IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 16, 2018, accepted November 12, 2018, date of publication December 7, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2882254

Person Re-Identification Using Hybrid Representation Reinforced by Metric Learning

NAZIA PERWIAZ[1], MUHAMMAD MOAZAM FRAZ[1,2,3], (Member, IEEE), AND MUHAMMAD SHAHZAD[1], (Member, IEEE)

[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad 44000, Pakistan
[2]The Alan Turing Institute, British Library, London NW1 2DB, U.K.
[3]Department of Computer Science, University of Warwick, Coventry CV47AL, U.K.

Corresponding author: Nazia Perwiaz (nazia.perwaiz@seecs.edu.pk)

ABSTRACT Person Re-Identification (Re-Id) is among the main constituents of an automated visual surveillance system. It aims at finding out true matches of a given query person from a large repository of non-overlapping camera images/videos. In this paper, we have proposed an efficient Re-Id approach that is based on a highly discriminative hybrid person representation which combines the low-level hand-crafted appearance based features together with the mid-level attributes and semantic based deep features. The low-level hand crafted features are extracted by using hierarchical Gaussian and local histogram distributions in different color spaces. These features incorporate discriminative texture, shape and color information which is invariant to distractors, e.g., variations in pose, viewpoint and illumination, and so on. The mid-level attribute based deep features are extracted to incorporate contextual- and semantic-based information. The feature space is optimized and self-learned using cross-view quadratic discriminant analysis and multiple metric learning, with the aim to reduce the intra-class differences and increase the inter-class variations for robust person matching. The proposed framework is evaluated on publicly available small scale (VIPeR, PRID450s, and GRID) and large scale (CUHK01, Market1501, and DukeMTMC-ReID) person Re-Id datasets. The experimental results show that the hybrid hand-crafted and deep features outperformed the existing state-of-the-art in approaches in the unsupervised paradigm.

INDEX TERMS Person re-identification, visual surveillance, hand-crafted features, hybrid person representation, metric learning, deep learning.

I. INTRODUCTION

The active research in the combined areas of computer vision, machine learning and internet of things (IoT) has formulated the concept of smart cities in the past few years [1]. Specifically the automated surveillance is one among the most important aspect of smart cities. A robust and automated visual surveillance system is very helpful in ensuring the public safety and security in the smart environments [1]. In this context, person Re-Identification [2] is one among the important integral components of visual surveillance system. Person Re-Identification (Re-Id) is the process of automatic matching of a human in the multi-camera network with non-overlapping filed-of-views.

It determines that the images acquired by disjoint camera views belongs to the same or different individual with the aim to assigning the same identifier to different instances of the same person.

In multi-camera network with disjoint views, when an individual is moved from one camera view to another, the person Re-Id system aims to recognize the detected person as the same individual which had been identified earlier in the former camera view. As illustrated in Figure 1, the Re-Id pipeline begins with person detection in the scene followed by the representation with a unique descriptor [3]. The scene may contain multiple persons and other objects as well. The detected person is treated as query image and is matched with the set of images stored in the repository (termed as gallery) for label assignment [2].

Typically, the surveillance videos captured in unconstrained real world scenarios are of low resolution, affected by

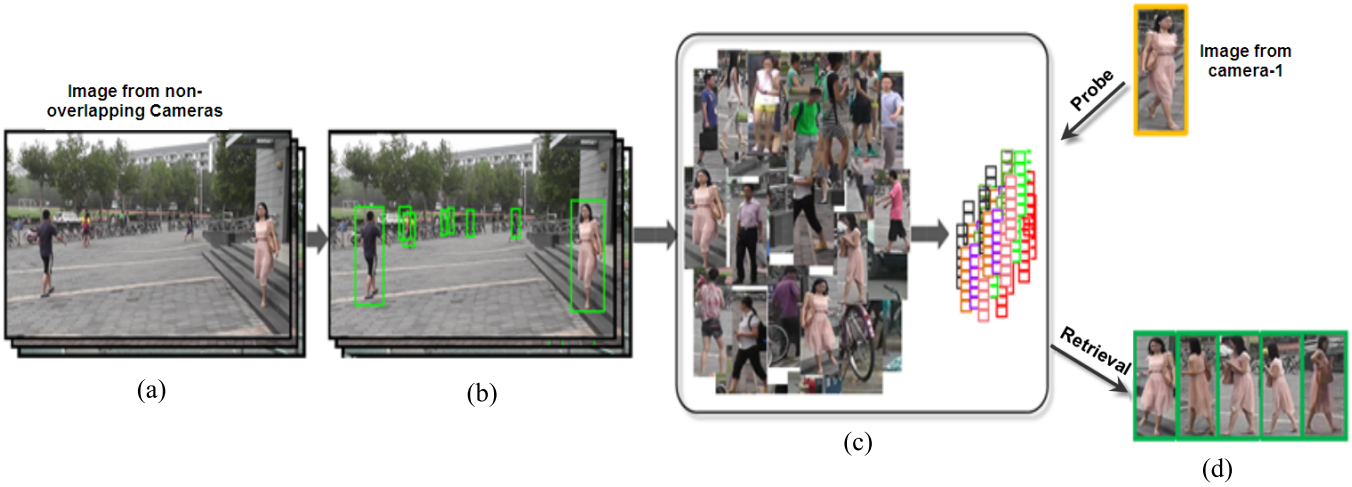

FIGURE 1. Person Re-Identification Pipeline: (a) Input Images/video feed from non-overlapping camera views; (b) Person detection in natural scene; (c) Person representation using discriminative features; (d) Person matching using similarity/dissimilarity learning.

environmental factors (e.g. rain, snow, storm etc.). Moreover, they are also influenced by temporal variation of lighting conditions in day/night timings and the change in viewing angles [4]. Person Re-id is not a trivial problem because of the challenges involved due to high intra-class variability, dramatic inter-class similarity, difference is illumination and contrast as well as pose variability [2]. The occlusions and cluttered background in natural scenes has made the robust re-id further difficult. These constraints make it even harder to compute discriminative biometric cues involving facial and gait features. In this context, the human visual appearance may be the most important clue to look for the robust person re-id task [4]. However, under the influence of complex unconstrained surveillance scenarios and non-rigid nature of human body, the visual appearance of same person looks dissimilar in multiple views. An efficient person re-id algorithm should handle such issues by accurate person detection, robust person representation (with highly discriminative features) and matching. Existing supervised learning approaches handle these problems efficiently [5]. Novel deep learning based methods performed even better [6], [7], but these approaches needs ample amount of labelled person images for training. In real world unconstrained scenarios, annotating (tens of) thousands of images is un-scalable, laborious and tedious task [8].

A number of semi-supervised and unsupervised approaches have been proposed to overcome this problem [9]–[13]. However due to non-availability of labelled data, these methodologies generally relies on feature variance, which often results in reducing the discriminativity among the features. Moreover, the high variance across disjoint camera views is rather difficult to model without labelled data.

To this end, the unsupervised methodologies should emphasize on extraction of robust and discriminative feature representation and employ efficient similarity/matching functions. In this paper, we have presented an unsupervised person re-id framework by introducing a hybrid person representation where the hand crafted features are combined with the features extracted from deep learning in order to formulate a highly discriminative person representation. Subsequently, the matching process is formulated in a metric learning framework such that the feature space is self-learned using Cross-view Quadratic Discriminant Analysis (XQDA) [14] and Multiple Metric Learning (MML) [15] approach. The aim is to reduce the intra-class differences and increase the inter-class variations. Overall, the main contribution of the proposed methodology are two-fold, that are summarized as follows:

- A highly discriminative person representation is presented by coalescing the hand crafted appearance based features, together with the hierarchical deep features involving person attributes and scene semantics.
- The high dimensional discriminative feature representation is optimized to retain the most distinctive features which are utilized in learning similarity/dissimilarity using state-of-the-art metric learning techniques.
- We evaluated our approach on both small scale and large scale public Re-Id datasets. Our proposed solution obtained significant improvement in person Re-Id performance over existing state-of-the-art Re-Id methodologies on the small scale datasets (VIPeR, PRID450s and GRID as well as on the large scale datasets (CUHK01, Market1501 and DukeMTMC-ReID). The code will be published with the article.

The paper is organized as follows: The existing mythologies of person Reid are illustrated in Section II. Section III presents the proposed framework in detail. The performance measures and the quantitative and qualitative results are presented in Section IV. The experimental evaluations are discussed and analyzed in Section V. In Section VI the conclusion and potential future research directions are elaborated.

II. RELATED WORK

Majority of research in the field of person Re-Id focuses on appearance based person representation [16]. Such representations include utilization of low level features information (e.g., color and/or texture histogram, geometric structures etc.) to provide solution to the Re-Id problem. Although they provide a viable solution especially in short term scenarios but tend to have lower accuracies owing to the issues including high variations in illumination, the changes in viewpoints/ poses in the cross-camera images with different settings and the back ground noise [16]. The solution of Re/Id thus requires robust person representation that is invariant to these problems. In this context, several feature descriptors have been proposed that further use either supervised [16], semi-supervised [17] or unsupervised machine learning techniques [18] for person matching.

Few approaches focused on the holistic view of the local features (color and pattern channels) of the images to address the Re-Id challenges, for example the view point variations are handled by the Ensembles of Local Features (ELF) [19] where a feature space of three colors and nineteen texture channels is defined and the AdaBoost algorithm is used for person matching. Zheng *et al.* [20] used the ELF for person representation and probabilistic relative distance comparison model for person matching. Moreover in order to extend the features discrimination, the covariance information is utilized along with the different color channels features and the descriptive sub-features (spatiograms) by Zeng *et al.* [21] and the L1 distance is used for the matching purpose. However such global representations result in the loss of local region wise discriminative information.

