10-2014

Online learning on incremental distance metric for person re-identification

Yuke SUN

Hong LIU

Qianru SUN
Singapore Management University, qianrusun@smu.edu.sg

Citation
SUN, Yuke; LIU, Hong; and SUN, Qianru. Online learning on incremental distance metric for person re-identification. (2014). *Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), Bali, December 5-10*. 1421-1426. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/4462

Proceedings of the 2014 IEEE
International Conference on Robotics and Biomimetics
December 5-10, 2014, Bali, Indonesia

Online Learning on Incremental Distance Metric for Person Re-identification

Yuke Sun, Hong Liu† and Qianru Sun

Abstract— **Person re-identification is to match persons appearing across non-overlapping cameras. The matching is challenging due to visual ambiguities and disparities of human bodies. Most previous distance metrics are learned by off-line and supervised approaches. However, they are not practical in real-world applications in which online data comes in without any label. In this paper, a novel online learning approach on incremental distance metric, OL-IDM, is proposed. The approach firstly modifies Self-Organizing Incremental Neural Network (SOINN) using Mahalanobis distance metric to cluster incoming data into neural nodes. Such metric maximizes the likelihood of a true image pair matches with a smaller distance than that of a wrong matched pair. Second, an algorithm for construction of incremental training sets is put forward. Then a distance metric learning algorithm called Keep It Simple and Straightforward Metric (KISSME) trains on the incremental training sets in order to obtain a better distance metric for the neural network. Aforesaid procedures are validated on three large person re-identification datasets and experimental results show the proposed approach's competitive performance to state-of-the-art supervised methods and self-adaption to real-world data.**

I. INTRODUCTION

Person re-identification handles person matching between a given probe query image and a set of candidate images captured from views of non-overlapping filed. Typically, the goal is to find the images in these candidates and return a list of probabilistic matched images ranked by degree of similarity. It is of great importance in visual surveillance and typical applications are not limited to criminal retrieval , analyzing crowd movements, multi-camera tracking in public places.

Nevertheless, person re-identification is a critical and challenging problem because a person's appearance often undergoes large variations under bad illumination, low resolution and in different poses. When observed under different camera views, different people look more alike than that of the same individual. A recent trend to tackle this problem is to use metric learning to minimize intra-class variation whilst maximize inter-class variation. Kostinger et al. [4] introduced an easy and efficient strategy called Keep It Simple and Straightforward Metric (KISSME) to learn a distance metric from equivalence constraints. Compared to other methods (LMNN [1], ITML [2], IDML [3]), KISSME does not rely on complex optimization and computationally expensive iterations. However, in a real situation, the performance of the machine learning model may deteriorate over time as new incoming data may deviate from the initial training data. Tradition methods are retrained in the batch mode using both existing and new data, which is time-consuming.

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025, 61340046), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Science and Technology Innovation Commission of Shenzhen Municipality (No.201005280682A, No.JCYJ20120614152234873), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20130001110011).

Mr. Yuke Sun is with the Engineering Lab on Intelligent Perception for Internet of Things(ELIP), Shenzhen Graduate School, Peking University, Shenzhen, 518055 CHINA. sunyuke@pku.edu.cn

Prof. Hong Liu (Corresponding Author) is with the Engineering Lab on Intelligent Perception for Internet of Things(ELIP), and the Key Laboratory of Machine Perception, Peking University, Beijing, 100087 CHINA. hongliu@pku.edu.cn

Ms. Qianru Sun is with the Engineering Lab on Intelligent Perception for Internet of Things(ELIP), Shenzhen Graduate School, Peking University, Shenzhen, 518055 CHINA. qianrusun@sz.pku.edu.cn

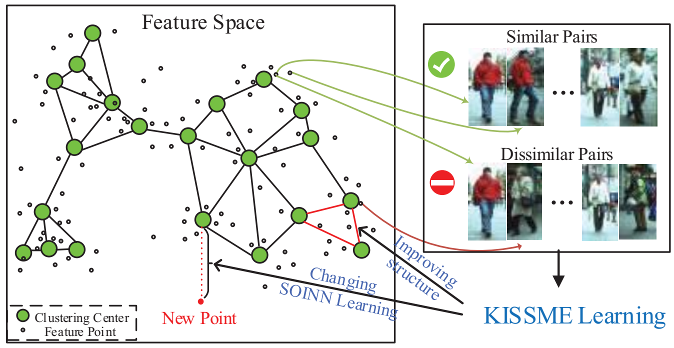


Fig. 1: An overview of the proposed approach.

