IEICE TRANS. INF. & SYST., VOL.E96–D, NO.11 NOVEMBER 2013

PAPER

Bi-level Relative Information Analysis for Multiple-Shot Person Re-Identification

Wei LI[†a)], **Yang WU**[††], ***Nonmembers***, **Masayuki MUKUNOKI**[††], ***Member***, ***and*** **Michihiko MINOH**[††], ***Fellow***

SUMMARY Multiple-shot person re-identification, which is valuable for application in visual surveillance, tackles the problem of building the correspondence between images of the same person from different cameras. It is challenging because of the large within-class variations due to the changeable body appearance and environment and the small between-class differences arising from the possibly similar body shape and clothes style. A novel method named "Bi-level Relative Information Analysis" is proposed in this paper for the issue by treating it as a set-based ranking problem. It creatively designs a relative dissimilarity using set-level neighborhood information, called "Set-level Common-Near-Neighbor Modeling", complementary to the sample-level relative feature "Third-Party Collaborative Representation" which has recently been proven to be quite effective for multiple-shot person re-identification. Experiments implemented on several public benchmark datasets show significant improvements over state-of-the-art methods.

key words: *Bi-level relative information analysis; multiple-shot person re-identification; visual surveillance*

1. Introduction

1.1 Background

Multiple-shot person re-identification tackles the problem of judging the re-appearance of the person by using sequential images acquired from distributed cameras. The difference between multiple-shot and single-shot person re-identification is whether spatial-temporal information on appearance cues can be used or not. Such inter-camera multiple-shot correspondence will be beneficial for tracking across cameras, but how to build the correct correspondence remains one of the most challenging issues in visual surveillance. The challenge primarily originates from both large within-class variations and small between-class differences, caused by pose varying, illumination changing, viewpoint altering, occlusion, body shape resemblance, clothes style similarity, and so forth. These difficulties are unavoidably mixed together. Most state-of-the-art methods incline to tackle them at the same time. For multiple-shot person re-identification, current approaches can be categorized into two main paradigms. The first paradigm attaches importance to reliable feature/signature designing or selecting. One typical method is Histogram Plus Epitome [1]. It focuses on the presence of overall chromatic content via histogram representation and recurrent local patches via epitomic analysis to effectively extract the complementary global and local features from human appearance. Another representative method is Haar-based and DCD-based Signature [2]. It takes advantage of the AdaBoost scheme to build a satisfactorily invariant and descriptive signature based on haar-like features and dominant color descriptors for each person. The second paradigm pays attention to robust dissimilarity/distance crafting or learning after feature/signature representation. One popular method is Mean Riemannian Covariance Grid (MRCG) [3]. It not only uses essential cues about spatial-temporal changes of the person's appearance by the Karcher mean based covariance grids but also crafts a suitable dissimilarity in Riemannian space for them. Another exemplary method is Set Based Discriminative Ranking (SBDR) [4]. It treats multiple-shot images per person from different cameras as one set, and then iteratively constructs the convex hulls for these sets and learns the discriminative set-to-set distance metric between these hulls.

1.2 Related Work

Most existing approaches exploit direct information for representation or measure, whereas, relative information is rarely considered. Currently, two novel methods have achieved remarkable results. They show the essence of relative information from different perspectives. One method is "Third-Party Collaborative Representation" (TPCR) [5], which focuses on the relative feature design towards the problem of multiple-shot person re-identification. TPCR resorts to the third-party data as more capable dictionary to encode the reconstructed coefficients of collaborative representation into a discriminative feature. This feature considers the linear combination relationship of each sample relatively with the third-party samples as words in the dictionary, thus it is different from traditional features/signatures. The other method is "Common-Near-Neighbor Analysis" (CNNA) [6], which concentrates on exploiting a relative dissimilarity modeling in a learned metric space towards the problem of single-shot person re-identification. CNNA makes use of the Rank-Order lists of each sample pair to form a reliable dissimilarity between them. This dissimilarity measures between each sample pair relatively by processing their neighborhood information, thus it is different from traditional dissimilarities/distances. Technically in-

Manuscript received February 12, 2013.
Manuscript revised June 20, 2013.

[†]The author is with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606–8501 Japan.
[††]The authors are with Academic Center for Computing and Media Studies, Kyoto University, Kyoto-shi, 606–8501 Japan.
a) E-mail: liwei@mm.media.kyoto-u.ac.jp

DOI: 10.1587/transinf.E96.D.2450

spired by the newly proposed TPCR and CNNA, we propose a novel idea named "Bi-level Relative Information Analysis" (BRIA) for multiple-shot person re-identification. It will integrate the advantages of two levels of relative information that are the sample-level relative feature instantiated by TPCR and the new set-level relative dissimilarity Set-level Common-Near-Neighbor Modeling (SCNNM) to be presented.

2. Problem Definition and Overview of Our Approach

This paper works on multiple-shot person re-identification, which attempts to build the correspondence between the person images obtained from non-overlapping cameras. We treat the multiple-shot images for each person in terms of set, and reformulate the issue into a set-based ranking problem.

In BRIA, two levels of relative information will be synthesized to handle the difficulty of the issue. As shown in Fig. 1, the person image will be represented by the sample-level relative feature from the source space, which will complement and be measured by the set-level relative dissimilarity in the target space. Collaborating two levels of relative information provides an effective solution to the problem without loss of methodological generality and elegance.

More specifically, BRIA is composed of two components: the relative feature TPCR originating from Collaborative Representation Classification (CRC) [7] and the relative dissimilarity SCNNM stemming from CNNA. In order to bring the best performance, our proposed method BRIA will take advantage of the complementarity between TPCR and SCNNM.

In procedure, firstly, BRIA extracts TPCR feature for each sample over the third-party data; then, since a group of TPCR features from the same class are treated within one set, set-to-set dissimilarities are measured by SCNNM; finally, set-based ranking is carried out according to these dissimilarities.

This paper is based on our accepted international conference papers [5], [6]. Even so, there are three obvious differences between this paper and our previous works: (1) SCNNM extends the scope of application for CNNA from the single-shot re-identification case to the multiple-shot issue. It enhances the reliability of the set-to-set dissimilarity by creatively exploring the set-level relative information; (2) BRIA takes advantage of the complementarity between the TPCR feature and the SCNNM dissimilarity for a good collaboration between them. It effectively overcomes the sensitivity of traditional set-to-set distances and simultaneously reduces the subjectiveness of TPCR feature; (3) extensive experiments on widely-used benchmark datasets show the substantial superiority of BRIA to state-of-the-art methods.

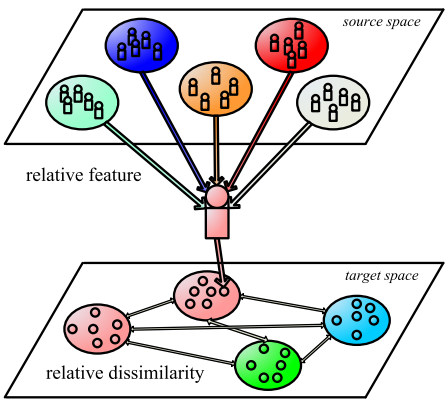


Fig. 1 Illustration of BRIA. Each feature point will be relatively represented by the third-party samples in the source space, and all set-to-set dissimilarities will be relatively measured in the target space.

3. Solution Statement and Analysis

3.1 Third-Party Collaborative Representation

In our early work, TPCR gains large performance enhancement for multiple-shot person re-identification. TPCR relies on the reconstructed coefficients from collaborative representation, which evolves from sparse representation. Thus, in order to explain TPCR, it is necessary to mention Sparse Representation for Classification (SRC) [8] and CRC at first.

