Hybrid Distance Metric Learning for Real-Time Pedestrian Detection and Re-identification

Xinyu Huang(✉), Jiaolong Xu, Gang Guo, and Ergong Zheng

Aviation University of Airforce, Changchun 130022, China
xinyu_huang121@126.com

Abstract. Cross-camera pedestrian re-identification (re-ID) is of paramount importance for surveillance tasks. Although considerable progress has been made to improve the re-ID accuracy, real-time pedestrian detection and re-ID remains a challeging problem. In this work, first, we proposed an enhanced aggregated channel features (ACF+) based on the ACF pedestrian detector [1] for real-time pedestrian detection and re-ID; Second, to further improve the representation power of the combined multiple channel features, we proposed a novel hybrid distance metric learning method. Extensive experiments have been carried on two public datasets, including VIPeR, and PRID2011. The experimental results show that our proposed method can achieve state-of-the-art accuracy while being computational efficient for real-time applications. The proposed hybrid distance metric learning is general, thus can be applied to any metric learning approaches.

Keywords: Pedestrian re-identification · Aggregated channel features · Hybrid distance metric learning

1 Introduction

Surveillance systems use the network of cameras to cover large public spaces (*i.e.* airport terminals, train stations, *etc.*). Cross-camera pedestrian re-identification (re-ID) is crucial for visual surveillance tasks. Due to the large variations in camera viewpoints, different background, human poses, illumination changes, partial occlusion, low resolution and motion-blur *etc.*, pedestrian re-ID is a challenging task.

Pedestrian re-ID can be divided into two steps: (1) feature extraction, *i.e.*, extracting features from detected pedestrian candidates; (2) matching, *i.e.*, matching a given *prob* pedestrian against a *gallery* of candidate ones. For feature representation, many delicate handcrafted features have been proposed, *e.g.* Hist-Moment feature [5], LOMO [11]. Recently, deep convolutional neural networks are used to learn robust feature representation [2,3]. Regarding to the matching, various metric learning methods have been proposed to improve matching accuracy, Liao *et al.* [12] proposed a weighting positive and negative samples differently with a positive semidefinite constraint. In [15], the differences

M. Liu et al. (Eds.): ICVS 2017, LNCS 10528, pp. 448–458, 2017.
DOI: 10.1007/978-3-319-68345-4_40

and commonness between image pairs are both considered and presented that the covariance matrices of dissimilar pairs can be inferred from similar pairs. Besides that, kLFDA [5] and XQDA [11] are very popular ones. Although continuous progress has been made in recent years, little focus has been paid on real-time re-ID. In this work, the re-ID computation cost is one of our major concerns.

For re-ID applications, pedestrian detection is usually regarded as a separate task or pre-processing step, or assuming the pedestrian bounding boxes have been provided. In this work, we consider the detection and re-ID in the joint pipeline and take advantages of detection to extract features for re-ID. As ACF pedestrian detector is known to be both fast and accurate [1], we build our re-ID framework based on it. Though features can be automatically extracted during detection, the obtained feature representations might be suboptimal for re-ID tasks. This is because the features are originally designed for classifying pedestrians and background. For example, the aggregated channel features extracted by ACF detector is consist of LUV channels (3 channels), gradient magnitude channel (1 channel) and histogram of oriented gradient channels (6 channels). These features are very discriminative for classifying pedestrians and background, but not for distinguishing different pedestrian instances.

In this paper, our contributions are threefolds: (1) We proposed an enhanced ACF (ACF+) by integrating HSV histogram and SILTP [11] features as additional channels, and it has already achieved comparable accuracy to the state-of-the-art LOMO feature; (2) We further explored better distance metric learning methods to leverage the representation power of different feature subspaces and then proposed a hybrid distance metric learning method which can find the best combination of distance metric for the aforementioned channel features; (3) We build a real-time pedestrian re-ID framework based on the proposed method. The experimental results on multiple public datasets show the efficacy and low computational cost of the proposed method. As the hybrid distance metric learning is general, it can be applied to any metric learning algorithm.

