Scale-Adaptive Low-Resolution Person Re-Identification via Learning a Discriminating Surface

Zheng Wang[1], **Ruimin Hu**[1,*], **Yi Yu**[2], **Junjun Jiang**[3], **Chao Liang**[1] **and Jinqiao Wang**[4]
[1]State Key Laboratory of Software Engineering, School of Computer, Wuhan University, China
[2]Digital Content and Media Sciences Research Division, National Institute of Informatics, Japan
[3]School of Computer Science, China University of Geosciences, China
[4]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
{wangzwhu, hrm, cliang}@whu.edu.cn, yiyu@nii.ac.jp, junjun0595@163.com, jqwang@nlpr.ia.ac.cn

Abstract

Person re-identification, as an important task in video surveillance and forensics applications, has been widely studied. But most of previous approaches are based on the key assumption that images for comparison have the same resolution and a uniform scale. Some recent works investigate how to match low resolution query images against high resolution gallery images, but still assume that the low-resolution query images have the same scale. In real scenarios, person images may not only be with low-resolution but also have different scales. Through investigating the distance variation behavior by changing image scales, we observe that scale-distance functions, generated by image pairs under different scales from the same person or different persons, are distinguishable and can be classified as feasible (for a pair of images from the same person) or infeasible (for a pair of images from different persons). The scale-distance functions are further represented by parameter vectors in the scale-distance function space. On this basis, we propose to learn a discriminating surface separating these feasible and infeasible functions in the scale-distance function space, and use it for re-identifying persons. Experimental results on two simulated datasets and one public dataset demonstrate the effectiveness of the proposed framework.

1 Introduction

Person re-identification (REID) is the task of visually matching images of the same person, obtained in different periods from different cameras distributed over non-overlapping locations of potentially substantial distances. It represents a valuable task in video surveillance scenarios [Li *et al.*, 2015a; Chen *et al.*, 2015]. Since classical biometric cues, such as face and gait, are usually unreliable or even infeasible in the uncontrolled surveillance environment [Zheng *et al.*, 2011], the appearance of individuals is mainly exploited for REID. Previous research efforts for solving the REID problem have primarily focused on the following two aspects: (1) feature representation [Farenzena *et al.*, 2010; Zhao *et al.*, 2013; An *et al.*, 2015a], which aims at constructing discriminative visual descriptions that can easily distinguish different persons in various cameras. (2) distance measure [Wang *et al.*, 2014a; Koestinger *et al.*, 2012; Wang *et al.*, 2016; An *et al.*, 2015b], which aims at learning a proper distance metric by using a group of labeled training data.

Although REID is a challenging task, where the capture environment changes a lot in various cameras [Liu *et al.*, 2013;

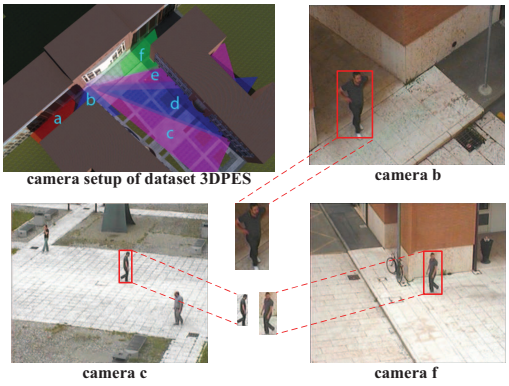


Figure 1: **An example illustrating image resolutions with different scales in the person re-identification task.** Three images of the same person are captured in three different camera views in the dataset 3DPES. The resolutions of these images are significantly different. The person image captured by camera b is relatively high-resolution, but the person images captured by camera c and camera f are relatively low-resolution. Meanwhile, the resolutions of any pair of these images are different. This situation is not investigated in previous person re-identification methods, where they assume that all person images are high-resolution with the uniform scale.

*R. Hu is the corresponding author, and is also with the National Engineering Research Center for Multimedia Software, the Collaborative Innovation Center of Geospatial Technology, and the Hubei Provincial Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University.

Wang *et al.*, 2014b], existing methods have shown effective results on some public datasets. Person images in related datasets are typically normalized to the same high-resolution (HR) and with a uniform scale. So, most of approaches generally assume that the individual image scale is constant, without considering low-resolution (LR) and scale mismatching. However, this assumption does not conform to practical situations. As Fig.1, using 3DPES dataset [Baltieri *et al.*, 2011] as an example, illustrates, a more practical situation is that person images are **not only LR, but also holding different scales**. We name this kind of task Scale-Adaptive Low-Resolution Person Re-identification (SALR-REID).