In order to preserve the discriminative information of the local regions/patches few methodologies focused on dividing the whole image into smaller regions, patches or strips before extraction of the features. For instance, In [22] the patch wise color features are extracted and are assembled into the hierarchical Gaussian features and for person verification the Locally Adaptive Decision Function (LADF) is proposed. Furthermore the correlated information among the local regions/patches is retained by preserving the local maximum occurrences information [14] using sliding windows mechanism and for person matching the metric learning is employed. However the accumulation of small overlapping patches features result in very high dimensional person representations.

The high dimensionality of the features is addressed either by focusing on the salient regions instead of whole feature space or by making use of mid-level semantic features or by proposing robust metric learning techniques. The use of salient regions/ features somehow resembles the human vision system. The discriminative features are enhanced in [22] by combining the unsupervised saliency learning with color and SIFT features. Moreover the RankSVM [23] approach is used to globally learn a subspace where the highest rank is given to the potential true match after matching of cross camera patches and saliency. Likewise the salient regions in an image are highlighted by the KEPLER in [15] and Multiple Metric Learning (MML) is applied for person matching [12]. The use of saliency is an implicit way to handle the background noise that is a very common Re-Id challenge.

The background noise is explicitly handled by the body parts based Re-Id approaches. For instance in the Symmetry Driven Accumulation of Local Features (SDALF) [9] the body parts (head, torso, legs) are segmented as image foreground using symmetry property of a person and the part wise local features resulted in view invariant feature descriptor and for matching Euclidean distance is used. Few other body-part techniques utilized pose estimation and human parsing techniques for foreground background segmentation. [24]–[26].

Scalability is another challenging aspect of the Re-Id system and is addressed by large-scale metric learning i.e. Keep It Simple and Straight forward Metric learning (KISSME) [27] and by Dual-Regularized KISS Metric Learning [28]. Yang et al. considered the Re-Id problem as a multitask distance metric learning problem and proposed the multitask maximally collapsing metric learning (MTMCML) [29]. Few Re-Id solutions addressed the scalability issue either by using unsupervised machine learning approaches for person matching i.e. simple nearest neighbor metric learning [27] and logistic metric learning approach using accelerated proximal gradient (MLAPG) [30] or by using unsupervised learning of visual co-occurrence descriptors for person representations [31]. Moreover by utilizing the saliency information to represent a person and then using KISS metric learning for person matching in [32] handled the background noise and scalability issues simultaneously.

Another alternative for scalable Re-Id is to make use of the middle level semantic attributes based person representation by utilizing the low-level information in images. The semantic attributes define the person in few parameters e.g. upper body color, hair color, shoe type etc. [33]–[35]. The mid-level filters (MLFL) [36] are learnt in an unsupervised way by clustering of patches and for person matching RankSVM is employed.

In contrast, the deep-learning based Re-Id solutions generally depend on the volume of available ground truth/ labeled training data (in addition to the architecture of deep network). If large amount of labeled training data is available, the deep learning based Re-Id approaches outperform conventional hand-crafted techniques with a large margin [37]–[41]. For end-to-end deep architectures, most commonly used loss functions are Siamese loss, Triplet loss and its variants [42], [43]. However due to the need of huge labelled training data the end-to-end deep approaches do not work well for small scale data. Likewise [43]–[46] proposed attributes based end-to-end deep learning solution for person Re-ID. For instance in [45], the Resnet is used as a pre-trained model and updated its parameters/ weights according to pedestrian attributes learning task. After training

of the model only deep features are used to perform person re-id. The results show that very good re-id performances are seen only for large scale benchmarks of person re-id i.e. Market1501 [47] and DukeMTMC-ReID [48]. However the foremost limitation of pure deep learning techniques is the dependency upon large amount of labeled data, that's why no small scale re-id benchmark is quoted in this paper. Whereas our work differs from [42] as our descriptor comprising complementary low-level hand crafted features is equally promising for the small scale as well as for the large scale re-id benchmarks.

Nevertheless the need of huge amount of labeled training data is an unavoidable limitation of the deep learning. Therefore hybrid approaches have been explored to combine the pros of both hand-crated and deep learning techniques. A hybrid approach [49] fed the fisher vectors (comprising of hand-crafted SIFT and color features) into deep neural network to learn the linearly separable deep space and the optimization of Linear Discriminant Analysis (LDA) based objective function is used to solve the person retrieval problem. Moreover Zhu *et al.* [50] proposed a hybrid similarity function learning mechanism (DHSL) where the features are extracted from a light weight (03 layered) convolutional neural network and the similarity score is learnt using hybrid similarity function. In an another hybrid approach [51] the low level feature i.e. Ensemble of Local Features (ELF) are embedded into CNN features before passing them to final softmax loss layer. During the back propagation phase of training process the weights of CNN are updated on the basis of combined loss of hand-crafted and deep features instead of single loss of deep features. Although the addition of low level features somehow optimized the final loss, but the vanilla CNN features for small scale re-id benchmarks cannot contribute significantly using this methodology. Another major limitation of vanilla CNN features is that the special coherency is lost during the training process. Our work differ from [51] in two folds: 1) instead of using vanilla CNN features, we customized the final layers of our CNN model and fine-tuned the weights of model using attributes and combination-attributes labels. Therefore the attained CNN features significantly contributed in our proposed person representation. 2) Instead of person representation based re-id through default softmax loss of CNN architectures we employed more sophisticated metric learning approaches to segregate similar and dissimilar images. Features fusion approach is also used in [52] and [53].

Inspired from the hybrid formulations, the proposed framework in this paper introduces a discriminative hybrid person representation by extracting hand-crafted and deep learning based person features. The robust hybrid person representation is followed by preservation of the most significant features to increase the performance of feature matching while maintaining the discriminativity. For robust person matching unsupervised metric learning approaches are employed to make the Re-Id solution scalable.

III. PROPOSED METHODOLOGY

The performance of Re-id solutions mainly depends on robust person representation and efficient person matching. The person matching is further reliant on efficient reduction of intra-person differences and the increase of Inter-person differences.

In this work our main focus is to develop a Re-Id solution by proposing highly discriminative person representation and its integration with powerful person matching techniques. The task is divided into three phases. The first phase is comprised of discriminative hybrid person representation by employing different hierarchical strategies to summarize hand-crafted and deep features to integrate the powers of different type of features as shown in Figure 2-b.1. In the next phase the highly significant and covariant dimensions of the original high dimensional representations are preserved by utilizing manifold learning mechanism as illustrated in Figure 2-b.2.

And finally two state-of-the-art metric learning techniques are used to learn the features sub-space for robust person matching by focusing on intra-person and inter-person differences as shown in Figure 2-c.

The architectural flow and the functional details of the hand-crafted features, deep features and the metric learning techniques are discussed in the respective sub-sections.

A. PERSON REPRESENTATION

After receiving the input of person images Figure 2-a., the aim of person representation phase if to extract the highly discriminative information from the images. The input image is simultaneously fed into three different channels to focus on the discriminative information in diverse styles. Two discriminative and complementary hand-crafted features are extracted by employing two different hierarchical strategies [10], [13], and the deep features are retrieved by summarizing the attribute based hierarchical information [54].

Among the two hand-crafted techniques, one comprehends of the hierarchical Gaussian distributions of different color channels by emphasizing on the mean information, whereas the other technique pools the strip-wise local features that include multiple color space and texture features in the form of weighted histogram. In addition to the hand-crafted features the attributes based deep features are extracted by fine-tuning of a convolutional neural network (CNN) based deep model [55] on pedestrian dataset.

After a comparative analysis of a large set of features, the selection of hand-crafted and deep features is made such that each of the selected feature contributes towards capturing the diverse aspects of a person in addition to addressing the background noise, variation in illumination, poses, and viewpoints at different scales. All three types of features are integrated to propose the final person representation. The detail of all three types of features is given in sub-sections.