SRC [8] has attracted many researchers due to its simultaneous effectiveness and efficiency for recognition tasks. SRC firstly codes a query as a sparse linear combination of all corpus. After that, it classifies the queries by judging which class leads to the minimum representation error. Suppose there is a dataset $D \in \mathbb{R}^{d\times n}$ composed of n samples with d dimension in K classes, and $X \subset D$ are corpus which will be used as the dictionary, where $X = [X_1, X_2, \ldots, X_K] \in \mathbb{R}^{d\times l}$, in which l is the number of all corpus samples, and $X_i (i = 1, \ldots, K)$ are the corpus each class. When a query $q \in (D - X)$ comes, where $q \in \mathbb{R}^{d\times 1}$, q can be classified by searching a sparse representation using the dictionary X, and finding the class $y \in \{1, \ldots, K\}$ which can best approximate q by the linear combination of its samples with their corresponding sparse coefficients. Typically, SRC solves the "l_1-regression with l_2-constraint" problem:

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \tag{1}$$

$$s.t.\ \ \|q - X\alpha\|_2^2 < \varepsilon, \tag{2}$$

where $\hat{\alpha} = [\hat{\alpha}_1^T, \ldots, \hat{\alpha}_K^T]^T$; $\hat{\alpha}$ is a column vector, consisting of $\hat{\alpha}_i$, which is the coding vector associated with class i; constant ε is used to balance the coding error of q and the sparsity of α. With $\hat{\alpha}_i$, we can judge the identity of q by $y(q) = \arg\min_i \|q - X_i\hat{\alpha}_i\|_2^2$, where $\forall i \in \{1, \ldots, K\}$.

There are two parts in SRC model: the sparsity part in Eq. (1) and the collaborative representation part in Eq. (2). For classification, minimization of the collaborative representation part of SRC, namely CRC, has been proved more effective than SRC itself, especially for the well-controlled face recognition problem [7]. CRC is formulated to solve the "l_2-regression with l_2-regularization" problem:

$$\hat{\alpha} = \arg\min_{\alpha}\{\|q - X\alpha\|_2^2 + \mu\|\alpha\|_2^2\}, \tag{3}$$

where μ is a trade-off parameter. Equation (3) has a closed-form solution $\hat{\alpha} = Pq$. $P = \left(X^T X + \mu \cdot I\right)^{-1} X^T$, where I denotes the identity matrix. Note that P can be pre-computed once the dictionary X has been given. Collaborative representation inclines to use a few words in the dictionary to represent each sample.

Essentially, for each sample, TPCR algorithm compacts a kind of relative information referring to the words in the dictionary into a feature vector. Those words with large weights tend to characterizes a kind of neighborhood information on the sample level. The algorithm of TPCR is detailed in Algorithm 1.

Algorithm 1 Third-Party Collaborative Representation (TPCR):

Require: The dataset $D \in \mathbb{R}^{d \times n}$ of corpus and queries; the third-party dataset $D_{tp} = [D_{tp}^1, \ldots, D_{tp}^L] \in \mathbb{R}^{d \times m}$ of L classes; the regularization parameter μ.

Ensure: A collaborative representation based description $\hat{\beta}'(s)$ for each sample $s \in D$ over D_{tp}.

1: Normalize the columns of D and D_{tp} to have unit l_2-norm.

2: Solve the "l_2-regression with l_2-regularization" problem:
$\hat{\alpha} = \arg\min_\alpha \{\left\|s - D_{tp}\alpha\right\|_2^2 + \mu\|\alpha\|_2^2\}$,
with a closed-form solution $\hat{\alpha} = P_{tp}s$, where $P_{tp} = \left(D_{tp}^T D_{tp} + \mu \cdot I\right)^{-1} D_{tp}^T$. Note that P_{tp} can be pre-computed once D_{tp} is given.

3: Compute the summed coefficients within each class:
$\hat{\beta}_i(s) = \sum_{j=1}^{n_i} \hat{\alpha}_{ij}, \forall i \in \{1, \ldots, L\}$, where n_i is the sample number of class i in the third-party data.

4: Normalize $\hat{\beta}(s)$: $\hat{\beta}'(s) = \hat{\beta}(s)/\sum_{i=1}^{L} \left|\hat{\beta}_i(s)\right|$, and return $\hat{\beta}'(s)$.

TPCR initially introduces the third-party data to enhance the descriptive and representative power of the dictionary, and further concatenates the reconstructed coefficients with intra-class sum pooling into a feature vector. Attributing to the usage of third-party data as the dictionary, TPCR does not need extensive training samples for each testing class. Its discriminative power is explored from the abundant third-party datasets which can be different from the training and testing datasets. Therefore, it enables using an existing dictionary for testing new data without time-consuming data annotation and model re-training. The third-party data dictionary covers enough information of pose altering, illumination varying, viewpoint changing, and localization errors. Owing to intra-class sum of reconstructed coefficients benefited from such dictionary in the process of vectorization, TPCR is robust to localization errors and large within-class variations so that it is applicable to real-world person re-identification tasks. As expected, the performance of TPCR evidently outstrips original features under traditional set-to-set distances [5].

3.2 Common-Near-Neighbor Modeling

In this paper, we recast the person re-identification problem into a set-based ranking problem that depends on set-to-set dissimilarity measured by SCNNM. SCNNM is extended from CNNA. Accordingly, in order to present SCNNM, it is necessary to describe CNNA beforehand.

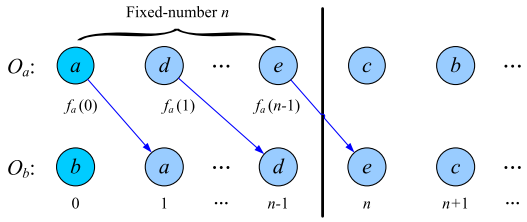


Fig. 2 O_a and O_b are two Rank-Order lists. Samples are denoted by a, b, c, d, e, and so on. $D^{\text{Fixed-number}}(a, b)$ is calculated from the 0^{th} to the $(n-1)^{th}$ nearest neighbor sample in a's Rank-Order list.

In our previous work, CNNA has been proposed to deal with the single-shot person re-identification problem [6]. Based on the assumption that most samples belonging to the same class will share more common near neighbors than those from different classes in a learned metric space [6], CNNA explores this kind of information to further make intra-class dissimilarities smaller than inter-class dissimilarities for all samples. Basically, CNNA contains two parts: Metric Learning and Common-Near-Neighbor Modeling (CNNM). As the core part, CNNM is composed of the symmetric dissimilarity and the asymmetric dissimilarity, thus can be expressed as:

$$D^{\text{CNNM}}(a, b) = D^{\text{Symmetric}}(a, b) + 2\lambda n D^{\text{Asymmetric}}(a, b), \tag{4}$$

where a and b are two samples; λ is the balancing parameter between the symmetric and asymmetric dissimilarities; n is the "Fixed-number".

As the symmetric dissimilarity of CNNM, $D^{\text{Symmetric}}(a, b)$ is given by:

$$D^{\text{Symmetric}}(a, b) = D^{\text{Fixed-number}}(a, b) + D^{\text{Fixed-number}}(b, a), \tag{5}$$

where

$$D^{\text{Fixed-number}}(a, b) = \sum_{i=0}^{n-1} O_b(f_a(i)), \tag{6}$$

$f_a(i)$ returns the i^{th} element in a's Rank-Order list O_a, where Rank-Order list of an assigned sample consists of the ranked sequence of all samples according to their distances to this sample; $D^{\text{Fixed-number}}(a, b)$ sums the rank orders of $f_a(i)$ over i in b's Rank-Order list under the setting of n, and $D^{\text{Fixed-number}}(b, a)$ is calculated in a similar way, as shown in Fig. 2.

As the asymmetric dissimilarity of CNNM, $D^{\text{Asymmetric}}(a, b)$ is given by:

$$D^{\text{Asymmetric}}(a, b) = min(O_a(b), O_b(a)), \tag{7}$$

where $O_b(a)$ is the rank order of a in b's Rank-Order list O_b, and $O_a(b)$ is defined in a similar way.

