SET-BASED CLASSIFICATION FOR PERSON RE-IDENTIFICATION UTILIZING MUTUAL-INFORMATION

Hao Liu[1], *Lei Qin*[2], *Zhongwei Cheng*[1], *Qingming Huang*[1,2]

[1]University of Chinese Academy of Sciences, Beijing 100049, China
[2]Key Lab. of Intell. Info. Process., Inst. of Comput. Tech., Chinese Academy of Sciences, China

ABSTRACT

Identifying individuals in multi-view camera network, known as person re-identification, becomes an emerging topic for video surveillance. In this paper, we address person re-identification as a set-based classification problem and introduce mutual-information to fully utilize gallery information. Firstly, we define a set-based structure that contains pairwise features between query image and gallery images. Then these features are fed into a set-class model, which exploits the relationship between set and class label (person identity) using mutual-information. Finally, we estimate and rank the mutual-information scores, and the corresponding label of the highest score is assigned to the query image. Our method has gained a superior performance compared with the state-of-the-art in the benchmark datasets i-LIDS and ETHZ.

Index Terms— Person re-identification, mutual information, video surveillance

1. INTRODUCTION

Matching people by adjacent cameras in visual surveillance scenarios, considered as person re-identification (Re-Id), has been largely attracted in computer vision. In the past few years, Re-Id remains primarily unsolved. This is due to mainly two reasons. First, person images undergo background variations in different view angles. Second, the visual appearance is not stable through the transition of multi-view cameras. For instance, a person carries with a yellow bag in one view but the bag may not be visible in another view.

In the literature, some Re-Id methods that rely on the visual information are addressed as appearance-based methods. These methods mainly focus on designing descriptive features such as low-dimensional discriminant features[1], viewpoint invariance features [2], accumulation of multiple features [3], combination of both local and global features [4], bio-inspired features [5] and fisher vector encoded features [6]. According to the verifications, the appearance-based techniques can be grouped into two categories: single-shot and multi-shot. The first group, such as [1] and [2], selects only one image for each person to build the candidate set called gallery. As to the single-shot methods, they do not work well when persons are observed in large variations across the multiple cameras. Differently, the multi-shot methods, i.e. [4, 5, 6], choose two or more images to model a person, which employs the multiple images as a signature. For the multi-shot case, the query image is matched with the different signatures (a signature represents a person with multiple images) and then the lowest distance is obtained. However, the proposed multi-shot approaches only use partially information from the images within gallery. In addition, metric learning based methods can be improved in Re-Id such as [7, 8, 9, 10]. This kind of methods aims at finding a global, linear transformation on image features so as to augment the discriminative dimensions meanwhile weaken the meaningless dimensions. For these methods, training stage is imperative and the metric should be estimated beforehand.

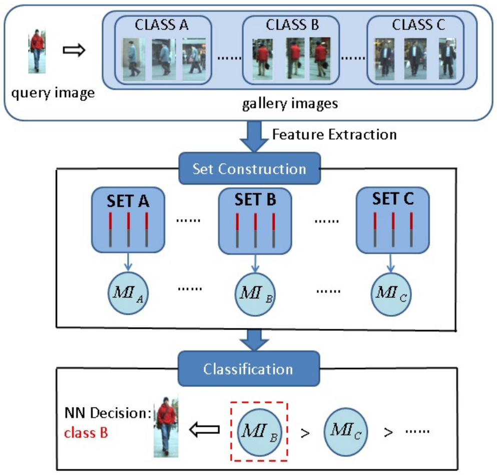

Fig. 1. The flowchart of our approach. (The strips combined with red and blue within *SET* are pairwise features and *MI* represents the mutual-information between the *SET* and the class.)

In this paper, we propose a set-based classification approach to settle Re-Id utilizing mutual-information. To overcome the limitation of the multi-shot methods, we define a set-based structure for each class called SET. For every class, each SET contains concatenated features of the query image

978-1-4799-2341-0/13/$31.00 ©2013 IEEE

feature and those from the gallery images. After constructing each SET, pairwise features are fed into a set-class model. The set-class model indicates the relationship between set-based structure and class label by mutual-information. Moreover, we use nearest neighbor method to approximate the mutual-information score for each SET. Finally, these scores are ranked in a descending order and the corresponding class label of the highest value is assigned to the query image. We evaluate the effectiveness of our approach using the benchmark datasets i-LIDS[11] and ETHZ[12]. With simple features, we obtain a better performance compared with the state-of-the-art.

