PERSON RE-IDENTIFICATION USING ASYNCHRONOUS LEARNING METHODS

[1]Giridhar M- P.G Student, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.Email: giridhar33mg@gmail.com

[2]Ruba Soundar K-Associate Professor, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.Email:rubasoundar@gmail.com

Abstract

Both the governmental and private sectors use video surveillance systems for safety and medical reasons. The effectiveness of a surveillance system relies heavily on how well it can re-identify individuals. A camera's field of vision is composed of three zones: the subject's body, the backdrop, and anything the subject could be carrying. Existing methods for re-identification either fail to take context into account or treat it as a physical trait. These three areas are given unequal weight in this paper's re-identification methodology. In our suggested method, we first apply a deep semantic segmentation strategy to the input image to semantically divide it into the three areas. The significance of a given location in the context of re-identification is then used to fine-tune the influence of that region on distinctive human traits. Using robust descriptors like the Gaussian of Gaussian (GOG) and Hierarchical Gaussian Descriptors (HGD), the suggested method may improve upon previous approaches to resolving difficult problems such partial occlusion induced by transported items and backdrop. The suggested strategy has been shown to improve the performance of state-of-the-art re-identification algorithms, as shown by experimental findings on widely-used people re-identification datasets.

Keywords: *Gaussian of Gaussian, Hierarchical Gaussian Descriptors, backdrop*

1. INTRODUCTION

Networked surveillance systems often use many video cameras with distinct fields of vision. Public and private transportation systems [1], government facilities [2], and "smart buildings" for the care of the elderly [3] all employ these types of systems to monitor and analyze the behavior of their patrons in order to spot any unusual occurrences. Re-identification of individuals is a difficult problem that has a direct impact on the efficiency of surveillance systems.

Both the gallery set and the probe set of photos of persons are used by re-identification systems. photographs of known individuals make up the former, while photographs of unknown or recently arrived individuals captured by a network camera make up the latter. The purpose of re-identification systems is to assign labels to the probe set members. To this end, we will first consider the degree of overlap between each probe and the audience participants take a poll. Then, the gallery set is ordered by how closely each individual matches the probe set. The probe member is considered to be a newcomer to the surveillance system and given a new label if the similarity between it and the most comparable member of the gallery set is below a threshold.

Re-identification systems' capabilities are hindered by obstacles such partial occlusion, position fluctuation, and lighting shifts. [4]. Data samples from the CUHK01 database [5] are shown in image 1; each row of the image represents data from the same subject captured by two separate cameras. The samples in the first and second columns of this figure were taken with the same camera, whereas the samples in the third and fourth columns were taken with a different camera. The effects of lighting and position changes on examples 1 and 2 are shown here. Carry-over occlusions partially obscure objectsandposevariationsleadtoappearancechangesinsamples3and4.Insample5,thepartialocclusioncausedby the carried object and variations in illumination andposechangedtheperson’sappearance.

According to the samples in Figure 1, appearancefeatures fromthe images cannot describe the personsappropriately.Inthesesamples,apersoncarriesanobjectand moves across the camera network. In this situation,the carried objects may be disappeared or occlude theperson’sbody,due topose variations.

Althoughbackgroundofapersonmaynotbethesamein different camera views of a network, there are somesimilarities between backgrounds of adjacent cameras inasurveillancesystem.Inthispaper,threeregions,includingperson’sbody,possiblecarriedobject, andbackground, are considered in a camera view, each ofwhich with different importance in re-identification.Adeepsemanticsegmentationapproachisemployedinthispapertosegmenteachinputimageintoregions correspondingtotheperson’sbody,possiblecarriedobjects, and background. We show that contributing thethree different regions with a significance factor improvesaccuracy of re-identification system.

Figure1.DatasamplesfromtheCUHK01database,eachrowrepresentsaperson attwo differentcameraview

2. RELATEDWORKS

Descriptors for people re-identification, such as GOG (Gaussian of Gaussian) and HGD (Hierarchical Gaussian Descriptors), were suggested by Matsukawa et al. [6, 7], who are resilient to posture and lighting variations. In these methods, the input image is broken down into seven overlapping horizontal sections in order to provide a local description of its structure. Furthermore, we think of each area as a collection of overlapping windows. The vertical position of each pixel in the window, the amount of the pixel intensity gradient in four directions, and the colour information in RGB, HSV, LAB, and nRGB are only some of the attributes used to represent each pixel. The Gaussian distribution in the given window may be calculated from these characteristics. Therefore, numerous Gaussian distributions are used to represent each area. Due of the non-linear nature of Gaussian distributions, the Euclidean operation cannot be applied directly to them. Therefore, in GOG and HGD, the window and region distribution parameters are translated onto the linear tangent space through the Symmetric Positive Definite (SPD) Riemannian manifold. Finally, the whole input image is described using the mapped Gaussian distributions of the areas. On the other hand, HGD employs a few feature norm normalization techniques to lessen the skewedness of

SPD matrix descriptors [8].

Local Maximal Occurrence (LOMO) is a description suggested by Liao et al. [9] that is insensitive to changes in lighting and perspective. A hybrid HSV histogram with 8x8x8 bins and histograms from the Scale Invariant Local Ternary Pattern (SILTP) [10] are used to characterize the image. The input image is first segmented into overlapping windows using this description. It is believed that the bins of the histograms correspond to the frequency with which the appropriate pattern occurs in each window. Then, to extract viewpoint-independent characteristics, we examine windows in the same horizontal position and maximize the frequency with which each pattern occurs locally (in the same histogram bin). In addition, a subspace and the cross-view quadratic discriminant analysis (XQDA) technique are presented, where the XQDA metric is concurrently learnt on the derived subspace, allowing for the learning of a discriminant low dimensional subspace.