To overcome this problem, well-known incremental learning algorithms, such as Self-Organizing Map [5], Growing Neural Gas [6], are designed based on neural network to represent unlabeled data's topological structure and cluster the data into different classes. To be suitable for processing online data, Shen et al. [7] proposed an incremental learning method called Self-Organizing Incremental Neural Network (SOINN). It outperforms aforementioned algorithms by learning the necessary number of neural nodes and representing the topological structure of input probability density. A salient disadvantage of SOINN is that it uses Euclidean distance to measure distance between input data and nodes. However, because of the intra-class and inter-class variation, a Mahalanobis metric is more suitable for person re-identification problem.

This paper proposes a incremental distance metric learning method based on clustering to solving the problem of real-world re-identification. Figure 1 shows an overview of the proposed approach. In this approach, new data is learned by a neural network to obtain a stable topological structure of input space. Then typical prototype nodes are output from

978-1-4799-7396-5/14/$31.00 © 2014 IEEE

SOINN. They are used to obtain several nearest samples and construct similar and dissimilar image pairs for further training. Then KISSME is done using such training set to update the metric matrix of SOINN. The reason for adopting KISSME is that it only involves computation of two small sized covariance matrices, which can be trained efficiently. The learning of neural network and updating of metric are performed iteratively. We refer the proposed approach as OL-IDM.

Performances are evaluated on the VIPeR [8], i-LIDS [9] and ETHZ [10] dataset. The results demonstrate that (1) when tested on on-line data, the approach shows its ability to fit the input space; (2) it achieves competitive results to state-of-the-art supervised distance metric learning methods.

II. RELATED WORK

Existing methods on person re-identification can generally fall into two categories: appearance-based methods and learning-based methods.

A. Appearance-based methods

There are a large number of feature types having been proposed, such as global or regional features and patch-based features. Many methods integrated some of these types of features in order to gain more robust features [12], [13]. Spatial information about the layout of these features is also an important cue. A typical example is the Symmetry-Driven Accumulation of Local Features (SDALF) [14] which located relevant body parts driven by asymmetry and symmetry principles to handle view variation.

Above methods focused on feature design while others considered an alternative perspective that feature saliency was valuable in describing each particular individual and feature importance mining could be achieved in an unsupervised way [15], [16].

B. Learning-based methods

Traditional feature learning methods for re-identification like Support Vector Machine and boosting [17], [18] are widely used. These methods generally cast the problem into two-class or multi-class classification problem.

However, Prosser et al. [19] formulated person re-identification as a ranking problem. It trained a primal RankSVM ranker and tried to find a linear function to weight the absolute difference between samples.

Relative Distance Comparison (RDC) [20] considered the joined effect between different features using a second-order feature quantification model. RDC could be viewed as specific variant of distance metric learning algorithm [1], [2], [3]. Metric learning algorithms learn a Mahalanobis metric which is more powerful to measure feature differences. Mignon et al. [21] introduced the Pairwise Constrained Component Analysis to learn the distance metric from sparse pairwise similarity/dissimilarity constraints in the high-dimensional input space. Moreover, Li et al. [22] learned specific metric under a transferred metric learning framework.

Recently, slow metric learning speed of early approaches has driven re-identification research toward the fast and light methods [4], [23]. Nevertheless, these methods require identity labels for training samples and generally trained off-line.

A related work of the proposed approach is [24] which performed incremental clustering and distance metric learning iteratively to produce an online incremental clustering algorithm for high-dimensional data. It used Adaptive Metric Learning algorithm [25] to help SOINN to separate chunk clusters by removing edges between neural nodes. A main difference between our method and the former one is that they focus on improving the quality of data clustering while ours works on incremental distance metric learning.

III. BASIS ALGORITHMS

A. Review of KISSME

KISSME [4] is a simple and effective strategy to learn a distance or similarity metric based on Mahalanobis distance functions. Generally, the Mahalanobis distance metric measures the squared distance between two features $\mathbf{x}_i$ and $\mathbf{x}_j$ as:

$$d^2_{\boldsymbol{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{M} (\mathbf{x}_i - \mathbf{x}_j). \tag{1}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}_d$ and $\boldsymbol{M}$ is a positive semidefinite. The method tries to obtain the decision whether a pair of image (i, j) is dissimilar or not by a likelihood ratio test:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log \left(\frac{p(\mathbf{x}_i, \mathbf{x}_j | H_0)}{p(\mathbf{x}_i, \mathbf{x}_j | H_1)} \right). \tag{2}$$

Hypothesis H_0, which means a pair is dissimilar, is accepted when a high value of $\delta(\mathbf{x}_i, \mathbf{x}_j)$ is got. By contrast, a low value means hypothesis H_1 is accepted and the pair is considered as similar.