2 Related Work

For pedestrian Re-ID, many works focus on how to construct effective feature representation and numerous features are proposed. The most effective features can be categoried into two groups: color-based and texture-based features [3]. Among them, HSV color histogram [4], LAB color histogram [16], LBP histogram [9] and their fusion are the most popular ones. Recently, LOMO is proposed to combine HSV histogram and SILTP [11], which acheives state-of-the-art re-ID accuracy. Other methods consider combining other cues of pedestrians. For example, [14] proposed to incoporate human part-based features and [4] proposed to use silhouette and symmertry structure of person. Salience based features is used in [16] which can deal with large pose variations. Notably, none of these work consider resuing the features extracted from prior detection process. As feature extraction are normally computational expensive, reusing the features extracted from detection step can save a lot of computation time and thus facilitate real-time re-ID. In this work, we build our re-ID framework based on ACF detector [1]. Using ACF features alone is not sufficient for effective re-ID representation. Inspired from the LOMO feature, we proposed an enhanced ACF features (ACF+) by combining HSV histogram and SILTP to form robust features.

Based on the extracted features, metric learning are commonly used for measuring the similarity of two samples due to its flexibility. Most researches used holistic metric for feature matching, others took part information into account and divide samples into several groups [3]. In [5], various kernel-based metric learning methods have been evaluated. Among them, kLFDA with linear or RBF kernel has shown robust performance. Recently, [11] proposed XQDA based on Mahalanobis distance and achieved state-of-the-art results. These metric learning algorithms handle the extracted features as a single vector without considering the underlying components. In this work, we argue that instead of combining the ACF+ as a single vector, we can learn different distance functions according to the type of channels. After that, we combine all these distances by a hybrid distance metric. Our experimental results show that the hybrid distance metric obtained significantly better results than simple concatenation.

3 Proposed Methods

In this section, we first give an overview of the detection and re-identification framework, then we introduce the novel feature representation ACF+. Finally, we illustrate the proposed hybrid distance metric learning method.

3.1 Pedestrian Detection and Re-identification Framework

The re-identification framework is as follows. We first apply the ACF detector to detect pedestrian candidates. The bounding boxes of pedestrians and the corresponding ACF features are obtained at this stage. Then HSV histogram and SILTP features are extracted from the bounding box locations and integrated with ACF to form ACF+. As the dimension of ACF+ is large and there might be redundant information among the channels, we make use of hybrid distance metric learning to reduce dimensions of ACF+ and improve its discriminative power.

3.2 Enhanced Aggregated Channel Features (ACF+)

As we will see from the experiments, the ACF extracted by detector has limited representation capability for re-ID task, thus we consider to enhance it with other discriminative features. HSV histogram and SILTP (Scale Invariant Local Ternary Pattern) [10] descriptors are known to be effective for re-ID task from the previous literatures. HSV histogram feature can improve performance with color information and SILTP contains texture information. In this work, we extract

HSV histogram and SILTP features as additional channels to the ACF. For HSV color feature, we compute histograms on each channel with bin size of 8, and there are 24 channels in total for the final feature map. SILTP is computed with two scales with radius [3,5], neighbour points of 4, and scale factor of 0.3, which constructs a 2-channel feature map. All these feature maps are concatenated with ACF and forms the enhanced ACF, which we call it ACF+. The final feature representation contains following feature channels: 3 LUV channels, 1 normalized gradient magnitude channels, 6 histogram of oriented gradient channels, 24 HSV histogram channels, and 2 SILTP channels. The entire feature has dimensions of 21642. The advantages of ACF+ are twofold: (1) saving the computation time of features for re-identification; (2) improving re-ID performance by fusing different types of features.

