Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11868-y

1200: MACHINE VISION THEORY AND APPLICATIONS FOR CYBER PHYSICAL SYSTEMS

# People re-identification under occlusion and crowded background

Zahra Mortezaie[1] · Hamid Hassanpour[1] · Azeddine Beghdadi[2]

Received: 31 January 2021 / Revised: 28 October 2021 / Accepted: 23 December 2021
Published online: 19 March 2022


## Abstract

The performance of video surveillance systems with network cameras depends on their accuracy in people re-identification. Body occlusion, crowded background, and variations in scene illumination and pose are challenging issues in people re-identification. In this paper, a technique is proposed to improve the performance of re-identification approaches using (a) a pre-processing step; and (b) a proposed weighing mechanism. In this approach, first, the input image is segmented into the person's body, background, and possible carried objects. Then, considering the image's segments, the occluded parts of the body are retrieved using their neighboring pixels. The processed image is transformed into the log chromatically color space which is robust to scene illumination changes. Using the transformed images along with descriptors which are robust to appearance changes such as Gaussian of Gaussian (GOG) and Hierarchical Gaussian Descriptor (HGD) can improve performance of the descriptors. In this paper, the GOG and HGD are used in a weighed form to represent the pre-processed images considering the importance of each segment of the images in people re-identification. The proposed re-identification system is evaluated using VIPeR and PRID450s datasets, where it respectively achieves 61.9% and 83.4% rank-1 matching rates. Experimental results show that our proposed approach outperforms other existing approaches in people re-identification.

**Keywords** Re-identification · Carried objects · Occlusion · Crowded background · Appearance characteristics

✉ Hamid Hassanpour
h.hassanpour@shahroodut.ac.ir

Zahra Mortezaie
zm.mortezaie@gmail.com

Azeddine Beghdadi
azeddine.beghdadi@univ-paris13.fr

Extended author information available on the last page of the article

## 1 Introduction

Human behavior is often analyzed in public and private places using video surveillance systems. The current trend is to move more towards intelligent systems, taking into account both acquisition conditions and complex scenarios by using artificial intelligence. These systems involve multiple cameras with overlapping or non-overlapping fields of view [7]. One of the first difficult problems to solve in video surveillance is the detection and estimation of the position of objects of visual interest in the scenes filmed by the camera(s) with non-overlapping fields of view [51, 50]. Another equally important aspect that is the subject of this paper is the person re-identification (Re-Id). The accuracy of Re-Id directly affects the performance of video surveillance system. Re-Id systems assign the same label to the images of the same person in different camera views by analyzing their appearance. However, the appearance may change across cameras and threaten the accuracy of re-identification systems. There are challenging issues for people tracking using multi-camera network, including variations in scene illumination, changes of the background, pose variations, the absence of a previously carried object, and occlusion. Hence, it is important to improve the performance of re-identification systems considering the issue of appearance changes.

The above-mentioned issues are depicted in Fig. 1 using a number of data samples from the VIPeR dataset [10]. A person at two different camera views is shown in each column of the figure. In these samples the background of the same person changes in the field of view of different cameras. Also, variations in person's pose across different cameras cause appearance changes as some parts of the person's body are occluded via carried objects. Besides, in Samples 7 to 9 variations in scene illumination affect the appearance characteristics.

Many approaches have been introduced in literature to boost performance of re-identification systems against the challenging issues mentioned earlier. These approaches can mainly be grouped into two categories. In the first category, the approaches do not consider the dramatic effects of body occlusion and crowded


**Fig. 1** Data samples from the VIPeR dataset representing the appearance changes in different cameras' view. Each column represents a person in two different camera views

background on the extracted characteristics. Some of these approaches attempt to improve the performance of re-identification systems using the discriminative appearance features extracted from deep learning-based methods [16, 24, 39, 44], while other approaches consider challenging issues such as appearance changes caused by variations in viewpoint, pose, and scene illumination [5, 11, 21, 25, 26, 31, 35, 43, 46, 47], different cameras conditions [29, 30], inconsistency of data distributions across camera views [45], as well as the issue of the lack of labeled training samples in person re-identification [15, 18, 48]. In the second category, the approaches stabilize re-identification systems against the issues of partial occlusion [14, 27, 49] and crowded background [36, 40]. These approaches do not consider both of the mentioned issues simultaneously in feature extraction.

Despite the importance of the issues discussed above (specially the issue of occlusion due to carried objects) in people re-identification, they were not implicitly considered in the existing approaches. Hence, in this paper a people re-identification approach is proposed using descriptors robust to appearance changes. In this approach, the input image is semantically segmented into three parts: the person's body, background and possible carried objects. Then, occluded parts of the person's body caused by the carried object are retrieved using a pre-processing step. We name this pre-processing step as unification process, since occluded parts of the body are unified with the color of their neighboring pixels. Also, the log chromatically color space [8] is used in order to stabilize the hand-crafted descriptors against variations in scene illumination. The proposed approach has two contributions to person re-identification as follows:

a) A part-based weighing mechanism is proposed, where the low-level features are extracted from each segment of the pre-processed image proportional to their importance in person re-identification. It will be addressed that the three segments may not have the same importance in re-identification.
b) Two commonly used descriptors i.e., Gaussian of Gaussian (GOG) [25] and Hierarchical Gaussian Descriptor (HGD) [26] are boosted against appearance changes caused by carried objects, crowded background, and also, scene illumination changes using unification process, weighing mechanism and also, the log chromatically color space respectively.

Also, experimental results demonstrate that our proposed approach outperforms the other existing approaches in people re-identification.

The rest of this paper is structured as follows: Some existing re-identification methods are reviewed briefly in Section 2. In Section 3, the proposed re-identification approach is introduced. Finally, the experimental results and conclusion are presented in Sections 4 and 5 respectively.

## 2 Related works

As mentioned above, the existing re-identification approaches are mainly grouped into two categories: the approaches that do not consider the dramatic effects of body occlusion and crowded background in feature extraction [5, 11, 15, 16, 18, 21, 24–26, 29–31, 35, 39, 43–48], and those approaches that stabilize re-identification

systems against the issues of partial occlusion and crowded background [14, 27, 36, 40, 49]. In this section, some of these re-identification approaches are reviewed briefly.

