Person Reidentification With Reference Descriptor

Le An, *Member, IEEE*, Mehran Kafai, *Member, IEEE*, Songfan Yang, *Member, IEEE*, and Bir Bhanu, *Fellow, IEEE*

***Abstract*— Person identification across nonoverlapping cameras, also known as person reidentification, aims to match people at different times and locations. Reidentifying people is of great importance in crucial applications such as wide-area surveillance and visual tracking. Due to the appearance variations in pose, illumination, and occlusion in different camera views, person reidentification is inherently difficult. To address these challenges, a reference-based method is proposed for person reidentification across different cameras. Instead of directly matching people by their appearance, the matching is conducted in a reference space where the descriptor for a person is translated from the original color or texture descriptors to similarity measures between this person and the exemplars in the reference set. A subspace is first learned in which the correlations of the reference data from different cameras are maximized using regularized canonical correlation analysis (RCCA). For reidentification, the gallery data and the probe data are projected onto this RCCA subspace and the reference descriptors (RDs) of the gallery and probe are generated by computing the similarity between them and the reference data. The identity of a probe is determined by comparing the RD of the probe and the RDs of the gallery. A reranking step is added to further improve the results using a saliency-based matching scheme. Experiments on publicly available datasets show that the proposed method outperforms most of the state-of-the-art approaches.**

***Index Terms*— Person reidentification, reference descriptor (RD), reranking, saliency, subspace, surveillance.**

I. Introduction

IMAGING sensors are being widely deployed for many real-world applications, such as video surveillance and access control. In particular, in relation to camera networks, there has been an increasing interest in person reidentification and considerable progress has been made recently [1]–[5]. Person reidentification is a recognition task that aims to match individuals across nonoverlapping cameras at different times and locations. Accurate person reidentification can help locate a target subject in video-monitored surroundings. The matching result of person reidentification can be used in other tasks such as tracklet association in a multicamera tracking system [6]. Due to the large amount of image data that contain persons of interest, it is not feasible to manually screen and identify every person in a video or image. Thus, automatic labeling or matching of people is highly desired.

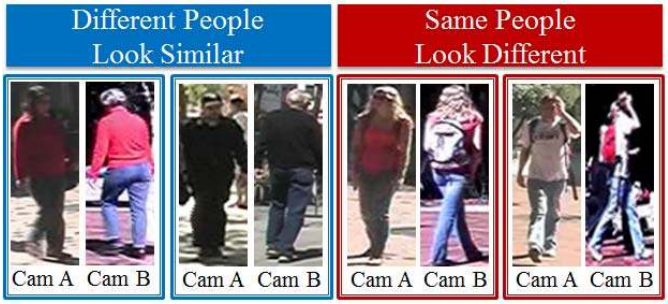


Fig. 1. In nonoverlapping camera views, different people may look very similar (left) while the same people's appearance may change dramatically due to variations in pose and illumination (right). Samples are from two cameras (Cam A and Cam B) in the VIPeR dataset [9].

Recently, many labeling techniques have been proposed for large-scale image data, such as the seminal works in [7] and [8] that can robustly annotate image even with noise. However, such methods cannot be directly applied to person reidentification since matching people in different cameras is intrinsically difficult due to the imaging condition disparity among different cameras. In particular, the following problems contribute to the complications of person reidentification in a camera network.

1) *Low Resolution:* Most of the surveillance cameras are not able to capture high-resolution images due to the low resolution of inexpensive cameras and large distance between camera and human subjects.
2) *Arbitrary Poses:* Since a subject is captured by surveillance cameras with nonoverlapping field of views, the poses of a subject in different camera are usually quite different.
3) *Changing Illumination:* The images are captured at different times and/or locations. As a consequence, the appearance of a person may change dramatically due to illumination changes.
4) *Occlusion:* A subject may carry accessories such as a backpack and briefcase, which may occlude distinctive features of the subject in a certain view.

Fig. 1 shows some image pairs of the same and different people in two cameras. Due to large variations in pose,

Manuscript received July 8, 2014; revised December 10, 2014 and February 5, 2015; accepted March 10, 2015. Date of publication March 25, 2015; date of current version April 1, 2016. This work was supported in part by the National Science Foundation under Grant 1330110 and in part by the Office of Naval Research, Arlington, VA, USA, under Grant N00014-12-1-1026. This paper was recommended by Associate Editor Q. Tian. *(Corresponding author: Le An.)*

L. An was with the Department of Electrical and Computer Engineering, University of California at Riverside, Riverside, CA 92521 USA. He is now with the Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mail: lan004@ucr.edu).

M. Kafai is with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail: mehran.kafai@hp.com).

S. Yang is with the College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China (e-mail: syang@scu.edu.cn).

B. Bhanu is with the Center for Research in Intelligent Systems, University of California at Riverside, Riverside, CA 92521 USA (e-mail: bhanu@cris.ucr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2015.2416561

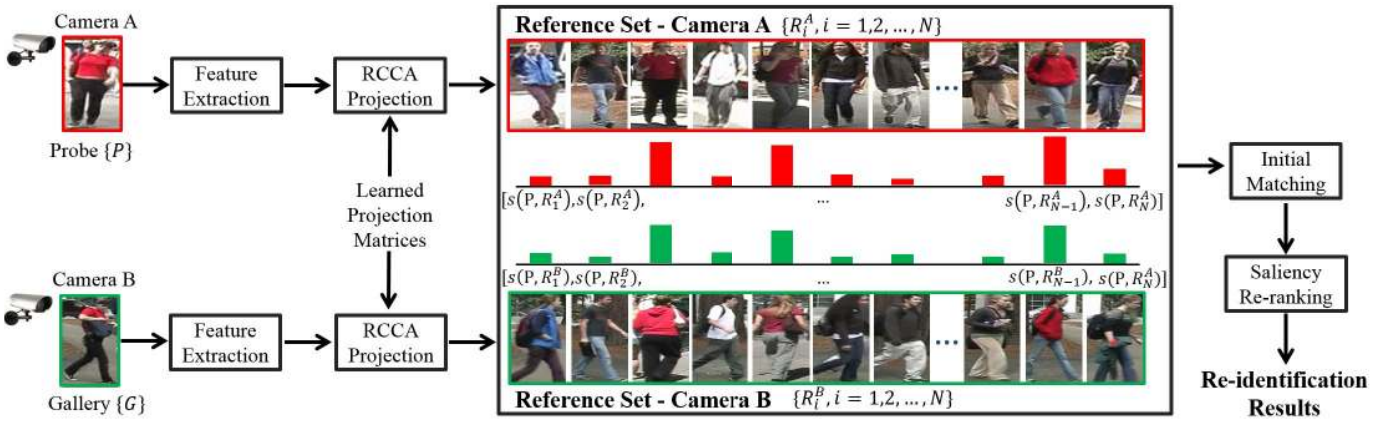


Fig. 2. Framework of the proposed reference-based reidentification. The appearance features are first extracted from probe and gallery images and are then projected onto RCCA subspace with learned projection matrices. An RD for a probe or gallery instance is generated by computing and concatenating its similarity scores with respect to a reference set. After the RDs for both probe and gallery are generated, initial matching is performed using the RDs. A saliency-based reranking scheme is included to further improve the reidentification accuracy.

illumination, and background, the appearance of the same subject may look very different in different cameras, while different people may highly resemble in appearance. The significant view and appearance changes across nonoverlapping cameras make person reidentification inherently difficult.

The gallery for reidentification usually contains images of known subjects in one camera view and the probes are subjects from another camera view. To recognize a given probe from a large gallery, the basic idea is to first extract a robust feature representation for both probe and gallery images, and then perform matching using this representation. This kind of approach is called *appearance-based* and it makes use of visual cues only.

Appearance-based methods can be categorized into two groups. The goal of the methods in the first group is to extract feature representations that have low intra-class variation for the same subject and high inter-class variation among different subjects [9]–[11]. However, due to the significant appearance change across different cameras, the intra-class variation is often larger than the inter-class variation. As a result, accurate matching is very difficult.

For the second group of methods, the goal is to learn the optimal distance metric for the image pairs from two different cameras [12]–[14]. These metric learning approaches learn a transformation for the original feature representation such that the intra-class distances are minimized while the inter-class distances are maximized. The drawback of the metric-learning-based methods is that the learned model tends to overfit the training data. In addition, some popular approaches [15]–[17] are computationally expensive due to complex optimization involved.