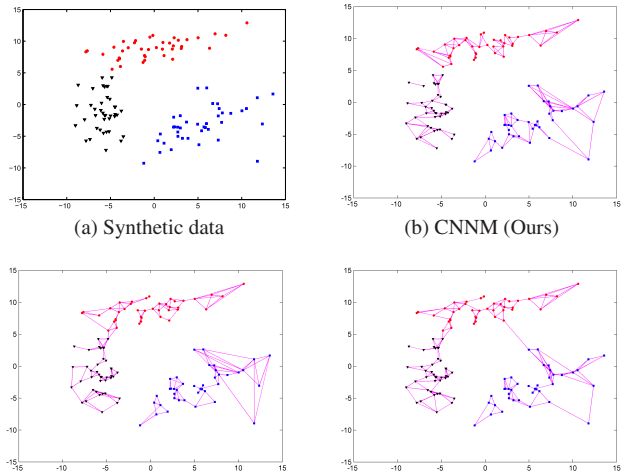
The procedure of CNNA is presented in Algorithm 2.

Algorithm 2 COMMON-NEAR-NEIGHBOR ANALYSIS (CNNA)

Require: Training samples x_ts and their labels l_ts; corpus x_cs and queries x_qs as the testing samples.

Ensure: Ranking y_q of all x_cs w.r.t. each x_q.

1: In the training stage, metric learning is performed on (x_ts,l_ts) to construct a ranking-optimized metric space $\mathcal{M}$.
2: In the testing stage, all x_qs and all x_cs are projected into $\mathcal{M}$ as $\mathbf{x}_q$s and $\mathbf{x}_c$s.
3: All $\mathbf{x}_q$s and all $\mathbf{x}_c$s are randomly mixed together into a sample list $\mathbf{x}$.
4: All samples of $\mathbf{x}$ are sorted according to their Euclidean distance to each $\mathbf{x}_q$ and each $\mathbf{x}_c$, respectively, to obtain the sample-level Rank-Order lists.
5: CNNM is performed based on the sample-level Rank-Order lists of each pair of ($\mathbf{x}_q$, $\mathbf{x}_c$) to measure the dissimilarity between them.
6: For each $\mathbf{x}_q$, according to the dissimilarities between itself and all $\mathbf{x}_c$s calculated in step 5, all $\mathbf{x}_c$s are re-ranked to return the result y_q.



(a) Synthetic data (b) CNNM (Ours)

(c) Rank-Order distance (Zhu et al.'s) (d) Euclidean distance

Fig. 3 Synthetic data are generated randomly to test the capability of CNNM. Classes are labeled by distinct colors. Magenta lines are used to connect queries and corpus which are ranked first w.r.t. these queries.

3.3 Bi-level Relative Information Analysis

3.3.1 Merits of CNNM

There are several merits of CNNM. Firstly, CNNM creatively uses rank orders to calculate the dissimilarity between samples, which can be regarded as a kind of quantized distance. This quantized distance can overcome the non-uniform sample distribution problem that may impair the effectiveness of ranking; secondly, the symmetric dissimilarity concerns the "Fixed-number" of nearest neighbors for each pair of samples other than themselves by means of summing the n rank orders symmetrically in both Rank-Order lists. Summing offers robustness due to its statistical averaging effect on the neighborhood information of the sample pair; last but not least, the asymmetric ranking problem is oftentimes obvious when the class size is small. This means a given pair of samples usually don't have the same rank order for each other in their own Rank-Order lists. Randomly considering one side of them to determine the rank order is too heuristic and unfair. This problem can be tackled with the asymmetric dissimilarity, which selects the smallest rank order. The cooperation between the symmetric and asymmetric dissimilarity makes CNNM more flexible and reliable.

Furthermore, for a fair comparison with Zhu et al.'s Rank-Order distance [9] on the clustering ability derived from the ranking results, we also provide the evaluation for CNNM on the synthetic data, similar to those in Zhu et al.'s work. We generate three different Gaussian-distributed data randomly as samples from three different classes (with class size 40). For fairness, queries and corpus from the data are assigned by random half splitting for ten times independently. We connect each query point to the corpus point which is ranked first w.r.t. this query point measured by CNNM, for which n is set as half of the average class size and λ is set to 1 tentatively. For conviction, we also compare CNNM with Euclidean distance, by using the same generated data and the same query-corpus splitting in each time. All the lines visualized in Fig. 3 are the accumulative results for ten times. From them, we can clarify that CNNM performs best for its results have the least false lines.

3.3.2 Disadvantages of Traditional Set-to-Set Distances

TPCR feature provides a kind of relative information on the sample level referring to the words represented by the basic features in the dictionary. The subjectiveness in the basic feature representation and dictionary building will inevitably lead to the limited discriminative power of the TPCR feature space, in which intra-class dissimilarities may still be larger than inter-class dissimilarities for some samples. Thus, traditional set-to-set distances such as Minimum Point-wise Euclidean Distance (MPD) [10] and Convex Hull based Image Set Distance (CHISD) [11] in the TPCR feature space are still far from being perfect. Though being impressive, either MPD or CHISD with TPCR has weakness due to the fact that they rely on the measure between only some local parts of the sets. More concretely, MPD depends on the nearest samples between the sets. In this case, outliers of each class may easily influence the measuring reliability. CHISD tries to improve it by considering the distance between convex hulls for the set pair, however, it is unavoidably influenced by the layout of nearest samples between the sets which support the convex hulls. The illustration of MPD and CHISD is shown in Fig. 4. Such sensitivity may easily cause the asymmetric ranking, which means, a pair of sets usually don't have the same rank order for each other in their own set-level Rank- Order lists. Thus, it is unfair to judge the rank order only considering one side of them.

3.3.3 Set-Level Common-Near-Neighbor Modeling

CNNM has been proven to be effective for the issue of person re-identification [6]. However, this method operates on the sample level and is designed for the target of single-shot person re-identification, which is much different from the

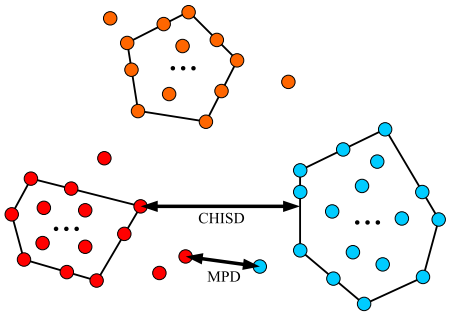


Fig. 4 Illustration of MPD and CHISD. For the set pair, MPD considers the minimum distance between points, while CHISD concerns the minimum distance between convex hulls.

multiple-shot case [10]. Though it's possible to directly apply it to multiple-shot problems, it is undesirable to do so. If we transform the multiple-shot problem into a single-shot problem to solve, the efficiency will be low. Because the dissimilarity between each pair of samples is required to measure in this case, when the sample number increases in each set, the computation will be combinatorial explosion. Furthermore, CNNM explores the relative information for every sample pair, so it can neither reflect the within-set variations as a whole, nor maintain the robustness to the noisy outliers for the set, nor accord with the evaluation criterion of multiple-shot person re-identification [10]. Thus, the effectiveness will be low as well.

Based on the robust CHISD and inspired by the recently presented CNNM, we propose a new set-to-set dissimilarity called "Set-level Common-Near-Neighbor Modeling (SCNNM)" to explore the relative information among sets instead of samples towards the multiple-shot person re-identification problem. When most sets of the same class stay closer to each other than those from different classes, the sets within the same class will share more common-near-neighbor sets than those from different classes. SCNNM utilizes such kind of information to further ensure inter-class dissimilarities to be larger than intra-class dissimilarities for all sets instead of samples.

Similar to CNNM, SCNNM dissimilarity is given by:

$$H^{\text{SCNNM}}(A,B) = H^{\text{Symmetric}}(A,B) + 2\Lambda N H^{\text{Asymmetric}}(A,B), \tag{8}$$

where A and B are two sets; Λ is the balancing parameter; N is the "Fixed-number".

The symmetric term $H^{\text{Symmetric}}(A,B)$ is as below:

$$H^{\text{Symmetric}}(A,B) = H^{\text{Fixed-number}}(A,B) + H^{\text{Fixed-number}}(B,A), \tag{9}$$

where

$$H^{\text{Fixed-number}}(A,B) = \sum_{i=0}^{N-1} O_B(F_A(i)), \tag{10}$$

$F_A(i)$ returns the i^{th} element in A's Rank-Order list O_A.