Some of the re-identification approaches use the Convolutional Neural Networks (CNNs) to extract discriminative appearance characteristics for people re-identification. In [24], a CNN is fine-tuned for improving the discriminative ability of appearance features extracted from the CNN. In [44], a multi-stage Region of Interest (ROI) pooling approach is used to guide the feature learning process in a CNN and further, to extract features from regions of person's body. In this method, the extracted features from each region are pooled out individually at different steps. Also, a tree-structured fusion network is used to fuse the learned features in a competitive manner. In [16], a parallel spatio-temporal attention model is proposed for extracting both the spatial and temporal characteristics simultaneously from the CNNs without losing space information. For learning discriminative characteristics from different body parts, a Part Loss Networks (PL-Net) module was proposed in [39], where the part loss and global classification loss are minimized simultaneously. In this module, first, a number of body parts are generated considering person part loss. Then, for each generated part, the classification loss is optimized individually. Many re-identification approaches try to face the issue of appearance changes caused by variations in viewpoint, pose, and scene illumination [5, 11, 21, 25, 26, 31, 35, 43, 46, 47]. A Graph Correspondence Transfer (GCT) method was proposed in [46] to deal with the issues of viewpoint and pose variation. In this approach, first, a patch-wise graph matching mechanism is used for learning a set of patch-wise correspondence templates from positive image pairs with various pose-pair configurations. Then, some training pairs with the most similar pose-pair configurations are selected as references for each pair of test images. The correspondences of the references are further transferred to the test pair to determine the similarity between images. Also, for improving the correspondence transfer used in [46], a pose context descriptor based on the topology structure of the estimated joint locations [4], is used in [47]. Also, to handle the issue of occlusion and pose variations, in [5], the HSV and Scale Invariant Local Ternary Pattern (SILTP) [20] features are extracted in the local form by dividing the images into sub-regions in horizontal and vertical directions. A Local Maximal Occurrence representation approach namely LOMO is proposed in [21]. In this approach, first, the images are divided into a number of overlapping windows. Then, two scales of SILTP [20] histograms, and an $8 \times 8 \times 8$-bin joint HSV histogram are extracted from each window for representing the local structure of the images. In [43], the color, shape, and texture of the input image are extracted in a weighed form. In this approach, a weight map is obtained using the Markov chain approach [12], where the weight of each pixel of the input image is computed considering its adjacent pixels. In [25, 26], first, the input image is divided into seven overlapping horizontal regions. Meanwhile, each region is considered as a number of overlapping windows. Further, a Gaussian distribution is computed for each window, using some low-level features of the window such as magnitudes of the pixel intensity gradient in different orientations, and the color values in R, G, and B components. In [25, 26], the obtained Gaussian distributions are mapped into the linear tangent spaces, as the Gaussian distributions are non-linear spaces and the Euclidean operations cannot be directly applied on these spaces. The mapped Gaussian distribution of the windows are further used for computing the Gaussian distribution of the

corresponding region. Also, the Gaussian distributions of the regions are mapped into the tangent spaces. The mapped Gaussian distributions of the regions are concatenated for representing the input image. In [52], based on the color of person body and carried objects, a re-ranking approach is proposed to improve the performance of GOG and HGD approaches. In this approach, using DeepLabv3+ [53], the images are semantically segmented into person's body, background, and carried objects. Then, the colors of person body and carried objects are categorized using the color categorization approach proposed in [54]. Finally, the most frequent colors of the body and carried objects are used to reordering the ranked results obtained from GOG and HGD approaches. The similar feature extraction manner used in [25, 26], namely a Multi-Level Gaussian Descriptor (MLGD), is proposed in [35]. In this approach, using some different color and texture information for depicting each pixel of the images, such as color moment values of RGB components and Schmid filter responses, makes this descriptor a little different from [25, 26]. In the re-identification approach proposed in [31], the extracted features from LOMO descriptor and also, the extracted features from CNN are combined to improve the performance of people re-identification systems. In [11], a density-adaptive smooth kernel re-ranking approach is proposed, where a smooth kernel function is applied to formulate the relation between data samples, using a density-adaptive parameter. This approach uses LOMO and GOG descriptors to represent images. In [29], for dealing with the issue of appearance changes caused by different camera conditions, the Kernel Cross-View Collaborative Representation based Classification (Kernel X-CRC) approach is proposed, where images are represented via cross-view discriminative information. In this approach, GOG descriptor is used for extracting images' appearance characteristics. Then, the extracted characteristics are mapped into the learned subspaces. Meanwhile, for computing the similarity between the images, the mapped features are further passed through the Kernel X-CRC. Also, a nonlinear regression model, namely Kernel Multiblock Partial Least Squares (Kernel MBPLS) was proposed in [30], for mapping data samples into a low-dimensional subspace considering multiple sources of data. Similar to [29], for re-identifying people, the extracted characteristics from GOG are mapped into the learned subspaces and are then used as the input of the Kernel X-CRC. A similarity learning approach was proposed in [45] for boosting re-identification systems against the issue of data distributions inconsistency between camera views. In this approach, by considering people re-identification as a consistent iterative multi-view joint transfer learning optimal problem, the optimal problem is solved using the Inexact Augmented Lagrange Multiplier (IALM) algorithm [37]. In order to improve the performance of re-identification systems, the issue associated with the lack of labeled training samples is considered in [15, 18, 48]. In [48], a body symmetry and part-locality-guided Direct Nonparametric Deep Feature Enhancement (DNDFE) module is proposed to handle the issue of the lack of training samples used in deep feature learning approaches. The DNDFE module involves two nonparametric layers including: the body symmetry average pooling and local normalization layers. This module is embedded between the deep feature learning and similarity learning modules. Also, in [18] a semi-supervised co-metric learning approach is proposed, where a discriminative Mahalanobis-like distance matrix is learned using a few annotated training samples. In this approach, first, both the hand-crafted features, i.e. salient color names-based color descriptor (SCNCD) [38], and the features extracted from Siamese convolutional

neural network [3] are used to represent single-view person images. Meanwhile, a binary-weight learning approach is further used to decompose single-view person features into pseudo binary views. The decomposed features are used for learning metric models. Also, the metric models are jointly updated using both the pseudo-labels and references for obtaining discriminative metrics. Besides, a View-Specific Semi-supervised Subspace Learning (VS-SSL) approach is proposed in [15] in order to learn specific projections for each camera view from a limited number of annotated training samples. Similar to [18, 48], this approach attempts to reduce the number of labeled training samples in person re-identification.

Meanwhile, the issue of partial occlusion in people re-identification is considered in some researches based on deep learning such as [14, 49]. In [14, 49], a number of occluded sample images are generated from original sample images by adding sliding windows to the person's body. Then, in order to train a re-identification model boosted against the occluded regions, both the original sample images and occluded images are used to train ResNet-50 [13]. Also, in the feature extraction approach proposed in [27], the attention maps guide the feature extractor model to ignore occluded regions, where the attention maps are generated using the extracted human landmarks. Wang et al. tried to handle the issue of occlusion in re-identification [36], where a framework is provided to learn high-order relation and topology information. In this approach, a CNN and a key-points estimation model are used in order to extract semantic local features. Then, an Adaptive Direction Graph Convolutional (ADGC) layer is introduced considering the local features as nodes of a graph in order to pass relation information between nodes. In this approach, aligning two groups of local features from two images is considered as a graph matching problem. Hence, in [36] topology information is jointly learnt and embedded into local features using a Cross-Graph Embedded Alignment (CGEA) layer. Besides, in [40], a Semantic aware Occlusion-Robust Network (SORN) is proposed in order to face the issue of occlusion in person re-identification. SORN involves three branches including: a local branch, a global branch, and a semantic branch. The local branch is used to extract part-based features from images. In the global branch, a Spatial-Patch Contrastive loss (SPC) is used for extracting occlusion-robust global features. Meanwhile, the non-occluded parts of the body are obtained using the semantic branch. In this approach, both

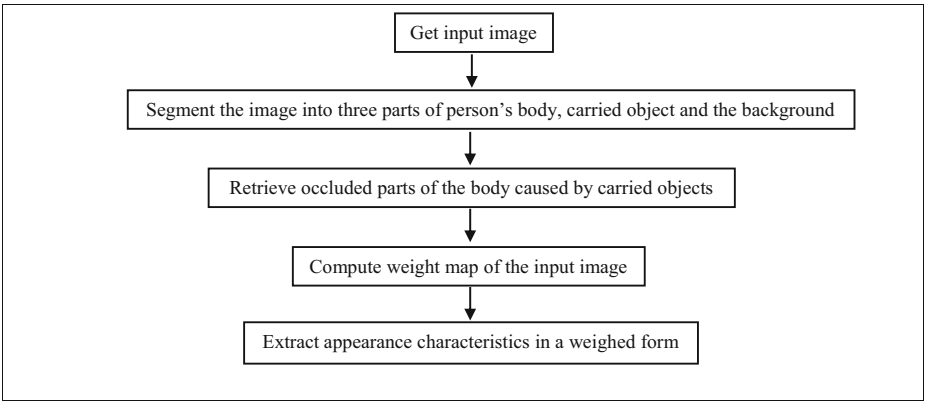


**Fig. 2** The main steps of the proposed approach applied on each image

the local features extracted from the non-occluded parts and the global features extracted from the whole image are used for matching between images.

