Iranian (Iranica) Journal of Energy & Environment
Journal Homepage: www.ijee.net
IJEE an official peer review journal of Babol Noshirvani University of Technology, ISSN:2079-2115

# Sensing Image Regions for Enhancing Accuracy in People Re-identification

Z. Mortezaie[1], H. Hassanpour[1]*, A. Beghdadi[2]

*[1] Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran*
*[2] Institut Galilée, Université Sorbonne Paris Nord, Villetaneuse, France*

*PAPER INFO*

*Paper history:*
Received 03 May 2022
Accepted in revised form 19 May 2022

*Keywords*:
Carried object
Crowded background
Partial occlusion
Re-identification

*ABSTRACT*

Video surveillance systems are widely used in the public and private sectors for maintaining security and healthcare purposes. Performance of surveillance systems directly depends on their accuracy in re-identification. There are three regions in a camera view, including person's body, background, and possible carried object by the person. Background, in existing approaches, is either overlooked or treated like a person's body in re-identification. In this paper, these three regions are considered in re-identification but with different importance. In our proposed technique, first, the input image is semantically segmented into the three regions using a deep semantic segmentation approach. Then, the effect of each region on characteristic features of people is tuned depending on the region's importance in re-identification. The proposed technique, leveraging robust descriptors, such as the Gaussian of Gaussian (GOG) and Hierarchical Gaussian Descriptors (HGD), can enhance existing methods in dealing with the challenging issues such as partial occlusion caused by carried objects and background in re-identification. Experimental results on commonly used people re-identification datasets demonstrate effectiveness of the proposed technique in improving performance of existing re-identification methods.

***doi****: 10.5829/ijee.2022.13.03.09*

## INTRODUCTION

Surveillance systems with a network of cameras often use several video cameras with non-overlapping views. These systems are used to analyze people's behavior and ultimately detect abnormal events in public and private places such as transport systems [1], government offices [2], and smart buildings in which elder citizens are cared after [3]. People re-identification is a challenging issue in surveillance systems, which directly affects their performance.

Re-identification systems deal with images of people in two sets, namely *gallery set* and *probe set*. The former includes the images of identified people and the latter involves the images of un-identified or newly arrived people viewed by a camera in the network. Determining labels of the probe set members is the goal of re-identification systems. To achieve this purpose, first, the similarity between each member of the probe set and the members of the gallery set is measured. Then, members of the gallery set are ranked depending on their similarity to the probe members. If the similarity of the most similar member of the gallery set to the probe member is less than a threshold, the probe member will be treated as a newly arrived person in the surveillance system, hence a new label is assigned to it.

There are a number of challenging issues that limit performance of re-identification systems, such as partial occlusion, pose variation, and illumination changes [4]. Figure 1 shows data samples from the CUHK01 database [5], where each row of the figure involves samples from the same person in two different camera views. In this figure, the samples shown in the first and the second columns were captured by the same camera; and the third and the fourth columns involve samples from another camera. As shown in this figure, the appearance of samples 1 and 2 change due to variations in illumination and pose. Partially occluded regions caused by carried

*Corresponding Author Email: h_hassanpour@yahoo.com (H. Hassanpour)

Please cite this article as: Z. Mortezaie, H. Hassanpour, A. Beghdadi, 2022. Sensing Image Regions for Enhancing Accuracy in People Re-identification, Iranian (Iranica) Journal of Energy and Environment, 13(3), pp. 295-304. Doi: 10.5829/ijee.2022.13.03.09

objects and pose variations lead to appearance changes in samples 3 and 4. In sample 5, the partial occlusion caused by the carried object and variations in illumination and pose changed the person's appearance.

According to the samples in Figure 1, appearance features from the images cannot describe the persons appropriately. In these samples, a person carries an object and moves across the camera network. In this situation, the carried objects may be disappeared or occlude the person's body, due to pose variations.

Although background of a person may not be the same in different camera views of a network, there are some similarities between backgrounds of adjacent cameras in a surveillance system. In this paper, three regions, including person's body, possible carried object, and background, are considered in a camera view, each of which with different importance in re-identification. A deep semantic segmentation approach is employed in this paper to segment each input image into regions corresponding to the person's body, possible carried objects, and background. We show that contributing the three different regions with a significance factor improves accuracy of re-identification systems.