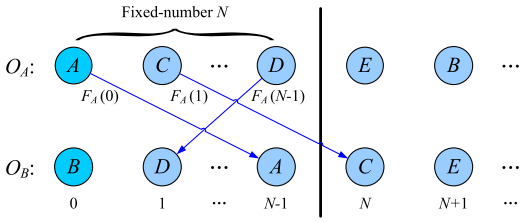


Fig. 5 Sets are denoted by A, B, C, D, E, and so on. $H^{\text{Fixed-number}}(A,B)$ is calculated from the 0^{th} to the $(N-1)^{th}$ nearest neighbor set in A's Rank-Order list.

Here, set-level Rank-Order list of an assigned set is formed by the ranking of all the other sets according to their set-to-set distance to this set. $H^{\text{Fixed-number}}(A,B)$ sums the rank orders of $F_A(i)$ over i in B's Rank-Order list under the setting of N, and $H^{\text{Fixed-number}}(B,A)$ is calculated in a similar way, as shown in Fig. 5.

The asymmetric term $H^{\text{Asymmetric}}(A,B)$ is as below:

$$H^{\text{Asymmetric}}(A,B) = min(O_A(B), O_B(A)), \tag{11}$$

where $O_B(A)$ is the rank order of A in B's Rank-Order list O_B, and $O_A(B)$ is defined similarly.

Undoubtedly, both symmetric term and asymmetric term in SCNNM encode a kind of relative information between the neighborhood of each set pair expressed by Rank-Order lists.

"Fixed-number" N is an important tunable parameter in SCNNM, which may influence the symmetric term, so it deserves in-depth explanation. N describes the neighborhood size concerned by the symmetric term. If the neighborhood size is too large, set pair from different classes may share many common near neighbors for the top "Fixed-number" elements in both Rank-Order lists, thus, the symmetric term will be reduced for sets from different classes w.r.t. the sets within the same class; if the neighborhood size is too small, set pair in the same class may share few common near neighbors for the top "Fixed-number" elements in both Rank-Order lists, then, the symmetric term will be enlarged for the sets in the same class w.r.t. the sets from different classes. Obviously, both cases have negative influence on dissimilarity-based ranking, thus should be avoided. In order to measure by a robust symmetric term, it is reasonable to propose the choice of "Fixed-number" N to be approximate half of the average set number in each class in a compromise, in case the neighborhood size is too large or too small, though we cannot give strict mathematical proof.

From the efficiency perspective, SCNNM treats the samples in the same class as one whole set, so it is much more computationally efficient than CNNM, especially when there are multiple-shot images in each set. From the effectiveness perspective, SCNNM inherits all the merits of CNNM, and develops them to the set level. The symmetric term can be treated as the robust quantized set-to-set distance. The asymmetric term can tackle the set-level asymmetric ranking problem, and a balance between the symmetric term and asymmetric term provides more flexibility and

reliability for SCNNM [12].

3.3.4 Collaboration of TPCR and SCNNM

SCNNM dissimilarity overcomes the weakness of sensitivity for MPD and CHISD. This modeling strategy incorporates the relative information on the set level, which complements the relative feature TPCR on the sample level. Such complementarity can be understood from two aspects. Firstly, sample-level and set-level information are simultaneously considered; secondly, the dissimilarity measurement is suitable for the feature representation, whereby the subjectiveness of TPCR is indirectly reduced, thus leading up to a remarkable performance for BRIA. The algorithm of BRIA is formulated in Algorithm 3, in which, step 1 to 2 are TPCR feature mapping, and step 3 to 5 are SCNNM dissimilarity measure.

Algorithm 3 BI-LEVEL RELATIVE INFORMATION ANALYSIS (BRIA)

Require: The labeled third-party dataset X_{tp}; the testing data of corpus sets X_cs and query sets X_qs.
Ensure: Ranking Y_q of all X_cs w.r.t. each X_q.
1: TPCR algorithm is performed using X_{tp}s as the dictionary to acquire the TPCR feature projection matrix P_s.
2: In the testing stage, all X_qs and all X_cs are mapped into the TPCR feature space by P_s as $\mathbf{X}_q$s and all $\mathbf{X}_c$s.
3: Every time, all $\mathbf{X}_q$s and all $\mathbf{X}_c$s are mixed together into the set list $\mathbf{X}$.
4: All sets of $\mathbf{X}$ are sorted by their CHISD to each $\mathbf{X}_q$ and each $\mathbf{X}_c$, respectively, to obtain the set-level Rank-Order lists.
5: SCNNM dissimilarity is measured between each pair of $(\mathbf{X}_q, \mathbf{X}_c)$ using the set-level Rank-Order lists.
6: For each $\mathbf{X}_q$, according to the set-to-set dissimilarities between itself and all $\mathbf{X}_c$s calculated in step 5, all $\mathbf{X}_c$s are re-ranked to returnY_q.

BRIA is different from conventional methods for multiple-shot person re-identification. Technically, BRIA considers for the issue from two levels: sample level and set level, while most of traditional methods focus one aspect. Methodologically, BRIA addresses the relative information, which is rarely concerned by current existing methods. Relative information considers the neighborhood topological information by encoding the relationship between the concerned sample/set and other several distributed samples/sets. It is robust especially when the size of each class is not large and samples/sets themselves are non-uniform distributed. Taking advantage of the collaboration between feature representation and dissimilarity measurement, BRIA enhances the performance as far as possible, which will also be experimentally demonstrated in Sect. 4.

The proposed method BRIA has some limitations as well. It requires the third-party data to build a dictionary for TPCR. Thus, the ability of TPCR is inevitably influenced by the quality of the dictionary. Currently, there is no optimization method on dictionary selection to maximize the ability of TPCR. Even so, TPCR has promising effectiveness with the recommended dictionary [5]. Furthermore, the SCNNM dissimilarity in BRIA is based on set-level Rank-Order lists, which are formed by low-level set-to-set distance measurements, thus the capability of it highly depends on the robustness of these measurements. If the low-level set-to-set distance is too sensitive to the noises of each set, the reliability of SCNNM dissimilarity might be reduced, which will give rise to the low performance of BRIA.

4. Experiments and Results

4.1 Dataset Description

We demonstrate the superiority of BRIA on several public benchmark datasets: ETHZ [13], iLIDS [14], iLIDS-MA [3], and iLIDS-AA [3]. All of them have multiple images of spatial-temporal variations for each person, as shown in Fig. 6.

The ETHZ dataset contains three video sequences of crowded street scenes captured by two moving cameras mounted on a carriage. We use three subsets of it extracted by Schwartz and Davis for person re-identification [15]. ETHZ1 has 83 persons within 4857 images, ETHZ2 has 35 persons within 1936 images, and ETHZ3 has 28 persons within 1762 images.

The iLIDS MCTS dataset is captured by a multi-camera CCTV network at an airport arrival hall in the busy time. From these videos, the i-LIDS dataset, which was extracted by Zheng *et al.*, is composed of 479 images for 119 individuals [14]; the i-LIDS-MA dataset [3] contains 40 individuals from two cameras, with 46 images annotated manually for each person; the i-LIDS-AA dataset [3] is made of 100 individuals obtained by the HOG-based human detector and tracker from both cameras. The noisy detection and tracking results make i-LIDS-AA more challenging. These three datasets are subject to more serious illumination changes and occlusions than ETHZ.

4.2 Experimental Settings

We normalize all the images into 128×48 pixels, and then randomly select 10 images per person for each query set and corpus set, respectively (coming from different cameras

Fig. 6 Exemplars from dataset ETHZ, i-LIDS, i-LIDS-MA, and i-LIDS-AA. For each person, four images are randomly selected.

if possible). For i-LIDS, each person has at most eight images, so we use them all by one half as the query set and the other half as the corpus set. For persuasiveness, we average the results for ten-fold cross validation with random corpus-query data splitting. Because each person only has two sets (one query set and one corpus set), the "Fixed-number" N and the balancing parameter Λ for SCNNM dissimilarity are suggested as "$N = 1$" and "$\Lambda = 1$", respectively.