In this paper, instead of designing a complex feature representation or learning a specialized distance metric as it has been done in the previous methods, we present a new framework for single-shot person reidentification in which the matching is performed using *reference descriptors* (RDs). Fig. 2 shows the framework of the proposed reference-based method. To match a probe and a gallery instance, appearance features are first extracted. Using learned projection matrices, the probe and gallery features are projected onto a lower dimensional subspace. We use regularized canonical correlation analysis (RCCA) to learn the projection matrices since the RCCA is able to maximize the correlation between the data from different views. After feature projection, the RDs of the probe and gallery are generated using a *reference set*. The reference set is a set of images of the subjects from different camera views and the identities in the reference set do not overlap with probe or gallery subjects. An RD of a probe or a gallery instance is formed by concatenating the similarity scores between this probe or gallery to the reference set in the RCCA feature space. Thus, the dimension of an RD is determined by the size of the reference set and is irrelevant to the size of the image features. The matching between the probe and gallery is performed by computing the similarity between their RDs. In this way, the probe and gallery from different views are indirectly compared using a reference set, instead of being matched directly. To improve the initial matching results, a saliency-based reranking stage is added to obtain the final reidentification results.

The rest of this paper is organized as follows. Section II introduces related work and summarizes our contributions. The details of the proposed method for person reidentification are presented in Section III. Section IV provides the experimental results, and finally Section V concludes this paper and states the future work.

II. Related Work and Contributions

A. Related Work

Two major directions to tackle person reidentification are to extract invariant feature representations and to learn specialized distance metrics across different camera views. We review the related work in these categories as well as some work in other areas related to our reference-based matching.

1) Feature-Driven Approaches: Cheng *et al.* [18] adopted pictorial structures to localize the human parts and search part-to-part correspondences to match subjects. Farenzena *et al.* [10] extracted features accounting for the overall chromatic content, the spatial arrangement, and the

presence of recurrent local motifs to match the individuals with appearance variation. Bak *et al.* [19] learned a model in a covariance metric space to select features based on the idea that different regions for each subject should be matched specifically. Gray and Tao [20] used AdaBoost to select the most discriminative features instead of handcrafted features. Prosser *et al.* [21] formulated the reidentification as a relative ranking problem instead of an absolute scoring problem. Hirzer *et al.* [22] proposed a two-step method by first using a descriptive model to obtain an initial ranking, which was refined in the second step by a discriminative model with human feedback. Kviatkovsky *et al.* [23] discovered the color intra-distribution structure and showed that this structure was invariant under certain illumination changes and could be combined with the covariance descriptor for person reidentification. Ma *et al.* [24] used both biologically inspired features and covariance descriptors to handle background and illumination variations. Martinel and Micheloni [25] presented an appearance-based approach by computing a novel discriminative signature from multiple local features.

Beyond low-level features, semantic features have been explored for improved reidentification results. Kuo *et al.* [26] applied semantic color names to describe an image of a person instead of using color histograms for better stability. Layne *et al.* [27] proposed midlevel semantic attributes to describe person for the purpose of reidentification. An *et al.* [28] used biometric attributes such as gender from images to rerank the initial reidentification results from low-level features. Zhao *et al.* [29], [30] proposed to use salient features for person reidentification. The saliency was estimated using unsupervised learning and was combined with existing methods [10] to improve the recognition performance.

Yang *et al.* [31] proposed a color descriptor based on salient color names, which can guarantee that a higher probability will be assigned to the color name that is closest to the intrinsic color. Zhao *et al.* [32] learned discriminative midlevel filters from automatically discovered patch clusters to identify specific visual patterns. Li *et al.* [33] proposed a neural network in which misalignment, pose difference, occlusions, and background clutter were jointly handled with abundant data. Zhang and Saligrama [34] tackled the appearance variation in different cameras using basis functions that encode cooccurrences of visual patterns. Specifically, locality sensitive cooccurrence measures were developed to incorporate semantically meaningful appearance changes. Liu *et al.* [35] proposed a postrank optimization method that allowed a human-in-the-loop to select negative samples. This improved the performance gain over 30% and compared with the exhaustive search, the time efficiency significantly improved. Liu *et al.* [36] provided extensive study of feature importance for person reidentification and proposed a method for on-the-fly mining of feature. For person reidentification on mobile devices, Vernier *et al.* [37] introduced a client-server system that improved the reidentification performance over time with reduced computation time.

2) Distance Learning-Based Approaches: Hirzer *et al.* [12] proposed a relaxed pairwise learned metric (RPLM) based on the Mahalanobis distance learning that took advantages of the structure of the data with reduced computational cost. It achieved state-of-the-art results with simple feature descriptors. Köstinger *et al.* [14] proposed a simple yet effective method to learn the distance metric called keep it simple and straightforward (KISS) metric (KISSME) from a statistical inference perspective. Tao *et al.* [38] extended the KISSME by introducing regularization to robustly estimate covariance matrices against the instability in calculating the inverse of a covariance matrix from a small size training set. A method termed minimum classification error-KISS (MCE-KISS) [39] was proposed to handle the small sample size problem in estimating eigenvalues of a covariance matrix and it was shown to be effective for person reidentification. Zheng *et al.* [13] formulated reidentification as a relative distance comparison problem. It maximized the likelihood such that the distance between a pair of images of the same person is smaller than a pair of images of different people. Liu *et al.* [40] incorporated attribute information into the framework of [13] to further improve the reidentification results by feature weighting. Li and Wang [41] jointly partitioned the image spaces of two camera views into different configurations based on the similarity of cross-view transforms. Image pairs with similar transforms were projected onto a common feature space for matching.

Standard metric learning techniques such as large margin nearest neighbor (LMNN) [15], information-theoretic metric learning (ITML) [17], and logistic discriminant metric learning (LDML) [16] were also applied to person reidentification. Dikmen *et al.* [42] developed a variant of LMNN by introducing a reject option to the unfamiliar matches (LMNN-R) and achieved improved results. Martinel *et al.* [43] extracted multiple features from image pairs and obtained a so-called distance feature vector. The reidentification was achieved by classifying this distance feature vector using a trained binary classifier. Pedagadi *et al.* [44] used local Fisher discriminant analysis (LFDA) to reduce feature dimensionality for person reidentification. It outperformed other metric learning-based methods. Mignon and Jurie [45] proposed pairwise constrained component analysis (PCCA) to learn a low-dimensional mapping in which distances between data points complied with a set of sparse training pairwise constraints. An *et al.* [46] performed matching in a common space where the same subjects from different cameras were maximally correlated through a robust feature mapping.

Loy *et al.* [47] reported a manifold ranking (MRank) approach in which the probe information was propagated along the data manifold in an unsupervised manner. It showed that the performance of existing metric-learning-based methods could be significantly improved by integrating the MRank. Xiong *et al.* [48] applied multiple kernel-based metrics in conjunction with histogram-based features and showed improvement over state-of-the-art on several datasets. Liu *et al.* [49] learned two coupled dictionaries jointly for gallery and probe using both labeled and unlabeled images to mitigate the appearance variation between different cameras. Recently, Liao *et al.* [50] considered the open-set person

reidentification problem that removed the assumption that a probe subject should belong to the gallery. In this more practical setting, the presence of the probe subject in the gallery is first determined, followed by an identification step using several metric learning methods as baselines. Comprehensive survey on person reidentification can be found in [1], [2], and [4].

3) Matching in Reference Space: Pattern matching using a reference set has been explored in different fields. Gyaourova and Ross [51] generated fixed-length codes for indexing biometric databases. The index codes were constructed by computing match scores between a biometric image and a fixed set of images. Duin and Pkalska [52] discussed the dissimilarity space to convert the structural representation of data to a dissimilarity representation using a representation set and some suggestions for prototype selection were provided. Guo *et al.* [53] proposed a prototype embedding of visual appearance using a representation set of model prototypes for vehicle matching. Recently, Chen *et al.* [54] developed a reference-based approach for tracking people across nonoverlapping cameras using a reference-based appearance model.

B. Contributions of This Paper

Compared with the previous work discussed in Section II-A, the major contributions of this paper are twofolded. First, we tackle the reidentification problem using a reference-based scheme in conjunction with subspace learning. Our framework avoids direct matching of image pairs with significant appearance variation and achieves superior performance compared with the state-of-the-art methods as validated by the experiments. Second, we use different methods to pursue optimality for reference set selection and the experiments show that the size of reference set can be reduced without a significant loss of accuracy. In addition, the proposed reference-based reidentification framework is compatible with any feature descriptor and can be extended to other applications.

A preliminary version of this paper appeared in [55]. In this paper, we have the following major changes and improvements compared with [55].

