COMMON-NEAR-NEIGHBOR ANALYSIS FOR PERSON RE-IDENTIFICATION

Wei Li

Graduate School of Informatics,
Kyoto University,
Kyoto 606-8501, Japan

Yang Wu, Masayuki Mukunoki, Michihiko Minoh

Academic Center for Computing and Media Studies,
Kyoto University,
Kyoto 606-8501, Japan

ABSTRACT

Person re-identification tackles the problem whether an observed person of interest reappears in a network of cameras. The difficulty primarily originates from few samples per class but large amounts of intra-class variations in real scenarios: illumination, pose and viewpoint changes across cameras. So far, proposals in the literature have treated this either as a matching problem focusing on feature representation or as a classification/ranking problem relying on metric optimization. This paper presents a new way called *Common-Near-Neighbor Analysis*, which to some extent combines the strengths of these two methodologies. It analyzes the commonness of the near neighbors of each pair of samples in a learned metric space, measured by a novel rank-order based dissimilarity. Our method, using only color cue, has been tested on widely-used benchmark datasets, showing significant performance improvement over the state-of-the-art.

Index Terms— Person re-identification, common-near-neighbor analysis, metric learning

1. INTRODUCTION

Person re-identification is one of the most important and challenging issues of video processing for intelligent visual surveillance application currently. Current methods can be mainly categorized into two paradigms. The first paradigm treats the issue as a matching problem, and highly depends on feature representation including design, selection, ensemble and so forth, which are mostly centralized on good statistical descriptions of person appearance. One of the most representative methods called SDALF (abbr. Symmetry-Driven Accumulation of Local Features) [1] partitions the human body into three parts according to the symmetry and asymmetry of body structure. SDALF adopts three powerful features and takes the Bhattacharyya metric into consideration. Nevertheless, its performance is still far from satisfaction.

This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Special Coordination Fund for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

The second paradigm pays more attention to learning a good metric. One of the most exemplary methods is LMNN (Large Margin Nearest Neighbor) [2, 3], which uses traditional SVM framework to obtain an optimized metric for nearest neighbor classification. OMRR (Optimizing Mean Reciprocal Rank) [4], resorting to MLR (abbr. Metric Learning to Rank) [5], relying on structural SVM framework and attaching importance to designing the loss function of it, achieves more satisfactory improvement than before. However, subject to the complexity of data, in the learned metric space, many classes are still inseparable.

Based on metric learning and inspired by a newly proposed Rank-Order distance by Zhu, et al. [6], this paper recasts the primal problem into a new model, called *Common-Near-Neighbor Analysis* (using *CNNA* for short). The method contains three main components. (1) Metric Space Construction: feature space is warped by the learned metric; (2) Common-Near-Neighbor Modeling: common near neighbors of each pair of samples are analyzed in the new metric space, using both relative and direct information. It can be regarded as a kind of feature representation after metric learning; (3) Re-ranking: re-ranking is performed using our new model to obtain the improved ranking results. There are two contributions of this paper: (1) different from traditional feature representation which operates in the original data space, we propose to extract new features in the learned metric space for matching, which combines the strengths of learning and matching; (2) a new rank-based dissimilarity has been provided, which describes and synthesizes the relative information and direct information of each pair of samples. This model has been proved to be more effective and efficient than Zhu's Rank-Order distance in the issue of person re-identification.

2. COMMON-NEAR-NEIGHBOR ANALYSIS

2.1. Metric Space Construction

Person re-identification can be treated as a ranking problem, for which, dissimilarity or distance measurement in a suitable metric space is critical, because high dimension feature space is usually non-Euclidean and noisy. Learned metric with respect to the data based on Mahalanobis distance has been

proved distinctly superior to the Euclidean metric. RankSVM [7] is a considerable framework for metric learning, while MLR [5] goes further to consider the relationship among different classes and makes the margin flexible. There are kinds of loss functions for MLR, and the one named "MRR" has been proved to be an effective choice for the issue of person re-identification [4].

Given a query set $\mathcal{Q} = \{q \mid q \in \mathbb{R}^d\}$ and a corpus set $\mathcal{X} = \{x_{qi} \mid x_{qi} \in \mathbb{R}^d\}$, suppose $\phi_{qi}(x_{qi}, q)$ is used to denote the relative feature representation of a corpus sample x_{qi} w.r.t. q and suppose w is the metric we intend to optimize. A desired ranking model could be $g_w(x_{qi}) = w^T\phi_{qi}(x_{qi}, q)$, which scores each x_{qi}. Let $y \in \mathcal{Y}$ be a ranking of $\mathcal{X}$ w.r.t the query q, and $\psi(q, y, \mathcal{X}) \in \mathbb{R}^d$ be a vector-valued joint feature map [4]. Then, optimizing w for the ranking model $g_w(x_{qi})$ is equivalent to optimizing the following model based on $\psi(q, y, \mathcal{X})$.