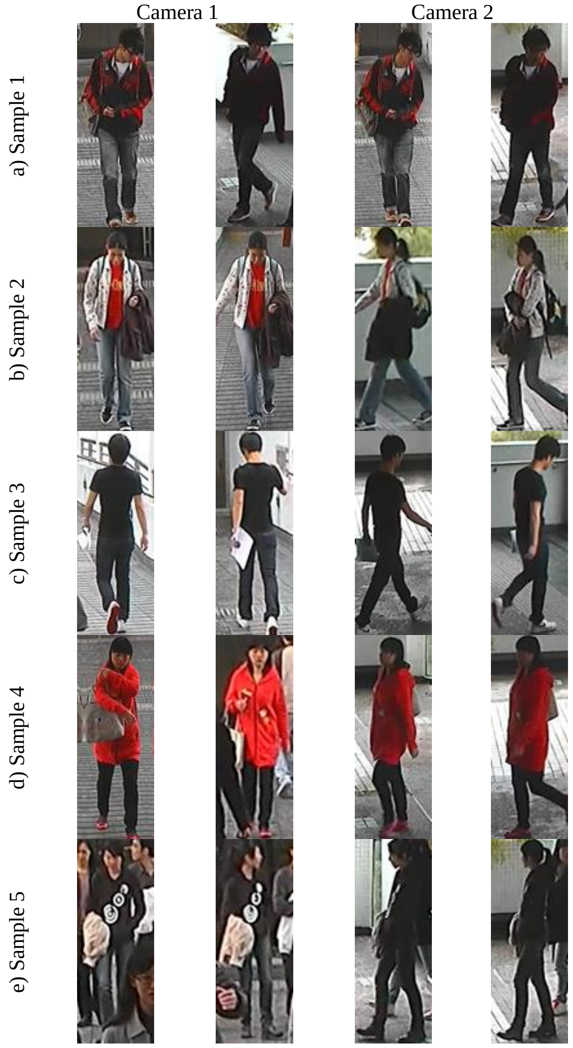

**Figure 1.** Data samples from the CUHK01 database, each row represents a person at two different camera views

## RELATED WORKS

Matsukawa et al. [6, 7], proposed descriptors robust to pose and illumination changes, namely GOG (Gaussian of Gaussian) and HGD (Hierarchical Gaussian Descriptors) for people re-identification. In these approaches, to locally describe structure of the input image, the image is considered as seven horizontal regions with 50% overlapping. Also, each region is considered as a number of overlapping windows. Each pixel in the window is depicted using several features, including its location in the vertical direction, magnitudes of the pixel intensity gradient in four different orientations, and the color information in RGB, HSV, LAB, and nRGB color spaces. These features are used for computing the Gaussian distribution in the corresponding window. Hence, each region is represented as a set of multiple Gaussian distributions. The Gaussian distributions are non-linear spaces and the Euclidean operation cannot be directly applied on these spaces. Hence, in GOG and HGD, distributions' parameters associated with the windows and regions are mapped into the linear tangent space via one of the Riemannian manifolds namely Symmetric Positive Definite (SPD). Finally, the mapped Gaussian distributions of the regions are used to describe the whole input image. However, in HGD some feature norm normalization approaches are used to decrease the bias of SPD matrix descriptors [8].

Liao et al. [9] proposed a descriptor robust to illumination and viewpoint changes, namely Local Maximal Occurrence (LOMO). In this method, histograms from the Scale Invariant Local Ternary Pattern (SILTP) [10] along with an 8×8×8-bin joint HSV histogram are used for describing the image. This descriptor initially divides the input image into a number of overlapping windows. For each window, bins of the histograms are assumed to represent the occurrence probability of the corresponding pattern in the window. Then, for extracting features robust to viewpoint changes, the windows at the same horizontal location are analyzed and the local occurrence of each pattern (i.e., the same histogram bin) is maximized. Also, in this approach, to learn a discriminant low dimensional subspace, a subspace and the cross-view quadratic discriminant analysis (XQDA) method are introduced where the XQDA metric is learned on the derived subspace simultaneously.

Vishwakarma et al. [11] proposed a Multi-Level Gaussian Descriptor (MLGD), where some low-level features such as color moment values of RGB components and Schmid filter responses are used for representing each pixel of the images. The feature extraction mechanism used in this approach is similar to work presented by Matsukawa et al. [6]; but the low-level features used by Vishwakarma and Upadhyay [11] are different from Matsukawa et al. [6].

Prates et al. [12] proposed a Kernel Cross-View Collaborative Representation based Classification (Kernel X-CRC) approach, in order to cover appearance changes issues caused by different cameras conditions. In this method, GOG descriptor was used to extract people appearance characteristics. Also, for determining similarity between the images, first, the extracted features were mapped into the learned subspaces, and then, they were passed through the Kernel X-CRC. Besides, Prates and Schwartz [13] [reposed a nonlinear regression model namely, Kernel Multiblock Partial Least Squares (Kernel MBPLS). In this approach, the features extracted from data samples are mapped onto a low-dimensional subspace considering multiple sources of data. The approach proposed by Prates and Schwartz [13] uses the extracted features from GOG descriptor, and maps them into the learned subspaces, similar to work presented by Prates et al. [12]. Finally, the mapped features are used as the input of the Kernel X-CRC for person re-identification.

A Graph Correspondence Transfer (GCT) people re-identification approach was proposed by Zhou et al. [14] for dealing with the issues associated with variations in viewpoint and pose. In this approach, first, positive image pairs with various pose-pair configurations are used to learn a set of patch-wise correspondence templates via a patch-wise graph matching mechanism. Then, for each pair of test images, some training pairs with the most similar pose-pair configurations are selected as references. Then, for computing the similarity between images, the correspondences of the references are transferred to test pair. Further, for empowering the correspondence transfer used by Zhou et al. [14]; in fact, Zhou et al. [15] used a pose context descriptor based on the topology structure of the estimated joint locations [16].

To overcome the issue of appearance changes across camera views, Fang et al. [17] proposed a Sample Specific Multi- Kernel (SSMK) approach. In this approach, first, the images are divided into six horizontal regions, and afterwards some features such as RGB, YUV, HSV, LAB and YCbCr color information, Dense SIFT [18, 19], color naming feature [20] and deep features [21] are extracted from each region. Then, the extracted features are mapped into the weighed multi-kernel feature space to learn a discriminative metric for re-identification.

For reducing the number of labeled training samples in person re-identification, by Jia et al. [22], a View-Specific Semi-supervised Subspace Learning (VS-SSL) approach was proposed. This approach uses GOG descriptor for representing the images, and also, learns specific projections for each camera view.

Zhao et al. [23] used Inexact Augmented Lagrange Multiplier (IALM) algorithm [24] for people re-identification by considering the re-identification as a consistent iterative multi-view joint transfer learning optimal problem. The goal of this approach is handling the issue of inconsistency in data distributions across various camera views.

Some approaches use both the low-level features (i.e., color and texture) and mid-level characteristics (carried objects and clothes with a specific color) in order to cover the issues of viewpoint and illumination variations in re-identification [25–28]. Layne et al. [25] selected mid-level characteristics such as carried objects, sunglasses, and logos, as distinctive characteristics. For each characteristic, a classifier is trained using the low-level features of the training samples. The trained classifiers are then used to investigate the existence of the mid-level characteristics in the probe images.

An unsupervised saliency learning approach was proposed by Zhao et al. [26]. In these methods, in order to obtain a saliency map for each image, the patches of the image are compared with a reference set without any distinctive characteristics. The comparisons are done based on the color and texture of the patches. Also, the *K*-Nearest Neighbors and one-class SVM are used for computing the saliency score of each patch [29]. After obtaining the saliency maps of the images, the probe image is re-identified by comparing its saliency map with the saliency maps of the gallery images.

Martinel et al. [27, 30] used the saliency maps of the images for weighing low-level features. In these approaches, the image saliency map is obtained via the Markov chain approach [28]. The value of each pixel in the image saliency map denotes the importance of the pixel in the final extracted features. Both the weighed and non-weighed features are then used in a pairwise-based multiple metric learning framework.

