Weighted-Diffusion Distance Measurement Ranking for Pedestrian Re-Identification

To cite this article: Hua Han et al 2018 IOP Conf. Ser.: Mater. Sci. Eng. 435 012039

View the article online for updates and enhancements.
Weighted-Diffusion Distance Measurement Ranking for Pedestrian Re-Identification

Hua Han¹⁺, Yujin Zhang¹ and Chunyuan Wang¹

¹School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, P. R. China

⁺2070967@mail.dhu.edu.cn

Abstract. Pedestrian re-identification (Re-ID) is a very difficult problem due to complex and changeable environment. Although we can extract discriminative features, simple metric algorithms tend to yield their unreliable matching results. In order to achieve higher matching rate, we propose a more reliable histogram comparison method called weighted-diffusion distance. When calculating the cumulative difference between spatial vector histograms on the basis of high-dimensional diffusion distance, lacking the representation of differences between internal elements results in less desired accuracy. Thus, greater weight should be given to those of the corresponding element values that are same or similar. After these operations, the proposed weighted-diffusion distance is computed. Then we use it to complete multi-feature fusion, and establish a similarity measurement function, which is used to measure the similarity between two pedestrians. At last, pedestrian Re-ID experiments with different cameras are performed and the results show that the proposed method can achieve much better results than some known methods.

1. Introduction
Pedestrian re-identification (Re-ID) is an important issue in multi-camera tracking applications [1-2]. Because of highly complex environment characterized by such as different illuminations, pedestrian gaits, and scales occlusion, it is usually very difficult to achieve excellent Re-ID matching rate. The fundamental question is whether a pedestrian detected in different views or at different time instants can be linked to a same individual [3]. The past decade has witnessed many outstanding solutions in this field. They can be divided into two classes: feature and metric ones.

Feature-class methods, from the appearance of targets (pedestrians in this paper), select or design a discriminating feature descriptor, which has a certain intra-class invariance, and strong inter-class distinction. The features used in pedestrian Re-ID are mainly color, structure, shape, gradient, texture, key point, and area descriptor. Before feature extraction, a pre-processing process, such as foreground segmentation and human body parts detection, can be carried out. The different features extracted can usually be cascaded or combined to form a more discriminative feature. Symmetry-Driven Accumulation of Local Features (SDALF) [4] divides a human body into head, torso and legs after extracting the foreground, and uses the weighted HSV (Hue, Saturation and Value) histogram, maximally stable color region and recurrent highly-structured patches to re-identify a person. Custom Pictorial Structures (CPS) [5] complete Re-ID after body parts detection by using a human body structure. Histogram Plus Epitome (HPE) [6] uses the appearance model that is similar to that in [4]. It is proposed to condense a set of frames of the same individual into a highly informative signature.
Hybrid Spatiogram and Covariance Descriptor (HSCD) [7] combines Multi-Channel based Spatio-Histogram (MCSH) and Multi-Statistics on Pyramid of Covariance (MSPC) to complete pedestrian Re-ID. Encoded Local Descriptors Fisher Vectors (eLDFV) [8] uses a Fisher vector to encode luminance and gradient information of an image and combines it with SDALF. Structural Constraints Enhanced Feature Accumulation (SCEFA) [9] uses a hierarchical weighted HSV histogram, bidirectional matching color region feature, and Gabor texture patterns to describe the key points. Effectiveness Salience and Dense Correspondence (eSDC) [10] searches distinct individual features by using unsupervised learning methods and combines it with SDALF. Mean Riemannian Covariance Grid (MRCG) [11] uses an appearance model that relies on no body parts. It is an extension of Spatial Covariance Regions (SCR) [12] that is the covariance of a vector of eleven cues derived from equalized RGB colors. An MRCG descriptor is computed as a mean of gallery examples but only applicable to multi-shot Re-ID modalities.

Metric-class methods focus on distance measurement. The essence of metric learning is to estimate the scale or weight of each component when calculating the distance between feature vectors, and to select or highlight the strongly distinguishing features. Ensemble of Localized Features (ELF) [13] selects color and texture features via AdaBoost classifiers. Ensemble RankSVM (ERSVM) [14] integrates multiple eigenvector weights by using SVM classifiers based on relative distance ordering. In [15] a metric learning framework is used to obtain a robust Mahalanobis metric for Large Margin Nearest Neighbor with Rejection (LMNN-R). Probabilistic Relative Distance Comparison (PRDC) [16] introduces a novel comparison model. Its goal is to make the probability of a correct match pair with a smaller distance greater than that of an incorrect match pair. Local Fisher (LF) [17] carries out principal-component-analysis to reduce dimensionality of a high dimension feature, then performs dimensionality reduction mapping based on local Fisher discriminant analysis, and finally obtains very high Re-ID rate.