3.3 Metric Learning

In this work, we apply a subspace metric learning method called Cross-view Quadratic Discriminant Analysis (XQDA) to the extracted features for pedestrian re-ID. XQDA has shown superior performance with LOMO feature. So we first investigate its application to the proposed ACF+. Next, we propose a hybrid distance metric learning to the split ACF+ feature spaces to further boost its matching accuracy.

Cross-View Quadratic Discriminant Analysis. In this section, we first briefly review the XQDA method. The XQDA algorithm is inspired from the Bayesian face recognition and the KISSME algorithm. These methods define the intra-class variation and inter-class variation as two classes of zero-mean Gaussian distribution, which can be formulated as the following:

$$\begin{aligned} p_s(\mathbf{x}_i, \mathbf{x}_j) &= \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_s|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_i-\mathbf{x}_j)^\top \mathbf{\Sigma}_s^{-1}(\mathbf{x}_i-\mathbf{x}_j)} \\ p_d(\mathbf{x}_i, \mathbf{x}_j) &= \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}_d|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_i-\mathbf{x}_j)^\top \mathbf{\Sigma}_d^{-1}(\mathbf{x}_i-\mathbf{x}_j)}, \end{aligned} \tag{1}$$

where $\mathbf{x}_i - \mathbf{x}_j$ is the difference of two samples, p_s is the intra-class distribution, *i.e.* the probability of *same* individual, p_d is the inter-class distribution, *i.e.* the probability of *different* individual. $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_d$ are the covariance matrices of intra-class and inter-class respectively. The log likelihood ratio is used to estimate the two Gaussian distributions:

$$\begin{aligned} d(\mathbf{x}_i, \mathbf{x}_j) &= log\left(\frac{p_d(\mathbf{x}_i,\mathbf{x}_j)}{p_s(\mathbf{x}_i,\mathbf{x}_j)}\right) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^\top \left(\mathbf{\Sigma}_s^{-1} - \mathbf{\Sigma}_d^{-1}\right)(\mathbf{x}_i - \mathbf{x}_j). \end{aligned} \tag{2}$$

In theory, the distance function of (2) can be directly used for pedestrian re-ID. However, the original feature dimension of $\mathbf{x}$ is usually large, and a lower dimension space is preferred for classification. XQDA extends the Bayesian face recognition and KISSME algorithm and proposes to learn a subspace and at the same time a distance function. So the distance function of (2) in the reduced subspace can be written as:

$$d_Q\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\mathbf{x}_i - \mathbf{x}_j\right)^\top \mathbf{Q}\left(\boldsymbol{\Sigma}_s^{'-1} - \boldsymbol{\Sigma}_d^{'-1}\right)\mathbf{Q}^\top\left(\mathbf{x}_i - \mathbf{x}_j\right), \tag{3}$$

where $\mathbf{Q}$ is the projection matrix which projects the $\mathbf{x}$ to a lower subspace, $\boldsymbol{\Sigma}_s^{'} = \mathbf{Q}^\top\boldsymbol{\Sigma}_s\mathbf{Q}$ and $\boldsymbol{\Sigma}_d^{'} = \mathbf{Q}^\top\boldsymbol{\Sigma}_d\mathbf{Q}$. In [11], it is shown that optimizing the projection matrix $\mathbf{Q}$ is equivalent to solve the following optimization problem:

$$\max_Q \mathbf{Q}^\top\boldsymbol{\Sigma}_d\mathbf{Q}, \quad s.t.\, \mathbf{Q}^\top\boldsymbol{\Sigma}_s\mathbf{Q} = 1, \tag{4}$$

which can be solved by the generalized eigenvalue decomposition method.

Hybrid Distance Metric Learning. In this section, we explain our hybrid distance metric learning based on ACF+ and XQDA. As ACF+ is a concatenation of different types of features, each of them may have different representation abilities. We propose to split the ACF+ into separate feature space according to the feature type and learn individual distance metric for each type. Each individual metric may have its specific discriminative ability, our goal is to find the optimal combination of them to boost the overall re-ID performance.