Some low-level information, including colour moment values of RGB components and Schmid filter responses, are employed for a Multi-Level Gaussian Descriptor (MLGD) developed by Vishwakarma et al. [11] encoding the values of individual image pixels. This method borrows its process for extracting characteristics from the work of Matsukawa et al. [6], however Vishwakarma and Upadhyay [11] use alternative low-level features.

To address the problem of image quality variations due to varying camera settings, Prates et al. [12] suggested a Kernel Cross-View Collaborative Representation based Classification (Kernel X-CRC) method. In this technique, we employed the GOG descriptor to determine people's physical traits. In addition, the Kernel X-CRC was used to compare the images by first mapping the retrieved features into the learnt subspaces. In addition, Kernel Multiblock Partial Least Squares (Kernel MBPLS) was proposed as a nonlinear regression model by Prates and Schwartz [13]. In this method, we take into account a wide variety of data sources and map the characteristics we extract from data samples onto a low-dimensional subspace. Prates and Schwartz [13] suggest a method that, like Prates et al. [12], employs the GOG descriptor's extracted features by mapping them into the learnt subspaces. When it comes time to re-identify an individual, the mapped characteristics are sent into a Kernel X-CRC.

To address the difficulties introduced by perspective and position shifts, Zhou et al. [14] suggested a Graph Correspondence Transfer (GCT) people re-identification method. First, a collection of patch-wise correspondence templates is learned using positive image pairings in a variety of pose-pair combinations using a patch-wise graph matching method. Next, reference training pairings that most closely match the pose-pair combinations of the test photos are chosen. The image similarity is then calculated by applying the reference pair's correspondences to the test pair. In addition, Zhou et al. [14] applied a pose context descriptor based on the topological structure of the predicted joint locations [16] to strengthen the correspondence transfer they used.

Sample Specific Multi- Kernel (SSMK) is a method introduced by Fang et al. [17] to address the problem of inconsistent visuals across cameras. First, the photos are split into six horizontal halves, and then features are collected from each section using methods such

Dense SIFT [18, 19], colour naming features [20], and deep features [21]. Next, a discriminative metric for re-identification is learned by mapping the extracted features into the weighted multi- kernel feature space.

Jia et al. [22] suggested a View- Specific Semi-supervised Subspace Learning (VS-SSL) strategy for lowering the number of labelled training samples in person re-identification. GOG descriptor is used to describe the photographs and the method also learns iterative multi-view joint transfer learning optimum problem, Zhao et al. [23] implemented the Inexact Augmented Lagrange Multiplier (IALM) method [24] for people re-identification. The aim of this method is to deal with the problem of data distribution inconsistency across different camera perspectives.

To account for difficulties associated with perspective and lighting differences in re-identification [25-28], some methods employ both the low-level elements (i.e., colour and texture) and the mid-level attributes (carried items and garments with a certain colour). The properties of carrying goods, sunglasses, and logos were chosen as intermediate qualities by Layne et al. [25]. A classifier is learned for each characteristic by analyzing the training samples' low-level characteristics. The probing images are then analyzed with the help of the trained classifiers to look for evidence of the intermediate features.

Zhao et al. [26] suggested an unsupervised saliency learning method. These techniques compare image patches to a generic reference set in order to generate a saliency map for each image. Patches are compared according to their colour and texture. The saliency score for each patch is calculated using the K-Nearest Neighbors algorithm and a one-class support vector machine [29]. The probing image is re-identified by comparing its saliency map to the saliency maps of the gallery images, which were obtained after the saliency maps were obtained.

Martinel et al. [27, 30] weighed low-level features using the images' saliency maps. The image saliency map is created using a Markov chain technique in these methods [28]. Each pixel's value in the image saliency map represents the pixel's relative significance in the final retrieved features. A pairwise-based multiple metric learning approach is then used to both the weighed and unweighted features.

Convolutional Neural Network (CNN) characteristics learnt and retrieved from images are employed in certain current re-identification methods. In order to extract global appearance characteristics from CNN, Sun et al. [31] suggested SVDNet by optimizing the deep representation learning process using SVD. Zheng et al. [32] introduced the pedestrian alignment network (PAN) using CNN feature maps to fix the problem of crooked photos. While training pedestrian descriptions, this network also aligns pedestrians inside boundary boundaries. By pairing the hand-crafted and deep learning-based features described in the literature [6, 34, 35] with a variety of metric learning approaches, Yu et al. [33] simulated many investigators for each probe image. Researchers' separate rankings were then combined using a crowdsourcing-based rating aggregation method.Linetal.[36]jointlytrainedmulti-tailored views from each camera's perspective.

In order to maximize the linked complementary information between attention selection and feature discrimination, a new module called the Harmonious Attention CNN (HA-CNN) has been developed.

