Article

Person Re-Identification by Discriminative Local Features of Overlapping Stripes

Fawad ^{1,*} , Muhammad Jamil Khan ¹ and MuhibUr Rahman ^{2,*}

¹ ACTSENA Research Group, Telecommunication Engineering Department, University of Engineering and Technology Taxila, Punjab 47050, Pakistan; muhammad.jamil@uettaxila.edu.pk

² Department of Electrical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada

* Correspondence: engr.fawad@students.uettaxila.edu.pk (F.); muhibur.rahman@polymtl.ca (M.R.)

Received: 23 February 2020; Accepted: 8 March 2020; Published: 17 April 2020

Abstract: The human visual system can recognize a person based on his physical appearance, even if extreme spatio-temporal variations exist. However, the surveillance system deployed so far fails to re-identify the individual when it travels through the non-overlapping camera's field-of-view. Person re-identification (Re-ID) is the task of associating individuals across disjoint camera views. In this paper, we propose a robust feature extraction model named Discriminative Local Features of Overlapping Stripes (DLFOS) that can associate corresponding actual individuals in the disjoint visual surveillance system. The proposed DLFOS model accumulates the discriminative features from the local patch of each overlapping strip of the pedestrian appearance. The concatenation of histogram of oriented gradients, Gaussian of color, and the magnitude operator of CjLBP bring robustness in the final feature vector. The experimental results show that our proposed feature extraction model achieves rank@1 matching rate of 47.18% on VIPeR, 64.4% on CAVIAR4REID, and 62.68% on Market1501, outperforming the recently reported models from the literature and validating the advantage of the proposed model.

Keywords: re-identification; texture; tracking; convolution; filter

1. Introduction

Surveillance cameras are mounted in critical geographical locations to ensure public safety and security. The visual scenes captured by these cameras require constant monitoring. Most often, the camera network deployed within the campus possesses blank spaces due to the non-overlapping field-of-view [1]. The object trackers [2] fail whenever the object travels through such non-overlapping areas of the surveillance coverage. To monitor the overall trajectory of a person, the tracker needs to employ a robust re-identification (Re-ID) system [3]. Person Re-ID is the task of associating individuals within the non-overlapping camera scenes [4,5]. The appearance variability of the same individual due to the non-uniform environmental conditions and deformable nature of the human body degrades the performance of the Re-ID system [6]. The viewpoint, object's scale, and illumination of the scene vary due to the camera mount position, depth, and light source of the surveillance environment [6,7]. Moreover, occlusions and background clutter also affect the performance of the Re-ID system [7]. The performance of the Re-ID model depends on the content of the probe and gallery set. The gallery can either be a single-shot or multi-shot with single or multiple images of the individual, respectively [8]. Moreover, the gallery image can be acquired online or offline. In the offline case, the gallery includes pre-registered candidates, while, in the online stage, gallery candidates are updated with time at various cameras across the network [9].

The probe and gallery images are described using features such as texture, color, shape, and motion. The representation of an individual in the camera scene requires a robust feature

descriptor, which is invariant to the geometric and photometric variations. The geometric changes are scale, orientation, and viewpoint. Scale variation that arises due to the continuous change in the camera depth is the origin of the difference in the resolution of the probe and gallery sample of the same candidate [10]. The viewpoint of the camera mount varies due to building architecture, which also creates variations in the pedestrian appearance [11]. The photometric variations are due to the noise in the electronics component of the camera, the difference in the sensor response, occlusions, and illumination variations. Face and iris are the most discriminative features of the human body. However, in the case of surveillance, they are useless due to the low resolution. The occlusion that is self-occlusion and occlusion due to other objects creates an increase in the intra-class variations [12]. The individual appears in the field of view of one camera that can have a different pose than another camera from the network. In-balance camera sensor response creates changes in the sensed color value for the same scene. The lesser amount of training samples in the gallery makes the problem inappropriate for the case of DNN based Re-ID approaches. The illumination variations due to the difference in the light conditions also increase the intra-class variations and decrease the inter-class changes. The performance of the person, the Re-ID system, degrades due to the effect of non-uniform illumination [13]. Due to intraclass differences in the dress code of the people, the automatic Re-ID system gets confused to recognize the same person in the non-overlapping field-of-views [14]. Robustness and distinctiveness are desired in the feature representation to tackle physical appearance variations [9]. The Re-ID accuracy also depends on the human detection mechanism. The human detector with an adequately aligned bounding box and low false-positive brings more precision to the overall performance of the system [15].