$$\arg\min_w \frac{1}{2}\|w\|^2 + \frac{C}{|\mathcal{Q}|}\sum_q \xi_q \tag{1}$$

s.t.

$$w^T\psi(q, y_q^*, \mathcal{X}) \geq w^T\psi(q, y, \mathcal{X}) + \Delta(y_q^*, y) - \xi_q, \quad \forall q, y \neq y_q^*;$$
$$\xi_q \geq 0, \forall q,$$

where y_q^* is the ground truth ranking for a given $q \in \mathcal{Q}$, ξ_q is the slack variable, C is the trade-off parameter, and $\Delta(y_q^*, y)$ is the "MRR" loss function to penalize predicting y instead of y_q^*, as defined in [4].

With cutting-plane algorithm, we can solve the optimizing model. After learning, a new metric space can be constructed, which has higher discriminative capability than the Euclidean one. Nevertheless, previous experimental results in person re-identification show that the learned metric is not impeccable, which may be due to the small training sample size per class, high complexity of data, heuristic selection of metric form and limited power of optimizing algorithm. That means after learning, the nearest neighbor might be an intruder. Thus, we aim at exploring new feature representations in the learned metric space.

2.2. Common-Near-Neighbor Modeling

In Zhu's method [6], for face tagging, a Rank-Order distance has been proposed to study the neighborhood structure. Firstly, two ranking order lists are generated by sorting the corpus to the given query utilizing absolute distance, and then an asymmetric Rank-Order distance is defined as:

$$D(a, b) = \sum_{i=0}^{O_a(b)} O_b(f_a(i)), \tag{2}$$

where $f_a(i)$ returns the i^{th} element in the ranking order list of a and $O_a(b)$ is the order of b in a's ranking order list. Asymmetric Rank-Order distance is calculated by the sum of rank orders of a's nearest neighbors in b's ranking order list, as shown in Fig. 1. The smaller Rank-Order distance they have, the higher possibility of sharing neighbors between them there will be.

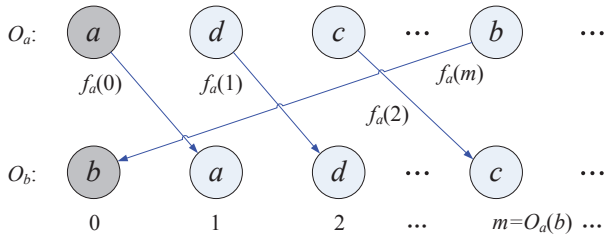

Fig. 1. O_a and O_b are ranking order lists. $D(a, b)$ is calculated from the nearest neighbor to b in a 's ranking order list.

Furthermore, the asymmetric Rank-Order distance is normalized and symmetrized for impartial comparability by:

$$D^{\text{Rank-Order}}(a, b) = \frac{D(a, b) + D(b, a)}{min(O_a(b), O_b(a))} \tag{3}$$

In Eq. 3, the original aim of normalization is to offset the bias of penalizing large $O_a(b)$ and $O_b(a)$. The numerator describes the dissimilarity of ranking order lists between samples. It is a kind of relative information, so we name it "Relative dissimilarity". The denominator reflects the dissimilarity between samples directly without relying on other samples, so we name it "Direct dissimilarity".

For unsupervised clustering in Zhu's method [6], Rank-Order distance is claimed to be able to solve the problem when data distribution has non-uniform sparse degree. In person re-identification, however, the flexible neighborhood size seems to be fragile. For two samples far from each other in the metric space, there will be a long ranking sequence between them in the ranking order list. The long ranking sequence is made up of both intra-class samples and inter-class samples. Effectiveness of Rank-Order distance will be impaired, because it is too noisy. Conversely, for two very close samples, Rank-Order distance will also lose effect, as it is too sensitive when too few samples are taken into account.

Therefore, we suggest to use a "Fixed-number n" of nearest neighbors instead of the flexible number determined by the rank of one sample in the ranking order list of another sample, as shown in Fig. 2. We will show that both effectiveness and efficiency can be enhanced when n is chosen properly.