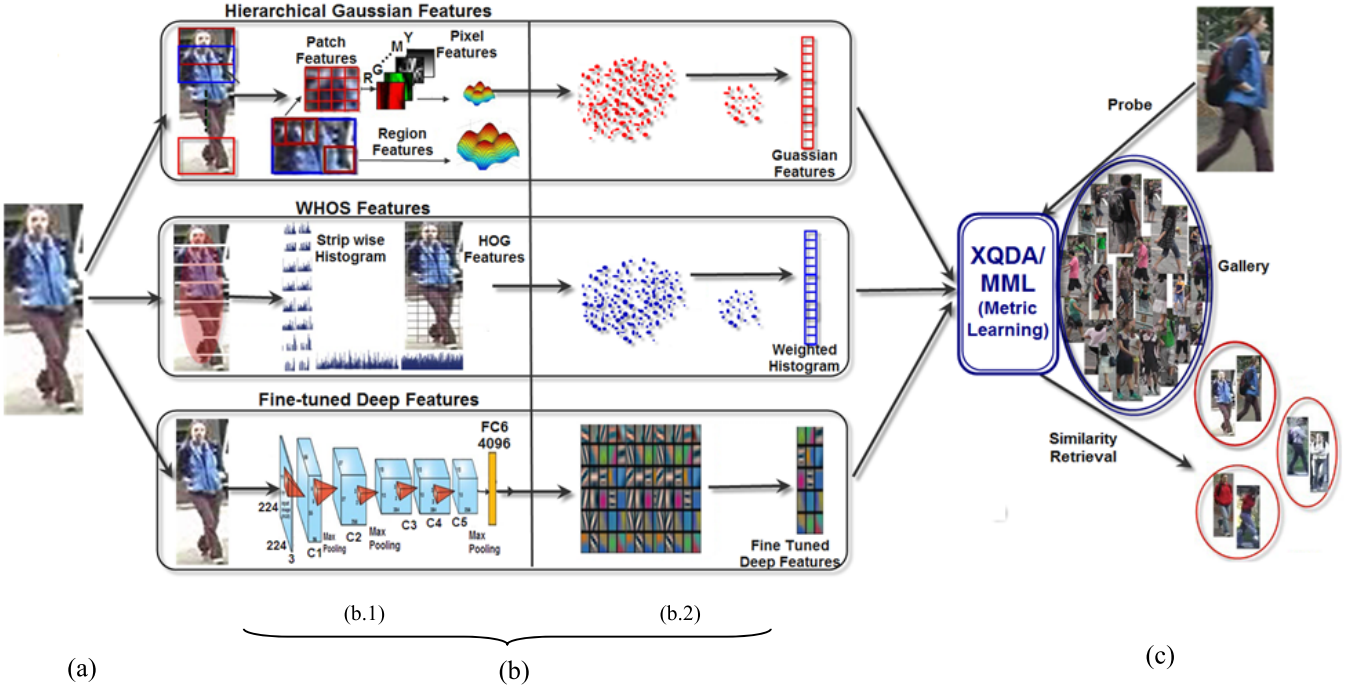

FIGURE 2. Proposed Architecture for Person Re-Identification; (a) Input Image (b) Person Representation (b.1) Features Extraction (b.2) Dimensionality Reduction (c) Unsupervised Distance Metric Learning and Person Matching.

1) HIERARCHICAL GUASSIAN FEATURES

The hierarchical Gaussian features are capable in capturing the local covariance information along with preserving the local mean of the color values [56]. Similar to LOMO [14], it handles the illumination variations by using Retinex algorithm [57] in the preprocessing step to equalize the different illumination conditions of cross-camera images. After illumination normalization, a person is divided into multiple overlapping regions and patches as illustrated in Figure 2-b.1. Moreover the detailed mechanism of local features extraction is also shown in the Figure 2. The features are extracted for different color channels i.e RGB, HSV, LAB. Each pixel i in a patch is described by eight different values that include the pixel location y, its gradient magnitude in four different orientations $M_\theta^\circ(0^\circ, 90^\circ, 180^\circ, 270^\circ)$ and the values of its respective color channels as shown in equation (1). (RGB color-space is shown in the equation, the same formula is used for rest of color spaces)

$$f_i = (y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B)^T \tag{1}$$

In the first stage pixel level features are extracted within the patch. After that the patch features are generated by summarizing the pixels features via covariance matrices and Gaussian distribution. The patch Gaussian computation is given in equation (2).

$$\mathcal{N}(f; \mu_s, \Sigma_s) = \frac{\exp\left(-\frac{1}{2}(f-\mu_s)^T \Sigma_s^{-1}(f-\mu_s)\right)}{(2\pi)^{\frac{d}{2}}|\Sigma_s|} \tag{2}$$

Where μ_s is the mean vector computed from all the pixels of given patch and $\sum_s$ is the covariance matrix of the given patch s.

The Gaussian distributions are flattened by converting the Riemannian manifold space into the Euclidean space for the feasible application of Euclidean algorithm in the Euclidean space. For this purpose the features are projected into tangent space through symmetric positive definite (SPD) matrix.

And finally the region features are generated by summarizing the patches features using same methodology through the covariance matrices and Gaussian distribution. To handle the background noise, different weightages are assigned to the patches based on their distance from the central axes. The closer patches are assigned higher weights whereas less weight is assigned to the patches that are away from axes and closer to the bounding boxes edges. For G regions, the G numbers of region-wise feature vectors z are obtained. The final feature vector is generated by concatenating all Gaussian regions as shown in equation (3).

$$z = (z_1, z_2, z_3, \ldots\ldots\ldots z_G) \tag{3}$$

This accumulation of hierarchical Gaussian features preserves the local distinctiveness along with region wise distinctiveness. The resultant feature descriptor of size 27,626 is attained to represent a person. Since the covariance information is preserved, the proposed technique provides better accuracy estimates. The effectiveness of the hierarchical Gaussian features extracted from different color channels will be discussed in the Section V.

2) WEIGHTED HISTOGRAM OF OVERLAPPING STRIPS

A sparse feature extraction scheme is adopted by pooling of the stripped features, where higher weights are assigned to the central part of person image and lower weights to the background region by employing center support Epanechnikov mask [10], [29].

The input images are divided into eight non-overlapping and seven overlapping horizontal strips. The non-overlapping strips highlight the strip wise discriminative information whereas the overlapping strips capture the correlation among the adjacent strips. For each strip the color and pattern features are encompassed. The spatial histograms of two color-spaces are extracted i.e. HS (Hue-Saturation) and RGB histograms. In addition to color spaces the local texture patterns are captured by Histogram of Oriented Gradient (HOG) and LBP (Local binary pattern) from each strip.

After assigning the weights to the pixels i.e. the pixels close to the central axis of the image are given more weightages as compared to the boundary pixels the stripped features are pooled together to generate the final feature descriptor of a person. Combining 2880 dimensional color features with the 2438 dimensional pattern features, the final descriptor resulted in the size of 5318 dimensions. Use of the sparse strips significantly reduced the person representation size.

3) FINE-TUNED DEEP FEATURES

A neural network is comprised of a bunch of neurons, activations and learnable parameters. Neural network comprising multiple hidden layers is termed as deep neural network. Training a huge size deep neural network from scratch is computationally very expensive. A feasible alternative is to use some similar pre-trained deep architecture and modify its weights according to the new problem. A similar approach is followed in this work. Generally, the top layer features of the pre-trained deep neural network models can be reliably used as off-the-shelf descriptors for the given task. For example as the pre-trained deep neural networks Alexnet [55], Resnet50 [47] and VGG19 [58] etc are pre-trained models for the task of object recognition therefore their top layers features can be reliably used for the recognition tasks. For the person Re-Id task these features won't work as the pre-trained models were not aimed for the re-identification task. In order to reuse the built-in structure of these pre-trained deep models (with minimal changes) for Re-Id task, the models need to be fine-tuned on person based dataset and the desired output function needs to be aligned with the default output of pre-trained models. For this purpose the retrieval task of the Re-Id is converted into multiple classification problems. For this purpose the top layer CNN features of Alexnet are fine-tuned on a pedestrian attribute dataset PETA [59]. Alexnet is an image classification based pre-trained deep network. The deep architecture has five convolutional layers, three max pooling layers and three fully connected (FC) layers. The convolution layers used the Rectified linear unit as their activation function to extract the low level features with different receptive fields at each layer. The max pooling layers reduced the network size while preserving the most significant information. The FC layers used dropout regularization technique to generalize the deep model outputs. The final FC layer is extended as per our desired outputs/ labels. The labels are in the form of pedestrian attributes combination to define a person. We divided all attributes into seven major groups i.e. gender, age, luggage, upper body dress style, upper body main color, lower body dress style, lower body main color. In the first phase an attribute is selected from each of the seven major groups through the customized FC layer. In the next phase combination-attributes labels are classified by minimizing the softmax loss. The customized loss functions replaced the original loss function of pre-trained Alexnet, hence turned the image classification based deep architecture into person re-id deep architecture. The parameters/ pre-trained weights are updated during the back propagation phase of the deep learning process according to customized cost functions. The major contribution is the use of two successive classification loss layers for fine tuning. The first loss layer is used for the detection of multiple attribute labels of pedestrians. The second loss layer works for the combination of different attribute labels to classify person specific discriminative attributes. The loss function used for the fine-tuning of the deep network is given in equation III-A.4).

$$L = \alpha L^c + (1 - \alpha) \frac{1}{G} \sum_{g=1}^{G} L^g \tag{4}$$

Where $\alpha (0 \leq \alpha \leq 1)$ is the parameter to control the contribution of two types of losses; one is the multiple attributes classification loss and the other is the classification loss for combination-attributes. L^C is the loss of combination attributes. Total G attributes groups are used and each attribute group is represented by g.

After fine-tuning of the Alexnet model, the modified weights of the pre-trained model are used to generate person specific deep features that can be used as off-the-shelf features descriptor. The deep features from first fully connected layer (FC6) of the modified pre-trained deep network are extracted for further use as off-line person descriptors in this work. The extracted deep features of significantly low dimensions (4096) performed equivalent to the very high-dimensional hand-crafted features.

4) INTEGRATED PERSON REPRESENTATION

The final step of person representation phase is the integration of all three types of extracted features. The hierarchical Gaussian features, strip-wise pooled features and the fine-tuned deep features are combined linearly by using simple concatenation function as shown in equation (5).

$$F_i = (z, w, df) \tag{5}$$

Where z is the hierarchical Gaussian features vector, w is the stripped pooled features vector explained in

Section III (A-2)) and *df* are the deep features extracted using modified pre-trained model. The original length of proposed hybrid feature descriptor is 37040 (27,626+5318+4096). The reduced version is also proposed for the hybrid feature descriptor.

B. REDUCED PERSON REPREDENTATION THROUGH MANIFOLD LEARNING

Prior to reducing the features dimensions we used the proposed hybrid feature descriptor with its original dimensions to analyze its robustness and effectiveness. The Re-Id results of original length hybrid feature descriptor are discussed in Section IV.