The aforementioned strategies for re-identification do not account for the problem of visual modifications due to props and backgrounds. To better the performance of GOG and HGD techniques, Mortezaie et al. [37] developed a re-ranking strategy based on the colour of the body and carried things. Before using this method of re-ranking, the input image is separated into its component parts (person, carried item, backdrop). Then, one of Parraga and Akbarinia's [38] colour categorization methodologies is applied to the colors of the person's body and carrying things. Mortazaie et al. [39] also proposed the unification process, a pre-processing phase for minimizing the visual impact of a person's belongings. To replicate the colour of the body parts that were obscured by the things being carried, our method took into account the colour of the neighboring, non-occluded body parts.

Both deep learning network features and manually constructed characteristics are used in the re-identification methods. In this part, we took a quick look at a few of the many methods already in use. Note that the problem of appearance alterations produced by carrying items and backdrops is not addressed by re-identification systems based on hand-crafted characteristics. In addition, a huge dataset is required for training deep learning re-identification methods like the ones we've just discussed. In addition, there are a large number of parameters to adjust, making the training phase of these methods exceedingly lengthy. Instead of recognizing occlusion caused by carrying items automatically, the current re-identification methods are taught to ignore the masked region of the body.

3.PERSON RE-IDENTIFICATION

Our suggested method begins by using a deep semantic segmentation algorithm called DeepLabv3+ [40] to the input image, which separates it into three regions: the person's body, any carrying items, and the backdrop. Next, visual features are collected from each of the segments and included into the re-identification in accordance with their relative relevance.

To perform semantic image segmentation, DeepLabv3+ uses an encoder-decoder architecture. The encoder component of this method is responsible for providing the necessary contextual data. The decoder component also helps in determining where one item ends and another begins.

To divide the images into the backdrop, the person, and the partly occluded areas created by carrying items, we trained DeepLabv3+ on the manually segmented masks of VIPeR [41] images.

The PRID450s database [42] has segmented masks that may be downloaded for free. Figure 2 displays Semantic Segmented Maps (SSM) generated by the trained DeepLabv3+ for data samples taken from the VIPeR (samples 1 and 2) and CUHK01 (samples 3 and 4) databases. The segmented photos show the backdrop in black, the individual in white, and the occluded areas in shades of grey.

The suggested method we've developed takes into account a significance factor (in the interval [0,1]) for each segment after getting SSM for the input image. During re-identification, the significance factors are multiplied by feature vectors collected from each area. Factors associated with the person's body are given the highest rating, 1, because of the crucial role appearance plays in re-identification. Due to its lesser importance in re-identification than the person's body, the significance factor linked with a carried item is set at 0.9. When trying to identify a person again, the backdrop is less crucial than the person themselves or any props they may be carrying. Therefore, we set the background significance factor as 0.5 based on our experiments.

As was previously indicated, the method's effectiveness may be enhanced by detecting visual areas and taking their relevance in re-identification into account. We suggest a factoring strategy to enhance the performance of three previously reported persons re-identification algorithms, which we briefly present below. The enhanced versions are then subjected to experimental analysis.

Figure2.DatasamplesfromtheVIPeRandCUHK01databases,andtheir corresponding SSM

Matsukawa et al.'s [6, 7] re-identification methods are resilient to changes in appearance caused by factors like scene lighting and position alterations. However, these methods are limited by issues with cluttered backgrounds and visual changes induced by partly obscured locations.

$$Z_{Fusion}=[Z^T_{RGB},Z^T_{LAB},Z^T_{HSV},Z^T_{nRGB}]^T. \quad (3)$$

The GOG and HGD descriptions both break the image down into seven overlapping horizontal sections. Furthermore, we think of each area as a collection of overlapping windows. When applied to Equation (1) by Matsukawa et al. [6, 7], these descriptors extract features for individual pixels (i) in each window without taking into account their connection to the backdrop, person's body, or perhaps partly occluded areas:

$$F_i=[v;D_{0^\circ};D_{90^\circ};D_{180^\circ};D_{270^\circ};x_R;x_G;x_B]^T \quad (1)$$

where v denotes pixel's vertical location; D_{0°,D_{90°,D_{180°,D_{270°,representthemagnitudesofthepixelintensitygradientin fourdifferentorientations;and x_R,x_G,x_B,arethepixelvalues inR,G,andB channels.

To make the GOG and HGD descriptors robust against appearance changes caused by crowded background and carried object, we propose to use a significance factoring scheme to bias the extracted feature set from each segment after segmenting the images into the person's body, carried objects, and background.

Using its extracted feature (biased feature), a Gaussian distribution of window pixels is calculated. The non-linear Gaussian distributions derived by Matsukawa et al. [6, 7] are translated into linear tangent spaces. The related area is then characterized by a collection of several Gaussian distributions based on the mapped Gaussian distributions of the windows. Distributions of windows may be thought of as being Gaussian. The Gaussiandistributionsoftheregionsaremappedintothetangent space.

By concatenating the mapped Gaussian distributionsof the regions, the whole input image is described byMatsukawaet al.[6, 7]asfollows:

$$Z_{RGB}=[z^T_{g1},z^T_{g2},z^T_{g3},z^T_{g4},z^T_{g5},z^T_{g6},z^T_{g7}]^T \quad (2)$$

where z_{gk}, $k=1,2,\dots,7$ represents the mapped Gaussiandistributionof region$k$.