Compared with feature-class methods, metric-class methods have lower requirements for feature selection, and they can usually achieve better Re-ID effects. But their training time complexity and space size can be tremendous. There is an over-fitting phenomenon when training samples are insufficient. When a scene and library change, they need to be re-trained. The feature-class methods require manual search and design of relatively better features, and their performance depend on the used features. This work designs excellent features and descriptions under the idea of feature-class methods. Achieving the desired results requires us to compute weighted-diffusion distance quickly and accurately.

2. Measurement Ranking by Weighted-Diffusion Distance
Most of traditional metric algorithms for similarity measurement are optimized based on Euclidean or cosine distance. However, either distance has its shortcoming. The former cannot be restored from a high-dimensional eigenvector to a low-dimensional geometric mechanism. It is sensitive to the deformation of an image, and lacks any spatial information between pixels. The cosine distance has some disadvantages due to the large difference in position determination and insensitivity to numerical values, which can result in many errors. Diffusion distance [18], to be defined next, does not bear any mentioned shortcomings. It is different from the case where the similarity is determined by Euclidean distance. The similarity based on diffusion distance depends on the speed of diffusion and number of diffusion paths. It appears to be superior in matching pedestrians. Its principle can refer to [19]. Assume that there are two $m$-dimensional distributions $h_i(x)$ and $h_j(x)$, and $x$ is an $m$-dimensional real matrix. High-dimensional diffusion distance can be defined as follows:

$$K(h_i, h_j) = \sum_{l=0}^{K} k|d_l(x)|$$

Where

$$d_l(x) = h_l(x) - h_l(x) , \quad l = 0$$
\[ d_i(x) = \left[ d_{i,j}(x) * \phi(x, \sigma) \right] _{2^{(L-i)}} , \quad (i=1, \cdots, L) \]  
\[ \phi(x, t) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{x^2}{2t}} \]  
(3) \hspace{1cm} (4)

In (3), \( \downarrow \) denotes downsampling. If the sampling value is too small, some useful information may be missing; If the sampling value is too large, it reduces computational efficiency. Based on the above considerations, the size of this sample is set to the empirical value \( \downarrow_{2^3} \), that is, two thirds of the sample value. \( \hat{L} \) is the number of Gaussian Pyramid layers and \( \sigma \) is the standard deviation in the Gauss function \( \phi(x, \sigma) \). Using \( \hat{L} \), norm, (1) can be simplified to:

\[ K(h_i,h_j) = \sum_{i=1}^{\hat{L}} |d_i(x)| \]  
(5)

2.1. Weighted-Diffusion Distance

From (3), the amount of computation of downsampling in \( d_i \) exponentially reduces; Secondly, a Gaussian filter convolution operation is a linear one, which has obvious advantages in terms of computational efficiency. Despite this, from (1)-(2), when representing a cumulative difference between spatial vector histograms \( h_i(x) \) and \( h_j(x) \), the optimized diffusion distance misses the differences between internal individual elements. Although (5) can be applied directly to the similarity measurement, it ignores the influence between the corresponding elements of a spatial vector on similarity, thus resulting in lower accuracy. Since each element in a different histogram represents a different characteristic attribute, for example, the \( z \)-th element in a wHSV histogram represents the total number of pixels in the \( z \)-th region or represents the color value in the image. Therefore, if similarity measurement is carried out, it is necessary to consider not only the effects of cumulative differences, but also the effects of similarity between different elements. Assume that \( h_i(x) = [3,7,10] \), \( h_j(x) = [1,1,5] \), and \( h_k(x) = [3,6,2] \). To simplify the calculation, let \( \phi(x, \sigma) = 1 \). Without using downsampling, we have \( K(h_i, h_j) = \sqrt{55} = K(h_i, h_k) \). It is not possible to distinguish which two histograms are more similar. Through observation we can find that \( h_i(x) \) and \( h_j(x) \) have more of the same or similar elements than \( h_i(x) \) and \( h_k(x) \). Intuitively, \( h_i(x) \) and \( h_j(x) \) should be more similar than that \( h_i(x) \) and \( h_k(x) \) are.

