INTEGRATED METRIC LEARNING WITH ADAPTIVE CONSTRAINTS FOR PERSON RE-IDENTIFICATION

Lei Hao, Wenbin Yao, Chao Pei, Yuesheng Zhu

Communication and Information Security Lab, Institute of Big Data Technologies,
Shenzhen Graduate School, Peking University

ABSTRACT

Person re-identification is an important technique to search a probe person against a set of gallery persons and metric learning methods have shown their effectiveness in matching person images. In this paper, an Integrated Metric Learning with Adaptive Constraints (IMLAC) method is proposed to promote the performance for person re-identification. In the method, the *difference* and *commonness* of an image pair are combined to define a novel integrated metric. Considering the complex variations of pedestrian images, a rule of adaptive pairwise constraints is extended for the integrated metric to further enhance separation and reunion between image pairs. Extensive experiments conducted on three person re-identification datasets including VIPeR, PRID450S and GRID indicate that the proposed method outperforms the state-of-the-art methods.

***Index Terms*—** Person re-identification, metric learning, adaptive constraints

1. INTRODUCTION

Person re-identification is a hot and challenging task, which is of great value in both vision research and video surveillance applications. This task is extremely challenging because there are large variations in person images caused by viewpoints, illuminations, human poses and background. In the past years, many methods have been proposed [1] for person re-identification.

Among existing methods, metric learning methods [2, 3, 4, 5, 6, 7, 8] have made impressive improvements and promoted the person re-identification research. Martin *et al*. [2] derived metric learning method from equivalence constraints (KISSME) by computing the difference between the intra-class and inter-class covariance matrix. As an improvement, Tao *et al*. [3] enhanced KISSME by regularizing the two covariance matrices to reduce overestimation of large eigenvalues of the two estimated covariance matrices. Liao *et al*. [4] suggested preserving with a positive semidefinite (PSD) constraint and asymmetric sample weighting strategy. Recently, Yang *et al*. [5] first put forward the concepts of the *difference* and *commonness* of an image pair, which are defined as the subtraction and summation between the pair feature vectors respectively in Eq.(1). Notably, the discriminativeness of the *commonness* is not exploited in most metric learning methods until Yang *et al*. [5] considered both the *difference* and *commonness* of image pairs to learn a similarity metric with two separated matrices and obtained excellent results for person verification. However, the learned similarity metric could be weakened by losing collaboration between the *difference* and *commonness* because of the separated matrices. Considering above, we propose a novel integrated metric based on commonly used Mahalanobis distance metric combining the *difference* and *commonness* by learning one matrix.

Pairwise constraints are widely used to learn a distance metric in person re-identification since the only known information is whether a pair of pedestrian images is matched or not. However, pairwise constrained metric learning methods [6, 7, 8] usually requires the distance of a positive pair lower than a fixed upper bound and meanwhile pushed the distance of a negative pair higher than a fixed low bound. Li *et al*. [7] learned a decision function for verification and both the lower and upper bounds are fixed. Yao *et al*. [8] limited the learned distance of negative pair to lower than one. But the complex variations of pedestrians could result in learning less effective metric under the bound-fixed constraints. To better handle this, we extend the locally adaptive constraints [9] to further enhance the separation and reunion between image pairs for the metric. Moreover, regularizing penalty terms are crucial in re-identification to avoid over-fitting when the training sets are small or medium-sized [10]. Regarding this, the diversity regularizer proposed in [11] is introduced to avoid over-fitting as well as promoting low-rankness of the target metric.