Assuming the Gaussian structure of the difference space is $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, we can relax the problem and rewrite Eq. (2) to

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \log \left(\frac{\frac{1}{\sqrt{2\pi |\sum_{y_{ij}=0}|}} \exp(-1/2 \mathbf{x}_{ij}^T \sum_{y_{ij}=0}^{-1} \mathbf{x}_{ij})}{\frac{1}{\sqrt{2\pi |\sum_{y_{ij}=1}|}} \exp(-1/2 \mathbf{x}_{ij}^T \sum_{y_{ij}=1}^{-1} \mathbf{x}_{ij})} \right), \tag{3}$$

where

$$\sum\nolimits_{y_{ij}=0} = \sum_{y_{ij}=0} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \tag{4}$$

$$\sum\nolimits_{y_{ij}=1} = \sum_{y_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \tag{5}$$

and $y_{ij} = 1$, if $\mathbf{x}_i$ and $\mathbf{x}_j$ share the same class label; otherwise, $y_{ij} = 0$. The maximum likelihood estimate of the Gaussian is equivalent to minimizing the Mahalanobis distances in a least-square manner. By taking the log, Eq. (3) can be changed into

$$\begin{aligned} \delta(\mathbf{x}_{ij}) &= \mathbf{x}_{ij}^T \sum\nolimits_{y_{ij}=1}^{-1} \mathbf{x}_{ij} - \mathbf{x}_{ij}^T \sum\nolimits_{y_{ij}=0}^{-1} \mathbf{x}_{ij} \\ &\quad + \log(|\sum\nolimits_{y_{ij}=1}|) - \log(|\sum\nolimits_{y_{ij}=0}|) \\ &\approx \mathbf{x}_{ij}^T (\sum\nolimits_{y_{ij}=1}^{-1} - \sum\nolimits_{y_{ij}=0}^{-1}) \mathbf{x}_{ij}. \end{aligned} \tag{6}$$

Finally, Mahalanobis distance metric $\boldsymbol{M}$ in Eq. (1) can be obtained by re-projection of $\hat{\boldsymbol{M}} = \sum_{y_{ij}=1}^{-1} - \sum_{y_{ij}=0}^{-1}$ onto the cone of positive semidefinite matrices and $\hat{\boldsymbol{M}}$ can be determined by clipping the spectrum of $\hat{\boldsymbol{M}}$ (details can be referred to [4]).

B. Review of SOINN

The SOINN [7] is an unsupervised incremental learning algorithm based on [5] and [6]. As with online non-stationary data which has a complex distribution, it can approximate the distribution of input data and estimate the number of classes in a self-organizing way.

A two-layer network is adopted in original SOINN and the training results of the first layer are used as the training set for the second layer. In each layer, if a new data is far away from its nearest nodes and the second nearest nodes, it is added as a new node, otherwise, the nearest and second nearest nodes learn the new data by adative learning rates. Details of SOINN algorithm are described elsewhere in the literature [7].

SOINN is effective for processing real-world data as it is not necessary to predefine its network structure. However, the original SOINN utilizes Euclidean distance to measure the distance between the input data and nodes. It is not suitable for the re-identification problem because of the intra-class and inter-class variation. Therefore, we present a revised version of SOINN with a Mahalanobis metric.

IV. OVERVIEW OF OL-IDM APPROACH

The proposed approach aims at learning an incremental distance metric to gain a reliable Mahalanobis distance metric. Considering the ever growing amount of data, learning a stable Mahalanobis metric is tough because of labored labeling work on a large scale dataset. Fortunately, the preocess of OL-IDM based on SOINN algorithm covers the high dimensional feature spaces with growing neural nodes. It can serve as a tool to predict whether two data points are similar or not. If two data points are similar, they own the same nearest node.

When assuming that every neural node is a good cluster of people with similar appearances, a distance metric learning algorithm can get enough training data for further incremental learning. These training data is acquired by picking up some points nearest to the same node as similar pairs, whilst some points from different nodes as dissimilar pairs. Section IV-A explains the details of how to construct a training set for KISSME.