As far as we know, two pioneer researches investigated the LR REID problem. In 2015, motivated by super-resolution restoration works [Jiang *et al.*, 2015; 2014], [Jing *et al.*, 2015] proposed a semi-coupled low-rank discriminant dictionary learning approach, intending to uncover the relationship between the features of LR and HR images. During the same time period, [Li *et al.*, 2015b] assumed that images of the same person should distribute intrinsically in a similar structure in a latent space, and then optimized the cross-scale image domain alignment simultaneously with discriminant distance metric modeling in a joint learning framework. In both of these researches, given a LR probe image, the algorithm is expected to match against normal or even HR gallery images. Specially, in [Jing *et al.*, 2015], the probe images are uniformly 1/8 down-sampled from the original HR images, while in [Li *et al.*, 2015b] the resized scale is a quarter of the original HR scale in common. Based on the relatively ideal assumption that **scales of LR are the same**, these two approaches show their effectiveness in the LR REID problem, through introducing relationship between HR and LR into traditional re-identification models. However, both of these two researches have neglected the fact that the scales of LR are always different. If there were 100 different scales in the dataset, the methods need to construct 100 different relationships, and it cannot be guaranteed the 100 relationships work perfectly matching.

For the new SALR-REID problem, the practical task is that given a HR probe image, the algorithm is expected to match against LR gallery images with different scales. So, traditional models, efficient and effective to re-identify gallery images with the same scale, may result in a significant loss of performance when the resolutions of gallery images are low and the scales vary unsteadily. In addition, traditional features using different resolutions, such as wavelet [Mallat, 1989] and SIFT [Ng and Henikoff, 2003], are proved not proper for the REID task with low-resolution images.

Since the feature distance of a probe-galley image pair varies with the resolution of the galley image, an important aspect of the problem is to understand how the feature distance of the image pair changes when the gallery image scale goes down gradually. In this way, we map the association of the gallery image scale and the distance of the probe-galley image pair, then generate a scale-distance function (SDF) for each image pair. Inspired by [Martinel *et al.*, 2015], which attempts to classify the feasible and infeasible warp functions to re-identify different image pairs, we propose to classify different SDFs formed by different image pairs, instead of exploiting the feature distance metric model.

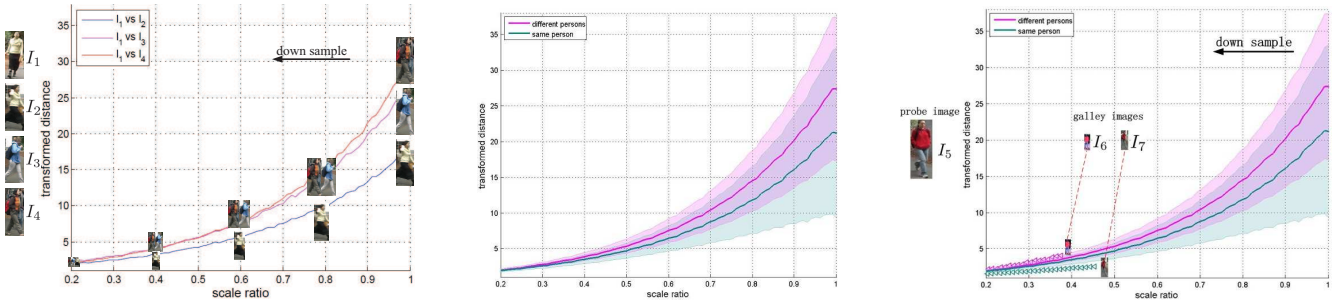
Actually, we observe that SDFs can be separated as feasible (positive) and infeasible (negative) ones as well. The feasible functions are generated by image pairs from the same targets, while the infeasible functions are generated by those from different targets. These feasible and infeasible functions comprise the scale-distance function space (SDFS). The observation in detail, which demonstrates a visual proof of the discriminating power of the feasible and infeasible SDFs, is described in Sec.2. The proposed work explores this discriminating power in the SDFS for SALR-REID. First, a SDFS composed of the collection of feasible and infeasible SDFs is built. Second, parameter vectors are obtained by the model fitting method to represent SDFs. Third, a random forest of decision trees is learned by the trained parameter vectors to discriminate whether a test pair of query-gallery images is feasible or not.