In some of the existing re-identification approaches, the learned and extracted features from Convolutional Neural Networks (CNNs) are used for people re-identification. Sun et al. [31] optimized the deep representation learning process using the Singular Vector Decomposition (SVD) and proposed SVDNet for extracting global appearance features using CNN. To overcome the issue of misaligned images, Zheng et al. [32], proposed the pedestrian alignment network (PAN) based on the CNN feature maps. In this network, the pedestrians are aligned within bounding boxes and simultaneously pedestrian descriptors are trained. Yu et al. [33] simulated a number of investigators for each probe image, by combining the hand-crafted and deep learning-based features introduced in literature [6, 34, 35] with various metric learning methods in pairs. Then, a crowd sourcing-based ranking aggregation approach was proposed to fuse the ranking lists obtained from the investigators. Lin et al. [36] jointly trained multigranularity attention selection and feature representation using a proposed Harmonious Attention CNN (HA-CNN) module, where the correlated complementary information between attention selection and feature discrimination is maximized.

The above-mentioned approaches do not handle the issue of appearance changes caused by carried objects and background in re-identification. Mortezaie et al. [37] proposed a re-ranking approach based on the color of person body and carried objects in order to improve the performance of GOG and HGD approaches. In this re-ranking approach, first, the input image is segmented into person's body, carried object, and background. Then, the colors of person body and carried objects are categorized using one of the color categorization approaches introduced by Parraga and Akbarinia [38]. Besides, a pre-processing step namely unification process was introduced by Mortazaie et al. [39] for reducing the effects of the carried objects on person's appearance. In this approach, the color of the occluded body parts caused by carried objects was simulated considering their non-occluded adjacent regions related to the body.

The re-identification approaches use either the hand-crafted features or the features from deep neural networks. In this section some of the existing approaches were reviewed briefly. Note that the re-identification approaches based on hand-crafted features do not address the issue of appearance changes caused by carried objects and backgrounds in re-identification. Also, deep learning re-identification approaches, such as those mentioned above, need a very large dataset for training. In addition, the training phase of these approaches are very time consuming as there are a considerable number of parameters to tune. The existing re-identification approaches are trained to disregard the masked area of the body rather than automatically recognizing occlusion due to carried objects.

## PEOPLE RE-IDENTIFICATION

In our proposed technique, the input image is initially segmented into three segments as person's body, possible carried objects, and background using a deep semantic segmentation approach namely DeepLabv3+ [40]. Then, appearance characteristics are extracted from each of the segments, and they are incorporated for the re-identification proportional to their importance.

In DeepLabv3+, an encoder-decoder structure is applied in order to semantically segmenting the images. In this approach, the contextual information is achieved via the encoder module. Besides, the object boundaries are obtained via the decoder module.

We trained DeepLabv3+ using the manually segmented masks of VIPeR [41] images to segment the images into three regions as background, person, and partially occluded areas caused by carried objects. The segmented masks are publicly available in the PRID450s database [42].

Samples of data from the VIPeR (i.e., samples 1 and 2) and CUHK01 (i.e., samples 3 and 4) databases, and their corresponding Semantic Segmented Maps (SSM) using the trained DeepLabv3+ are shown in Figure 2, respectively in the first and second rows. In the segmented images, the background, person, and partially occluded regions are depicted using black, white, and grey colors respectively.

In our proposed technique, after obtaining SSM for the input image, we consider a significance factor (in the interval (0-1]) for each segment. The significance factors are then multiplied with the feature vectors extracted from each region in re-identification. Since appearance of a person has the major role in re-identification, the maximum value, i.e., 1, is assigned to factor related to the person's body. The significance factor associated with a carried object is set to 0.9 as it is less important compared to the person's body in re-identification. The background has the lowest importance in person re-identification comparing to both the person's body and carried objects. Hence, 0.5 is assigned experimentally to the significance factor of the background.

As mentioned before, sensing image regions and considering their significance in re-identification can improve performance of the technique. Below, we briefly introduce three published people re-identification methods and incorporate the proposed factoring scheme to improve their performance. Then their improvements are experimentally evaluated.

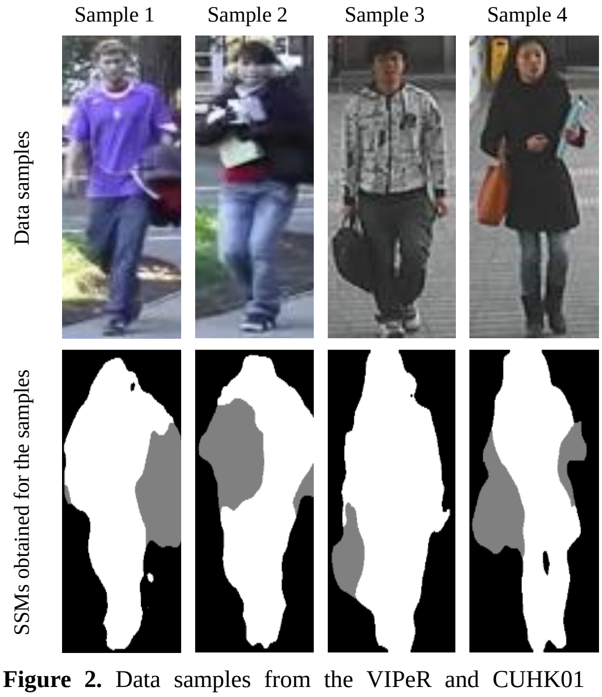

**Figure 2.** Data samples from the VIPeR and CUHK01 databases, and their corresponding SSM

The re-identification approaches proposed by Matsukawa et al. [6, 7] are robust to appearance changes such as scene illumination and pose variations. But these approaches cannot overcome the appearance changes caused by partially occluded regions and the issue of crowded backgrounds.

Both the GOG and HGD descriptors consider the image as seven horizontal regions with 50% overlapping. Also, each region is considered as a number of overlapping windows. Then these descriptors extract features for individual pixels ($i$) of each window using Equation (1), by Matsukawa et al. [6, 7], without considering their association with the background, person's body and possible partially occluded regions:

$$F_i = [v; D_{0^\circ}; D_{90^\circ}; D_{180^\circ}; D_{270^\circ}; x_R; x_G; x_B]^{\boldsymbol{T}}, \tag{1}$$

where $v$ denotes pixel's vertical location; $D_{0^\circ}, D_{90^\circ}, D_{180^\circ}, D_{270^\circ}$, represent the magnitudes of the pixel intensity gradient in four different orientations; and $x_R, x_G, x_B$, are the pixel values in R, G, and B channels.

