MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

People Re-identification Using Two-Stage Transfer Metric Learning

Guanwen Zhang, Jien Kato, Yu Wang, Kenji Mase
Graduate Shcool of Information Science, Nagoya University, Japan
zhang@mv.ss.is.nagoya-u.ac.jp

Abstract

With assumptions that people usually do not change their clothes during an observation period, people appearance data are easily outdated in re-identification applications. This raises the over-fitting problem because only a few training data are available for learning statistical models. In this paper, we propose a two-stage transfer metric learning approach for multiple-shot people re-identification to tackle this small training data problem. In the first stage, we transfer the generic knowledge from a large existing dataset, and in the second stage, we transfer the learned distance metric for each probe-specific person using the side-information. Experimental results on several public benchmark datasets show that our proposed approach is superior over conventional approaches.

1 Instructions

People re-identification is to recognize the people from different camera views. It is an important task to understand people behavior across camera network for surveillance purpose. However because of the low camera resolutions and the long distance to the camera, people biological information (e.g., face or gait) is generally unavailable. Therefore the existing researches in this filed mainly focus on people appearance with an acceptable assumption that people will not change their clothing during the observation period. This leads to another unpleasant situation that people appearance data are easily outdated and there is only a few training data available. The high level statistical models learned on the limited data usually suffer from over-fitting and will eventually affect the re-identification performance.

Although it has not been widely addressed, there are several interesting works trying to tackle this issue by using the transfer learning methods for re-identification propose. In Wu et al.'s work [10], they sample several images as a "third party" from another dataset. On the basis of the "third party", they utilize the sparse coding method to express the probe and gallery images as a collaborative representation. They aim to discover an intermediate feature representation that bridges the gap between the query and the gallery sets. Zheng et al. [14] formulate the re-identification as a set-based verification problem under a transfer learning framework. The useful relationships between non-target data (unlabeled people) and target data (labeled people) are introduced as additional constrains for learning process, which results in a more robust model than the one learned on target data only. Zhang et al. [13] propose an adaptive metric learning method, which transfer the information from both the source and target set. By utilizing the generic and target metrics, the adaptive metric learning method can not only acquire the overall knowledge about the re-identification, but also be closely associated with the target task. The concept of transfer learning is also applied in single-shot re-identification scenario. For example, Li et al. [6] learn a generic distance metric on the entire target training set. According to the similarities of probe person, they select and re-weight samples as the training data. The learned generic metric is then transferred to a candidate-set-specific distance metric.

Despite the great efforts on this topic, it still leaves much room to further improve. Inspired by Zhang et al.'s work [12], we propose a two-stage transfer metric learning approach to tackle the small training data problem for multiple-shot people re-identification. The proposed approach consists of two transfer distance metric learning stage. In the first stage, we perform the adaptive metric learning. The adaptive metric learning can leverage the generic knowledge from existing datasets (source set) to integrate with specific information of target task (target set). The generic knowledge, encoded in generic distance metric, is able to furnish the missing information for target task, while specific task information, encoded in target distance metric, will lead the generic knowledge to avoid the sub-optimal solution. On the basis of adaptive metric learning, we carry out a second stage transfer metric learning for each probe person during the online re-identification. Given a probe person, the side-information, i.e., the image data of other person captured in the same camera view, is available. We intend to transfer the adaptive distance metric toward probe-specific person by utilizing such side-information. Since such side-information will certainly differ from probe person, we are able to impose more constraints into the metric learning. On the other hand, the image data in probe and gallery set are captured in different camera views, emphasizing much on the side-information (only contains probe image data) will result in over-fitting. Therefore, we regularize the probe-specific distance metric with the adaptive metric to improve the metric quality. Integrated with adaptive metric and more reliable constraints, the second stage metric learning is able to increase the ability of probe-specific metric to distinguish the probe person from others.

The organization of this paper is as follow. Section 2 first briefly introduces the adaptive metric learning method for the first stage metric learning. Section 3 then proposes the second stage transfer metric learning method for the probe-specific distance metric. The experimental results are discussed in Section 4, and conclusion are presented in Section 5.