In this module the proposed hybrid descriptor is exposed to the dimensionality reduction in order to reduce the computational complexity and to improve the overall performance of the proposed framework. The original length feature vectors are exposed to manifold learning that employs the Principal Component Analysis (PCA). In order to get an intuition about the number of highly significant and covariant dimensions, the architecture of PCA is explored in detail. The PCA utilizes the covariance information of the given dimensions to rearrange the whole features space by preserving highly covariant dimensions on higher orders. More technically the higher Eigen values determine the high covariance among the dimensions. With the results of dimensions reduction process, it is analyzed that for all types of the features, the top 100 dimensions performed almost equivalent to the performance of whole length features [27].

Further to validate the number of selected dimensions, the proposed model is passed through the cross-validation. The cross-validation results shown that for given hand-crafted and deep features there are different number of effective dimensions but the number is lesser than 100 for all of the features. In order to increase the robustness of proposed model we incorporated the cross-validation process in our model and the respective number of dimensions is used for each type of feature to propose the reduced hybrid person descriptor. For the resultant low dimensional feature descriptor, a comparatively complex but more effective person matching technique is employed and will be discussed in next section.

C. PERSON MATCHING

For person matching, two different state-of-the-art metric learning approaches are utilized to increase the intra-class similarities and the inter-class differences. The Cross-view Quadratic Discriminant Analysis [14] and Multiple Metric Learning [15] are applied separately to the original length proposed feature descriptor and the reduced proposed descriptor respectively to analyze the robustness of proposed descriptors.

1) CROSS-VIEW QUADRATIC DISCRIMINANT ANALYSIS

With the original length feature descriptor the state-of-the-art XQDA metric learning methodology is used. XQDA metric learning is the extension of Bayesian face [60] and KISS metric learning [27]. Original length features descriptor is used as this metric learning approach is not sensitive to the higher dimensions. The XQDA learns the features subspace with cross-view data and simultaneously learns a distance function for cross-view similarity measure in the reduced dimensional subspace (i.e. r-dimensional subspace). In this method the intra-personal variations Ω_I and extra-personal variations Ω_E are used to solve the multi-class classification problem. The distance function is computed for r- dimensional subspace using equation III-C.2).

$$d_w(x, z) = (x - z)^T W \left(\Sigma_I^{'-1} - \Sigma_E^{'-1}\right) W^T (x - z) \quad (6)$$

Where W is r - dimensional subspace, Σ_I and Σ_E are the covariance matrices of Ω_I and Ω_E (the intra-personal variations and extra-personal variations) and the x and z are the cross-view training sets of c number of classes.

Only the highly discriminative features space is retained by XQDA and the rest of features with little or no impact on Re-Id task are discarded during the training phase. It learns the global distance metric to ensure that the selected feature space is reducing the intra-class variations and increasing the inter-class variations to perform person specific Re-Identification.

2) MULTIPLE METRIC LEARNING

The reduced dimensional hybrid descriptor is received from the manifold learning module. It is then fed into the multiple metric learning module [15] to learn the non-Euclidean metric for all of the extracted feature types. MML is a self-learning based approach that learns the discriminative subspace of the features by modeling different transformation functions for each feature type separately instead of using a single transformation function for the whole feature space. In this approach the highly discriminative features are given more weightages to increase their contribution significantly towards final similarity and dissimilarity measures. The weight vector β is computed using equation (7).

$$\beta_f = \frac{1}{\tau} \sum_{i=1}^{|\tau|} R_f^i \quad (7)$$

where τ represents the query images, i represents the *ith* person in the given list of probe images, the function R_f^i results in highest value i.e. 1 in case the query image has the lowest dissimilarity (hence highest similarity) with its true match from the given gallery images. In short, the β defines Re-Id performance for each feature type. The MML module computes the final dissimilarity using equation (8).

$$D\left(I^A, I^B\right) = \sum_f \beta_f d_f^2\left(I^A, I^B\right) \quad (8)$$

Where I^A and I^B are the probe and gallery images, d_f^2 is the dissimilarity measure of the feature type f from both given images. Therefore the best features are given higher weightages while computing the final dissimilarity by fusing individual similarities.

The experimental results have shown that the proposed hybrid person representations outperformed the existing state-of-the-art approaches in both cases i.e. with the use of original length feature descriptor and with the reduced descriptor. The experimental details are discussed in next section.

IV. EXPERIMENTAL EVALUATION

A. PERFORMANCE MEASURES

Following performance measures are used to evaluate the effectiveness of the proposed re-identification framework.

1) RANK-1 ACCURACY

The output of a Re-Id system is the retrieval of ranked gallery images for a given query/ probe image. The closest match is assigned the top most rank i.e. rank-1. The ratio of true matches at rank-1 measures the rank-1 accuracy of a Re-Id system.

2) CUMULATIVE MATCHING CHARACTERISTICS CURVE

Keeping in view the Re-Id challenges a little compromise in the assigned ranks is acceptable that is characterized by another performance measure called Cumulative Matching Characteristics (CMC) curve. CMC curve is the precision curve and is used to measure the ranking rate at rank-r for the given gallery data. The horizontal axis shows the ranks of Re-Id and the vertical axis plots the Re-Id accuracy percentage for the given ranks.

In case of multiple ground truth matches only the first true match is counted in the CMS calculations as CMS curve is basically concerned to rank the true match in the top few ranks [51].

B. DATASETS

The proposed work is evaluated on following challenging Re-Id datasets.

1) VIPeR

Viewpoint Invariant Pedestrian Recognition Re-Id dataset [61] is a widely used small scale Re-Id benchmark. The VIPeR dataset has 632 unique identities captured by two disjoint cameras in outdoor scenes. For each person two images are captured, one image is from the front view and the other image is from some any other orientation i.e. 45°, 30°, 90° etc. The main challenges of this dataset include background noise, pose and viewpoint variations, varying illumination conditions, these all variation type of variations are even observed within single camera images that make it more challenging to define some generic solution.

2) PRID450s

In the PRID450s dataset [62] the images are taken from two disjoint cameras. The dataset contains significant poses variations, illumination and viewpoint variations and the occlusions. It contains 450 unique identities two images of a single entity therefore Re-Id methodology encompassing single verses single matching can sufficiently handle Re-Id task for this dataset.

3) GRID

The Grid dataset [63] contains the images taken from an underground station. The images are captured by 8 disjoint cameras. This dataset is very challenging as its images have very low resolution with pose variations and occlusions. 250 unique identities are available in it. In addition of 250 pairs of images the dataset contains 775 unpaired identities having no true match that caused increased number of images in the gallery set. Due to the very low resolution images and a large number of unpaired images, the accuracy of Re-Id methods is considerably low for this dataset.

4) CUHK01

The campus dataset CUHK01 [64] is a medium sized Re-Id dataset contains 971 unique identities with two images per person. The images are captures by two disjoint camera views. Inter-camera variations are higher for this dataset as one camera focus on the frontal view and the back view of a person in majority of images whereas the other camera contains more variations in poses and viewpoints.

5) Market-1501

The Market-1501 dataset [47] is a large scale person Re-Id dataset that contains 32,668 images of 1501 unique identities. The training set contains 751 identities and the test set contains 750 unique identities. The images are annotated with a deformable part model to generate the bounding boxes for people in the images. The images are taken using six camera views installed in front of a supermarket, hence resulting in a collection of diverse poses and viewpoints in the dataset.

6) DukeMTMC-ReID

The DukeMTMC-ReID dataset [48] is a large scale person re-identification dataset. The dataset contains 36,411 images of 1404 unique identities, out of which 702 identities are included in training set and 702 are in test set. 16522 images are the training images, 17661 images are the test images and 2228 images are query images. The images are captured in Duke University campus using 8 cameras to include a variety of poses and viewpoints in the dataset to make it a complex dataset.

C. RESULTS

1) EXPERIMENTAL SETUP

In order to extract hierarchical Gaussian features (GOG) of all color channels, the images are resized into 128 x 48 for all of the datasets. For stripped pooled features (WHOS) the same dimensions are used to resize the images. The aim of keeping the same size of the images for GOG and WHOS features is to preserve the complementary low level features at same level. The fine-tuned CNN features (FT-CNN) are

TABLE 1. Re-Id results of the proposed methodology-small scale benchmarks.

Rank Accuracy	VIPeR		PRID450s		GRID	
	Hybrid+ XQDA	R-Hybrid+ MML	Hybrid+ XQDA	R-Hybrid+ MML	Hybrid+ XQDA	R-Hybrid+ MML
Rank 1	50.5	66.0	66.5	68.6	24.5	28.0
Rank 5	82.5	88.0	87.9	89.0	46.6	47.5
Rank 10	91.5	95.8	93.6	93.8	57.0	57.4
Rank 20	97.2	98.9	97.0	97.5	68.2	68.8

TABLE 2. Re-Id results of the proposed methodology-large scale benchmarks.

Rank Accuracy	CUHK01		Market-1501		DukeMTMC-Re-ID	
	Hybrid +XQDA	R-Hybrid +MML	Hybrid +XQDA	R-Hybrid +MML	Hybrid +XQDA	R-Hybrid +MML
Rank 1	68.3	70.2	60.17	61.22	73.45	75.2
Rank 5	88.0	90.6	73.73	73.9	81.43	85.65
Rank 10	93.2	94.9	77.57	80.1	84.2	89.88
Rank 20	96.6	98.1	80.23	90.5	86.43	95.1

extracted by fine-tuning of the pre-trained model Alexnet. The pre-trained deep model Alexnet receives the input images in size 227 $\times$ 227 therefore the images are resized to 227 $\times$ 227 before extraction of fine-tuned deep features.

