Optimization of Person Re-Identification through Visual Descriptors

Naima Mubariz, Saba Mumtaz, M. M. Hamayun and M. M. Fraz
School of Electrical Engineering and Computer Science,
National University of Sciences and Technology, Islamabad, Pakistan

Keywords: Public Safety and Security (PSS), Person Re-identification, Metric Learning, Visual Surveillance, Biometrics.

Abstract: Person re-identification is a complex computer vision task which provides authorities a valuable tool for maintaining high level security. In surveillance applications, human appearance is considered critical since it possesses high discriminating power. Many re-identification algorithms have been introduced that employ a combination of visual features which solve one particular challenge of re-identification. This paper presents a new type of feature descriptor which incorporates multiple recently introduced visual feature representations such as Gaussian of Gaussian (GOG) and Weighted Histograms of Overlapping Stripes (WHOS) latest version into a single descriptor. Both these feature types demonstrate complementary properties that creates greater overall robustness to re-identification challenges such as variations in lighting, pose, background etc. The new descriptor is evaluated on several benchmark datasets such as VIPeR, CAVIAR4REID, GRID, 3DPeS, iLIDS, ETHZ1 and PRID450s and compared with several state-of-the-art methods to demonstrate effectiveness of the proposed approach.

1 INTRODUCTION

Person re-identification (re-id) is a computer vision domain which has the potential to be a powerful security tool in surveillance applications (Vezzani et al., 2013). It offers features such as tracking individuals of interest within a network of cameras or retrieving video sequences containing targeted individuals etc. In person re-id task, person descriptors are used to find the true match of a query subject over a range of candidate targets which may appear significantly different in captured images (Yang et al., 2014) (see Figure 1).

Person re-id is performed in two steps: (a) descriptor generation and (b) similarity matching (Yang et al., 2014). The most challenging aspect of person re-id is the process of generating descriminative features. Many feature representation strategies currently exist and can be classified into two modes, single shot and multi-shot. In single shot mode, single image per person is used to describe a person while multi-shot mode uses multiple images per person to generate descriptors. In both modes, the extracted features should be discriminative in nature and robust to changes in light, pose, background noise, occlusion etc. Single shot methods usually exhibit low accuracies since single image is insufficient to correctly represent an identity. On the other hand, multi-shot methods are able to achieve high accuracies since multiple images for each individual are utilized to build a robust descriptor. However, the number of images used per ID for descriptor generation cannot be too high since accuracy of the descriptor needs to be balanced against computation costs. Several methods have already been proposed that are able to achieve reasonably high accuracies for person re-id. For example, Local Maximal Occurence (LOMO) features (Liao et al., 2015), ensembles of local features (ELF)

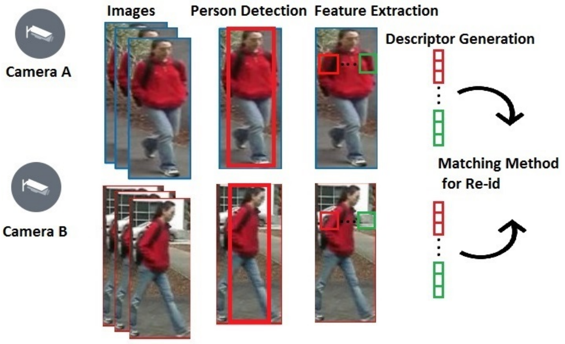

Figure 1: In person re-identification multiple images of a person are collected by camera A and B. Person detection algorithms are applied on captured images to separate person from background. Then features are extracted from images to represent a person. All the extracted features are concatenated to generate a descriptor. Then descriptors matching is performed for identification.

Mubariz, N., Mumtaz, S., Hamayun, M. and Fraz, M.
Optimization of Person Re-Identification through Visual Descriptors.
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 4: VISAPP*, pages 348-355
ISBN: 978-989-758-290-5

(Gray and Tao, 2008), interative sparse ranking (ISR) (Lisanti et al., 2015), mid level filters (MLFL) (Zhao et al., 2014) etc. Most of these feature extraction methods target a specific re-id challenge. LOMO tried to resolve pose estimation issues while the ELF approach handles viewpoint changes. The need to deal with multiple challenges simultaneously in a single descriptor therefore still exists and is open for improvement.