In [34] the issue of crowded background in person re-identification, is considered in terms of the background bias issue. This approach attempts to stabilize re-identification systems against background variations by proposing a deep neural network based on human parsing for learning more discriminative features from foreground regions. In this approach, the background and foreground regions are separated using a human parsing approach. Then, the network is forced to focus on informative regions of the input image using the separated regions. Besides, a patch selection approach based on parsing and saliency detection is proposed in [22], for eliminating the patches corresponding to the background. This approach first uses Deep Decompositional Network (DNN) to semantically segment images. Then, both the sliding window and color matching approaches are used to eliminate the background patches from the segmented images. In this approach, appropriate patches are selected from the remaining patches using saliency detection approach. Also, some hand-crafted features such as Pyramid Histogram of Oriented Gradients (PHOG [42]), HSV and Scale Invariant Feature Transform (SIFT [2]) features are further extracted from the selected patches. Then, the local extracted features are combined with the global feature extracted using LOMO. In [41], a Relation-Aware Global Attention (RGA) module is proposed to learn discriminative features from human body regions, where the structural information (i.e., clustering-like information) is found by investigating the global scope relations. RGA models the pairwise relations between feature vectors. The relations are considered as a vector, where it depicts the global structural information. The appearance features and the global scope relations are further used for determining the importance of features from a global view.

The above-mentioned approaches do not consider both the issues of crowded background and the issue of occluded regions caused by carried objects. Also, to reduce the dramatic effects of carried objects on visual characteristics, the existing approaches do not attempt to retrieve and simulate the body occlusion. Hence, in this paper we propose a people re-identification approach which is robust to appearance changes caused by carried objects, crowded background, and also, illumination variations. To achieve this goal, a unification pre-processing step and a weighing mechanism are proposed. Using the proposed approach along with a color space which is independent of illumination changes, improves the performance of two commonly used hand-crafted descriptors i.e., GOG and HGD.

## 3 Proposed method

As mentioned before, carried objects may occlude some parts of person's body and change the person's appearance. Hence, they disrupt the distinctive characteristics of the appearance. To deal with this issue, in our proposed re-identification approach, the input image is manually segmented into three parts of person's body, carried object and the background. The segmented image involves three values as 1, 0.5, and 0 for representing body, carried objects, and the background respectively. Considering the segmented image, a pre-processing step namely unification process is used for retrieving occluded parts of the body caused by carried objects. In addition to the unification process, a weighing mechanism is proposed in this paper to indicate the significance of each segment in re-identification. Appropriate features are

extracted from each of the three segments, and they are incorporated for the re-identification depending to their significance value in the weighing mechanism. Figure 2 represents the main steps of the proposed approach applied on each image of the gallery and probe sets:

In Sub-sections 3.1 and 3.2, the details of the unification process and the weighing mechanism are described respectively. Also, in Sub-section 3.3, we show how the pre-processing step and the weighing mechanism are used for re-identification.

### 3.1 Pre-processing step

In pre-processing step, pixels of person's body occluded by carried objects are retrieved considering the colors of their neighboring pixels. Assuming that the input image matrix ($I$) and its segmented image i.e., ($I_s$) involve $m$ rows and $n$ columns, the following steps are taken to obtain the retrieved image ($I_u$), i.e., output of the unification process:

1) $r = 1$; $I_u = I$;

2) Put row $r$ of the matrix ($I$) and its corresponding $I_s$ into vectors $X$ and $Y$ respectively as follows:

$$X(j) = I(r, j), \qquad j = 1, 2, 3, \dots, n;$$

$$Y(j) = I_s(r, j), \qquad j = 1, 2, 3, \dots, n;$$

3) If vector $Y$ contains both the carried object label (i.e., 0.5) and the person's body label (i.e., 1), then:

3-1) Put the values of vector $X$ which are related to the person's body into vector $U$ with the size of $1 \times T$, where $T$ is the number of the body pixels in vector $X$:

$$Y(j) = 1 \rightarrow U(i) = X(j), \qquad j = 1,2,3, \dots, n \; and \; i = 1, 2, 3, \dots, T;$$

3-2) Compute the number of the carried object pixels in vector $X$:

$$counter = 0;$$

$$Y(j) = 0.5 \rightarrow counter = counter + 1, \qquad j = 1,2,3, \dots, n;$$

3-3) If $counter > T$, then:

$$\widehat{U} = upsample(U),$$

where, the size of vector $\widehat{U}$ is $1 \times counter$, and $T = counter$;

else $\widehat{U} = U$;

3-4) In row $r$ of $I_u$, substitute the pixels' values of the carried object with the pixels in vector $\widehat{U}$:

$$Y(j) = 0.5 \rightarrow I_u(r, j) = \widehat{U}(i), \qquad j = 1, 2, 3, \dots, n \; and \; i = 1, 2, 3, \dots, T;$$

4) $r = r + 1$; if $r \leq m$, then go to Step (2);

5) Finish.

Note that in the unification process, if the carried object mask in row $r$ has no horizontal neighbor, the above-mentioned steps will be applied on the columns of the image ($I$) and its corresponding $I_s$.

A number of data samples from the VIPeR dataset along with their corresponding manually segmented image ($I_s$) and results of the unification process ($I_u$) are respectively shown in rows 1 to 3 of Fig. 3. In this figure, each column refers to a data sample. As shown in this figure, the


**Fig. 3** Data samples from the VIPeR database, their corresponding segmented image $I_s$, the unified samples, and the proposed weight map ($W$)

masking effect of the carried object is reduced from the appearance using the unification process. Indeed, the occluded pixels of the body are properly simulated by substituting them with their neighboring pixels.

### 3.2 Weighing mechanism

In the proposed re-identification approach, a weight map $W$ with the same size as the input image is employed considering the segmented image $I_s$. Pixels' values in the weight map are in the interval (0–1], representing significance of the corresponding pixel in the re-identification. The person's appearance involves the most important characteristics for re-identification. Hence, the maximum value in the interval (0–1], i.e., 1, is assigned to those pixels of $W$ which are related to the person's body. However, the retrieved regions may not be accurate, as the unification process approximately simulates the occluded parts of the body using their neighboring pixels. Hence, the retrieved parts should not have the same importance as other parts of the body in re-identification. Meanwhile, the carried objects that do not occlude the body, can be used as an auxiliary characteristic for re-identification, where the carried object is obvious in at least a few of the cameras' view. But we should consider the dramatic effects of the absence of a previously carried objects on the extracted characteristics. Besides,

background pixels which are closer to the person's body should be more important than the other ones.

