodal neural network for attention model and overhead was proposed in [15, 16]. In [15], Lejbolle et al. developed a CNN network which takes RGB and depth images as input and finally provides fused features. These combined features are used to recognize an individual from image sequences. In [16], authors proposed an attention network to extract local features from RGB and depth modalities and finally fused these features at globally.

Uddin et al. [17] introduced a depth guided attention model for person re-identification where they extract foreground part of an image using a mask, which removes the background part. Removing the background region from the image helps to extract robust features and increase the recognition accuracy. A triplet-based deep learning method for retail application was developed in [18], where overhead RGB-D based video data was collected for Re-id and background part was removed to separate an individual from whole scene. In [19], authors calculated the scores separately which were taken from 3-channel and 4-channel image-based Re-id, and finally fused both score in dissimilarity space to get final score for recognition of an individual. Uddin et al. [20] utilized local shape information of RGB and depth modalities and proposed a cross-modal person re-identification method.

Apart from the above works, some state-of-the-art methods utilize the RGB image and skeleton information for Re-id [21-24]. In Ref. [21], authors fuse RGB appearance features with anthropometric features extracted from depth domain. Liu et al. [22] proposed a feature funnel model for person Re-id which fully uses the RGB appearance features and skeleton information. A point cloud based Re-id framework was developed in [23], where a skeleton standard posture was defined to create a partitioned grid of the point cloud on different positions of an individual. In [24], authors proposed a top view configuration for person Re-id method to reduce the occlusion problem. Some re-identification approaches tried to address the clothing changes problem which can occur in long-term monitoring systems.