In this paper, we propose an Integrated Metric Learning with Adaptive Constraints (IMLAC) method for person re-identification, where both the *difference* and *commonness* of an image pair are exploited by learning one matrix to increase its discriminativeness. It's shown that the *commonness* improved the metric in matching person images. Meanwhile, the extended rule to the method can adaptively determine pairwise constraints, thus can effectively distinguishes between positive pairs and negative pairs in training with complexly distributed data. In addition, the introduced diversity regularizer is proven to be effective in distance measuring with a few

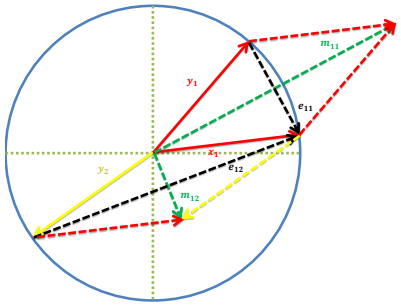


Fig. 1. An illustration of the *difference* e and *commonness* m in 2-dimensional Euclidean space. The pair (x_1, y_1) and (x_1, y_2) denote the normalized feature vectors of a positive pair and a negative pair respectively. The circle is a unit one.

numbers of latent factors.

2. INTEGRATED METRIC LEARNING WITH ADAPTIVE CONSTRAINTS

2.1. Combination of the *difference* and *commonness*

Suppose we have an image pair (x, y), where x,y are d-dimensional feature vectors and normalized with the l_2 -norm. According to [5], the *difference* e and *commonness* m are defined as:

$$\begin{cases} e = x - y \\ m = x + y \end{cases} \tag{1}$$

From illustration by Fig.1, (x_1, y_1) denotes a positive pair and (x_1, y_2) denotes a negative pair. It can be clearly observed that for a positive pair, the value of $\|e_{11}\|_2$ is small but $\|m_{11}\|_2$ is big, while the value of $\|e_{12}\|_2$ is big but that of $\|m_{12}\|_2$ is small for a negative pair. Therefore, both the m and e of an image pair contribute to the discriminativeness. To our knowledge, most distance metric learning methods only considered the *difference* of the pair to learn a Mahalanobis metric:

$$D_M(x, y) = (x - y)^T M(x - y) = e^T M e \tag{2}$$

where $M \succeq 0, M \in R^{d \times d}$. The learned matrix M does not exploit the discriminative ability of the *commonness* of the pair. Considering this, we expect more discriminative information by forming the metric with both e and m. Accordingly, we propose to learn an integrated metric by one matrix based on commonly used Mahalanobis distance function [12] as:

$$D_M(x, y) = e^T M e - \lambda m^T M m \tag{3}$$

where λ is an empirical parameter and used to balance the effect of the *difference* and *commonness*. By Re-parametrizing M with $M = W^T W$ [13], $D_W(x, y) = \|Wx - Wy\|_2^2 - \lambda\|Wx + Wy\|_2^2$, in which $W \in R^{k \times d}$ $(k \ll d)$. Note that the D_W does not meet the strict definition of distance in mathematics and we just treat it as a numerical measure.

2.2. Locally Adaptive Constraints

Pairwise constraints are pervasive in metric learning for person re-identification task. In practice, conventional distance metric based on pairwise constraints [7] usually requires fixed bounds as the Eq.(4) shows:

$$\begin{array}{ll} D_W(x, y) \le u & (x, y) \in P \\ D_W(x, y) \ge l & (x, y) \in N \end{array} \tag{4}$$

where P is a set of positive pairs belonging to the same pedestrians and N is a set of negative pairs belonging to the different pedestrians. Additionally, u and l are constants for all different pairs.

However, the above bound-fixed constraints fail to learn effective metric from data with complex distributions discussed in [9]. To handle this, we extend the rule in [9] to compute the adaptive bounds by replacing the fixed u and l with a local adaptive function $f(D_{xy})$. Given a pair (x, y), the D_{xy} is defined as $D_{xy} = \|e\|_2^2 - \lambda\|m\|_2^2$ to guide the changes in original space. One similar principle we can obviously obey is: the larger D_{xy} of a positive pair is, the more $f(D_{xy})$ shrinks, while the smaller D_{xy} of a negative pair is, the more $f(D_{xy})$ expands. Hence, the adaptive constraints for each pair can be designed as:

$$\begin{array}{l} u \leftarrow f_p(D_{xy}) = D_{xy} - (D_s)^\alpha / (D_{max} - D_{min}) \\ l \leftarrow f_n(D_{xy}) = D_{xy} + (D_{max} - D_{min}) / (D_s)^\alpha \end{array} \tag{5}$$

where D_{max} is the maximal D_{xy} and D_{min} is the minimal D_{xy} among all pairs used in the training set. In this paper, we set $D_s = D_{xy} - D_{min} + 1$ and $\alpha = \log(D_{\max} - D_{\min})$.