Considering the above-mentioned points, in our re-identification technique, assuming that the person's body is located in the center of the image, we consider retrieved parts of the person's body, carried objects, and background with less importance compared to actual pixels associated with the person's body, where the closer pixels to the center of the image have more importance than the other ones. To achieve this goal, first, we compute a distance matrix ($D$) with the same size as the input image, where each pixel of this matrix denotes the Euclidian distance of the corresponding pixel in the image from the center of the image. Assuming that segment $S$ is one of the corresponding segments, (i.e., retrieved regions, carried object, and background) of the image and it contains $K$ pixels, the distance matrix ($D$) for this segment is computed as follows:

$$\hat{y}_k = |y_{center} - y_k| \tag{1}$$

$$\hat{x}_k = |x_{center} - x_k| \tag{2}$$

$$D_k = \sqrt{[2]\hat{y}_k^2 + \hat{x}_k^2} \tag{3}$$

where, $y_{center}$ and $x_{center}$ are the $y$ and $x$ coordinates of the central pixel in the image respectively; $y_k$ and $x_k$ respectively denote the $y$ and $x$ coordinates of pixel $k$ in segment $S$; $D_k$ is the Euclidian distance of the corresponding pixel (i.e., pixel $k$).

Also, the values of $D$ are normalized, with $M$ as its maximum value, in the interval (0, 1] as follow:

$$\hat{D}_k = \frac{D_k}{\mathrm{M}}, k = 1, 2, 3, , K \tag{4}$$

Using $\hat{D}$ and the manually segmented map, we assign $(1 - \hat{D})$ to those pixels of $W$ which are related to segment $S$.

The fourth row in Fig. 3 shows the weight maps (i.e., $W$) of the corresponding data samples from the VIPeR dataset. As can be seen in this figure, the proposed weighing mechanism assigns lower values to pixels of the regions which are far from middle of the image.

### 3.3 Unification process and weighing mechanism in re-identification

As mentioned earlier, the appearance changes leads to miss the person during the tracking task. Hence, it is necessary to use descriptors which are robust to appearance changes due to illumination changes in re-identification. In our proposed people re-identification approach, the characteristics of the unified images are extracted using robust and discriminant descriptors i.e., the GOG [25] and HGD [26]. Also, our proposed weighing mechanism and the log chromatically color space are used to improve the performance of these descriptors.

In GOG and HGD, first, the input image is divided into seven overlapping horizontal regions. The regions are also divided into a number of overlapping windows. Then each pixel of the windows is represented as follow [25, 26]:

$$F_t = [v; D_{0^\circ}; D_{90^\circ}; D_{180^\circ}; D_{270^\circ}; x_R; x_G; x_B]^{\boldsymbol{T}} \tag{5}$$

where, $F_t$ is a feature vector which is used to represent each pixel $t$ in the window. In this feature vector, $v$ represents the pixel location in the vertical direction in the image, $D_{0^\circ}, D_{90^\circ}$

, $D_{180^\circ}$ , $D_{270^\circ}$ are the magnitudes of the pixel intensity gradient in four different orientations, and $x_R$, $x_G$, and $x_B$ respectively denote the color values in R, G, and B channels.

In our proposed re-identification approach, each value of the weight map ($W$) is used to tune the effect of its corresponding raw extracted features, i.e., $F_t$, as follow, where $W_t$ is the weight of the pixel $t$:

$$\hat{F}_t = W_t \times F_t \tag{6}$$

In GOG and HGD, for each window $Q$ with $n$ pixels, a Gaussian distribution is computed using the low-level features (i.e., $F_t$) of its pixels as follow [25, 26]:

$$\mathcal{N}\left(\{\emptyset\,\mu_Q, \Sigma_Q\right) = \frac{\exp\left(-\frac{\infty}{\in}\left(\{-\mu_Q\right)^T \Sigma_Q^{-\infty}\left(\{-\mu_Q\right)\right)}{(\in\pi)^{\frac{\lceil}{\in}}|\Sigma_Q|} \tag{7}$$

where $f$ is the corresponding raw features (i.e., $f = \{F_1, F_2, \ldots F_n\}$) and $\mu_Q$ and $\Sigma_Q$ respectively denote the mean and covariance matrix of the feature vectors associated with window $Q$. Also, $d$ is the size of the feature vector and $|.|$ denotes the determinant operator.

As the Gaussian distribution is nonlinear, the Euclidean operations cannot be directly applied on this distribution. Hence, in GOG and HGD, the Gaussian distributions of the windows are mapped into the linear tangent space using a Riemannian manifold which is a Symmetric Positive Definite (SPD) matrix [19, 33]. However, in HGD some feature normalization approaches are used to decrease the bias of SPD matrix descriptors. In GOG and HGD, following the work in [23] the space of $d$-dimensional multivariate Gaussians is embedded into the space of $(d + 1)$-dimensional SPD matrices, using the mean vector and covariance matrix of each window $Q$ as follow:

$$P_Q = \left|\Sigma_Q\right|^{-\frac{1}{d+1}} \begin{bmatrix} \Sigma_Q + \mu_Q \mu_Q^T & \mu_Q \\ \mu_Q^T & 1 \end{bmatrix} \tag{8}$$

Then, the SPD matrix $P_Q$ in Eq. (8) is mapped to the tangent space using matrix logarithm as follow:

$$\begin{aligned} V_Q &= vec\left(\log\left(P_Q\right)\right) \\ &= \left[\hat{p}_{Q(1,1)}, \sqrt{2}\hat{p}_{Q(1,2)}, , \sqrt{2}\hat{p}_{Q(1,d+1)}, \hat{p}_{Q(2,2)}, \sqrt{2}\hat{p}_{Q(2,3)}, , \hat{p}_{Q(d+1,d+1)}\right]^T \end{aligned} \tag{9}$$

where $V_Q$ denotes a vector with $\frac{\left(d^2+3d\right)}{2}+1$ elements and $log(.)$ is the matrix logarithm operator. As $log(P_Q)$ is a symmetric matrix, using the $vec(.)$ operator in Eq. (9), the upper triangular part of the mapped matrix (i.e., $log(P_Q)$) is obtained as a vector. Note that in this equation, $\hat{p}_{Q(q,r)}$ denotes the element $(q, r)$ of $log(P_Q)$ matrix. Meanwhile, in [25, 26], the Gaussian distribution of each region are then computed using the mapped Gaussian distribution of its corresponding windows. Similar to the windows, the Gaussian distribution of each region is then mapped into a linear tangent space.