Currently, researchers are developing descriptors that remain robust to noise and invariant to geometric and photometric variations [16]. Various combinations of color and texture descriptors have employed to improve the performance of the Re-ID models [17]. The biometric-based methods used the most discriminative visual cues of the human body that are the face, iris, and fingerprint. However, in the surveillance system, such biometric cues are not available due to low resolution, low scale, and considerable depth from the camera [13]. The appearance-based models depend on the overall physical appearance of the subject. The performance of such models depends only on the quality of the texture and color descriptors. The biometric features are highly sensitive to view angle, scale, illumination, and orientation of the person; therefore, the overall appearance is used for the case of Re-ID. In surveillance applications, the person's appearance also varies continuously with space and time [2]. In comparison to biometric, appearance-based features are more useful for a wide range of camera networks. The performance of the appearance-based methods degrades due to profound geometric and photometric changes in the person's appearance [18]. The appearance variability due to many sources, such as scale, viewpoint, pose, illumination, and occlusions in the non-overlapping field-of-view of the camera network, degrades the performance of an automatic Re-ID system [19].

The cumulative match curve is used as an evaluation criterion for the comparison of various techniques. The probe image is ranked through the matching method with the gallery of the already known individuals [8]. The unknown person images are labeled using the gallery of previously known individuals. There is no guarantee that the image of an unknown individual captured by a camera would have the same background, depth, view angle, illumination, and pose as the image already exists in the gallery for the same individual. The researchers so far have developed techniques that are robust to the geometric and photometric variations in the person's appearance. The feature extraction technique developed so far through handcraft and deep learning are dataset-specific. We propose a robust feature extraction technique that will provide better rank 1 matching accuracy on many commonly Re-ID datasets both in single-shot and multi-shot approaches.

2. Related Work

The person Re-ID problem raised when the research on the tracking started in the multi-camera surveillance system. Person Re-ID in non-overlapping field-of-view has been broadly

reviewed [8–14,18–22] in the last decade. The existing ReID techniques depend mostly on the features extraction and distance metric learning process. The Re-ID system consists of a gallery and probe set [11]. The gallery includes the sample images of the candidates whose identity is already known. The probe set consists of the test images that need to be identified based on gallery information. The feature extraction process used in the Person Re-ID approach is broadly categorized into two classes that are hand-crafted and non-hand-crafted.

The hand-crafted approach is one in which what features to extract from the input image to get better accuracy are manually decided. The drawback of this approach is that it is often fine-tuned toward the dataset. It works better for one dataset while its performance degrades with another dataset provided [8]. The most used features in this approach are color and texture. The spatial structure is divided into strips and grid; the features extracted from each grid cell are concatenated to get a global description. Many hand-crafted feature descriptors have been developed, consisting of various color and texture combinations. In [20], RGB, HSV, and YCbCr have been fused to describe the appearance based on color information within the detected bounding box. The Gabor and covariance related information are used in [23], and [24] to describe the visual cues within the pedestrian appearance. In [25], Symmetry-driven Accumulation of Local Features (SDLF) is produced, which decides the vertical symmetry before feature extraction. The symmetric axis is determined based on chrominance value; the texture features near the vertical symmetry are used to represent the pedestrian appearance. The SDLF has low complexity, but the matching rate is lower for rank-1 [26]. The Weighted Histogram of Overlapping Strips (WHOS) [27] descriptor is developed to extract a combined histogram of oriented gradients with the color feature of horizontally overlapping strips. The hierarchical Gaussian descriptor (GOG) has been reported in [28], which employs the localized color information to describe the human visual appearance. The GOG has better rank-1 accuracy; however, this method is highly sensitive to bounding box alignment problems [29]. The Bio-inspired Features (gBiCov) [30], is developed by combining the biologically inspired features with the covariance descriptor. The gBiCov has a higher facial Re-ID accuracy but a lower matching rate on the full-body bounding box of the pedestrian. The Local Maximal Occurrence (LOMO), along with a novel distance metric learning approach Cross-view Quadratic Discriminant Analysis (XQDA) [21], was developed in [21] for Pr-ID in the non-overlapping camera views. The Discriminative Accumulation of Local Features (DALF) [31] weights the local histogram representing the discriminative positions of the bounding box. A cross-view adaptation framework known as Camera coRelation Aware Feature augmentation (CRAFT) is introduced in [7] for person re-id. The CRAFT adaptively presents the feature augmentation by measuring the cross-camera correlation of the visual scenes.