For each query image the proposed model generates the similarity score of all gallery images and the rankings are generated on the basis of similarity scores. For all Re-Id benchmarks, the 50% - 50% split of train and test data is used. The experiments for all benchmark datasets are conducted 10 times with a different random splits and their results are averaged to quote the final results. A deep analysis is performed by using color features, texture features and annotation based deep features in different combinations for all of the datasets. The analysis will be discussed in Section V.

2) QUANTITATIVE PERFORMANCE MEASURES

The experimental results comprising of Re-Id accuracies attained for all given datasets at different ranks i.e. rank-1, rank 5, rank-10 and rank-20 are in Table 1 for small scale datasets (VIPeR, PRID450s and GRID) and in Table 2 for large scale dataset (CUHK01, Market-1501 and DukeMTMC-ReID). For the original length proposed hybrid feature descriptor exposed to XQDA metric learning the Re-Id results are listed in the columns titled as Hybrid+XQDA. Whereas for the reduced version of the proposed hybrid descriptor exposed to multiple metric learning, the results are list in the columns with titles R-Hybrid+MML. The results highlighted that the reduced version of proposed hybrid features outperformed the original length descriptor for Re-Id accuracies of all benchmarks. Moreover the reduced dimensions ensured the obvious reduction in the computational complexity and the time complexity.

In the Table 3, a detailed comparison of Re-Id results of the proposed framework with multiple existing state-of-the-art approaches is shown. The results show that both versions of the proposed framework outperformed the Re-Id accuracies of all existing relevant methodologies with a great margin. The overview of existing listed methodologies is already discussed in the Related Work section. The proposed approach not only leaded the various hand-crafted Re-Id solutions and the metric learning techniques but it also left behind the various hybrid approaches.

The cumulative matching characteristics curves of the proposed Re-Id results along with the CMC curves of existing approaches are shown in Figure 3(a-f) for the VIPeR, PRID450s, GRID, CUHK01, Market-1501 and DukeMTMC-ReID datasets respectively.

V. DISCUSSION

A. VIPeR

For the VIPeR dataset [48] the random split of training and test with 50 - 50 ratio resulted in 316 identities in the training set and the remaining 316 in the test set. The experimental results on publically available dataset are given in Table 3. The results of two types of proposed hybrid features are listed at serial 1 and 2, the accuracy rates of other hybrid approaches discussed in related work section are given at serial 3-4, the Re-Id accuracies claimed by all baseline incorporated approaches for this work are available at serial 5-7 and 9-10 and the Re-Id accuracy results for rest of few relevant solutions including state-of-the-art approaches are also recorded in the table.

Inclusion of the mean color information in the hierarchical Gaussian features leaded the existing pixels-based features and attained the rank-1 accuracy of 42.3% for the most common RGB color space. It is analyzed that RGB and LAB (rank-1 accuracy of 44.2%) color spaces based hierarchical features performed better than HS color based Gaussian features having the rank-1 accuracy of 38.9%. Although the

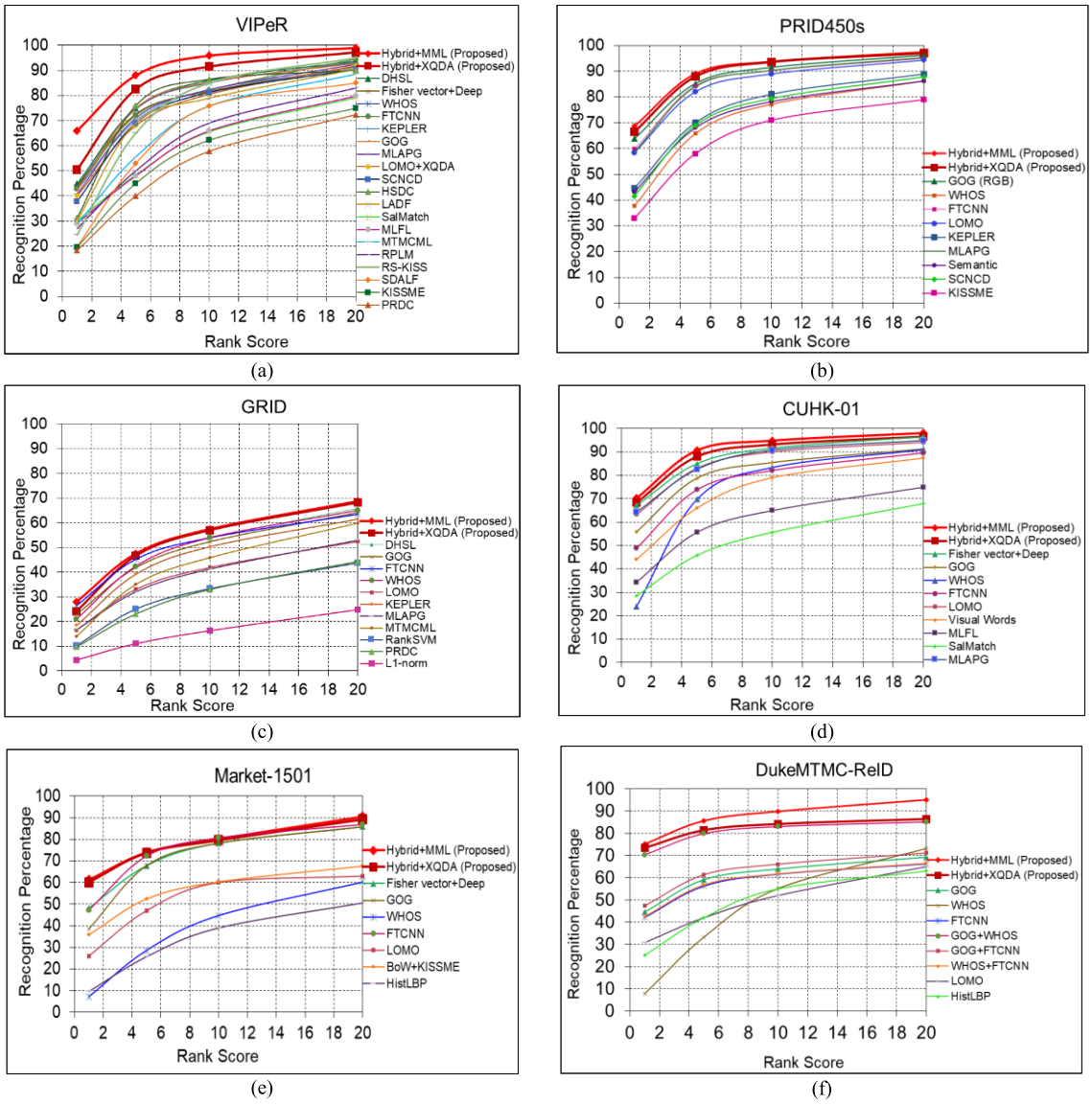

FIGURE 3. **Comparison of Cumulative Matching Characteristics Curves of Proposed Approach with the Existing Re-Id Solutions; (a) CMC Curves for VIPeR dataset (b) CMC Curves for PRID450s dataset (c) CMC Curves for GRID dataset (d) CMC Curves for CUHK01 dataset (e) CMC Curves for Market-1501 dataset (f) CMC Curves for DukeMTMC-ReID dataset.**

combination of all color channels of hierarchical features improved the overall accuracy to 49% however it resulted in a very high dimensional feature descriptor.

The stripped pooled histograms and fine-tuned CNN features are comparatively low-dimensional features as compared to the hierarchical Gaussian features and performed almost equivalent with the rank-1 accuracy rates are 43.3% and 42.9% respectively. State-of-the-art LOMO could achieve 40% rank-1 accuracy for VIPeR dataset. The hybrid approaches however performed better than the purely hand-crafted techniques due to incorporation of deep learning benefits and resulted in rank-1 accuracies of 44.1% hybrid approaches however performed better than the purely hand-crafted techniques due to incorporation of deep learning benefits and resulted in rank-1 accuracies of 44.1% for the hand-crafted fisher vector fed to deep model and 44.9% for the hybrid similarity learning mechanism. The use of saliency with the multiple metric learning by the KEPLER claimed the rank-1 accuracy of 44.5%.

For the proposed framework, the use of manifold learning and the selection of metric learning techniques handled the high dimensionality of the features efficiently therefore at the first phase of person representation the main focus is to accumulate the hand-crafted and deep learning based features which can maximize the capture of person discrimination information in diverse styles.

A deep analysis is done for the effectiveness of different color spaces and pattern features combination, when stripped histograms are used along with different channels of hierarchical features the hierarchical Gaussian with RGB channel,

TABLE 3. Comparison of experimental results of proposed methodology with existing approaches.