After batch metric learning, the updated distance metric is adopted by SOINN. OL-IDM does not retrain the neural network using the updated metric. Actually, the influence of old data decreases with removing of nerual nodes. The updated metric gradually improves the topological structure of neural network, which makes those existing nodes have better ability of representation.

The remaining question is how often KISSME is done in the learning process. Section IV-B explains a method to decide when metric matrix will be updated. At last, we explain the OL-IDM algorithm in Section IV-C.

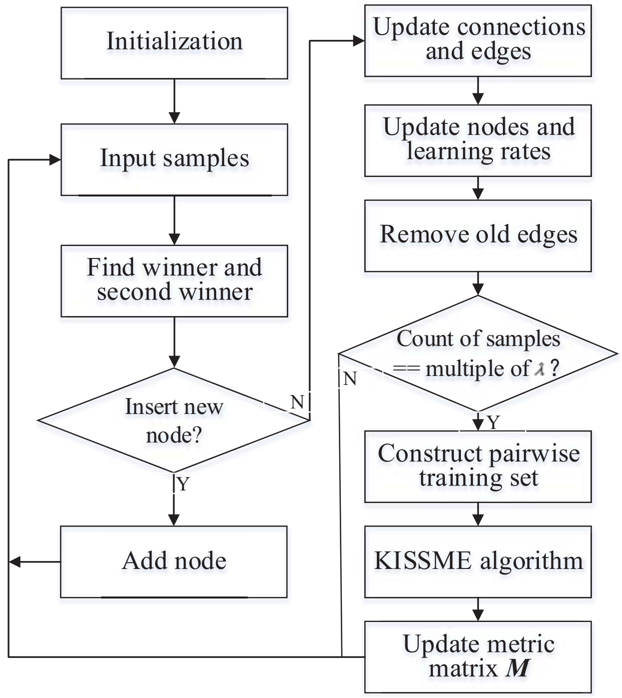


Fig. 2: Flowchart the OL-IDM approach

A. Construction of incremental training set

The incremental training set can be equally divided into two sets. One contains those pairs that belong to similar classes, the other one is composed of mismatched pairs of points that belong to dissimilar classes.

Assuming there is a neural network with a node set N trained by SOINN. Given a node $\mathbf{i} \in N$, we search its n nearest neighbors points $N_n^{\mathbf{i}}$. Let $\mathbf{x}_k^{\mathbf{i}} \in N_n^{\mathbf{i}}$ represents k-th element of the neighbor set of node $\mathbf{i}$. Then training set of matched points S_{sim} are defined as below:

$$S_{sim} = \{S_{sim}^{\mathbf{i}} | \mathbf{i} \in N\}, \tag{7}$$

$$S_{sim}^{\mathbf{i}} = \{(\mathbf{x}_a^{\mathbf{i}}, \mathbf{x}_b^{\mathbf{i}}) | a = 1, ..., n, b = rand(n), b \neq a\}. \tag{8}$$

Likewise, the training set of mismatched points S_{dis} can be defined as:

$$S_{dis} = \{S_{dis}^{\mathbf{i}} | \mathbf{i} \in N\}, \tag{9}$$

$$S_{dis}^{\mathbf{i}} = \{(\mathbf{x}_a^{\mathbf{i}}, \mathbf{x}_b^{\mathbf{j}}) | a = 1, ..., n, b = rand(n), \mathbf{i} \neq \mathbf{j}\}, \mathbf{j} \in N \backslash \{\mathbf{i}\}. \tag{10}$$

B. Decision to update $\boldsymbol{M}$

The number of nodes keeps changing during the training. With the purpose of reducing the frequency of updating we gather a count C for the number of new nodes added and noise nodes removed. The distance metric $\boldsymbol{M}$ is retrained when the equation is satisfied as:

$$C \geq ||N||/\alpha, \tag{11}$$

where α is a constant tradeoff between time cost and efficiency in the online learning.

C. Procedure of OL-IDM Approach

In this section, the approach is described following the flowchart in Fig. 2. As limited by the scope of this paper, we briefly separate the procedure into four parts.