Unlike Zhu's method, the normalization is not necessary in our model because we use a "Fixed-number" of nearest neighbors. Considering symmetry, our "Relative dissimilarity" is given by:

$$D^{\text{Relative}}(a, b) = D_{\text{Fixed-number}}(a, b) + D_{\text{Fixed-number}}(b, a) \tag{4}$$

$$D_{\text{Fixed-number}}(a, b) = \sum_{i=0}^{n} O_b(f_a(i)) \tag{5}$$

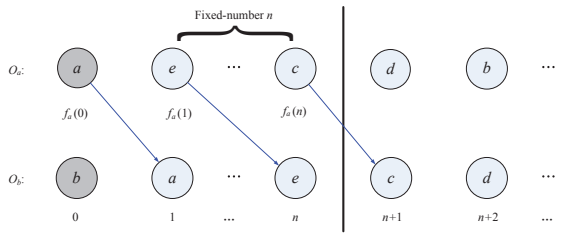

Fig. 2. $D_{\text{Fixed-number}}(a, b)$ is calculated from the 1^{th} to the n^{th} nearest neighbor in a's ranking order list.

Besides of it, we also use the "Direct dissimilarity" which is defined as:

$$D^{\text{Direct}}(a, b) = min(O_a(b), O_b(a)), \quad (6)$$

since it has its own measuring power, which is unlike Zhu's treatment of it as only a normalizer. It is also different from traditional distance computation between two sample points in using the ranking order list for high-level measurement.

In consideration of complementary advantages between "Relative dissimilarity" and "Direct dissimilarity", we present a new model named *Common-Near-Neighbor* (using *CNN* for short) dissimilarity, to combine them, as shown in Eq. 7:

$$D^{\text{CNN}}(a, b) = D^{\text{Relative}}(a, b) + K \cdot D^{\text{Direct}}(a, b), \quad (7)$$

where K is a trade-off parameter, making the model more flexible and reasonable.Re-ranking process will be performed using $D^{\text{CNN}}(a, b)$ to attain the final results.

3. EXPERIMENTAL RESULTS

We demonstrate our method *CNNA* on public available datasets VIPeR [8] and ETHZ respectively (with representative samples shown in Fig. 3) and averaging the results for ten times random training-test data splitting. For fairness, each time, we use the same selected samples in the assigned dataset for comparison.

The VIPeR dataset consists of images of 632 unique pedestrians and a total of 1264 images. Each pedestrian image pair has been taken from arbitrary viewpoints under varying illumination conditions. Complicated variations of viewpoint, illumination and pose enable "VIPeR" to be one of the most challenging datasets for the issue of person re-identification. The ETZH dataset is composed of three video sequences of crowed street scenes captured by two moving cameras mounted on a chariot. We utilize three subsets of it extracted by Schwartz and Davis for person re-identification [9]. It has smaller pose and viewpoint variations, yet more occlusions than "VIPeR". There are 83 pedestrians within 4857 images in "SEQ1", 35 pedestrians within 1936 images in "SEQ2" and 28 pedestrians within 1762 images in "SEQ3". We normalize all the images to 64×32 pixels.

VIPeR ETHZ

Fig. 3. Sample images from the VIPeR dataset and the ETHZ dataset. For "VIPeR", each column represents the matching pair of the same person. For "ETHZ", which contains multiple instances per person, three of them are shown here.

We compare our method *CNNA* with typical state-of-the-art methods on both "VIPeR" and "ETHZ". Feature representation of *CNNA* is "DCHs+wHSV", which is concatenated by "DCHs" (abbr. densely sampled color histograms) and "wHSV" (abbr. weighted HSV color histogram) [4]. The learned metric space for *CNNA* is constructed by MLR using structural SVM with loss function "MRR". Parameter-settings for *CNN* modeling are tunable. We suggest to use the average number of samples in each class to minus 1, as the "Fixed-number n" to calculate *CNN* dissimilarity in case any two samples stay so far or so close that the reliability of *CNN* dissimilarity may decrease. Since *CNN* concerns the common near-neighbors, intuitively, we expect such neighbors to be most beneficial. Too large "n" will statistically reduce the inter-class dissimilarity w.r.t. the intra-class dissimilarity, while too small "n" will increase the later w.r.t. the former. According to analysis and cross-validation, we recommend a suitable group of parameter-settings as follows: for the VIPeR dataset, $K = 1, n = 1$; for the ETHZ dataset, $K = 40, n = 60$.

Method Comparison: We compare our method with S-DALF, LMNN, OMRR, which are typical state-of-the-art methods for the issue of person re-identification. Original S-DALF uses three types of features denoted by "wHSV", "M-SCR" and "RHSP", and does matching in the Bhattacharyya metric space. Original LMNN uses "DCHs" as feature representation. Original OMRR applies "DCHs+wHSV". For conviction, besides the original ones, we add experiments using "DCHs+wHSV" for LMNN and SDALF.

Modeling Comparison: We compare our modeling with Zhu's Rank-Order distance modeling in the Euclidean space. For equitableness, we use the same feature representation "DCHs+wHSV" for Zhu's model and tentatively put Zhu's model into the same learned metric space as ours.