The D_{xy} defined in our work can take the discriminativeness of both the *difference* and *commonness* to guide the shrinkage and expansion. The setting of α ensures the rapid shrinkage and expansion depending on intrinsic structure information of the data by the use of D_{max} and D_{min}. Therefore, it can help distinguish between positive pairs and negative pairs effectively.

2.3. Integrated Metric Learning

The study in [11] has shown that the diversity regularizer can achieve good performance in retrieval and clustering task. Given k latent factors in $W \in R^{k \times d}$, the diversity regularizer $\Omega(W)$ proposed in [11] indicates how diverse the factors are and it reaches the global maximum when the factors are orthogonal to each other.

To diversify the latent factors of the matrix W, we define the Integrated Metric Learning with Adaptive Constraints (IMLAC) method as:

$$\begin{array}{l} \min\limits_{W} -\phi\Omega(W) \\ s.t.\ D_W(x, y) \le f_p(D_{xy}) \quad (x, y) \in P \\ \qquad D_W(x, y) \ge f_n(D_{xy}) \quad (x, y) \in N \end{array} \tag{6}$$

where $\phi \geq 0$ is a tradeoff parameter between the distance loss and the diversity regularizer. The latent factor k of W and ϕ needs to be properly chosen to achieve the optimal balance.

2.4. Optimization

In this section, we present an algorithm to solve the problem defined in Eq.(6). A strategy similar to [13] is used to remove the constraints and a hinge loss function is adopted. Then Eq.(6) can be reformulated as:

$$\begin{aligned}\min_{W} \tfrac{1}{|P|} \sum_{(x,y)\in P} \max\left(0, D_W(x,\mathrm{y}) - f_p(D_{xy})\right) \\ + \tfrac{1}{|N|} \sum_{(x,y)\in N} \max(0, f_n(D_{xy}) - D_W(x,\mathrm{y})) - \phi\Omega(W)\end{aligned} \tag{7}$$

For the ease of optimization, let $W = diag(g)\widetilde{W}$, where g is a vector and g_i denotes the l_2-norm of the i-th row of W, and the l_2-norm of each row vector in $\widetilde{W}$ is 1. According to the definition of $\Omega(W)$, we have $\Omega(W) = \Omega(\widetilde{W})$. A smooth and convex low bound $\Gamma(\widetilde{W})$ of $\Omega(\widetilde{W})$ proposed in [11] is applied here since $\Omega(\widetilde{W})$ is hard to optimize directly. It has been proven that maximizing the $\Gamma(\widetilde{W})$ can increase $\Omega(\widetilde{W})$. Please refer to [11] for details.

Therefore, we instead optimize the $\Gamma(\widetilde{W})$ and the problem over $\Omega(W)$ can be further reformulated as:

$$\begin{aligned}&\min_{\widetilde{W}} \tfrac{1}{|P|} \sum_{(x,y)\in P} \max\left(0, D_{\widetilde{W}}(x,\mathrm{y}) - f_p(D_{xy})\right) \\ &\quad + \tfrac{1}{|N|} \sum_{(x,y)\in N} \max\left(0, f_n(D_{xy}) - D_{\widetilde{W}}(x,\mathrm{y})\right) - \phi\Gamma(\widetilde{W}) \\ &s.t. \ \left\|\widetilde{W}_i\right\| = 1, \ \forall i = 1, ..., k\end{aligned} \tag{8}$$

where $\widetilde{W}_i$ denotes the i-th row of $\widetilde{W}$ and $D_{\widetilde{W}}(x,\mathrm{y}) = \left\|diag(g)\widetilde{W}(x-y)\right\|_2^2 - \lambda\left\|diag(g)\widetilde{W}(x+y)\right\|_2^2$. This problem can be solved by alternatively optimizing g and $\widetilde{W}$ with projected subgradient method: optimizing g with $\widetilde{W}$ fixed and optimizing $\widetilde{W}$ with g fixed.