2 The First Metric Learning Stage

In the first transfer metric learning stage, we employ adaptive metric learning [12] to transfer the information from source set to the target set. The adaptive metric learning is able to leverage the knowledge from existing datasets to assist target task learning, which is perfectly consistent with our need. In this section, we briefly introduce the adaptive metric learning.

Given two appearance instances, x_i and x_j, the Mahalanobis distance can be expressed as $d_M(x_i, x_j) = (x_i - x_j)^T M(x_i - x_j)$, where $M \in \Re^{d\times d}$ is the distance metric, a symmetric and definite matrix that completely parameterizes the distance function. Let D_s and D_t denote the source and target dataset, and let M_s and $M_{t'}$ denote the generic and target distance metric learned on D_s and D_t, respectively. The adaptive metric learning tries to integrate the two metrics together for the desired adaptive distance metric M_t. The objective function is defined as:

$$f_{M_t} = L(M_t, D_t) + \gamma_1 R(M_t, M_s) + \gamma_2 R(M_t, M_{t'}). \quad (1)$$

In Eq. (1), L term is the loss function on the target task, while the two R terms are regularization, which are divergence functions for measuring the nearness between generic and target distance metrics.

The basic idea of defining loss term L is to keep similar appearance instances close and keep dissimilar appearance instances far apart. In a spirit similar to that of local distance comparison [11], L term is defined to punish the invasion in the local neighbors from the invaders with different person identity. The regularization R terms in Eq. (1) encode M_s and $M_{t'}$ into the M_t. In order to preserve the local structure, the von Neumann divergence is utilized in Bregman divergence. Therefore M_t is learned by minimizing the objective function Eq. (1) with constraints as:

$$\begin{aligned} \min \sum_{i,j\rightsquigarrow i} & d_{M_t}(x_i, x_j) + C \sum_{i,j\rightsquigarrow i,k} (1 - \xi_{ik})\zeta_{ijk} \\ & + \gamma_1 tr(M_t log M_t - M_t log M_s - M_t + M_s) \\ & + \gamma_2 tr(M_t log M_t - M_t log M_{t'} - M_t + M_{t'}), \\ \textbf{s.t.}\ & d_{M_t}(x_i, x_k) - d_{M_t}(x_i, x_j) \geq 1 - \zeta_{ijk}, \\ & \zeta_{ijk} \geq 0, M_t \geq 0, \end{aligned} \quad (2)$$

In Eq. (2), $j \rightsquigarrow i$ indicates that j is i's local neighbor, C is a predefined positive constant. ξ_{ik} is the indicator function that if x_i and x_k have the same person identity, $\xi_{ik} = 1$; otherwise, $\xi_{ik} = 0$. Since M_t is positive definite, Eq. (2) is convex and can be efficiently solved by the semi-definite programming. In particular, the M_t is initialized with M_s to preserve the structure information of the generic knowledge. By taking steps of gradient descent, M_t moves toward $M_{t'}$ until it reaches convergence. M_t can achieve a equilibrium point between M_s and $M_{t'}$, and make Eq. (1) minimal. The adaptive metric M_t, with capability of generic knowledge and specific information of the target task, is able to overcome over-fitting problem and have the better performance on the target re-identification task.

3 The Second Metric Learning Stage

In order to further improve the quality of the distance metric, we attend to transfer the adaptive distance metric for each probe person. During the online re-identification, we carry out a second stage metric learning to learn a probe-specific distance metric M_p. Similar as the works in [6, 7], we utilize the side-information as the training data for the metric learning propose. The side-information refers to the remaining image data in the probe and gallery side. Particularly, we only employ the data from the probe side in this paper, i.e., the image data in the gallery side are not involved during the learning process.

Without loss of the generality, consider the probe person P is captured in camera view A with N instances, i.e., $P = \{x_1, x_2, \ldots, x_N\}$. And all the left persons captured in this same camera view comprise R set, where $R = \{y_1, y_2, \ldots, y_M\}$ and M indicates the instance number. The objective function of the the second stage transfer learning is defined as:

$$f_{M_p} = L(M_p, D_t) + \eta R(M_p, M_t). \quad (3)$$

Similar as adaptive metric learning, Eq. (3) also consists of loss and regularization. A trade-off parameter η is introduced to balance the two terms.