The deep learning model automatically decides through backpropagation what features to be extracted. Various deep neural models reported in the literature [6,32,33] can re-identify the individual in the presence of extreme distortion. Various pre-defined models that are AlexNet, CaffeNet, GoogLeNet, VGG networks, ResNet, and SVDnet have also been used as feature extraction criteria for the person Re-ID. These models can easily be retrained on the data available in the surveillance model.

Many researchers have employed distance metric learning in their Re-ID models to improve the rank-1 matching accuracy. In addition to feature extraction, the distance metric learning can bring more improvement in the rank-1 performance of the Re-ID system. An overview of distance metric learning approaches is reported in [34]. Instead of using Euclidean or Manhattan distance, a supervised metric learning approach known as Mahalanobis [34] is used, which keeps the feature of the same class closer while that of a different type apart. Fisher discriminant analysis (FDA) and its local variant LFDA and KLFDA [35] have been developed in [18] to learn and reduce the raw features. The discriminative null space learning (NFST) [13], marginal Fisher analysis (MFA) [36], and kernel-based MFA (KMFA) [35] all employ Fisher optimization criteria to reduce the inter-class and increase the intra-class feature variations. The metric learning method Keep It Simple and Straightforward MEtric learning (KISSME) [37], Probabilistic Relative Distance Comparison

(PRDC) [20], and Pairwise Constrained Component Analysis (PCCA) [38] all learn the Mahalanobis distance with the constraints principle.

3. Proposed Re-ID Model

The Discriminative Local Features of Overlapping Stripes (DLFOS) model employs the robust color and texture information to achieve better accuracy. A novel texture descriptor OMTLBP_M is fused to HoG and Gaussian of color features to describe the appearance of the person with more robustness to variations. The details of the proposed DLFOS are as follows.

Features Extraction

The proposed Discriminative Local Features of Overlapping Stripes (DLFOS) consist of various color and texture descriptors. The Histogram of oriented Gradients (HoG) descriptor concatenated with the Gaussian of color features and the OMTLBP_M in the proposed Re-ID model.

The input bounding box is scaled to a size of 126×64 , then 36×64 pixel size horizontal strips with 50% overlap among them are extracted as shown in Figure 1. The HoG feature [15] vector is extracted for each horizontal strip as shown in Figure 1a. Each strip is transformed into a grid, where each grid cell has a size of 4×4 . The cells of the grid are described through a 32-dimensional feature vector. The nine orientations overall are described with the help of 27 variables. Moreover, one truncation and four texture variables also include the complete representation. Let f_1 be the HoG feature with dimensions of 3780 variables. The color-naming feature (CN) [39] of 11-variable vocabulary dimensions is collected from each grid's cell of the input segment. Suppose f_2 is the feature vector representing the color-naming. The Gaussian of Gaussian (GOG) [28] for the color feature is extracted for the 7×7 patch of each horizontal strip. The RGB, HSV, and YCbCr color values of the pixels are used to extract the Gaussian distribution of pixel features, as shown in Equation (2). The Gaussian collected from each patch is flattened and then vectorized according to the geometry of Gaussian. Moreover, the local Gaussians of each region are collected into a Gaussian region. After that, those Gaussian regions are further flattened to create a feature vector. Lastly, the feature vectors belonging to each region are concatenated into a single vector:

$$g = \{R, G, B, H, S, V, Y, Cb, Cr\} \quad (1)$$

For each 7×7 patch p , the patch of Gaussian $G(g; \mu_p \lambda_p)$

$$G(g; \mu_p \lambda_p) = \frac{\exp(-\frac{1}{2}(g - \mu_p)^T \lambda_p^{-1} (g - \mu_p))}{(2\pi i)^{d/2} |\lambda_p|} \quad (2)$$