The straightforward way is to apply ensemble learning methods, *i.e.* combining multiple hypothesis to form a (hopefully) better hypothesis. For example, voting mechanism can be used to choose the best distance metric among the individual distances. We denote it by $\mathbf{d}_{ij}^k, i \in M, j \in N$ the distance of sample $\mathbf{x}_i$ and $\mathbf{x}_j$, where $k \in K$ is the kth feature type in ACF+ and K is the number of types, M is the size of gallery set and N is the size of probe set. The maximum voting can be expressed as: $\mathbf{d}_{ij}^* = \max_k \mathbf{d}_{ij}^k$, where $\mathbf{d}_{ij}^*$ is the final distance for re-ID. Similarly, we can calculate minimum and average voting $\mathbf{d}_{ij}^* = \min_k \mathbf{d}_{ij}^k$ and $\mathbf{d}_{ij}^* = \frac{1}{K}\sum_{k=1}^K \mathbf{d}_{ij}^k$. These three methods are denoted by MAX, MIN and MEAN respectively.

However, such methods might be suboptimal as they are not able to learn hybrid distance from the training data. To address this problem, we propose to learn the combination weights of individual distance function from the training data. We denote $\mathbf{D}^k = \{\mathbf{d}_{ij}^k | i \in M, j \in N, k \in K\}$ the distance matrix of individual feature. Our goal is to learn parameters $\mathbf{w} = \{\omega_k | k \in K\}$ to obtain final distance matrix $\mathbf{D}^*$:

$$\mathbf{D}^* = \sum_k \omega_k * \mathbf{D}^k, \tag{5}$$

which can be explicitly written as:

$$\begin{bmatrix} \omega_1 d_{11}^1 + \omega_2 d_{11}^2 + \cdots + \omega_K d_{11}^K & \cdots & \omega_1 d_{1N}^1 + \omega_2 d_{1N}^2 + \cdots + \omega_K d_{1N}^K \\ \vdots & \ddots & \vdots \\ \omega_1 d_{M1}^1 + \omega_2 d_{M1}^2 + \cdots + \omega_K d_{M1}^K & \cdots & \omega_1 d_{MN}^1 + \omega_2 d_{MN}^2 + \cdots + \omega_K d_{MN}^K \end{bmatrix}. \tag{6}$$

Because in the training set, we can obtain the ground truth identities of the pedestrian, and thus define the perfect matching distances of $\mathbf{D}^*$. In this practice, we let $\mathbf{d}_{ij}^* = 0$ if sample i and sample j are from the same identity, otherwise

$\mathbf{d}^*_{ij} = 1$. Note that $\mathbf{D}^*$ and $\mathbf{D}^k$ are matrices of size $M \times N$. We can flatten $\mathbf{D}^*$ and $\mathbf{D}^k$ into column vectors, denoted by $\mathbf{d}^*$ and $\mathbf{d}^k$ respectively. Equation (5) then can be written as:

$$\mathbf{d}^* = \mathbf{w}^\top \mathbf{z}, \tag{7}$$

where $\mathbf{z} = [\mathbf{d}^1, \mathbf{d}^2 \ldots \mathbf{d}^K]$. As we have already known the perfect distance matching values of $\mathbf{d}^*$, (7) can be converted into to a linear regression problem, where $\mathbf{z}$ is the input variables, $\mathbf{d}^*$ is the measured variables, and $\mathbf{w}$ is the K-dimension parameter vector. Equation (7) can be easily solved by off-the-shelf linear regression solvers. At testing time, we can apply the learned parameters $\mathbf{w}$ to the input $\bar{\mathbf{d}} = \mathbf{w}^\top \mathbf{z}$, where $\bar{\mathbf{d}}$ is the predicted final distances. Compare to the simple ensemble methods, the hybrid distance learning method can learn optimal parameters from the training data. This is verified in the experiment section.