3. EXPRIMENT

3.1. Datasets and Setup

We evaluate the proposed method on three widely used person re-identification datasets: VIPeR [14], PRID450S [15] and GRID [16]. VIPeR is a widely used person re-identification dataset for benchmark evaluation, which contains 632 person image pairs captured at outdoor by two different camera views. PRID 450S consists of 450 person image pairs from two different surveillance cameras. GRID contains 250 image pairs captured on underground station from 8 disjoint camera views. Besides, it includes additional 775 images that do not belong to the 250 image pairs, which makes the match process much misleading.

We follow the widely adopted experimental protocol of single shot settings commonly used in [4, 17]. To be specific, we randomly divide each dataset into half for training and the other half for testing. With one exception, we add additional 775 images into the gallery set for GRID dataset. The average performance is reported by Cumulative Matching Characteristic (CMC) according to 10 repeated times of evaluation procedure.

3.2. Feature Representation

Local Maximal Occurrence (LOMO) feature proposed in [12] is used for baseline evaluation in our work. To test the ability of our method to fuse features, we also consider another Convolutional Neural Network (CNN) based image feature proposed in [24]. The fusion feature of LOMO and CNN results in 31056 dimensions. For all the experiments, PCA is first applied to reduce dimensions of the feature but all energies are reserved similar to [4].

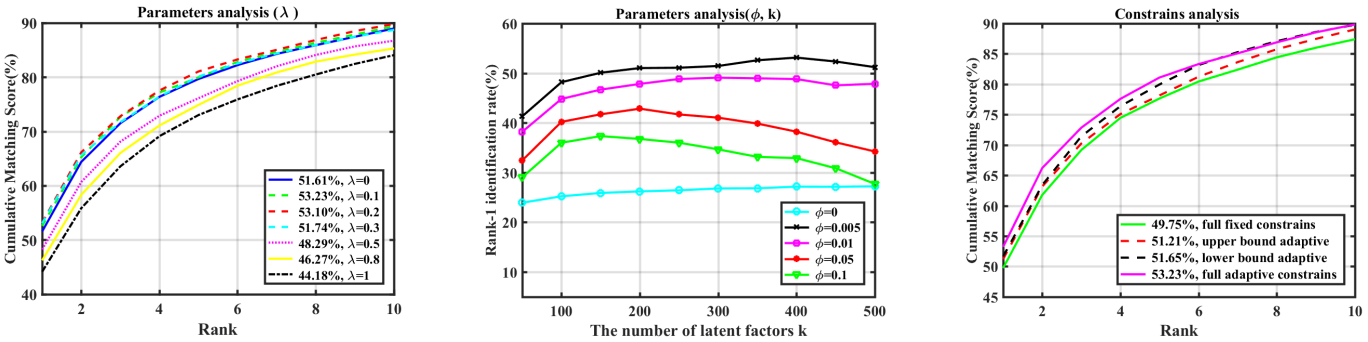
3.3. Parameters analysis

There are three important parameters in the method. The parameter λ plays an important role in balancing the *difference* and *commonness* in Eq.(2). The ϕ and the number of latent factors k are tradeoff parameters in the regularizer in Eq.(6). We conduct experiments on VIPeR with the fusion feature to analyze the parameters. One obvious principle we follow is that, when we test a parameter, the others will be fixed at their best value.

λ analysis: It can be seen from Fig.2(a) that our method performs best when $\lambda = 0.1$. Also, when λ ranges from 0 to 0.3, the *commonness* takes positive effect in matching person images compared with that of $\lambda = 0$ which only considers the *difference*. It can be concluded that the *commonness* mo-