$$\mu_p = \frac{1}{n_p} \sum_{i=1}^{n_s} (g^i) \quad (3)$$

$$\lambda_p = \frac{1}{n_p - 1} \sum_{i=1}^{n_s} (g - \mu_p)(g - \mu_p)^T \quad (4)$$

where $|\cdot|$ is the determinant of the covariance matrix λ_p . The μ_p denote the mean of the feature g extracted from the sample patch p . The n_p is the total number of pixels in the patch p . The OMTLBP_M_{P,R}^{riu2} [40] is also employed to bring rotation invariance in the proposed model. The sampling value $S = 8$ along with a set of radius $R = \{1, 2, 3\}$ are employed to extract the OMTLBP_M_{P,R}^{riu2} as presented in Equation (5). The OMTLBP_M operator from each 7×7 segment of the corresponding strip has a dimensionality of 200 variables to describe each input strip:

$$OMTLBP_M_{S,R}^{riu2} = \begin{cases} \omega(m_{r,c}^k, \mu_m), & \text{if } \chi_{S,R} \leq 2 \\ S + 1, & \text{otherwise} \end{cases} \quad (5)$$

$$\chi_{S,R} = |\omega(m_{r,c}^{S-1}, \mu_m) - \omega(m_{r,c}^0, \mu_m)| + \sum_{k=1}^{S-1} |\omega(m_{r,c}^k, \mu_m) - \omega(m_{r,c}^{k-1}, \mu_m)| \quad (6)$$

where $\omega(x, y) = 0$ for x less than y and 1 otherwise. The $m_{r,c}^k$ is the k th magnitude component with a mean value of μ_m , which is the mean value at each (r, c) coordinate of the images' segments. The S and R in Equation (5) are the sample points and radius with values 8 and 3, respectively. The final feature vector is a concatenated feature of HoG , GoG , and $OMTLBP_M_{S,R}^{riu2}$.

Figure 1. Framework of the proposed method.

The proposed feature extraction model is evaluated on several distance metric learning methods. However, better results have been achieved using Cross-view Quadratic Discriminant Analysis (XQDA) [21]. The multi-shot method is evaluated using SRID [41].

4. Experimental Results

The proposed DLFOS model in combination with XQDA and SRID is evaluated on VIPeR, CAVIARR4REID, and the Market1501 database, and the results have been summarized in the shape of the CMC curve and the table as follows.

4.1. Dataset

There are several publicly available datasets for person re-ID [42]. Many datasets have been developed and made publicly available by the researchers working in the field of security and surveillance. These datasets can be categorized into single-shot and multi-shot, depending on the scenario of the research. The single-shot are those databases that have a single probe and a unique gallery image. On the other hand, a multi-shot database has multiple images in the probe set and various pictures in the gallery. The probe and the gallery images may have a difference in camera view angle, illumination, scale, pose, and resolution. The proposed model is evaluated on VIPeR, CAVIARR4REID, and Market1501, which have been shown in Figure 2.

Figure 2. Sample images from the dataset used to test the proposed descriptor; Row 1 is the samples from VIPeR, Row 2 is the samples from CAVIARR4REID, and Row 3 is the samples from Market1501.

4.1.1. VIPeR

The viewpoint invariant pedestrian recognition (VIPeR) dataset contains 632 pedestrian image pairs. These images were collected from two cameras with a difference in view angle, pose, and lighting conditions. The images are scaled up to 128×48 pixel size. Performance is assessed by matching each test image in Cam A against Cam B in the gallery. Some of the pictures given as an example of the dataset are given as follows.