According to the current research situation, we suggest the basic feature for TPCR to be concatenated by Dense-Sampled-Color-Histograms (DSCH), Schmid-Filter-Bank (SFB), and Gabor-Transform (GT) [15], to capture the color, texture, and edge information consistently. By contrast with the relative feature TPCR, this concatenated original feature, denoted by "Ori", can be considered as a kind of absolute feature, which is also valuable for demonstrating and comparing.

We emphasize the flexibility of the third-party data for building a descriptive and representative dictionary. Although the dictionary selection based on optimization seems more mathematically strict, it is not the focus of this paper. According to [5], even some heuristic selection of the third-party data as the dictionary would not reduce the capability of TPCR. So we just follow the suggestions in [5], and after some trials, we recommend the third-party data for each ETHZ dataset to be the rest two similar ETHZ datasets together with a very different dataset i-LIDS-AA. Taking ETHZ1 for example, we use ETHZ2, ETHZ3, and i-LIDS-AA together, denoted by "ETHZ2 + ETHZ3 + i-LIDS-AA". We also recommend the third-party data for i-LIDS to be "ETHZ3 + i-LIDS-MA + i-LIDS-AA", for i-LIDS-MA to be "i-LIDS-AA", and for i-LIDS-AA to be "i-LIDS-MA". If there are labeled data from the same dataset as the corpus belong to, we may also involve these data together with the third-party data in the dictionary. In the process of dictionary building, we limit the image number to be no larger than 46 for each person, which is the largest class size in i-LIDS-MA. Honestly, in our experiments, the third-party data selection are not guaranteed to be the best, but it will not influence the validation of the effectiveness of BRIA. Moreover, such tolerance to the flexible representation of TPCR feature mirrors the stability and reliability of BRIA to a certain degree.

4.3 Method Comparison

To expound the reasonability of BRIA, which can also be denoted by "TPCR_SCNNM(CHISD)", we compare it with typical related methods for person re-identification, including original feature and TPCR feature under different set-to-set distances like MPD and CHISD. In order to further validate the capability of SCNNM dissimilarity itself, we conduct experiments on its cooperation with several possible combinations of features and low-level set-to-set distances, such as "Ori_SCNNM(MPD)", "Ori_SCNNM(CHISD)", and "TPCR_SCNNM(MPD)". Moreover, we demonstrate the capability of our method by comparing with typical state-of-the-art methods as well, including the unsupervised method MRCG, and the supervised methods MCC [16], RankSVM [17], RDC [18], and SBDR [4]. Experimental results are illustrated by the "Cumulative Matching Characteristic" (CMC) curve, which visualizes the expectation of the correct match at each rank based on the ranking of each of the corpus w.r.t. the query [17].

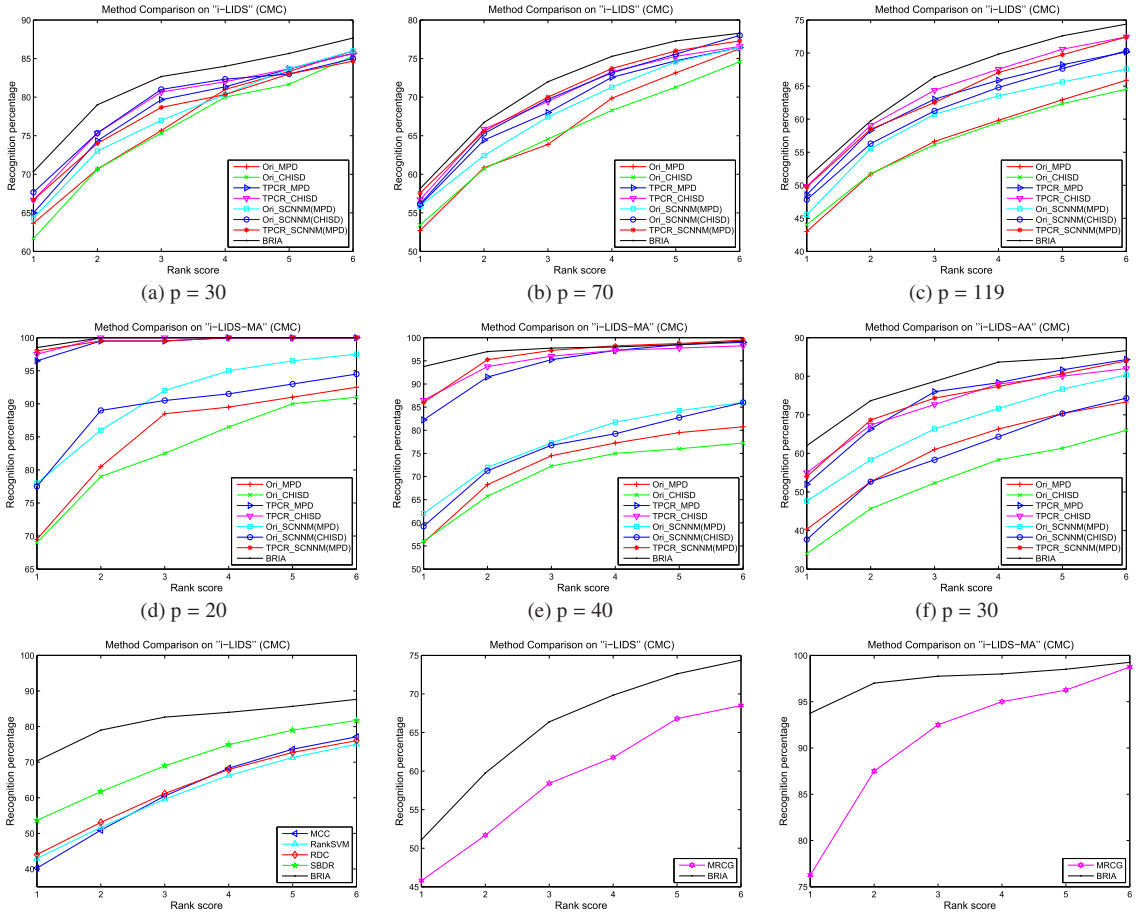
Results are illustrated in Fig. 7 except those on ETHZ, because for ETHZ1, ETHZ2, and ETHZ3, BRIA approaches 100% recognition rate on Rank-1 of CMC for all persons, which are superior to any other state-of-the-art methods. In Fig. 7, p denotes the number of persons. For i-LIDS, we test on 30 persons, 70 persons, and 119 persons; for i-LIDS-MA, we experiment on 20 persons and 40 persons; for i-LIDS-AA; we demonstrate on 30 persons. Overall, significantly, BRIA outperforms all the other concerned methods, as the evidence for the effectiveness of complementarity between the sample-level relative feature TPCR and the set-level relative dissimilarity SCNNM.

Different datasets may have different difficulties to address, though other challenges may also exist. BRIA is proposed to handle these difficulties simultaneously rather than separately. From the experimental results, we can see BRIA is robust to:

- Viewpoint variation and illumination variation (ETHZs, i-LIDS, i-LIDS-MA, and i-LIDS-AA);
- Pose variation and occlusion (i-LIDS and i-LIDS-AA);
- Localization errors (i-LIDS-AA).

The third-party data based dictionary covers kinds of variations, occlusion, and localization errors for different persons. Benefited from such a dictionary, TPCR can be robust to these similar variations, occlusion, and localization errors. For example, i-LIDS-AA covers the viewpoint variation, illumination variation, pose variation, occlusion, and localization errors, as shown in Fig. 6. If it is used as a dictionary, it may help to handle such difficulties in another similar dataset i-LIDS-MA. Even if TPCR feature cannot guarantee the enough ability to satisfactorily discriminate the person image sets, SCNNM will further exploit set-level common-near-neighbor information to make up for TPCR, so as to ensure intra-class dissimilarities are smaller than inter-class dissimilarities for all sets.