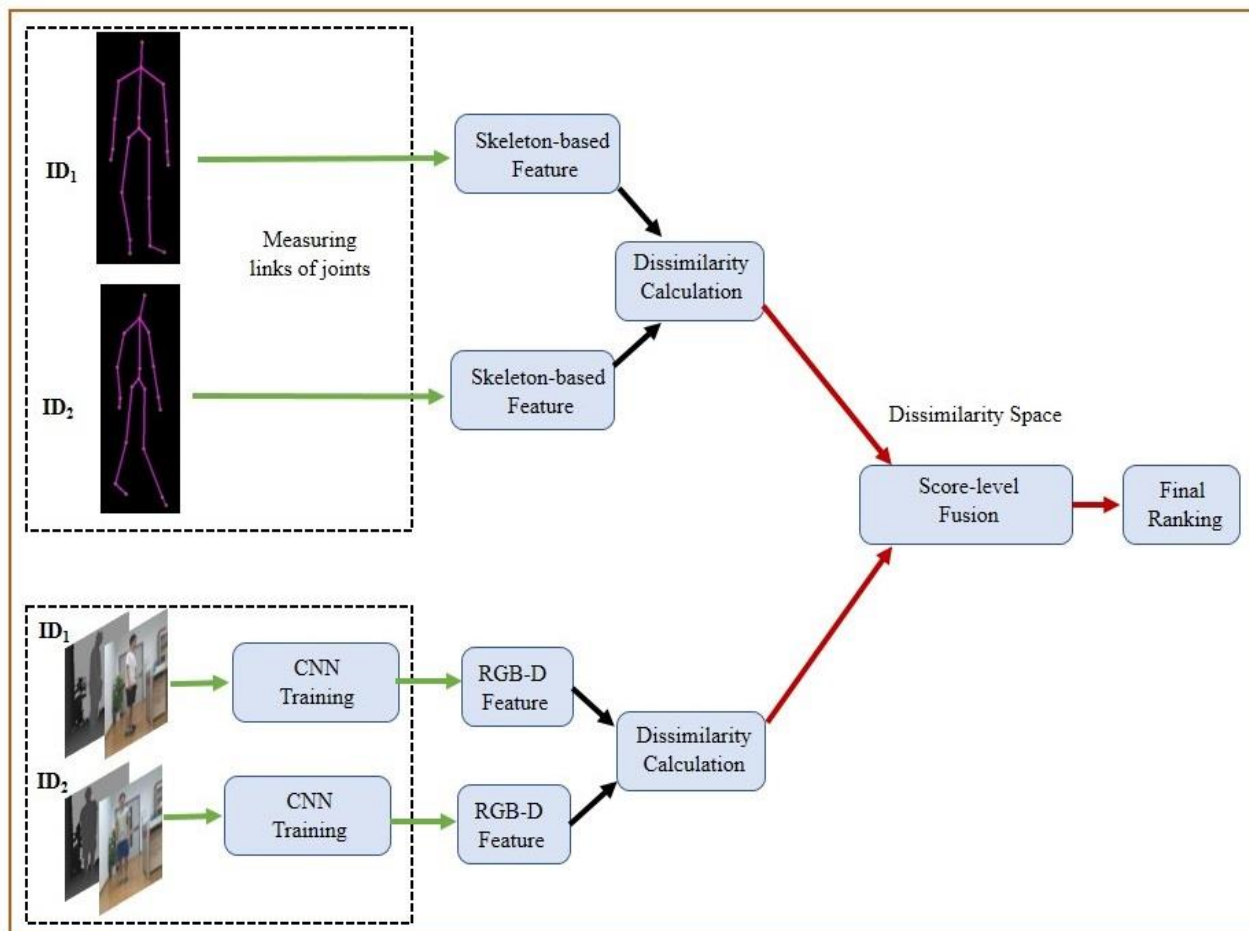


Fig. 3. Our proposed Re-id framework. This framework has three parts: 1) Skeleton information is taken as input and calculate the skeleton-based features in the first part. 2) In the second part, 4-channel RGB-D images are fed into the CNN model to train the whole RGB-D image-based network and 3) Dissimilarity scores from both parts are calculated among the individuals, and then finally fused both scores to get final ranking.

In some situation, when some cameras are installed in deplorable lighting conditions, in this scenario, color information is not reliable at all because traditional RGB-based camera fails to capture color cues. To address these obstacles, some state-of-the art approaches used illumination and color invariant depth and skeleton information [5-10]. Imani et al. [25] proposed a short term person re-identification method. They extract local vector pattern both from RGB and depth modalities. Skeleton features are extracted from different geometric measurement

by using body joint points. Finally, the result of the triplet combination is calculated using score-level fusion.

In contrast to the above existing methods, we proposed a 4-channel image-based and anthropometric feature-based person re-identification method, which effectively addresses the problems mentioned above and gains high recognition rates.

III. PROPOSED APPROACH

In this section, we describe our proposed skeleton and 4-channel RGB-D image-based person re-identification method. The overall framework of our proposed approach is illustrated in Fig. 3. Our proposed framework is divided into three parts: a) Geometric feature extraction from skeleton joint points, and then dissimilarity score calculation between the query and gallery images on the same testing dataset. b) 4-channel RGB-D feature extraction from trained network, and then calculate the dissimilarity score on the testing set and b) finally, fuse both scores in dissimilarity space using score-level fusion technique.

A. Skeleton-based Geometric Feature Calculation

3D location of body joint points are tracked by Microsoft SDK tracker [7]. The length of different joint points and the ratio between them carries physical information of a person. Therefore, we can easily distinguish the individuals. Though it is difficult to calculate the length of joint points when a pedestrian appears in side-view, but we can easily estimate the Euclidean distances between limb joint points and their ratios when appears in frontal view. We calculate the skeleton-based feature vector, which contains 13 distance values and ratios, where ratios are calculated from the positions of joint points (see in Fig. 4).