The proposed DLFOS model is evaluated on the VIPeR dataset, with numerous combinations of metric learning methods, and the results have been shown in Figure 3 and Table 1. Figure 3 presents the matching rate in percentage for the range of rank value within the range 1 and 30. The proposed feature extraction model, in combination with the XQDA metric learning method, provides 47.18 rank-1 performance. The DLFOS-XQDA results in 26.59, 18.18, 18.1, 17.42, 2.34, 1.91, 1.71, 1.69, 1.47, and 0.01 percent better rank-1 than PRDC [20], RPCCA [38], KPCCA [38], KISSME [37], KMFA [35], KLFDA [35], FDA [18], LFDA [18], MFA [36], and NFST [13]. The performance of the proposed DLFOS feature extraction model in comparison with other descriptors and the results have been summarized in Figure 4 and Table 3. The results show that the best combination of the descriptor and metric learning are WHOS-NFST, GOG-XQDA, gBiCov-NFST, LOMO-NFST, Color-Texture-NFST,

and DALF-KISSME with rank-1 performance 38.51, 41.20, 15.97, 33.92, 26.95, and 35.35 respectively. The proposed DLFOs-XQDA provides 31.21, 20.23, 13.26, 11.83, 8.67, and 5.98 from gBiCov-NFST, Color-Texture-NFST, LOMO-NFST, DALF-KISSME, WHOS-NFST, and GOG-XQDA, respectively.

Figure 3. CMC curves of various metric learning methods in combination with the proposed descriptor on a single-shot VIPeR dataset.

Figure 4. CMC comparison of various descriptors and metric learning on a single-shot VIPeR dataset.

Table 1. Top-performing metric learning combined with our proposed descriptor on the single-shot VIPeR dataset.

Descriptor	Metric Learning	Rank 1	Rank 5	Rank 10
Proposed	XQDA [21]	47.18	75.90	86.58
	MFA [36]	45.71	77.34	87.58
	FDA [18]	45.47	77.52	87.58
	LFDA [18]	45.49	77.50	87.60
	KLFDA [35]	45.27	76.19	86.53
	KMFA [35]	44.84	76.58	86.69
	RPCCA [38]	29.00	59.73	73.22
	KPCCA [38]	29.08	60.01	73.48
	NFST [13]	47.17	75.90	86.55
	KISSME [37]	29.76	54.54	67.21
	PRDC [20]	20.59	38.82	48.51

4.1.2. CAVIAR4REID

Context-Aware Vision using the Image-based Active Recognition (CAVIAR4REID) dataset contains pedestrians' images in a shopping center. It consists of 72 different people with 1221 images taken from two cameras from various views. From these 72 people, 22 come from the same camera and 50 from both of them. Each person's images vary from two to five images with various sizes from 17×39 to 72×144 . The proposed DLFOS feature extraction model, along with numerous combinations of distance metric learning, is tested on the CAVIAR4REID dataset, and the results have been presented in Figure 5 and Table 2. Figure 6 shows the matching rate in percentage for the range of rank value within the range 1 and 25. The proposed feature extraction model, in combination with the XQDA metric learning method, provides a 64.4% matching rate at rank-1. The DLFOS-XQDA results in 28.4, 26, 26, 25.6, 22, 18, 13.6, 11.2, 8.8, and 8.0 percent better matching rate vs. rank-1 than MFA [36], KMFA [35], KPCCA [38], KLFDA [35], NFST [13], PCCA [38], KISSME [37], PRDC [20], FDA [18], and LFDA [18], respectively. The DLFOS-XQDA is tested on the SRID ranking method for the multi-shot CAVIAREID dataset. The results in Table 3 show that the best combination of the descriptor and metric learning are gBiCov-KISSME, Color-Texture-XQDA, WHOS-XQDA, LOMO-KISSME, and GOG-KISSME. The proposed DLFOS-XQDA provides 34, 14, 9.2, 6, and 4.8 percent higher matching rates than gBiCov-KISSME, Color-Texture-XQDA, WHOS-XQDA, LOMO-KISSME, and GOG-KISSME.

Table 2. Top performing metric learning combination with our proposed descriptor on multi-shot shot CAVIAR4REID dataset.

Descriptor	Metric Learning	Rank 1	Rank 5	Rank 10
Proposed	XQDA [21]	64.4	90.8	98.0
	LFDA [18]	56.4	86.0	95.6
	PCCA [38]	46.4	75.6	93.2
	KISSME [37]	50.8	78.0	90.4
	FDA [18]	55.6	83.6	95.6
	PRDC [20]	53.2	77.2	89.6
	MFA [36]	36.0	63.2	81.6
	KMFA [35]	38.4	70.0	89.6
	KPCCA [38]	38.4	65.6	78.4
	NFST [13]	42.4	78.4	92.8
	KLFDA [35]	38.8	68.4	88.0

Figure 5. CMC comparison of the proposed descriptor in combination with multiple metric learning methods on a multi-shot Caviar dataset.