Table 1. CMC results on VIPeR dataset

Method/Rank	1	10	20
ITML(LOMO)[6]	24.65	63.04	78.39
KISSME(LOMO)[2]	34.81	77.22	86.71
KLFDA(LOMO)[18]	38.58	80.44	89.15
XQDA(LOMO)[12]	40.00	80.51	91.05
MLAPG(LOMO)[4]	40.73	82.34	92.37
DRML(LOMO)[8]	40.19	78.80	88.61
LSSL(LOMO)[5]	36.46	82.75	**92.73**
IMLAC(LOMO)	**42.53**	**83.45**	92.12
MED_VL[19]	41.1	83.2	91.7
LSSCDL[20]	42.66	87.28	94.84
GOG[21]	49.7	88.7	94.5
MCPB-CNN[22]	47.8	84.8	91.1
TMA[23]	48.19	87.65	93.54
IMLAC(Fusion)	**53.23**	**88.75**	**95.63**



(a) λ analysis (b) ϕ and k analysis (c) Constrains analysis

Fig. 2. Parameters analysis on VIPeR.

Table 2. CMC results on PRID 450S dataset

Method/Rank	1	10	20
XQDA(LOMO)[12]	61.4	**91.0**	**95.3**
MLAPG(LOMO)[4]	58.18	88.80	94.67
LSSL(LOMO)[5]	51.20	87.78	94.32
IMLAC((LOMO)	**63.02**	89.07	94.31
Shen[25]	44.4	82.2	89.8
MED_VL[19]	45.9	82.9	91.1
LSSCDL[20]	60.49	88.58	94.84
GOG[21]	68.4	94.5	93.60
TMA[23]	54.22	83.11	90.22
IMLAC((Fusion)	**73.82**	**95.07**	**97.64**

Table 3. CMC results on GRID dataset

Method/Rank	1	10	20
KISSME(LOMO)[2]	10.64	31.60	43.20
XQDA(LOMO)[12]	16.56	41.44	52.48 R
MLAPG(LOMO)[4]	16.64	41.20	52.96
LSSL(LOMO)[5]	14.40	40.16	52.08
IMLAC((LOMO)	**19.84**	**45.28**	**56.64**
PolyMap[17]	16.3	46.0	57.6
LSSCDL[20]	22.40	51.28	61.20
GOG[21]	24.7	58.4	69.0
SSDAL[26]	22.4	48.0	58.4
IMLAC((Fusion)	**33.36**	**60.64**	**70.56**

tivates the *difference* to get more discriminativeness to some degree.

ϕ and k analysis: Fig.2(b) illustrates how rank-1 accuracy varies with different ϕ as k increases. It is shown that a lager ϕ makes the metric more diverse but loses the effectiveness in measuring. The larger k could make the metric get more information for distinguishing pairs, but too large k could lead to data over-fitting.

In general, the above parameters can be properly set to achieve the optimal results. What's more, Fig.2(c) shows how adaptive constraints improve the results compared with the fixed upper and low bounds (The fixed bound are set as the average of all adaptive bounds). Better result can be obtained when full adaptive constraints are taken.

3.4. Comparison with the state-of-the-art methods

We first compared our method with state-of-the-art metric learning on the same features LOMO on the three datasets and then with state-of-the-art published results with the fusion features. As shown in Table 1 on VIPeR, our metric learning method outperforms all the other state-of-the-art metric learning methods on Rank-1. With a feature fusion, our method achieves comparative performance compared with the state-of-the-art results. Furthermore, in Table 2 and Table 3, we also get improving results on PRID 450S and GRID datasets. The reasonable explanations of outperforming results are: the proposed metric learning method gets more discriminative information combining the *difference* and *commonness*, and the extended adaptive constraints improve the ability of distinguishing between positive pairs and negative pairs during training. Additionally, the diversity regularizer captures more unique information individually by the uncorrelated latent factors.

4. CONCLUSION

In this paper, we propose a novel metric learning method which enhances discriminating power to match images for person re-identification. The *difference* and *commonness* of an image pair are combined to obtain more ability of discriminativeness. Also, the extended adaptive constraints and diversity regularizer show their effectiveness in the task. Extensive experiments on three datasets indicate that our method achieves the state-of-the-art performance.

5. ACKNOWLEDGEMENT

This work is supported by the Shenzhen Municipal Development and Reform Commission (Disciplinary Development Program for Data Science and Intelligent Computing), and the Shenzhen Engineering Laboratory of Broadband Wireless Network Security.