Similarity matching is another serious challenge for re-id since appearances and pose captured by disjoint cameras can make subjects appear significantly different. Moreover, a large subject gallery can further reduce uniqueness of descriptors. Devising methods that can correctly match descriptors can therefore be a critical challenge. The complexity further increases in case of multi-shot methods due to computation costs, as discussed earlier.

In this paper, an effective new descriptor is presented for re-id algorithms. In order to deal with several re-id challenges simultaneously, the complementary benefits of multiple feature types are utilized. Our feature descriptor incorporates the recently introduced second version of Weighted Histograms of Overlapping Stripes (WHOS) (Lisanti et al., 2015) as well as Gaussian of Gaussian (GOG) (Matsukawa et al., 2016) features into a single descriptor. WHOS features were first introduced in 2014 and the authors later proposed an extension to them in June, 2015. Extensive experiments show that our descriptor performed noticeably better in comparison to other state-of-the-art approaches on seven benchmark datasets. Experiments were further carried out to compare performance of each individual feature type to the performance achieved by the proposed descriptor.

The paper is organized as follows. Section 2 discusses related work in the domain of person re-id. Section 3 presents the proposed feature descriptor while section 4 provides details on datasets, experimental results and state-of-the-art comparison.

2 RELATED WORK

Many efforts have been made to solve person re-id challenge that are mostly focused on feature representation and feature matching (Zheng et al., 2016).

2.1 Feature Representation

This process focuses on representing meaningful patches in an image into numerical vectors which then act as descriptors for that image. Extensive efforts have been made for extracting features robust to changes in illumination, viewpoints, occlusion and intra-personal appearances (Bazzani et al., 2013; Ma et al., 2012). (Bazzani et al., 2013) exploited perceptual principles of symmetry and asymmetry to extract three different sets of features invariant to viewpoint. This method is simple yet efficient and can easily be fused with other similarity matching methods. A model is proposed by (Datta et al., 2012) that adjusts variations in illumination settings between two non-overlapping cameras. Weights are assigned to the test set based on how closely they resemble the training set. Color (HSV, RGB, YCbCr) histogram and HOG features are used to build unique descriptors. (Yang et al., 2014) proposed a novel method to solve the challenge of illumination variations by employing a color naming scheme instead of color histograms. Probabilities are assigned to each color name on the basis of the closeness of that color to a particular color name. (Zheng et al., 2015) used an unsupervised bag-of-words approach that describes each person in visual words. The descriptor has then been fused with other compatible techniques such as Gaussian masks to further improve accuracy. A fusion method is introduced in (Mumtaz S et al., 2017), where existing feature descriptors are fused together into a novel descriptor.

Many recent methods (Martinel et al., 2014; Martinel et al., 2015; Zhao et al., 2013a; Zhao et al., 2013b; Zhao et al., 2017) have explored saliency information as a novel feature for re-id. Saliency is defined as the distinctive patches of an image that differentiate it from others. The pixels or regions unique in nature define local salience for any particular image, this approach is adopted by (Zhao et al., 2013a) to help avoid incorporation of background information into feature descriptors. Saliency features proposed by (Zhao et al., 2013a) are invariant to pose and viewpoint variations. (Nguyen et al., 2016) used global salience to separate background clutter from foreground and later used local salience for re-id. Their two step approach ensures that salient patches are extracted purely from person appearance and ignores any background noise.

2.2 Feature Matching

The second step of person re-id is devising methods that will find correct feature matches through distance formulas and metric learning. LADF model proposed in (Li et al., 2013) aims to learn an adaptive decision function. Most metric learning methods aim to reduce intra class differences but the idea proposed in (Zheng et al., 2011) focuses on relative distance comparison optimization. However, this approach can become unmanageable on large datasets. Since camera settings vary from camera to camera in a multi camera network scenario, using a generalized metric learning approach is a compromise on accuracy. This problem of variations in multi camera network is tackled by (Ma et al., 2014b). Multiple metric learning strategy is proposed by (Ma et al., 2014b) where Mahalanobis distances are calculated from images of a single ID taken from the same camera as well as from different cameras in the network. The method proposed by (Xing et al., 2003) is simple yet efficient and has been utilized extensively in other person re-id techniques.

3 METHODOLOGY

The architectural work flow of the proposed re-id approach is presented in Figure 2. Features and similarity learning employed in this method to optimize person re-id are discussed below.