In our proposed approach, after segmenting the images into the person's body, carried objects, and background, to make the GOG and HGD descriptors robust against appearance changes caused by crowded background and carried object, the extracted feature set from each segment is biased using the proposed significance factoring scheme.

The Gaussian distributions of pixels in the windows are computed using its extracted featured (biased feature). Matsukawa et al. [6, 7] obtained Gaussian distributions which are non-linear spaces, are mapped on the linear tangent spaces. The mapped Gaussian distributions of the windows are then used to describe the corresponding region as a set of multiple Gaussian distributions. Similar to Gaussian distributions of the windows, the Gaussian distributions of the regions are mapped into the tangent space.

By concatenating the mapped Gaussian distributions of the regions, the whole input image is described by Matsukawa et al. [6, 7] as follows:

$$Z_{RGB} = \left[z_{g1}^T, z_{g2}^T, z_{g3}^T, z_{g4}^T, z_{g5}^T, z_{g6}^T, z_{g7}^T\right]^T, \tag{2}$$

where $z_{gk}, k = 1,2,\ldots,7$ represents the mapped Gaussian distribution of region $k$.

In GOG and HGD, in addition to RGB color space, the LAB, HSV, and nRGB color spaces are also used by substituting their color channels with RGB information (i.e., $x_R, x_G$, and $x_B$) used in Equation (1). We name the extracted feature vectors using LAB, HSV, and nRGB[1], as $Z_{LAB}, Z_{HSV}$, and $Z_{nRGB}$ respectively. By concatenating $Z_{RGB}, Z_{LAB}, Z_{HSV}$, and $Z_{nRGB}$, the final feature vector is discussed by Matsukawa et al. [6, 7] as follows:

$$Z_{Fusion} = [Z_{RGB}^T, Z_{LAB}^T, Z_{HSV}^T, Z_{nRGB}^T]^T. \tag{3}$$

Matsukawa et al. [6, 7], $Z_{Fusion}$, are then used to learn the XQDA distance metric. The XQDA learns both a discriminative subspace and a distance metric simultaneously by extending Bayesian face [43] and KISSME [44] approaches. This problem in XQDA is treated as a Generalized Rayleigh Quotient [45]. Besides, the generalized eigenvalue decomposition is used to achieve a closed form solution.

As the third approach to incorporate our proposed technique, we consider the feature extraction method proposed by Martinel et al. [27] for re-identification. Martinel et al. [27] proposed a KErnelized saliency-based Person re-identification through multiple metric LEaRning, namely KEPLER. In this approach, first, the saliency maps ($\omega$) of the images are computed in a kernelized saliency detection module which is based on the Markov chain approach [28]. The saliency maps are further used for weighing the extracted features. In feature extraction step, first, the input image is transformed into HSV, Lab, YUV, rgs[1], RGB, and gray color spaces. Each color channel ($l$) and the saliency map ($\omega$) of the input image, are divided into a number of overlapping windows. Then, the color mean, 128-dimensional SIFT descriptor, and Haar-like sparse-compressive features [46], are extracted from each window ($j$) in each color channel ($l$), and also, the Local Binary Pattern (LBP) [47] is extracted from each window ($j$) of the grayscale image. Martinel et al. [27] mentioned features are extracted in the non-weighed form. Meanwhile, in this approach, a histogram feature, namely the saliency weighed histogram ($H$) is extracted in the weighed form using $\omega$ as follow:

$$H_{a,b}^{j,l} = \sum\nolimits_{(m,n)\in T^{j,l}} \begin{cases} \omega_{m,n}^{j}, & a < T_{m,n}^{j,l} \le b \\ 0 & o.w \end{cases} \tag{4}$$

where, $\omega_{m,n}^{j}$ and $T_{m,n}^{j,l}$ are the saliency value and the pixel intensity at location $(m,n)$ for window $j$ and color channel $l$ respectively. Also, $a$ and $b$ denote the lower and upper bin limits.

In this paper, we incorporate our proposed significance factor in KEPLER approach by substituting $\omega$ in Equation (4) with the significance factor.

## EXPRIMENTAL RESULTS

One of the commonly measures used for evaluating the performance of re-identification systems, is Rank-*k*, where *k* indicates the number of top matches with correct answer. Rank-*k* is the strictest measure for *k* = 1, whereas, this measure permits some errors for k > 1 [48]. In this paper, Rank-*k* (*k* = 1, 5, 10, 20) is used in order to evaluating performance of the proposed technique on CUHK01 [5], VIPeR [41], PRID450s [42], and CUHK03 [49] datasets.

[1] r = R/(R + G + B), g = G/(R + G + B), s = (R + G + B)/3

The VIPeR and PRID450s datasets respectively involve 1,264 images of 632 persons, and 900 images of 450 persons, captured in two different camera views. As these datasets contain one image of each person in each camera view, we evaluate the performance of our proposed approach with single-shot matching. The CUHK01 dataset involves 3,884 images of 971 persons, where two images of each person are captured in each camera view. Hence, in Table 1, the performance of our proposed approach on this dataset is reported with single-shot matching (M=1) and multi-shot (M=2) matching. The CUHK03 dataset contains 13,164 images of 1,360 persons, where, averagely 4.8 images of each person are captured in each camera view. We use the manually cropped images (labeled) of this dataset and evaluate the performance of the comparing approaches with multi-shot matching.

Table 1 summarized the effect of our proposed approach on GOG, HGD, and KEPLER re-identification methods. In this table, each ranking is the obtained results using our proposed approach applied on GOG descriptor (i.e., the enhanced GOG using the significance factor), the classic GOG method; the obtained results using our proposed approach applied on HGD descriptor (i.e., the enhanced HGD using the significance factor), the classic HGD method; the obtained results using our proposed approach applied on KEPLER (i.e., the enhanced KEPLER), and its classic version respectively. For each ranking order, this table shows the obtained results from the enhanced and classic version of the corresponding method in a row where the bolded results are the most accurate results from the corresponding database.

In this paper, the XQDA distance metric used in GOG and HGD, and the KEPLER method are learned and tested using 10 different train and test sets of data samples. Hence, the average of the obtained accuracy using the test sets were reported in Table 1.