Initialization. In order to get a good initial guess of $\hat{\boldsymbol{M}}_0$, a small labeled training set S_0 for KISSME is required. The training set S_0 contains the images from a small number of people captured in various sites. It is noted that SOINN's initialization is data hungry. It requires thousands of data points to learn a stable neural network. Therefore, we duplicate S_0 for k times ($k = 100$ in our experiments) to complete the initialization.

Input samples and node insertion. When an input vector $\mathbf{x}$ is given to SOINN, it finds the nearest node (winner) $\mathbf{c}_1 \in \mathbb{R}_d$ and the second-nearest node (second winner) $\mathbf{c}_2 \in \mathbb{R}_d$ of $\mathbf{x}$ as follows:

$$\begin{aligned} \mathbf{c}_1 &= \arg\min_{\mathbf{c}\in N}((\mathbf{x}-\mathbf{c})^T\boldsymbol{M}(\mathbf{x}-\mathbf{c})), \\ \mathbf{c}_2 &= \arg\min_{\mathbf{c}\in N\setminus\{\mathbf{c}_1\}}((\mathbf{x}-\mathbf{c})^T\boldsymbol{M}(\mathbf{x}-\mathbf{c})), \end{aligned} \tag{12}$$

where N is the set of all nodes.

$\mathbf{x}$ is added as a new node if the distance between $\mathbf{x}$ and $\mathbf{c}_1$ or $\mathbf{c}_2$ is greater than the similarity threshold $T_{\mathbf{c}_1}$ or $T_{\mathbf{c}_2}$. Calculation of such thresholds can refer to [7].

Updating process of SOINN. Different from original algorithm [7], we only adopt the first layer of SOINN for a consideration of efficiency. If $\mathbf{x}$ is not added as a new node, then it is used to optimize the structure of network. In the optimization, the edge between $\mathbf{c}_1$ and $\mathbf{c}_2$ is connected; the ages of the other edges emanating from $\mathbf{c}_1$ are incremented by 1; vectors of the winner nodes and its direct topological neighbors $N^{\mathbf{c}_1}$are updated as follows:

$$\begin{aligned} \mathbf{c}_1 &= \mathbf{c}_1 + \varepsilon_1(\mathbf{x}-\mathbf{c}_1), \\ \mathbf{i} &= \mathbf{i} + \varepsilon_2(\mathbf{x}-\mathbf{i}), \mathbf{i}\in N^{\mathbf{c}_1}, \end{aligned} \tag{13}$$

where ε_1 and ε_2 are dynamic learning rates. At last, edges with age greater than a predefined threshold a_g are removed.

Incremental Distance Metric Learning. Calculation of $\delta(\mathbf{x}_{ij})$ in KISSME algorithm fails when $y_{ij=0}$ or $y_{ij=1}$ is not full rank. Noise nodes may expose the incremental learning to such a risk. Hence, OL-IDM removes noise nodes periodically as SOINN does.

If the number of input vectors is an integer multiple of parameter λ, the apporach finds the nodes whose neighbors are less than or equal to 1 and deletes such nodes based on the idea that such nodes are "noisy". After denoising we refer to section IV-B for deciding whether it should update the distance metric matrix $\boldsymbol{M}$. When updating is confirmed, incremental metric learning is carried out using incremental pairs of samples constructed from section IV-A and ones from initial labeled training set S_0. After that, the learning continues. Algorithm 1 shows the main procedure of incremental learning.

Algorithm 1 Process of incremental learning.

Input:
Nodes set N;
Count C.
Output:
Updated distance metric $\boldsymbol{M}$.
Remove d "noise" nodes;
C = C + d;
if $C \geq ||N||/\alpha$
  $S_{sim} = S_{dis} = \{\}$;
  for Node $\mathbf{i} \in N$ **do**
    Search the nearest neighbor points $N_n^{\mathbf{i}}$ for $\mathbf{i}$;
    $S_{sim}^{\mathbf{i}} = \{(\mathbf{x}_a^{\mathbf{i}}, \mathbf{x}_b^{\mathbf{i}}) | a = 1, ..., n, b = rand(n), b \neq a\}$;
    Randomly select a node $\mathbf{j}$ from N, $\mathbf{j} \in N \setminus \{\mathbf{i}\}$;
    $S_{dis}^{\mathbf{i}} = \{(\mathbf{x}_a^{\mathbf{i}}, \mathbf{x}_b^{\mathbf{j}}) | a = 1, ..., n, b = rand(n), \mathbf{i} \neq \mathbf{j}\}$;
    $S_{sim} = S_{sim} \bigcup S_{sim}^{\mathbf{i}}$, $S_{dis} = S_{dis} \bigcup S_{dis}^{\mathbf{i}}$;
  end for
  $\boldsymbol{M} = KISSME(S_{sim} \bigcup S_{dis} \bigcup S_0)$;
  C = 0;
end