Figure 6. CMC comparison of the proposed descriptor in combination with multiple metric learning methods on a multi-shot Caviar dataset.

Table 3. Comparison of various descriptor and metric learning combination on VIPeR, Caviar, and Market1501 dataset.

Descriptor	Metric Learning	VIPeR Dataset			Caviar Dataset			Market1501 Dataset		
		Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10	Rank 1	Rank 5	Rank 10
WHOS [27]	XQDA [21]	28.78	54.91	67.63	55.20	79.60	89.60	42.34	63.45	71.20
	NFST [13] _{exp}	38.51	68.53	81.84	50.00	77.20	88.00	56.68	77.46	84.56
	KISSME [37]	23.43	44.91	56.50	50.00	74.00	85.20	10.72	19.15	23.55
GOG [28]	XQDA [21]	41.20	71.01	82.10	59.20	82.40	96.00	49.05	69.66	78.86
	NFST [13] _{exp}	36.53	67.15	79.48	52.00	75.60	93.20	52.79	73.55	80.70
	KISSME [37]	17.09	35.85	47.20	59.60	85.60	96.80	-	-	-
gBiCov [30]	XQDA [21]	08.26	23.29	34.07	10.80	40.40	61.20	02.17	08.19	13.21
	NFST [13] _{exp}	15.97	39.26	54.40	16.80	36.80	65.20	20.99	39.96	48.60
	KISSME [37]	11.25	26.55	36.69	30.40	54.00	69.20	17.87	35.36	45.28
LOMO [21]	XQDA [21]	32.86	62.44	75.22	56.40	85.20	95.60	-	-	-
	NFST [13] _{exp}	33.92	63.47	77.26	56.00	87.60	97.60	-	-	-
	KISSME [37]	14.84	34.64	46.25	58.40	86.80	93.60	-	-	-
Color Texture	XQDA [21]	18.75	43.58	56.25	50.40	76.80	89.20	34.53	56.00	65.53
	NFST [13] _{exp}	26.95	57.59	72.12	45.20	75.20	90.00	24.52	43.05	53.95
	KISSME [37]	13.31	31.99	43.75	47.60	73.60	88.80	17.84	30.55	37.29
DALF [31]	KISSME [37]	35.35	62.03	73.48	-	-	-	-	-	-
Proposed	XQDA [21]	47.18	75.90	86.58	64.40	90.80	98.00	62.68	81.18	86.76

4.1.3. Market1501

The Market1501 database includes a diverse set of test subjects with multiple snaps belonging to six disjoint cameras. This dataset also contains 2793 false positives of the deformable part model. The bounding boxes are more challenging than CUHK03. Later on, 500K distractors included bringing diversity within the database. The authors employed mAP to evaluate the test Re-ID models on the proposed database. The proposed DLFOS-XQDA combination, when evaluated on the Market1501 database, achieved a 62.68% matching rate over rank-1. Figure 7 presents the percentage matching rate over the rank values ranges between 1 and 25. Table 3 shows that WHOS-NFST, GOG-NFST, Color-Texture-XQDA, and gBiCov-NFST combinations give high matching rates in comparison to other combinations. The DLFOS-XQDA provides 41.96%, 28.15%, 9.89%, and 6% higher matching rates than gBiCov-NFST, Color-Texture-XQDA, GOG-NFST, and WHOS-NFST for rank-1.

Figure 7. CMC comparison of the proposed descriptor in combination with multiple metric learning methods on the Market1501 dataset.

5. Conclusions

This paper mainly proposes a novel feature extraction for Person Re-Identification (ReID). In this method, we accumulated the distinctive features representing the silent regions in the overall appearance. We offer a fused feature extraction model consisting of HoG, Gaussian of color, and the discriminative novel texture descriptor CJLBP_M. Inspired by the WHOS framework, the detected bounding box is divided into horizontally overlapped strips before the feature extraction process. The selection of rich features in the proposed DLFOS model has brought robustness to noise and invariance against variations in illumination, scale, and orientation of the person's appearance. To validate the performance of our proposed DLFOS, we have conducted several experiments on three publicly available datasets. The outcome of the experiments indicates that the proposed DLFOS model, along with XQDA metric learning, offers a 5.98%, 4.8%, and 6% high rank-1 matching rate compared to many recently reported works when tested on publicly available ViPeR, CAVIAR, and Market1501 databases, respectively. In the future, we would like to employ the feature reduction model to improve the computational complexity by reducing the dimensionality of the extracted feature vector.