Assuming that each region $\mathcal{G}$ is finally described using a vector $z_{\mathcal{G}}$, for each input image, the output of the GOG and HGD descriptors is obtained by concatenating the $z_{\mathcal{G}}$ associated with

each of its seven regions as follow [25, 26]:

$$Z_{RGB} = \left[ z^T_{\mathcal{G}_\infty}, z^T_{\mathcal{G}_\in}, z^T_{\mathcal{G}_\ni}, z^T_{\mathcal{G}_\triangle}, z^T_{\mathcal{G}_\triangledown}, z^T_{\mathcal{G}_\gamma} z^T_{\mathcal{G}_\wr} \right]^T \tag{10}$$

For each input image, four different feature vectors are computed by substituting the RGB information (i.e., $x_R$, $x_G$, and $x_B$) used in Eq. (5) with the components of LAB, HSV, and nRGB color spaces. Assuming that the extracted feature vectors using RGB, LAB, HSV, and nRGB, are respectively named as $Z_{RGB}$, $Z_{LAB}$, $Z_{HSV}$, and $Z_{nRGB}$, the input image is represented by concatenating $Z_{RGB}$, $Z_{LAB}$, $Z_{HSV}$, and $Z_{nRGB}$ as follow [25, 26]:

$$Z_{Fusion} = \left[ Z^T_{RGB}, Z^T_{LAB}, Z^T_{HSV}, Z^T_{nRGB} \right]^T \tag{11}$$

Note that according to [8, 9, 17], the variations in scene illumination can be modeled using a linear transformation namely diagonal model as follow:

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{bmatrix} \times \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix} \tag{12}$$

where, $R_2$, $G_2$, and $B_2$ are the RGB values in illumination condition 2, which are computed using multiplication of a diagonal matrix by RGB values in illumination condition 1 (i.e., $R_1$, $G_1$, and $B_1$). Indeed, the $x$, $y$, and $z$ values of the diagonal matrix simply apply the transformation. Considering the diagonal model for scene illumination changes, one of the color spaces which can be used to stabilize the hand-crafted descriptors against the variations in scene illumination, is log chromatically color space [1]. This color space is obtained from RGB color space as follows:

$$Ł_r = \ln\left(\frac{R}{G}\right)$$

$$Ł_b = \ln\left(\frac{B}{G}\right)$$

According to [1, 17], the log chromatically color space is robust to variations in scene illumination as follow:

$$Ł^2_r = \ln\left(\frac{xR}{yG}\right) = Ł^1_r + \ln\left(\frac{x}{y}\right)$$

$$Ł^2_b = \ln\left(\frac{zB}{yG}\right) = Ł^1_b + \ln\left(\frac{z}{y}\right)$$

where, $Ł^1_r$, $Ł^1_b$, $Ł^2_r$, and $Ł^2_b$ are the log chromatically components in illumination conditions 1 and 2 respectively. According to Eqs. (15) and (16), mapping illumination changes using

diagonal model, only causes a shift in the coordinates, and the pixels' values of log chromatically color space are independent from scene illumination changes.

Considering the mentioned points about log chromatically color space, we apply this color space on GOG and HGD for improving the performance of these descriptors. To achieve this goal, we substitute the RGB information (i.e., $x_R$, $x_G$, and $x_B$) used in Eq. (5) with the components of log chromatically color space. Assuming that the extracted feature vector using this color space is named as $Z_{LC}$, the input image is represented by concatenating $Z_{RGB}$, $Z_{LAB}$, $Z_{HSV}$, $Z_{nRGB}$, and $Z_{LC}$ as follow:

$$\hat{Z}_{Fusion} = \left[Z_{RGB}^T, Z_{LAB}^T, Z_{HSV}^T, Z_{nRGB}^T, Z_{LC}^T\right]^T \tag{17}$$

In [25, 26] the extracted feature vectors, i.e., $Z_{Fusion}$ are passed throughout a distance metric learning approach proposed in [21], namely XQDA distance metric, to learn a discriminant low dimensional subspace. Assuming N images in the probe set and M images in the gallery set, the output of XQDA is a score matrix $S$, in the size of N × M. Each element of the score matrix denotes the distance between the corresponding images in probe and gallery sets. Finally, by sorting each row of the score matrix, gallery images are ranked proportional to their distance to the corresponding probe image.

## 4 Experimental results

In this section we use the Rank-$k$ measure for evaluating the performance of re-identification approaches. In this measure, $k$ determines the number of top matches with the correct answer [6]. Hence, for $k = 1$, Rank-$k$ is the strictest measure, whereas for $k > 1$, this measure permits some error [28].

In this paper, Rank-$k$ ($k = 1, 5, 10, 20$) is used for comparing the results of our proposed re-identification approach on the VIPeR [10] and PRID450s [32] datasets with the reported results in the existing re-identification approaches. More details about the obtained results on VIPeR and PRID450s will be discussed in Sub-sections 4.1 to 4.3.

### 4.1 Experimental results on viper

The VIPeR dataset involves 1264 images of 632 persons, captured in two different camera views. This dataset contains one image from a person in each camera view. Hence, the performance of our proposed approach is evaluated on VIPeR with single-shot matching. This dataset is a challenging dataset as it contains the low-resolution images with pose, orientation and scene illumination changes. Most of the images in VIPeR suffer from crowded backgrounds and occluded body parts by carried objects.

In Table 1, the reported results from the classic GOG [25] and HGD [26] methods on VIPeR, as well as, the obtained results using our proposed re-identification approach applied on GOG (i.e., EGOG) and HGD (i.e., EHGD) methods are shown. The obtained results using the proposed approach in this table are respectively obtained using:

(I) only the unification process (i.e., *UI*),
(II) unification process along with the weight map (i.e., $UI + W$), and.

**Table 1** Performance of our proposed approach and the methods proposed in [25, 26] on VIPeR

| | | Re-identification approach | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Rank | Classic GOG | EGOG using: | | | Classic HGD | EHGD using: | | |
| | | *UI* | *UI+W* | *UI+W+*log *_chrom* | | *UI* | *UI+W* | *UI+W+*log *_chrom* |
| 1% | 49.7 | 53.4 | 59.4 | **61.9** | 50.0 | 54.8 | 61.5 | **61.8** |
| 5% | 79.7 | 82.7 | 86.0 | **86.2** | 79.5 | 83.2 | 86.7 | **86.7** |
| 10% | 88.7 | 90.3 | 92.0 | **92.7** | 88.9 | 90.4 | 92.5 | **92.8** |
| 20% | 94.5 | 95.5 | 96.5 | **97.4** | 94.6 | 95.9 | 96.6 | **96.8** |

(III) unification process, weight map, and log chromatically color space (*UI* + *W* + log *_chrom*).

Note that, 10 different training and test sets of data samples were used for training and testing the XQDA distance metric used in GOG and HGD. Hence, the average of the obtained accuracy using the test sets were reported in Table 1.

As shown in Table 1, applying the unification process, weighing mechanism, and also, the log chromatically color space on GOG and HGD improves the performance of the descriptors on VIPeR dataset. The GOG and HGD descriptors extract appearance characteristics in a local form considering some raw features such as: pixels location in the vertical direction, the magnitudes of the pixels intensity gradient in four different orientations, and the color values in RGB, LAB, HSV, and nRGB color spaces. The mentioned raw features respectively stabilize the GOG and HGD descriptors against variations in pose, orientation, and scene illumination. But GOG and HGD do not consider the dramatic effects of crowded backgrounds and occluded body parts by carried objects on the extracted characteristics. To deal with these issues, in our proposed approach we try to improve the performance of GOG and HGD descriptors using unification process, weighing mechanism, and the log chromatically color space. Indeed, in the proposed approach, retrieving occluded parts of the body by carried objects, weighing each region of the image proportional to their importance, and also, using the log chromatically color space, make GOG and HGD descriptors robust against crowded backgrounds, occluded body parts, as well as scene illumination changes.