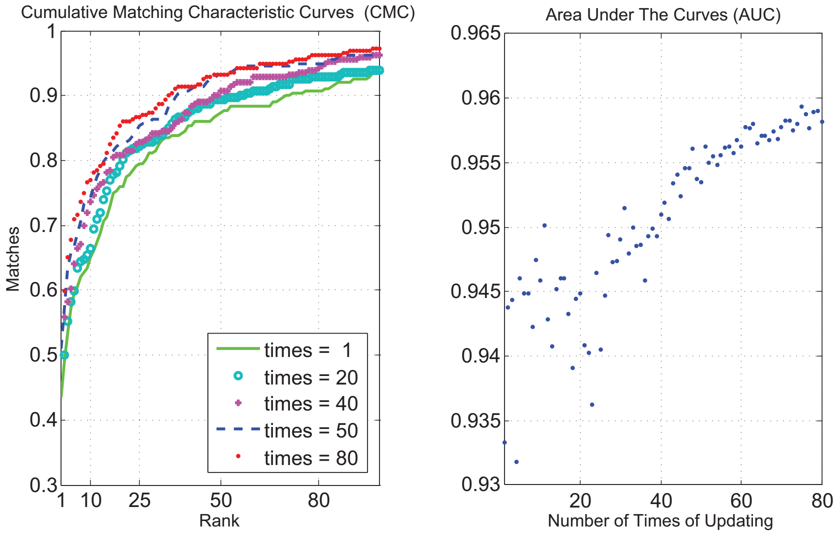
V. EXPERIMENT AND DISCUSSION

The OL-IDM approach[1] is evaluated on three person re-identification datasets, VIPeR dataset [8], i-LIDS [9] and ETHZ [10]. The VIPeR dataset is a person re-identification dataset captured outdoor consisting of 632 people with two images for each person. For i-LIDS dataset, there are 119 people with a total 476 images captured by multiple non-overlapping cameras at a busy airport arrival hall. It has an average of 4 images for each person. The ETHZ dataset was originally designed for person detection and tracking in image sequences captured from a moving camera in a busy street scene. It contains 146 people and 8555 images. See Fig. 3 for some examples of datasets.

Fig. 3: Examples from the VIPeR, i-LIDS and ETHZ.

In order to obtain a generic and representative descriptor, we adopt the way of feature extraction in [20]. In [20], Zheng et al. divided a person image into six horizontal stripes. Then RGB, YCbCr, HSV color features, Schmid and Gabor filters are computed. In total 29 feature channels are constructed for each stripe and each feature channel is represented by a 16 dimensional histogram. Each person image is thus represented by a feature vector in a 2784

[1]Matlab code can be found at https://github.com/mocosun/IDM .



(a) (b)

Fig. 4: Evaluation of OL-IDM with online data. (a) ROC curves show the performance improves with more times of updating of distance metric; (b) AUC results of every updating of $\boldsymbol{M}$ are presented.

	Train: VIPeR Test: ETHZ				Train: i-LIDS Test: VIPeR			
Rank(%)	**r=1**	**r=10**	**r=25**	**r=50**	**r=1**	**r=10**	**r=25**	**r=50**
RDC	40.4	62.3	72.3	83.2	1.4	7.9	11.6	19.6
KISSME	51.0	**73.0**	81.9	91.1	2.7	10.8	17.3	25.8
ITML	43.8	64.4	75.3	82.5	1.7	9.1	14.2	20.0
IDML	48.7	70.0	80.5	89.4	2.5	9.7	15.0	22.6
LMNN	47.7	70.9	80.5	90.1	2.2	9.7	15.2	21.7
OL-IDM*	**52.3**	**73.0**	**83.4**	**95.3**	**4.5**	**15.0**	**23.7**	**34.5**

TABLE I: **Scalability comparison to state-of-the-art approaches.** Top ranked matching rate on different datasets. OL-IDM has also learned the test data in an unsupervised way (*).

dimensional. Finally, to be compared to [4], we project the concatenated descriptor into a 40 dimensional subspace by PCA . PCA helps to reduce the large memory requirement by RDC, especially when training on ETHZ.