Author Contributions: Conceptualization, F.; Data curation, F. and M.J.K.; Formal analysis, F., M.J.K. and M.R.; Funding acquisition, M.R.; Investigation, F. and M.R.; Methodology, M.R.; Resources, M.R.; Supervision, M.J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding this publication.

References

1. Gao, C.; Wang, J.; Liu, L.; Yu, J.G.; Sang, N. Superpixel-Based Temporally Aligned Representation for Video-Based Person Re-Identification. *Sensors* **2019**, *19*, 3861. [[CrossRef](#)] [[PubMed](#)]
2. Fawad; Jamil Khan, M.; Rahman, M.; Amin, Y.; Tenhunen, H. Low-Rank Multi-Channel Features for Robust Visual Object Tracking. *Symmetry* **2019**, *11*, 1155. [[CrossRef](#)]
3. Borgia, A.; Hua, Y.; Kodirov, E.; Robertson, N.M. Cross-view discriminative feature learning for person re-identification. *IEEE Trans. Image Process.* **2018**, *27*, 5338–5349. [[CrossRef](#)] [[PubMed](#)]
4. Wu, L.; Hong, R.; Wang, Y.; Wang, M. Cross-entropy adversarial view adaptation for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**. [[CrossRef](#)]
5. Liu, F.; Zhang, L. View Confusion Feature Learning for Person Re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–3 November 2019; pp. 6639–6648.
6. Bai, X.; Yang, M.; Huang, T.; Dou, Z.; Yu, R.; Xu, Y. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognit.* **2020**, *98*, 107036. [[CrossRef](#)]
7. Chen, Y.C.; Zhu, X.; Zheng, W.S.; Lai, J.H. Person re-identification by camera correlation aware feature augmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 392–408. [[CrossRef](#)] [[PubMed](#)]
8. Gou, M.; Wu, Z.; Rates-Borras, A.; Camps, O.; Radke, R.J. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 523–536.
9. Shen, Y.; Xiao, T.; Li, H.; Yi, S.; Wang, X. End-to-end deep kronecker-product matching for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6886–6895.
10. Li, D.; Chen, X.; Zhang, Z.; Huang, K. Learning deep contextaware features over body and latent parts for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 384–393.
11. Li, Z.; Chang, S.; Liang, F.; Huang, T.S.; Cao, L.; Smith, J.R. Learning locally-adaptive decision functions for person verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3610–3617.
12. Zhong, Z.; Zheng, L.; Cao, D.; Li, S. Re-ranking person reidentification with k-reciprocal encoding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1318–1327.