4 Experiments and Analysis

In this section we describe the set of experiments used to evaluate efficiency of the proposed feature representations as well as the hybrid distance metric learning method.

4.1 Datasets and Experimental Protocol

Our methods were evaluated on two public datasets which are commonly used throughout the literatures. We repeat the experimental process 10 times and average the performance for the final report.

The **VIPeR** [13] dataset is a challenging test bed for person re-id. It contains 632 identities and each has two images captured outdoor from two cameras with different views and illumination intensity. Meanwhile the whole images are scaled to 128×48 pixels. The 632 image-pairs are randomly divided into two parts, one is used as training set and the other for testing.

The **PRID2011** [6] dataset consists of person images captured from two cameras. We use the singleshot version of this dataset in our experiments. Specifically, Camera A captures 385 persons and camera B captures 749 persons. Only 200 people appear in both of them. We follow the data splitting of [7]. For training set, we randomly choose 100 identities from camera A and their counterparts from camera B, while the remaining 100 identities of camera A are used as the probe set, and the remaining 649 samples of camera B are used as the gallery set.

Evaluation Protocol. Our experiments follow the protocol in [11]. The dataset is divided into training set and test set, and test set further divided into two parts, one is the gallery set, and another is the probe set. Then we match each probe image with every image in the gallery set and rank the similarity scores. The experimental results is evaluated by *CMC* (Cumulative Matching Characteristic) curves [11], which is an estimation of the expectation of finding the correct match in the top n matches.

4.2 Performance of the ACF+ Feature

First we conduct experiments to evaluate the performance of the proposed ACF+ feature representation. We compared ACF+ to four different feature representations, namely ACF, Hist-HSV, HistMoment-6Patch and LOMO, where ACF is the original features extracted from ACF detector, Hist-HSV is the channel feature of HSV histograms, HistMoment-6Path is the patch based features in [5], and LOMO is the state-of-the-art feature proposed in [11]. Table 1 shows the matching scores on top 1 to top 20, and Fig. 1 depicts the CMC curves. It can be seen that using ACF only results in poorest performance as ACF is not discriminative for intra-class variations of the pedestrians. Hist-HSV alone shows much better results and even outperforms a patch based feature representation HistMoment-6Path, which integrates LBP, RGB and HSV histograms. From this comparison we can see that channel-feature based representation like Hist-HSV is competitive to the path-based ones, while being much faster as can be seen from the following experiments. When integrating Hist-HSV and SILTP with ACF to generate ACF+, we achieved comparable accuracy with the state-of-the-art LOMO feature.

Table 1. CMC scores on VIPeR (higher the better)

Method	Rank1	Rank5	Rank10	Rank15	Rank20
ACF	2.88	8.70	14.27	18.92	22.91
Hist-HSV	20.25	42.37	55.89	64.87	71.80
HistMomet-6Patch [8]	19.37	41.39	53.35	61.08	66.46
LOMO [11]	40.28	68.32	80.89	87.12	91.20
ACF+	36.39	66.68	79.34	86.04	90.60