6. REFERENCES

[1] Liang Zheng, Yi Yang, and Alexander G Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016.

[2] Martin Köstinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *CVPR*, 2012, pp. 2288–2295.

[3] Dapeng Tao, Yanan Guo, Mingli Song, Yaotang Li, Zhengtao Yu, and Yuan Yan Tang, "Person re-identification by dual-regularized kiss metric learning," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2726–2738, 2016.

[4] Shengcai Liao and Stan Z Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in *ICCV*, 2015, pp. 3685–3693.

[5] Yang Yang, Shengcai Liao, Zhen Lei, and Stan Z Li, "Large scale similarity learning using similar pairs for person verification," in *AAAI*, 2016, pp. 3655–3661.

[6] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon, "Information-theoretic metric learning," in *ICML*, 2007, pp. 209–216.

[7] Zhen Li, Shiyu Chang, Feng Liang, Thomas S. Huang, Liangliang Cao, and John R. Smith, "Learning locally-adaptive decision functions for person verification," in *CVPR*, 2013, pp. 3610–3617.

[8] Wenbin Yao, Zhenyu Weng, and Yuesheng Zhu, "Diversity regularized metric learning for person re-identification," in *ICIP*, 2016, pp. 4264–4268.

[9] Qilong Wang, Wangmeng Zuo, Lei Zhang, and Peihua Li, "Shrinkage expansion adaptive metric learning," in *ECCV*, 2014, pp. 456–471.

[10] Cijo Jose and François Fleuret, "Scalable metric learning via weighted approximate rank component analysis," in *ECCV*, 2016, pp. 875–890.

[11] Pengtao Xie, Yuntian Deng, and Eric Xing, "Diversifying restricted boltzmann machine for document modeling," in *SIGKDD*, 2015, pp. 1315–1324.

[12] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015, pp. 2197–2206.

[13] Kilian Q Weinberger and Lawrence K Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. Feb, pp. 207–244, 2009.

[14] Douglas Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *PETSW*, 2007, vol. 3.

[15] Peter M Roth, Martin Hirzer, Martin Koestinger, Csaba Beleznai, and Horst Bischof, "Mahalanobis distance learning for person re-identification," in *Person Re-Identification*, pp. 247–267. 2014.

[16] Chen Change Loy, Tao Xiang, and Shaogang Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *International Journal of Computer Vision*, vol. 90, pp. 106–129, 2010.

[17] Dapeng Chen, Zejian Yuan, Gang Hua, Nanning Zheng, and Jingdong Wang, "Similarity learning on an explicit polynomial kernel feature map for person re-identification," in *CVPR*, 2015, pp. 1565–1573.

[18] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier, "Person re-identification using kernel-based metric learning methods," in *ECCV*, 2014, pp. 1–16.

[19] Yang Yang, Zhen Lei, Shifeng Zhang, Hailin Shi, and Stan Z Li, "Metric embedded discriminative vocabulary learning for high-level person representation," in *AAAI*, 2016, pp. 3648–3654.

[20] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan, "Sample-specific svm learning for person re-identification," in *CVPR*, 2016, pp. 1278–1287.

[21] Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato, "Hierarchical gaussian descriptor for person re-identification," in *CVPR*, 2016, pp. 1363–1372.

[22] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng, "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," in *CVPR*, 2016, pp. 1335–1344.

[23] Niki Martinel, Abir Das, Christian Micheloni, and Amit K Roy-Chowdhury, "Temporal model adaptation for person re-identification," in *ECCV*, 2016, pp. 858–877.

[24] Tetsu Matsukawa and Einoshin Suzuki, "Person re-identification using cnn features learned from combination of attributes," in *ICPR*, 2016, pp. 2429–2434.

[25] Yang Shen, Weiyao Lin, Junchi Yan, Mingliang Xu, Jianxin Wu, and Jingdong Wang, "Person re-identification with correspondence structure learning," in *ICCV*, 2015, pp. 3200–3208.

[26] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian, "Deep attributes driven multi-camera person re-identification," in *ECCV*, 2016, pp. 475–491.