InGOGandHGD,inadditiontoRGBcolorspace,theLAB, HSV, and nRGB color spaces are also used bysubstituting their color channels with RGB information(i.e., x_R, x_G,and x_B) used in Equation (1). We name theextracted feature vectors using LAB, HSV, and nRGB[1],asZ_{LAB},Z_{HSV},andZ_{nRGB}respectively.Byconcatenating Z_{RGB},Z_{LAB},Z_{HSV},andZ_{nRGB},thefinalfeaturevectoris discussedbyMatsukawaetal.[6,7]asfollows:

The XQDA distance measure is then learned with the help of Matsukawa et al.'s [6, 7] ZFusion. By combining the Bayesian face [43] and KISSME [44] methods, the XQDA is able to concurrently learn a discriminative subspace and a distance metric. In XQDA, this is modelled as a Generalized Rayleigh Quotient issue (see [45]). Additionally, a closed form solution is obtained by the generalized eigenvalue decomposition.

Martinel et al.'s [27] feature extraction method for re-identification is the third method we examine incorporating our suggested strategy into. KErnelized saliency-based Person re-identification by multiple metric LEaRning (KEPLER) was suggested by Martinel et al. [27]. First, a kernelized saliency detection module [28] based on the Markov chain technique is used to construct saliency maps () of the images. Further, the retrieved characteristics are weighed using the saliency maps. The input image is converted to the HSV, Lab, YUV, rgs1, RGB, and grayscale colour spaces before the feature extraction phase begins. Multiple overlapping windows are created using the input image's colour channels (l) and saliency map (). Each window (j) in each colour channel (l) and each window (j) of the grayscale image are then extracted to provide a colour mean, 128-dimensional SIFT descriptor, and Haar-like sparse- compressive features [46, 47]. Features specified by Martinel et al. [27] are retrieved in their unweighted form. Meanwhile, the saliency weighted histogram (H) is retrieved in the weighed form utilizing in the following way:

$$H^{j,l}_{a,b} = \sum_{(m,n) \in T^{j,l}} \begin{cases} \square^{j}_{m,n}, & a < T^{j,l}_{m,n} \leq b \\ 0 & o.w \end{cases} \qquad (4)$$

where, $\square^{j}$ and $T^{j,l}$ 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.

Inthispaper,weincorporateourproposedsignificancefactorinKEPLERapproachbysubstituting □ in Equation(4)with thesignificancefactor.

4. EXPRIMENTALRESULTS

When assessing the efficacy of a re-identification system, one popular metric is Rank-k, where k is the number of top matches that provide a right response. For k = 1, rank-k is the most stringent metric, but for k > 1, it allows for certain inaccuracies [48]. In this study, the suggested method is tested on the CUHK01 [5], VIPeR [41], PRID450s [42], and CUHK03 [49] datasets using the Rank-k (k = 1, 5, 10, 20) metric.

One thousand two hundred and sixty-four photos of 632 people and nine hundred images of four hundred and fifty people are included in the VIPeR and PRID450s databases, respectively. Since these datasets only include a single image of each individual from each camera angle, we use single-shot matching to test how well our method works. There are 3,884 photos of 971 people in the CUHK01 collection, with two photographs of each subject collected in each camera position. We provide the results of our proposed method on this dataset for both single-shot (M=1) and multi-shot (M=2) matching in Table 1. On average, 4.8 photos of each individual were acquired by each camera in the CUHK03 collection, which has 13,164 total photographs of 1,360 people. We compare methods using multi-shot matching using this dataset's manually cropped photos (with labels).

Our suggested method's impact on GOG, HGD, and KEPLER re-identification techniques was summarized in Table 1. Each position in this table is the result of applying our suggested method to the GOG description.

Specifically, we compare the results obtained using our proposed approach applied to the GOG descriptor (i.e., the enhanced GOG using the significance factor) with those obtained using the classic GOG method, the HGD descriptor (i.e., the enhanced HGD using the significance factor), and the KEPLER descriptor (i.e., the enhanced KEPLER) with its classic version. Each ranking order in the table is followed by the results acquired using the improved and traditional versions of the respective approach; the bolded results represent the best estimates available from the related database.

In this study, 10 separate sets of training and testing data are utilized to master the XQDA distance measure, which is used in both GOG and HGD, and the KEPLER technique. Therefore, Table 1 displays an average of the test set accuracy results.

Table1. Performanceofourproposed approachand themethodsreportedinliterature [6,7,27]

	Re-identificationapproach	**VIPeR**	**CUHK01 (M=1)**	**CUHK01 (M=2)**	**PRID 450s**
Rank1%	EnhancedGOGusing $\hat{W}$	**59.4**	**62.5**	**72.1**	**78.7**
	ClassicGOG [6],(2016)	49.7	57.8	67.3	68.4
	EnhancedHGDusing $\hat{W}$	**60.9**	**63.3**	**74.0**	**80.5**
	ClassicHGD [7],(2019)	50.0	59.0	70.3	70.4
	Enhanced KEPLERusing $\hat{W}$	**41.7**	**42.0**	**57.0**	**52.0**
	ClassicKEPLER[27],(2015)	40.2	42.0	54.8	51.9
Rank5%	EnhancedGOGusing $\hat{W}$	**85.0**	**82.4**	**89.5**	**93.5**
	ClassicGOG [6],(2016)	79.7	79.1	86.9	88.8