### 4.2 Experimental results on PRID450S

The PRID450s dataset involves 900 images of 450 persons, captured in two different camera views. Similar to VIPeR, this dataset contains one image from a person in each camera view. Hence, we evaluate the performance of our proposed approach on PRID450s with single-shot matching. Compared with VIPeR dataset, the images of PRID450s have simpler backgrounds. Also, PRID450s contains fewer images with occluded body parts in comparison with VIPeR.

The reported results from the classic GOG and HGD methods on PRID450s, as well as, the obtained results using our proposed re-identification approach applied on GOG (i.e., EGOG) and HGD (i.e., EHGD) methods are shown in Table 2. In this table, similar to Table 1, the obtained results using the proposed approach are respectively obtained using:

(I) only the unification process (i.e., *UI*),

**Table 2** Performance of our proposed approach and the methods proposed in [25, 26] on PRID450s

| | | Re-identification approach | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Rank | Classic GOG | EGOG using: | | | Classic HGD | EHGD using: | | |
| | | *UI* | *UI*+*W* | *UI*+*W*+log _*chrom* | | *UI* | *UI*+*W* | *UI*+*W*+log _*chrom* |
| 1% | 68.0 | 68.3 | 80.4 | **80.9** | 70.4 | 70.7 | 82.9 | **83.4** |
| 5% | 88.7 | 88.6 | 94.8 | **94.9** | 91.2 | 90.1 | **95.6** | 95.2 |
| 10% | 94.4 | 94.1 | 97.6 | **97.6** | 94.8 | 95.1 | 97.7 | **97.8** |
| 20% | 97.6 | 97.2 | 99.1 | **99.3** | 97.6 | 97.7 | 98.8 | **99.0** |

(II) unification process along with the weight map (i.e., $UI + W$), and.

(III) unification process, weight map, and log chromatically color space ($UI + W + \log\_chrom$).

Also, for training and testing the XQDA distance metric on PRID450s dataset, 10 different training and test sets of data samples were used. Hence, as in Table 1, the average of the obtained accuracy using the test sets were reported in Table 2.

As it is shown in Table 2, the performance of GOG and HGD on PRID450s dataset is improved by applying the unification process, weighing mechanism, and also, the log chromatically color space on these descriptors.

Note that, applying the unification process on VIPeR dataset induces better results comparing with PRID450s dataset, as the images in PRID450s dataset involves less body occlusion than the images of VIPeR dataset.

Also, images of PRID450s mostly involve carried objects that do not occlude the person's body and, also, the persons' appearance are mostly similar. In this situation, these kind of carried objects can be used as an auxiliary characteristic for re-identification. Hence, applying weighing mechanism induces more improvement to the accuracy of people re-identification on PRID450s dataset comparing with VIPeR dataset. The mentioned points are illustrated in Fig. 4 using a number of data samples from the PRID450s dataset, where, each column refers to a data sample. In these samples, the bodies' characteristics are mainly similar to each other and also, the objects do not occlude the person's body. Hence, the characteristics of the carried objects should be used to distinguish between the persons.

Besides, the VIPeR dataset suffers from illumination changes more than PRID450s dataset. Hence, using log chromatically color for stabilizing descriptors against illumination changes, mainly causes more improvement on VIPeR dataset comparing with PRID450s.


**Fig. 4** Data samples from the PRID450s dataset

**Table 3** Performance of our proposed approach and the some of the existing people re-identification approaches

| Approaches | Ranks on VIPeR | | | | Ranks on PRID450s | | | |
|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 20% | 1% | 5% | 10% | 20% |
| Matsukawa et al. [24], (2016) | 52.1 | 79.6 | 89.2 | 95.0 | 71.5 | 90.6 | 94.7 | 97.5 |
| Leng [18], (2018) | 32.9 | 60.3 | 73 | – | 38.2 | 67.1 | 76.0 | – |
| Vishwakarma et al. [35], (2018) | 47.5 | – | 87.9 | 93.7 | 62.4 | – | 93.5 | 96.9 |
| Tian et al. [34], (2018) | 51.9 | 74.4 | 84.8 | 90.2 | – | – | – | – |
| Zhou et al. [46], (2018) | 49.4 | 77.6 | 87.2 | 94.0 | 58.4 | 77.6 | 84.3 | 89.8 |
| Zhou et al. [47], (2019) | 53.5 | 81.3 | 89.1 | 94.7 | 70.9 | 89.1 | 93.5 | 96.5 |
| Ren et al. [31], (2019) | 42.1 | 64.0 | 73.4 | – | 60.6 | 82.8 | 90.8 | – |
| Chu et al. [5], (2019) | 49.0 | 74.1 | 84.4 | 93.1 | – | – | – | – |
| Prates et al. [11], (2019) | 51.6 | 80.5 | 89.5 | 95.2 | 71.3 | 91.7 | 96.0 | 98.1 |
| Yao et al. [39], (2019) | 56.6 | 82.6 | 89.9 | – | – | – | – | – |
| Prates et al. [30] (2019) | 51.2 | 79.9 | 89.9 | – | 68.1 | 90.7 | 95.0 | – |
| Zhu et al. [48], (2020) | 55.3 | – | 90.5 | 95.3 | – | – | – | – |
| Jia et al. [15] (2020) | 43.9 | 72.2 | 80.9 | 87.8 | 68.2 | 90.2 | 94.9 | 98.0 |
| Guo et al. [11] (2020) | 55.9 | – | 91.9 | 96.9 | 75.3 | – | 97.4 | 99.0 |
| Liu et al. [22] (2020) | 56.8 | – | 92.0 | **97.3** | 72.5 | – | 96.4 | 98.7 |
| Zhao et al. [45] (2020) | 56.3 | 83.0 | 90.0 | 95.8 | 72.1 | – | 94.6 | – |
| EGOG | **61.9** | **86.2** | **92.7** | **97.4** | **80.9** | **94.9** | **97.6** | **99.3** |
| EHGD | **61.8** | **86.7** | **92.8** | 96.8 | **83.4** | **95.6** | **97.8** | **99.0** |

### 4.3 Comparison with existing re-identification approaches

In this Sub-section, the performance of the proposed approach is compared with some of the existing re-identification approaches introduced in [5, 11, 15, 18, 22, 24, 29–31, 34, 35, 39, 45–48]. Hence, the best results obtained using our proposed re-identification approach are compared in Table 3 with the best results reported in other existing people re-identification approaches on VIPeR and PRID450s datasets. In this table, the bolded results denote the first (the under lined number) and the second most accurate results in each rank respectively.

As shown in Table 3, the proposed re-identification approach is more accurate in all ranks on VIPeR and PRID450s datasets, comparing with the existing re-identification approaches. Note that, the compared approaches do not consider both the issues of crowded backgrounds and the occluded regions caused by carried objects. Meanwhile, the existing approaches do not attempt to retrieve the body occlusion to reduce the dramatic effects of carried objects on visual characteristics.