Furthermore, we can see TPCR_SCNNM(MPD), which is also the collaboration of two levels of relative information, but joined by a sensitive low-level set-to-set distance MPD, cannot work as well as BRIA which relies on CHISD, though TPCR_SCNNM(MPD) has potentiality to enhance the performance compared with other analogues and competitors. It justifies the argument that SCNNM is not only necessarily dependent on but also inevitably influenced by the robustness of low-level set-to-set distance measure. CHISD is more robust to noisy outliers, so it can provide a better platform for SCNNM. Furthermore, when the person number p increases, though in i-LIDS, the original feature collaborating with SCNNM may be competitive with the TPCR feature collaborating with SCNNM, our proposed



(a) p = 30 (b) p = 70 (c) p = 119

(d) p = 20 (e) p = 40 (f) p = 30

(g) p = 30 (h) p = 119 (i) p = 40

Fig. 7 CMC performance comparison on dataset i-LIDS, i-LIDS-MA, and i-LIDS-AA.

BRIA still have significant advantages, and such advantages are especially remarkable on i-LIDS-MA and i-LIDS-AA. From another perspective, we can clarify the strong adaptability of our proposed SCNNM dissimilarity according to the facts that SCNNM dissimilarity not only works very well with the TPCR feature, but also has good cooperation with the original feature. Analytically, original feature can be seen as a kind of sample-level absolute feature in a sense, potentially complementary to the set-level relative information as well.

We can also see that MPD performs better than CHISD in some results on i-LIDS-MA and i-LIDS-AA, especially in case of the original features. As mentioned in Sect. 3.3.2, MPD depends on the nearest sample points between the sets. Therefore, outliers of each class may easily influence the measuring reliability. CHISD tries to improve it by considering the distance between convex hulls for the set pair. However, it is unavoidably influenced by the layout of nearest sample points between the sets which support the convex hulls.

Actually, the convex hulls play a role to produce other interpolated points on them. The distribution of these interpolated points is determined by the existing sample points in the feature space. If the feature space is discriminative, existing sample points will be well distributed. In this situation, the interpolated points on the convex hulls will be reliable. Otherwise, existing sample points will be unsatisfactorily distributed. On this case, the convex hulls will be unreliable, and the interpolated points will bring more noises. Obviously, as demonstrated, TPCR can provide a more discriminative feature space than the original one. Therefore, in the TPCR feature space, CHISD can bring its superiority into play. But in the original feature space, CHISD loses its effect to a certain level, so that it may be outperformed by MPD. As for i-LIDS, sample points per set are very few and unsatisfactorily distributed in the original feature space. Thus, CHISD cannot either exploit its advantages or interpolate more noisy points, so will stay at a similar capability level to MPD.

In the i-LIDS dataset, there are less than 10 images per person. Because the expected performance of MRCG relies on the enough number of images per person to effectively extract the Karcher mean based covariance descriptors to condense the within-class correlations, image number per person in i-LIDS stays as a bottleneck for MRCG. However, our method remarkably outperforms it. Indeed, adequate image number per person can display the superiority of our method, but even when the image number per person

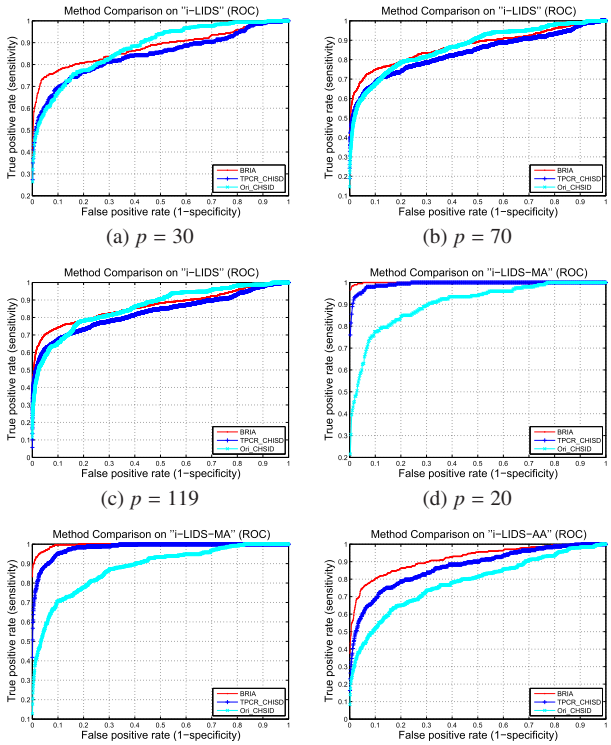
is comparatively small, the usage of relative information can offset the degrading of performance to some extent as well.

MCC, RankSVM, RDC, and SBDR are supervised approaches. The performance of them are unavoidably influenced by whether training samples and testing samples are independently and identically distributed, which cannot be ensured always in fact. Consequently, in i-LIDS, the insufficiency and complexity of person images in each class limit the performance of these methods. In i-LIDS-MA, it is difficult to carry out training and testing because there are merely 40 persons together. If we randomly split it into training data and testing data, too few persons for training will easily cause overfitting, and too few persons for testing will be unconvincing for method comparison as well. By contrast, BRIA doesn't require implementing learning using extensive training samples which need the matched people to be tediously annotated across camera views in the real scene. Taking advantage of the third-party data as the dictionary to represent the relative feature TPCR on the sample level, and making use of Rank-Order lists to model the relative dissimilarity SCNNM on the set level, BRIA has obvious superiority to all the concerned methods.

In Fig. 7, results are drawn by CMC. It is different from the ROC (Receiver Operating Characteristic) curve. CMC judges the ranking capability of the 1:m identification system, and has been widely used for person re-identification. By contrast, ROC is used for the verification system, and expresses the quality of the 1:1 matcher. ROC plots the fraction of true positives out of the positives (TPR = True Positive Rate) versus the fraction of false positives out of the negatives (FPR = False Positive Rate), at various threshold settings. This paper discusses the topic on re-identification, which is a multiple-class ranking problem. The ranking results shown by CMC directly illustrate the re-identification performance. Even so, we can transfer the 1:m ranking problem into a match/no-match binary verification problem, and draw the ROC curve based on that.

We match all the sets between the query side and corpus side. We treat the set pairs belonging to the same class as positive, and those from different classes as negative. In greater details, we form a list that contains all the set pairs ordered by the distance between the query set and corpus set within each pair. Then, we vary the threshold from the first correctly matched pair to the final correctly matched pair, and evaluate the TPR and FPR at each threshold.

The results are drawn in Fig. 8. For many methods, the differences on ROC curves are too small, so we only draw some representative ones. The transferred problem is quite unbalanced, in which the number of negative samples (set pairs) are much larger than the positive ones (correctly matched set pairs), so it is more meaningful to check the TPR at low FPRs, but not the global AUC measure. Using this criterion, we can see that BRIA performs significantly better than the others, which coincides with CMC.



(a) $p = 30$ (b) $p = 70$ (c) $p = 119$ (d) $p = 20$ (e) $p = 40$ (f) $p = 30$

Fig. 8 ROC performance comparison on dataset i-LIDS, i-LIDS-MA, and i-LIDS-AA.

4.4 Parameter Evaluation

As mentioned in Sect. 3.3.3, the Fixed-number N may influence the performance of the symmetric term in SCNNM. To show the robustness of our proposal, we display the results by changing N for both single-set cases and multiple-set cases. Results are evaluated by a condensed measure called MRR (Mean Reciprocal Rank), which has been suggested for usage in [19] with detailed definition and explanation. MRR can be calculated from CMC, and is able to provide an overall evaluation for the ranking. Here, p denotes the class number, S denotes the set number in each class, and q denotes the average sample number per set.

4.4.1 Single-Set Cases

From Table 1, we can see the overall decreasing trend of the results when N grows, and the symmetric term achieves its best performance with $N = 1$. With a suitable N, we further study the balancing parameter Λ of the model. Results are shown in Table 2. Generally, joining the symmetric term and asymmetric term for SCNNM can bring a better performance. In most cases, having $\Lambda \in [1, 10]$ is a good choice. Even so, tuning Λ does not substantially change the results, which shows the stability of SCNNM as well.