1) We have studied and discussed more recent advances in person reidentification. We have provided more comparisons to our method.
2) We have conducted more in-depth experiments on more datasets, and included detailed performance analysis. In addition, we have shown that the performance of the proposed method can be further improved by incorporating a modified cosine similarity measure and a saliency-based reranking step.
3) We have explored different methods for reference set selection and provided recommendations about how to select reference set based on empirical validations.

III. Person Reidentification in Reference Space

The proposed method involves an offline process and an online reidentification process. In the offline process, the RCCA projection matrices are learned and the RDs of the gallery are generated. During online reidentification process, the RD of a probe is generated and is compared with the RDs of the gallery to obtain the initial matching result. Reranking is then performed to improve the initial results based on image saliency. The details are explained as follows.

A. Offline Process

1) CCA Subspace Learning: Canonical correlation analysis (CCA) is a multivariate statistical analysis technique, which was first introduced in [56]. It aims to explore the relationship between two sets of random variables from the different observations on the same data (e.g., images of subjects from different views). CCA finds projections such that the correlation between these two sets of random variables is maximized after projection.

Mathematically, given two sets of data observations, $D^A = \{d_i^A \in \mathbb{R}^m, i = 1, 2, \ldots, N\}$ and $D^B = \{d_i^B \in \mathbb{R}^n, i = 1, 2, \ldots, N\}$, CCA aims at obtaining two sets of basis vectors $W_A \in \mathbb{R}^m$ and $W_B \in \mathbb{R}^n$ such that the correlation coefficient ρ of $W_A^T D^A$ and $W_B^T D^B$ is maximized. The objective function to be maximized is

$$\begin{aligned}\rho &= \frac{\mathrm{Cov}\left(W_A^T D^A, W_B^T D^B\right)}{\sqrt{\mathrm{Var}\left(W_A^T D^A\right)}\sqrt{\mathrm{Var}\left(W_B^T D^B\right)}} \\ &= \frac{W_A^T C_{AB} W_B}{\sqrt{W_A^T C_{AA} W_A W_B^T C_{BB} W_B}}\end{aligned} \tag{1}$$

where C_{AA} is the covariance matrix of D^A, C_{BB} is the covariance matrix of D^B, and C_{AB} is the cross-covariance matrix between D^A and D^B.

Equivalently, the CCA can be formulated as a constrained optimization problem by

$$\underset{W_A, W_B}{\operatorname{argmax}} W_A^T C_{AB} W_B \tag{2}$$

subject to $W_A^T C_{AA} W_A = 1$ and $W_B^T C_{BB} W_B = 1$.

Using the Lagrange multiplier, the solution of (2) is equivalent to solving the following generalized eigenvalue problems:

$$\begin{aligned}C_{AB} W_B &= \lambda C_{AA} W_A \\ C_{BA} W_A &= \lambda C_{BB} W_B\end{aligned} \tag{3}$$

where $C_{BA} = C_{AB}^T$. CCA is performed in an unsupervised manner and both correlation maximization and dimensionality reduction can be achieved simultaneously by choosing the number of basis vectors to use.

Often in practice, the feature dimension of the data is significantly larger than the number of data samples. In this case, the covariance matrices C_{AA} and C_{BB} may be singular and their inverse would be ill conditioned. RCCA has been proposed to solve this problem and it prevents overfitting [57]. In the solution of RCCA, the generalized eigenvalue problem becomes

$$\begin{aligned}C_{AB} W_B &= \lambda (C_{AA} + \lambda_1 I_A) W_A \\ C_{BA} W_A &= \lambda (C_{BB} + \lambda_2 I_B) W_B\end{aligned} \tag{4}$$

where λ_1 and λ_2 are the two nonnegative regularization parameters. I_A and I_B are the two identity matrices. Usually λ_1 and λ_2 are determined by cross validation.

2) Gallery Data in Reference Space: The reference set contains images $\{I_i^A, i = 1, 2, \ldots, N\}$ and $\{I_i^B, i = 1, 2, \ldots, N\}$ of N subjects from two different cameras A and B. The features such as color histograms and texture descriptors from each image are extracted and two feature sets $\{F_i^A, i = 1, 2, \ldots, N\}$ and $\{F_i^B, i = 1, 2, \ldots, N\}$ are obtained. Since the features are from images in different views, we first learn a RCCA subspace in which the correlations between the projected feature sets $\{W_A^T F_i^A, i = 1, 2, \ldots, N\}$ and $\{W_B^T F_i^B, i = 1, 2, \ldots, N\}$ are maximized. The RCCA projection matrices W_A and W_B are learned as in (4). By projecting the original features onto the RCCA subspace, we obtain the projected features of the reference set denoted by $\{f_i^A, i = 1, 2, \ldots, N\}$ and $\{f_i^B, i = 1, 2, \ldots, N\}$ with reduced dimensionality and enhanced correlation.

Suppose we have a gallery of M subjects from camera A, the features of the gallery subjects are first extracted and then projected onto the RCCA subspace using the learned projection matrix W_A. The RCCA feature for the jth subject in the gallery set is denoted by f_j^g. From f_j^g, its RD R_j^g, as a new representation, is generated by

$$R_j^g = \left[s\left(f_j^g, f_1^A\right), s\left(f_j^g, f_2^A\right), \ldots, s\left(f_j^g, f_N^A\right)\right]^T \tag{5}$$

where $s(a, b)$ denotes the similarity between the features a and b. We use the cosine similarity to compute $s(a, b)$. In this process, the representation of the gallery subject is transformed to a descriptor of length N regardless of the original feature dimension and each element in R_j^g indicates the similarity between this gallery subject and a reference subject. The projected features of the reference set from camera A $\{f_i^A, i = 1, 2, \ldots, N\}$ are similar to basis functions and in the reference space, they jointly describe the appearance of a gallery subject in terms of its similarity to individuals in the reference set. Fig. 2 shows the basic idea of how the RDs are generated.

The rationale for first projecting the features onto the RCCA subspace is to better couple the features $\{f_i^A, i = 1, 2, \ldots, N\}$ and $\{f_i^B, i = 1, 2, \ldots, N\}$. In the reidentification, a probe image is described using $\{f_i^B, i = 1, 2, \ldots, N\}$. Since $\{f_i^A, i = 1, 2, \ldots, N\}$ and $\{f_i^B, i = 1, 2, \ldots, N\}$ are maximally correlated after RCCA projection, the matching between the probe and the gallery becomes meaningful and reliable.

B. Online Reidentification

1) Initial Matching: Suppose the probe is from camera B and the detection of a subject (I_p) is given, the appearance features F^p are first extracted. The projected feature f^p of the probe in the RCCA subspace is given by

$$f^p = W_B^T F^p. \tag{6}$$

The RD of the probe R^p is computed in a similar manner as in (5) using the projected features of the reference set from camera B $\{f_i^B, i = 1, 2, \ldots, N\}$ by

$$R^p = \left[s\left(f^p, f_1^B\right), s\left(f^p, f_2^B\right), \ldots, s\left(f^p, f_N^B\right)\right]^T \tag{7}$$

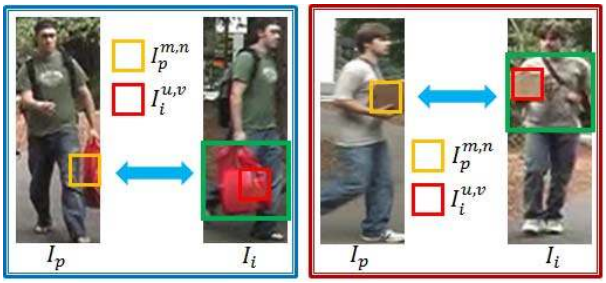


Fig. 3. Samples of saliency detection in two-camera views. To estimate the saliency of a patch $I_p^{m,n}$ (in yellow bounding box) in the image I_p in one camera view, a constraint space (in green bounding box) is searched in each image I_i in the reference set in the other camera view. The patch $I_i^{u,v}$ (in red bounding box) is found out as the most similar patch to $I_p^{m,n}$, which will be used to calculate the patch saliency as in (10).

where f_i^B is the projected features in the RCCA subspace of the reference subject i in camera B.