SN	Methods	VIPeR		PRID450s		GRID		CUHK01		Market1501		DukeMT MC-ReID	
		R=1	R=20	R=1	R=20	R=1	R=20	R=1	R=20	R=1	R=20	R=1	R=20
1	**Hybrid+MML (Proposed)**	**66.0**	**98.9**	**68.6**	**97.5**	**28.0**	**68.8**	**70.2**	**98.1**	**61.22**	**90.5**	**75.2**	**95.1**
2	**Hybrid+XQDA (Proposed)**	**50.5**	**97.2**	**66.5**	**97.0**	**24.5**	**68.2**	**68.3**	**96.6**	**60.17**	**89.23**	**73.45**	**86.43**
3	DHSL (Zeng et al.) [51]	44.9	93.7	-	-	21.2	65.8	-	-	-	-	-	-
4	Fisher vector +Deep (Shen et al.) [50]	44.1	91.5	-	-	-	-	67.1	96.5	48.15	85.8	-	-
5	WHOS (Lisanti et al.) [30]	43.3	93.2	37.8	86.2	21.0	65.0	23.9	91.0	7.19	60.23	7.8	73.29
6	FT-CNN (Matsukawa et al.) [55]	42.9	94.6	59.8	95.0	26.4	63.4	48.8	89.6	47.31	86.88	42.12	66.24
7	KEPLER (Martinel et al.) [15]	42.4	90.7	44.6	88.9	18.4	61.4	-	-	-	-	-	-
8	MLAPG (Liao et al.) [31]	40.7	92.4	58.8	95.3	16.6	52.9	64.2	94.9	-	-	-	-
9	LOMO (Liao et al.) [14]	40.0	91.8	58.4	94.3	16.5	52.4	63.2	94.1	26.07	63.2	30.75	65.1
10	GOG-RGB (Matsukawa et al.) [57]	42.3	92.8	63.8	96.2	22.9	64.1	55.8	91.3	38.53	85.8	44.6	69.27
11	SCNCD (Yang et al.) [33]	37.8	90.4	41.6	87.8	-	-	-	-	-	-	-	-
12	HSDC (Wu et al.) [21]	31.2	90.0	-	-	-	-	-	-	33.54	-	-	-
13	LADF (Li et al.) [22]	30.2	90.4	-	-	-	-	-	-	-	-	-	-
14	SalMatch (Zhao et al.) [23]	30.2	79.2	-	-	-	-	28.4	67.9	-	-	-	-
15	MLFL (Zhao et al.) [37]	29.1	79.8	-	-	-	-	34.3	74.9	-	-	-	-
16	MTMCML (Lisanti et al.) [30]	28.8	88.5	-	-	14.1	59.8	-	-	-	-	-	-
17	RPLM (Hirzer et al.) [28]	27.0	83.0	-	-	-	-	-	-	-	-	-	-
18	RS-KISS (Jin et al.) [66]	24.5	95.0	-	-	-	-	-	-	35.84	-	-	-
19	SDALF (Farenzena et al.) [9]	19.8	85.0	-	-	-	-	-	-	20.53	-	-	-
20	KISSME (Hirzer et al.) [28]	19.6	74.9	33.0	79.0	-	-	-	-	44.24		25.13	-
21	PRDC (Gong et al.) [20]	18.3	72.4	-	-	9.68	44.3	-	-	-	-	-	-
22	Semantic (Hospedales et al.) [67]	-	-	43.1	86.2	-	-	-	-	-	-	-	-
23	RankSVM (Chunheng et al.) [24]	-	-	-	-	10.2	43.7	-	-	-	-	-	-
24	L1-norm (Liu et al.) [68]	-	-	-	-	4.40	24.8	-	-	8.28	-	-	-
25	VisualWords (Chen et al.) [32]	-	-	-	-	-	-	44.0	87.4	-	-	-	-

the system resulted in highest rank-1 accuracy i.e. 48.7% in comparison with the fusion of stripped features with the rest of Gaussian color-spaces. The improvement in accuracy is due to the complementary features captured by WHOS. The local mean information is preserved by Gaussian arrangements at different hierarchical levels and the accumulation of weighted histograms focused on local level highly discriminant information in terms of colors and patterns. Moreover the use of attribute based fine-tuned deep features improved the overall Re-Id results with the rank-1 accuracy of 50.5% as these features combined the person attributes along with the contextual information in order to summarize the person data. Therefore it is observed that the proposed hybrid feature descriptor prior to the dimensionality reduction when used with the XQDA metric learning approach outperformed the LOMO by 10.5% for rank-1 accuracy and 5.4% for rank-20 accuracy. Moreover the proposed original length descriptor leaded the rank-1 accuracies of hierarchical Gaussian descriptor RGB channel, weighted histogram descriptor and fine-tuned CNN features by 8.2%, 7.2% and 7.6% respectively. With the comparative analysis the discriminative dominancy of the proposed hybrid features is proved.

Moreover the effectiveness of the reduced hybrid descriptor is analyzed by passing it through multiple metric learning for person matching. It is observed that the use of separate global distance metrics for the given features channels resulted in the significant boost in Re-Id performance when compared with the single global distance metric applied to whole feature space using XQDA. The PCA applied reduced length hybrid feature descriptor outperformed the original length descriptor by a large margin and achieved unprecedented rank-1 and rank-20 accuracies of 66% and 98.9% respectively.

B. PRID450s

PRID450s Re-Id dataset [62] is a relatively smaller dataset and with 50-50 training-test split, 225 identities are included in each of the set in our experimental settings LOMO reported the rank-1 accuracy as 58.44%. Among all channels of GOG features, the GOG-RGB performed the best with rank-1 accuracy of 63.8%. Individual stripped pooled features and FT-CNN performed 47.4% and 59.8% respectively for rank-1 accuracy. It is observed that deep learning based FT-CNN features performed almost equivalent to the hand-crafted state-of-the-art features for this smaller dataset. When GOG-RGB features are combined with the WHOS and FT-CNN features a little improvement in rank-1 accuracy is observed with the result of 65.2% accuracy. Combining GOG-LAB and GOG-RGB with the WHOS features also performed better than their individual usage. The proposed descriptor with original length of extracted features improved the rank-1 accuracy of LOMO by 8% by achieving

66.5% rank-1 accuracy. However the best ever Re-Id performance is seen when the hierarchical features are combined with stripped pooled features and FT-CNN features in their reduced form and resulted in rank-1 accuracy of 68.6%. The proposed methods improved the rank-20 accuracy of the LOMO by around 3.5% and the rank-1 accuracy of LOMO by around 10%.

The PRID450s dataset has not been evaluated by other hybrid approaches discussed earlier but the proposed approach achieved significant lead over the hand-crafted approaches as well as it left behind the deep learning based approach [66] based on the transfer learning for semantic person representation, the results as shown in Table 3.

C. GRID

For Grid Re-Id dataset [63] rank-1 accuracy claimed by LOMO is significantly low (16.56%) and the rank-20 accuracy is 52.40% due to the huge number of deviated images in the gallery set. In the proposed work in addition to hierarchical Gaussian features, the use of stripped pool of features that took advantage of the spatial patterns and the fine-tuned deep features that utilized the semantic attributes information resulted in better performance than existing hand-crafted methodologies. The proposed original length feature descriptor outperformed LOMO for rank-1 accuracy by 8% and achieved the rank-1 accuracy of 24.5% and the reduced length proposed descriptor exposed to multiple metric learning performed even better than single global distance metric learning and improved the rank-1 accuracy of LOMO descriptor by 11.5% with the attained rank-1 accuracy of 28%. After analyzing the Gaussian features of different color spaces it is observed that the rank-1 accuracy achieved by RGB color channel is 22.9% and for the stripped pooled features it is 21%. Moreover when RGB channel features are combined with the stripped pooled features the overall performance is improved to rank-1 accuracy of 24.2%. A slight improvement is seen when attribute based deep features are added to design the proposed hybrid descriptor with the rank-1 accuracy of 24.5%. However the independent use of fine-tuned deep features is better than combined features with the rank-1 accuracy of 26.4%. The use of hybrid similarity approach DHSL claimed its rank-1 accuracy as 21.2% hence got improvement over LOMO however our proposed hybrid descriptor outperformed the rank-1 accuracy of hybrid DHSL by 2.3% with its original length version. Moreover when the reduced hybrid descriptor is analyzed, it proved that the individual feature type wise learning of distance metric performed better than single global distance metric. A significant improvement is observed with the rank-1 accuracy of 28%. This way the proposed hybrid descriptor claimed clear lead of 11.5% rank-1 accuracy in comparison with the LOMO descriptor.

The reduced version also outperformed the Re-id performance over fine-tuned deep features by 1.6% for rank-1 accuracy. However the size of reduced descriptor is significantly smaller than the fine-tuned deep descriptor size therefore overall contribution of the proposed framework is still higher. The reduced version leaded the hybrid DHSL by 6.8% for rank-1 accuracy.

D. CUHK01

The experimental results for the campus dataset CUHK01 [64] have shown that among all channels of hierarchical Gaussian features the RGB channel achieved maximum rank-1 accuracy of 55% than the rest of color channels. Moreover when RGB Gaussian features are combined with the stripped pool of histograms, it gave a rise of more than 5% accuracy with resulting 60.2% rank-1 accuracy. The integration of fine-tuned deep features with Gaussian features resulted in improvement of 2% in Re-Id results of original Gaussian results. In comparison with Gaussian features, in individual performance of the stripped histograms and fine-tuned CNN based resulted the rank-1 accuracy as 47.7% and 48.8% respectively.

The hybrid approach i.e. fisher vector with deep features reported the higher Re-Id results with rank-1 accuracy of 67%. The main reason for this improvement is comparatively larger size of the Re-Id dataset that could result in good training of the deep network hence performed better than the pure hand-crafted approaches. However as the proposed hybrid descriptor combined the hand-crafted and deep learning based benefits so the best results are achieved. The proposed original length hybrid descriptor significantly improved the Re-Id performance by 12.5% over Gaussian features and achieved the rank-1 accuracy of 68.3%. The proposed descriptor also outperformed the hybrid technique proposed in [49] by 2.9%. Moreover, the reduced descriptor performed even better than the original length descriptor and achieved the rank-1 accuracy of 70.2%. The comparative analysis of the results is given in Table 3.