As shown in Table 1, generally the enhanced GOG, and the improved HGD using the significance factoring scheme, achieve more accurate results compared to the original approaches, for ranks 1, 5, 10, and 20 on all comparing databases. Also, applying our proposed

**Table 1.** Performance of our proposed approach and the methods reported in literature [6, 7, 27]

| | **Re-identification approach** | **VIPeR** | **CUHK01<br>(M=1)** | **CUHK01<br>(M=2)** | **PRID<br>450s** |
|---|---|---|---|---|---|
| Rank1 % | Enhanced GOG using $\widehat{W}_S$ | **59.4** | **62.5** | **72.1** | **78.7** |
| | Classic GOG [6], (2016) | 49.7 | 57.8 | 67.3 | 68.4 |
| | Enhanced HGD using $\widehat{W}_S$ | **60.9** | **63.3** | **74.0** | **80.5** |
| | Classic HGD [7], (2019) | 50.0 | 59.0 | 70.3 | 70.4 |
| | Enhanced KEPLER using $\widehat{W}_S$ | **41.7** | **42.0** | **57.0** | **52.0** |
| | Classic KEPLER [27], (2015) | 40.2 | 42.0 | 54.8 | 51.9 |
| Rank5 % | Enhanced GOG using $\widehat{W}_S$ | **85.0** | **82.4** | **89.5** | **93.5** |
| | Classic GOG [6], (2016) | 79.7 | 79.1 | 86.9 | 88.8 |
| | Enhanced HGD using $\widehat{W}_S$ | **85.9** | **83.5** | **90.0** | **94.7** |
| | Classic HGD [7], (2019) | 79.5 | 79.7 | 87.9 | 91.2 |
| | Enhanced KEPLER using $\widehat{W}_S$ | **71.3** | **66.0** | **79.3** | 77.2 |
| | Classic KEPLER [27], (2015) | 70.0 | 65.0 | 78.9 | **77.6** |
| Rank10 % | Enhanced GOG using $\widehat{W}_S$ | **91.7** | **88.9** | **94.1** | **97.1** |
| | Classic GOG [6], (2016) | 88.7 | 86.2 | 91.8 | 94.5 |
| | Enhanced HGD using $\widehat{W}_S$ | **92.2** | **89.8** | **94.1** | **97.7** |
| | Classic HGD [7], (2019) | 88.9 | 86.2 | 92.2 | 94.8 |
| | Enhanced KEPLER using $\widehat{W}_S$ | **81.5** | **75.3** | **86.0** | **84.6** |
| | Classic KEPLER [27], (2015) | 81.4 | 74.5 | 85.3 | 84.5 |
| Rank20 % | Enhanced GOG using $\widehat{W}_S$ | **95.9** | **94.0** | **97.3** | **98.8** |
| | Classic GOG [6], (2016) | 94.5 | 92.1 | 95.9 | 97.8 |
| | Enhanced HGD using $\widehat{W}_S$ | **96.7** | **94.3** | **97.5** | **99.2** |
| | Classic HGD [7], (2019) | 94.6 | 92.0 | 95.8 | 97.6 |
| | Enhanced KEPLER using $\widehat{W}_S$ | **90.7** | **83.6** | **91.0** | **91.0** |
| | Classic KEPLER [27], (2015) | 90.4 | 83.0 | 90.8 | 90.7 |

approach on KEPLER, leads to achieve more accurate results comparing to the classic KEPLER, for ranks 1, 5, 10, and 20 on VIPeR, and CUHK01 databases, as well as, for ranks 1, 10, and 20 on PRID450s.

Note that, the VIPeR database includes many images with a crowded background and a partially occluded region. The obtained results on this database show that the extracted feature vectors, biased using the proposed approach, are more distinctive and more accurate comparing to the original non-biased feature vectors.

Note that, in our proposed technique, for segmenting images into person's body, carried objects, and background regions, first, DeepLabv3+ is trained using the manually segmented masks of the VIPeR and PRID450s datasets; also, the trained network is then used to segment the images of the other datasets (i.e., CUHK01 and CUHK03). Hence, the overhead of training DeepLapv3+ can be ignored as the parameters of the network is tuned once during training. According to Table 1, we can apply our proposed biasing scheme on descriptors by multiplying the factors on the extracted features. Hence, biasing the descriptors imposes the constant overhead as $O(1)$.

In Tables 2 to 5, the performance of our proposed approach is compared using state-of-the-art methods on the CUHK03, VIPeR, CUHK01, and PRID450s datasets, respectively. Note that most of the re-identification approach reported their accuracy on CUHK03 dataset only in rank 1; hence, in Table 2, we compared the performance of our people re-identification approach with the other methods in rank 1.

As mentioned in section 2, the re-identification approaches proposed in literature [16, 31–33, 36] used deep neural networks to learn and extract the appearance features. In these approaches tuning deep networks to learn appropriate features is space and time consuming.

In addition to deep learning-based approaches, some hand-crafted based approaches involve training phase in feature extraction. Zhou et al. [14, 15] trained a set of patch-wise correspondence templates using a patch-wise graph matching mechanism. In the mentioned training

**Table 2.** Comparison of the performance of our proposed approach with state-of-the-art methods on CUHK03 (labeled)

| **Approach** | **Rank1 (%)** |
|---|---|
| Sun et al., [31], (2017) | 40.9 |
| Zheng et al, [32], (2018) | 36.9 |
| Lin et al, [36], (2018) | 44.4 |
| Yu et al, [33], (2020) | 53.9 |
| Classic GOG [6], (2016) | 67.3 |
| Classic HGD [7], (2019) | 68.9 |
| Enhanced GOG | **69.6** |
| Enhanced HGD | 68.5 |

**Table 3.** Comparison of the performance of our proposed approach with state-of-the-art methods on VIPeR

| | **Ranks%** | | | |
|---|---|---|---|---|
| Approaches | **1** | **5** | **10** | **20** |
| Layne et al., [25], (2012) | 18.8 | 40.9 | 54.9 | - |
| Zhao et al., [26], (2013) | 26.7 | 50.7 | 62.4 | 76.4 |
| Martinel et al., [30], (2014) | 33.0 | - | 75.6 | 86.9 |
| Liao et al., [9] (2015) | 40.0 | - | 80.5 | 91.1 |
| Vishwakarma et al., [11], (2018) | 47.5 | - | 87.9 | 93.7 |
| Prates et al. [12], (2019) | 51.6 | 80.5 | 89.5 | 95.2 |
| Prates et al. [13] (2019) | 51.2 | 79.9 | 89.9 | - |
| Fang et al., [17], (2019) | 43.8 | 79.2 | 87.2 | 94.9 |
| Jia et al., [22], (2020) | 44.8 | 72.3 | 79.3 | 86.1 |
| Mortezaie et al., [37], (2021) | 53.0 | 82.7 | 90.7 | 95.7 |
| Enhanced GOG | 59.4 | 85.0 | 91.7 | 95.9 |
| Enhanced HGD | **60.9** | **85.9** | **92.2** | **96.7** |