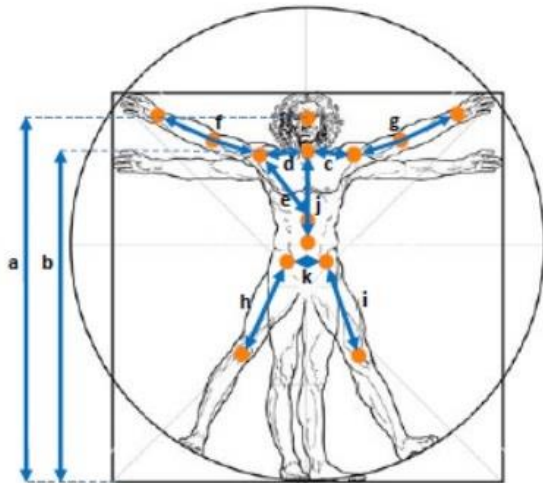


Fig. 4. Illustration of body joint points tracked by Kinect SDK [7].

The feature vector contains the following length and ratio information: 1) the height of head (a), 2) the height of neck (b), 3) distance between neck to left shoulder (c), 4) neck to right shoulder (d), 5) right shoulder to torso (e), 6) the length of left arm (g), 7) the length of right arm (f), 8) the length of left upper leg (i), 9) the length of right upper leg (h), 10) the length of torso (j), 11) distance between right hip and left hip (k), 12) the ratio between torso and right upper leg (j/h), 13) the ratio between torso and left upper leg (j/i). All distances are measured in cm. The feature vector is defined as

$$X = [d_1, d_2, d_3, \dots, d_{13}]$$

B. RGB-D Image-based Model Training

We follow the same strategy for RGB-D model training in [19]. In this method, 4-channel adapting CNN was proposed for RGB-D images. To train this model triplet loss function is used because triplet loss function has the ability to pull

distance of feature vectors (i.e. F_i^a and F_i^p) for same person (i.e. anchor image 'a' and positive image 'p') and push the distance for different persons (i.e. anchor 'a' and negative image 'n'). Triplet loss function is defined as

$$L_{trp} = \max\left\{0, \|F_i^a - F_i^p\|_2^2 - \|F_i^a - F_i^n\|_2^2 + m\right\} \quad (1)$$

Where m is predefined margin for regularizing the distance and $\|\cdot\|_2^2$ is square Euclidean distance. In our training procedure, particularly we follow the batch-hard triplet generation strategy [35] for choosing right combination of triplet inputs.

C. Fusion Rule

The scores obtained from skeleton-based Re-id and RGB-D image-based Re-id are fused using score level-fusion strategy using summing rule. Most the existing methods for multi-modal fusion follow this rule which is denoted as

$$Dist_{Fusion}(q, G) = Dist_{rgb-d}(q, G) + \omega Dist_{skl}(q, G) \quad (2)$$

where $Dist_{rgb-d}(q, G)$ and $Dist_{skl}(q, G)$ are dissimilarity scores which are estimated from skeleton-based features and RGB-D based features respectively between query (q) and gallery (G) set, and where $Dist_{fusion}(q, G)$ is the final score. And ω is fusion weight.

IV. EXPERIMENTS

In this section, we describe our experimental evaluations. We evaluated our proposed re-identification framework on two publicly available datasets RGBD-ID [5] and RobotPKU [22], which were collected by Microsoft Kinect camera.

A. Datasets

There are some publicly available datasets, such as IAS-Lab RGBD-ID [8], KinectREID [21] and BIWI RGBD-ID [7]. Still, we choose RGBD-ID and RobotPKU datasets for our experiments because of the size of the dataset. As we train our CNN model, we need a large dataset containing decent number of people to train a good model. In this case, RGBD-ID and RobotPKU datasets meet our experimental requirements. RobotPKU dataset was captured by Microsoft Kinect camera in two disjoint locations, has two groups, still and walking. Each group contains 180 video sequences of 90 people. The video frames that cannot capture depth image and skeleton information properly, were discarded from our experiments. RGBD-ID dataset has four groups such as walking1, walking2, collaborative and backwards. Each group has RGB, 3D point cloud of depth frame, and skeleton information of 79 people. In this dataset, some people wore different dresses in different acquisitions.