E. MARKET-1501

The experimental results on Market-1501 [47] have shown that the hierarchical Gaussian features performed better than hand-crafted features LOMO and bag of words features and attained 38.53% accuracy on the rank-1. However the fine-tuned deep features and fisher vector containing both hand-crafted and deep features performed much better than Gaussian features with the rank-1 accuracies of 47.31 and 48.15 respectively. Weighted histograms did not perform well for Re-Id at its own however while combined with the Gaussian features, the overall performance is increased with rank-1 accuracy of 58.52%. Our proposed hybrid descriptor stood at the top place with rank-1 accuracy of 60.17% and the rank-20 accuracy of around 90% for its full length version and 61.22% rank-1 accuracy for the reduced version. The integration of complementary features outperformed the individual hand-crafted features and few hybrid techniques. Many of the existing hand-crafted techniques evaluated their significance only on small scale Re-Id benchmarks. However the comparison of our proposed methodology

TABLE 4. Significance of features (comparison of rank-1).

Feature Type	Rank-1 Accuracy					
	VIPER	PRID450s	GRID	CUHK01	Market-1501	DukeMTMC-ReID
GOG	42.3	63.8	22.9	55.8	38.53	44.6
WHOS	43.3	47.4	21	23.9	7.19	7.8
FTCNN	42.9	59.8	26.4	48.8	47.31	42.12
GOG+WHOS	48.7	67.2	24.2	61.8	58.52	70.34
GOG+FTCNN	46.6	66.8	25.4	56.3	38.94	47.37
WHOS+FTCNN	43.5	48.1	21.9	27	7.26	42.53
GOG+ WHOS+FTCNN	50.5	67.4	24.5	68.3	60.17	73.45

TABLE 5. Significance of features (detailed comparison for rank-1 to rank-20 accuracies).

	VIPER				GRID			
	R1	R5	R10	R20	R1	R5	R10	R20
GOG	42.3	75	85.3	92.8	22.9	41.8	52.3	64.1
WHOS	43.3	74.6	85.1	93.2	21.0	42.4	53.9	65.0
FTCNN	42.9	72.1	82.2	91.6	26.4	45.2	53.9	63.4
GOG+WHOS	48.7	81.3	89.9	95.2	24.2	46.6	57.0	68.2
GOG+FTCNN	46.6	75.1	85.5	92.9	25.4	46.4	57.8	69.0
WHOS+FTCNN	43.5	75.3	85.3	93.8	21.9	44.4	54.8	65.8
GOG+ WHOS+FTCNN	50.5	82.2	91.5	97.2	24.5	47.3	57.0	68.2
	PRID450s				CUHK01			
	R1	R5	R10	R20	R1	R5	R10	R20
GOG	63.8	85.2	91.5	96.2	55.8	78.8	85.5	91.3
WHOS	47.4	75.5	85.8	92.4	23.9	69.7	83.4	91.0
FTCNN	59.8	84.3	90.2	95.0	48.8	73.8	82.1	89.6
GOG+WHOS	67.2	89.1	94.4	97.6	61.8	82.9	88.8	94.1
GOG+FTCNN	66.8	88.6	94.0	97.6	56.3	79.0	85.6	91.4
WHOS+FTCNN	48.1	85.6	91.6	96.2	27.0	72.7	85.2	92.0
GOG+ WHOS+FTCNN	67.4	89.3	94.2	97.7	68.3	88.0	93.2	96.6
	Market-1501				DukeMTMC-ReID			
	R1	R5	R10	R20	R1	R5	R10	R20
GOG	38.53	67.81	78.2	85.8	44.6	59.13	63.94	69.27
WHOS	7.19	28.63	44.74	60.23	7.8	33.28	55.14	73.19
FTCNN	47.31	72.24	80.67	86.88	42.12	56.25	61.58	66.24
GOG+WHOS	58.52	67.97	70.6	72.81	70.34	79.88	83.03	85.1
GOG+FTCNN	38.94	68	78.3	85.84	47.37	61.39	66.12	71.15
WHOS+FTCNN	7.26	28.61	47.24	62.42	42.53	56.8	61.76	66.49
GOG+ WHOS+FTCNN	60.17	73.73	77.57	89.23	73.45	81.43	84.2	86.43

results with the results of pure hand-crafted and hybrid methods found in the literature is given in the Table 3.

F. DUKEMTMC-ReID

Experimental results on a large scale Re-Id benchmark DukeMTMC-ReID [48] also proves the significance of our proposed hybrid features. Rank-1 accuracy of hand-crafted LOMO descriptors is 30.8%. Gaussian features performed better than LOMO and WHOS features with the rank-1 accuracy of 44.6%. Combination of Gaussian features with weighted histograms boosted the overall performance and attained 70.3% rank-1 accuracy due to the accumulation of complementary features. However our proposed descriptor performed the best among all individual features and different combinations with the rank-1 accuracy of 73.45% for full version and above 75% for the reduced one. This significant raise in the Re-Id accuracy by some hybrid methodology is indeed comparable with the latest end-to-end deep methods.

A detailed performance comparison of Gaussian features, WHOS features, off the shelf fine-tuned features and their different combinations is given in Table 5. Our proposed methodology increased the rank-1 accuracy over the well-known LOMO by more than 40% and the fine-tuned features by more than 30% for rank-1 accuracy.

FIGURE 4. Difficult true matches by proposed method.

FIGURE 5. Robustness for similar appearances.

Our proposed methodology increased the rank-1 accuracy over the well-known LOMO by more than 40% and the fine-tuned features by more than 30% for rank-1 accuracy.

G. FEATURES IMPORTANCE

Additionally, the contribution of individual and different combinations is analyzed through the experiments. It is observed that hierarchical Gaussian features contributed mostly for the performance of proposed work. Table 4 provides the performance comparison of different features for the evaluation measure Rank-1 accuracy for the proposed full-length hybrid descriptor with the renowned XQDA metric learning. A detailed comparison of different ranks accuracy for individual features and their different combinations is given in Table 5.

It is observed that hierarchical Gaussian features contributed mostly for the performance of proposed work. Table 4 provides the performance comparison of different features for the evaluation measure Rank-1 accuracy for the proposed full-length hybrid descriptor with the renowned XQDA metric learning. A detailed comparison of different ranks accuracy for individual features and their different combinations is given in Table 5.

H. ROBUSTNESS

The experimental results showed that the proposed methodology performed better than multiple state-of-the-art Re-Id methods. The robustness of the proposed method is multifold. First the selection of complementary features played significant role. Secondly the assignment of more weight to the central/human part of the image excluded the back ground noise to high extent.

The removal of background noise improved the robustness of the propose methodology for cluttered background and for the multiple people appearances in the image background. Addition of attribute b ased deep features that implicitly highlighted the salient regions of the image increased the robustness. Few of the cases having cluttered back ground and multiple people in the image, the proposed method performed true identification on higher ranks are shown in Figure 4. However it is observed that the exactly same appearance of different people having blurred images the attribute based features did not contribute to a significant level. In such case the proposed methodology mainly depends on the spatial and pattern features to perform the re- id task and it is observed that even in such case a person is re-identified in top new ranks if not on to the most rank, as shown in Figure 5.

For fine-tuning of attribute based deep features, the experiments are performed on NVIDIA TitanX GPU. It took five hours to converge the model after 50,000 iterations. WHOS hand-crafted features were provided by [29]. Hierarchical features were extracted using Intel Core i5 @ 3.1 GHz with 12GB RAM and it took 0.01 sec per image. The major time is consumed for different kinds of features extraction. Afterwards the features integration time is negligible (below 0.001 sec per image).

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a hybrid person representation technique by utilizing the benefits of low-level hand-crafted and mid-level deep features, to formulate a highly discriminative person representation. For hand-crafted features, the hierarchical Gaussian color features utilized the mean value of a patch in the features extraction, hence explicitly highlighted the discriminative patch of a certain region. Further the horizontal stripped pooling of weighted local features from different color spaces and patterns resulted in the generation of pose invariant feature descriptor while maintaining the correlation between the strips. Moreover the attributes based deep learned features aided the Re-Id solution by focusing on salient attributes of a person. The hierarchical approach is opted to get the global view of a person by accumulation of the information from all hierarchical levels. The proposed hybrid representation is learned with XQDA and MML metric learning techniques in order to validate the effectiveness of proposed descriptor. The hybrid descriptor had contributed well to cope with the Re-Id challenges that include illumination and viewpoint variations, the misalignments of the people in respective images and handling of background noise for the person Re-Id process. The proposed approach is equally good for small scale and large scale Re-Id scenarios. The experimental results revealed that the proposed hybrid person representation and the integrated distance learning approaches outperformed the existing state-of-the-art Re-Id solutions with a large margin.

Although the proposed approach handled various Re-Id challenges effectively and achieved the significant Re-Id performance. However our work is based on the cropped images of the persons. In the future we aim to develop an end-to-end deep learning architecture for person Re-identification such that the person detection would also be integrated into the model. Our focus would be to jointly learn the attention based feature representation and to summarize the multi-level hierarchical discrimination through the attention oriented convolutional neural networks.

ACKNOWLEDGEMENTS

The authors acknowledge the support of NVIDIA Corporation for the provision of the Titan X GPU that speeded up the non-trivial part of our research i.e., extraction of mid-level deep features.