3.1 Feature Representation

3.1.1 Hierarchical Gaussian Features

(Matsukawa et al., 2016) proposed a novel descriptor, namely Gaussian of Gaussian (GOG), that utilizes appearance based features. It is a pixel wise feature extraction method which incorporates hierarchical distribution of features. Many other such descriptors have been proposed but they lack mean information of the extracted features which makes them less discriminative.

In this method, a part based approach is employed where the whole image is divided into G regions. The following are the main components of GOG feature extraction:

- *Dense patch extraction* Given a region $G1$, a window of $k \times k$ size is moved along the region to extract dense features (intensity, gradient and color) in RGB space from the selected patches. The pixel features defined as:

$$f_p = [y, M^\circ_0, M^\circ_{90}, M^\circ_{180}, M^\circ_{270}, R, G, B]^T \quad (1)$$

where f_p is a pixel feature for every pixel p in the region, y is vertical position of pixel, $M_\theta \in 0^\circ, 90^\circ, 180^\circ, 270^\circ$ are orientation magnitudes in four directions and R, G, B are color channel values in RGB color space.

- *Patch Gaussian* Gaussian distribution is applied on the features extracted to outline the details within each patch.
- *Flatten patch Gaussian* Gaussian patches are projected to a tangent space where application of Euclidean algorithm is feasible.
- *Region Gaussian* The same distribution is applied on the summarized patches to get an overall understanding of a particular region. To avoid incorporating background clutter in the feature vector, patches are assigned weights. Patches aligned near to center are assigned higher weights in comparison to other patches since persons are generally located in the center of an image.
- *Feature vector* All the region Gaussians $\{z_g\}_{g=1}^{G}$ are concatenated into an efficient image descriptor, which is defined as:

$$z = [z_1^T, z_2^T, \ldots, z_G^T]^T \quad (2)$$

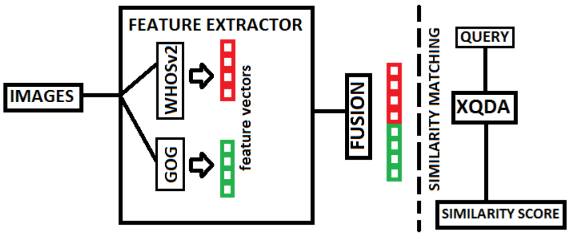

Figure 2: Both GOG and WHOSv2 features are extracted from input images. Feature vectors generated are then concatenated to form a single descriptor. In descriptor matching stage, similarity of query image to gallery image is calculated using XQDA metric learning and used to determine the correct ID match.

3.1.2 Weighted Histograms of Overlapping Stripes Version2

Weighted histogram of overlapping stripes (WHOS) (Lisanti et al., 2015) is another successful state-of-the-art person image descriptor. An improved version of the WHOS descriptor was later released by the authors in 2015, refered to as WHOSv2. Where the WHOS descriptor only extracts HS, RGB and HOG features, WHOSv2 extention adds Lab color histograms and local-binary-pattern (LBP) (Guo et al., 2010).

- *Feature Extraction*
  - The color histograms are extracted from an image in two levels: 1) The image is segmented into eight horizontal regions. 2) The image is segmented into seven overlapping horizontal regions.
  - For HOG and LBP features 8 pixels are removed from the sides and features are extracted from a grid of $n \times n$ cell.
  - The color and texture histogram based feature vector H is defined as:

$$H = [HS, RGB, Lab, HOG, LBP] \quad (3)$$

- *Epanechnikov mask* In both levels, color histograms are weighted using Epanechnikov filters and later concatenated with texture (HOG, LBP) features. The color histograms make the descriptor invariant to changes in brightness and Epanechnikov kernel helps exclude background information.

3.1.3 Fusion

In fusion step (Figure 2), GOG in RGB space (GOG_{RGB}) and WHOSv2 descriptors are simply concatenated to form an effective descriptor. The proposed descriptor is defined as:

$$F = [z, H] \quad (4)$$

where, z represent GOG_{RGB} features and H represent WHOSv2 features. The dimension of GOG_{RGB} and WHOSv2 is 7567 and 5138 respectively. Thus, the proposed descriptor has 12,705 dimensions.

For GOG_{RGB}, the image is divided into 7 regions (G=7) and 8-dimensional features are extracted from a window of 5×5 (k=5). Features are extracted from a total of 15 horizontal sections (8 non-overlapping and 7 overlapping) for WHOSv2 descriptor. The HS histograms are quantized to 8×8 bins, whereas RGB and Lab contains $4 \times 4 \times 4$ bins. The HOG histograms are extracted from a grid of 2×2 cells (n=2) and LBP histogram contains 58 bins.