Through the above analysis, in the calculation of the distance between two space vector histograms, we should not give the same weight to all internal elements, but greater weight to those same or similar corresponding elements. Therefore, in this work, we re-optimize the diffusion distance based on the former one optimized by Gaussian pyramid, and propose a new distance concept called weighted-diffusion distance. Hence, (2) is rewritten as

\[ d_i(x) = \sqrt{\sum_{p=1}^{n} (h_{ip}(x) - h_{jp}(x))^2} \]  
(6)

Where \( ip \) and \( jp \) represent the \( p \)-th element in histograms \( i \) and \( j \), respectively. \( p \in \{0,1, \cdots, n\} \). Equations (2) and (6) represent similar meanings that the distance between two histogram differences is equivalent to the accumulation of the corresponding element differences in each pair of histograms. Then we rewrite (6) as:

\[ d_i(x) = \sqrt{\sum_{p=0}^{R} \frac{1}{R} (h_{ip}(x) - h_{jp}(x))^2} \]  
(7)

\[ r_p = e^{-|h_{ip} - h_{jp}|/\sigma} \]

Where \( R \) is a normalization factor, \( R = \sum_{p} r_p \), \( r_p \) is the weight, \( \sigma \) is the adjustment factor. According to (10) given the prior \( h_i(x) \rightarrow h_k(x) \), we have \( K(h_i, h_j) = 2.337 \) and \( K(h_i, h_k) = 1.025 \) via (5). Intuitively, \( h_i(x) \)
and $h_i(x)$ should be more similar than $h_i(x)$ and $h_i(x)$. It can be seen that the weighted-diffusion distance based on the quadratic optimization of the proposed diffusion distance enables the similarity measurement to exhibit better discriminability. Next we use the weighted-diffusion distance to measure and rank the distance among features. The greater distance indicates greater difference among individuals. On the contrary, the smaller distance, the smaller difference. Therefore, we can choose the foremost ranked candidate to re-identify him/her as the same individual.

2.2. Multi-Feature Fusion Based on Weighted-Diffusion Distance

In our work, the features we used include local RGB, local wHSV, Histogram of Oriented Gradient (HOG) [20], and global Maximally Stable Color Regions (MSCR) [21]. The similarity measurement function constructed by any of the above mentioned features alone fail to achieve desired matching effects. Therefore, we propose a multi-feature fusion method based on weighted-diffusion distance to complete pedestrian Re-ID. Extract the histograms of the above features, and then calculate the corresponding weighted-diffusion distance. In order to prevent big difference in the weighted-diffusion distance calculated from different features. The calculated weighted-diffusion distance of each feature will be normalized. In pedestrian Re-ID, there are usually two sets of images: target $I_A$ to be re-identified and candidate target $I_B$.

We use $I_A$ and $I_B$ to denote two image sets, calculate the weighted-diffusion distance based on the above mentioned features and normalize them. Use (5) and (7) to calculate the matching distance $d(I_A, I_B)$ between the pedestrian targets of $I_A$ and $I_B$:

$$d(I_A, I_B) = \alpha \cdot d_{MSCR}(I_A, I_B) + \beta \cdot d_{RGB}(I_A, I_B) + \gamma \cdot d_{wHSV}(I_A, I_B) + \lambda \cdot d_{HOG}(I_A, I_B)$$

(8)

Where $\alpha$, $\beta$, $\gamma$, and $\lambda$ represent the weights of the different feature histograms, respectively, satisfying $\alpha + \beta + \gamma + \lambda = 1$. Rank all matching distances, and choose the candidate target with the smallest matching distance as the final re-identified one.

3. Experiments and Results

VIPeR is a challenging pedestrian Re-ID database that has been widely used for benchmark evaluation [2]. It contains 632 pairs of person images, captured by a pair of cameras in an outdoor environment. Some pairs contain large variations in background, illumination, and viewpoint. All are scaled to 128x48 pixels. A widely adopted experimental protocol is to randomly divide all pairs into one half for training and another half for test, and repeat the procedure several times to obtain average performance. Although our method has achieved a desired result, we have to do some specific numerical comparison experiments to test its performance. Gray et al. [22] propose a standard protocol to compare results by using cumulative match curve (CMC) by introducing the VIPeR dataset for Re-ID. Next, we use CMC to evaluate the performance of our method.

3.1. Statistical Evaluation of Experimental Results

In order to verify the performance of the proposed distance in pedestrian Re-ID, we have done many experiments based on VIPeR. In the experiment, 50 pairs of pedestrian targets are randomly selected. Five trials are conducted. Finally, an average CMC curve is given. The results are shown in Figure 1.

Figures 1(a)-(e) show the CMC curves generated after the first to fifth experiments, where the horizontal axis in Figure 1 represents a rank score, and the vertical axis is matching rate. The matching rate of the first point on the average curve can be up to 63.6%, the second one 72.4%, and the one 77.6%. The results are in line with our expectations. Yet we need to know how our methods performs in comparison with the existing algorithms. The comparison with the state-of-the-art is given next.