To summarize, the contributions of this paper are the followings. (1) We raise a new issue, Scale-adaptive Low-resolution Person Re-identification, which has not been investigated before as far as we know. (2) We observe the discriminating power of the feasible and infeasible SDFs respectively generated by positive and negative image pairs. (3) Then, the SALR-REID problem is addressed by mapping a SDF onto the SDFS and classifying it as either the feasible or infeasible SDF. Experimental results on two simulated datasets and one public dataset demonstrate the effectiveness of the proposed framework.

2 Motivation

As the image resolution affects the feature distance, we attempt to comprehend the distance variation rule as the resolution changes.

To investigate the issue, we made a preliminary experiment. Four images were selected from the VIPeR dataset [Gray *et al.*, 2007]. They were respectively denoted as I_1, I_2, I_3, I_4, where I_1 acted as the probe image, and the other three acted as the gallery images. To simplify the process, an easy and uniform method was used to obtain the feature distance of every image pair. At first, we divided each image, regardless of its resolution, into 24 patches (3 columns * 8 rows). For each patch, 64 dimension HSV feature were extracted. Then, each image was represented by a 1536 dimension feature, and the feature distance of an image pair was calculated by Euclidean distance. In this way, we could easily obtain a sequence of feature distances between the probe image I_1 and one of the gallery images $I_2 - I_4$, when we down-sampled the gallery image step by step. We used d to denote each Euclidean distance, and k to represent the scale ratio of the down-sampled gallery image resolution to the original one. To visualize the changes of d with k, we drew curves with sequences of $k - d$ value pairs in Fig.2, where the horizontal axis represents the scale ratio, and the vertical axis stands for the transformation value of distance. It should be noted that, to make observation easier, the distance was transformed by $d' = exp(d * k)$. We named each curve as the scale-distance function. Fig.2(a) shows the three SDFs respectively generated by I_1 and $I_2 - I_4$. We found



(a) (b) (c)

Figure 2: In the above three figures, the horizontal axis represents k, the scale ratio of the down-sampled gallery image resolution to the original probe image resolution, and the vertical axis denotes the transformation value of the distance of image pair. The transformation formulation is $d' = exp(d * k)$, where d is the feature distance between two images, and d' is the transformation form. (a) **Three scale-distance functions.** I_1 - I_4 are four images in the VIPeR dataset [Gray *et al.*, 2007]. I_1 and I_2 are from the same person, and I_1, I_3 and I_4 are from different persons. Each curve describes how the transformed distance between each probe-gallery image pair changes with the scale of gallery image resolution. We name this curve scale-distance function. (b) **The distribution of scale-distance functions.** 100 randomly chosen samples of feasible and infeasible scale-distance functions are averaged to get the mean scale-distance functions (in bold line). The shaded areas show the corresponding spread of the variances (as $\pm$ standard deviation value). This figure shows that feasible and infeasible scale-distance functions, respectively for same persons and different persons, can be discriminative and used for re-identification. (c) **New testing probe-gallery image pairs in the distribution.** I_5 - I_7 are three images in the VIPeR dataset, where I_5 is the HR probe image, and I_6 and I_7 are LR with different scales. I_5 and I_7 are from the same person, while I_5 and I_6 are from different persons. As I_6 and I_7 are down-sampled gradually, we draw the part of scale-distance functions. This figure shows that the scale-distance function generated by I_5 and I_7 is more likely to be the feasible one, while that generated by I_5 and I_6 is more likely to be the infeasible one.

that, even though for each curve, the transformed distance value decreases with the ratio k, the three functions can be separated for different image pairs.

Furthermore, we randomly selected 100 image pairs of the same persons and 100 image pairs of different persons from the VIPeR dataset. Then, 100 SDFs for the same persons and different persons were respectively obtained, following the SDF generation method described above. As Fig.2(b) shows, the red bold line stands for the mean SDF for different persons, and the red shaded area demonstrates the corresponding spread of the variance. Meanwhile, the blue bold line and shaded area together represent the SDFs for the same persons. We name the SDFs for the same persons as feasible ones, and that for different persons as infeasible ones. This figure shows that feasible and infeasible SDFs can be discriminative and used for re-identification. The proposed method explores this discriminating power of the feasible and infeasible SDFs in the SDFS for SALR-REID. Directly exploiting general feature-distance models may not be possible to discriminate persons well enough in this kind of resolution mismatching situation. So we computed the SDF between a pair of resolution mismatching images in the actual experimentations to deal with these challenges. Discrimination between the two classes of SDFs is further enhanced in a classification framework which finds a complex discriminating surface in a higher dimensional SDFS. In Fig.2(c), an example was introduced to show how to match I_6 and I_7 with different low resolutions against the probe image I_5. In the same way, we gradually down-sampled the gallery images and calculated their distances, then formed two SDFs. We can see from the figure that the SDF generated by I_5 and I_7 lies in the blue shaded area, and is more likely to be the feasible one. In comparison, the SDF generated by I_5 and I_6 lying in the red shaded area is more likely to be the infeasible one.