3.2 Metric Learning

3.2.1 XQDA

XQDA (Liao et al., 2015) has achieved state-of-the-art performances in person re-id and face recognition. The metric learning is an extension to Bayesian-face algorithm (Moghaddam et al., 2000) and KISSME (Koestinger et al., 2012) and is not sensitive to dimensions rather it learns to reduce dimensions automatically. XQDA project features into a discriminative subspace $W = (w_1, w_2, ..., w_n) \in R^{d \times n}$ and also learns an efficient distance metric. Lets assume C number of classes with $\{X, Z\}$ training set, here $X = (x_1, x_2, ..., x_n) \in R^{d \times r}$ represent samples from one view in a space with dimension d and $Z = (z_1, z_2, ..., z_m) \in R^{d \times m}$ represent samples from other view in the same space. The goal of learning subspace W is to reduce intra-personal and increase extra-personal differences. The XQDA metric learning solve the multi-class classification problem by differentiating between two classes that is intra-personal variations and extra-personal variations.

4 EXPERIMENTAL EVALUATION

4.1 Performance Measure

Rank matching is a commonly used performance metrics for person re-id. The task is formulated as a ranking problem that gives r matching rates. A Rank n score therefore shows that the correct match exists within the first n ranked entries. CMC (cumulative matching characteristic) curves are also computed to represent rank matching rates. In experiments, average rank scores are computed over 10 trials. In each trial, probe images are matched with gallery images and rankings are generated based on similarity scores.

4.2 Datasets

Sample images of dataset are presented in Figure 3. The following benchmark datasets have been used to test the proposed feature descriptor and re-id approach:

4.2.1 VIPeR

VIPeR (Gray and Tao, 2008) is the most widely used dataset for person re-id. Image characteristics such as low resolutions, pose variations, background noise and illumination conditions make it a challenging dataset. It consists of 632 identities taken from two non-overlapping cameras in an outside setting where two images are captured per person, one from front view and other from side view.

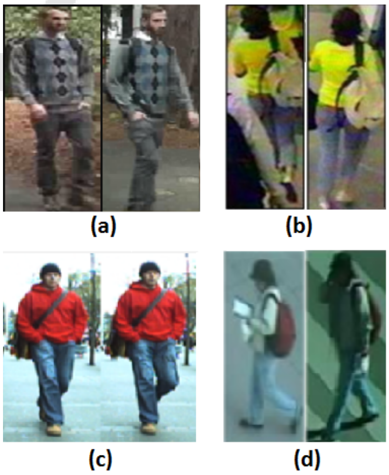

Figure 3: Example images show variations in (a) background (b) viewpoint (c) pose and (d) illumination. Images are from VIPeR, GRID, ETHZ1 and PRID450s respectively.

4.2.2 GRID

The GRID (Loy et al., 2013) captures an underground station scenario and contains 250 identities and 1275 images in total. There are 250 paired and 775 unpaired identities with no actual image match. On average 10 images are collected per person from 8 disjoint cameras. The images are of poor quality with background clutter, occlusion and pose variations.

4.2.3 PRID450s

PRID450s (Roth et al., 2014) is extracted from PRID2011 (Hirzer et al., 2011) and contains 450 identities captured from two static disjoint cameras. It is a relatively new dataset and is more realistic compared to the VIPeR dataset due to significant pose, illumination, viewpoint and occlusion variations. The dataset is only suitable for single shot testing as one sample per person is available.

4.2.4 CAVIAR4REID

CAVIAR4REID (Cheng et al., 2011) is generated by two disjoint cameras in a shopping mall capturing a real world surveillance scenario. It contains 72 identities and each identity has on average 10 images per person. The images collected are of very poor quality with large occlusion conditions.

4.2.5 3DPeS

The 3DPeS dataset (Baltieri et al., 2011a) is introduced in 2011. This dataset is an extension of Sarc3D dataset (Baltieri et al., 2011b) with total of 1012 images of 193 identities captured from 8 non-overlapping calibrated cameras in an outdoor campus setting. Multiple images, varying from 2 to 26 for each person are available except for one identity which has only one image. The dataset also includes masks for all images which assists in removing background noise in feature extraction process.