	EnhancedHGDusing $\hat{W}$	**85.9**	**83.5**	**90.0**	**94.7**
	ClassicHGD [7],(2019)	79.5	79.7	87.9	91.2
	Enhanced KEPLERusing $\hat{W}$	**71.3**	**66.0**	**79.3**	77.2
	ClassicKEPLER [27],(2015)	70.0	65.0	78.9	**77.6**
Rank10%	EnhancedGOGusing $\hat{W}$	**91.7**	**88.9**	**94.1**	**97.1**
	ClassicGOG[6],(2016)	88.7	86.2	91.8	94.5
	EnhancedHGDusing $\hat{W}$	**92.2**	**89.8**	**94.1**	**97.7**
	ClassicHGD[7],(2019)	88.9	86.2	92.2	94.8
	Enhanced KEPLERusing $\hat{W}$	**81.5**	**75.3**	**86.0**	**84.6**
	ClassicKEPLER [27],(2015)	81.4	74.5	85.3	84.5
Rank20%	EnhancedGOGusing $\hat{W}$	**95.9**	**94.0**	**97.3**	**98.8**
	ClassicGOG[6],(2016)	94.5	92.1	95.9	97.8
	EnhancedHGDusing $\hat{W}$	**96.7**	**94.3**	**97.5**	**99.2**
	ClassicHGD[7],(2019)	94.6	92.0	95.8	97.6
	Enhanced KEPLERusing $\hat{W}$	**90.7**	**83.6**	**91.0**	**91.0**
	ClassicKEPLER [27],(2015)	90.4	83.0	90.8	90.7

Table 1 demonstrates that for rankings 1,5,10 and 20 across all comparison databases, the enhanced GOG and the improved HGD utilizing the significance factoring technique provide more accurate findings than the original approaches. In addition, if you use our suggested.

When applied to KEPLER, this method yields more precise findings than the original for rankings 1, 5, 10, and 20 on the VIPeR and CUHK01 databases, and for levels 1-10 on the PRID450s.

Keep in mind that many photographs in the VIPeR database have a busy backdrop or an area that is obscured by other elements. Based on the findings obtained with this database, it is clear that the extracted feature vectors biased using the suggested technique are more unique and accurate than the original non-biased feature vectors.

In our proposed method, we first train DeepLabv3+ on the manually segmented masks from the VIPeR and PRID450s datasets, and then we use the trained network to segment images from the CUHK01 and CUHK03 datasets, separating out the human body, carried objects, and background regions. Since DeepLapv3+ only needs to have its parameters modified once during training, the time spent training it has little overhead. Multiplying the components in Table 1 by the retrieved features uses our suggested biassing approach on descriptors. Therefore, there is a continuous cost of O (1) when using skewed descriptors.

State-of-the-art approaches are used to compare our proposed approach's performance on the CUHK03, VIPeR, CUHK01, and PRID450s datasets in Tables 2 through 5. Most re-identification techniques only reported their accuracy on the CUHK03 dataset at rank 1, thus we only compared our persons re-identification method to those at rank 1 in Table 2.

Section 2 said that deep neural networks were utilized to train and extract the appearance characteristics using the re-identification methodologies provided in the literature [16, 31-33, 36]. It takes a lot of time and resources to fine-tune deep networks in these methods so that they can learn the right characteristics.

The training step in feature extraction is used by both deep learning-based and hand-crafted methods. Using a patch-wise graph matching method, Zhou et al. [14, 15] trained a collection of patch-wise correspondence templates.

Table2.Comparisonoftheperformanceofourproposedapproachwith state-of-the-artmethodson CUHK03(labeled)

Approach	**Rank1 (%)**
Sunetal.,[31],(2017)	40.9
Zheng etal,[32],(2018)	36.9
Linetal,[36],(2018)	44.4
Yuetal,[33],(2020)	53.9
ClassicGOG[6],(2016)	67.3
ClassicHGD[7],(2019)	68.9
EnhancedGOG	**69.6**
EnhancedHGD	68.5

Table3.Comparisonoftheperformanceofourproposedapproachwithstate-of-the-artmethodsonVIPeR

Approaches	**Ranks%**			
	1	**5**	**10**	**20**
Layneetal.,[25],(2012)	18.8	40.9	54.9	-
Zhaoetal.,[26],(2013)	26.7	50.7	62.4	76.4
Martineletal.,[30],(2014)	33.0	-	75.6	86.9
Liaoetal.,[9](2015)	40.0	-	80.5	91.1
Vishwakarmaetal.,[11],(2018)	47.5	-	87.9	93.7
Pratesetal. [12],(2019)	51.6	80.5	89.5	95.2
Pratesetal. [13](2019)	51.2	79.9	89.9	-

Fangetal., [17],(2019)	43.8	79.2	87.2	94.9
Jiaetal., [22], (2020)	44.8	72.3	79.3	86.1
Mortezaieetal., [37],(2021)	53.0	82.7	90.7	95.7
EnhancedGOG	59.4	85.0	91.7	95.9
EnhancedHGD	**60.9**	**85.9**	**92.2**	**96.7**

Table4.Comparisonoftheperformanceofourproposedapproach withstate-of-the-artmethodsonCUHK01(m=2)

Approaches	**Ranks%**			
	1	**5**	**10**	**20**
Liaoetal.,[9](2015)	63.2	-	90.8	94.9
Vishwakarmaetal.,[11],(2018)	54.5	-	83.5	90.5
Fangetal., [17],(2019)	69.2	87.8	93.2	97.1
Pratesetal.,[12],(2019)	63.1	82.7	89.0	94.6
Zhaoetal., [23],(2020)	68.4	86.3	93.6	96.8
Mortezaieetal., [37],(2021)	70.7	88.2	92.3	96.2
EnhancedGOG	72.1	89.5	94.1	97.3
EnhancedHGD	**74.0**	**90.0**	**94.1**	**97.5**