Note that, the approaches introduced in [5, 11, 15, 22, 29, 30, 35, 45–47] use hand-crafted features for person re-identification. These approaches mainly focus on appearance changes caused by viewpoint, pose variations, and different camera conditions. In [46], a patch-wise graph matching mechanism is used to deal with the issues of viewpoint and pose variation. Zhou et al. used a pose context descriptor based on the topology structure of the estimated joint locations to handle the issues of viewpoint and pose variation [47]. To handle the issue of occlusion and pose variations, Chu et al. applied the HSV and SILTP features in the local form [5]. Also, some re-identification approaches are based on GOG and HGD descriptors [11, 29, 30, 35]. Hence, similar to GOG and HGD, these approaches are robust against some issues such as variations in pose, orientation, and scene illumination. In [35], MLGD descriptor is introduced with the similar feature

extraction manner used in GOG and HGD. Using color moment values of RGB components and Schmid filter responses, makes MLGD a little different from GOG and HGD. For representing images, the LOMO, GOG, and HGD descriptors are used in [11], where, a density-adaptive smooth kernel re-ranking approach is proposed to formulate the relation between data samples. Meanwhile, the GOG descriptor is used in [29], where, Kernel X-CRC approach is introduced to map the features extracted form GOG into the learned subspaces. Using Kernel X-CRC approach, the images are represented via cross-view discriminative information in order to handle the issue of appearance changes caused by different camera conditions. Also, based on GOG descriptor, a nonlinear regression model, i.e., Kernel MBPLS is introduced in [30]. Kernel MBPLS maps the extracted features into a low-dimensional subspace considering multiple sources of data. The mapped features are then used as the input of the Kernel X-CRC.

In [45], in order to stabilize re-identification systems against the issue of data distributions inconsistency between camera views, people re-identification is considered as a consistent iterative multi-view joint transfer learning optimal problem. Also, the lack of labeled training samples issue is focused in [15, 18, 48]. The VS-SSL approach introduced in [15] tries to learn specific projections for each camera view using limited number of annotated training samples. In [48], the DNDFE module is proposed, where, two nonparametric layers (i.e., the body symmetry average pooling and local normalization layers) are embedded between the deep feature learning and similarity learning modules. Leng tries to learn a discriminative Mahalanobis-like distance matrix using a few annotated training samples [18]. In this approach, for representing single-view images, the salient color name, and the features extracted from Siamese convolutional neural network are used. These features are then decomposed into pseudo binary views in order to use in learning metric models. Meanwhile, in [24, 31, 39], CNNs are applied to only extract discriminative appearance characteristics for people re-identification.

Also, the issue of crowded backgrounds is addressed in [22, 34]. Tian et al. used a human parsing approach for separating the background and foreground regions [34]. Then, a deep neural network is forced to focus on informative regions of the input image using the separated regions. Besides, for eliminating the patches corresponding to the background, Liu et al. used DNN in order to semantically segment images. Then, the background patches are eliminated from the segmented images using the sliding window and color matching approaches. Also, a saliency detection approach is used to select salient patches from the remaining patches. Finally, some hand-crafted features such as PHOG, HSV and SIFT are extracted from the selected patches. Meanwhile, the extracted features are then combined with the global feature extracted using LOMO.

Despite the efforts of the above-mentioned approaches to improve the performance of re-identification systems, these approaches do not address both the issues of crowded backgrounds and occlusion caused by carried objects in people re-identification. Note that, in our proposed re-identification approach, robust and discriminant descriptors i.e., the GOG and HGD are used in a weighed form to extract characteristics of the unified images. Indeed, in proposed approach, GOG and HGD descriptors are stabilized against the issues of occlusion, crowded background, and illumination changes, using unification process, weighing mechanism, and the log chromatically color space respectively. Hence, the performance of re-identification approaches [11, 29, 30, 35], which are based on the GOG and HGD, can be improved using our proposed EGOG and EHGD descriptors.

Also, despite using the hand-crafted features in our approach, the accuracy of the proposed re-identification approach is higher than all the comparing approaches especially deep learning-based approaches proposed in [18, 24, 31, 34, 39, 48]. It means that embedding the proposed unification process, weighing mechanism, and log chromatically color space in the deep feature learning modules can improve the performance of these people re-identification approaches.

Assuming that the input image involves $m$ rows and $n$ columns, and the processing overhead of the operations used in Steps (1) to (5) of the unification process is constant as $O(1)$, the processing overhead of the proposed unification process in re-identification systems is $O(n + m)$. Besides, assume that segment $S$ is one of the retrieved regions, carried object, and background of the image and it contains $K$ pixels. Meanwhile, the overhead of the operations used in Eqs. (1) to (4) is constant as $O(1)$, the overhead of the proposed weighing mechanism in re-identification systems is $O(K)$. Also, the overhead of using the log chromatically color space in re-identification approaches, directly depends on the algorithm used in each approach. Consequently, applying our proposed unification process and also, the weighing mechanism on existing re-identification approaches imposes a little overhead in these approaches.

Note that, the accuracy of both the unification process and the proposed weighing mechanism depends on the accuracy of segmentation approach used for extracting person's body, carried objects, and background. Besides, resolution of the images used in re-identification systems is mainly low. Also, the existing semantic segmentation approaches cannot segment these images with an appropriate accuracy. Hence, in this paper we use the manually segmented images of the VIPeR and PRID450s datasets. For making our proposed re-identification approach more applicable on other datasets, a semantic segmentation approach can be introduced in future works to automatically segment the low-resolution images with a high accuracy.

## 5 Conclusion

The performance of video surveillance systems directly depends on the accuracy of people re-identification. People re-identification approaches face challenges such as variations in person's appearance across the cameras network caused by body occlusion, crowded background, and variations in scene illumination and pose. Despite the destructive effects of appearance changes on the performance of re-identification systems, many existing re-identification approaches do not consider this issue. To deal with person's appearance changes caused by body occlusion, crowded background, and variations in scene illumination and pose, in this paper a pre-processing step namely unification process, a proposed weighing mechanism, and also, a color space robust to scene illumination changes (i.e., log chromatically color space) are used along with the commonly used descriptors namely GOG and HGD. Experimental results show the superiority of the proposed re-identification approach compared to other existing methods in people re-identification.

**Declarations** The authors have no funding, and potential conflicts of interest (financial or non-financial). Also, the research does not involve human participants, and animals.