**Table 4.** Comparison of the performance of our proposed approach with state-of-the-art methods on CUHK01 (m=2)

| | **Ranks%** | | | |
|---|---|---|---|---|
| Approaches | **1** | **5** | **10** | **20** |
| Liao et al., [9] (2015) | 63.2 | - | 90.8 | 94.9 |
| Vishwakarma et al., [11], (2018) | 54.5 | - | 83.5 | 90.5 |
| Fang et al., [17], (2019) | 69.2 | 87.8 | 93.2 | 97.1 |
| Prates et al., [12], (2019) | 63.1 | 82.7 | 89.0 | 94.6 |
| Zhao et al., [23], (2020) | 68.4 | 86.3 | 93.6 | 96.8 |
| Mortezaie et al., [37], (2021) | 70.7 | 88.2 | 92.3 | 96.2 |
| Enhanced GOG | 72.1 | 89.5 | 94.1 | 97.3 |
| Enhanced HGD | **74.0** | **90.0** | **94.1** | **97.5** |

**Table 5.** Comparison of the performance of our proposed approach with state-of-the-art methods on Prid450s

| | **Ranks%** | | | |
|---|---|---|---|---|
| Approaches | **1** | **5** | **10** | **20** |
| Liao et al., [9] (2015) | 62.6 | 85.6 | 92.0 | 96.6 |
| Vishwakarma et al., [11], (2018) | 62.4 | - | 93.5 | 96.9 |
| Zhou et al., [14], (2018) | 58.4 | 77.6 | 84.3 | 89.8 |
| Zhou et al., [15], (2019) | 70.9 | 89.1 | 93.5 | 96.5 |
| Prates et al. [12], (2019) | 71.3 | 91.7 | 96.0 | 98.1 |
| Prates et al. [13] (2019) | 68.1 | 90.7 | 95.0 | - |
| Jia et al., [22], (2020) | 68.2 | 90.2 | 94.9 | 98.0 |
| Zhao et al., [23], (2020) | 72.1 | - | 94.6 | - |
| Mortezaie et al., [37], (2021) | 74.9 | 93.0 | 96.6 | 99.1 |
| Enhanced GOG | 78.7 | 93.5 | 97.1 | 98.8 |
| Enhanced HGD | **80.5** | **94.7** | **97.7** | **99.2** |

step using several positive image pairs with various pose-pair configurations increases computational complexity of these approaches. Zhao et al. [23] considered the re-identification process as a consistent iterative multi-view joint transfer learning optimal problem. Also, Layne et al., [25] used some distinctive mid-level characteristics such as carried objects, sunglasses, and logos, where for each characteristic, a classifier is trained using the low-level features of the training samples. But, training a number of mid-level features is time consuming compared to only using the raw low-level features. Besides, Zhao et al. [26] computed a patch-wise saliency map for each image by comparing the color and texture of images' patches with the patches of the reference images. These comparisons bring many computational overheads on the approach proposed by Zhao et al. [26]. Note that, the GOG and HGD descriptors do not involve any learning process. Hence, the computational complexity of the enhanced GOG and enhanced HGD is lower than the mentioned approaches.

Also, Martinel et al. [30], computed a weight map for each image using Markov chain approach. Note that, in our proposed approach semantically segmenting the images with a pre-trained network and assigning a constant number to each segment needs simpler computations than using Markov chain approach for each input image. Meanwhile, the LOMO descriptor proposed by Liao [9], is simpler than GOG and HGD as it is only based on two hand-crafted features (i.e., HSV and SILTP). But, according to Tables 3 to 5, the accuracy of the enhanced GOG and enhanced HGD, are considerably more than the accuracy of LOMO.

Meanwhile, the feature extraction mechanism used by Vishwakarma and Upadhyay [11] is similar to GOG and HGD, where, it only used different low-level features from Matsukawa et al. [6, 7] work. Consequently, the computational complexity of approach introduced by Vishwakarma and Upadhyay [11] is similar to GOG and HGD, whereas, according to Tables 3 to 5, the performance of the enhanced GOG and enhanced HGD are better than this approach.

Also, the re-identification approaches proposed by Prates and Schwartz [12, 13], Jia et al. [22] and Mortezaie et al. [37] are based on GOG descriptor. Hence, the computational complexity of these approaches in feature extraction is similar to the enhanced GOG and enhanced HGD.

Besides, according to Table 2, despite using hand-crafted features in enhanced GOG and enhanced HGD the performance of these approaches is better than the deep learning-based approaches on CUHK03. Meanwhile, as it is shown in Tables 3 to 5, and Figures 3 to 5, our proposed approach outperforms the comparing methods in all comparing ranks on the VIPeR, CUHK01, and PRID450s datasets. Indeed, in our proposed technique, reducing the effects of the partial occlusion caused by carried objects, as well as background, on the person's appearance leads to superiority of enhanced GOG and enhanced HGD comparing to other approaches. Consequently, applying our proposed technique on appearance-based descriptors can improve the accuracy of the re-identification without imposing further processing overhead.

## CONCLUSION

The performance of the surveillance systems directly depends on the performance of re-identification systems. Extracting characteristics from different regions of the images considering their significance in re-identification, can improve the performance of appearance-based re-identification approaches. To achieve this goal, in this paper, a technique is proposed, where, the effect of each pixel of the images on their extracted feature vectors is tuned considering its association with background, person's body and partially occluded regions caused by carried objects. The experimental results on various databases show effectiveness of the proposed technique in improving performance of existing re-identification methods. In our proposed approach the significance factor for each region of the image was determined exprimentally. In future, the significance factors can be automatically computed considering the region and its size in the image.