The identity of the subject is determined by the similarity $\text{sim}(R^p, R_i^g)$ between the probe R^p and each gallery R_i^g and then the top match R_k^g is found in the gallery such that

$$k = \underset{i}{\text{argmax}}\, \text{sim}\left(R^p, R_i^g\right). \tag{8}$$

To compute similarity, we use the modified cosine similarity [58] defined as

$$\text{sim}\left(R^p, R_i^g\right) = \frac{\left|(R^p)^T \cdot R_i^g\right|}{\|R^p\| \left\|R_i^g\right\| \left(\left\|R^p - R_i^g\right\|_p + \epsilon\right)} \tag{9}$$

where $\|\cdot\|_p$ is the l_p norm and ϵ is a small positive number to prevent division by zero. The reason to apply the modified cosine similarity is that the standard cosine similarity does not take into consideration the actual distance between two vectors, while the modified cosine similarity is able to address both the distance measure and angular measure and has improved performance in recognition tasks [58].

2) Saliency Detection: To improve the reidentification accuracy, we opt to high-level image information to rerank the initially returned results. Specifically, we use image saliency [29], [30] to improve the rank of the correct match. Image saliency, such as carrying item, is a discriminative visual feature to match subjects across different views. Fig. 3 shows two examples of saliency correspondence across different cameras.

Given the reference set $\mathbf{I} = \{I_i, i = 1, 2, \ldots, N\}$ in one camera view, to compute the saliency of an image I_p in another view, image patches are first densely sampled. For each patch $I_p^{m,n}$ in I_p, where m and n denote the row and column location of this patch, a constrained search for similar patches in each image in $\mathbf{I}$ is performed in the search space $D(I_p^{m,n}, I_i) = \{I_i^{x,y} | x = m - l, \ldots, m + l\}$, where l is a small integer that defines the half width of the search space. In other words, the search space for the patch $I_p^{m,n}$ is a strip in I_i located between row $m - l$ and $m + l$. This search space tolerates saliency shift in the horizontal direction due to the change in camera views and misalignment in the vertical direction.

For each image I_i in $\mathbf{I}$, the most similar patch $I_i^{u,v}$ is found from the search space $D(I_p^{m,n}, I_i)$. The distance $d_i(I_p^{m,n}, I_i^{u,v})$

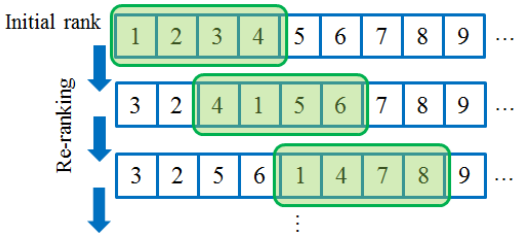


Fig. 4. Illustration of the reranking process. The initial returned ranked list is reranked based on the saliency similarity of the probe and gallery. In this example, a local sliding window of size $\alpha = 4$ with a step size of $\beta = 2$ is shown.

is calculated by the Euclidean distance between feature vectors from the two patches $I_p^{m,n}$ and $I_i^{u,v}$. The distances $\{d_i(I_p^{m,n}, I_i^{u,v}) | i = 1, 2, \ldots, N\}$ are then sorted and the saliency score for patch $I_p^{m,n}$ is defined as

$$\text{sal}\left(I_p^{m,n}\right) = 1 - e^{\left(-\frac{d_{i_k}\left(I_p^{m,n}, I_i^{u,v}\right)}{\sigma_1^2}\right)} \tag{10}$$

where $d_{i_k}(I_p^{m,n}, I_{i_k}^{u,v})$ is the Euclidean distance of the kth nearest neighbor (kNN) of $I_p^{m,n}$ from the search space of I_{i_k} and σ_1 controls the bandwidth of Gaussian function. k is set to $N/2$ in the experiments and only the kNN is involved in saliency computation. In this way, the saliency scores for each patch in the probe and gallery images are calculated. The saliency of a patch $I_p^{m,n}$ is computed from this kNN perspective such that the uniqueness of a patch is approximated by its distance to the samples in the reference set. The interpretation is that the more distinct a patch $I_p^{m,n}$, the larger its distance to the patches in the search space of images in $\mathbf{I}$, and thus, the saliency score $\text{sal}(I_p^{m,n})$ will be high. In this way, the saliency is calculated without supervision.

3) Reranking: Once the saliency of the probe and gallery is detected, the reranking of the initial reidentification results is based on the saliency similarity between the probe I_p from camera B and a returned gallery match I_t at rank t from camera A, which is defined as

$$\text{sim}_{\text{sal}}(I_p, I_t) = \sum_{m,n} \text{sal}\left(I_p^{m,n}\right) \times \text{sal}\left(I_t^{u,v}\right) \times e^{-\frac{d\left(I_p^{m,n}, I_t^{u,v}\right)}{\sigma_2^2}} \tag{11}$$

where $I_t^{u,v}$ is the NN of $I_p^{m,n}$ found in the search space and σ_2 is a Gaussian parameter. $d(I_p^{m,n}, I_t^{u,v})$ is the Euclidean distance between the features of $I_p^{m,n}$ and $I_t^{u,v}$.

Given a probe image, the reference-based method returns the matching results in descending order based on the similarity between the probe RD and gallery RDs. Based on the saliency similarity sim_{sal} between the probe image and the returned matching candidate, the initial ranked list is reranked using a local sliding windows of size α and a step size of β, and the candidate with a higher saliency similarity to the probe is moved forward in the local window. The reranking process is shown in Fig. 4.

C. Selection of Reference Set

The reference set can be optimized by selecting the most discriminative reference subjects and removing any redundant data. The goal is to select the basis reference subjects that will span the reference space. In other words, dissimilar subjects are preferred to form a more definitive reference set. Different reference selection rules are suggested in [51] and [52]. Three different methods are considered to select $\hat{N}$ reference subjects out of a total of N available ones.

1) *Random Selection:* We randomly sample $\hat{N}$ reference subjects out of the reference pool of size N.
2) *Max-Variation Rule:* In this rule, for each image I_i in the reference set $\{I_i, i = 1, 2, \ldots, N\}$, the similarity $s(f_i, f_j)$ between I_i and $I_{j,j\neq i}$ is computed for all j. The variation score v_i is $\text{Var}\{s(f_i, f_j)\}_{j=1,j\neq i}^{N}$. By ranking v_i values in a descending order, top $\hat{N}$ images are chosen.
3) *Min-Correlation Rule:* This rule is a backward selection process. Starting with the entire reference set $\{I_i, i = 1, 2, \ldots, N\}$, the sample I_i is removed whose average correlation with other samples $I_{j,j\neq i}$ is the highest. This process is repeated until $\hat{N}$ samples are left.

In Section IV-G, we evaluate the effectiveness of these reference set selection methods on the matching rate of the reidentification task.

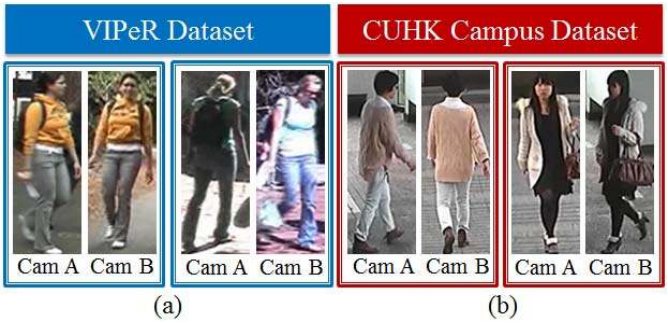


Fig. 5. Sample images from (a) VIPeR dataset [9] and (b) CUHK Campus dataset [59].

IV. Experimental Results

A. Datasets

1) VIPeR Dataset: The VIPeR dataset[1] is one of the most popular benchmark datasets for person reidentification [9]. It contains image pairs of 632 pedestrians. The images are taken by two nonoverlapping cameras with a significant view change. For most of the subjects, the view change is more than 90°. In addition, the illumination may also change dramatically. Other aspects such as cluttered background and occlusions further make this dataset more challenging. It is considered as the most challenging dataset currently available for pedestrian reidentification. For each person, a single image is available from each camera view. All of the images in the VIPeR dataset are normalized to 128×48. Some sample images are shown in Fig. 5(a).

2) CUHK Campus Dataset: The CUHK Campus dataset[2] contains images of 971 subjects from two nonoverlapping

[1]The VIPeR dataset is available at http://vision.soe.ucsc.edu/?q=node/178.
[2]The CUHK Campus dataset is available at http://www.ee.cuhk.edu.hk/~xgwang/CUHK_identification.html.

camera views [59]. One camera captures the frontal or back view of the subjects and the other camera captures profile views. Each person in each view has two images. The image quality of CUHK Campus dataset is higher compared with that of the VIPeR dataset. All of the images in the CUHK Campus dataset are resized to 128×48 in our experiments. Some sample images are shown in Fig. 5(b).