Observing the rule above, to better re-identify persons, we propose to learn a discriminating surface separating these two sets of functions in SDFS, and then classify a test function as feasible or infeasible.

3 Our Approach

The overall scheme of the proposed person re-identification process is shown in Fig.3. It is divided into two stages. In the offline training stage, given positive and negative training pairs we learn a discriminative model in the SDFS to get the probability indicating whether a sample SDF comes from the same person or not. For each image pair, parameter vector valued SDFs are computed by the SDF representation module. The set of all feasible and infeasible SDFs forms the SDFS. Given the SDFS, a decision surface discriminating the two sets of SDFs is learned using a Random Forest (RF) [Breiman, 2001]. In the online re-identifying stage, test image pairs are input to the SDF representation module to compute the SDFs. Finally, the RF classifies the SDFs in the SDFS as feasible or infeasible.

We first divide the training data into positive pairs and negative pairs, where each pair of images come from different cameras. We denote a probe-gallery image pair as $(\mathbf{I}_i, \mathbf{I}_j)$. Regardless of the resolution difference, we extract visual fea-

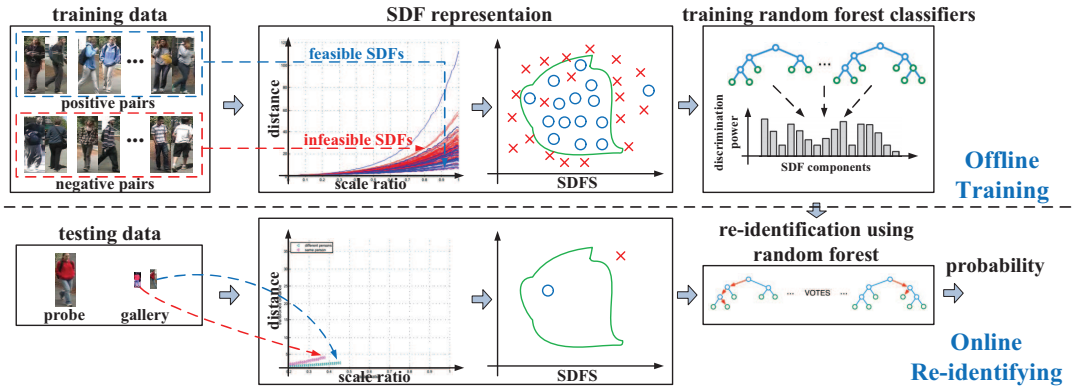


Figure 3: *The framework of our method.* It includes the offline training stage and onine re-identifying stage. The scale-distance function representation module takes each image pair, resizes one image scale step by step, obtains a series of distances, generates the SDF, and transforms the representation to a parameter vector. In the figure, we use two dimension space as an example to demonstrate SDFS. A random forest classifier is trained to discriminate between the feasible and the infeasible SDFs in the SDFS. The trained classifier is used to classify the test SDFs.

ture $\mathbf{x}$ for each image $\mathbf{I}$, with the same dimension m. Sec.2 gives an example, that can be named as a patch-based feature representation method, to represent different resolution images with the same dimension.

3.1 SDF representation

To generate the scale-distance function, holding on the size of $\mathbf{I}_i$, we represent its visual feature as $\mathbf{x}_i^1$, where the superscript 1 stands for the scale ratio between the sampled image resolution and the original image resolution. Then, we down-sample $\mathbf{I}_j$ step by step, and thus obtain a series of visual features $\mathbf{x}_j^1, \mathbf{x}_j^{0.99}, \mathbf{x}_j^{0.98}, ..., \mathbf{x}_j^{0.06}, \mathbf{x}_j^{0.05}$. Here, it should be noted that the scale ratio stands for the ratio of two image heights or widths. That is to say, if the scale ratio $k = 0.5$, it means that the height and the width of the down-sampled image $\mathbf{I}_j^{0.5}$ is half of those of the image $\mathbf{I}_i$. In our approach, we decrease the scale of the image $\mathbf{I}_j$ by 0.01 in each down-sampling step. And the minimum scale ratio is 0.05, because we find that the feature can not be extracted if the resolution is too low. Generally, we obtain 96 different scales of each image.