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References


1. Berwick D, Lee S (1998) A chromaticity space for specularity, illumination color- and illumination pose-invariant 3-D object recognition. 6th International Conference on Computer Vision. Bombay, pp. 165–170
2. Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In Proceedings of the 6th ACM international conference on image and video retrieval. Amsterdam, pp. 401–408
3. Bromley J, Bentz JW, Bottou L, Guyon I, LeCun Y, Moore C, Sackinger E, Shah R (1993) Signature verification using a "Siamese" time delay neural network. Int J Pattern Recognit Artif Intell 7(4):669–688
4. Cao Z, Simon T, Wei SE, 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. Honolulu, pp 7291–7299
5. Chu H, Qi M, Liu H, Jiang J (2019) Local region partition for person re-identification. Multimed Tools Appl 78:27067–27083
6. Delac K, Grgic M, Grgic CS. (2005) Statistics in face recognition: analyzing probability distributions of PCA, ICA and LDA performance results. In: ISPA 2005. Proceedings of the 4th IEEE International Symposium on Image and Signal Processing and Analysis. Zagreb, pp. 289–294
7. Feizi A (2019) Convolutional gating network for object tracking. Int J Eng 32(7):931–939
8. Finlayson G, Drew M, Funt B (1993) Diagonal transforms suffice for color constancy. 4th International Conference on Computer Vision. Berlin, pp 164–171
9. Finlayson G, Drew M, Funt B (1994) Spectral sharpening: sensor transformations for improved color Constancy. J Optic Soc Am A 11(5):1553–1563
10. Gray D, Tao H (2008) Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European Conference on Computer Vision. Springer, Berlin, pp 262–275
11. Guo R, Li CG, Li Y, Lin J, Guo J (2020) Density-adaptive kernel based efficient reranking approaches for person reidentification. Neurocomputing 411:91–111
12. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. In: Advances in neural information processing systems. Vancouver, pp 545–552
13. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, pp. 770–778
14. Huang H, Li D, Zhang Z, Chen X, Huang K (2018) Adversarially occluded samples for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, pp 5098–5107
15. Jia J, Ruan Q, Jin Y, An G, Ge S (2020) View-specific subspace learning and re-ranking for semi-supervised person re-identification. Pattern Recogn 108:107568
16. Kong J, Teng Z, Jiang M, Huo H (2020) Video-based person re-identification with parallel spatial–temporal attention module. J Electron Imaging 29(1):013001
17. Kviatkovsky I, Adam A, Rivlin E (2012) Color invariants for person reidentification. IEEE Trans Pattern Anal Mach Intell 35(7):1622–1634
18. Leng Q (2018) Co-metric learning for person re-identification. Adv Multimed. https://doi.org/10.1155/2018/3586191
19. Li P, Wang Q, Zhang L (2013) A novel earth mover's distance methodology for image matching with gaussian mixture models. In: Proceedings of the IEEE International Conference on Computer Vision. Sydney, (pp 1689–1696)
20. Liao S, Zhao G, Kellokumpu V, Pietikäinen M, Li SZ (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. San Francisco, pp 1301–1306
21. Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, pp 2197–2206
22. Liu Y, Zhang Y, Coleman S, Bhanu B, Liu S (2020) A new patch selection method based on parsing and saliency detection for person re-identification. Neurocomputing 374:86–99
23. Lovri'c M, Min-Oo M, Ruh EA (2000) Multivariate normal distributions parametrized as a Riemannian symmetric space. J Multivar Anal 74(1):36–48
24. Matsukawa T, Suzuki E (2016) Person re-identification using CNN features learned from combination of attributes. In: 2016 23rd International Conference on Pattern Recognition (ICPR). Cancun, pp 2428–2433

25. Matsukawa T, Okabe T, Suzuki E, Sato Y (2016) Hierarchical gaussian descriptor for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, pp. 1363–1372.
26. Matsukawa T, Okabe T, Suzuki E, Sato Y (2019) Hierarchical gaussian descriptors with application to person re-identification. IEEE Trans Pattern Anal Mach Intell 42(9):2179–2194
27. Miao J, Wu Y, Liu P, Ding Y, Yang Y (2019) Pose-guided feature alignment for occluded person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision. Seoul, pp 542–551
28. Mortezaie Z, Hassanpour H (2019) A survey on age-invariant face recognition methods. Jordan J Comput Inform Technol (JJCIT) 5(02):87–96
29. Prates R, Schwartz WR (2019) Kernel cross-view collaborative representation based classification for person re-identification. J Vis Commun Image Represent 58:304–315
30. Prates R, Schwartz WR (2019) Matching people across surveillance cameras. In: Anais Estendidos da XXXII Conference on Graphics, Patterns and Images. SBC, pp. 84–90
31. Ren QQ, Tian WD, Zhao ZQ (2019) Person re-identification based on feature fusion. In: International Conference on Intelligent Computing. Springer, Cham, pp 65–73
32. Roth PM, Hirzer M, Köstinger M, Beleznai C, Bischof H (2014) Mahalanobis distance learning for person re-identification. In: Advances in computer vision and pattern recognition. Springer, London, pp 247–267
33. Sadatnejad K, Shiry Ghidari S, Rahmati M (2018) A geometry preserving kernel over Riemannian manifolds. J AI Data Min 6(2):321–334
34. Tian M, Yi S, Li H, Li S, Zhang X, Shi J, Yan J, Wang X (2018) Eliminating background-bias for robust person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, pp. 5794–5803
35. Vishwakarma DK, Upadhyay S (2018) A deep structure of person re-identification using multi-level gaussian models. IEEE Trans Multi-Scale Comput Syst 4(4):513–521
36. Wang GA, Yang S, Liu H, Wang Z, Yang Y, Wang S, Yu G, Zhou E, Sun J (2020) High-order information matters: learning relation and topology for occluded person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, pp 6449–6458
37. Xu Y, Fang X, Wu J, Li X, Zhang D (2015) Discriminative transfer subspace learning via low-rank and sparse representation. IEEE Trans Image Process 25(2):850–863
38. Yang Y, Yang J, Yan J, Liao S, Yi D, Li SZ (2014) Salient color names for person re-identification. In: European Conference on Computer Vision. Springer, Cham, pp 536–551
39. Yao H, Zhang S, Hong R, Zhang Y, Xu C, Tian Q (2019) Deep representation learning with part loss for person re-identification. IEEE Trans Image Process 28(6):2860–2871
40. Zhang X, Yan Y, Xue JH, Hua Y, Wang H (2020) Semantic-aware occlusion-robust network for occluded person re-identification. IEEE Trans Circ Syst Video Technol 31(7):2764–2778
41. Zhang Z, Lan C, Zeng W, Jin X, Chen Z (2020) Relation-aware global attention for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, pp. 3186–3195
42. Zhao R, Ouyang W, Wang X (2014) Learning mid-level filters for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, pp. 144–151
43. Zhao R, Oyang W, Wang X (2016) Person re-identification by saliency learning. IEEE Trans Pattern Anal Mach Intell 39(2):356–370
44. Zhao H, Tian M, Sun S, Shao J, Yan J, Yi S, Wang X, Tang X (2017) Spindle net: person re-identification with human body region guided feature decomposition and fusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, pp 1077–1085
45. Zhao C, Wang X, Zuo W, Shen F, Shao L, Miao D (2020) Similarity learning with joint transfer constraints for person re-identification. Pattern Recogn 97:107014
46. Zhou Q, Fan H, Zheng S, Su H, Li X, Wu S, Ling H (2018) Graph correspondence transfer for person re-identification. In: Thirty-second AAAI Conference on Artificial Intelligence (AAAI-18). New Orleans, pp 7599–7606
47. Zhou Q, Fan H, Yang H, Su H, Zheng S, Wu S, Ling H (2019) Robust and efficient graph correspondence transfer for person re-identification. IEEE Trans Image Process 30:1623–1638
48. Zhu J, Zeng H, Huang J, Zhu X, Lei Z, Cai C, Zheng L (2020) Body symmetry and part-locality-guided direct nonparametric deep feature enhancement for person reidentification. IEEE Internet Things J 7(3): 2053–2065
49. Zhuo J, Chen Z, Lai J, Wang G (2018) Occluded person re-identification. In: IEEE International Conference on Multimedia and Expo (ICME). San Diego, pp 1–6

## Affiliations

**Zahra Mortezaie**[1] · **Hamid Hassanpour**[1] · **Azeddine Beghdadi**[2]

[1] Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

[2] Institut Galilée Université Sorbonne Paris Nord, Villetaneuse, France