Different from the existing methods upon Re-Id, the contributions of our approach are: 1) we propose a set-based classification solution to utilize the gallery information, 2) we introduce mutual-information into Re-Id to depict the relationship between SET and class label. As to our proposed approach, we have obtained a significant performance with simple and common features. In addition, we can also deal with the new class occurrence without retraining procedure unlike the learning based methods. To our knowledge, we are the first to utilize mutual-information theory to solve Re-Id problem.

The remainder of this paper is organized as follows: our new approach details are described in Sec.2. The experimental performance and results are presented in Sec.3. Finally, we draw conclusions and put forward future work.

2. THE PROPOSED APPROACH

In this section, we describe the proposed approach (See Fig.1 for a shortcut). Firstly, we construct the SET between the query image and the gallery images, and then provide an algorithm to estimate the mutual information score between the SET and the class label. Finally, the query image is assigned to the class label with the highest mutual-information score.

2.1. Modeling SET-Class Relationship

Given a query image, we want to find its class label c. We use the pairwise features to enhance the discrimination of query images for classification. To model the relationship between the query image and the gallery images, we combine the query feature $\mathbf{x}_q$ with those $\mathbf{x}_j^c$ for class c within gallery into pairwise features $\mathbf{x}_{qj}^c$, where $j \in \{1, 2, ..., N\}$, N generally represents the same size of the gallery for each class, and c is the label of the class. Such features $\mathbf{x}_{qj}^c$ constitute a SET w.r.t class label c, $\mathbf{x}_{qj}^c \in \mathbf{S}^c$. Fig.2 demonstrates how these sets are formed.

Mutual-information can measure the mutual dependence of the two random variables. Upon Re-Id, we could apply the dependence between SET and class label towards classification. From the information theory of view, we introduce the mutual-information approach to model the relationship between SET and class label. Thus, we re-formulate Re-Id as a common classification problem:

$$\widehat{c} = \arg \max_{c \in \{1,2,..,|\mathbf{C}|\}} MI\left(\mathbf{C} = c; \mathbf{S}^c\right) \tag{1}$$

where $MI(\cdot)$ denotes the mutual-information between SET and class label, $\mathbf{S}^c$ denotes the SET for class label c and $|\mathbf{C}|$ is the number of classes.

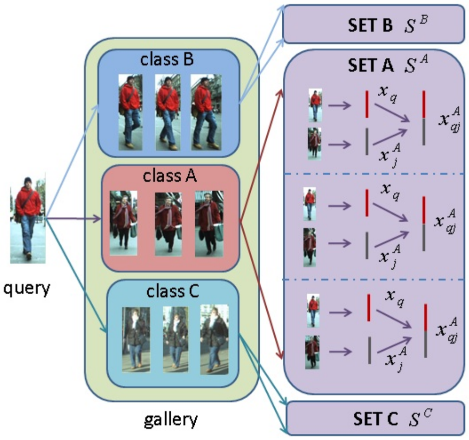

Fig. 2. The schematic diagram of SET construction: given a query image, it should be paired with all images in gallery to form a structure defined SET for each class

2.2. The approximating method for mutual-information

Assuming the independence of each pairs within one SET, we define the mutual-information formulation in a probability form as:

$$MI\left(\mathbf{C} = c; \mathbf{S}^c\right) = \sum_{\mathbf{x}_{qj} \in \mathbf{S}^c} \log \frac{P\left(\mathbf{x}_{qj}, \mathbf{C} = c\right)}{P\left(\mathbf{x}_{qj}\right) P\left(\mathbf{C} = c\right)} \tag{2}$$