After obtaining a series of visual features from $\mathbf{I}_j$, we calculate a range of distances $d_{i,j}(\mathbf{x}_i^1, \mathbf{x}_j^k), k \in [0.05, 1]$ between $\mathbf{I}_i$ and the series of down-sampled $\mathbf{I}_j$. Then each distance $d_{i,j}(\mathbf{x}_i^1, \mathbf{x}_j^k)$ is transformed into a new form $d'_{i,j}(\mathbf{x}_i^1, \mathbf{x}_j^k) = \exp(d_{i,j}(\mathbf{x}_i^1, \mathbf{x}_j^k) * k)$, where k stands for the scale ratio. That is to say, for each image pair $(\mathbf{I}_i, \mathbf{I}_j^k)$, the transformed distance d' varies with k. By this means, the SDF is generated as the curves shown in Fig.3, where the horizontal axis stands for the scale ratio k, and the vertical axis stands for the transformed distance d'. In the offline training stage, those positive pairs produce the feasible SDFs, and those negative pairs produce the infeasible SDFs.

To separate these two kinds of functions, the model fitting method is exploited to learn a set of parameters, and then we use those parameter vectors to represent SDFs. We denote $f(\cdot)$ as the model, $\mathbf{w}$ as the parameter vector of the model. For each training image pair $(\mathbf{I}_i, \mathbf{I}_j)$, $K = 96$ pairs of observation data $(k, d'_{i,j}(k))$ are used for the model fitting process. Let f be the uniform model from k to $d'_{i,j}$, that is

$$d'_{i,j}(k) = f(k, \mathbf{w}), k \in [0.05, 1]. \quad (1)$$

Supposed that the dimension of the parameter vector $\mathbf{w} = (w_0, ..., w_n, ..., w_{N-1})$ is N, the parameter vector $\mathbf{w}_{i,j}$ for image pair $(\mathbf{I}_i, \mathbf{I}_j)$ is generated as

$$\mathbf{w}_{i,j} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{K} \sum_{k \in [0.05,1]} |d'_{i,j}(k) - f(k, \mathbf{w})|^2 + \lambda \sum_{n=0}^{N-1} |w_n| \quad (2)$$

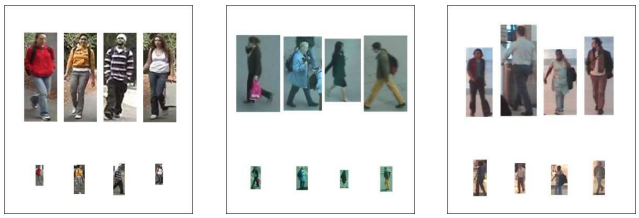
In the online re-identifying stage, different LR gallery images hold different resolutions. This resolution mismatching makes it non-reliable to rank images directly. Following the procedure in the offline training stage, the test gallery image is resized from 1 to 0.05 step by step as well. Then, the SDF representation module obtains a series of observation data. If the resolution of gallery image is too low, the number of observation data may be limited for the regression process. So we also up-sampled the gallery image by bilinear interpolation method to get more observation data. Then, we get the same number of observations in the re-identifying stage as in the training stage. By regressing the observation data, a test parameter vector for the probe-gallery image pair is generated.

In this paper, the fitting model we use is the polynomial model as follows,

$$f(k, \mathbf{w}) = \sum_{n=0}^{N-1} w_n k^n \quad (3)$$

3.2 Training random forest classifiers

The parameter vector may not be discriminating enough between the feasible and infeasible SDFs. Thus, a classifier giving more importance to the more discriminative dimension is



(a) SALR-VIPeR (b) SALR-PRID (c) CAVIAR

Figure 4: *Example image pairs from three datasets.* Each column shows two images of the same identity from two different cameras with different resolutions, where images in the bottom row are LR. (a) the SALR-VIPeR dataset; (b) the SALR-PRID dataset; (c) the CAVIAR dataset.

desirable. A random forest (RF) [Breiman, 2001] is a popular and efficient classifier based on bootstrapped aggregation ideas. It is a combination of many binary decision trees built using several bootstrap samples. At each node of each tree a subset of the SDF dimensions is randomly chosen and the best split is calculated only within this subset. This randomization of the SDF dimensions effectively chooses the dimensions according to their importance in separating the feasible and the infeasible functions in the SDFS.