4.2.6 iLIDS

The iLIDS (Zheng et al., 2009) is taken from larger iLIDS MCTS collected in busy public place. Most of the images have huge illumination variations along with large occlusion conditions like baggage and crowds. The dataset contains 119 identities with total of 476 images and each person has 4 images on average with a resolution of 128 × 64.

4.2.7 ETHZ1

ETHZ dataset (Ess et al., 2007; Schwartz and Davis, 2009) is collected in an outdoor street scenario through a single non-static camera and gives a range of changes in human appearances. The non-static camera is an extra challenge of ETHZ dataset that has 3 different sequences and each of them have multiple images per person making it suitable for multi-shot testing. There are 83 identities with 4857 images in ETHZ1 dataset with variations in illumination and scale.

4.3 Results

For the experiments, half of the dataset images are randomly labeled gallery while the remaining half are labeled as probe. Images are selected randomly for train and test phases and evaluation is performed in 10 trials to get an average result. Table 1 and 2 present comparison of proposed method on the benchmark dataset with state-of-the-art methods.

VIPeR is compared with state-of-the-art methods such as LOMO (Liao et al., 2015), SCNCD (Yang et al., 2014), kBiCov (Ma et al., 2014a) and others. The proposed approach at rank 1 recognition rate of 44.97% outperforms the other methods by more than 4%. The performance of proposed feature descriptor is compared with other methods using CMC as shown in Figure 4. The comparison show that the proposed approach gives the best performance on the VIPeR dataset. In the GRID dataset, 250 paired and 755 unpaired images make up gallery set and remaining 250 makes probe set. The best results reported on GRID are 16.56% by LOMO (Liao et al., 2015) but the proposed method outperforms LOMO by 7.12% as reported in Table 2. LOMO (Liao et al., 2015) report high accuracy on PRID450s dataset. Comparison with the proposed method shows that it performs even better with rank 1 rate of 60.93%.

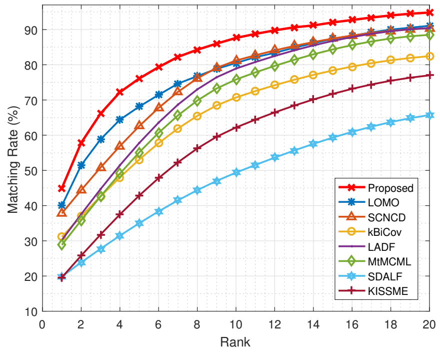

Figure 4: Comparing performance using CMC curves on VIPeR dataset.

Table 1: State-of-the-art comparison on benchmark datasets.

	VIPeR		GRID		PRID450s		CAVIAR		3DPeS		iLIDS		ETHZ1
	R#1	R#20	R#1	R#20	R#1	R#20	R#1	R#20	R#1	R#20	R#1	R#20	R#1
Proposed (GOG+WHOSv2)	44.97	94.84	23.68	68.24	60.93	94.89	35.14	84.81	67.13	96.08	52.64	95.03	97.69
LOMO (Liao et al., 2015)	40.00	91.08	16.56	52.40	58.44	92.31	31.21	82.43	56.46	91.77	41.20	93.51	96.02
SCNCD (Yang et al., 2014)	37.80	90.40	-	-	41.6	87.8	-	-	-	-	-	-	-
kBiCov (Ma et al., 2014a)	31.11	82.45	-	-	-	-	-	-	-	-	-	-	-
LADF (Li et al., 2013)	30.22	90.44	6.00	41.28	-	-	33.28	-	27.0	83.2	-	-	-
MtMCML (Ma et al., 2014b)	28.83	88.51	14.08	59.84	-	-	-	-	-	-	-	-	-
SDALF (Bazzani et al., 2013)	19.87	65.73	-	-	-	-	12.25	-	-	-	-	-	90.2
KISSME (Koestinger et al., 2012)	19.60	77.00	10.64	43.20	36.31	83.69	33.45	93.26	33.0	78.8	-	-	-

⁻Results are not reported.

Table 2: Improvement on state-of-the-art methods.