B. Evaluation Protocol

We evaluated our result using rank-k accuracy, and Cumulative Matching Characteristic (CMC) [26] curve as re-identification literature follow this evaluation rule [3, 6, 27-29, 34]. Rank-k accuracy is measured with cumulative recognition of the correct match of an individual at k-th position in the rank. We repeat our experiments 10 times, and the table reports the average results of rank 1, 5 and 10 accuracies and mean average precision (mAP).



C. Implementation Details

When we trained our model with 4-channel RGB-D image, we set some initial parameter values, including momentum (0.9), initial learning rate (0.01), weight decay (5×10^{-4}) and margin value $m=0.3$ in (1). We follow the data augmentation strategy and horizontal flip technique when train our network [31]. We resized all the images to 256×192 before feeding to the network and set the batch size $20 \times 4=80$. According to the batch-hard triplet loss strategy, every iteration takes images from 20 different persons with 4 instances per person to form a mini-batch. In all our experiments, we randomly select half of the people for training and half of the people for testing from each dataset. We set fusion weight $\omega=1$ in (2). We implement our model on Pytorch platform [30].

D. Experimental Evaluations

We performed all our experiments on the RobotPKU and RGBD-ID datasets. First, we compare our Re-id framework with baseline method, RGB-D feature-based Re-id. Then we compare our proposed approach with existing state-of-the-art re-identification methods.

Table-I. Comparison of our score-level fusion method with baseline method on RGBD-ID dataset.

| Methods | RGBD-ID | | | |
|---|--------------|--------------|--------------|--------------|
| | rank 1 | rank 5 | rank 10 | mAP |
| RGB-D CNN feature-based Re-id | 78.65 | 83.33 | 86.86 | 68.32 |
| Our proposed approach (RGB-D + Skeleton) | 80.67 | 86.45 | 90.20 | 73.48 |

Here we consider RGB-D CNN feature-based re-identification as our baseline method. We compare this baseline method with our proposed score-level fusion of RGB-D and skeleton based Re-id. Table I and II report the comparison results for RGBD-ID and RobotPKU datasets, respectively.

Table-II. Comparison of our proposed RGB-D and Skeleton based score-level fusion method with the baseline method on RobotPKU dataset.

| Methods | RobotPKU | | | |
|---|--------------|--------------|--------------|--------------|
| | rank 1 | rank 5 | rank 10 | mAP |
| RGB-D CNN feature-based Re-id | 89.63 | 93.45 | 95.80 | 84.27 |
| Our proposed approach (RGB-D + Skeleton) | 92.95 | 96.45 | 97.39 | 88.45 |

Table I and II clearly show that our proposed approach outperforms the rank 1, 5 and 10 results, and mAP for RGBD-ID and RobotPKU dataset, respectively. Table I reports that Rank 1 accuracy and mAP of our proposed method for RGBD-ID dataset is 80.67% and 73.48%, and 78.65% and 68.32% for our baseline method, so total gain 2.02% and 5.16%, respectively. Table II shows that the mAP and rank 1 accuracy are 88.45% and 92.95% for our proposed approach, and 84.27% and 89.63% for baseline method, with

gain 4.18% and 3.32%, respectively. However, total gain for mAP is high compared to rank 1 accuracy for both datasets.

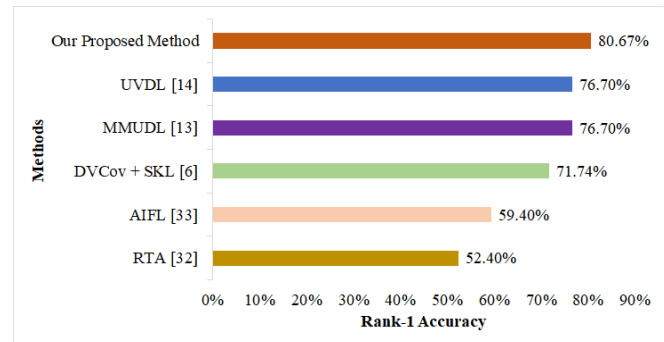


Fig. 5. Comparison results of our Re-id approach with state-of-the-art methods for RGBD-ID dataset.