## REFERENCES

1. Hammoudi, K., Abu Taha, M., Benhabiles, H., Melkemi, M., Windal, F., El Assad, S., and Queudet, A., 2020. Image-Based Ciphering of Video Streams and Object Recognition for Urban and Vehicular Surveillance Services. In Fourth International Congress on Information and Communication Technology, pp 519–527. Doi: 10.1007/978-981-32-9343-4_42
2. Goyal, A., Anandamurthy, S.B., Dash, P., Acharya, S., Bathla, D., Hicks, D., Bhan, A., and Ranjan, P., 2020. Automatic Border Surveillance Using Machine Learning in Remote Video Surveillance Systems. In Emerging Trends in Electrical, Communications, and Information Technologies, pp.751-760. Springer, Singapore. Doi:10.1007/978-981-13-8942-9_64
3. Huang, Z., Liu, Y., Fang, Y., and Horn, B.K.P., 2018. Video-based Fall Detection for Seniors with Human Pose Estimation. In: 2018 4th International Conference on Universal Village (UV). IEEE, pp 1–4. Doi: 10.1109/UV.2018.8642130
4. Gawande, U., Hajari, K., and Golhar, Y., 2020. Pedestrian detection and tracking in video surveillance system: issues, comprehensive review, and challenges. Recent Trends in Computational Intelligence, pp.1–24. London, United Kingdom: IntechOpen
5. Li, W., Zhao, R., and Wang, X., 2013. Human Reidentification with Transferred Metric Learning. Asian conference on computer vision, Berlin, Heidelberg, pp: 31-44. Doi: 10.1007/978-3-642-37331-2_3
6. Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y., 2016. Hierarchical Gaussian Descriptor for Person Re-identification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp: 1363-1372. Doi: 10.1109/CVPR.2016.152

7. Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y., 2020. Hierarchical Gaussian Descriptors with Application to Person Re-Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9), pp.2179–2194. Doi: 10.1109/TPAMI.2019.2914686

8. Li, P., Wang, Q., and Zhang, L., 2013. A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models. In: 2013 IEEE International Conference on Computer Vision. pp: 1689–1696. Doi: 10.1109/ICCV.2013.212.

9. Liao, S., Hu, Y., Xiangyu Zhu, and Li, S.Z., 2015. Person re-identification by Local Maximal Occurrence representation and metric learning. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp: 2197-2206. Doi: 10.1109/CVPR.2015.7298832

10. Liao, S., Zhao, G., Kellokumpu, V., Pietikainen, M., and Li, S.Z., 2010. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2197-2206. Doi: 10.1109/CVPR.2015.7298832

11. Vishwakarma, D.K., and Upadhyay, S., 2018. A Deep Structure of Person Re-Identification Using Multi-Level Gaussian Models. *IEEE Transactions on Multi-Scale Computing Systems*, 4(4), pp.513–521. Doi: 10.1109/TMSCS.2018.2870592

12. Prates, R., and Schwartz, W.R., 2019. Kernel cross-view collaborative representation based classification for person re-identification. *Journal of Visual Communication and Image Representation*, 58, pp.304–315. Doi: 10.1016/j.jvcir.2018.12.003

13. Prates, R., and Schwartz, W.R., 2019. Matching People Across Surveillance Cameras. In: Anais Estendidos da Conference on Graphics, Patterns and Images (SIBGRAPI). Sociedade Brasileira de Computação - SBC, pp. 84-90. Doi: 10.5753/sibgrapi.est.2019.8306

14. Zhou, Q., Fan, H., Zheng, S., Su, H., Li, X., Wu, S., and Ling, H., 2018. Graph Correspondence Transfer for Person Re-Identification | Proceedings of the AAAI Conference on Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). Doi: 10.5555/3504035.3504966

15. Zhou, Q., Fan, H., Yang, H., Su, H., Zheng, S., Wu, S., and Ling, H., 2021. Robust and Efficient Graph Correspondence Transfer for Person Re-Identification. *IEEE Transactions on Image Processing*, 30, pp.1623–1638. Doi: 10.1109/TIP.2019.2914575

16. Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 7291–7299.

17. Fang, J., Zhang, R., and Jiang, F., 2019. Sample Specific Multi-Kernel Metric Learning for Person Re-identification. In: 2nd International Conference on Electrical and Electronic Engineering (EEE 2019). Atlantis Press, pp 226–241. Doi: 10.2991/eee-19.2019.38

18. Liu, Y., Liu, S., and Wang, Z., 2015. Multi-focus image fusion with dense SIFT. *Information Fusion*, 23, pp.139–155. Doi: 10.1016/j.inffus.2014.05.004

19. An, L., Kafai, M., Yang, S., and Bhanu, B., 2013. Reference-based person re-identification. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, pp 244–249. Doi: 10.1109/AVSS.2013.6636647

20. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., and Li, S.Z., 2014. Salient Color Names for Person Re-identification. pp 536–551. Doi: 10.1007/978-3-319-10590-1_35

21. Ahmed, E., Jones, M., and Marks, T.K., 2015. An improved deep learning architecture for person re-identification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 3908–3916. Doi: 10.1109/CVPR.2015.7299016

22. Jia, J., Ruan, Q., Jin, Y., An, G., and Ge, S., 2020. View-specific subspace learning and re-ranking for semi-supervised person re-identification. *Pattern Recognition*, 108, pp.107568. Doi: 10.1016/j.patcog.2020.107568

23. Zhao, C., Wang, X., Zuo, W., Shen, F., Shao, L., and Miao, D., 2020. Similarity learning with joint transfer constraints for person re-identification. *Pattern Recognition*, 97, pp.107014. Doi: 10.1016/j.patcog.2019.107014

24. Xu, Y., Fang, X., Wu, J., Li, X., and Zhang, D., 2016. Discriminative Transfer Subspace Learning via Low-Rank and Sparse Representation. *IEEE Transactions on Image Processing*, 25(2), pp.850–863. Doi: 10.1109/TIP.2015.2510498

25. Layne, R., Hospedales, T.M., and Gong, S., 2012. Towards Person Identification and Re-identification with Attributes. pp 402–412. . Doi: 10.1007/978-3-642-33863-2_40

26. Zhao, R., Ouyang, W., and Wang, X., 2013. Unsupervised Salience Learning for Person Re-identification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 3586–3593. Doi: 10.1109/CVPR.2013.460

27. Martinel, N., Micheloni, C., and Foresti, G.L., 2015. Kernelized Saliency-Based Person Re-Identification Through Multiple Metric Learning. *IEEE Transactions on Image Processing*, 24(12), pp.5645–5658. Doi: 10.1109/TIP.2015.2487048

28. Harel, J., Koch, C., and Perona, P., 2006. Graph-based visual saliency. *Advances in neural information processing systems*, 19, pp. 545-552.