3.2. Comparison with the State-of-the-Art

We have done many comparison tests with the state-of-the-art methods. These experimental comparisons are done for the algorithms in both feature-class and metric-class.

4
Figure 1. CMC curves of experiments based on VIPeR. (a) CMC curve generated after the first experiment, (b) CMC curve generated after the second experiment, (c) CMC curve generated after the third experiment, (d) CMC curve generated after the fourth experiment, (e) CMC curve generated after the fifth experiment, and (f) The average CMC curve after 5 trials.

1) Comparison with feature-class algorithms

In this paper, we compare the proposed with HSCD [7], eSDC [10], SCEFA [9], eLDFV [8], CPS [5], and SDALF [4]. The experimental results can be found in Figure 2.

From Figure 2, the matching rate of SDALF is relatively low. After dividing the pedestrian body into three parts: head, torso, and leg, SDALF combines wHSV, MSCR and repetitive texture block to describe the pedestrian target. Finally, the Euclidean distance is used to measure the similarity. The advantage of this widely used algorithm is the better feature selection, but it uses the traditional Euclidean distance for the measurement part. This is the reason of its lower matching rate. CPS has the advantage of being able to detect various parts of a body according to the pedestrian body structure. It can extract the foreground features better. eLDFV, combined with SDALF, uses a Fisher vector to encode the brightness and gradient information of an image. Its matching rate at the first point in its CMC curve can reach 21.42%. But its metric algorithm needs to be improved. SCEFA uses key point detection based on Gabor texture and two-way color feature matching. The use of multiple features is
its biggest advantage. eSDC is also a combination of SDALF and unsupervised learning methods to find better features of an individual. But it lacks the consideration of how to improve the metric algorithm. HSCD combines two excellent statistical descriptions. They are spatial histogram and regional covariance. A variety of mutually complementary statistical vectors are extracted from several statistical areas. This method is the best among the other feature-class algorithms. Its matching rate at the first point can reach 29.32%. Nevertheless, the matching rate of the above algorithms is still lower than ours. Based on multi-feature fusion and diffusion distance similarity measurement, the first point matching rate of our method can reach 63.6%, which is much higher than the best one (29.32%) from all the others, the second one is 72.4%, and the third one is 77.6%, which all beat the state-of-the-art methods. This represents a very significant improvement in this field. Next, we compare the proposed method with algorithms of the metric-class.

2) Comparison with metric-class algorithms

Our approach well outperforms the feature-class algorithms. Now we compare it with the state-of-art metric-class algorithms. They are ELF [13], ERSVM [14], PRDC [16], LMNN-R [15], COS [23], and Histogram Intersection (HI) [24]. The experimental results can be found in Figure 3.

From Figure 3, the matching rate of ELF is relatively low. It uses AdaBoost classifiers to select color and texture features. This method makes full use of the selection of classifiers, but it does not improve greatly in terms of matching rate. ERSVM has a certain similarity to PRDC. They use SVM classifiers, which are sorted by relative distance, to integrate multiple weights of feature vectors. This method has obvious advantages in the measurement algorithm, but it needs improvement in terms of feature selection. LMNN-R also uses PRDC to do relative distance comparison. It uses an LMNN classifier. But, in the feature design, it needs to be improved. In the metric-class, the best performance algorithm is HI. It is a kind of kernel functions based on an implicit correspondence relation. It solves the classification problem of unordered, variable-length vectors. It is positive definite. Because of its advantages, the matching rate at the first point is as high as 47.38% in pedestrian Re-ID. Therefore, under the same conditions, the metric-class methods have some certain advantages in matching rate compared with the mentioned feature-class methods. From Figure 3, our algorithm is able to beat them in terms of matching rate.

![Figure 3. CMC curves of comparison with metric-class algorithms.](image)

In summary, our algorithm is clearly outstanding in comparison with both feature-class and metric-class methods. This is because of proper features' fusion and the advantage of the weighted-diffusion distance in similarity measurement.

4. Conclusion

This paper has presented a new method to measure the similarity among histograms of pedestrians. This resultant measurement is called weighted-diffusion distance. It can be used to perform histogram comparison well. In order to obtain higher matching rate, we assign different internal elements of the histograms with different weights in the defined high-dimensional diffusion distance. In order to obtain more discriminative features, we extract multiple features of pedestrians. Then we use the
proposed distance to do multi-feature fusion, and establish the similarity measurement function, which is in turn used to measure the similarity of two pedestrians. Experimental results show that the proposed method can achieve drastically better results than its six peers, thereby representing one of major breakthrough results in the field.