A. Evaluation of OL-IDM with online data

Since ETHZ dataset has more images than other two datasets, we treat it as a sequence of samples which are learned one by one. 283 samples from the first 4 persons in ETHZ are used to construct a training set with 486 pairs of samples. Then the initial metric matrix $\hat{\boldsymbol{M}}_0$ is trained using this small training set. SOINN is also initialized using the same 283 samples according to section VI-C. We set the maximum age is $a_g = 30$, $\lambda = 100$, $\alpha = 10$ for updating , nearest neighbor number $n = 7$ for constructing the incremental training set.

Figure 4 shows the results. It is clear that with the data coming, the performance of the approach improves and the peek of AUC curve is reached at the 80th incremental learning. Rank-1 rate grows from 43.5% to 52.4%, while Rank-10, Rank-25 and Rank-50 increase by 18.4% on average.

B. Scalability comparison to state-of-the-art approaches

We evaluated the scalability of different state-of-the-art approaches that we train on ViPeR and test on ETHZ. The second experiment trains on i-LIDS and test on VIPeR. The reason for this experiment is to simulate a real world situation. In such situation, unseen samples are coming without knowing which classes the samples are belong to. Note that because the OL-IDM approach has also learned the data of the testing set, this comparison is not fair enough. A fair comparison will be done in the next section.

In TABLE I the performance of OL-IDM is compared to state-of-the-art methods in a range of the first 50 ranks. It is observed that OL-IDM outperforms all the five state-of-the-art approaches across all ranks. The approach achieves a slight rise in the first experiment. Moreover, since the images in the VIPeR are more similar to ETHZ than i-LIDS, the ranking rates in the first experiment are much higher than the second one, which proves the proposed approach's ability of self-adaption to the changing data.

C. Comparisons on wholly labeled datasets

LMNN [1], ITML [2] and LDML [3] are used for performance evaluation on EHTZ and i-LIDS. In particular, we examine results with Euclidean distance (L2 method) for demonstration of better performance with Mahalanobis metric.

Figure 5 shows that comparing to the state-of-the-art

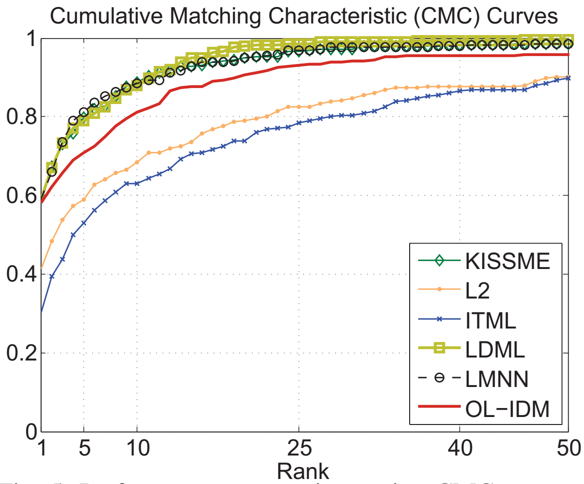


Fig. 5: Performances comparison using CMC curves on the ETHZ dataset.

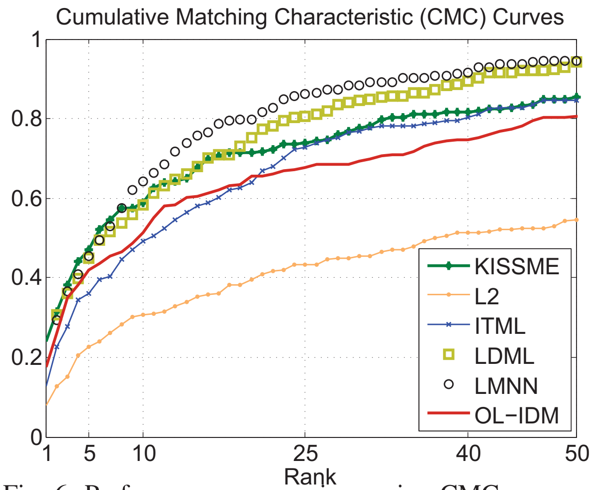


Fig. 6: Performances comparison using CMC curves on the i-LIDS dataset.

methods, the proposed approach achieves competitive results without significant loss of performance (8.2% on average compared to the best LDML). The loss is bigger in Fig. 6 while controlled in an acceptable range. It is mainly because the approach only learns 476 samples on the i-LIDS. The quality of representation of nodes in the neural network depends on the size of training set. Nevertheless, the OL-IDM still outperforms L2 across all ranks.