29. Heller, K., Svore, K., Keromytis, A.D., and Stolfo, S., 2003. One class support vector machines for detecting anomalous windows registry accesses. In: ICDM Workshop on Data Mining for Computer Security. Doi: 10.7916/D85M6CFF

30. Martinel, N., Micheloni, C., and Foresti, G.L., 2015. Saliency Weighted Features for Person Re-identification. pp 191–208. Doi: 10.1007/978-3-319-16199-0_14

31. Sun, Y., Zheng, L., Deng, W., and Wang, S., 2017. SVDNet for Pedestrian Retrieval. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 3820–3828. Doi: 10.1109/ICCV.2017.410

32. Zheng, Z., Zheng, L., and Yang, Y., 2019. Pedestrian Alignment Network for Large-scale Person Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(10), pp.3037–3045. Doi: 10.1109/TCSVT.2018.2873599

33. Yu, Y., Liang, C., Ruan, W., and Jiang, L., 2020. Crowdsourcing-Based Ranking Aggregation for Person Re-Identification. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 1933–1937. Doi: 10.1109/ICASSP40776.2020.9053496

34. Sun, Y., Zheng, L., Yang, Y., Tian, Q., and Wang, S., 2018. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline). pp 501–518. Doi: 10.1007/978-3-030-01225-0_30

35. He, K., Zhang, X., Ren, S., and Sun, J., 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 770–778. Doi: 10.1109/CVPR.2016.90

36. Chen, L., Yang, H., Xu, Q., and Gao, Z., 2021. Harmonious attention network for person re-identification via complementarity between groups and individuals. *Neurocomputing*, 453, pp.766–776. Doi: 10.1016/j.neucom.2020.07.118

37. Mortezaie, Z., Hassanpour, H., and Beghdadi, A., 2021. A Color-Based Re-Ranking Process for People Re-Identification : Paper ID 21. In: 2021 9th European Workshop on Visual Information Processing (EUVIP). IEEE, pp 1–5. Doi: 10.1109/EUVIP50544.2021.9484056

38. Parraga, C.A., and Akbarinia, A., 2016. NICE: A Computational Solution to Close the Gap from Colour Perception to Colour Categorization. *PLOS ONE*, 11(3), pp.e0149538. Doi: 10.1371/journal.pone.0149538

39. Mortezaie, Z., Hassanpour, H., and Beghdadi, A., 2022. People re-identification under occlusion and crowded background. *Multimedia Tools and Applications*, pp.1–21. Doi: 10.1007/s11042-021-11868-y

40. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. Proceedings of the European conference on computer vision (ECCV), pp: 801-818. Doi: 10.1007/978-3-030-01234-2_49

41. Gray, D., and Tao, H., 2008. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. pp 262–275. Doi: 10.1007/978-3-540-88682-2_21

42. Roth, P.M., Hirzer, M., Köstinger, M., Beleznai, C., and Bischof, H., 2014. Mahalanobis Distance Learning for Person Re-identification. In: Person Re-Identification. Springer London, London, pp 247–267. Doi: 10.1007/978-1-4471-6296-4_12

43. Moghaddam, B., Jebara, T., and Pentland, A., 2000. Bayesian face recognition. *Pattern Recognition*, 33(11), pp.1771–1782. Doi: 10.1016/S0031-3203(99)00179-X

44. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., and Bischof, H., 2012. Large scale metric learning from equivalence constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 2288–2295. Doi: 10.1109/CVPR.2012.6247939

45. Alipanahi, B., Biggs, M., and Ghodsi, A., 2008. Distance metric learning vs. fisher discriminant analysis. In: Proceedings of the 23rd national conference on Artificial intelligence. pp 598–603. Doi: 10.5555/1620163.1620164

46. Zhang, K., Zhang, L., and Yang, M.-H., 2012. Real-Time Compressive Tracking. European conference on computer vision (ECCV), pp 864–877. Doi: 10.1007/978-3-642-33712-3_62

47. Ojala, T., Pietikainen, M., and Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp.971–987. Doi: 10.1109/TPAMI.2002.1017623

48. mortezaie, zahra, and Hassanpour, H., 2019. A survey on age invariant face recognition methods. *Jordanian Journal of Computers and Information Technology*, 5(2), pp.87–96. Doi: 10.5455/jjcit.71-1554841475

49. Li, W., Zhao, R., Xiao, T., and Wang, X., 2014. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 152–159. Doi: 10.1109/CVPR.2014.27

## Persian Abstract

### چکیده

سیستم‌های نظارت ویدیوئی به‌طور گسترده در مکان‌های عمومی و خصوصی برای حفظ امنیت و مراقبت‌های بهداشتی استفاده می‌شود. عملکرد سیستم‌های نظارت ویدیوئی به‌طور مستقیم به دقت این سیستم‌ها در بازشناسایی انسان بستگی دارد. در نمای یک دوربین سه ناحیه شامل بدن فرد، پس‌زمینه و اشیاء در حال حمل توسط افراد، وجود دارد. در رویکردهای بازشناسایی موجود، پس‌زمینه تصاویر یا نادیده گرفته می‌شود یا ویژگی‌های استخراجی از آن مانند ویژگی‌های استخراجی از بدن فرد در نظر گرفته می‌شود. در این مقاله، سه ناحیه مورد نظر در بازشناسایی با اهمیت متفاوت در نظر گرفته شده است. در روش پیشنهادی، ابتدا تصویر ورودی با استفاده از یک رویکرد قطعه‌بندی معنایی عمیق، به سه ناحیه تقسیم می‌شود. سپس تاثیر هر ناحیه بر ویژگی‌های استخراجی متناسب با اهمیت ناحبه در بازشناسایی تنظیم می‌شود. روش پیشنهادی، می‌تواند با استفاده از توصیفگرهای قوی، مانند توصیفگرهای گوسی گوسی  و گوسی سلسله مراتبی، روش‌های بازشناسایی موجود را در برابر مسائل چالش‌برانگیز مانند انسداد جزئی ناشی از حمل اشیاء و پس‌زمینه‌های شلوغ بهبود دهد. نتایج تجربی روی برخی از مجموعه داده‌های موجود در زمینه بازشناسایی انسان، تاثیر روش پیشنهادی را در بهبود عملکرد روش‌های بازشناسایی موجود نشان می‌دهد..