B. Feature Extraction and Parameters

Both color and texture features are extracted as in [12]. Specifically, the HSV (hue, saturation, value) and Lab color features are used to describe the color appearance of a subject. For the texture feature, we use local binary patterns (LBP) [60]. The image is divided into blocks of size 8×16. The blocks are overlapped by 50% in both the horizontal and vertical directions. Thus, the total number of blocks for one image of size 128×48 is $31 \times 5 = 155$. For each block, the quantized mean values of the HSV and Lab color channels are computed and the 8-bit LBP histogram is extracted. The final feature representation for one block is the concatenation of the means of the color channels and the LBP histogram with dimension $3 + 3 + 256 = 262$.

In the RCCA projection, the first 50 eigenvectors in the projection matrices W_A and W_B are used (i.e., the RCCA reduces the dimensions of the original features to 50). λ_1 and λ_2 are set to $10^{-1.6}$. For reranking, a local sliding window of size $\alpha = 4$ with a step size of $\beta = 2$ is used. The RCCA parameters as well as α and β are chosen based on cross-validation on the training data. In saliency detection, both Gaussian parameters [σ_1 in (10) and σ_2 in (11)] are set to 2.8, which is similar to the setting in [30]. Based on various experiments, we find that the matching performance is not sensitive to the choice of σ_1 and σ_2.

For the discovery of saliency, we use the same feature and parameter settings as in [30]. Specifically, the color histogram in Lab channels as well as 128-D SIFT features are extracted from 10×10 overlapping local patches with a step size of 5 pixels. Additional color histograms are extracted from two downsampled scales of each patch to more robustly retain the color information. Readers are referred to [30] for more details on the features used for saliency detection.

C. Evaluation Protocol

In our experiments, we follow the experimental protocols in [10], [14], and [30]. We randomly partition each dataset into two sets of equal size. Half of the data are used for training and constructing the reference set, and the other half of the data are used for testing. In the testing, the images from one camera are used as gallery and the images from the other camera are used as probes. The recognition rates at major top ranks and the cumulative matching characteristic (CMC) curves are reported. The CMC curve represents the expectation of finding the correct match in the top r matches. In other words, a rank-r recognition rate shows the percentage of the probes that are correctly recognized from the top r matches in the gallery. The experiments are performed 10 times and the average results are reported.

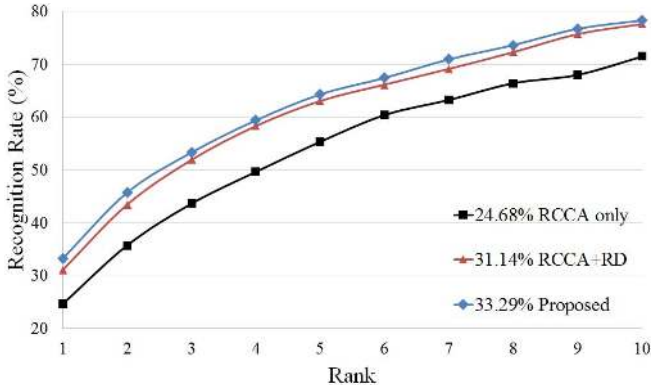


Fig. 6. CMC curves for the VIPeR dataset. The results by the proposed method, the method using RCCA only, and the method using RCCA and RD without reranking are shown.

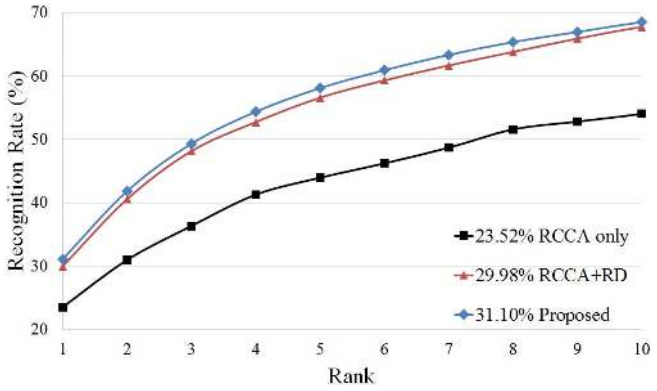


Fig. 7. CMC curves for the CUHK Campus dataset. The results by the proposed method, the method using RCCA only, and the method using RCCA and RD without reranking are shown.

D. Reidentification Performance

1) VIPeR Dataset: The recognition performance on the VIPeR dataset is shown in the CMC plots in Fig. 6 from ranks 1 to 10. The results from intermediate steps in the proposed method are also shown. When RCCA is used followed by direct matching only, the rank-1 recognition rate is 24.68%. When the matching is performed in the reference space, the rank-1 recognition rate rises to 31.14%, with an improvement of 26%. The reranking step further improves the rank-1 recognition rate to 33.29%. The gain by reranking is 7% compared with the results before reranking. At each rank in Fig. 6, the reference-based matching with reranking achieves the highest recognition rate.

2) CUHK Campus Dataset: Fig. 7 shows the recognition performance as CMC plots for the CUHK Campus dataset. Compared with the rank-1 recognition rates of 23.52% using RCCA only and 29.98% in the reference space after RCCA projection, the reference-based matching with saliency-based reranking as proposed achieves a rank-1 recognition rate of 31.10%. Figs. 6 and 7 show that each step in the proposed method contributes to the recognition performance.

E. Comparison With the Current Methods

1) VIPeR Dataset: The VIPeR dataset is the most popular benchmark dataset for person reidentification, and hence,

TABLE I
COMPARISON OF THE TOP-RANKED RECOGNITION RATES (IN PERCENTAGE) ON THE VIPeR DATASET. BEST RESULTS ARE IN BOLD

Rank→	$r = 1$	10	20	50	100
Proposed	33.29	78.35	88.48	**97.53**	**99.36**
SCNCD [31]	**33.70**	74.80	85.00	93.80	—
kLFDA [48]	32.33	**79.72**	**90.95**	—	—
SalMatch [30]	30.16	65.54	79.15	91.49	98.10
MidFilter [32]	29.11	65.95	79.87	92.47	98.04
MCE-KISS [39]	28.20	72.10	85.50	95.60	99.00
RPLM [12]	27.34	69.02	82.69	94.56	98.54
SSCDL [49]	25.60	68.10	83.60	95.20	—
RS-KISS [38]	24.50	66.60	81.70	93.50	98.00
CPS [18]	21.84	57.21	71.00	87.00	91.77
BiCov [24]	20.66	56.18	68.00	81.56	88.66
KISSME [14]	20.03	62.39	77.46	92.81	98.19
LMNN-R [42]	20.00	66.00	79.00	92.50	95.18
SDALF [10]	19.87	49.37	65.73	84.84	90.43
MRank [47]	19.34	55.51	70.44	87.69	96.90
PCCA [45]	19.27	64.91	80.28	95.00	97.01
DDC [22]	19.00	52.00	65.00	80.00	91.00
LMNN [15]	17.41	53.86	67.88	88.13	96.23
aPRDC [40]	16.14	50.98	65.95	88.00	93.00
PRDC [13]	15.66	53.86	70.09	87.79	92.84
ITML [17]	15.54	53.13	69.05	88.54	96.93
RankSVM [21]	14.00	51.00	67.00	85.00	94.00
ELF [20]	12.00	43.00	60.00	81.00	93.00

a lot of recent progress in reidentification reports results on this dataset. We compare our approach with the following methods: salient color-name-based color descriptors (SCNCD) with four color models [31], kernel local Fisher discriminant classifier (kLFDA) [48], saliency matching (SalMatch) [30], midlevel filters (MidFilter) [32], MCE-KISS [39], RPLM [12], semisupervised coupled dictionary learning [49], regularized smoothing KISSME learning [38], custom pictorial structures [18], biologically inspired features and covariance descriptors (BiCov) [24], KISSME [14], LMNN with rejection (LMNN-R) [42], symmetry-driven accumulation of local features (SDALF) [10], MRank [47], PCCA [45], descriptive and discriminative classification [22], LMNN [15], attribute-based probabilistic relative distance comparison (aPRDC) [40], probabilistic relative distance comparison (PRDC) [13], ITML [17], support vector ranking (RankSVM) [21], and ensemble of localized features [20]. For a fair comparison with the other methods, the results on the VIPeR datasets are either provided by the authors or cited from the corresponding papers directly.