Acknowledgments
This work was supported in part by the National Nature Science Foundation of China (No. 61305014, 61473077, 61473078), China Scholarship Council (No. 201508310033), “Chen Guang” project supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (No. 13CG60).

References
[1] D. Figueira, M. Taiana, J. C. Nascimento, and A. Bernardino 2016 A Window-Based Classifier for Automatic Video-Based Reidentification IEEE Transactions on Systems, Man & Cybernetics: Systems, vol. 46, no. 12, pp. 1736-1747.
[2] D. Tao, L. Jin, Y. Wang, and X. Li, “Person reidentification by minimum classification error-based KISS metric learning,” IEEE Transactions on Cybernetics, vol. 45, no. 2, pp. 242-252, Feb, 2015.
[3] T. Zhao and R. Nevatia 2004 Tracking multiple humans in complex situations, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1208 -1221.
[4] N. Martinel and C. Micheloni 2012Re-identify people in wide area camera network, 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, USA, pp. 31-36.
[5] I. Kviatkovsky, A. Adam, and E. Rivlin 2013 Color invariants for person reidentification, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no.7, pp. 1622-1634.
[6] L. Kai, Z. Zhao, and A. Cai 2015 Datum-adaptive local metric learning for person re-identification, IEEE Signal Processing Letters, vol. 22, no. 9, pp. 1457-1461.
[7] Y. Cai, W. Chen, K. Huang, and T. Tan 2007 Continuously tracking objects across multiple widely separated cameras, 8th Asian Conference on Computer Vision, Tokyo, Japan, pp. 843-852.
[8] C. Motamed and O. Wallart 2007 A temporal fusion strategy for cross-camera data association, Pattern Recognition Letters, vol. 28, no.2, pp. 233-245.
[9] S. F. Tahir and A. Cavallaro 2014 Cost-effective features for reidentification in camera networks, IEEE Transactions on Circuits and System for Video Technology, vol. 24, no. 8, pp. 1362-1374.
[10] W.-S. Zheng, S. Gong, and T. Xiang 2013 Reidentification by relative distance comparison, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 3, pp. 653-668.
[11] N. Martinel and C. Micheloni 2012 Re-identify people in wide area camera network, IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), Providence, USA, pp. 31-36.
[12] I. Kviatkovsky, A. Adam, and E. Rivlin 2013 Color invariants for person reidentification, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1622-1634.
[13] B. Wang, G. Wang, K. L. Chan, and L. Wang 2014 Tracklet association with online target-specific metric learning, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, pp. 1234-1241.
[14] A. Li, L. Liu, S. Liu, S. Yan, and K. Wang 2015 Clothing attributes assisted person reidentification, IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 5, pp. 869-878.
[15] X. Wang, W.-S. Zheng, X. Li, and J. Zhang 2016 Cross-Scenario Transfer Person Reidentification, IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 8, pp. 1447-1460.
[16] X. Liu, H. Wang, J. Wang, and X. Ma 2017 Person re-identification by multiple instance metric learning with impostor rejection, *Pattern Recognition*, vol. 67, pp. 287-298.
[17] J. Lin 1991 Divergence measures based on the Shannon entropy, *IEEE Trans. on Information Theory*, vol. 37, no. 1, pp. 145-151.
[18] H. Ling and K. Okada 2006 Diffusion distance for histogram comparison, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, New York, pp. 246-253.
[19] Q. J. Wang and R. B. Zhang 2008 LPP-HOG: A new local image descriptor for fast human detection, *IEEE International Symposium on Knowledge Acquisition and Modeling Workshop*, Wuhan, China, pp. 640-643.
[20] M.-Y. Zeng, Z.-M. Wu, C. Tian, Y. Fu, and F.-R. Jie 2014 Fusing appearance statistical features for person re-identification, *Journal of Electronics & Information Tech.*, vol. 36, no. 8, pp. 1844-1851.
[21] A. Mignon and F. Jurie 2012 PCCA: A new approach for distance learning from sparse pairwise constraints,” *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, pp. 2666-2672.
[22] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, and H. Bischof 2012 Large scale metric learning from equivalence constraints, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, pp. 2288-2295.
[23] H. Zhao, et al. 2015 Single Image Fog Removal Based on Local Extrema, *IEEE/CAA Journal of Automatica Sinica*, 2(2), 158-165.
[24] X. Feng, et al. 2016 Coupled Cross-correlation Neural Network Algorithm for Principal Singular Triplet Extraction of a Cross-covariance Matrix, *IEEE/CAA Journal of Automatica Sinica*, 3(2), 147-156.