4.4.2 Multiple-Set Cases

In order to explain the importance of the Fixed-number N,

Table 1 MRR scores with different Ns for single-set cases.

i-LIDS	$N = 1$	$N = 5$	$N = 10$	$N = 15$	$N = 20$
$p = 30, 1 \le q \le 4$	**0.7697**	0.6932	0.6755	0.6443	0.6478
$p = 70, 1 \le q \le 4$	**0.6625**	0.5962	0.5429	0.5378	0.5311
$p = 119, 1 \le q \le 4$	**0.6007**	0.5492	0.4933	0.4767	0.4725
i-LIDS-MA	$N = 1$	$N = 5$	$N = 10$	$N = 15$	$N = 20$
$p = 20, q = 10$	**0.9925**	0.9033	0.8084	0.8092	0.7537
$p = 40, q = 10$	**0.9802**	0.8899	0.8378	0.8121	0.8050
i-LIDS-AA	$N = 1$	$N = 5$	$N = 10$	$N = 15$	$N = 20$
$p = 30, q = 10$	**0.7101**	0.6565	0.5912	0.5710	0.5566

Table 2 MRR scores generated by tuning Λ for single-set cases.

Λ	i-LIDS $p = 30$	i-LIDS $p = 70$	i-LIDS $p = 119$	i-LIDS -MA $p = 20$	i-LIDS -MA $p = 40$	i-LIDS -AA $p = 30$
0	0.7697	0.6625	0.6007	0.9925	0.9802	0.7101
0.001	0.7713	0.6647	0.6027	0.9925	0.9802	0.7121
0.01	0.7713	0.6647	0.6027	0.9925	0.9802	0.7121
0.1	0.7713	0.6647	0.6022	0.9925	0.9802	0.7164
1	**0.7775**	0.6691	0.6069	0.9925	**0.9804**	**0.7240**
10	0.7759	**0.6754**	**0.6095**	**0.9950**	0.9768	0.7180
100	0.7758	0.6743	0.6075	0.9950	0.9754	0.7146
1000	0.7758	0.6743	0.6075	0.9950	0.9754	0.7146
∞	0.7331	0.6331	0.5649	0.9842	0.9672	0.7087

Table 3 MRR scores with different Ns for multiple-set cases.

i-LIDS-MA	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$
$S = 4, q = 5$	**0.9338**	0.9294	0.9069	0.8795	0.8618
$S = 8, q = 5$	0.9464	**0.9495**	0.9464	0.9442	0.9317
i-LIDS-MA	$N = 1$	$N = 3$	$N = 5$	$N = 7$	$N = 9$
$S = 2, q = 5$	**0.9376**	0.8953	0.8547	0.8146	0.8062
$S = 6, q = 5$	0.9358	**0.9423**	0.9392	0.9285	0.9151
$S = 10, q = 5$	0.9444	0.9445	**0.9455**	0.9446	0.9418

we further evaluate this parameter on multiple-set cases with different set numbers in each class.

Experiments are carried out on i-LIDS-MA, because this dataset is not only challenging, but also has enough samples to conduct experiments for multiple-set cases. By contrast, the results on ETHZs are saturated, and the image number for each identity in i-LIDS and i-LIDS-AA are not enough to satisfy the required experimental condition. We simply separate this dataset into several sets without separating the data from different cameras. Results are displayed in Table 3. Obviously, when the Fixed-number N is set to approximate half of the set number in each class, the symmetric term achieves its best performance. This phenomenon not only shows the importance of the Fixed-number N, but also supports the recommendation for it. Besides the Fixed-number N, we also test the balancing parameter Λ, as described in Table 4. Generally, being coincident with the results in single-set cases, the proposed modeling plays its best performance with $\Lambda \in [1, 10]$ for multiple-set cases as well.

4.5 Computation Time Analysis

Though the proposed method BRIA can get satisfactory results, yet it cannot work in real time. BRIA consists of two primary steps: TPCR feature extraction and SCNNM(CHISD) dissimilarity measure. The computation time of BRIA is influenced by them. Here, we provide the actual running time for them and some other related comparative set-based distance measures.

Table 4 MRR scores generated by tuning Λ for multiple-set cases.

Λ	$S = 2$, $q = 5$	$S = 4$, $q = 5$	$S = 6$, $q = 5$	$S = 8$, $q = 5$	$S = 10$, $q = 5$
0	0.9376	0.9338	0.9423	0.9495	0.9455
0.001	0.9366	0.9334	0.9424	0.9496	0.9457
0.01	0.9366	0.9334	0.9424	0.9496	0.9458
0.1	0.9366	0.9335	0.9430	0.9501	0.9457
1	0.9382	**0.9355**	**0.9448**	**0.9518**	**0.9488**
10	**0.9406**	0.9329	0.9389	0.9452	0.9429
100	0.9405	0.9308	0.9347	0.9378	0.9337
1000	0.9405	0.9308	0.9347	0.9377	0.9336
∞	0.9281	0.9210	0.9294	0.9332	0.9304

Table 5 Computation time evaluation (unit:second).

	TPCR	SCNNM (CHISD)	SCNNM (MPD)	CHISD	MPD
i-LIDS, $p = 30$	310.8	6.2	0.3	1.6	0.1
i-LIDS, $p = 70$	286.4	35.6	1.9	9.1	0.4
i-LIDS, $p = 119$	285.9	98.8	5.6	25.1	1.2
i-LIDS-MA, $p = 20$	77.0	242.6	3.5	58.6	1.2
i-LIDS-MA, $p = 40$	60.5	46.0	3.3	11.8	0.5
i-LIDS-AA, $p = 30$	21.4	26.0	1.1	6.4	0.3

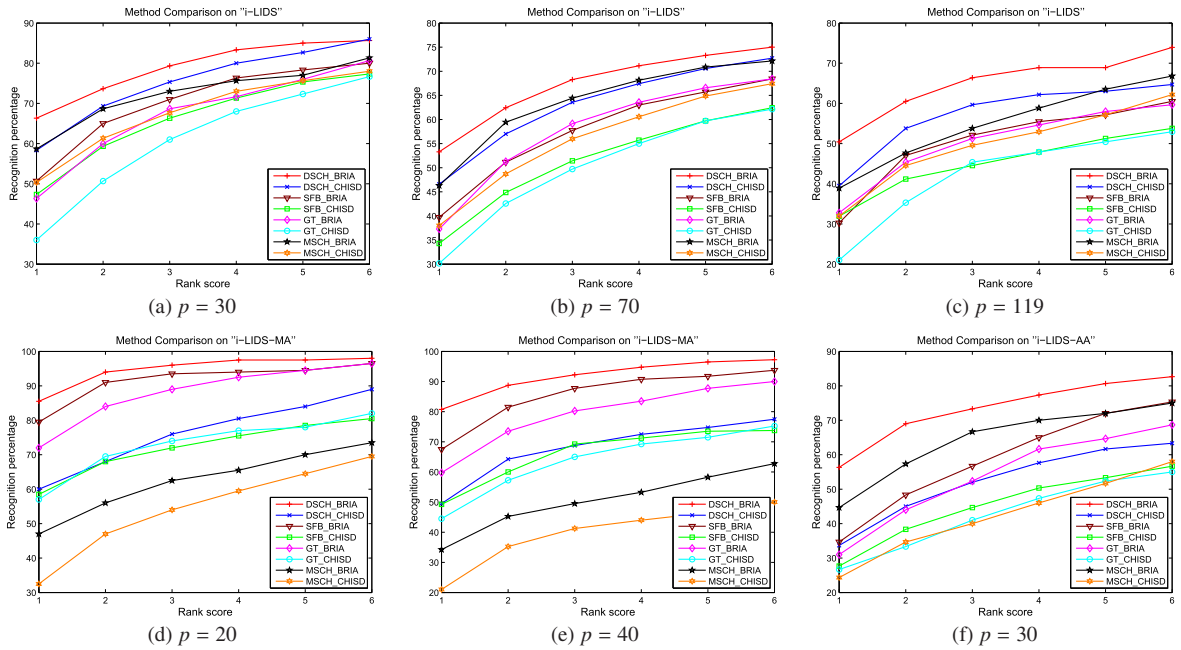
The computation time to extract TPCR feature depends on the dimension and the size of the dictionary. With the extracted TPCR feature, SCNNM measure doesn't need additional training process. Thus, generally, it is much faster than other supervised learning methods, such as MCC, RankSVM, RDC, and SBDR.