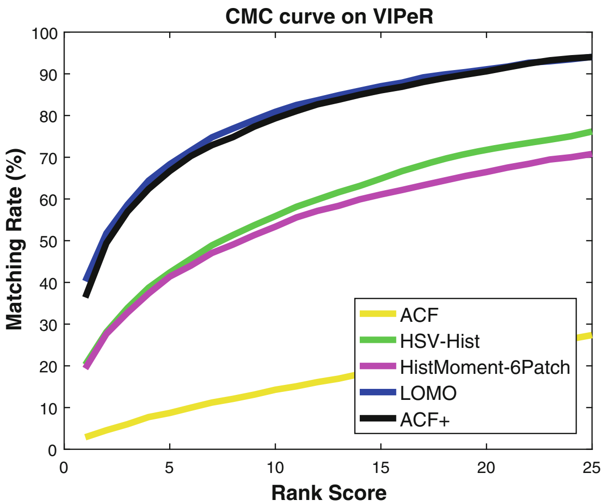


Fig. 1. CMC curves of different features on VIPeR

Next we compare the feature computation time of different features. Table 2 summarizes the feature dimension and frame rate on 128×48 size images. As we can see from the results, ACF is the fastest due to the simplicity of channel feature computation. Hist-HSV is slower than ACF but the discriminative capability is excellent, so we integrate it into the ACF+. Moreover, ACF+ can share some computation with ACF detector and achieves frame rate of 20 fps, which can be used for real-time applications. Note that all these experiments are conducted on laptop with Intel core i5 CPU. The code are written in MATLAB and thus speed can be further improved by optimizing the implementation.

Table 2. Feature computation time on 128×48 size images.

Feature type	ACF	Hist-HSV	HistMoment-6Path	LOMO	ACF+
Dimension	3840	9216	6966	26960	21642
FPS	131	25	<1	17	20

4.3 Evaluation of the Hybrid Distance Metric Learning

In this section, we evaluate the proposed hybrid distance metric learning method. Our *baseline* method is built on XQDA [11] metric learning method using ACF+ as feature representation. Three ingredients of ACF+, namely ACF, HSV-Hist and SILTP are used to learn different distance metrics using XQDA, and then get a jointed distance metric. To evaluate the performance of the proposed hybrid distance metric learning method, we compared different ensemble methods. Table 3 describes the details of the compared methods in our experiments.

$$PUR = \frac{log\left(G\right) - \sum_{r=1}^{G} \boldsymbol{M}\left(r\right) log\left(\boldsymbol{M}\left(r\right)\right)}{log\left(G\right)} \tag{8}$$

The Proportion of Uncertainty Removed (PUR) score is defined in (8), where G is the size of the gallery set, $\boldsymbol{M}\left(r\right)$ is the accumulated match characteristic

Table 3. The compared methods

Method	Description
Baseline	ACF+ with XQDA without hybrid distance metric
MIN	Take the minimum value among different distance metric
MAX	Take the maximum value among different distance metric
MEAN	Take the mean value among different distance metric
PUR-weight	The weight factors are computed by PUR score (Eq. 8)
FIT-weight	The weight factors are learned from training set

matrix at rank $r(r \leq 20)$. $r = 1$ is preferred in the literatures and also used in our experiments.

As PUR has been widely used in Re-ID literature as an performance indicator because of its invariant quality for the size of the gallery set, we use it to rank the different distance metrics and compute the weight factors. The weight factors are computed as following: $\omega_i = \frac{PUR_i}{PUR_1 + \cdots + PUR_i + \cdots PUR_n}, i \in (1, 2, \cdots n)$, where ω_i is the weight of distance metric i. The larger score of PUR, the better accuracy of the distance metric. Thus the weight of distance metric ω_i should be counted more, if its PUR score is larger.

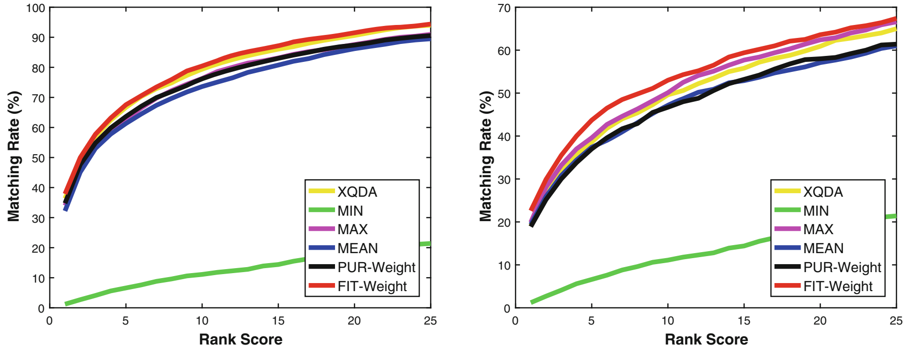
Table 4. CMC scores on VIPeR (%)