In the offline training stage, the SDF representation module produces positive parameter vectors from positive pairs, and negative parameter vectors from negative pairs, as discussed above. Then, the RF module exploits these vectors to train random forest classifiers.

3.3 Re-identification in SDFS

In the online re-identifying stage, the trained RF classifies a SDF as coming from the same target or not according to whether it lies in the positive or the negative region. For example, a new coming probe-gallery image pair first produces a series of distances by down-sampling and up-sampling the gallery image. Then, the SDF representation module transforms the generated SDF into a parameter vector. The RF module predicts the probability that the parameter vector is positive, which is the probability that the gallery image and the probe image come from the same target. Actually, the computation of the proposed method can be divided into two parts, (i) offline: changing the scale of a gallery image and obtaining a series of visual features of scaled gallery images, (ii) online: computing the visual feature of the query image, and a series of distances. Therefore, the proposed method has almost the same the online computation time as traditional methods.

4 Experiments

In this section, we first demonstrate the influence brought by the scale-adaptive low-resolution problem. Then, the proposed method is evaluated. The evaluation is run on two simulated person datasets SALR-VIPeR and SALR-PRID, which are based on the VIPeR dataset [Gray *et al.*, 2007] and the PRID450S dataset [Roth *et al.*, 2014] respectively, and one public dataset the CAVIAR dataset [Cheng *et al.*, 2011].

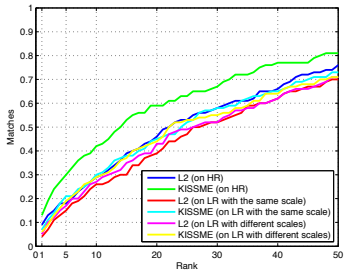


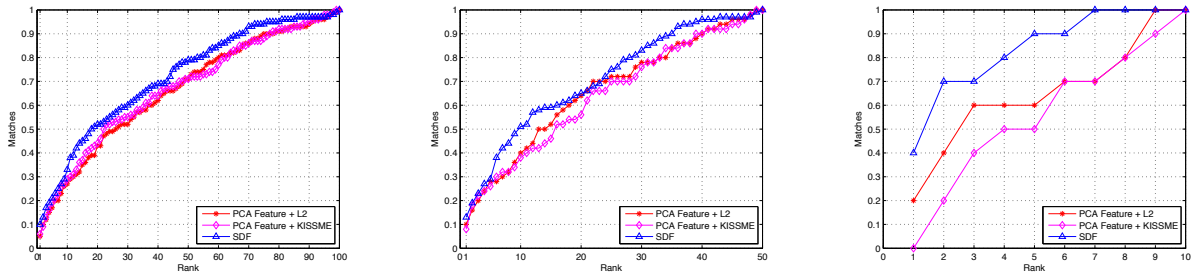
Figure 5: *Evaluation on three different resolution situations of the VIPeR dataset.* The curves of evaluation on HR denote that both the probe and gallery sets are HR. The curves of evaluation on LR with the same scale denote that the gallery set turns to be LR, and the scale of the resolution is the same. The curves of evaluation on LR with different scales denote that the gallery set is LR, but the scales of resolutions are different. L2 stands for L2 distance. KISSME [Koestinger *et al.*, 2012] stands for a metric learning method.

4.1 Experimental Datasets

SALR-VIPeR. The widely used VIPeR dataset [Gray *et al.*, 2007] contains 1264 outdoor images obtained from two views of 632 persons. Each person has a pair of images taken from two different cameras respectively. View changes are the most significant cause of appearance change. Other variations are also considered, such as illumination conditions. However, resolution differences are ignored and all images of individuals are normalized to a size of $128*48$ pixels. Images from camera A are set as the HR probe set, whose resolution remains unchanged. While images from camera B are set as the LR gallery set, which are down-sampled randomly to different scales. The scale ratios range from 0.1 to 0.25. Some example images are shown in Fig.4(a).

SALR-PRID. The PRID450S [Roth *et al.*, 2014] is a challenge dataset, particularly there is camera characteristics variation. It contains 450 singleshot image pairs captured over two spatially disjoint camera views. All images are normalized to $168 * 80$ pixels. Different from the VIPeR dataset, this dataset has significant and consistent lighting changes and chromatic variation. Images from camera A are set as the HR probe set, whose resolution remains unchanged, while images from camera B are set as the LR gallery set, which are down-sampled randomly to different scales. The scale ratios range from 0.1 to 0.25. Some example images are shown in Fig.4(b).