	VIPeR R#1 improve. %	GRID R#1 improve. %	PRID450s R#1 improve. %	CAVIAR R#1 improve. %	3DPeS R#1 improve. %	iLIDS R#1 improve. %	ETHZ1 R#1 improve. %
LOMO (Liao et al., 2015)	4.97	7.12	2.49	3.93	10.67	11.44	1.67
SCNCD (Yang et al., 2014)	7.17	-	19.33	-	-	-	-
kBiCov (Ma et al., 2014a)	13.86	-	-	-	-	-	-
LADF (Li et al., 2013)	14.75	17.68	-	1.86	40.13	-	-
MtMCML (Ma et al., 2014b)	16.14	9.6	-	-	-	-	-
SDALF (Bazzani et al., 2013)	25.1	-	-	22.89	-	-	7.49
KISSME (Koestinger et al., 2012)	25.37	13.04	24.62	1.69	34.13	-	-

Multi-shot evaluation is performed on full CAVIAR4REID dataset. Images are resized to 128 × 48 and 36 identities are assigned for both training and testing. The proposed method outperforms LOMO by 3.93%. Comparisons with several state-of-the-art person re-id methods including LADF (Li et al., 2013), LOMO (Liao et al., 2015), SDALF (Bazzani et al., 2013) and KISSME (Koestinger et al., 2012) are also presented in Table 1. The Table 2 represent improvement of proposed method and it give high accuracy compared to state-of-the-art methods. The proposed approach gives best performance with rank 1 rate of 35.14%. In experiments on 3DPeS, identity 139 was excluded because re-id cannot be performed on single image. Gallery images are picked from camera view 1 and probe images from camera view 2. As observed in Table 1, the proposed descriptor gives the best results. The proposed descriptor was also compared with LOMO since both employ same matching technique. Rank 1 rate of LOMO is 56.46% whereas the proposed approach improves this result by more than 10%. The resulting multi-shot CMC are illustrated in Figure 5 on 3DPeS dataset.

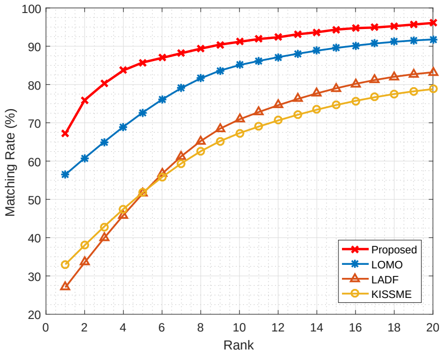

Figure 5: Comparing performance using CMC curves on 3DPeS dataset.

For iLIDS, 60 identities are randomly selected for training phase and tests are performed on the remaining 59 identities. LOMO (Liao et al., 2015) rank 1 results on iLIDS is 41.20%. Table 1 presents state-of-the-art methods results and comparison shows that proposed method have achieved significant improvement on others. The rank 1 rate is more than 50% and all the methods in comparison report results less than 42%.

In ETHZ1, total images vary from 7 to 226 per person. In evaluation, 10 images per person were selected except for 2 identities which have less than 10 images in dataset. Images are selected randomly with 42 identities in training set and 41 identities in test set. As seen in Table 1, performance of proposed method is better than both SDALF (Bazzani et al., 2013) and LOMO (Liao et al., 2015) with recognition accuracy of 97.69%.

5 CONCLUSIONS

In this paper, we analyzed performances of state-of-the-art person re-id features. We created a new feature descriptor that incorporates some of the best performing feature extraction methods. The recently introduced GOG and WHOSv2 features were used to compute a discrminative feature descriptor that incorporates both color and texture information. Extensive experiments on a large number of datasets demonstrate that the proposed descriptor and similarity matching works better than the most state-of-the-art re-id algorithms. Re-id accuracy has been improved on VIPeR, GIRD, PRID450s, CAVIAR4REID, 3DPeS, iLIDS and ETHZ1 to 44.97%, 23.68%, 60.93%, 35.14%, 67.13%, 52.64% and 97.69% respectively. While experiments have only been conducted on closed set re-id datasets, the same descriptor should perform well in an open set setting as well. For the future, we plan to perform multimodal person re-id incorporating anthropometric measures and thermal features. We hope to explore their impact on accuracy rates and evaluating their fusion with visual descriptors.

REFERENCES

Baltieri, D., Vezzani, R., and Cucchiara, R. (2011a). 3dpes: 3d people dataset for surveillance and forensics. In *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, pages 59–64. ACM.

Baltieri, D., Vezzani, R., and Cucchiara, R. (2011b). Sarc3d: a new 3d body model for people tracking and re-identification. *Image Analysis and Processing–ICIAP 2011*, pages 197–206.