REFERENCES

[1] F. Behmann and K. Wu, *Collaborative Internet of Things (C-IoT): For Future Smart Connected Life and Business*, vol. 21. Chichester, U.K.: IEEE, 2015, p. 282.
[2] S. Gong, M. Cristani, S. Yan, and C. C. Loy, *Person Re-Identification*. New York, NY, USA: Springer, 2013.
[3] L. Zheng, H. Zhang, S. Sun, M. Chandraker, Y. Yang, and Q. Tian, "Person re-identification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3346–3355.
[4] Q. Zhou, S. Zheng, H. Ling, H. Su, and S. Wu, "Joint dictionary and metric learning for person re-identification," *Pattern Recognit.*, vol. 72, pp. 196–206, Dec. 2017.
[5] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Computer Vision—ACCV*. Berlin, Germany: Springer, 2011.
[6] E. Ahmed, M. Jones, T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3908–3916.
[7] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.
[8] X. Wang, W.-S. Zheng, X. Li, and J. Zhang, "Cross-scenario transfer person reidentification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 8, pp. 1447–1460, Aug. 2016.
[9] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR*, Jun. 2010, pp. 2360–2367.
[10] Kodirov, E., "Person re-identification by unsupervised ℓ_1 graph learning," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016.
[11] Z. Shi, T. M. Hospedales, and T. iang, "Transferring a semantic representation for person re-identification and search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4184–4193.
[12] S. Mumtaz, N. Mubariz, S. Saleem, and M. M. Fraz, "Weighted hybrid features for person re-identification," in *Proc. 7th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov./Dec. 2017, pp. 1–6.
[13] N. Mubariz, S. Mumtaz, M. M. Hamayun, and M. M. Fraz, "Optimization of Person Re-Identification through Visual Descriptors," in *Proc. 13th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP)*, 2018, pp. 348–355.
[14] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2197–2206.
[15] N. Martinel, C. Micheloni, and G. L. Foresti, "Kernelized saliency-based person re-identification through multiple metric learning," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5645–5658, Dec. 2015.
[16] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, vol. 32, no. 4, pp. 270–286, 2014.
[17] F. Chen, J. Chai, D. Ren, X. Liu, and Y. Yang, "Semi-supervised distance metric learning for person re-identification," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 733–738.
[18] H.-X. Yu and A. W.-S. Wu Zheng. (2017). "Cross-view asymmetric metric learning for unsupervised person re-identification." [Online]. Available: https://arxiv.org/abs/1708.08062
[19] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2008.
[20] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.
[21] M. Zeng, Z. Wu, C. Tian, L. Zhang, and L. Hu, "Efficient person re-identification by hybrid spatiogram and covariance descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 48–56.
[22] R. Zhao, W. Ouyang, X. Wang, "Person re-identification by salience matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2528–2535.
[23] A. Jiang, C. Wang, and Y. Zhu, "Calibrated rank-SVM for multi-label image categorization," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1450–1455.
[24] A. Bedagkar-Gala and S. K. Shah, "Multiple person re-identification using part based spatio-temporal color appearance model," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1721–1728.
[25] P. K. Sathish and S. Balaji, "Person re-identification using part based hybrid descriptor," in *Proc. 2nd Int. Conf. Cogn. Comput. Inf. Process. (CCIP)*, Aug. 2016, pp. 1–4.
[26] S. Li, C. Gao, H. Yu, and J. Zhang, "Person re-identification via person DPM based partition," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 3856–3861.
[27] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2288–2295.
[28] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang, "Person re-identification by dual-regularized KISS metric learning," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2726–2738, Jun. 2016.
[29] G. Lisanti, I. Masi, A. D. Bagdanov, and A. D. Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1629–1642, Aug. 2015.
[30] S. Liao and S. Z. Li, "Efficient PSD constrained asymmetric metric learning for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3685–3693.
[31] Z. Zhang, Y. Chen, and V. Saligrama, "A novel visual word co-occurrence model for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 122–133.
[32] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3656–3670, Aug. 2014.
[33] R. Layne, T. M. Hospedales, and S. Gong, *Towards Person Identification and Re-identification With Attributes*. Berlin, Germany: Springer, 2012.
[34] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," in *Proc. CVPR*, Jun. 2011, pp. 801–808.
[35] K. Yu, B. Leng, Z. Zhang, D. Li, and K. Huang. (2016). "Weakly-supervised learning of mid-level features for pedestrian attribute recognition and localization." [Online]. Available: https://arxiv.org/abs/1611.05603
[36] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 144–151.
[37] W. Liao, M. Y. Yang, N. Zhan, and B. Rosenhahn, "Triplet-based deep similarity learning for person re-identification," in *Proc. ICCV*, 2017, pp. 385–393.
[38] X. Bai, M. Yang, T. Huang, Z. Dou, R. Yu, Y. Xu. (2017). "Deep-person: Learning discriminative deep features for person re-identification." [Online]. Available: https://arxiv.org/abs/1711.10658
[39] W. Chen, X. Chen, J. Zhang, K. Huang. (2017). "Beyond triplet loss: A deep quadruplet network for person re-identification." [Online]. Available: https://arxiv.org/abs/1704.01719
[40] C. Shen *et al.*, "Deep siamese network with multi-level similarity perception for person re-identification," in *Proc. 25th ACM Int. Conf. Multimedia*, 2017, pp. 1942–1950.
[41] W. L, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2194–2200.

[42] A. Hermans, L. Beyer, and B. Leibe. (2017). "In defense of the triplet loss for person re-identification." [Online]. Available: https://arxiv.org/abs/1703.07737
[43] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned CNN embedding for person reidentification," *ACM Trans. Multimedia Comput., Commun., Appl.*, 14, no. 1, pp. 13:1–13:20, 2018.
[44] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Deep attributes driven multi-camera person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 475–491.
[45] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang. (2017). "Improving person re-identification by attribute and identity learning." [Online]. Available: https://arxiv.org/abs/1703.07220
[46] D. Tao, Y. Guo, B. Yu, J. Pang, and Z. Yu, "Deep multi-view feature learning for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2657–2666, Oct. 2018.
[47] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.
[48] E. Ristani, F. Solera, R. S. Zou, R. Cucchiara, C. Tomasi. (2016). "Performance measures and a data set for multi-target, multi-camera tracking." [Online]. Available: https://arxiv.org/abs/1609.01775
[49] L. Wu, C. Shen, A. van den Hengel. (2016). "Deep linear discriminant analysis on Fisher networks: A hybrid architecture for person re-identification." [Online]. Available: https://arxiv.org/abs/1606.01595
[50] J. Zhu, H. Zeng, S. Liao, Z. Lei, C. Cai, and L. Zheng. (2017). "Deep hybrid similarity learning for person re-identification." [Online]. Available: https://arxiv.org/abs/1702.04858
[51] S. Wu, Y.-C. Chen, X. Li, A.-C. Wu, J.-J. You, and W.-S.Zheng , "An enhanced deep feature representation for person re-identification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–8.
[52] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1846–1855.
[53] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1741–1750.
[54] T. Matsukawa and E. Suzuki, "Person re-identification using CNN features learned from combination of attributes," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2428–2433.
[55] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1. Lake Tahoe, NV, USA: Curran Associates, 2012, pp. 1097–1105.
[56] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1363–1372.
[57] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
[58] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[59] Y. Deng, P. Luo, C. C. Loy, and X. Tang, "Pedestrian attribute recognition at far distance," in *Proc. 22nd ACM Int. Conf. Multimedia.*, Orlando, FL, USA, 2014, pp. 789–792.
[60] B. Moghaddam, T. Jebara, and A. Petland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, 2000.
[61] D. Gray and S. H. Brennan Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. 10th IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, 2007, pp. 1–7.
[62] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. Scand. Conf. Image Anal.*, 2011, pp. 91–102.
[63] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1988–1995.
[64] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 31–44.
[65] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," in *Proc. IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.
[66] Z. Shi, T. M. Hospedales, and T. Xiang. (2017). "Transferring a semantic representation for person re-identification and search." [Online]. Available: https://arxiv.org/abs/1706.03725
[67] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 3567–3571.

NAZIA PERWIAZ received the M.S. degree in computer software engineering from the College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad, Pakistan, where she is currently pursuing the Ph.D. degree at the School of Electrical Engineering and Computer Science. Her research interests include visual surveillance and automated person re-identification.

MUHAMMAD MOAZAM FRAZ (M'17) received the B.E. and M.S. degrees in software engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, and the Ph.D. degree from Kingston University, London, U.K., in 2013. In 2003, he has started his career as a Software Development Engineer at Elixir Technologies Corporation, a California-based software Company, where he served at various roles and capacities, including a Software Developer, Development Manager, and Program Manager, until 2010. He is a PMI (www.pmi.org) certified Project Management Professional. He held a post-doctoral position at Kingston University. He is currently an Assistant Professor with NUST-SEECS, Islamabad. His research interests include medical image analysis, computational pathology, visual recognition, and automated visual surveillance. He is currently a Rutherford Fellow at The Alan Turing Institute, London, which is the U.K.'s National Center for Data Science and AI. He is an active member of the Tissue Image Analytics Lab, University of Warwick, U.K.

MUHAMMAD SHAHZAD (S'12–M'16) received the B.E. degree in electrical engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, the M.Sc. degree in autonomous systems (robotics) from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany, and the Ph.D. degree in radar remote sensing & image analysis from the Department of Signal Processing in Earth Observation, Technische Universität München, Munich, Germany, in 2004, 2011, and 2016, respectively. His Ph.D. topic was automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic-aperture-radar image stacks. Besides, he has also attended twice two weeks professional thermography training course at the Infrared Training Center, North Billerica, MA, USA, in 2005 and 2007, respectively. He was a Guest Scientist at the Institute for Computer Graphics and Vision, Technical University of Graz, Austria, from 2015 to 2016. Since 2016, he has been an Assistant Professor at the School of Electrical Engineering & Computer Science, NUST. His research interests include machine (deep) learning applied to process both unstructured/structured 3D point clouds, optical RGBD data, and very high resolution radar images.