Table5.Comparisonoftheperformanceofourproposedapproach withstate-of-the-artmethodsonPrid450s

Approaches	**Ranks%**			
	1	**5**	**10**	**20**
Liaoetal.,[9](2015)	62.6	85.6	92.0	96.6
Vishwakarmaetal.,[11],(2018)	62.4	-	93.5	96.9
Zhouetal.,[14], (2018)	58.4	77.6	84.3	89.8
Zhouetal.,[15], (2019)	70.9	89.1	93.5	96.5
Pratesetal. [12],(2019)	71.3	91.7	96.0	98.1
Pratesetal. [13](2019)	68.1	90.7	95.0	-

Jiaetal., [22], (2020)	68.2	90.2	94.9	98.0
Zhaoetal., [23],(2020)	72.1	-	94.6	-
Mortezaieetal., [37],(2021)	74.9	93.0	96.6	99.1
EnhancedGOG	78.7	93.5	97.1	98.8
EnhancedHGD	**80.5**	**94.7**	**97.7**	**99.2**

The computational cost of these methods grows when several positive image pairings are used, each with a unique pose-pair configuration. The process of re-identification was seen by Zhao et al. [23] as an ideal issue in multi-view joint transfer learning that required iterative solutions. Mid-level features including carrying items, sunglasses, and logos were employed by Layne et al. [25], with classifiers trained on the low-level features of the training samples for each feature. However, unlike when just the raw low-level features are used, training a large number of mid-level features takes a lot of time. In addition, Zhao et al. [26] used colour and texture comparisons between regions in images and patches in the reference photos to generate a saliency map for each image. The method provided by Zhao et al. [26] incurs significant processing overheads due to these comparisons. Keep in mind that there is no learning process involved with the GOG and HGD descriptions. Therefore, the computational complexity of the improved GOG and HGD is less than that of the aforementioned methods.

In addition, Martinel et al. [30] used a Markov chain to calculate a weight map for each image. It's important to keep in mind that our suggested method uses fewer complex calculations than the Markov chain technique since it uses a pre-trained network to semantically split the images and assigns a constant number to each segment. On the other side, Liao's [9] LOMO descriptor is less complex than GOG and HGD since it relies on only two manually-created characteristics (namely, HSV and SILTP). Tables 3–5 show, however, that the improved GOG and enhanced HGD are both much more accurate than LOMO.

Vishwakarma and Upadhyay's [11] feature extraction technique is comparable to GOG and HGD; however, it employs distinct low-level characteristics from those used by Matsukawa et al.According to Tables 3–5, the improved GOG and enhanced HGD have greater performance than the method described by Vishwakarma and Upadhyay [11], although having a computational complexity that is comparable to that of GOG and HGD.

Prates and Schwartz [12, 13], Jia et al. [22, 37], and their own re-identification methods are all based on the GOG descriptor. As a result, the computational cost of these methods in feature extraction is close to that of the improved GOG and improved HGD.

Furthermore, Table 2 shows that improved GOG and enhanced HGD outperform deep learning-based methods on CUHK03, while using hand-crafted features. Our suggested technique also beats the competing approaches in every ranking category on the VIPeR, CUHK01, and PRID450s datasets, as demonstrated in Tables 3–5 and Figures 3–5. Indeed, in our suggested method, the impacts of partial occlusion brought by carried items and backdrop are reduced, leading to an improved depiction of the individual.

5. CONCLUSION

Successful re-identification is crucial to the efficiency of surveillance systems. The performance of appearance-based re-identification methods may be enhanced by extracting features from various parts of the images while taking into account their value in re-identification. In this study, we suggest a method for doing this by tuning the influence of each image pixel on the derived feature vectors by taking into account the pixel's relationship with the backdrop, the body of the person, and partly occluded areas created by held items. The suggested strategy has been shown to improve the efficiency of previous re-identification strategies, as shown by testing findings on a number of different datasets. Each area's relevance factor was calculated based on extensive experimentation in our suggested method. Possible future automated computation of significant factors based on area and size in image highlighted the advantages of improved GOG and improved HGD over competing methods. Therefore, our suggested method may increase the accuracy of the re-identification using appearance-based descriptors with little to no additional processing cost.

REFERENCES

1. Hammoudi, K., Abu Taha, M., Benhabiles, H., Melkemi, M.,Windal, F., El Assad, S., and Queudet, A., 2020. Image-BasedCipheringofVideoStreamsandObjectRecognitionforUrbanandVehicularSurveillanceServices.InFourthInternationalCongresson 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 BorderSurveillanceUsingMachineLearninginRemoteVideoSurveillanceSystems.InEmergingTrendsinElectrical,Communications,andInformationTechnologies,pp.751-760.Springer,Singapore.Doi:10.1007/978-981-13-8942-9_64.
3. Huang,Z.,Liu,Y.,Fang,Y.,andHorn,B.K.P.,2018.Video-basedFall Detection for Seniors with Human Pose Estimation. In: 20184th International Conference on Universal Village (UV). IEEE, pp1–4.Doi:10.1109/UV.2018.8642130.
4. Gawande,U.,Hajari,K.,andGolhar,Y.,2020.Pedestriandetectionandtrackinginvideosurveillancesystem:issues,comprehensivereview,andchallenges.RecentTrendsinComputational Intelligence, pp.1–24. London, United Kingdom:IntechOpen.
5. Li, W., Zhao, R., and Wang, X., 2013. Human Reidentificationwith Transferred Metric

Learning. Asian conference on computervision, Berlin, Heidelberg.