To be consistent with the local distance comparison framework, we focus on instances' local neighbors. Given an instance x_i of P, we can find out a closet neighbor x_i^c in P. If there is an instance y_j from R and $d_M(x_i, y_j) < d_M(x_i, x_i^c)$, y_j cause a loss to local neighbors of the specific probe person P. The objective of the second stage transfer metric learning is to punish such loss. The invasion loss is supposed to be smaller in the induced feature space after metric learning. The second stage transfer learning is expected to punish such loss. To strengthen the punishment on the invasion and to improve the discriminate ability of M_p, we add a margin as additional penalties on the local neighbors. Therefore, the loss term is defined as:

$$L(M_p, D_t) = \sum_{i,j} [1 + d_{M_p}(x_i, x_i^c) - d_{M_p}(x_i, y_j)]_+. \quad (4)$$

In Eq. (6), $[\cdot]_+$ is a standard hing loss function, and $[a]_+ = \max(a, 0)$. The minimizing Eq. (4) will result in that in the local neighbors, the instances with the same person identity stay close, while the those with different identifies are far apart. In order to avoid the over-fitting on probe data, we add regularization to associate M_p with M_t. The regularization term R is essentially divergence function measuring the distance between two metrics. In the second transfer metric learning stage, we choose Bregman divergence for R term. Since the second stage metric learning will be performed during the online re-identification, it is close related to the online metric learning methods. Different from the regularization defined in Eq. (2), we choose LogDet divergence [5] for Bregman matrix divergence as most online metric learning methods [4, 8]. The regularization divergence R is then defined as:

$$R(M_p, M_t) = tr(M_t M_p^{-1}) - \log\det(M_t M_p^{-1}) - d, \quad (5)$$

where, d is the dimension of the distance metric.

On the basis of above discussion, the second stage transfer metric learning can be formulated as following optimization problem:

$$\begin{aligned} \min \sum_{ij} & \zeta_{i,j} + \eta\{tr(M_t M_p^{-1}) - \log\det(M_t M_p^{-1}) - d\} \\ \textbf{s.t.}\ & d_{M_p}(x_i, y_j) - d_{M_t}(x_i, x_i^c) \geq 1 - \zeta_{ij}, \\ & \zeta_{ij} \geq 0, M_p \geq 0. \end{aligned} \quad (6)$$

Algorithm 1 The second stage transfer metric learning

Input:

probe set: $P = \{x_1, x_2, \ldots, x_N\}$
side-information: $R = \{y_1, y_2, \ldots, y_M\}$
adaptive metric: M_t, target metric: $M_{t'}$
parameters: η, stop criterion: ε

Output:

probe-specific distance metric: M_p

initialize $M_p^{(0)} \leftarrow M_{t'}$;
initialize $l \leftarrow 1$;
repeat
  $\forall x_i \in P$, search x_i^c using $M_p^{(l)}$
  compute $L(M_p^{(l)}, D_t)$ term as in Eq. (4)
  compute $R(M_p^{(l)}, M_t)$ term as in Eq. (5)
  solve Eq. (6) for $M_p^{(l+1)}$ using the SDP solver
  $l \leftarrow l + 1$
until $\|M_p^{(l)} - M_p^{(l-1)}\|_2 < \varepsilon$

With the positive semi-definite constraints of M_p, Eq. (6) is convex. It can be solved by using SDP for a global solution. During the optimization, we use the new M_p update the nearest neighbor x_i^c for every x_i at each iteration. Particularly, we initialize the M_p with the target distance metric $M_{t'}$. Begin at a good starting point, the optimization procedure convergences fast. During experiments, we observed training a probe-specific metric with 500 instances takes less than 40 seconds. The whole procedure is summarized in Algorithm 1.

4 Experimental Results

We evaluate the proposed approach on the three public datasets: ETHZ [2], CAVIAR4REID [1] and Person Re-ID 2011 [3]. These datasets are considered as benchamarking for evaluating people re-identification approaches. The varying illumination, heavy occlusion, and different camera resolutions make them very challenging. Some selected images are shown in Fig. 1. For each dataset, images of half size of the people are used as the training set; the remaining people are used as the testing set. M images of each person are selected to generate the probe set, while different M images are selected to generate the gallery set. In order perform the adaptive metric learning, we employ an experiment setting similar to [12], i.e., when one dataset is evaluated as the target set, the left two dataset are utilized as the source set.