13. Zhang, L.; Xiang, T.; Gong, S. Learning a discriminative null space for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1239–1248.
14. Shen, Y.; Li, H.; Xiao, T.; Yi, S.; Chen, D.; Wang, X. Deep groupshuffling random walk for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2265–2274.
15. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [[CrossRef](#)] [[PubMed](#)]
16. Fawad; Rahman, M.; Khan, M.J.; Asghar, M.A.; Amin, Y.; Badnava, S.; Mirjavadi, S.S. Image Local Features Description Through Polynomial Approximation. *IEEE Access* **2019**, *7*, 183692–183705. [[CrossRef](#)]
17. Saeed, A.; Fawad; Khan, M.J.; Riaz, M.A.; Shahid, H.; Khan, M.S.; Amin, Y.; Iqbal, J.; Tenhunen, H. Robustness-driven hybrid descriptor for noise-deterrent texture classification. *IEEE Access* **2019**, *7*, 110116–110127. [[CrossRef](#)]
18. Pedagadi, S.; Orwell, J.; Velastin, S.A.; Boghossian, B.A. Local fisher discriminant analysis for pedestrian re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3318–3325.
19. Chen, D.; Yuan, Z.; Chen, B.; Zheng, N. Similarity learning with spatial constraints for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1268–1277.
20. Zheng, W.; Gong, S.; Xiang, T. Person re-identification by probabilistic relative distance comparison. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 649–656.
21. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2197–2206.
22. Asghar, M.A.; Khan, M.J.; Fawad; Amin, Y.; Rizwan, M.; Rahman, M.; Badnava, S.; Mirjavadi, S.S. EEG-Based Multi-Modal Emotion Recognition using Bag of Deep Features: An Optimal Feature Selection Approach. *Sensors* **2019**, *19*, 5218. [[CrossRef](#)] [[PubMed](#)]
23. Ma, B.; Su, Y.; Jurie, F. Discriminative image descriptors for person reidentification. In *Person Re-Identification*; Springer: London, UK, 2014; pp. 23–42.
24. Ma, B.; Su, Y.; Jurie, F. BiCov: A novel image representation for person re-identification and face verification. In Proceedings of the British Machine Vision Conference, Guildford, UK, 3–7 September 2012; p. 11.
25. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
26. Yao, H.; Zhang, S.; Hong, R.; Zhang, Y.; Xu, C.; Tian, Q. Deep representation learning with part loss for person re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 2860–2871. [[CrossRef](#)] [[PubMed](#)]
27. Lisanti, G.; Masi, I.; Bagdanov, A.D.; Del Bimbo, A. Person re-identification by iterative re-weighted sparse ranking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1629–1642. [[CrossRef](#)] [[PubMed](#)]
28. Matsukawa, T.; Okabe, T.; Suzuki, E.; Sato, Y. Hierarchical gaussian descriptor for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1363–1372.
29. Dai, J.; Zhang, P.; Wang, D.; Lu, H.; Wang, H. Video person re-identification by temporal residual learning. *IEEE Trans. Image Process.* **2018**, *28*, 1366–1377. [[CrossRef](#)] [[PubMed](#)]
30. Ma, B.; Su, Y.; Jurie, F. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vis. Comput.* **2014**, *32*, 379–390. [[CrossRef](#)]
31. Matsukawa, T.; Okabe, T.; Sato, Y. Person re-identification via discriminative accumulation of local features. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 3975–3980.
32. Chen, T.; Ding, S.; Xie, J.; Yuan, Y.; Chen, W.; Yang, Y.; Wang, Z. Abd-net: Attentive but diverse person re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–3 November 2019; pp. 8351–8361.

33. Shen, Y.; Li, H.; Yi, S.; Chen, D.; Wang, X. Person re-identification with deep similarity-guided graph neural network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 486–504.
34. Wu, L.; Wang, Y.; Gao, J.; Li, X. Deep adaptive feature embedding with local sample distributions for person re-identification. *Pattern Recognit.* **2018**, *73*, 275–288. [[CrossRef](#)]
35. Xiong, F.; Gou, M.; Camps, O.; Sznai, M. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 1–16.
36. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *29*, 40–51. [[CrossRef](#)]
37. Koestinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Large scale metric learning from equivalence constraints. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2288–2295.
38. Mignon, A.; Jurie, F. Pcca: A new approach for distance learning from sparse pairwise constraints. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2666–2672.
39. Danelljan, M.; Khan, F.S.; Felsberg, M.; Weijer, J.V.D. Adaptive color attributes for real-time visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
40. Fawad; Khan, M.J.; Riaz, M.A.; Shahid, H.; Khan, M.S.; Amin, Y.; Loo, J.; Tenhunen, H. Texture Representation through Overlapped Multi-oriented Tri-scale Local Binary Pattern. *IEEE Access* **2019**, *7*, 66668–66679. [[CrossRef](#)]
41. Karanam, S.; Li, Y.; Radke, R. Sparse re-id: Block sparsity for person re-identification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2015, Boston, MA, USA, 7–12 June 2015.
42. Gong, S.; Liu, C.; Ji, Y.; Zhong, B.; Li, Y.; Dong, H. Image Understanding-Person Re-identification. In *Advanced Image and Video Processing Using MATLAB*; Springer: Cham, Switzerland, 2019; pp. 475–512.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).