CAVIAR. The CAVIAR dataset [Cheng *et al.*, 2011] is widely used for evaluating person re-identification, containing images of 72 individuals captured from 2 cameras in a shopping mall. This dataset is suitable for testing SALR-REID, as the resolution of images captured from the second camera is much lower than that in the first camera (Fig.4(c)). Among the 72 people, 18 were only captured in a single camera view with no low resolution images, and they were thus removed. The remaining persons were used in our experiments, where a HR image of each person is selected to form



(a) SALR-VIPeR (b) SALR-PRID (c) CAVIAR

Figure 6: *Experimental results on three datasets.* For each dataset, we compare the proposed method (SDF) with general feature distance method (PCA feature+L2), and a distance learning method (PCA feature+KISSME). (a) the SALR-VIPeR dataset; (b) the SALR-PRID dataset; (c) the CAVIAR dataset.

the probe set, and a LR image of each person is selected to form the gallery set.

4.2 Experimental Settings

Feature Representation. In order to compare the distance of two images with different resolutions, we represent each image feature with the same dimension. As Sec.2 does, we divide each image, regardless of its resolution, into 24 patches (3 columns * 8 rows). For each patch, a 64 dimension HSV feature is extracted. Then, each image is represented by a 1536 dimension feature. To accelerate the process and reduce noise, we conducted principal component analysis (PCA) to obtain a relatively low dimensional representation, i.e. 300 for the SALR-VIPeR and SALR-PRID datasets, and 100 for the CAVIAR dataset.

Settings. All datasets are randomly divided into training set and testing set. Persons for training and testing are respectively 532 and 100 (SALR-VIPeR), 400 and 50 (SALR-PRID), and 44 and 10 (CAVIAR). The probe set consists of all HR images per person. LR images are randomly downsampled and selected to construct the gallery set. The entire evaluation procedure was repeated 5 times. Cumulative Matching Characteristic (CMC) curves [Wang *et al.*, 2007] were used to calculate the average performance, and the value of CMC@r indicates the percentage of the real match ranked in the top r.

4.3 Observation on the Influence by Scale-Adaptive Low-Resolution

In this subsection, we demonstrate that the traditional feature-distance model will gradually lose its effectiveness, as the resolution of images transforms from HR to LR with the same scale, then to LR with different scales. We evaluated on three different resolution situations using the VIPeR [Gray *et al.*, 2007] dataset. Holding on the resolution of the probe set, we respectively tested on the HR gallery set, the LR gallery set with the same scale, and the LR gallery set with different scales. The scale ratio of image resolution of the LR gallery set with the same scale was 0.15, while that of the LR gallery set with different scales ranged from 0.1 to 0.25. For each situation, the L2 distance and a metric learning method (KISMME [Koestinger *et al.*, 2012]) were exploited to obtain the distance of each probe-gallery image pair. We listed the top-50 CMC curves of these three situations in Fig.5, each generating two curves.

From the Fig.5, the conclusions are made as follows. (1) Comparing L2 distance results in the three situations, we conclude that more resolution mismatching makes results worse. (2) Comparing the results respectively on the HR and the LR gallery set with the same scale, we conclude that the metric learning method is suitable for resolution matching situation or constructing the relationship of resolutions with two different scales. (3) The promotion of the metric learning method is limited, when it is exploited on the LR gallery set with different scales.

4.4 Evaluation on the Datasets

In this subsection, we prove that the proposed method is suitable for the SALR-REID task. We evaluated the effectiveness of the proposed method by comparing with general feature-distance model regardless of resolution mismatching, and the distance metric learning method KISSME, which tries to construct the resolution relationship, on the SALR-VIPeR dataseet, the SALR-PRID dataset and the CAVIAR dataset, respectively. The obtained results are shown in Fig.6. As can be seen, our approach has improvements on all the three datasets for the SALR-REID task, compared with the PCA Feature+L2 and PCA Feature+KISSME methods, and the effectiveness of the metric learning method is not obvious when encountering this multiple resolution mismatching problem.