Comparison with Existing Methods. This experimental analysis aims to compare our method with existing state-of-the-art methods. We compare our Re-id framework with state-of-the-art methods for both experimental datasets. Fig. 5 and 6 show the comparison result with existing methods for RGBD-ID and RobotPKU datasets, respectively.

Here, rank 1 accuracy is considered for comparison because rank 1 accuracy has most significant impact on re-identification system.

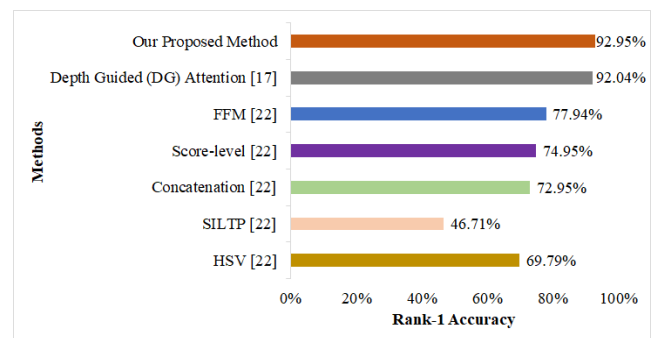


Fig. 6. Comparison results of our method with existing re-identification approaches for RobotPKU dataset.

The result from Fig. 5, we see that our proposed re-identification method outperforms the state-of-the-art methods. Our method achieves 80.67% rank 1 accuracy for RGBD-ID dataset. However, Fig. 6 shows that the rank 1 accuracy for Robot PKU dataset is 92.95%, which is the highest of all existing methods. From the result, it can be said that our score-level fusion of RGB-D and skeleton based Re-id method achieves superior result compared to state-of-the-art approaches.

V. CONCLUSION

This paper presents an RGB-D and skeleton-based person re-identification method to address the some constraints, such as illumination and clothing changes problem. To construct robust features, we considered 3-channel RGB and single-channel depth images to form a 4-channel RGB-D image.

From 4-channel RGB-D images, we extract CNN features. These features are used to calculate the recognition score. On the other hand, color invariant skeleton-based anthropometric features are extracted from skeleton information by calculating Euclidean distance from body joint points. Skeleton features are used to calculate the score, and then finally fused both scores in dissimilarity space to gain final ranking for person re-identification. Our proposed method achieves superior result using multi-modal features.