The recognition results of the proposed method at ranks 1, 10, 20, 50, and 100 are compared with those of the other methods in Table I. Compared with SCNCD [31] that yields the best rank-1 result, our method achieves very competitive result at rank-1. In addition, our method consistently outperforms SCNCD [31] at higher ranks. Compared with SalMatch [30] with a rank-1 recognition rate of 30.16%, our method achieves a rank-1 recognition rate of 33.29%, which indicates a relative improvement of over 10%. Note that in [30] saliency is used for matching, while in our method, we use saliency for reranking only. Even without using saliency for reranking, our reference-based method (i.e., RCCA + RD only), with a rank-1 recognition rate of 31.14% (Fig. 6),

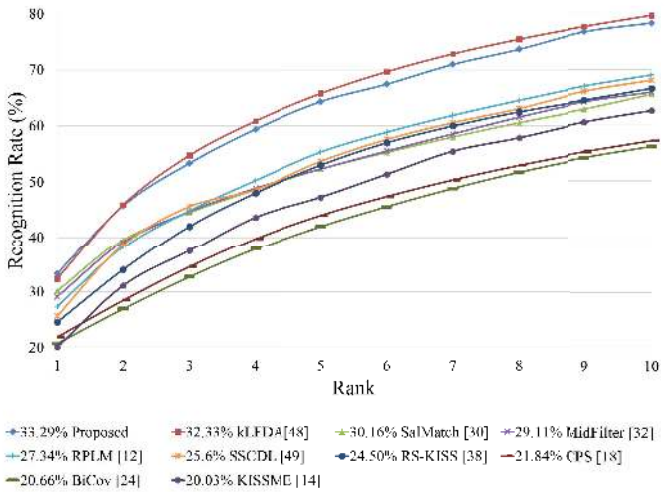


Fig. 8. Comparison of the CMC curves on the VIPeR dataset for the proposed method and the other methods.

TABLE II
COMPARISON OF THE RECOGNITION RATES (IN PERCENTAGE) WITH DIFFERENT TRAINING (REFERENCE) SET SIZES. BEST RESULTS ARE IN BOLD

Training size→	N=200			N=100		
Rank→	$r = 1$	10	20	$r = 1$	10	20
PRDC [13]	12.64	44.28	59.95	9.12	34.4	48.55
RPLM [12]	19.51	56.44	71.09	10.88	37.69	51.64
Proposed	**25.93**	**62.73**	**77.31**	**17.86**	**49.44**	**64.29**

still outperforms SalMatch [30] as well as most of the other methods in Table I. Fig. 8 shows the CMC curves of our method and some of the other top performers in Table I. The performance of our method is close to that of kLFDA [48], and both methods show a significant improvement over the other methods.

In Table II, we evaluate the performance of our method with reduced training/reference set size. All the data from the VIPeR dataset are used. As the size of the reference set decreases, the number of subjects in the gallery and probe data increases, which makes the reidentification more difficult. We compare our results with the reported results by RPLM [12] and PRDC [13]. From the comparison in Table II, it can be observed that with a smaller reference set, the proposed method performs significantly better, with rank-1 recognition rates of 25.93% and 17.86%, when reduced reference sets of sizes 200 and 100 are used, respectively.

2) CUHK Campus Dataset: For the CUHK Campus dataset, we compare the proposed approach with the following methods: MidFilter [32], SalMatch [30], SDALF [10], LMNN [15], ITML [17], as well as baseline methods using L1 norm and L2 norm as reported in [30]. The results of the comparing methods are provided by the authors or cited from the corresponding papers directly.

Table III reports the recognition rates at different ranks. The highest rank-1 accuracy is achieved by MidFilter [32], while at higher ranks, our method consistently performs better than all of the other methods including MidFilte [32]. Compared with

TABLE III
COMPARISON OF THE TOP-RANKED RECOGNITION RATES (IN PERCENTAGE) ON THE CUHK CAMPUS DATASET. BEST RESULTS ARE IN BOLD

Rank→	$r = 1$	10	20	50	100
Proposed	31.10	**68.55**	**79.18**	**90.38**	**95.86**
MidFilter [32]	**34.30**	64.96	74.94	86.89	94.26
SalMatch [30]	28.45	55.68	67.95	83.53	92.10
ITML [17]	15.98	45.60	59.81	76.61	88.32
LMNN [15]	13.45	42.25	54.11	73.29	86.65
SDALF [10]	9.90	30.33	41.03	55.99	67.39
L2-norm [30]	9.84	26.42	33.13	46.98	63.48
L1-norm [30]	10.33	26.34	33.52	45.62	61.95

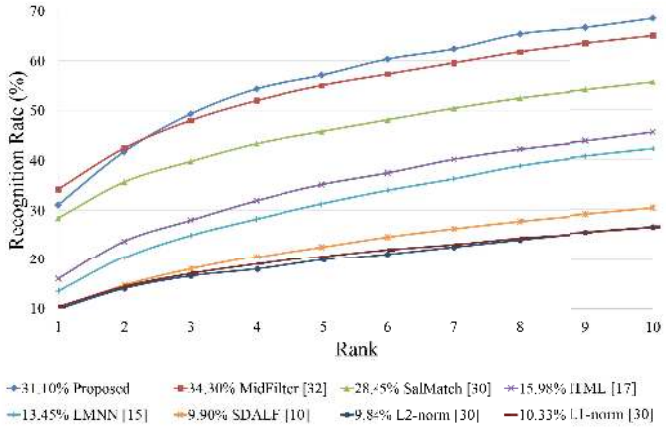


Fig. 9. Comparison of the CMC curves on the CUHK Campus dataset for the proposed method and the other methods.

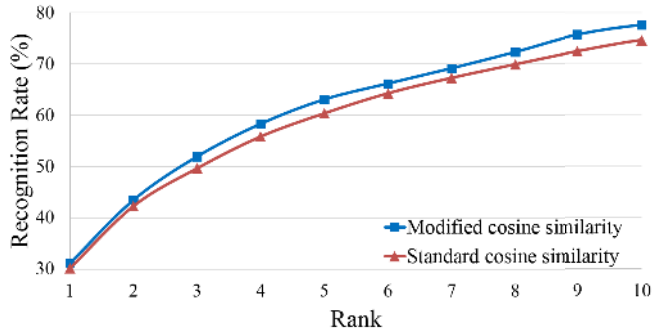


Fig. 10. Comparison between using the modified cosine similarity and standard cosine similarity on the VIPeR dataset.

SalMatch [30], our method has over a 9% relative improvement with a rank-1 recognition rate of 31.10%. In the case that a reranking step is not included, the proposed reference-based method (RCCA + RD) achieves a rank-1 recognition rate of 29.98% (Fig. 7), which is also higher than all of other methods except MidFilter [32]. Fig. 9 shows the CMC curves of different methods. Even the rank-1 result of our method is lower than that of MidFilter [32], at the other ranks, our method always achieves better accuracy and both our method and MidFilter [32] outperform the other methods by a large margin.

F. Effects of Modified Cosine Similarity

Figs. 10 and 11 compare the matching performance using the modified cosine similarity [58] (9) and the standard

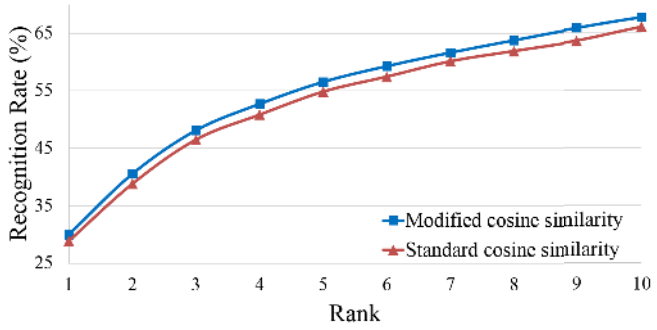


Fig. 11. Comparison between using the modified cosine similarity and standard cosine similarity on the CUHK Campus dataset.