5 Conclusion

This paper raises a new issue, which has not been investigated before as far as we know. Traditional features under multi-resolutions and feature-distance models may not be proper for this task, due to the resolution mismatching. We propose to learn a discriminating surface to address the new problem, by mapping a SDF onto the SDFS and classifying it as either the feasible or infeasible SDF. Experimental results illustrate that the traditional model has a significant loss of performance when the resolutions of gallery images are low and the scales vary unsteadily, and demonstrate the effectiveness of the proposed framework.

Acknowledgments

The research was supported by the National High Technology Research and Development Program of China (2015AA016306, 2014AA015104), the National Nature Science Foundation of China (61231015, 61303114, 61273034, 61332016), the Specialized Research Fund for the Doctoral Program of Higher Education (20130141120024), the Nature Science Foundation of Hubei Province (2014CFB712), and the Technology Research Project of Ministry of Public Security (2014JSYJA016).

References

[An *et al.*, 2015a] Le An, Mehran Kafai, Songfan Yang, and Bir Bhanu. Person re-identification with reference descriptor. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2015.

[An *et al.*, 2015b] Le An, Songfan Yang, and Bir Bhanu. Person re-identification by robust canonical correlation analysis. *Signal Processing Letters, IEEE*, 2015.

[Baltieri *et al.*, 2011] Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *Joint ACM workshop on Human gesture and behavior understanding*, 2011.

[Breiman, 2001] Leo Breiman. Random forests. *Machine learning*, 2001.

[Chen *et al.*, 2015] Ying-Cong Chen, Wei-Shi Zheng, and Jianhuang Lai. Mirror representation for modeling view-specific transform in person re-identification. In *IJCAI*, 2015.

[Cheng *et al.*, 2011] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011.

[Farenzena *et al.*, 2010] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[Gray *et al.*, 2007] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007.

[Jiang *et al.*, 2014] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhen Han. Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning. *Image Processing, IEEE Transactions on*, 2014.

[Jiang *et al.*, 2015] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhihua Cai. Cdmma: Coupled discriminant multi-manifold analysis for matching low-resolution face images. *Signal Processing*, 2015.

[Jing *et al.*, 2015] Xiao-Yuan Jing, Xiaoke Zhu, Fei Wu, Xinge You, Qinglong Liu, Dong Yue, Ruimin Hu, and Baowen Xu. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In *CVPR*, 2015.

[Koestinger *et al.*, 2012] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[Li *et al.*, 2015a] Sheng Li, Ming Shao, and Yun Fu. Cross-view projective dictionary learning for person re-identification. In *IJCAI*, 2015.

[Li *et al.*, 2015b] Xiang Li, Wei-Shi Zheng, Xiaojuan Wang, Tao Xiang, and Shaogang Gong. Multi-scale learning for low-resolution person re-identification. In *ICCV*, 2015.

[Liu *et al.*, 2013] Chunxiao Liu, Chen Change Loy, Shaogang Gong, and Guijin Wang. Pop: Person re-identification post-rank optimisation. In *ICCV*, 2013.

[Mallat, 1989] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1989.

[Martinel *et al.*, 2015] Niki Martinel, Aruneema Das, Christian Micheloni, and AK Roy-Chowdhury. Re-identification in the function space of feature warps. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[Ng and Henikoff, 2003] Pauline C Ng and Steven Henikoff. Sift: Predicting amino acid changes that affect protein function. *Nucleic acids research*, 2003.

[Roth *et al.*, 2014] Peter M Roth, Martin Hirzer, Martin Köstinger, Csaba Beleznai, and Horst Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*. 2014.

[Wang *et al.*, 2007] Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu. Shape and appearance context modeling. In *ICCV*, 2007.

[Wang *et al.*, 2014a] Yimin Wang, Ruimin Hu, Chao Liang, Chunjie Zhang, and Qingming Leng. Camera compensation using a feature projection matrix for person reidentification. *Circuits and Systems for Video Technology, IEEE Transactions on*, 2014.

[Wang *et al.*, 2014b] Zheng Wang, Ruimin Hu, Chao Liang, Qingming Leng, and Kaimin Sun. Region-based interactive ranking optimization for person re-identification. In *PCM*. 2014.

[Wang *et al.*, 2016] Zheng Wang, Ruimin Hu, Chao Liang, Yi Yu, Junjun Jiang, Mang Ye, Jun Chen, and Qingming Leng. Zero-shot person re-identification via cross-view consistency. *Multimedia, IEEE Transactions on*, 2016.

[Zhao *et al.*, 2013] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.

[Zheng *et al.*, 2011] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by probabilistic relative distance comparison. In *CVPR*, 2011.