REFERENCES

1. S. Ding, L. Lin, G. Wang and H. Chao, "Deep feature learning with relative distance comparison for person re-identification", *Pattern Recognition*, 48(10), pp.2993-3003, 2015.
2. D.S. Cheng, M. Cristani, M. Stoppa, L. Bazzani and V. Murino, "Custom pictorial structures for re-identification", In *Proceedings of Machine Vision Conference (BMVC)*, 2011.
3. D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features", In *European conference on computer vision (ECCV)*, pp. 262-275, Springer, Berlin, Heidelberg, 2008.
4. M. Hirzer, C. Beleznai, P.M. Roth and H. Bischof, "Person re-identification by descriptive and discriminative classification", In *Scandinavian conference on Image analysis*, pp. 91-102, Springer, Berlin, Heidelberg, 2011.
5. I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani and V. Murino, "Re-identification with rgb-d sensors", In *European conference on computer vision*, Springer, Berlin, Heidelberg, pp. 433-442, 2012.
6. A. Wu, W.S. Zheng and J.H. Lai, "Robust depth-based person re-identification", *IEEE Transactions on Image Processing*, 26(6), pp.2588-2603, 2017.
7. M. Munaro, A. Fossati, A. Basso, E. Menegatti and L. Van Gool, "One-shot person re-identification with a consumer depth camera", In *S. Gong, M. Cristani, S. Yan, C. Loy (eds) Person Re-Identification. Advances in Computer Vision and Pattern Recognition*, Springer, London, pp. 161-181, 2014.
8. M. Munaro, A. Basso, A. Fossati, L. Van Gool and E. Menegatti, "3D reconstruction of freely moving persons for re-identification with a depth sensor", In *IEEE international conference on robotics and automation*, pp. 4512-4519, 2014.
9. Z. Imani and H. Soltanizadeh, "Histogram of the node strength and histogram of the edge weight: two new features for RGB-D person re-identification", *Science China Information Sciences*, 61(9), p.092108, 2018.
10. Z. Imani, H. Soltanizadeh, Person reidentification using local pattern descriptors and anthropometric measures from videos of kinect sensor, *IEEE Sensors Journal*, 16(16), pp.6227-6238, 2016.
11. R. Satta, F. Pala, G. Fumera and F. Roli, "Real-time Appearance-based Person Re-identification Over Multiple KinectTM Cameras", In *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)*, (2), pp. 407-410, 2013.
12. L. Wei, S. Zhang, W. Gao and Q. Tian, "Person transfer gan to bridge domain gap for person re-identification", In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 79-88, 2018.
13. L. Ren, J. Lu, J. Feng and J. Zhou, "Multi-modal uniform deep learning for RGB-D person re-identification", *Pattern Recognition*, 72, pp. 446-457, 2017.
14. L. Ren, J. Lu, J. Feng and J. Zhou, "Uniform and variational deep learning for RGB-D object recognition and person re-identification", *IEEE Transactions on Image Processing*, 28(10), pp.4970-4983, 2019.
15. A.R. Lejbolle, K. Nasrollahi, B. Krogh, T.B. Moeslund, Multimodal neural network for overhead person re-identification, in: *International Conference of the Biometrics Special Interest Group (BIOSIG)*, Darmstadt, pp.1-5, 2017.
16. A.R. Lejbolle, B. Krogh, K. Nasrollahi, T.B. Moeslund, Attention in multimodal neural networks for person re-identification, in: *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 179-187, 2018.
17. M. K. Uddin, A. Lam, H. Fukuda, Y. Kobayashi and Y. Kuno, "Depth Guided Attention for Person Re-identification", In *DS. Huang, P. Premaratne (eds) Intelligent Computing Methodologies, ICIC*, Springer, vol. 12465, pp.110-120, 2020.
18. M. Martini, M. Paolanti and E. Frontoni, "Open-World Person Re-Identification with RGBD Camera in Top-View Configuration for Retail Applications", *IEEE Access*, 8, pp. 67756-67765, 2020.
19. M. K. Uddin, A. Lam, H. Fukuda, Y. Kobayashi and Y. Kuno, "Fusion in Dissimilarity Space for RGB-D Person Re-identification", *Array*, vol. 12, p. 100089, 2021.
20. M. K. Uddin, A. Lam, H. Fukuda, Y. Kobayashi and Y. Kuno, "Exploiting Local Shape Information for Cross-Modal Person Re-identification", In *Huang DS., Huang ZK., Hussain A. (eds) Intelligent Computing Methodologies, ICIC*, vol 11645, pp. 74-85, Springer, 2019.
21. F. Pala, R. Satta, G. Fumera and F. Roli, "Multimodal person reidentification using RGB-D cameras", *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4), pp. 788-799, 2015.
22. H. Liu, L. Hu and L. Ma, "Online RGB-D person re-identification based on metric model update", *CAAI Transactions on Intelligence Technology*, 2(1), pp. 48-55, 2017.
23. C. Patrino, R. Marani, G. Cicirelli, E. Stella and T. D'Orazio, "People re-identification using skeleton standard posture and color descriptors from RGB-D data", *Pattern Recognition*, 89, pp.77-90, 2019.
24. Liciotti, D., Paolanti, M., Frontoni, E., Mancini, A. and P. Zingaretti, "Person re-identification dataset with rgb-d camera in a top-view configuration", In *Video Analytics. Face and Facial Expression Recognition and Audience Measurement*, vol. 10165, pp. 1-11, Springer, 2017.
25. Z. Imani, H. Soltanizadeh and A.A. Orouji, "Short-Term Person Re-identification Using RGB, Depth and Skeleton Information of RGB-D Sensors", *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 44, pp. 669-681, 2020.
26. H. Moon and P. J. Phillips, "Computational and performance aspects of pca-based face-recognition algorithms," *Perception*, vol. 30, no. 3, pp.303-321, 2001.
27. Bhuiyan, A., Perina, A., V. Murino, "Person re-identification by discriminatively selecting parts and features", In *European conference on computer vision (ECCV)*, 2014.
28. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In *IEEE conference on computer vision and pattern recognition (CVPR)*, 2015.
29. Panda, R., Bhuiyan, A., Murino, V., Roy-Chowdhury, A.K.: Unsupervised adaptive re-identification in open world dynamic camera networks. In *IEEE conference on computer vision and pattern recognition IEEE conference on computer vision and pattern recognition (CVPR)*, 2017.
30. A. Paszke, S. Gross, S. Chintala and G. Chanan, Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration, 2017. [online] Available: <https://pytorch.org/>
31. E. Ahmed, M. Jones and T.K. Marks, "An improved deep learning architecture for person re-identification", In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3908-3916, 2015.
32. N. Karianakis, Z. Liu, Y. Chen, S. Soatto, Reinforced temporal attention and split-rate transfer for depth-based person re-identification, In *European conference on computer vision*, pp. 715-733, 2018.
33. Z. Yu, Y. Zhao, B. Hong, Z. Jin, J. Huang, D. Cai, X. He, X.S. Hua, Apparel-invariant Feature Learning for Apparel-changed Person Re-identification. arXiv preprint arXiv:2008.06181 (2020).
34. M.K. Uddin, A. Bhuiyan and M. Hasan, "Impact of Kernel-PCA on Different Features for Person Re-identification", *International Journal of Innovative Technology and Exploring Engineering*, 10(11), pp. 76-81, 2021.
35. A. Hermans, L. Beyer and B. Leibe, "In defense of the triplet loss for person re-identification", *arXiv preprint arXiv: 1703.07737* (2017).