Metric Space Comparison: For more convincing, we also evaluate *CNNA* in different metric spaces, including the non-Euclidean ones constructed by state-of-the-art methods and the Euclidean one.

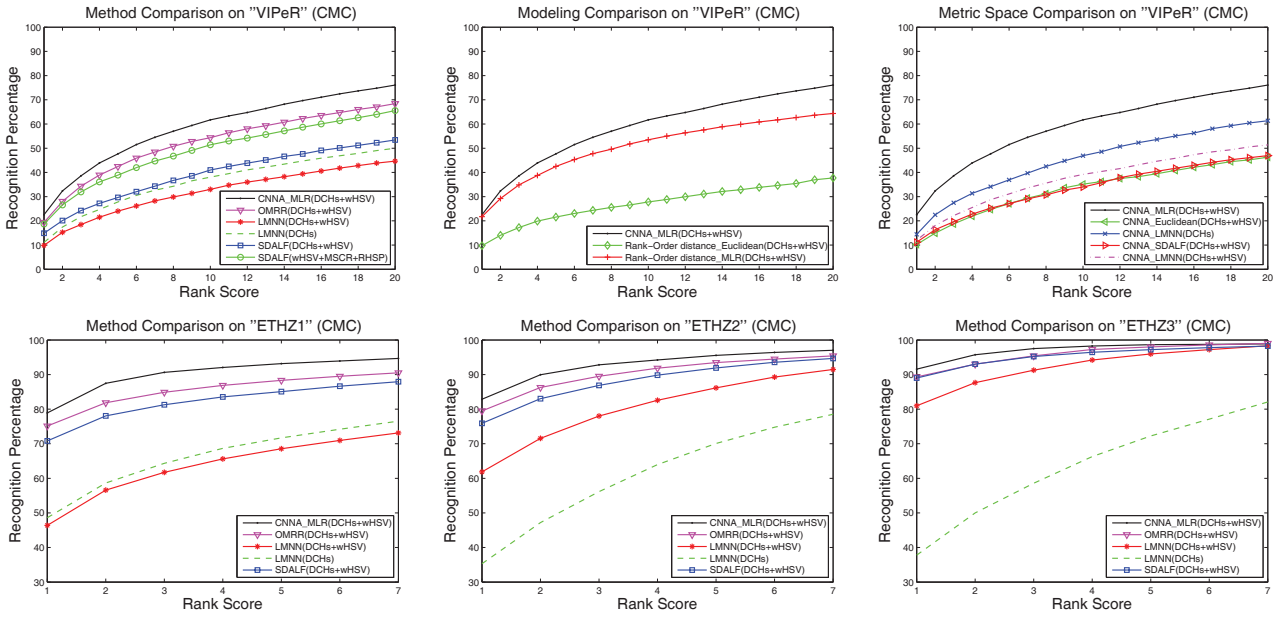

Fig. 4. CMC performance comparison on the VIPeR dataset and the ETHZ dataset.

Results: Experimental results are illustrated by the CMC (Cumulative Matching Characteristic) curves, as shown in Fig. 4. In summary, our proposed method evidently outstrips other representative methods on public available datasets VIPeR and ETHZ, even if the same feature has been applied. Our new modeling, made up of "Relative dissimilarity" and "Direct dissimilarity", prevails over Zhu's "Rank-Order distance" modeling for the issue of person re-identification. With the feature "DCHs+wHSV" and in the learned metric space constructed by MLR with loss function "MRR", our method generates currently the best performance.

4. CONCLUSION

This paper has proposed a new method *CNNA* for the issue of person re-identification. Experimental results have testified the superiority of our method to other typical state-of-the-art methods. Possible future work includes applying *CNNA* to the issues of human detection and tracking.

5. REFERENCES

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *IEEE CVPR*, Jun. 2010, pp. 2360–2367.

[2] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, Feb. 2009.

[3] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *IEEE ACCV*, Nov. 2010, pp. 501–512.

[4] Y. Wu, M. Mukunoki, T. Funatomi, and M. Minoh, "Optimizing mean reciprocal rank for person re-identification," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Aug. 2011, pp. 408–413.

[5] B. McFee and G. Lanckriet, "Metric learning to rank," in *International Conference on Machine Learning (ICML)*, Jun. 2010, pp. 775–782.

[6] C. Zhu, F. Wen, and J. Sun, "A rank-order distance based clustering algorithm for face tagging," in *IEEE CVPR*, Jun. 2011, pp. 481–488.

[7] B. Prosser, W. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *British Machine Vision Conference (BMVC)*, Aug. 2010.

[8] D. Gray and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, Oct. 2007, pp. 41–47.

[9] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, Oct. 2009, pp. 322–329.