In the experiment, we compute a mixture of histogram features with dense color, SIFT, and HOG, LBP. These features are normalized to zero mean and unit variance, and then concatenated to form a single feature vector. A principal component analysis (PCA) is performed to preserve 90% energy. The LMNN [9] algorithm is performed on source set and target set for learning generic metric M_s and target metric $M_{t'}$. The two stages distance transfer metric learning are initialized with the two types of metrics. We perform the local distance comparison [11] for re-identification purpose. We use the cumulative matching characteristic (CMC) curve to show the re-identification matching rates. The area under curve (AUC) and proportion

Figure 1. Selected images from the three datasets. From the left to right is ETHZ, CAVIAR4REID, and Person Re-ID dataset.

of uncertain removed (PUR) scores are also utilized to indicate the overall performance. In the following experiment, we empirically set $\gamma_1 = 0.6$, $\gamma_2 = 0.4$ for adaptive metric, and set $\eta = 0.5$ for the second stage metric learning. We repeat the experiment 20 times and determine the average re-identification rate.

In the experiments, we evaluate the re-identification performance with the target metric (Target), generic metric (Generic), and adaptive metric (Adaptive). We compare the performances with that of the popular transfer learning method, transfer metric learning (TML) and multi-task large margin nearest neighbor (mtLMNN), to demonstrate the effectiveness of the proposed approach. We also compare our approach with the state-of-the-art methods: mutual subspace method (MSM), sparse approximated nearest points (SANP), and symmetry-driven accumulation of local features (SDALF). The CMC curves are plotted in Fig. 2. The normalized AUC and PUR scores are summarized in Table 1.

As shown in Fig. 2, the performance of our approach outperforms that of target, generic, and adaptive distance metric on all the three datasets. Although the adaptive metric underperform the target metric on the Person Re-ID dataset due to the poor generic metric, our approach still achieve 2% gains over that of the target metric. On the basis of the adaptive distance metric, the proposed approach performs the second stage transfer metric learning for each probe-specific person. On one hand, it can integrate the generic knowledge of re-identification and the specific information of target task into adaptive metric. On the other hand, it further transfers the adaptive metric for each probe-specific person. Comparing with the popular transfer distance metric learning methods mtLMNN and TML, our approach achieves significant improvements. The two-stage transfer metric learning obtains 20% and 10% superior of AUC over TML and mtLMNN. Different from TML, our proposed approach focuses on optimization in local neighbors instead of searching global solution. It is therefore consistent with the local distance comparison framework for multiple-shot re-identification purpose. Comparing with mtLMNN, the proposed approach transfer the information from source set to target set, and then towards probe-specific person. By contrast, the mtLMNN attempts to balance the performance on all the source sets. Benefiting from above considerations, our proposed approach consequently outperforms TML and mtLMNN. Comparing with state-of-the-art methods: MSM, SANP, and SDALF, the CMC results in Fig. 2 and normalized

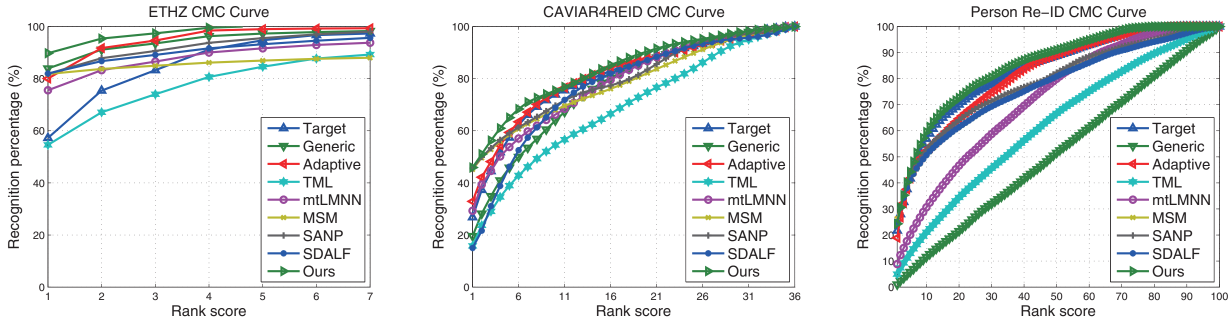

Figure 2. The CMC curves of different methods on the ETHZ, CAVIAR4REID and Person Re-ID datasets.