Based on the conditional probability, we obtain a further derivation as:

$$\begin{aligned} \log \frac{P\left(\mathbf{x}_{qj}, \mathbf{C} = c\right)}{P\left(\mathbf{x}_{qj}\right) P\left(\mathbf{C} = c\right)} &= \log \frac{P\left(\mathbf{x}_{qj} | \mathbf{C} = c\right)}{P\left(\mathbf{x}_{qj}\right)} \\ &= \log \frac{1}{P\left(\mathbf{C} = c\right) + \dfrac{P\left(\mathbf{x}_{qj} | \mathbf{C} \neq c\right)}{P\left(\mathbf{x}_{qj} | \mathbf{C} = c\right)} P\left(\mathbf{C} \neq c\right)} \end{aligned} \tag{3}$$

where $\dfrac{P\left(\mathbf{x}_{qj} | \mathbf{C} \neq c\right)}{P\left(\mathbf{x}_{qj} | \mathbf{C} = c\right)}$ is considered as the likelihood ratio test.

In common sense, the mutual-information described above can be defined over the probability densities. However, we will not directly access mutual-information but use an alternative algorithm to approximate it. From Equation(3), it is crucial to evaluate the likelihood ratio with a known prior for the probability of class c. We calculate the likelihood ratio by using the NBNN algorithm proposed in [13]. The NBNN algorithm applies the very simple form $\log P\left(\mathbf{x}_{qj}|\mathbf{C}\right) \propto -\|\mathbf{x}_{qj} - \mathbf{x}^{c}\|^{2}$ to approximate the Gaussian kernel without dependence on the variance. Thus, the likelihood ratio item can be estimated as:

$$\frac{p\left(\mathbf{x}_{qj}|\mathbf{C} \neq c\right)}{p\left(\mathbf{x}_{qj}|\mathbf{C} = c\right)} \propto \exp^{-\left(\|\mathbf{x}_{qj}-\mathbf{x}^{c}_{NN-}\|^{2}-\|\mathbf{x}_{qj}-\mathbf{x}^{c}_{NN+}\|^{2}\right)} \quad (4)$$

here $\mathbf{x}^{c}_{NN-}$, $\mathbf{x}^{c}_{NN+}$ are the nearest neighbors of $\mathbf{x}_{qj}$ in target pairs set and non-target pairs set, respectively. As demonstrated, target pairs are pairwise combinations in the same class c, whereas non-target pairs are combinations for different classes.

As Equation(4) shows, the value of the likelihood ratio is determined by the distance between the query pair $\mathbf{x}_{qj}$ and the reference sets (target pairs and non-target pairs). The ratio value is considered into classification, i.e. if the query pair $\mathbf{x}_{qj}$ is closer to one of the samples in target pair set, that is, it is far from the non-target pairs set, the relevance proves that $\mathbf{x}_{qj}$ is more likely to belong to class c. Furthermore, the query pair $\mathbf{x}_{qj}$ indicates a low likelihood ratio and high mutual-information score. On the contrary, if $\mathbf{x}_{qj}$ is closer to samples in non-target pairs set, it is likely that it does not belong to the class c.

Based on the above derivation, we obtain the ultimate form of the mutual-information as:

$$MI\left(\mathbf{C} = c; \mathbf{S}^{c}\right) \approx \sum_{\mathbf{S}^{c}} \log \frac{|C|}{1 + \exp^{-\omega\left(\mathbf{x}_{qj}\right)}\left(|C| - 1\right)} \quad (5)$$

where $\omega\left(\mathbf{x}_{qj}\right)$ denotes $\|\mathbf{x}_{qj}-\mathbf{x}^{NN-}_{ij}\|^{2}-\|\mathbf{x}_{qj}-\mathbf{x}^{NN+}_{ij}\|^{2}$, and $P(\mathbf{C} = c) = \frac{1}{|C|}$, $P(\mathbf{C} \neq c) = 1 - \frac{1}{|C|}$ ($|C|$ is the number of classes). For each SET, the final decision of classification is determined by the highest one of all the mutual-information scores. After constructing the sets, the query features have been augmented for classification, because each query pair $\mathbf{x}^{c}_{qj}$ within one SET can provide a positive or negative vote for class c. From Equation(5), if the mutual-information score is positive, it is indicated that SET $\mathbf{S}^{c}$ votes a positive score for class c. According to this, the query image is considered to belong to class c with a high probability. Otherwise, when the mutual-information is negative, the query image seems not belonging to class c. As to the analysis, the sign of the mutual-information suffers the hardship that interference of the negative pairs should be considered. Thus, we make a decision by ranking all of the mutual-information scores instead. Finally, the query is classified to the class c based on the highest score.