6. Matsukawa,T.,Okabe,T.,Suzuki,E.,andSato,Y.,2016.Hierarchical Gaussian Descriptor for Person Re-identification. In:2016IEEEConferenceonComputerVisionandPattern152.Doi: 10.1109/CVPR.2016.152.
7. Matsukawa,T.,Okabe,T.,Suzuki,E.,andSato,Y.,2020.HierarchicalGaussianDescriptorswith ApplicationtoPersonRe-Identification.IEEETransactionsonPatternAnalysisandMachineIntelligence,42(9),pp.2179–2194.Doi:10.1109/TPAMI.2019.2914686.
8. Li, P., Wang, Q., and Zhang, L., 2013. A Novel Earth Mover'sDistanceMethodologyforImageMatchingwithGaussianMixtureModels.In:2013IEEEInternationalConferenceonComputerVision.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 andmetric learning. In: 2015 IEEE Conference on Computer VisionandPatternRecognition(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 patternsfor background subtraction in complex scenes. In: 2010 IEEEComputer Society Conference on Computer Vision and PatternRecognition.IEEE, pp.2197-2206.Doi:10.1109/CVPR.2015.7298832.
11. Vishwakarma,D.K.,andUpadhyay,S.,2018.ADeepStructureofPersonRe-IdentificationUsingMulti-LevelGaussianModels.IEEETransactionsonMulti-ScaleComputingSystems,4(4),pp.513–521.Doi:10.1109/TMSCS.2018.2870592.
12. Prates,R.,andSchwartz,W.R.,2019.Kernelcross-viewcollaborative representation-based classification for person re-identification.JournalofVisualCommunicationandImageRepresentation, 58,pp.304–315. Doi:10.1016/j.jvcir.2018.12.003.
13. Prates, R., and Schwartz, W.R., 2019. Matching People AcrossSurveillance Cameras. In: Anais Estendidos da Conference onGraphics,PatternsandImages(SIBGRAPI).SociedadeBrasileiradeComputação-SBC,pp.84-90.Doi:10.5753/sibgrapi.est.2019.8306.
14. Zhou,Q.,Fan,H., Zheng,S.,Su, H.,Li,X.,Wu,S.,andLing,H.,2018.GraphCorrespondenceTransferforPersonRe-Identification | Proceedings of the AAAI Conference on ArtificialIntelligence. *Proceedings of the AAAI Conference on ArtificialIntelligence*,32(1). Doi:10.5555/3504035.3504966.
15. Zhou, Q., Fan, H., Yang, H., Su, H., Zheng, S., Wu, S., and Ling,H.,2021.RobustandEfficientGraphCorrespondenceTransferforPersonRe-Identification.IEEETransactionsonImageProcessing,30,pp.1623–1638.Doi:10.1109/TIP.2019.2914575.

16. Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y., 2017. Realtimemulti-person 2d pose estimation using part affinity fields. In:ProceedingsoftheIEEEconferenceoncomputervisionandpattern recognition.pp7291–7299.
17. Fang, J., Zhang, R., and Jiang, F., 2019. Sample Specific Multi-KernelMetricLearningforPersonRe-identification.In:2ndInternationalConferenceonElectricalandElectronicEngineering(EEE2019).AtlantisPress,pp226–241.Doi:10.2991/eee-19.2019.38.
18. Liu, Y., Liu, S., and Wang, Z., 2015. Multi-focus image fusionwithdenseSIFT.InformationFusion,23,pp.139–155.Doi:10.1016/j.inffus.2014.05.004.
19. An,L.,Kafai,M.,Yang,S.,andBhanu,B.,2013.Reference-basedpersonre-identification.In:201310thIEEEInternationalConference on Advanced Video and Signal Based Surveillance.IEEE,pp244–249.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 deeplearning architecture for person re-identification. In: 2015 IEEEConferenceonComputerVisionandPatternRecognition(CVPR).IEEE,pp3908–3916.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.PatternRecognition,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 personre-identification.PatternRecognition,97,pp.107014.Doi:10.1016/j.patcog.2019.107014.
24. Xu,Y.,Fang,X.,Wu,J.,Li,X.,andZhang,D.,2016.Discriminative Transfer Subspace Learning via Low-Rank andSparse Representation. IEEE Transactions on Image Processing,25(2),pp.850–863. Doi:10.1109/TIP.2015.2510498.
25. Layne,R.,Hospedales,T.M.,andGong,S.,2012.TowardsPersonIdentificationandRe-identificationwithAttributes.pp402–412.Doi:10.1007/978-3-642-33863-2_40.
26. Zhao,R.,Ouyang,W.,andWang,X.,2013.UnsupervisedSalience Learning for Person Re-identification. In: 2013 IEEEConference on Computer Vision and Pattern Recognition. IEEE,pp3586–3593. Doi:10.1109/CVPR.2013.460.
27. Martinel, N., Micheloni, C., and Foresti, G.L., 2015. KernelizedSaliency-Based Person Re-Identification Through Multiple MetricLearning.IEEETransactionsonImageProcessing,24(12),pp.5645–5658.Doi:10.1109/TIP.2015.2487048.
28. Harel, J., Koch, C., and Perona, P., 2006. Graph-based visualsaliency. Advances in neural information processing systems, 19,pp.545-552.