The actual time for TPCR extraction, SCNNM measure, CHISD measure, and MPD measure is shown in Table 5. The code has been implemented by Matlab 2010. We admit other programming environments, such as C and C++, may much accelerate the running speed. On the whole, though SCNNM takes longer time than direct set-to-set distance (MPD and CHISD), this is in an acceptable efficiency range. Besides, because we involve the remaining data in the same dataset as testing data together with the third-party data in the dictionary, the length of dictionary will decrease when the testing person number increases. Therefore it is a little faster to extract TPCR for the larger testing person number in i-LIDS and i-LIDS-MA.

4.6 Robustness to Various Features

As is well known, rising tide would lift the boat. Introducing a better feature to our model will probably result in a better performance. The reason we choose the signature combined by DSCH, SFB, and GT is that this signature has been adopted in several state-of-the-art methods [4], [5], [18], and been proved quite suitable and effective for the issue of person re-identification. For the proposed method BRIA, this signature can not only bring a good performance, but also ensure a fair comparison with the state-of-the-art methods.

Honestly, BRIA can be combined with various types



(a) $p = 30$ (b) $p = 70$ (c) $p = 119$

(d) $p = 20$ (e) $p = 40$ (f) $p = 30$

Fig. 9 Feature validation and comparison on dataset i-LIDS, i-LIDS-MA, and i-LIDS-AA.

of image features. Therefore, we tentatively demonstrate it by discussing the performance changes when using different suitable features for the issue. These features include DSCH, SFB, GT, and another representative feature, named Multiple Space Color Histogram (MSCH) [18]. Results are illustrated in Fig. 9. Clearly, though these features may have different performance on different datasets, equipped with BRIA, they can obtain remarkable performance enhancement. This validates the robustness of BRIA.

5. Conclusion

This paper has proposed a new method named "BRIA" for multiple-shot person re-identification, which integrates two levels of relative information. As the set-level relative dissimilarity, standing on the shoulders of CNNA, a reliable and effective set-to-set dissimilarity SCNNM has been presented, which is complementary to the newly proposed sample-level relative feature TPCR. Their integration leads to an encouraging performance compared with state-of-the-art methods. Possible future work will include applying BRIA to the issues of human tracking across cameras.

Acknowledgements

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.

References

[1] B. Loris, C. Marco, P. Alessandro, F. Michela, and M. Vittorio, "Multiple-shot person re-identification by hpe signature," Proc. 20th International Conference on Pattern Recognition, ICPR, Istanbul, Turkey, pp.1413–1416, Aug. 2010.

[2] B. Slawomir, C. Etienne, B. Francois, and T. Monique, "Person re-identification using haar-based and dcd-based signature," Proc. 7th International Conference on Advanced Video and Signal Based Surveillance, AVSS, Boston, USA, pp.1–8, Aug. 2010.

[3] B. Slawomir, C. Etienne, B. Francois, and T. Monique, "Boosted human re-identification using riemannian manifolds," Image Vis. Comput., vol.30, pp.443–452, June 2012.

[4] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao, "Set based discriminative ranking for recognition," Proc. 12th European Conference on Computer Vision, ECCV, pp.497–510, Florence, Italy, Oct. 2012.

[5] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao, "Robust object recognition via third-party collaborative representation," Proc. 21st International Conference on Pattern Recognition, ICPR, Tsukuba, Japan, Nov. 2012.

[6] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analsis for person re-identification," Proc. 19th International Conference on Image Processing, ICIP, Florida, USA, Sept. 2012.

[7] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," Proc. 13th International Conference on Computer Vision, ICCV, pp.471–478, Barcelona, Spain, Nov. 2011.

[8] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, pp.210–227, 2009.

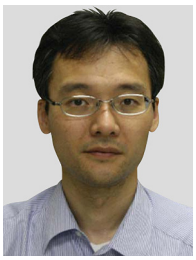
[9] C. Zhu, F. Wen, and J. Sun, "A rank-order distance based clustering algorithm for face tagging," Proc. 24th International Comference on Computer Vision, CVPR, pp.481–488, Colorado Springs, USA, June 2011.

[10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," Proc. 23rd IEEE Conference on Computer Vision and Pat-

tern Recognition, CVPR, pp.2360–2367, San Francisco, USA, June 2010.
[11] H. Cevikalp and B. Triggs, "Face recognition based on image sets," Proc. 23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp.2567–2573, San Francisco, USA, June 2010.
[12] W. Li, Y. Wu, Y. Kawanishi, M. Mukunoki, and M. Minoh, "Riemannian set-level common-near-neighbor analysis for multiple-shot person re-identification," Proc. 13th IAPR Internatianal Conference on Machine Vision Applications, pp.9–12, Kyoto, Japan, May 2013.
[13] A. Ess, B. Leibe, and L.V. Gool, "Depth and appearance for mobile scene analysis," Proc. 11th International Conference on Computer Vision, ICCV, Rio de Janeiro, Brazil, Oct. 2007.
[14] W.S. Zheng, S.G. Gong, and T. Xiang, "Associating groups of people," Proc. 20th British Machine Vision Conference, BMVC, London, UK, pp.23.1–23.11, Sept. 2009.
[15] W.R. Schwartz and L.S. Davis, "Learning discriminative appearance-based models using partial least squares," Proc. XXII Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI, pp.322–329, Rio de Janeiro, Brazil, Oct. 2009.
[16] A. Globerson and S. Roweis, "Metric learning by collapsing classes," Advances in Neural Information Processing Systems, vol.18, pp.451–458, 2006.
[17] B. Prosser, W. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," Proc. 21st British Machine Vision Conference, BMVC, pp.21.1–21.11, Aberystwyth, UK, Aug. 2010.
[18] W.S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," IEEE Trans. Pattern Anal. Mach. Intell., vol.35, pp.653–668, March 2013.
[19] Y. Wu, M. Mukunoki, T. Funatomi, and M. Minoh, "Optimizing mean reciprocal rank for person re-identification," Proc. 8th International Conference on Advanced Video and Signal-Based Surveillance, AVSS, pp.408–413, Klagenfurt, Austria, Aug. 2011.

Wei Li is a Ph.D candidate in Department of Intelligence Science and Technology at Kyoto University currently. He received a B.S. degree in Measuring and Control Technology and Instrumentations and a M.S. degree in Instrument Science and Technology from Southeast University in 2007 and 2010, respectively. His research interests include computer vision, pattern recognition, and convex optimization.

Yang Wu is currently a post-doctoral researcher of Academic Center for Computing and Media Studies, Kyoto University. He received a BS degree in information engineering and a Ph.D degree in pattern recognition and intelligent systems from Xi'an Jiaotong University in 2004 and 2010, respectively. From Sep. 2007 to Dec. 2008, he was a visiting student in the General Robotics, Automation, Sensing and Perception (GRASP) lab at University of Pennsylvania. His research is in the fields of computer vision and pattern recognition, with particular interests in the detection, tracking and recognition of humans and also generic objects. He is also interested in image/video search and retrieval, along with machine learning techniques.



Masayuki Mukunoki received the bachelor, master and doctoral degrees in information engineering from Kyoto University. He is now an Associate Professor in the Academic Center for Computing and Media Studies and a faculty member in the Graduate School of Informatics, in Kyoto University. His research interests include computer vision, video media processing, lecture video analysis and human activity sensing with camera.

Michihiko Minoh is a professor at Academic Center for Computing and Media Studies (ACCMS), Kyoto University, Japan. He received the B.Eng., M.Eng. and D.Eng. degrees in Information Science from Kyoto University, in 1978, 1980 and 1983, respectively. He served as director of ACCMS from April 2006 to March 2010 and concurrently served as vice director in the Kyoto University Presidents Office from October 2008 to September 2010. Since October 2010, he has been vice-president, chief information officer at Kyoto University, and director-general at Institute for Information Management and Communication, Kyoto University. His research interest includes a variety area of Image Processing, Artificial Intelligence and Multimedia Applications, particularly, model centered framework for the computer system to help visual communication among humans and information media structure for human communication. He is a member of Information Processing Society of Japan, Institute of Electronics, Information and Communication Engineers of Japan, the IEEE Computer Society and Communication Society, and ACM.