Method	Rank1	Rank5	Rank10	Rank15	Rank20
XQDA	36.39	66.68	79.34	86.04	90.60
MIN	2.53	11.30	19.91	26.30	32.53
MAX	34.02	62.66	76.30	83.13	87.59
MEAN	32.22	61.33	73.61	80.76	86.20
PUR-weight	34.78	63.64	76.23	83.01	87.41
FIT-weight	**37.88**	**67.56**	**80.44**	**87.22**	**91.42**

Table 5. CMC scores on PRID2011(%)

Method	Rank1	Rank5	Rank10	Rank15	Rank20
XQDA	19.00	38.60	49.60	55.80	60.90
MIN	1.20	6.60	11.10	14.40	18.50
MAX	19.90	39.60	50.10	57.70	62.40
MEAN	19.40	37.40	47.20	52.90	57.10
PUR-weight	18.90	37.00	46.70	53.20	58.00
FIT-weight	**22.60**	**43.70**	**53.00**	**59.40**	**63.60**

We conduct experiments on two public datasets. Tables 4 and 5 show the results on VIPeR and PRID2011 respectively. Figure 2 depicts the CMC curves on VIPeR and PRID2011. From the results, we can see that FIT-weight achieves the best performance on both datasets. Especially on PRID2011, it outperforms the baseline by a large margin (up to 5 percentage points). Among all these methods, PUR-weight obtains second best results. However, it still gets lower accuracy than the baseline. Other methods obtained worse results than baseline, showing that simple ensemble methods fail to improve the ACF+. This further verified the efficacy of the proposed hybrid distance metric learning method, which learns optimal weight factors from the training data.



(a)VIPeR (b)PRID2011

Fig. 2. CMC curves on different datasets.

5 Conclusion

In this work, we proposed a novel feature representation called ACF+ for pedestrian re-identification. ACF+ can share feature computation with ACF detector, and perform pedestrian detection and re-identification in real-time. To further boost the accuracy of the ACF+, we proposed a hybrid distance metric learning method which learns optimal weight factors for different metrics and outperforms the simple feature concatenation of ACF+. In the future work, we would like to explore the proposed method with other metric learning algorithms and further improve the computation time.

Acknowledgments. This work is supported by National Natural Science Foundation of China (project no. 6160011396).

References

1. Appel, R., Belongie, S., Perona, P., Doll, P.: Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014)
2. Chen, S.Z., Guo, C.C., Lai, J.H.: Deep ranking for person re-identification via joint representation learning. IEEE Trans. Image Process. **25**(5), 2353–2367 (2016)
3. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: International Conference on Pattern Recognition (2014)
4. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
5. Xiong, F., Gou, M., Camps, O., Sznaier, M.: Person re-identification using kernel-based metric learning methods. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 1–16. Springer, Cham (2014). doi:10.1007/978-3-319-10584-0_1

6. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21227-7_9
7. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7577, pp. 780–793. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33783-3_56
8. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
9. Li, W., Wang, X.: Locally aligned feature transforms across views. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
10. Liao, S., Zhao, G., Kellokumpu, V., Pietikainen, M., Li, S.Z.: Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
11. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition (2015)
12. Liao, S., Li, S.Z.: Efficient PSD constrained asymmetric metric learning for person re-identification. In: International Conference on Computer Vision (2015)
13. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Proceedings of IEEE International Workshop on PETS (2007)
14. Xu, Y., Lin, L., Zheng, W.S., Liu, X.: Human re-identification by matching compositional template with cluster sampling. In: International Conference on Computer Vision (2013)
15. Yang, Y., Liao, S., Lei, Z., Li, S.Z.: Large scale similarity learning using similar pairs for person verification. In: AAAI (2016)
16. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)