3. EXPERIMENTS

In this section, we show our experimental results on multi-shot datasets i-LIDS and ETHZ. The challenging aspects of the two datasets are illumination changes, image blurring, low resolution, and occlusions. For evaluation, we use the standard measurement named Cumulative Match Characteristic (CMC) curve, which exploits correct matches (vertical coordinates) at ranking top k (horizontal coordinates).

3.1. Settings and feature representation

In our approach, we randomly select N images for each class to build gallery and the remaining images are used for testing. Different from both the appearance based methods and ours, the learning based approaches, such as [7, 9, 10, 14], put one part of persons into training and the other part is used for testing. We repeat the whole experiment with 10 times and average the results of CMC. We compared with the state-of-the-art [1, 3, 5, 6] under non-learning multi-shot setting, and the metric learning methods are not considered. The feature representation was extracted by Zheng et al in [7, 15]. These images in both datasets are not in the same size, therefore, we normalize them to the same size of 128×64 pixels. Each image is divided into 6 horizontal stripes. For each stripe, histogram features are extracted by 8 color (RGB, YCbCr, HSV) channels and 2 kinds of texture (Schmid and Gabor). A single feature is represented by a 2784 dimensional vector and we reduce the high dimension to 150 by PCA.

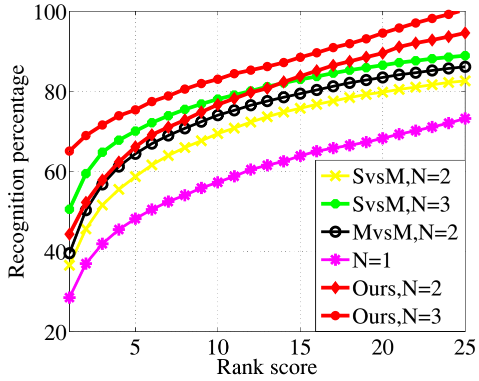

Fig. 4. CMC performance on dataset i-LIDS compared to the SDALF[3].

3.2. Evaluation on i-LIDS

The i-LIDS dataset contains 119 persons and 476 images in all. All images are captured by multiple non-overlapping cameras at a busy airport. We compare our approach with the popular method SDALF[3] in several settings such as SvsM (single test image and multiple gallery images), MvsM (equivalent multiple images for test and gallery), and more details can be found in the relevant reference.

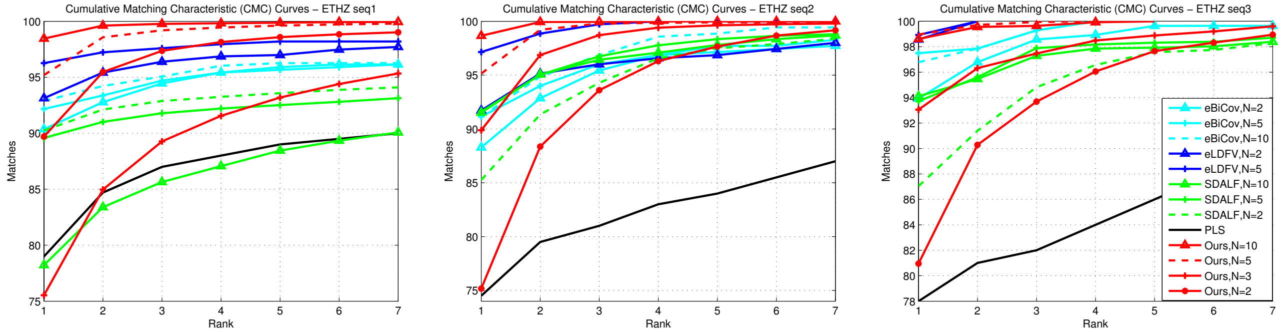

Fig. 3. CMC performance on the dataset ethz compared with the state-of-the-art methods.