VI. CONCLUSIONS

In this work we introduce an online learning method on incremental distance metric (OL-IDM) for person re-identification. The learning of neural network and updating of metric are performed iteratively. Typical prototype nodes output from SOINN are used to obtain nearest samples which constitute similar and dissimilar pairs of a incremental training set. Then KISSME is trained using such training set to update the metric matrix of SOINN. The performances on three datasets demonstrate that the proposed approach outperform the state-of-the-art methods with in-line data and achieves competitive results to the off-line supervised distance metric learning algorithms.

REFERENCES

[1] K. Q. Weinberger and L. K. Saul. Fast solvers and efficient implementations for distance metric learning. In Proc. IEEE Intern. Conf. on Machine Learning, pp. 1160-1167, 2008.

[2] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In Proc. IEEE Intern. Conf. on Machine Learning, pp. 209-216, 2007.

[3] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In Proc. IEEE Intern. Conf. on Computer Vision, pp. 498-505, 2009

[4] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth and H. Bischof. Large scale metric learning from equivalence constraints. In Proc. IEEE Intern. Conf. on Computer Vision and Pattern Recognition, pp. 2288-2295, 2012.

[5] T. Kohonen. Self-organized formation of topologically correct feature maps. Biological Cybernetics, vol. 43, pp. 59-69, 1982.

[6] B. Fritzke. A growing neural gas network learns topologies. Adv. in neural information processing systems. vol. 7, pp. 625-632, 1995.

[7] F. Shen, O. Hasegawa. An Incremental Network for On-line Unsupervised Classification and Topology Learning. Neural Networks, vol. 19, pp. 90-106, 2006.

[8] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In Proc. IEEE Intern. Workshop on Performance Evaluation of Tracking and Surveillance, 2007.

[9] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In Proc. British Machine Vision Conf., 2009.

[10] A. Ess, B. Leibe, and L. Van Gool. Depth and appearance for mobile scene analysis. In Proc. IEEE Intern. Conf. on Computer Vision, pp. 1-8, 2007.

[11] F. Shen and O. Hasegawa. A fast nearest neighbor classifier based on self-organizing incremental neural network. Neural Networks, vol. 21, no. 10, pp. 1537-1547, 2008.

[12] L. Bazzani, M. Cristani, A. Perina, and V. Murino. Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recogn. Lett., vol. 33, no. 7, pp. 898-903, 2012.

[13] Y. Lu, L. Lin, and W.-S. Zheng. Human re-identification by matching compositional template with cluster sampling. In Proc. IEEE Intern. Conf. on Computer Vision, pp. 3152-3159, 2013.

[14] L. Bazzani , M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. Comput. Vis. Image Underst., vol. 117, no. 2, pp. 130-144, 2013.

[15] C. Liu, S. Gong, and C. C. Loy. On-the-fly feature importance mining for person re-identification. Pattern Recognition, vol. 47, no. 4, pp. 1602-1615, 2014.

[16] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In Proc. IEEE Intern. Conf. on Computer Vision and Pattern Recognition, pp. 3586-3593, 2013.

[17] D.Gray and H.Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In Proc. European Conf. on Computer Vision, pp. 262-275, 2008.

[18] M. Hirzer , C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. Image Analysis, vol. 6688, pp. 91-102, 2011.

[19] B. Prosser, W. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In Proc. British Machine Vision Conf., 2010.

[20] W. Zheng, S. Gong, and T. Xiang. Re-identification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 653-668, 2013.

[21] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. IEEE Intern. Conf. on Computer Vision and Pattern Recognition, pp. 2666-2672, 2012.

[22] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In Proc. Asian Conf. on Computer Vision, pp. 31-44, 2012.

[23] M. Hirzer, P. Roth, M. Kostinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In Proc. European Conf. on Computer Vision, pp. 780-793, 2012.

[24] S. Okada and T. Nishida. Online incremental clustering with distance metric learning for high dimensional data. In Proc. Intern. Joint Conf. on Neural Networks, pp. 2047-2054, 2011.

[25] J. Ye, Z. Zhao, and H. Liu. Adaptive distance metric learning for clustering. In Proc. IEEE Intern. Conf. on Computer Vision and Pattern Recognition, pp.1-7, 2007