AUTHORS PROFILE



Md Kamal Uddin, is currently working as an Assistant Professor in the department of Computer Science and Telecommunication Engineering at Noakhali Science and Technology University, Noakhali, Bangladesh. His main research interests include computer vision, pattern recognition, machine learning, person re-identification and video surveillance techniques.



Amran Bhuiyan, received the Bachelor degree in Applied Physics, Electronic & Communication Engineering from the University of Dhaka, Bangladesh in 2009, the M.Sc. degree in Computer Engineering and Information Technology from the Lucian Blaga University of Sibiu, Romania under the Erasmus Mundus external window in 2011 and the Ph.D. degree in Pattern Analysis and Computer Vision from the Istituto Italiano di Tecnologia, Genova, Italy. Previously, he was a Mitacs Elevate Postdoctoral Researcher with LIVIA, École de Technologie Supérieure, Université du Québec, Montréal, Canada and the industrial partner organization is SPORTLOGiQ INC. Currently, he is working as Associate Professor in the department of Computer Science and Telecommunication Engineering at Noakhali Science and Technology University, Bangladesh. His main research interests include computer vision, machine learning techniques for image and video processing, with applications such as video surveillance, summarizations, and sports analytics, and seeks to match or recognize individuals across non-overlapping views in a multi-camera system. Apart from these, he involves himself in designing simulation tools for the application of parameter variations in micro and macro electronic devices.



Mahmudul Hasan, is currently working as a computer science and engineering faculty member at Comilla University, Cumilla, Bangladesh. He obtained an M.Sc. (Thesis) in Computer Science and Engineering from the University of Rajshahi, Bangladesh. His research interest includes Machine Learning, Computer Vision and Human-Computer Interaction (HCI).