Table 1. Normalized AUC and PUR scores (%) on the three datasets using different methods.

Methods		Our	Target	Generic	Adaptive	TML	mtLMNN	MSM	SANP	SDALF
ETHZ	AUC	**97.66**	85.11	94.04	94.63	76.85	87.67	85.60	92.09	90.40
	PUR	**16.11**	13.66	15.39	15.51	12.05	14.13	13.71	15.00	14.66
CAVIAR	AUC	**84.04**	80.66	76.68	81.28	68.93	78.72	78.69	80.01	77.05
	PUR	**37.56**	35.82	33.81	36.12	29.75	34.79	34.68	35.39	34.04
Re-ID	AUC	**85.26**	83.50	51.75	82.19	62.46	72.36	77.60	79.00	77.45
	PUR	**82.76**	80.73	46.33	79.27	57.58	68.48	73.84	75.45	73.66

AUC and PUR in Table 1 show the advantages of our proposed approach. The proposed approach not only shows the better re-identification performance (CMC results) but also is able to be generic across different rank accuracy (normalized AUC and PUR). Since all the methods in the experiments utilize the same feature representation, we can confirm that the proposed approach is able to perform people re-identification well in different and difficult real-world scenarios.

5 Conclusion and Future Work

In this paper, we present a two-stage transfer metric learning method against small training data problem for multiple-shot people re-identification. The proposed approach consists of two stage transfer metric learning. In the first stage, we employ the adaptive metric learning to transfer the generic knowledge from source set to assist the learning task in the target set. In the second stage, we transfer the adaptive metric learning towards probe-specific metric for each probe person by utilizing the side-information. The experiments on the public datasets confirm the superior to the conventional approaches. In the further work, we intend to accelerate the optimization of the second stage transfer metric learning for the real-world re-identification applications.

Acknowledge

This research is supported by the Grant-in-Aid for Scientific Research B (No. 26280057) and the Grant-in-Aid for Challenging Exploratory Research (No. 26540081). Yu Wang is also supported by the JSPS Postdoctoral Fellowship for Foreign Researchers (No. 2604043).

References

[1] D. S. Cheng, M. Cristani, M. Stoppa, Loris Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVA*, 2011.

[2] A. Ess, B. Leibe, and L. J. V. Gool. Depth and appearance for mobile scene analysis. In *ICCV*, 2007.

[3] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Proceedings of SCIA*, 2011.

[4] Prateek Jain, Brian Kulis, Inderjit S Dhillon, and Kristen Grauman. Online metric learning and fast similarity search. In *NIPS*, pages 761–768, 2009.

[5] Brian Kulis, M. Sustik, and Inderjit Dhillon. Learning low-rank kernel matrices. In *Proceedings of the 23rd ICML*, pages 505–512. ACM, 2006.

[6] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Asian Conference on Computer Vision*, 2012.

[7] Wei Li, Yang Wu, Masayuki Mukunoki, and Michihiko Minoh. Common-near-neighbor analysis for person re-identification. In *ICIP*, 2012.

[8] J. Mei, M. Liu, H. Karimi, and H. Gao. Logdet divergence based metric learning with triplet constraints and its applications. *TIP*, 2014.

[9] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 2009.

[10] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao. Robust object recognition via third-party collaborative representation. In *ICPR*, 2012.

[11] G-W. Zhang, J. Kato, Y. Wang, and K. Mase. People re-identification with local distance comparison using learned metric. *IEICE Trans. on Inf. and Syst.*, E97-D(9), Sep. 2014.

[12] Guanwen Zhang, J. Kato, Yu Wang, and Kenji Mase. Adaptive metric learning for people re-identification. *IEICE Trans. on Inf. and Syst.*, 97(11), 2014.

[13] Guanwen Zhang, Jien Kato, Yu Wang, and Kenji Mase. Adaptive metric learning in local distance comparison for people re-identification. In *ACPR 2013*.

[14] W.-S. Zheng, S. Gong, and T. Xiang. Transfer re-identification: From person to set-based verification. In *IEEE Conference on CVPR*, 2012.