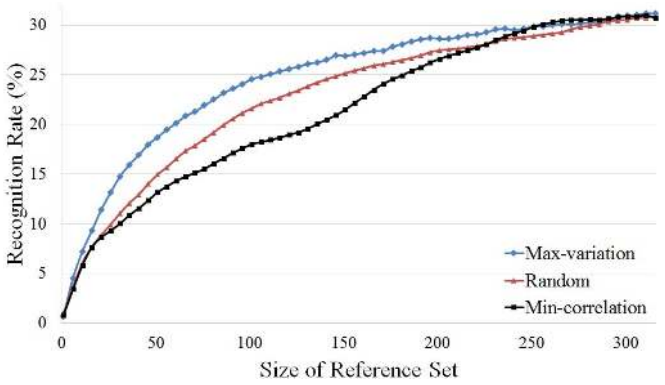


Fig. 12. Rank-1 reidentification accuracy using reference set with different sizes on the VIPeR dataset.

cosine similarity for the VIPeR and CUHK Campus datasets, respectively. Note that the reported results are the matching rates before reranking. As observed in Figs. 10 and 11, for both datasets, using the modified cosine similarity computed in (9) consistently achieves higher matching accuracy compared with the results using the standard cosine similarity measure. At higher ranks, more performance gain is observed. This suggests that for a better matching accuracy, the modified cosine similarity can be utilized.

G. Effects of Reference Set Selection

1) VIPeR Dataset: Fig. 12 shows the rank-1 recognition accuracy on the VIPeR dataset with varying reference set size using the reference selection strategies described in Section III-C. For the random selection, as the size of the reference set increases, the recognition rate keeps improving. For the maximum variation rule, when the number of selected reference subjects is small, the recognition performance is higher than the results by random selection. As the reference set size reaches above 250, both rules result in a similar performance, with only marginal improvement by adding more reference samples. Compared with random selection and maximum variation selection, the minimum correlation rule does not select better reference samples when the size of the reference set is not sufficiently large. Note that the size of the reference set can be reduced without a significant loss in performance. Using the maximum variation rule, the size of

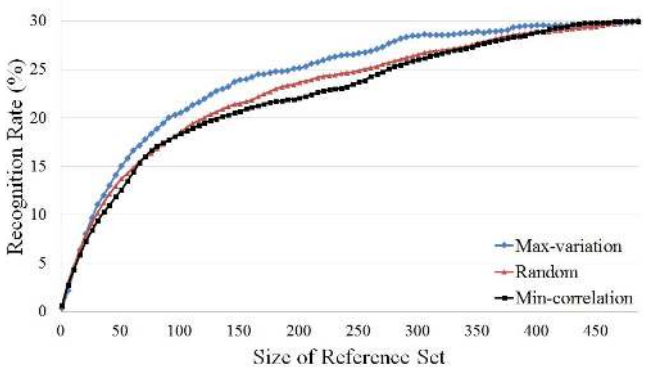


Fig. 13. Rank-1 reidentification accuracy using reference set with different sizes on the CUHK Campus dataset.

the reference set is reduced by over 40% from 316 to 180 with a performance drop of less than 10%. While keeping sufficient accuracy with less than 4% degradation, the size of the reference set can be reduced by over 20% from 316 to 250. With constraint such as computational efficiency on reference set, the size of the reference set may be chosen where the improvement in recognition rate starts changing slowly.

2) CUHK Campus Dataset: Fig. 13 shows the results with varying reference set size on the CUHK Campus dataset. A trend similar to Fig. 12 is observed. The maximum variation rule is able to select a better subset of the reference samples. As suggested by the experimental results, maximum variation is an effective strategy for reference set selection. For the CUHK Campus dataset, using maximum variation rule for selection, the size of the reference set can be reduced by over 40% from 486 to 286 with a performance drop of $\sim$5%.

H. Computational Cost

The computational cost mainly consists of the following parts: 1) feature extraction; 2) RCCA subspace learning; 3) RD generation; 4) initial matching; and 5) reranking. The experiments are performed using MATLAB implementation without optimization on a laptop with Intel i7 2.4-GHz CPU and 8-GB RAM. For each image, the feature extraction takes about 0.37 s. On the VIPeR dataset, learning RCCA projection matrices takes about 4.2 s. For the CUHK Campus dataset, this procedure takes slightly longer of about 4.4 s, due to more data involved. However, the projection learning is done during the offline process and need to be performed only once. The generation of a RD is very efficient and it takes less than 2×10^{-6} s. Initial matching on the VIPeR dataset for one query takes about 0.8×10^{-4} s, and this goes up to 1.1×10^{-4} s for the CUHK Campus dataset. Saliency-based reranking for one probe takes about 0.81 s on the VIPeR dataset and about 0.96 s on the CUHK Campus dataset. The efficiency of reranking can be improved using fast patch match technique [61].

I. Limitation

Our current framework operates in a two-view scenario where images from one camera are probes to be matched to the gallery data captured from the other camera, assuming that the identities of the probes are enrolled in the gallery. In addition, RCCA as used in our approach only handles data from two views (i.e., images from two cameras). Using the reference set idea, Chen *et al.* [54] have developed an approach for tracking people across a network of nonoverlapping cameras. This approach can be considered reidentification in a video network subject to motion, appearance, and time constraints. In practice, it would be desired to perform unconstrained reidentification across the entire camera network in an optimal manner. For this purpose, a global optimization approach can be used to uniquely label individuals in a network of cameras at a given point in time. A variety of optimization approaches such as stochastic relaxation [62], [63], dynamic programming [64], branch and bound [65], and integer programming [66] can be used for this purpose. However, these techniques are not practical at video rates for a large number of people, but they may provide initialization of the individuals in different cameras that may result in obtaining pairwise relationships among different cameras. After this, the proposed approach described in this paper can be effectively used. In addition, there could be various generalizations of our approach to a variety of practical applications, a topic left for future research.

V. Conclusion

In this paper, the use of a reference set for person reidentification is proposed. Compared with the previous methods in which either invariant features are extracted or a distance metric is explored, in this paper, a reference set is utilized to transfer the matching problem from an appearance space to a reference space. The reidentification is achieved by matching the RDs generated with the reference set and the matching results are improved by a reranking step using image saliency information. The experiments on different datasets showed that the proposed method using RCCA in conjunction with the reference set outperformed 17 current approaches on the VIPeR dataset and six recently published techniques on the CUHK Campus dataset.

The proposed method avoids a direct comparison between the gallery and the probe using appearance features. Reference-based matching with reranking significantly improved upon RCCA-based matching as a baseline method ($\sim$35% improvement on the VIPeR dataset and $\sim$32% improvement on the CUHK Campus dataset). The proposed method can be combined with any advanced feature representation to further improve the reidentification accuracy, and the dimension of RDs is determined only by the size of the reference set, which can be optimized based on the analysis presented in this paper.

The color and texture features that we have used for both matching and saliency detection are not explicitly designed to be illumination invariant. Thus, seeking more robust features for both matching and saliency detection is one of our ongoing research topics. In addition, we plan to extend the proposed reference-based framework to multishot person reidentification. Currently, our method focuses on reidentifying people on pairwise cameras, and generalizing our framework to multicamera scenarios will be the focus of our future research for real-world applications.

REFERENCES

[1] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey," *ACM Comput. Surv.*, vol. 46, no. 2, Nov. 2013, Art. ID 29.

[2] S. Gong, M. Cristani, S. Yan, and C. C. Loy, Eds., *Person Re-Identfication* (Advances in Computer Vision and Pattern Recognition). London, U.K.: Springer-Verlag, 2014.

[3] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: Problem overview and current approaches," *J. Ambient Intell. Humanized Comput.*, vol. 2, no. 2, pp. 127–151, 2011.

[4] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image Vis. Comput.*, vol. 32, no. 4, pp. 270–286, 2014.

[5] L. Zhang, D. V. Kalashnikov, S. Mehrotra, and R. Vaisenberg, "Context-based person identification framework for smart video surveillance," *Mach. Vis. Appl.*, vol. 25, no. 7, pp. 1711–1725, 2014.

[6] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, and S. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 663–671, Apr. 2006.

[7] J. Tang, R. Hong, S. Yan, T.-S. Chua, G.-J. Qi, and R. Jain, "Image annotation by *k*NN-sparse graph-based label propagation over noisily tagged Web images," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, Feb. 2011, Art. ID 14.

[8] J. Tang, Z.-J. Zha, D. Tao, and T.-S. Chua, "Semantic-gap-oriented active learning for multilabel image annotation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2354–2360, Apr. 2012.

[9] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, Sep. 2007, pp. 1-7.

[10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2360–2367.

[11] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2007, pp. 1–8.

[12] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 780–793.

[13] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 653–668, Mar. 2013.

[14] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2288–2295.

[15] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Dec. 2009.

[16] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Sep./Oct. 2009, pp. 498–505.

[17] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 209–216.

[18] D. S. Cheng, M. Cristan, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2011, pp. 68.1–68.11.

[19] S. Bak, G. Charpiat, E. Corvee, F. Bremond, and M. Thonnat, "Learning to match appearances by correlations in a covariance metric space," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2012, pp. 806–820.