For the dataset has averaging 3-4 instances for each person, we set N to {2, 3} as [3]. Fig.4 shows the final comparisons under multi-shot case. We can infer from the figure that the MvsM setting outperforms the SvsM for the additional test candidates, however, when add more images to gallery, the performance can be further improved even than MvsM. Compared to SDALF, our method has outperformed to the SvsM setting when N is assigned to 2. We also test our algorithm by N=3, and the performance is improved and even better than SvsM in SDALF when N=3. This can be explained that adding a second instance will provide more information, which validates the comparable results between SvsM and MvsM in SDALF. Although the dataset undergoes some occlusions and quite crowded scenes, we conquer the hinders and gain a evident performance.

3.3. Evaluation on ETHZ

ETHZ dataset contains 3 video sequences captured by a moving camera in a busy street. The dataset consists of 146 people and 8555 images in all, including SEQ.#1 with 4857 images for 83 persons; SEQ.#2 with 35 persons in 1936 images; SEQ.#3 with 28 persons in 1762 images. We compare our approach with the multi-shot methods including PLS[1], SDALF[3], eBiCov[5] and eLDFV[6].

We set N to {2,3,5,10} for the multi-shot cases as i-LIDS. The results are reported in Fig.3. For all the sequences, when setting N to 2, we perform better performance than the baseline PLS[1]. However, when N is assigned to {3,5} for each person, we have obtained a big improvement. According to the results, 5 images is enough to achieve in outperforming. While N reaches 10, we perform the highest performance compared with the state-of-the-art. Especially, our CMC curve is close to 100% after rank 3 with N=10. For SEQ.#3, we have a comparable performance because most images are similar for each person with small variation in different view angels. As we know, [6] exhibits the best performance on ETHZ dataset in the literature for it utilizes the powerful fisher vector representation. Our approach uses the simple HSV Histogram and we obtain an comparable even better performance than [6], which shows the effectiveness of our proposed mutual-information based approach. As our approach proposed, multiple images can cover more appearance variations and gallery images provide the total information for the classification.

4. CONCLUSION

In this paper, we re-formulate Re-Id as a set-based classification problem. Specifically, we define set-based structure between the query image and the gallery images, and exploit the relationship between set and the class label by mutual-information. And then the classification according to the mutual-information scores is achieved. Our preferable performances on the popular datasets are reasonable as we make full use of the association between query image and gallery images. However, our approach encounters high computational cost problem since we have to compute all positive and negative pairwise neighbors for every class. In the future, we will optimize the algorithm and extend our method with a proper metric by metric learning method.

5. ACKNOWLEDGMENT

This work was supported in part by National Basic Research Program China (973 Program): 2012CB316400, in part by National Natural Science Foundation of China: 61025011, 61035001, 61003165 and 61133003, and in part by Beijing Natural Science Foundation: 4111003

6. REFERENCES

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

[2] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *Computer Vision–ECCV 2008*, pp. 262–275, 2008.

[3] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2360–2367.

[4] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-shot person re-identification by hpe signature," in *Proc. ICPR*, 2010, pp. 1413–1416.

[5] B. Ma, Y. Su, and F. Jurie, "Bicov: a novel image representation for person re-identification and face verification," in *British Machine Vision Conference*, 2012.

[6] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *International Workshop conjunction with European Conference on Computer Vision*, 2012, pp. 413–422.

[7] W.S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 649–656.

[8] W.S. Zheng, S. Gong, and T. Xiang, "Transfer re-identification: From person to set-based verification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2650–2657.

[9] M. Hirzer, P. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," *Computer Vision–ECCV 2012*, pp. 780–793, 2012.

[10] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2288–2295.

[11] U. H. Office, "i-lids multiple camera tracking scenario definition," 2008.

[12] B-Leibe A.Ess and L.V.Gool, "Depth and appearance for mobile scene analysis," in *International Conference on Computer Vision*. IEEE, 2007.

[13] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[14] Michael Lindenbaum Tamar Avraham, Ilya Gurvich and Shaul Markovitch, "Learning implicit transfer for person re-identification," in *1st International Workshop on Re-Identification (Re-Id 2012) In conjunction with ECCV 2012*, 2012, p. 381C390.

[15] B. Prosser, W.S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *British Machine Vision Conference*, 2010, vol. 10.