29. Heller, K., Svore, K., Keromytis, A.D., and Stolfo, S., 2003. Oneclass support vector machines for detecting anomalous windowsregistryaccesses.In:ICDMWorkshoponDataMiningforComputerSecurity. Doi:10.7916/D85M6CFF.
30. Martinel, N., Micheloni, C., and Foresti, G.L., 2015. SaliencyWeightedFeaturesforPersonRe-identification.pp191–208.Doi:10.1007/978-3-319-16199-0_14.
31. Sun, Y., Zheng, L., Deng, W., and Wang, S., 2017. SVDNet forPedestrian Retrieval. In: 2017 IEEE International Conference onComputerVision(ICCV).IEEE,pp3820–3828.Doi:10.1109/ICCV.2017.410.
32. Zheng, Z., Zheng, L., and Yang, Y., 2019. Pedestrian AlignmentNetworkforLarge-scalePersonRe-Identification.IEEETransactionsonCircuitsandSystemsforVideoTechnology,29(10),pp.3037–3045.Doi:10.1109/TCSVT.2018.2873599.
33. Yu,Y.,Liang,C.,Ruan,W.,andJiang,L.,2020.Crowdsourcing-BasedRankingAggregationforPersonRe-Identification.In:ICASSP2020-2020IEEEInternationalConferenceonAcoustics,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(andAStrongConvolutionalBaseline).pp501–518.Doi:10.1007/978-3-030-01225-0_30.
35. He, K., Zhang, X., Ren, S., and Sun, J., 2016. Deep ResidualLearning for Image Recognition. In: 2016 IEEE Conference onComputerVisionandPatternRecognition(CVPR).IEEE,pp770–778.Doi:10.1109/CVPR.2016.90.
36. Chen, L., Yang, H., Xu, Q., and Gao, Z., 2021. Harmoniousattention network for person re-identification via complementaritybetween 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-BasedRe-RankingProcessforPeopleRe-Identification:PaperID 21.In:20219thEuropeanWorkshoponVisualInformationProcessing(EUVIP).IEEE,pp1–5.Doi:10.1109/EUVIP50544.2021.9484056.
38. Parraga, C.A., and Akbarinia, A., 2016. NICE: A ComputationalSolution to Close the Gapfrom Colour Perception toColourCategorization.*PLOSONE*,11(3),pp. e0149538.Doi:10.1371/journal.pone.0149538.
39. Mortezaie,Z.,Hassanpour,H.,andBeghdadi,A.,2022.Peoplere-identificationunderocclusionandcrowdedbackground.MultimediaToolsandApplications,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

forSemanticImageSegmentation.ProceedingsoftheEuropeanconferenceoncomputervision(ECCV),pp:801-818.Doi:10.1007/978-3-030-01234-2_49.

41. Gray,D.,andTao,H.,2008.ViewpointInvariantPedestrianRecognitionwithanEnsembleofLocalizedFeatures.pp262–275.Doi:10.1007/978-3-540-88682-2_21.
42. Roth, P.M., Hirzer, M., Köstinger, M., Beleznai, C., and Bischof,H.,2014.MahalanobisDistanceLearningforPersonRe-identification.In:PersonRe-Identification.SpringerLondon,London,pp247–267.Doi: https://doi.org/10.1007/978-1-4471-6296-4_12.
43. Moghaddam,B.,Jebara,T.,andPentland,A.,2000.Bayesianfacerecognition.PatternRecognition,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.Largescalemetriclearningfromequivalenceconstraints.In:2012IEEEConferenceonComputerVisionand PatternRecognition.IEEE,pp2288–2295.Doi:10.1109/CVPR.2012.6247939.
45. Alipanahi, B., Biggs, M., and Ghodsi, A., 2008. Distance metriclearning vs. fisher discriminant analysis. In: Proceedings of the23rd national conference on Artificial intelligence. pp 598–603.Doi:10.5555/1620163.1620164.
46. Zhang,K.,Zhang,L.,andYang,M.-H.,2012.Real-TimeCompressive Tracking. European conference on computer vision(ECCV),pp864–877.Doi:10.1007/978-3-642-33712-3_62.
47. Ojala,T.,Pietikainen,M.,andMaenpaa,T.,2002.Multiresolutiongray-scale and rotation invariant texture classification with localbinarypatterns.IEEETransactionsonPatternAnalysisandMachineIntelligence,24(7), pp.971–987.Doi:10.1109/TPAMI.2002.1017623.
48. mortezaie,zahra, and Hassanpour, H., 2019. A survey on ageinvariantfacerecognitionmethods.JordanianJournalofComputers andInformation Technology, 5(2), pp.87–96. Doi:10.5455/jjcit.71-1554841475.
49. Li, W., Zhao, R., Xiao, T., and Wang, X., 2014. DeepReID: DeepFilter Pairing Neural Network for Person Re-identification. In:2014IEEEConferenceonComputerVisionandPatternRecognition.IEEE,pp152–159.Doi:10.1109/CVPR.2014.27.