[20] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2008, pp. 262–275.

[21] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2010, pp. 21.1–21.11.

[22] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. 17th Scandin. Conf. Image Anal. (SCIA)*, 2011, pp. 91–102.

[23] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1622–1634, Jul. 2013.

[24] B. Ma, Y. Su, and F. Jurie, "BiCov: A novel image representation for person re-identification and face verification," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2012, pp. 57.1–57.11.

[25] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2012, pp. 31–36.

[26] C.-H. Kuo, S. Khamis, and V. Shet, "Person re-identification using semantic color names and RankBoost," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 281–287.

[27] R. Layne, T. Hospedales, and S. Gong, "Person re-identification by attributes," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2012, pp. 24.1–24.11.

[28] L. An, X. Chen, M. Kafai, S. Yang, and B. Bhanu, "Improving person re-identification by soft biometrics based reranking," in *Proc. ACM/IEEE 7th Int. Conf. Distrib. Smart Cameras (ICDSC)*, Oct./Nov. 2013, pp. 1–6.

[29] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3586–3593.

[30] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2528–2535.

[31] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. L. Li, "Salient color names for person re-identification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 536–551.

[32] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 144–151.

[33] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 152–159.

[34] Z. Zhang and V. Saligrama. (2014). "Person re-identification via structured prediction." [Online]. Available: http://arxiv.org/abs/1406.4444

[35] C. Liu, C. C. Loy, S. Gong, and G. Wang, "POP: Person re-identification post-rank optimisation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 441–448.

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

[37] M. Vernier, N. Martinel, C. Micheloni, and G. L. Foresti, "Remote feature learning for mobile re-identification," in *Proc. ACM/IEEE 7th Int. Conf. Distrib. Smart Cameras (ICDSC)*, Oct./Nov. 2013, pp. 1–6.

[38] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.

[39] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 242–252, Feb. 2015.

[40] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important?" in *Proc. Eur. Conf. Comput. Vis. Workshops Demonstrations*, 2012, pp. 391–401.

[41] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3594–3601.

[42] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2011, pp. 501–512.

[43] N. Martinel, C. Micheloni, and C. Piciarelli, "Learning pairwise feature dissimilarities for person re-identification," in *Proc. ACM/IEEE 7th Int. Conf. Distrib. Smart Cameras (ICDSC)*, Oct./Nov. 2013, pp. 1–6.

[44] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3318–3325.

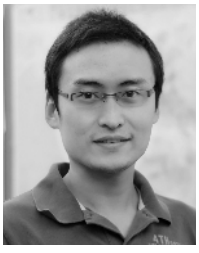
[45] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2666–2672.

[46] L. An, S. Yang, and B. Bhanu, "Person re-identification by robust canonical correlation analysis," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1103–1107, Aug. 2015.

[47] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2013, pp. 3567–3571.

[48] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 1–16.

[49] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu, "Semi-supervised coupled dictionary learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3550–3557.
[50] S. Liao, Z. Mo, J. Zhu, Y. Hu, and S. Z. Li. (2014). "Open-set person re-identification." [Online]. Available: http://arxiv.org/abs/1408.0872
[51] A. Gyaourova and A. Ross, "Index codes for multibiometric pattern retrieval," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 518–529, Apr. 2012.
[52] R. P. W. Duin and E. Pkalska, "The dissimilarity space: Bridging structural and statistical pattern recognition," *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 826–832, May 2012.
[53] Y. Guo, Y. Shan, H. Sawhney, and R. Kumar, "PEET: Prototype embedding and embedding transition for matching vehicles over disparate viewpoints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.
[54] X. Chen, L. An, and B. Bhanu, "Reference set based appearance model for tracking across non-overlapping cameras," in *Proc. ACM/IEEE 7th Int. Conf. Distrib. Smart Cameras (ICDSC)*, Oct./Nov. 2013, pp. 1–6.
[55] L. An, M. Kafai, S. Yang, and B. Bhanu, "Reference-based person re-identification," in *Proc. 10th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2013, pp. 244–249.
[56] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
[57] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman, "Canonical correlation analysis when the data are curves," *J. Roy. Statist. Soc. Ser. B (Methodological)*, vol. 55, no. 3, pp. 725–740, 1993.
[58] C. Liu, "Discriminant analysis and similarity measure," *Pattern Recognit.*, vol. 47, no. 1, pp. 359–367, 2014.
[59] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, 2013, pp. 31–44.
[60] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[61] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized PatchMatch correspondence algorithm," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 29–43.
[62] B. Bhanu and O. D. Faugeras, "Shape matching of two-dimensional objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 2, pp. 137–156, Mar. 1984.
[63] B. Bhanu, "Representation and shape matching of 3-D objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 3, pp. 340–351, May 1984.
[64] S. Calderara, R. Cucchiara, and A. Prati, "A dynamic programming technique for classifying trajectories," in *Proc. 14th Int. Conf. Image Anal. Process. (ICIAP)*, Sep. 2007, pp. 137–142.
[65] C. Picus, R. Pflugfelder, and B. Micusik, "Branch and bound global optima search for tracking a single object in a network of non-overlapping cameras," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1825–1830.
[66] A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent re-identification in a camera network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 330–345.



Le An (M'15) received the B.Eng. degree in telecommunications engineering from Zhejiang University, Hangzhou, China, in 2006; the M.Sc. degree in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 2008; and the Ph.D. degree in electrical engineering from University of California at Riverside, Riverside, CA, USA, in 2014.

He is a Post-Doctoral Research Associate with University of North Carolina, Chapel Hill, NC, USA. His research interests include image processing, computer vision, pattern recognition, and machine learning.

Dr. An was a recipient of the best paper award from the 2013 IEEE International Conference on Advanced Video and Signal-Based Surveillance.

Mehran Kafai (M'13) received the M.Sc. degree in computer engineering from Sharif University of Technology, Tehran, Iran, in 2005; the M.Sc. degree in computer science from San Francisco State University, San Francisco, CA, USA, in 2009; and the Ph.D. degree in computer science from Center for Research in Intelligent Systems, University of California at Riverside, Riverside, CA, USA, in 2013.

He is a Research Scientist with Hewlett Packard Laboratories, Palo Alto, CA, USA. His research interests include secure information retrieval and large-scale in-memory analytics.

Songfan Yang (M'14) received the B.S. degree in electrical engineering from Sichuan University, Chengdu, China, in 2009, and the M.S. and Ph.D. degrees in electrical engineering from University of California at Riverside, Riverside, CA, USA.

He is an Associate Professor with the College of Electronics and Information Engineering, Sichuan University. His research interests include computer vision, pattern recognition, and affective computing.

Dr. Yang was a recipient of the Best Entry Award of the FG 2011 Facial Expression Recognition and Analysis emotion challenge competition.

Bir Bhanu (F'95) received the S.M. and E.E. degrees in electrical engineering and computer science from Massachusetts Institute of Technology, Cambridge, MA, USA; the Ph.D. degree in electrical engineering from the Image Processing Institute, University of Southern California, Los Angeles, CA, USA; and the M.B.A. degree from University of California at Irvine, Irvine, CA, USA.

He is currently a Distinguished Professor of Electrical and Computer Engineering and a Co-operative Professor of Computer Science and Engineering and Mechanical Engineering and the Interim Chair of the Department of Bioengineering with the University of California at Riverside, Riverside, CA, USA. He is also the Director of the Center for Research in Intelligent Systems, the Visualization and Intelligent Systems Laboratory, and the NSF IGERT on Video Bioinformatics. He has been the Principal Investigator of various programs for National Science Foundation, Defense Advanced Research Projects Agency, National Aeronautics and Space Administration, Air Force Office of Scientific Research, Office of Naval Research, Army Research Office, and other agencies and industries in the areas of video networks, video understanding, video bioinformatics, learning and vision, image understanding, pattern recognition, target recognition, biometrics, autonomous navigation, image databases, and machine-vision applications. He has published seven authored and three edited books. He holds 18 (five pending) patents. He has authored over 475 reviewed technical publications, including over 135 journal papers and 45 book chapters. His current research interests include computer vision, pattern recognition and data mining, machine learning, artificial intelligence, image processing, image and video database, graphics and visualization, robotics, human-computer interactions, biological, medical, military, and intelligence applications.

Dr. Bhanu is a fellow of the American Association for the Advancement of Science, the American Institute for Medical and Biological Engineering, the International Association for Pattern Recognition, and the International Society for Optics and Photonics.