Bazzani, L., Cristani, M., and Murino, V. (2013). Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144.

Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L., and Murino, V. (2011). Custom pictorial structures for re-identification. In *Bmvc*, volume 2, page 6.

Datta, A., Brown, L. M., Feris, R., and Pankanti, S. (2012). Appearance modeling for person re-identification using weighted brightness transfer functions. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2367–2370. IEEE.

Ess, A., Leibe, B., and Van Gool, L. (2007). Depth and appearance for mobile scene analysis. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

Gray, D. and Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. *Computer Vision–ECCV 2008*, pages 262–275.

Guo, Z., Zhang, L., and Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6):1657–1663.

Hirzer, M., Beleznai, C., Roth, P. M., and Bischof, H. (2011). Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, pages 91–102. Springer.

Koestinger, M., Hirzer, M., Wohlhart, P., Roth, P. M., and Bischof, H. (2012). Large scale metric learning from equivalence constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE.

Li, Z., Chang, S., Liang, F., Huang, T. S., Cao, L., and Smith, J. R. (2013). Learning locally-adaptive decision functions for person verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3610–3617.

Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206.

Lisanti, G., Masi, I., Bagdanov, A. D., and Del Bimbo, A. (2015). Person re-identification by iterative re-weighted sparse ranking. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1629–1642.

Loy, C. C., Liu, C., and Gong, S. (2013). Person re-identification by manifold ranking. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 3567–3571. IEEE.

Ma, B., Su, Y., and Jurie, F. (2012). Local descriptors encoded by fisher vectors for person re-identification. In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 413–422. Springer.

Ma, B., Su, Y., and Jurie, F. (2014a). Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390.

Ma, L., Yang, X., and Tao, D. (2014b). Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing*, 23(8):3656–3670.

Martinel, N., Micheloni, C., and Foresti, G. L. (2014). Saliency weighted features for person re-identification. In *ECCV Workshops (3)*, pages 191–208.

Martinel, N., Micheloni, C., and Foresti, G. L. (2015). Kernelized saliency-based person re-identification through multiple metric learning. *IEEE Transactions on Image Processing*, 24(12):5645–5658.

Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y. (2016). Hierarchical gaussian descriptor for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1363–1372.

Moghaddam, B., Jebara, T., and Pentland, A. (2000). Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782.

Mumtaz S, M. N., S, S., and M, F. M. (2017). Weighted hybrid features for person re-identification. In *International Conference on Image Processing Theory, Tools and Applications*, pages 1–8. IEEE.

Nguyen, T. B., Pham, V. P., Le, T.-L., and Le, C. V. (2016). Background removal for improving saliency-based person re-identification. In *Knowledge and Systems Engineering (KSE), 2016 Eighth International Conference on*, pages 339–344. IEEE.

Roth, P. M., Hirzer, M., Koestinger, M., Beleznai, C., and Bischof, H. (2014). Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, pages 247–267. Springer.

Schwartz, W. R. and Davis, L. S. (2009). Learning discriminative appearance-based models using partial least squares. In *Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on*, pages 322–329. IEEE.

Vezzani, R., Baltieri, D., and Cucchiara, R. (2013). People reidentification in surveillance and forensics: A survey. *ACM Computing Surveys (CSUR)*, 46(2):29.

Xing, E. P., Jordan, M. I., Russell, S. J., and Ng, A. Y. (2003). Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*, pages 521–528.

Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., and Li, S. Z. (2014). Salient color names for person re-identification. In *European conference on computer vision*, pages 536–551. Springer.

Zhao, R., Ouyang, W., and Wang, X. (2013a). Person re-identification by salience matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2528–2535.

Zhao, R., Ouyang, W., and Wang, X. (2013b). Unsupervised salience learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593.

Zhao, R., Ouyang, W., and Wang, X. (2014). Learning mid-level filters for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 144–151.

Zhao, R., Oyang, W., and Wang, X. (2017). Person re-identification by saliency learning. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):356–370.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015). Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124.

Zheng, L., Yang, Y., and Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*.

Zheng, W.-S., Gong, S., and Xiang, T. (2009). Associating groups of people. In *BMVC*, volume 2.

Zheng, W.-S., Gong, S., and Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. In *Computer vision and pattern recognition (CVPR), 2011 IEEE conference on*, pages 649–656. IEEE.